Identifying the fingerprint of permanganate oxidizable carbon as a measure of labile soil organic carbon using Fourier transform mid-infrared photoacoustic spectroscopy

Huang, Jing; Rinnan, Åsmund; Bruun, Thilde Bech; Engedal, Tine; Bruun, Sander

Published in: European Journal of Soil Science

DOI: 10.1111/ejss.13085

Publication date: 2021

Document version
Peer reviewed version

Document license: Unspecified

Identifying the fingerprint of permanganate oxidizable carbon as a measure of labile soil organic carbon using Fourier transform mid-infrared photoacoustic spectroscopy

J. Huang\textsuperscript{a}, Å. Rinnan\textsuperscript{b}, T. B. Bruun\textsuperscript{c}, T. Engedal\textsuperscript{a}, S. Bruun\textsuperscript{a*}

\textsuperscript{a}Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark

\textsuperscript{b}Department of Food Science, University of Copenhagen, Rolighedsvej 26, 1958 Frederiksberg, Denmark

\textsuperscript{c}Department of Geosciences and Natural Resource Management, University of Copenhagen, Øster Voldgade 10, 1350 Copenhagen, Denmark

*Corresponding Author: Sander Bruun. (E-mail: sab@plen.ku.dk).

Running Title: POXC ASSESSMENT USING FTIR-PAS
Summary

Permanganate oxidizable carbon (POXC) is a sub pool of labile soil organic carbon (SOC) and is defined as the carbon (C) that can be oxidized by potassium permanganate (KMnO₄). Determination of the content of POXC has been used as a way to measure labile C and may be related to soil microbial biomass and particulate organic C. Fourier transform mid-infrared photoacoustic spectroscopy (FTIR-PAS) has been used for characterizing the chemical composition of soil organic matter. The objective of the current study was to identify the chemical information in the FTIR-PAS spectra that is unique to POXC and therefore associated with labile C. FTIR-PAS was used for characterizing SOC using a set of soil samples collected in four different countries (Laos, Malaysia, Peru and Thailand, n=575). Partial least squares (PLS) regression was used to correlate the collected FTIR-PAS spectra with both POXC and total SOC. The FTIR-PAS spectra combining with PLS method was useful for predicting both SOC and POXC. However, SOC and POXC were also highly correlated ($R^2=0.84$), and since POXC can be seen as a fraction of total SOC, a method for extracting the FTIR-PAS fingerprint of the POXC model which is independent of SOC was applied for the first time. The results showed that the POXC prediction was mainly based on its strong correlation with SOC and there was little information unique to POXC. However, the results do indicate four dominant peaks at 1574 cm⁻¹, 1495 cm⁻¹, 1138 cm⁻¹ and 1041 cm⁻¹ to be specific for POXC. These peaks can be ascribed to amide II, aromatics, polysaccharides and carbohydrates, respectively.

Highlights

- Permanganate oxidizable carbon (POXC) was predicted using FTIR-photoacoustic spectroscopy (FTIR-PAS).
• The POXC prediction using FTIR-PAS is mainly based on its high correlation with SOC.

• A method extracting the unique spectral features of POXC independently of SOC was applied.

• Around 11% of the information used to predicting POXC was identified as unique.

• Four dominant peaks were found to represent the POXC fingerprint.

**Keywords:** FTIR-photoacoustic spectroscopy; Permanganate oxidizable carbon (POXC); Labile C; PLS regression; Correlated references

**Introduction**

Permanganate oxidizable carbon (POXC) is suggested to represent a labile soil C pool and has been used as a sensitive indicator of the effect of management on soil quality by providing an early indication of degradation or buildup of soil organic matter in response to management practices (Bruun., 2017; Culman et al., 2012; Weil et al., 2003). Since Weil et al. (2003) developed the method where SOC is oxidized by diluted 0.02M KMnO₄, the determination of the content of POXC was heralded as a rapid and inexpensive method to study dynamic labile C in soil. However, oxidation of SOC using permanganate not only targets labile C, but also affects more recalcitrant forms of C (Skjemstad et al., 2006; Tirol-Padre and Ladha, 2004).

Although the chemical oxidation (using a reagent of a certain concentration) is supposed to mimic biological oxidation, it is a different process compared to biological degradation which is the natural process that degrades soil organic matter (Gruver, 2015). In addition, it has been shown several times that chemical fractionation methods like POXC designed to determine a labile fraction produces results that are highly correlated with total SOC (Calderón et al., 2017; Thomsen et
Therefore, more evidence that POXC actually represent a labile C pool is needed.

