



Cite this: *J. Anal. At. Spectrom.*, 2026, **41**, 416

## *In situ* multi-element soil analysis using laser-induced breakdown spectroscopy (LIBS)

Minghui Gu,<sup>ab</sup> Huansong Huang,<sup>ab</sup> Qingbin Jiao,<sup>\*,a</sup> Ding Ma,<sup>a</sup> Yuxing Xu,<sup>c</sup> Chao Liu,<sup>ab</sup> Jiguo Li,<sup>ab</sup> Xin Zhang,<sup>ab</sup> Mingyu Yang,<sup>a</sup> Liang Xu,<sup>a</sup> Sijia Jiang,<sup>a</sup> Hong Li,<sup>d</sup> Jiahui Qi,<sup>d</sup> Junbo Zang<sup>d</sup> and Xin Tan<sup>\*,a</sup>

Soil heavy metal contamination poses a serious threat to agricultural product safety and public health, which urgently calls for the development of rapid and accurate *in situ* detection techniques. LIBS enables simultaneous multi-element analysis and requires minimal sample preparation, and has been widely applied in the field of elemental analysis. However, under practical field conditions, moisture in soil significantly interferes with the stability and intensity of LIBS signals, thereby limiting its capability for large-area, *in situ*, and accurate detection in real-world environments. To address this issue, this study proposes a novel approach for simultaneous multi-element quantitative analysis by integrating neural networks with physical correction strategies. Adaptive Iteratively Reweighted Penalized Least Squares (airPLS) and Random Forest were employed to optimize spectral data and screen characteristic spectral fingerprints. An ablation factor model was established to correct spectral intensity under moisture interference, and a Multi-Task Convolutional Attention Network (MT-CAN) was constructed to predict both moisture content and multiple heavy metal concentrations. The results demonstrated that the root mean square error for moisture prediction reached 0.83%, and the relative errors for simultaneous quantification of Zn, Cr, Cu, and Pb were all below 8%. Finally, a transfer learning strategy based on model parameters was adopted to further enhance the cross-regional generalization capability of the model. This study provides an effective technical foundation for achieving *in situ* heavy metal detection in field soil environments.

Received 10th September 2025  
 Accepted 27th October 2025

DOI: 10.1039/d5ja00355e  
[rsc.li/jaas](http://rsc.li/jaas)

## 1 Introduction

Soil, as a fundamental natural resource essential for sustaining agricultural production, ecological balance, and human survival, has attracted extensive attention regarding its quality status. However, accelerated industrial and agricultural development, along with the continuous application of agrochemicals, has led to the introduction of heavy metals into soils through multiple pathways. Over the past half-century, the global environment has accumulated more than 3 million tons of chromium (Cr) and 80 million tons of lead (Pb),<sup>1</sup> with over 10% of agricultural soils exceeding safe thresholds for heavy metal concentrations.<sup>2</sup> These potentially toxic metallic pollutants exhibit persistent toxicity and a propensity for bioaccumulation,<sup>3</sup> posing severe threats to soil functionality and public health. There is an urgent need to develop rapid

detection methods which are capable of multi-element analysis, large-area coverage, and *in situ* measurement under field soil conditions.

Laser-induced breakdown spectroscopy (LIBS), recognized for its rapid analysis, capability for simultaneous multi-element detection, and minimal sample preparation requirements, is widely regarded as a promising tool for on-site elemental analysis.<sup>4</sup> It has been successfully applied in various fields including geological exploration.<sup>5–7</sup> For instance, Han *et al.* converted LIBS spectral intensities into RGB images and employed clustering algorithms to map the distribution of Cu, Cr, and Pb in contaminated soil with high spatial resolution, thereby achieving heavy metal imaging and spatial distribution analysis of the polluted areas.<sup>8</sup> Li *et al.* integrated graphite enhancement with a machine learning model (LWNet), which enabled the accurate quantification of Cd across mixed soil types and significantly improved the sensitivity and accuracy of LIBS for detecting trace toxic heavy metals.<sup>9</sup> In contrast to traditional soil detection methods that rely on complex wet chemical digestion procedures, such as AAS,<sup>10</sup> ICP-AES,<sup>11</sup> and ICP-MS,<sup>12</sup> LIBS substantially reduces the analytical cycle time and dependency on laboratory settings, rendering it more suitable for rapid analysis in field

<sup>a</sup>Chinese Academy of Sciences, Changchun Institute of Optics, Fine Mechanics and Physics, Changchun, 130033, China

<sup>b</sup>University of the Chinese Academy of Sciences, Beijing, 100049, China

<sup>c</sup>School of Physics and Electronic Information, Yantai University, Yantai, Shandong 264005, China

<sup>d</sup>Jilin Province Product Quality Supervision and Inspection Institute, Changchun, 130022, China



environments. However, moisture in natural soils (ranging from 2 to 25%)<sup>13</sup> severely interferes with the formation and evolution of the laser-induced plasma, markedly reducing signal stability and intensity,<sup>14,15</sup> which limits the technique's ability to achieve multi-element *in situ* detection in field soil environments. Consequently, heavy metal detection in soils still largely requires laboratory drying pretreatment to ensure analytical accuracy.

Notably, LIBS technology itself offers potential pathways to overcome moisture interference. Studies have indicated that LIBS spectral signals contain information closely related to the moisture state of samples, which can be utilized for the quantitative assessment of water content.<sup>16</sup> To date, several researchers have employed LIBS to investigate moisture content in samples and the plasma excitation characteristics of wet specimens. For instance, M. Chen *et al.* studied the influence of moisture content in coal powder on laser-induced plasma properties, revealing a nonlinear relationship between moisture variation and plasma electron density;<sup>17</sup> however, no effective signal correction method was proposed. Y. Liu *et al.* utilized laser-induced breakdown spectroscopy (LIBS) to analyze moisture content in cheese by normalizing the ratio of the oxygen signal to the CN signal,<sup>18</sup> yet the generalization capability of this single-ratio approach in complex real-world matrices remains limited. Chen *et al.* developed an artificial neural network (ANN) prediction model based on low-moisture coal samples and proposed a stochastic spectral attenuation method to mitigate moisture-induced perturbations,<sup>19</sup> but simultaneous multi-element detection was not achieved. Meanwhile, Wudil *et al.* achieved non-destructive analysis of soil moisture using support vector regression combined with adaptive boosting (SVR-ADB) for feature selection,<sup>20</sup> although they did not extend the method to quantify heavy metals under moisture influence.

Building upon this, the present study is designed to develop a method for simultaneous and accurate multi-element quantitative analysis that is suitable for field soils by integrating moisture content prediction, spectral correction strategies, and neural network-based advanced data processing techniques. Initially, a two-dimensional convolutional neural network with an attention mechanism will be constructed based on extracted spectral fingerprint features to predict soil moisture content. Subsequently, incorporating an analysis of the laser ablation physical mechanism, an ablation factor will be introduced to correct the intensity of spectral signals affected by moisture. On this basis, a multi-task convolutional attention network will be established to achieve synchronous and high-precision quantitative analysis of multiple target heavy metal elements (such as Zn, Cr, Cu, and Pb) in moist soil. Finally, to enhance the model's generalizability and practicality, transfer learning of model parameters will be introduced. Using a small amount of spectral data from the target region (or soil type), the pre-trained model will be fine-tuned to rapidly adapt to detection requirements in new environments. This research is expected to provide technical support for real-time and *in situ* monitoring of heavy metal contamination in agricultural soils (moisture content 0–25%).

