Dinesh B. Madhavan, Jeff A. Baldock, Zoe J. Read, Simon C. Murphy, Shaun C. Cunningham, Michael P. Perring, Tim Herrmann, Tom Lewis, Timothy R. Cavagnaro, Jacqueline R. England, Keryn I. Paul, Christopher J. Weston, Thomas G. Baker

Rapid prediction of particulate, humus and resistant fractions of soil organic carbon in reforested lands using infrared spectroscopy

Journal of Environmental Management, 2017; 193:290-299

Rapid prediction of particulate, humus and resistant fractions of soil organic carbon in reforested lands using infrared spectroscopy

Dinesh B. Madhavan[a,*], Jeff A. Baldock[b], Zoe Read[c], Simon C. Murphy[a], Shaun C. Cunningham[d,e,^], Michael P. Perring[f,g], Tim Herrmann[h], Tom Lewis[i], Timothy R. Cavagnaro[j], Jacqueline R. England[k], Keryn I. Paul[l], Christopher J. Weston[m], Thomas G. Baker[a]

[a] School of Ecosystem and Forest Sciences, The University of Melbourne, Richmond VIC 3121, Australia;
[b] CSIRO Agriculture, Glen Osmond SA 5064, Australia;
[c] Fenner School of Environment and Society, Australian National University, Canberra ACT 0200, Australia;
[d] School of Life and Environmental Sciences, Deakin University, Burwood VIC 3125, Australia;
[e] Institute for Applied Ecology, University of Canberra, Bruce ACT 2617, Australia;
[f] School of Plant Biology, The University of Western Australia, Crawley WA 6009 Australia;
[g] Forest & Nature Laboratory, Ghent University, BE-9090, Gontrode-Melle, Belgium;
[h] Department of Environment, Water and Natural Resources, Adelaide SA 5001, Australia;
[i] Agri-Science Queensland, Department of Agriculture and Fisheries, Sippy Downs QLD 4556, Australia;
[j] School of Agriculture, Food and Wine, University of Adelaide, Waite Campus, PMB 1, Glen Osmond SA 5064, Australia;
[k] CSIRO Agriculture and CSIRO Land and Water, Clayton South VIC 3169, Australia;
[l] CSIRO Agriculture and CSIRO Land and Water, Canberra ACT 2601, Australia;
[m] School of Ecosystem and Forest Sciences, The University of Melbourne, Creswick, VIC 3363, Australia.

*Corresponding author: Dinesh Madhavan
Postal address: 500 Yarra Boulevard, Richmond VIC 3121, Australia.
Email: dineshbm@unimelb.edu.au
Mobile: +61 433 481 500

^Author deceased (September 2016)

Abstract

Reforestation of agricultural lands with mixed-species environmental plantings can effectively sequester C. While accurate and efficient methods for predicting soil organic C content and composition have recently been developed for soils under agricultural land uses, such methods under forested land uses are currently lacking. This study aimed to develop a method using infrared spectroscopy for accurately predicting total organic C (TOC) and its fractions (particulate, POC; humus, HOC; and resistant, ROC organic C) in soils under environmental plantings. Soils were collected from 117 paired agricultural-reforestation sites across Australia. TOC fractions were determined in a subset of 38 reforested soils using physical fractionation by automated wet-sieving and ^{13}C nuclear magnetic resonance (NMR) spectroscopy. Mid- and near-infrared spectra (MNIRS, 6000-450 cm^{-1}) were acquired from finely-ground soils from environmental plantings and agricultural land. Satisfactory prediction models based on MNIRS and partial least squares regression (PLSR) were developed for TOC and its fractions. Leave-one-out cross-validations of MNIRS-PLSR models indicated accurate predictions ($R^2 > 0.90$, negligible bias, ratio of performance to deviation > 3) and fraction-specific functional group contributions to beta coefficients in the models. TOC and its fractions were predicted using the cross-validated models and soil spectra for 3109 reforested and agricultural soils. The reliability of predictions determined using *k*-nearest neighbour score distance indicated that > 80% of predictions were within the satisfactory inlier limit. The study demonstrated the utility of infrared spectroscopy (MNIRS-PLSR) to rapidly and economically determine TOC and its fractions and thereby accurately describe the effects of land use change such as reforestation on agricultural soils.

Key words: C sequestration; Biodiverse environmental plantings; Mid-infrared spectroscopy; Near-infrared spectroscopy; NMR spectroscopy; Partial least squares regression

1. Introduction

The soil organic carbon (SOC) pool is one of the largest terrestrial stores of carbon (C), containing about 1500 Gt C (Schlesinger, 1986). Consequently, even small changes in the amounts of soil C in response to land use, management or climate may have large effects on global C cycling and climate change. Reforestation is implemented around the world to sequester C and improve environmental conditions (e.g., water quality and habitat availability, Cunningham et al., 2015b). The total area of world's planted forest in 2010 was estimated to be 264 million ha, making up 6.6% of the world's forest area (FAO, 2010). In Australia, mixed-species environmental plantings are established on previous agricultural land for C sequestration and other environmental outcomes accounting up to 20% of the 1.14 million ha of reforestation between 1990 and 2012 (e.g., Paul et al., 2013). Studies have focused on improving measurement and modelling of biomass C following reforestation with environmental plantings (Paul et al., 2015; Perring et al., 2015), and measured associated changes in SOC (e.g., Cunningham et al., 2015a).

Conventional measurement of SOC as total organic C (TOC) using chemical oxidation (Walkley and Black, 1934) or dry combustion (Merry and Spouncer, 1988) is inadequate to explain changes in SOC in terms of soil physical, chemical and biological activity. Partitioning TOC into fractions that are related to active, intermediate or slow and passive or inert conceptual pools used in soil C turnover models (e.g., RothC, Jenkinson et al., 1992; CENTURY, Parton et al., 1987), can help elucidate their role in soil processes (Skjemstad et al., 2004). Such fractions may represent labile, humified and inert C with turnover times respectively of the order of annual (<3 yr), decadal (20-50 yr) and millennial (>1000 yr, Jenkinson and Coleman, 1994).

Several functional pools are known which are accessible by different fractionation methods. Physical fractions are such as aggregates, particle sizes and density fractions, chemical fractions are usually extracts (DOM, soil microbial biomass, organic matter soluble in alkali and acid, etc.) and also combinations of fractionation methods are used (von Lützow et al., 2007).

Skjemstad et al. (2004) demonstrated that the RothC model could be initialised and changes in soil C stock simulated by replacing the conceptual stocks of resistant plant material (RPM), humus (HUM) and inert organic matter (IOM), respectively, with operationally measured particulate (POC), humus (HOC) and resistant (ROC) organic C. Subsequently, Janik et al. (2007) showed that the concentrations of POC, HOC and ROC in < 2 mm sieved soil could be predicted from mid-infrared (MIR) spectra acquired from finely ground samples.

Diffuse reflectance mid-infrared spectroscopy (MIRS) is a rapid, non-destructive and low-cost technique, demonstrated to be suitable for routine analysis of a variety of soil properties (Soriano-Disla et al., 2014). The MIRS technique requires minimal sample preparation (i.e. air drying and fine grinding) and no use of hazardous chemicals. MIR spectra and corresponding analytical data in multivariate analyses such as partial least squares regression (PLSR) can be combined to develop prediction models for soil attributes. The predictive ability of MIRS-PLSR techniques for total, organic and inorganic C of soils has been well investigated and reported (Grinand et al., 2012; Madari et al., 2006; Madhavan et al., 2016). However, the use of MIRS-PLSR to predict the concentrations of fractions of TOC is limited (Bornemann et al., 2010; Janik et al., 2007; Zimmermann et al., 2007). MIRS-PLSR prediction models have been developed for total, organic and inorganic C, and TOC fractions

104 (Baldock et al. 2013a), and these models were applied to predict the content of TOC fractions
105 in subsequent agricultural soil C studies (Karunaratne et al., 2014; Rabbi et al., 2014).
106 However, the applicability of such predictive models to soils under woody vegetation has not
107 been investigated. Further, there is increasing interest in reforestation of agricultural lands to
108 mitigate greenhouse gas emissions through sequestering C in woody biomass (e.g., Canadell
109 and Raupach, 2008) and in soil (e.g., Lal, 2005). Thus, there is a need to develop accurate and
110 efficient measurement techniques and prediction models applicable to reforested land to
111 understand and predict their potential to sequester C and mitigate greenhouse gas emissions.
112
113 An extensive study was conducted to investigate the changes in TOC and its fractions
114 following reforestation with mixed-species environmental plantings at 117 sites from
115 temperate, Mediterranean-type and tropical climatic regions of Australia, for the purpose of
116 calibrating a soil C accounting model (FullCAM, Brack and Richards, 2002) developed from the
117 RothC model (Paul et al. 2015b), and to provide measurements and predictions of TOC
118 fractions. Prediction models for TOC fractions have been reported in agricultural soils by
119 Baldock et al., 2013a, but these models are unlikely to be accurate for reforested soils because
120 of the difference in plant inputs, chemistry and decomposition rates between land uses, and
121 thereby changes in TOC fractions (Del Galdo et al., 2003; Berthrong et al., 2012; Cunningham
122 et al., 2015a). This warrants a need to develop infrared spectroscopic prediction models for
123 TOC fractions that are suitable for soils under reforestation. This work presents the first
124 attempt to develop prediction models for TOC fractions in a treed ecosystem whereas
125 previous work has focused on agricultural soils (Baldock et al., 2013a). Our objectives were to
126 measure soil TOC fractions (POC, HOC and ROC) in a representative set of reforested soils;
127 develop robust infrared and PLSR prediction models for TOC and its fractions for reforested

soils; and predict TOC and its fractions for both environmental planting and reference agricultural soils (i.e. pastures and cropping).

2. Methods

2.1. Soils

Soils were collected from 117 sites, each comprising a mixed-species environmental planting paired with an adjacent agricultural land use (Fig. 1). Details of site characteristics and sampling methods are detailed in England et al., 2016 and summarised here.

The sites were across southern and eastern Australia and covered the range of rainfall zones where environmental plantings occur (380 – 1150 mm per year). Generally the sites represented a range of planting age (…), productivity (as above-ground biomass increment …), surface soil texture (…) and previous land use. The plantings comprised species … .

141 Agricultural land use at the sites included grazing, cropping, and rotational cropping and
142 grazing … .

2.1. Study sites

New and existing data were collated from 117 mixed-species environmental plantings (subsequently termed 'environmental plantings') established on agricultural land. Plantings were across southern and eastern Australia (latitude −30.9 to −38.7 S, longitude 117.4–150.3 E), and covered the range of rainfall zones where planting occurs (380–1147 mm y^{-1}; Table 1, Fig. 1). Thirty-six new sites were measured to improve the representativeness of plantings with respect to age, previous land use, productivity (aboveground biomass increment) and soil texture (Table 1). Environmental plantings are often established along stream banks to reduce erosion or in areas with shallow water tables to mitigate dryland salinity by minimising recharge (George et al., 1999). Thus sites were selected to include both dryland plantings across a range of landscape positions ($N = 97$) and riparian plantings ($N = 20$).

Previous land use at the sites included grazing, cropping, and rotational cropping and grazing (Table 1). For analysis, we combined cropping with rotational cropping and grazing as 'cropping'. There were fewer previously-cropped sites ($N = 32$) than previously-grazed sites ($N = 85$) reflecting current reforestation activity in Australia. Because environmental plantings are a relatively new land use, mean age of the plantings was 14 y, with 95% of plantings aged between 1 and 28 y. Biomass productivity varied greatly among sites (0.2–31.7 Mg DM ha^{-1} y^{-1}, calculated using methods of Paul et al., 2015) due to differences in site characteristics (e.g. soil type, rainfall), planting structure (e.g. stand density, species-mix, planting configuration) and age.

143

144 For the existing soils, at each site and for each land use, 1-8 plots or transects and 5-18 cores
145 (0 to 30 cm depth), either 0-5 cm and 5-30 cm or 0-10 cm and 10-30 cm depth increments,
146 per plot or transect were collected and analysed separately or after compositing. For the new
147 sites, for each land use, 40 cores (0 to 30 cm depth) were collected for 2-4 increments (0-10,
148 10-30 cm; 0-5, 5-10, 10-30 cm; or 0-5, 5-10, 10-20 and 20-30 cm from a 0.40 ha area divided
149 into 40 sampling units of 10 m x 10 m blocks (one random sampling position in each unit), and
150 randomly composited into 8 replicate samples for analysis.

