Geoderma 461 (2025) 117502

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

# Comparison of pyrogenic carbon abundance in coarse-textured soil by hydrogen pyrolysis, NMR and dichromate oxidation and MIR-PLSR

Jonathan Sanderman [a,1], Jordahna Haig [b], Sourav Das [c,2], Colleen Partida [a], Christina Asanopoulos [d], Michael I. Bird [e,*]

[a] Woodwell Climate Research Center, 149 Woods Hole Road, Falmouth, MA 02540, United States
[b] College of Science and Engineering, James Cook University, Cairns, QLD 4870, Australia
[c] School of Electrical Engineering, Computing and Mathematical Sciences, Science & Engineering, Building 314, Wark Avenue, Curtin University Bentley, WA 6102, Australia
[d] CSIRO Agriculture & Food, PO Box 200, Glenside SA 5065, Australia
[e] College of Science and Engineering and ARC Centre of Excellence for Indigenous and Environmental Histories and Futures, James Cook University, Cairns, QLD 4870, Australia

ARTICLE INFO

Handling Editor: Dr Cornelia Rumpel

*Keywords:*
Pyrogenic carbon
Soil carbon
Charcoal
Hydrogen pyrolysis
Nuclear magnetic resonance
Mid-infrared spectroscopy
Partial least squares regression

ABSTRACT

Soil pyrogenic carbon (PyC) is of considerable significance to the global carbon cycle as a carbon pool which is resistant to mineralization and thus offers opportunities to facilitate net carbon sequestration. Quantifying the size and dynamics of the soil PyC pool is hampered by the large number of techniques that yield a wide range of abundances even when applied to the same sample. We used hydrogen pyrolysis to quantify stable polycyclic aromatic carbon (SPAC) of pyrogenic origin ($PyC_{SPAC}$) in a globally distributed set of coarse-textured soils, in which the percentage of particles finer than 53 μm ranged from 0.1 to 24.1 % (mean = 7.2 ± 5.8 % 1σ). $PyC_{SPAC}$ values ranged from 0 to 0.37 % (mean = 0.08 ± 0.06 %). We compared the $PyC_{SPAC}$ values with estimates derived from nuclear magnetic resonance spectroscopy ($PyC_{NMR}$) and found a strong correlation between the two (r = 0.90). However, the $PyC_{NMR}$ estimates were ~7 times higher than $PyC_{SPAC}$ values, attributed partly to NMR measuring a wider range of pyrogenic molecules but also likely due to the inclusion of aromatic 'resistant' soil carbon of non-pyrogenic origin. In contrast, there was little correspondence between either $PyC_{SPAC}$ or $PyC_{NMR}$ and abundances determined by dichromate oxidation ($PyC_{OREC}$). Partial least squares modelling of the mid-infrared (MIR) spectra was able to predict both $PyC_{SPAC}$ and $PyC_{NMR}$ values with high confidence (r = 0.77 and 0.94 respectively). The study suggests that, with appropriate scaling factors, $PyC_{SPAC}$ and $PyC_{NMR}$ can be directly compared, and both can be predicted by MIR.

## 1. Introduction

Pyrogenic carbon (PyC; charcoal, soot, black carbon, biochar, Inert or Resistant Organic Matter) refers to pyrolysed carbon that is a solid residue resulting from the incomplete combustion of organic matter during biomass burning and the consumption of fossil fuels. PyC ranges in size from macroscopic fragments to individual pyrogenic molecules, and is almost ubiquitously present in the atmosphere, soils, sediments, ice, terrestrial water bodies and the ocean (Bird et al., 2015, Santín et al., 2015). A variable fraction of global PyC production annually is incorporated into the Total Organic Carbon (TOC) pool in the soil. This soil PyC is increasingly recognized to be of considerable significance to the global carbon cycle (Bird et al., 2015; Santín et al., 2015). Pyrogenic carbon sequestration is a 'legacy' carbon sink (Bowring et al., 2022) and

*Abbreviations:* PyC, Pyrogenic Carbon; TOC, Total Organic Carbon; HyPy, Hydrogen pyrolysis; SPAC, Stable Polycyclic Aromatic Carbon, as measured by hydrogen pyrolysis; NMR, Nuclear Magnetic Resonance spectroscopy; OREC, Oxidation Resistant Elemental Carbon as measured by acid dichromate oxidation; fPyC, the fraction of total organic carbon that is pyrogenic carbon; MIR-PLSR, Mid InfraRed spectroscopy Partial Least Squares Regression; ROC, Resistant Organic Carbon; RMSE, Root Mean Square Error; RPD, Ratio of Performance to Deviation.

* Corresponding author.
*E-mail address:* michael.bird@jcu.edu.au (M.I. Bird).
[1] ORCID: 0000-0002-3215-1706.
[2] ORCID: 0000-0002-4956-6895.

https://doi.org/10.1016/j.geoderma.2025.117502
Received 24 March 2025; Received in revised form 27 August 2025; Accepted 1 September 2025
Available online 9 September 2025

the resistance to mineralization of (some) PyC in the soil has led to interest in the possibility of long-term carbon sequestration of manufacturing 'biochar' PyC for addition to agricultural soils (Lehmann, 2007; Lehmann et al., 2008).

Uncertainty in both the stocks and residence time of PyC in the soil translates directly into uncertainty in modelling the soil organic carbon pool (Bowring et al., 2022), especially when the existence of a PyC pool is not explicitly recognized. Lehmann et al. (2008) demonstrated that failing to account for slow cycling PyC in soil carbon models for Australia leads to an overestimate of modelled $CO_2$ emissions from a three-degree warming over 100 years by 18.3–24.4 %. Lavallee et al., (2019) found that including estimates of PyC stocks in models of TOC dynamics, changed turnover time estimates for non-PyC TOC in soil fractions by 1–21 %.

There have been numerous attempts to estimate the fraction of TOC that is PyC (fPyC). Forbes et al. (2006) estimated an fPyC of 1–35 % of TOC depending on vegetation and climate. Preston & Schmidt (2006) reviewed estimates of PyC in soils available at that time, noting a large range of values (fPyC < 1–60 %) due to natural variability associated with environmental and edaphic factors, but also due to the range of techniques used in quantification. Hockaday et al. (2007) concluded that fPyC was generally 5–15 %. More recently, Bird et al. (2015) estimated that ~7.7 % of the global soil carbon pool to 30 cm depth is PyC (~54 Pg), while Reisser et al. (2016) compiled 560 measurements from 55 studies and estimated ~200 Pg of PyC is sequestered in the upper 2 m of soil.

Many techniques – physical, chemical, thermal and spectroscopic – have been developed to quantify PyC in soil (see Bird, 2015 for a review). There have also been several efforts to both standardise individual techniques across laboratories and compare results for the same samples across multiple techniques (Schmidt et al., 2001; Hammes et al., 2007; Meredith et al., 2012; Cotrufo et al., 2016). These studies have tended to demonstrate a wide range of abundance measurements (up to orders of magnitude) on the same samples depending on the technique used. This is partly because different techniques target different ranges of the 'combustion' or 'pyrogenic carbon' continuum (Hedges et al., 2000; Bird et al., 2015). Some techniques quantify a broad range of pyrogenic components while some operationally define PyC as the carbon remaining after a chemical or thermal treatment under specified conditions. Reisser et al. (2016) compiled results obtained using six different techniques, these techniques having been shown to produce results that varied widely from each other in inter-comparison studies (Hammes et al., 2007). Techniques such as solid-state $^{13}C$ NMR are used to semi-quantify the proportion of pyrogenic-like material in soils but do not explicitly isolate PyC. Meanwhile, Fourier Transform mid-infrared spectroscopy (MIR), combined with multivariate chemometrics (e.g., partial least-squares regression – PLSR), uses measured analytical data, such as the NMR-derived resistant organic carbon (ROC) fraction, to predict the PyC-like fraction of soils. In the Australian context, the NMR approach, later combined with MIR-PLSR, has been used extensively to estimate the size of the ROC pool, a slow cycling pool of soil organic carbon that is assumed to be dominated by PyC (e.g. Skjemstad et al., 1996; Skjemstad et al., 1999; Janik et al., 2007; Baldock et al., 2013a,b).

There is clearly a need to benchmark commonly used techniques against each other using a range of natural samples to facilitate (i) direct comparison between studies and (ii) enable calibration of cheaper, faster techniques against analytically intensive and relatively costly techniques. Here we present a comparison of three techniques for pyrogenic carbon quantification in soils in Australia, that target different parts of the PyC continuum (i) hydrogen pyrolysis (HyPy, reported as $PyC_{SPAC}$), (ii) $^{13}C$ nuclear magnetic resonance ($PyC_{NMR}$), and (iii) dichromate oxidation ($PyC_{OREC}$). These techniques all have advantages and disadvantages in terms of cost and precision (Bird, 2015), and all use separate operational definitions for the portion of the PyC continuum they measure. All are in common use, so the need to be able to compare results across these techniques is urgent. We additionally utilize MIR spectroscopy on the bulk soil samples to both assess the ability of the rapid MIR spectral measurement to predict PyC estimated by these different methods and as a diagnostic tool to better understand the chemistry of the isolated PyC.