## 2 Materials and methods

### 2.1 Sample preparation

The experiment utilized National Standard Soil for Composition Analysis (GBW07552, acquired from the China National Resource Platform for Certified Reference Materials, CNRM), the matrix of which was derived from typical farmland soil in Anyang City, Henan Province, China. After being oven-dried, five portions (each 5 g) of dried soil labeled 1 to 5 were weighed. A standard solution containing Zn (1000  $\mu\text{g mL}^{-1}$ ), Cr (1000  $\mu\text{g mL}^{-1}$ ), Cu (1000  $\mu\text{g mL}^{-1}$ ), and Pb (1000  $\mu\text{g mL}^{-1}$ ) (purchased from the Beijing General Research Institute of Nonferrous Metals, Beijing, China) was diluted with ultrapure water to form a working solution with concentrations of Zn 200  $\mu\text{g mL}^{-1}$ , Cr 150  $\mu\text{g mL}^{-1}$ , Cu 50  $\mu\text{g mL}^{-1}$ , and Pb 75  $\mu\text{g mL}^{-1}$ .<sup>21</sup> Standard samples were prepared using the standard addition method. Aliquots of 0, 1, 2, 3, and 4 mL of the mixed solution were added to the standard soil samples, respectively. After constant volume adjustment, stirring, and soaking, the mixtures were sealed and allowed to stand for 6 hours to ensure complete saturation of the soil. Subsequently, the samples were oven-dried at 105 °C until constant weight was achieved in accordance with the Chinese National Standard (HJ 613-2011). From each of the five samples, 2 g was taken, and the elemental concentrations of Zn, Cr, Cu, and Pb were analyzed using ICP-MS as reference values.

To prepare moist soil samples, five standard soil materials (4 g) were oven-dried and subdivided into 2.0 g aliquots. Each subsample was placed in a 50 mL beaker, and 2.0 g of ultrapure water was added. To alleviate moisture gradients within the soil particles, the samples were periodically stirred and sealed to maintain equilibrium. After 6 hours, the samples were transferred to a drying oven to undergo slow and more uniform moisture loss. By controlling the duration in the drying chamber, soil samples with final moisture contents of 0%, 9.09%, 13.79%, 18.03%, and 23.08% were obtained. This gradient was designed to cover the typical soil moisture range and, based on observations from actual farmland sampling (where soil moisture at 15 cm depth seldom falls below 10%) focuses on characterizing the higher moisture range. Prior to analysis, each soil sample was stirred and mixed again to further reduce heterogeneity, ensuring that the local moisture content at any random laser ablation site reasonably approximated the overall average moisture content determined gravimetrically. One set of samples was used for LIBS analysis, while parallel samples were tested according to HJ 613-2011 to determine the actual moisture content, which served as the benchmark truth value for the LIBS moisture prediction model. The concentrations of elements Zn, Cr, Cu, and Pb, along with the moisture content, are presented in Table 1.

### 2.2 LIBS setup and measurement

A schematic diagram of the mobile LIBS analysis instrument is shown in Fig. 1. The laser source was a Q-switched Nd:YAG laser (Quantel CFR200, France) which operated at a wavelength of 1064 nm with a pulse width of 7 ns and a single-pulse energy of



Table 1 Concentrations of Zn, Cr, Cu, and Pb and moisture content

Sample		1	2	3	4	5
Dry/ppm	Zn	65.8	102.7	140.6	178.5	214.2
	Cr	34.6	66.3	95.2	121.7	158.6
	Cu	15.1	28.1	32.8	45.1	54.7
	Pb	39.5	56.2	72.4	85.5	102.6
	MC	0.00	9.09	13.79	18.03	23.08
Wet/%	MC	0.00	9.09	13.79	18.03	23.08

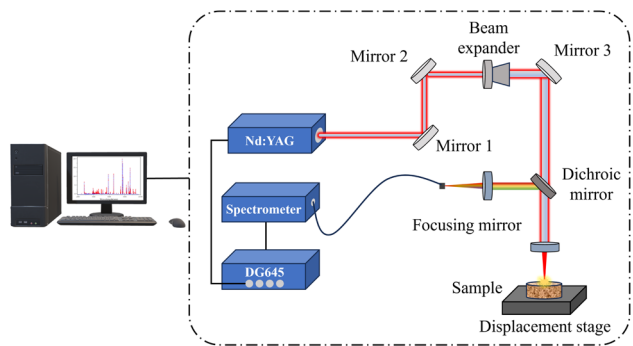


Fig. 1 Schematic diagram of the mobile LIBS analysis instrument.

up to 200 mJ. A beam expander was employed in the beam path to reduce the laser divergence angle and adjust the spot size, thereby ensuring a smaller and more stable focused spot on the sample surface while also protecting the focusing lens from damage. Subsequently, a dichroic mirror was used. It has high transmissivity at the laser wavelength (1064 nm), allowing the beam to reach the sample and generate plasma, while also having high reflectivity in the range of 200–990 nm, effectively transmitting the plasma emission light to the optical fiber probe. The transmitted beam was focused 1 mm below the sample surface using a lens with a focal length of 35 mm to avoid air breakdown. The spectrometer (LIBS2500plus, Ocean Optics, UK) is equipped with a 7-channel CCD array, covering a spectral range of 190–990 nm with a resolution of 0.1 nm (FWHM). It effectively captures characteristic emission lines of typical trace metal elements, including Zn at 213.8 nm, Cr at 427.4 nm, Cu at 327.4 nm, Pb at 406.3 nm, and Mg at 279.6 nm.

To obtain high-quality plasma signals, the optimized experimental parameters were set as follows: a single-pulse energy of 50 mJ, an integration time of 1.2 ms, a delay time of 1  $\mu$ s, and a repetition rate of 10 Hz. The samples were placed on an X–Y–Z motorized translation stage for precise positioning. To ensure that the spectral data used for modeling represented stable and representative signals, 100 sampling points were randomly selected from the surface of each soil sample, with five spectra collected per point. By calculating the average cosine similarity of each spectrum to the others,<sup>22</sup> the 400 spectral data points with the smallest differences were ultimately selected as representative signals, thereby minimizing the potential impact of local moisture heterogeneity and other transient fluctuations on model reliability. The total spectral acquisition time per sample was  $\leq 8$  minutes, ensuring that moisture content

variation remained within 1% (see Fig. S3). A total of 4000 LIBS spectra were collected, of which 2000 were used for moisture content analysis and 2000 for multi-element concentration prediction.