Infrared (IR) spectroscopy has been widely used as a rapid and non-destructive method for characterizing a large variety of materials. In terms of evaluation of C in soils, near-infrared (NIR) and mid-infrared (MIR) spectroscopy have provided accurate determinations of SOC (Barthes et al., 2006; Chang et al., 2001; Reeves et al., 2001) as well as several fractions of carbon such as particulate organic carbon, charcoal C (Janik et al., 2007; Calderón et al., 2011; Madhavan et al., 2016) and biodegradable C as assessed in incubation studies (Peltre et al., 2014). Some studies have also used spectroscopic methods to predict POXC (Calderón et al., 2017; Margenot et al., 2020). Calderón et al. (2017) used both MIR and NIR spectroscopy for predicting POXC and SOC using a relatively large soil sample set (n=496) with different soil types. The results showed that both MIR and NIR spectroscopy gave acceptable predictions of POXC with a slightly better prediction using MIR than NIR spectroscopy. The fact that POXC can be predicted from the IR spectra indicates that the method is isolating a fraction of SOC which somehow reflects the chemical composition of labile C in the soil. However, Calderón et al. (2017) also found that SOC was highly correlated with POXC ($R^2 = 0.71$). This indicates that prediction of POXC could to some extent be based on features in the spectra that are associated with the signals from SOC.

There is no doubt that the IR spectroscopy can be used for predicting carbon fractions in soils such as total C, organic C and POXC. However, these predictions may rely on the covariance structures between these parameters. Therefore, reported predictions of POXC from IR measurements are likely to rely on indirect correlations, which are confined to covariance structures in the data set rather than absorption bands directly
associated with POXC. Therefore, there is a current lack of knowledge to the 
existence of unique information in the IR spectra with regards to POXC, or if POXC 
can be predicted just because it is correlated with SOC.

This challenge was termed the *cage of covariance* by Eskildsen et al. (2014) who 
discussed the inherent correlation between individual fatty acids in milk and the effect 
of breed on the whole fatty acid profile. Eskildsen et al. (2016) expanded on this 
work, in a study of prediction of individual proteins and the total protein content in 
milk. They investigated the correlation between the different regression coefficients in 
the final model, as well as the correlation between the different predicted values.

Berhe et al. (2016) went further, and divided the data matrix (X) into an orthogonal 
and a predictive part based on the regression coefficients, and used this to indicate the 
unique and common parts for each PLS models. Bevilacqua et al. (2017) and Løkke et 
al. (2016) also discussed the problem with the *cage of covariance*, and the importance 
of being aware of this for any multivariate model. Instead of investigating the close 
relations between the different regression coefficients in the PLS models, Rinnan et 
al. (2017) investigated this by selecting a smaller sample set with greatly reduced 
correlation between cellulose and hemicellulose in straw in order to identify the 
information which was unique to the two components.

In this study, FTIR-PAS (Fourier transform infrared spectroscopy with photoacoustic 
detection) was applied, which has been used in many different areas and provided 
useful applications in soil analysis (Du et al., 2009; Du and Zhou, 2009; Peltre et al., 
2014). This method can be used with highly absorbing samples without pretreatment. 
Moreover, FTIR-PAS offers the advantage that sample opacity has little effect on the 
PAS signal (Peltre et al., 2014).
The purpose of this study is to identify the information in the FTIR-PAS spectra which is unique to POXC, i.e. the information which is only useful for predicting POXC and does not rely on signals in the spectra produced by SOC. If any such information can be extracted, it is an indication that POXC is providing separate information in addition to SOC. Subsequently, we can analyze this unique information and identify chemical bonds and compounds as identified by the FTIR-PAS spectra, which are associated with the POXC and therefore supposedly labile C.

Materials and methods

Soil sample collection

A total of 575 soil samples were collected from four different studies in Peru (287), Laos (162), Malaysia (66) and Thailand (60). The soils at all four sites are classified as Ultisols (USDA Soil Taxonomy System). The study site in Peru was located in the Northern Peruvian Amazon in the district of Yurimaguas in the borderland between the two regions of Loreto and San Martin (5°57’-6°8’S, 76°10’-76°17’W). Sites in this study were selected to investigate the changes in soil carbon stocks following conversion from primary forest to swidden cultivation and oil palm plantations under different management practices (Engedal, 2018). The study site in Laos was located in Navene, a remote upland village located in the northern province of Louangphabang (20°13’29.37”N, 103°06’30.68”E). Sites in this study were selected to investigate soil fertility change during the fallow period of a shifting cultivation system and soil samples were collected from sites under different fallow years (Hepp et al., 2018). The study site of Malaysia was located in the state of Sarawak in Malaysian Borneo. Soil was sampled from plots that had been converted from shifting cultivation to small-scale oil palm plantations with a sandy clay loam texture (Bruun et al., 2013, 2006). The study site of Thailand was a rural village, Ban Huai Ra Pee (18°54’N;
100°37’E) located in the northern part of Thailand used for maize cultivation in traditional shifting cultivation system (Burup, 2015).

