Pearson EJ, Juggins S, Tyler J.

Ultra-high resolution total organic carbon analysis using Fourier-Transform Near Infrared Reflectance Spectroscopy (FT-NIRS).

Geochemistry, Geophysics, Geosystems 2014, 15(1), 292-301.

Copyright:

©2013. The Authors. Geochemistry, Geophysics, Geosystems published by Wiley Periodicals, Inc. on behalf of American Geophysical Union.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI link to article:

http://dx.doi.org/10.1002/2013GC004928

Date deposited: 20th March 2014

This work is licensed under a Creative Commons Attribution 3.0 Unported License

http://eprint.ncl.ac.uk
Ultrahigh resolution total organic carbon analysis using Fourier Transform Near Infrared Reflectance Spectroscopy (FT-NIRS)

Emma J. Pearson and Steve Juggins
School of Geography, Politics and Sociology, Newcastle University, Newcastle-upon-Tyne NE1 7RU, UK (emma.pearson@ncl.ac.uk)

Jonathan Tyler
School of Earth Science, University of Melbourne, Carlton, Victoria, Australia

[1] Fourier transform near infrared reflectance spectroscopy (FT-NIRS) is a cheap, rapid, and nondestructive method for analyzing organic sediment components. Here, we examine the robustness of a within lake FT-NIRS calibration using a data set of almost 400 core samples from Lake Suigetsu, Japan, as a means to rapidly reconstruct % total organic carbon (TOC). We evaluate the best spectra pretreatment, examine different statistical approaches, and provide recommendations for the optimum number of calibration samples required for accurate predictions. Results show that the most robust method is based on first-order derivatives of all spectra modeled with partial least squares regression. We construct a TOC model training set using 247 samples and a validation test set using 135 samples (for test set $R^2 = 0.951$, RMSE = 0.280) to determine TOC and illustrate the use of the model in an ultrahigh resolution (e.g., 1 mm/annual) study of a long sediment core from a climatically sensitive archive.

Components: 6313 words, 3 figures, 1 table.

Keywords: near infrared spectroscopy (NIR); total organic carbon (TOC); Lake Suigetsu; high resolution; paleoclimate

Index Terms: 1094 Instruments and techniques: Geochemistry; 1055 Organic and biogenic geochemistry: Geochemistry; 1694 Instruments and techniques: Global Change; 9320 Asia: Geographic Location.

Received 5 July 2013; Revised 13 December 2013; Accepted 22 December 2013; Published 31 January 2014.


1. Introduction

[2] High resolution studies of sediment archives are becoming increasingly important for understanding the rates and timings of abrupt climate changes, and for the detailed reconstruction of change in different areas that provides important information about causal links between regions, the understanding of which is a key issue in climate studies and Quaternary geology.

[3] Total organic carbon (TOC) is a basic and fundamental property of lake sediments and component of the carbon cycle [e.g., Boyle, 2001]. Sediment TOC is an important indicator of environmental and surface water quality [e.g., Ouyang et al., 2006] and has a major influence on biogeochemical processes, nutrient cycling, and redox potential that take place in sediments. Inland waters are a significant component of the global carbon cycle and sedimentary TOC provides
information on changes in lake productivity in response to environmental and climate change, and is used to quantify carbon storage in lakes, which are increasingly recognized as important carbon sinks [e.g., Larsen et al., 2011]. Changes in climate result in changes in the burial and release of carbon and it is therefore of relevance and importance to study these processes in the past.

