Shrestha, G., Calvelo Pereira, R., Anderson, C., Jeyakumar, P., Poggio, M., and Kereszturi, G. 2020. Predicting cadmium concentration in New Zealand agricultural soil using mid-infrared spectroscopy. In: *Nutrient Management in Farmed Landscapes*. (Eds. C.L. Christensen, D.J. Horne and R. Singh). http://flrc.massey.ac.nz/publications.html. Occasional Report No. 33. Farmed Landscapes Research Centre, Massey University, Palmerston North, New Zealand. 12 pages.

PREDICTING CADMIUM CONCENTRATION IN NEW ZEALAND AGRICULTURAL SOIL USING MID-INFRARED SPECTROSCOPY

Gautam Shrestha[1]*, Roberto Calvelo Pereira[1], Chris Anderson[1], Paramsothy Jeyakumar[1], Matteo Poggio[2], and Gabor Kereszturi[1]

[1]Environmental Sciences Group, School of Agriculture and Environment, Massey University, Manawatu Campus
[2]Manaaki Whenua-Landcare Research, Palmerston North
**Email: g.shrestha@massey.ac.nz*

Abstract

Spectroscopy based soil analysis have gained popularity over traditional wet chemistry methods in the recent years. Wet chemistry techniques are precise, but highly technical and expensive while being time consuming. Spectroscopy based analysis techniques are cheap, fast and easy to use while maintaining reasonable accuracy. In the context of increased concern of soil cadmium (Cd) level in New Zealand agricultural soil requiring regular monitoring, this research used mid-infrared (MIR: 7498 to 600 cm^{-1}) spectroscopy to develop a robust statistical model to accurately predict agricultural soil Cd concentration.

Eighty-seven topsoil (0-15 cm depth) samples obtained from 30 dairy farms were scanned using MIR spectroscopy; soil characterisation was also done through traditional wet chemistry methods. Data were used to develop spectroscopy based statistical models to predict soil total Cd concentration. Two spectroscopic data pre-processing techniques including first derivative with Savitzky-Golay smoothing and continuum removal, and two machine learning algorithms including partial least square (PLS) and random forest (RF) regression were tested to generate meaningful calibration and validation models.

Soil total Cd concentration ranged between 0.10 and 2.03 mg Cd/ kg soil, and higher Cd values were characteristically found in Allophanic soils (0.76 mg Cd/kg) than non-allophanic soils (0.35 mg Cd/kg). Soil total Cd was significantly positive in correlation (r = 0.77) with total phosphorus (P). Spectral data pre-processing using first derivative transformation with Savitzky-Golay smoothing improved the outcome of both regression models than continuum removal. PLS regression validation model predicted soil total Cd variations with relatively high coefficient of determination (R^2_{Val}=0.72) and ratio of performance to deviation (RPD_{Val} = 1.82) and relatively low root mean square error ($RMSE_{Val}$=0.12 mg Cd/kg). RF-based validation model showed less ideal performance (R^2_{Val}=0.64, RPD = 1.60, and $RMSE_{Val}$= 0.14 mg Cd/kg). These results indicated that MIR spectroscopy based soil Cd dry analysis can help agricultural soil Cd monitoring cheap and fast contributing to the effective management.

Keywords: mid-infrared spectroscopy, soil total cadmium, wet chemistry, pastoral soil, machine learning algorithms

Introduction

Trace elements are present in the soil naturally in small amounts and some of them are required for plant nutrition (Marschner, 2012). However, all trace elements at high concentration are potentially toxic in nature (Hooda, 2010). In particular, cadmium (Cd) is present in almost all soil types in the world with background natural concentration of soil total Cd upto 0.6 mg Cd/kg in the case of New Zealand (Cavanagh et al., 2015). However, continuous use of phosphorus (P) fertiliser for plant production containing traces of Cd (Abraham, 2018) has unintentionally augmented Cd concentrations in agricultural soils during last century. Plants can uptake and accumulate the soil Cd in their root and shoot parts (Cavanagh et al., 2019). Consequently, pasture grazing management could expose animals to dietary accumulation of Cd with time, raising health (Godt et al., 2006) and trade concerns (Kim, 2005).

Europe and Canada have monitored and regulated these high Cd levels in agricultural soils, establishing soil guideline values around 1.5 mg/kg soil (CCME, 1999; MIEF, 2007). In New Zealand, the National Cadmium Management Strategy (CMS) was issued in 2011 (CWG, 2011) to manage ongoing soil Cd accumulation. As a consequence of CMS, the Tiered Fertiliser Management System (TFMS) was implemented (McDowell et al., 2013) to help managing agricultural soil Cd level. With four main threshold levels of soil total Cd: 0.6, 1.0, 1.4 and 1.8 mg/kg, TFMS places limits on the amount of P fertiliser to be applied depending on actual soil total Cd concentrations (Fertiliser Association, 2019). The TFMS emphasises on an assessment of soil Cd through a detailed soil monitoring at the farm level (Fertiliser Association, 2019).

At present, soil total Cd analyses are based on reference methods including digestion using strong acids (wet chemistry) prior to the determination of Cd in the extraction (Roberts et al., 1994). Intensive soil monitoring of Cd concentrations under the TMFS using wet chemistry methods will involve time-consuming sample processing leading to increase in total analysis cost to the farmer. As an alternative, soil spectroscopy has been established as a reliable soil testing procedure (Shepherd & Walsh, 2007). Soil spectroscopy is the study of interaction of soil with electromagnetic radiation to quantify soil properties applying statistical tools (Nocita et al., 2015). Already near-infrared (NIR) spectroscopy is used to routinely quantify soil carbon and nitrogen in New Zealand (Hill Laboratories, 2019). In the end, measuring multiple soil properties is carried out faster, cheaper with high reproducibility (Poggio et al., 2018), and the real time monitoring of soil characteristics is possible using soil spectroscopy (Rouillon et al., 2017).

