Canadian Journal of Soil Science
Revue canadienne de la science du sol

Infrared spectroscopy prediction of organic carbon and total nitrogen in soil and particulate organic matter from diverse Canadian agricultural regions

Journal:	<i>Canadian Journal of Soil Science</i>
Manuscript ID	CJSS-2017-0070.R1
Manuscript Type:	Article
Date Submitted by the Author:	26-Sep-2017
Complete List of Authors:	Zhang, Lei; Institute of Applied Ecology Chinese Academy of Sciences, Yang, X.; Agriculture and Agri-Food Canada, 2585 County Road 20, Drury, Craig; Agriculture & Agri-Food Canada, Chantigny, M.; Agriculture Canada, Gregorich, Ed; Agriculture Canada, Miller, Jim; Agriculture and Agri-Food Canada, Bittman, Shabtai; AAFC, Reynolds, W.; Agriculture & Agri-Food Canada, Research Branch Yang, Jingyi; Agriculture and Agri-Food Canada, Greenhouse and Processing Crops Research Centre, Canada,
Keywords:	Carbon, Nitrogen, Infrared Spectroscopy, Bulk soil, Particulate organic matter

SCHOLARONE™
Manuscripts

**Infrared spectroscopy prediction of organic carbon and total nitrogen in soil and particulate organic matter
from diverse Canadian agricultural regions**

Lei Zhang^{1, 6, 7}, Xueming Yang^{1,*}, Craig Drury¹, Martin Chantigny², Ed. Gregorich³, Jim Miller⁴, Shabtai Bittman⁵, Dan Reynolds¹, Jingyi Yang¹

¹ Harrow Research and Development Centre, Agriculture and Agri-Food Canada (AAFC), Harrow, Ontario N0R 1G0, Canada

² Quebec Research and Development Centre, AAFC, Quebec, Quebec G1V 2J3, Canada

³ Ottawa Research and Development Centre Central Experimental Farm, AAFC, Ottawa, Ontario K1A 0C6, Canada

⁴ Lethbridge Research and Development Centre, AAFC, Lethbridge, Alberta T1J 4B1, Canada

⁵ Agassiz Research and Development Centre, AAFC, Agassiz, British Columbia V0M 1A0, Canada

⁶ Institute of Applied Ecology, Chinese Academy of Sciences, 72 Wenhua Road, Shenhe District, Shenyang, Liaoning 110016, China

⁷ University of Chinese Academy of Sciences, 19 Yuquan Road, Shijingshan, Beijing 100049, China

*Corresponding author. E-mail: xueming.yang@agr.gc.ca

Abstract

Infrared spectroscopy has the capacity to predict soil organic carbon (SOC) and total nitrogen (TN) at local/regional scales, but no studies have been conducted to evaluate this technique at a large (cross-regional) scale in Canada. In this paper, mid-infrared (MIR) and near-infrared (NIR) spectroscopy in combination with partial least squares regression (PLSr) were used to predict SOC and TN in whole soil and in particulate organic matter (POM) fractions on cross-regional, regional and local scales. Both MIR- and NIR-PLSr models well estimated SOC (Coefficient of determination (R^2) = 0.79-0.92, Residual Prediction Deviation/ Ratio of Prediction to Deviation (RPD) = 2.19-3.47), TN (R^2 = 0.70-0.92, RPD = 1.83-3.50), POM-C (R^2 = 0.76-0.96, RPD = 2.04-5.25) and POM-N (R^2 = 0.70-0.97, RPD = 1.83-5.78). The prediction efficiency of cross-regional models (R^2 = 0.90-0.96, RPD = 3.13-5.49) were similar to or better than the prediction of regional (R^2 = 0.70-0.97, RPD = 1.83-5.78) and local models (R^2 = 0.70-0.96, RPD = 1.83-5.33) and overall MIR-PLSr models (R^2 = 0.90-0.96, RPD = 1.98-3.47) yielded similar predictions for SOC and TN relative to NIR-PLSr models (R^2 = 0.70-0.92, RPD = 1.83-3.50) at cross-regional scale. Hence, it may be possible to develop MIR and/or NIR spectral models to estimate and monitor SOC, TN, POM-C and POM-N, and therefore soil quality, in a rapid and cost-efficient manner across regions with diverse soil types, climate and cropping history.

Keywords: Carbon, Nitrogen, Infrared spectroscopy, Bulk soil, Particulate organic matter

Introduction

Soil organic carbon (SOC) changes in agricultural soils are of great concern because of its beneficial roles in enhancing soil fertility, plant growth, environmental quality as well as its role being a sink/source for atmospheric carbon dioxide. The quantity and quality of SOC can be influenced by the intensity and type of soil management, including tillage, crop rotation, crop type and/or amendment use (Degryze et al., 2004; Lal, 2004), making SOC as an amenable indicator of soil quality change. Therefore, methods that measure SOC accurately, quickly, inexpensively and non-destructively will help advance studies to improve understanding of the impacts of soil management on SOC. Carbon (C) or nitrogen (N) sequestration or loss under different management practices occur slowly over time and it is difficult to detect changes in these pools over the short term. However, the labile fractions of soil C and N are usually sensitive to management changes and can respond rapidly. Soil particulate organic matter (POM) is often used as an indicator of active C and N (Christensen, 1992; Gregorich and Beare., 2008), because it is more sensitive and responds faster to management-induced changes than total SOC, total N (TN) or the biochemically recalcitrant organic fractions associated with the silt and clay particles which are stabilized and/or protected from decomposition by physicochemical mechanisms (Six et al., 2002).

Although POM indicator can quickly respond to soil management changes, the separation and determination of POM is labor-intensive and time-consuming, and often involves the use of chemicals (e.g. hexametaphosphate) to disperse the soil (Cambardella and Elliott, 1992; Gregorich and Beare, 2008). Carbon and N determinations in soil and POM are often conducted using wet digestion or dry combustion methods. These analytical methods have disadvantages including the use of hazardous chemicals and their subsequent release to the environment, incomplete oxidation, underestimation of SOC associated with wet chemistry methods (Hussain and Olson, 2000; Mebius, 1960; Meersmans et al., 2009; Walkley, 1946; Walkley and Black, 1934). Hence, there is a need to develop ways to accurately, efficiently, and inexpensively measure C and N in soil samples and in labile SOC fractions.

To overcome the above-mentioned problems of traditional chemical methods, infrared (IR) spectroscopy combined with multivariate analysis techniques, such as partial least-squares regression (PLSR), has increasingly been used for determining numerous soil physical, chemical and biological properties (e.g., Cécillon et al., 2008; Dick et al., 2016; Janik et al., 2009; Janik et al., 1998; Soriano-Disla et al., 2014; Viscarra Rossel et al., 2016; Zornoza et al., 2008). Infrared spectroscopy is based upon the principle that molecules absorb specific electromagnetic frequencies that represent their structural characteristics (Bellon-Maurel and McBratney, 2011; Soriano-Disla et al., 2014). Compared with conventional wet-chemistry methods, IR (including near-IR (NIR) and mid-IR(MIR)) spectroscopic

techniques are time- and cost-efficient, non-destructive, require less sample preparation, and often result in similar accuracy without use of chemical reagents (McCarty et al., 2002; Shepherd and Walsh, 2002; Viscarra Rossel et al., 2006). In order to assure the robustness and reliability of IR spectroscopic technique for determination of soil total and labile C fractions, calibration models should include a large number of samples with a wide range of soil properties (Du and Zhou, 2009; Zornoza et al., 2008). The usefulness of the calibration models at different scales (field, national or global) is directly related to what basis (e.g., analytic levels, texture, soil type, etc. or combinations thereof) soil samples can be included in a single calibration (Reeves III, 2010). Thus, the performance of IR models with different soil region at the same scale are necessary to be investigated. Also, since soils can be extremely different in composition and properties, there is a trend to develop large spectral libraries using larger, more diverse soil, and more dispersed sample sets, to ensure the IR spectroscopy being a robust analytical technique for determination of SOC and TN concentrations and their spatial variations (Brown et al., 2005; Islam et al., 2003; Shepherd and Walsh, 2002; Stevens et al., 2008). Extending local or regional spectral libraries for development of a dynamic and easily updatable database with better global covering spectral library has been encouraged (Viscarra Rossel et al., 2016). Nevertheless, Stenberg et al. (2010) stated that valid prediction could be possible if the characteristics of the spectra contained in the spectral libraries are similar to the estimated samples.

