



**Infrared spectroscopy estimation methods for water-dissolved carbon and amino sugars in diverse Canadian agricultural soils**

Journal:	<i>Canadian Journal of Soil Science</i>
Manuscript ID	CJSS-2018-0027.R1
Manuscript Type:	Article
Date Submitted by the Author:	17-May-2018
Complete List of Authors:	Zhang, Lei; Institute of Applied Ecology Chinese Academy of Sciences, Yang, X.; AAFC Drury, Craig; Agriculture & Agri-Food Canada, Chantigny, M.; Agriculture Canada, Gregorich, E.; Agriculture Canada, Miller, Jim; Agriculture and Agri-Food Canada, Bittman, Shabtai; AAFC, Reynolds, W.D.; Agriculture & Agri-Food Canada, Research Branch Yang, Jingyi; Agriculture and Agri-Food Canada, Greenhouse and Processing Crops Research Centre, Canada,
Keywords:	Organic carbon, Soil water dissolved carbon, Amino sugars, Infrared spectroscopy, Modeling prediction
Is the invited manuscript for consideration in a Special Issue?:	Not applicable (regular submission)

SCHOLARONE™  
Manuscripts

**Infrared spectroscopy estimation methods for water-dissolved carbon and amino sugars  
in diverse Canadian agricultural soils**

Lei Zhang<sup>1,6,7</sup>, Xueming Yang<sup>1,\*</sup>, Craig Drury<sup>1</sup>, Martin Chantigny<sup>2</sup>, Edward Gregorich<sup>3</sup>, Jim Miller<sup>4</sup>, Shabtai Bittman<sup>5</sup>, W. Dan Reynolds<sup>1</sup>, Jingyi Yang<sup>1</sup>

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

<sup>2</sup> Quebec Research and Development Centre, AAFC, Quebec, Quebec G1V 2J3, Canada

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

<sup>4</sup> Lethbridge Research and Development Centre, AAFC, Lethbridge, Alberta T1J 4B1, Canada

<sup>5</sup> Agassiz Research and Development Centre, AAFC, Agassiz, British Columbia V0M 1A0, Canada

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

<sup>7</sup> University of Chinese Academy of Sciences, 19 Yuquan Road, Shijingshan, Beijing 100049, China

\*Corresponding author. E-mail: [xueming.yang@agr.gc.ca](mailto:xueming.yang@agr.gc.ca)

**Abstract**

Infrared spectroscopy has the potential to rapidly analyse soil water-dissolved carbon and amino sugars. In this study, mid-infrared (MIR) and near-infrared (NIR) spectra collected from soil water extracts or from bulk soils were analysed with partial least squares regression (PLSr) to estimate the concentrations of water-dissolved carbon and amino sugars in diverse agricultural soils collected from 5 field sites in two Western and two Eastern Canadian provinces. The MIR-PLSr models developed from soil water extract spectra estimated hot-water (100 °C) dissolved carbon (HWDC) [ $R^2 = 0.97-0.70$ , Ratio of Prediction to Deviation (RPD) = 6.13-1.83] well but the MIR-PLSr models did not estimate cold-water (21 °C) dissolved carbon (CWDC) well ( $R^2=0.82-0.50$ , RPD=2.35-1.42). The model estimates

of HWDC at the multi-site scale (all samples together) and for the two Western Canada sites ( $R^2=0.97-0.93$ , RPD = 6.13-3.68) surpass the modal estimates for the three Eastern Canadian sites ( $R^2=0.80-0.70$ , RPD=2.28-1.83). The MIR- and NIR-PLSr models derived from bulk soil spectra both estimated HWDC well at the multi-site scale ( $R^2=0.91-0.88$ , RPD=2.90-2.32) and for the Western Canada sites ( $R^2=0.90-0.87$ , RPD=3.18-2.89). Models developed from hot-water extract spectra and bulk soil spectra resulted in poor estimates of soil amino sugars ( $R^2=0.74-0.21$ , RPD=1.89-1.12), except for the approximate quantitative estimation of muramic acid by models based on soil spectra at the Western and the multi-site scale ( $R^2=0.80-0.82$ , RPD=2.21-2.33). We concluded that MIR/NIR models at regional and multi-site scales can be used as a tool to monitor HWDC but that additional research is required for estimating soil amino sugars.

**Keywords:** Infrared spectra, organic carbon, soil water-dissolved carbon, amino sugars, modeling prediction.

## Introduction

Soil organic matter (SOM) has been a focus of scientific research because of its pivotal role for assessing soil quality (Gregorich et al., 1994; Reeves, 1997; Shukla et al., 2006) and climate change (Lal, 2004). Soil management practices including tillage, crop rotation, and/or amendment use could significantly change the quality and quantity of SOM. However, these changes occur slowly over time and are difficult to detect in the short term (Evans et al., 2001; Haynes, 2005). Nevertheless, the labile fractions of soil carbon (C) and nitrogen (N) species have been suggested as sensitive indicators of management changes (Liang et al., 1998; Bending et al., 2000; Haynes, 2000; Haynes, 2005). Soil water-dissolved organic C (OC) is one of the most active and labile organic matter fractions in soil (Bu et al., 2011), and is extremely sensitive to soil management change (Chantigny, 2003; Ghani et al., 2003). Water-dissolved OC (equal to the difference between water-dissolved total carbon (TC) and water-dissolved

inorganic carbon (IC)) in soils can be extracted using either hot water (i.e. soil extracted with 80°C or boiling water) or cold water (i.e. soil extracted with room temperature water) (Evans et al., 2001; Gregorich et al., 2003; Landgraf et al., 2006; Bu et al., 2011). The cold water-dissolved C (CWDC) is closely correlated to dissolved OC measured directly in soil using lysimeters or suction cups (Rees and Parker, 2005). The hot water-dissolved C (HWDC) consists of more stable components that function as a reserve of nutrients and energy for plants and microorganisms (Bu et al., 2011). Compared with the CWDC, the HWDC pool is bigger (5- to 10-fold) and represents a more heterogeneous C pool (Hamkalo and Bedernichek, 2014). This is because hot ( $\geq 70$  °C) water kills vegetative cells of microorganisms and extracts many components from microbial biomass, as well as many nonmicrobial soil organic substances (Landgraf et al. 2006).

Soil amino sugars are components of soil organic matter that are predominantly of microbial origin (Parsons 1981; Stevenson 1983; Amelung, 2001; Amelung et al., 2018) and are relatively stable over time (Chantigny et al. 1997). Soil amino sugars contribute 5%-12% of soil organic N (Stevenson, 1983; Amelung, 2001) and have been considered as plant available N. Amino sugars have been taken into account when developing N recommendations for crops (Khan et al., 2001; Mulvaney et al., 2001). The three most important amino sugars in soil are muramic acid (MurA), glucosamine (GluN), and galactosamine (GalN) (Amelung, 2001; Joergensen and Wichern, 2008). Muramic acid originates from bacterial cell walls (Parsons, 1981; Amelung, 2001), GluN is the most abundant amino sugar in soil and it is primarily derived from the chitin of fungal cell walls (Parsons, 1981; Amelung et al., 1999). In contrast little is known about the origin of GalN although it accounts for 20-40% of the total amino

sugar (TAS) pool (Amelung et al., 1999; Joergensen et al., 2010). The total amino sugar (TAS) content of soil is commonly used for estimating the microbial origin of N in soil, and the GluN and MurA concentrations are used to separate the contributions of fungi and bacteria to microbial-derived SOM (Guggenberger et al., 1999; Joergensen et al., 2010). Therefore, both the concentrations and types of amino sugars could be useful parameters to elucidate the fate of microbial-derived SOM and reflect the SOM status/changes under different management practices.

The routine method for analyzing soil water-dissolved C includes water extraction followed by C analysis using either wet digestion or dry combustion methods. However, amino sugar analysis is considerably more involved as it involves a hydrolysis step with 6 M HCl for 3 to 8 h (Zelles, 1988; Zhang and Amelung, 1996) and a complex derivatization procedure (Zelles, 1988; Zhang and Amelung, 1996; Appuhn et al., 2004) or an extra purification process (Kaiser and Benner, 2000). These steps involve using and disposing of hazardous chemicals including dichloromethane (Zhang and Amelung, 1996). Gas chromatography, high-performance liquid chromatography or high-performance anion-exchange chromatography can be used to separate and determine amino sugars in hydrolysates. All of these analytical methods are very tedious, time consuming and costly, and need specialized equipment and experienced laboratory technicians. Therefore, these assays have been considered too complicated and time-consuming for use in routine soil testing (Khan et al., 2001). A simple soil test for amino sugars and soluble organic C would be extremely helpful in evaluating the impacts of changes to soil and crop management on soil health and environmental quality.