[5] TOC is also commonly used in normalization during calculation and evaluation of organic biomarker proxies in sediments in environmental and paleoenvironmental studies [e.g., Pearson et al., 2007] or in contamination studies [e.g., Malley et al., 2000]. As part of the Lake Suigetsu 2006 Project (www.suigetsu.org), the analysis of organic geochemical biomarkers is being carried out at decadal resolution during key climatic periods which, combined with the long (~150 ka) and climatically significant nature of the Lake Suigetsu sediments, initiated our interest in developing a robust FT-NIRS method to obtain high resolution TOC data for this site. Near infrared reflectance spectroscopy (NIRS) is an inexpensive, rapid (sample analysis in triplicate takes approximately three minutes), nondestructive, quantitative, technique that requires small sample sizes (0.01–0.1 g dry weight), and minimal sample pretreatment (drying and grinding/homogenizing) meaning samples can be used for later analyses (e.g., biomarkers). NIRS works through vibrational excitation in the form of stretching and bending of bonds between atoms within organic materials. Differences in the chemical structure and composition of organic molecules corresponding to sediment composition display characteristic vibrations which are expressed as absorption at particular wavelengths in the NIR spectrum. By using multivariate statistics, it is possible to explore relationships between NIRS spectra and sediment properties for use in paleoenvironmental studies [e.g., Korsman et al., 2001]. In Fourier transform near infrared reflectance spectroscopy (FT-NIRS) all frequencies are measured simultaneously which enables rapid measurement of multiple scans per second, higher signal/noise ratios, and enhanced repeatability via an internal reference laser.

[5] FT-NIRS has been used in a wide range of environmental applications, including the monitoring of river mining contamination [Kemper and Sommer, 2002; Malley et al., 1996; Persson et al., 2007; Xia et al., 2007], determining the composition of peat [McTiernan et al., 1998] and lake sediments [Malley et al., 1999, 2000], and the reconstruction of lake-water chemistry (e.g., pH [Korsman et al., 1992]; pH, TP, and TOC [Nilsson et al., 1996]; TOC [Rosén, 2005]) and climate [Inagaki et al., 2012; Rosén, 2005; Rosén et al., 2001]. It has also been used, along with near ultraviolet and visible reflectance spectroscopy to examine TOC, carbonate, and opal content of marine sediments [Balsam and Deaton, 1996; Chang et al., 2005; Jarrard and Vanden Berg, 2006; Leach et al., 2008]. These applications of NIRS make use of a calibration data set of either surface or sediment core samples which are used to model the relationship between NIR spectra and the parameter of interest measured using traditional, often time-consuming, methods. Choice of calibration method is thus crucial to the success of the method but most NIRS studies do not report on spectral data treatment comparisons or examine the potential and merits of different statistical approaches. In addition, no studies to date have examined NIR at different resolutions or assessed the number of samples required for optimal model construction.

[5] The aim of this study was to examine these questions and construct a robust calibration model for Lake Suigetsu sediments using FT-NIRS to enable rapid, high resolution reconstruction of TOC via development of a model that can be used in on-going studies at this site. In addition, the methods and approaches developed here should be applicable to a range of sedimentary environments including lakes, peats, and marine sediments. In order to construct the most robust model possible, we examine (i) the effects of different data pretreatment, (ii) the most appropriate wavelength range to include in the calibration model, (iii) the most robust statistical method to employ for model construction, and (iv) the effect of calibration data set size on the accuracy of the TOC model outputs. We also employ rigorous statistical techniques for model cross-validation. This is the first extensive within-lake FT-NIRS calibration study which explores all these features to obtain the most robust model for use to reconstruct TOC at ultra-high (1 mm) resolution.

2. Materials and Methods

2.1. Study Site: Lake Suigetsu

[7] Lake Suigetsu (35°35′N, 135°53′E, 0 m above present sea level) is a meromictic, tectonic lake located on the Sea of Japan coast, Honshu Island, central Japan. The lake is 34 m deep and covers an area of c. 4.3 km². It is one of the “five Mikata
lakes” (within the Mikata-goko Ramsar site), with a regional climate typically characterized by both summer and winter monsoons. The location of the lake in relation to the Asian monsoon boundary front renders it ideal for providing important palaeoclimatic information about changes in this boundary in the past.

