IRAQI JOURNAL OF APPLIED PHYSICS
Vol. 21, No. 2, April-June 2025, pp. 263-269

Essam M. Rasheed [1]
Zainab K. Nasser Allah [1]
Suhaib Q. Abdullah [1]
Ali A. Al Maliki [2]
Kareem K. Mohammad [3]

[1] Department of Physics,
College of Science,
University of Al-Nahrain,
Baghdad, IRAQ
[2] Scientific Research Commission,
Research and Technology
of Environment,
Water and Renewable
Energy Center,
Baghdad, IRAQ
[3] Al-Nahrain Renewable
Energy Research Center,
Al-Nahrain University,
Jadiriyah, Baghdad, IRAQ

Spectroscopy and Regression Methods to Predict of Heavy Metals Contamination in Soil: Case Study in Baghdad

Heavy metal contamination in soil poses a significant environmental challenge. Spectroscopy provides a cost-effective, rapid, and reliable method for predicting soil properties, including color, texture, and temperature. This study assesses heavy metal contamination in Baghdad's soil using spectral data analysis. Forty soil samples were analyzed with an ASD Spectro-radiometer to determine concentrations of As, Cr, Cu, and Zn. X-ray fluorescence (XRF) validated the regression analysis results. Partial least squares regression (PLSR) models effectively predicted Zn and As contamination, achieving R^2 values of 0.65 and 0.61, with RMSE values of 7 and 94, respectively. These findings confirm the effectiveness of spectroscopy in detecting heavy metals in soil, supporting its use as a reliable tool for environmental monitoring and contamination assessment.

***Keywords*:** Soil contamination; NIRS; Partial least square; Spectrum analysis
Received: 15 December 2024; **Revised:** 20 February; **Accepted:** 27 February 2025

1. Introduction

Near infrared reflectance spectroscopy (NIRS) has been widely proposed to quickly determine soil quality and without complicated sample preparations, spectroscopy has been employed as an alternative way for determining soil properties. Heavy metals are harmful element that can contaminate and destroy soil, posing significant threats to both the environment and human health. Understanding the unique spectral signatures of heavy metals is crucial for their detection using spectroscopic methods. While some elements are inherently easier to detect due to their strong spectral features and higher concentrations in the environment, others pose more challenges [1]. Heavy metals in soil can be detected using spectral data through both direct and indirect methods, depending on the nature of the interaction between the spectral signal and the metal [2].

In recent years, soil contaminated was studied in Baghdad, for example samples were collected and analyzed to determine the concentrations of heavy metals [3]. Environmental maps were created using a Geographic Information System (GIS) to display the concentrations of these heavy metals [1,4].

Contaminants in soil, air, and water are randomly distributed, making localized sampling unreliable. Spectroscopy enables the rapid scanning of large areas, improving the detection of spatial distribution patterns of heavy metals. Several researchers have suggested that spectroscopy is effective for estimating heavy metal levels in soils, for example, Wu et al. [5] and Kemper and Sommer [6] have explored the use of reflection spectroscopy combined with statistical prediction models to estimate mineral concentrations. The authors suggested to apply remote sensing in specific technique for monitoring heavy metal contamination. Other research works, such as Smith et al. [7] and Liu et al. [8], have indicated that heavy metals can regularly identified indirectly through their effects on other soil components, including iron oxides, organic matter, and clay minerals. These components have distinct spectral characteristics that can be detected using spectral reflectance analysis.

Detecting heavy metals in soil using spectroscopy methods involve identifying the concentration levels at which these metals can be reliably detected. However the sensitivity depends on multiple factors, including the calibration model, soil properties, and the specific metal being analyzed. Each heavy metal has a unique spectral signature and some elements are inherently easier to detect than others. The previous literatures confirm that NIRS data provide an effective and accurate way to predict the presence of heavy metals in soil [4,9,10]. The results demonstrated that some heavy metals, such as As, Cr, Cu, Fe Zn, and Ni, exhibit distinct spectral features with minimum range concentrations of 10, 10, 10, 100, 50, and 10 mg/kg, respectively.