Soil spectroscopy has been used to quantify trace element concentration in contaminated and agricultural soils (Nawar et al., 2019). Visible-NIR (25000 – 4000 cm^{-1}) spectroscopy has been used to quantify Cd in agricultural soil (Chen et al., 2015). In the context of New Zealand pastoral soils, soil total Cd concentration was estimated using direct relation of Cd with Vis-NIR predicted soil total carbon and total nitrogen (Stafford et al., 2018). Mid-infrared (MIR: 4000 – 400 cm^{-1}) spectroscopy was also used to quantify Cd in soils with good accuracy (Soriano-Disla et al., 2014). Further development of prediction model relies on the use of representative datasets (Tasumi & Sakamoto, 2014).

Following on previous research on Cd quantification using soil spectroscopy, the objective of this work was to develop a robust soil Cd prediction model using MIR spectroscopy in conjuction with reference wet chemistry methods applied to New Zealand pastoral topsoil samples.

Materials and methods

Soil sampling and chemical analyses

Topsoil (0 – 15 cm) samples of 50 Allophanic and 37 non-allophanic soils (other soil types) were collected from 30 dairy farms by Aaron Stafford during his PhD research in 2014 (Stafford, 2017). Sub-samples of those soil samples were used for this study, and analysed following reference methods. Soil pH was measured in 1:10 soil:water extract. Acid oxalate extractable aluminium (Al_o), iron (Fe_o) and silicon (Si_o) were extracted using 1:10 soil:acid oxalate solution (Blakemore et al., 1987) then analysed using microwave-plasma atomic emission spectrometer (MP-AES 4200, Agilent, USA). Total P was extracted using 1 g soil in the 4 ml digestion mixture of concentrated sulphuric acid, potassium sulphate and selenium and analysed through autoanalyzer using nitric-vanadomolybdate acid colouring reagent (McKenzie & Wallace, 1954). Cation exchange capacity (CEC) was determined using ammonium acetate extraction (Blakemore et al., 1987) and analysed through MP-AES for cations: sodium (Na), magnesium (Mg), potassium (K) and calcium (Ca). Total Cd was extracted with 1 g soil in 10 ml concentrated nitric (HNO_3) acid (Gray et al., 1999) and analysed through graphite furnace atomic absorption spectrometer (GFAAS PinAAcle 900Z, PerkinElmer, UK). A representative subsample of 0.5 mm sieve passed soil was prepared to determine total carbon (C) and total nitrogen (N) using a Vario MACRO cube elemental system (Elementar Analysensysteme GmbH, Hanau, Germany).

Spectral measurement and model development

For MIR scanning, a representative subsample of 2 mm sieved soil was ground (below 0.5 mm diameter) and homogenised. An aluminium microtiter plate (48 wells) was filled with four replicates of each sample. For reference, along with gold embedded in the plate, three samples with known characteristics were scanned for long term consistency or drift of the machine measurements. Soil samples were scanned through a microplate reader extension for high throughput screening in infrared spectroscopy equipment (HTS-XT, Tensor II, Bruker, Germany) to get raw MIR spectra (7498 – 600 cm^{-1}) using OPUS Lab version 7 software (Bruker, Germany). Total 3578 wavebands were recorded for each sample within 40 seconds of scanning.

Rstudio version 1.2.5033 (RStudio Team, 2015) along with R version 3.6.1 (R Core Team, 2013) were used for data analysis. Raw spectra file was later converted to csv format using R code developed by Sila & Terhoeven-Uselmans (2013). Four spectral replicates of each soil samples were averaged to get a single spectral dataset for each soil sample (r package: *readxl* (Wickham et al., 2019a)). Spectra data were pre-processed using two techniques: continuum removal using r packages: *prospectr* (Stevens & Ramirez Lopez, 2014) and *resemble* (Ramirez Lopez & Stevens, 2016) and first derivative transformation with Savitzky Golay smoothing. For first derivative transformation window size of 9 bands and the order of 2 polynomial was used.

Pre-processed data were inspected for outliers using Mahalanobis distance. Samples having Mahalanobis distance score larger than 1 were considered as outlier. Three soil were removed from the analysis as they were outliers containing higher Cd concentration (1.59, 1.75 and 2.03 mg Cd/kg soil). Rest of the samples were distributed at the ratio of 80:20 resulting 68 calibration set and 16 validation set using Kennard-Stone sampling method.

Partial least square (PLS) and random forest (RF) regression algorithms were compared to develop prediction models. PLS regression is multivariate analysis with the assumption of dependent variables can be estimated using linear combination of explanatory variables. It is a sum of regression analysis, principle component analysis and correlation analysis (Adams, 2007). RF regression is an ensemble learning method which combines Brieman's bagging approach and random selection of waveband to construct multiple decision trees (Breiman, 2001). PLS regression has been used successfully to develop prediction models for soil properties using spectral measurement (Riedel et al., 2018) while others found RF more powerful in developing accurate prediction model (Wang et al., 2017a). For PLS regression model development r packages: *caret* (Kuhn et al., 2020) and *pls* (Mevik et al., 2019) were used; repeated cross validation method with five repeats was also employed. Pre-processed data were again passed through centering and scaling pre-processing. For RF regression model development r package: *randomForest* (Breiman et al., 2018) was used and 500 trees were selected based on importance of the tree.

Relationship between MIR spectral response of soil with soil total Cd and total P concentration were analysed using pearson's correlation coefficient determined between waveband values for pre-processed spectra and wet chemistry results of soil element concentration (r packages: *reshape2* (Wickham, 2017) and *ggplot2* (Wickham et al., 2019b)).