151

152 Soil samples were collected from each land-use at each site from formal plots or along
153 transects, and to 30 cm depth variously in depth increments (cm): 0-5 & 5-30; 0-10 & 10-30;

154 0-5, 5-10 & 10-30; or 0-5, 5-10, 10-20 & 20-30. Individual samples for analysis (site x land use
155 x depth, *N*= 3109) were generated by compositing (typically) 5 cores.
156
157 2.2. Soil processing and C analysis
158 Soil samples were air dried and manually sieved (≤ 2 mm) without crushing > 2 mm organic
159 matter. The existing soils had been similarly air-dried and processed, and stored in air-tight
160 containers at room temperature. Subsamples (approx. 20 g) of air dried ≤ 2 mm soil were
161 finely ground in a vibratory 10 cm bowl steel puck mill (Rocklabs, Auckland, New Zealand) for
162 120 s (an additional 60 s for coarse textured soils, if required) to obtain a fine talc-like
163 consistency, so as to ensure sub-sample homogeneity for chemical analysis and acquisition of
164 consistent spectra across subsamples. TOC concentrations were determined on the finely
165 ground ≤ 2 mm soil by Dumas high-temperature combustion (Rayment and Lyon, 2011) using
166 a LECO CNS-2000 dry combustion analyser (LECO Corporation, St Joseph, MI, USA). Prior to
167 TOC analyses, the finely ground ≤ 2 mm soil samples were tested for the presence of
168 carbonates using 1 M HCl acid fizz test (Rayment and Lyon, 2011). Samples with carbonate
169 were treated with 1 mL of 5-6% H_2SO_3 at 100 ^oC until effervescence ceased (Nelson and
170 Sommers, 1996) before drying and TOC analysis.
171
172 2.3. Acquisition of infrared spectra
173 Diffuse reflectance spectra (7800-450 cm^{-1} at 4 cm^{-1} resolution), which included the mid-
174 infrared region and some of the near-infrared region, were acquired using a PerkinElmer
175 Frontier FT-NIR/MIR Spectrometer (PerkinElmer Inc., Waltham, MA, USA) equipped with a
176 KBr beam-splitter, a DTGS detector and AutoDiff automated diffuse reflectance accessory

177 (Pike Technologies, Madison, WI, USA). Finely-ground soil samples (approx. 150-250 mg) were
178 uniformly packed and levelled in 9 mm diameter stainless steel cups. Sixty cups and a
179 background disc (silicon carbide) were loaded onto the Autodiff accessory and scanned. At
180 the start of each scanning cycle, a background signal was collected from the silicon carbide
181 disk (average of 240 scans) and subtracted from all sample spectra. For each soil sample, 64
182 co-added scans were acquired and averaged to obtain a representative reflectance spectrum
183 (R). These reflectance spectra were converted to absorbance spectra (A) by the formula, A =
184 log (1/R) using PerkinElmer instrument control Spectrum 5.0.1 Software, which contained
185 peaks corresponding to molecular vibrations of organic matter and minerals (absorbance
186 units along Y axis) against specific wavenumbers (cm^{-1} along X axis, Fig. 2a). The PerkinElmer
187 specific spectral file format (.sp) was converted to the thermo galactic file format (.spc) using
188 GRAMS/AI 9.1 (Thermo Fisher Scientific Inc., Waltham, MA USA). The absorbance spectra
189 were later trimmed to 6000-450 cm^{-1} for analyses.
190
191 2.4. Soil TOC fractionation

192 Environmental plantings soils from 19 sites (0-5 and 5-10 cm depth, n = 38) were selected
193 across the geographical regions of the study area for fractionation, containing TOC
194 concentrations of > 25 g kg^{-1} in 0-5 cm and > 15 g kg^{-1} in 5-10 cm and thereby containing
195 discernible quantities of TOC fractions. The TOC fractionation methodology of Baldock et al.
196 (2013b) was used to measure POC, HOC and ROC concentrations. Three 10 g replicates of ≤ 2
197 mm soil were dispersed using 45 mL of 5 g L^{-1} sodium hexametaphosphate by shaking the
198 tubes overnight. These samples were wet-sieved for 3 min through a 50 µm mesh using an
199 automated vibratory sieve shaker (Analysette 3 PRO, FRITSCH GmbH, Idar-Oberstein,

200 Germany) with an amplitude of 2.5 mm at 20 sec intervals and water flow rate of approx. 100
201 mL min^{-1} to separate coarse (2000-50 μm) and fine fractions (≤ 50 μm). The fractions were
202 lyophilized, coarse fractions were finely ground and analysed for total C concentration by
203 Dumas high temperature combustion. Additional ≤ 2 mm soil was similarly fractionated where
204 necessary to accumulate sufficient coarse and fine fraction material (equivalent of > 20 mg of
205 organic C) for measurement of ROC using solid-state ^{13}C NMR spectroscopy. Prior to NMR
206 analysis, the coarse fractions were finely ground using a small pestle and mortar, and the fine
207 fractions were treated with 2% HF (Skjemstad et al., 1994) to remove paramagnetic matter
208 and concentrate organic C.
209
210 2.5. ^{13}C NMR spectroscopy

211 Spectra of ^{13}C NMR were acquired for the physically-fractionated coarse and fine fractions
212 using a Bruker 200 Avance 200 MHz spectrometer (Bruker Corporation, Billerica, MA, USA),
213 equipped with 4.7 T superconducting magnet operating at 50.33 MHz resonance (Baldock et
214 al. 2013b). Samples (approx. 100-300 mg) of the coarse and fine fractions were packed in
215 small zirconia rotors (7 mm diameter rotors fitted with Kel-F end caps) and spun at 5 KHz. For
216 a few coarse fractions insufficient material was accumulated to completely fill the rotors, so
217 1 or 2 Kel-F inserts were used. All samples were subjected to a cross-polarisation (CP) ^{13}C NMR
218 analysis in which 10,000 scans were acquired by applying a pulse of 3.2 μs, 195 W and 90°,
219 with 1 ms contact time and 1 s recycle decay time. The length of the recycle delay time was
220 confirmed to be more than 5 times the T_1H value of the samples calculated using an inversion
221 recovery pulse sequence. A variable spin lock experiment using various spin-lock times, 1 ms
222 contact time and 1 s recycle delay determined sample specific T1ρH values. Glycine was used

as an external intensity standard for comparing the observability of C in samples. Background signal correction was done by subtracting the signal intensity acquired for an empty rotor from the sample spectra. The NMR spectra were processed using the Bruker Topspin 3 software. Signal intensity found in the aryl (110-145 ppm) and O-aryl regions (145-165 ppm) of the NMR spectra (Fig. 2b) were used to calculate the proportions of lignin and ROC present in the coarse and fine fraction using the formula ROC = (aryl C – 1.77 *O-aryl C)/0.45 (Baldock et al. 2013b).

POC concentration was calculated by multiplying the TOC concentration of the coarse fraction by (1 - the proportion of ROC in the coarse fraction). Similarly HOC concentration was calculated by multiplying the TOC concentration of fine fraction by (1 – the proportion of ROC in the fine fraction). The recovery of TOC fractions was calculated as the percentage recovery of TOC using the formula, Recovery% = [(POC + HOC + ROC)/TOC]*100. The absolute deviation of recovery was calculated as the difference in the measured concentrations between TOC and the sum of its fractions, and expressed in g kg^{-1}.

2.6. Infrared-PLSR calibrations and chemometrics

The PLSR models were developed as described by Haaland and Thomas (1988) using the infrared spectra and measured data from 38 environmental plantings soils and 130 agricultural soil samples included from the Australian Soil Carbon Research Program (SCaRP, Baldock et al. 2013b). These prediction models were then used to predict TOC, POC, HOC and ROC in 3109 environmental plantings and agricultural soils.

246 Prior to multivariate analyses, the absorbance spectra were pre-processed using
247 multiplicative scatter correction (Martens and Naes, 1989) and mean centring in Matlab
248 R2013a (The MathWorks, Natick, MA, USA). Principal component analysis (PCA) and PLSR
249 analysis were performed using PLS_Toolbox 7.0 (Eigenvector Research Inc., Wenatchee, WA,
250 USA). The mid-and near-infrared spectra (MNIRS) used for PCA and PLSR included the MIR
251 spectral region (4000-450 cm^{-1}) corresponding to absorbance from fundamental bands of
252 molecular vibrations, and a portion of near-infrared (NIR) region (6000-4000 cm^{-1}) containing
253 overtones and combinations of fundamental bands but with less band specificity (Bellon-
254 Maurel and McBratney, 2011). Also, baseline offset pre-processing of the MIR 4000-450 cm^{-1}
255 region produced inconsistent results by offsetting either in the middle or at the either end of
256 spectral range, which was improved when using 6000-450 cm^{-1} MNIRS spectral range to
257 produce consistent offsetting around 6000-5500 cm^{-1}. Therefore spectral information from
258 both mid and near-infrared regions was used to improve and develop the MNIRS-PLSR
259 prediction models.
260
261 A square root transformation was used to normalise the distributions of TOC and its fraction
262 concentrations and minimise non-linearity in the resultant PLSR prediction algorithm (Janik
263 et al. 2007). A user-defined Matlab pre-processing function (*preprouser.m*) was used for the
264 data transformation and back-transformation processes – *calibrate* and *apply* (applying
265 square-root), and *undo* (squaring the model predictions). A PCA was performed using the pre-
266 processed soil spectra to be used to construct the PLSR prediction models to quantify their
267 spectral variability. This PCA model was used to ascertain spectral homogeneity and detect
268 potential outliers from the spectra acquired here for all reforested and agricultural soil
269 samples in the present study (Viscarra Rossel et al., 2008).

The predictive abilities of the MNIRS-PLSR models were evaluated using a range of statistical parameters; coefficient of determination (R^2), root mean square error (RMSE), bias, ratio of performance to deviation (RPD) and ratio of error range (RER) as commonly reported for assessing the quality of predicted soil attributes in spectroscopic studies (Bellon-Maurel and McBratney, 2011). These statistical analyses were made using PLS_Toolbox and reported in back-transformed values. R^2 indicates the model's fit, and root-mean-square error of calibration (RMSEC) and cross-validation (RMSECV) are measures of standard deviation of the residuals in calibration and in cross-validation respectively. Bias is the mean value of the difference between predicted and measured values. RPD is the ratio of standard deviation of measured values to standard error of prediction (Bellon-Maurel and McBratney, 2011). The RER is the ratio of the range of the predicted data (maximum – minimum) to RMSE. Generally, strong prediction models are expected to contain high values for R^2, minimum values for RMSE and RPD > 2 (Grinand et al., 2012), and RER > 10 (Williams and Sobering, 1996). The number of principal components (PCs) in a given PCA model, and factors or latent variables (LV) contributing to each PLSR model was restricted when the following PC or LV did not reduce the RMSE of calibration by more than 1%. The robustness of the MNIRS-PLSR models were also investigated by examining the variations in beta coefficients and Variable Importance in Projection (VIP) scores (Chong and Jun, 2005), which reveal the infrared bands and specific functional groups that contributed the most to the prediction models of TOC fractions, and helped to distinguish fraction-specific organic C composition at the functional-group level.