Infrared spectroscopy is widely used for estimating C and N concentrations in bulk soil (e.g. Reeves III et al., 2001; Viscarra Rossel et al., 2006; Yang and Mouazen, 2012) and in POM fraction (Bornemann et al., 2010; Reeves III et al., 2006; St. Luce et al., 2014; Zimmermann et al., 2007) at different scales (local, regional, national or global). In Canada, MIR and NIR spectroscopy have been successfully used to estimate C and N in bulk soils (Martin et al., 2002; Nduwamungu et al., 2009; St. Luce et al., 2012; St. Luce et al., 2014; Xie et al., 2011), and C and N in POM fractions (St. Luce et al., 2014; Yang et al., 2012) in agricultural soils, but most of these calibration models are based upon field scale (Martin et al., 2002; Nduwamungu et al., 2009; Yang et al., 2012) or local/provincial scales (Reeves III and Smith, 2009; St. Luce et al., 2014; Xie et al., 2011; Yang et al., 2012). No broader calibration models of SOC, TN, POM-C, -N have been tested across geographical regions with diverse soil types in Canada.

In this study, we applied MIR- and NIR-PLSr modeling techniques to a set of diverse soils collected from different cropping zones across Canada's primary agricultural regions to test the efficacy of the models with different soil samples scales on the predictions of SOC, TN, and POM-C and -N concentrations in bulk soil. The models were constructed on different scale levels, including cross-regional (all samples included), regional (western or eastern Canada) and local (field site) scale. The two specific objectives of this study were to: i) evaluate the performance of cross-regional IR-PLSr models on the

prediction of SOC, TN, POM-C and POM-N concentrations in bulk soil compared to the performance of regional and local IR-PLSr models, and ii) compare the performance of MIR-, relative to NIR-PLSr models for these predictions.

Materials and methods

Soil sampling

Soil samples (0-10cm or 0-15cm depths, depending on locations) were collected from five Agriculture & Agri-Food Canada (AAFC)'s long-term field study sites (at least 20 years) in fall 2014. The different sites encompass a wide diversity of geographical and climatic regions and include numerous management practices, which should ensure that a wide range of C and N concentrations are included at the different scales of the study (Table 1). The experimental sites were located (i) in Agassiz, British Columbia (49°10' N, 125°15' W) including 80 samples from a manure experiment, (ii) in Lethbridge, Alberta (49°38' N, 112°48' W) including 56 samples from a fertilizer and organic amendment study, (iii) in Woodslee, Ontario (42°13' N, 82°44' W) including 50 samples from a compost study, (iv) in Ottawa, Ontario (45°18' N, 75°43' W) including 30 samples from a rotation and fertilization study, and (v) in Québec City, Québec (46°48' N, 71°23' W) including 60 samples from a tillage and organic-mineral fertilization study.

Soil analysis using conventional methods

The bulk soil samples collected from the surface layer were air-dried, passed through a 2-mm sieve and homogeneously mixed. A subsample (~20 g) of each sample (< 2-mm) was further ball-milled to < 0.15-mm sieve for the determination of organic carbon (OC) and TN in bulk soil. Soil POM was analyzed using the method of Cambardella and Elliott (1992). Briefly, ten-gram subsamples of soils (< 2-mm sieve) were dispersed in 30 mL of 5 g L⁻¹ sodium hexametaphosphate by shaking for 16 h on a reciprocal shaker. The dispersed samples were filtered through a 53-µm sieve and, after rinsing with water until all fine particles (< 53-µm) were washed away, the materials retained on the sieve were collected as POM and oven dried at 50 °C. The dried POM was ball-milled and sieved (< 0.15-mm). The concentrations of OC and TN both in ground bulk soil (< 0.15-mm) and ground POM (< 0.15-mm) were determined using a LECO CN-2000 analyzer (LECO Corp., St Joseph, MI, USA). For the soils containing free carbonates (mainly the soils from Lethbridge, Alberta), OC content was assessed as the difference between total C and mineral C, and the latter was determined on samples ashed at 475 °C for 5 h.

MIR and NIR spectral collection and processing

Stainless steel sample cups were filled with air-dried ground ($< 0.15\text{-mm}$) soil samples ($\sim 1\text{-g}$), and the sample surface in the cup was leveled with a flat spatula. Three separate MIR spectra and NIR spectra were obtained from each soil sample using a Bruker-TENSOR 37 spectrometer (Bruker Optik GmbH, Germany) equipped with diffuse-reflectance accessory (Easidiff, Pike Technologies, Madison, USA). The ambient air MIR spectrum was measured on potassium bromide (KBr) sample and ambient air NIR was measured using an alignment mirror coated with silicon monoxide. The MIR spectra were recorded in the range of 4000 to 400 cm^{-1} ($2500\text{-}25000\text{ nm}$) with 64 scans and 2 cm^{-1} resolution, while the NIR spectra were recorded in the range of 8000 to 4000 cm^{-1} ($1250\text{-}2500\text{ nm}$) with 16 scans and 1 cm^{-1} resolution. Data were displayed as pseudo-absorbance ($\log [1/\text{reflectance}]$). The MIR spectrum was baseline corrected and the original spectrum was used for NIR, and the average MIR (or NIR) spectra of three scans were used in model development to fit the concentrations of SOC and TN, POM-C and POM-N.

Model development and statistics

Model calibration was performed using PLSr with the OPUS QUANT 2 software by related the pre-processed spectra data with the measured SOC, TN, POM-C and POM-N. In the process of model development, spectra data was pre-processed using the OPUS QUANT 2 software (Bruker Optik GmbH, Germany) with various mathematical approaches to optimize the calibration model, including (1) original spectra without data processing, (2) constant offset elimination, (3) straight line subtraction, (4) vector normalization, (5) min–max normalization, (6) multiplicative scatter correction, (7) first derivative, (8) second derivative, (9) first derivative + straight line subtraction, (10) first derivative + vector normalization, and (11) first derivative + multiplicative scatter correction.

The optimized calibration models from all mathematical approaches were ranked, and the best calibration model with the highest coefficient of determination (R^2) and lowest root mean square error of calibration (RMSEC) for each optimization was selected. For each soil parameter, the calibration models were validated employing both a “leave-one-out” full-cross validation as well as a test-set validation which 70% of the total samples was randomly used in model calibration and the other 30% of samples was used as independent test-set in model validation. We reported and discussed the “leave-one-out” cross-validation because this validation not only yields an optimistic statistic of the models, but it also avoids model over-fitting (Hemmateenejad et al., 2005; Escandar et al., 2006; Nduwamungu et al., 2009; Brunet et al., 2007). The performance of independent test-set validation showed similar results as the “leave-one-out” validation which was demonstrated in supporting information (Tables S1 and S2). Models were tested on three levels: a cross-regional model (developed by using the complete dataset), regional models (developed by pooling samples from western or eastern region) and local models (developed by using site-specific samples).