### 3 Data analysis

#### 3.1 Data preprocessing

The overall flow of the study is shown in Fig. 2. The raw LIBS signals contain significant background interference, which necessitates baseline correction. Traditional baseline correction methods fall into two major categories. The first category employs machine learning for automated or semi-automated spectral processing. Although this approach can achieve high precision, it requires hundreds of thousands of spectra for model training. The second category involves selecting specific points on the spectral line as anchor points for fitting a baseline curve. However, the selection of these points greatly influences the effectiveness of background correction, and noticeable troughs may appear on both sides of the peaks, adversely affecting subsequent quantitative analysis. This study adopts the adaptive iteratively reweighted penalized least squares (airPLS) method, which iteratively adjusts the weight of the sum of squared errors (SSEs) between the fitted baseline and the original signal. In the characteristic peak regions, the weight approaches zero, while in the baseline regions, the weight coefficients are updated based on the fitting error, thereby effectively separating the target signal from the background baseline.<sup>23</sup> This method requires no prior knowledge and offers strong portability, making it highly suitable for practical engineering applications.<sup>24</sup> Subsequently, discrete wavelet transform (DWT) denoising was applied to the baseline-corrected spectral data using the 'db5' wavelet basis,<sup>25</sup> and the results are illustrated in Fig. 3.

#### 3.2 Feature selection

Soil spectral data often suffer from noise, feature redundancy, and complex nonlinear interactions, which increase computational burden and reduce model generalizability. This study employed the Random Forest (RF) algorithm for dimensionality reduction and feature selection of LIBS data. Compared to

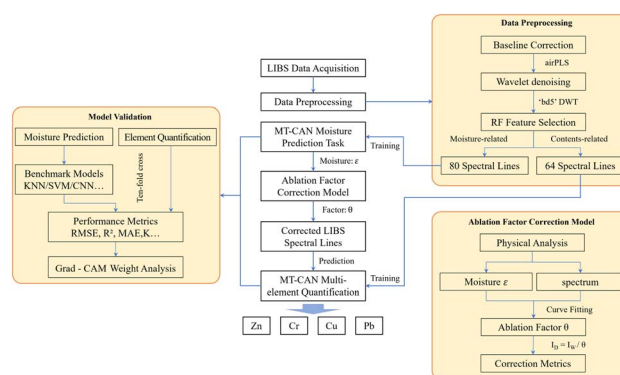


Fig. 2 Flow chart of multi-element prediction in moisture soil.



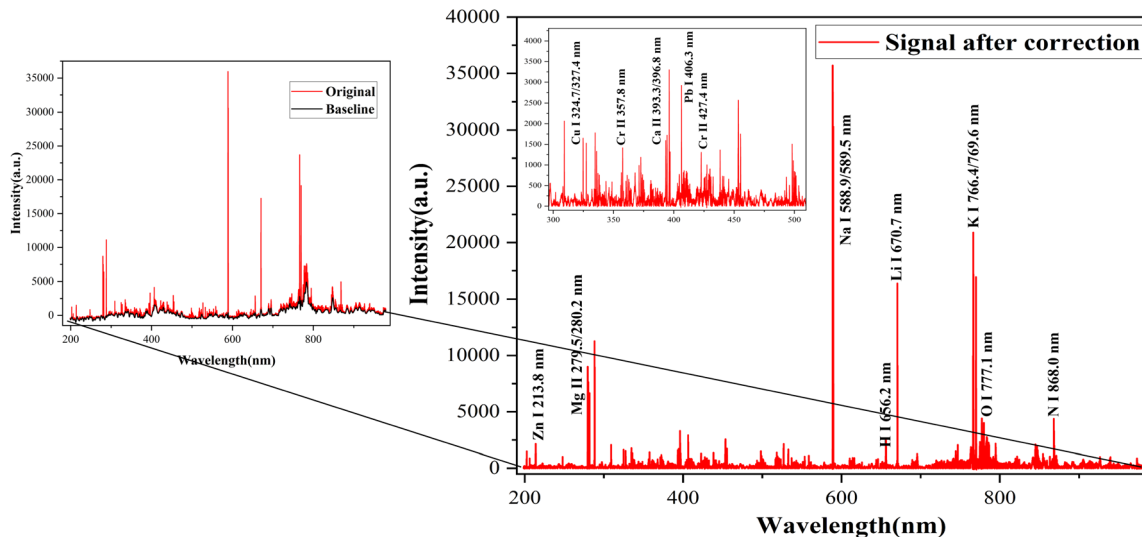


Fig. 3 Preprocessed LIBS signal.

unsupervised linear dimensionality reduction methods such as Principal Component Analysis (PCA), RF demonstrates superior performance in preserving original feature interpretability, handling nonlinear relationships, and improving predictive accuracy.<sup>26–28</sup> First, the spectral data were standardized to eliminate scale differences. By optimizing the tree number and maximum depth based on OOB error, a random forest model consisting of 100 trees with a maximum depth of 10 for each tree was constructed to balance model complexity and generalization ability. Furthermore, based on the relationship between the ‘inflection point’ on the feature number curve and cumulative importance, the spectral line with cumulative importance exceeding 99.5% was selected as the feature spectral fingerprint (shown in Data Preprocessing module in Fig. 2). This method combines the efficiency of filtering approaches with the predictive orientation of wrapper methods, significantly reducing dimensionality while retaining critical discriminative information. The selected LIBS spectral lines and their importance analysis are shown in Fig. 4.

Among the 80 spectral lines correlated with moisture content, high-weight features are concentrated in characteristic hydrogen and oxygen emission regions such as H $\alpha$  656.3 nm, O I 777.2/844.6 nm, and the OH molecular band at 308–320 nm. The remaining weights are mainly concentrated in the 200–500 nm band, where the emission lines of metallic elements are dense, due to the rapid vaporization and dissociation of water molecules in moist soil when irradiated with laser pulses. A part of the incident laser energy is consumed, thereby reducing the overall plasma temperature. This suppressed plasma state directly attenuates the emission intensity of certain metallic elements. Consequently, the intensity variations of these metallic spectral lines indirectly reflect plasma state modifications induced by moisture content. Furthermore, from the LIBS data of dried soils with varying concentrations, 64 core spectral lines related to elemental quantification were densely distributed in the 200–500 nm region, which contains characteristic

metal emission peaks including Zn I 213.8 nm, Cr I 427.4 nm, Cu I 324.7 nm, and Pb I 406.3 nm, showing high consistency with the characteristic excitation bands of the target elements.

