Catena 162 (2018) 72–79

Contents lists available at ScienceDirect

Catena

journal homepage: www.elsevier.com/locate/catena

# Feasibility of using visible and near-infrared reflectance spectroscopy to monitor heavy metal contaminants in urban lake sediment

Qinghu Jiang[a], Minxia Liu[a,b], Jun Wang[a], Feng Liu[a,*]

[a] Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China
[b] College of Forestry, Shanxi Agricultural University, Taigu 030801, China

ARTICLE INFO

*Keywords:*
Heavy metal
Urban lake sediments
Visible and near-infrared reflectance spectroscopy
Spectral feature selection
Predictive mechanism

ABSTRACT

Contamination of urban lake sediments by heavy metals is a major threat to environmental safety and human health due to its high toxicity and persistence. This paper aims to examine the feasibility of using visible and near-infrared reflectance spectroscopy (VNIRS) to rapidly quantify heavy metals (i.e., As, Cd, Cr, Cu, Hg, Ni, Pb and Zn) in urban lake sediments. Lake sediment samples ($n$ = 103) were collected from the East Lake in Wuhan, China. Partial least squares regression (PLSR) calibration models were developed for the heavy metals estimation. Genetic algorithm (GA) and competitive adaptive reweighted sampling (CARS) were compared to test whether spectral feature selection could improve heavy metals predications or not. Correlation analysis was carried out to gain better understanding of the predictive mechanism for the assessment of the heavy metals in sediments using VNIRS. PLSR calibration models showed that Cd, Hg, Ni and Pb had acceptable model prediction (with $r^2_{CV}$ values range from 0.32 to 0.40), while model results for As, Cr, Cu and Zn were unsatisfactory (with $r^2_{CV}$ values of 0.01–0.06). These different accuracies were likely caused by the different relationships between heavy metals and spectrally-active constitutes (e.g., total organic carbon, TOC). When compared with full-spectrum PLSR models, GA-PLSR and CARS-PLSR models slightly increased the accuracies by remove uninformative spectral variables. Given these undesired practices, this study demonstrated that VNIRS should not be recommended for heavy metals estimation in urban lake sediment. The inherent correlations between heavy metals and spectrally-active constitutes (e.g., TOC) rather than modeling method (e.g., feature selection algorithms) were critical for the limited heavy metals estimation using VNIRS.

## 1. Introduction

Over the past decades, with urbanization and economic development, accumulation of heavy metals in aquatic environments (e.g., lake sediment) has become an important issue worldwide related to environmental risk and human health (Farias et al., 2007). The distribution and retention of heavy metals in lake sediments is of crucial importance for the understanding of biogeochemical processes in aquatic systems and lake management (Arain et al., 2008; Santos-Echeandia et al., 2009; Smolders et al., 2006). Spatial and temporal variation of heavy metals can be used to indicate fast environmental change. Therefore, it is necessary to regularly monitor and evaluate heavy metal contaminant in surface sediment (i.e., the top layer of lake sediment) and assess its potential ecological risks for aquatic systems.

Heavy metals in lake sediment are usually analyzed by wet chemical methods. These conventional methods are expensive, time consuming, and labor intensive. Meanwhile lake surface sediments are very sensitive to peripheral land-cover change and seasonal precipitation, so the heavy metal content in lake sediments shows high spatial and temporal variation (Kleinebecker et al., 2013). Thus, an integrated assessment and regularly monitoring of the heavy metals in lake surface sediments is necessary, but often lacking. Development of an alternative assessment method that is rapid, inexpensive, reproducible and eco-friendly for the determination of heavy metals in sediments would be of great value.

Visible and near-infrared reflectance spectroscopy (VNIRS), as a fast and low-cost tool for soil analysis, has been widely used to measure physical and chemical properties (e.g., organic carbon (OC), total nitrogen (TN) and clay) in soils with relatively reliable accuracy (Viscarra Rossel et al., 2006). The measurements of organic components, water and clay contents are based on their direct spectral response in the VNIR range. However, unlike them, heavy metals (such as As, Cd, Cu, Hg, and Zn) do not possess direct spectral responses in the VNIR region (Chodak, 2008; Malley and Williams, 1997) unless the concentration exceeds 4000 mg kg$^{-1}$ (Wu et al., 2005). Thus, VNIRS measurements of heavy metal are not based on their direct spectral responses. Instead,

* Corresponding author.
*E-mail address:* liufeng@wbgcas.cn (F. Liu).

https://doi.org/10.1016/j.catena.2017.11.020
Received 3 March 2017; Received in revised form 27 September 2017; Accepted 22 November 2017

