metals

MDPI

Article

Accurate Detection of Cd^{2+} and Pb^{2+} Concentrations in Soils by Stripping Voltammetry Peak Areas under the Mutual Interference of Multiple Heavy Metals

Wenshuai Ye [1,†], Ning Liu [2,†], Guo Zhao [3] and Gang Liu [1,2,*]

[1] Key Laboratory of Smart Agriculture Systems, Ministry of Education, China Agricultural University, Beijing 100083, China
[2] Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture and Rural Affairs of China, China Agricultural University, Beijing 100083, China
[3] College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China
* Correspondence: pac@cau.edu.cn
† These authors contributed equally to this work.

Abstract: The accurate detection of Cd^{2+} and Pb^{2+} in soils by square-wave anodic stripping voltammetry (SWASV) faces great challenges because the interaction between multiple heavy metal ions (HMIs) interferes seriously with their SWASV signals. To detect Cd^{2+} and Pb^{2+} by SWASV with high accuracy, an overlooked but informative signal, i.e., stripping current peak area, was employed and combined with chemometric methods to suppress the above mutual interference. An easy-to-prepare electrode, i.e., in-site electroplating bismuth film modified glassy carbon electrode, was used to sense the multiple HMIs. Two machine learning algorithms, including SVR and PLSR, were used to establish the detection models of Cd^{2+} and Pb^{2+}. In addition, this study developed a homemade algorithm to automatically acquire the stripping peak heights and stripping peak areas of Zn^{2+}, Cd^{2+}, Pb^{2+}, Bi^{3+}, and Cu^{2+}, which acted as the inputs of machine learning models. Then, the detection performance of various SVR and PLSR models were compared based on the R^2 and *RMSE* values of the validation dataset. Results showed that the SVR detection models established by the algorithmically acquired peak areas presented the best stability and accuracy for detecting both Cd^{2+} and Pb^{2+} concentrations under the existence of Zn^{2+} and Cu^{2+}. The R^2 and *RMSE* values of the SVR models built using the peak heights of HMIs acquired by electrochemical workstation control software (I_{manu}-SVR) were 0.7650 and 5.3916 μg/L for Cd^{2+}, and 0.8791 and 20.0015 μg/L for Pb^{2+}, respectively; the R^2 and *RMSE* values of the SVR models built using the peak area automatically acquired by the developed algorithm (A_{algo}-SVR) were 0.9204 and 2.9906 μg/L for Cd^{2+}, and 0.9756 and 13.1574 μg/L for Pb^{2+}, respectively. More importantly, the detection results of the proposed method in real soil extracts for Cd^{2+} and Pb^{2+} concentrations were close to those of ICP-MS, verifying its practicability. This study provides a new solution for the accurate detection of targeted heavy metals under the co-existence of multiple HMIs by the SWASV method.

Keywords: SWASV; chemometrics; machine learning; electrochemistry; soil extracts; interference factors

check for updates

Citation: Ye, W.; Liu, N.; Zhao, G.; Liu, G. Accurate Detection of Cd^{2+} and Pb^{2+} Concentrations in Soils by Stripping Voltammetry Peak Areas under the Mutual Interference of Multiple Heavy Metals. *Metals* **2023**, *13*, 270. https://doi.org/10.3390/met13020270

Academic Editor: Antonije Onjia

Received: 14 December 2022
Revised: 20 January 2023
Accepted: 25 January 2023
Published: 29 January 2023

1. Introduction

Pb^{2+} and Cd^{2+} in soils can accumulate in the human body through the food chain, eventually posing a serious threat to human health [1–3], which has become the focus of attention. Therefore, the accurate detection of the concentrations of Pb^{2+} and Cd^{2+} in soil is of great necessity for evaluating the levels of heavy metal pollution. Compared with conventional spectroscopic techniques for the detection of Cd^{2+} and Pb^{2+}, the square-wave anodic stripping voltammetry (SWASV) technique as an electrochemical analysis technique has been considered a promising method due to its advantages of high sensitivity, easy operation, rapid response, and cost-effectiveness, as well as the simultaneous detection

of multiple heavy metals [4–8]. However, the accurate detection of Cd^{2+} and Pb^{2+} in soil using SWASV still has problems to be solved such as the interactive interference between multiple heavy metal ions (HMIs).

Soil contains multiple heavy metals, such as Zn^{2+}, Cu^{2+}, Pb^{2+}, Cd^{2+}, Ag^{+}, and As^{3+}/As^{5+}. Among them, the content of Cu^{2+} and Zn^{2+} are highly enriched in soil compared to Pb^{2+} and Cd^{2+} [9,10]. Serious interactions between multiple ions (such as forming complex and various alloys) occur in electrochemical analysis, which will severely interfere with the detection accuracy of SWASV to Cd^{2+} and Pb^{2+} in soils. It was reported [5,11,12] that the peak currents of Cd^{2+} firstly decreased and, then, leveled off with the increase in the externally added Pb^{2+}, however, it decreased sharply or even disappeared in spite of encountering the same concentration of Cu^{2+} [13]. The above phenomenon may be because of two aspects. On one hand, Pb^{2+} and Cu^{2+} compete for the active sites on the surface of the working electrode with Cd^{2+}, which decreases the electro-deposited amount of Cd^{2+}. On the other hand, multiple ions can form the Cd–Pb, Cd–Cu, and Cd–Pb–Cu alloys on the surface of the working electrode during the electrochemical deposition process, which inhibits the stripping of Cd^{2+} to obtain the stripping peak currents [14,15]. In addition, the existence of low concentration Zn^{2+} is beneficial to the determination of Cd^{2+} by forming the favorable Cd–Zn intermetallic film [11,16], but high-concentration Zn^{2+} suppresses the peak current of Cd^{2+} [17]. In our previous work, two-dimensional correlation spectroscopy was used to explore the interactive interference characteristics of multiple heavy metal ions (HMIs), which showed that Cu^{2+} was the most serious interference ion on the peak currents of Cd^{2+} and Pb^{2+}, followed by Zn^{2+}, while Cd^{2+} and Pb^{2+} would interfere with each other [18]. In summary, the interactive interference of various HMIs is a momentous problem in the detection of Cd^{2+} and Pb^{2+} concentrations in soil by SWASV.

To overcome the interactive interference between HMIs, some studies [19–23] prepared complex biochemical materials to modify the working electrode. However, the electrode modified using specific recognition materials required expensive costs, complicated steps, and harsh storage conditions. In comparison, other studies [5,24,25] established machine learning models, which took the manually acquired stripping peak currents of each heavy metal as the input and the Cd^{2+} and Pb^{2+} concentrations as the output. Model inputs, after all, needed to be acquired manually, which was time-consuming, had large human error, and could not satisfy automated detection. Therefore, characteristic stripping currents (i.e., multiple stripping currents selected by machine learning algorithms) were proposed to act as input variables, which offered more comprehensive information about the interactions among Zn^{2+}, Cd^{2+}, Pb^{2+}, and Cu^{2+} [18]. However, the stripping potential of SWASV would drift due to the difference in detection environments or the reference electrodes, leading to the disability of the characteristic stripping currents [5]. The peak area of the stripping currents is defined as the integral value of the stripping current on the scale of stripping potential time, which physically represents the amount of charges. Theoretically, the peak area of the stripping current directly reflects the number of stripping HMIs under a fixed electrode surface area. Therefore, building a machine learning model using the peak area instead of peak current as the input may be a good approach to improve the SWASV detection accuracy of Cd^{2+} and Pb^{2+} under interactive interference between multiple HMIs.

The accurate acquisition of SWASV signals, including the stripping peak currents and the stripping current peak areas of HMIs, is crucial for the detection of heavy metal concentrations. The peak current signals of Cd^{2+} and Pb^{2+} were acquired, under the premise that the background currents of Cd^{2+} and Pb^{2+} are the same, by development algorithms [26]. These algorithms cannot automatically acquire the SWASV signals of multiple HMIs and are helpless for overlapping peak signals. Therefore, developing an algorithm that can accurately acquire the stripping current peak areas of Zn^{2+}, Cd^{2+}, Pb^{2+}, Bi^{3+}, and Cu^{2+} is of great significance for the automatic and accurate detection of Cd^{2+} and Pb^{2+} concentrations by SWASV. In addition, the stripping current peak heights of HMIs are also acquired to highlight the advantages of stripping current peak areas in modeling.

In this paper, a combination of chemometrics, machine learning, and homemade algorithms was used to improve the detection accuracy of Cd^{2+} and Pb^{2+} concentrations under the interaction interference of Zn^{2+} and Cu^{2+}. On the one hand, this study attempted to build machine learning models for accurately detecting the concentrations of Cd^{2+} and Pb^{2+} by using a simple bismuth-film modified electrode, which not only saves financial, material, and human resources [8,19,20,23], but also avoids secondary pollution caused by modified materials to the environment. On the other hand, this study investigated the difference between the HMIs peak area and the traditionally used peak height as the model inputs for the accurate detection of Cd^{2+} and Pb^{2+}. Importantly, the practicality of the selected optimal detection models was tested using real soil samples.

2. Materials and Methods

2.1. Reagents and Apparatus

All chemical reagents were of an analytical grade. The working solutions of Zn^{2+}, Cd^{2+}, Pb^{2+}, Bi^{3+}, and Cu^{2+} were prepared from 1 mg/mL of standard stock solutions of $Zn(NO_3)_2$, $Cd(NO_3)_2$, $Pb(NO_3)_2$, $Bi(NO_3)_3$, and $Cu(NO_3)_2$, respectively. Acetic acid acted as the electrolyte buffer for the electrochemical measurement of HMIs. Millipore–Q water (18.2 MΩ·cm) was used for diluting reagents and the cleaning of containers for all experiments.

A three-electrode system including an Ag/AgCl reference electrode, a platinum wire counter electrode, and a Bi/GCE working electrode was used to collect the stripping currents of Zn^{2+}, Cd^{2+}, Pb^{2+}, Bi^{3+}, and Cu^{2+}. SWASV measurements were performed using the EmStat3 electrochemical workstation from PalmSens (PSTrace 5.9, EmStat3, Houten, The Netherlands).

2.2. Preparation of the Bi/GCE

The $Bi(NO_3)_3$ was used to prepare the in-site electroplating bismuth film-modified glassy carbon electrode (Bi/GCE) as a working electrode because bismuth possessed excellent electrocatalysis ability towards Cd^{2+} and Pb^{2+} [27–29]. The specific preparation of the Bi/GCE electrode is presented in the Supplementary Materials.

2.3. SWASV Measurement

The optimal experimental conditions for the detection of Cd^{2+} and Pb^{2+} concentrations using Bi/GCE by SWASV had been explored in our previous study [24]; the optimal concentration of Bi^{3+} was 300 μg/L, the optimal pH value was 5.0, the optimal deposition potential was −1.3 V, and the optimal deposition time was 180 s. In addition, the repeatability, stability and electroanalytical performance of Bi/GCE had been validated, as presented in the Supplementary Materials, which demonstrated that the obtained SWASV dataset was reliable to build machine learning detection models.