Evaluation of model performance

To evaluate the model performance, results of prediction model were compared with wet chemistry results and determined coefficient of determination (R^2), root mean square error (RMSE) and ratio of performance to deviation (RPD) (r package: chillR (Luedeling, 2019)). The R^2 explains total variability in the dependent variable possible to predict by the independent variable. R^2 has values ranged from 0 to 1; with less than 0.5 considered as unreliable prediction model and higher value means more reliable prediction model (Williams & Norris, 2001). The RMSE value is interpreted as inaccuracy of prediction model; lesser the value meaning more accurate prediction. The RPD is the ratio between standard deviation of measured value to standard error of prediction (Wang et al., 2017b). In general, a prediction model with RPD value below 1.5 is considered as a poor model, range of 1.5 – 2.0 is considered as a fair model. A prediction model with RPD value above 2.0 is considered a good model (Rossel & Webster, 2012). Hence, a prediction model with largest R^2 and RPD value with smallest RMSE value can be considered as an optimal model.

Other statistical methods

Mean, minimum and maximum were calculated and tabulated for all soil parameters analysed. Pearson's correlation coefficient for soil parameters were calculated using r package corrplot (Wei et al., 2017). Mean value of soil parameters were compared for Allophanic and non-allophanic soils using Student's two sample t-test in R.

Results and discussion

Soil chemical characteristics

Mean values of pH, total C, total N and CEC were non-significantly different between Allophanic and non-allophanic soils (Table 1). Both soil types were in acidic (pH 4.81 – 6.44) condition, and total C concentration ranged widely from 2.21 – 42.6% (Table 1). On an average, total P, oxalate extractable Al, Fe and Si concentrations were significantly ($p<0.001$) higher in Allophanic soils when compared to non-allophanic soils (Table 1). Total P concentration in Allophanic soils was higher (2.36 g P/kg soil) compared to non-allophanic soils (1.30 g P/kg soil). As expected, Allophanic soils had higher oxalate extractable Al (3.09 g Al/kg) compared to non-allophanic soils (0.78 g Al/kg soil).

Soil total Cd concentration was on an average 0.76 mg Cd/kg soil in Allophanic soils whereas only 0.35 mg Cd/kg soil in non-allophanic soils. In general, soil total Cd concentrations reported in this study were in agreement with the soil Cd concentration ranges found in similar New Zealand soils (Abraham, 2018).

Table 1 Soil chemical characteristics determined for 50 Allophanic and 37 non-allophanic topsoil samples collected from 30 New Zealand dairy farms. Results (P values) from two sample *t*-tests comparing both groups are included for each variable.

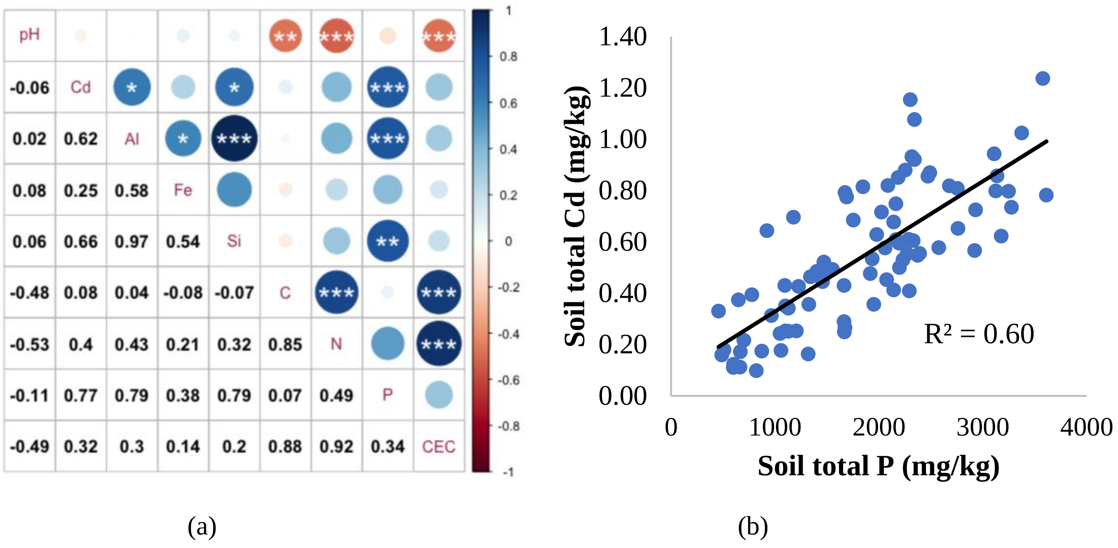
Soil characteristics	Allophanic soil (50)			Non-allophanic soil (37)			P value
	Maximum	Mean	Minimum	Maximum	Mean	Minimum	
Soil pH	6.08	5.61	4.95	6.44	5.51	4.81	0.126
Total Cd (mg/kg)	2.03	0.76	0.34	0.87	0.35	0.10	< 0.001
Total P (g/kg)	3.61	2.36	7.78	2.49	1.30	4.57	< 0.001
Total C (%)	15.0	8.57	3.18	42.6	10.5	2.21	0.215
Total N (%)	1.31	0.83	0.34	1.70	0.72	0.21	0.107
Al_o (g/kg)	5.10	3.09	0.16	1.48	0.78	0.18	< 0.001
Fe_o (g/kg)	1.47	0.83	0.31	1.30	0.52	0.15	< 0.001
Si_o (g/kg)	1.10	0.60	0.02	0.41	0.08	0.01	< 0.001
CEC (cmol/kg)	48.2	33.4	15.0	70.8	30.7	7.84	0.376

Al_o, Fe_o, Si_o= oxalate extractable Al, Fe, Si; CEC = cation exchange capacity

Relationship between soil Cd and other chemical characteristics

Soil total Cd concentration showed a positive and significant ($p<0.05$) correlation with oxalate extractable Al and Si, and total P concentrations (Figure 1a). Soil total P was also significantly ($p<0.001$) correlated with oxalate extractable Al (Figure 1a). In contrast to Stafford et al. (2018), no significant correlation was found between soil total C and total Cd or between soil total C and total P. This lack of agreement can be related to the use of only topsoil samples with a very wide range of total C (Table 1), masking the relationship with total P values. Significant correlation of total Cd with oxalate extractable Al could be related with sorption of Cd in mineral surfaces containing Al (Loganathan et al., 2012).