[5] In 2006, as part of the “Lake Suigetsu Varves 2006” Project (www.suigetsu.org), a 74 m long core (SG06) covering the last c.150 ka was collected from the lake. The upper 46 m of sediment consists of varves (covering c. 60 ka), contains abundant fossil leaves, more than 800 of which have been 14C dated [Bronk Ramsey et al., 2013], and offers a remarkably well-dated, high resolution multiproxy archive of palaeoenvironmental change [e.g., Kossler et al., 2011; Nakagawa et al., 2003, 2012]. Lake Suigetsu is an auxiliary Global Stratotype Section and Point (GSSP) for the onset of the Holocene Epoch [Walker et al., 2009] which also highlights the climatic and global significance of this site.

2.2. Conventional TOC Analyses

[5] For conventional TOC analyses samples were extracted at 15 cm intervals throughout core SG06 (n = 382). Samples were dried, ground, and weighed into Ag capsules, acidified using 20% HCl, then redried before placing the Ag capsule into a larger Sn capsule. C and N concentrations were measured by combustion EA (FlashEA 1112) calibrated against a sulfanilamide standard. Analytical reproducibility of ±0.5% was calculated through repeat analyses of an in-house standard (marine sediment).

2.3. FT-NIRS Analyses

[10] For the calibration exercise dried and ground samples as used for conventional analyses were analyzed in triplicate and spectra averaged before statistical analysis. For the high-resolution core study samples were analyzed at 1 cm and also 1 mm resolution across specific boundaries of interest such as the lateglacial-Holocene onset as a means to examine the robustness of the calibration at different resolutions. Each sample was scanned 64 times over a wavelength range of 835–2500 nm (12,000–4000 cm⁻¹) at 1 nm resolution using a Fourier transform spectrometer (ThermoNicolet Nexus 870) equipped with a Quartz beamsplitter, InGaAs (indium gallium arsenide) detector and a top loading diffuse reflectance Smart Near-IR UpDRIFT sampling accessory. The Updrift accessory and method maximizes diffusely scattered radiation whilst minimizing scatter and specular reflection/interference. A Spectralon® white reference, collected every 30 min, was used to standardize sample spectra. Reflectance values range between 0.569 and 0.913 and triplicate samples had a mean error across all wavelengths of 0.38%, indicating high sample reproducibility. Diffuse reflectance (R) was transformed to absorbance (A) using the equation A = log(1/R) and the spectra for each sample smoothed using a running mean of width 4 to remove spectral noise and give a final spectral data set of 416 wavelengths at 4 nm intervals.

2.4. Spectral Preprocessing and Statistical Analyses

[11] NIR spectra of lake sediments are complex and contain hundreds or thousands of wavelengths which are highly collinear. Predicting sediment properties from spectral data is thus a problem of multivariate calibration. One common solution is to use a data reduction technique such as partial least squares regression (PLS) [Naes et al., 2002] to reduce the spectral data to a small number of components that maximize the covariance between spectra and the sediment property of interest. PLS is particularly suited to modeling NIR data where the spectra are highly multicollinear and the number of variables (wavelengths) often greatly exceeds the number of samples. Most PLS calibrations use all near infrared wavelengths as predictor variables [e.g., Rosén et al., 2010] although some have argued that noninformative wavelengths should be omitted as they may increase random error [Zou et al., 2010], or have found that the main absorption bands influencing the calibrations lie within a narrower wavelength range (e.g., Malley et al., [1996] use 1500–2498 nm and Leach et al., [2008] use 2100–2500 nm). We therefore compare PLS using all and selected wavelengths, omitting those that have an absolute value of the PLS regression coefficient of <10% of the maximum [Garrido Frenich et al., 1995]. The optimal number of PLS components for each model was determined using a randomization t-test [van der Voet, 1994].

