



# NIRS quantification of lake sediment composition by multiple regression using end-member spectra

Fiona E. Russell · John F. Boyle · Richard C. Chiverrell

Received: 12 April 2018 / Accepted: 21 March 2019  
© The Author(s) 2019

**Abstract** Here we develop a novel method for quantifying sediment components, e.g. biogenic silica, organic or mineral matter, from near infrared (NIR) spectra based on fitting by multiple regression of measured spectra for end-member materials. We show that with suitable end-members our new open-source multiple regression routine gives excellent simultaneous quantification of the major components of a sediment, the concentrations comparing well with independent methods of quantification. Widely used partial least squares regression approaches rely on large environmental training data sets; our method produces comparable results, but with the advantages of negating the need for a training dataset and with greater simplicity and theoretical robustness. We demonstrate that component NIR spectra are additive, a prerequisite for use of multiple regression to un-mix the compound spectra, and show that a number of environmental materials make suitable end-members

for this analysis. We show that spectral mixing is not conservative with respect to mass proportion, but rather to the relative chromatic intensity of contributing sediment components. Concentrations can be calculated using the measured spectra by correction using a chromatic intensity factor, the value of which can be measured independently. We have applied our approach to a postglacial sediment sequence from Loch Grannoch (SW Scotland) and reveal a down-core pattern of varying dominance by biogenic silica, organic and mineral content from the late glacial to present. With isolation and measurement of appropriate end-members this multivariate regression approach to interrogating NIR spectra has utility across a wide range of sedimentary environments and potentially for other spectral analytical methods.

**Keywords** NIRS · End-member · Multiple regression · Sediment characteristics · Organic matter · Palaeolimnology

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10933-019-00076-2>) contains supplementary material, which is available to authorized users.

F. E. Russell · J. F. Boyle (✉) · R. C. Chiverrell  
University of Liverpool, Liverpool, UK  
e-mail: jfb@liverpool.ac.uk

F. E. Russell  
e-mail: frussell@liverpool.ac.uk

R. C. Chiverrell  
e-mail: rchiv@liverpool.ac.uk

## Introduction

Near infrared diffuse reflectance spectroscopy (NIRS) has proven to be a valuable tool for quantifying the components present in lake sediments, such as type and quantity of organic and mineral matters (Malley et al. 1999; Malley and Williams 2014; Pearson et al. 2014). NIRS has the advantages of being quick,

simple, and non-destructive and is consequently used in many different fields including environmental research, agriculture, and pharmaceutical industries. Applications of NIRS in palaeolimnology have focused on extracting key environmental parameters or proxies for inorganic/organic markers (e.g. % total organic carbon: Pearson et al. 2014). The overlapping absorption bands in near infrared (NIR) spectra make identification or quantification of signals attributed to individual materials or characteristics difficult (Brown et al. 2006; Zornoza et al. 2008; Korsman et al. 2001). Development in chemometric methods, particularly partial least squares (PLS) regression, during the 1980s (Workman et al. 1996) improved and simplified the interpretation of spectra. Though highly successful applications have been made possible by these developments, the PLS approach has two disadvantages. First, a substantial data training set is required for a component to be quantified from the IR or NIR spectra, which must be based either on large numbers of samples with independent measurement of that given component (for example, % total organic carbon studied by Pearson et al. 2014) or on artificial mixtures requiring large quantities of purified end-member materials (for example, biogenic silica studied by Meyer-Jacob et al. 2014, in this case by FTIRS). Second, because PLS methods cannot be applied simultaneously to a number of components, it is difficult to evaluate potential interference in the NIR spectra between components, which instead may remain hidden in the PLS numerical processing.

Here we propose and present an entirely novel approach to generating palaeoenvironmental data from NIRS, based on the assumption that the NIR spectra of mixtures comprise linear combinations of the spectra of sediment components. This allows each individual sample to be analysed by regressing its NIR spectrum on to the spectra of a chosen set of end-member materials, the regression coefficients quantifying the end-member mixing proportions. This is fundamentally different from traditional methods that use regression not to analyse samples, but to develop a statistical model using an extensive training data set. For example, in the widely applied weighted-average PLS methods a set of concentrations known independently are regressed onto entire or partial NIR spectra to evaluate coefficients that can then be applied to unknown samples. The training sets required for this process are based typically on either a range of modern

environmental samples or on parallel independent measurements obtained for the same palaeorecord. Our method circumvents the need for such training sets, using instead regression to un-mix each sample spectrum from a library of end-member spectra. Consequently, our method provides simultaneous quantitative reconstruction for multiple components without requiring a training stage.

Here, we present and test our end-member regression methodology, address a number of key issues with applying multiple regression to un-mix spectra. Important among these are: the isolation and measurement of appropriate end-members; sensitivity testing of end-member choice; and comparison with the PLS approach. To have confidence in our multiple regression-based reconstructions some comparison is necessary with independently measured equivalent data, but critically the same issue applies to reconstructions by PLS. We report the results of an application of our approach tested on binary mixtures and on a full late-glacial to present lake sediment profile from Loch Grannoch, SW Scotland. We also test the generality of the procedure and end-member materials by applying them to the sediments of three additional lakes from differing regions (Wales, Norway and Sweden).

## Methods

### Sampling and the end-member library

To develop the end-member approach we collected and measured a library of organic and inorganic materials. Table 1 describes the locations and materials collected that form this end-member library. This includes materials that are homogeneous (e.g. individual minerals) and more heterogeneous materials (e.g. plant materials, rocks and sediments). To apply our end-member approach we obtained a postglacial sediment core from Loch Grannoch in the Galloway Hills (SW Scotland). Loch Grannoch is a small (1.14 km<sup>2</sup>) upland (210 m O.D.) oligotrophic lake, with a granite bedrock (Cairnsmore of Fleet intrusion) (14 km<sup>2</sup>) catchment area (Flower et al. 1987). The Loch Grannoch core was sampled on 31 October 2016 from the central part of the lake (54.9954°N, 4.2832°W) from an anchored floating platform in ~ 16 m of water. The cores comprise 4 overlapping