Soil samples were collected from different depth intervals (Table 1). After the collection, all the soil samples were dried and finely ground (<1mm) before analysis.

**POXC and SOC analyses**

Concentrations of POXC were determined using the same experimental setup in all the studies deploying the KMnO₄ oxidation protocol as described by Weil et al. (2003). Briefly, 2.5 g of dry and crushed soil was weighed into a Falcon tube, to which 18 ml of milliQ water and 2 ml of 0.2 M KMnO₄ were added. The samples were shaken for 2 min. After a settling time of 10 min, 1 ml of the supernatant was transferred to a new Falcon tube with 19 ml milliQ water. Absorbance was measured at 550 nm with a spectrophotometer (Genesys 10S UV-VIS) and the POXC values were calculated as follows:

\[
POXC \text{ (mg kg}^{-1}\text{ soil}) = [0.02 \text{ mol L}^{-1} - (a + b \times \text{Abs})] \cdot (9000 \text{ mg C mol}^{-1}) \cdot (0.02 \text{ L solution} / 0.0025 \text{ kg soil})
\]

Where 0.02 mol L⁻¹ is the solution concentration used for the oxidation, \(a\) is the intercept and \(b\) is the slope of the standard curve, \(\text{Abs}\) is the absorbance of the unknown soil sample, 9000 mg is mg C oxidized by 1 mol of MnO₄ changing from Mn⁷⁺ to Mn⁴⁺, 0.02 L is the volume of added KMnO₄ solution and 0.0025 kg is the amount of the soil.

Total SOC was measured by Dumas combustion on an elemental analyser for soil samples in Laos (Elementar Analysensysteme GmbH, Hanau, Germany) and Peru (CE1110, Thermo Electron, Milan, Italy) while soil samples from Malaysia and Thailand were determined by Dumas combustion and an elemental analyzer coupled
to an isotope ratio mass spectrometer (ANCA SL & 20–20, Europa Scientific, Crewe, United Kingdom).

FTIR-PAS analyses

FTIR-PAS spectra of dried and archived soil samples were recorded using a PA301 photoacoustic detector (Gasera Ltd. Turku, Finland) connected to a Nicolet 6700 FTIR spectrometer (ThermoScientific, USA). A helium gas purging flow was used to reduce the noise produced by the evaporation of moisture from the samples. The samples were packed in small cups of 10 mm in diameter. For each sample, 64 scans were recorded in the mid-infrared region between 4000 and 600 cm\(^{-1}\) at a resolution of 4 cm\(^{-1}\) and averaged.

Data analysis

The relationship between POXC and SOC was studied by fitting different models to the data. This was done by minimizing the sum of squared errors between observations and model predictions of POXC as a function of SOC. The tested models included linear, logistic, Monod, and a linear model with a maximum level (Supplement Figure 1).

The FTIR-PAS spectra were pre-processed by a combination of a smoothing (Savitzky–Golay algorithm (Savitzky and Golay, 1964) using four points on each side and a zeroth order polynomial (in effect a moving average)), a Standard Normal Variate (Barnes et al., 1989), and finally mean-centering. A one-component Principal Component Analysis (PCA) was made on this pre-processed data. The dataset was prepared for cross-validation (Stähle and Wold, 1987) by segmenting the dataset according to the ordered score-values from this PCA, by a five-segmented Venetian Blind, ensuring an even distribution of the samples in each of the segments. Subsequently, a PLS model using this cross-validation scheme was calculated. The
number of PLS components was estimated based on the root-mean-squared-error (RMSE) of cross-validation, as well as the difference between the RMSE of calibration and RMSE of cross-validation. Subsequently the process to isolate the information unique to POXC in the spectra was applied (see “Cage of covariance” section below). A jack-knifing uncertainty test (Martens and Martens, 2000) was performed on all estimated regression coefficients: the one from the SOC model, the one from the POXC model, as well as the unique POXC information. We decided to be very conservative in finding significant regression coefficients, and thus decided for five standard deviations, instead of the common three. All data analysis was performed in Matlab 2017b (Matworks Inc., Natwick, USA) using well established algorithms for all calculations.

**Cage of covariance**

Cage of covariance, refers to the challenge of understanding if a relationship between a set of independent variables, $X$, with a set of dependent variables, $Y$, is due to chance correlations or there actually is an underlying causal relationship between the two datasets. This is particularly challenging if the attributes of interest in $Y$ are highly correlated. If only one of the $Y$-variables are causally related to the set of independent variables, the regression of all $Y$-variables will behave well. However, deeper data analysis is needed in order to better understand this underlying cage of covariance between the $Y$-variables, and unravel the problem that if this covariance is broken in future samples, the $Y$-variables that are not related to $X$ will fail to give adequate predictions.