2.7. Prediction assessments

294 The reliability of predictions from MNIRS-PLSR models were evaluated by using *k*-nearest
295 neighbour (KNN) score distances (Sharaf et al., 1986), calculated in PLS_Toolbox software.
296 The KNN score distance is the distance to the nearest neighbour (shortest distance) in the
297 three-dimensional score space of the samples (Ripley, 1996), and scalar value '*k*' indicates the
298 number of neighbours to which distance should be calculated and averaged over. The scalar
299 was set to *k*=1 as described in ASTM D6122-06 Standard Practice for Validation of the
300 Performance of Multivariate Process Infrared Spectrophotometers, (ASTM International,
301 West Conshohocken, PA, USA). The scalar setting was performed in Matlab using the
302 command: *setplspref ('plotscores','knnscoredistance',1)*. The maximum KNN score distance
303 observed for samples in the calibration is called the inlier limit. Prediction samples outside
304 the inlier limit suggest that they fall within a sparsely populated region of the calibration space
305 (ATSM International, 2006). Classification and neighbour selection studies generally use a
306 Mahalanobis distance of 3 or more to determine non-members or spectral outliers (e.g., Gogé
307 et al., 2012). As KNN score distance and Mahalanobis distance are highly correlated (Wise et
308 al., 2006), predicted samples with higher KNN score distance, i.e. more than 3 times of inlier
309 limit were considered as having unreliable predictions. Hotelling's T^2 represents the measure
310 of the variation in each sample within the model and indicates how far each sample is from
311 the center of the model. Hotelling T^2 contribution was used to reveal how individual variables
312 (spectral wavenumbers) contributed to unusual variation in predicted outlier samples as
313 determined by inlier test (Kourti, 2005).
314
315 **3. Results and Discussion**
316 3.1. Soil TOC fractions from environmental plantings

317 The concentration ranges of TOC and its fractions in the environmental plantings soils used
318 for fractionation are provided in Table 1. The recovery of TOC after fractionation was similar
319 to that obtained by Baldock et al. (2013b) and ranged between 87 and 112%, with a mean
320 recovery of 94%, and a standard deviation of 5.6%. The absolute deviation of recovery of TOC
321 fractions indicated that for > 90% of the samples, TOC recovery was within ± 2 g kg^{-1} C with
322 the variation in absolute recovery increasing with increasing TOC concentration. When
323 expressed as percentage of TOC, POC ranged between 8.1% and 54.0% (mean ± SD = 25.8 ±
324 12.5%), with 35.5% and 16.9% in the 0-5 cm and 5-10 cm layers, respectively. HOC ranged
325 between 30.3 and 72.5% (mean ± SD = 52.1 ± 10.8%), and ROC ranged between 15.5 and
326 31.7% (mean ± SD = 20.4 ± 3.44%). By comparison, in agricultural soils representing the range
327 of soil types and climate across the productive land in Australia (organic C ranging from 10.0
328 to 91.0 g kg^{-1} soil), the mean proportions were 19.2 ± 12.3% for POC, 56.1 ± 14.5% for HOC
329 and 26.2 ± 9.6% for ROC in TOC (Baldock et al. 2013b).
330
331 The concentrations of TOC fractions were correlated with soil TOC concentration (R^2 = 0.83
332 for POC; R^2= 0.73 for HOC; R^2 = 0.84 for ROC, $P < 0.05$). Relationship slopes of 0.55, 0.30 and
333 0.16 for POC, HOC and ROC, respectively, were higher for POC, lower for HOC and similar for
334 ROC in 0-10 cm environmental plantings soils when compared to those obtained by Baldock
335 et al. (2013b) for agricultural soils having similar TOC concentrations (1.20-95.0 g kg^{-1}). These
336 results suggest that on average soils under environmental plantings contain more POC and
337 less HOC compared to agricultural soils.
338
339 3.2. PCA of samples

340 The first three PCs of the PCA of spectra acquired for the environmental plantings and the
341 agricultural soils explained 94% of the model's variation. The distribution of PC scores
342 indicated that the spectra from environmental planting soils differed from those of the
343 agriculture soils, largely due to differences in the PC2 and PC3 scores (Fig. 3a). The PC1 loading
344 spectra which accounted for 69.7% of spectral variance showed absorbance intensity mostly
345 from soil mineral components (Fig. 3b). In contrast, PC2 (19% of spectral variance) and PC3
346 (4.1% of spectral variance) showed distinct features associated with lignin and polysaccharide
347 molecules with signals from aliphatic C-H stretch at 2930, 2858 and 1484 cm^{-1}, carboxylic acid
348 -COOH stretch at 1710 cm^{-1} and C-O stretch at 1224 cm^{-1}, amide I C=O stretch at 1666 cm^{-1},
349 and amide II C-N stretch and potentially aromatic C=C stretch at 1556 cm^{-1} that differentiated
350 the environmental planting soils from the agricultural soils (Table 2). Projection of PC1, PC2
351 and PC3 scores for all other soil samples collected in this project (*n* = 3109) onto the 3D scores
352 plot of the calibration samples (Fig. 3c) showed a large cluster of samples within a three
353 dimensional space defined by the calibration samples, indicating the similarity of spectra in
354 both sets. A few samples were distributed away from the main cluster along the negative axis
355 of PC1 and PC2 scores. These 16 samples (upper 5%) originated from a site having relatively
356 high TOC concentration (> 150 g kg^{-1} soil) and had regular spectra with spectral features
357 similar to other spectra collected here, and were therefore retained and included in
358 subsequent analyses.

359 3.3. MNIRS-PLSR models and cross-validations

360 The predictive ability of MNIRS-PLSR models was tested using leave-one-out cross-validation
361 (Table 3). Previous studies have used 10 to 13 latent variables for TOC fraction models
362 (e.g., Janik et al. 2007) but the PLSR models developed here required only 6 or 7 latent

363 variables to explain the maximum variation in spectra and the measured analytical data (TOC
364 and its fractions). Scatter plots of measured and predicted data showed strong linear
365 relationships for all prediction models (R^2 > 0.92). The PLSR prediction model for TOC in the
366 range of 0.20 to 100 g kg^{-1} resulted in an RMSE of 2.68, negligible bias and an RPD of 5.59.
367 Similarly, the PLSR models developed for TOC fractions showed RMSE values of approximately
368 5% of the data range used in their derivation, negligible bias and RPD values > 3.42.
369
370 The results obtained from cross-validation for TOC and its fractions were similar to calibration.
371 Cross-validation of TOC resulted in excellent prediction (R^2 = 0.95,
372 bias = -0.07, RMSE = 3.04 and RPD = 4.84), consistent with other reports for soils (e.g., Janik
373 et al., 2007; Viscarra Rossel et al., 2008). Similar predictions (R^2 = 0.95) have been found for
374 TOC in Mediterranean soils with a comparable TOC range (5-150 g kg^{-1}, D'Acqui et al. 2010).
375 Janik et al., 2007) reported POC calibrations under native vegetation of Australia with R^2 =
376 0.71 (range 0.20-16.8 g kg^{-1}) and suggested that the variations in the chemistry of plant inputs
377 may influence the development of accurate generalised prediction models. Here, POC
378 predictions within the range of 0.20-16.8 g kg^{-1} were relatively stronger for both calibration
379 (R^2 = 0.88) and cross-validation (R^2 = 0.86). This is consistent with the results reported by
380 (Baldock et al., 2013a). The RPD values for TOC and its fractions derived in this study were all
381 > 3. Although RPD in cross-validation results were slightly lower than the prediction model,
382 all values were > 2.89, which indicated excellent predictive ability of models, as suggested by
383 Grinand et al. (2012).
384
385 3.4. Functional group analysis of models

386 Beta coefficients of the derived PLSR prediction models plotted against the wavenumbers
387 illustrate the major functional groups and relative magnitude of their contributions to each
388 prediction model (Fig. 4). For TOC, the dominant organic peaks were aliphatic C-H stretch
389 (2930 and 2878 cm^{-1}), carboxylic -COOH stretch (1708 cm^{-1}), carbonyls and amide I and II C=O
390 or C=N stretch (1666 and 1568 cm^{-1}), aliphatic C-H deformations (1440 cm^{-1}), carboxylic acid
391 (1204 cm^{-1}) and carbohydrates C-O or –COH stretch (1080 cm^{-1}). Janik et al. (2007) reported
392 similar spectral features for their TOC model in average beta coefficients. For POC, in addition
393 to some features noted for TOC, strong peaks of amides (C=O stretch, 1666 cm^{-1}),
394 polysaccharides and silicate absorbance (C-O and Si-O stretch, 1200-1080 cm^{-1}), and distinct
395 peaks of aromatic skeletal C=C stretching vibrations (1520 cm^{-1}) and aromatic C=C stretch
396 (1598 cm^{-1}) were evident, indicating functional groups from proteins and carbohydrates. The
397 correlation of beta coefficients derived for TOC with those derived for POC and ROC (R^2 ~
398 0.50) showed significant scatter indicating differences in spectral features and magnitudes
399 used to predict these components and suggested a degree of independence between the
400 derived prediction models (Fig. S1). The beta coefficients of TOC and HOC were highly
401 correlated (R^2 = 0.75) as reported by Baldock et al., (2013a). However, the beta coefficient for
402 HOC had subtle differences in magnitudes and shifts in peaks, with higher contributions from
403 carboxylic acids (1740 and 1240 cm^{-1}). For ROC, most of the regions differed from that of other
404 fractions, such as absence of aliphatic C-H (3000-2800 cm^{-1}) and phenols (~1380 cm^{-1})
405 ascribed to lignin, and increased intensities from aromatic C=C (1570 cm^{-1}), -COOH (1708 cm^{-1})
406 and C-O (1228 cm^{-1}) from carboxylic acid, aromatic C-H stretch (840 cm^{-1}) and kaolinite
407 (3626 cm^{-1}).
408

409 Peaks with VIP scores > 1 significantly contributed to the model, while wavenumber < 1 was
410 less important (Chong and Jun, 2005). Here, VIP scores in the MNIRS-PLSR models included C-
411 H (2930 and 2878 cm^{-1}) and C=O (1666 cm^{-1}) for TOC and all fractions (VIP score > 2.8);
412 carboxylic acid C-O stretch, O-H deformation, ester and phenol C-O (1280-1200 cm^{-1}),
413 carbohydrate (1160-1080 cm^{-1}), and aromatic and mineral absorbances (< 1000 cm^{-1}) for POC
414 (VIP score > 4); and aromatic C=C absorbance (1568 cm^{-1}) for ROC (VIP score = 3.85) (Fig. S2).
415 These functional groups were similar to the inferences from beta coefficients specific to TOC
416 and its fractions in their models.
417
418 3.5. Predictions of TOC, POC, HOC and ROC contents
419 Using the leave-one-out cross-validated MNIRS-PLSR prediction models, TOC and its fractions
420 were predicted in 3109 soils from environmental plantings and reference agricultural lands.
421 The KNN score distance for the predicted samples indicated that > 75% of the predictions fell
422 within the inlier limit (TOC-88.2%, POC-84.8%, HOC-75.8% and ROC-84.5%), while the
423 remaining samples were within the satisfactory limits of 3 x inlier limit, except for a few
424 samples (TOC = 4, POC = 3, HOC = 6, ROC = 5, Fig. S3). KNN score distance is higher for samples
425 that fall in lower-populated regions of score space (Wise et al., 2006). Predicted samples with
426 high C content that were scattered in the PCA were within the inlier limits, so their inclusion
427 was reasonable.
428
429 Hotelling T^2 described > 96% of sample distribution within all models, hence it was useful in
430 identifying the spectral regions contributing to unusual variations in the samples containing
431 highest KNN scores (Wise et al., 2006). Hotelling T^2 contributions for outlier samples showed

increased intensities from mineral bands at 4000-3500 and 1080-450 cm^{-1} and inorganic C and quartz bands at 2200-1800 cm^{-1} (Fig. S4). These results were consistent with dominant site characteristics, i.e. a site with sandy texture from Western Australia and another site containing inorganic C from Victoria.

3.6. Application of the models

An example of the application of the MNIRS-PLSR models indicates that concentrations of POC and ROC in 0-5 cm soil (*n* = 79) were on-average greater under environmental plantings (_ and _ g kg^{-1} respectively) than under agricultural land-use (_ and _ g kg^{-1} respectively), but that there is much variability from site to site (Fig. S5). A more detailed analysis of differences in TOC and in MNIRS-PLSR predicted values of TOC fractions between environmental planting and agriculture soils was reported in England et al. (2016).

The site to site variability demands that at least for landscape level studies and C accounting rapid analytical methodologies such as developed here are essential. While the models developed here were satisfactorily cross-validated, they remain specific calibrations for the environments (climatic, edaphic) and vegetation from which the soils were sampled, notwithstanding that the samples were from a range of environments across southern and eastern Australia. We might expect that marked differences in climate (e.g arid zones), in soil parent material and soil mineralogy, in properties of plant organic matter added to the soil, and in management practices (e.g. tillage, prescribed burning) might challenge the model calibrations. While models for TOC seem relatively robust, those for the POC and HOC at least are less so and likely will always require very specific local based calibrations.

4. Conclusions

We developed accurate MNIRS-PLSR predictive models for TOC and TOC fractions under reforested environments for application on a national/continental scale. The study demonstrated the potential of MNIRS-PLSR technique to rapidly and cost-effectively predict soil TOC and its fractions and thereby assess the magnitude of changes following reforestation. The technique has the potential for use in regional to global carbon accounting in reforested ecosystems. There is a need to build a broad database of soil TOC and TOC fractions for a wider range of climate, soils, vegetation and management to enable local to general calibration and validation of models.

Acknowledgements

The study was funded by the Australian Government's Filling the Research Gap program. TRC is supported by an Australian Research Council (ARC) Future Fellowship (FT120100463). MPP was partially supported through an ARC Laureate Fellowship to R. Hobbs. We acknowledge M. Bullock, T. Fairman, R. Greene, P. Holliday, R. Law, J. McGowan, T. Morald, T. Murshed, A. Schapel, G. Szegedy and A. Wherrett, for provision of soil samples and field and laboratory support.