Before performing SWASV measurements, dissolved oxygen was removed from the test solution by blowing N_2 gas for 2 min. The main steps of SWASV were as follows. The HMIs were electro-deposited into the GCE surface for 180 s at the potential of −1.3 V, accompanied by stirring at a speed of 300 r/min. After an equilibration period for 10 s, the deposited HMIs were stripped off from the GCE surface by applying a square-wave excitation potential in the range of −1.4 to 0.2 V to obtain stripping currents. The square-wave frequency, potential amplitude, and potential increment of the excitation potential were 25 Hz, 25 mV, and 5 mV, respectively. A total of 320 data points were collected in the stripping potential range of −1.4 to 0.2 V for one SWASV data curve due to the potential increment of 5 mV.

The detailed steps of the SWASV measurements are described in the Supplementary Materials.

2.4. Establishment of Experimental Datasets

To explore the interactive influence of multiple HMIs on the SWASV signals of Cd^{2+} and Pb^{2+}, eight concentration gradients of four kinds of HMIs were set. The Zn^{2+} and Cu^{2+} acted as the interference ions, their concentration gradients started from 0 μg/L. Specifically, the Zn^{2+} concentrations included 0, 50, 100, 150, 200, 250, 300, and 350 μg/L, while the Cu^{2+} concentrations included 0, 25, 50, 75, 100, 125, 150, and 175 μg/L, the Pb^{2+} concentrations included 5, 25, 50, 75, 100, 125, 150, and 175 μg/L, and the Cd^{2+} concentrations included 2, 5, 10, 15, 20, 25, 30, and 35 μg/L. The orthogonal experimental method was adopted to conduct the experiment's design on the SPSS software (SPSS 26, IBM SPSS, Chicago, IL, USA). A total of 64 experiments were designed according to the L64(8^4) orthogonal table under four factors and eight levels. The 64 sets of orthogonal experiments were listed in Table S1. Namely, 64 sets of SWASV data were collected to develop the algorithm for the automatic acquisition of peak heights and peak areas of multiple HMIs and to establish the models for the accurate detection of Cd^{2+} and Pb^{2+} concentrations under the existence of multiple HMIs. The SWASV measurement of each sample was repeated three times to guarantee the reliability of the dataset.

2.5. Development of Peak Height and Peak Area Acquisition Algorithms

First, each SWASV curve was pre-processed and divided into several segments according to the extremum points. Then, the homemade algorithm was used to obtain the peak high and peak area of each segment. Finally, the type of each HMI was identified according to the position of the stripping potential corresponding to the peak current.

2.5.1. Segmentation of SWASV Curves

The Savitzky-Golay (S-G) algorithm was used to smooth the SWASV curve as shown in Figure 1A. The extreme points in the SWASV curve after S-G smoothing were accurately identified by the peakdet function. The strategy of the peakdet function was to look for the highest point, around which there were lower points on both sides. It was defined as the extreme point in the SWASV curve by finding the closest point in the data points to the voltage value of the extreme point in the SWASV curve after S-G smoothing, as shown in Figure 1B. It is worth noting that the number of local minimum points should be one more than the number of local maximum points to ensure the existence of local minimum points on the left and right of each local maximum. However, one point located at the end or front of the SWASV curve could not be detected as the local minimum point using the peakdet function. Thus, it was set as the minimum point to ensure that every segment of the SWASV curve contained one local maximum and two local minimum points, as shown in Figure 1C. Assuming that the coordinates of 320 data points were $(V_1, I_1), (V_2, I_2) \ldots (V_{320}, I_{320})$, among them, the coordinates of the local maximum points were (V_{maxy}, I_{maxy}), and the coordinates of the local minimum points were (V_{minz}, I_{minz}). The y and z were the number of local maximum and minimum points, respectively. Then, each SWASV curve was divided into several segments according to the minimum points, as shown in Figure 1D.

2.5.2. Calculation of Valid Data Segments

The valid data segment needed to be extracted from the segmented curve due to the presence of background currents in the SWASV curve. The core principle was to calculate the secant line of the segmented curve with only two intersections, which were located to the left and right of the local maximum point. As an example, the valid data segment i of the SWASV curve was calculated based on Section 2.5.1, as follows.

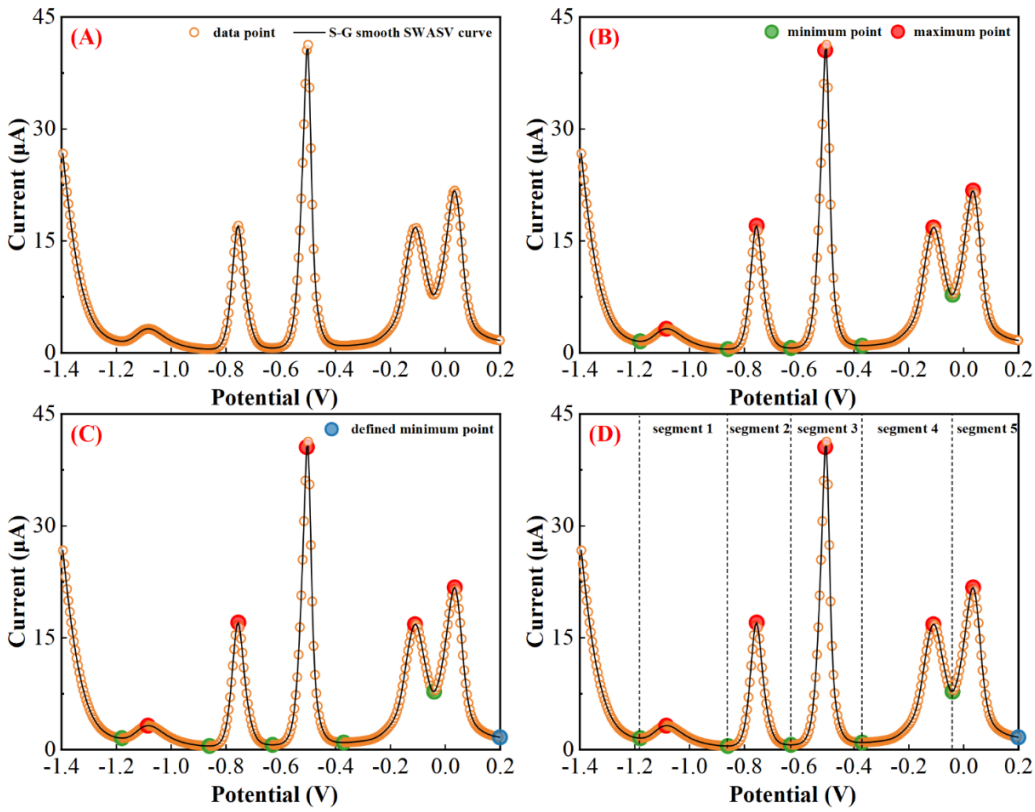

Figure 1. Schematic representation of the SWASV curve smoothed by S-G algorithm (**A**), the extreme points identified by the peakdet function (**B**), the end of the SWASV curve set as the minimum point (**C**), and the segmentation of SWASV curve (**D**). (SWASV response curve of 150 μg/L Zn^{2+}, 35 μg/L Cd^{2+}, 175 μg/L Pb^{2+}, 300 μg/L Bi^{3+}, and 100 μg/L Cu^{2+}).

Firstly, the slope matrix K_{L1}, i.e., Equation (1), was calculated by the slope between every coordinate point from the local maximum point (V_{maxi}, I_{maxi}) to the local minimum point $\left(V_{min(i+1)}, I_{min(i+1)}\right)$ and the local minimum point (V_{mini}, I_{mini}) as shown in Figure 2A.

$$K_{L1} = \begin{bmatrix} k_{10} \\ k_{11} \\ \vdots \\ k_{1u} \\ \vdots \\ k_{1p} \end{bmatrix} \tag{1}$$

where $p = \frac{V_{min(i+1)} - V_{maxi}}{0.005}$ is the number of elements in K_{L1}, i.e., the data points number from the local maximum point to the right local minimum point, $k_{1u} = \frac{I_{maxi+u} - I_{mini}}{V_{maxi+u} - V_{mini}}$ $(u \in [0, p])$ is the u-th element in K_{L1}, and $k_{1m} = min\{K_{L1}\}$ is the global minimum value in K_{L1}.

The coordinate $P_{i_1}(V_{maxi+1m}, I_{maxi+1m})$ was acquired by k_{1m}.

Secondly, the slope matrix K_{R1}, i.e., Equation (2), was calculated by the slope between every coordinate point from the local maximum point (V_{maxi}, I_{maxi}) to the local minimum point (V_{mini}, I_{mini}) and $P_{i_1}(V_{maxi+1m}, I_{maxi+1m})$ as shown in Figure 2B.

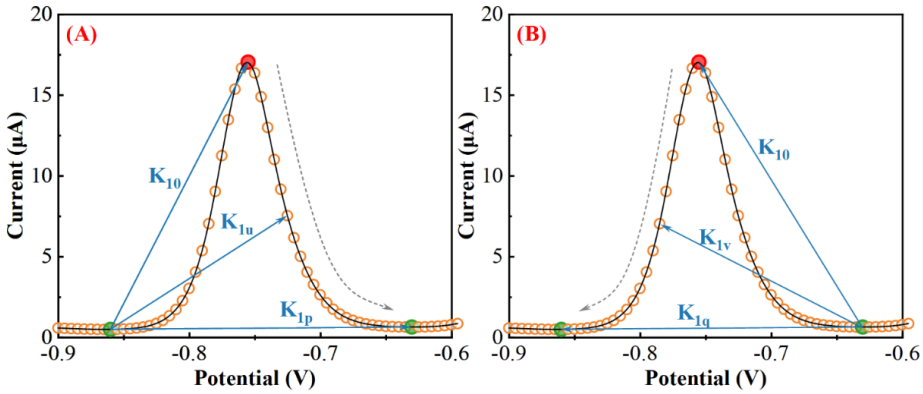

Figure 2. Schematic representation of the calculation of valid data segments (**A**,**B**). (The *i*-th segment of the SWASV response curve for 150 μg/L Zn^{2+}, 35 μg/L Cd^{2+}, 175 μg/L Pb^{2+}, 300 μg/L Bi^{3+}, and 100 μg/L Cu^{2+}).

$$K_{R1} = \begin{bmatrix} k_{10} \\ k_{11} \\ \vdots \\ k_{1v} \\ \vdots \\ k_{1q} \end{bmatrix} \tag{2}$$

where $q = \frac{V_{maxi} - V_{mini}}{0.005}$ is the number of elements in K_{R1}, i.e., the data points number from the local maximum point to the left local minimum point, $k_{1v} = \frac{I_{maxi-v} - I_{mini}}{V_{maxi-v} - V_{mini}}$ $(v \in [0, q])$ is the v-th element in K_{R1}, and $k_{1n} = max\{K_{R1}\}$ is the global maximum value in K_{R1}.

The coordinate Q_{i_1} $(V_{maxi-1n}, I_{maxi-1n})$ was acquired by k_{1n}.

