

The prediction of soil carbon fractions using mid-infrared-partial least square analysis

L. J. Janik^A, J. O. Skjemstad^A, K. D. Shepherd^B, and L. R. Spouncer^A

^ACSIRO, Land and Water, and CRC for Greenhouse Accounting, PMB No. 2, Glen Osmond, SA 5064, Australia.

^BWorld Agroforestry Centre (ICRAF), PO Box 30677 – 00100, Nairobi, Kenya.

^CCorresponding author. Email: jan.skjemstad@csiro.au

Abstract. This paper describes the application of mid-infrared (MIR) spectroscopy and partial least-squares (PLS) analysis to predict the concentration of organic carbon fractions present in soil. The PLS calibrations were derived from a standard set of soils that had been analysed for total organic carbon (TOC), particulate organic carbon (POC), and charcoal carbon (char-C) using physical and chemical means. PLS calibration models from this standard set of soils allowed the prediction of TOC, POC, and char-C fractions with a coefficient of determination (R^2) of measured v. predicted data ranging between 0.97 and 0.73. For the POC fraction, the coefficient of determination could be improved ($R^2 = 0.94$) through the use of local calibration sets. The capacity to estimate soil fractions such as char-C rapidly and inexpensively makes this approach highly attractive for studies where large numbers of analyses are required. Inclusion of a set of soils from Kenya demonstrated the robustness of the method for total organic carbon and charcoal carbon prediction.

Additional keywords: mid infrared, MIR, PLS, organic carbon, TOC, carbon fractions, charcoal.

Introduction

Soil organic carbon content is normally measured using some form of chemical oxidation (e.g. Walkley and Black 1934) or by dry combustion (e.g. Leco carbon determinator, Merry and Spouncer 1988) as total carbon. TOC values, however, may not adequately describe the role of carbon in many soil processes, and it may be more appropriate to partition it into several pools or fractions with varying degrees of reactivity or biological stability (van Veen *et al.* 1984; Parton *et al.* 1987; Jenkinson and Coleman 1994). These pools can be modelled with computer simulations, and range from labile organic matter derived from plant debris, to moderately to highly resistant carbon from humified organic matter, and inert or highly protected organic matter (Skjemstad *et al.* 1998).

For carbon modelling purposes, it would be useful to have a set of defined soil fractions that could be closely related to the conceptual pools used in particular carbon turnover models. Skjemstad *et al.* (1996) showed that a significant proportion of soil organic carbon in most Australian soils was present as charcoal (char) and that this fraction probably represented a soil pool which was highly resistant to microbial decomposition. This fraction can be determined routinely using high-energy ultraviolet (UV) photo-oxidation and ¹³C NMR spectroscopy and was considered to be a good candidate as a useful soil pool for modelling purposes. This fraction also appeared to fit well with the RothC model for C turnover (Jenkinson and Coleman 1994), which contained an inert pool as part of its pool structure. Subsequently, Skjemstad *et al.* (2004) used the char-C fraction along with the >53 µm particulate organic carbon (POC) fraction reported by Cambardella and Elliot (1992) to

initialise, calibrate, and verify the RothC model. In this model, the POC fraction is equated to the resistant plant material pool, char-C to the inert pool, and the difference between total organic carbon (TOC) and char-C plus POC pools to the humic pool.

The TOC and POC fractions can be measured easily, but the char-C analytical procedure cannot. The soil must be separated to isolate the <53 µm size fraction which normally contains >90% of the char, which is then subjected to high-energy UV photo-oxidation to remove labile and humified OC, and treated with HF. The residual fraction is then analysed by ¹³C nuclear magnetic resonance (NMR) spectroscopy to determine its char-C content (Skjemstad *et al.* 1999). NMR analysis is required because the material remaining after photo-oxidation represents a mixture of char and other organic carbon which is strongly protected by inclusion within microaggregates against the photo-oxidation process. The latter is usually very small, but with some soils can represent up to 50% of the remaining OC and would therefore cause an overestimation of the char-C content if NMR analysis was not performed. This fractionation scheme proposed by Skjemstad *et al.* (1999) is therefore very time-consuming and expensive due to the difficulty in separating enough material to obtain an acceptable NMR spectrum (~50 mg C).

For the routine analysis for carbon fractions of a large number of soils, a simpler and cheaper method of acceptable accuracy is needed. The fractionation scheme discussed above relies on the availability of photo-oxidation equipment, access to a solid-state ¹³C NMR spectrometer, and hazardous chemical treatments with HF, and is also laborious. Infrared spectroscopy offers a simple, rapid, and low-cost alternative, with a further advantage that it is sensitive to the chemistry of both organic and mineral

components in the soil. MIR spectroscopy is characterised by fundamental vibrations of molecules associated with particular chemical functional groups. For example, the technique enables the identification of specific soil minerals and of organic matter functional groups such as alkyl, carboxylic (protonated and non-protonated), carbohydrates, amide, amine, and most importantly aromatic functional groups (Van der Marel and Beutelspacher 1976; Skjemstad and Dalal 1987; Theng and Tate 1989; Theng *et al.* 1992; Piccolo 1994; Janik and Skjemstad 1995; Wander and Traina 1996; Janik *et al.* 1998).

The advantages of MIR spectroscopy, particularly when combined with multivariate chemometric techniques such as partial least-squares (PLS) regression for the prediction of sample properties, are its analytical speed and simplicity. No fractionation is required, hazardous chemical reagents are avoided, and many soil components can be predicted from a single spectrum, although air-drying and grinding is advantageous. The MIR-PLS technique has already been shown to be suitable for the prediction of TOC from spectra of whole soils (Janik and Skjemstad 1995; Janik *et al.* 1998) and because it is sensitive to the various organic functional groups contributing to soil organic matter should, in principle, be sensitive to the distribution of different carbon pools, offering a realistic and practical approach for large-scale carbon pool analysis. The increased potential of MIR spectroscopy has resulted from the widespread availability of PLS software, which provides a powerful and robust quantification tool (Holmgren and Nordén 1988; Janik and Skjemstad 1995; Janik *et al.* 1998; Hazel *et al.* 1997).