### 3.3 Multi-task convolutional attention network

Previous studies have confirmed that machine learning can effectively correct chemical matrix effects in LIBS analysis.<sup>29–31</sup> Convolutional neural networks (CNN), as an efficient deep learning method, have demonstrated significant potential in spectral data analysis. However the one-dimensional nature of LIBS data restricts convolutional kernels to capturing only local features between adjacent wavelengths. Thus, this study first transformed the one-dimensional spectral features of each

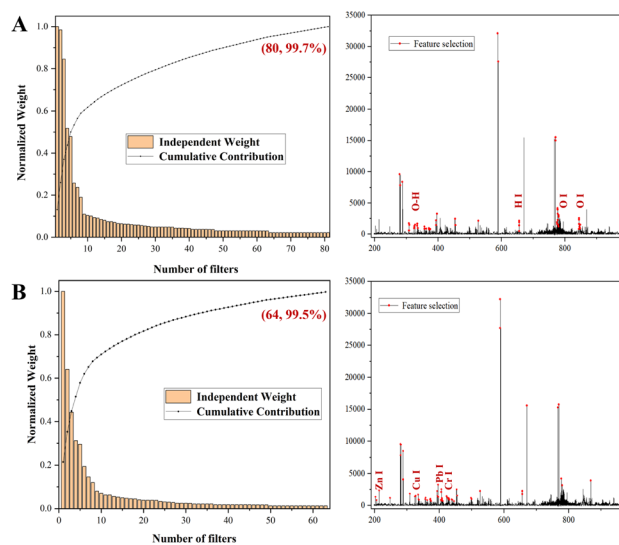


Fig. 4 RF feature selection weights and spectral line distribution: (A) pertaining to moisture content; (B) pertaining to element concentrations.



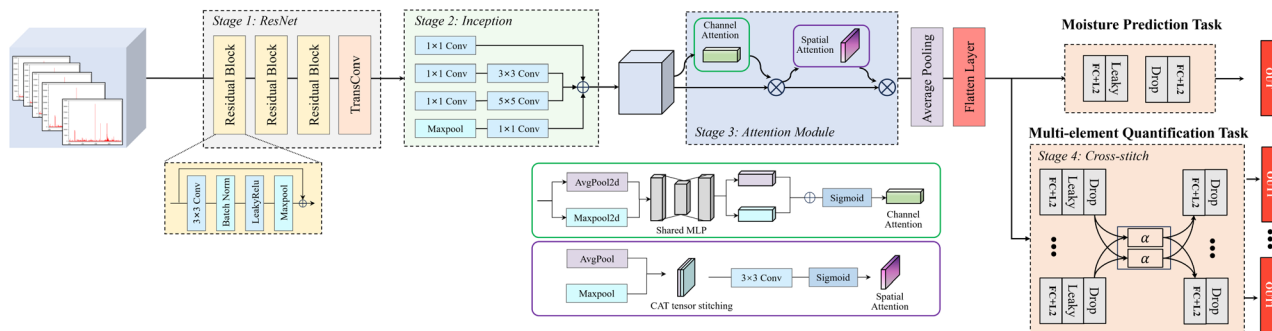


Fig. 5 MT-CAN flow diagrams.

sample into a square matrix format with missing elements padded with zeros, thereby increasing the dimensionality of feature interactions while preserving the original sequence order.

Based on this, we propose a Multi-Task Convolutional Attention Network (MT-CAN, Fig. 5), which integrates two key tasks: moisture content prediction and multi-element concentration regression, with the aim of achieving rapid and accurate *in situ* detection of various heavy metal elements in humid soils. The core structure of MT-CAN consists of multi-level residual feature extraction, parallelized multi-scale Inception modules, dual attention feature calibration mechanisms, cross-stitch multitask interaction units, and a deep regression output module. First, the network employs a three-level residual module to construct the base feature extractor. This structure, through multi-layer convolution and non-linear hierarchical stacking of the original spectrum, preliminarily extracts spectral features related to elemental content or moisture, while effectively alleviating the gradient vanishing problem in deeper networks, thus providing robust low-level feature representations for subsequent multi-scale analysis. Subsequently, an improved parallelized Inception module is introduced for multi-scale feature reorganization. This module includes four independent paths, each employing  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$  convolution kernels along with max-pooling operations, further capturing subtle peak variations and global spectral trends within the spectral data from different samples. Residual units are embedded at the terminus of each path which enhances the stability of model training and convergence efficiency. To further strengthen the representation capability of critical features, a dual-attention feature calibration mechanism is employed in this work that adaptively re-weights feature responses across both channel and spatial dimensions. The channel attention submodule integrates average-pooling and max-pooling information and learns the significance weights of each channel *via* a shared multilayer perceptron whereas the spatial attention submodule generates spatial weight maps based on double-polarity pooled features through a standard convolutional layer thereby enhancing the focus on key wavelength regions and suppressing redundant responses. The collaborative weighting operation of both submodules effectively improves the model's selectivity toward useful spectral signals and its anti-interference capability. In response to the

inherent correlations among chemical substrates in multi-element regression prediction tasks, a cross-stitch multi-task interaction mechanism<sup>32</sup> is adopted which connects every two elemental prediction tasks *via* a  $4 \times 4$  learnable weight matrix thereby enabling adaptive feature sharing across tasks and allowing the network to autonomously learn synergistic and constrained relationships among different elements.

To mitigate the risk of overfitting in the complex model, we adopted a hierarchical regularization strategy: L2 regularization (weight decay coefficient  $\lambda = 0.005$ ) was introduced in the fully connected layers, batch normalization layers were embedded after each convolutional and fully connected layer, and a progressive dropout mechanism is employed (task layer dropout rate = 0.3; output layer dropout rate = 0.2). The loss function was formulated in a multitask weighted form (eqn (1)), considering both the regression errors of each task and the regularization constraint of the cross-stitch matrix. The optimal task weight combination [0.3, 0.2, 0.3, 0.2] was determined through grid search on the validation set, in order to balance the learning process of the different subtasks.

$$\mathcal{L} = \sum_{i=1}^4 w_i \|y_i - \hat{y}_i\|_2 + \lambda \|A - I\|_F \quad (1)$$

## 4 Results and discussion

### 4.1 Prediction of moisture content

The 80 characteristic spectral lines related to moisture content selected by RF were used as input to the moisture prediction module of the MT-CAN network, with the gravimetrically determined actual moisture content serving as the ground truth for training. To comprehensively evaluate model performance, key metrics including the Root Mean Square Error (RMSE), slope ( $K$ ), and Mean Absolute Error (MAE) were calculated using ten-fold cross-validation, as summarized in Fig. 6 presents the scatter plot of predicted *versus* true moisture content values during the validation phase, along with the distribution of absolute errors.

Furthermore, comparative experiments were conducted with other commonly used methods, including KNN, SVM and 1DCNN. Table 2 summarizes the performance metrics of all comparative models. The MT-CAN model consistently outperformed all other conventional models across every



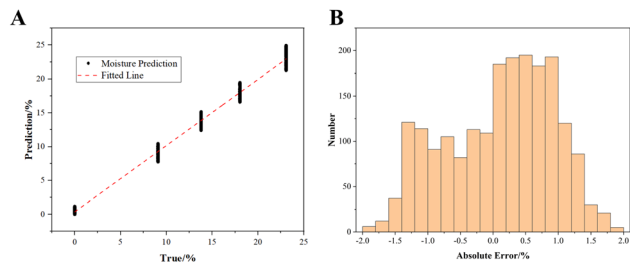


Fig. 6 (A) Scatter plot of predicted and true values; (B) the distribution of absolute errors.

performance metric. With a  $K$  of 0.978 and an RMSE of 0.831, it demonstrates high predictive accuracy. Moreover, the model maintained robust predictive capability even in extreme moisture content ranges (<2% or >20%), with the maximum absolute error not exceeding 2%, thereby meeting practical detection requirements.