Then, the slope matrix K_{L2} was calculated by the slope between every coordinate point from the local maximum point (V_{maxi}, I_{maxi}) to $P_{i_1}(V_{maxi+1m}, I_{maxi+1m})$ and $Q_{i_1}(V_{maxi-1n}, I_{maxi-1n})$ to acquire the coordinate $P_{i_2}(V_{maxi+2m}, I_{maxi+2m})$. Therefore, the above calculations were repeatedly iterated until the conditions in Equation (3) were satisfied. Finally, the coordinates of $P_{i_s}(V_{maxi+sm}, I_{maxi+sm})$ and $Q_{i_s}(V_{maxi-sn}, I_{maxi-sn})$ (s was the final number of iterations) were obtained. The segment between P_{i_s} and Q_{i_s} was the valid data segment that this study required.

$$\begin{cases} P_{i_{sm}} = P_{i_{(s-1)m}} \\ Q_{i_{sn}} = Q_{i_{(s-1)n}} \end{cases} \tag{3}$$

2.5.3. Acquirement of Peak Heights and Peak Areas

The straight line l_i from points $P_{i_{sm}}$ to $Q_{i_{sn}}$ and the curve C_i fitted to the data points between points $P_{i_{sm}}$ and $Q_{i_{sn}}$ could be determined. The area between the straight line l_i and the curve C_i was the peak area S_i, which could be calculated by Equation (4). The line connected by the local maximum point, (V_{maxi}, I_{maxi}) and $(V_{maxi}, 0)$, had an intersection, $D_i(V_{maxi}, I_{inti})$, with l_i. Thus, the peak height (h_i) could be calculated with the local maximum points and D_i according to Equation (5). An example is given in Figure 3.

$$S_i = \int_{Q_{i_{sn}}}^{P_{i_{sm}}} C_i - l_i dv \tag{4}$$

$$h_i = I_{maxi} - I_{inti} \tag{5}$$

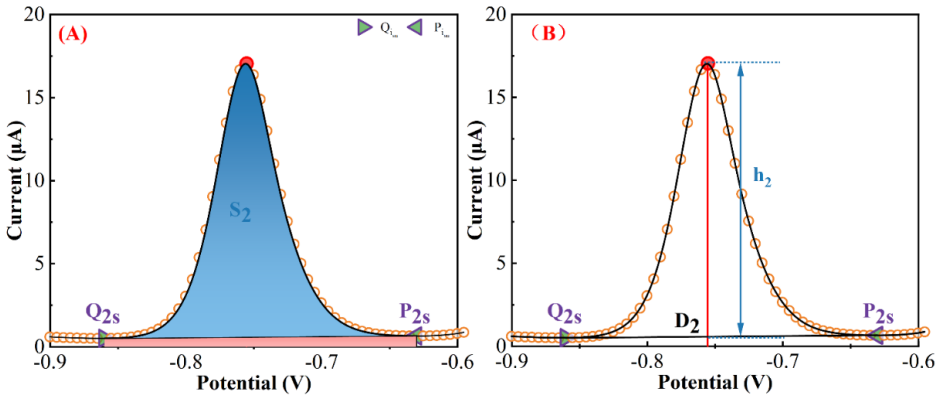

Figure 3. Schematic representation of peak area (**A**) and peak height (**B**) calculation. (The *i*-th segment of the SWASV response curve for 150 μg/L Zn^{2+}, 35 μg/L Cd^{2+}, 175 μg/L Pb^{2+}, 300 μg/L Bi^{3+}, and 100 μg/L Cu^{2+}).

The type of HMIs could be identified according to the potential corresponding to the maximum current of every segment. In this study, Zn^{2+} corresponded to a potential range of −1.2 to −0.85 V, Cd^{2+} corresponded to a potential range of −0.85 to −0.65 V, Pb^{2+} corresponded to a potential range of −0.65 to −0.35 V, Bi^{3+} corresponded to a potential range of −0.35 to −0.1 V, and Cu^{2+} corresponded to a potential range of −0.1 to 0.1 V.

2.6. Detection Models

A total of 64 SWASV curves were collected, which corresponded to the 64 groups of concentration combinations of Zn^{2+}, Cd^{2+}, Pb^{2+}, and Cu^{2+}. For each SWASV curve, there were four kinds of variables, which were the stripping current peak heights (I_{auto}) and areas (A_{auto}) acquired automatically by the homemade algorithm, and the stripping current peak heights (I_{manu}) and areas (A_{manu}) acquired manually using electrochemical workstation control software. The dataset (64 samples) was divided into a calibration set (48 samples) and a validation set (16 samples) according to the ratio of 3:1 by sample set partitioning based on joint x–y distance (SPXY) [30–33], which is described in the Supplementary Materials. In addition, partial least squares regression (PLSR) and support vector regression (SVR) were employed in this study to build detection models of Cd^{2+} and Pb^{2+} concentrations.

The PLSR algorithm is a widely used linear regression method, which can solve multicollinearity questions among multiple variables [34–37]. PLSR was used to project predictors and observed variables into a new space to find a linear regression model. However, the interference between HMIs was complex and nonlinear. The SVR, as a kind of machine learning algorithm about nonlinear regression, had recently become popular to model small size samples [25,38]. The use of the SVR kernel function could fit the nonlinear relationship into a linear or a nearly linear regression hypersurface in the high-dimensional feature space. Compared with other functions, the radial basis function (RBF), which was a kind of kernel function, had better performance due to its strong relationship between input and output. The SVR penalty parameter *c* and the kernel function parameter *g* could greatly influence the performance of the SVR regression [39]. Therefore, the particle swarm optimization (PSO) algorithm was used to automatically optimize the parameters of *c* and *g* in the SVR model. The optimization results of parameters *c* and *g* are presented in the Supplementary Materials.

2.7. Evaluation Parameters of Data and Model Performance

The I_{manu}, I_{algo}, A_{manu}, and A_{algo} of five HMIs were obtained from 64 samples. The cosine similarity (cos θ) [40,41] and determination coefficient (R_θ^2) were used to evaluate the

differences between the peak heights and peak areas obtained by the homemade algorithm and those acquired manually by the electrochemical workstation control software.

$$\cos\theta = \frac{\sum_{i=1}^{n} \alpha_i \beta_i}{\sqrt{\sum_{i=1}^{n} \alpha_i^2}\sqrt{\sum_{i=1}^{n} \beta_i^2}} \tag{6}$$

$$R_\theta^2 = 1 - \frac{\sum_{i=1}^{n} \left(\alpha_i - \beta_i^*\right)^2}{\sum_{i=1}^{n} \left(\alpha_i - \overline{\alpha}\right)^2} \tag{7}$$

where α_i and β_i refer to the five manually acquired HMI signals (I_{manu} or A_{manu}) and algorithmically acquired HMI signals (I_{algo} or A_{algo}) for sample i, respectively, and $\overline{\alpha}$ is the average value of the five manually acquired HMI signals (I_{manu} or A_{manu}) for n samples. The n is 64 in this study. The closer the values of $\cos\theta$ and R_θ^2 are to 1, the difference between the algorithm and the manually acquired values is smaller.

The calibration set (48 samples) was used to build the detection model, and the validation set (16 samples) was used to validate the model performances. The stability and accuracy of the detection model were evaluated separately using the determination coefficient (R^2) and the root−mean−square error ($RMSE$) of the validation set.

$$R^2 = 1 - \frac{\sum_{i=1}^{j} \left(y_i - y_i^*\right)^2}{\sum_{i=1}^{j} \left(y_i - \overline{y}\right)^2} \tag{8}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{j} \left(y_i - y_i^*\right)^2}{j}} \tag{9}$$

where y_i and y_i^* refer to the actual and detected concentrations of HMIs (Cd^{2+} or Pb^{2+}) for the i-th sample, respectively, and $\overline{y}$ is the average value of actual concentrations of HMIs (Cd^{2+} or Pb^{2+}) for the j-th validation set sample. A value of R^2 close to 1 indicated the high stability of the model, and the small value of $RMSE$ indicated the high detection accuracy of the model.

2.8. Preparation of Real Soil Extracts

Soil samples (National sharing platform of reference materials, China) were collected from two different provinces in China. Detailed information about the soil samples was presented in Table S2. Bioavailable HMIs were extracted during sample preparation procedures, which include ion-exchange, carbonate-bound, and soluble humus-bound fractions. The detailed preparation procedures of the two soil extracts is described in the Supplementary Materials.

3. Results and Discussion

3.1. Analysis of SWASV Signals and Input Variables

3.1.1. Analysis of Interactive Interference of Multiple HMIs by SWASV Signals

The 64 SWASV curves were plotted by the concentration of Cd^{2+} and Pb^{2+}, as shown in Figure 4. In Figure 4A, the stripping voltammetry signals of Cd^{2+} showed different shapes at the same concentration, which suggested that Cd^{2+} could be interfered with by different types and concentrations of HMIs. In addition, although it was interfered with by different concentrations of non-target ions, the stripping voltammetry signals of Cd^{2+} also gradually increased with the increase in Cd^{2+} concentration. In Figure 4B, it could be observed that Pb^{2+} possessed the same situation as Cd^{2+}, which indicated that the SWASV signals of both Cd^{2+} and Pb^{2+} could be interfered with by other HMIs. That was also consistent with our previous research [24]. Therefore, the SWASV signals of Cd^{2+} and Pb^{2+} were severely interfered with by the interaction of Zn^{2+}, Cd^{2+}, Pb^{2+}, and Cu^{2+}. It was necessary to inhibit the interactive interferences between multiple HMIs for the accurate detection of Cd^{2+} and Pb^{2+} concentrations.

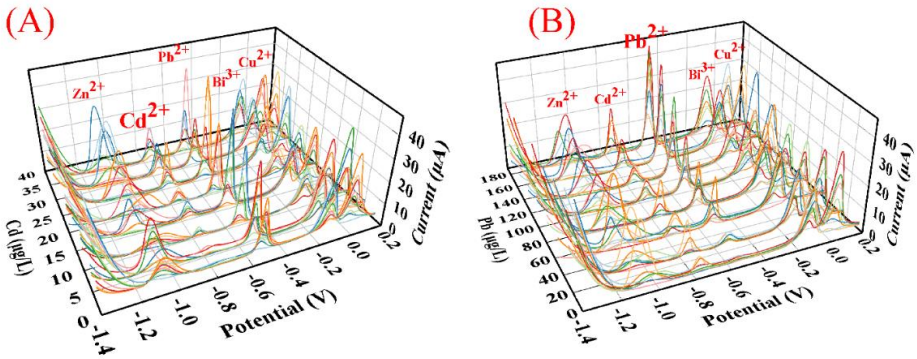

Figure 4. The SWASV curves of 64 experiments sorted by Cd^{2+} (**A**) and Pb^{2+} (**B**) concentration.