The minimum concentration required to detect a specific heavy metal using spectroscopy depends on spectroscopic technique. Different techniques have varying sensitivities, for example, inductively-coupled plasma mass spectrometry (ICP-MS) is highly sensitive and can detect heavy metals in the low ppt to high ppb range, while x-ray fluorescence (XRF) might be limited to parts per million (ppm) for some elements, but often fall in the ppm to low ppm range [11,12]. Moreover, atomic absorption spectroscopy (AAS) detection limits are typically in the ppm to low ppm range, making it an effective method for detecting even trace amounts of heavy metals in soil samples. This sensitivity allows for precise quantification of contaminants such as Pb, Cd, As, and Hg, which are of significant environmental and public health concern [12]. There are many factors affecting detection limits:

- Device Detection limit: The minimum concentration or amount of a substance that the NIRS device can reliably detect [13]. Each portable spectrometer has limit spectral resolution and sensitivity. Lower detection limits can be achieved with higher resolution devices, and vice versa, but this also depends on calibration and environmental conditions.
- Soil content: Fractions, moisture content, organic matter content and mineral composition influence the detection limits of heavy metals. Soil texture and mineral composition play a crucial role in determining spectral reflectance [14].
- Validation: Calibration models typically incorporate the soil compositions to enhance the accuracy of heavy metal detection [15]. Cross-validation with traditional laboratory methods is crucial for validation.

This study aims to assess the effectiveness of soil reflectance spectra as a reliable and non-destructive method for accurately estimating the concentrations of As, Cr, Cu, and Zn in suburban soils.

2. Materials and Methods

Previous studies and data on soil contamination in Baghdad provide a strong foundation for comparative analysis, making it an ideal location for spectroscopy and regression-based predictive modeling. Baghdad is located at the center of Iraq, between longitude 44°2157.96 E; latitude 33°1854.72 N, as in Fig (1). The general weather for Baghdad is characterized by semi-arid, hot dry summer, and cold, dry winter. The annual rainfall rates range from 100 to 150 mm. Baghdad, the capital of Iraq, has a diverse range of soil types and textures due to its geographical location, climate, and historical agricultural practices. The land on which the Baghdad city was built is almost flat and low-lying, being of alluvium soil origin due to the periodic massive overflow of the Tigris banks river. Forty soil samples were collected from seven main districts, distributed according to the four main geographic directions from the center. These districts included urban, rural, agricultural, industrial, and residential areas. The primary sub districts within Baghdad exhibit varying soil characteristics. The type and texture of soil in Baghdad influence the accumulation and mobility of heavy metals:

a) Adsorption and Retention: Clay soils and soils with high organic matter content have a greater capacity to adsorb and retain heavy metals. Metals like Pb, Cd, and As bind to clay particles and organic matter, reducing their mobility but potentially increasing their concentration in the soil [16].
b) Leaching Potential: Sandy loam soils, due to their coarser texture and better drainage, may facilitate the leaching of heavy metals into groundwater. This can lead to lower concentrations in the surface soil but poses a risk of contaminating water sources [17]
c) pH and Soil Chemistry: The pH level of Baghdad's soils can also affect the solubility and mobility of heavy metals. In more acidic soils, heavy metals tend to be more mobile and bioavailable, increasing the risk of plant uptake and entry into the food chain [18]
d) Urban and Industrial Activities: Baghdad's urbanization and industrial activities contribute to heavy metal contamination. Soils near industrial zones, traffic areas, and waste disposal sites show higher levels of heavy metals due to deposition from atmospheric sources and runoff [18].