We hypothesize that $PyC_{SPAC}$ isolates a smaller PyC fraction than other techniques but one uniquely composed of PyC and so will not be directly related to TOC content, whereas $PyC_{NMR}$ quantifies a broader spectrum of PyC (but potentially also other TOC components). Further, we hypothesis that MIR calibration against a method that isolates only PyC, such as HyPy, should yield robust estimates for PyC and also fPyC since PyC is represented by a distinct subset of organic matter chemistry identifiable by MIR. OREC is measured by a technique that has been shown to not uniquely isolate PyC (Meredith et al., 2013) and so may, or may not, be well correlated to results obtained by the other techniques.

## 2. Materials and methods

### 2.1. Samples

The majority of samples used in this study have previously been used in studies of soil organic carbon encompassing Australian biomes from desert through tropical savanna, temperate and boreal biomes (Bird et al., 2004; Wynn et al., 2006). The samples were collected in the field using the stratified sampling protocol designed to encapsulate the variability in carbon stocks and carbon isotope composition in mixed tree-grass ecosystems. All were collected from relatively coarse textured soils and where particle size distribution was determined (Australian samples only), the percentage of particles finer than 53 μm ranged from 0.1 to 24.1 % (mean of 7.2 ± 5.8 % 1σ). Limiting the sample set to coarse textured soils minimises the potential additional confounding influence of soil texture on TOC stocks but necessarily limits the degree to which the results may be applicable to all soils. The reported dry bulk densities across all samples averaged 1.14 ($g/cm^3$) ± 0.37 (1σ). Where pH in $H_2O$ was recorded it ranged from 3.06 to 8.25 (mean = 5.59 ± 1.2 1σ). See Bird et al. (2004), Wynn et al. (2006) and Haig et al. (2025) for original data and detailed methods.

Four bulked samples represent each region, comprising 200 individual samples from each region collected in the field. The four samples are −5T and −30T (0–5 cm and 5–30 cm, collected at half canopy distance from individual trees) and −5G and −30G (collected at the midpoint between trees in open canopy environments such as tropical savannas). In closed forests, the −T and −G samples are equivalent. In all cases samples were dried, sieved and the <2 mm fraction was ball milled to a powder for all subsequent analyses. In the few cases where carbonate was present this was removed during pre-treatment by acidification and mass balance used to account for the loss of carbonate (Wynn and Bird, 2007). An aliquot of each sample was also sieved at 53 μm and the fine fraction also analysed to apportion PyC into fine and coarse fractions.

In total the initial sample set comprises 171 samples from 43 widely dispersed regions. Not all samples could be analysed by all techniques (particularly NMR; n = 65) due to low carbon content. A total of only 44 samples were analysed by all techniques and the statistical treatment below focuses on only those samples. Results for all samples are listed in Dataset S1 and their spatial distribution is shown in Fig. 1.

### 2.2. Analytical

#### 2.2.1. Hydrogen pyrolysis (HyPy)

The stable polycyclic aromatic carbon (SPAC, hereafter called '$PyC_{SPAC}$') component was separated from labile soil carbon by hydrogen pyrolysis. This is an established method (Meredith et al., 2012, 2017; Wurster et al., 2012, 2013; Haig et al., 2024, 2025) that yields the abundance of pyrogenic carbon that is resistant to degradation on centennial timescales (with >7 condensed aromatic rings), rather than all components of the pyrogenic carbon continuum, some of which are

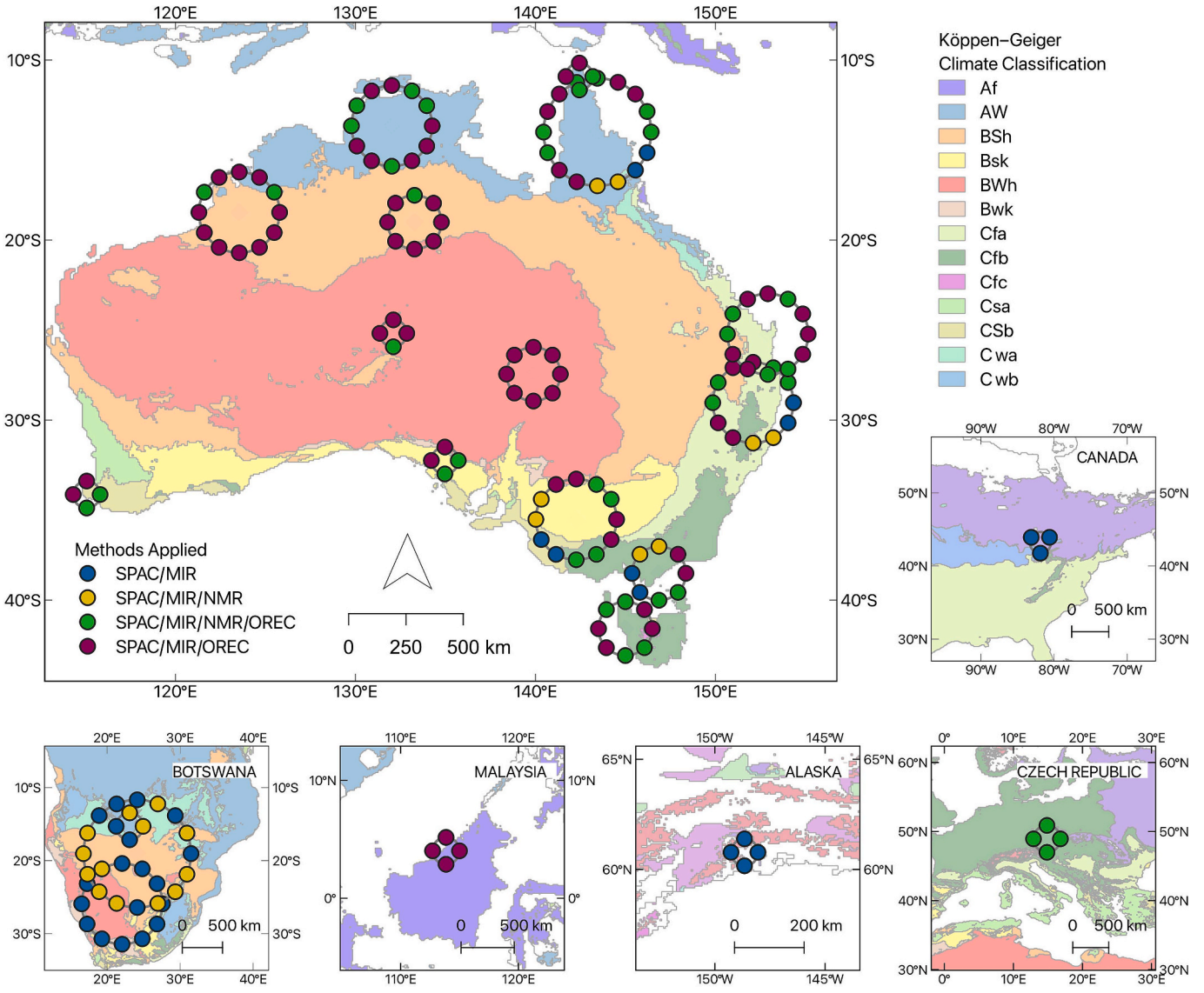


**Fig. 1.** Spatial distribution of the dataset indicating the PyC extraction/estimation methods applied to each sample. For simplicity, proximate samples (i.e. within a 400 km radius) are grouped graphically. Samples which received all treatments are shown in green; those with HyPy and MIR in blue; HyPy, MIR and NMR in yellow; and HyPy, MIR and OREC in purple. Köppen-Geiger Climate classification for coarse textured mineral soils is also shown. In this, the world's climate is classified based on 'main climate' (first letter: equatorial (A), arid (B), warm temperate (C), snow (D), polar (E); 'precipitation' (second letter: desert (W), steppe (S), fully humid (f), summer dry (s), winter dry (w), monsoonal (m); 'temperature' (last letter: hot summer (a), warm summer (b), cool summer (c), extremely continental (d), hot arid (h), cold arid (k), polar frost (F), polar tundra (T)). See Kottek et al. (2006) for more detail. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

also labile (Bird et al., 2015). Hypy has an advantage over some other techniques in that it physically separates PyC from other carbon in a sample that can then be analysed for stable isotope composition (e.g. Haig et al., 2025) and used for radiocarbon dating (e.g. Schiedung et al., 2024).

