



# Mid-infrared spectra predict nuclear magnetic resonance spectra of soil carbon



Mohsen Forouzanoghar<sup>a,\*</sup>, Jeffrey A. Baldock<sup>b</sup>, Ronald J. Smernik<sup>c</sup>, Bruce Hawke<sup>b</sup>, Lauren T. Bennett<sup>a</sup>

<sup>a</sup> School of Ecosystem and Forest Sciences, The University of Melbourne, 4 Water Street, Creswick, Victoria 3363, Australia

<sup>b</sup> Commonwealth Scientific and Industrial Research Organisation (CSIRO) Agriculture Flagship, Glen Osmond, South Australia 5064, Australia

<sup>c</sup> School of Agriculture, Food and Wine, and Waite Research Institute, The University of Adelaide, Urrbrae, South Australia 5064, Australia

## ARTICLE INFO

### Article history:

Received 18 November 2014

Received in revised form 16 February 2015

Accepted 18 February 2015

Available online 28 February 2015

### Keywords:

Soil organic carbon

Mid-infrared spectroscopy

Solid-state C-13 NMR spectroscopy

Partial least-squares

## ABSTRACT

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for characterising the complex chemistry of soil organic carbon (SOC), but is prohibitively expensive, time-consuming and technically-demanding. Diffuse reflectance mid-infrared (MIR) spectroscopy is an attractive alternative because it is a high-throughput, cost-effective and easy-to-use technique that provides information on the amount and nature of soil mineral and organic components. However, interpretation of complex MIR spectra can be challenging due to difficulties with distinguishing SOC peaks from overlapping mineral-related peaks. We present a novel approach to predict the entire NMR spectra of SOC from corresponding MIR spectra using partial least-squares regression (PLSR) in an R environment. We developed a multi-response MIR-PLSR prediction model by regressing corresponding NMR and MIR spectra of 99 HF-treated <50 μm fractions of soils using the pls package. The model was validated using (set-aside) test sets in four model iterations. The model provided accurate predictions of the entire average NMR spectra. Average Euclidean distance values between spectra in the training set were at least 3.5 fold greater than those between average reference and predicted NMR spectra, indicating that prediction errors were small relative to between-soil variation. Our approach accurately predicted intricate NMR spectra, demonstrating new potential for routine analysis of complex SOC chemistry.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

The large quantity of organic carbon stored in soil (Trumbore et al., 1996) and the strong coupling of SOC to atmospheric CO<sub>2</sub> mean that SOC is extremely important to support life on earth. Conversion of natural ecosystems to cropping lands has resulted in up to 75% reductions in SOC stocks (Lal, 2004). This has been estimated as a loss of 78 ± 12 Pg of carbon from soils to the atmosphere at a global scale (Lal et al., 2004). Extensive release of carbon from soils has adversely affected soil and water quality and sustainable food production and has been linked to worsening air quality and global warming (Lal, 2004). In routine soil tests, SOC is usually reported only in terms of its total gravimetric content (g C/g soil) or volumetric stock (Mg C/ha to a specific depth). However, since SOC is a complex mixture of diverse organic molecular structures (Baldock and Broos, 2012; Kogel-Knabner, 2002), quantification of total SOC alone can only ever provide a simplified view of the complex processes controlling its accumulation or loss.

It is acknowledged that the chemical composition of SOC affects a range of soil physical, chemical and biological properties, and especially its susceptibility to microbial mineralisation, and thereby affects the

properties of an entire ecosystem (Baldock and Skjemstad, 1999). Moreover, the variability of organic carbon chemistry between soils has been linked to their potential for carbon sequestration (Krull et al., 2006; Singh et al., 2009), and to the sorption of dissolved organic carbon (Kaiser et al., 2000) and pesticides (Ahmad et al., 2001; Kile et al., 1999) onto soil particles. Therefore, understanding the nature and changes of SOC chemical composition provides us with valuable insights into the dynamics of soil carbon, and into the health and productivity of soils.