References

476 Baldock, J.A., Hawke, B., Sanderman, J., Macdonald, L.M., 2013a. Predicting contents of
477 carbon and its component fractions in Australian soils from diffuse reflectance mid-
478 infrared spectra. Soil Res. 51, 577-595.
479 Baldock, J.A., Sanderman, J., Macdonald, L.M., Puccini, A., Hawke, B., Szarvas, S., McGowan,
480 J., 2013b. Quantifying the allocation of soil organic carbon to biologically significant
481 fractions. Soil Res. 51, 561-576.
482 Bellon-Maurel, V., McBratney, A., 2011. Near-infrared (NIR) and mid-infrared (MIR)
483 spectroscopic techniques for assessing the amount of carbon stock in soils – Critical
484 review and research perspectives. Soil Biol. Biochem. 43, 1398-1410.
485 Berthrong, S.T., Piñeiro, G., Jobbágy, E.G., Jackson, R.B., 2012. Soil C and N changes with
486 afforestation of grasslands across gradients of precipitation and plantation age. Ecol.
487 Appl. 22, 76-86.
488 Bornemann, L., Welp, G., Amelung, W., 2010. Particulate organic matter at the field scale:
489 Rapid acquisition using mid-infrared spectroscopy. Soil Sci. Soc. Am. J. 74, 1147-1156.
490 Brack, C.L., Richards, G.P., 2002. Carbon accounting model for forests in Australia. Environ.
491 Pollut. 116 Suppl 1, S187-194.
492 Canadell, J.G., Raupach, M.R., 2008. Managing forests for climate change mitigation. Science
493 320, 1456-1457.
494 Chong, I.G., Jun, C.H., 2005. Performance of some variable selection methods when
495 multicollinearity is present. Chemometr. Intell. Lab. 78, 103-112.
496 Cunningham, S.C., Cavagnaro, T.R., Mac Nally, R., Paul, K.I., Baker, P.J., Beringer, J., Thomson,
497 J.R., Thompson, R.M., 2015a. Reforestation with native mixed-species plantings in a
498 temperate continental climate effectively sequesters and stabilizes carbon within
499 decades. Glob. Change Biol. 21, 1552-1566.

500 Cunningham, S.C., Mac Nally, R., Baker, P.J., Cavagnaro, T.R., Beringer, J., Thomson, J.R.,
501 Thompson, R.M., 2015b. Balancing the environmental benefits of reforestation in
502 agricultural regions. Perspect. Plant Ecol. Evol. and Syst. 17, 301-317.
503 D'Acqui, L.P., Pucci, A., Janik, L.J., 2010. Soil properties prediction of western Mediterranean
504 islands with similar climatic environments by means of mid-infrared diffuse reflectance
505 spectroscopy. Eur. J. Soil Sci. 61, 865-876.
506 Del Galdo, I., Six, J., Peressotti, A., Francesca Cotrufo, M., 2003. Assessing the impact of land-
507 use change on soil C sequestration in agricultural soils by means of organic matter
508 fractionation and stable C isotopes. Glob. Change Biol. 9, 1204-1213.
509 England, J.R., Paul, K.I., Cunningham, S.C., Madhavan, D.B., Baker, T.G., Read, Z., Wilson, B.R.,
510 Cavagnaro, T.R., Lewis, T., Perring, M.P., Herrmann, T., Polglase, P.J., 2016. Previous
511 land use and climate influence differences in soil organic carbon following reforestation
512 of agricultural land with mixed-species plantings. Agric. Ecosyst. Environ. 227, 61-72.
513 FAO, 2010. Global Forest Resources Assessment 2010, Main Report Rome.
514 Gogé, F., Joffre, R., Jolivet, C., Ross, I., Ranjard, L., 2012. Optimization criteria in sample
515 selection step of local regression for quantitative analysis of large soil NIRS database.
516 Chemometr. Intell. Lab. 110, 168-176.
517 Grinand, C., Barthes, B.G., Brunet, D., Kouakoua, E., Arrouays, D., Jolivet, C., Caria, G.,
518 Bernoux, M., 2012. Prediction of soil organic and inorganic carbon contents at a national
519 scale (France) using mid-infrared reflectance spectroscopy (MIRS). Eur. J. Soil Sci. 63,
520 141-151.
521 Haberhauer, G., Rafferty, B., Strebl, F., Gerzabek, M.H., 1998. Comparison of the composition
522 of forest soil litter derived from three different sites at various decompositional stages
523 using FTIR spectroscopy. Geoderma 83, 331-342.

524 Janik, L.J., Skjemstad, J.O., Shepherd, K.D., Spouncer, L.R., 2007. The prediction of soil carbon
525 fractions using mid-infrared-partial least square analysis. Aust. J. Soil Res. 45, 73-81.
526 Jenkinson, D.S., Coleman, K., 1994. Calculating the annual input of organic matter to soil from
527 measurements of total organic matter and radiocarbon. Eur. J. Soil Sci. 45, 167-174.
528 Jenkinson, D.S., Harkness, D.D., Vance, E.D., Adams, D.E., Harrison, A.F., 1992. Calculating net
529 primary production and annual input of organic matter to soil from the amount and
530 radiocarbon content of soil organic matter. Soil Biol. Biochem. 24, 295-308.
531 Karunaratne, S.B., Bishop, T.F.A., Baldock, J.A., Odeh, I.O.A., 2014. Catchment scale mapping
532 of measureable soil organic carbon fractions. Geoderma 219–220, 14-23.
533 Kourti, T., 2005. Application of latent variable methods to process control and multivariate
534 statistical process control in industry. Int. J. Adapt. Control Signal Process. 19, 213-246.
535 Lal, R., 2005. Forest soils and carbon sequestration. For. Ecol. Manage. 220, 242-258.
536 Ludwig, B., Nitschke, R., Terhoeven-Urselmans, T., Michel, K., Flessa, H., 2008. Use of mid-
537 infrared spectroscopy in the diffuse-reflectance mode for the prediction of the
538 composition of organic matter in soil and litter. J. Plant Nutr. Soil Science-Zeitschrift Fur
539 Pflanzenernahrung Und Bodenkunde 171, 384-391.
540 Madari, B.E., Reeves Iii, J.B., Machado, P.L.O.A., Guimarães, C.M., Torres, E., McCarty, G.W.,
541 2006. Mid- and near-infrared spectroscopic assessment of soil compositional
542 parameters and structural indices in two Ferralsols. Geoderma 136, 245-259.
543 Madhavan, D.B., Kitching, M., Mendham, D.S., Weston, C.J., Baker, T.G., 2016. Mid-infrared
544 spectroscopy for rapid assessment of soil properties after land use change from
545 pastures to Eucalyptus globulus plantations. J. Environ. Manage. 175, 67-75.
546 Martens, H., Naes, T., 1989. Multivariate calibration. John Wiley & Sons, New York, USA.

547 Merry, R.H., Spouncer, L.R., 1988. The measurement of carbon in soils using a microprocessor-
548 controlled resistance furnace. Commun. Soil Sci. Plant Anal. 19, 707-720.
549 Nelson, D.W., Sommers, L.E., 1996. Total carbon, organic carbon, and organic matter, in:
550 Sparks, D.L. (Ed.), Methods of Soil Analysis. Part 3 - Chemical Methods. Soil Science
551 Society of America Inc. and American Society of Agronomy Inc., Madison, Wisconsin,
552 USA, pp. 961-1010.
553 Nguyen, T.T., Janik, L.J., Raupach, M., 1991. Diffuse Reflectance Infrared Fourier-Transform
554 (Drift) Spectroscopy in Soil Studies. Aust. J. Soil Res. 29, 49-67.
555 Parton, W.J., Schimel, D.S., Cole, C.V., Ojima, D.S., 1987. Analysis of Factors Controlling Soil
556 Organic Matter Levels in Great Plains Grasslands. Soil Sci. Soc. Am. J. 51, 1173-1179.
557 Paul, K.I., Reeson, A., Polglase, P., Crossman, N., Freudenberger, D., Hawkins, C., 2013.
558 Economic and employment implications of a carbon market for integrated farm forestry
559 and biodiverse environmental plantings. Land Use Policy 30, 496-506.
560 Paul, K.I., Roxburgh, S.H., England, J.R., de Ligt, R., Larmour, J.S., Brooksbank, K., Murphy, S.,
561 Ritson, P., Hobbs, T., Lewis, T., Preece, N.D., Cunningham, S.C., Read, Z., Clifford, D.,
562 John Raison, R., 2015. Improved models for estimating temporal changes in carbon
563 sequestration in above-ground biomass of mixed-species environmental plantings. For.
564 Ecol. Manage. 338, 208-218.
565 Perring, M.P., Jonson, J., Freudenberger, D., Campbell, R., Rooney, M., Hobbs, R.J., Standish,
566 R.J., 2015. Soil-vegetation type, stem density and species richness influence biomass of
567 restored woodland in south-western Australia. For. Ecol. Manage. 344, 53-62.
568 Rabbi, S.M.F., Tighe, M., Cowie, A., Wilson, B.R., Schwenke, G., McLeod, M., Badgery, W.,
569 Baldock, J., 2014. The relationships between land uses, soil management practices, and
570 soil carbon fractions in South Eastern Australia. Agric. Ecosyst. Environ. 197, 41-52.

571 Rayment, G.E., Lyon, D.J., 2011. Soil Chemical Methods - Australasia. CSIRO Publishing,
572 Melbourne, Victoria, Australia.
573 Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press,
574 Cambridge
575 Schlesinger, W.H., 1986. Changes in soil carbon storage and associated properties with
576 disturbance and recovery, in: Trabalka, J.R., Reichle, D.E. (Eds.), The Changing Carbon
577 Cycle. Springer, New York, pp. 194-220.
578 Sharaf, M.A., Illman, D.L., Kowalski, B.R., 1986. Chemometrics. John Wiley & Sons, New York.
579 Skjemstad, J., Clarke, P., Taylor, J., Oades, J., Newman, R., 1994. The removal of magnetic
580 materials from surface soils - a solid state 13C CP/MAS NMR study. Soil Res. 32, 1215-
581 1229.
582 Skjemstad, J.O., Spouncer, L.R., Cowie, B., Swift, R.S., 2004. Calibration of the Rothamsted
583 organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon
584 pools. Aust. J. Soil Res. 42, 79-88.
585 Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., MacDonald, L.M., McLaughlin, M.J., 2014.
586 The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for
587 Prediction of Soil Physical, Chemical, and Biological Properties. Appl. Spectrosc. Rev. 49,
588 139-186.
589 Viscarra Rossel, R.A., Jeon, Y.S., Odeh, I.O.A., McBratney, A.B., 2008. Using a legacy soil sample
590 to develop a mid-IR spectral library. Aust. J. Soil Res. 46, 1-16.
591 von Lützow, M., Kögel-Knabner, I., Ekschmitt, K., Flessa, H., Guggenberger, G., Matzner, E.,
592 Marschner, B., 2007. SOM fractionation methods: Relevance to functional pools and to
593 stabilization mechanisms. Soil Biol. Biochem. 39, 2183-2207.

Walkley, A., Black, I.A., 1934. An Examination of the Degtjareff Method for Determining Soil Organic Matter, and A Proposed Modification of the Chromic Acid Titration Method. Soil Sci. 37, 29-38.

Williams, P.C., Sobering, D.C., 1996. How do we do it?: a brief summary of the methods we use in developing near infrared calibrations. NIR Publications, Chichester, UK.

Wise, B.M., Gallagher, N.B., Bro, R., Shaver, J.M., Windig, W., Koch, R.S., 2006. Chemometric Tutorial for PLS_Toolbox and Solo. Eigenvector Research Inc., Wenatchee, WA, USA.

Zimmermann, M., Leifeld, J., Fuhrer, J., 2007. Quantifying soil organic carbon fractions by infrared-spectroscopy. Soil Biol. Biochem. 39, 224-231.