The procedure for PLS analysis adopted here closely follows that detailed by Haaland and Thomas (1988), and later by Janik and Skjemstad (1995) and Janik *et al.* (1998) for soils. MIR spectra typically contain >2000 data points, many of which can potentially correlate with the soil carbon data. For predictions of soil properties, MIR spectra, and the corresponding analytic data are transformed into a smaller set of orthogonal PLS loadings and loading scaling terms (scores), thus combining the spectral and concentration data into a multivariate calibration model (Haaland and Thomas 1988). This calibration model can then be used for predictions of property values from spectra of unknown samples. A further advantage of the PLS analysis method is that it can provide regression coefficient 'signatures' of frequencies with high correlation with the analyte concentrations. The PLS method also results in spectral components of the calibration (loading weights) where the first few loading weights are the spectral signatures of the 'pure component' related to the property of interest, e.g. TOC, POC, or char-C in this case (Haaland and Thomas 1988; Janik and Skjemstad 1995; Janik *et al.* 1998).

This paper describes the use of MIR-PLS for the prediction of TOC and the 2 soil carbon fractions, POC and char-C, which had previously been determined by a combination of physical and chemical separations and ^{13}C NMR spectroscopy.

Materials and methods

Soils and charcoal

A set of 177 soil samples from varying depths within the 0–0.50 m layer was selected to cover a large range of different

soil types and parent materials from all States in Australia, ensuring that the proposed MIR method would be generally representative of Australian soils. Most samples were from under native vegetation; however, samples from cultivated soils were also included. Samples ranged in texture from <1 to >80% clay and included calcareous soils. This set of samples was used for the TOC calibration, and of these, 141 samples were selected for the POC calibration and 121 samples for char-C calibration. To test the robustness of the TOC calibration, 2 further sets of soils were selected as test samples for validating the prediction for TOC by MIR and tested for prediction—a set of 557 Australian samples not used in the calibration, and a set of 235 soils from Kenya. The Kenyan samples were topsoils (0–0.15 m depth) selected to represent the spectral diversity in a diverse collection of >1000 soils taken from western Kenya, covering a wide range of parent materials (mudstones, claystones, acid igneous rocks, basic igneous rocks, and alluvium) and landscape positions in dry sub-humid to humid climatic zones. The subset of 235 Kenya soils was selected to sample the spectral diversity in soils with TOC ranging from 2.5 to 94.2 g C/kg soil and a mean of 27.7 g C/kg soil. Subsamples of the 2 mm air-dried material from all soils were used for chemical analysis and further ground for 60 s in an orbital mill (TEMATM) for MIR analysis.

A sample of wood (species unknown) charcoal was collected from the soil surface after a recent fire in the Northern Territory. The sample showed no contamination from soil and had a TOC content of 84.9%. A Vertosol (Isbell 1996) used in previous charcoal studies (SS8) (Skjemstad *et al.* 1999) was also used in this study for spectral subtraction studies as described under the section on infrared methodology.

Soil fractionation procedures

The procedure used was essentially that reported by Skjemstad *et al.* (1999). Briefly, the soils were sodium-saturated with sodium hexametaphosphate by shaking overnight. The sodium saturated soil suspensions were passed through a 200 μm sieve and a 53 μm sieve in series. The fractions remaining in the sieves were gently worked with a spatula to assist passing of soil aggregates through the sieves. The POC fraction was calculated as the combined organic carbon remaining on the sieves. The <53 μm fractions were further treated using high energy ultraviolet (UV) photo-oxidation (Skjemstad *et al.* 1996, 1999). The residue remaining following the photo-oxidation procedure was treated with HF and then analysed by solid-state cross polarisation with magic angle spinning (CPMAS) ^{13}C NMR spectroscopy to determine the char-C content (Skjemstad *et al.* 1999).

This scheme, therefore, provided the following 3 soil organic carbon fractions including TOC, POC (>200 μm plus 200–53 μm fractions), and the char-C fraction, and was applied to the Australian soils. The humic (HUM) fraction could be calculated as TOC minus POC and minus char-C. The TOC, POC, HUM, and char-C fractions are the same as those used by Skjemstad *et al.* (2004) for initialisation of the RothC carbon turnover model.

Chemical analyses

The TOC content of the soils and POC fractions were determined by dry combustion in a LECO furnace (Merry and

Spouncer 1988). Where appropriate, samples were corrected for carbonate carbon determined by Collins Calcimeter (Loveday and Reeve 1974). Organic carbon on other fractions such as the <53 μm fractions were determined in suspension by dichromate oxidation using the procedure of Heanes (1984).

Infrared spectroscopy and PLS analysis

Subsamples of the ground soil samples were analysed as neat powders using the MIR diffuse reflectance technique reported by Janik and Skjemstad (1995) and Janik *et al.* (1998). The equipment consisted of a rapid scanning Fourier Transform spectrometer (Bio-Rad 175C) scanning at 1 scan/s with an extended range KBr beamsplitter and DTGS detector, with a spectral range of 8300–470 cm^{-1} at 8 cm^{-1} resolution. Spectral frequencies were referenced against an internal He-Ne laser to give a precision and accuracy of 0.01 cm^{-1} and each sample was scanned for 60 scans using a diffuse reflectance accessory (HarrickTM DRS-3SO) with off-axis geometry. Only the MIR portion of the spectra between 4000 and 500 cm^{-1} was used for the PLS analysis. An initial KBr blank spectrum was run to test the spectrometer performance and as a reference for calculating the sample spectra in absorbance units. Sample absorbance spectra were acquired in GRAMS (*.SPC) format (Thermo ElectronTM NH).

The PLS calibrations of TOC, POC, and char-C were carried out using full 'leave-one-out' cross validation with the PLSplus/IQTM (Thermo-Electron GRAMSTM) software. Haaland and Thomas (1988) had previously described the algorithms used in this study for an earlier version of the software (PLS-PLUSTM GalacticTM NH). Optimum frequency ranges and pre-processing parameters were determined empirically for each component to give maximum R^2 values. Of the 2000 data points available for analysis, the spectra were averaged over successive 2-point segments, mean centred, and auto-baseline corrected. The optimum analytical frequency range used for all fractions was 4000–1030 cm^{-1} . All OC reference data were transformed to a square root of the data before calibration, in order to minimise non-linearity in the calibration. The resulting cross-validation and test sample predictions were back-transformed by squaring the PLS predicted data.