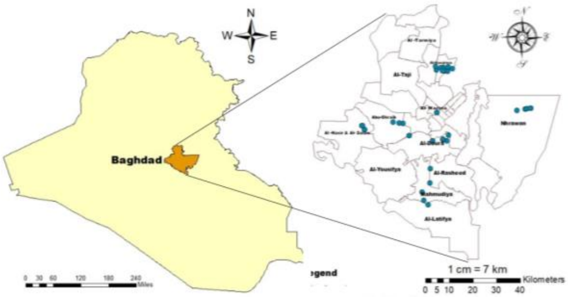


Fig. (1) Study area and sampling locations

Soil samples were analyzed to determine some heavy metals concentrations (As, Cr, Cu, and Zn) in soil samples. The x-ray fluorescence (XRF) was used to measure and examine and identify the presence of heavy metals in the soil samples at ppm levels as in table (1).

The soil samples pass through a series of pre-processing operations. The spectroscopy detection of soil contamination involves measuring the reflectance or absorption spectra of the soil and correlating these spectral features with known concentrations. The typical detection range for heavy metals in soil can vary based on several factors, including the soil properties, the calibration model used, and the sensitivity of the device.

Preparation of samples: the preparation of sample started with drying field samples, where samples are dried at 150°C, grinding and homogenizing the samples to ensure uniformity for accurate spectroscopic measurements.

Table (1) Amounts of heavy metals in soil samples* in ppm

Sample	Cr	Cu	Zn	As
1	40.4	11	23.8	11.0
2	97.9	11	31.6	11.0
3	64.0	11	32.9	11.0
4	36.0	11	38.6	11.0
5	70.02	11	25.4	11.0
6	67.1	15.2	48.8	11.0
7	61.02	11	36.4	11.0
8	94.4	11	35.0	11.0
9	71.9	13.5	24.1	11.0
10	54.8	16.21	13.0	11.0
11	53.6	11	44.2	11.0
12	63.6	11	47.1	11.0
13	65.1	11	48.0	11.0
14	58.4	36.4	13.0	65.0
15	49.3	32.05	13.0	41.1
16	41.1	27.7	13.0	11.0
17	28	19.99	13.0	35.5
18	28	21.5	13.0	11.0
19	28	20.6	13.0	32.5
20	28	17.7	13.0	40.8
21	28	11	13.0	47.4
22	34.8	11	13.0	44.5
23	28	15.241	13.0	37.1
24	28	11	34.9	180.0
25	35.7	18.73	13.0	33.1
26	28	11.6	13.0	84.7
27	28	11	13.0	11.0
28	37.6	16.6	13.0	37.2
29	28	21.5	20.3	47.2
30	78.4	23.04	13.0	48.1
31	37.6	11	13.0	44.8
32	28	19.5	13.0	93.6
33	65.8	17.6	13.0	720.0
34	28	16.34	13.0	96.3
35	37. 3	11	13.0	29.5
36	38.49	20	13.0	51.5
37	28.94	10.9	13.0	11.6
38	28	26.3	14.6	38.6
39	29.98	27.82	13.0	45.0
40	28	46.4	22.9	230.0

** Each value was obtained as the average of four samples collected from points at each corner of a square meter*

Data analysis: the Spectrora-diometer ASD Field Spec3 with a spectral range of 350–2500 nm, was used to obtain the spectral signatures [20]. The ASD spectral arrangement is separated into three spectral regions involve:

1. Visible Range (Visible) (400-700 nm)

Heavy metal like Pb, Cu, Fe and Zn compounds may show subtle absorption features in the visible range, though these are generally less pronounced than in the infrared ranges [9,20-22].

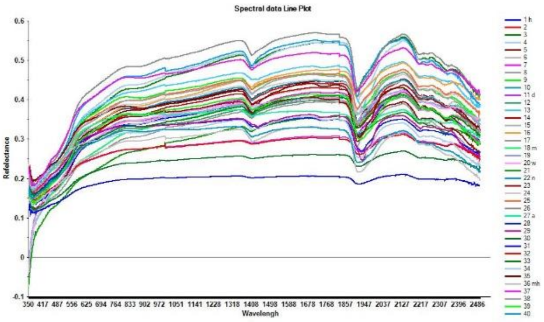
2. Near-Infrared Range (NIR) (700-1400 nm)

In the NIR range, heavy metal such as Cd, Cr and Ni can exhibit specific absorption features that are useful for its detection. These features are often more distinct than in the visible range [10,23].