Estimation models using infrared (IR) spectroscopy with chemometrics such as partial least-squares regression (PLSr) may be an alternative approach to estimating soil properties in a quick and inexpensive manner (Shepherd and Walsh, 2002; Brown et al., 2006). These techniques have been used for determining numerous soil physical, chemical and biological properties (Janik et al., 1998; Malley et al., 2004; Rinnan and Rinnan, 2007; Terhoeven-Urselmans et al., 2008; Zornoza et al., 2008; Janik et al., 2009; Ludwig et al., 2015). The main advantage of this technique is its ability to quickly obtain quantitative data for the property of interest without the use of chemical reagents and associated waste disposal issues. In addition, this technique is less costly and requires less labor for sample preparation than the traditional wet-chemistry based techniques. Before this technique can be used on a regular basis, a large number of samples with a wide range in soil properties of interest are required for model calibration (Stenberg et al., 2010; Soriano-Disla et al., 2014) to assure the robustness and applicability of the IR spectroscopy models (Zornoza et al., 2008; Du and Zhou, 2009). Reeves III (2010) noted that the usefulness of the calibration models at different scales (field, regional, global) is directly related to what basis (e.g. analyte concentrations, soil texture, or combinations thereof) soil samples are included in a single calibration. Soils from different agricultural zones under different management practices have extremely different characteristics which could be associated with significant variations in the IR spectra. Thus, there is a need to develop models using more diverse sample sets (Stevens et al., 2008; Stenberg et al., 2010; Soriano-Disla et al., 2014). Such models for soil property estimations have been developed for the soils in Africa (Shepherd and Walsh, 2002), Australia (Rossel and Webster, 2012), and for an even larger global set of samples (Brown et al., 2006; Viscarra

Rossel et al., 2016). However, a large number of samples require huge sampling effort and produce expensive analytical costs of reference data, and the model based on a large number of samples may be not accurately suitable to relatively homogeneous areas or to small-scale applications (Soriano-Disla et al., 2014).

Until recently, limited efforts have been made to test if IR-PLSr modeling can estimate water-dissolved C using NIR spectra collected from bulk soils (Vasques et al., 2009; Vohland and Emmerling, 2011). The use of this technique in estimating soil amino sugar were only reported in two studies (1) using NIR spectroscopy collected from bulk soils (Dick et al., 2013) and (2) using MIR spectroscopy collected from soil hydrolysates (Zhang et al., 2013). Since water-dissolved C and amino sugar are both labile organic fractions in soils that could also exist in water extracts. The existence of these materials could be reflected in the IR spectra of bulk soils or soil water extracts. Therefore, we hypothesise that both spectra collected from bulk soils and water extracts could be used to estimate water-dissolved C and amino sugars by IR-PLSr techniques quickly and efficiently. To extend the scope of IR-PLSr from a local scale to a large scale, the objectives of this study were to: 1) evaluate the reliability of spectral models for estimating the C concentrations in hot- and cold-water extracts by using the MIR spectra collected from water extracts and the MIR/NIR spectra collected from bulk soils from different sites and ecoregions in Canada; 2) test the possibility that the concentrations of amino sugars could be estimated by using MIR spectra collected from hot-water extracts and by using NIR/MIR spectra collected from bulk soils and 3) compare how the efficiency and accuracy of estimate vary at different sampling scales, i.e., from the Western- or Eastern-Canada regions to the multi-sites (pooling all samples from 5 sites).

## Materials and methods

### 2.1 Soil sampling

Soil samples (0-10 cm or 0-15 cm, vary with locations) were collected from five Agriculture and Agri-Food Canada (AAFC)'s long-term field study sites in the fall of 2014. The two experimental sites located in the Western Canada were (i) Agassiz, British Columbia (BC, 49°10' N, 125°15' W) including 80 samples from a manure experiment, (ii) Lethbridge, Alberta (AB, 49°38' N, 112°48' W) including 56 samples from a fertilizer and organic amendment study. The other three experimental sites located at the Eastern Canada were (iii) Woodslee, Ontario (ON, 42°13' N, 82°44' W) including 50 samples from a compost study, (iv) Ottawa, Ontario (ON, 45°18' N, 75°43' W) and included 30 samples from a rotation and fertilization study, and (v) Québec City, Québec (QC, 46°48' N, 71°23' W) including 60 samples from a tillage and organic-mineral fertilization study. Altogether, there were a total of 276 samples. The experimental sites encompassed a wide diversity of soil types and included numerous management practices, which ensured that a wide range of SOM concentrations and a wide range of water-dissolved C and amino sugar concentrations were included in the dataset for developing IR-PLSr models (Table 1).

### 2.2 Water-dissolved C and amino sugar analysis

Soil hot or cold water-dissolved C were obtained by shaking 5 g of air dry soil (< 2 mm) with 50 mL distilled water bath at either 100 °C or at room temperature (21 °C) at 120 oscillations min<sup>-1</sup> for 1 hour. The suspension was then centrifuged at relative centrifugal force (RCF) = 3951 g for 30 minutes. The supernatant was vacuum filtered, then passed through a



0.45  $\mu\text{m}$  filter and the concentrations of water-dissolved TC and IC were determined using a Shimadzu TOC-LCPH analyser (Shimadzu TOC-L). Water-dissolved OC content was assessed as the difference between TC content and IC content.

Soil amino sugars including GluN, GalN and MurA, were hydrolyzed, purified, and analysed using the method described by Zhang and Amelung (1996) and Zhang et al. (2013). Briefly, 300 mg finely ground sample was hydrolyzed with 10 mL 6 M HCl at 105 °C for 8 h. The solution was filtered, evaporated, and then purified by neutralization and centrifugation (RCF = 988 g for 10 min). The supernatant was decanted and freeze-dried, and the residue was washed with methanol and then centrifuged again. The methanol supernatant was transferred to a 3 mL conical reaction vial and dried by purging with a gentle stream of purified N<sub>2</sub> gas. The residue was then re-dissolved with 1 mL double distilled water. The freeze-dried sample set was transformed into aldononitrile derivatives by reacting at 75-80 °C with derivatization reagent (containing 32 mg mL<sup>-1</sup> hydroxylamine hydrochloride and 40 mg mL<sup>-1</sup> 4-dimethylamino-pyridine in a 4:1 pyridine/methanol solvent) for 35 minutes and then with acetic anhydride for another 25 minutes. The aldononitrile derivatives were extracted from the aqueous solution with dichloromethane. Excess acetic anhydride was destroyed with 1 M HCl and double-distilled water. After drying the extracts with a N<sub>2</sub> stream, the amino sugar derivatives were re-dissolved in a hexane and ethyl acetate solvent (v/v=1:1) and separated on a Varian 450 GC equipped with an HP-5 capillary column (30 m  $\times$  0.32 mm  $\times$  0.25  $\mu\text{m}$ ) and quantified using a flame ionization detector. Amino sugars were quantified based on the internal standard, Myoinositol, which was added prior to purification. Methyl-glucosamine was used as a recovery standard to monitor recovery efficiency. Total

amino sugar concentration was calculated as the sum of the concentrations of GluN, GalN, and MurA. The concentration of MurA in many soils from Eastern Canada (Quebec, Ottawa and Woodslee sites) were not detected, so the samples from the Western Canada plus a small number of samples (30 samples) from the Eastern Canada sites were used to develop the calibration models for MurA estimation. The low concentration of MurA in soils has also been reported in previous studies (Miliar and Casida, 1970; Balkwill et al., 1988).

### 2.3 MIR and NIR spectral measurements

Soil hot- and cold-water extracts were used for MIR transmission spectra collection. Two drops of the solution sample (30  $\mu\text{L}$ ) were applied to the center of imprinted marks on the standardized microtiter plates (Bruker Optik GmbH, Ettlingen, Germany) with a Thermo scientific pipette. Three replicates were prepared for each sample. Two drops of double distilled water (30  $\mu\text{L}$ ) were prepared and used for background spectrum. Due to the high surface tension of specifically designed plate (Bruker Optik GmbH, Ettlingen, Germany), solution drops could be applied precisely on the imprinted mark and formed uniform films when they were dry (40  $^{\circ}\text{C}$ ) (Fan et al., 2012). MIR transmission spectra were obtained using a TENSOR 37 spectrometer (Bruker Optik GmbH, Germany) equipped with HTS-XT micro-plate reader (Bruker Optik GmbH, Ettlingen, Germany) in the range of 4000-400 $\text{cm}^{-1}$  with 64 scans and 8  $\text{cm}^{-1}$  resolution (Fig. 1). The MIR spectra were baseline corrected and three scans of each sample were averaged. The MIR spectra of hot- and cold-water extracts were used in model development to fit the concentrations of OC, IC and TC in hot- and cold-water extracts determined with the Shimadzu TOC-LCPH Analyser (Shimadzu, Japan). The MIR spectra of hot-water extracts were also used in model development to fit the

concentrations of GluN, GalN, MurA and TAS in bulk soil determined with the GC method (Zhang and Amelung, 1996).

