SPECTRAL PREPROCESSING FOR HYPERSPECTRAL REMOTE SENSING OF HEAVY METALS IN WATER

M. Lee*, X.-Y. Chen, H.-C. Lee

Department of Safety, Health and Environmental Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan – (mlee, 0613824, sunmao)@nkust.edu.tw

KEY WORDS： Hyperspectral Data; First Derivative Transformation; Logarithm Transformation; Multivariate Linear Regression; Environmental Monitoring

ABSTRACT:

This study aims to investigate the feasibility of using hyperspectral remote sensing technique by visible-near infrared spectroradiometer (VNIR, FieldSpec HandHeld 2) for rapid water monitoring of heavy metal, followed by comparison of different spectral preprocessing methods for the development of quantitative predictive model. The water samples evaluated in this study were prepared in our laboratory by dilution of stock solutions. Heavy metals of lead (Pb), zinc (Zn) and copper (Cu) in the range of concentration between 100 to 2000 mg/L were selected as the target samples in this study. The sensitive bands for the target metals were characterized in the range from 800 nm to 1075 nm, based on the reflectance spectral data. Spectral data for developing of the quantitative predictive model was first preprocessed with first derivative and logarithm transformation, followed by establishing of the prediction model using multivariate linear regression (MLR). It was observed that increase in the number of sensitive bands for the MLR can significantly improve the adjusted R^2 for the model. The prediction model for Cu was found to have the highest adjusted R^2 of 0.92 and least normalized root mean square error (NRMSE) of 0.065, while using the reflectance values of 7 sensitive bands. This result could be attributed to the blue color characteristic of the solution, whereas the others remain clear. Additionally, the first derivative transformation was determined as the best method for predicting Pb, whereas the logarithm transformation provided the best outcomes for predicting Cu and Zn.

1. INTRODUCTION

Current analytical methods for determining heavy metals in water, including atomic absorption/emission spectroscopy (AAS), inductively coupled plasma mass spectrometry (ICP-MS) and cold vapour atomic fluorescence spectrometry (CV-AFS), require use of chemicals for sample processing and pretreatment as well as long labor time for sampling and analysis. Thus, use of non-destructive method for rapid analysis of the metal content in water has drawn considerable attention, but there has been very little research reported on the effectiveness of such use.

Remote sensing techniques are widely applied to provide environmental quality information based on the spectral data from optical properties of the sensing objects. The use of optical properties as key parameters has been widely used in environmental monitoring. Recent efforts were therefore made in the development of emerging analytical approach using non-destructive spectroscopic techniques, including laser induced breakdown spectrometry (LIBS), X-ray fluorescence (XRF) and visible-near infrared spectroradiometer (VNIR), for analysing metal content in samples. Use of hyperspectral technique for determination of heavy metal in environmental samples were seldom reported, especially for water bodies. Dong et al. (2016) revealed the use of spectral data for estimating heavy metals in mining reclamation areas. Aldabaa et al. (2015) demonstrated the feasibility of using proximal (potable-XRF and VNIR) sensing for rapid soil salinity quantification; while other studies also shown comparable potential for using the abovementioned techniques for prediction of organic matter (Morona et al., 2017) and pH value (Sharma et al, 2014) in soil.

All of the proposed approaches are expected to be used as alternatives for conventional analytical techniques to provide reliable environmental monitoring results. But chemometric treatment, such as multi-linear regression, stepwise multi-linear regression, partial least squares regression, on the spectral data all played an important role in improving the accuracy on the prediction models (Gredilla et al., 2016; Dörnhöfer & Oppelt, 2016).

Development of statistical methods for the spectral data in characterization and measuring of sensing objects is therefore a key focus in the field of environmental monitoring using remote sensing. In particular, research interests in statistical analysis of remote sensing data to provide measurements for sustainable developments goals are continuously increasing (Holloway and Mengersen, 2018). The demand for good quality, real-time, non-destructive and high resolution water quality information therefore has been growing in research areas of environmental monitoring, in order to comply with the concept of green chemistry.

This study aims to investigate the feasibility of using hyperspectral remote sensing technique by visible-near infrared spectroradiometer (VNIR, FieldSpec HandHeld 2) for rapid water monitoring of heavy metal, followed by comparison of different spectral preprocessing methods for the development of quantitative predictive model. Clarification on the applicability of VNIR for field study is also addressed. Results of this technique are expected to support regional water quality monitoring, in particular for area that are inconvenient for aircraft or satellite-based remote sensing.