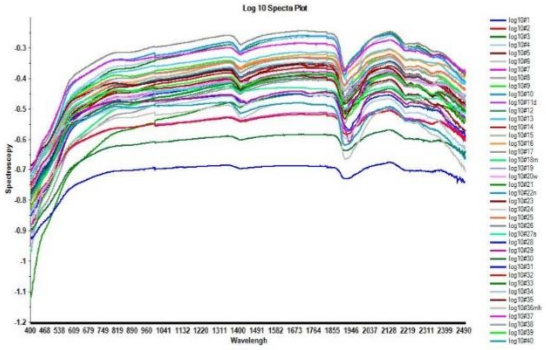
3. Shortwave Infrared Range (SWIR) (1400-2500 nm)

The shortwave infrared range is particularly useful for detecting heavy metal, as many heavy metal such as Zn and Cu compounds have characteristic absorption features in this region. Peaks and troughs in the reflectance spectrum within this range are indicative of heavy metal presence as explained in [24].

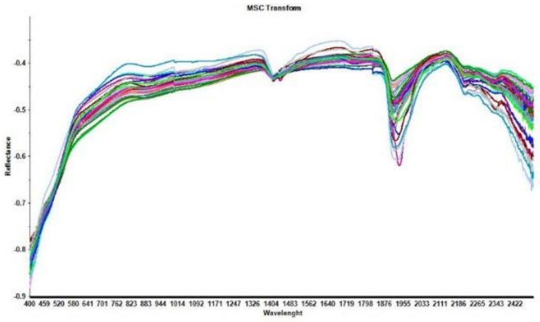
Data processing was used to reduce the physical effects, remove variation in spectral data, and to treat scattering in light. Figure (2) shows three pre-processing methods used in this study comparing to origin spectral data. Log10 was utilized to convert spectral reflectivity values to more accurate data, free of distortions and more reasonable for dealing with statistical operations. Data were log10 transformed prior to import and the new matrix was analyzed. Converting all data to a log scale reduced differences between the variables due to the measurement units and result ranges. The second processing on data is Orthogonal Signal Correction (OSC), which is a transformation and pre-processing technique for analysis operations that deals with spectral data to ensure the quality and accuracy of the results, the OSC transformation is applied to the test and training data,. It works to remove the discrepancy between the x matrix variables that are orthogonal to the y variables. The main idea is to remove the variance that is not related to the main variables in the work, which It leads to clarity and stability in regression models and is used to make the partial least squares (PLS) model more accurate as it is used in applications on NIR data [25], When light interacts with a sample during measurement, light scatters and spreads in different directions.



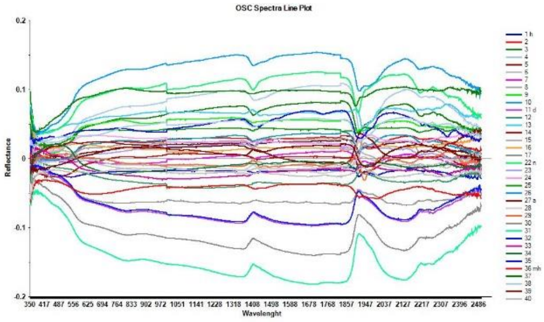
(a)



(b)



(c)



(d)

Fig. (2) Three pre-processing transform tasks (a) the original spectra line plot, (b) Log10 spectra line plot, (c) MSC spectra line plot, and (d) OSC spectra line plot

Another processing method applied in this study was the multiplicative scatter correction (MCS) process, which is used to analyze multivariate data that deals with chemical data and spectroscopic data, especially IR and NIR. It works to remove differences associated with scattering in the data that affect the quality (Fig. 2b).