604 **Table 1.** Concentration range, mean and standard error (SE) of soil (0-5, 5-10 cm) TOC and

605 TOC fractions in 19 environmental plantings sites

Depth (cm)	($g\ kg^{-1}$ soil)											
	TOC			POC			HOC			ROC		
	Range	Mean	SE	Range	Mean	SE	Range	Mean	SE	Range	Mean	SE
0-5	26.2-59.9	40.9	2.5	5.6-29.5	14.8	1.8	11.8-25.1	17.2	0.9	4.8-10.7	7.8	0.4
5-10	15.1-29.8	20.4	1.0	1.4-6.6	3.0	0.3	8.0-19.2	11.1	0.8	2.9-6.9	4.4	0.3

607

608

609 **Table 2**. Major peaks in the soil infrared spectra and loadings and their functional group

610 assignments

Wavenumber (cm^{-1})	Functional group	Vibration	References
3696, 3630	Kaolinite, montmorillonite and illite	O-H stretching	Viscarra Rossel et al., 2008
2930, 2878	Alkyl	C-H stretching	Janik et al., 2007
2000, 1870, 1784	Quartz	Si-O stretching	Nguyen et al., 1991
1740 to 1710	Carboxylic acid	-COOH stretching	Janik et al., 2007
1666	Carbonyls of amide-I	C=O stretching	Janik et al., 2007; Bornemann et al., 2010
1610 to 1590	Aromatic; carboxylates	C=C stretching; COO- symmetric stretching	Haberhauer et al., 1998
1568	Amide II	C=N stretching	Janik et al., 2007; Ludwig et al., 2008
1520 to 1510	Aromatic	C=C stretching	Haberhauer et al., 1998
1484 to 1440	Alkyl	C-H deformation	Janik et al., 2007; Viscarra Rossel et al., 2008
1380 to 1360	Phenol	C-O stretching, O-H deformation	Bornemann et al., 2010
1280-1200	Carboxylic acid and phenol; ester	-COOH and C-O stretching; O-H deformation	Bornemann et al., 2010
1200 to 950	Carbohydrates; alumina-silicate of clay minerals	C-O, -COH and C-O-C stretching; Al-O-Si, Si-O-Si and Al-O-Al lattice vibrations	Ludwig et al., 2008; Bornemann et al., 2010;
915, 816, 712,	Aromatics; clay minerals and quartz	C-H deformation; O-H deformation, Si-O stretching	Nguyen et al., 1991; Haberhauer et al., 1998

611

612

613 **Table 3.** Statistics for MNIRS-PLSR calibrations (Cal) and leave-one-out cross-validations (CV)
614 for TOC, POC, HOC and ROC. Samples (*n*) included 38 environmental planting soils and 53-130
615 agricultural soils. Spectra (6000-450 cm^{-1}) were multiplicative scatter corrected and mean
616 centered prior to PLSR analysis. RMSE = Root mean square error, RPD = Ratio of performance
617 to deviation and RER = Ratio of error range.

618

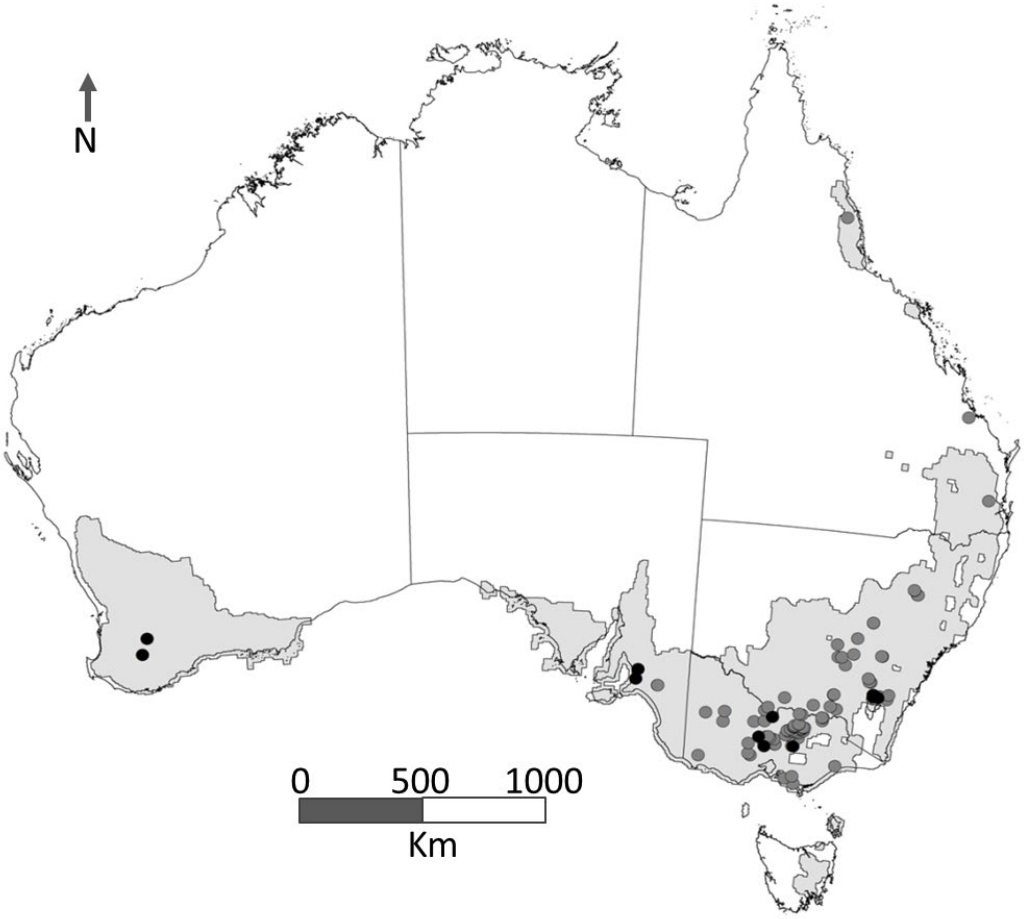
	TOC		POC		HOC		ROC	
n	153		91		154		168	
Minimum (g kg^{-1} soil)	2.1		1.3		1.4		0.4	
Maximum (g kg^{-1} soil)	90.2		30.5		34.5		17.8	
Mean (g kg^{-1} soil)	23.5		10.0		11.4		5.63	
Standard deviation	14.9		5.7		6.9		3.3	
Latent variables	6		7		6		7	
	Cal	CV	Cal	CV	Cal	CV	Cal	CV
Slope	0.95	0.93	0.88	0.86	0.91	0.89	0.91	0.89
Intercept	1.17	1.55	0.96	1.38	0.95	1.10	0.46	0.58
R^2	0.97	0.96	0.92	0.88	0.93	0.92	0.92	0.90
RMSE	2.68	3.04	1.67	1.97	1.75	1.97	0.93	1.05
Bias	-0.08	-0.07	-0.07	-0.05	-0.07	-0.06	-0.04	-0.04
RPD	5.59	4.84	3.42	2.89	3.90	3.47	3.57	3.15
RER	32.8	28.4	17.4	14.8	18.9	17.3	18.7	16.6

619

620

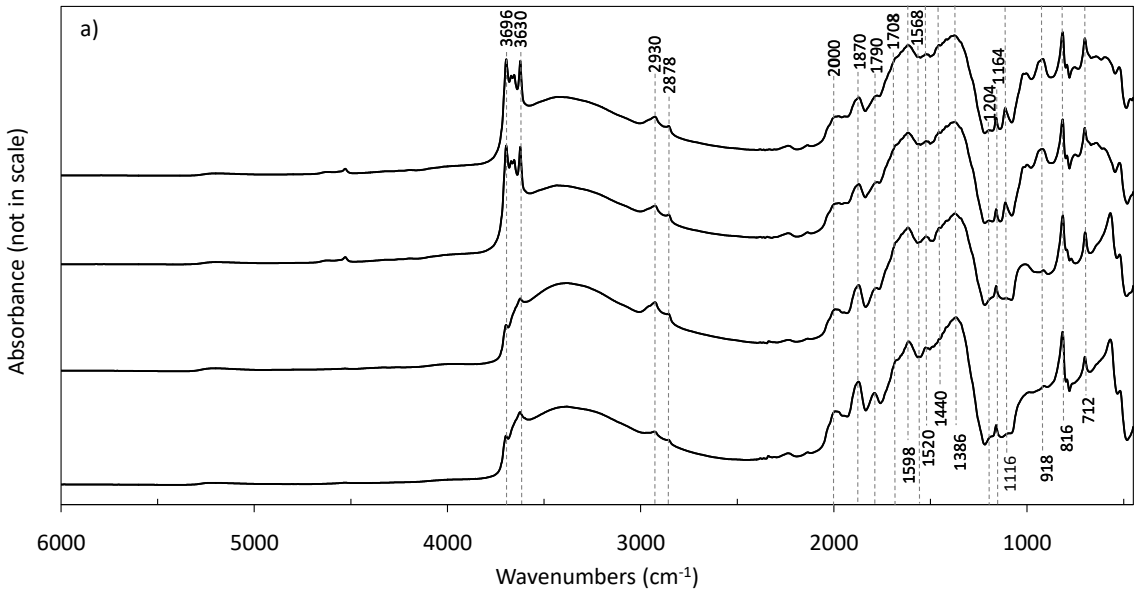
621

622 **Fig. 1.** Distribution of environmental planting sites and adjacent agricultural references sites
623 (*n* = 117, all circles). Soils from a subset of these plantings (*n* = 19, black circles) were physically
624 fractionated. The shaded area represent the geographical regions of application of calibration
625 for the environmental planting study area (after Paul et al., 2015).



626

Fig. 2. Representative a) infrared spectra (6000-450 cm^{-1}) acquired from finely-ground soils and b) ^{13}C NMR spectra acquired from coarse (2000-50 µm) and fine fractions (< 50 µm) of environmental planting soils.



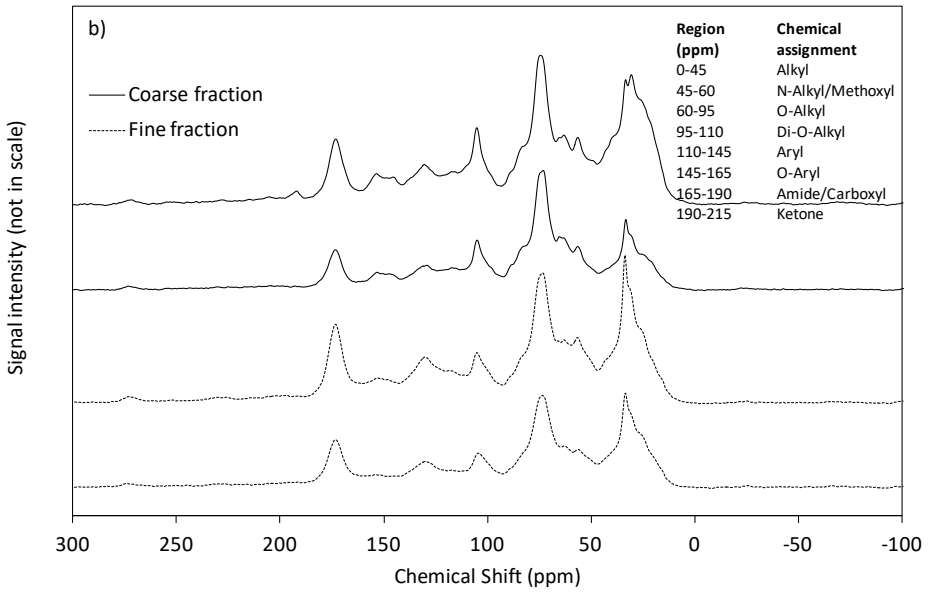
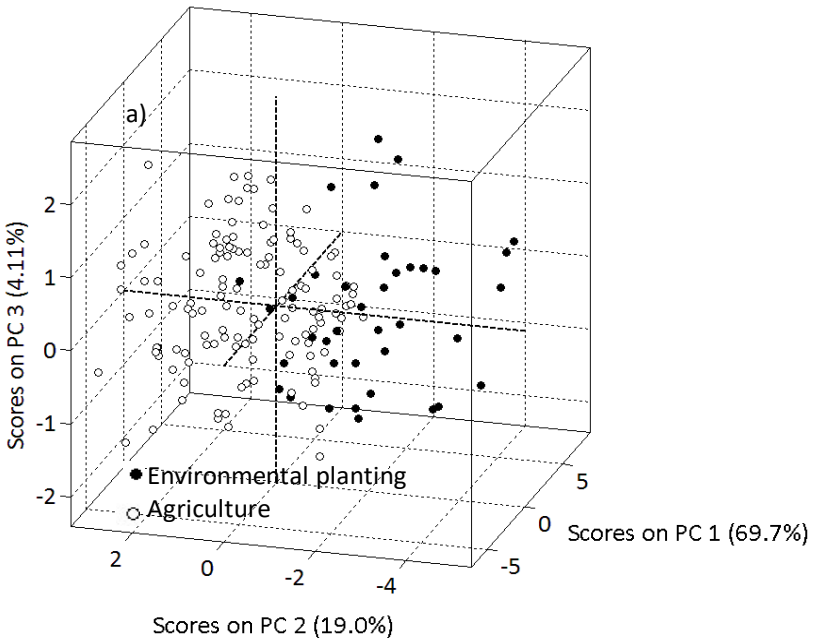
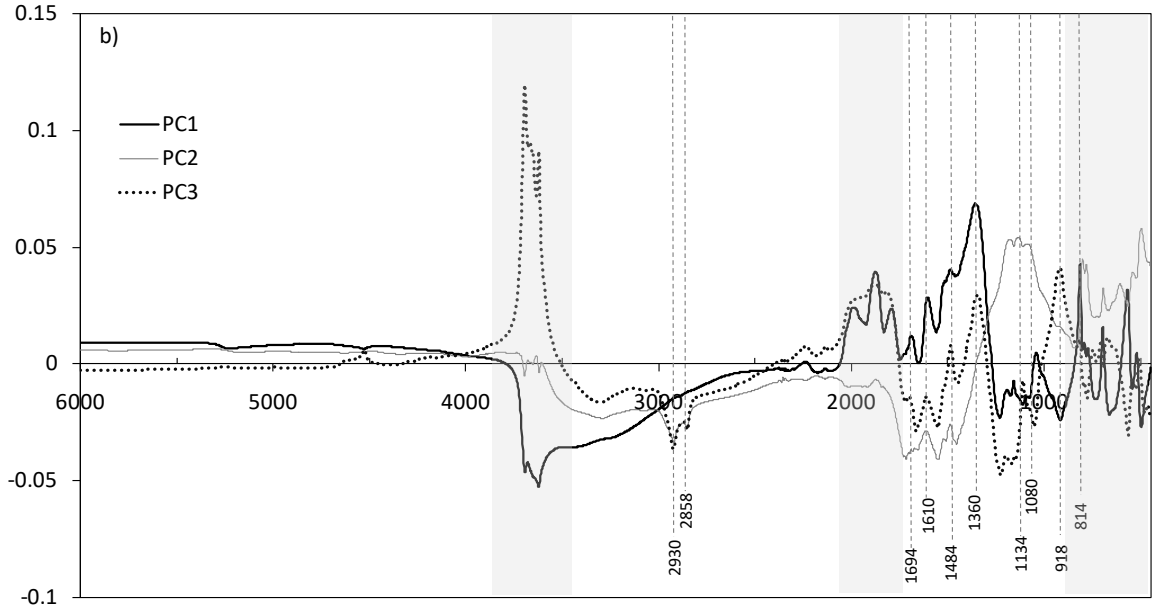
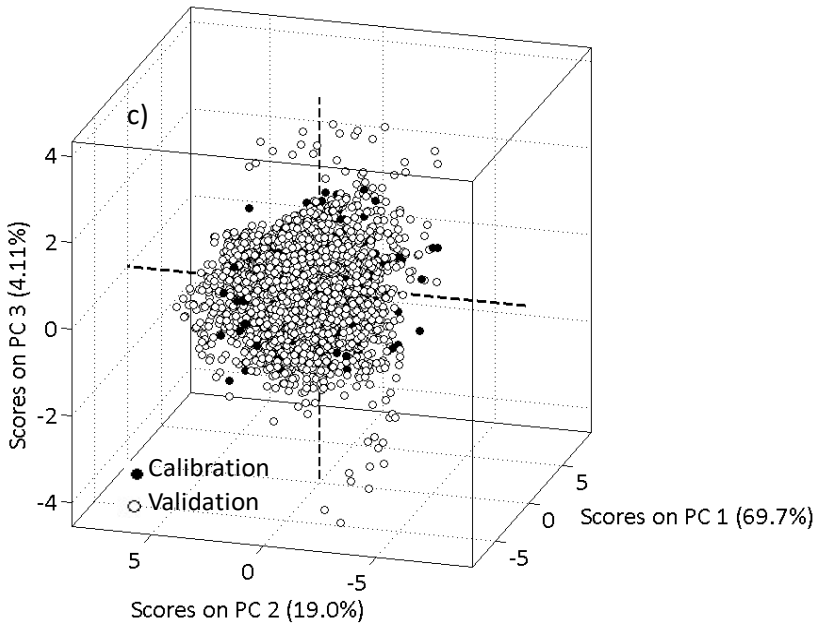


