

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Biological Systems Engineering: Papers and Publications

Biological Systems Engineering

2020

Predicting total petroleum hydrocarbons in field soils with Vis–NIR models developed on laboratory-constructed samples

Nuwan Wijewardane, Post Doc research associate

Yufeng Ge


Natasha Sihota

Thomas Hoelen

Toni Miao

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/biosysengfacpub>

 Part of the [Bioresource and Agricultural Engineering Commons](#), [Environmental Engineering Commons](#), and the [Other Civil and Environmental Engineering Commons](#)

This Article is brought to you for free and open access by the Biological Systems Engineering at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Biological Systems Engineering: Papers and Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Authors

Nuwan Wijewardane, Post Doc research associate; Yufeng Ge; Natasha Sihota; Thomas Hoelen; Toni Miao; and David C. Weindorf

TECHNICAL REPORTS

Ecosystem Restoration

Predicting total petroleum hydrocarbons in field soils with Vis–NIR models developed on laboratory-constructed samples

Nuwan K. Wijewardane¹  | Yufeng Ge² | Natasha Sihota³ | Thomas Hoelen³ | Toni Miao⁴ | David C. Weindorf⁵ 

¹ Dep. of Biological Systems Engineering, Univ. of Nebraska-Lincoln, 158 Chase Hall, East Campus, Lincoln, NE 68583, USA

² Dep. of Biological Systems Engineering, Univ. of Nebraska-Lincoln, Chase Hall, East Campus, Lincoln, NE 68583, USA

³ Chevron Energy Technology Company, San Ramon, CA 94583, USA

⁴ Chevron Energy Technology Company, Richmond, CA 94801, USA

⁵ Dep. of Plant and Soil Science, Texas Tech Univ., Lubbock, TX 79409, USA

Correspondence

Nuwan K. Wijewardane, Dep. of Biological Systems Engineering, Univ. of Nebraska-Lincoln, 158 Chase Hall, East Campus, Lincoln, NE 68583, USA.

Email: nwijewardane2@unl.edu

Assigned to Associate Editor Jun-Jian Wang.

Abstract

Accurate quantification of petroleum hydrocarbons (PHCs) is required for optimizing remedial efforts at oil spill sites. While evaluating total petroleum hydrocarbons (TPH) in soils is often conducted using costly and time-consuming laboratory methods, visible and near-infrared reflectance spectroscopy (Vis–NIR) has been proven to be a rapid and cost-effective field-based method for soil TPH quantification. This study investigated whether Vis–NIR models calibrated from laboratory-constructed PHC soil samples could be used to accurately estimate TPH concentration of field samples. To evaluate this, a laboratory sample set was constructed by mixing crude oil with uncontaminated soil samples, and two field sample sets (F1 and F2) were collected from three PHC-impacted sites. The Vis–NIR TPH models were calibrated with four different techniques (partial least squares regression, random forest, artificial neural network, and support vector regression), and two model improvement methods (spiking and spiking with extra weight) were compared. Results showed that laboratory-based Vis–NIR models could predict TPH in field sample set F1 with moderate accuracy ($R^2 > .53$) but failed to predict TPH in field sample set F2 ($R^2 < .13$). Both spiking and spiking with extra weight improved the prediction of TPH in both field sample sets (R^2 ranged from .63 to .88, respectively); the improvement was most pronounced for F2. This study suggests that Vis–NIR models developed from laboratory-constructed PHC soil samples, spiked by a small number of field sample analyses, can be used to estimate TPH concentrations more efficiently and cost effectively compared with generating site-specific calibrations.

Abbreviations: ANN, artificial neural networks; DCM, dichloromethane; GC-FID, gas chromatography with flame ionized detector; MIR, mid-infrared; NIR, near-infrared; PHC, petroleum hydrocarbon; PLSR, partial least squares regression; RF, random forests; RPD, ratio of performance to deviation; RPIQ, ratio of performance to interquartile range; SVR, support vector regression; TPH, total petroleum hydrocarbons; Vis–NIR, visible and near-infrared reflectance spectroscopy.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Journal of Environmental Quality* published by Wiley Periodicals LLC on behalf of American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America

1 | INTRODUCTION

Petroleum hydrocarbons (PHCs) are one of the most important energy sources for modern society. However, because of accidental releases at production and distribution sites, PHCs in soil are often a focus for remedial actions. To design an effective and efficient plan to address PHC-impacted soil remediation objectives, it is critical to obtain an accurate understanding of PHC (or TPH) concentration levels in impacted soil.

Conventional methods for TPH analysis in soils usually involve multiple steps, including sample collection, followed by laboratory extraction, purification, and analyses. TPH analyses are therefore often time consuming, costly, and mostly limited to the laboratory environment. While laboratory approaches provide accurate measurements, in situ or field-based methods (which might have lower analytical accuracy) are also highly valuable. For example, an approach which can provide rapid, real-time quantification of soil TPH would enable timely and cost-effective decisions for site management and remediation. Petroleum hydrocarbons in the soil are usually distributed in a highly heterogeneous manner (Chakraborty et al., 2012a; Franco et al., 2006), thereby necessitating analysis of a significant number of samples to sufficiently delineate the soil impacts for remedial decision-making. Consequently, the potential to streamline this process is attractive.