300 mg of each crushed sample was loaded at 10 % by weight with a Mo catalyst using an aqueous methanol (80:20) solution of ammonium dioxydithiomolybdate, sonicated and dried at 60 °C overnight. The sample/catalyst mixture was loaded into the reactor of the instrument, pressurised with $H_2$ to 150 bar with a purge gas flow of 5 L $min^{-1}$, heated at 300 °C $min^{-1}$ to 250 °C, then at 8 °C $min^{-1}$ to 550 °C, with the final temperature held for 5 min.

The carbon abundance of samples before and after HyPy were measured by elemental analyser isotope ratio mass spectrometry using a Costech Elemental Analyzer (Costech Analytical Technologies Inc., Valencia, CA, USA) fitted with a zero-blank autosampler coupled via a ConFloIV (Thermo Fisher Scientific, Waltham, MA, USA) to a ThermoFinnigan DeltaVPLUS using Continuous-Flow Isotope Ratio Mass Spectrometry at the Advanced Analytical Unit at James Cook University, Cairns. $PyC_{SPAC}$ abundance measurements have an average standard error of 2.2 % of the value, which we corrected for possible in situ production of stable polycyclic aromatic carbon during the hydrogen pyrolysis reaction (Wurster et al., 2012).

### 2.2.2. Nuclear magnetic resonance

Soil organic matter composition for 65 samples with sufficient TOC were characterised through solid-state $^{13}C$ cross-polarisation (CP) NMR with magic angle spinning (MAS). Samples were selected on the basis of having sufficient organic carbon content (≥2 %) that enabled for robust NMR spectral acquisition. Solid-state $^{13}C$ CPMAS NMR spectra were acquired using a 200 MHz Avance spectrometer (Bruker Corporation, Billerica, MA, USA) equipped with a 4.7 T, wide-bore superconducting magnet running at a $^{13}C$ resonance frequency of 50.33 MHz. Soils (200–600 mg) were spun at 5 kHz in uniformly packed 7-mm zirconia

rotors with Kel-F® end caps and inserts, used where required to ensure sample placement in the centre of the rotor. A standard 3.5 μs, 195 W and 90° pulse sequence with 1 ms contact time, and 1 s recycle delay CP NMR experiment was used for collection of 10,000–30,000 scans for all samples, the number of scans increasing in relation to decline of carbon in the rotor. Chemical shift values were calibrated to the methyl resonance at 17.36 ppm of hexamethylbenzene, and Lorentzian line broadening of 50 Hz was applied to all spectra. All soils were pretreated with hydrofluoric acid (2 %) prior to NMR analysis to remove paramagnetic material and improve spectral resolution (Skjemstad et al., 1996). Bruker TopSpin 3.5 software was used for processing all $^{13}C$ NMR spectra acquired. See Baldock et al. (2013b) for further details.

For estimates of broad carbon types, the spectra were integrated using the chemical shift limits defined by Baldock et al. (2013b): 45 – 0 ppm (alkyl C), 60 – 45 ppm (methoxyl and N-alkyl C), 90 – 60 ppm (O-alkyl C), 110 – 90 ppm (di-O-alkyl C), 145 – 110 ppm (aryl and unsaturated C), 165 – 145 ppm (O-aryl C), 190 – 165 ppm (carbonyl and amide C), and 215 – 190 ppm (ketone C). Glycine was used as an external standard for NMR observable OC (Baldock et al., 2013b; Baldock and Smernik, 2002). NMR estimation of the proportion of PyC-like materials was determined using a two-step model whereby the proportion of aryl-C and O----–aryl C measured for each sample were then corrected for: 1. the proportion of carbon estimated to be lignin; and 2. the fraction undetectable by CP analysis. The corrected proportions are defined as the ROC fraction which is assumed to be dominated by charcoal and charred plant residues. Supporting information and in-depth descriptions surrounding the calculations are detailed in Baldock et al. (2013b).

### 2.2.3. Dichromate oxidation

Dichromate oxidation is a commonly used technique for quantifying pyrogenic carbon as oxidation resistant elemental carbon, or OREC (Bird and Gröcke, 1997). Wynn and Bird (2007) report the results for the samples in this study, we report the methodology here because oxidation times and strengths vary between studies. An aliquot of ground soil was weighed into a 50 mL centrifuge tube filled with 0.1 M $K_2Cr_2O_7$ in a 2 M solution of $H_2SO_4$. The tubes were shaken for 72 h in the solution. Samples that consumed the oxidant before the full period was reached were replenished with fresh solution. The soil material remaining after oxidation was rinsed twice in deionized water with centrifugation between steps, dried at 105 °C and powdered for measurement of carbon abundance. 117 analyses of samples from 29 sample regions are available, summarized in Dataset S1.

### 2.2.4. Fourier-transform mid-infrared spectroscopy

Fourier-transform mid-infrared spectroscopy (mid-DRIFTS) was utilized as a relatively inexpensive, rapid method to simultaneously estimate TOC and PyC content of the soil samples (Baldock et al., 2013a; Sanderman et al., 2021a,b). Briefly, finely milled subsamples of each bulk soil were scanned on a Bruker Vertex 70 FTIR (Bruker Optics, Billerica, MA, USA) with a wide-range beamsplitter/detector combination and a Pike AutoDiff (Pike Technologies, Fitchburg, WI, USA) diffuse reflectance accessory. Sixty scans were collected and co-added for each sample and an Al mirror was used to collect the background single channel scan.

### 2.2.5. Statistical analyses

While 171 samples were analyzed by the different methods, only 44 of these samples had data for all three PyC methods (Fig. S1). We focused most of the statistical analyses on this smaller complete dataset but provide results for the full datasets in the supplement. Basic summary statistics were generated for TOC and all three PyC methods. Given the skewed nature of the data, TOC and PyC concentration data were normalized using natural log (1 + value), shortened hereafter to logp1 transformation following the name of the Python function in the NumPy package (Harris et al., 2020). The fraction of TOC that is PyC (fPyC = $PyC_{method}$/TOC) was also calculated for all three PyC methods. Pearson correlation coefficients were calculated between logp1 TOC and PyC concentration data as well as between logp1(TOC) and fPyC data for all three PyC methods.

Partial least squares regression (PLSR, Geladi and Kowalski, 1986) models were used to understand the relationship between the MIR spectra and the different PyC methods both to better understand what each method is isolating and to see if MIR can be used as a rapid estimation method for inferring PyC content. PLSR models work by trying to find orthogonal factors that simultaneously maximize the explained in the X (spectra) and Y (TOC or PyC or fPyC) data. PLSR models were built to predict logp1 transformed TOC and PyC concentration data using 10-fold cross validation with standard normal variate (SNV) pre-processing applied to the 4000–628 $cm^{-1}$ spectral range.

PLSR models were also built using the same 10-fold cross validation approach as above to predict fPyC data using both the MIR and NMR data as predictors. The solid-state $^{13}C$ NMR data was included in this analysis because this analytical method only quantifies organic matter composition whereas MIR spectroscopy simultaneously measures most mineral and organic constituents in the soil, often with a lot of overlap between features. Model goodness-of-fit was assessed using root mean square error (RMSE), coefficient of determination ($R^2$), and ratio of performance to deviation (RPD):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad (1)$$

where x and y are the measured and predicted values, respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (2)$$

where x and y are the measured and predicted values, respectively, and $\bar{x}$ is the mean of the measured values.

$$RPD = \frac{s_x}{RMSE} \quad (3)$$

where $s_x$ is the standard deviation of measured values and RMSE is the root mean square error.

Loading weights and variable importance in projection (VIP) scores were used to understand what parts of the MIR and NMR spectra were responsible for predicting PyC concentration and fPyC. In PLSR, the loading weights, in soil spectroscopy sometimes refered to as the loading spectra, are the weights of the MIR or NMR spectra at each wavenumber (or chemical shift region for NMR) that maximizes the explained variance in both the predictor (e.g. spectra) and response variable (e.g. PyC or fPyC). The VIP score, which quantifies the importance of each predictor variable in explaining the response variable, for the $j^{th}$ variable is defined following Farrés et al. (2015) as:

$$VIP_j = \sqrt{\frac{\sum_{f=1}^{F}(w_{jf}^2 \times SSY_f \times J)}{SSY_{total} \times F}} \quad (4)$$

where $w_{jf}$ is the weight value for j variable and f component, $SSY_f$ is the sum of squares of explained variance for the $f^{th}$ component and J number of X variables. $SSY_{total}$ is the total sum of squares explained by the dependent variable, and F is the total number of components.

All analyses were performed in Python using the sckit-learn (Pedregosa et al., 2011), pandas (McKinney, 2010), NumPy (Harris et al., 2020), seaborn (Waskom, 2021) and Matplotlib (Hunter, 2007) packages.