To determine the effect of the soil matrix on charcoal spectral features, a finely ground subsample of wood charcoal was mixed with a finely ground soil sample (SS8) to produce a mixture of 5% wt/wt charcoal/soil mixture. MIR spectra were obtained for the charcoal, soil, and mixture. Another spectrum of the charcoal was then obtained by subtracting the spectrum of the soil from that of the mixture so that no soil spectral features remained. This spectrum was then compared with the original charcoal spectrum to ascertain if there were any effects, such as changes in refractive index, of the soil matrix on the charcoal spectrum.

B-coefficients

For prediction, the regression model can be written as:

$$Y = b_0 + b_1 X_1 + \dots + b_k X_k + e$$

where the observed response values (Y) (e.g. char concentration) are approximated by a linear combination of the values of the spectral intensities (X), coefficients (b) called B-coefficients, and

an error term (e). The estimated B-coefficients are determined from the observed Y values and PLS scores for the optimum number of PLS factors and indicate the correlation between the Y values and each of the X spectral intensities. The B-coefficients can help identify the important X -variables in relationship with a given Y -variable. The B-coefficients contain the spectral information that is most important in driving the PLS model and indicate which variables are important in explaining both the X - and the Y -data. The coefficients indicate which X -variables are contributing to the modelling of the structure in the Y -variables.

Results and discussion

Soil TOC and fractions

The distributions of TOC, POC, and char-C as g C/kg soil for the samples used in the calibration are depicted in Fig. 1. The POC and char fractions showed a considerable skew towards low carbon values. The TOC distribution was similar to that reported by Spain *et al.* (1983) for a much larger set of >3000 Australian soils. Spain *et al.* (1983) found that two-thirds of the soils in their

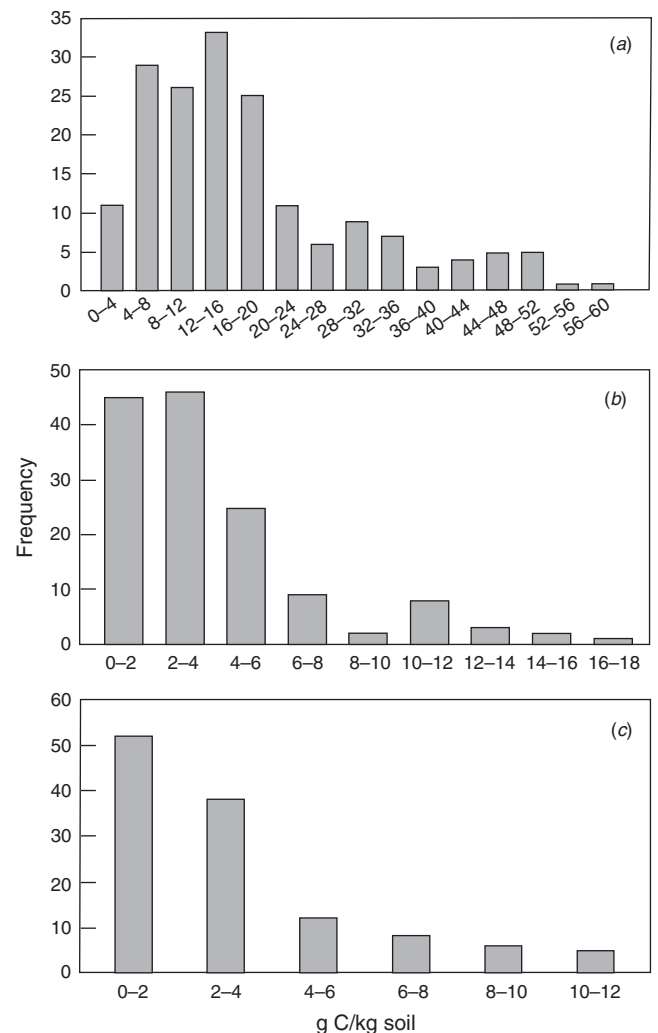


Fig. 1. Frequency distribution of (a) TOC, (b) POC, and (c) char-C for samples used for cross-validation.

database had TOC levels <20 g C/kg soil and only 6% exceeded 60 g C/kg soil. The average values for the TOC, POC, HUM-C, and char-C were 18.1, 4.1, 10.8, and 3.2 g C/kg soil, respectively.

The relationships between TOC and the 141-sample POC set and the 121-sample char-C set, respectively, are given in Fig. 2a and b. The correlations between TOC and the 2 fractions were both poor ($R^2 = 0.25$ and 0.04, respectively), showing considerable scatter. Figure 2c shows the equally poor correlation between POC and char-C for the 121-sample char-C

soil set with an $R^2 = 0.02$. It is clear, therefore, that there is no covariance between the carbon contents of the fractions and that simple relationships between the fractions could not be used to predict one from the other.

PLS analysis

The PLS cross-validation data for TOC and the 2 fractions are presented in Table 1. PLS cross-validation of TOC resulted in a standard error of cross-validation (SECV) of 0.09 and an $R^2 = 0.94$, similar to data already published for Australian soils by Janik and Skjemstad (1995) and Janik *et al.* (1998). A plot of predicted against reference TOC is presented in Fig. 3a. Although 14 factors were found to be optimal for prediction of TOC, the first 4 were sufficient to account for 85% of the explained variance.

Cross-validation of the POC fractions resulted in $R^2 = 0.71$ and $SECV = 1.13$. Figure 3b shows that there is considerable scatter about the line, illustrating the difficulty of predicting the POC fraction with MIR. As discussed earlier, the POC fraction represents on average about 23% of the OC in this set of soils and consists largely of plant residues in various states of decomposition. As a result, its chemistry is similar to that of the bulk TOC, with the largest difference being due to a slightly higher carbohydrate content in the POC fraction (Skjemstad *et al.* 1999). Since the POC fraction is representative of the plant input, variations in the nature of the vegetation between different ecosystems may also contribute to the scatter.