In this section, we will show how we have performed the analysis in order to investigate the amount of unique POXC information present in the FTIR-PAS data through a set of linear equations. The method is similar, but simpler, to the one
proposed by Berhe et al. (2016). The whole idea behind the procedure is to estimate how much the two outputs (POXC and SOC) are related, and thus to indicate the dependence of the two within the current dataset. As POXC is seen as a part of the SOC, this is also the direction of relationship we have decided to investigate in this study. A schematic overview of the whole operation is shown as below (Figure 1).

**Figure 1** A schematic overview of how to divide the regression coefficient of the PLS model towards POXC ($b_{POXC}$) into the part that is common to the PLS model towards SOC ($b_{||POXC}$) and the unique part ($b_{\perp POXC}$). PLS models for both constituents are made, and the regression coefficient of the POXC model is projected onto the regression coefficient for the SOC model, efficiently finding the part of the regression coefficient common between the two models. The remaining part of the regression coefficient of POXC is thus the unique part of this PLS model (towards SOC).

$$b_{POXC} = b_{||POXC} + b_{\perp POXC} = k \cdot b_{SOC} + b_{\perp POXC}$$

1 “Projected onto”, can also be seen as “regressed onto”, but projected is more correct from a mathematical point of view.
The general equation of any regression model can be written in matrix notation as:

\[ y = X \cdot b + e_y \]

where \( y \) is the mean-centered response value (POXC or SOC in this case), \( X \) is the pre-processed FTIR-PAS data, \( b \) is the regression coefficients, and \( e_y \) is the prediction errors of the model. Please note that \( X \) is the same for both the POXC and SOC regression model, as long as the same pre-processing is used, and the same samples have been included. \( y \), \( b \) and \( e_y \) are all dependent on what response value that is predicted.

If we now assume that we have used the same pre-processing and have included the same samples for both models, we can now write the regression model for POXC:

\[ y_{POXC} = X \cdot b_{POXC} + e_{y_{POXC}} \]

And for SOC:

\[ y_{SOC} = X \cdot b_{SOC} + e_{y_{SOC}} \]

However, these two equations are very general, and we have not yet done an effort to investigate the POXC fingerprint of the regression model. In order to investigate if the POXC model is just a scaled version of the SOC model, we can divide the \( b_{POXC} \) into a part that is common with \( b_{SOC} \) (\( b_{\parallelPOXC} \)) and one that is unique to \( b_{POXC} \) (\( b_{\perpPOXC} \)):

\[ b_{POXC} = b_{\parallelPOXC} + b_{\perpPOXC} = k \cdot b_{SOC} + b_{\perpPOXC} \]

where \( k \) is a scaling factor between \( b_{SOC} \) and \( b_{\parallelPOXC} \), as the units of SOC and POXC are different. The solution to Equation 4 is performed by projecting \( b_{POXC} \) onto \( b_{SOC} \). The part of \( b_{POXC} \) that is common with \( b_{SOC} \) is then found, and the residual between these is the unique POXC information. In other words, we here show how to isolate

---

1 \( \parallel \) Refers to the mathematical term parallel to, while \( \perp \) means orthogonal to. In common language, this translates into common and unique information (with regards to SOC in this context).
the unique information of POXC from the information that is common with SOC.

This further means that:

Equation 5 \[ \mathbf{y}_{\text{POXC}} = \mathbf{X} \cdot \mathbf{b}_{\text{POXC}} + \mathbf{e}_{\text{POXC}} = \mathbf{X} \cdot (\mathbf{b}_{\parallel \text{POXC}} + \mathbf{b}_{\perp \text{POXC}}) + \mathbf{e}_{\text{POXC}} \]

Equation 5 \[ \mathbf{y}_{\text{POXC}} = \mathbf{y}_{\text{POXC}} = \mathbf{y}_{\parallel \text{POXC}} + \mathbf{e}_{\text{POXC}} = \mathbf{y}_{\parallel \text{POXC}} + \mathbf{y}_{\perp \text{POXC}} + \mathbf{e}_{\text{POXC}} \]

This means that we now can split the prediction of POXC (\(\mathbf{y}_{\text{POXC}}\)) up into a part that is dependent on SOC (\(\mathbf{y}_{\parallel \text{POXC}}\)), and one that is unique to POXC (\(\mathbf{y}_{\perp \text{POXC}}\)). Please be aware that since \(\mathbf{y}_{\parallel \text{POXC}}\) only includes part of the total POXC regression coefficient, these predictions will be worse at predicting POXC, than the full model (\(\mathbf{y}_{\text{POXC}}\)). In order to investigate the amount of commonality and uniqueness between the SOC and POXC regression, we can deduce (from Figure 1 and Equation ) that:

Equation 6 \[ SSQ(\mathbf{b}_{\text{POXC}}) = SSQ(\mathbf{b}_{\parallel \text{POXC}}) + SSQ(\mathbf{b}_{\perp \text{POXC}}) \]

This means that by calculating the sum-of-squares (SSQ) of the three regression coefficients, we can also calculate the commonality and uniqueness as a percentage of the total, just as one does in order to calculate the explained variance of a PCA model, or the significance of a treatment in an ANOVA model.