A significant relationship between soil total Cd and total P concentrations was found in this set of samples (Figure 1b), in agreement with other studies in New Zealand agricultural soils suggesting P fertiliser as a main source of Cd accumulation above background concentrations (Abraham, 2018).



pH					**	***		***
-0.06	Cd	*		*			***	
0.02	0.62	Al	*	***			***	
0.08	0.25	0.58	Fe					
0.06	0.66	0.97	0.54	Si			**	
-0.48	0.08	0.04	-0.08	-0.07	C	***		***
-0.53	0.4	0.43	0.21	0.32	0.85	N		***
-0.11	0.77	0.79	0.38	0.79	0.07	0.49	P	
-0.49	0.32	0.3	0.14	0.2	0.88	0.92	0.34	CEC

Figure 1 (a) Correlation matrix for selected soil chemical properties measured for all the topsoil samples (i.e. Allophanic and non-allophanic soils together) obtained from New Zealand dairy farms. p value <0.05 = *, <0.01 = ** and <0.001 = *. See Table 1 for details on each variable; (b) regression analysis of soil total Cd in relation to soil total P concentration.**

MIR prediction of soil Cd concentration and other chemical characteristics: effect of spectra pre-processing

Mean Allophanic and non-allophanic soils MIR reflectance spectra, as well as spectra after and pre-processing using either continuum removal or first derivative transformation with Savitzky Golay smoothing are depicted in Figure 2. Raw spectra (2a) showed relatively small reflectance differences in Allophanic soil compared to non-allophanic soil. With raw spectra, pointing the exact location of spectral response was not easy which was maybe due to non-relevant spectral information masking the relevant ones, scattering effect, chemical interference or instrumental drift requiring pre-processing of spectra before developing a robust prediction model (Sila et al., 2016). Continuum removal pre-processed spectra (2b) normalised the spectra with a common baseline (Clark & Roush, 1984), thus absorption features more prominently displayed (Wu et al., 2007). With first derivative transformation and Savitzky Golay smoothing pre-processing (2c), changes in the spectral response are featured sharp, intensified and distinct (Todorova et al., 2014).

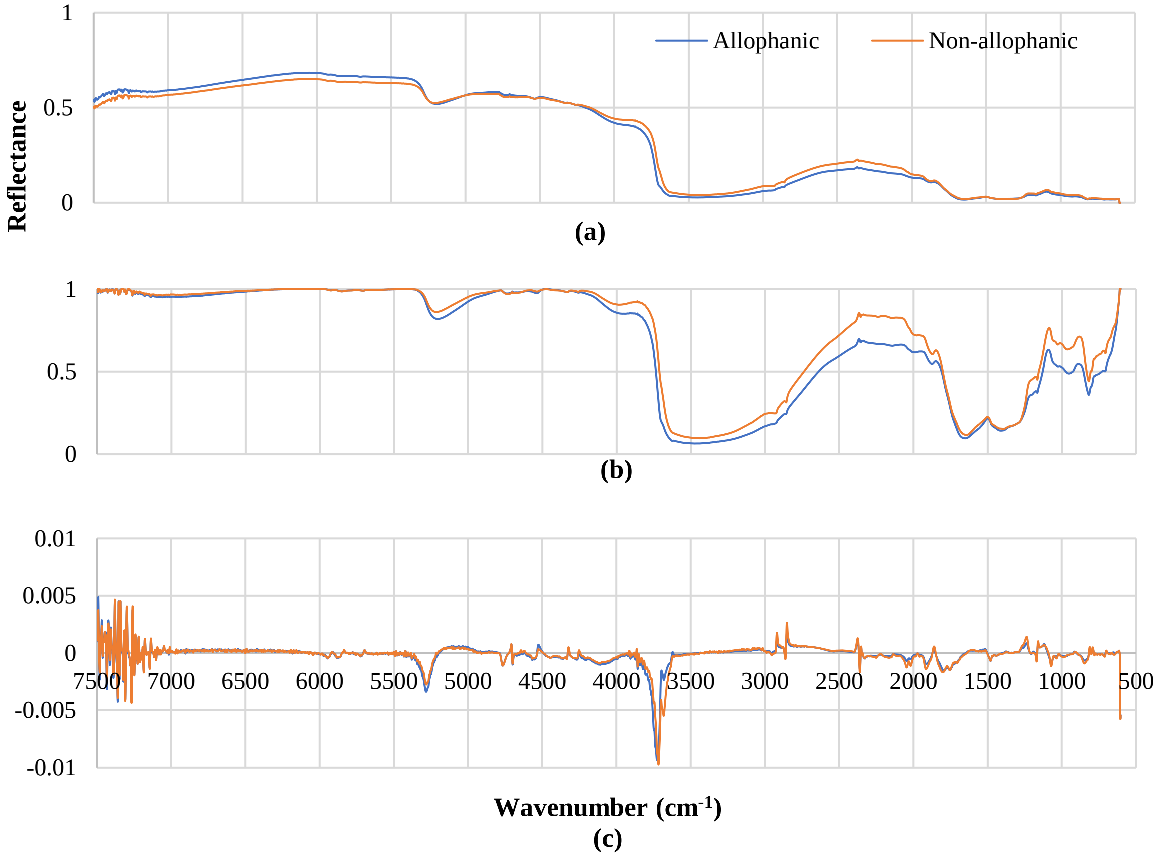


Figure 2 Averaged Allophanic (blue colour) and non-allophanic (orange colour) soils MIR spectra: (a) reflectance, (b) continuum removal pre-processed, (c) first derivative transformation with Savitzky Golay smoothing pre-processed.