## 3. Results

### 3.1. TOC and PyC data for samples analysed by all methods

The common set of 44 samples that were analyzed by all PyC methods contained a mean TOC content of 3.24 % (Interquartile Range; IQR = 1.05–3.36 %). For these samples, $PyC_{SPAC}$ content was lowest with a mean value of 0.12 % (IQR = 0.07–0.16 %), while $PyC_{NMR}$ content (mean = 0.77 %, IQR = 0.32–0.89 %) and $PyC_{OREC}$ content (mean = 0.52 %, IQR = 0.10–0.55 %) were 4–6 times higher. All PyC methods were well correlated with TOC content with Pearson's correlation of 0.96, 0.80 and 0.79 for $PyC_{NMR}$, $PyC_{SPAC}$ and $PyC_{OREC}$ methods, respectively (Fig. 2). All PyC methods were also correlated with each other, although the strength of these correlations varied with $PyC_{OREC}$ being least similar to the other two methods (Fig. 2). Examples of the raw NMR spectra high PyC/low TOC and low PyC/High TOC samples are provided in Fig. S2.

Examining the ratio of PyC to TOC – the fraction of PyC (fPyC) – for each method can remove the overall strong influence of TOC content over PyC content and potentially better highlight differences between methods. There remained a strong correlation between $fPyC_{SPAC}$ and $fPyC_{NMR}$ (r = 0.63) but much weaker correlations between $fPyC_{OREC}$ and the two other measures (Fig. 3). Interestingly, there was a moderately strong non-linear negative correlation between TOC and $fPyC_{SPAC}$ (r = −0.73) but only weak correlations with TOC for the other two methods (Fig. 3).

### 3.2. PyC regression analysis

Given the demonstrated ability for MIR spectroscopy on bulk soil samples to estimate TOC and PyC concentration in other studies (Baldock et al., 2013a; Cotrufo et al., 2016; Sanderman et al., 2021a), we employed it here using PLSR models as a diagnostic tool to better understand the underlying PyC methods. Examples of the raw MIR spectra high PyC, low TOC and low PyC high TOC samples are provided in Fig. S2. All models had an $R^2$ of at least 0.70 but RPD values suggest only moderate predictive ability in predicting $PyC_{OREC}$, good predictive ability for $PyC_{SPAC}$ and excellent predictive ability for $PyC_{NMR}$ and TOC

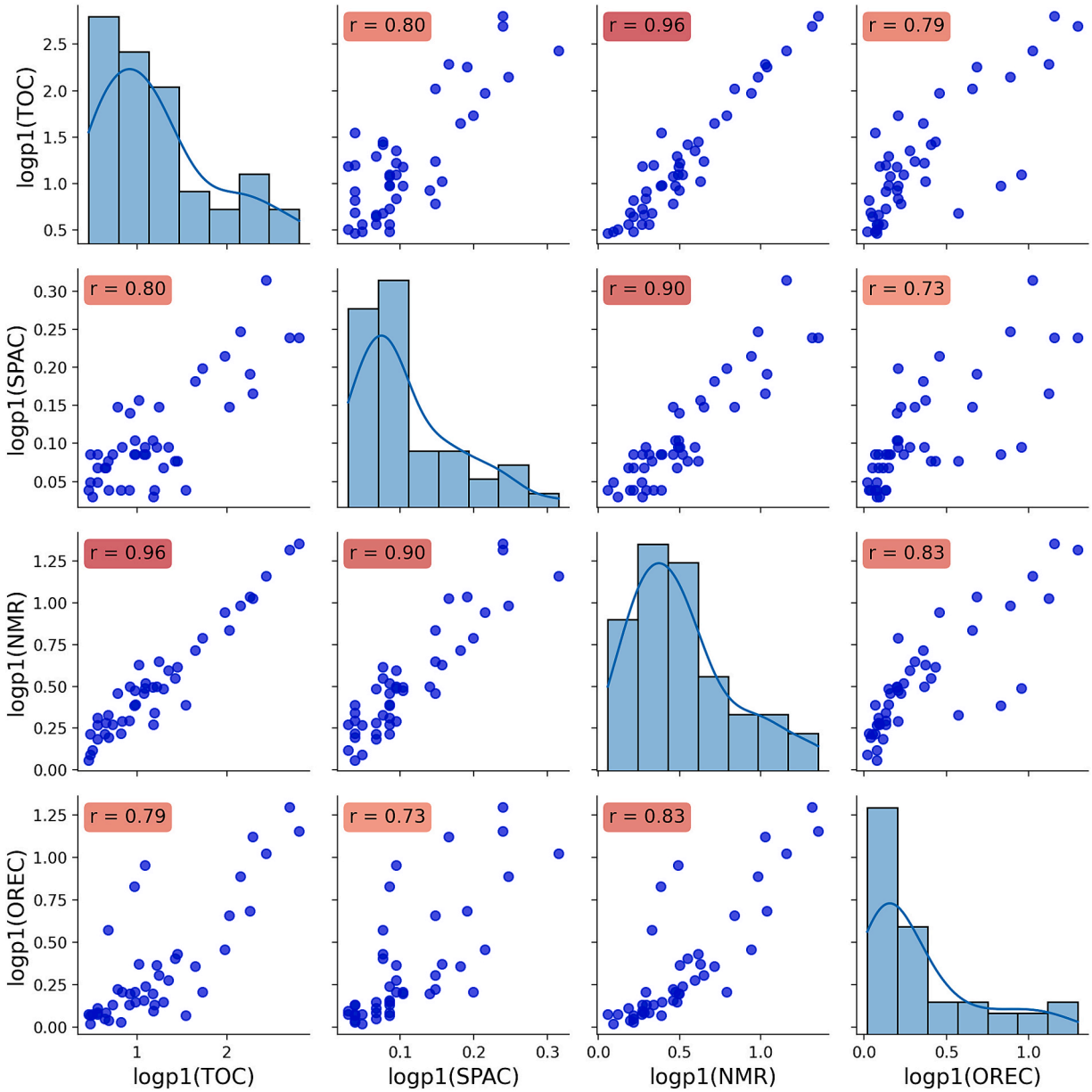


**Fig. 2.** Correlations between logp1-transformed TOC content and PyC content measured by SPAC, NMR and OREC methods.

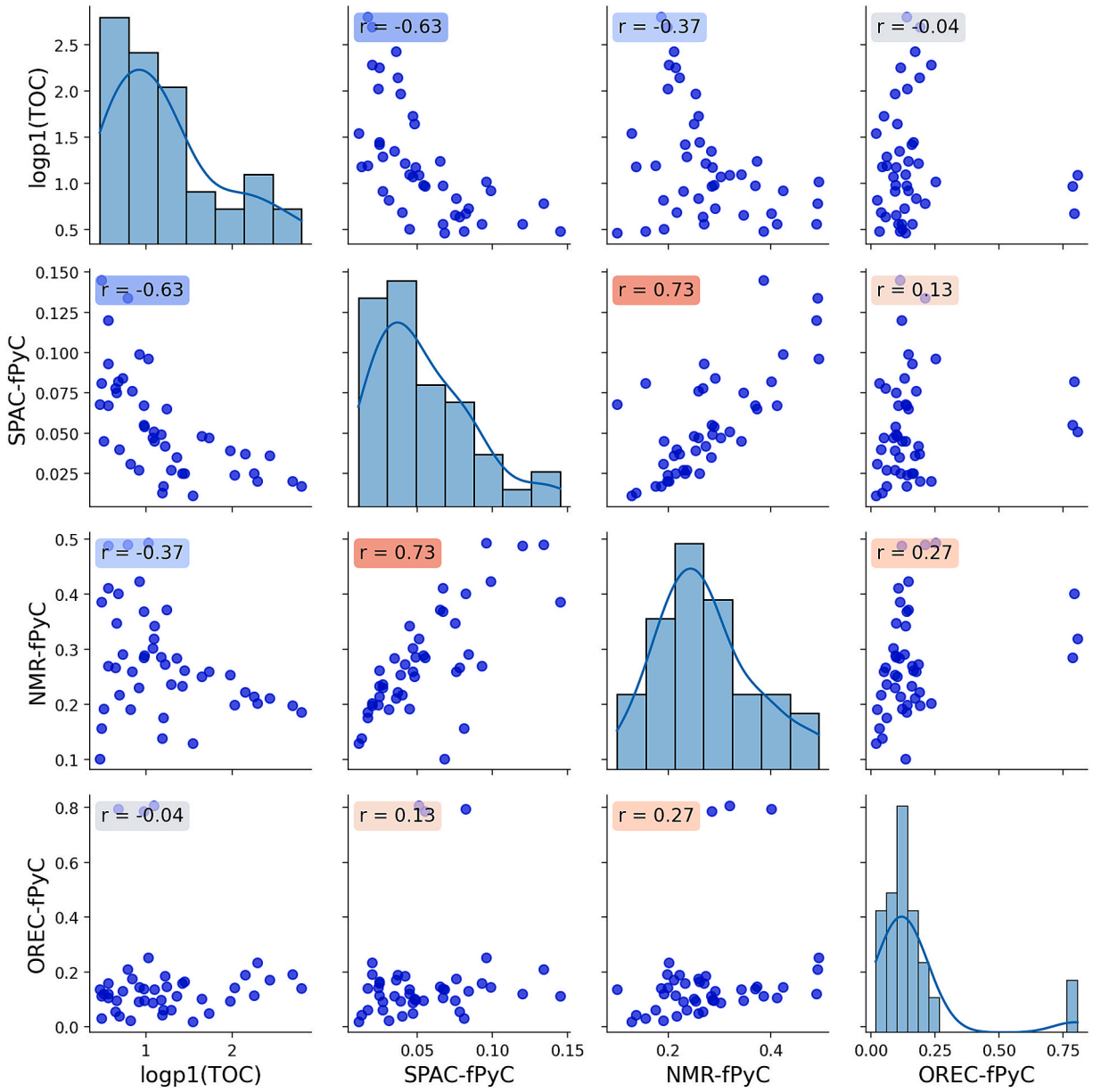


**Fig. 3.** Correlations between logp1-transformed TOC content and fraction TOC that is PyC (fPyC) for SPAC, NMR and OREC-based PyC estimation methods.