This whole procedure is readily imbedded in any validation method, such as cross-validation (Ståhle and Wold, 1987), and thereby in an uncertainty test (i.e. jack-knifing (Martens and Martens, 2000)) giving direct information into the variables contributing most to the unique information.

However, this methodology is not implemented into any chemometric software, and thus, unfortunately, needs to be programmed.

Results

Soil characteristics

The average contents of both POXC and SOC in four different sites are summarized by different depth intervals (Table 1). The distributions of POXC and SOC within
The soils used in this study are all classified as Ultisols, with a lower pH range of 3.3-5.0 in Peru and a larger pH range of 5.0-6.5 in Thailand. It is not easy to compare POXC and SOC contents between the different sites because the soil was not sampled in the same depth intervals in the different studies included. In general, the average contents of POXC and SOC from upper soil layers were higher than in soil samples from the lower layers for all the soil samples. Particularly, the average contents of POXC and SOC at 0-5 cm depth in Laos were the highest of 1107 mg kg\(^{-1}\) and 4.9% respectively. On the contrary, the lowest average contents of POXC was 36 mg kg\(^{-1}\) in the soil from Thailand at 30-50 cm depth interval, while the lowest average contents of SOC was 0.6% in the soil from Peru at 20-50 cm depth interval.

**Correlation between POXC and SOC**

Of the different models tested, a linear model with a maximum level was the best fit of POXC and SOC contents and the model gave rise to an \(R^2\) of 0.88 (Supplement Figure 1). The fit showed that when SOC < 4.6%, the correlation of POXC and SOC can be well described using a linear model. Above this level POXC did not increase even though SOC did. As discussed later the reason for this relates to the maximum detection limit of the POXC method. Therefore, to give the best correlation between these two parameters, in this study, soil samples with SOC < 4.6% were only included for further analysis (Figure 2).
Figure 2 A linear model between POXC and SOC using the soil samples with SOC<4.6%.

PLS models for SOC, POXC, and the unique POXC information relation

To avoid inconsistencies, 12 samples from Malaysia with a very different IR profile compared with the other soil samples were removed. Moreover, 7 samples detected by Hotelling’s $T^2$ were also removed from the dataset.

Figures of merit for the PLS regression models for SOC, POXC, as well as the part of POXC prediction that is dependent on SOC model ($||POXC$) are given (Table 2).

It can be appreciated that the unique POXC information in the FTIR-PAS spectra is relatively small, as the difference between model 2 and model 3 is relatively small, with an improvement of RMSEC from 137 mg kg$^{-1}$ with the SOC model down to 121 mg kg$^{-1}$ in the POXC model, with the values of RMSECV being very similar (from 138 mg kg$^{-1}$ down to 123 mg kg$^{-1}$). By using Equation it can be calculated that the amount of unique information in the regression coefficient ($SSQ(b_{LPOXC})/SSQ(b_{POXC})$) equals 11.3% ± 1.2%. This is very similar to the decrease in RMSE values (13.9% and 11.8% for respectively RMSEC and RMSECV). As there appears to be some uniqueness in the models, it would be of interest to investigate what
wavenumbers and chemical bonds the corresponding unique POXC information in the FTIR-PAS data reflects.

**Regression coefficients analysis**

**POXC vs SOC**

The regression coefficients of the POXC, ||POXC and ⊥POXC are shown, among which ||POXC is a scaled version of SOC (Figure 3). Therefore, the comparison of POXC and ||POXC is equal to the comparison of POXC and SOC.

![Figure 3](normalized_regression_coefficients.png)

**Figure 3** Normalized regression coefficients of POXC, ||POXC (common with SOC) and ⊥POXC (unique). The black dotted line is the zero line. The four vertical black lines indicate the most important wavenumbers of the ⊥POXC (unique) part of the regression coefficients.