The char-C cross-validation predictions were better than that of the POC, with an $R^2 = 0.86$ and a $SECV = 0.04$. Figure 3c shows that there were a few outliers, but removing these from the calibration did not significantly improve the regression. Considering the many extractions and chemical treatments required to calculate the reference laboratory values for char-C, $R^2 = 0.86$ can be considered an excellent result. Prediction of char-C would be expected to be relatively robust, since unlike POC, charcoal has a specific chemistry different to that of most other carbon species contributing to the TOC, being dominated by condensed aromatic structures (Skjemstad and Taylor 1999; Schmidt and Noack 2000) rather than aliphatic species. One possible source of error in the PLS prediction of char-C is due to the uncertainty of not having accounted for all of the soil char-C in the reference data, since the analytical method used only determines char-C in the <53 μm fraction. In reality, some unmeasured char-C may also occur in the >53 μm fraction, whereas the MIR analyses presented here was applied to the soil as a whole, possibly leading to higher results in some samples. In order to minimise this issue, the POC fractions were observed under a binocular microscope to determine if any charcoal was present in this fraction. In almost all cases, no charcoal could be

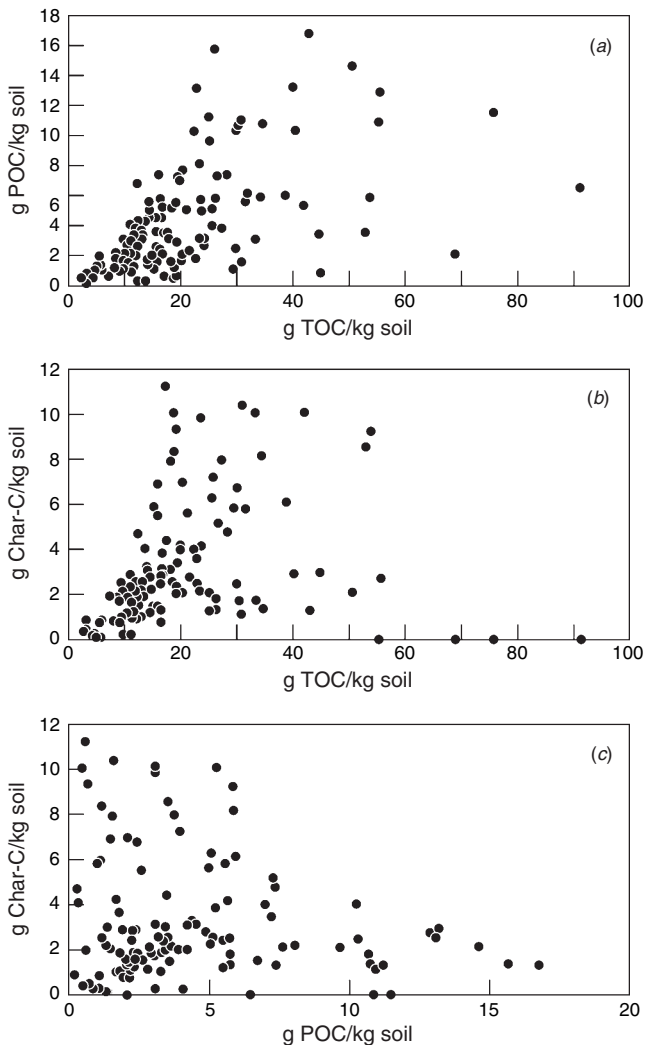


Fig. 2. The relationship between (a) TOC and POC, (b) TOC and char-C, and (c) POC and char-C.

Table 1. Calibration data for prediction of TOC, POC, and char-C

Analyte	No. of samples	R^2	SECV ^A (g C/kg soil)	Factors used for prediction	Range (g C/kg soil)
TOC	177	0.94	0.09	14	0.8–62.0
POC	141	0.71	1.13	10	0.2–16.8
Char-C	121	0.86	0.04	13	0.0–11.3

^AStandard error of cross validation.

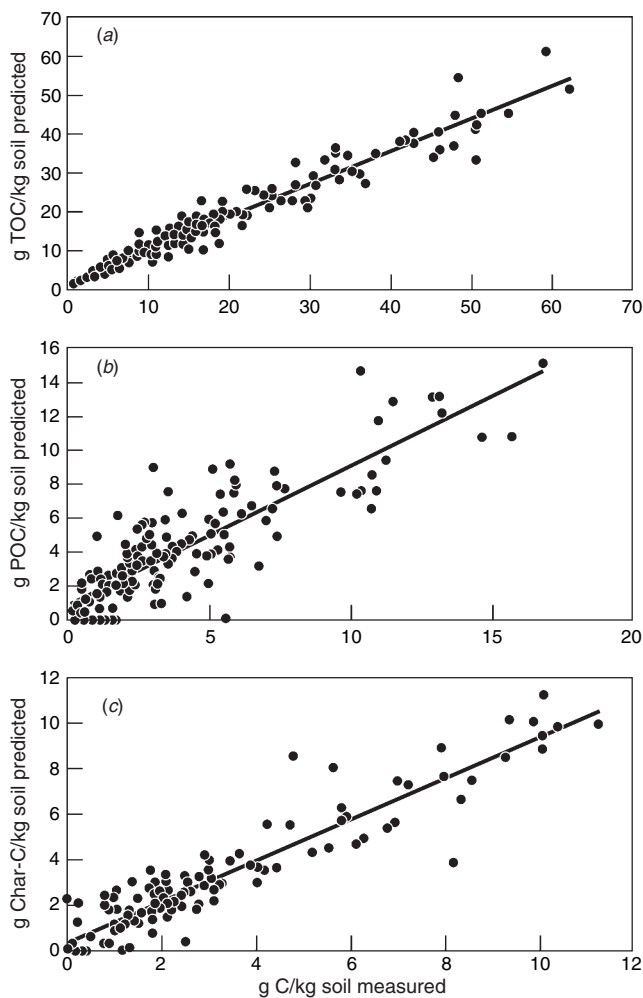


Fig. 3. PLS cross validation plots for (a) TOC, (b) POC, and (c) char-C.

observed, or very little. For one sample, a large amount of char-C was observed and this sample was removed from both POC and char-C calibration sets. Small contributions from charcoal in the POC fraction may explain some of the high positive prediction errors at char-C values <3.0 g C/kg soil (Fig. 3c).