Both MIR and NIR spectra were also collected from the bulk soil samples. The stainless steel sample cups were filled with air-dried and ground ( $< 0.15$ -mm) soil samples ( $\sim 1$ -g), and the sample surface in the cup was leveled using a flat spatula. Three separate MIR 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 spectra were recorded in the range of 4000 to  $400\text{ cm}^{-1}$  (2500 - 25000 nm) with 64 co-added scans and  $2\text{ cm}^{-1}$  resolution for MIR and in the range of  $8000 - 4000\text{ cm}^{-1}$  (1250 - 2500 nm) with 16 co-added scans and  $1\text{ cm}^{-1}$  resolution for NIR. Data were displayed as pseudo-absorbance ( $\log [1/\text{reflectance}]$ ). Three spectra for each sample were collected and the averaged spectrum from these three spectra was used in model development to fit the measured concentrations of hot-water-dissolved C (OC, IC and TC) and amino sugars (GluN, GalN, MurA and TAS). In addition, the MIR spectrum was baseline corrected and the original NIR spectrum was used in model calibration and validation.

#### 2.4 Model development and statistics

Model calibration was performed using PLSr with the OPUS QUANT 2 software by relating the pre-processed spectra data with the measured concentrations of hot- and cold-water-dissolved C (OC, IC and TC) and amino sugars (GluN, GalN, MurA and TAS). In the process of model development, spectra data were 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 minimum root mean square error of calibration (RMSEC) for each optimization was selected and validated using the “leave-one-out” cross-validation procedure. We report and discuss 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 (Brunet et al., 2007; Cozzolino et al., 2018; Wenger and Olden, 2012). The IR-PLS<sub>r</sub> models were tested at two scales: a multi-sites model (developed by using the complete dataset) and regional models (developed by pooling samples from the Western Canada sites or the Eastern Canada sites).

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:

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

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

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

where  $y_i$ ,  $\bar{y}$  and  $\hat{y}_i$  represent measured values, mean of measured values and estimated values of soil attributes of interest, respectively, of water-dissolved C and amino sugars;  $n$  is the number of the estimated/measured values and  $SD$  is the standard deviation of the measured data.

To evaluate the robustness of the 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 poor for a semi-quantitative prediction, a value for RPD between 2 and 2.5 can be considered 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.

### 3 Results

#### 3.1 Descriptive statistics of water-dissolved C and amino sugar concentrations in soil

Descriptive statistics of the concentrations of HWDCs and CWDCs (OC, IC and TC) and amino sugars (GluN, GlaN, MurA and TAS) in soil are presented in Table 1. Hot-water-dissolved OC, IC and TC concentrations varied from 183 to 5840 mg C kg<sup>-1</sup>, 5.48 to 422 mg C kg<sup>-1</sup> and 194 to 6010 mg C kg<sup>-1</sup>, respectively, among all samples. The cold-water-dissolved OC, IC and TC concentrations were considerably less than those in the hot-water extracts, ranging from 64.1 to 1520 mg C kg<sup>-1</sup>, 0.200 to 35.2 mg C kg<sup>-1</sup> and 66.1 to 1530 mg C kg<sup>-1</sup>, respectively, for the corresponding OC, IC and TC species. The soils in Western Canada had higher concentrations of OC, IC and TC in hot-water extracts (mean 1060, 68.3 and 1130 mg C kg<sup>-1</sup>, respectively) than those in Eastern Canada (mean 394, 16.9 and 410 mg C kg<sup>-1</sup>, respectively), and also the soils in Western Canada had greater

concentrations of OC, IC and TC in cold-water extracts (mean 331, 4.52 and 336 mg C kg<sup>-1</sup>, respectively) than those in Eastern Canada (mean 138, 2.27 and 140 mg C kg<sup>-1</sup>, respectively). Soil TAS concentrations varied from 700 to 6600 mg N kg<sup>-1</sup>, and GluN concentrations had the highest values (485-4030 mg N kg<sup>-1</sup>), followed by GalN at intermediate concentrations (188-2051 mg N kg<sup>-1</sup>), whereas MurA was either undetectable in many soil samples or presented at relatively low levels (44.5-1430 mg N kg<sup>-1</sup>) amongst other samples. Greater concentrations of GluN, GalN, MurA and TAS were observed for the soil samples from BC and AB in Western Canada (mean 1591, 818, 345 and 2750 mg N kg<sup>-1</sup>, respectively) compared to the soils from ON and QC in Eastern Canada (mean 1110, 623, 141 and 1770 mg N kg<sup>-1</sup>, respectively). In addition, the coefficients of variation (CVs) of soil water-dissolved C (OC, IC and TC) and amino sugar (GluN, GalN, MurA and TAS) concentrations were higher for the soils from the Western region than from Eastern Canada.

### 3.2 IR-PLSr Models for estimating water-dissolved C in soil

Hot- and cold-water-dissolved C (OC, IC and TC) concentrations were fitted to MIR-PLSr models, respectively. Mid-IR spectral bands and spectra processing procedures employed to model calibration were the same for water-dissolved OC and TC which differed from those for water-dissolved IC. The statistical parameters of MIR-PLSr models (water extracts) for the estimations of water-dissolved C varied among C species (OC, IC or TC) and sampling scales (multi-sites, Western region or Eastern region) (Table 2). The models based on hot-water extract spectra yielded better calibration results ( $R^2 = 0.81-0.98$ , RPD = 2.29-7.96) than the models based on cold-water extract spectra ( $R^2 = 0.66-0.85$ , RPD = 1.73-2.60) (Table 2). Excellent calibrations were obtained for soil HWDCs at each scale ( $R^2 > 0.89$ , RPD > 3), except the calibration for hot-water-dissolved IC in soils from Eastern Canada which had a good calibration ( $R^2 = 0.81$ , RPD = 2.29). The calibration models for

estimating hot-water-dissolved OC and TC ( $R^2 = 0.89-0.98$ , RPD = 3.04-7.96) were similar to or better than the model for hot water-dissolved IC ( $R^2 = 0.81-0.95$ , RPD = 2.29-4.78). For HWDCs estimations, multi-site models ( $R^2 = 0.95-0.98$ , RPD = 4.78-7.26) were functioning similarly to the models based on the soils from the Western region ( $R^2 = 0.94-0.98$ , RPD = 4.12-7.96), and were superior to the models based on the soils from the Eastern region ( $R^2 = 0.81-0.89$ , RPD = 2.29-3.10). Approximate quantitative calibrations were obtained for CWDCs in soils at each scale ( $R^2 > 0.76$ , RPD > 2), except for the poor calibration for cold water-dissolved IC in the soil from Eastern region ( $R^2 = 0.66$ , RPD = 1.73). Similar calibration models were found for cold water-dissolved OC and TC ( $R^2 = 0.76-0.83$ , RPD = 2.04-2.44). Multi-site calibration models ( $R^2 = 0.82-0.83$ , RPD = 2.35-2.44) and the Western region models ( $R^2 = 0.83-0.85$ , RPD = 2.42-2.60) performed similarly well for the CWDCs, and were superior to the calibration models using the soils from Eastern region ( $R^2 = 0.67-0.76$ , RPD = 1.73-2.05).

In general, the performance of hot-water extraction spectra derived models ( $R^2 = 0.70-0.97$ , RPD = 1.83-6.13) were better than the cold-water extraction spectra derived models ( $R^2 = 0.50-0.82$ , RPD = 1.42-2.35) for water-dissolved C estimates (Fig. 2). The models for the Eastern region provided poor or approximate quantitative estimates for HWDCs ( $R^2 = 0.70-0.81$ , RPD = 1.83-2.28) compared with the excellent estimations for HWDCs by hot-water spectra models on the multi-site or the Western region ( $R^2 = 0.93-0.98$ , RPD = 3.68-6.13), cold-water spectra derived models yielded approximate quantitative estimates of the CWDCs at multi-site scale or for the soils from the Western region ( $R^2 = 0.78-0.82$ , RPD = 2.13-2.35), however, the estimates were poor for the soils from the Eastern region ( $R^2 = 0.50-0.60$ , RPD = 1.42-1.58).

Calibration models derived from MIR and NIR soil spectra for estimates of HWDCs fitted the hot-water-dissolved data very well ( $R^2 = 0.91-0.96$ , RPD = 3.37-4.93), except for the hot-water-dissolved IC of the Eastern soils ( $R^2 = 0.72-0.77$ , RPD = 1.90-2.07, Table 3) which was not surprising as IC was generally low in Eastern Canadian soils. No other clear differences in the statistical parameters of calibration models were found among different C

species (OC, IC or TC) and sampling scales (multi-site, Western region or Eastern region, Table 3). Nevertheless, the NIR-PLSr models with  $R^2 = 0.72-0.96$  and  $RPD = 1.90-4.93$  were slightly better than the MIR-PLSr models with  $R^2 = 0.77-0.93$  and  $RPD = 2.07-3.84$  for predictions of HWDCs.

Models developed using soil MIR and NIR spectra both estimated the hot water-dissolved OC, IC and TC in soils very well at each sampling scale ( $R^2 = 0.82-0.90$ ,  $RPD = 2.34-3.18$ ) except for the hot water-dissolved IC in soils from the Eastern region ( $R^2 = 0.47-0.72$ ,  $RPD = 1.37-1.89$ ) (Fig. 3). Generally, the MIR-PLSr models provided similar estimates for hot water-dissolved TC and OC ( $R^2 = 0.86-89$ ,  $RPD = 2.64-3.17$ ) as the NIR-PLSr models ( $R^2 = 0.82-0.88$ ,  $RPD = 2.39-2.94$ ).