To enhance the interpretability of the MT-CAN model, full-spectrum data were input into the model, and Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to generate feature importance heatmaps.<sup>33</sup> This technique computes gradient weights of the feature maps from the final convolutional layer and visually highlights key spectral regions critical to the model's decision-making process. As shown in Fig. 7, the highest activation intensities consistently correspond to the characteristic emission regions of hydrogen (H $\alpha$  656.3 nm) and oxygen (OI 777.2/844.6 nm). This correlation occurs because water molecules (H<sub>2</sub>O) dissociate in the high-temperature plasma, and the resulting hydrogen (H) and oxygen (O) atoms become excited by the plasma, enhancing these characteristic spectral lines. Meanwhile, other significant weights are concentrated in the 200–500 nm region, which is rich in emission lines from metallic elements. This is attributed to the substantial consumption of laser energy in vaporizing and dissociating water, leading to a temperature drop in the plasma core, thereby attenuating their spectral line intensities. The visualization results confirm that MT-CAN successfully captures the interaction between moisture and soil matrix components, which is consistent with the features selected by RF. Therefore, employing RF feature pre-selection can improve the training efficiency of the network.

#### 4.2 Correction model of the ablation factor

Variations in humidity can lead to divergent spectral intensities being collected from samples with similar elemental concentrations, thereby preventing the establishment of a reliable linear regression model.<sup>34</sup> Consequently, it is essential to

Table 2 Performance of different models

	MT-CAN	KNN	SVM	1D-CNN
RMSE/%	0.83	1.07	1.21	1.98
$K$	0.98	0.96	0.95	0.89
MAE/%	0.71	0.93	1.01	1.73

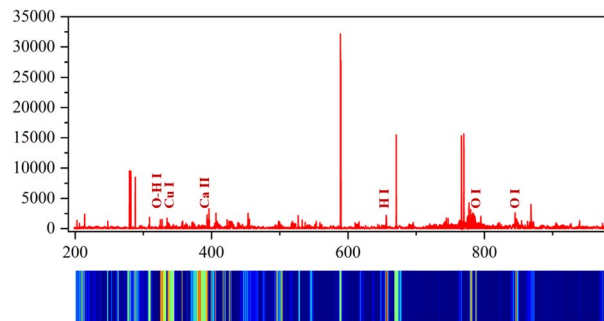


Fig. 7 The feature visualization based on Grad-CAM; different colors denote different contribution values.

perform intensity correction on *in situ* field LIBS data obtained from soils. In 2022, Xu *et al.* proposed a correlation model between the ablation factor and moisture content.<sup>35</sup> When a laser irradiates a moist sample, the ideal ablation mass of the soil is the total ablated mass minus the mass of water. However, in practice, some samples are not effectively ablated by the laser due to the absorption of energy by water and the rapid expansion after vaporization. In this study, this effect is quantified using a reduction factor  $K$  to represent the influence of moisture on the actual ablated mass. Thus, the ablation factor  $\theta$  for moist samples can be calculated using eqn (2).

$$\theta = \frac{m_a}{m_d} = \frac{m_w}{m_d} \cdot \frac{m_a}{m_w} = \frac{m_w}{m_d} \cdot \frac{1}{\frac{m_a + m_s}{m_a}} = \frac{1 - \varepsilon}{1 + K(\varepsilon)} \quad (2)$$

Here,  $m_a$  denotes the actual ablated mass of the moist soil sample,  $m_d$  is the ablated mass of the dry soil sample under identical conditions,  $m_w$  is the ideal ablated mass of the moist soil sample,  $m_s$  is the mass left unablated due to the presence of moisture,  $\varepsilon$  is the moisture content of the sample, and  $K(\varepsilon)$  is the moisture-dependent reduction factor. Furthermore, according to the Lomakin–Scheibe formula,<sup>36</sup> which states that the intensity of an elemental spectral line exhibits a linear relationship with the concentration of the analyte, the following expression can be derived:

$$I_d = \frac{I_w}{\theta} = \frac{I_w K(\varepsilon)}{(1 - \varepsilon)} \quad (3)$$

$I_w$  denotes the measured spectral line intensity of an element in the moist sample, while  $\bar{I}_d$  represents the characteristic spectral line intensity acquired by exciting a dry sample of identical mass under laboratory conditions (shown in the ablation factor correction model in Fig. 2). This study established an ablation factor correction model using LIBS data obtained from the five moist soil samples listed in Table 1. To ensure the representativeness of the model, characteristic spectral bands of elements including carbon, hydrogen, oxygen, nitrogen, aluminum, calcium, magnesium, sodium, and potassium, as well as 64 spectral bands selected by RF associated with the concentrations of Zn, Cr, Cu, and Pb, were employed as corrected data.

The reduction factor for each spectral band across samples with varying moisture contents was calculated according to eqn (3) and fitted exponentially, as shown in Fig. 8 (with fitting



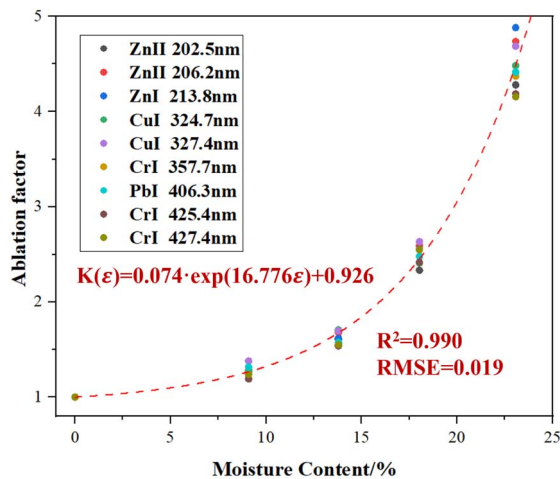


Fig. 8 Fitting result of the reduction factor.

results for other bands provided in Fig. S4). A correlation coefficient of 0.99 was achieved, indicating a strong correlation between the ablation factor and moisture content. The empirically derived fitting function was incorporated into the calculation of the ablation factor in eqn (3), which was then used to correct the peak intensities of characteristic spectral lines for samples with different moisture levels. In Fig. 10, the uncorrected intensities are shown in black, and the corrected intensities are shown in red. It can be observed that the corrected spectral intensities increased significantly across all bands and converged toward those obtained under dry conditions. To quantify the overall correction effect, cosine similarity was employed to evaluate the spectral similarity between the uncorrected and corrected spectra and the dry spectra. (calculated only within the corrected spectral bands), as shown in Table 3. The results demonstrate that the spectra corrected using the ablation factor exhibit substantially higher consistency with those acquired under dry conditions.