Three statistics were used to evaluate the performance of models (Eqs. 1 to 3). The calibration models or validation results reported in this paper were those with the highest coefficient of determination (R^2), the lowest values of root mean square error of calibration (RMSEC) or prediction (RMSEP), and the highest ratio of the standard deviation (SD) of measured data to RMSEC or RMSEP, called the ratio of standard deviation of prediction (or calibration) to standard deviation (of the property to be estimated) (RPD). RMSEC represented the root mean square error of prediction between measured values and fitted values in model calibration. RMSEP represented the root mean square error of prediction between measured values and estimated values in model validation. The equations describing the statistics are:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \dots\dots\dots (1)$$

$$RMSEP = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \dots\dots\dots (2)$$

$$RPDc(RPDp) = \frac{SD}{RMSEC(RMSEP)} \dots\dots\dots (3)$$

where y_i , $\bar{y}$ and $\hat{y}_i$ represent measured values, mean of measured values and estimated values, respectively, of SOC, TN, POM-C and -N; n , is the number of the estimated or measured values and SD , is the standard deviation of the measured data.

To evaluate the accuracy of models, the R^2 and the RPD statistic were used. According to Saeys et al. (2005) and Janika et al. (2009), a R^2 value between 0.66 and 0.81 indicates approximate quantitative predictions, whereas a R^2 value between 0.82 and 0.90 reveals good prediction. Calibration models having $R^2 > 0.91$ are considered to be excellent. Regarding RPD statistic, an $RPD < 2$ is considered insufficient for applications, a value for RPD between 2 and 2.5 means approximate quantitative predictions, a value for RPD between 2.5 and 3 can be classified as good semi-quantitative predictions, and an $RPD > 3$ indicates an excellent quantitative prediction.

Results

Descriptive statistical data for SOC, soil TN, POM-C and POM-N

Summery statistics of the concentrations of SOC and TN determined using traditional dry combustion method are listed in Table 1. Across all sites, SOC and TN concentrations varied from 10.3 to 162 g C kg⁻¹ (Coefficient of Variation (CV) 58.8%) and 0.950 to 12.1 g N kg⁻¹ (CV 58.18%). Across the sites from western Canada, the SOC and TN concentrations varied from 20.8 to 162 g C kg⁻¹ (CV 53.6%) and 1.71 to 12.1 g N kg⁻¹ (CV 58.1%). Across the sites from eastern Canada, the SOC and TN concentrations varied from 10.3 to 63.8 g C kg⁻¹ (CV 34.9%) and 0.950 to 4.40 g N kg⁻¹ (CV 27.8%).

Lethbridge soils averaged the highest concentrations of SOC (mean 72.1 g C kg^{-1}) and TN (mean 5.84 g N kg^{-1}), followed by the soils in Agassiz with intermediate concentrations of SOC and TN (42.2 g C kg^{-1} and 3.30 g N kg^{-1}), and the soils from other sites (Quebec city, Ottawa and Woodslee) had the similar low concentrations of SOC ($24.0\text{-}34.8 \text{ g C kg}^{-1}$) and TN ($2.18\text{-}2.38 \text{ g N kg}^{-1}$).

Descriptive statistics of the POM-C and -N concentrations are listed in Table 1. The POM-C and -N concentrations varied from <0.001 to 109 g C kg^{-1} soil and <0.001 to 7.43 g N kg^{-1} soil, respectively, among all samples. The POM-C and -N concentrations varied from 2.81 to 109 g C kg^{-1} soil and 0.149 to 7.43 g N kg^{-1} soil, respectively, among the soil samples from western Canada. The POM-C and -N concentrations varied from <0.001 to 21.7 g C kg^{-1} soil and <0.001 to 1.61 g N kg^{-1} soil, respectively, among the soil samples from eastern Canada. Lethbridge soils had the highest POM-C (mean 34.5 g C kg^{-1} soil) and -N (mean 2.64 g N kg^{-1} soil), followed by the soils in Agassiz and Ottawa with intermediate concentrations of POM-C ($5.91\text{-}9.04 \text{ g C kg}^{-1}$ soil), and the soils in Woodslee and Quebec had similar low concentrations of POM-C ($3.34\text{-}3.47 \text{ g C kg}^{-1}$ soil). The soils from Agassiz, Quebec, Ottawa and Woodslee had similar low concentrations of POM-N ($0.224\text{-}0.594 \text{ g N kg}^{-1}$ soil).

MIR and NIR spectra

Averaged MIR and NIR spectra for each site are presented in Fig. 1. Soil spectral features, either MIR or NIR, are results of absorption and combination of bond vibrations of functional groups which are attributed to the mineral and organic components in soil (Bellon-Maurel and McBratney, 2011; Soriano-Disla et al., 2014). Based on literature (e.g. Bellon-Maurel and McBratney, 2011; Soriano-Disla et al., 2014), soil organic components might contain following MIR absorption bands: (i) a broad band at 3420 cm^{-1} due to H bonded to OH groups (also from the minerals); (ii) bands between 2950 and 2850 cm^{-1} attributed to aliphatic C-H stretches; (iii) bands between 1454 and 1157 cm^{-1} associated to C-H bending in alkane (1454 cm^{-1}), a symmetric C=O stretch of COO^- and C-H bending of aliphatics bands near 1370 cm^{-1} , and C-O bond in aromatic ester (1280 cm^{-1}), respectively. In addition, obvious absorbance bands also occurred at 979 cm^{-1} (di-substituted alkene), 809 cm^{-1} (vinylidene alkene), 701 cm^{-1} (di-substituted alkene) and 682 cm^{-1} (di-substituted alkene) frequencies. The soil NIR spectra demonstrate distinct fingerprints related to several functional groups such as absorption bands at the 7070 cm^{-1} , 5220 cm^{-1} and 4530 cm^{-1} frequencies, which are likely caused by the following functional groups: aliphatic C-H and water O-H (7070 cm^{-1} , 5220 cm^{-1}), phenolic O-H, amine N-H, aliphatic C-H and amide N-H (4530 cm^{-1}). The absorbance peaks at these frequencies of, both MIR and NIR, differed in intensity and wavenumbers among soils from the five sites. These IR spectral characteristics have been widely used as bases for classifying soils (Viscarra Rossel et al., 2016; Xie et al., 2015) and estimating the SOC

and TN concentrations in soil samples (Bellon-Maurel and McBratney, 2011; St. Luce et al., 2014; Soriano-Disla et al., 2014).