Both POXC and SOC were mainly showing positive regression coefficients with the band at 1556 cm\(^{-1}\) attributed to amide II (Janik et al., 2007), with the band at 1442-
1385 cm$^{-1}$ attributed to alkyl–CH$_2$ deformation, which could also be ascribed to methyl compounds considered to be easily degradable (Peltre et al., 2014). Positive coefficients were also found for the band at 1230 cm$^{-1}$, attributed to aromatic–CH in plane deformation, with the band at 1706 cm$^{-1}$ attributed to carboxylic acid–COOH, with the band at 1655 cm$^{-1}$ attributed to amide I and aromatic C=C (Janik et al., 2007; Calderón et al., 2013; Peltre et al., 2014; Margenot et al., 2015; Barthès et al., 2016). A positive correlation was also observed at 1508 cm$^{-1}$, corresponding to aromatic compounds (Smidt and Meissl, 2007). In addition, both POXC and SOC were positively correlated with the band at 1041 cm$^{-1}$ may attributed to carbohydrates instead of Si-O from silicates or quartz (Soriano-Disla et al., 2014). The peaks at 2925 cm$^{-1}$ and 2850 cm$^{-1}$ showed weakly positive coefficients for both POXC and SOC and are attributed to aliphatic C–H stretching (Calderón et al., 2017).

Both POXC and SOC were generally negatively correlated with the band at 1872 cm$^{-1}$ ascribed to the quartz overtone, and with the band at 3700-3500 cm$^{-1}$ centered at 3693 cm$^{-1}$ attributed to kaolinite and gibbsite mineral phases (Janik et al., 2007). Negative correlation were also observed with the band at 849 cm$^{-1}$ of SOC and 821 cm$^{-1}$ of POXC respectively, representing quartz-like Si-O (Margenot et al., 2020).

**Unique POXC ($\perp$POXC)**

In the regression coefficients of $\perp$POXC, we can interpret the functional groups of the POXC model that is unique to POXC independently from SOC (Figure 3). Based on the jack-knifing result on the unique part of the POXC model, we have been able to identify four wavenumber-regions that are most important for the unique part (largest ratio between the average regression coefficient and the standard deviation of the same regression coefficient, based on the results from the cross-validation).
A positive correlation of unique POXC information (∟POXC) was found with the band at 1574 cm\(^{-1}\), which could be attributed to the stretching vibration of N-H and C-N from Amide II (Peltre et al., 2014). As for the band at 1495 cm\(^{-1}\), which may be ascribed to aromatic rings and amide II vibration, a positive correlation with ∟POXC was found. It could also be attributed to N-H stretching of proteic amides (Laudicina et al., 2015).

Positive coefficients were also found with the bands at 1138 cm\(^{-1}\) and 1041 cm\(^{-1}\), which could be ascribed to polysaccharides and C-O in other carbohydrates (Janik et al., 2007b; Peltre et al., 2014). Accepting the premise that POXC represent a labile fraction of C, this may indicate that these carbohydrates belongs to the easily degradable fraction. Peltre et al. (2014) found positive regression coefficients around 1110 cm\(^{-1}\) which could correspond to the first peak, but not the second peak.

**Discussion**

**Correlation between POXC and SOC**

In general, there was a strong correlation between POXC and SOC contents in the soil dataset with R\(^2\) from 0.84 to 0.88 (Supplement Figure 1). It was clear that at high SOC levels the relationship between POXC and SOC leveled off, however this is not due to changes in the proportions of SOC versus POXC (Figure 2). In theory, the maximum value of POXC is 1440 mg kg\(^{-1}\) when 2 ml of 0.2M KMnO\(_4\) is used (Calderón et al., 2017), which means there is a detection limitation of the POXC method (Weil et al., 2003). Therefore the measured values become unreliable when we approach this value. This justifies the choice to remove samples from the analysis containing more than 4.6% of SOC.
The same observation was made by Calderón et al. (2017), who found a better fit of POXC and SOC was hyperbolic rather than linear ($R^2=0.71$) due to the detection limit of the POXC method.

Comparison of PLS performance and regression coefficients between POXC and SOC

Our results show a comparable performance of POXC prediction ($R_{cv}^2$ of 0.848 and RMSECV of 123 mg kg$^{-1}$) with other studies using diffuse reflectance mid-infrared spectra (Calderón et al., 2017; Margenot et al., 2020; Veum et al., 2014). In addition, the prediction of SOC was better than POXC in our case, which may indicate that the prediction of POXC could be dependent on SOC. However, it could also simply indicate the reference measurement of POXC is more difficult to perform, and thus include more noise. On the contrary, in the study of Calderón et al. (2017), a better performance of POXC prediction than SOC was found. As POXC can be seen as a sub-part of SOC, and not the other way around, the soil dataset in our study is more suitable for investigating if there is some unique spectra features of POXC independently of SOC.