Visible and near-infrared reflectance spectroscopy (Vis-NIR) is a proven technology to rapidly detect and quantify TPH in soils (Chakraborty et al., 2010; Douglas, Nawar, Alamar, Mouazen, & Coulon, 2018; Forrester et al., 2013; Malley, Hunter, & Webster, 1999; Masakorala et al., 2014; Okparanma, Coulon, & Mouazen, 2014; Schwartz, Ben-Dor, & Eshel, 2012). Petroleum hydrocarbons are a complex mixture of saturated aliphatic hydrocarbons (with trace concentrations of unsaturated and polycyclic aromatic hydrocarbons) that are abundant in C-H and C-C chemical bonds. These bonds have strong diagnostic absorption bands in the mid-infrared (MIR) region from 4,000 to 400 cm^{-1} , such as symmetric and asymmetric stretch vibrations of C-H bonds in the methyl and methylene groups (Horta et al., 2015). These fundamental vibrational bands (when a vibrational mode is excited from $\nu = 0$ to $\nu \pm 1$, Banwell & McCash, 1994) have overtones (i.e., when a vibrational mode is excited from $\nu = 0$ to $\nu > 1$; Banwell & McCash, 1994) and combinational bands (i.e., combination of two or more fundamental bands; Struve, 1989) in the Vis-NIR region from 400 to 2,500 nm. Compared with MIR, these overtones and combinational bands in Vis-NIR are usually weak and overlap with spectral absorptions from other constituents

Core Ideas

- A set of petroleum hydrocarbon-containing soil samples was created in the laboratory.
- Vis-NIR TPH models were developed from laboratory set and applied to field samples.
- Four modeling techniques (PLSR, RF, ANN, and SVR) were compared.
- Spiking and spiking with extra weight were tested for model improvement.
- SVR with spiking with extra weight allowed best TPH prediction in field samples.

in soils (such as clay). Previously multivariate techniques (such as partial least squares regression [PLSR]) have been demonstrated to overcome these issues, enabling the development of empirical models to quantitatively relate TPH with Vis-NIR spectra responses at all wavelengths (Douglas, Nawar, Alamar, Coulon, & Mouazen, 2019; Viscarra Rossel & Behrens, 2010).

Malley et al. (1999) undertook a feasibility study to determine whether NIR spectrometry could be used to predict TPH in diesel-impacted soils. They used multiple linear regressions to develop NIR calibrations and found R^2 of .68 and .72. In another study, the researchers conducted a series of experiments to investigate the utility of Vis-NIR to rapidly identify crude oil-impacted soils (Caddo silt loam [fine-silty, siliceous, active, thermic Typic Glossaqualfs], Guyton silt loam [fine-silty, siliceous, active, thermic Typic Glossaqualfs], Malbis fine sandy loam [fine-loamy, siliceous, subactive, thermic Plinthic Paleudults], and Ruston fine sandy loam [fine-loamy, siliceous, semiactive, thermic Typic Paleudults]) in Louisiana (Chakraborty et al., 2010, 2012a, 2012b). They also studied the spectral variability of PHC-contaminated soils and mapped the variation of TPH concentration in an 80-ha field using Vis-NIR. Under laboratory conditions, Schwartz et al. (2012) created 30 soil samples with three soil types (Typic Xerofluven, Typic Xerocherept, and Typic Chromoxerert) in Israel and mixed the uncontaminated samples with diesel, kerosene, and 95% octane at targeted concentrations from 0 to 12,000 mg kg^{-1} . Partial least squares regression models were developed and compared with the TPH analysis from three certified laboratories that used Method EPA 418.1 to estimate TPH. In this method, the sample is acidified to a low pH (<2) and serially extracted with fluorocarbon-113 in a separatory funnel followed by the removal of interferences with silica gel adsorbent. Then, the infrared analysis of the extract is performed by direct comparison with standards to obtain TPH content.

The TPH estimates from Vis–NIR reflectance spectroscopy were in good agreement with the commercial laboratories (45, 25, and 60% deviation for the different laboratory estimations and 45% deviation for Vis–NIR spectroscopy method), making the Vis–NIR models a viable field-screening method for TPH. Forrester et al. (2013) tested whether diffuse reflectance spectroscopy in both NIR and MIR regions could be used to accurately estimate soil (Agrixeroll and Calcixerollic Xerochrept soils) TPH concentration. Both laboratory-spiked and field samples were used. They found excellent correlation between laboratory-measured and Vis–NIR-estimated TPH ($R^2 = .92$).

In summary, use of laboratory-constructed TPH-contaminated soil samples to predict for field samples is sporadically reported in the literature. Although the ability of nonlinear modeling techniques to outperform the state-of-art technique, PLSR, has been demonstrated in soil spectroscopy (Viscarra Rossel & Behrens, 2010; Wijewardane, Ge, Wills, & Loecke, 2016), a comparison study with multiple nonlinear modeling techniques was not reported for TPH estimation. When the number of field samples for model calibration are limited due to cost or access restrictions, laboratory-constructed samples can be spiked with extra weight (Guerrero et al., 2014) to calibrate models that can be applied for field samples.

The overall hypothesis of this study is that Vis–NIR TPH models developed from laboratory-constructed samples can be used to estimate TPH from Vis–NIR spectral responses of field soil samples. This is an important method optimization aspect because the most time-consuming and expensive part of developing a Vis–NIR method for accurately estimating TPH concentration is generating a site-specific sample set for calibrating chemometric models. Therefore, if the resultant data are of sufficient quality, creation of laboratory-based calibration samples could provide significant time and cost savings. The three specific objectives of this study were (a) to construct a set of PHC-containing samples by mixing PHCs with uncontaminated soil samples under laboratory conditions, (b) to develop Vis–NIR TPH estimation models from the laboratory-constructed samples and apply them to field samples, and (c) to compare two different strategies to improve model performance—spiking and spiking with extra weight.