For the predictions to be considered sufficiently robust to be transferable to samples outside of the calibration set, the calibration spectra must be characterised by spectral features consistent with the various forms of organic matter expected to contribute to the carbon pools, rather than covariance due to other features such as clay content or mineralogy. To test the correlation between carbon pool chemistry and spectral dependence of the PLS model, average beta coefficients of the first 4 PLS loadings for the TOC, POC, and char-C were developed. These are presented in Fig. 4 and represent the major input to the prediction model. For TOC (Fig. 4a), the dominant peaks are all characteristic of organic matter representing alkyl- CH_2 stretching modes (2922 and 2853 cm^{-1}), carbohydrate overtones of the $-\text{COH}$ stretch (2200 – 2000 cm^{-1}), carboxylic acid- COOH (1708 cm^{-1}), amide I and II bands (1660 and 1556 cm^{-1}), alkyl- CH_2 deformation (1450 – 1400 cm^{-1}), aromatic- CH in

plane deformation (1238 cm^{-1}), and carbohydrate- COH stretch (1051 cm^{-1}). The beta coefficient for POC (Fig. 4b) is similar to TOC except that the alkyl, carbohydrate, and carboxylic acid spectral peaks are stronger and the amide peaks are weaker. As well, the peaks in the <1400 cm^{-1} region are different and are more suggestive of cellulose, although lignin can also contribute to this region. The 1230 cm^{-1} peak attributed to aromatic structures is also absent. The beta coefficient for char-C (Fig. 4c) shows 3 broad bands near 1581 , 1393 , and 1223 cm^{-1} that can be attributed to $\text{C}=\text{C}$ skeletal vibrations, phenolic, or COO stretching vibrations and ring $\text{C}-\text{H}$ in plane deformations, respectively, and are very similar to spectra of finely divided and slightly oxidised black carbon (BC) reported by Smith *et al.* (1975). The presence of the 1393 cm^{-1} band (carboxylate anion) is consistent with observations of soil charcoals through NMR studies, which show that they are associated with carboxyl groups (Skjemstad *et al.* 1996).

The average beta coefficient spectra given in Fig. 4 are therefore consistent with organic materials and show differences that are also consistent with the expected correlations with the chemistry of the different carbon pools. Some small, or negative, peaks representing kaolinite and gibbsite mineral phases are present near 3700 – 3500 cm^{-1} . The loading weights for the prediction of TOC should reflect the broad range of organic materials found in soils, and Fig. 4a reflects the major groupings of lipids, polysaccharides, proteins, and charcoal. The comparison of a spectrum of charcoal, the spectrum of charcoal obtained by spectral subtraction from a soil admixture, and the first 4 and 6 beta coefficients for char-C prediction are given in Fig. 5. There appears to have been some influence by the soil matrix, particularly with the 1607 cm^{-1} peak (Fig. 5b), which has been broadened to a secondary peak at 1581 cm^{-1} . The appearance of a peak at 1199 cm^{-1} is in agreement with a peak in the same region in the beta coefficient spectra. The band between 1750 and 1550 cm^{-1} becomes more resolved as more coefficients are included and clear peaks are now resolved at 1718 and 1583 cm^{-1} . The similarities strongly indicate that aromatic structures (charcoal) are characteristic of the charcoal average beta coefficients. Smith *et al.* (1975) oxidised BC slightly by grinding samples for 18 h and report major peaks for these materials to be at 1720 , 1580 , and 1230 cm^{-1} , which agrees well with the peaks observed in the average beta coefficient spectra. That the MIR spectra should reflect oxidised charcoal is logical, since the MIR calibration was made against charcoal content of the <53 μm fraction only, which Skjemstad *et al.* (1996, 1999) demonstrated to be high in carboxyl groups. The average beta coefficient spectrum for POC prediction is similar to TOC but does not contain the charcoal signals and is more typical of less decomposed materials, being higher in alkyls and cellulose/lignin and lower in protein.

The general applicability of the TOC calibration was tested against the large set of 734 (177 plus 577) soil samples. Several thousand soil samples were available but these samples were more biased towards low TOC values, and were thus chosen to represent the soil types in Australia and were also similar to the TOC distribution reported by Spain *et al.* (1983). Calibration models using the entire 734 samples did not improve the regression ($R^2 = 0.95$). However, predicting TOC in the 557 soils from the smaller calibration set gave excellent

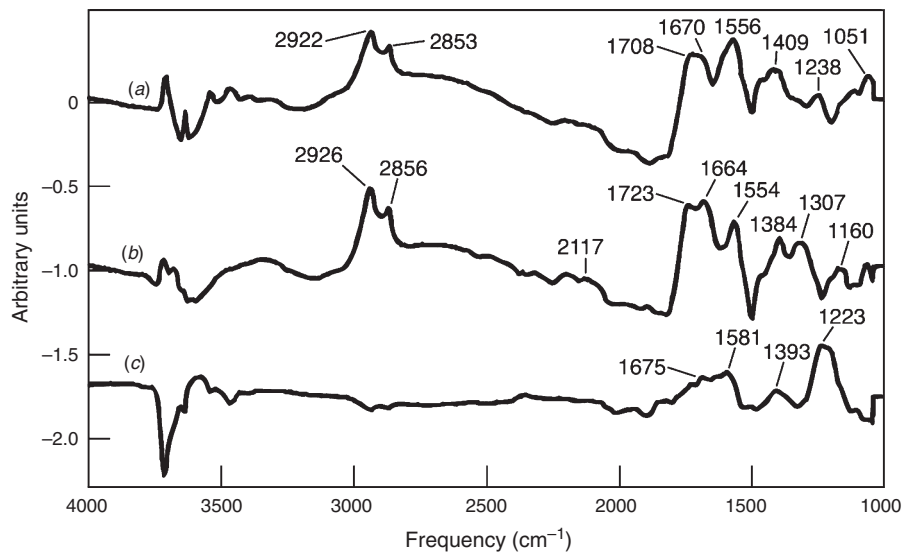


Fig. 4. Beta coefficient spectra of the average of the first 4 loadings for (a) TOC, (b) POC, and (c) char-C.