Fig. 3. Principal components analyses of near- and mid-infrared spectra (6000-450 cm^{-1}) for soils: (**a**) 3D scores plot for fractionated soils used to develop PLSR prediction models (38 environmental planting soils, present study; 130 agricultural soils, Baldock et al. 2013a); (**b**) Principal component (PC) loadings against wavenumbers (cm^{-1}) for the fractionated soils showing major organic bands in broken lines and regions dominated by mineral bands in grey; and (**c**) 3D scores plot for all soils (calibration, *n* = 168, predicted, *n* = 3109). For the 3D scores plots, values in parentheses are percentage of spectral variance explained by each component.







640

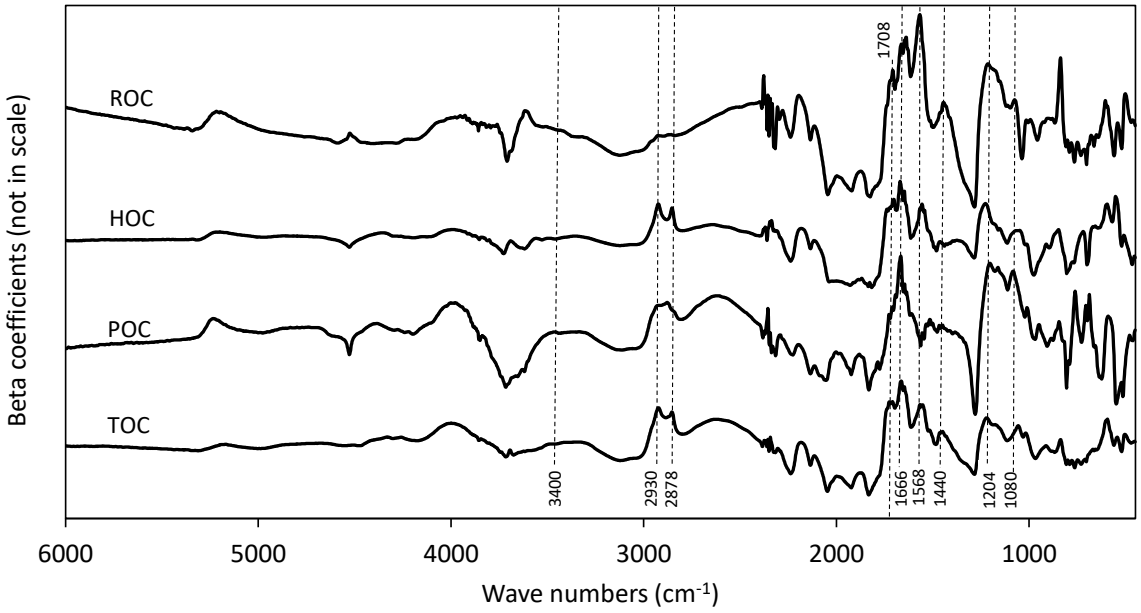
641

642

643

644 **Fig. 4.** Beta coefficients from MNIRS-PLSR prediction models for TOC, POC, HOC and ROC

645 showing functional group-specific peaks (dotted lines)



646

647

Supplementary material

Fig. S1. Correlations of TOC beta coefficients and those obtained for (**a**) POC, (**b**) HOC and (**c**) ROC in the MNIRS-PLSR models ($P < 0.001$). The magnitude of scatter in each figure indicates the degree of independence between models, i.e. more scatter indicates more difference in spectral features contributing to respective prediction model.

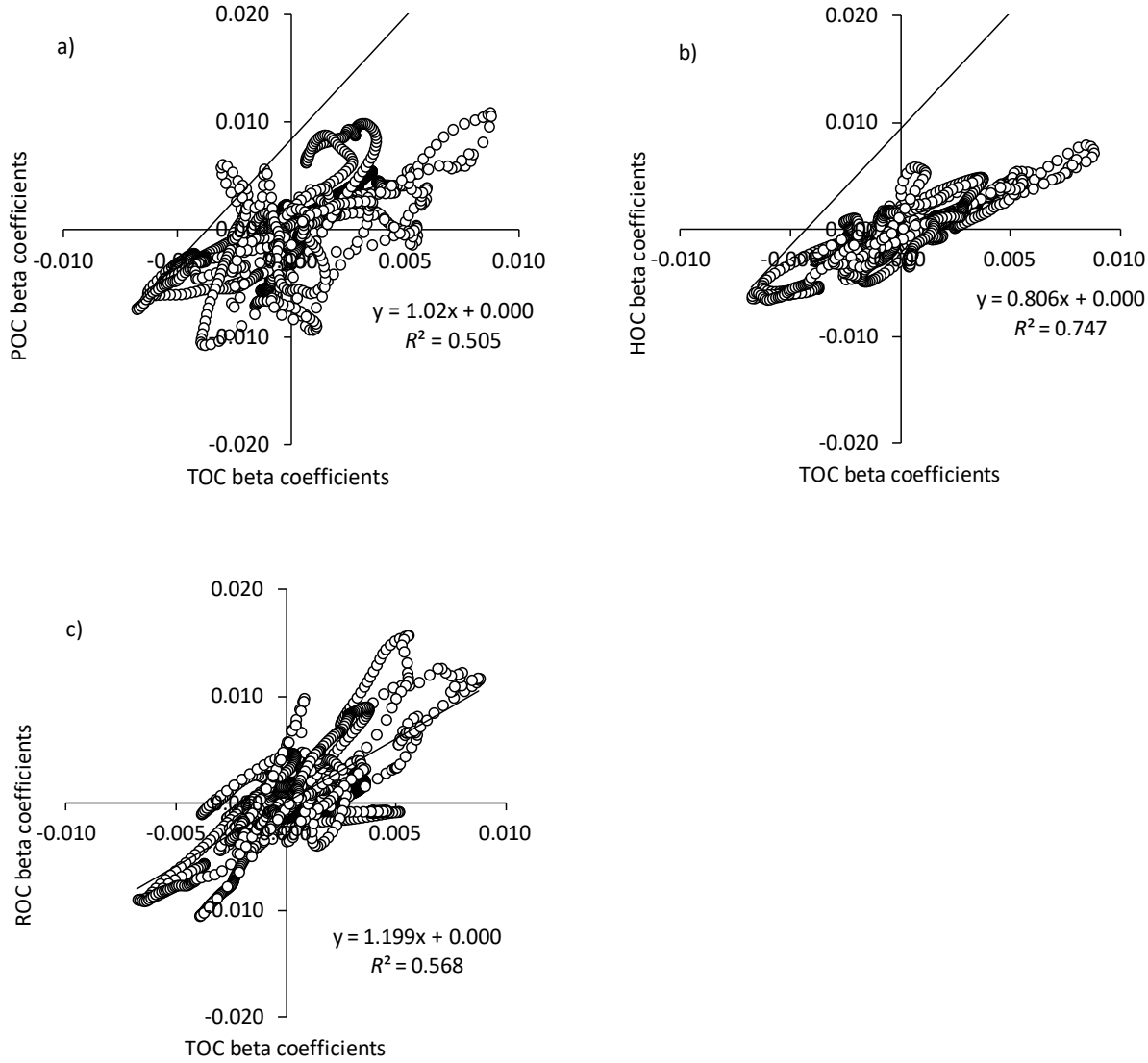


Fig. S2. Variable Importance in Projection (VIP) scores from MNIRS-PLSR models for TOC, POC, HOC and ROC. The functional group-specific peaks with VIP scores > 1 have major contribution to the model's prediction. Grey shaded regions of the wavenumbers are dominated by mineral bands.