### 4.3 Multi-element quantitative analysis

The 64 LIBS spectral lines associated with elemental concentrations, selected by RF, were first used as inputs for the MT-CAN. The elemental concentrations of Zn, Cr, Cu, and Pb, as determined by ICP-MS, served as ground truth values. The dataset was divided into training and validation sets in an 8 : 2 ratio to thoroughly train the model. To simulate realistic field conditions involving moist soil, five dried soil samples with different concentration levels (Table 1) were prepared by adding ultrapure water following a standardized procedure, resulting in moist samples with moisture contents of 0%, 8.7%, 12.8%,

Table 3 Cosine similarity between uncorrected and corrected spectra and drying spectra

MC/%	0.00	9.09	13.79	18.03	23.08
uncorrected	1.00	0.98	0.94	0.85	0.68
Corrected	1.00	0.99	0.98	0.99	0.99

16.6%, and 24.1%. For each sample, 120 LIBS spectra were collected, and the same set of 64 characteristic spectral lines was extracted. These spectral data were intensity-corrected using the model described in Section 4.2, and the corrected data were then input into the pre-trained MT-CAN network for elemental concentration prediction. The prediction results for Zn, Cr, Cu, and Pb contents on the independent validation set are presented in Fig. 9 and Table 4.

To evaluate the contribution of each key component in the MT-CAN architecture and explore potential simplifications, we systematically designed seven ablation models: Model A (without residual connections), Model B (without Inception-style multi-branch blocks), Model C (without attention mechanisms), Model D (without CrossStitch units), Model E (retaining only residual and attention modules), Model F (retaining only attention and CrossStitch modules), and Model G (retaining only residual and CrossStitch modules). As summarized in Table S1 (see SI), the performance of the complete model was compared with that of these ablated versions. The results demonstrate that (1) removing the dual attention mechanism caused the most significant performance degradation (average RMSE increased by 62.9%), confirming its crucial role in calibrating feature responses and suppressing noise. (2) Disabling the CrossStitch units, thereby isolating the tasks, led to substantial performance decline (average RMSE increased by 40.4%), highlighting the importance of leveraging intrinsic correlations among elemental concentrations for synergistic prediction. (3) Replacing the multi-scale Inception modules with standard convolutional layers resulted in notable deterioration (average RMSE increased by 39.4%), validating the necessity of capturing spectral features at different scales. (4) Moreover, eliminating residual connections not only reduced prediction accuracy (average RMSE increased by 25.1%) but also led to training instability and slower convergence.

The results demonstrate that the integrated LIBS analytical framework, which combines a dynamic moisture content correction model with multi-task learning, achieves relative errors of less than 8% for the quantitative analysis of all four

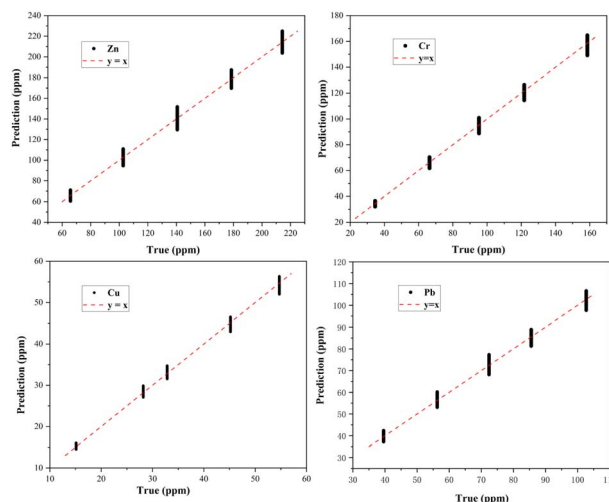


Fig. 9 Scatter plot of multi-element quantitative prediction results.



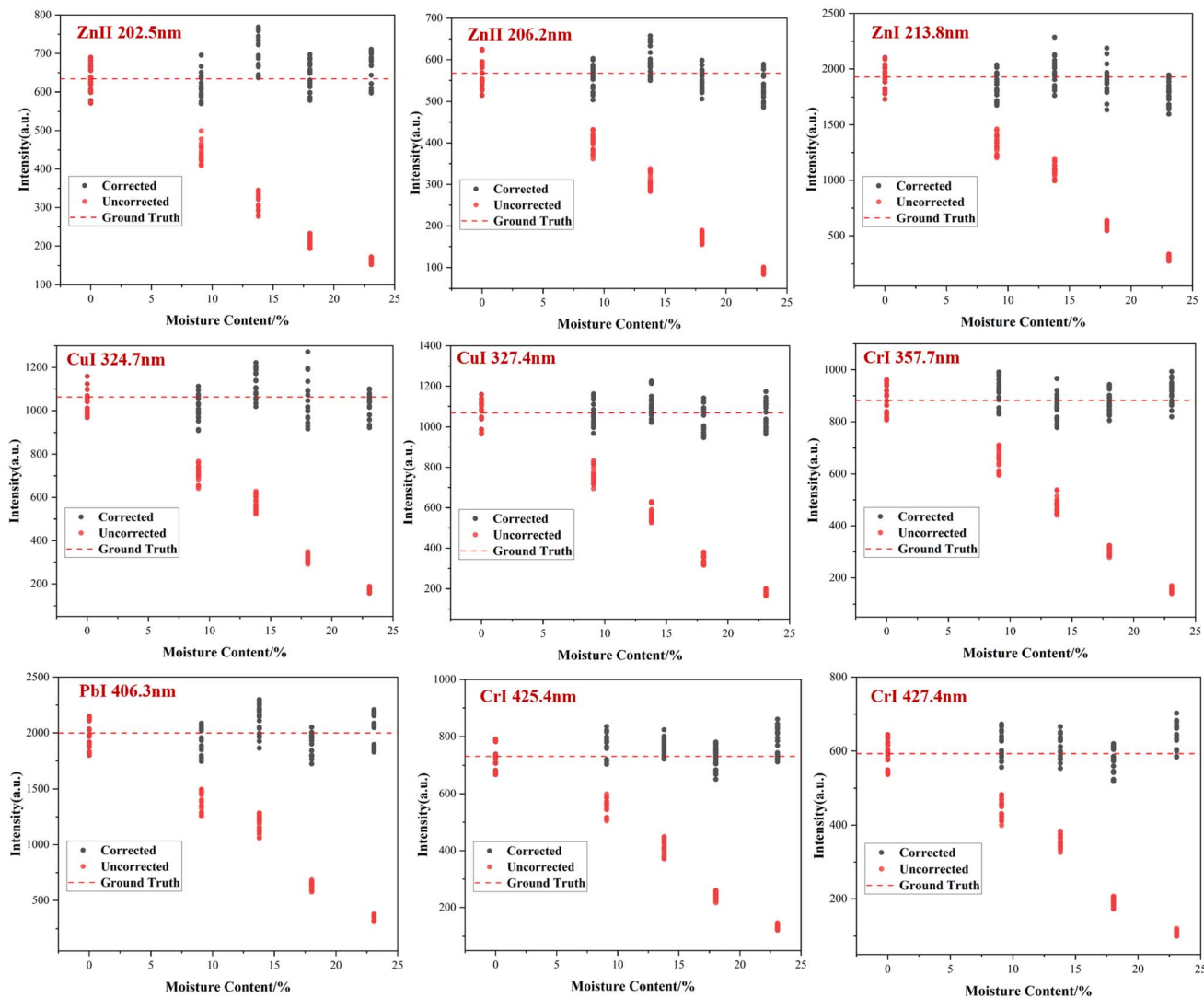


Fig. 10 Correction effect of typical spectral bands of Zn, Cu, Pb and Cr.

elements, fulfilling the detection requirements in moist soil environments and providing effective technical support for the *in situ* monitoring of heavy metal contamination in soils.