3.1.2. Comparison of Peak Heights and Peak Areas of Multiple HMIs Acquired by Different Methods

Chemometrics and machine learning provided a promising platform to inhibit the interactive interference of multiple HMIs for further accurate detection of Cd^{2+} and Pb^{2+} concentrations. Appropriate input variables could enhance the accuracy of detection models. Therefore, the I_{manu}, I_{algo}, A_{manu}, and A_{algo} of Cd^{2+} and Pb^{2+} were acquired by the homemade algorithm and the electrochemical workstation control software, as shown in Figure 5 and Figures S3–S5. The distribution of scattered points was consistent with the results of the analysis in Figure 4A,B, and indirectly confirmed the effectiveness of the homemade algorithm. For I_{manu} and I_{algo} of Zn^{2+}, Cd^{2+}, and Pb^{2+}, the R_θ^2 and $\cos\theta$ were all greater than 0.9997. For I_{manu} and I_{algo} of Bi^{3+} and Cu^{2+}, the R_θ^2 and $\cos\theta$ were all greater than 0.9980. For M_{manu} and M_{algo} of Zn^{2+}, Cd^{2+}, and Pb^{2+}, the R_θ^2 and $\cos\theta$ were all greater than 0.9999. For M_{manu} and M_{algo} of Bi^{3+} and Cu^{2+}, the R_θ^2 and $\cos\theta$ were all greater than 0.9970. Detailed information about R_θ^2 and $\cos\theta$ values was presented in Table S3. Compared with Zn^{2+}, Cd^{2+}, and Pb^{2+}, the peak heights and peak areas of manually acquired Bi^{3+} and Cu^{2+} differed significantly from those acquired by the homemade algorithm, which was attributed to the overlapping peaks between Bi^{3+} and Cu^{2+}. It was prone to human errors when manually obtaining the peak height and area of overlapping peaks.

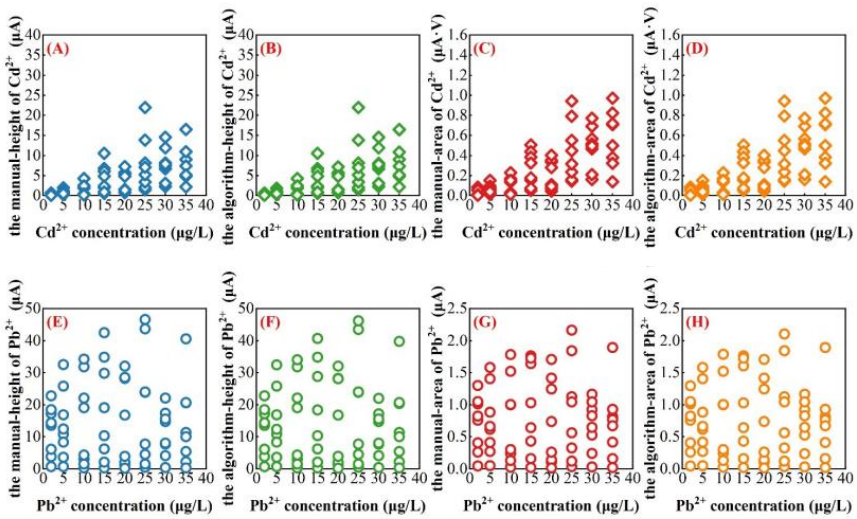

Figure 5. I_{manu} (**A**,**E**), I_{algo} (**B**,**F**), A_{manu} (**C**,**G**), and A_{algo} (**D**,**H**) of Cd^{2+} and Pb^{2+} acquired from the SWASV curves of 64 experiments. (Diamond scatters represent the peak heights or peak areas of Cd^{2+}, circular scatters represent the peak heights or peak areas of Pb^{2+}. Blue scatters represent the I_{manu}, green scatters represent the I_{algo}, red scatters represent the A_{manu}, orange scatters represent the A_{algo}).

3.2. Detection of Cd^{2+} and Pb^{2+} Concentrations by Peak Heights

The peak heights of Zn^{2+}, Cd^{2+}, Pb^{2+}, Bi^{3+}, and Cu^{2+} obtained by manual acquisition (I_{manu}) and algorithmic acquisition (I_{algo}) were used as input for PLSR and SVR models to detect the concentrations of Cd^{2+} and Pb^{2+}. The performance of the above models was compared using the *RMSE* and R^2 values of the validation set. All results of the calibration set were presented in Figure S6. Detailed information about the PLSR and SVR models was presented in Table S4.

3.2.1. PLSR Models Established Using Peak Heights

The PLSR models built using the I_{manu} and I_{algo} of the HMIs were named I_{manu}-PLSR and I_{algo}-PLSR, respectively. As shown in Figure 6A,B, the I_{manu}-PLSR model (*RMSE* = 5.4663 μg/L, R^2 = 0.7375) had close accuracy to the I_{algo}-PLSR model (*RMSE* = 5.4838 μg/L, R^2 = 0.7323) for the detection of Cd^{2+} concentrations. However, the I_{manu}-PLSR model (*RMSE* = 19.2392 μg/L, R^2 = 0.8645) presented a smaller *RMSE* and a larger R^2 value than I_{algo}-PLSR model (*RMSE* = 21.9047 μg/L, R^2 = 0.7766) for detecting Pb^{2+} concentrations, as shown in Figure 6C,D. Although the accuracy and stability of I_{algo}-PLSR model was inferior to that of the I_{manu}-PLSR model for both Cd^{2+} and Pb^{2+}, the discrepancies between them were not significant, which illustrated the developed algorithm could accurately acquire the peak heights.

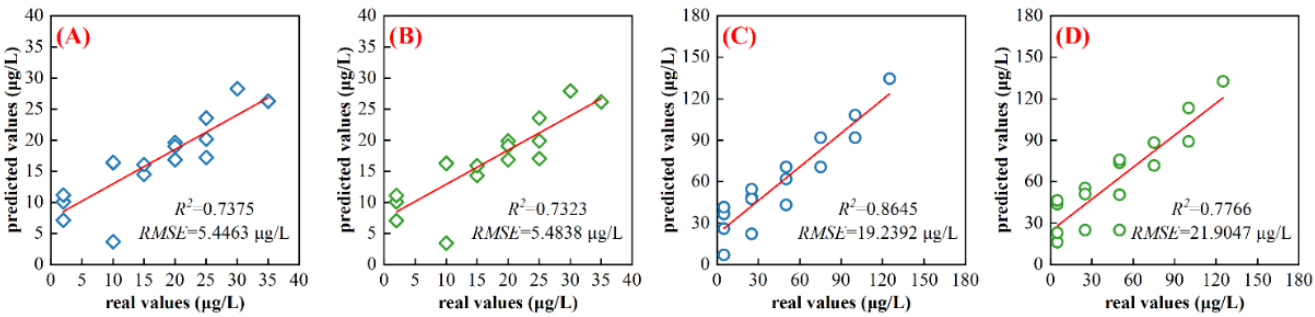

Figure 6. PLSR model results for Cd^{2+} (**A**,**B**) and Pb^{2+} (**C**,**D**) concentration detection in the validation dataset. (Diamond scatters represent Cd^{2+} concentrations, circular scatters represent Pb^{2+} concentrations, blue scatters represent the model input as I_{manu}, green scatters represent the model input as I_{algo}).

However, the PLSR models established using peak heights had the larger *RMSE* values for Cd^{2+} and Pb^{2+} concentrations and needed to be further improved. This might be because PLSR could not be adequate for resolving the interactive interference among HMIs. Hence, it became a possibility to use of a nonlinear model to improve the model detection accuracy of Cd^{2+} and Pb^{2+}.

3.2.2. SVR Models Established Using Peak Heights

The SVR models built using the I_{manu} and I_{algo} of the HMIs were named I_{manu}-SVR and I_{algo}-SVR, respectively. As shown in Figure 7A,B, for the detection of Cd^{2+} concentration, the *RMSE* and R^2 values of the I_{manu}-SVR model were 5.3916 μg/L and 0.7650, respectively, and that of the I_{algo}-SVR model were 3.9922 μg/L and 0.8424, respectively. As shown in Figure 7C,D, for the detection of Pb^{2+} concentration, the *RMSE* and R^2 values of the I_{manu}-SVR model were 20.0015 μg/L and 0.8791, respectively, and that of the I_{algo}-SVR model were 21.3431 μg/L and 0.8229, respectively. Compared with the I_{manu}-SVR model, the I_{algo}-SVR model had a lower *RMSE* value for the detection of Cd^{2+} concentrations. However, for the Pb^{2+} concentrations, the *RMSE* values of the I_{manu}-SVR and the I_{algo}-SVR models were approximate, which illustrated that the homemade algorithm could accurately acquire the peak heights of HMIs.

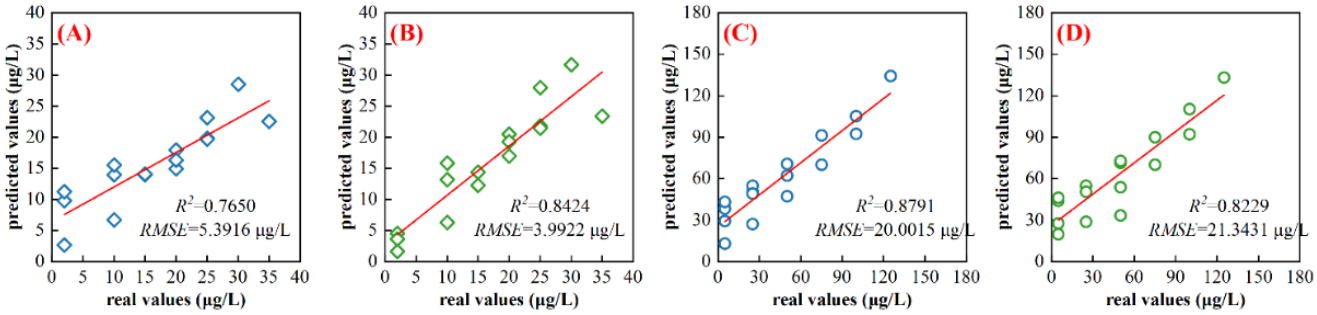

Figure 7. SVR model results for Cd^{2+} (**A**,**B**) and Pb^{2+} (**C**,**D**) concentration detection in the validation dataset. (Diamond scatters represent Cd^{2+} concentrations, circular scatters represent Pb^{2+} concentrations, blue scatters represent the model input as I_{manu}, green scatters represent the model input as I_{algo}).

In addition, for Cd^{2+} concentration, the I_{manu}-SVR model had a lower *RMSE* value of 0.0547 μg/L than the I_{manu}-PLSR model and a higher R^2 value of 0.0275, while the I_{algo}-SVR model had a lower *RMSE* value of 1.4916 μg/L and a higher R^2 value of 0.1101 than the I_{algo}-PLSR model. The SVR model built using algorithm-acquiring peak heights (I_{algo}) as its input significantly improved the detection accuracy of Cd^{2+} concentration. This was because the signal of Cd^{2+} was much weaker than Pb^{2+} due to the low concentration of Cd^{2+}, which caused the large relative errors of Cd^{2+} peak heights acquired manually by the electrochemical control software, while the homemade algorithm in this study could exactly capture the weak Cd^{2+} peak heights.