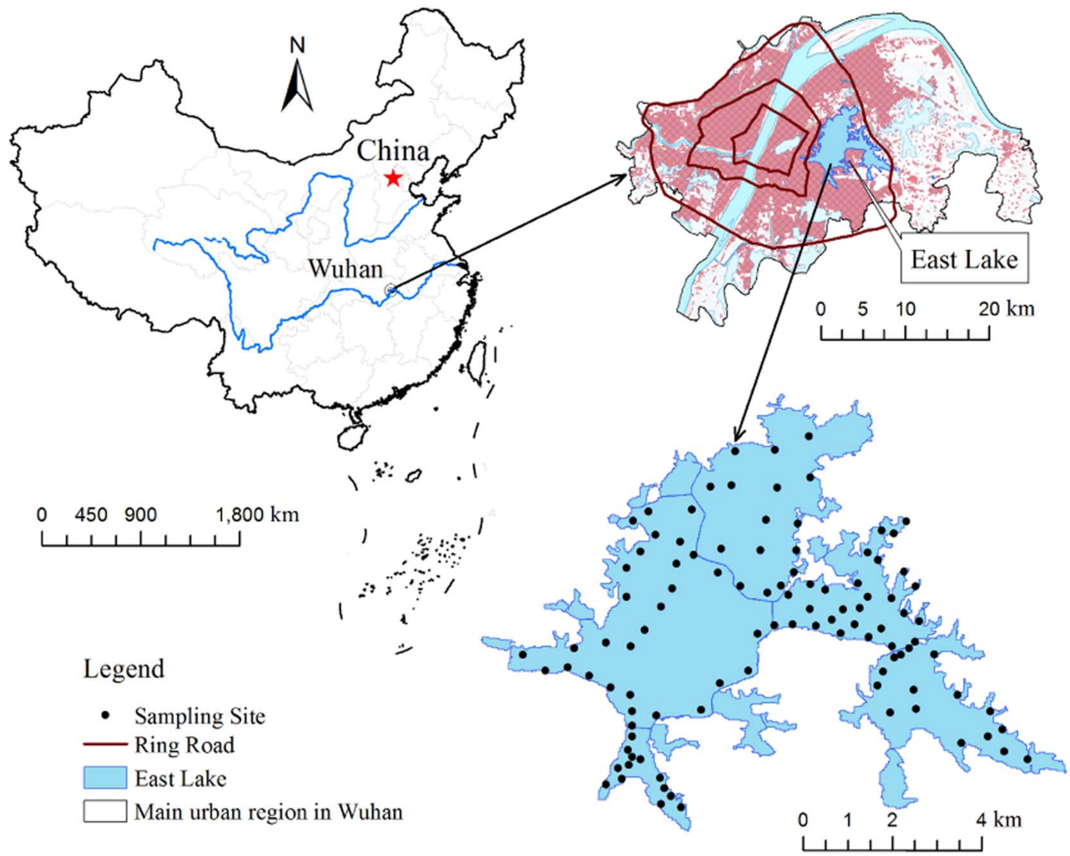


**Fig. 1.** Geographic location of East Lake and spatial distribution of the collected sediment samples.

some studies successfully predicted heavy metal concentrations in soils rely on the "surrogate" correlations between heavy metals and spectrally-active properties of soils, including soil organic matter, clay, and iron oxide ($Fe_2O_3$) (Chen et al., 2016). For example, Malley and Williams (1997) demonstrated that the $r^2$ values of the prediction models for Cu, Zn, Pb and Ni in a natural lake were 0.91, 0.93, 0.81 and 0.88, respectively. Siebielec et al. (2004) reported that heavy metals in surface soils were detectable ($r^2$ values were for Fe = 0.87, Cu = 0.61, Ni = 0.84 and Zn = 0.67) using NIRS. Based on these successful practices, VNIRS were shown to be a promising tool to predict heavy metals concentrations in soils.

However, there are also studies showing poor predictions. They indicated that the persistent VNIRS "myth" that VNIRS could be used to monitor the presence of heavy metals in soils should be rest. For instance, Baveye and Laba (2015) pointed out that the correlations between heavy metals and soil-borne spectral sensitive compounds may or may not exist. They believed that the correlations depend on the formation and accumulation mechanism of heavy metal pollution and the specific circumstance (e.g., soil types, land use and management). Thus, which soil variable they are related to and the actual correlations are site specific.

Given these contradictory results, to clarify the confusion, we wanted to evaluate the possibility of using VNIRS to estimate heavy metal concentrations in urban lake sediments. Since heavy metals are spectrally featureless elements, when and under what circumstances can we estimate their concentration in lake sediments using VNIRS? Can we use wavelength feature selection techniques to improve model prediction accuracies? To answer these questions, we collected lake surface sediment samples from the East Lake in Wuhan China to evaluate the feasibility of using VNIRS to estimate heavy metal concentrations. The East Lake is an internal lake of 33 $km^2$ in Wuhan, a large city with population size over 10 million. Heavy metal pollution in the East Lake is becoming a widespread concern by both the government and local communities.

Soil VNIR spectra are nonspecific, multi-collinear owing to the spectral overlap of spectrally-active properties and redundant noise, which may weaken the performance of VNIRS in estimating soil properties (Cai et al., 2008; Vohland et al., 2011). By selecting the most informative feature wavelengths, the performance of VNIRS in identification of soil constituents can be improved. Nowadays, many studies have shown that the partial least squares regression (PLSR) coupled with feature selection algorithms, such as genetic algorithm (GA), successive projection algorithm (SPA), variable importance of projection (VIP), competitive adaptive reweighted sampling (CARS) (Y.L. Liu et al., 2014; Shi et al., 2014; Vohland et al., 2014), can be used to select optimal informative bands to better estimate soil OC, TN and pH values, etc. However, few studies have focused on the application of feature selection algorithm in the prediction of soil heavy metals contents. Wang et al. (2014) predicted low concentrations of As, Cu, Pb and Zn in agricultural soils using VNIRS, and the GA-PLSR models achieved better accuracies than the PLSR models.

Based on these backgrounds, we want to answer two main questions: (1) is it feasible to use VNIRS to estimate heavy metal concentration in urban internal lake sediments? (2) Can we use spectral feature selection techniques (i.e., CARS algorithm compared to GA algorithm) to establish more accurate and concise VNIRS models? By answering this two questions, we hope to evaluate the potential of VNIRS in measurement of heavy metals (i.e., As, Cd, Cr, Cu, Hg, Ni, Pb and Zn) in urban lake sediment and to better understand the prediction mechanism.