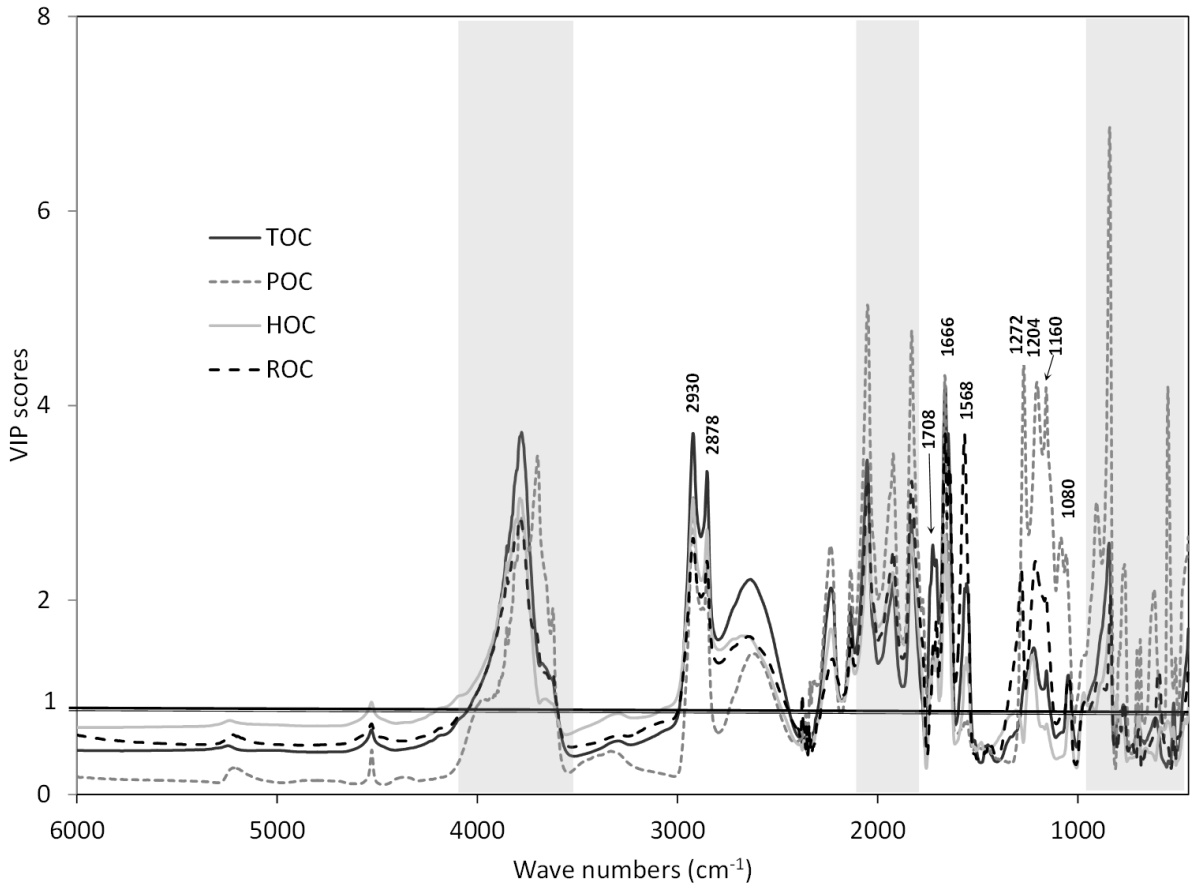
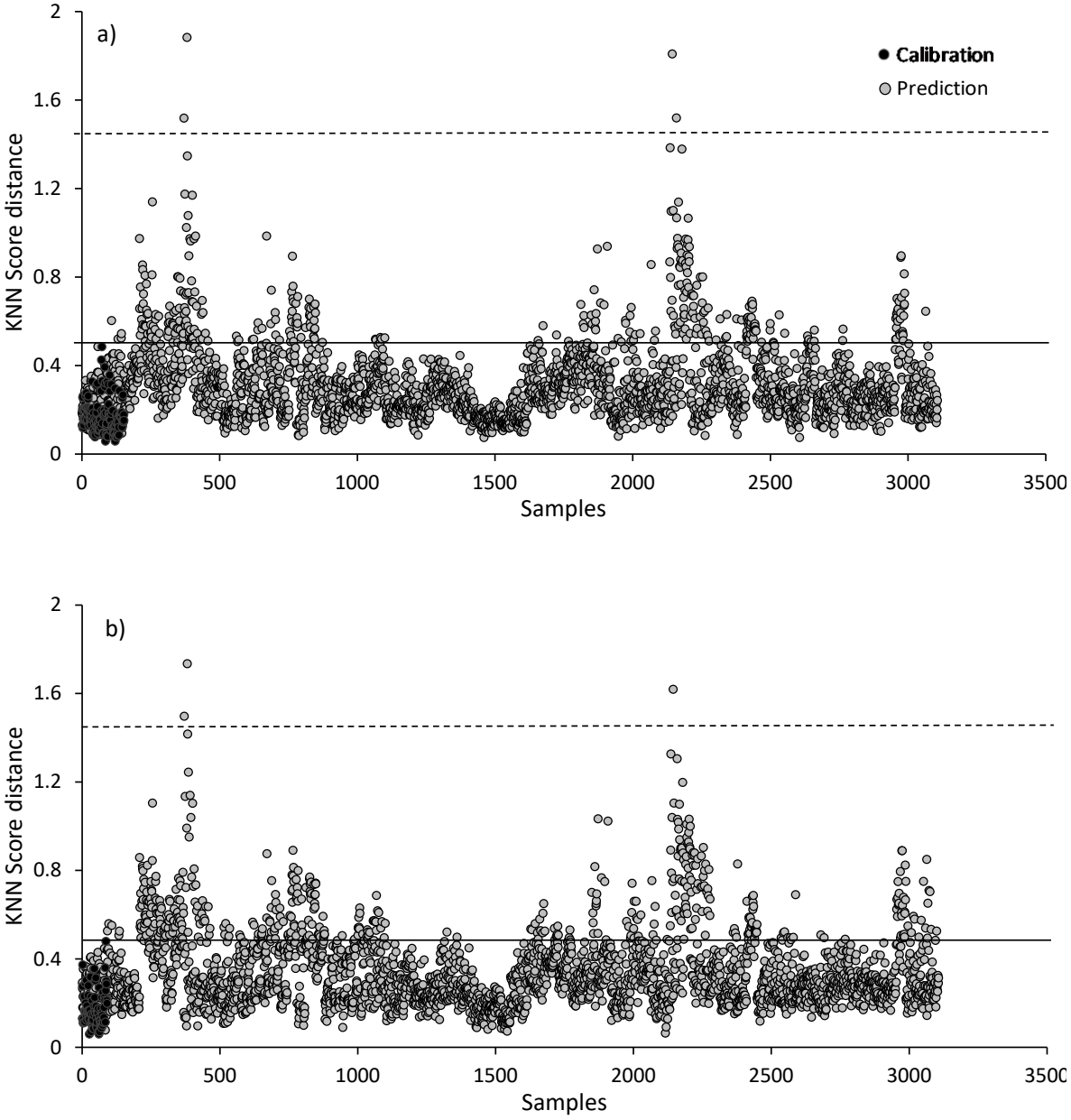
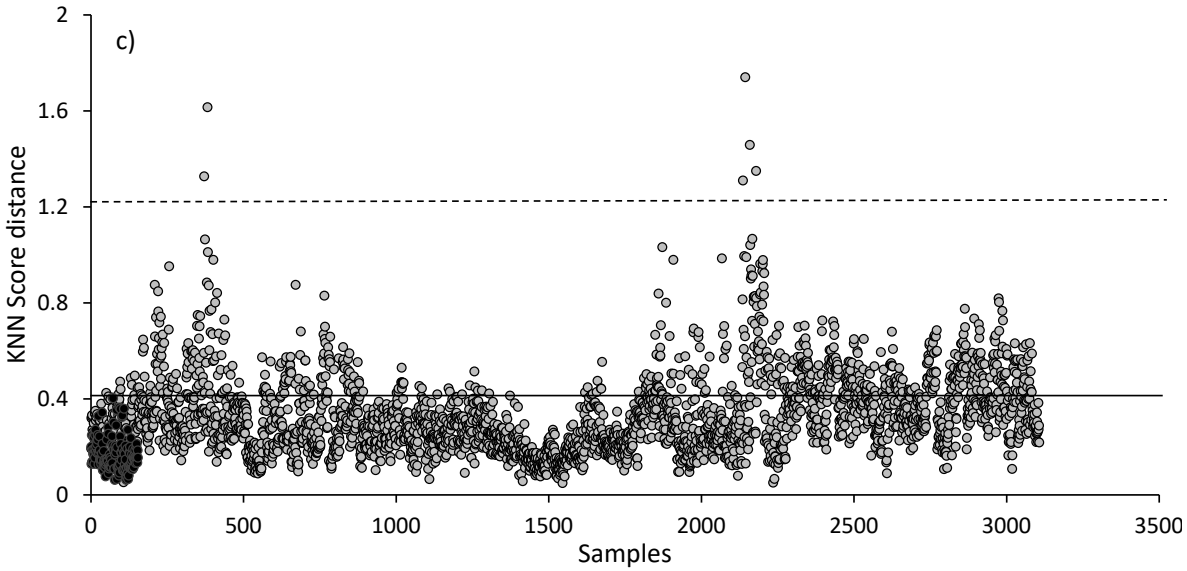
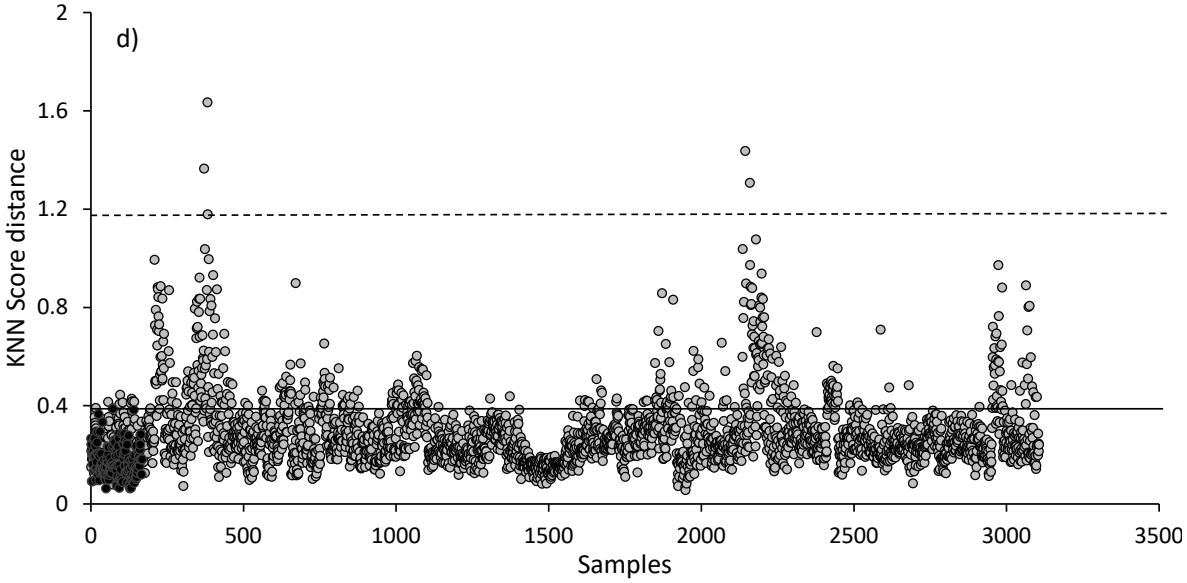


Fig. S3. The *k*-nearest neighbour (KNN) score distances from (**a**) TOC, (**b**) POC, (**c**) HOC, and (**d**) ROC predictions showing both MNIRS-PLSR model samples (Calibration, black circles) and 3109 predicted samples (Prediction, grey circles). Solid line indicates the inlier limit (MNIRS-PLSR model limit) and the broken line indicates 3 x inlier limit. Samples with KNN score distances more than 3 x inlier limit are considered as having less reliable predictions.



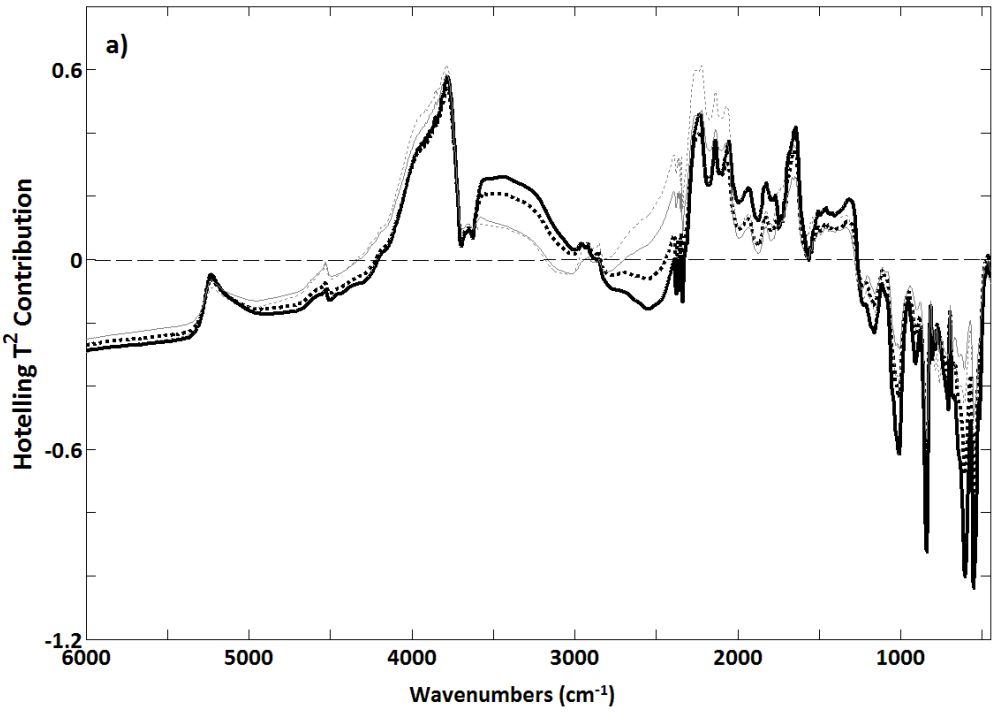

c)
KNN Score distance
2
1.6
1.2
0.8
0.4
0
0
500
1000
1500
2000
2500
3000
3500
Samples