IR-PLSr prediction of SOC and TN in bulk soil

The statistical parameters for the SOC and TN model calibration showed good reliabilities at different sample scales, as indicated by high R^2 (> 0.90) and RPD (> 3), except for the MIR-PLSr at Agassiz ($R^2 = 0.80$ - 0.89 , RPD = 2.26 - 2.98), MIR-PLSr (TN) in eastern Canada ($R^2 = 0.81$, RPD = 2.30), NIR-PLSr in eastern Canada ($R^2 = 0.79$ - 0.87 , RPD = 2.16 - 2.82), and at Woodslee ($R^2 = 0.06$ - 0.13 , RPD = 1.03 - 1.17) (Table 2). Both MIR- and NIR-PLSr derived calibration models for SOC and TN predictions were more consistent on the cross-regional level ($R^2 = 0.92$ - 0.95 , PRD = 3.61 - 4.42) than on the regional level ($R^2 = 0.79$ - 0.93 , PRD = 2.16 - 3.85), and were better for soils from western region ($R^2 = 0.91$ - 0.93 , PRD = 3.29 - 3.85) than from eastern region ($R^2 = 0.79$ - 0.93 , PRD = 2.16 - 3.73) (Table 2). When models were developed for SOC and TN predictions on the local level, the best MIR-PLSr models were found for the soils from Ottawa ($R^2 = 0.98$, PRD > 6), and the best NIR-PLSr models were found for the soils from Quebec and Ottawa ($R^2 = 0.99$, PRD > 10) (Table 2). As for the difference between IR types, the MIR- and NIR-PLSr derived calibration models for the predictions of SOC and TN yielded similar results at cross-regional and regional scales, however, the NIR-PLSr models were better than the MIR-PLSr models at the local scale (Table 2). The infrared spectral bands and/or the spectra processing procedures employed to model calibration for SOC were different from those for TN, and this was found for both MIR and NIR spectra. The statistical parameters of IR-PLSr model calibration also varied among properties (SOC or TN), sampling scales (cross-regional, regional or local) and IR types (MIR or NIR) (Table 2).

The performance of the optimal MIR- and NIR-PLSr models varied among properties (SOC or TN), sample scales (cross-regional, regional or local) and IR types (MIR or NIR) (Fig. 2). The MIR- and NIR-PLSr models performed well for estimating SOC and TN at the cross-regional scale with $R^2 = 0.90$ - 0.91 and RPD = 3.13 - 3.29 (Fig. 2). For the soils from western Canada, the MIR- and NIR-PLSr models also performed well for SOC and TN ($R^2 = 0.89$ - 0.91 , RPD = 3.01 - 3.42). For the soils from eastern Canada, the prediction of MIR-PLSr models were good for SOC ($R^2 = 0.85$, RPD = 2.85) and reasonably good for TN ($R^2 = 0.77$, RPD = 2.10), and the prediction of NIR-PLSr model was reasonably good for SOC ($R^2 = 0.81$, RPD = 2.32) but poor for TN ($R^2 = 0.70$, RPD = 1.83) (Fig. 2). At a local scale, MIR-PLSr model yielded excellent prediction for SOC in the soils from Quebec ($R^2 = 0.92$, RPD = 3.47). The predictions of MIR-PLSr models were good for SOC at the Lethbridge, Agassiz and Ottawa sites ($R^2 = 0.84$ - 0.88 , RPD = 2.52 - 2.86) and for TN at the Lethbridge and Quebec sites ($R^2 = 0.87$ - 0.89 , RPD = 2.80 -

2.96), and were reasonably good for TN at the Ottawa site ($R^2 = 0.83$, RPD = 2.44). NIR-PLSr resulted in an excellent prediction for SOC at the Quebec site ($R^2 = 0.91$, RPD = 3.28) and for TN at the Lethbridge site ($R^2 = 0.92$, RPD = 3.50). The predictions of NIR-PLSr models were good for SOC at the Lethbridge site ($R^2 = 0.88$, RPD = 2.91), and reasonably good for SOC at the Agassiz and Ottawa sites ($R^2 = 0.79$ -0.83, RPD = 2.19-2.43) and for TN at the Quebec and Ottawa sites ($R^2 = 0.81$ -0.83 and RPD = 2.31-2.39). Both MIR- and NIR-PLSr models yielded poor predictions for SOC and TN at the Woodslee site ($R^2 < 0.35$, RPD < 1.25) (Fig. 2). The MIR models outperformed the NIR models for the predictions of SOC and TN at cross-regional, regional and local scales, except for the Woodslee site where both MIR- and NIR-PLSr models failed (Fig. 2).

IR-PLSr predictions for C and N in POM

For the predictions of POM-C and -N concentrations, both MIR- and NIR-PLSr models were developed using the spectra collected from bulk soil samples (Table 3, Fig. 3). The spectral bands and/or the spectra processing procedures used in the model optimization process varied between POM-C and POM-N, and this difference applied to both MIR- and NIR models (Table 3). The statistics for the model calibration of POM-C and -N showed satisfactory results at different scales, except for the models at Woodslee (Tables 3). Both the MIR- and NIR-PLSr models calibrated for POM-C and -N predictions were better on the cross-regional level ($R^2 = 0.96$ -0.97, PRD = 4.77-6.31) and for the soils from western region ($R^2 = 0.97$ -0.98, PRD = 6.06-6.86) than for the soils from eastern region ($R^2 = 0.83$ -0.95, PRD = 2.43-4.67) (Table 3). Better MIR-PLSr models for POM-C and -N predictions were calibrated for the soils from Lethbridge and Ottawa ($R^2 = 0.97$ -0.99, PRD > 6) than for soils from the other sites; the NIR-PLSr models for POM-C and -N predictions were particularly good for soils from Lethbridge, Agassiz, and Ottawa, with $R^2 = 0.99$, PRD > 10 (Table 3). Considering IR types, the models for POM-C and -N predictions calibrated with NIR-PLSr were better than models developed with MIR-PLSr at the cross-regional, regional and local scales (Table 3).

The performance (validation) of IR-PLSr models (both MIR and NIR) for POM-C and -N predictions were very good with $R^2 > 0.95$ and PRD values > 4 at the cross-regional scale (Fig. 3). The predictions of MIR- and NIR-PLSr derived models were excellent for POM-C and -N in western region soils ($R^2 = 0.96$ -0.97, RPD > 5), and were good for POM-C ($R^2 = 0.87$ -0.86, RPD = 2.82-2.67) and reasonable for POM-N ($R^2 = 0.78$ -0.79, RPD = 2.11-2.19) in eastern region soils (Fig. 3). At the local scale, MIR-PLSr models yielded excellent predictions for POM-C in Lethbridge and Ottawa soils ($R^2 = 0.93$ -0.96, RPD = 3.79-4.76) and for POM-N in Lethbridge soils ($R^2 = 0.96$, RPD = 5.3), good predictions for POM-C in Agassiz and Quebec soils ($R^2 = 0.85$ -0.86, RPD = 2.1-2.64) and for POM-N in Agassiz and

Ottawa soils ($R^2 = 0.84-0.85$, RPD = 2.54-2.63), and reasonably good predictions for POM-N in Quebec soils ($R^2 = 0.79$, RPD = 2.19). NIR-PLSr models yielded excellent predictions for POM-C ($R^2 = 0.94-0.96$, RPD = 4.06-4.87) and POM-N ($R^2 = 0.92-0.96$, RPD = 3.46-5.33) in Lethbridge and Ottawa soils, good predictions for POM-C in Quebec soils ($R^2 = 0.85$, RPD = 2.60), reasonably good predictions for POM-C and -N in Agassiz soils ($R^2 = 0.75-0.76$, RPD = 2.01-2.04), and poor predictions for POM-N in Quebec soils ($R^2 = 0.70$, RPD = 1.83). The predictions for both MIR- and NIR-PLSr models were poor for POM-C and -N in Woodslee soils (Fig. 3).