2 | MATERIALS AND METHODS

2.1 | Construction of petroleum hydrocarbon-containing laboratory soil samples

A total of 203 PHC-containing soil samples were constructed in the laboratory. Uncontaminated soil samples

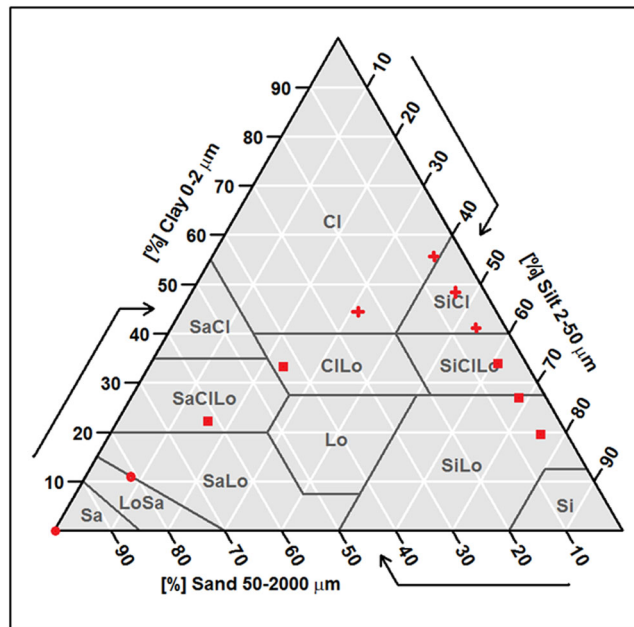


FIGURE 1 Eleven textural categories of the laboratory-constructed soil samples mapped on the USDA soil classification triangle. These 11 textural categories were divided into three classes. Sand class is denoted by circles, silt class is denoted by squares, and clay class is denoted by crosses. Cl, clay; Sa, sand; Lo, loam; Si, silt

were made by mixing ground sand, a silt loam, and a clay at different proportions to achieve 11 textural categories (Figure 1). The silt loam (14% sand, 68% silt, and 18% clay) and clay (2% sand, 28% silt, and 70% clay) were collected from a farm at Louisiana State University AgCenter. The sand (95% sand, 2% silt, and 3% clay) was purchased from a local store.

The construction of PHC-containing soil samples was carried out by mixing the uncontaminated soil samples with a mid-range crude oil at predetermined mass ratios. The targeted TPH concentrations ranged from 0 to 100,000 mg kg⁻¹ (10%), with a number of intermediate levels (see the supplemental material) to facilitate development of Vis–NIR models to estimate TPH concentration. The highest concentration of 100,000 mg kg⁻¹ (10%) was based on our experience that PHC-impacted field samples rarely exceed this TPH concentration.

The following procedures were used to construct PHC impacted soil samples. First, crude oil was dissolved in dichloromethane (DCM) to make standard solutions at concentrations of 0.25, 1.25, 2.5, 5, 12.5, 25, 125, and 250 g L⁻¹. The DCM solution was then added at a predetermined ratio to the soil and mixed thoroughly to achieve a homogenized sample with the targeted TPH concentration. The homogenized samples were then placed under a ventilation hood for ~5 h and stirred three to four times to expedite the evaporation of DCM, and “weather” the oil to develop a mixture more representative of field

TABLE 1 Summary statistics of total petroleum hydrocarbon concentrations in the three sample sets

Dataset	No. of samples	Min.	First quartile	Median	Mean	Third quartile	Max.	Skewness	Kurtosis
mg kg ⁻¹									
Laboratory	203	42	360	2,900	3,778	2,400	30,000	3.37	13.32
Field 1	107	0	2,600	2,600	6,781	14,000	45,000	1.79	8.43
Field 2	63	6	66	740	6,577	12,050	23,000	1.02	2.04

conditions. A numeric example of constructing a sample with TPH of 10,000 mg kg⁻¹ is provided in the supplemental material. Although synthetic weathering in the laboratory can be different from natural weathering of contaminated soils and have differing effects on oil constituents, TPH measurements can still be used as a single reliable parameter for crude oil contamination estimations as used in Bejarano, Levine, and Mearns (2013), Dincer Kirman et al. (2016), and Trindade, Sobral, Rizzo, Leite, and Soriano (2005). All laboratory-constructed soil samples were sealed in air-tight glass jars to prevent further hydrocarbon volatilization and stored at 4 °C. Upon analysis, each sample was divided into two roughly equal parts, one for Vis-NIR spectral scanning and the other part for TPH analysis with GC-FID (gas chromatography with a flame ionization detector) in a commercial laboratory. Method USEPA SW-846 3350B was used for extraction, and Method 8015B was used for GC-FID analysis (see Supplemental Figure S1). Method USEPA SW-846 3350B is an ultrasonic extraction method where the sample was mixed with anhydrous sodium sulfate to form a free-flowing powder. Then, the mixture was extracted with solvent three times, using ultrasonic extraction followed by separation from the sample by vacuum filtration. Per Method 8015B, the sample extract was analyzed by capillary gas chromatography using flame ionization detection (HP 5890 with hydrogen at constant flow of 10 ml min⁻¹ in constant flow mode as the carrier, 320 °C detector temperature, N₂ as the makeup gas, and 1-μl injection size). Quantitation in the analysis specific range was performed by comparing total peak area of the sample pattern with the total area of C8 to C40 hydrocarbons and No. 2 fuel oil.

2.2 | Field samples

Two field sample sets were collected for use in this study. The first field set (referred to as F1) consisted of 107 soil samples impacted with weathered crude oil. The second field set (F2) consisted of 63 soil samples collected from a sand beach in Pensacola, FL, and Grand Isle National Park (Grand Isle, LA) in June and September of 2010 respectively, when the areas were impacted

by British Petroleum's Deepwater Horizon oil spill (Kerr, 2010). The recorded spill contamination levels were as high as 500,000 mg kg⁻¹ (Lin & Mendelsohn, 2012). The samples were randomly collected, homogenized, sealed in glass jars, and stored at 4 °C until analysis. The same approach was used to analyze the reference TPH value for the field samples as the laboratory-constructed samples. Table 1 shows the summary statistics of TPH for the three sample sets.