[12] An alternative to data reduction methods are data-mining tools that seek to extract relationships from high-dimensional data in the form of predictive models, or rules, based on a reduced set of the original predictors. These methods have the advantage that they omit the data reduction step
and select directly the best wavelengths for prediction. Here, we compare the performance of PLS to two data mining techniques, random forests (RF) and regression rules (RR). Random forests are an extension of classification and regression trees (CART) that builds an ensemble model based on many regression trees [Breiman, 2001]. Each individual tree is trained on a bootstrap sample of the training set data consisting of a random subset of predictor variables. By averaging predictions over a large number of trees random forests overcome the problem of instability in individual trees and reduce model uncertainty [Minasny and McBratney, 2008]. Random forest models were developed using the default to 500 trees. In regression trees, the prediction is a value at each node of the tree. Regression rules are similar to regression trees except that each node in the tree represents a rule consisting of a linear regression model based on the predictors used in previous splits [Quinlan, 1992]. Regression rules can also use a boosting-like scheme in which the final prediction is averaged over a number of trees or committees (25 in our case, chosen by cross-validation), to reduce model bias and variance.

NIR spectra of particulates often contain uninformative information resulting from light scattering effects that vary between samples due, for example, to differences in particle size. It is therefore usual to apply one or more spectral pretreatments to remove these unwanted effects [e.g., Dåbakk et al., 1999b; Rosén and Hammarlund, 2007]. A range of numerical pretreatments have been applied to spectra from lake and marine sediments [e.g., Balsam and Deaton, 1996; Korsman et al., 2001; Rosén et al., 2000] but with no clear guidelines on which is most appropriate. We therefore test a number of different spectral pretreatments and compare them to no pretreatment: (1) standard normal variate (SNV) which centers and scales each spectrum by its mean and standard deviation, respectively (so called autoscaling), (2) multiplicative scatter correction (MSC), (3) first-order derivative using the Savitsky-Golay algorithm with an order two and interval width of seven, (4) MSC followed by first-order derivative, and (5) second-order derivative with order two and interval width of 11 variables.

The performance of the above models and pretreatments was assessed using the squared correlation between observed and predicted TOC values, and the root mean squared error (RMSE), a measure of the average error of the prediction. RMSE is usually estimated using leave-one-out cross-validation [e.g., Inagaki et al., 2012]. However, down-core calibration data may be temporally auto-correlated, meaning leave-one-out cross-validation does not properly simulate an independent test data set, and ultimately leads to an underestimation of the likely prediction error when the calibration is applied to new data [Burman et al., 1994]. Correlograms and partial correlograms of the Suigetsu TOC calibration data (not shown) demonstrate significant temporal autocorrelation at lags of up to five samples. We therefore use a pseudo-h-block cross-validation scheme and split the data into two parts: approximately two thirds \((n = 247)\) for calibration development and one third \((n = 135)\) for validation, with the latter split into three groups of 45 samples each, arranged in equal-spaced blocks down-core (Figure 1). Tuning of model parameters was determined using 10-fold leave-out using the calibration set.

[15] All analyses were performed using R software [R Core Team, 2013] with the packages rioja [Juggins, 2012], Cubist [Kuhn et al., 2013], randomForest [Liaw and Wiener, 2002], and caret [Kuhn, 2013].

3. Results and Discussion

3.1. Spectral Pretreatment, Wavelength Selection, and Model Performance

[16] Table 1 compares the performance of PLS, random forest (RF), and regression rules (RR) models with different spectra pretreatments for the validation data. Prediction errors for the training set derived using leave-one-out (LOO: not shown) underestimate the validation set errors by between 6% and 70% (mean 20%) which suggests that published errors using LOO are actually underestimates of the true prediction error. Prediction errors for our final model (see below) using LOO and h-block cross-validation are 0.235 and 0.280, respectively, highlighting the importance of using an appropriate h-block cross-validation design that properly accounts for the temporal dependency in the training set data [Arlot, 2010].