**Table 1** Materials for which  $CIF_M$  values have been measured

Sample code	Type	Material	Location	$CIF_M$	Latitude	Longitude
PeatMMoss	Organic	Peat	May Moss, N Yorkshire, UK	1.00	54.3559	− 0.6532
Platanus	Organic	Platanus leaf	Liverpool, UK	2.70	53.4020	− 2.9652
Sphag	Organic	Sphanum sp.	May Moss, N Yorkshire, UK	0.77	54.3559	− 0.6532
Erioph	Organic	Eriophyllum	May Moss, N Yorkshire, UK	1.46	54.3559	− 0.6532
Calluna	Organic	Calluna	May Moss, N Yorkshire, UK	1.55	54.3559	− 0.6532
HAMMoss	Organic	Humic acid	May Moss, N Yorkshire, UK	4.51	54.3500	− 0.6426
HARiv	Organic	Humic acid	Rivington Moor, Lancashire, UK	4.73	53.6204	− 2.5380
FAMMoss	Organic	Fulvic acid	May Moss, N Yorkshire, UK	1.56	54.3500	− 0.6426
DOCWhix	Organic	DOC	Whixhall Moss, Cheshire, UK	0.84	52.9175	− 2.7596
DOCRiv	Organic	DOC	Rivington Moor, Lancashire, UK	0.78	53.6207	− 2.5381
DOCMig	Organic	DOC	Migneint, N Wales, UK	0.87	52.9767	− 3.8347
LGClayLOR	Sediment/soil	Late glacial silty clay	Lilla Öresjön, Sweden	1.26	57.5531	12.3170
LGClayMyn	Sediment/soil	Late glacial silty clay	Llyn Cwm Mynach, W Wales, UK	1.94	52.7960	− 3.9609
LGClayGran	Sediment/soil	Late glacial silty clay	Loch Grannoch, Galloway, UK	2.26	54.9967	− 4.2838
TillOgGrey	Sediment/soil	Till clay	Ogwen Valley, N Wales, UK	1.56	53.2301	− 4.0789
TillOgBrown	Sediment/soil	Till clay	Ogwen Valley, N Wales, UK	1.79	53.2301	− 4.0789
TillAfon	Sediment/soil	Till clay	Afon Gain, Gwynedd, W Wales, UK	2.82	52.8717	− 3.8730
Rhyo	Rock	Rhyolite	Ogwen Valley, N Wales, UK	1.40	53.1289	− 4.0316
RhyoWeath	Rock	Rhyolite, weathered	Ogwen Valley, N Wales, UK	2.60	53.1289	− 4.0316
Gneiss	Rock	Gneiss	Vågsøy, Norway	3.03	62.0302	5.0034
MicrGabbr	Rock	Microgabbro	Ogwen Valley, N Wales, UK	3.09	53.1235	− 4.0266
MicrGrani	Rock	Microgranite	Ogwen Valley, N Wales, UK	3.90	53.1221	− 4.0095
Tuff	Rock	Tuff	Ogwen Valley, N Wales, UK	4.00	53.1246	− 3.9961
LST	Rock	Limestone	Great Orm, N Wales, UK	6.73	53.3236	− 3.8517
SiltST	Rock	Siltstone	Ogwen Valley, N Wales, UK	0.98	53.1405	− 4.0371
SlatePurp	Rock	Slate, purple	Ogwen Valley, N Wales, UK	1.14	53.1730	− 4.0572
SlateGrey	Rock	Slate, grey	Ogwen Valley, N Wales, UK	2.82	53.1730	− 4.0572
SlateGreen	Rock	Slate, green	Ogwen Valley, N Wales, UK	5.22	53.1730	− 4.0572
SST	Rock	Permotriassic Sst	Wirral, UK	1.35	53.3535	− 3.1357
RockBWR0	Rock	Andesitic volcanoclastics	Dovedale, Cumbria, UK	3.74	54.4949	− 2.9582
RockBWR5b	Rock	Andesitic volcanoclastics	Dovedale, Cumbria, UK	3.90	54.4971	− 2.9505
RockBWR6	Rock	Andesitic volcanoclastics	Dovedale, Cumbria, UK	3.70	54.4967	− 2.9569
RockBWR7	Rock	Andesitic volcanoclastics	Dovedale, Cumbria, UK	3.53	54.4950	− 2.9584
DiatomMar	Biogenic silica	Marine diatom	California, USA	3.09		
DiatomEd	Biogenic silica	Diatom	Not known	3.79		
DiatomGM	Biogenic silica	Diatom	Grasmere, Cumbria, UK	2.50	54.4551	− 3.0296

lengths sampled using a  $0.075 \times 1.5$  m capacity hand-percussive Russian corer. Cores were wrapped and sealed in polythene and stored refrigerated until required for analysis. The sediments comprised 3.19 m of largely organic limnic muds and 0.6 m of inorganic muds that extend to deglaciation including

the late glacial oscillations (Greenland Interstadial and Stadial 1), with regional deglaciation of the Galloway Hills dated to  $\sim 15$  k years ago (Ballantyne et al. 2013).

Specific end-member mineral samples were selected and ground to a fine powder using a pestle

and mortar. The aim was to isolate single materials but in reality some minor contaminant minerals potentially remain. Rocks were sampled from various catchments to reflect detrital sediment sources to lakes. These were cut to expose fresh surfaces, and fine powders were obtained using a diamond drill. In formerly glaciated environments the basal lacustrine muds lain down while ice was still affecting the lake could be regarded as a partially homogenised sample of catchment sediment sources, except where glacial ice external to the lake catchment was likely. Dried and powdered samples of these deglacial inorganic muds are used as end-members reflecting bedrock of the catchments.

Organic materials include specific plant samples (e.g. *Sphagnum*, *Calluna vulgaris* (L.) Hull), a sample of UK ombrotrophic peat (from May Moss, a site of varying composition in terms of plant remains and the degree of peat humification), natural dissolved organic matter from streams draining peatlands (Rivington Moor, Whixall Moss, and Migneint) isolated by freeze-drying, and separated humic and fulvic acid extracted from May Moss peat (collected March 2017). Humic acid was also collected from May Moss and Rivington Moor stream water by acidification to pH 2 and filtration.

Biogenic silica end-members were obtained using (1) the > 90 µm fraction of diatom-rich sediment taken from the edge of Grasmere (Cumbria), visually inspected to confirm the absence of mineral matter, (2) a cultured marine diatom (*Thalassiosira pseudonana*, supplied by Reed Mariculture, Campbell, California, USA), and (3) commercial food supplement diatomaceous earth (Ultra-fine freshwater diatomaceous earth, Diatom Retail Ltd, Leicester, UK). All three samples were treated with hot acidified hydrogen peroxide to remove organic matter.

Humic and fulvic acid were extracted from May Moss peat following the method of Hayes et al. (1975) but excluding the acid prewash (our samples lack carbonate). The extraction comprised (Stage 1): 10 g peat reacted with 100 ml of 1 M NaOH, swirled for 16 h at room temperature and then decanted following centrifugation. Stage 2 involved careful acidification of the NaOH extract with concentrated AnalR HCl, adjusting to pH 1. After 3 days at 4 °C, the humic acid precipitate was collected by centrifugation, and repeatedly washed in deionised water. Stage 3 involved neutralisation of the remaining solution with

1 M NaOH. The neutral solution was then freeze dried to recover a mixture of fulvic acid and NaCl. Stage 4 involved removal of the NaCl using dialysis tubing.

We also created artificial binary and ternary admixtures of materials to assess the impact of varying compositions on the quantitative information acquired from the NIRS analyses. All such mixtures were homogenised by grinding in a pestle and mortar. A ternary equal mass proportion admixture was created using the late glacial muds from Loch Grannoch, Llŷn Cwm-mynach (N Wales) and Lilla Öresjön (S Sweden). May Moss peat and late glacial sediment (Loch Grannoch) were prepared in quantity for these synthetic admixtures. The raw materials were homogenised by repeated sieving (63 µm for the mineral matter, 125 µm for the peat). The binary admixtures were created on a mass proportional basis (0, 7, 20, 33, 47, 60, 73, 87, 93, 100%) of Loch Grannoch late glacial muds to May Moss peat to assess any deviation of fitting coefficients from a linear relationship with respect to mass proportions.

#### Analytical methodology

NIR spectra for both Loch Grannoch sediment and the end-member materials, processed as outlined in this section, are available in the University of Liverpool Data Repository as tab delimited text files (<http://dx.doi.org/10.17638/datacat.liverpool.ac.uk/550>). These files are formatted for use in the R code, also provided.

NIR spectra were measured by diffuse reflectance using an integrating sphere on a Bruker MPA Fourier-Transform NIRS for both the end-member data set and for 65 discrete evenly spaced 5-mm-thick subsamples from the 3.19 m Loch Grannoch core. All samples measured were freeze-dried, homogenised by grinding in a mortar, and lightly hand pressed (Korsman et al. 2001), with the NIR spectra based on combining 64 scans collected at 8 cm<sup>-1</sup> intervals across the range 3595–12,500 cm<sup>-1</sup>.

To compare partial least squares (PLS) regression methods applied conventionally to interrogate IR spectra (Pearson et al. 2014; Meyer-Jacob et al. 2014) with our new multiple regression end-member approach, PLS-WA analysis was undertaken using the Bruker OPUS software (Quant Package). A range of numerical processing procedures were used to systematically vary the numerical processing of the NIR spectra including various normalisation procedures

and derivatives. We also used a Principal Components Analysis (PCA) on a correlation basis to examine the overall spectral structure. This approach has enabled assessment of the most appropriate numerical methods and wavelength range in the NIR spectra for determining organic and mineral components in the sediments. Similar to previous work (Burns and Ciurczak 2001; Korsman et al. 2001; Pearson et al. 2014) we found using the 1st derivative of the NIR spectra was most appropriate. 1st derivatives for all NIR spectra were calculated using a centrally-weighted Savitzky–Golay smoothing (SGA) algorithm, and our analysis focuses on the wavelengths 8000–3800  $\text{cm}^{-1}$ , minimising noise whilst containing the key spectral structure diagnostic of organic and mineral components.