### 3.3 IR-PLSr Models for predicting soil amino sugar concentrations

Three types of IR spectra, including MIR spectra from hot-water extracts and MIR/NIR spectra from bulk soil samples, were used to fit PLSr models to estimate the concentrations of amino sugar species (GluN, GlaN, MurA and TAS) (Table 4). Among the models, the ones derived from soil NIR spectra yielded better calibrations with greater  $R^2$  and RPD values ( $R^2 = 0.63-0.84$ ,  $RPD = 1.64-2.52$ ) for amino sugar species than the models derived from soil MIR spectra ( $R^2 = 0.50-0.87$ ,  $RPD = 1.41-2.80$ ) and from hot-water extract MIR spectra ( $R^2 = 0.32-0.87$ ,  $RPD = 1.22-2.80$ ). The multi-site models based on samples from all 5 sites had similar performance ( $R^2 = 0.40-0.87$ ,  $RPD = 1.30-2.80$ ) as the models based on the soils using only the Western Canada sites ( $R^2 = 0.32-0.86$ ,  $RPD = 1.22-2.72$ ). The model calibrated for the soil samples collected from Eastern Canada showed wide variations in  $R^2 = 0.46-0.94$  and  $RPD = 1.36-4.22$ .

Multi-site results showed that the model estimates of amino sugar species and TAS were similar for NIR- and MIR-PLSr models based upon soil spectra, with averaged  $R^2 =$



0.60 for both models, which were better than the estimations by the MIR-PLSr models derived from hot-water extract spectra ( $R^2 = 0.45$ ) (Fig. 4). The performance of the multi-site models and the models for the soils in BC and AB, with averaged  $R^2 = 0.55$  and  $0.63$ , respectively, were better than the models for the ON and QC soils, with averaged  $R^2 = 0.47$ . All models were unable to quantitatively estimate the amounts of GluN, GalN and TAS ( $R^2 < 0.7$ ,  $RPD < 1.90$ ) in soils, nevertheless, the predictions for MurA were approximate quantitative at the multi-site scale ( $R^2 = 0.79-0.81$ ,  $RPD = 2.21-2.32$ ) and for the Western region soils ( $R^2 = 0.74-0.82$ ,  $RPD = 1.99-2.33$ ).

## Discussion

To our knowledge, MIR spectra from soil water extracts (hot and cold) have not been used to estimate water-dissolved soil OC, IC and TC. Our models were found to work better for estimating HWDCs than for CWDCs. This might be due to: (i) the quantity of water-dissolved organic matter in soil are affected by the extraction temperature (Bu et al., 2011; Ghani et al., 2003; Gregorich et al., 2003; Landgraf et al., 2006) and (ii) the compositions and properties of HWDC are distinct from the CWDC, with more carbohydrates, phenols, and lignin monomers in hot-water extracts relative to cold-water extracts (Landgraf et al., 2006). All these factors make the IR spectroscopy more sensitive to the HWDCs than to the CWDCs (Fig. 1), which also suggests that HWDC would be a better indicator than the CWDC when using the IR spectroscopy technique to monitor changes in SOM. For the estimation of soil HWDCs, MIR and NIR spectra from bulk soils were tested and the models developed using bulk soil MIR spectra performed better than the models using the bulk soil NIR spectra. Models used to measure HWDC developed using hot-water extraction generally outperformed

the models using the bulk soils. However, there is an extra step (i.e. additional lab work) involved in extracting soils using water at 100 °C.

Infrared-PLSr models developed hot- and cold-water extraction spectra (MIR) and soil spectra (MIR and NIR) could estimate water-dissolved OC and TC at multi-site and regional scales very well. However, water-dissolved IC was not well predicted in the soils from ON and QC provinces. These results may reflect the greater ranges of water-dissolved IC concentrations in soils from the Western Canada compared to the soils from the Eastern Canada (Table 1). In a study with 141 calibration samples collected from a single site, Vohland and Emmerling (2011) reported a poor prediction for hot water-dissolved OC (extracted at 100°C) using NIR-PLSr model with IR spectra collected from soil samples ( $R^2 = 0.71$  and RPD = 1.84), and the hot water-dissolved OC concentrations (42 - 1666 mg kg<sup>-1</sup>) were much smaller than those in our study (183-5840 mg kg<sup>-1</sup>, Table 1). In addition, Vasques et al. (2009) also reported a poor prediction of hot water-dissolved OC (extracted at 70°C) using NIR-PLSr model with 141 soil samples collected from one river watershed ( $R^2 = 0.69$ , RPD = 1.68). The hot water-dissolved OC concentrations in Vasques et al. (2009)'s study were ranging from 221 to 8990 mg kg<sup>-1</sup>, but the Log<sub>10</sub> units of hot water-dissolved OC concentrations (2.34-3.59 log(mg kg<sup>-1</sup>)) were utilized to develop the model. Soil samples employed in our study were collected from five different agricultural sites across Canada accompanying with distinct IR spectra and extensive ranges in water-dissolved C concentrations, which were different from the previous studies with a smaller number of soil samples collected from a single site that had a more limited range in soil chemical and physical properties (Vasques et al., 2009; Vohland and Emmerling, 2011). Site-specific models, such as those used for the Woodslee site in our

study, have been calibrated with both MIR- and NIR-PLSr for estimating water-dissolved OC and the results for a single site were poor with  $R^2 = 0.44-0.57$  and  $RPD = 1.34-1.53$  (unpublished data). This clearly implies the need for a wide range of properties in soil samples used in the calibration sets of interest in order to assure the accuracy and robustness of a model (Zornoza et al., 2008).

The models developed with soil spectra (both MIR and NIR) and hot-water spectra all failed to semi-quantitatively estimate the concentrations of GluN, GalN, MurA and TAS in soil ( $R^2 = 0.20 - 0.82$ ,  $RPD = 1.12 - 2.33$ ). The results thus suggest it is not feasible to use this approach to estimate amino sugar species directly using the IR-PLSr models developed with spectra collected from either bulk soils or from soil water extracts. Hence our hypotheses that amino sugars could be estimated using the IR spectra collected from bulk soils and soil water extracts was rejected. This failure might be due to that fact that, in contrast to the water soluble characteristic of water-dissolved C in soil, the free amino sugars in soil solution turn over rapidly (Roberts et al., 2007) and this would impacts on model performance because the amounts of water extracted free amino sugars in the moment of IR collection more likely differ from the amounts of amino sugars in solution when the amino sugars were extracted. The contents of soil amino sugar are relatively low and they exist in soil as a stabilized organic fraction attached to the soil matrix (Glaser et al., 2004) or present in a polymeric form (Roberts et al., 2007). The low concentrations of stabilized and polymeric amino sugar in bulk soil and hot-water extracts may have contributed to their low detectability using IR spectroscopy, and thus resulted in poor IR-PLSr model calibrations. The performance of IR-PLSr models in terms of estimating soil amino sugar contents were not adequate in the current study relative to a

previous study in which the MIR-PLSr models were developed using the MIR spectra collected from soil hydrolysates (Zhang et al., 2013). This difference is likely due to the fact that almost all stabilized amino sugar in bulk soil can be captured and condensed in soil hydrolysates after acid hydrolysis with 6 M HCl (Zhang and Amelung, 1996). Therefore, the resulting MIR-PLSr models developed using the soil hydrolysates spectra yielded good estimations of soil amino sugar concentrations. However, the opposite result occurred with poor estimation for MurA ( $R^2 = 0.61$ , RPD = 1.61) in Zhang et al.'s study (2013). It is interesting to note that the IR models calibrated for MurA in the current study were generally better than the models calibrated for GluN, GalN and TAS, which may reflect the greater diversity (higher CVs) in the values of MurA concentration compared to the smaller diversity (lower CVs) in the values of other amino sugar concentrations. A good estimation ( $R^2 = 0.90$ ) for total amino sugar-N in a set of diverse Ohio soils was obtained using NIR-PLSr with wavelength region of 1100-2498 nm (Dick et al., 2013). The amino sugar-N concentrations in Dick's study were estimated using the Illinois Nitrogen Soil Test by incubating samples at 55 °C for 5 h in 2 M NaOH (Khan et al., 2001) which was very different from the method used in the present study (Zhang and Amelung, 1996). The amino sugar-N concentrations and range in Dick's study (9.04-614 mg kg<sup>-1</sup>) were much smaller than those in our study (1214-6600 mg kg<sup>-1</sup>). There are other aspects that could affect model fitting and performance, such as the selection of wavelength ranges and the mathematic methods for spectra data pre-processing. These could all have affected the accuracy and reliability of the models developed for estimating some specific soil attributes (e.g. soil amino sugar). Discrepancies among studies remind us that more studies are required

to explore the feasibility of MIR- and NIR-PLSr technique in estimations of soil properties, especially the soil properties which have low concentrations.