**Table 1**
Performance of partial least squares regression models to predict TOC and each PyC method from the MIR spectra of the bulk soils.

| Response variable | n | Factors | $R^2$ | RMSE | RPD |
|---|---|---|---|---|---|
| logp1(TOC) | 44 | 4 | 0.97 | 0.105 | 5.99 |
| logp1($PyC_{SPAC}$) | 44 | 8 | 0.77 | 0.033 | 2.07 |
| logp1($PyC_{NMR}$) | 44 | 6 | 0.94 | 0.075 | 4.31 |
| logp1($PyC_{OREC}$) | 44 | 5 | 0.70 | 0.187 | 1.86 |

(Table 1). Results were consistent for the full data set for each method with all models converging on the same number of factors (Table S1).

Given the strong correlations between TOC content and PyC content (Fig. 2), we further explored the ability of a simple linear regression with TOC to predict PyC content using the same cross-validation approach. The ability to predict $PyC_{NMR}$ from just TOC content was nearly as good as using the full MIR spectrum but slightly lower for $PyC_{OREC}$ and much lower for $PyC_{SPAC}$ (Table S2).

Plots of important variables, defined by loading weights and VIP scores, in the PLSR models can reveal which regions in the MIR spectrum are contributing to the ability to predict TOC and the different forms of PyC (Fig. 4). While the loading weights for the first factor, which explained the majority of total variance explained for each model, were nearly identical in all four models (Fig. 4a), the VIP scores varied more between models (Fig. 4b). The VIP scores were nearly identical between TOC and $PyC_{NMR}$ ($r = 0.97$), very similar between TOC and $PyC_{OREC}$ ($r = 0.86$) and moderately similar between TOC and $PyC_{SPAC}$ ($r = 0.74$). The largest dissimilarities between models were found in regions associated with aromatic compounds (Fig. 4b).

### 3.3. *fPyC regression analysis*

We also explored the ability to predict fPyC as a function of MIR and NMR spectra. Examples of the raw MIR spectra high fPyC and low fPyC samples are provided in Fig. S2. There was no ability to predict $fPyC_{OREC}$ using either MIR or NMR spectra (Table 2). The $fPyC_{NMR}$ estimates were poorly predicted from MIR spectra ($R^2 = 0.18$ and $RPD = 1.11$) but well predicted using NMR data. This second finding was to be expected since

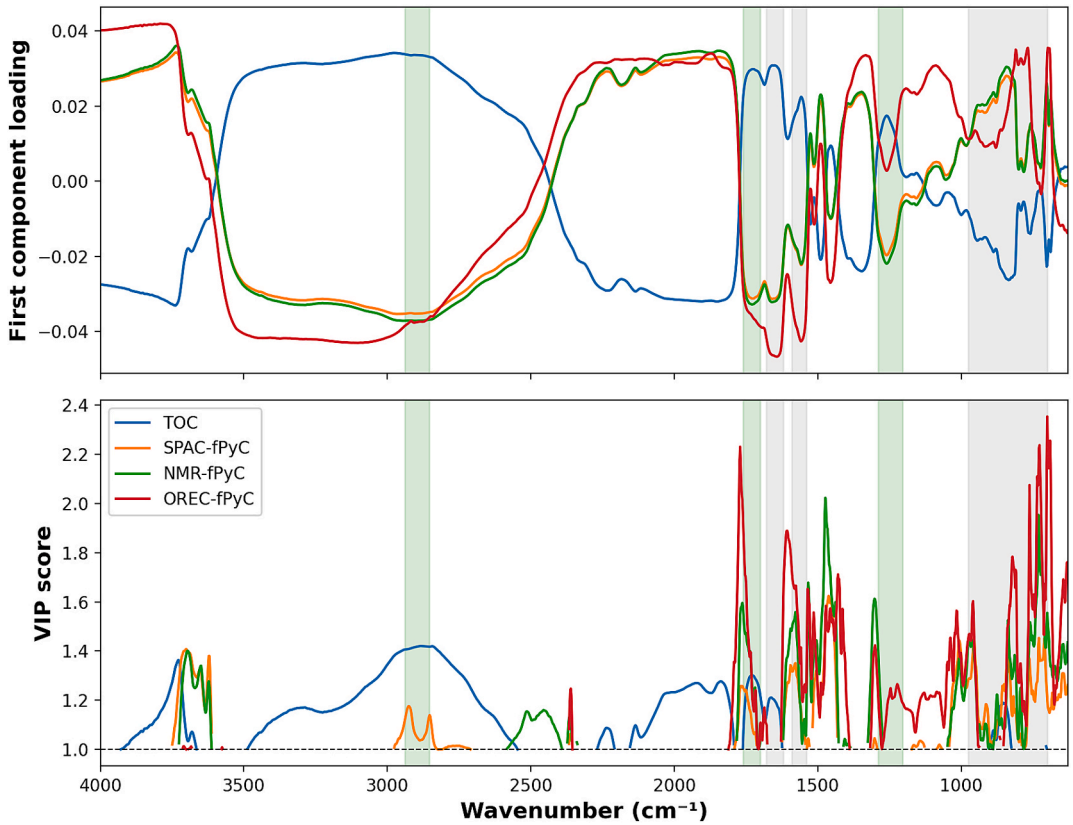


**Fig. 4.** (a) Loading weights for the first component and (b) variable importance of projection (VIP) scores for partial least squares regression models for logp1 of TOC, $PyC_{SPAC}$, $PyC_{NMR}$ and $PyC_{OREC}$. VIP Scores only shown when >1 meaning they are more important than average for a particular model. Green shading represents major non-aromatic organic carbon regions (2940–2855: aliphatic C–H stretching; 1760–1700: carbonyl C═O stretching, and 1290–1205: carboxylic C–O stretching) and grey shading represents some important aromatic regions (1680–1620: C═C stretching; 1590–1540: aromatic C–H deformation and C═C stretching; and 975–700: aromatic C–H bending). Peak assignments from Parikh et al. (2014). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**
Partial least squares regression results for predicting the fraction of pyrogenic carbon (fPyC) using MIR and NMR spectra predictors. A common set of 44 samples were utilized in all models using a cross-validation approach for estimating model performance.

| Response variable | Predictor | Factors | $R^2$ | RMSE | RPD |
|---|---|---|---|---|---|
| $fPyC_{SPAC}$ | MIR | 13 | 0.65 | 0.020 | 1.62 |
| $fPyC_{SPAC}$ | NMR | 5 | 0.71 | 0.018 | 1.83 |
| $fPyC_{NMR}$ | MIR | 4 | 0.18 | 0.086 | 1.11 |
| $fPyC_{NMR}$ | NMR | 4 | 0.87 | 0.036 | 2.66 |
| $fPyC_{OREC}$ | MIR | 1 | −0.09 | 0.190 | 0.95 |
| $fPyC_{OREC}$ | NMR | 1 | 0.06 | 0.176 | 1.02 |

the $fPyC_{NMR}$ is derived directly from the NMR data itself. The $fPyC_{SPAC}$ estimates were predicted moderately well using both MIR and NMR data ($R^2$ = 0.65 and 0.71, respectively).

Despite the varying success in the NMR data for predicting the different estimates of fPyC (Table 2), the shape of the loading weights for the first component of the PLSR models were nearly identical (Fig. 5a). This suggests that there are some commonalities in the chemical attributes that each PyC method is isolating. In Fig. 5a, it is clear that only the aryl-C (145 – 110 ppm) region is positively correlated with increasing fPyC and this region has the largest VIP scores for all three fPyC models (Fig. 5b). The remaining chemical shift regions generally have near neutral or strong negative loadings. The O-alkyl-C (90 – 60 ppm) region shows up as important (Fig. 5b) but with a negative loading indicating that when O-alkyl-C dominates the OM chemistry there is little PyC in the sample. Given the poor fits of the MIR model for predicting fPyC for the NMR and OREC methods (Table 2), the loadings and VIP scores for those models should be interpreted with caution.