For Pb^{2+} concentration, the I_{manu}-SVR model had a higher *RMSE* value of 0.7623 μg/L and a higher R^2 value of 0.0146 than the I_{manu}-PLSR model, while the I_{algo}-SVR model had a lower *RMSE* value of 0.5616 μg/L and a higher R^2 value of 0.0463 than the I_{algo}-PLSR model. However, the increase in detection accuracy of Pb^{2+} concentrations was either negligible or failed by the SVR and algorithm-acquired peak heights (I_{algo}). A more important reason for the above results was that the peak heights contained too little information to reflect the serious interaction between multiple HMIs. Therefore, the stripping current peak areas were employed as the inputs of SVR and PLSR models in the next section in expectation of improving the detection accuracy of Cd^{2+} and especially Pb^{2+} concentrations.

3.3. Detection of Cd^{2+} and Pb^{2+} Concentrations by Peak Areas

Same as Section 3.2, the peak areas of Zn^{2+}, Cd^{2+}, Pb^{2+}, Bi^{3+}, and Cu^{2+} obtained by manual acquisition (A_{manu}) and algorithmic acquisition (A_{algo}) were used as input for the PLSR and SVR models to detect the concentrations of Cd^{2+} and Pb^{2+}. The performance of the above models was compared the *RMSE* and R^2 values of the validation set. All of the results of the calibration set were presented in Figure S7. Detailed information about the PLSR and SVR models was presented in Table S5.

3.3.1. PLSR Models Established Using Peak Areas

The PLSR models built using the A_{manu} and A_{algo} of HMIs were named A_{manu}-PLSR and A_{algo}-PLSR, respectively. As shown in Figure 8A,B, the A_{manu}-PLSR model (*RMSE* = 3.8996 μg/L, R^2 = 0.8828) had close accuracy to the A_{algo}-PLSR model (*RMSE* = 3.7414 μg/L, R^2 = 0. 8910) for the detection of Cd^{2+} concentration. As shown in Figure 8C,D, the A_{manu}-PLSR model (*RMSE* = 15.1333 μg/L, R^2 = 0.9272) presented a larger *RMSE* value and a smaller R^2 value than the A_{algo}-PLSR model (*RMSE* = 12.9264 μg/L, R^2 = 0.9668) for the detection of Pb^{2+} concentration. For both Cd^{2+} and Pb^{2+} concentration detection, the A_{algo}-PLSR model was more stable and accurate than the A_{manu}-PLSR model. The use of A_{algo} as a model input contributed to the improvement of the detection accuracy and stability of the models, which implied that the homemade algorithm could accurately acquire the peak areas of HMIs.

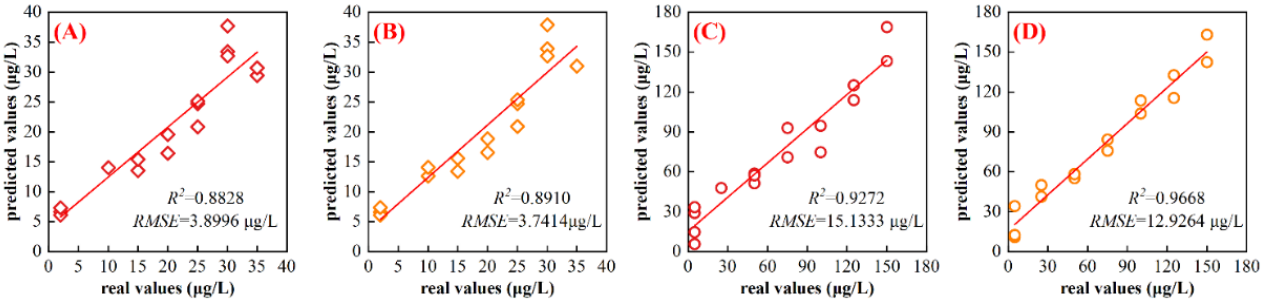

Figure 8. PLSR model results for Cd^{2+} (**A**,**B**) and Pb^{2+} (**C**,**D**) concentration detection in the validation dataset. (Diamond scatters represent Cd^{2+} concentrations, circular scatters represent Pb^{2+} concentrations, red scatters represent the model input as A_{manu}, orange scatters represent A_{algo} as the model input).

In addition, for Cd^{2+} concentration, A_{manu}-PLSR model had a lower *RMSE* value of 1.5467 μg/L and a higher R^2 value of 0.1453 than the I_{manu}-PLSR model, while the A_{algo}-PLSR model had a lower *RMSE* value of 1.7424 μg/L and a higher R^2 value of 0.1587 than the I_{algo}-PLSR model. For Pb^{2+} concentration, the A_{manu}-PLSR model had a lower *RMSE* value of 4.1059 μg/L and a higher R^2 value of 0.0627 than I_{manu}-PLSR model, while the A_{algo}-PLSR model had a lower *RMSE* value of 8.9785 μg/L and a higher R^2 value of 0.1902 than the I_{algo}-PLSR model. In the case of modeling with the PLSR algorithm, using peak areas instead of peak heights as the model input greatly improved the stability and accuracy of the Cd^{2+} and Pb^{2+} concentration detection models. In addition, for both the *RMSE* and R^2 values, the improvement of the I_{manu}-SVR models and the I_{algo}-SVRs models over the I_{manu}-PLSR models and the I_{algo}-PLSR models, respectively, was significantly less than the improvement of the A_{manu}-PLSR models and the A_{algo}-PLSR models over the I_{manu}-PLSR models and the I_{algo}-PLSR models, respectively. This might be related to the fact that the peak areas contained more comprehensive SWASV information of HMIs than the peak heights. The peak area of HMIs was defined as the integral value of the stripping current on the scale of the stripping potential time, which physically represented the amount of ions. The above analysis provided a new direction for the accurate detection of Cd^{2+} and Pb^{2+} under the interactive interference of multiple HMIs, i.e., using peak areas as a model input and building models with the SVR algorithm.

3.3.2. SVR Models Established Using Peak Areas

The SVR models built using the A_{manu} and the A_{algo} of HMIs were named A_{manu}-SVR and A_{algo}-SVR, respectively. As shown in Figure 9A,B, for the detection of Cd^{2+} concentration, the *RMSE* and R^2 values of the A_{manu}-SVR model were 3.9824 μg/L and 0.8785, respectively, and that of the A_{algo}-SVR model were 2.9906 μg/L and 0.9204, respectively. As shown in Figure 9C,D, for the detection of Pb^{2+} concentration, the *RMSE* and R^2 values of the A_{manu}-SVR model were 13.3444 μg/L and 0.9359, respectively, and that of the A_{algo}-SVR model were 13.1574 μg/L and 0.9756, respectively. For both Cd^{2+} and Pb^{2+} concentration detection, A_{algo}-SVR was more stable and accurate than A_{manu}-SVR, which again implied that the homemade algorithm could accurately acquire the peak areas of the HMIs.

For the detection of Cd^{2+} concentrations, the A_{manu}-SVR model had a lower *RMSE* value of 1.4092 μg/L and a higher R^2 value of 0.1135 than I_{manu}-SVR model, while the A_{algo}-SVR model had a lower *RMSE* value of 1.0016 μg/L and a higher R^2 value of 0.078 than the I_{algo}-SVR model. For the detection of Pb^{2+} concentration, the A_{manu}-SVR model had a lower *RMSE* value of 6.6571 μg/L and a higher R^2 value of 0.0568 than the I_{manu}-SVR model, while the A_{algo}-SVR model had a lower *RMSE* value of 8.1857 μg/L and a higher R^2 value of 0.1527 than the I_{algo}-SVR model. In the case of modeling with SVR algorithm, using peak areas instead of peak heights as model input greatly improved the detection

stability and accuracy for Cd^{2+} and Pb^{2+} concentrations. In addition, for the detection of Cd^{2+} concentration, the A_{manu}-SVR model had a higher *RMSE* value of 0.0828 μg/L and a lower R^2 value of 0.0043 than the A_{manu}-PLSR model, while the A_{algo}-SVR model had a lower *RMSE* value of 0.7508 μg/L and a higher R^2 value of 0.0294 than the A_{algo}-PLSR model. For the detection of Pb^{2+} concentration, the A_{manu}-SVR model had a lower *RMSE* value of 1.7889 μg/L and a higher R^2 value of 0.0087 than the A_{manu}-PLSR model, while the A_{algo}-SVR model had a higher *RMSE* value of 0.2310 μg/L and a higher R^2 value of 0.0088 than the A_{algo}-PLSR model. Compared with the models built by the PLSR algorithm, the accuracy of the A_{manu}-SVR model for Cd^{2+} concentration detection and the A_{algo}-SVR model for Pb^{2+} concentration detection decreased slightly after modeling by the SVR algorithm. Additionally, the stability and accuracy of the A_{algo}-SVR model for the detection of Cd^{2+} concentration and the A_{manu}-SVR model for detection of Pb^{2+} concentration were further improved. In summary, the A_{algo}-SVR model was the most stable and accurate model for detecting Cd^{2+} concentration as shown in Figure 8B. The A_{algo}-PLSR model was the most accurate and the A_{algo}-SVR model was the most stable for detecting Pb^{2+} concentration as shown in Figures 7D and 8D.

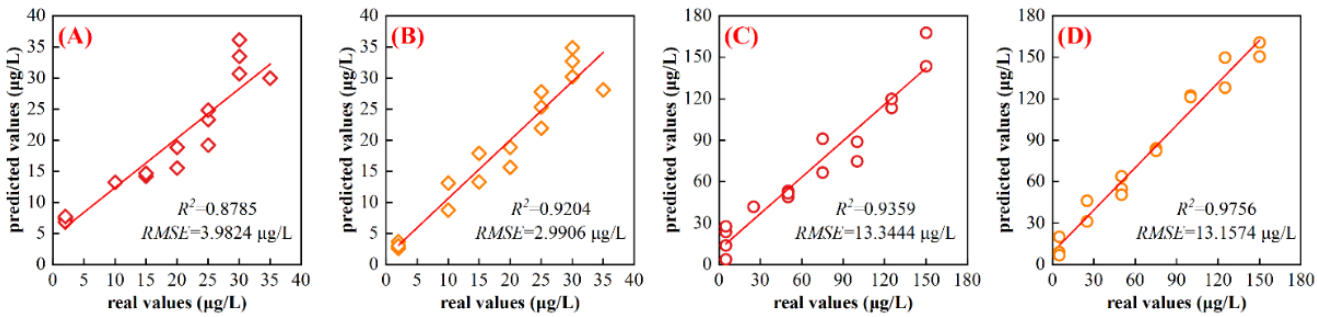

Figure 9. PSO-SVR model results for Cd^{2+} (**A**,**B**) and Pb^{2+} (**C**,**D**) concentration detection in the validation dataset. (Diamond scatters represent Cd^{2+} concentrations, circular scatters represent Pb^{2+} concentrations, red scatters represent the model input as A_{manu}, orange scatters represent the model input as A_{algo}).