Finally, to facilitate practical application, this study developed a standalone software tool based on Python 3.8 and the PyQt5 module, which integrates and encapsulates the entire workflow described above—including spectral preprocessing, moisture content analysis, the ablation factor correction model,

multi-element quantitative prediction, and result analysis. A representative example of the software's user interface is provided in the Fig. S5.

#### 4.4 Transfer learning

In practical application scenarios, field soil samples often exhibit regional variations (*e.g.*, mineral composition, organic matter content, and physicochemical properties), which can degrade prediction performance of models trained on data from a single region when applied to new target regions. To enhance the cross-regional generalization capability of the MT-CAN model, a model parameter transfer learning strategy was introduced.<sup>37</sup> This approach leverages the universal spectral features learned from the source domain (typical farmland soil from Anyang City, Henan Province) and fine-tunes the model with a limited number of samples from the target domain (farmland soil from Changchun, Jilin Province, prepared with identical moisture content gradients; 200 LIBS spectra collected

Table 4 Performance of the multi-element quantitative prediction model

MC = 16.6%		Zn	Cr	Cu	Pb
Corrected	RE <sub>MAX</sub>	7.89%	6.97%	5.13%	6.27%
	RMSE	5.27	3.43	1.04	2.34
	MAE	4.41	2.75	0.84	1.97
Uncorrected	RE <sub>MAX</sub>	49.29%	44.20%	37.53%	49.10%
	RMSE	36.16	24.92	8.09	18.81
	MAE	29.35	19.65	7.16	15.56



Table 5 Performance of multi-element quantitative prediction model

		Zn	Cr	Cu	Pb
Strategy A	RE	10.31%	9.41%	8.98%	12.89%
	RMSE	8.24	6.25	2.11	3.65
	MAE	7.03	5.47	1.83	3.25
Strategy B	RE	15.62%	14.63%	15.33%	13.51%
	RMSE	11.96	11.08	3.72	3.98
	MAE	10.65	9.52	3.19	3.42
Strategy C	RE	4.25%	3.32%	3.49%	5.04%
	RMSE	3.31	1.79	1.07	1.33
	MAE	2.90	1.51	0.90	1.10

in total), enabling rapid adaptive optimization of model parameters.

The specific implementation of the transfer learning framework includes the following steps: (1) the MT-CAN model fully trained on the source domain was used as the initialized network; (2) the front-end shared feature extraction layers (multi-level residual layer and Inception layer) were frozen, which preserved the knowledge of common spectral features learned from the source model; (3) only the task-specific layers (dual-attention layer, cross-stitch interaction layer, and regression output layer) were unfrozen and fine-tuned. The fine-tuning process utilized target domain data with a lower initial learning rate ( $5 \times 10^{-4}$ ), employed the loss function with weight calibration, and applied early stopping to prevent overfitting. This approach effectively reused source domain knowledge while adapting to the target domain characteristics through targeted adjustments of task-specific parameters.

To validate the effectiveness of transfer learning, three modeling strategies were compared: Strategy A (training from scratch using only 200 target domain samples), Strategy B (direct application of the source domain pre-trained model without fine-tuning), and Strategy C (the transfer learning strategy proposed in this study). As shown in Table 5, Strategy C demonstrated superior predictive performance. This confirms that transfer learning significantly reduces the reliance on target domain data volume (requiring only approximately 15% of the source domain data size) by reusing shared spectral features and incorporating targeted adjustments with target domain samples, thereby offering an effective technical solution for LIBS-based monitoring in complex and variable agricultural soils.

## Conclusions

This study addresses the technical challenge of real-time monitoring of heavy metal elements in moist soils by proposing a simultaneous multi-element quantitative analysis method based on LIBS. By integrating deep learning with physical correction strategies, it effectively overcomes moisture interference in LIBS signals and achieves high-precision analysis of multiple heavy metal elements such as Zn, Cr, Cu, and Pb.

The spectral data were optimized and characteristic spectral fingerprints were screened using airPLS combined with random forest; a multi-task convolutional attention network (MT-CAN) was constructed to achieve simultaneous prediction of soil

moisture content and multi-element concentrations. For moisture prediction, the MT-CAN model demonstrated superior performance ( $k = 0.98$ ,  $RMSE = 0.83\%$ , and  $MAE = 0.71\%$ ). Even under extreme moisture conditions, the prediction error remained within 2%, indicating significantly enhanced stability compared to other benchmark models. Furthermore, the introduction of an ablation factor model for intensity correction of moist sample spectra resulted in corrected spectra that exhibited high consistency with the reference dry spectra. All evaluation metrics ( $RE_{MAX}$ ,  $RMSE$ , and  $MAE$ ) for multi-element quantitative analysis were reduced to approximately one-seventh of their pre-correction levels, with relative errors consistently below 8%, demonstrating the method's strong capability to correct moisture interference.

Finally, by employing a transfer learning strategy that fine-tunes a source domain pre-trained model with a limited number of target domain samples, the generalization capability of the model across different regional soils was significantly enhanced. Compared with other strategies, the transfer learning fine-tuned model achieved the lowest mean absolute error in predicting target elements, validating the effectiveness and practicality of the proposed method for cross-regional applications.

Although the transfer learning strategy and the proposed MT-CAN model in this study have demonstrated good adaptability between two distinct soil types (the yellow-brown soil from Henan and the black soil/chernozem from Jilin), it must be acknowledged that the robustness of the proposed method when applied to other soils (such as sandy/clayey textures or markedly different mineral compositions) still requires further systematic evaluation. Future work will focus on constructing a large-scale spectral library encompassing a wider variety of soil types and exploring the incorporation of soil physico-chemical properties as auxiliary input variables to develop more universal LIBS *in situ* detection solutions.

## Author contributions

Minghui Gu: writing – original draft, writing – review & editing, formal analysis, software, investigation, methodology; Huan-song Huang: investigation, visualization; Qingbin Jiao: supervision, writing – review & editing; Ding Ma: funding acquisition; Yuxing Xu: resources; Chao Liu: validation; Jiguo Li: investigation; Xin Zhang: formal analysis; Mingyu Yang: data curation; Liang Xu: funding acquisition; Sijia Jiang: methodology; Hong Li: investigation, validation; Jiahui Qi: investigation, validation; Junbo Zang: investigation, validation; Xin Tan: project administration, writing – review & editing, funding acquisition.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The partial dataset (including 100 sets of LIBS spectra under varying moisture and elemental conditions) and the custom



code (airPLS preprocessing, quantitative analysis, and ablation factor correction models) generated during this study are available in the [Github] repository at [<https://github.com/gmh472/LIBS>]. The complete dataset will be made publicly available upon completion of the project and relevant academic qualification.

Supplementary information: sample preparation methods, soil moisture content trends, ablation factor correction model, software interface description, and ablation test results for the MT-CAN model. See DOI: <https://doi.org/10.1039/d5ja00355e>.