## 4. Discussion

In this study we take the measured $PyC_{SPAC}$ as the most reliable indicator of the abundance of PyC that is likely to be resistant to decomposition as the benchmark to assess PyC content by other techniques. This is an assumption but is reasonable because hydrogen pyrolysis has been shown to directly isolate and measure only a very specific and recalcitrant fraction of PyC (Ascough et al., 2010; Meredith et al., 2012, 2017; Lavallee et al., 2019; Schiedung et al., 2024). We acknowledge that there are other techniques that access a suitably specific analytical window that would enable benchmarking that we could not apply in this study.

Fig. 2 shows that while there is excellent correlation between $PyC_{NMR}$ and the measured $PyC_{SPAC}$ content (r = 0.90), the NMR technique over-estimates PyC in the samples by a factor of ~7 on average compared to the Hydrogen pyrolysis technique. This means that, on average, only ~15 % of the carbon identified by $PyC_{NMR}$ is the same slow cycling PyC component that is quantified as $PyC_{SPAC}$ by hydrogen pyrolysis. Part of this difference is likely due to $PyC_{NMR}$ measuring carbon that is of pyrogenic origin but is not captured in the analytical window of hydrogen pyrolysis. However, it is also possible that the overestimation of $PyC_{NMR}$, in comparison to $PyC_{SPAC}$ is the result of the NMR technique including other aromatic compounds, for example lignin moieties that are potentially refractory in nature but are not necessarily of pyrogenic origin (e.g. Hammes et al., 2008).

Fig. 3 indicates there is a negative relationship between TOC and

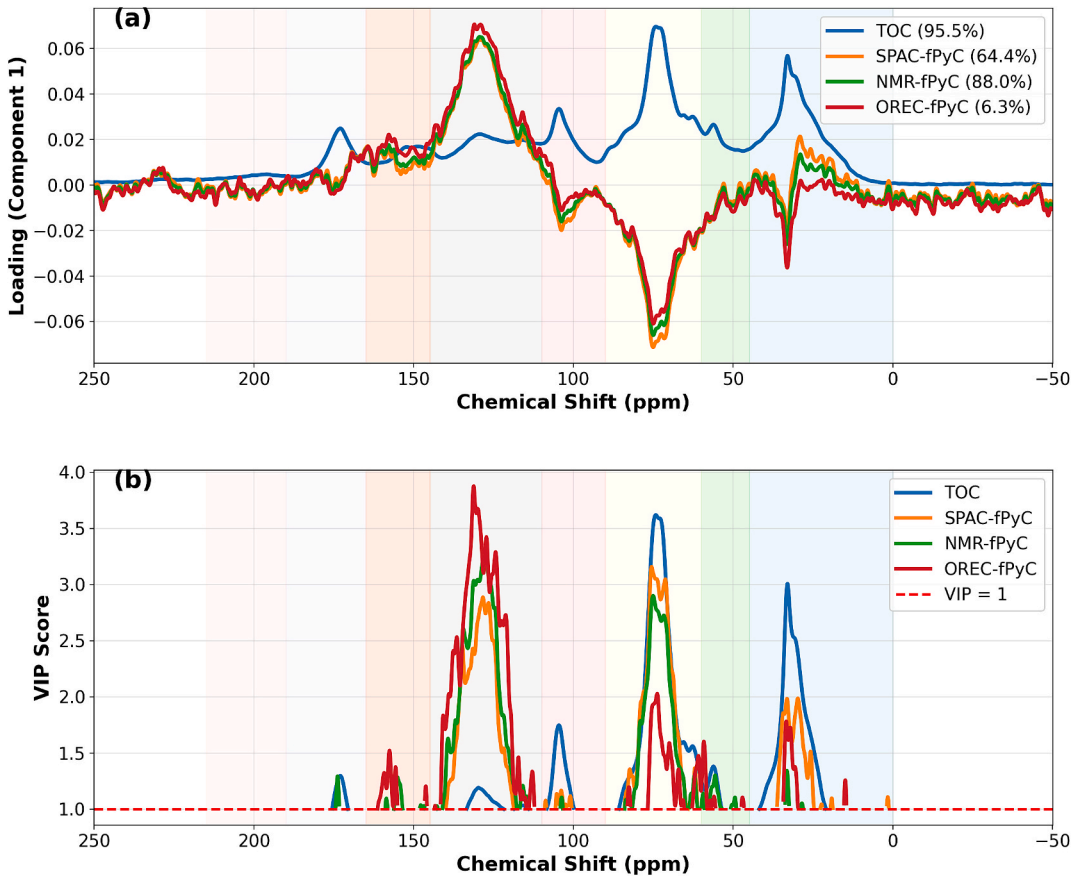


**Fig. 5.** Loading weights for the first component (a) and variable importance in projection (VIP) scores for partial least squares regression models built using NMR to predict fPyC based on SPAC, NMR and OREC methods. Model for logp1(TOC) also shown for comparison. Values in parentheses are the validation Y variance explained by this first component. Shading indicates major chemical shift regions from right to left: 45 – 0 ppm (alkyl C), 60 – 45 ppm (methoxyl and N-alkyl C), 90 – 60 ppm (O-alkyl C), 110 – 90 ppm (di-O-alkyl C), 145 – 110 ppm (aryl and unsaturated C), 165 – 145 ppm (O-aryl C), 190 – 165 ppm (carbonyl and amide C), and 215 – 190 ppm (ketone C).

$fPyC_{SPAC}$ in ($r = -0.63$), where $fPyC_{SPAC}$ is higher in lower TOC soils and decreases in high TOC soils. This relationship occurs because, soil type being similar, TOC preservation in the soil is favoured by high rainfall and/or low temperatures (Stockmann et al., 2013), whereas the conditions favouring PyC production are strongly seasonal precipitation, and seasonally high temperatures (Haig et al., 2025). Thus the drivers of TOC and PyC abundance are related but partly decoupled. This expected negative relationship is evident, but considerably weaker ($r = -0.37$) for $fPyC_{NMR}$.

Fig. 2 also compares the $PyC_{OREC}$ and $PyC_{SPAC}$ values. These are both direct measurements of the abundance of carbon left in a sample after chemical treatment, in both cases assumed to be PyC after the removal of 'non-PyC' from the sample. The comparison suggests that at low TOC, for some samples, there is reasonable agreement between two techniques. As TOC increases, and even for some samples with relatively low TOC, $PyC_{OREC}$ overestimates the PyC content of the samples by an order of magnitude or more. In one extreme case $PyC_{OREC}$ is estimated at ~1 % when hydrogen pyrolysis removes essentially all carbon, indicating there is very little carbon of pyrogenic origin in the sample. This suggests that while it is certainly the case that the two techniques are isolating different windows with the pyrogenic carbon continuum, dichromate oxidation removes TOC to a variable degree, and this becomes increasingly problematic as TOC increases. This conclusion is evident in the absence of any relationship between $fPyC_{SPAC}$ and $fPyC_{OREC}$ ($r = -0.13$) seen in Fig. 3 and the inability of the MIR and NMR spectra to predict $fPyC_{OREC}$ (Table 2). It has previously been shown that dichromate oxidation overestimates PyC relative to other techniques (Hammes et al., 2007), partly due to the inability of the aqueous dichromate solution to access and oxidise hydrophobic organic compounds (Meredith et al., 2013) and also because and dichromate will react with other interferences including reduced iron, manganese and chloride (Schumacher, 2002).

MIR can be used to predict TOC and PyC concentrations estimated by all PyC methods (Tables 1 and S1). The wavenumber region ascribed to carbonyl C═O stretching (1760–1700 $cm^{-1}$) is important to all predictions. The regions of importance to the OREC predictions are very similar to those driving the TOC predictions, with subtle differences. The regions driving the NMR predictions are also broadly similar to those driving TOC predictions, but also much more similar to the regions of importance to the $PyC_{SPAC}$ predictions than the $PyC_{OREC}$ predictions. The regions driving the $PyC_{SPAC}$ predictions are the most distinct from the TOC model, driven primarily by wave numbers associated with carbonyl C═O stretching (1760–1700 $cm^{-1}$), and with no importance ascribed to carboxylic acid C–O stretching (1280–1200 $cm^{-1}$). Cotrufo et al. (2016) used MIR-PLSR to predict PyC abundances in a range of matrices from a burnt watershed, measured by the SPAC and the Benzene Carboxylic Acid (BPCA) techniques. While Cotrufo et al. (2016) found that BPCA-estimated PyC content was five times lower than $PyC_{SPAC}$ estimates, only small differences in the loading spectra of PLSR models built to predict these two forms of PyC were found.