2.3 | Vis-NIR spectral scanning

Both the laboratory-constructed and field samples were scanned with an ASD spectroradiometer (formerly Analytical Spectral Devices, now part of Malvern PANalytical and Spectris). The spectral range of the instrument is from 350 to 2,500 nm, with a spectral sampling interval of 1 nm. To prevent further volatilization of PHCs, these samples were handled gently but quickly during spectral scanning. About 20 g of the sample was placed on a flat lid and scanned with ASD's contact probe. Before each scanning, the instrument and the light source were allowed to warm up for at least 1 h to stabilize the output intensity of the light source and the instrument dark current. A 99% calibrated reflectance Spectralon panel (Labsphere) was used for white referencing. White reference scans were taken every 15 min during sample scanning. To improve the signal-to-noise ratio, each output scan by the instrument was set as the average of 15 instantaneous scans. Each sample was scanned with the contact probe twice, with a 90° turn between the two scans to partially account for directional effect of diffuse reflectance.

2.4 | Spectral preprocessing and total petroleum hydrocarbon modeling

All spectra were preprocessed with averaging along the wavelengths using a window size of 10 nm to reduce the dimensionality of spectral data. Averaging also reduced the computational time for spectroscopic modeling and prevented overfitting caused by the need to estimate too

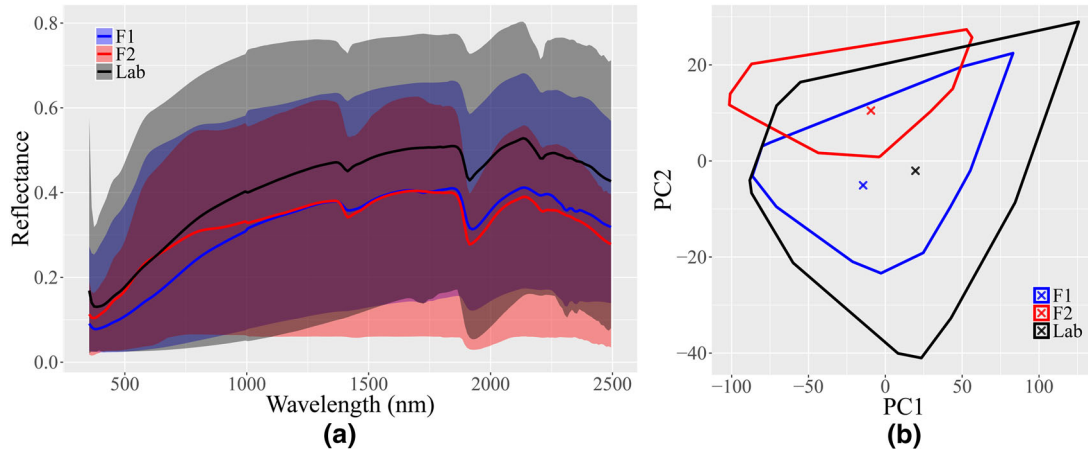


FIGURE 2 (a) Mean spectra of the laboratory-constructed and two field sample sets and their ranges (bounding boxes of the minimum and maximum value at each wavelength), and (b) convex hulls derived from the first and second principal component (PC) scores of the spectra

many model parameters (such as in artificial neural networks [ANN]). All the modeling techniques were implemented on the laboratory sample set, and the calibrated models were used to predict TPH for the F1 and F2 field sample sets.

Four modeling techniques—namely, PLSR, ANN, random forests (RF), and Support Vector Regression (SVR)—were used to calibrate models from the laboratory samples. Partial least squares regression (PLSR) is a linear modeling approach and is widely used in multivariate spectroscopic modeling, whereas ANN, RF, and SVR are nonlinear (or data mining) techniques that are less restrictive on the distributional assumptions of data and often lead to better modeling performance over PLSR (Wijewardane et al., 2016). For each modeling technique, different tuning parameters were tested to develop a sequence of models (explained in the paragraph below), and cross-validation with 10-fold random segments was used to select the best model with the lowest RMSE in cross-validation ($RMSE_{CV}$).

For PLSR, the number of latent variables (n_{LV}) ranged from 1 to 30. For ANN, the number of nodes in the hidden layer was allowed to change from three to nine and decay of weight at 0.01, 0.1, and 0.3. For RF, the number of variables randomly sampled as candidates at each tree node split (m_{try}) was tested between 10 and 114. For SVR, a grid search was carried out for severity of the violations to the margin (C) from 8 to 80, and inverse kernel width for radial basis kernel function (σ) at 0.0005 and 0.001.

The performance of these regression models were evaluated by calculating R^2 (coefficient of determination between predicted and reference values in the validation

set), RMSE, ratio of performance to deviation (RPD), and ratio of performance to inter-quartile range (RPIQ; Bellon-Maurel, Fernandez-Ahumada, Palagos, Roger, & McBratney, 2010).

2.5 | Spiking and spiking with extra weight

Thirty samples were randomly selected from each field sample set (F1 and F2) to implement (a) spiking (i.e., incorporation of the selected samples for the model calibration library) and (b) spiking with extra weight (i.e., incorporation of the selected samples with repetition to match the number of calibration samples). The selected samples were incorporated to the laboratory sample set and the regression models were recalibrated (using the same procedures as described in Section 2.4). The new models were then applied to the remaining samples in the two field sample sets to evaluate the model performance. Similarly, for spiking with extra weight, 30 selected samples from field dataset were replicated six times (to achieve more balanced numbers of field vs. laboratory samples in model calibration) and then incorporated to the laboratory sample set. Models were calibrated for the combined dataset and used to predict for the corresponding field set to evaluate model performances. A more detailed account of spiking and spiking with extra weight can be found in Guerrero et al. (2014). It should be noted that, although this approach does require some sample collection from the field, the burden is significantly reduced when starting with the laboratory-based calibration (i.e., ~200 samples are needed to build an initial calibration).