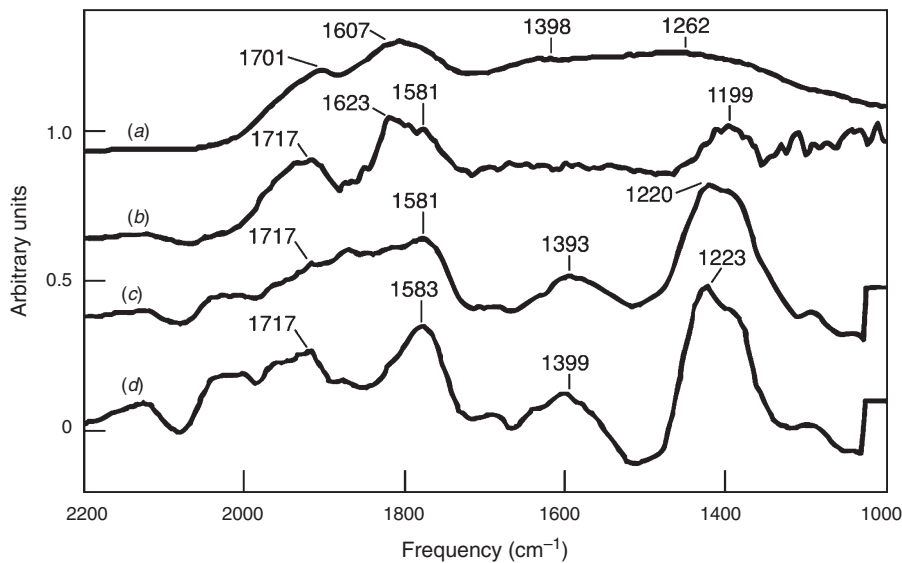


Fig. 5. MIR spectra of (a) a charcoal sample, (b) spectrum of the same charcoal after subtraction of an admixture with soil, and char-C beta coefficient spectra of the average of the (c) first 4 loading weights and (d) first 6 loading weights.

results ($R^2 = 0.94$) with a standard error of prediction (SEP) of 0.62 g C/kg soil.

To determine if the prediction of POC values could be improved using a local calibration set, 36 extra samples from a series of 9 profiles (0–0.05, 0.05–0.10, 0.10–0.20, and 0.20–0.30 m) taken randomly from within a forested site in eastern Australia were analysed by both the laboratory method and MIR/PLS. Prediction of these samples using the calibration set of 141 samples gave $R^2 = 0.90$ with SEP = 0.32 between measured and predicted values, demonstrating that the POC calibration was robust for the prediction of new unknowns.

Calibrating on this local set alone gave an $R^2 = 0.94$ for POC, which is a considerable improvement on the global calibration. Again, such an improvement would be expected because at the local level, changes in soil type and vegetation are minimised. For analyses such as POC, where a large number of analyses are required within a restricted geographical region or within a single soil type, the use of local calibrations may be more accurate.

For char-C analysis, only a limited number of analyses using the high-energy UV photo-oxidation method were available and all of these were included in the calibration set for cross-validation. To assess the robustness of the MIR-PLS approach

to char-C analysis, 15 of the samples were selected at random but included samples from across the range of char-C content and soil type. These were removed from the calibration for a test set and a new calibration undertaken on the remainder. For the remaining 106 samples, the new calibration PRESS analysis gave an $R^2 = 0.82$, which is slightly less than the $R^2 = 0.86$ obtained for the full 121 samples. Prediction of the 15 excluded samples using this new calibration gave acceptable data with an R^2 of measured against predicted of 0.78 with an SEP of 0.04 g C/kg soil—a reasonable result considering the small number of calibration samples (Fig. 6).

Kenyan soils

The Australian calibration set (177 samples) was used to predict TOC values for the Kenyan soil samples. The MIR-predicted values provided a strong relationship against the measured values ($R^2 = 0.95$) but the slope of the line was only 0.74 and exhibited bias and curvature (Fig. 7a). Therefore, although there is a strong relationship, there is also bias in the data with the MIR/PLS, predicting lower values than the Leco measured data. A new cross-validation was set up using the Kenyan soils, resulting in an excellent correlation between measured and predicted TOC ($R^2 = 0.98$, Fig. 7b).

Other experiments were run where half of the Australian soils were mixed with half of the Kenyan soils to produce an alternative calibration set. Cross-validation resulted in an $R^2 = 0.96$, and predicting the remaining Australian and Kenyan soils gave R^2 values of 0.94 and 0.96, respectively, with slopes close to 1.0. This illustrates the importance of the inclusion of local soils in any calibration set used for predicting TOC.

Char-C was also predicted on these soils but no measured char-C data were available. Skjemstad *et al.* (1996, 1999) and Smernik *et al.* (2000) previously demonstrated that the presence of char-C can be detected through ^{13}C NMR spectroscopy, because the char-C consists almost entirely of aryl-C in fused rings and so its presence makes a large contribution to the 130 ppm peak in ^{13}C NMR spectra. Although the size of the 130 ppm peak alone cannot be used to directly quantify char-C in samples that have not been photo-oxidised, its variation is a strong indicator of the contribution of char-C to the spectrum. To test if the char-C predictions from the Kenyan soils were reasonable, 7 soils predicted to have varying amounts of char-C