3.4. Comparison and Analysis of Different Detection Models

To accurately detect Cd^{2+} and Pb^{2+} concentrations in the presence of multiple HMIs interactive interference, four kinds of inputs and two modeling algorithms were used to explore the excellent detection model. Additionally, in order to improve the acquisition efficiency of peak heights and peak areas, and to eliminate human error, a homemade peak heights and peak areas acquisition algorithm was designed. The performances of various models for Cd^{2+} and Pb^{2+} concentration detection were compared by using the *RMSE* and R^2 values of the validation sets, as shown in Figure 10. The *RMSE* values of the models detecting Cd^{2+} concentration decreased in the order of I_{algo}-PLSR > I_{manu}-PLSR > I_{manu}-SVR > I_{algo}-SVR > A_{maun}-SVR > A_{manu}-PLSR > A_{algo}-PLSR > A_{algo}-SVR, and the R^2 values of those increased in the order of I_{manu}-PLSR > I_{algo}-PLSR > I_{manu}-SVR > I_{algo}-SVR > A_{manu}-SVR > A_{manu}-PLSR > A_{algo}-PLSR > A_{algo}-SVR. The *RMSE* values of the models detecting Pb^{2+} concentration decreased in the order of I_{algo}-PLSR > I_{algo}-SVR > I_{manu}-SVR > I_{manu}-PLSR > A_{manu}-PLSR > A_{manu}-SVR > A_{algo}-SVR > A_{algo}-PLSR, and the R^2 values of those increased in the order of I_{algo}-PLSR > I_{algo}-SVR > I_{manu}-PLSR > I_{manu}-SVR > A_{manu}-PLSR > A_{manu}-SVR > A_{algo}-PLSR > A_{algo}-SVR. All models for detecting Cd^{2+} and Pb^{2+} concentrations built using peak areas as input had smaller *RMSE* values and larger R^2 values than those models built using peak heights as input. This suggested that the peak areas of the HMIs could better reflect the comprehensive information of interactive interference between multiple heavy metals. The A_{algo}-SVR model had the best stability and accuracy for the detection of Cd^{2+} concentration. For detecting Pb^{2+}

concentration, the A_{algo}-SVR model had a slightly higher *RMSE* value of 0.231 μg/L than the A_{algo}-PLSR model. However, the difference of *RMSE* values between the calibration set and the validation set of the A_{algo}-PLSR and the A_{algo}-SVR models were 6.9065 μg/L and 2.6930 μg/L, respectively, which suggested that A_{algo}-SVR had a higher stability than A_{algo}-PLSR. As a result, A_{algo}-SVR had the best detection performance toward Pb^{2+} concentration. Interestingly, the change in model input (from peak heights to peak areas) greatly improved the model accuracy compared with the change in modeling algorithm (from PLSR to SVR). This validated the idea that peak areas as model input could significantly improve the detection accuracy of Cd^{2+} and Pb^{2+} concentrations.

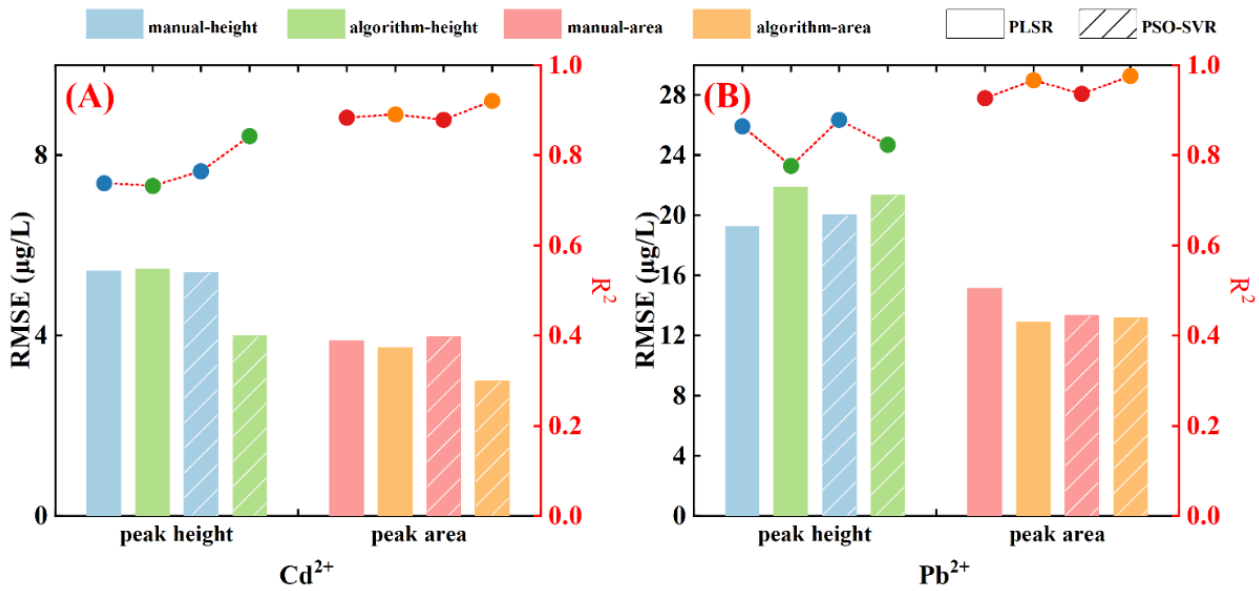

Figure 10. Model results of the validation sets for Cd^{2+} (**A**) and Pb^{2+} (**B**) concentration detection.

In all models using peak heights as input, the detection ability of those models built by I_{algo} was slightly inferior to I_{manu}, except for Cd^{2+} concentration detected by I_{algo}-SVR. This might be related to the relatively lower peak heights acquired by the homemade algorithm compared to those acquired manually by electrochemical control software, as shown in Figure 11B. In Section 2.5.1, the extreme points in the SWASV curves were recognized by the S-G algorithm and peakdet function. In this process, some noise points were eliminated, such as the recognition of the local extreme points of the curve in the dotted box in Figure 11A. Therefore, the extreme point obtained by the homemade algorithm was the red point in Figure 11B, while the point with the largest current value in dotted box was the point to the right of the red point. In addition, the peak heights were also influenced by the background currents. The manually acquired peak heights might have some deviations.

Moreover, in comparison with PLSR models, SVR models enhanced the detection performance of Cd^{2+} and Pb^{2+} concentrations. This illustrated that the interactive interference between HMIs was perfectly nonlinear in presentation. Previous studies [5,11,12] had reported that Zn^{2+}, and especially Cu^{2+}, could seriously interfere with the SWASV signals of Cd^{2+} and Pb^{2+}. The Cd^{2+} would not be detected by SWASV when Cu^{2+} concentration exceeded a certain concentration [13], thus the peak heights and peak areas could not be acquired by handwork or algorithms. However, the interference of Zn^{2+} on Cd^{2+} was negligible until Zn^{2+} reached a relatively high concentration [18]. In addition, there was mutual interference between Cd^{2+} and Pb^{2+}, and this interference was also nonlinear [12]. The nonlinear interference of multiple HMIs might be due to the following three aspects. (1) Although the introduction of Bi^{3+} could form alloys with most of the HMIs to reduce the activation energy required for electrodeposition [27,29], the concentrations of Zn^{2+}, Cd^{2+}, Pb^{2+}, and Cu^{2+} were ever-changing in this study and they required different concentrations of Bi^{3+}. Therefore, it was unknown whether the number of alloys formed by Bi^{3+} and

HMIs exhibited a linear variation with the increase in HMI concentrations. (2) During the deposition process, the number of deposition HMIs might not be in line with the increasing of HMI concentrations due to the limited deposition sites of GCE. (3) The alloys formed by multiple HMIs (especially the existence of Cu^{2+}) prevented the stripping of Cd^{2+} and Pb^{2+} off the GCE surface in the stripping process. Therefore, the detection accuracy of the PLSR models for Cd^{2+} and Pb^{2+} concentrations was lower than that of the SVR models.

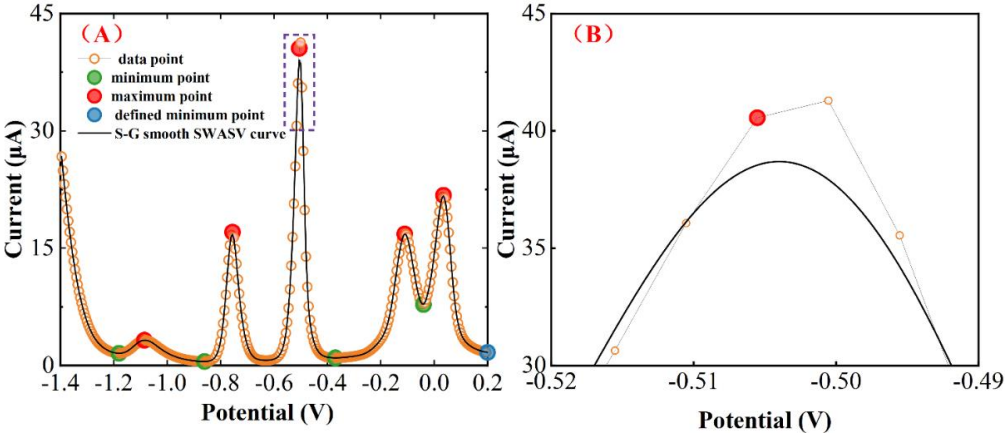

Figure 11. Schematic representation of the local maximum points calculated by homemade algorithm (**A**), and the deviator between the local extreme points of the curve in the dotted box and the extreme point (red point) obtained by the homemade algorithm (**B**).

From the above analysis, it could be concluded that the peak areas acquired by designed algorithm and SVR model significantly improved the SWASV detection accuracy of Cd^{2+} and Pb^{2+} concentrations by inhibiting the interactive interference of Zn^{2+}, Cd^{2+}, Pb^{2+}, and Cu^{2+}. Previous studies were also carried out based on Bi/GEC without other chemical or biological material modifications. Table 1 presents the comparison of this study with previous studies that built Cd^{2+} and Pb^{2+} detection models using peak heights rather than peak areas to suppress the interactive interference of multiple HMIs. Compared with previous studies, the machine learning models that used peak heights as input exhibited large *RMSE* values and small R^2 values for Cd^{2+} and/or Pb^{2+} concentrations, or only the interference of less than two non-target ions was resolved so that the models had small *RMSE* values and larger R^2 values. However, the models built in this study had the powerful ability to inhibit the interactive interferences among multiple HMIs for the detection of Cd^{2+} and Pb^{2+} concentrations.

Table 1. Comparison of the detection accuracy of various inputs and machine learning models for Cd^{2+} and/or Pb^{2+} under the interference of non-target HMIs.