MIR prediction of soil Cd concentration and other chemical characteristics: evaluation of model performance

MIR-based prediction models were obtained for five soil parameters (total Cd, total P, total C, oxalate extractable Al and CEC) by using two pre-processing methods (continuum removal and first derivative transformation with Savitzky Golay smoothing) and two regression algorithms (PLS and RF) (Table 2).

Models developed exhibited contrasting predictive performance for selected soil chemical characteristics (Table 2). Spectroscopy based model predictions of chemical characteristics as C or Al are feasible because they are considered spectrally active (Xia et al., 2007). Despite the fact that Cd is not expected to be spectrally active, its prediction has been associated with other spectrally active soil properties such as clay minerals (Kooistra et al., 2001), Fe oxides (Wu et al., 2007) and/or organic matter (Stafford et al., 2018). Exchangeable cations (Na, Mg, K and Mg) are spectrally active (Soriano-Disla et al., 2014) but their spectral response can be overlapped by other soil properties as clay and organic matter (Janik et al., 1998). Phosphorus is considered to have variable response to spectra based on which fraction of P is being compared (Stenberg et al., 2010). Hence, model performance of the soil property is influenced by whether it is associated with spectrally active or inactive component in the soil (Xia et al., 2007).

Pre-processing methods improved the performance of prediction models developed (Table 2). In general, models developed following spectra pre-processing using first derivative transformation with Savitzky Golay smoothing showed relatively high values for R^2 and RPD than models developed following continuum removal pre-processing technique (Table 2). PLS regression models showed better performace (i.e. high R^2 and RPD) for characteristics as soil total Cd, total P and total C than models based on RF following both pre-processing methods (Table 2). The RF prediction model had higher R^2 value for oxalate extractable Al and CEC than PLS using both pre-processing methods (Table 2).

For soil total Cd, PLS regression model developed using first derivative transformation with Savitzky-Golay smoothing pre-processed MIR spectra had higher R^2_{Val} (0.72), larger RPD_{Val} (1.81) and smaller $RMSE_{Val}$ (0.12 mg Cd/kg soil) than RF regression model (R^2_{Val} = 0.64, RPD_{Val} = 1.60, $RMSE_{Val}$ = 0.14 mg Cd/kg soil) (Table 2). Similar to these results, Siebielec et al. (2004) also found that PLS regression using MIR spectra improved the accuracy of models quantifying Cd in soil.

Table 2 Model performance of two regression algorithms (PLS and RF) using two pre-processing methods (continuum removal and first derivative transformation with Savitzky Golay smoothing) for calibration (n = 68) and validation (n = 16) models obtained using MIR soil spectra for soil chemical characteristics.

Pre-processing technique	Soil parameter	Partial least square regression model						Random forest regression model					
		Calibration			Validation			Calibration			Validation		
		R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD
Continuum removal	Total Cd (mg/kg)	0.70	0.14	1.82	0.59	0.18	1.56	0.45	0.19	1.35	0.45	0.21	1.39
	Total P (g/kg)	0.83	3.36	2.46	0.46	6.07	1.27	0.60	5.19	1.59	0.37	6.69	1.16
	Total C (%)	0.87	2.62	2.67	0.37	2.09	1.16	0.78	3.63	1.93	0.38	2.29	1.06
	Al_o (g/kg)	0.83	0.68	2.28	0.78	0.71	2.11	0.88	0.55	2.83	0.69	0.87	1.71
	CEC (cmol/kg)	0.99	1.36	9.91	0.42	5.46	1.25	0.73	7.11	1.89	0.57	5.02	1.36
First derivative	Total Cd (mg/kg)	0.74	0.14	1.95	0.72	0.12	1.81	0.48	0.20	1.40	0.64	0.14	1.60
	Total P (g/kg)	0.63	5.46	1.60	0.44	3.91	1.36	0.62	5.33	1.64	0.42	4.49	1.89
	Total C (%)	0.79	3.25	2.08	0.94	1.43	3.14	0.58	4.42	1.53	0.81	2.32	1.94
	Al_o (g/kg)	0.96	0.32	4.82	0.52	1.33	1.11	0.83	0.65	2.78	0.96	0.36	4.11
	CEC (cmol/kg)	0.83	5.29	2.46	0.84	5.58	1.67	0.53	8.88	1.47	0.96	2.80	3.33

Al_o = oxalate-extractable Al, CEC = cation exchange capacity

PLS-based calibration and validation regression models using MIR soil spectra following first derivative transformation and Savitzky Golay smoothing is shown in Figure 3. For the dataset measured, samples spread along the 1:1 line was with high positive fit (Figure 3). PLS-based calibration model was developed for soil total Cd concentration ranging between 0.1 and 1.4 mg Cd/kg soil, with a dataset including 68 samples. Prediction model of Cd using PLS was developed with only 16 soil samples (Cd concentration range: 0.24 – 0.87 mg Cd/kg soil) randomly selected as a validation set. The prediction models developed in this study were suitable for the quantification of relatively low soil total Cd concentrations, as those present in New Zealand agricultural soils (Abraham, 2018).