Both the NMR and MIR spectra were used to predict the fraction of PyC to TOC in the samples (fPyC). This approach has the advantage of assessing carbon quality rather than quantity by removing the overall influence of the correlation between TOC and PyC. While the MIR-based models were less informative, the NMR-based models – remembering that NMR is a direct measurement of OM chemistry – indicated that all

PyC techniques have nearly identical loading spectra with positive loadings for the Aryl-C region and negative loadings for the rest of the chemical shift regions (Fig. 5). The quality of the models, indicated by goodness-of-fit metrics listed in Table 2, suggest that while similar OM chemistries were driving the prediction of fPyC, only $fPyC_{SPAC}$ could be well predicted by these patterns suggesting that this technique is in fact isolating a chemically consistent PyC fraction. The $fPyC_{OREC}$ data could not be predicted at all from the NMR spectra suggesting this methodology is not isolating a consistent fraction.

The fact that the hydrogen pyrolysis techique is likely isolating a more pure fraction of PyC, as suggested by the limited features in the VIP scores plot (Figs. 4b and 5b) may also be the reason why the MIR-based PLSR models underperformed in predicting $PyC_{SPAC}$ compared to $PyC_{NMR}$ ($R^2$ = 0.77 v. 0.94, respectively). Essentially, there is less information directly related to $PyC_{SPAC}$ than to $PyC_{NMR}$. While the model may have more uncertainty, which may or may not be able to overcome with more samples (Ng et al., 2020) or smarter algorithms (Padarian et al., 2019), unlike the $PyC_{NMR}$ models the $PyC_{SPAC}$ model is predicting a fraction of TOC that is distinguishable from the bulk TOC model.

It has long been known that the techniques available for the quantification of PyC yield highly variable results, even when applied to the same samples (Hammes et al., 2008). This is due in part to the fact that different techniques quantify different analytical windows along the pyrogenic carbon continuum (Bird, 2015). It is also due in part to some techniques not being able to quantify carbon of uniquely pyrogenic origin (Hammes et al., 2008). This is particularly a problem in mineral soils where there can be hundreds to thousands of different C-containing compounds (Tfaily et al., 2015) that interact to varying degrees with charged mineral surfaces (Kleber et al., 2021; Newcomb et al., 2017).

Even when soil particle size in the sample set is constrained to some degree, as in this study, dichromate oxidation (OREC technique) performs poorly in many cases, particularly when TOC is high, as oxidation of the non-PyC organic component is likely incomplete. This does not mean that dichromate oxidation is not a useful technique. In our study, we found fairly strong linear relationships between $PyC_{OREC}$ and the two other methods (Fig. 2). In cases where TOC is low and relatively invariant, and the sedimentary/soil matrix is also similar, the OREC technique can provide a useful indication of variability in PyC content across a landscape and thereby enable an understanding of the drivers of that variability. However, in general it will be difficult to reliably integrate $PyC_{OREC}$ measurements into larger scale regional or global compilations because the degree of overestimation of PyC is highly variable but can be up to an order of magnitude in comparison to the $PyC_{SPAC}$ measurements we report here (Fig. 2).

We do observe a good correlation between $PyC_{NMR}$ and the $PyC_{SPAC}$ in the same samples but with $PyC_{NMR}$ estimating significantly higher abundances than $PyC_{SPAC}$. Part of this difference in PyC quantity between SPAC and NMR is due to the NMR technique likely including a wider range of likely more degradable pyrogenic carbon in the analytical window (as the “ROC pool”), but it also seems that aromatized soil organic matter is also measured. This can at least be partially observed by examining the VIP scores from the PLSR models (Figs. 4b and 5b) where there are less features present in the SPAC estimates of PyC and fPyC than the NMR-based estimates. To a first order at least, it is possible to estimate that $PyC_{SPAC}$ is approximately 15 % of $PyC_{NMR}$ suggesting that data from these techniques can be compared in regional or global studies. Combined with the study of Cotrufo et al. (2016), who compared SPAC with BPCA estimates of PyC abundance, it seems that all three techniques can be quantitatively benchmarked against each other.

## 5. Conclusions

The data presented here suggests that MIR-PLSR can be used to predict both $PyC_{NMR}$ and $PyC_{SPAC}$ abundances. Cotrufo et al. (2016) has demonstrated that the same approach can be used to predict PyC abundance determined by the BPCA technique. In combination then, we conclude that MIR-PLSR might ultimately represent a cost-effective technique that can be used to infer both the ‘slow cycling’ ROC pool, and an ‘inert’ PyC pool within the ROC pool, estimated by HyPy, or by BPCA. The results of this study will be of use in attempting to compile an updated global budget for soil PyC (e.g. Reisser et al., 2016) and also for modelling studies that rely on the identification of discrete soil organic carbon pools with different turnover times (e.g. Lehmann et al., 2008; Abramoff et al., 2022).

## CRediT authorship contribution statement

**Jonathan Sanderman:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Jordahna Haig:** Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Sourav Das:** Writing – review & editing, Validation, Investigation, Funding acquisition, Conceptualization. **Colleen Partida:** Writing – review & editing, Methodology, Investigation. **Christina Asanopoulos:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis. **Michael I. Bird:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

Funding: This work was funded by the Australian Research Council (grant number DP210100881).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2025.117502.

## Data availability

Add data is provided in the Supplementary information.

## References

Abramoff, R.Z., Guenet, B., Zhang, H., Georgiou, K., Xu, X., Rossel, R.A.V., Yuan, W., Ciais, P., 2022. Improved global-scale predictions of soil carbon stocks with Millennial Version 2. Soil Biol. Biochem. 164, 108466.
Ascough, P.L., Bird, M.I., Meredith, W., Wood, R.E., Snape, C.E., Brock, F., Higham, T.F. G., Large, D.J., Apperley, D.C., 2010. Hydropyrolysis: Implications for radiocarbon pretreatment and characterization of black carbon. Radiocarbon 52, 1336–1350.
Baldock, J.A., Hawke, B., Sanderman, J., Macdonald, L.M., 2013a. Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. Soil Res. 51 (8), 577–595.
Baldock, J.A., Sanderman, J., Macdonald, L.M., Puccini, A., Hawke, B., Szarvas, S., McGowan, J., 2013b. Quantifying the allocation of soil organic carbon to biologically significant fractions. Soil Res. 51, 561–576.
Bowring, S.P., Jones, M.W., Ciais, P., Guenet, B., Abiven, S., 2022. Pyrogenic carbon decomposition critical to resolving fire’s role in the Earth system. Nat. Geosci. 15 (2), 135–142.
Bird, M.I., Gröcke, D.R., 1997. Determination of the abundance and carbon isotope composition of elemental carbon in sediments. Geochim. Cosmochim. Acta 61 (16), 3413–3423.
Bird, M.I., Veenendaal, E.M., Lloyd, J.J., 2004. Soil carbon inventories and $\delta^{13}C$ along a moisture gradient in Botswana. Glob. Chang. Biol. 10 (3), 342–349.
Bird, M.I., 2015. Chapter 25: test procedures for biochar analysis in soils. In: Biochar for Environmental Management. Routledge, pp. 677–714.
Bird, M.I., Wynn, J.G., Saiz, G., Wurster, C.M., McBeath, A., 2015. The pyrogenic carbon cycle. Annu. Rev. Earth Planet. Sci. 43, 273–298.