TABLE 2 Validation results of the visible and near-infrared (Vis–NIR) total petroleum hydrocarbon models calibrated on the laboratory-constructed sample set and tested on the F1 and F2 field sample set

Validation set	Modeling technique ^a	R^2	mg kg ⁻¹		RPD ^b	RPIQ ^c
			RMSE	Bias		
F1	PLSR	.53	14,857	7,734	1.08	1.43
	ANN	.61	12,928	−6,127	1.24	1.65
	RF	.67	10,830	−3,786	1.48	1.97
	SVR	.72	9,507	−3,819	1.69	2.24
F2	PLSR	.13	19,125	1,389	1.01	0.23
	ANN	.03	19,445	−4,315	0.99	0.23
	RF	.03	21,268	−3,823	0.91	0.21
	SVR	.01	24,433	−13,025	0.79	0.18

^aPLSR, partial least squares regression; ANN, artificial neural networks; RF, random forests; SVR, support vector regression. ^bRPD, ratio of performance to deviation. ^cRPIQ, ratio of performance to interquartile range.

3 | RESULTS AND DISCUSSION

3.1 | Spectral variations among the laboratory-constructed and field sample sets

Figure 2a shows the mean Vis–NIR spectra of the laboratory and field sample sets, as well as the range of the spectra at each wavelength. Principal component analysis was also conducted on spectra data, and the first two principal components were used to draw the convex hull plot for the three sample sets (Figure 2).

The mean spectra of the laboratory-constructed and F1 sample set showed similar shapes along the wavelengths, whereas the F2 sample set deviated from the others in the region of 350–1,000 nm. Laboratory samples appeared to have the largest variation in the spectral domain, as indicated by the larger bounding box along the wavelengths in Figure 2a and the larger convex hull in Figure 2b. Furthermore, the convex hull of F1 was completely contained in that of the laboratory sample set, whereas F2 was only partially contained. Overall, Figure 2 indicated a high spectral similarity between F1 and the laboratory set, whereas F2 was spectrally more dissimilar. This spectral pattern could be explained by the textural properties of the three sample sets: both the laboratory and F1 set were predominantly silty or loamy, whereas the F2 set had predominantly sandy samples (which were collected from two sand beaches).

3.2 | Comparison of the four modeling techniques

As expected, validation of the Vis–NIR TPH models calibrated with the laboratory-constructed sample set on the

field sets showed large variation in performance (Table 2). The prediction of TPH on the F1 sample set was moderately accurate, with R^2 ranging from .53 to .72 and RPD from 1.08 to 1.69. This may be due to the differences in sample preparation or composition. For example, the calibration laboratory samples were prepared under the laboratory conditions by mixing ground sand, silt, and clay at different proportions. In contrast, the field samples contain natural soils, which likely have a more complex composition (e.g., additional organic matter) that may cause spectral interferences. Moreover, the field samples were impacted with hydrocarbons that were weathered under the natural conditions that differ from the synthetic weathering conducted in the laboratory. Consequently, a model calibrated on laboratory data will have higher prediction errors on field data, particularly when the differences between the soil and/or oil type among the sample sets become more pronounced.

Using the models developed from the laboratory samples, TPH prediction on the F2 sample set was very poor, with $R^2 < .13$ and RPD < 1.01 for all four modeling techniques. This distinct difference of the model performance on F1 and F2 was in agreement with the degree of spectral dissimilarity among the three sample sets as visualized in Figure 2. The laboratory-constructed and F1 sets were spectrally more similar, and the model performance was therefore better when applied to F1. Applying the Vis–NIR TPH model to F2 meant extrapolation in the spectral domain, which would explain the poor prediction of F2. When the different modeling techniques were compared in F1 prediction, SVM showed the best performance, followed by RF, ANN, and PLSR. This was also consistent with the literature, indicating superior modeling performance by nonlinear modeling techniques (SVM, RF and ANN) to the linear one (PLSR) (Viscarra Rossel & Behrens, 2010; Wijewardane et al., 2016).

TABLE 3 Validation results of the visible and near-infrared (Vis–NIR) total petroleum hydrocarbon models calibrated on the laboratory-constructed sample set with two model improvement schemes: spiking and spiking with extra weight

Improvement technique	Validation set	Modeling technique ^a	R^2	—mg kg ⁻¹ —			
				RMSE	Bias	RPD ^b	RPIQ ^c
Spiking	F1	PLSR	.66	9,810	390	1.69	2.17
		SVR	.79	8,208	-2,191	2.02	2.59
	F2	PLSR	.63	14,239	5,451	1.28	0.30
		SVR	.69	11,525	-1,866	1.58	0.37
Spiking with extra weight	F1	PLSR	.78	7,826	325	2.12	2.72
		SVR	.88	5,661	-577	2.92	3.75
	F2	PLSR	.64	13,964	6,089	1.30	0.30
		SVR	.74	10,377	827	1.76	0.41

^aPLSR, partial least squares regression; SVR, support vector regression. ^bRPD, ratio of performance to deviation ^cRPIQ, ratio of performance to interquartile range.

3.3 | Comparison of model improvement techniques

Table 3 shows that, with spiking and spiking with extra weight, the performances of the Vis–NIR TPH models on both the F1 and F2 field sample sets were improved. Only the results from PLSR and SVR are shown in Table 3 because they were the two extremes in accuracy among the original models (Table 2), and the results from RF and ANN followed the same trend. For the F1 sample set and PLSR, the validation R^2 increased from .53 to .66 (spiking) and .78 (spiking with extra weight), whereas RMSE decreased from 14,857 to 9,810 (spiking) and 7,826 mg kg⁻¹ (spiking with extra weight). The same trend was also observed with SVR.

The improvement on the F2 test was more pronounced. With PLSR, the R^2 increased from .13 to .63 (spiking) and .64 (spiking with extra weight), and RMSE decreased from 19,125 to 14,239 (spiking) and 13,964 mg kg⁻¹ (spiking with extra weight). The SVR models exhibited even greater improvement, with R^2 increased from .01 to .69 (spiking) and .74 (spiking with extra weight), and RMSE decreased from 24,433 to 11,525 (spiking) and 10,377 mg kg⁻¹ (spiking with extra weight). Overall, these results suggest that using a laboratory-developed sample set, augmented with a small number of field samples, may provide a time and cost-effective approach for rapid field screening of TPH in soils.