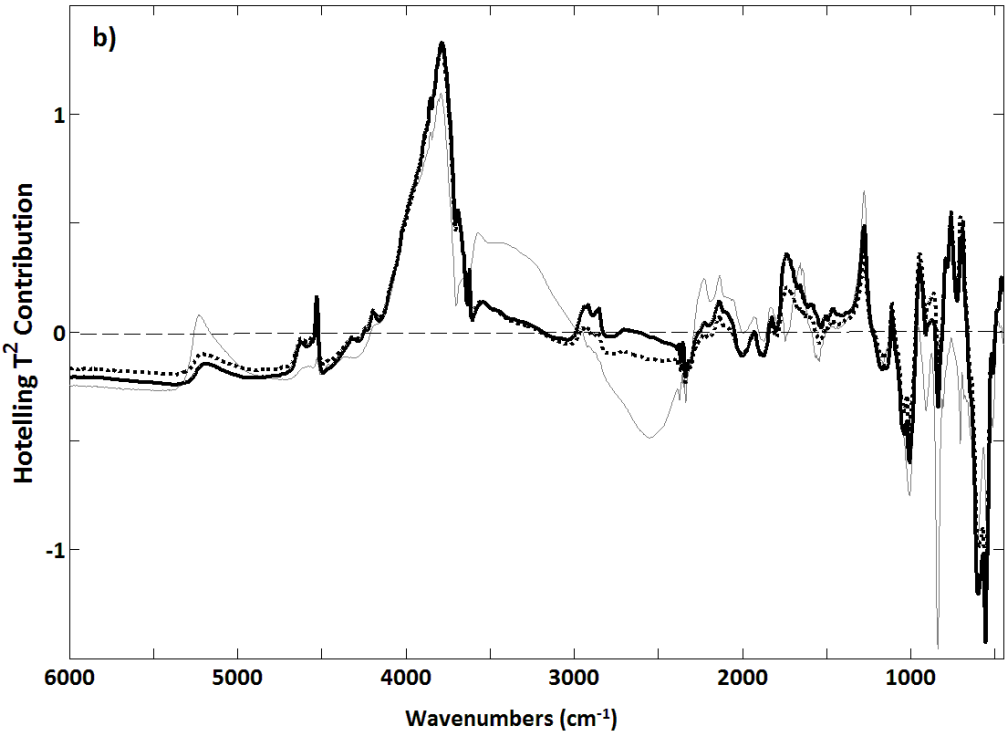

d)
KNN Score distance
2
1.6
1.2
0.8
0.4
0
0
500
1000
1500
2000
2500
3000
3500
Samples

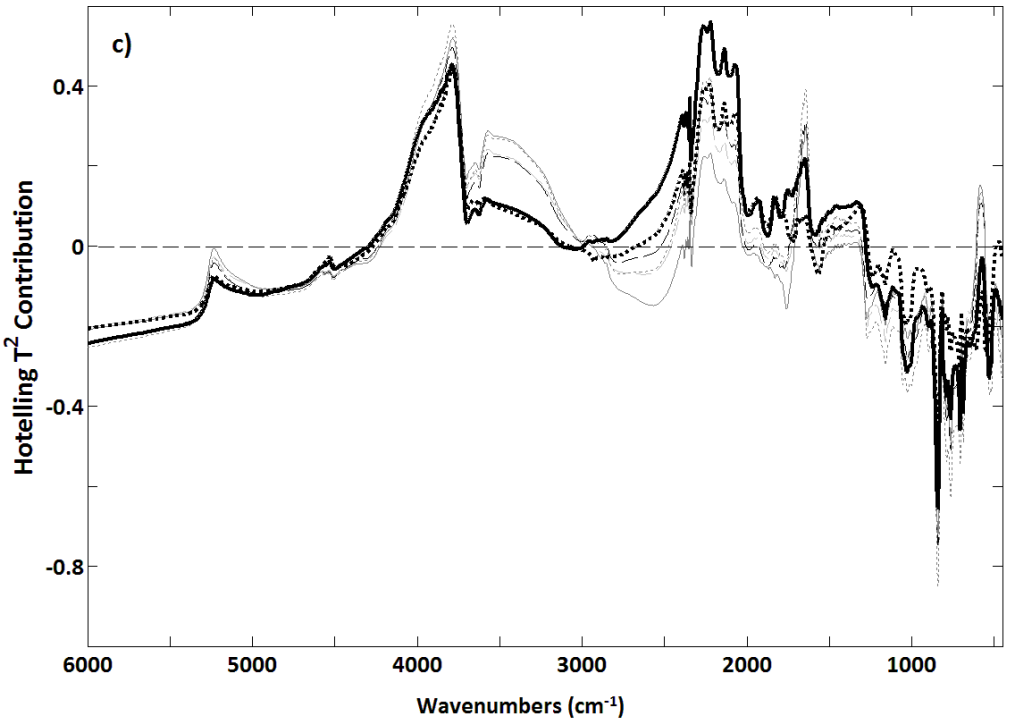
666

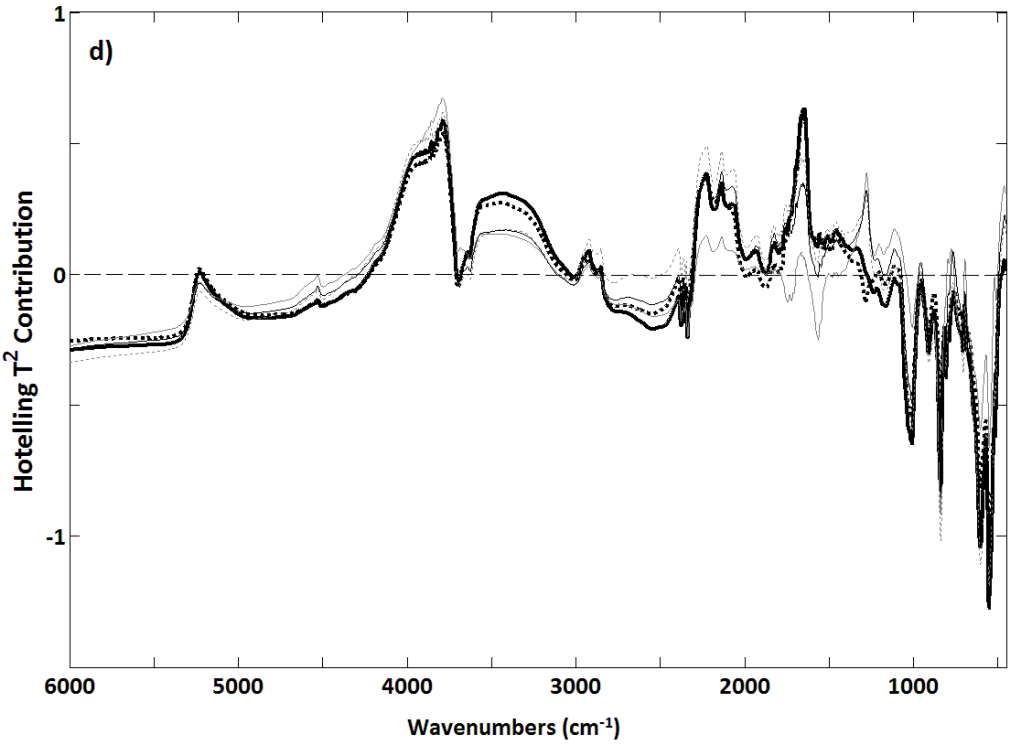
667

Fig. S4. Hotellings T^2 contributions of predicted (**a**) TOC, (**b**) POC, (**c**) HOC and (**d**) ROC samples with less reliable predictions (KNN score distances > 3 x inlier limit in Fig. S3). The wavenumbers show sharp mineral (4000-3600 cm^{-1} and 2000-1800 cm^{-1}) and inorganic C (~2300 cm^{-1}) peaks influencing their predictions. Bold lines represent the first two samples (sample numbers ~400, Sandy site - Katanning, WA), and the grey lines represent the subsequent samples (sample numbers ~2200, Site - Nhill, Vic with few calcareous soils) above the broken line (3 x inlier limit) in Fig. S3.




b)
1
0
-1
Hotelling T2 Contribution
6000
5000
4000
3000
2000
1000
Wavenumbers (cm-1)


c)
0.4
0
-0.4
-0.8
Hotelling T2 Contribution
6000
5000
4000
3000
2000
1000
Wavenumbers (cm-1)

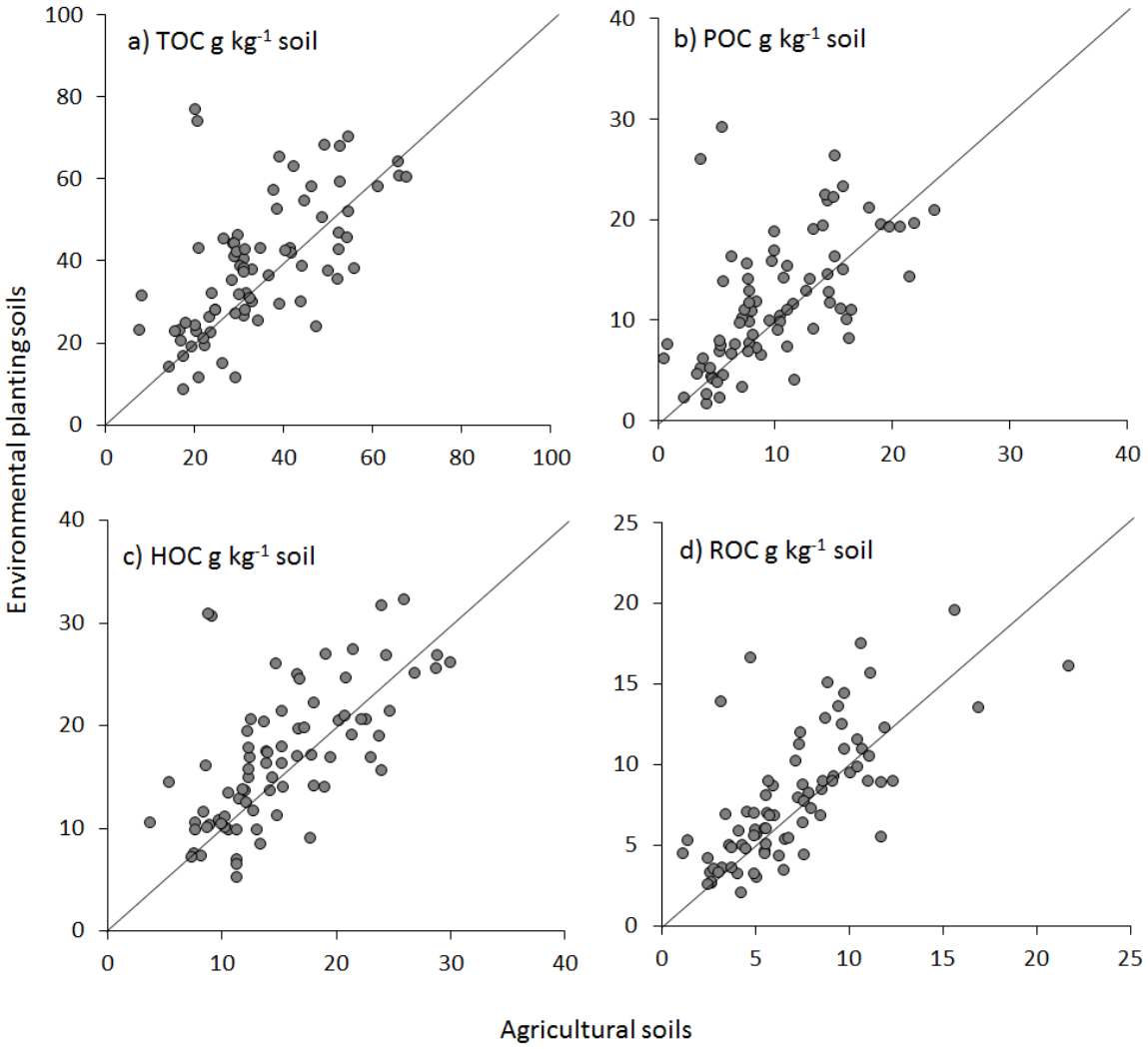

d)
1
0
-1
Hotelling T2 Contribution
6000
5000
4000
3000
2000
1000
Wavenumbers (cm-1)

675

676

677 **Fig. S5.** MNIRS-PLSR predicted concentrations of (**a**) TOC, (**b**) POC, (**c**) HOC and (**d**) ROC in
678 environmental planting soils and respective agricultural reference soils (site means across 79
679 sites, 0-5 cm). The diagonal line in each graph represents the 1:1 line.

680



681

682