### **Conclusions**

The IR-PLSr models developed using bulk soil MIR and NIR spectra could estimate hot water-dissolved C very well, and the MIR-PLSr models outperformed the NIR-PLSr models. The estimates of soil HWDC contents performed well at both the multi-scale and Western region scales. Additional research is required to improve the model performance for soils from the Eastern region. In addition, we speculate that low IR absorbance associated with low IC and amino sugar concentration in soil extract and bulk soil could be main reason for poor prediction and this may limit the infrared method could only be semi-quantitative on these soil properties.

### **Acknowledgement**

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

### **Reference**

- Amelung, W., 2001. Methods using amino sugars as markers for microbial residues in soil. In: R. Lal et al., editors, *Assessment methods for soil carbon*. Lewis Publishers, Boca Raton, FL. p. 233-272.
- Amelung, W., Brodowski, S., Sandhage-Hofmann, A., Bol, R., 2008. Combining biomarker with stable isotope analyses for assessing the transformation and turnover of soil organic matter. *Adv. Agron.* 100, 155-250.
- Amelung, W., Zhang, X., Zech, W., 1999. Amino sugars in native grassland soils along a climosequence in North America. *Soil Sci. Soc. Am. J.* 63, 86-92.

- Appuhn, A., Joergensen, R.G., Raubuch, M., Scheller, E., Wilke, B., 2004. The automated determination of glucosamine, galactosamine, muramic acid, and mannosamine in soil and root hydrolysates by HPLC. *J. Plant Nutr. Soil Sci.* 167, 17-21.
- Balkwill, D.L., Leach, F.R., Wilson, J.T., McNabb, J.F., White, D.C., 1988. Equivalence of microbial biomass measures based on membrane lipid and cell wall components, adenosine triphosphate, and direct counts in subsurface aquifer sediments. *Micro. Ecol.* 16, 73-84.
- Bending, G.D., Putland, C., Ranys, F., 2000. Changes in microbial community metabolism and labile organic matter fractions as early indicators of the impact of management on soil biological quality. *Biol. Fertil. Soils* 31, 78-84.
- Brown, D.J., Shepherd, K.D., Walsh, M.G., Dewayne Mays, M., Reinsch, T.G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* 132, 273-290.
- Brunet, D., Barthès, B.G., Chotte, J.-L., and Feller, C. 2007. Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity. *Geoderma* 139:106-117.
- Bu, X., Ding, J., Wang, L., Yu, X., Huang, W., Ruan, H., 2011. Biodegradation and chemical characteristics of hot-water extractable organic matter from soils under four different vegetation types in the Wuyi Mountains, southeastern China. *Eur. J. Soil Biol.* 47, 102-107.
- Chantigny, M.H., 2003. Dissolved and water-extractable organic matter in soils: a review on the influence of land use and management practices. *Geoderma* 113, 357-380.
- Cozzolino, D., Chandra, S., Roberts, J., Power, A., Rajapaksha, P., Ball, N., Gordon, R., and Chapman, J. 2018. There is gold in them hills: Predicting potential acid mine drainage events through the use of chemometrics. *Science of the Total Environment* 619–620: 1464–1472.
- Dick, W.A., Thavamani, B., Conley, S., Blaisdell, R., Sengupta, A., 2013. Prediction of  $\beta$ -glucosidase and  $\beta$ -glucosaminidase activities, soil organic C, and amino sugar N in a diverse population of soils using near infrared reflectance spectroscopy. *Soil Bio. Biochem.* 56, 99-104.
- Du, C., Zhou, J., 2009. Evaluation of soil fertility using infrared spectroscopy: a review. *Environ. Chem. Lett.* 7, 97-113.
- Evans, J.L., Fernandez, I.J., Rustad, L.E., Norton, S.A., 2001. Methods for evaluating carbon fractions in forest soils: A review. *MAFES Technical Bulletin* 178, 1-42.
- Fan, R.Q., Yang, X.M., Xie, H.T., Reeb, M.A. 2012. Determination of nutrients in hydroponic solutions using mid-infrared spectroscopy. *Sci. Hortic.* 144, 48-54.
- Ghani, A., Dexter, M., Perrott, K.W., 2003. Hot-water extractable carbon in soils: a sensitive measurement for determining impacts of fertilisation, grazing and cultivation. *Soil Bio. Biochem.* 35, 1231-1243.
- Glaser, B., Turrión, M.a.-B., Alef, K., 2004. Amino sugars and muramic acid–biomarkers for soil microbial community structure analysis. *Soil Bio. Biochem.* 36, 399-407.
- Gregorich, E.G., Beare, M.H., Stoklas, U., St-Georges, P., 2003. Biodegradability of soluble organic matter in maize-cropped soils. *Geoderma* 113, 237-252.

- Gregorich, E.G., M.R.Carter, Angers, D.A., Monreall, C.M., Ellerta, B.H., 1994. Towards a minimum data set to assess soil organic matter quality in agricultural soils. *Can. J. Soil Sci.* 74, 367-384.
- Guggenberger, G., Frey, S.D., Six, J., Paustian, K., Elliott, E.T., 1999. Bacterial and Fungal Cell-Wall Residues in Conventional and No-Tillage Agroecosystems. *Soil Sci. Soc. Am. J.* 63, 1188-1198.
- Hamkalo, Z., Bedernichek, T., 2014. Total, cold and hot water extractable organic carbon in soil profile: impact of land-use change. *Zemdirbyste-Agriculture*, 101:125-132.
- Haynes, R.J., 2005. Labile organic matter fractions as central components of the quality of agricultural soils: An overview. *Adv. Agron.* 85, 221-268.
- Haynes, R.J., 2000. Labile organic matter as an indicator of organic matter quality in arable and pastoral soils in New Zealand. *Soil Bio. Biochem.* 32, 211-219.
- Hemmateenejad, B., Safarpour, M.A., Mehranpour, M. 2005. Net analyte signal-artificial neural network (NAS-ANN) model for efficient nonlinear multivariate calibration. *Anal. Chim. Acta* 535:275-285.
- Janik, L.J., Forrestera, S.T., Rawsonb, 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.
- Joergensen, R., Wichern, F., 2008. Quantitative assessment of the fungal contribution to microbial tissue in soil. *Soil Bio. Biochem.* 40, 2977-2991.
- Joergensen, R.G., Mäder, P., Fließbach, A., 2010. Long-term effects of organic farming on fungal and bacterial residues in relation to microbial energy metabolism. *Biol. Fertil. Soils* 46, 303-307.
- Kaiser, K., Benner, R., 2000. Determination of amino sugars in environmental samples with high salt content by high-performance anion-exchange chromatography and pulsed amperometric detection. *Anal. Chem.* 72, 2566-2572.
- Khan, S. A., R. L. Mulvaney, and R. G. Hoeft. 2001. A simple soil test for detecting sites that are nonresponsive to nitrogen fertilization. *Soil Sci. Soc. Am. J.* 65:1751-1760.
- Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. *Science* 304, 1623-1627.
- Landgraf, D., Leinweber, P., Makeschin, F., 2006. Cold and hot water-extractable organic matter as indicators of litter decomposition in forest soils. *J. Plant Nutr. Soil Sci.* 169, 76-82.
- Liang, B.C., MacKenzie, A.F., Schnitzer, M., Monreal, C.M., Voroney, P.R., Beyaert, R.P., 1998. Management-induced change in labile soil organic matter under continuous corn in eastern Canadian soils. *Biol. Fertil. Soils* 26, 88-94.
- Ludwig, B., Sawallisch, A., Heinze, S., Joergensen, R.G., Vohland, M., 2015. Usefulness of middle infrared spectroscopy for an estimation of chemical and biological soil properties – Underlying principles and comparison of different software packages. *Soil Bio. Biochem.* 86, 116-125.