Currently, <sup>13</sup>C NMR spectroscopy is considered to be the most reliable technique to determine the chemistry of SOC (Simpson et al., 2011). However, the instrumentation is expensive to purchase and to run, and given that the concentration of organic carbon in mineral soils is usually <50 g C/kg soil, single samples typically require 5 h or more of instrument time. As such, the use of NMR spectroscopy to characterise SOC chemistry is far from routine, and currently beyond the reach of many studies. Diffuse reflectance mid-infrared (MIR) spectroscopy offers an attractive alternative because it is a high-throughput, cost-effective and easy-to-use technique that can quantify the nature of all chemical bonds present in a sample of soil. The downside of MIR analysis of mineral soils is that, in contrast to NMR analysis, chromophores for different organic structures are complex, overlap with each other, and with signals derived from mineral soil components and may represent only a small component of the total MIR signal. Thus,

\* Corresponding author.

E-mail address: [mohsenf@unimelb.edu.au](mailto:mohsenf@unimelb.edu.au) (M. Forouzanoghar).

the challenge with MIR analysis is one of spectral interpretation, in particular to selectively identify and quantify signals derived from organic carbon against a background dominated by signals derived from soil mineral components.

Through the developments of multivariate regression methods (Haaland and Thomas, 1988), infrared (IR) spectroscopy has been used to quantitatively predict a variety of soil chemical parameters (Janik et al., 1998) including SOC concentration (Forouzanoghar et al., 2009; Zimmermann et al., 2007), its biochemical components (Cecillon et al., 2012), and NMR integrals, i.e. the area under an NMR peak (Leifeld, 2006; Terhoeven-Urselmans et al., 2006). In this context, multivariate calibrations are developed between IR spectra and the reference soil property of interest. The IR calibration methods utilize multivariate regressions to reduce and to harness the complexity of IR data and to provide predictions. In a previous study (Forouzanoghar et al., 2013), using two-dimensional heterospectral correlation analysis, we demonstrated how major NMR spectral peaks of SOC could be cross-correlated with the corresponding MIR peaks. Accordingly, we hypothesized that MIR spectra coupled with partial least-squares regression (PLSR) could be used to produce a prediction of NMR spectra of SOC. Here, we describe the development and validation of such an approach in an R environment (R Core Team, 2013) using the *pls* package (Mevik and Wehrens, 2007), and demonstrate its utility to reliably predict the entire NMR spectra, and to thus enhance capability for routine analysis of complex SOC chemistry.

## 2. Material and methods

### 2.1. Soils

The samples used in this study were fine fractions ( $\leq 50 \mu\text{m}$ ) isolated from 99 soils included in the Australian Soil Carbon Research Program (Baldock et al., 2013a). The  $\leq 50 \mu\text{m}$  fraction was isolated as described by Baldock et al. (2013b). Briefly, the soils were dispersed and wet sieved to  $\leq 50 \mu\text{m}$  with the soil material passing through the sieve being frozen, lyophilised and treated with 2% hydrofluoric acid (HF) according to the methodology of Skjemstad et al. (1994).

### 2.2. Solid-state $^{13}\text{C}$ nuclear magnetic resonance spectroscopy

Solid state  $^{13}\text{C}$  cross polarisation (CP) NMR spectra were acquired with magic angle spinning (MAS) on a Bruker 200 Avance spectrometer equipped with a 4.7 T wide-bore superconducting magnet operating at a  $^{13}\text{C}$  resonant frequency of 50.33 MHz. Weighed samples (100–600 mg) with known carbon contents were packed into 7 mm diameter zirconia rotors with Kel-F end caps and spun at 5 kHz. A standard cross polarisation pulse sequence using a 3.2  $\mu\text{s}$ , 195 W, 90° pulse, a contact time of 1 ms and a recycle delay of 1 s was applied to all samples. Between 10,000 and 30,000 scans were collected for each sample with more scans collected for samples with lower carbon content so as to limit variation in signal-to-noise ratios among samples. Chemical shift was calibrated to the methyl resonance of hexamethylbenzene at 17.36 ppm and a 50 Hz Lorentzian line broadening was applied to all spectra. Bruker TopSpin 3.2 software was used to derive absolute signal intensities for all samples and for an empty rotor. The signal obtained from the empty rotor was subtracted from all spectra.