Independently quantified sediment component concentrations in the Loch Grannoch core (Organic matter, biogenic silica, and mineral matter) were needed both to test the results of the multiple regression and as a training dataset for the PLS comparison. A subset of 22 samples from the Loch Grannoch core was used to train the PLS-WA regression. Organic matter concentrations were quantified by loss on ignition (LOI), with weight loss measured after 1 h of ignition at 550 °C on sediment previously dried at 105 °C (Boyle 2004). Element concentrations from which to calculate normative mineral matter and biogenic silica were measured using an Energy Dispersive X-ray Fluorescence Analyser (ED-XRF). Dried samples were hand-pressed in 20 mm pots, and measured under a He atmosphere using a Spectro XEPOS 3 ED-XRF that emits a combined binary Pd and Co excitation radiation and uses a high resolution, low spectral interference silicon drift detector. The XRF analyser undergoes a daily standardization procedure and has accuracy verified using 18 certified reference materials (Boyle et al. 2015).

Garrels and Mackenzie (1971) demonstrated that mineral concentrations present in soil or sediment can be calculated using a method borrowed from igneous petrology. Idealised, or ‘normative’, mineral concentrations are calculated based standard compositions for the constituent minerals, via application of a series of steps that allow for the elements in the soil or sediment to be fully accounted for. Details are provided in Boyle (2001). The following steps were applied:

- Recalculate major elements as oxides ( $\text{SiO}_2$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{Fe}_2\text{O}_3$ ,  $\text{CaO}$ ,  $\text{MgO}$ ,  $\text{Na}_2\text{O}$ ,  $\text{K}_2\text{O}$ ,  $\text{TiO}_2$ ,  $\text{MnO}_2$ ,  $\text{P}_2\text{O}_5$ ), and together with LOI normalise these to unity
- Mineral matter is calculated as the sum of: measured oxides excluding  $\text{SiO}_2$ ; calculated  $\text{SiO}_2$  associated with silicate minerals (chlorite, albite, orthoclase, anorthite); and quartz
- Biogenic silica is calculated as measured  $\text{SiO}_2$  minus quartz, and minus calculated  $\text{SiO}_2$  associated with silicate minerals.
- Silicate-associated  $\text{SiO}_2$  is calculated by assuming it to be present only as chlorite, albite, orthoclase, and anorthite, considered to be the sole sources of  $\text{MgO}$ ,  $\text{Na}_2\text{O}$ ,  $\text{K}_2\text{O}$  and  $\text{CaO}$ , respectively. Silica/oxide ratios, from Deer et al. (1966), were taken to be 1.68, 6.13, 5.53, and 2.27, respectively.
- Quartz is not measured directly, but is taken to be 30 times  $\text{TiO}_2$ . This value is the average ratio of “free”  $\text{SiO}_2$  to  $\text{TiO}_2$  for the late glacial sediment (assumed to contain negligible biogenic silica), where free  $\text{SiO}_2$  is total measured  $\text{SiO}_2$  minus silicate associated  $\text{SiO}_2$ . This approach is most reliable where mineral matter concentrations are low, and least reliable when they are high (where quartz uncertainty is maximal and biogenic silica is lowest).

#### Data handling and statistical methods

The IR intensity arising from a mixture depends on the type and concentration of chromophores (the regions of a molecule that interact with photons) contributed by its component parts (Boroumand et al. 1992) with intensity related to sample component mass concentration. The combined intensity is not expected to be linearly related to the mass proportions, as the chromophore density will vary among materials. However, the component IR intensities (mass proportion scaled by chromophore density) should be additive, such that their mixing proportions can be found by multiple regression (Eq. 1).

$$I_{M,k} = b_0 + b_1 I_{C1,k} + b_2 I_{C2,k} + \dots + b_n I_{Cn,k} \quad (1)$$

where  $I$  is the signal intensity at wavenumber  $k$  ( $\text{cm}^{-1}$ ) for the mixture ( $M$ ) and components ( $C_1$  to  $C_n$ ),  $b_0$  to  $b_n$  are the regression coefficients

If chromophore density was equal for all materials, then the regression coefficient would yield the mass proportions (concentrations) of the components in a mixture. However, chromophore densities vary according to material, so the coefficients instead quantify what may be described as the chromatic proportions. To calculate mass proportions from this information we need to know something of the chromatic properties of the components. This need not be known in detail because we have measured spectra for the components, and are fitting these to the measured spectrum of mixtures. Instead, provided we assume (after Boroumand et al. 1992) that the component spectra are additive, then we simply need a coefficient that represents the average chromatic intensity of each component, which we term the Chromatic Intensity Factor (CIF). This allows mass proportions to be calculated (Eq. 2).

$$w_x = \frac{b_1 \text{CIF}_1}{\sum_{x=1}^n b_x \text{CIF}_x} \quad (2)$$

where shown for component 1:  $w_x$  is mass fraction of component  $x$ ,  $\text{CIF}_x$  if the chromatic intensity factor of component  $x$

The CIF value for each component may be found by parameterisation, if independent information is available for the composition of the mixture. However, it can also be measured using synthetic mixtures, if one component is chosen as a reference and assigned a value of 1, a 50:50 mixture of May Moss peat (our chosen reference) with component X yields a compound spectrum that can be regressed onto the spectra of its two components. The measured CIF value for component X relative to May Moss peat is given by Eq. 3.

$$\text{CIF}_x = \frac{b_{\text{MMPeat}}}{b_x} \quad (3)$$

Multiple regression was used for each sample to fit end-member spectra to the sample spectrum. This is done using the linear regression model (LM) function in R (R Core Team, 2013), with regression model coefficients and confidence intervals returned using the COEF and CONFINT functions. Mass normalisation following correction for chromatic intensity was undertaken using Eqs. 2 and 3. The R code reports and plots mixing proportions for each component included in the multiple regressions with 95% confidence intervals for each sample, and a measure ( $R^2$ ) of the

proportion of the variance in the sample NIR spectra explained by the selected component end-member NIR spectra. The R code is available in University of Liverpool Data Repository (<http://dx.doi.org/10.17638/datacat.liverpool.ac.uk/550>).

CIF values may also be found by optimisation if component concentrations are independently known. In such cases CIF values can be adjusted to minimise the mean squared difference between modelled and known component concentrations.