The analysis allows for more accurate analysis and facilitates the identification of relevant patterns. The preprocessed spectral data were then used to calibrate the PLSR models for each heavy metal. Khaleel et al. [26] observed that using of the OSC was enhancing calibration models. This can improve the predictive accuracy of models and for quantitative soil analysis. The MSC filter was utilized to reduce scattering effects in reflectance spectroscopy for assessing soil contamination. Stenberg et al. [27] applied MSC to spectral data collected from soil samples to correct for scattering effects due to soil texture and composition variations. This can improve the detection and quantification of contaminants such as heavy metals and organic pollutants.

Heavy metals show absorption properties in the different spectral ranges between 400-2400 nm [28-34]. The Unscrambler X 10.5.1 (Camo Software, 2022) was used in this work for applied multivariate data analyses and used to create an analysis model to link it with the spectral data to know the contamination of heavy metals in the soil through the reflectance spectra of each sample, as the models indicate a link between two matrices, the first representing the spectral data of the samples, which is complex, large, and a second matrix of chemical analysis data to form a relationship linking the smallest number of equivalent factors between X and Y [35].

Accurate detection of heavy metals requires robust calibration models that correlate spectral data with known heavy metals concentrations. The partial least squares regression (PLSR) as a multivariate statistical method is commonly used for correlating spectral data with aluminum concentrations. Perform cross-validation with independent samples to ensure the robustness of the calibration model. PLSR is a powerful technique utilized for modeling relationships between sets of variables.

Test set-validation and random cross-validation were used during the regression modelling. PLSR technology is assist to predict heavy metals in soil. It is one of the modern methods for dealing with linear data, as it works to simplify the relationship between PLSR consists of an algorithm that designs a matrix between input variables x and output variables y. It is based on modelling data relationships.

The important statistics toll were used to verify and validate the regression model, included the lowest root mean square error (RMSE), standard errors (SE), correlation R2 (Pearson), which were used as indicators of the most accurate regressions [22].

3. Results and Discussion

Using Unscrambler X software, PLSR models were developed to correlate the spectral data with the heavy metal concentrations obtained from XRF analysis. The correlation and validation plot which observe the relationship between spectral data and heavy metals concentration of As, Cr , Zn and Cu values are shown in Fig. (3).

The typical spectral range for detecting heavy metals generally falls within the visible to near-infrared (VNIR) and short-wave infrared (SWIR) regions. The spectral reflectance of heavy metals features can often be identified in the VNIR to SWIR range, particularly around the bands in table (1). There may be additional weaker absorption features in the VNIR range depending on the specific compounds present in the soil.

In evaluating PLS regression models, many statistics factors were observed in (Fig. 2) used as indicated to verify and validate the regression model. However, both the R² (coefficient of determination) and RMSE are critical metrics (table 3).

The acceptability of an R² value and RMSE depends on the field of study and the complexity of the model. In some fields, an R² of 0.60 might be considered acceptable, while in others, higher values may be expected. For environmental data, where there can be significant variability, an R² of 0.50 might be considered reasonable. Lower RMSE values generally indicate a more accurate model, but what's considered "low" can vary depending on the specific application and data range. An acceptable PLS regression model typically balances a sufficiently high R² with a low RMSE. Neither metric alone should be the sole determinant of model quality. A model could have a high R² but still have a high RMSE if there are outliers or non-linear relationships not well captured by the model. Therefore, both R² and RMSE should be considered together when determining the acceptability of a PLS regression model.

The best result found in this study is that the PLSR models effectively predicted both Zn and As contamination, achieving determination coefficients (R²) of 0.65 and 0.61 with a RMSE value of 7 and 94, respectively for correlation model. While the validation models recorded R² equal to 0.39 and 0.28 with RMSE values of 9.5 and 12, respectively.

Thus, the classification of this result as the "best" is justified based on the model's comparative predictive performance and the environmental relevance of Zn and As contamination in the study area. However, PLS model value were recording smaller values for calibration and validation for both Cr and Cu as in table (4). These results suggest that PLSR models can be a valuable tool for environmental monitoring and assessment, enabling rapid and non-destructive estimation of heavy metal contamination in soils.