[17] Random forests are consistently poor with the highest RMSEP (Table 1). Plots of variable (i.e., wavelength) importance (not shown) reveal models that are dominated by only a small number of wavelengths and predictions that are overestimated or underestimated at low and high TOC
values, respectively. For PLS and regression rules, preprocessing the spectra prior to analysis gave variable results: MSC and second-order derivatives reduced model performance in all cases. MSC in particular is frequently used in mid-infrared (FTIR) studies [e.g., Cunningham et al., 2011; Hahn et al., 2011; Rosén et al., 2010, 2011; Rouillard et al., 2011; Vogel et al., 2008] and has also been used in visible-NIR [Cunningham et al., 2011; Rouillard et al., 2011]. For near infrared (FT-NIR), it can improve predictions for some variables and has been used in some studies to eliminate light scattering effects arising from variables such as differences in particle size [e.g., Dåbakk et al., 1999a; Rosén and Hammarlund, 2007], but has also been found not to enhance performance [Dåbakk et al., 1999b] and it is not uniformly recommended [e.g., Minasny and McBratney, 2008]. Our results highlight the need to evaluate particular pretreatments for each study.

[19] First-order or second-order derivatives are another way of treating the data to achieve similar

Figure 1. Experimental setup: (a) number of samples in (top) calibration data set and (bottom) test set and range of TOC covered; (b) measured and predicted TOC using final PLS model using all wavelengths with a combination of SNV and first derivative pretreatment; (c) application of model down-core: top ticks indicate calibration samples, bottom ticks indicate validation samples, dotted black line = measured TOC, solid gray line = FT-NIRS predicted TOC.
outcomes to MSC since it removes baseline offset and has been used in some NIR studies [e.g., Korsman et al., 2001; Malley et al., 1996, 2000]. One problem with derivatives is that they amplify spectral curvature which may also increase random noise and result in poor models [e.g., Dåbakk et al., 1999b] although it appears that our initial spectral smoothing has reduced this effect. In our study, first-order derivatives improves predictions with PLS and give the lowest overall prediction errors when used in conjunction with SNV (standardized normal variance) scaling (autoscaling).

In our data set regression rules with spectral pretreatment using first or second derivatives builds poor ensemble models that utilize only a small number of wavelengths. Regression rules with no-spectral pretreatment yields a model that is competitive with the best PLS model. It produces very similar predictions (median difference 0.11% TOC) but at the expense of increased complexity [cf. Minasny and McBratney, 2008]. Straightforward PLS on raw data has also been found to perform best in a study by Dåbakk et al. [1999b] who examined modeling performance in filtered seston.

Similarly, in our data set, PLS in conjunction with wavelength selection generally performed no better than a similar model/pretreatment with all wavelengths. The region between 2100 and 2500 nm has been associated with organic carbon bonding [Osborne and Fearn, 1986] and reduced spectra data sets have been used in some FT-NIRS studies (e.g., 2100–2500 nm [Leach et al., 2008]; 1100–2500 nm [Korsman et al., 1999]; 1500–2498 nm [Balsam and Deaton, 1996]). Rouillard et al. [2011] point out in their study of visible-NIRS (VNIRS) investigations how different regions of spectra influence PLS analysis and could improve model performance, as well as reduce complexity, and that implementing the technique for wavelength weight on multivariate analyses should improve overall reliability of VNIRS-based models. Although wavelength selection has been used in some mid-IR studies [e.g., Vogel et al., 2008], in near IR [Das, 2007], and in the marine sediment NIR study of Leach et al. [2008], generally studies do not examine effects of different wavelengths or outline why spectral ranges were chosen, presumably choosing to use the NIR range available. In our study, we observe no benefit to model performance in including the additional complexity of a wavelength selection step and opt for a PLS using all wavelengths (835–2500 nm) with a combination of SNV and first derivative pretreatment as the most parsimonious model with the lowest prediction error.