In our study, POXC and SOC were mainly positively correlated with organic functional groups. Positive correlations of POXC and SOC with the band at 1556 cm$^{-1}$ and 1442-1385 cm$^{-1}$ can be ascribed to labile organic compounds (amide and methyl compounds). Moreover, positive correlations of POXC and SOC with aromatic compounds were also found at 1230 cm$^{-1}$ attributed to aromatic–CH and at 1655 cm$^{-1}$ and 1508 cm$^{-1}$ ascribed to aromatic C=C respectively. Other organic compounds (carboxylic acid–COOH and carbohydrates) also showed a positive correlation with POXC and SOC.

The peaks at 2925 cm$^{-1}$ and 2850 cm$^{-1}$ showed weakly positive coefficients for both POXC and SOC representing aliphatic C–H stretching (Calderón et al., 2017), but
could also be ascribed to carbonates (Legodi et al., 2001; Tatzber et al., 2007). Since all of the soils are highly weathered Ultisols with low pH values, it is not likely that the samples contain carbonates. In the study of Calderón et al. (2017), a highly positive correlation of POXC and SOC at 2925 cm$^{-1}$ to 2850 cm$^{-1}$ was found, which can be used as indicators for characterizing the condition of soil decomposition (Veum et al., 2014). However, both POXC and SOC were highly correlated with the bands at 2925 cm$^{-1}$ to 2850 cm$^{-1}$ and for this reason, they cannot be easily used to distinguish chemical groups uniquely representing POXC. The FTIR-PAS sensor provides high quality spectra in the lower range (< 1500 cm$^{-1}$) with less noise and clearer peaks compared with the spectra in Calderón et al. (2017). Therefore, more peaks related to organic matter were found by FTIR-PAS spectra in the lower range compared to characterization of POXC and SOC using conventional diffuse reflectance mid-infrared spectroscopy.

Considering the difference in the regression coefficients of POXC and SOC, a higher positive contribution to POXC than SOC was found at the band of 1556 cm$^{-1}$ representing amide as labile C compound. However, a higher positive regression coefficient for POXC than SOC were also found at 1655 cm$^{-1}$ and 1508 cm$^{-1}$ attributed to aromatic C=C, which is generally considered to be not easily degradable organic compounds. Moreover, a higher positive contribution to POXC than SOC was found at the band of 1041 cm$^{-1}$ representing carbohydrates. These could also be regarded as a proof that the prediction of POXC using mid-infrared spectra may contain some spectral features relating to SOC because the positive correlation of POXC with related function groups are not all representing easily degradable C fraction. At the band of 3700-3500 cm$^{-1}$, a slightly more negative correlation to POXC was observed while a more negative contribution of SOC was found at the
band at 1872 cm$^{-1}$ and 912-790 cm$^{-1}$, representing the quartz overtone and Si-O respectively. In general, the high degree of similarity of the regression makes it hard to know which of these differences are statistically significant.

**Importance of unique information in POXC**

The current paper is the first paper to discuss the implications of the high correlation between POXC and SOC. As discussed above this high correlation creates problems when interpreting what information in the spectra that are associated with POXC. Furthermore, it also creates problems for the interpretations of POXC as a labile fraction of SOC. Indeed, if the two variables are perfectly correlated then there is absolutely no additional information in the POXC values. Certainly, they are highly correlated, but the question is if the POXC conveys any additional information. We have here presented a new approach to identify unique information in the spectra of FTIR-PAS. The analysis indicates there is indeed additional information in the POXC spectra. This unique information constitutes 11.3% to the total variation in POXC.

**Interpreting the characteristic peaks for POXC independently of SOC**

In general, there are four characteristic peaks that were identified as positively correlated with the unique POXC information at the band of 1574 cm$^{-1}$, 1495 cm$^{-1}$, 1138 cm$^{-1}$ and 1041 cm$^{-1}$, respectively. The peak at 1574 cm$^{-1}$ can be attributed to amide, which is regarded as easily degradable. In the study of Peltre et al. (2014), the same positive correlation of total organic carbon was found at the band around 1570 cm$^{-1}$. The band at around 1570 cm$^{-1}$ could also be attributed to aromatics (Janik et al., 2007; Soriano-Disla et al., 2014), which is generally resistant. However, it is possible that there are compounds with aromatic groups that are chemically and microbially labile. The interpretation of peak at 1495 cm$^{-1}$ need to be more careful since it could be attributed to aromatic and amide compounds, which more likely to be ascribed to
aromatic compounds in this study. This is surprising as especially the aromatic 
compounds are usually not considered easily degradable. The other two peaks that 
were positively correlated with unique POXC were the bands at 1138 cm$^{-1}$ and 
1041 cm$^{-1}$, which could be ascribed to polysaccharides and carbohydrates. It is worth 
to mention that the characteristic peak at 1041 cm$^{-1}$ for the unique POXC prediction 
has also been identified by POXC prediction model.