Discussion

Our results, consistent with the above-mentioned studies, showed that the MIR- and NIR-PLSr models can estimate C and N concentrations in soil very well (e.g., Madari et al., 2006; Xie et al., 2011; Reeves III et al., 2006; Yang et al., 2012;) and POM very well (e.g., Bornemann et al., 2010; Reeves III et al., 2006; Yang et al., 2012; Zimmermann et al., 2007; St. Luce et al., 2014;). In this study, soil samples were collected from five different agricultural geographic sites accompanying with distinct IR spectra and extensive ranges in SOC, TN, POM-C and -N concentrations, the models at cross-regional and regional scales differ from our previous works with soil samples collected from a single site and more limited range in soil chemical and physical traits (Xie et al., 2011; Yang et al., 2012). This provides us an opportunity to test the efficacy of MIR- and/or NIR-PLSr models to estimate C and N concentrations in soil and POM at difference scales (cross-regional, regional, or local), encompassing the key agriculture areas across Canada.

Generally, for SOC and TN estimation the performance of both MIR- and NIR-PLSr models in cross-regional scale were similar or superior to the performance of the regional models. The models for western Canada always performed better than the models for eastern Canada, which may reflect the wider range of C and N concentrations found in soil and POM from the western soils compared to those from the eastern soils. This indicates that the variability of soils and soil types employed to the model directly influences the performance of model.

Sankey et al. (2008) proposed that soil calibration models based on local field sites generally yielded the most accurate predictions but required a substantial number of soil samples at each site, whereas global to regional calibration models were more economically efficient, but didn't provide sufficient accuracy as compared to the local models. However, similar or poorer accuracy for predictions of C and N in soil and POM were found with local models in comparison to cross-regional models. The weaker predictions for C and N in soil and POM in local models may be ascribed to the small number of soil samples in the local calibration models because the minimum detectable difference increases with the decrease in sample number for a given sample population (Stevens et al., 2008). Nevertheless, factors other

than sample size may have contributed, given that the accuracy of local models in estimating C and N in soil and POM differed among field sites with similar small sample numbers in the present study. For example, good model predictions occurred for all soil properties at the Ottawa (30 samples) and Lethbridge (56 samples) field sites, whereas poor model predictions were found with the Woodslee field site (50 samples). One presumed cause in different accuracy of local models was the big difference in the CV (%) of SOC and TN concentrations occurred between the Woodslee site (SOC: 9.13%, TN: 10.6%) and the Ottawa (SOC: 45.3%, TN: 40.7%) and Lethbridge (SOC: 53.2%, TN: 46.9%) sites. Another possible cause in different predictions of local models was that the distinct characteristics of soil organic matter resulted from different fertilization practices, such as continuous application of animal manure in Lethbridge and Agassiz soils versus single application of yard- and food-waste compost in Woodslee soils, which had not differentiated the contents of SOC and not shaped different nature of organic matter among the treatments. This clearly reflects the need to include soil samples from different geographic and climatic zones with diverse cropping, tillage and fertilization histories, and with a wide range in the values of soil properties for a robust model (Zornoza et al., 2008).

Previous studies at field scale in Canada reported excellent estimation by NIR-PLSr for SOC ($R^2 = 0.86-0.98$, RPD = 2.9-3.6) and TN ($R^2 = 0.83-0.93$, RPD = 3.4-4.8) (Nduwanangu et al., 2009), or reliable estimation for SOC ($R^2 = 0.875$, RPD = 2.16) and poor estimation for soil organic N ($R^2 = 0.326$, RPD = 1.22) (Martin et al., 2002). Reliable estimation by NIR-PLSr for SOC ($R^2 = 0.86$, RPD = 2.65) and TN ($R^2 = 0.85$, RPD = 2.56) were found for soils from the eastern Canada (St Luce et al., 2012), and excellent estimation by VNIR-PLSr for SOC ($R^2 = 0.92$, RPD = 3.70) and TN ($R^2 = 0.93$, RPD = 3.83) were found for the soils from the western Canada (St Luce et al., 2014). A previous study carried out at the same site (Woodslee), including the 50 samples used in this study and 245 more samples from other field trials at the same experimental farm (CV for SOC = 60%; CV for TN 40%; ranging 3.4-36.5 g C kg⁻¹, 0.6-2.8 g N kg⁻¹), provided excellent predictions for SOC and TN ($R^2 = 0.96-0.99$ for MIR and $R^2 = 0.95-0.97$ for NIR, the RPD values > 4) (Xie et al., 2011). A recent study conducted at the same site with the 50 samples used in this study and 110 additional samples (0.32-16.0 g C kg⁻¹, 0.08-0.94 g N kg⁻¹), found that POM-C and POM-N were predicted very well using both MIR- and NIR-PLSr ($R^2 = 0.83-0.78$ for MIR and $R^2 = 0.87-0.84$ for NIR, the RPD values > 2) (Yang et al., 2012). And St Luce et al., (2014) also reported reliable estimation by VNIR-PLSr for POM-N ($R^2 = 0.87$, RPD = 2.82) in soils from the western Canada. These discrepancies among this study and previous studies may imply that the sample diversity and/or sample number had a profound influence on the reliability of the calibration model. Considering the worse performance of regional models compared to national models and the relatively small number of samples used for the local models (site based), we consider that cross regional models were better choices than the regional or local models to predict SOC, TN, POM-C and POM-N in bulk soil.

In terms of the effectiveness of MIR- and NIR-PLSr models for SOC and TN predictions, the MIR-PLSr derived calibration models were similar to the NIR-PLSr derived models at cross-regional and regional scales or weaker than the NIR-PLSr derived models at a local scale. Nevertheless, the MIR-PLSr models provided better predictions for SOC and TN than the NIR-PLSr models at various scales at the model validation stage. This is consistent with the conclusion that MIR was more suitable and produced more robust predictions for SOC and TN than the VIS or NIR when analyzing ground, dry soils under laboratory conditions (Reeves III, 2010; Viscarra Rossel et al., 2006). For POM-C and POM-N predictions, the MIR-PLSr models were generally poorer than the NIR-PLSr models and provided poorer predictions for POM-C and -N at a cross-regional scale and at the Lethbridge and Ottawa sites; however, MIR-PLSr models provided better predictions for POM-C and -N at a regional scale and at the three other field sites. Yang et al. (2012) reported that MIR models were better than NIR models for POM-C and POM-N predictions. An opposite result was found by Reeves III et al. (2006) whereby the MIR model yielded a poorer prediction for POM-C relative to the NIR model. No explanation has been offered to explain such inconsistency between MIR and NIR models for the predictions of POM-C and POM-N. However, this implies that more work is required to study the efficacy of MIR- and NIR-PLSr models for predicting/determining soil properties. Both MIR and NIR spectroscopy methods have the potential to be used in the PLSr model to estimate C and N in soil and POM, and MIR-PLSr produces more robust predictions for SOC and TN than NIR-PLSr, however, the performance of these two models types in predicting POM-C and -N appears variable depending on scale.

Conclusions

Calibration models developed using MIR and NIR spectra collected from bulk soils well estimated the SOC and TN concentrations in both whole soil and POM. This infrared spectroscopy technique significantly saved the time and cost for SOC and TN analysis and eliminates the need to pre-treat the soil to separate POM prior to analysis. The cross-regional models developed with both MIR- and NIR-PLSr performed better than the regional and local models for the predictions of C and N in both soil and POM. Mid-infrared spectroscopy was better at predicting SOC and TN than NIR spectroscopy and the robustness for predicting POM-C and -N varied with scales. The excellent predictions for C and N in soil and POM provided by either MIR- or NIR-PLSr at a cross regional scale implies that a large spectral library for determining and/or monitoring the change of soil organic matter in a timely, rapid and cost-saving manner is feasible across the key agriculture areas of Canada. The large set of soil samples would enable models to accurately predict C and N concentrations in whole soil and POM across different climatic zones and under widely differing agricultural practices.