<table>
<thead>
<tr>
<th>Method</th>
<th>Pretreatment</th>
<th>All Wavelengths</th>
<th>Reduced Wavelengths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r²</td>
<td>RMSEP</td>
<td>N</td>
</tr>
<tr>
<td>PLS</td>
<td>None</td>
<td>0.918</td>
<td>0.364</td>
</tr>
<tr>
<td></td>
<td>MSC</td>
<td>0.836</td>
<td>0.534</td>
</tr>
<tr>
<td></td>
<td>SG1</td>
<td>0.942</td>
<td>0.306</td>
</tr>
<tr>
<td></td>
<td>SG2</td>
<td>0.947</td>
<td>0.350</td>
</tr>
<tr>
<td></td>
<td>MSC + SG1</td>
<td>0.819</td>
<td>0.640</td>
</tr>
<tr>
<td>PLS + SNV</td>
<td>None</td>
<td>0.917</td>
<td>0.365</td>
</tr>
<tr>
<td></td>
<td>MSC</td>
<td>0.859</td>
<td>0.485</td>
</tr>
<tr>
<td></td>
<td>SG1</td>
<td>0.951</td>
<td>0.280</td>
</tr>
<tr>
<td></td>
<td>SG2</td>
<td>0.914</td>
<td>0.421</td>
</tr>
<tr>
<td></td>
<td>MSC + SG1</td>
<td>0.873</td>
<td>0.542</td>
</tr>
<tr>
<td>RF</td>
<td>None</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>MSC</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>SG1</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>SG2</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>MSC + SG1</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>RR</td>
<td>None</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>MSC</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>SG1</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>SG2</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>MSC + SG1</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

*aPLS, partial least squares regression; SNV, standard normal variate; RF, random forest; None, no spectra pretreatment; MSC, data multiplicative scatter corrected before analysis; SG1, first-order derivative using the Savitsky-Golay algorithm with an order two and interval width of seven; SG2, second-order derivative with order two and interval width of 11; MSC + SG1, MSC followed by first-order derivative; N, number of wavelengths used; n/a, wavelengths not applicable for these methods.
as biogenic silica and clays which are negatively correlated with TOC. These minerals have their strong, primary absorptions in the mid-IR but also have weak overtone bands in the NIR range [Stenberg et al., 2010]. Our results suggest that any such effects are either negligible or can be ignored because if the composition of inorganic material varied down-core and had an adverse effect on model predictions we would expect a reduced-wavelength model (that is, one that dropped wavelengths associated with inorganic material) to have superior performance. This is not the case. However, it is possible that the confounding effects of inorganic material might restrict the development of a global, rather than site-specific, model for TOC prediction. Figures 1a and 1c show the experimental setup for the calibration and validation sets and Figure 1b shows the relationship between conventional and NIR-predicted TOC for the validation set using the optimal PLS model (PLS with first derivative spectral pretreatment applied to autoscaled data). Figure 1c shows both conventionally measured and FT-NIRS predicted TOC for the calibration data set down-core and highlights the fact that FT-NIRS modeled TOC very closely tracks the conventionally measured TOC. There is only a slight systematic trend to increasing error at higher values (>6% TOC; Figure 1b), probably as a result of fewer calibration samples in this range.

3.2. Application to High-Resolution Samples

[23] Selecting the best method and pretreatment, as listed in Table 1, we constructed a model and test set applying partial least squares (PLS) standard normal variate (SNV) regression to first-order derivative spectra. We apply our model to our Lake Suigetsu sediments at a range of resolutions (Figure 2). Figure 2a shows the application of our chosen model to a section of the Lake Suigetsu core at 1 cm resolution (solid line), highlighting the good match between the 1 and 15 cm (dots and dashes) resolution samples. Figure 2b magnifies a small section of the core from Figure 2a and shows the 1 cm samples superimposed on ultrahigh resolution (1 mm, c. annual) TOC modeled results. The results demonstrate that FT-NIRS can realistically be used to reconstruct TOC at a range of resolutions, including at 1 mm intervals, which in this study is the limit of subsampling that is possible. We also see that at higher resolution there are also many finer details in the reconstructed TOC values and our results highlight the internal variation that can occur in such high resolution studies (Figure 2), and therefore the potential importance of using FT-NIRS to examine high-resolution changes. These changes in terms of paleoenvironmental significance are being investigated further and are beyond the scope of this paper. The small sample size (0.01–0.1