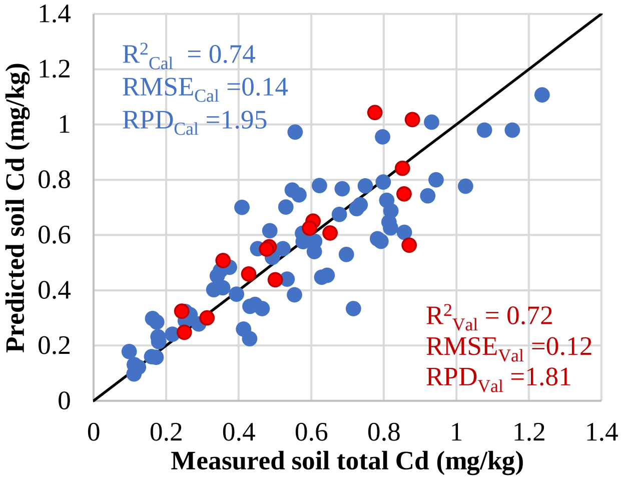


Figure 3 Measured soil total Cd concentration compared with predicted soil Cd from PLS regression based calibration (blue dots) and validation (red dots) models developed using MIR soil spectra pre-processed with first derivative transformation and Savitzky Golay smoothing. Continuous line is 1:1 line.

Correlation between MIR spectra and soil chemical characteristics

Pearson's correlation coefficient value of spectral response and soil properties (soil total Cd and oxalate extractable Al) were similar in shape with difference in magnitude as per their concentration in the soil (Figure 4). Pre-processing method influenced number of spectral bands showing positive relationdhip with soil properties and strength of correlation. First derivative pre-processing increased the number of spectral bands with positive correlation with soil Cd (Xia et al., 2007) than continuum removal. These relations can be used to select appropriate bands for prediction model development allowing improvement in accurate quantification of Cd (Wang et al., 2017a).

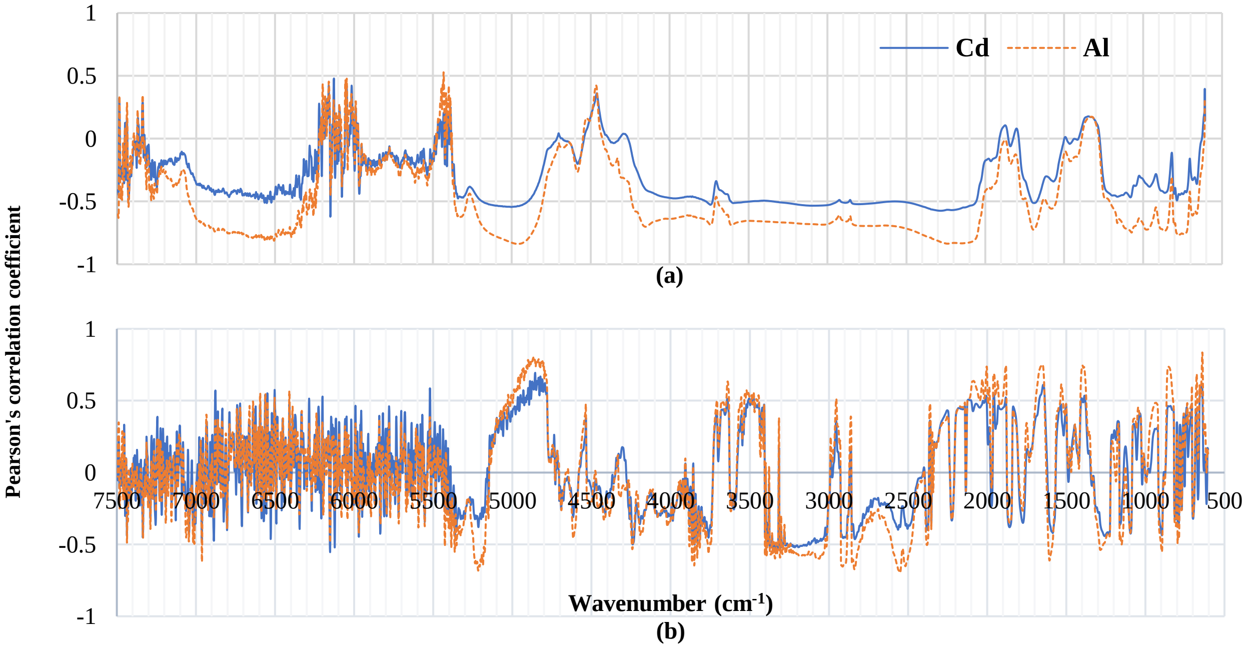


Figure 4 Soil total Cd (blue coloured smooth line) and oxalate extractable Al (orange coloured dotted line) pearson's correlation coefficient with MIR spectra response after pre-processing using continuum removal (a) and first derivative transformation with Savitzky Golay smoothing (b).

Summary and conclusion

This work used MIR spectroscopy in conjucation with reference wet chemistry methods to produce prediction models allowing quantitative estimation of Cd concentration in a set of agricultural soils (Cd concentration range: 0.1 and 1.4 mg Cd/kg soil). Using pre-processed MIR spectra with first derivative transformation and Savitzky Golay smoothing for PLS regression prediction model improved the model performace to predict soil Cd concentration. Spectroscopy-based quantification of Cd can be further improved by: 1) feeding more soil data into models and exploring other statistical modelling techniques, 2) understanding how other soil properties may influence soil Cd (e.g. soil Cd fractionation) and incorporate the information in the development of prediction models, and 3) integrating information obtained from different sensor technologies (e. g. vis-NIR, portable x-ray fluorescence spectroscopy), either by data or model fusion. The spectroscopy-based quantification of Cd has the potential to allow a rapid and cost-effective measurement of agricultural soil Cd concentration contributing to a better monitoring of this element.

Acknowledgements

Authors would like to acknowledge Aaron David Stafford for making available soil samples collected from 30 New Zealand dairy farms during his PhD work in 2014. We also thank Ian Furkert, Senior technician at soil science laboratory for teaching the main author soil laboratory techniques. This doctoral study is funded by Massey Doctoral Scholarship-2019.