Target Ions (Concentration: μg/L)	Interference Ions (Concentration: μg/L)	Inputs	Models	*RMSE* (μg/L)	R^2	Reference
Cd^{2+} (1–35)	Cu^{2+} (1–35), Pb^{2+} (1–35)	Peak currents [1]	SVR	0.9299	0.995	[12]
Cd^{2+} (1–35)	Pb^{2+} (0–150)	Peak currents	PLS SVR	5.3207 4.2277	0.746 0.867	[5]
Cd^{2+} (2–35)	Cu^{2+} (0–350), Zn^{2+} (0–175), Pb^{2+} (5–175)	Peak currents	PLS SVR	11.281 4.655	0.361 0.832	[18]
Cd^{2+} (2–35)	Cu^{2+} (0–350), Zn^{2+} (0–175), Pb^{2+} (5–175)	Peak areas [2]	SVR	2.9906	0.9204	This work
Pb^{2+} (1–110)	Cd^{2+} (0–110)	Peak currents	BP-ANN	1.69	0.998	[24]
Pb^{2+} (1–45)	Cu^{2+} (0–25)	Peak currents	SVR	1.1204	0.994	[25]
Pb^{2+} (5–175)	Cu^{2+} (0–350), Zn^{2+} (0–175), Cd^{2+} (2–35)	Peak currents	PLS SVR	25.234 25.119	0.811 0.811	[18]
Pb^{2+} (5–175)	Cu^{2+} (0–350), Zn^{2+} (0–175), Cd^{2+} (2–35)	Peak areas	SVR	13.1574	0.9756	This work

Abbreviations: SVR, support vector regression; PLSR, partial least squares regression; BP-ANN, backpropagation-artificial neural network. **Notes:** [1], Peak currents refer to the stripping peak currents of target and non-target HMIs; [2], Peak areas refer to the stripping peak area of target and non-target HMIs.

3.5. Application of A_{algo}-SVR for Detecting Cd^{2+} and Pb^{2+} Concentrations in Soil Extract Solution

To examine the applicability of the developed Cd^{2+} and Pb^{2+} detection models, two soil extract solution samples were prepared according to Section 2.8. As shown in Figure S8, the SWASV measurement results indicated that both soil extract solution samples contained Zn^{2+} and Cu^{2+}.The peak areas of five HMIs were acquired by the homemade algorithm (Section 2.5) and, then, substituted into the A_{algo}-SVR model to detect Cd^{2+} and Pb^{2+} concentrations, respectively. As shown in Table 2, the detection results of the two soil extract solution samples showed that the model A_{algo}-SVR was 0.094 μg/L and 0.0313 μg/L lower than ICP-MS in the detection of Cd^{2+} concentration, respectively, and A_{algo}-SVR was 0.1845 μg/L and 0.4731 μg/L lower than ICP-MS in the detection of Pb^{2+} concentration, respectively. The detection error for Pb^{2+} was slightly larger compared to Cd^{2+}, which was due to the fact that there was more Pb^{2+} in the soil than Cd^{2+}. However, the relative standard deviations for Cd^{2+} and Pb^{2+} were 5.69% and 1.83%, respectively, verifying the good practicality of the method proposed in this study.

Table 2. Detection results of Cd^{2+} and Pb^{2+} concentrations in soil extract solutions.

Soil Extracts	Cd^{2+} Detection (μg/L)		Pb^{2+} Detection (μg/L)	
	A_{algo}-SVR	ICP-MS	A_{algo}-SVR	ICP-MS
Sample No. 1	1.0067	1.1007	18.5140	18.6985
Sample No. 2	1.0698	1.1011	17.2347	17.7078

In summary, in situ bismuth modified GCE combined with the homemade peak area extraction algorithm, chemometrics, electrochemistry and machine learning can replace ICP-MS to detect Cd^{2+} and Pb^{2+} in soil. This method saves detection time, compresses costs and is more suitable for rapid on-site detection requirements.

3.6. Limitations and Prospects

Soil is a complex environment containing multiple substances. There are many interfering factors that will decrease the detection accuracy of Cd^{2+} and Pb^{2+} in soil extracts, such as the mutual interactions of multiple HMIs and the soluble humus [6,7,42]. This study combined simple Bi-film modified electrode and machine learning models to accurately detect Cd^{2+} and Pb^{2+}, which alleviated to some extent the interactive interference of multiple HMIs. However, the concentrations of Zn^{2+}, Cd^{2+}, Pb^{2+} and Cu^{2+} in soil may be more or less than the concentration ranges of the established dataset; Hg^{2+} and As^{3+} in soils also interfere with the detection of Cd^{2+} and Pb^{2+}. Therefore, the generality and universality of the built detection models need further validation [42]. In this regard, it is necessary to build a dataset with a wider range of HMI concentrations and more heavy metals types for the accurate detection of Cd^{2+} and Pb^{2+} in real soil extracts, and it is hopeful to further enhance the model detection ability by using deep learning algorithms. On the other hand, soluble humus (such as humic acid and fulvic acid) in soil extracts will complex with heavy metal ions and interfere with the stripping voltammetry signals of Cd^{2+} and Pb^{2+} [6,7]; thus, it is necessary to suppress the humus substance interference in further studies for accurately determining Cd^{2+} and Pb^{2+} in real soil extracts by the SWASV method.

4. Conclusions

The purpose of this study is to investigate the low accuracy of Cd^{2+} and Pb^{2+} concentration detection when using peak current and SWASV to detect heavy metals in soil. A novel algorithm based on extreme points and slopes for the automatic and accurate acquisition of peak heights and peak areas was designed first in this study. Then, Cd^{2+} and Pb^{2+} detection models were built by combining chemometrics and machine learning algorithms. The peak heights and peak areas of the Zn^{2+}, Cd^{2+}, Pb^{2+}, Bi^{3+}, and Cu^{2+} acquired by both the homemade algorithm and the electrochemical workstation software were used

acting as input variables to establish the PLSR and SVR detection models. The results of the modelling analysis showed that the SVR models established using algorithmically acquired peak areas presented the highest accuracy for the detection of Cd^{2+} and Pb^{2+} concentrations. The *RMSE* and R^2 values of the optimal Cd^{2+} concentration detection model (A_{algo}-SVR) were 2.9906 μg/L and 0.9204, respectively. The *RMSE* and R^2 values of the optimal Pb^{2+} concentration detection model (A_{algo}-SVR) were 13.1574 μg/L and 0.9756, respectively. The optimal A_{algo}-SVR model was used for real soil extracts to verify the practicability of the model for detecting Cd^{2+} and Pb^{2+} concentrations, respectively. This work provides a new method in the accurate detection of Cd^{2+} and Pb^{2+} concentrations under the co-existence of multiple HMIs in soils.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/met13020270/s1, Figure S1: CV curve of the bare GCE in the 5 mM $[Fe(CN)_6]^{3-/4-}$ solution with 0.1 M KCl; Figure S2: SWASV responses of (**a**) 20 μg/L Zn^{2+}, Cd^{2+}, and Pb^{2+} on GCE and Bi/GCE, (**b**) 20 μg/L Zn^{2+}, Cd^{2+}, Pb^{2+}, and Cu^{2+} on GCE and Bi/GCE, and (**c**,**d**) 40 μg/L Zn^{2+}, Cd^{2+}, Pb^{2+}, and Cu^{2+} on Bi/GCE for ten repetitive measurements; Figure S3: I_{manu} (A), I_{algo} (B), A_{manu} (C) and A_{algo} (D) of Zn^{2+} acquired from the SWASV curves of 64 experiments. (Blue scatters represent the I_{manu}, green scatters represent the I_{algo}, red scatters represent the A_{manu}, orange scatters represent the A_{algo}); Figure S4: I_{manu} (A), I_{algo} (B), A_{manu} (C) and A_{algo} (D) of Bi^{3+} acquired from the SWASV curves of 64 experiments. (Blue scatters represent the I_{manu}, green scatters represent the I_{algo}, red scatters represent the A_{manu}, orange scatters represent the A_{algo}); Figure S5: I_{manu} (A), I_{algo} (B), A_{manu} (C) and A_{algo} (D) of Cu^{2+} acquired from the SWASV curves of 64 experiments. (Blue scatters represent the I_{manu}, green scatters represent the I_{algo}, red scatters represent the A_{manu}, orange scatters represent the A_{algo}); Figure S6: PLSR models (A, B, C and D) and SVR models (E, F, G and H) results for Cd^{2+} and Pb^{2+} concentrations detection in the calibration dataset. (Diamond scatters represent Cd^{2+} concentrations, circular scatters represent Pb^{2+} concentrations, blue scatters represent the model input as I_{manu}, green scatters represent the model input as I_{algo}.); Figure S7: PLSR models (A, B, C and D) and SVR models (E, F, G and H) results for Cd^{2+} and Pb^{2+} concentrations detection in the calibration dataset. (Diamond scatters represent Cd^{2+} concentrations, circular scatters represent Pb^{2+} concentrations, red scatters represent the model input as A_{manu}, orange scatters represent A_{algo} as the model input.); Figure S8: SWASV response curves of soil extract samples;Table S1: the excogitation of 64 sets of orthogonal experiments by the L64(8^4) orthogonal table; Table S2: Detailed information on real soil samples; Table S3: Detailed R_θ^2 and $\cos\theta$ of the peak heights and peak areas; Table S4: The R^2 values and *RMSE* values of the models built by peak heights; Table S5: The R^2 values and *RMSE* values of the models.

Author Contributions: Data curation, W.Y. and N.L.; funding acquisition, G.L.; investigation, N.L.; methodology, W.Y. and G.Z.; project administration, G.L.; software, W.Y.; supervision, G.L.; validation, N.L.; visualization, W.Y.; writing—original draft, W.Y.; writing—review and editing, N.L. and G.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by the National Natural Science Foundation of China (Grant No. 32071898) and the Natural Science Foundation of Jiangsu Province (No. BK20200546).

Data Availability Statement: The data presented in this study are available in Supplementary Materials.

Acknowledgments: This work was assisted by Yinghui Li (Shanghai Jiao Tong University), who provided me with software support and writing advice.

Conflicts of Interest: The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported on in this paper.