![Figure 2. Application of model to Suigetsu core samples at different temporal resolutions. (top) Gray = measured TOC, black = modeled TOC, 1 cm resolution; (bottom) modeled TOC, gray = 1 cm data set, black = 1 mm resolution.](image-url)
g) required for FT-NIRS analysis is minimal compared to conventional techniques (0.5–1 g), meaning this technique is invaluable and has huge potential for use in ultrahigh resolution and multiproxy studies and where sediment availability is limited.

3.3. Optimum Calibration Data Set Size

[24] The final part of our study was to examine the optimum calibration data set size and to provide a guideline for future studies. Previously published NIR-TOC calibration studies range in sample number from a small down-core data set of 30 samples (20 in model and 10 in test set under external cross-validation to 65 samples with internal cross-validation) [Leach et al., 2008] and c. 20 samples [Korsman et al., 1992] to over 100 for surface sediment calibration data sets [e.g., Hahn et al., 2011; Cunningham et al., 2011]. None of these studies examine the effect calibration sample size has on prediction errors or what is the minimum number of samples required to develop an acceptable model.

[25] Our NIR-TOC data set contains a total of 382 samples (247 used for calibration and 135 used for validation). In order to examine the minimum number of samples needed to develop an accurate TOC calibration, we developed models of different sizes by randomly selecting samples from our calibration data set and plotted RMSE for the validation set versus number of samples in the calibration set (Figure 3). Results show the mean value and standard deviation for 50 random calibration sets of varying size. Error reaches a minimum with a calibration set of c. 120 samples and that there is little to be gained by expanding the training set beyond this number. This is larger than sample numbers generally used in calibration studies and suggests that the development of NIR calibrations are particularly useful for long core studies where large numbers of analyses are required such as in our study of Lake Suigetsu sediments. Likewise, it is applicable to other terrestrial (e.g., lake, peat) sites with long records, as well as marine sediment cores.

4. Conclusions

[26] The best NIR-TOC model with the lowest prediction error was obtained using PLS standard normal variate regression with first-order derivatives and using all wavelengths in the data set (835–2500 nm). Our data set consisted of a total of 382 samples but our analysis suggests that a calibration data set of c. 120 samples is sufficient, provided it covers the range of TOC values likely to be observed in the core. We highlight the use of FT-NIRS as a rapid and cheap method to reconstruct TOC at a range of resolutions up to ultrahigh resolution (1 mm/annual) in long sediment cores from climatically sensitive and significant archives.

Acknowledgments

[27] FT-NIRS analyses were carried out at Newcastle University during Natural Environment Research Council (NERC)
grant NE/G011001/1. Conventional TOC analyses were carried out at the Kochi Institute for Core Sample Research, Japan, during Japan Society for Promotion of Science (JSPS) fellowship PE07622. We thank Mathew Brown and Hayley McDowall for assisting with FT-NIRS, Minoru Ikehara and Yusuke Yokoyama for help in facilitating conventional TOC analyses, Takeshi Nakagawa for providing samples, and Tsuyoshi Haraguchi, Katsuya Gotanda, Hitoshi Yonenobu, Yusuke Yokoyama, Ryuji Tada, and Takeshi Nakagawa for help in sub-sampling the core. We also thank Hitoshi Yonenobu and an anonymous reviewer for their comments on the original manuscript. We thank Hitoshi Yonenobu and an anonymous reviewer for their comments on the original manuscript. This work contributes to the Suigetsu Varves 2006 Project (www.suigetsu.org).

References

Juggins, S. (2012), rioja: Analysis of Quaternary Science Data, R package version 0.8-2. [Available at http://cran.r-project.org/package=rioja.]

300