In chemometric modeling, RPD is commonly used as a criterion for model usability. Prediction models with RPD < 1.4 are regarded as poor, RPD between 1.4 and 2.0 are fair models with potential for improvement, and RPD > 2.0 are stable and accurate models (Chakraborty et al., 2010). By this criterion, the PLSR model applied to F1 was improved from poor (RPD = 1.08) to fair (RPD = 1.69)

and accurate (RPD = 2.12), with spiking and spiking with extra weight, respectively. The modeling improvement on the F2 field set was also substantial. For example, for SVR, RPD increased from 0.79 to 1.58 and 1.76, elevating the model from “poor” to “fair with potential for improvement.” To provide visual assessment of model prediction, the scatterplots of Vis–NIR-predicted versus measured TPH in F1 samples are shown in Figure 3, with PLSR and SVR modeling technique, as well as spiking and spiking with extra weight for model improvement.

In practical settings, remediation requirements for addressing PHC-containing soils are based on a risk-based concentration screening criterion. For example, some regulatory policies require remediation to reduce TPH concentrations when they exceed 1% (i.e., 10,000 mg kg⁻¹), which corresponds to federal regulations of hazard materials (such as Occupational Safety and Health Administration Regulations 29 CFR §1910.1200; U.S. Department of Labor, 2012). To assess the utility of these models for making field decisions, we created classification scenarios above and below the remediation threshold. In this case, knowing the exact concentration of TPH is not needed. Instead, samples only need to be placed into one of the two classes to effectively guide remediation decision (Figure 3).

For each scatterplot, vertical and horizontal dashed lines at 10,000 mg kg⁻¹ were drawn to divide the plot into four quadrants. Points in upper right and lower left quadrants (colored black) indicated that their laboratory-measured and Vis–NIR-predicted TPH values fell into the same categories and correct detections were made. Points in upper left (colored blue) and lower right (colored red) were incorrectly detected, and they presented the false positive and false negative errors made by Vis–NIR prediction. As a baseline of comparison, the laboratory PLSR TPH model

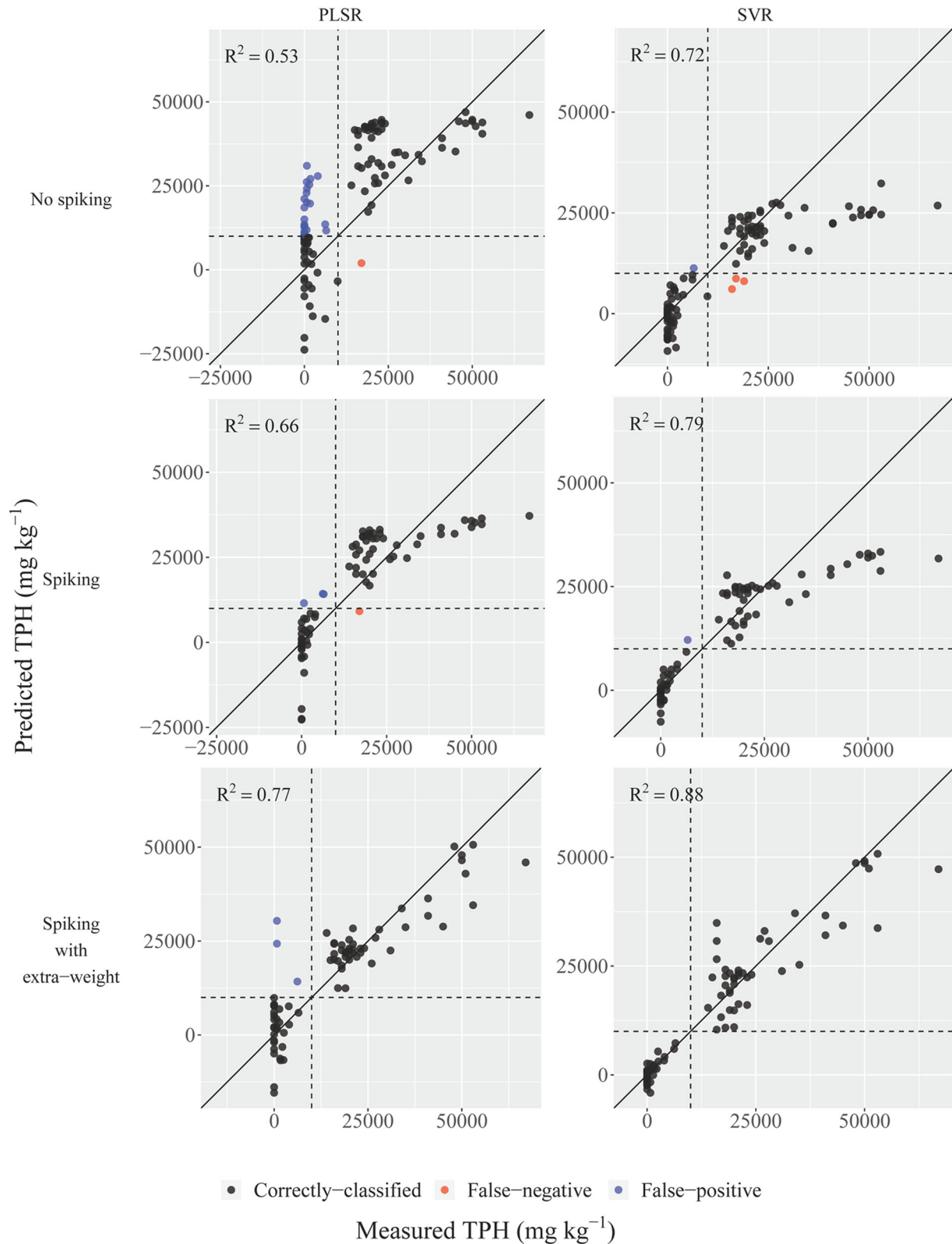


FIGURE 3 Scatterplots of the laboratory-measured vs. visible and near-infrared (Vis-NIR)-predicted total petroleum hydrocarbons (TPH) for the F1 field set. The left and right columns of the panel were the partial least squares regression (PLSR) and support vector regression (SVR) models, respectively. First, second, and third rows of the panel show the results without spiking, with spiking, and with spiking with extra weight, respectively. R^2 values are shown in Figure 3, and other model statistics are provided in Tables 2 and 3. The dashed lines represent a decision criterion of 1% TPH

directly applied to the F1 samples caused quite a few misclassifications, with the overall accuracy of 78.5% (including both false positive and false negative). The SVR models improved the classification accuracy, so also spiking and spiking with extra weight. Most notably, the SVR model together with spiking with extra weight achieved 100% classification accuracy.