### 2.3. Mid-infrared spectroscopy

Diffuse reflectance MIR spectra were acquired using a Nicolet 6700 FTIR spectrometer (Thermo Fisher Scientific Inc., MA, USA) equipped with a KBr beam-splitter, a DTGS detector and a Pike AutoDiff-Automated diffuse reflectance accessory (Pike Technologies, WI, USA). Approximately 100 mg of ground, air-dried sample was placed into 9 mm stainless steel auto sampler cups and the surfaces levelled. The samples were loaded onto a 60-sample automated wheel with a silicon

carbide background disk located in the centre. Spectra were acquired over 8000–400  $\text{cm}^{-1}$  with a resolution of 8  $\text{cm}^{-1}$ . In order to quantify and correct the background signal intensity, 240 scans were applied on a silicon carbide disk prior to analysing each set of 60 soil samples. A total of 60 scans was acquired and averaged to produce a reflectance spectrum for each individual soil sample. Omnic software (Version 8.0) was used to convert the acquired reflectance spectra into absorbance spectra (log transform of the inverse of reflectance).

### 2.4. Spectral data pretreatment

The noise in NMR and MIR spectra of HF-treated soils was reduced using principal component analysis (PCA) as proposed and described in detail by Jung (2003). In brief, PCA was performed separately on the NMR and MIR datasets to identify the spectral principal components (PCs) within each set of data. Then the spectra were reconstructed using the PCs. This noise reduction method has been shown to be effectively and rationally reduce the noise in spectral data (Jung, 2003). GRAMS/AI software (Thermo Galactic, Thermo Fisher Scientific, Waltham, MA) was used for PCA. Baseline correction was implemented for both NMR and MIR spectra.

### 2.5. Creating datasets for R computations

NMR and MIR spectral data were saved in separate comma delimited (\*.csv) datasheets to be used in R calculations with matrix or data frame structures. In the datasheets, each row corresponded to a single NMR or MIR spectrum, and each column corresponded to an NMR or MIR spectral variable. The first row was a header holding the NMR or MIR spectral variable labels. The first column contained the row (i.e. sample) names under id in the first cell.

### 2.6. Fitting a PLSR multi-response calibration model in R

The complete R code is described in detail in the Appendices, and here we provide a brief description. The matrices of corresponding NMR and MIR spectra were regressed, and a multi-response prediction model was fit using the *pls* function. Calibrations were developed using leave-one-out (LOO) cross-validation (Arlot and Celisse, 2010) to determine the optimal number of PLSR factors required for the predictive model. Specifically, the following function was used to develop the calibration model, here called *MIR-PLSR*,

```
> MIR-PLSR<-pls ( NMR ~ MIR, ncomp = 10, data = spectra.
train, validation = "LOO")
```

This function fitted a multi-response PLSR model with 10 components. The data argument introduces the dataset (here called *spectra.train*) to the function. The *spectra.train* dataset ( $n = 99$ ) had a data frame structure in R, consisting of  $n$  rows (where  $n$  is the number of spectra in the dataset) and of two variables, each being a matrix, of NMR and MIR spectra. The left-hand side of the formula *NMR ~ MIR* is the matrix of response variables (i.e. NMR spectra) and the right-hand side is the matrix of predictor variables (i.e. MIR spectra). We then used the resulting model to predict the NMR spectra from the acquired MIR spectra using the *predict* function, as follows:

```
> predict (MIR-PLSR, ncomp = 6, newdata = spectra.test)
```

The *newdata* argument introduces the MIR spectral data acquired for the unknown soil samples to the function. The *ncomp* argument defines the number of PLSR factors to be used by the model to make predictions.

## 3. Results and discussion

Being able to rapidly and inexpensively obtain NMR-derived information related to the composition of SOC will markedly improve current capacity to understand and monitor SOC dynamics including

Download English Version:

<https://daneshyari.com/en/article/6408567>

Download Persian Version:

<https://daneshyari.com/article/6408567>

[Daneshyari.com](https://daneshyari.com)