Acknowledgement

Primary funding for this research was provided by the Agriculture and Agri-Food of Canada. Mary-Anne Reeb, John Goerzen, Dan Dilliot, Santiago Giron, Lucas Koechlin provided invaluable laboratory assistance.

References

- Bellon-Maurel, V., McBratney, A., 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. *Soil Bio. Biochem.* **43**, 1398-1410.
- Bornemann, L., Welp, G., Amelung, W., 2010. Particulate organic matter at the field scale: rapid acquisition using mid-infrared spectroscopy. *Soil Sci. Soc. Am. J.* **74**, 1147-1156.
- Brown, D.J., Brickley, R.S., Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma* **129**, 251-267.
- Cambardella, C.A., Elliott, E.T., 1992. Particulate Soil Organic-Matter Changes across a Grassland Cultivation Sequence. *Soil Sci. Soc. Am. J.* **56**, 777-783.
- Cécillon, L., Cassagne, N., Czarnes, S., Gros, R., Brun, J.-J., 2008. Variable selection in near infrared spectra for the biological characterization of soil and earthworm casts. *Soil Bio. Biochem.* **40**, 1975-1979.
- Christensen, B.T., 1992. Physical Fractionation of Soil and Organic Matter in Primary Particle Size and Density Separates, In: Stewart, B.A. (Ed.), *Advances in Soil Science*. Springer New York, New York, NY, pp. 1-90.
- Degryze, S., Six, J., Paustian, K., Morris, S.J., Paul, E.A., Merckx, R., 2004. Soil organic carbon pool changes following land-use conversions. *Global Change Biol.* **10**, 1120-1132.
- Dick, W.A., Culman, S.W., Chatterjee, A., Clay, D., 2016. Biological and Biochemical Tests for Assessing Soil Fertility. *Soil Fertility Management in Agroecosystems*. Amitava Chatterjee and David Clay, Editors.
- Du, C., Zhou, J., 2009. Evaluation of soil fertility using infrared spectroscopy: a review. *Environ. Chem. Lett.* **7**, 97-113.
- Gregorich, E.G., Beare, M.H., 2008. Physically uncomplexed organic matter. In M.R. Carter and E.G. Gregorich (ed.) *Can. Soc. Soil Science, Lewis Publisher, Soil sampling and methods of analysis*. 2nd ed. CRC Press, Boca Raton, FL., 607-616.
- Hussain, I., Olson, K.R., 2000. Recovery rate of organic c in organic matter fractions of Grantsburg soils. *Commun. Soil Sci. Plant Anal.* **31**, 995-1001.
- Islam, K., Singh, B., McBratney, A., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Aust. J. Soil Res.* **41**, 1101-1114.
- Janik, L.J., Forrester, S.T., Rawson, A., 2009. The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks (PLS-NN) analysis. *Chemometr. Intell. Lab. Syst.* **97**, 179-188.
- Janik, L.J., Merry, R.H., Skjemstad, J.O., 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? *Aust. J. Exp. Agr.* **38**, 681-696.
- Lal, R., 2004. Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. *Science* **304**, 1623-1627.
- Madari, B.E., Reeves, J.B., Machado, P.L.O.A., Guimarães, C.M., Torres, E., McCarty, G.W., 2006. Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. *Geoderma* **136**, 245-259.
- Martin, P.D., Malley, D.F., Manning, G., Fuller, L., 2002. Determination of soil organic carbon and nitrogen at the field level using near-infrared spectroscopy. *Can. J. Soil Sci.* **82**, 413-422.

- McCarty, G.W., Reeves III, J.B., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-Infrared and Near-Infrared Diffuse Reflectance Spectroscopy for Soil Carbon Measurement. *Soil Sci. Soc. Am. J.* **66**, 640-646.
- Mebius, L.J., 1960. A rapid method for the determination of organic carbon in soil. *Anal. Chim. Acta* **22**, 120-124.
- Meersmans, J., Van Wesemael, B., Van Molle, M., 2009. Determining soil organic carbon for agricultural soils: a comparison between the Walkley & Black and the dry combustion methods (north Belgium). *Soil Use Manage.* **25**, 346-353.
- Nduwamungu, C., Ziadi, N., Tremblay, G.F., Parent, L.-É., 2009. Near-infrared reflectance spectroscopy prediction of soil properties: effects of sample cups and preparation. *Soil Sci. Soc. Am. J.* **73**, 1896-1903.
- Reeves III, J.B., 2010. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma* **158**, 3-14.
- Reeves III, J.B., Follett, R.F., McCarty, G.W., Kimble, J.M., 2006. Can near or mid-Infrared diffuse reflectance spectroscopy be used to determine soil carbon pools? *Commun. Soil Sci. Plant Anal.* **37**, 2307-2325.
- Reeves III, J.B., McCarty, G.W., Reeves, V.B., 2001. Mid-infrared Diffuse Reflectance Spectroscopy for the Quantitative Analysis of Agricultural Soils. *J. Agric. Food Chem.* **49**, 766-772.
- Reeves III, J.B., Smith, D.B., 2009. The potential of mid-and near-infrared diffuse reflectance spectroscopy for determining major-and trace-element concentrations in soils from a geochemical survey of North America. *Appl. Geochem.* **24**, 1472-1481.
- Rinnan, R., Rinnan, Å., 2007. Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil. *Soil Bio. Biochem.* **39**, 1664-1673.
- Saeyens, W., Mouazen, A.M., Ramon, H., 2005. Potential for Onsite and Online Analysis of Pig Manure using Visible and Near Infrared Reflectance Spectroscopy. *Biosyst. Eng.* **91**, 393-402.
- Sankey, J.B., Brown, D.J., Bernard, M.L., Lawrence, R.L., 2008. Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. *Geoderma* **148**, 149-158.
- Shepherd, K.D., Walsh, M.G., 2002. Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Sci. Soc. Am. J.* **66**, 988-998.
- Six, J., Conant, R.T., Paul, E.A., Paustian, K., 2002. Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. *Plant Soil* **241**, 155-176.
- 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. *Appl. Spectrosc. Rev.* **49**, 139-186.
- St. Luce, M., Ziadi, N., Nyiraneza, J., Tremblay, G.F., Zebarth, B.J., Whalen, J.K., Laterrière, M., 2012. Near Infrared Reflectance Spectroscopy Prediction of Soil Nitrogen Supply in Humid Temperate Regions of Canada. *Soil Sci. Soc. Am. J.* **76**, 1454-1461.
- St. Luce, M., Ziadi, N., Zebarth, B.J., Grant, C.A., Tremblay, G.F., Gregorich, E.G., 2014. Rapid determination of soil organic matter quality indicators using visible near infrared reflectance spectroscopy. *Geoderma* **232-234**, 449-458.
- Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and Near Infrared Spectroscopy in Soil Science. *Adv. Agron.* **107**, 163-215.
- Stevens, A., van Wesemael, B., Bartholomeus, H., Rosillon, D., Tychon, B., Ben-Dor, E., 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma* **144**, 395-404.
- Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez,