- Malley, D.F., Martin, P.D., Ben-Dor, E., 2004. Application in analysis of soils. In: Roberts, C.A., Workman Jr., J., Reeves III, J.B. (Eds.), *Near-Infrared Spectroscopy in Agriculture*. American Society of Agronomy, Madison, WI, pp. 729-784. Chapter 26.
- Miliar, W.N., Casida, J., L.E., 1970. Evidence for muramic acid in soil. *Can. J. Soil Sci.* 16, 299-304
- Mulvaney, R. L., S. A. Khan, R. G. Hoefl, and H. M. Brown. 2001. A Soil Organic Nitrogen Fraction that Reduces the Need for Nitrogen Fertilization. *Soil Sci. Soc. Am. J.* 65, 1164-1172.
- 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.
- Cozzolino, D., Chandra, S., Roberts, J., Power, A., Rajapaksha, P., Ball, N., Gordon, R., and Chapman, J. 2018. There is gold in them hills: Predicting potential acid mine drainage events through the use of chemometrics. *Science of the Total Environment* 619–620: 1464–1472.
- Parsons, J.W., 1981. Chemistry and distribution of amino sugars in soils and soil organisms. In: E.A. Paul and J.N. Ladd, editors, *Soil Biochemistry*. Vol. 5. Marcel Dekker, New York. p. 197-227.
- Rees R. M., Parker J. P., 2005. Filtration increases the correlation between water extractable organic carbon and soil microbial activity. *Soil Bio. Biochem.*, 37 (12): 2240-2248.
- Reeves, D.W., 1997. The role of soil organic matter in maintaining soil quality in continuous cropping systems. *Soil Tillage Res.* 43, 131-167.
- 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.
- 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.
- Roberts, P., Bol, R., Jones, D.L., 2007. Free amino sugar reactions in soil in relation to soil carbon and nitrogen cycling. *Soil Bio. Biochem.* 39, 3081-3092.
- Rossel, R.A.V., Webster, R., 2012. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database. *Eur. J. Soil Sci.* 63, 848-860.
- Saeys, 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.
- 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.
- Shukla, M.K., Lal, R., Ebinger, M., 2006. Determining soil quality indicators by factor analysis. *Soil Tillage Res.* 87, 194-204.
- 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.
- 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.
- Stevenson, F.J., 1983. Isolation and identification of amino sugars in soil. *Soil Sci. Soc. Am. J.* 47, 61-65.
- Terhoeven-Urselmans, T., Schmidt, H., Georg Joergensen, R., Ludwig, B., 2008. Usefulness of near-infrared spectroscopy to determine biological and chemical soil properties: Importance of sample pre-treatment. *Soil Bio. Biochem.* 40, 1178-1188.
- Vasques, G.M., Grunwald, S., Sickman, J.O., 2009. Modeling of Soil Organic Carbon Fractions Using Visible-Near-Infrared Spectroscopy. *Soil Sci. Soc. Am. J.* 73, 176-184.
- 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.
- Vohland, M., Emmerling, C., 2011. Determination of total soil organic C and hot water-extractable C from VIS-NIR soil reflectance with partial least squares regression and spectral feature selection techniques. *Eur. J. Soil Sci.* 62, 598-606.
- Wenger, S.J., and Olden, J.D. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods Ecol. Evol.* 2012: 260-267.
- Zelles, L., 1988. The simultaneous determination of muramic acid and glucosamine in soil by high-performance liquid chromatography with precolumn fluorescence derivatization. *Biol. Fertil. Soils* 6, 125-130.
- Zhang, B., Yang, X., Drury, C.F., Zhang, X., 2013. A Mid-infrared spectroscopy method to determine the glucosamine, galactosamine, and muramic acid concentrations in soil hydrolysates. *Soil Sci. Soc. Am. J.* 77, 842-849.
- Zhang, X.D., Amelung, W., 1996. Gas chromatographic determination of muramic acid, glucosamine, mannosamine, and galactosamine in soils. *Soil Bio. Biochem.* 28, 1201-1206.
- 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. Transmission spectra (mid-infrared) collected from soil cold and hot water extracts and reflectance spectra (mid- and near-infrared) collected from bulk soil samples.

Fig. 2 Predicted values of organic carbon (OC, mg kg<sup>-1</sup> soil), inorganic carbon (IC, mg kg<sup>-1</sup> soil) and total carbon (TC, mg kg<sup>-1</sup> soil) in hot and cold water extracts (HWE and

CWE) plotted against measured values and the statistics of model validation. The PLSr models were fitted using the MIR spectra collected from hot and cold water extracts.

Fig. 3. Predicted values of hot water dissolved organic carbon (OC, mg kg<sup>-1</sup> soil), inorganic carbon (IC, mg kg<sup>-1</sup> soil) and total carbon (TC, mg kg<sup>-1</sup> soil) plotted against measured values and the statistics of model validation. The PLSr models were fitted using the MIR and NIR spectra collected from bulk soil samples (Soil-MIR Spectra and Soil-NIR Spectra).

Fig. 4. Predicted values of glucosamine (GluN, mg kg<sup>-1</sup> soil), galactosamine (GalN, mg kg<sup>-1</sup> soil), muramic acid (MurA, mg kg<sup>-1</sup> soil) and total amino sugar (TAS, mg kg<sup>-1</sup> soil) plotted against measured values and the statistics of model validation. The PLSr models were fitted using the MIR spectra collected from hot water extracts (HWE) and the MIR and NIR spectra collected from bulk soil samples (Soil-MIR Spectra and Soil-NIR Spectra).

For Review Only

Table 1. Descriptive statistics of concentrations of the hot and cold water dissolved organic carbon (OC), inorganic carbon (IC) and total carbon (TC) and the amino sugars in soil.

Scale	Mean	Range	SD	CV (%)	Mean	Range	SD	CV (%)
	Hot water dissolved OC (mg kg <sup>-1</sup> soil)				Cold water dissolved OC (mg kg <sup>-1</sup> soil)			
Western	1060	215-5840	1230	116	331	64.1-1520	326	98.4
Eastern	394	183-1460	153	38.9	138	78.2-587	51.4	37.2
Multi-site	724	183-5840	932	129	233	64.1-1520	250	107
	Hot water dissolved IC (mg kg <sup>-1</sup> soil)				Cold water dissolved IC (mg kg <sup>-1</sup> soil)			
Western	68.3	5.56-422	83.4	122	4.52	0.200-35.2	4.93	110
Eastern	16.9	5.48-48.1	7.98	47.1	2.27	0.210-24.4	2.33	103
Multi-site	42.2	5.48-422	64.1	152	3.37	0.200-35.2	3.98	119
	Hot water dissolved TC (mg kg <sup>-1</sup> soil)				Cold water dissolved TC (mg kg <sup>-1</sup> soil)			
Western	1133	228-6010	1310	115	336	66.1-1530	329	98.1
Eastern	410	194-1500	158	38.5	140	78.9-588	51.9	37.0
Multi-site	766	194-6010	991	129	236	66.1-1530	253	107
	Glucosamine (GluN, mg kg <sup>-1</sup> soil)				Galactosamine (GalN, mg kg <sup>-1</sup> soil)			
Western	1501	716-4030	594	37.4	818.1	188-2050	330	40.4
Eastern	1110	485-2100	315	28.4	623.2	202-1790	223	35.8
Multi-site	1350	485-4030	530	39.4	718.9	188-2050	297	41.3
	Muramic acid (MurA, mg kg <sup>-1</sup> soil)				Total amino sugars (TAS, mg kg <sup>-1</sup> soil)			
Western	345	44.5-1430	250	72.4	2754.0	1210-6600	990	36.0
Eastern	141	96.3-223	38.0	26.9	1772.6	700-3470	522	29.5
Multi-site	308	44.5-1430	240	77.9	2254.4	700-6600	927	41.1
	Soil organic carbon (g C kg <sup>-1</sup> soil)				Total nitrogen (g N kg <sup>-1</sup> soil)			
Western	53.7	20.8-1160	28.8	53.6	4.28	1.71-12.1	2.19	51.1
Eastern	30.1	10.3-27.8	10.5	34.9	2.26	0.950-4.40	0.63	27.8
Multi-site	41.1	10.3-1160	24.4	58.8	3.23	0.950-12.1	1.88	58.1

**Note:** Min: Minimum; Med: Median; Max: Maximum; SD: Standard Deviation; CV: Coefficient of Variation, the ratio of standard deviation to mean. Western, Eastern and Cross stand for the western region, eastern region and cross region of Canada.

Table 2. Calibration models developed for prediction of water dissolved organic carbon (OC), inorganic carbon (IC) and total carbon (TC) using mid-infrared (MIR) spectra collected from hot water extracts and cold water extracts.

	Pre-processing	Greatest influence IR wavenumber range (cm <sup>-1</sup> )	Rank	R <sup>2</sup>	RMSEC (mg kg <sup>-1</sup> )	RPD
<b>Hot water extracts spectra:</b>		<b>Hot Water Dissolved OC</b>				
Western	FD	3282-2916,1840-1477,759-400	9	0.98	160	7.95
Eastern	FD	3641-3278,2920-2198,1481-400	10	0.89	52.8	3.01
Multi-site	NO	3641-2916,1840-1477,759-400	9	0.98	132	7.18
		<b>Hot Water Dissolved IC</b>				
Western	COE	1840-1477,759-400	3	0.94	20.5	4.12
Eastern	FD	4000-3278,1840-400	9	0.81	3.61	2.29
Multi-site	SLS	2561-1477	5	0.96	13.5	4.78
		<b>Hot Water Dissolved TC</b>				
Western	FD	3282-2916,1840-1477,759-400	9	0.98	170	7.96
Eastern	FD	3641-3278,2920-2198,1481-400	10	0.89	53.9	3.04
Multi-site	NO	3641-2916,1840-1477,759-400	9	0.98	139	7.26
<b>Cold water extracts spectra:</b>		<b>Cold Water Dissolved OC</b>				
Western	FD+VN	2920-2557,1118-400	4	0.83	125	2.42
Eastern	FD+SLS	4000-3637,2561-1836	6	0.76	25.6	2.05
Multi-site	FD+VN	3641-3278,2920-2198,1118-400	5	0.83	95.6	2.40
		<b>Cold Water Dissolved IC</b>				
Western	VN	759-399	3	0.85	1.54	2.60
Eastern	FD+SLS	2198-1836	10	0.67	0.58	1.73
Multi-site	FD+VN	2920-1836,1118-400	6	0.82	1.38	2.35
		<b>Cold Water Dissolved TC</b>				
Western	FD+VN	2920-2557,1118.6-400	4	0.83	125	2.43
Eastern	FD+SLS	4000-3637,2561.3-1836	6	0.76	26.0	2.04
Multi-site	FD+VN	3282-1836,1118.6-400	5	0.83	95.3	2.44

**Note:** R<sup>2</sup> = 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 prediction of water dissolved organic carbon (OC), inorganic carbon (IC) and total carbon (TC) using mid-infrared (MIR) and near-infrared (NIR) spectra collected from the bulk soil samples.