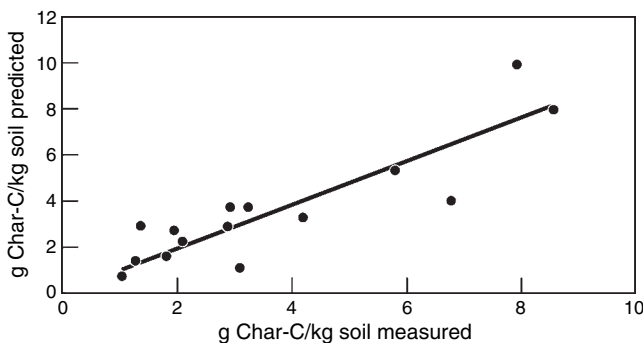


Fig. 6. Plot of measured versus predicted values for char-C from 15 soils not included in the calibration set.

were selected for validation, treated with HF, and analysed by CPMAS ^{13}C NMR spectroscopy. The spectra were integrated and the contribution of aryl region in each was calculated. Plotting MIR-predicted char-C against the NMR-measured aryl-C expressed in g C/kg soil gave an $R^2 = 0.95$ (Fig. 8a). The slope of the line was 0.8 with an intercept of 1.55, which is in keeping with the fact that organic materials other than char-C also contribute to the aryl region (positive intercept) and the CP process is not as sensitive to char-C as other organic materials found in soil (Skjemstad *et al.* 1999; Smernik and Oades 2000), resulting in a slope < 1.0 . Although a good correlation was achieved when data were expressed on a g C/kg soil basis, this can be affected by the TOC. For example, if char-C made a constant percentage contribution to each soil, then the relationship would be due solely to changes in TOC. To determine that this was not the case, the data were recalculated as a percentage of the TOC. This relationship was not as good ($R^2 = 0.81$) but demonstrated that the MIR was sensitive to aryl-C and was reflecting the presence of char-C (Fig. 8b).

Conclusions

The accurate determination of the pool structure of soil carbon is an important prerequisite for the application of conceptual carbon simulation models. Traditional techniques for separation and measurement of the distribution of carbon within the carbon pools are difficult, lengthy, and/or not particularly precise, precluding their widespread use. The PLS technique based on MIR spectra overcomes many of the problems associated with the determination of soil carbon pool chemistry. It allows for rapid analysis, often as quick as for a single Leco TOC

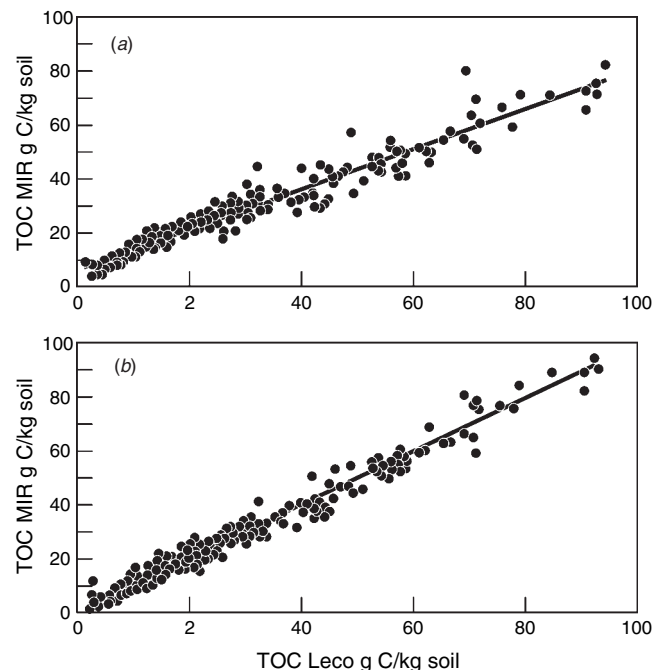


Fig. 7. Plot of the measured versus predicted TOC values for 244 Kenyan soils (a) using the Australian calibration set and (b) cross validation plot using data from Kenyan soils.

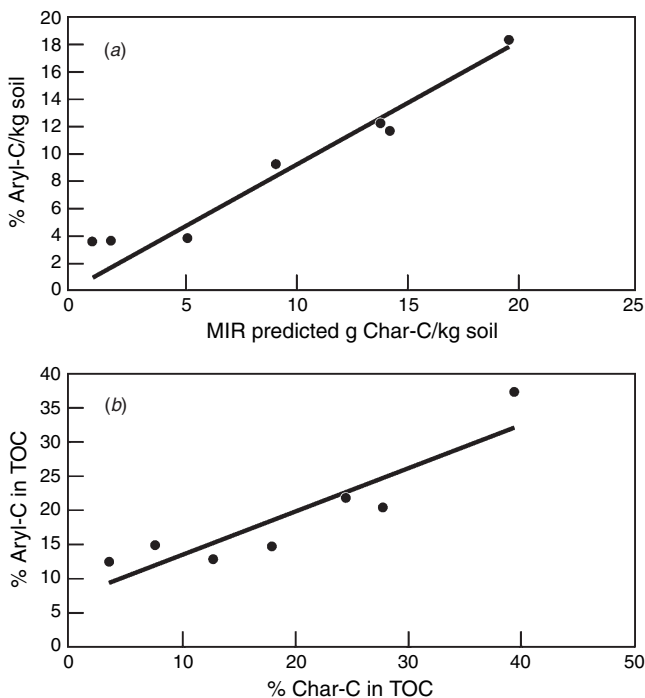


Fig. 8. Plot of MIR predicted char-C v. aryl-C measured by ^{13}C NMR as (a) g C/kg soil and (b) as a percentage of the TOC.

determination, not only for TOC but also for the other pools from a single spectrum.

Predictive accuracy for TOC was excellent and that for char-C good. The determination of carbon within the POC pool was less precise, although still acceptable for carbon modelling applications and could be improved with the use of local calibration sets. The MIR/PLS method clearly demonstrates the feasibility of using it as a surrogate for the traditional methods of laboratory analysis. Not only is the MIR/PLS method accurate for several fractions, but the simplicity of its use has the potential to expand the range of carbon analyses able to be carried out in many laboratories. Analysis of samples from another continent with different soil types and environments with acceptable results demonstrates the robustness of the TOC and char-C analyses, particularly if samples from those areas are included in the calibration.