- C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. *Earth Sci. Rev.* **155**, 198-230.
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **131**, 59-75.
- Walkley, A., 1946. A critical examination of a rapid method for determining organic carbon in soils- Effect of variations in digestion conditions and of inorganic soil constituents. *Soil Sci.* **63**, 251-264.
- Walkley, A., Black, I.A., 1934. An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci.* **37**, 29-38.
- Xie, H.T., Yang, X.M., Drury, C.F., Yang, J.Y., Zhang, X.D., 2011. Predicting soil organic carbon and total nitrogen using mid- and near-infrared spectra for Brookston clay loam soil in Southwestern Ontario, Canada. *Can. J. Soil Sci.* **91**, 53-63.
- Xie, H.T., Zhao, J., Wang, Q., Sui, Y., Wang, J., Yang, X.M., Zhang, X.D., Liang, C. 2015. Soil type recognition as improved by genetic algorithm-based variable selection using near infrared spectroscopy and partial least squares discriminant analysis. *Scientific Reports* **5**, 10930. <http://doi.org/10.1038/srep10930>
- Yang, H., Mouazen, A.M., 2012. Vis/near- and mid- infrared spectroscopy for predicting soil C and N at a farm scale. *Infrared Spectroscopy - Life and Biomedical Sciences*, Prof. Theophanides Theophile (Ed.), ISBN: 978-953-51-0538-1, InTech, 185-210.
- Yang, X.M., Xie, H.T., Drury, C.F., Reynolds, W.D., Yang, J.Y., Zhang, X.D., 2012. Determination of organic carbon and nitrogen in particulate organic matter and particle size fractions of Brookston clay loam soil using infrared spectroscopy. *Eur. J. Soil Sci.* **63**, 177-188.
- Zimmermann, M., Leifeld, J., Fuhrer, J., 2007. Quantifying soil organic carbon fractions by infrared-spectroscopy. *Soil Bio. Biochem.* **39**, 224-231.
- Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K.M., Arcenegui, V., Mataix-Beneyto, J., 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Bio. Biochem.* **40**, 1923-1930.

Figure Captions

Fig. 1. Averaged spectra of mid-infrared (MIR) and near-infrared (NIR) in each site, MIR spectra were base-line corrected.

Fig. 2. Predicted soil organic carbon (SOC) and total nitrogen (TN) values (g kg^{-1} soil) plotted against measured values (g kg^{-1} soil) and the statistics of model validation. The PLRs models were fitted using the MIR and NIR spectra collected from bulk soil samples.

Fig. 3. Predicted values (g kg^{-1} soil) of C and N in particulate organic matter (POM) plotted against measured values (g kg^{-1} soil) and the statistics of model validation. The PLRs models were fitted using the MIR and NIR spectra collected from bulk soil samples.

Table 1 Descriptive statistics of the concentrations of soil organic carbon (SOC), total nitrogen (TN), particulate organic matter carbon (POM-C) and nitrogen (POM-N)

Site	N	Mean	Min	Med	Max	SD	CV (%)	Mean	Min	Med	Max	SD	CV (%)
SOC (g C kg ⁻¹ soil)								TN (g N kg ⁻¹ soil)					
Lethbridge (L)	56	72.1	20.8	65.7	162	38.3	53.2	5.84	1.78	5.56	12.1	2.74	46.9
Agassiz (A)	80	42.2	25.3	42.4	63.3	9.53	22.6	3.30	1.71	3.36	5.77	0.811	24.5
Quebec (Q)	60	34.8	15.9	33.0	63.8	12.3	35.4	2.19	0.980	2.31	3.64	0.684	31.3
Ottawa (O)	30	24.0	10.3	20.7	50.8	10.9	45.3	2.18	0.950	1.91	4.40	0.887	40.7
Woodslee (W)	50	28.1	24.3	27.8	34.4	2.57	9.15	2.38	1.99	2.34	3.08	0.252	10.6
West Region (L+A)	136	53.7	20.8	46.0	161.7	28.8	53.6	4.28	1.71	3.55	12.1	2.19	51.1
East Region (W+O+Q)	140	30.1	10.3	27.8	63.8	10.5	34.9	2.26	0.950	2.27	4.40	0.628	27.8
Cross Region (All Sites)	276	41.4	10.3	34.6	162	24.4	58.8	3.23	0.950	2.68	12.1	1.88	58.1
POM-C (g C kg ⁻¹ soil)								POM-N (g N kg ⁻¹ soil)					
Lethbridge (L)	56	34.5	2.81	30.4	109	26.9	78.0	2.64	0.149	2.22	7.43	2.07	78.5
Agassiz (A)	80	9.04	4.62	8.32	16.8	2.62	29.0	0.594	0.275	0.533	1.17	0.211	35.6
Quebec (Q)	60	3.34	<0.001	2.94	7.57	1.80	53.9	0.408	<0.001	0.433	0.643	0.152	37.2
Ottawa (O)	30	5.91	1.46	4.08	21.7	4.92	83.2	0.350	0.016	0.223	1.61	0.358	102
Woodslee (W)	50	3.47	2.03	3.27	5.65	0.846	24.3	0.224	0.097	0.215	0.397	0.0678	30.2
West Region (L+A)	136	19.5	2.81	10.4	109	21.4	110	1.44	0.149	0.700	7.43	1.67	116
East Region (W+O+Q)	140	3.94	<0.001	3.26	21.7	2.78	70.6	0.330	<0.001	0.279	1.61	0.212	64.2
Cross Region (All Sites)	276	11.6	<0.001	6.18	109	17.0	146	0.875	<0.001	0.454	7.43	1.30	149

L: Lethbridge; A: Agassiz; Q: Quebec; O: Ottawa; W: Woodslee; N: Numbers of soil sample; Min: Minimum; Med: Median; Max: Maximum; SD: Standard Deviation; CV: Coefficient of Variation, the ratio of standard deviation to mean.

Table 2 Calibration models developed for prediction of SOC and TN using MIR and NIR spectroscopy