Pre-processing		Greatest influence IR wavenumber range (cm <sup>-1</sup> )	Rank	R <sup>2</sup>	RMSEC (mg kg <sup>-1</sup> )	RPD
<b>Soil-MIR spectra:</b>		<b>Hot Water Dissolved OC</b>				
Western	FD	4003-3641,3282-2919,1481-1118	6	0.91	354	3.26
Eastern	MSC	3644-3282,2561-1839,1481-400	10	0.92	45.7	3.48
Multi-site	SLS	3282-2198,1839-1118	10	0.92	251	3.45
		<b>Hot Water Dissolved IC</b>				
Western	MSC	3644-2198,1839-1477	6	0.92	22.9	3.63
Eastern	SD	4003-2919,1839-1477	5	0.77	3.71	2.07
Multi-site	MSC	4003-3641,3282-2919,1839-1477	10	0.92	18.0	3.51
		<b>Hot Water Dissolved TC</b>				
Western	SLS	3644-2919,2561-2198,1839-1477	10	0.93	326	3.84
Eastern	MSC	3644-3282,2561-1839,1481-400	10	0.92	47.4	3.46
Multi-site	SLS	3282-2198,1839-1118	10	0.92	264	3.51
<b>Soil-NIR spectra:</b>		<b>Hot Water Dissolved OC</b>				
Western	FD+VN	6100-5450,4600-4249	10	0.94	298	3.93
Eastern	FD	5450-4249	10	0.93	42.4	3.75
Multi-site	FD+SLS	7998-7500,450-4599	10	0.94	219	3.96
		<b>Hot Water Dissolved IC</b>				
Western	FD	5450-4249	10	0.96	17.1	4.93
Eastern	FD+SLS	4600-4249	10	0.72	4.13	1.90
Multi-site	FD	6100-4599	10	0.94	15.1	4.18
		<b>Hot Water Dissolved TC</b>				
Western	FD+VN	4600-4249	4	0.91	362	3.37
Eastern	FD	5450-4249	10	0.93	44.0	3.73
Multi-site	FD+SLS	7998-7500,6100-5175	10	0.94	226	4.09

**Note:** R<sup>2</sup> = 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 4. Calibration models developed for predicting amino sugar using mid-infrared (MIR) spectra collected from the hot water extracts and the mid-infrared (MIR) and near-infrared (NIR) spectra collected from the bulk soil samples.

	Pre-processing	Greatest influence IR wavenumber range (cm <sup>-1</sup> )	Rank	R <sup>2</sup>	RMSEC (mg g <sup>-1</sup> )	RPD
<b>Hot water extracts spectra:</b>		<b>Glucosamine (GluN)</b>				
Western	FD+SLS	2920-2557	2	0.54	380	1.48
Eastern	SD	3641-1477,1118-756	10	0.68	184	1.78
Multi-site	SD	2920-2198	10	0.64	308	1.66
		<b>Galactosamine (GalN)</b>				
Western	FD+MSC	4000-3278,2920-2557,1481-756	4	0.32	262	1.22
Eastern	NO	3282-2557,2198-1477	6	0.46	151	1.36
Multi-site	VN	1840-397	10	0.40	220	1.30
		<b>Muramic acid (MurA)</b>				
Western	FD	3641-3278,2920-2557,1481-397	10	0.86	86.5	2.72
Eastern	MSC	3641-3278,2561-1114	10	0.79	21.7	2.16
Multi-site	FD	3641-1836,1481-756	10	0.87	71.5	2.80
		<b>Total amino sugars (TAS)</b>				
Western	SD	2920-2557	10	0.57	635	1.53
Eastern	SD	3641-3278,1840-756	5	0.56	353	1.51
Multi-site	FD+VN	4000-3637,3282-2557,2198-1836	10	0.67	524	1.73
<b>Soil-MIR spectra:</b>		<b>Glucosamine (GluN)</b>				
Western	SD	2923-2561,2202-1839	9	0.73	318	1.93
Eastern	FD+SLS	4003-2198,1839-1118	10	0.81	141	2.32
Multi-site	FD+SLS	3282-2198,1839-1477	10	0.71	292	1.85
		<b>Galactosamine (GalN)</b>				
Western	VN	3644-3282,2923-2561,2202-1839	10	0.56	228	1.51
Eastern	NO	2561-1839,1481-1118	5	0.50	160	1.42
Multi-site	VN	4003-3282,2923-2561,2202-1839	2	0.50	211	1.41
		<b>Muramic acid (MurA)</b>				
Western	MMN	4003-3641,3282-2919	3	0.84	81.4	2.45
Eastern	COE	4003-3641,2923-2561	4	0.51	28.7	1.42
Multi-site	FD+VN	3644-3282,1839-1118	6	0.87	71.2	2.73
		<b>Total amino sugars (TAS)</b>				
Western	MMN	4003-3641,3282-2561	3	0.63	608	1.65
Eastern	COE	3282-2561,2202-756	6	0.69	299	1.79
Multi-site	SLS	4003-3282,2923-2198,1839-1477.2	9	0.73	490	1.92
<b>Soil-NIR spectra:</b>		<b>Glucosamine (GluN)</b>				
Western	FD	7998-7500,4600-4249	10	0.83	254	2.43
Eastern	NO	7749-7500,5450-4599	10	0.76	159	2.06
Multi-site	FD+VN	7500-6099,4600-4249	10	0.77	248	2.08
		<b>Galactosamine (GalN)</b>				

Western	MSC	7500-6099	10	0.63	209	1.64
Eastern	FD+SLS	7998-6099	10	0.94	54.9	4.22
Multi-site	FD+MSC	7049-6099,5450-4599	10	0.68	172	1.76
<b>Muramic acid (MurA)</b>						
Western	FD+VN	7998-7500,4600-4249	10	0.90	64.2	3.20
Eastern	FD+VN	7500-6799,4600-4249	4	0.84	16.3	2.52
Multi-site	FD+VN	7749-7500,4600-4249	2	0.83	79.6	2.42
<b>Total amino sugars (TAS)</b>						
Western	VN	6800-6099,5450-4249	10	0.71	559	1.84
Eastern	FD	7500-6099,5450-4599	10	0.90	176	3.09
Multi-site	FD+MSC	7998-6099,5450-4249	10	0.77	421	2.07

**Note:**  $R^2$  = 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.

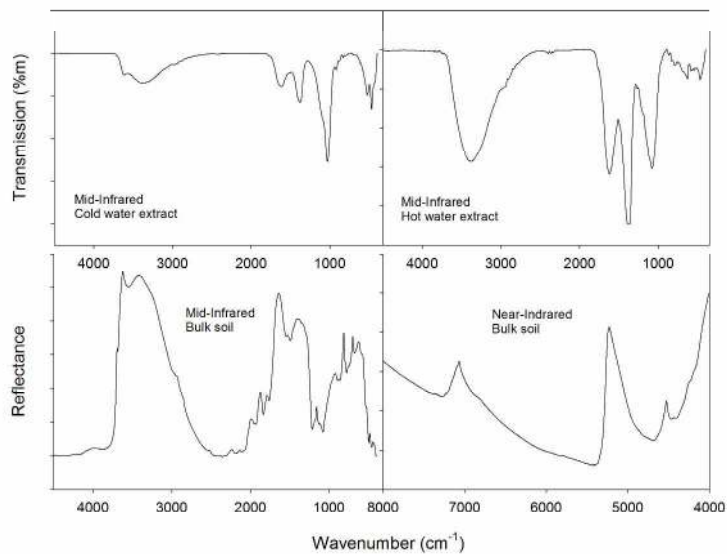


Fig. 1.