References

Abraham, E. (2018). Cadmium in New Zealand agricultural soils. *New Zealand Journal of Agricultural Research,* 1-18.

Adams, M. J. (2007). *Chemometrics in analytical spectroscopy*. Royal Society of Chemistry, UK.

Blakemore, L. C., Searle, P. L., & Daly, B. K. (1987). *Methods for chemical analysis of soils*. NZ Soil Bureau, Department of Scientific and Industrial Research.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32.

Breiman, L., Cutler, A., Liaw, A., & Wiener, M. (2018). randomForest: Breiman and Cutler's random forests for classification and regression. R package version 4.6-14.

Cavanagh, J.-A. E., Yi, Z., Gray, C. W., Munir, K., Lehto, N., & Robinson, B. H. (2019). Cadmium uptake by onions, lettuce and spinach in New Zealand: Implications for management to meet regulatory limits. *Science of The Total Environment, 668,* 780-789.

Cavanagh, J., McNeill, S., Arienti, C., & Rattenbury, M. (2015). Background soil concentrations of selected trace elements and organic contaminants in New Zealand. *Lincoln, NZ: Landcare Research. Contract Report LC2440*.

CCME. (1999). *Canadian soil quality guidelines for the protection of enviornmental and human health: Cadmium. Canadian environmental quality guidelines, 1999. Canadian Council of Ministers of the Environment. Winnipeg, Cananda*.

Chen, T., Chang, Q., Clevers, J. G. P. W., & Kooistra, L. (2015). Rapid identification of soil cadmium pollution risk at regional scale based on visible and near-infrared spectroscopy. *Environmental Pollution, 206,* 217-226.

Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth, 89*(B7), 6329-6340.

CWG. (2011). *Cadmium and New Zealand agriculture and horticulture: a strategy for long term risk management.* A report prepared by the Cadmium Working Group for the Chief Executives Environmental Forum. (Accessed 24 August 2015).

Fertiliser Association. (2019). *Tiered fertiliser management system for cadmium: for the management of soil cadmium accumulation from phosphate fertiliser applications* (Developed by the Fertiliser Association of New Zealand as part of the work programme undertaken for the Cadmium Working Group, which was established by the Regional Councils' Chief Executives' Environment Forum. Version 3.).

Godt, J., Scheidig, F., Grosse-Siestrup, C., Esche, V., Brandenburg, P., Reich, A., & Groneberg, D. A. (2006). The toxicity of cadmium and resulting hazards for human health. *Journal of occupational medicine and toxicology, 1*(1), 22.

Gray, C.W., McLaren, R.G., Roberts, A.H.C. & Condron, L.M. (1999). The effect of long-term phosphatic fertiliser applications on the amounts and forms of cadmium in soils under pasture in New Zealand. *Nutrient Cycliing in Agroecosystems,* 54(3), 267-277.

Hill Laboratories. (2019). *Summary of methods: sample type: soil. R J Hill Laboratories Limited, Hamilton, New Zealand*.

Hooda, P. S. (Ed.). (2010). *Trace elements in soils*. John Wiley & Sons.

Janik, L. J., Merry, R. H., & Skjemstad, J. O. (1998). Can mid infrared diffuse reflectance analysis replace soil extractions? *Australian Journal of Experimental Agriculture, 38*(7), 681-696.

Kim, N. (2005). *Cadmium accumulation in Waikato soils. Environment Waikato Regional Council, Hamilton, New Zealand. Environment Waikato Technical Report 2005/51*.

Kooistra, L., Wehrens, R., Buydens, L. M. C., Leuven, R. S. E. W., & Nienhuis, P. H. (2001). Possibilities of soil spectroscopy for the classification of contaminated areas in river floodplains. *International Journal of Applied Earth Observation and Geoinformation, 3*(4), 337-344.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2020). caret: classification and regression training. R package version 6.0-85.

Loganathan, P., Vigneswaran, S., Kandasamy, J., & Naidu, R. (2012). Cadmium sorption and desorption in soils: a review. *Critical Reviews in Environmental Science and Technology, 42*(5), 489-533.

Luedeling, E. (2019). chillR: statistical methods for phenology analysis in temperate fruit. R package version 0.70.21.

Marschner, H. (2012). *Marschner's mineral nutrition of higher plants*. Academic press.

McDowell, R. W., Taylor, M. D., & Stevenson, B. A. (2013). Natural background and anthropogenic contributions of cadmium to New Zealand soils. *Agriculture, Ecosystems and Environment, 165*, 80-87.

McKenzie, H., & Wallace, H. (1954). The Kjeldahl determination of nitrogen: A critical study of digestion conditions-temperature, catalyst, and oxidizing agent. *Australian Journal of Chemistry, 7*(1), 55-70.

Mevik, B.-H., Ron Wehrens, R., Liland, K. H., & Hiemstra, P. (2019). pls: partial least squares and principal component regression. R package version 2.7-1.

MIEF. (2007). *Government Decree on the Assessment of Soil Contamination and Remediation Needs (214/2007).* Ministry of the Environment Helsinki, Finland.

Nawar, S., Cipullo, S., Douglas, R. K., Coulon, F., & Mouazen, A. M. (2019). The applicability of spectroscopy methods for estimating potentially toxic elements in soils: state-of-the-art and future trends. *Applied Spectroscopy Reviews*, 1-33.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben Dor, E., Brown, D. J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J. A. M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J. M., Shepherd, K. D., Stenberg, B., Towett, E. K., Vargas, R., & Wetterlind, J. (2015). Soil spectroscopy: an alternative to wet chemistry for soil monitoring. In D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. 132, pp. 139-159). Academic Press.