	Pre-processing	Greatest influence IR range / cm^{-1}	Rank	R^2	RMSEC / mg g^{-1}	RPD
Mid-infrared spectra:		Soil Organic Carbon (SOC)				
Lethbridge (L)	SLS	4003.6-3282.3, 2923.6-1477.2	10	0.95	9.92	4.33
Agassiz (A)	VN	4003.6-2919.7, 2561-2198.5, 1839.8-1477.2	4	0.89	3.28	2.98
Quebec (Q)	FD	3282.3-2919.7, 1839.8-756	3	0.93	3.26	3.87
Ottawa (O)	SD	4003.6-2198.5	10	0.98	2.05	6.54
Woodslee (W)	MSC	3644.9-2919.7, 1481.1-1118.5	3	0.13	2.48	1.07
West Region (L+A)	FD	4003.6-3282.3, 1839.8-1118.5	6	0.92	8.30	3.55
East Region (W+O+Q)	MSC	3644.9-3282.3, 1839.8-1118.5, 759.8-397.3	10	0.93	2.92	3.73
Cross Region (All Sites)	FD	3644.9-2561, 1839.8-756	10	0.92	6.87	3.61
		Total Nitrogen (TN)				
Lethbridge (L)	NO	4003.6-1477.2, 1118.5-756	10	0.96	0.61	4.81
Agassiz (A)	MMN	4003.6-2561, 2202.3-1477.2	6	0.80	0.37	2.26
Quebec (Q)	MSC	4003-3641, 3282-2919, 2561-2198, 1484.1-1118.5	4	0.90	0.22	3.21
Ottawa (O)	SD	4003.6-2198.5	10	0.98	0.15	7.10
Woodslee (W)	MSC	4003.6-3282.3, 1481.1-1118.5	3	0.06	0.24	1.03
West Region (L+A)	FD+VN	4003.6-3282.3, 2561-756	5	0.92	0.58	3.65
East Region (W+O+Q)	MSC	1839.8-1118.5, 759.8-397.3	4	0.81	0.28	2.30
Cross Region (All Sites)	FD+SLS	3644.9-2561, 1839.8-756	10	0.93	0.51	3.76
Near-infrared spectra:		Soil Organic Carbon (SOC)				
Lethbridge (L)	SLS	5450.9-4599.5	10	0.95	9.85	4.37
Agassiz (A)	FD+VN	7998.5-7048.7, 5450.9-4249.5	10	0.99	1.20	8.48
Quebec (Q)	FD	7998.5-7500, 6100.8-5450, 4600.4-4249.5	10	0.99	0.41	29.3
Ottawa (O)	MMN	7749.7-7500, 4425-4249.5	10	0.99	1.28	10.5
Woodslee (W)	SD	5775.9-5450, 4425-4249.5	2	0.57	1.72	1.52
West Region (L+A)	MSC	7500.9-6099.9, 4600.4-4249.5	6	0.91	8.95	3.29
East Region (W+O+Q)	VN	7998.5-7748.7, 5450.9-4249.5	10	0.87	3.86	2.82
Cross Region (All Sites)	FD+SLS	7749.7-7500, 5450.9-4249.5	10	0.93	6.36	3.90
		Total Nitrogen (TN)				
Lethbridge (L)	COE	5450.9-4249.5	7	0.95	0.62	4.54
Agassiz (A)	FD+VN	7500.9-6099.9, 5450.9-4249.5	10	0.99	0.09	8.76
Quebec (Q)	FD+MSC	7998.5-6723.7, 4600.4-4249.5	10	0.99	0.03	27.2
Ottawa (O)	MMN	7749.7-7500, 4425-4249.5	10	0.99	0.10	10.7
Woodslee (W)	SD	7998.5-7500, 4600.4-4249.5	3	0.84	0.10	2.50
West Region (L+A)	VN	7500.9-6099.9, 4850.2-4249.5	8	0.93	0.56	3.85
East Region (W+O+Q)	VN	7998.5-6099.9, 5450.9-4249.5	10	0.79	0.30	2.16
Cross Region (All Sites)	FD+SLS	7998.5-4249.5	10	0.95	0.43	4.42

Rank is the number of partial least square vectors used in models; R^2 = Coefficient of determination; RMSEC = Root Mean Square Error of Calibration; RPD = Residual Prediction Deviation; COE = Constant Offset Elimination; FD = First Derivative; NO = No Spectral Pre-processing; MMN = Min-Max Normalization; MSC = Multiplicative Scattering Correction; SD = Second Derivative; SLS = Straight Line Subtraction; VN = Vector Normalization.

Table 3 Calibration models developed for predicting POM-C and POM-N using MIR and NIR spectra collected from the bulk soil samples

	Pre-processing	Greatest influence IR range / cm^{-1}	Rank	R^2	RMSEC / mg g^{-1}	RPD
Mid-infrared spectra:		Particulate organic carbon (POM-C)				
Lethbridge (L)	FD+MSC	3282.3-2198.5, 1839.8-1477.2	10	0.98	3.51	7.94
Agassiz (A)	SD	3644.9-2198.5, 1839.8-1477.2	10	0.94	0.67	4.25
Quebec (Q)	MMN	3282.3-2919.7, 2561-756	10	0.93	0.52	3.77
Ottawa (O)	FD+SLS	4003.6-2561, 1839.8-1477.2	10	0.99	0.69	8.79
Woodslee (W)	FD+VN	3282.3-2198.5	2	0.29	0.73	1.19
West Region (L+A)	FD+VN	3282.3-2919.7, 2561-1839.8	7	0.97	3.38	6.06
East Region (W+O+Q)	FD+VN	4003.6-3282.3, 2923.6-2561, 2202.3-1477.2	8	0.91	0.86	3.34
Cross Region (All Sites)	FD+VN	3644.9-2919.7, 2561-1839.8	9	0.96	3.08	5.23
		Particulate organic nitrogen (POM-N)				
Lethbridge (L)	FD+MSC	4003.6-3641, 3282.3-2198.5, 1839.8-1477.2	10	0.99	0.25	9.08
Agassiz (A)	SD	3644.9-3282.3, 2923.6-2198.7, 1839.8-1118.5	10	0.94	0.06	3.95
Quebec (Q)	FD+MSC	4003.6-3641, 2923.6-2198.5, 1481.1-1118.5	10	0.90	0.05	3.25
Ottawa (O)	FD+VN	3282.3-2919.7, 2202.3-1477.2	10	0.97	0.07	6.26
Woodslee (W)	FD+SLS	3282.3-2919.7, 2561-1839.8	3	0.29	0.06	1.19
West Region (L+A)	FD+MSC	3282.3-2198.5, 1839.8-1477.2	8	0.98	0.25	6.81
East Region (W+O+Q)	FD	3282.3-2561, 2202.3-1477.2	7	0.83	0.09	2.43
Cross Region (All Sites)	FD	4003.6-3641, 2923.6-2198.5	10	0.96	0.27	4.77
Near-infrared spectra:		Particulate organic carbon (POM-C)				
Lethbridge (L)	FD+VN	7749.7-7500, 6100.8-5774.9, 4425-4249.5	10	0.99	1.73	16.1
Agassiz (A)	FD+MSC	7998.5-6099.9, 4600.4-4249.5	10	0.99	0.168	16.7
Quebec (Q)	VN	7500.9-6099.9, 5450.9-4599.5	10	0.96	0.410	4.81
Ottawa (O)	COE	7049.7-6099.9, 4600.4-4249.5	10	0.99	0.263	23.1
Woodslee (W)	FD	7500.9-6099.9	4	0.31	0.70	1.20
West Region (L+A)	FE+VN	4600.4-4249.5	10	0.97	3.28	6.30
East Region (W+O+Q)	FD	6100.8-4249.5	10	0.95	0.619	4.67
Cross Region (All Sites)	FD+SLS	5450.9-4249.5	9	0.97	2.70	5.96
		Particulate organic nitrogen (POM-N)				
Lethbridge (L)	FD+VN	5775.9-5450, 4600.4-4249.5	10	0.99	0.21	11.0
Agassiz (A)	FD+MSC	7998.5-6099.9, 4600.4-4425	10	0.99	0.02	15.1
Quebec (Q)	VN	7998.5-7500, 4600.4-4249.5	10	0.93	0.04	3.72
Ottawa (O)	FD+MSC	4006.4-4249.5	10	0.99	0.03	15.7
Woodslee (W)	FD+SLS	7500.9-6099.9	3	0.44	0.05	1.34
West Region (L+A)	FD+VN	4600.4-4249.5	7	0.98	0.24	6.86
East Region (W+O+Q)	FD+VN	4600.4-4249.5	10	0.89	0.07	3.07
Cross Region (All Sites)	FD	5450.9-4249.5	9	0.97	0.20	6.31

Rank is the number of partial least square vectors used in models; R^2 = Coefficient of determination; RMSEC = Root Mean Square Error of Calibration; RPD = Residual Prediction Deviation; COE = Constant Offset Elimination; FD = First Derivative; MMN = Min-Max Normalization; MSC = Multiplicative Scattering Correction; SD = Second Derivative; SLS = Straight Line Subtraction; VN = Vector Normalization.

148x209mm (300 x 300 DPI)

148x209mm (300 x 300 DPI)

148x209mm (300 x 300 DPI)