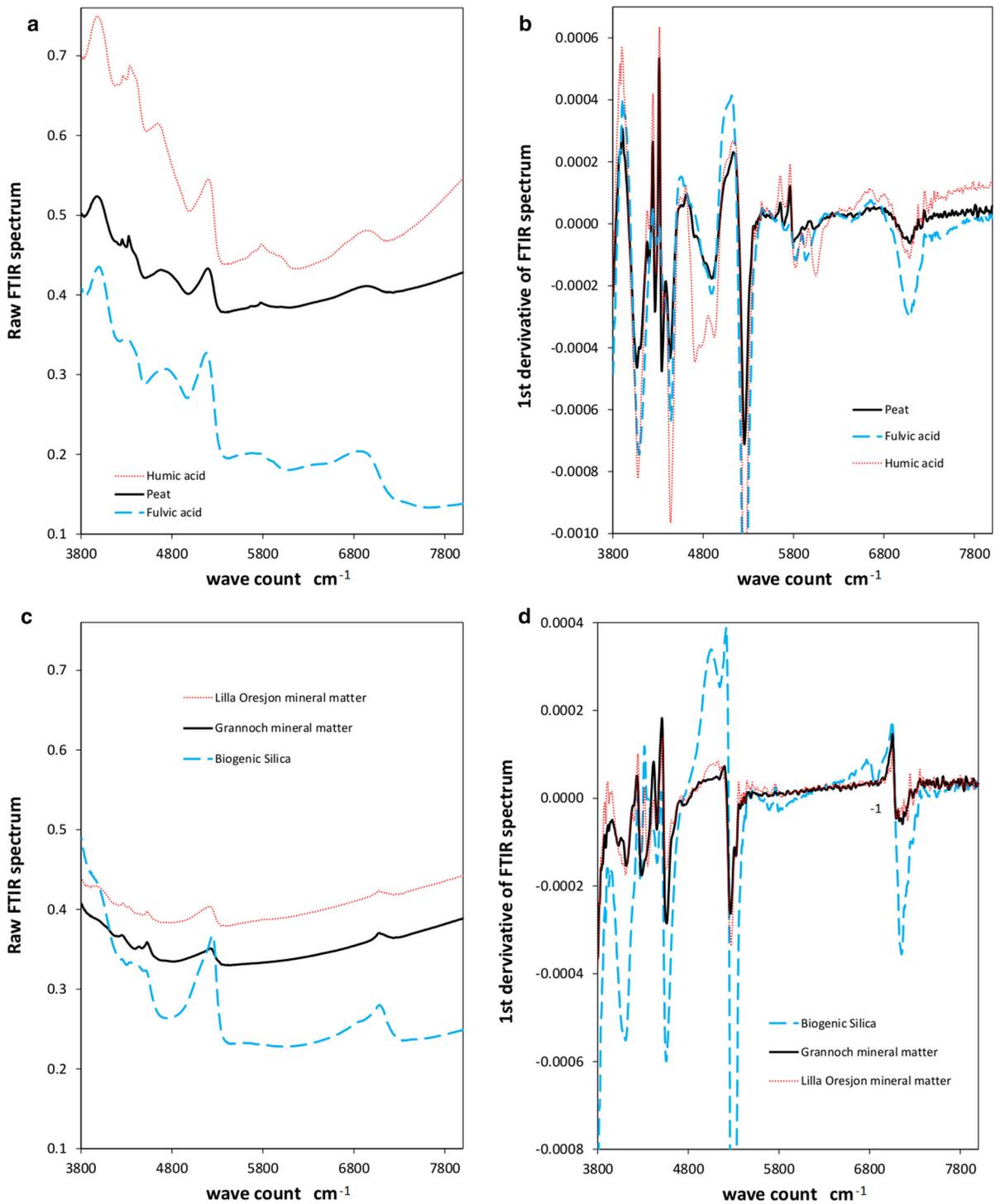
## Results

### End-member NIR spectra

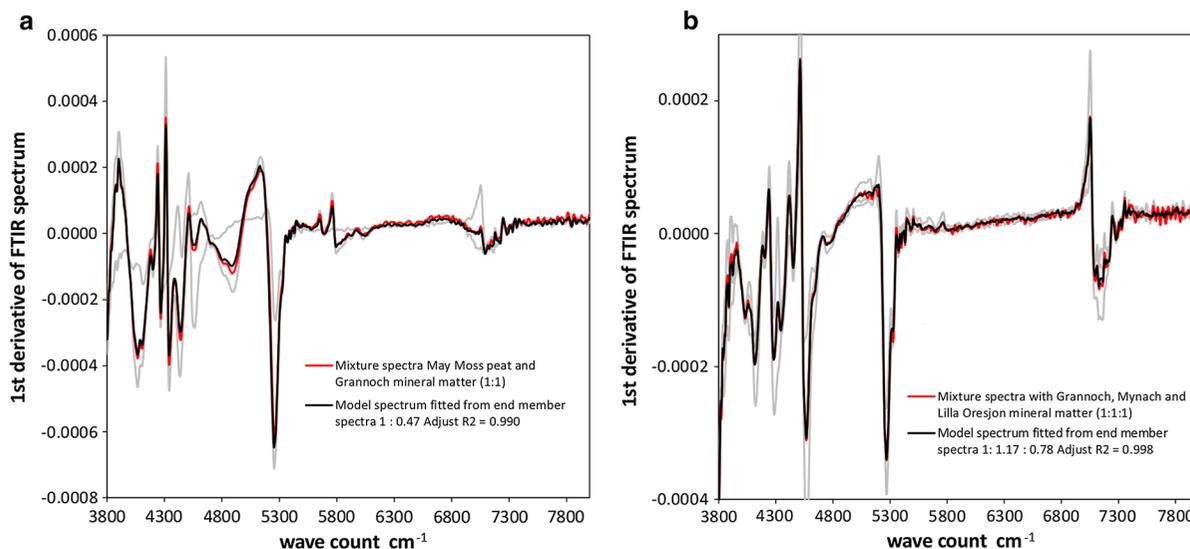
A selection of raw unprocessed spectra for potential end-members (Fig. 1a, c) illustrates the high degree of similarity between widely different materials. Comparing the same materials, a greater distinction is achieved using 1st derivative NIR spectra, with clearer differences between organic and mineral end-members (Fig. 1b, d).

Comparing the spectra of a 50:50 binary synthetic mixture (Red: Fig. 2a) with May Moss Peat and Loch Grannoch mineral component end-member spectra (Grey: Fig. 2a) shows that the 50:50 admixture spectrum lies between, but not midway between, its end-members at all wave counts. For an equal ternary admixture of Llŷn Cwm-mynach, Lilla Öresjön and Loch Grannoch mineral matter, the admixture spectrum is also bracketed by the end-members (Fig. 2b), but the distinction between the original end-member 1st derivative spectra is less clear. In both cases (Fig. 2), multiple regression has been applied to fit the component end-member spectra to the admixture spectrum, and a very high degree of fit is obtained for both the binary (adjusted  $R^2 = 0.991$ ,  $F = 60,723$ ) and ternary mixtures (adjusted  $R^2 = 0.996$ ,  $F = 93,022$ ). The regression coefficients, however, do not conform to the known mass mixing proportions of the end-member components (values shown on Figure).

The non-linear NIR signal response to end-member concentration is further illustrated by a range of binary mixtures of Loch Grannoch mineral matter with May Moss peat (Fig. 3). The raw multiple regression coefficients (Fig. 3, filled symbols) do not lie on a mass-proportion mixing line, but do lie on a

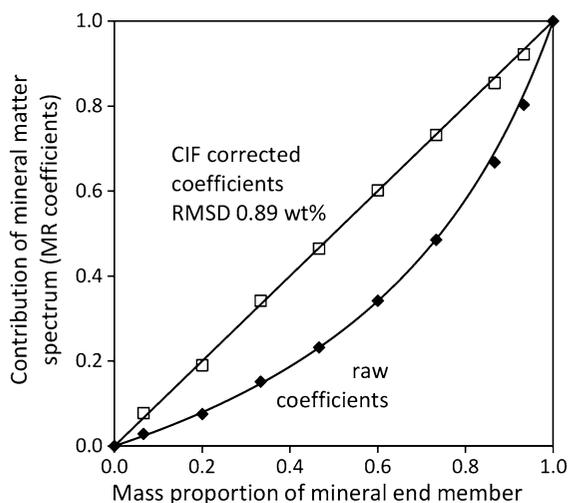


**Fig. 1** Exemplar NIR spectra showing raw NIR spectra for **a** selected mineral and biogenic materials, **b** selected organic matter end-members, and 1st derivative NIR spectra for **c** mineral and **d** organic end-member materials. Colour figure available online



**Fig. 2** Mixing of end-member components showing the 1st derivative NIR spectra for **a** a binary mixture of May Moss peat and Loch Grannoch mineral matter and **b** an equal ternary mixture of Llyn Cwm Mynach, Lilla Öresjön and Loch

Grannoch mineral matter. The NIR spectra shown include: the original raw end-members, the measured admixture and the modelled fitted admixture spectrum from the end-members derived by multiple regression. Colour figure available online



**Fig. 3** Known versus NIRS quantified mineral matter end-member proportions for synthetic binary mixtures (0, 7, 20, 33, 47, 60, 73, 87, 93, 100%) of May Moss peat and Loch Grannoch mineral matter. CIF corrected proportions were calculated from the MR coefficient using Eq. 2 with CIF = 1 for the peat, and 2.26 for the mineral matter