* Corresponding author

2. RESEARCH METHODS

2.1 Research Development

Figure 1 presents the research flowchart of this study. In order to elucidate the feasibility of using VNIR-based hyperspectral remote sensing technique for rapid water monitoring of heavy metal, the study first worked on identification of spectral characteristics of the study metals, followed by comparison of different spectral preprocessing methods for the development of quantitative predictive model.

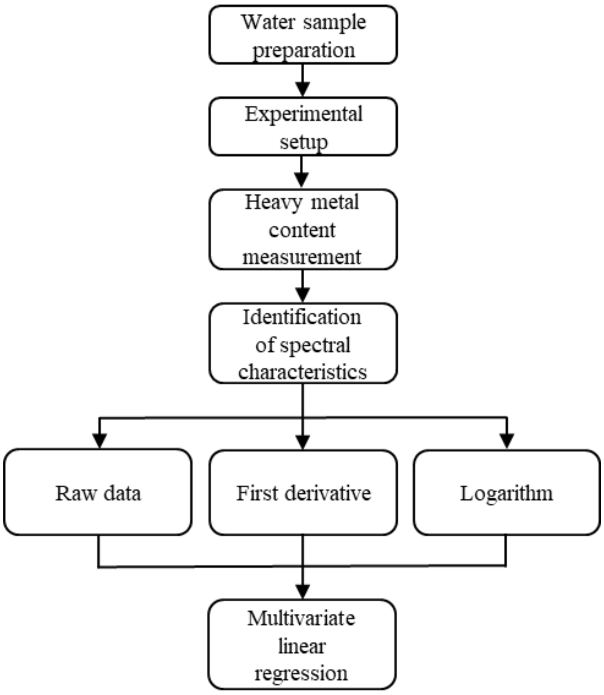

Figure 1. Research flowchart of this study

2.2 Water Sample Preparation

The water samples evaluated in this study were prepared in our laboratory by diluting of stock solutions. Heavy metals of lead (Pb), zinc (Zn) and copper (Cu) were selected as the target water constitutes in the study. Concentrations of the heavy metals in water are in the range of 100-3000 mg/L. The water samples were stored in a quartz beaker during sensing measurements, in order to minimize the reflective interference from the beaker.

2.3 Experimental Setup

Figure 2 reveals the experimental setup for the remote sensing technique using visible-near infrared spectroradiometer (VNIR, FieldSpec HandHeld 2). The sensing experiments were conducted in an open field (i.e., roof of our research building) on regular sunny days. The distance between the VNIR and water samples was set to be 30 cm. Each of the sensing experiment was repeated for at least 10 times. The FieldSpec VNIR has spectral range between approximately 325 and 1075 nm.

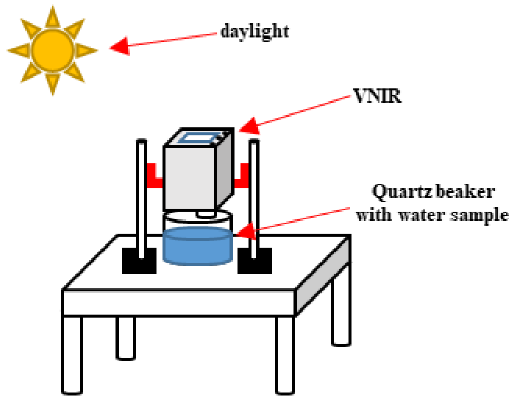

Figure 2. Setup of light booth for environmental simulation for sensing measurements

2.4 Development of Prediction Model

Spectral data for developing of the quantitative predictive model was first preprocessed with first derivative or logarithm transformation, followed by establishing of the prediction model using multivariate linear regression (MLR). Stability and accuracy of the prediction model were evaluated by the coefficient of determination (adjusted R^2) and the normalized root mean square error (NRMSE), respectively.

3. RESULTS AND DISCUSSION

3.1 Identification of Spectral Characteristics

Figure 3 shows the spectral data (from 800 nm to 1075 nm) for the studied heavy metal using the environmental experiment setup proposed in this study. The spectral curves without pretreatment were relatively smooth with no abnormal reflection peaks, which were directly used to identify the sensitive bands for the target metals. The sensitive bands were characterized as 984 nm, 968 nm and 1024 nm for Cu, Zn and Pb, respectively. Spectral characteristics on other sensitive bands for the studied metals are provided in Table 1. It was shown that most of the sensitive bands for the studied heavy metals fall in the range from 800 nm to 1075 nm, which were in line with the findings in Rostom et al. (2017).