Poggio, M., Roudier, P., Blaschek, M., & Hedley, C. (2018). Integration of NIR on a multi-sensor platform to improve soil resource assessments. *NIR news, 29*(5), 15-18.

R Core Team. (2013). *R: A language and environment for statistical computing.* R Foundation for statistical computing, Vienna, Austria. *URL http://www.R-project.org/.*

Ramirez Lopez, L., & Stevens, A. (2016). resemble: regression and similarity evaluation for memory based learning in spectral chemometrics. R package version 1.2.2.

Riedel, F., Denk, M., Muller, I., Barth, N., & Glasser, C. (2018). Prediction of soil parameters using the spectral range between 350 and 15,000 nm: A case study based on the permanent soil monitoring program in Saxony, Germany. *Geoderma, 315*, 188-198.

Roberts, A. H. C., Longhurst, R. D., & Brown, M. W. (1994). Cadmium status of soils, plants, and grazing animals in New Zealand. *New Zealand Journal of Agricultural Research, 37*(1), 119-129.

Rossel, R. A. V., & Webster, R. (2012). Predicting soil properties from the Australian soil visible–near infrared spectroscopic database. *European Journal of Soil Science, 63*(6), 848-860.

Rouillon, M., Taylor, M. P., & Dong, C. (2017). Reducing risk and increasing confidence of decision making at a lower cost: In-situ pXRF assessment of metal-contaminated sites. *Environmental Pollution, 229*, 780-789.

RStudio Team. (2015). *RStudio: Integrated Development for R. RStudio Inc., Boston, MA . Version: 1.2.1335. URL http://www.rstudio.com/.*

Shepherd, K. D., & Walsh, M. G. (2007). Infrared spectroscopy—enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries. *Journal of Near Infrared Spectroscopy, 15*(1), 1-19.

Siebielec, G., McCarty, G. W., Stuczynski, T. I., & Reeves, J. B. (2004). Near-and mid-infrared diffuse reflectance spectroscopy for measuring soil metal content. *Journal of environmental quality, 33*(6), 2056-2069.

Sila, A. M. & Terhoeven-Uselmans, T. (2013). soil.spec: soil spectral file conversion, data exploration and regression functions. R package ver. 2.1.4.

Sila, A. M., Shepherd, K. D., & Pokhariyal, G. P. (2016). Evaluating the utility of mid-infrared spectral subspaces for predicting soil properties. *Chemometrics and Intelligent Laboratory Systems, 153*, 92-105.

Soriano-Disla, J. M., Janik, L. J., Viscarra Rossel, R. A., Macdonald, L. M., & McLaughlin, M. J. (2014). The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Applied Spectroscopy Reviews, 49*(2), 139-186.

Stafford, A., Kusumo, B., Jeyakumar, P., Hedley, M., & Anderson, C. (2018). Cadmium in soils under pasture predicted by soil spectral reflectance on two dairy farms in New Zealand. *Geoderma Regional, 13,* 26-34.

Stafford, A. D. (2017). *Distribution of cadmium in long-term dairy soils, its accumulation in selected plant species, and the implications for management and mitigation* Massey University, Manawatu Campus, New Zealand].

Stenberg, B., Rossel, R. A. V., Mouazen, A. M., & Wetterlind, J. (2010). Visible and near infrared spectroscopy in soil science. In *Advances in agronomy* (Vol. 107, pp. 163-215). Elsevier.

Stevens, A., & Ramirez Lopez, L. (2014). *prospectr: miscellaneous functions for processing and sample selection of vis-NIR diffuse reflectance data. R package version 0.1.3.*

Tasumi, M., & Sakamoto, A. (2014). *Introduction to experimental infrared spectroscopy: Fundamentals and practical methods*. John Wiley & Sons.

Todorova, M., Mouazen, A. M., Lange, H., & Atanassova, S. (2014). Potential of near-infrared spectroscopy for measurement of heavy metals in soil as affected by calibration set size. *Water, Air, & Soil Pollution: An International Journal of Environmental Pollution*(8), 1.

Wang, C., Li, W., Guo, M., & Ji, J. (2017a). Ecological risk assessment on heavy metals in soils: Use of soil diffuse reflectance mid-infrared fourier-transform spectroscopy. *Scientific reports, 7,* 40709.

Wang, F., Li, C., Wang, J., Cao, W., & Wu, Q. (2017b). Concentration estimation of heavy metal in soils from typical sewage irrigation area of Shandong province, China using reflectance spectroscopy. *Environmental Science and Pollution Research, 24*(20), 16883-16892.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). corrplot: visualization of a correlation matrix. R package version 0.84.

Wickham, H. (2017). reshape2: flexibly reshape data: a reboot of the reshape package. R package version 1.4.3.

Wickham, H., Bryan, J., Kalicinski, M., Leitienne, C., Colbert, B., Hoerl, D., & Miller, E. (2019a). readxl: read excel files. R package version 1.3.1.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., & Yutani, H. (2019b). ggplot2: Create elegant data visualisations using the grammer of graphics. R package version 3.2.1.

Williams, P., & Norris, K. H. (2001). *Near-infrared technology in the agricultural and food industries* (2nd ed.). American Association of Cereal Chemists.

Wu, Y., Chen, J., Ji, J., Gong, P., Liao, Q., Tian, Q., & Ma, H. (2007). A mechanism study of reflectance spectroscopy for investigating heavy metals in soils. *Soil science society of America journal, 71*(3), 918-926.

Xia, X. Q., Mao, Y. Q., Ji, J. F., Ma, H. R., Chen, J., & Liao, Q. L. (2007). Reflectance spectroscopy study of Cd contamination in the sediments of the Changjiang river, China. *Environmental Science & Technology, 41*(10), 3449-3454.