4. Conclusion

In conclusion, this study highlights the effectiveness of reflectance spectroscopy combined with PLSR modeling for predicting heavy metal concentrations in Baghdad’s soils. The models demonstrated reasonable predictive performance, particularly for Zn and As, though, further improvements are needed to enhance accuracy. Expanding calibration datasets and integrating complementary analytical techniques could refine model reliability and broaden their applicability. Ultimately, PLSR models serve as a valuable, non-destructive tool for rapid soil contamination assessment, contributing to more efficient environmental monitoring and sustainable land management strategies.

References

[1] H.A. Hussain, S.A. Abdullah and A.A. Al Maliki, “Predicting of heavy metals in some areas of Iraq using spectral analysis techniques”, *IOP Conf. Ser.: Earth Environ. Sci.*, 1202 (2023) 012009.

[2] A. Al Maliki, G. Owens and D. Bruce, “Capabilities of Remote Sensing Hyperspectral Images for the Detection of Lead Contamination: A Review”, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Part 7(1) (2012) 55-60.

[3] M.A. Fadhel and F.M. Abdulhussein, “Assessment of the Contamination of Baghdad Soils withLead Element”, *The Iraqi Geol. J.*, 55(1F) (2022) 166-177.

[4] A.A. Ahmed and A. Mohammed, “Spatial distribution and risk assessment of heavy metal contamination in urban soils in Baghdad, Iraq”, *Enviro. Monitor. Assess.*, 192 (2020) 374.

[5] Y. Wu et al., “Estimating heavy metal concentrations in soils using reflectance spectroscopy: A case study of copper”, *Enviro. Pollut.*, 221 (2017) 161-170.

[6] T. Kemper and S. Sommer, “Reflectance spectroscopy as a tool for mapping heavy metal contamination in soils”, *J. Enviro. Quality*, 49(5) (2020) 1264-1274.

[7] A.T. Smith, R.P. Johnson and E.M. Thompson, “Remote sensing of soil contamination: Spectral signatures of heavy metals in agricultural lands”, *Remote Sens. Enviro.*, 256 (2021) 112318.

[8] Z. Liu, Y. Zhang and X. Huang, “Indirect detection of heavy metals in soils using spectral reflectance: The role of soil organic matter and iron oxides”, *Geoderma*, 406 (2022) 115403.

[9] R.A.V. Rossel et al., “Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties”, *Geoderma*, 131(1-2) (2006) 59-75.

[10] T. Shi et al., “Estimation of heavy metal concentrations in agricultural soils using reflectance spectroscopy and partial least squares regression”, *Sci. Total Enviro.*, 776 (2021) 145968.

[11] J.M. Mermet, M. Otto and M. Valcárcel, “Atomic Spectrometry: Introduction and Fundamentals” *J. Anal. Atom. Spectro.*, 20 (2005) 11-16.

[12] D.A. Skoog, F.J. Holler and S.R. Crouch, **“Principles of Instrumental Analysis”**, 7th ed., Cengage Learning (2017), pp. 267-297.

[13] D.F. Malley, P.D. Martin and E. Ben-Dor, “Application in Agricultural and Environmental Analysis”, *J. Near Infrared Spectro.*, 12(3) (2004) 165-176.

[14] E. Choe, D. van Zyl and N.C. Woo, “Assessment of heavy metal contamination in urban soils using reflectance spectroscopy”, *Enviro. Pollut.*, 153(1) (2008) 112-12.

[15] J.B. Reeves, D.B. Smith and K. Gallagher, “Developing soil reflectance spectral libraries and calibration models for quantitative analysis of soil properties”, *Geoderma*, 241-242 (2015) 150-159.

[16] B.J. Alloway, **“Heavy Metals in Soils: Trace Metals and Metalloids in Soils and their Bioavailability”**, Springer Science & Business Media (2013).

[17] A. Kabata-Pendias, **“Trace Elements in Soils and Plants”**, CRC Press (2010).

[18] Z. Li et al., “A review of soil heavy metal pollution from mines in China: Pollution and

health risk assessment", *Sci. Total Enviro.*, 468-469 (2014) 843-853.