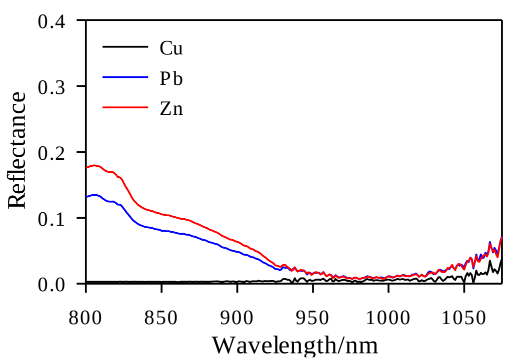

Figure 3. Spectral data for the studied metal elements (both of the samples are at 2000 mg/L)

Metal	Spectral characteristics(nm)
Cu	942、984、986、998、1063、1065、1073
Pb	951、983、998、1024、1053、1066、1072
Zn	930、968、999、1037、1040、1062、1073

Table 1. Spectral data for the studied metal elements

Figure 4 presents the spectra for the studied metal elements at different concentration from 500 mg/L to 3000 mg/L. It was observed that the reflectance on sensitive bands improved as increase of metal concentrations. Detail spectral data for the metals are provided in Table 2 to Table 4. All of the findings suggested the need of data preprocessing or select of appropriate chemometric approach to improve the accuracy and reliability on prediction model (Dörnhöfer & Oppelt, 2016)。

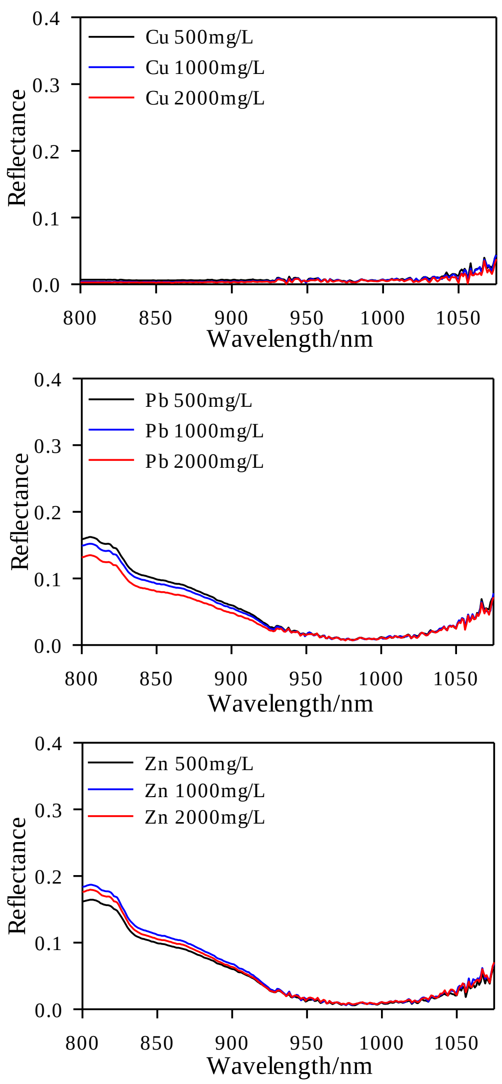

Figure 4. Spectra of the studied metal elements under different concentration (500 to 2000 mg/L)

Conc.(mg/L) / Spectral(nm)	500	1000	2000
942	0.0100	0.0086	0.0081
984	0.0056	0.0063	0.0071
986	0.0077	0.075	0.0072
998	0.0058	0.0053	0.0056
1052	0.0223	0.0166	0.0162
1065	0.0199	0.0190	0.0137
1073	0.0304	0.0293	0.0216

Table 2. Reflectance characteristics at sensitive bands for Cu

Conc.(mg/L) / Spectral(nm)	500	1000	2000
951	0.0167	0.0181	0.0163
983	0.0086	0.0081	0.0086
998	0.0097	0.0093	0.0091
1024	0.0120	0.0136	0.0150
1052	0.0359	0.0374	0.0347
1066	0.0546	0.0531	0.0494
1072	0.0535	0.0488	0.0457