Cotrufo, M.F., Boot, C., Abiven, S., Foster, E.J., Haddix, M., Reisser, M., Wurster, C.M., Bird, M.I., Schmidt, M.W., 2016. Quantification of pyrogenic carbon in the environment: an integration of analytical approaches. Org Geochem. 100, 42–50.
Farrés, M., Platikanov, S., Tsakovski, S., Tauler, R., 2015. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. J. Chemom. 29 (10), 528–536.
Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1–17.
Haig, J., Sanderman, J., Zwart, C., Smith, C., Bird, M.I., 2024. Impact of fire return interval on pyrogenic carbon stocks in a tropical savanna, North Queensland, Australia. Int. J. Wildland Fire 33 (8).
Haig, J., Das, S., Sanderman, J., Bird, M.I., 2025. Rainfall and seasonality drive pyrogenic carbon stocks in coarse-textured mineral soils. Global Biogeochem. Cycles 39 (2), e2024GB008240.
Hammes, K., Schmidt, M.W., Smernik, R.J., Currie, L.A., Ball, W.P., Nguyen, T.H., Louchouarn, P., Houel, S., Gustafsson, Ö., Elmquist, M., Cornelissen, G., 2007. Comparison of quantification methods to measure fire-derived (black/elemental) carbon in soils and sediments using reference materials from soil, water, sediment & the atmosphere. Global Biogeochem. Cycles 21 (3), GB3016.
Hammes, K., Smernik, R.J., Skjemstad, J.O., Schmidt, M.W., 2008. Characterisation and evaluation of reference materials for black carbon analysis using elemental composition, colour, BET surface area and $^{13}C$ NMR spectroscopy. Appl. Geochem. 23 (8), 2113–2122.
Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., 2020. Array programming with NumPy. Nature 585 (7825), 357–362.
Hedges, J.I., Eglinton, G., Hatcher, P.G., Kirchman, D.L., Arnosti, C., Derenne, S., Evershed, R.P., Kögel-Knabner, I., de Leeuw, J.W., Littke, R., Michaelis, W., 2000. The molecularly-uncharacterized component of nonliving organic matter in natural environments. Org Geochem. 31 (10), 945–958.
Hunter, J.D., 2007. Matplotlib: a 2D graphics environment. Comput. Sci. Eng. 9 (03), 90–95.
Janik, L.J., Skjemstad, J.O., Shepherd, K.D., Spouncer, L.R., 2007. The prediction of soil carbon fractions using mid-infrared-partial least square analysis. Soil Res. 45 (2), 73–81.
Kleber, M., Bourg, I.C., Coward, E.K., Hansel, C.M., Myneni, S.C., Nunan, N., 2021. Dynamic interactions at the mineral–organic matter interface. Nat. Rev. Earth Environ. 2 (6), 402–421.
Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15, 259–263.
Lavallee, J.M., Conant, R.T., Haddix, M.L., Follett, R.F., Bird, M.I., Paul, E.A., 2019. Selective preservation of pyrogenic carbon across soil organic matter fractions and its influence on calculations of carbon mean residence times. Geoderma 354, 113866.
Lehmann, J., 2007. A handful of carbon. Nature 447 (7141), 143–144.
Lehmann, J., Skjemstad, J., Sohi, S., Carter, J., Barson, M., Falloon, P., Coleman, K., Woodbury, P., Krull, E., 2008. Australian climate–carbon cycle feedback reduced by soil black carbon. Nat. Geosci. 1 (12), 832–835.
McKinney, W., 2010. Data structures for statistical computing in Python. SciPy 445 (1), 51–56.
Meredith, W., Ascough, P.L., Bird, M.I., Large, D.J., Snape, C.E., Sun, Y., Tilston, E.L., 2012. Assessment of hydropyrolysis as a method for the quantification of black carbon using standard reference materials. Geochim. Cosmochim. Acta 97, 131–147.
Meredith, W., Ascough, P.L., Bird, M.I., Large, D.J., Snape, C.E., Song, J., Sun, Y., Tilston, E.L., 2013. Direct evidence from hydropyrolysis for the retention of long alkyl moieties in black carbon fractions isolated by acidified dichromate oxidation. J. Anal. Appl. Pyrol. 103, 232–239.
Meredith, W., McBeath, A., Ascough, P., Bird, M.I., 2017. Chapter 17 – analysis of biochars by hydropyrolysis. In: Singh, B., Camps-Arbestain, M., Lehmann, J. (Eds.), Biochar: A Guide to Analytical Methods. CRC Press, pp. 187–198.
Newcomb, C.J., Qafoku, N.P., Grate, J.W., Bailey, V.L., De Yoreo, J.J., 2017. Developing a molecular picture of soil organic matter–mineral interactions by quantifying organo–mineral binding. Nat. Commun. 8 (1), 396.
Ng, W., Minasny, B., Mendes, W.D.S., Demattê, J.A.M., 2020. The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data. Soil 6 (2), 565–578.
Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning to predict soil properties from regional spectral data. Geoderma Reg. 16, e00198.
Parikh, S.J., Goyne, K.W., Margenot, A.J., Mukome, F.N., Calderón, F.J., 2014. Soil chemical insights provided through vibrational spectroscopy. Adv. Agron. 126, 1–148.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Reisser, M., Purves, R.S., Schmidt, M.W., Abiven, S., 2016. Pyrogenic carbon in soils: a literature-based inventory and a global estimation of its content in soil organic carbon and stocks. Front. Earth Sci. 4, 80.
Sanderman, J., Baldock, J.A., Dangal, S.R., Ludwig, S., Potter, S., Rivard, C., Savage, K., 2021a. Soil organic carbon fractions in the Great Plains of the United States: an application of mid-infrared spectroscopy. Biogeochemistry 156 (1), 97–114.
Sanderman, J., Savage, K., Dangal, S.R., Duran, G., Rivard, C., Cavigelli, M.A., Gollany, H.T., Jin, V.L., Liebig, M.A., Omondi, E.C., Rui, Y., 2021b. Can agricultural management induced changes in soil organic carbon be detected using mid-infrared spectroscopy? Remote Sens. (Basel) 13 (12), 2265.
Santín, C., Doerr, S.H., Preston, C.M., González-Rodríguez, G., 2015. Pyrogenic organic matter production from wildfires: a missing sink in the global carbon cycle. Glob. Chang. Biol. 21 (4), 1621–1633.
Schiedung, M., Ascough, P., Bellè, S.-L., Bird, M.I., Bröder, L., Haghipour, N., Hilton, R. G., Lattaud, J., Abiven, S., 2024. Millennial-aged pyrogenic carbon in high-latitude mineral soils. Commun. Earth Environ. 5, 177.
Schmidt, M.W., Skjemstad, J.O., Czimczik, C.I., Glaser, B., Prentice, K.M., Gelinas, Y., Kuhlbusch, T.A., 2001. Comparative analysis of black carbon in soils. Global Biogeochem. Cycles 15 (1), 163–167.
Schumacher, B.A., 2002. Methods for the determination of total organic carbon (TOC) in soils and sediments. US Environmental Protection Agency, Office of Research and Development, Ecological Risk Assessment Support Center. Washington, DC. pp. 1–23.
Skjemstad, J.O., Clarke, P., Taylor, J.A., Oades, J.M., Mcclure, S.G., 1996. The chemistry and nature of protected carbon in soil. Soil Res. 34, 251–271.
Skjemstad, J.O., Taylor, J.A., Smernik, R.J., 1999. Estimation of charcoal (char) in soils. Commun. Soil Sci. Plant Anal. 30 (15–16), 2283–2298.
Stockmann, U., Adams, M.A., Crawford, J.W., Field, D.J., Henakaarchchi, N., Jenkins, M., Minasny, B., McBratney, A.B., De Courcelles, V.D.R., Singh, K., Wheeler, I., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agric. Ecosyst. Environ. 164, 80–99.
Tfaily, M.M., Chu, R.K., Tolić, N., Roscioli, K.M., Anderton, C.R., Paša-Tolić, L., Robinson, E.W., Hess, N.J., 2015. Advanced solvent-based methods for molecular characterization of soil organic matter by high-resolution mass spectrometry. Anal. Chem. 87 (10), 5206–5215.
Waskom, M.L., 2021. Seaborn: statistical data visualization. J. Open Source Software 6 (60), 3021.
Wynn, J.G., Bird, M.I., Vellen, L., Grand-Clement, E., Carter, J., Berry, S.L., 2006. Continental-scale measurement of the soil organic carbon pool with climatic, edaphic, and biotic controls. Global Biogeochem. Cycles 20 (1).
Wynn, J.G., Bird, M.I., 2007. $C_4$-derived soil organic carbon decomposes faster than its C3 counterpart in mixed $C_3/C_4$ soils. Glob. Chang. Biol. 13 (10), 2206–2217.
Wurster, C.M., Lloyd, J., Goodrick, I., Saiz, G., Bird, M.I., 2012. Quantifying the abundance and stable isotope composition of pyrogenic carbon using hydrogen pyrolysis. Rapid Commun. Mass Spectrom. 26 (23), 2690–2696.
Wurster, C.M., Saiz, G., Schneider, M.P., Schmidt, M.W., Bird, M.I., 2013. Quantifying pyrogenic carbon from thermosequences of wood and grass using hydrogen pyrolysis. Org Geochem. 62, 28–32.
Forbes, M.S., Raison, R.J., Skjemstad, J.O., 2006. Formation, transformation and transport of black carbon (charcoal) in terrestrial and aquatic ecosystems. Sci. Total Environ. 370, 190–206.
Preston, C.M., Schmidt, M.W.I., 2006. Black (pyrogenic) carbon: a synthesis of current knowledge and uncertainties with special consideration of boreal regions. Biogeosciences 3, 397–420.
Hockaday, W.C., Grannas, A.M., Kim, S., Hatcher, P.G., 2007. The transformation and mobility of charcoal in a fire-impacted watershed. Geochim. Cosmochim. Acta 71, 3432–3445.
Baldock, J.A., Smernik, R.J., 2002. Chemical composition and bioavailability of thermally altered Pinus resinosa (Red pine) wood. Org. Geochem. 33 (9), 1093–1109.