In summary, the identical functional groups of POXC contains amides, aromatics, 
polysaccharides and carbohydrates at four dominant peaks, which could be used as 
characteristic peaks for POXC independently of SOC. However, interpreting the 
peaks as labile functional groups is a challenge as some of them, such as aromatic C is 
usually considered non-labile components and rests on an assumption that POXC is 
actually oxidizing primarily labile compounds.

To the best of our knowledge, this is the first time the unique POXC information 
in IR spectra has been extracted and investigated. Further studies are definitely needed 
in order to provide a more precise fingerprint of POXC in different soil types and 
under different soil management.

Conclusions

FTIR-PAS spectroscopy can be used as an efficient method for predicting both SOC 
and POXC. However, a large part of the POXC prediction is dependent on the SOC 
model as the two variables are highly correlated. This means that the signal from total 
SOC and POXC cannot be separated easily. Therefore, direct interpretations of 
chemical functional groups identified in the spectra that are associated with POXC are 
problematic. Based on a method for extracting the information from the FTIR-PAS 
spectra, which is unique to POXC ($\perp$POXC), it was found that the information in 
POXC directly related to SOC ($\parallel$POXC) turned out to constitute the vast majority of
the total POXC regression coefficient (88.7%), while the unique information describing POXC constitutes a relatively small proportion (11.3%). Although this is limited, it indicates that POXC is conveying some additional information not captured by total SOC. The regression coefficients in the model of the unique information in the FTIR-PAS spectra of POXC (\( \trianglepoxc \)) showed four dominant peaks at 1574 cm\(^{-1}\), 1495 cm\(^{-1}\), 1138 cm\(^{-1}\) and 1041 cm\(^{-1}\) respectively, which could be seen as characteristic peaks for POXC which are independent of SOC.

Acknowledgements

We gratefully acknowledge the financial support from the China Scholarship Council for Jing Huang (201606350206). We thank Catherine Maria Hepp and Maria Libach Burup for making soil samples available for this study.

Author contributions

Study concept and design: S.Bruun and J.Huang

Drafting of the manuscript: J.Huang

Data analysis and chemometrics: Å.Rinnan and J.Huang

Sampling and laboratory analysis: J Huang, T.B.Bruun and T.Engedal

All authors discussed the results and revised the manuscript.

Conflict of interest

No conflict of interest to declare

References


Environment. Land Degradation & Development, 26, 45–53.

doi:10.1002/ldr.2293


Table 1 Average values of POXC and SOC at different depths.

<table>
<thead>
<tr>
<th>Site</th>
<th>pH (cm)</th>
<th>Depth (cm)</th>
<th>POXC (mg kg(^{-1}))</th>
<th>SOC (%)</th>
<th>n</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laos</td>
<td>4.6-4.9</td>
<td>0-5</td>
<td>1107±179</td>
<td>4.9±1.3</td>
<td>81</td>
<td>Hepp et al., 2018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0-10</td>
<td>444±127</td>
<td>2.5±0.7</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td>4.8-6.0</td>
<td>0-10</td>
<td>613±310</td>
<td>2.2±0.9</td>
<td>33</td>
<td>Bruun et al., 2006, 2013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10-20</td>
<td>251±155</td>
<td>1.0±0.4</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Peru</td>
<td>3.3-5.0</td>
<td>0-10</td>
<td>489±233</td>
<td>2.0±0.7</td>
<td>95</td>
<td>Engedal, 2018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10-20</td>
<td>147±109</td>
<td>0.9±0.4</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20-50</td>
<td>66±74</td>
<td>0.6±0.2</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0-10</td>
<td>795±213</td>
<td>3.0±0.8</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>5.0-6.5</td>
<td>10-20 (20-30)</td>
<td>293±177</td>
<td>1.5±0.5</td>
<td>30</td>
<td>Burup, 2015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30-50</td>
<td>36±42</td>
<td>0.9±0.2</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
Table 2 PLS regression calibration of SOC, POXC and ||POXC.

<table>
<thead>
<tr>
<th>#</th>
<th>Description</th>
<th>Factors</th>
<th>R$_{cal}^2$</th>
<th>RMSEC</th>
<th>R$_{CV}^2$</th>
<th>RMSECV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SOC</td>
<td>5</td>
<td>0.934</td>
<td>0.28$^a$</td>
<td>0.933</td>
<td>0.28$^a$</td>
</tr>
<tr>
<td>2</td>
<td>POXC</td>
<td>5</td>
<td>0.854</td>
<td>121$^b$</td>
<td>0.848</td>
<td>123$^b$</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>POXC</td>
<td></td>
<td>0.811</td>
<td>137$^b$</td>
</tr>
</tbody>
</table>

a: the unit is %. b: the unit is mg kg$^{-1}$. 
FIGURE CAPTIONS

Figure 1

Figure 2