Table 3. Reflectance characteristics at sensitive bands for Pb

Conc.(mg/L) / Spectral(nm)	500	1000	2000
930	0.0301	0.0309	0.0472
968	0.0104	0.0101	0.0121
999	0.0072	0.0106	0.0122
1037	0.0149	0.0191	0.0247
1040	0.183	0.0225	0.0299
1062	0.0338	0.0436	0.0573
1073	0.0552	0.0544	0.0636

Table 4. Reflectance characteristics at sensitive bands for Zn

3.2 Spectral Estimation Results

Results from the spectral analysis were then be used for correlation analysis between the measured content and the spectral reflectance or elemental results.

Figure 5 shows the correlation between the measured metal concentration in the validation samples (by ICP-OES) compared to the proposed non-destructive approach using VNIR. It was found that the correlation was the highest for Cu (R^2=0.949), followed by Zn and Pb (both of the metals were less significant but still at moderate levels). The relatively higher correlation found for Cu could be attributed to the blue color characteristic of the solution, whereas the others remain clear. Future research efforts are therefore suggested to apply various multivariate calibrations or data preprocessing techniques for improving the accuracy of prediction models.

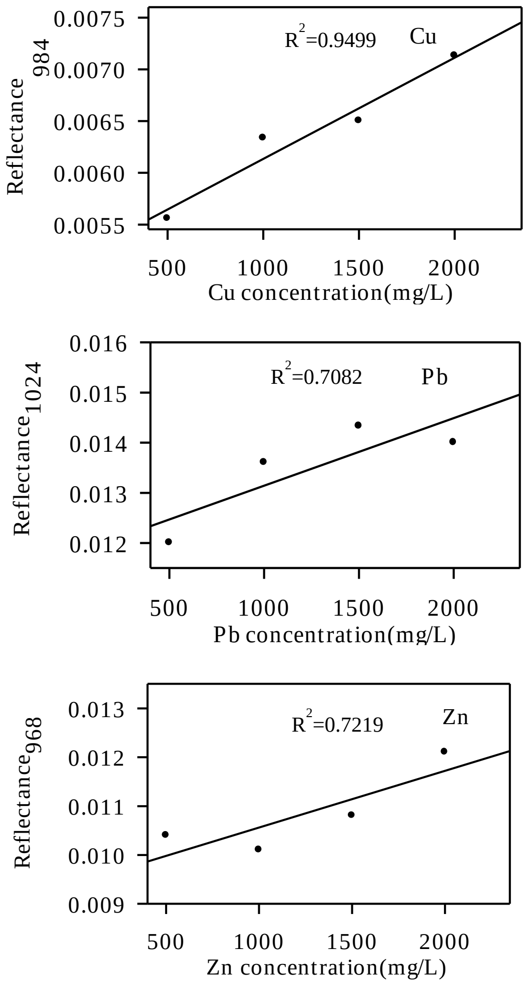

Figure 5. The measured metal concentration in the validation samples (by ICP-OES) compared to the proposed non-destructive approach using VNIR

3.3 MLR Prediction Model

Spectral data for developing of the quantitative predictive model was first preprocessed with first derivative and logarithm transformation, followed by establishing of the prediction model using multivariate linear regression (MLR).

Table 5 compares the results of adjusted R^2 and NRMSE for the prediction models with and without data preprocessing. The prediction model for Cu with logarithm transformed spectral data was found to have the highest adjusted R^2 of 0.92 and least normalized root mean square error (NRMSE) of 0.065, while using the reflectance values of 7 sensitive bands. This result could be attributed to the blue color characteristic of the solution, whereas the others remain clear. Similarly, the logarithm transformed approach significantly improved the accuracy of the prediction model for Zn, giving an adjusted R^2 of 0.83 and NRMSE of 0.097. Therefore, the first derivative transformation was determined as the best method for predicting Pb, whereas the logarithm transformation provided the best outcomes for predicting Cu and Zn.

model	Preprocessing	adjusted R^2	NRMSE
Cu	Raw data	0.91	0.070
	first derivative	0.69	0.120
	logarithm	0.92	0.065
Pb	Raw data	0.64	0.090
	first derivative	0.78	0.070
	logarithm	0.69	0.134
Zn	Raw data	0.80	0.108
	first derivative	0.65	0.143
	logarithm	0.83	0.097