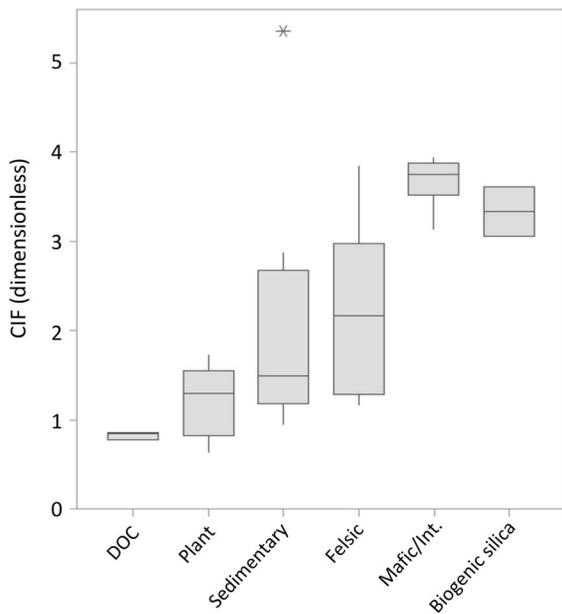
theoretically constrained line (form obeying Eq. 2, with a fitted CIF value of 2.26 for Loch Grannoch mineral obtained by minimising the squared differences). If this CIF value is used to correct the regression coefficients using Eq. 2 (rearranged to

yield chromatic proportions), then samples (Fig. 3, open symbols) do lie close to the ideal 1:1 mixing line.

#### Evaluating chromatic intensity factors

The method used above to determine a CIF value for Loch Grannoch mineral matter could be applied to any other material. However, a logistically easier alternative is to calculate the value from a single point, for which we use the 50:50 mass admixture of each end-member (Table 1) with May Moss peat. This was done for all end-member materials (Table 1) to obtain measured CIF values, which range from 0.77 for Sphagnum and up to 6.73 for a limestone sample. For the materials we have measured to date, organic materials typically show lower values (except humic acids), biogenic silica are higher, and rock materials range widely (Fig. 4).

While testing the applicability of these measured CIF values, we observed some cases where mixtures were non-ideal. For example, where fine rock powder coated larger fibres of peat, low concentrations of the rock powder yielded exaggerated chromatic intensities, presumably because fibre surfaces were preferentially measured. Consequently, it is desirable to take an additional approach to CIF estimation, whereby values are found for natural admixtures (soil or



**Fig. 4**  $CIF_M$  values plotted against broad categories of material type. Lowest  $CIF_M$ s were encountered for organic matter, particularly DOC recovered from water. Measured rock types show considerable variation, with highest values in mafic and intermediate igneous rocks. The biogenic silica samples show high values. The asterisk indicates an isolated extreme value

sediment) by adjusting  $CIF$  values to optimise agreement with independently quantified component concentrations. We demonstrate this in the case of the Loch Grannoch sediment record (Figs. 5, 6). We thus have two classes of  $CIF$  value, here distinguished as measured  $CIF_M$  and optimised  $CIF_O$ .

#### Sensitivity to the choice of end-members

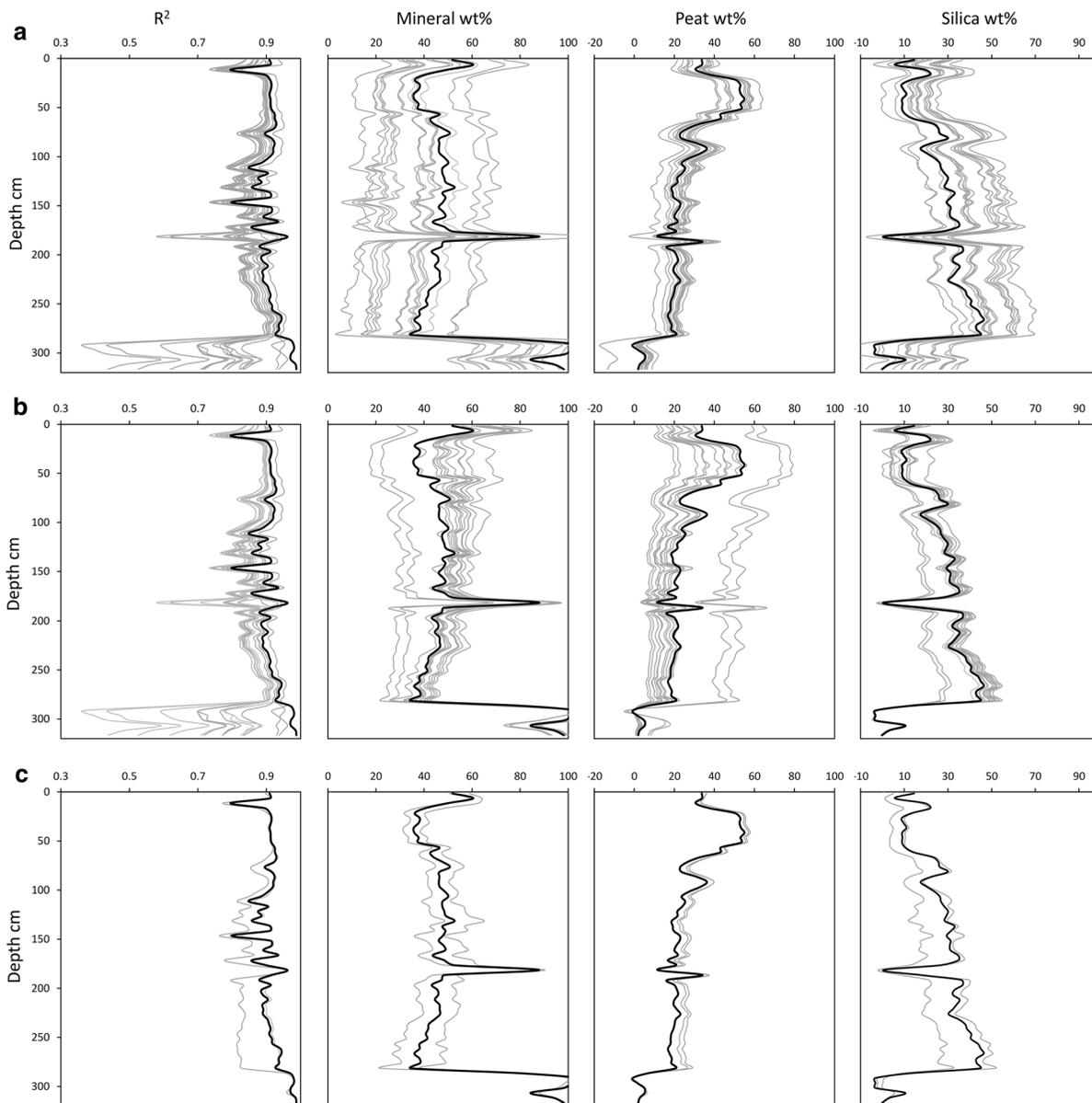
To explore the sensitivity of our multiple regression approach to the choice of end-members, two further experiments were conducted holding two of the end-members constant and varying a third using a range of related materials (Table 1). The experiment uses the NIR spectra obtained for the Loch Grannoch sediment record (Fig. 5). The primary end-member materials (the ones found to give the best overall fit) were May Moss peat (organic matter), marine diatom (biogenic silica) and Loch Grannoch late-glacial sediment (mineral matter). The first experiment (Fig. 5a) uses May Moss peat and marine diatom as fixed end-members, and cycles through a range of differing minerogenic sediments and rock types in place of

Loch Grannoch late-glacial sediment. The greatest impacts are on the quantification of the mineral matter fraction with organic and biogenic silica content less affected by choice of mineral matter end-member. Whilst the principal down-core pattern of mineral variation is captured in all cases, the mineral component shows widely varying values. This variation shows some association with rock type. Thus, slates give low values, while quartz-rich rock types give high. However, there are exceptions, so predicting the outcome based on the known local rock type is not reliable. Instead, incorporating a local catchment specific mineral sediment end-member appears important in order to capture both the pattern and critically the magnitude of down-core variation in all end-members. Powdered local bedrock or glaciogenic lake sediments of late glacial age both appear to successfully account for catchment mineral matter. The second experiment (Fig. 5b) holds the Loch Grannoch late-glacial mineral matter and marine diatom as fixed components, but cycles through various organic matter fractions. The impact of varying the organic component is substantial, affecting the fit for all three components. For example, humic acid yields low values for mineral matter and biogenic silica, and high values for organic matter. That said, the down core patterns remain broadly similar for all the end-members assessed. We currently have only three samples of biogenic silica in conducting the third experiment (Fig. 5c). Very similar patterns are obtained with slightly varying magnitude.