Visible and near-infrared spectral modeling for soil TPH quantification is a data-driven approach. It relies on the development of calibration models requiring the collection and analysis of PHC-containing soil samples from the field. This limits the practical use of Vis-NIR, because developing a calibration set for each and every field campaign can be labor intensive and costly. The results of our study suggested that it is feasible to use laboratory-constructed PHC soil samples for calibrating Vis-NIR models and subsequently predicting TPH in field samples. This could expand the practical use of Vis-NIR for soil TPH quantification, as efforts to build the calibration set could be greatly reduced.

However, laboratory-constructed soil samples and field samples can vary in many aspects: intrinsic soil physical and chemical properties, moisture content, structure and aggregation, all of which introduce significant variability to Vis-NIR spectra in addition to PHC type and concentration. For this reason, Vis-NIR models developed from laboratory samples may not perform well on field samples, in particular when there was large spectral difference between the two sets (this is the case for F2 prediction in our study). The spectral difference observed in this study was mainly caused by the difference in the textural properties among the sample sets. But more practically, many other factors could also give this spectral dissimilarity (e.g., moisture content, organic content, oil type). Spiking solved this problem effectively, by incorporating a small number of field samples to the laboratory samples to form a spiked calibration set. Spiking with extra weight made additional copies of the field samples such that the numbers of laboratory and field samples in the calibration set are roughly equal, therefore increasing the influence of field samples in model calibration to improve prediction.

The approach in this study presented a potentially rapid and cost-effective method for TPH analysis of field soil samples. A laboratory PHC-containing soil sample set can be constructed as a base calibration set. For each field campaign, a small set of field samples should be analyzed with the reference TPH method and incorporated into the base set to develop a Vis-NIR calibration specific for that field. The need to analyze field samples to build Vis-NIR calibration is therefore greatly reduced. Our results showed that this approach was successful in predicting TPH of two

field sample sets. The results were particularly satisfactory when SVR was used along with spiking with extra weight for model calibration.

4 | CONCLUSIONS

In this paper, we successfully demonstrated the use of laboratory-constructed PHC-containing soil samples to predict TPH concentration of field samples. This approach was different from previous research of TPH prediction using Vis-NIR spectroscopy, and indicated a more cost-effective framework for rapid detection and quantification of TPH in field samples. The major conclusions drawn from this study are as follows:

1. Among the four multivariate modeling techniques tested, SVR showed the best performance in predicting soil TPH, followed by ANN, RF, and PLSR. The superiority of nonparametric modeling (SVR, ANN, and RF) to parametric modeling (PLSR) was due to highly skewed distribution of TPH in all three sample sets.
2. Models developed from the laboratory samples predicted TPH in one field (F1) with moderate accuracy, but poor in the other field (F2). This was explained by the larger spectral difference between the laboratory and the F2 sample set, which was attributed to the textural and mineralogical differences between the two sets.
3. Both spiking and spiking with extra weight improved the TPH prediction accuracy for F1 and F2. Improvement for F2 was most pronounced with SVR along with spiking with extra weight.
4. The Vis-NIR TPH prediction models demonstrated the capability to accurately classify field samples into one of the two categories using $10,000 \text{ mg kg}^{-1}$ as a threshold, which suggested a rapid and cost-effective TPH detection method for remediation decision-making.

ACKNOWLEDGMENTS

We would like to thank the support from Chevron Energy Technology Company for this work.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Nuwan K. Wijewardane  <https://orcid.org/0000-0001-8962-9451>