Table 5. Comparison of prediction results for raw data and preprocessing of spectral data for the studied heavy metals

Table 6 shows the established prediction model for the studied heavy metals using the reflectance values of 7 sensitive bands for MLR. It was observed that increase in the number of sensitive bands for the MLR can significantly improve the adjusted R^2 for the model.

model	equation
Cu	$Y=1614.94+10186.75\cdot R_{w\_1065}-16482.5\cdot R_{w-1073}-35529.6\cdot R_{w-942}+83296.38\cdot R_{w-947}-50347.7\cdot R_{w-1063}+14756.31\cdot R_{w-984}-12220\cdot R_{w-986}$
Pb	$Y=2582.42-13158.3\cdot R_{w-1072}+39143.3\cdot R_{w-1066}-45110.8\cdot R_{w-1052}-210393\cdot R_{w-1024}-544868\cdot R_{w-998}+1374160\cdot R_{w-983}+1995240\cdot R_{w-951}$
Zn	$Y=-10960.1-1912.83\cdot R_{w-1073}+2383.259\cdot R_{w-1062}+9398.775\cdot R_{w-1040}+1133.273\cdot R_{w-1037}-598.672\cdot R_{w-999}-7251.19\cdot R_{w-968}-5814.01\cdot R_{w-930}$

Note: R_{w-1065} is reflectance at wavelength 1065 nm

Table 6. MLR prediction model for the studied heavy metals (100-2000 mg/L)

4. CONCLUSIONS

The study demonstrated the use of non-destructive analytical method of VNIR for quantification of metal contents (Cu, Pb, Zn) in water samples. Reasonable agreements between the measurements from validated samples and moderate correlations were perceived for simple linear regression model using spectral information from VNIR. With respect to the establishment of prediction models using MLR, data preprocessing was shown to be effective in improving the accuracy and stability of the prediction models. First derivative transformation was determined as the best method for predicting Pb, whereas the logarithm transformation provided the best outcomes for predicting Cu and Zn. Results from this study are expected to provide useful information on rapid identification of metal-polluting sources. Overall, this study demonstrated the feasibility of using nondestructive alternative of hyperspectral method for monitoring of the metal content in water environment. All the information would be useful for identification of pollution sources as well as for improvement of pollution management strategies.

ACKNOWLEDGEMENTS

The authors are grateful for the research funding from the Ministry of Science and Technology in Taiwan through project number 107-2218-E-992 -305.

REFERENCES

Aldabaa, A.A.A., Weindorf, D.C., Chakraborty, S., Sharma, A., Li, B., 2015. Combination of proximal and remote sensing methods for rapid soil salinity quantification. *Geaderma*, 239-240, 36-46.

Dong, J., Dai, W., Xu, J., Li, S., 2016. Spectral estimation model construction of heavy metals in mining reclamation areas. *International Journal of Environmental Research and Public Health*, 13, 640.

Dörnhöfer, K., Oppelt, N., 2016. Remote sensing for lake research and monitoring – Recent advances. *Ecological Indicators*, 64, 105-122.

Gredilla, A., Fdez-Ortiz de Vallejuelo, S., Elejoste, N., de Diego, A., Madariaga, J.M., 2016. Non-destructive Spectroscopy combined with chemometrics as a tool for Green Chemical Analysis of environmental samples: A review. *TrAC Trends in Analytical Chemistry*, 76, 30-39.

Holloway, J., Mengersen, K., 2019. Statistical machine learning methods and remote sensing for sustainable development goals: A review. *Remote Sensing*, 10, 1365.

Morona, F., dos Santos, F.R., Brinatti, A.M., Melquiades, F.L., 2017. Quick analysis of organic matter in soil by energy-dipersive X-ray fluorescence and multivariate analysis. *Applied Radiation and Isotopes*, 130, 13-20.

Sharma, A., Weindorf, D.C., Man, T., Aldabaa, A.A.A., Chakraborty, S., 2014. Characterizing soils via portable X-ray fluorescence spectrometer: 3. Soil reaction (pH). *Geoderma*, 232-234, 141-147.

Rostom, N.G., Shalaby, A.A., Issa, Y.M., Afifi, A.A., 2017. Evaluation of Mariut Lake water quality using Hyperspectral Remote Sensing and laboratory works. *The Egyptian Journal of Remote Sensing and Space Science*, 20, S39-S48.