## Acknowledgements

This study was supported by the Jilin Province Science and Technology Development Plan Project (20240402029 GH, 20240302001GX, YDZJ202401310ZYTS, and 20250203092SF); the Tianjin Science and Technology Plan Project (23YFYSHZ00300); the National Natural Science Foundation of China (NSFC) (42377037); the Changchun Science and Technology Development Plan Project (24GXYSZZ34); and the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDB1290202).

## References

- Q. Yang, Z. Li, X. Lu, Q. Duan, L. Huang and J. Bi, *Sci. Total Environ.*, 2018, **642**, 690–700.
- D. Hou, X. Jia, L. Wang, S. P. McGrath, Y.-G. Zhu, Q. Hu, F.-J. Zhao, M. S. Bank, D. O'Connor and J. Nriagu, *Science*, 2025, **388**, 316–321.
- B. Hu, S. Shao, T. Fu, Z. Fu, Y. Zhou, Y. Li, L. Qi, S. Chen and Z. Shi, *J. Geochem. Explor.*, 2020, **210**, 106443.
- D. S. Ferreira, D. V. Babos, M. H. Lima-Filho, H. F. Castello, A. C. Olivieri, F. M. V. Pereira and E. R. Pereira-Filho, *J. Anal. At. Spectrom.*, 2024, **39**, 2949–2973.
- F. F. Fontana, B. van der Hoek, S. Tassios, C. Tiddy, J. Stromberg, N. Francis, Y. A. Uvarova and D. G. Lancaster, *J. Geochem. Explor.*, 2023, **246**, 107160.
- J. Yan, J. Ma, K. Liu, Y. Li and K. Li, *J. Anal. At. Spectrom.*, 2025, **40**(6), 1447–1468.
- N. Li, J. Guo, J. Song, W. Ye, Y. Lu, Y. Tian and R. Zheng, *J. Anal. At. Spectrom.*, 2021, **36**, 2660–2668.
- B. Han, W. Gao, J. Feng, A. Iroshan, J. Yang, G. Chen, Y. Zhang, N. Aizezi and Y. Liu, *J. Hazard. Mater.*, 2025, 139284.
- X. Li, R. Chen, F. Liu, Z. You, J. Huang, J. Peng and G. Li, *Comput. Electron. Agric.*, 2025, **229**, 109831.
- R. S. Malikula, C. C. Kaonga, H. W. Mapoma, F. G. Thulu and P. Chiipa, *Water*, 2022, **14**, 121.
- W. Guanghui, L. Heng and F. Liping, *Phys. Test. Chem. Anal., Part B*, 2025, **61**, 331–337.
- I. Guagliardi, N. Ricca and D. Cicchella, *Toxics*, 2025, **13**, 314.
- H. Chen, L. Li, M. Awais, M. I. Abdulraheem, W. Zhang, V. Raghavan and J. Hu, *J. Soil Sci. Plant Nutr.*, 2024, **24**, 8137–8150.
- J. Peng, Y. He, L. Ye, T. Shen, F. Liu, W. Kong, X. Liu and Y. Zhao, *Anal. Chem.*, 2017, **89**, 7593–7600.
- Y. Wang, J. Li, G. Xue, K. Pan, Y. Fan, Y. Xue, S. Zhong, C. Zhang and M. Liu, *Talanta*, 2024, **275**, 126086.
- O. Al-Najjar, Y. Wudil, U. Ahmad, O. Al-Amoudi, M. Al-Osta and M. Gondal, *Appl. Spectrosc. Rev.*, 2023, **58**, 1–37.
- M. Chen, T. Yuan, Z. Hou, Z. Wang and Y. Wang, *Spectrochim. Acta, Part B*, 2015, **112**, 23–33.
- Y. Liu, M. Baudelet and M. Richardson, *J. Lab.*, 2012, **5–6**, 16–17.
- J. Chen, Q. Li, K. Liu, X. Li, B. Lu and G. Li, *J. Anal. At. Spectrom.*, 2022, **37**, 1658–1664.
- Y. Wudil, M. A. Al-Osta, M. Gondal and S. Kunwar, *Arabian J. Sci. Eng.*, 2024, **49**, 10021–10034.
- W. Lu, H. Luo, L. He, W. Duan, Y. Tao, X. Wang and S. Li, *Comput. Electron. Agric.*, 2022, **197**, 106923.
- M. Gu, C. Liu, H. Huang, X. Zhang, J. Li, Q. Jiao, L. Xu, M. Yang and X. Tan, *Results Chem.*, 2025, 102446.
- Z.-M. Zhang, S. Chen and Y.-Z. Liang, *Analyst*, 2010, **135**, 1138–1146.
- Y. Bai, H. Luo, Z. Li, *et al.*, *Acta Opt. Sin.*, 2024, **44**, 0730001.
- N. Zhao, Z. Luo, J. Li, Q. Ma, L. Guo, C. Shan and Q. Zhang, *At. Spectrosc.*, 2023, **44**, 219–226.
- H. Sun, C. Song, X. Lin and X. Gao, *Spectrochim. Acta, Part B*, 2022, **194**, 106456.
- P. Wang, N. Li, C. Yan, Y. Feng, Y. Ding, T. Zhang and H. Li, *Anal. Methods*, 2019, **11**, 3419–3428.
- Y. Chu, T. Chen, F. Chen, Y. Tang, S. Tang, H. Jin, L. Guo, Y. feng Lu and X. Zeng, *J. Anal. At. Spectrom.*, 2018, **33**, 2083–2088.
- T. F. Boucher, M. V. Ozanne, M. L. Carmosino, M. D. Dyar, S. Mahadevan, E. A. Breves, K. H. Lepore and S. M. Clegg, *Spectrochim. Acta, Part B*, 2015, **107**, 1–10.
- C. Sun, Y. Tian, L. Gao, Y. Niu, T. Zhang, H. Li, Y. Zhang, Z. Yue, N. Delepine-Gilon and J. Yu, *Sci. Rep.*, 2019, **9**, 11363.
- T. Chen, T. Zhang and H. Li, *TrAC, Trends Anal. Chem.*, 2020, **133**, 116113.
- I. Misra, A. Shrivastava, A. Gupta and M. Hebert, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3994–4003.
- R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, 618–626.
- P. Paris, K. Piip, A. Lepp, A. Lisovski, M. Aints and M. Laan, *Spectrochim. Acta, Part B*, 2015, **107**, 61–66.
- Y. Xu, B. Han, X. Tan, Q. Jiao, Z. Ma, B. Lv, Y. Li, H. Li, Y. Zou and L. Yang, *Eur. J. Soil Sci.*, 2022, **73**, e13213.
- R. Hark, R. Harmon, S. Musazzi and P. Perini, *Laser-Induced Breakdown Spectroscopy - Theory and Applications*, ed. R. R. Hark, R. S. Harmon, S. Musazzi and P. Perini, Springer, Berlin Heidelberg, 2014, vol. 182.
- J. Chen, W. Yan, L. Kang, B. Lu, K. Liu and X. Li, *Anal. Methods*, 2023, **15**, 5157–5165.