References

- Cambardella CA, Elliot ET (1992) Particulate soil organic matter changes across a grassland cultivation sequence. *Soil Science Society of America Journal* **56**, 777–783.
- Haaland DM, Thomas VT (1988) Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical Chemistry* **60**, 1193–1202. doi: 10.1021/ac00162a020
- Hazel G, Bucholtz F, Aggarwal GN, Ewing KJ (1997) Multivariate analysis of mid-IR spectra of hydrocarbon-contaminated wet soils. *Applied Spectroscopy* **51**, 984–989. doi: 10.1366/0003702971941359
- Heanes DL (1984) Determination of total organic-C in soils by an improved chromic acid digestion and spectrophotometric procedure. *Communications in Soil Science and Plant Analysis* **15**, 1191–1213.
- Holmgren A, Nordén B (1988) Characterization of peat samples by diffuse reflectance FT-IR spectroscopy. *Applied Spectroscopy* **42**, 225–262.
- Isbell RF (1996) 'The Australian soil classification.' (CSIRO Publishing: Melbourne)
- Janik LJ, Merry RH, Skjemstad JO (1998) Can mid-infrared diffuse reflectance analysis replace soil extractions? *Australian Journal of Experimental Agriculture* **38**, 681–696. doi: 10.1071/EA97144
- Janik LJ, Skjemstad JO (1995) Characterization and analysis of soils using mid-infrared partial least squares. II. Correlations with some laboratory data. *Australian Journal of Soil Research* **33**, 637–650. doi: 10.1071/SR9950637
- Jenkinson DS, Coleman K (1994) Calculating the annual input of organic matter to soil from measurements of total organic matter and radiocarbon. *European Journal of Soil Science* **45**, 167–174. doi: 10.1111/j.1365-2389.1994.tb00498.x
- Loveday J, Reeve R (1974) Carbonate determination. In 'Methods of analysis of irrigated soils'. Technical Communications, Commonwealth Agricultural Bureau No. 54. (Ed. J Loveday) pp. 131–134. (CAB International)
- Merry RH, Spouncer LR (1988) The measurement of carbon in soils using a microprocessor-controlled resistance furnace. *Communications in Soil Science and Plant Analysis* **19**, 707–720.
- Parton WJ, Schimel DS, Cole CV, Ojima DS (1987) Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Science Society of America Journal* **51**, 1173–1179.
- Piccolo A (1994) Advanced infrared techniques (FT-IR, DRIFT, and ATR) applied to organic and inorganic soil materials. In '15th World Congress of Soil Science'. Acapulco, Mexico. Transactions 3a, pp. 3–22. (IUSS)
- Schmidt MWI, Noack AG (2000) Black carbon in soils and sediments: Analysis, distribution, implications, and current challenges. *Global Biogeochemistry Cycles* **14**, 777–793. doi: 10.1029/1999GB001208
- Skjemstad JO, Clarke P, Taylor JA, Oades JM, McClure SG (1996) The chemistry and nature of protected carbon in soil. *Australian Journal of Soil Research* **34**, 251–271. doi: 10.1071/SR9960251
- Skjemstad JO, Dalal RC (1987) Spectroscopic and chemical differences in organic matter of two Vertisols subjected to long periods of cultivation. *Australian Journal of Soil Research* **25**, 323–335. doi: 10.1071/SR9870323
- Skjemstad JO, Janik LJ, Taylor JA (1998) Non-living organic matter: what do we know about it? *Australian Journal of Experimental Agriculture* **38**, 667–680. doi: 10.1071/EA97143
- Skjemstad JO, Spouncer LR, Cowie B, Swift RS (2004) Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools. *Australian Journal of Soil Research* **42**, 79–88. doi: 10.1071/SR03013
- Skjemstad JO, Taylor JA (1999) Does the Walkley-Black method determine soil charcoal? *Communications in Soil Science and Plant Analysis* **30**, 2299–2310.
- Skjemstad JO, Taylor JA, Smirnik RJ (1999) Estimation of charcoal (char) in soils. *Communications in Soil Science and Plant Analysis* **30**, 2283–2298.
- Smernik RJ, Oades JM (2000) The use of spin counting for determining quantitation in solid state ^{13}C NMR spectra of natural organic matter 2. HF-treated soil fractions. *Geoderma* **96**, 159–171. doi: 10.1016/S0016-7061(00)00007-0
- Smernik RJ, Skjemstad JO, Oades JM (2000) Virtual fractionation of charcoal from soil organic matter using solid state ^{13}C NMR spectral editing. *Australian Journal of Soil Research* **38**, 665–683. doi: 10.1071/SR99115
- Smith DM, Griffin JJ, Goldberg ED (1975) Spectroscopic method for the quantitative determination of elemental carbon. *Analytical Chemistry* **47**, 233–238. doi: 10.1021/ac60352a032

- Spain AV, Isbell RF, Probert ME (1983) Soil organic matter. In 'Soils: An Australian viewpoint'. pp. 551–563. CSIRO Division of Soils. (CSIRO: Melbourne/Academic Press: London)
- Theng BKG, Tate KR (1989) Interactions of clays with soil organic constituents. *Clay Research* **8**, 1–10.
- Theng BKG, Tate KR, Becker-Heidmann P (1992) Towards establishing the age, location, and identity of the inert soil organic matter of a spodosol. *Zeitschrift für Pflanzenernährung und Bodenkunde* **155**, 181–184.
- Van der Marel HW, Beutelspacher H (1976) Clay and related minerals. In 'Atlas of infrared spectroscopy of clay minerals and their admixtures'. (Elsevier Scientific: Amsterdam)
- van Veen JA, Ladd JN, Frissel MJ (1984) Modelling C and N turnover through the microbial biomass of soil. *Plant and Soil* **76**, 257–274. doi: 10.1007/BF02205585
- Walkley A, Black LA (1934) An examination of the Dgtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science* **37**, 29–38. doi: 10.1097/00010694-193401000-00003
- Wander MM, Traina SJ (1996) Organic matter fractions from organically and conventionally managed soils: II. Characterization of composition. *Soil Science Society of America Journal* **60**, 1087–1094.

Manuscript received 6 July 2006, accepted 19 December 2006