#### Applying end-member multiple regression to lake sediments

To test the utility of the methods presented here in discerning evidence for environmental change from the NIRS analysis of lake sediments, three different approaches have been applied using NIR spectra obtained for the Loch Grannoch sediment record (Fig. 6). The three approaches were:

1. End-member multiple regression using the NIR spectra for three materials with measured  $CIF$  values ( $CIF_M$ ) (Fig. 6a).
2. End-member multiple regression using the NIR spectra for three materials with optimised  $CIF$  values ( $CIF_O$ ) (Fig. 6b)

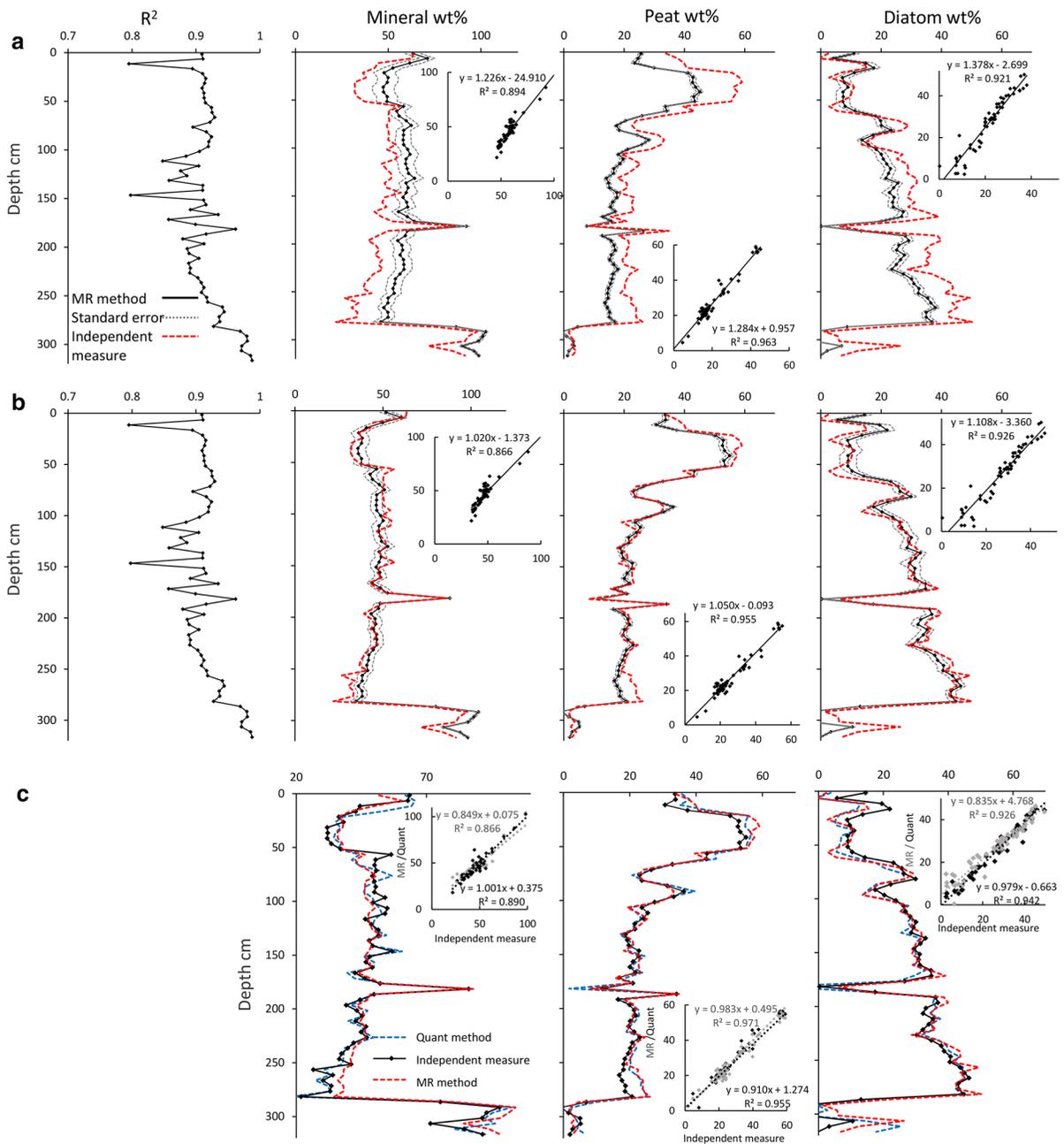


**Fig. 5** Example fitted concentrations for rocks, organic and biogenic silica end-members to Loch Grannoch. **a** Cycles through a range of differing minerogenic sediments and rock types, with May Moss peat and marine diatom as fixed end-members. **b** Cycles through a range of differing organic matter types, with Loch Grannoch mineral matter and marine diatom as

fixed end-members. **c** Uses May Moss peat and Loch Grannoch mineral matter as fixed end-members, together with each of the three biogenic silica samples. The thick black line represent the preferred case with marine diatom, Loch Grannoch mineral matter, and May Moss peat

- Using a PLS method relating the NIR spectra to independently quantified environmental parameters based on a training set that comprised a third of the samples from the Grannoch core (Fig. 6c).

All three experiments attempt to reconstruct the concentrations of biogenic silica, organic and mineral matter for the sediment core. The results are compared with independently quantified measures of the three parameters. The end-member materials were May Moss peat representing natural organic matter; marine



**Fig. 6** Fitted concentrations for Loch Grannoch using Loch Grannoch mineral matter, marine diatom, and May Moss peat end-members with **a** measured  $CIF_m$  values (2.26, 3.09, 1 respectively) and **b** optimised  $CIF_o$  values (1.4, 3.09, 1

respectively), and **c** values quantified by PLS. These are compared with independently quantified concentrations. Colour figure available online

diatom representing biogenic silica; and Loch Grannoch late glacial sediment representing catchment mineral matter. When using the measured  $CIF_M$  values to estimate concentration of mineral matter, organic matter and biogenic silica in the Loch Grannoch

sediment, we observe high correlations with the independently quantified concentrations, and good agreement in the depth of peaks and troughs (Fig. 6a). However, the absolute values differ particularly for

mineral and organic matter and the regression lines significantly deviate from the origin.

With optimised  $CIF_O$  values (Fig. 6b), absolute magnitudes are constrained to be similar to those independently quantified. However, there is also a substantial improvement in that the regression lines now pass through the origin for all three components, and in the case of organic matter the  $R^2$  value is substantially increased. The poor fit obtained for the biogenic silica in the basal samples need not indicate a failure of our method; a likely explanation is the poor independent quantification in these highly minerogenic sediments (Fig. 6).

The PLS method, trained using a third of the samples from the Loch Grannoch core with independently quantified concentrations, was successful in predicting the concentrations of these components. Correlation between independently quantified and PLS estimates should be and is strong (Red curve: Fig. 6c). While the PLS approach is effective, equivalently good results are obtained using CIF corrected end-member multiple regression (Fig. 6b), despite not using a training data set, or indeed being trained at all.