[19] ASD Inc., "FieldSpec 3 User Manual", Analytical Spectral Devices, Boulder (CO, USA) (2008).

[20] E. Ben-Dor and A. Banin, "Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties", *Soil Sci. Soc. America J.*, 59(2) (1995) 364-372.

[21] D.F. Malley, L. Yesmin and R.G. Eilers, "Rapid analysis of soil organic carbon, nitrogen, and phosphorus using near infrared reflectance spectroscopy", *Soil Sci. Soc. America J.*, 66(5) (2002) 1682-1688.

[22] A. Al Maliki, D. Bruce and G. Owens, "Prediction of lead concentration in soil using reflectance spectroscopy", *Enviro. Technol. Innov.*, 8(15) (2014) 1.

[23] H. Xie et al., "Mapping heavy metals in agricultural soils using reflectance spectroscopy and machine learning", *Remote Sensing*, 12(5) (2020) 830.

[24] Z. Shi and Q. Wang, "Characterization of heavy metals in soil by reflectance spectroscopy in the range of 350–2500 nm", *Int. J. Appl. Earth Observ. Geoinform.*, 28 (2014) 165-179.

[25] R. Laref et al., "Orthogonal Signal Correction to Improve Stability Regression", *J. Sensors,* 2017 (2017) 8.

[26] Z.A. Khaleel, A.H. Shaban and A.A. Al Maliki, "Investigations for Soil Contamination with Hydrocarbon Compounds near Oil Stations", *J. Phys.: Conf. Ser.*, 2754 (2024) 012025.

[27] B. Stenberg et al., "Visible and near infrared spectroscopy in soil science", *Adv. Agronomy*, 107 (2010) 163-215.

[28] H.S. Bingari et al., "Application of near infrared spectroscopy in sub-surface monitoring of petroleum contaminants in laboratory-prepared soils", *Soil Sedim. Contam.: Int. J.*, 32(4) (2023) 400-416.

[29] R.N. Clark, "Spectroscopy of rocks and minerals, and principles of spectroscopy", in *Remote Sensing for the Earth Sciences, Manual of Remote Sensing*, John Wiley & Sons, vol. 3 (1999) pp. 3-58.

[30] C.W. Chang et al., "Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties", *Soil Sci. Soc. America J.*, 65(2) (2001) 480-490.

[31] C. Pasquini, "Near infrared spectroscopy: fundamentals, practical aspects and analytical applications", *J. Brazilian Chem. Soc.*, 14(2) (2003) 198-219.

[32] E. Choe and W. Lee, "Detection and quantification of heavy metals in soil using reflectance spectroscopy and machine learning", *J. Hazard. Mater.*, 394 (2020) 122555.

[33] R.N. Okparanma and A.M. Mouazen, "Visible and Near-Infrared Spectroscopy Analysis of a Polycyclic Aromatic Hydrocarbon in Soils", *The Sci. World J.*, 2013 (2013) 9.

[34] A.A. Maliki, D. Bruce and G. Owens, "Spatial distribution of Pb in urban soil from Port Pirie, South Australia", *Enviro. Technol. Innov.*, 4 (2015) 123-136.

[35] A.A. Ajeel and A.K. Owaid, "Study of contrast between satellite image data and ground data", *Baghdad Sci. J.*, 8(3) (2011) 717-722.

Table (4) Validation statistics parameters related to PLS regression models which were utilized to anticipate heavy metals concentration in soil samples, the certainty level of $p<0.005$ was utilized in all prediction models

Elements	No. of samples	PLS Factor	Calibration		Validation	
			R^2	RMSE	R^2	RMSECV
As,	40	5	0.61	94	0.28	12
Cr,	39	7	0.57	8.8	0.27	133
Zn,	40	6	0.65	7	0.39	9.5
Cu,	40	3	0.47	5.4	0.40	5.9