References

1. Singh, N.; Gupta, V.K.; Kumar, A.; Sharma, B. Synergistic Effects of Heavy Metals and Pesticides in Living Systems. *Front. Chem.* **2017**, *5*, 70. [CrossRef] [PubMed]
2. Pande, A.; Mun, B.; Methela, N.J.; Rahim, W.; Lee, D.; Lee, G.; Hong, J.K.; Hussain, A.; Loake, G.; Yun, B. Heavy metal toxicity in plants and the potential NO-releasing novel techniques as the impending mitigation alternatives. *Front. Plant Sci.* **2022**, *13*, 1019647. [CrossRef]

3. Wang, L.; Zeraatpisheh, M.; Wei, Z.; Xu, M. Heavy metal pollution and risk assessment of farmland soil around abandoned domestic waste dump in Kaifeng City. *Front. Environ. Sci.* **2022**, *10*, 1440. [CrossRef]
4. Liao, J.; Song, Y.; Wang, R.; Zeng, Y.; Si, H.; Ge, C.; Lin, S. A synergistic promotion strategy for selective trapping and sensing of lead(II) by oxygen-vacancy and surface modulation of MnO_2 nanoflowers. *Sens. Actuators B Chem.* **2021**, *345*, 130384. [CrossRef]
5. Liu, N.; Zhao, G.; Liu, G. Accurate SWASV detection of Cd(II) under the interference of Pb(II) by coupling support vector regression and feature stripping currents. *J. Electroanal. Chem.* **2021**, *889*, 115227. [CrossRef]
6. Liu, N.; Zhao, G.; Ye, W.; Wang, J.; Liu, G. VUV-H_2O_2 photolysis as a pretreatment method for improving the SWASV detection accuracies of Cd^{2+} and Pb^{2+} in soil extracts. *J. Environ. Chem. Eng.* **2022**, *10*, 107813. [CrossRef]
7. Liu, N.; Ye, W.; Zhao, G.; Liu, G. Release of free-state ions from fulvic acid-heavy metal complexes via VUV/H_2O_2 photolysis: Photodegradation of fulvic acids and recovery of Cd^{2+} and Pb^{2+} stripping voltammetry currents. *Environ. Pollut.* **2022**, *315*, 120420. [CrossRef]
8. Wang, Y.; Wu, X.; Sun, J.; Wang, C.; Zhu, G.; Bai, L.; Jiang, Z.; Zhang, W. Stripping voltammetric determination of cadmium and lead ions based on a bismuth oxide surface-decorated nanoporous bismuth electrode. *Electrochem. Commun.* **2022**, *136*, 107233. [CrossRef]
9. Meng, Z.H.; Zheng, Y.F.; Xiao, H.F. Distribution and Ecological Risk Assessment of Heavy Metal Elements in Soil. *Adv. Mater. Res.* **2011**, *183–185*, 82–87. [CrossRef]
10. Liu, R.; Xu, Y.; Zhang, J.; Wang, W.; Elwardany, R.M. Effects of heavy metal pollution on farmland soils and crops: A case study of the Xiaoqinling Gold Belt, China. *China Geol.* **2020**, *3*, 402–410. [CrossRef]
11. Zhang, Q.; Wen, H.; Peng, D.; Fu, Q.; Huang, X. Interesting interference evidences of electrochemical detection of Zn(II), Cd(II) and Pb(II) on three different morphologies of MnO_2 nanocrystals. *J. Electroanal. Chem.* **2015**, *739*, 89–96. [CrossRef]
12. Zhao, G.; Liu, G. Interference Effects of Cu(II) and Pb(II) on the Stripping Voltammetric Detection of Cd(II): Improvement in the Detection Precision and Interference Correction. *J. Electrochem. Soc.* **2018**, *165*, H488–H495. [CrossRef]
13. Jiang, T.; Guo, Z.; Ma, M.; Fang, L.; Yang, M.; Li, S.; Liu, J.; Zhao, N.; Huang, X.; Liu, W. Electrochemical laser induced breakdown spectroscopy for enhanced detection of Cd(II) without interference in rice on layer-by-layer assembly of graphene oxides. *Electrochim. Acta* **2016**, *216*, 188–195. [CrossRef]
14. Tibbetts, D.F.; Davis, J.; Compton, R.G. Sonoelectroanalytical detection of lead at a bare copper electrode. *Fresenius' J. Anal. Chem.* **2000**, *368*. [CrossRef] [PubMed]
15. Jiang, M.; Ma, M.; Lin, C.; Yang, M.; Fang, L.; Liu, J.; Zhao, N.; Huang, X. Uniform manganese-loaded titanium dioxide nanotube arrays for accurate detection of trace Cd^{2+} in water, soil and tea: Enhanced stability and sensitivity. *Chem. Eng. J.* **2020**, *400*, 125972. [CrossRef]
16. Xiao, L.; Wang, B.; Ji, L.; Wang, F.; Yuan, Q.; Hu, G.; Dong, A.; Gan, W. An efficient electrochemical sensor based on three-dimensionally interconnected mesoporous graphene framework for simultaneous determination of Cd(II) and Pb(II). *Electrochim. Acta* **2016**, *222*, 1371–1377. [CrossRef]
17. Hwang, G.; Han, W.; Park, J.; Kang, S. Determination of trace metals by anodic stripping voltammetry using a bismuth-modified carbon nanotube electrode. *Talanta* **2008**, *76*, 301–308. [CrossRef]
18. Liu, N.; Ye, W.; Liu, G.; Zhao, G. Improving the accuracy of stripping voltammetry detection of Cd^{2+} and Pb^{2+} in the presence of Cu^{2+} and Zn^{2+} by machine learning: Understanding and inhibiting the interactive interference among multiple heavy metals. *Anal. Chim. Acta* **2022**, *1213*, 339956. [CrossRef]
19. Liu, N. Sensitive Stripping Voltammetric Determination of Pb(II) in Soil Using a Bi/single-walled Carbon Nanotubes-Nafion/ionic Liquid Nanocomposite Modified Screen-Printed Electrode. *Int. J. Electrochem. Sci.* **2020**, *15*, 7868–7882. [CrossRef]
20. Malakootian, M.; Hamzeh, S.; Mahmoudi-Moghaddam, H. A new electrochemical sensor for simultaneous determination of Cd(II) and Pb(II) using $FeNi_3$/CuS/BiOCl: RSM optimization. *Microchem. J.* **2020**, *158*, 105194. [CrossRef]
21. Sawan, S.; Maalouf, R.; Errachid, A.; Jaffrezic-Renault, N. Metal and metal oxide nanoparticles in the voltammetric detection of heavy metals: A review. *TrAC Trends Anal. Chem.* **2020**, *131*, 116014. [CrossRef]
22. Lin, W.; Zhai, W.; Yan, Y.; Liu, Y. Highly sensitive Pb^{2+} sensor based on rod-like poly-tyrosine/Bi modified glassy carbon electrode combined with electrodeposition to eliminate Cu^{2+} interference. *Microchem. J.* **2021**, *160*, 105664. [CrossRef]
23. Zhou, J.; Pan, K.; Qu, G.; Ji, W.; Ning, P.; Tang, H.; Xie, R. rGO/MWCNTs-COOH 3D hybrid network as a high-performance electrochemical sensing platform of screen-printed carbon electrodes with an ultra-wide detection range of Cd(II) and Pb(II). *Chem. Eng. J.* **2022**, *449*, 137853. [CrossRef]
24. Zhao, G.; Wang, H.; Liu, G.; Wang, Z. Optimization of Stripping Voltammetric Sensor by a Back Propagation Artificial Neural Network for the Accurate Determination of Pb(II) in the Presence of Cd(II). *Sensors* **2016**, *16*, 1540. [CrossRef]
25. Liu, N.; Zhao, G.; Liu, G. Coupling Square Wave Anodic Stripping Voltammetry with Support Vector Regression to Detect the Concentration of Lead in Soil under the Interference of Copper Accurately. *Sensors* **2020**, *20*, 6792. [CrossRef] [PubMed]
26. Zhao, G.; Liu, G. A Portable Electrochemical System for the On-Site Detection of Heavy Metals in Farmland Soil Based on Electrochemical Sensors. *IEEE Sens. J.* **2018**, *18*, 5645–5655. [CrossRef]
27. Gich, M.; Fernández-Sánchez, C.; Cotet, L.C.; Niu, P.; Roig, A. Facile synthesis of porous bismuth–carbon nanocomposites for the sensitive detection of heavy metals. *J. Mater. Chem. A* **2013**, *1*, 11410. [CrossRef]
28. Xu, C.; Liu, J.; Bi, Y.; Ma, C.; Bai, J.; Hu, Z.; Zhou, M. Biomass derived worm-like nitrogen-doped-carbon framework for trace determination of toxic heavy metal lead(II). *Anal. Chim. Acta* **2020**, *1116*, 16–26. [CrossRef] [PubMed]

29. Liuzhu, Z.; Sekar, S.; Chen, J.; Lee, S.; Kim, D.Y.; Manikandan, R. A polyrutin/AgNPs coated GCE for simultaneous anodic stripping voltammetric determination of Pb(II) and Cd(II) ions in environmental samples. *Colloids Surf. A Physicochem. Eng. Asp.* **2022**, *648*, 129082. [CrossRef]
30. Li, X.; Fan, P.; Li, Z.; Chen, G.; Qiu, H.; Hou, G. Soil Classification Based on Deep Learning Algorithm and Visible Near-Infrared Spectroscopy. *J. Spectrosc.* **2021**, *2021*, 1508267. [CrossRef]
31. Ferreira, R.D.A.; Teixeira, G.; Peternelli, L.A. Kennard-stone method outperforms the random sampling in the selection of calibration samples in snps and nir data. *Ciência Rural* **2022**, *52*. [CrossRef]
32. He, Z.; Song, S.; Shen, K.; Zhang, X. Performance enhancement-based active learning sample selection method. *J. Chemometr.* **2022**, *36*, e3386. [CrossRef]
33. Wang, S.F.; Han, P.; Cui, G.L.; Wang, D.; Liu, S.S.; Zhao, Y. The nir detection research of soluble solid content in watermelon based on spxy algorithm. *Spectrosc. Spect. Anal.* **2019**, *39*, 738–742. [CrossRef]
34. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17. [CrossRef]
35. Herrero, A.; Ortiz, M.C. Modelling the background current with partial least squares regression and transference of the calibration models in the simultaneous determination of Tl and Pb by stripping voltammetry. *Talanta* **1998**, *46*, 129–138. [CrossRef]
36. Huang, Q.T.; Zhou, L.Q.; Shi, Z.; Li, Z.Y.; Gu, Q. Analysis of Lead in Soil with Partial Least Square Regression (PLS) Method and Field Portable X-ray Fluorescence (FPXRF) Analyzer. *Spectrosc. Spect. Anal.* **2009**, *29*, 1434–1438. [CrossRef]
37. Chen, F.; Chen, C.; Chen, C.; Yan, Z.; Gao, R.; Han, H.; Li, W.; Lv, X. Application of PLSR in rapid detection of glucose in sheep serum. *Optik* **2020**, *224*, 165734. [CrossRef]
38. Yazir, S.M.; Dhas, J.E.R.; Darwins, A.K.; Lewise, K.A.S.; Gupta, M.S. Modelling of weld residual stress parameters by SVR approach. *Mater. Today Proc.* **2022**, *64*, 338–344. [CrossRef]
39. Liang, H.; Zou, J.; Li, Z.; Khan, M.J.; Lu, Y. Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm. *Future Gener. Comput. Syst.* **2019**, *95*, 454–466. [CrossRef]
40. Han, J.; Kamber, M.; Pei, J. 2—Getting to Know Your Data. In *Data Mining*, 3rd ed.; Han, J., Kamber, M., Pei, J., Eds.; Morgan Kaufmann: Boston, MA, USA, 2012; pp. 39–82.
41. Metcalf, L.; Casey, W. Chapter 2—Metrics, Similarity, and Sets. In *Cybersecurity and Applied Mathematics*; Metcalf, L., Casey, W., Eds.; Syngress: Boston, MA, USA, 2016; pp. 3–22.
42. Ning, L.; Guo, Z.; Gang, L. Research advances of influencing factors and weakening methods to determine Pb^{2+} and Cd^{2+} in soils by anodic stripping voltammetry. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 232–243. [CrossRef]