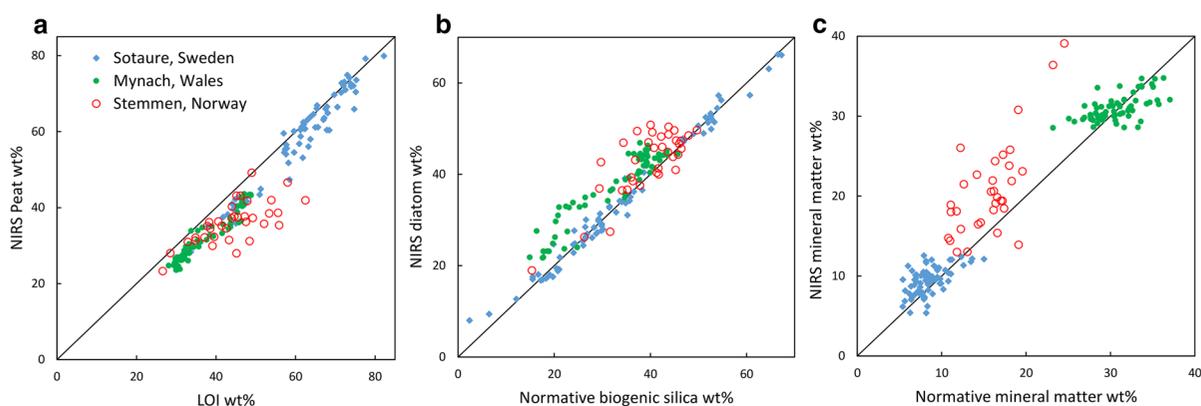
To further test the generality of our approach, and of our organic and biogenic silica end-member materials, we have measured NIRS spectra from three additional lake sediment cores from different regions and with differing climate and bedrock and applied our method (Fig. 7). We fitted our two general end-members (May Moss peat and marine diatom) using the  $CIF_O$  values optimised for Loch Grannoch. Local

mineral matter end-members were used to reflect the differing geology. No optimisation was undertaken. Site details are shown in Table 2, and the chosen mineral matter end-members are listed in the caption to Fig. 7.

The results reveal good fits with low bias and noise for May Moss peat and marine diatom for the sites in Wales and Sweden, with rather noisier results at Stemmen which has a rather unusual bedrock type. At all three sites mineral matter is the least well fitted. The results are therefore fully consistent with our experiments at Loch Grannoch; fitted organic matter is least impacted by choice of mineral matter end-member, and fitted mineral matter most so.

## Discussion

The approach demonstrated here affirms previous work showing that taking the 1st derivative of NIR spectra is the most effective way of reducing unwanted physical information, for example the effects of particle size (Rinnan et al. 2009). Alternative normalisation procedures can also reduce or remove these signals, but have been shown not to enhance performance and are not recommended for pre-treatment (Dåbakk et al. 1999). Here the 1st derivative is calculated using a centrally-weighted 9-point Savitzky–Golay smoothing algorithm combined to reduce unwanted noise, remove the baseline offset and amplify the spectral curvature, all of which contribute



**Fig. 7** Fitted concentration of May Moss peat (a), marine diatom (b), and local mineral matter (c) for three additional sites. Site details in Table 2. Mineral matter end-member used: schist for Sotaure, local late glacial sediment at Llyn Cwm-

mynach, and orthoclase plus crushed anorthosite at Stemmen. 1:1 lines are plotted to emphasise any bias. Colour figure available online

**Table 2** Site details for additional cores shown on Fig. 7

Site name	Sotaure	Tarn at Stemmen	Llyn cwm Mynach
Country	Sweden	Norway	Wales
Latitude°	66.7333	58.4534	52.7958
Longitude°	20.5734	6.0870	-3.9608
Rock type	Granitic gneiss and schist	Anorthosite	Ordovician metasediments
Core code	SO-2-1	N17-Lake1-GC1	MYN15-SC2
Date of coring	15/07/2012	21/08/2017	25/06/2015
Lake area km <sup>2</sup>	0.016	0.015	0.057
Catchment area km <sup>2</sup>	0.042	0.064	1.19
Coring depth m	6.2	9.5	11
Core length m	0.95	2.57	0.35
Core type	Russian	Russian	8 cm diameter, Gravity

to it being regarded as the best mathematical pre-treatment choice for handling NIR spectra (Burns and Ciurczak 2001; Korsman et al. 2001; Pearson et al. 2014).

The 1st derivative spectra exhibit well-defined peaks, some of which may be attributed to specific chemical bonds (Terhoeven-Urselmans et al. 2006; Brown et al. 2006; Zornoza et al. 2008), with the intervals of the spectrum 4100–4500 cm<sup>-1</sup> and 5100–5300 cm<sup>-1</sup> dominated by organic chemical bonds (Korsman et al. 2001). This observation appears to be in contradiction with empirical evidence (Pearson et al. 2014) that model prediction skill is not improved by restricting analysis to regions of the spectrum. Our results, however, offer a simple explanation. If component spectra are additive, then each influences all parts of the spectrum, and all parts of the spectrum would contain information about each component of the mixture.

The question of whether spectra are fully additive, in essence that the spectrum of a mixture comprises a linear combination of its component spectra, is fundamentally important to quantitative interpretation of NIR spectra. This behaviour is expected on the basis of theory (i.e., Boroumand et al. 1992), but our analysis of artificial mixtures demonstrates this to be the case in practice too. For the spectra of synthetic binary mixtures of mineral and organic matter (Fig. 3) 99% of the variance can be explained by fitting the component spectra using multiple regression, which yields root mean squared differences between known

and inferred concentrations in the order of 1% for both mineral and organic matter.

Success in component fitting is harder to assess for natural heterogeneous mixtures, because of the uncertainty in the nature of the end-member materials. Fitting three different materials (May Moss peat, Loch Grannoch late glacial mud, and marine diatom) to the Loch Grannoch core produced adjusted R<sup>2</sup> values that ranged 0.85–0.97. The best fits are unsurprisingly for the late glacial mineral matter, where our end-member has been specifically tailored to be suitable. The failure to explain the mineral matter more fully can be attributed to variability in its composition, with varying proportions of the constituent minerals (quartz, feldspars, micas and chlorite). Through the Holocene there are a number of events and intervals of higher or lower R<sup>2</sup>, which may represent periods in which the component materials were slightly different, or periods during which additional materials different from the selected end-member materials were present. A crucial question is whether periods of poorer fit as measured by the R<sup>2</sup> of regression represent periods of poor quantification of the components or simply dilution by other materials. This cannot be fully answered, but the squared differences between the NIRS inferred and independently quantified values for the Loch Grannoch core do not co-vary with the R<sup>2</sup> value (Fig. 6). Thus, although more work is needed, it appears that the fitted concentrations are relatively insensitive to the proportion of total variance (R<sup>2</sup>) explained.

The additive nature of the 1st derivative NIR spectra (Fig. 2; Boroumand et al. 1992) means that it is possible to separate end-members quantitatively using multiple regression. However, we have shown that the multiple regression coefficients are not linearly related to mass concentration (Fig. 3). This result is consistent with the theoretical treatment of Boroumand et al. 1992, who show that varying chromophore density among materials means that conservative mixing by mass would be the exception rather than the rule. We demonstrate in the case of the binary mixture (Fig. 3) that application of a simple chromatic intensity factor (CIF) corrects for this effect, such that we observe conservative mixing of what may be termed the chromatic proportions. This is fully quantifiable, such that mass concentration (mass/mass) may be calculated from the chromatic proportions provided the CIF values are known. However, we cannot extract end-members from natural mixtures, so our measured CIF values must be obtained instead using proxies, the appropriateness of which cannot be fully assessed. In the case of the Loch Grannoch core, we have used (Fig. 6a) the measured  $CIF_M$  values for May Moss peat (definitively 1), Loch Grannoch mineral matter (2.26), and marine diatom (3.09). Yet, when we find  $CIF_O$  values by optimisation (minimising the squared differences from known values), a very different value is found for the mineral matter (1.4). The  $CIF_O$  for marine diatom is identical to the  $CIF_M$  value. The differing CIF values for mineral matter likely reflect the differences in the mineral matter mixture between the late glacial and the Holocene. This suggests that the issue of CIF magnitude is rather less important than the problem of being unable to choose end-members on a truly a priori basis. However, the sensitively testing (Fig. 5) shows that useful results are obtained even when an imperfect end-member material is used. And, when using optimal CIF values, it is apparent (Fig. 6b) that multiple regression very successfully explains both magnitude and variation in the independently quantified components (biogenic silica, organic and mineral matter). Quantification of major sediment components in the Loch Grannoch core using multiple regression compares well with the independently quantified values despite not being trained in the manner that the PLS methods are. Indeed, this favourable comparison is still more encouraging given that the PLS method was trained using the independently quantified variable data set

and would therefore be expected to predict it successfully, while our method did not include a training procedure.