209x296mm (300 x 300 DPI)



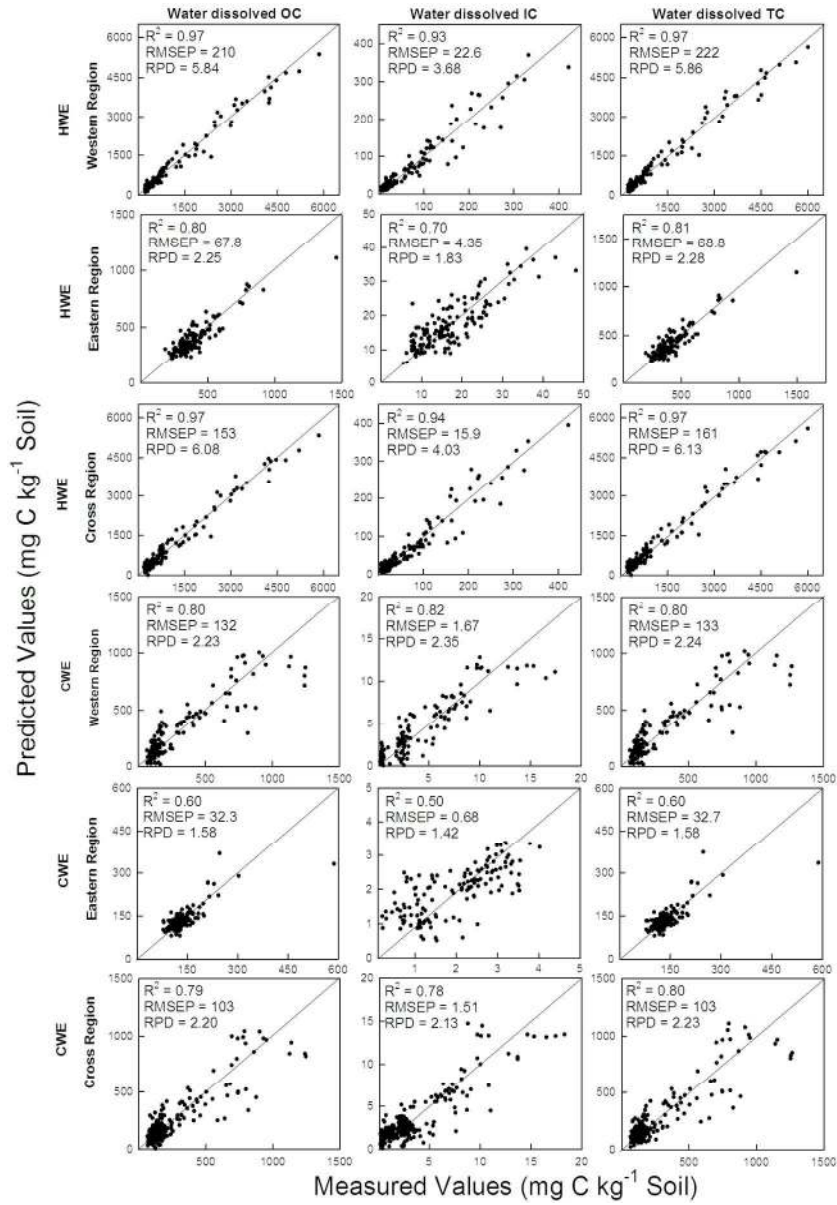


Fig. 2.

141x199mm (300 x 300 DPI)

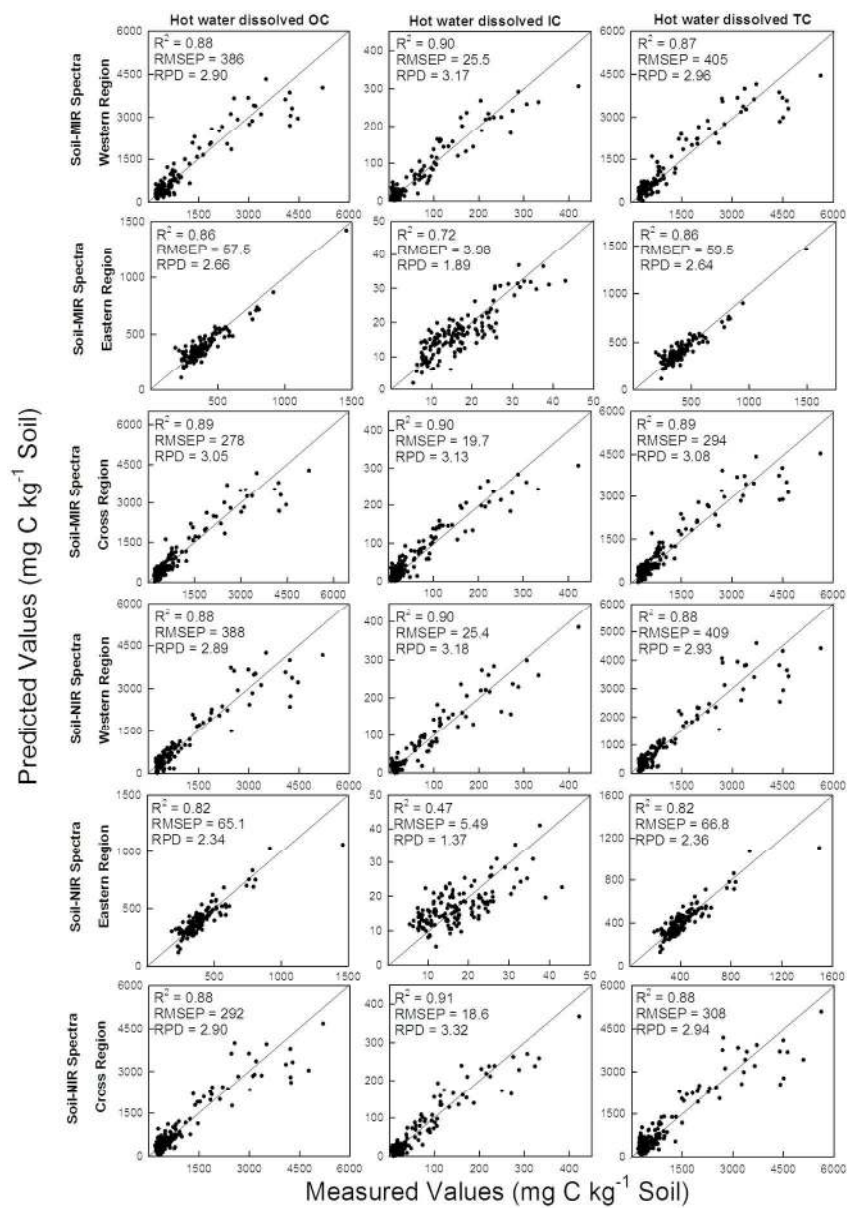


Fig. 3.

141x199mm (300 x 300 DPI)

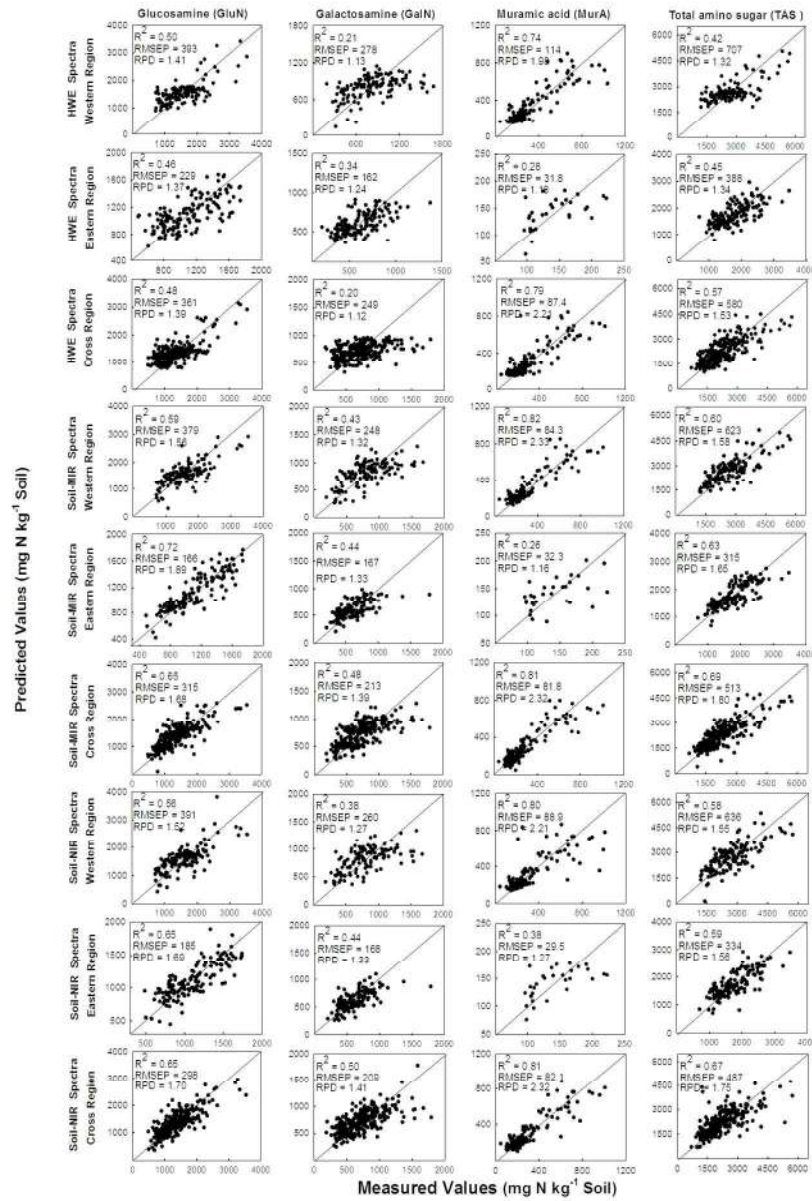


Fig. 4.

141x199mm (300 x 300 DPI)