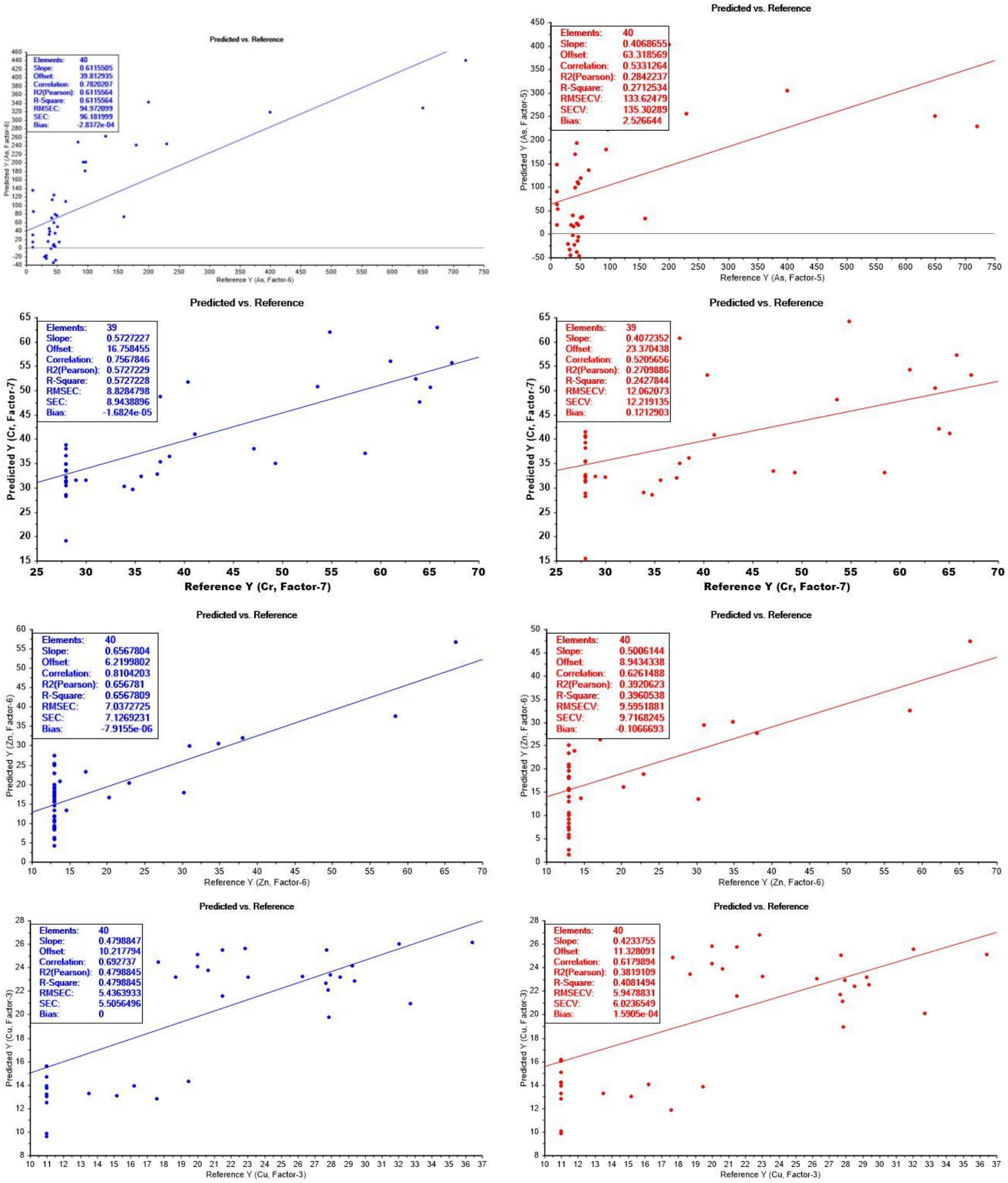


Fig. (3) PLSR models illustrate the relationship between the expected (predictive) and measured values (references) of As, Cr, Zn and Cu respectively: Calibration (blue) and Validation (red)