Choice of organic materials can affect the fit for other major sediment components. In the case of the Loch Grannoch core, May Moss peat provides the best proxy for lake sediment organic matter. This is likely due to it comprising a mixture of organic materials which would be similar to those found in the lake sediment that are allochthonous in origin. It is possible that other combinations of organic materials would replicate this natural mixture. This has not been fully explored, but other examples of good fits for Loch Grannoch organic matter are combinations of humic acid, fulvic acid and *Sphagnum* spp. Generally, with the exception of May Moss peat, precision of fitting is improved when pairs or greater combinations of organic matter end-members, are used. Some combinations, owing to the similarity of spectra for some types of organic matter with consequent high multicollinearity, lead to excessively high ( $\gg 1$ ) or low ( $\ll 0$ ) regression coefficients. However, with favourable combinations, such as humic acid, fulvic acid and *Sphagnum*, coefficients generally lie between 0 and 1 even when both biogenic silica and mineral matter are also included as independent variables. This is highly promising, though further work is needed to test the validity of such results.

Due to a limited end-member library of biogenic silica materials, we cannot generalise with confidence. However, the similarity of results found using three widely differing sources of biogenic silica, and the good agreement with independently quantified biogenic silica for the Loch Grannoch core, are very promising.

The broader generality of our method is illustrated by its successful application, without optimisation, to sites with different characteristics and climate from Wales, Norway and Sweden. It is particularly encouraging that good results were obtained when applying May Moss peat and marine diatom, with only mineral matter choice adjusted for local conditions. On the other hand, based on the results from these three sites, and the experiments at Loch Grannoch, we can expect poorer results where the sediment mineral component is poorly represented by the selected end-member, so at new sites, particularly those with unknown or unusual bedrock types, some validation of the end-member fits is recommended.

## Conclusions

With suitable end-member materials selected, our new open-source multiple regression procedure gives simultaneous quantification of the major components (biogenic silica, mineral and organic matter content) of lake sediments, with excellent performance compared with results obtained by independent methods. Poor choice of end-member can lead to bias in the quantification, thus tailoring the choice of mineral material, for example selecting a similar bedrock type to that of the sample environment, is the preferred option. An advantage of the multiple regression method is that it does not require a training data set of materials which have been independently quantified using alternative methods. A library of end-member materials has been created relatively simply with little time investment and is easily expanded. Fitting of end-member spectra by multiple regression is a valuable alternative to the various PLS methods (or other 'trained' methods), that do not transparently allow assessment of potential interfering factors. It has the further advantage of resting in theory rather than on chemometric statistical methods.

**Acknowledgements** Thank you to Forestry Commission Scotland for providing us with access to the coring site at Loch Grannoch, and to Poonperm Vardhanabindu, Jenny Bradley and Maddy Moyle for assistance in the field. FER PhD is supported by the School of Environmental Sciences, University of Liverpool.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Ballantyne CK, Rinterknecht V, Gheorghiu DM (2013) Deglaciation chronology of the Galloway Hills Ice Centre, southwest Scotland. *J Quat Sci* 28:412–420. <https://doi.org/10.1002/jqs.2635>
- Boroumand F, Moser JE, Van Den Bergh H (1992) Quantitative diffuse reflectance and transmittance infrared spectroscopy of nondiluted powders. *Appl Spectrosc* 46:1874–1886. <https://doi.org/10.1366/0003702924123502>

- Boyle JF (2001) Inorganic geochemical methods in palaeolimnology. In: Last WM, Smol JP (eds) *Tracking environmental change using lake sediments*. Kluwer Academic Publishers, Dordrecht, pp 83–141
- Boyle J (2004) A comparison of two methods for estimating the organic matter content of sediments. *J Paleolimnol* 31:125–127. <https://doi.org/10.1023/B:JOPL.0000013354.67645.df>
- Boyle J, Chiverrell R, Schillereff D (2015) Approaches to water content correction and calibration for  $\mu$ XRF core scanning: comparing x-ray scattering with simple regression of elemental concentrations. In: Croudace IW, Rothwell RG (eds) *Micro-XRF studies of sediment cores*. Springer, Dordrecht, pp 373–390
- Brown DJ, Shepherd KD, Walsh MG, Mays MD, Reinsch TG (2006) Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* 132:273–290. <https://doi.org/10.1016/j.geoderma.2005.04.025>
- Burns DA, Ciurczak EW (2001) *Handbook of near-infrared analysis*. M. Dekker, New York
- Dåbakk E, Nilsson M, Geladi P, Wold S, Renberg I (1999) Sampling reproducibility and error estimation in near infrared calibration of lake sediments for water quality monitoring. *J Near Infrared Spectrosc* 7:241–250. <https://doi.org/10.1255/jnirs.254>
- Deer WA, Howie RA, Zussman J (1966) *An introduction to the rock-forming minerals*. Longmans, Harlow
- Flower RJ, Battarbee RW, Appleby PG (1987) The recent paleolimnology of acid lakes in Galloway, Southwest Scotland—diatom analysis, pH trends and the role of afforestation. *J Ecol* 75:797–824. <https://doi.org/10.2307/2260207>
- Garrels RM, Mackenzie FT (1971) *Evolution of sedimentary rocks*. Norton, New York
- Hayes MHB, Swift RS, Wardle RE, Brown JK (1975) Humic materials from an organic soil: a comparison of extractants and of properties of extracts. *Geoderma* 13:231–245. [https://doi.org/10.1016/0016-7061\(75\)90020-8](https://doi.org/10.1016/0016-7061(75)90020-8)
- Korsman T, Renberg I, Dabakk E, Nilsson MB (2001) Near-infrared spectrometry (Nirs) in palaeolimnology. In: Last WM, Smol JP (eds) *Tracking environmental change using lake sediments*. Kluwer Academic Publishers, Dordrecht, pp 299–317
- Malley DF, Williams P (2014) Analysis of sediments and suspended material in lake ecosystems using near-infrared spectroscopy: a review. *Aquat Ecosyst Health Manage* 17:447–453. <https://doi.org/10.1080/14634988.2014.979311>
- Malley DF, Rönické H, Findlay DL, Zippel B (1999) Feasibility of using near-infrared reflectance spectroscopy for the analysis of C, N, P, and diatoms in lake sediments. *J Paleolimnol* 21:295–306. <https://doi.org/10.1023/A:1008013427084>
- Meyer-Jacob C, Vogel H, Boxberg F, Rosén P, Weber ME, Bindler R (2014) Independent measurement of biogenic silica in sediments by FTIR spectroscopy and PLS regression. *J Paleolimnol* 52:245–255. <https://doi.org/10.1007/s10933-014-9791-5>
- Pearson EJ, Juggins S, Tyler J (2014) Ultrahigh resolution total organic carbon analysis using Fourier Transform Near Infrared Reflectance Spectroscopy (FT-NIRS). *Geochem*

- Geophys Geosyst 15:292–301. <https://doi.org/10.1002/2013GC004928>
- Rinnan Å, van den Berg F, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal Chem* 28:1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- R Core Team (2013). R: a language and environment for statistical computing. R Foundation for statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Terhoeven-Urselmans T, Michel K, Helfrich M, Flessa H, Ludwig B (2006) Near-infrared spectroscopy can predict the composition of organic matter in soil and litter. *J Plant Nutr Soil Sci* 169:168–174. <https://doi.org/10.1002/jpln.200521712>
- Workman JJ, Mobley PR, Kowalski BR, Bro R (1996) Review of chemometrics applied to spectroscopy: 1985–1995, part I. *Appl Spectrosc Rev* 31:73–124. <https://doi.org/10.1080/05704929608000565>
- Zornoza R, Guerrero C, Mataix-Solera J, Scow KM, Arcenegui V, Mataix-Beneyto J (2008) Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol Biochem* 40:1923–1930. <https://doi.org/10.1016/j.soilbio.2008.04.003>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.