David C. Weindorf  <https://orcid.org/0000-0002-3814-825X>

REFERENCES

- Banwell, C. N., & McCash, E. M. (1994). *Fundamentals of molecular spectroscopy* (Vol. 851). New York: McGraw-Hill.
- Bejarano, A. C., Levine, E., & Mearns, A. J. (2013). Effectiveness and potential ecological effects of offshore surface dispersant use during the Deepwater Horizon oil spill: A retrospective analysis of monitoring data. *Environmental Monitoring and Assessment*, *185*, 10281–10295. <https://doi.org/10.1007/s10661-013-3332-y>
- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., & McBratney, A. B. (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends in Analytical Chemistry*, *29*, 1073–1081. <https://doi.org/10.1016/j.trac.2010.05.006>
- Chakraborty, S., Weindorf, D. C., Morgan, C. L. S., Ge, Y., Galbraith, J. M., Li, B., & Kahlon, C. S. (2010). Rapid identification of oil-contaminated soils using visible near-infrared diffuse reflectance spectroscopy. *Journal of Environmental Quality*, *39*, 1378–1387. <http://doi.org/10.2134/jeq2010.0183>
- Chakraborty, S., Weindorf, D. C., Zhu, Y., Li, B., Morgan, C. L. S., Ge, Y., & Galbraith, J. (2012a). Assessing spatial variability of soil petroleum contamination using visible near-infrared diffuse reflectance spectroscopy. *Journal of Environmental Monitoring*, *14*, 2886–2892. <http://doi.org/10.1039/c2em30330b>
- Chakraborty, S., Weindorf, D. C., Zhu, Y., Li, B., Morgan, C. L. S., Ge, Y., & Galbraith, J. (2012b). Spectral reflectance variability from soil physicochemical properties in oil contaminated soils. *Geoderma*, *177*–*178*, 80–89. <https://doi.org/10.1016/j.geoderma.2012.01.018>
- Dincer Kirman, Z., Sericano, J. L., Wade, T. L., Bianchi, T. S., Marcantonio, F., & Kolker, A. S. (2016). Composition and depth distribution of hydrocarbons in Barataria Bay marsh sediments after the Deepwater Horizon oil spill. *Environmental Pollution*, *214*, 101–113. <https://doi.org/10.1016/j.envpol.2016.03.071>
- Douglas, R. K., Nawar, S., Alamar, M. C., Coulon, F., & Mouazen, A. M. (2019). The application of a handheld mid-infrared spectrometry for rapid measurement of oil contamination in agricultural sites. *Science of the Total Environment*, *665*, 253–261. <https://doi.org/10.1016/j.scitotenv.2019.02.065>
- Douglas, R. K., Nawar, S., Alamar, M. C., Mouazen, A. M., & Coulon, F. (2018). Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-NIR spectroscopy and regression techniques. *Science of the Total Environment*, *616*–*617*, 147–155. <https://doi.org/10.1016/j.scitotenv.2017.10.323>
- Forrester, S. T., Janik, L. J., McLaughlin, M. J., Soriano-Disla, J. M., Stewart, R., & Dearman, B. (2013). Total petroleum hydrocarbon concentration prediction in soils using diffuse reflectance infrared spectroscopy. *Soil Science Society of America Journal*, *77*, 450–460. <http://doi.org/10.2136/sssaj2012.0201>
- Franco, M. A., Viñas, L., Soriano, J. A., De Armas, D., González, J. J., Beiras, R., ... Albaigés, J. (2006). Spatial distribution and ecotoxicity of petroleum hydrocarbons in sediments from the Galicia continental shelf (NW Spain) after the Prestige oil spill. *Marine Pollution Bulletin*, *53*, 260–271. <http://doi.org/10.1016/j.marpolbul.2005.10.004>
- Guerrero, C., Stenberg, B., Wetterlind, J., Viscarra Rossel, R. A., Maestre, F. T., Mouazen, A. M., ... Kuang, B. (2014). Assessment of soil organic carbon at local scale with spiked NIR calibrations: Effects of selection and extra-weighting on the spiking subset. *European Journal of Soil Science*, *65*, 248–263. <https://doi.org/10.1111/ejss.12129>
- Horta, A., Malone, B., Stockmann, U., Minasny, B., Bishop, T. F. A., McBratney, A. B., Pallasser, R., & Pozza, L. (2015). Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. *Geoderma*, *241*–*242*, 180–209. <http://doi.org/10.1016/j.geoderma.2014.11.024>
- Kerr, R. A. (2010). A lot of oil on the loose, not so much to be found. *Science*, *329*, 734–735. <https://doi.org/10.1126/science.329.5993.734>
- Lin, Q., & Mendelsohn, I. A. (2012). Impacts and recovery of the Deepwater Horizon oil spill on vegetation structure and function of coastal salt marshes in the northern Gulf of Mexico. *Environmental Science & Technology*, *46*, 3737–3743. <https://doi.org/10.1021/es203552p>
- Malley, D. F., Hunter, K. N., & Webster, G. R. B. (1999). Analysis of diesel fuel contamination in soils by near-infrared reflectance spectrometry and solid phase microextraction-gas chromatography. *Journal of Soil Contamination*, *8*, 481–489. <http://doi.org/10.1080/10588339991339423>
- Masakorala, K., Yao, J., Chandankere, R., Liu, H., Liu, W., Cai, M., & Choi, M. M. F. (2014). A combined approach of physicochemical and biological methods for the characterization of petroleum hydrocarbon-contaminated soil. *Environmental Science and Pollution Research*, *21*, 454–463. <http://doi.org/10.1007/s11356-013-1923-3>
- Okparanma, R. N., Coulon, F., & Mouazen, A. M. (2014). Analysis of petroleum-contaminated soils by diffuse reflectance spectroscopy and sequential ultrasonic solvent extraction-gas chromatography. *Environmental Pollution*, *184*, 298–305. <http://doi.org/10.1016/j.envpol.2013.08.039>
- Schwartz, G., Ben-Dor, E., & Eshel, G. (2012). Quantitative analysis of total petroleum hydrocarbons in soils: Comparison between reflectance spectroscopy and solvent extraction by 3 certified laboratories. *Applied and Environmental Soil Science*, *2012*, 1–11. <http://doi.org/10.1155/2012/751956>
- Struve, W. S. (1989). *Fundamentals of molecular spectroscopy*. New York: Wiley.
- Trindade, P. V. O., Sobral, L. G., Rizzo, A. C. L., Leite, S. G. F., & Soriano, A. U. (2005). Bioremediation of a weathered and a recently oil-contaminated soils from Brazil: A comparison study. *Chemosphere*, *58*, 515–522. <https://doi.org/10.1016/j.chemosphere.2004.09.021>
- U.S. Department of Labor. (2012). *Hazard Communication: 1910.1200*. U.S. Department of Labor, Occupational Safety and

Health Administration. Retrieved from https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=standards&p_id=10099

Viscarra Rossel, R. A., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158, 46–54. <https://doi.org/10.1016/j.geoderma.2009.12.025>

Wijewardane, N. K., Ge, Y., Wills, S., & Loecke, T. (2016). Prediction of soil carbon in the conterminous United States: Visible and near infrared reflectance spectroscopy analysis of the Rapid Carbon Assessment Project. *Soil Science Society of America Journal*, 80, 973–982. <https://doi.org/10.2136/sssaj2016.02.0052>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Wijewardane NK, Ge Y, Sihota N, Hoelen T, Miao T, Weindorf DC. Predicting total petroleum hydrocarbons in field soils with Vis–NIR models developed on laboratory-constructed samples. *J. Environ. Qual.* 2020;1–10. <https://doi.org/10.1002/jeq2.20102>