Chapter 21
Prediction of Geochemical Composition from XRF Core Scanner Data: A New Multivariate Approach Including Automatic Selection of Calibration Samples and Quantification of Uncertainties

G. J. Weltje, M. R. Bloemsma, R. Tjallingii, D. Heslop, U. Röhl and Ian W. Croudace

Abstract A multivariate log-ratio calibration (MLC) model for XRF-core-scanning devices is presented, based on a combination of basic XRF-spectrometry theory and principles of compositional data analysis. The performance of the MLC model is evaluated in comparison with other empirical calibration procedures for XRF core scanner data using two data sets acquired with two different XRF core scanners. The quality of calibration models is assessed by calculating the uncertainties associated with predicted concentrations using cross-validation techniques. Results show that

G. J. Weltje (✉)
Department of Earth and Environmental Sciences, University of Leuven, Celestijnenlaan 200E, 3001 Leuven-Heverlee, Belgium
e-mail: gertjan.weltje@ees.kuleuven.be

M. R. Bloemsma
Department of Geoscience and Engineering, Delft University of Technology, Stevinweg 1, 2628CN Delft, The Netherlands

R. Tjallingii
Section 5.2 Climate Dynamics and Landscape Evolution, GFZ, German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany

Department of Marine Geology, NIOZ Royal Netherlands Institute for Sea Research, Den Burg, P.O. Box 59, 1790 AB Texel, The Netherlands

D. Heslop
Research School of Earth Sciences, The Australian National University, Canberra, ACT 0200, Australia

U. Röhl
MARUM—Centre for Marine Environmental Sciences, University of Bremen, Leobener Strasse, 28359 Bremen, Germany

I. W. Croudace
Ocean and Earth Science, National Oceanography Centre, University of Southampton, Waterfront Campus, European WaySouthampton SO14 3ZH, UK

I. W. Croudace, R. G. Rothwell (eds.), *Micro-XRF Studies of Sediment Cores*, Developments in Paleoenvironmental Research 17, DOI 10.1007/978-94-017-9849-5_21

(1) the commonly used direct linear calibration (DLC) methods, which are based on the questionable assumption of a unique linear relation between intensities and concentrations and do not acknowledge the compositional nature of the calibration problem, give poor results; (2) the univariate log-ratio calibration (ULC) model, which is consistent with the compositional nature of the calibration problem but does not fully incorporate absorption and enhancement effects on intensities, and permits estimation of "relative" concentrations only, is markedly better, and (3) the MLC algorithm introduced in this contribution, which incorporates measurement uncertainties, accommodates absorption and enhancement effects on intensities, and exploits the covariance between and among intensities and concentrations, is the best by far. The predictive power of the MLC model may be further increased by employing automatic sample selection based on the multivariate geometry of intensity measurements in log-ratio space. The precision attained by MLC in conjunction with automatic sample selection is comparable to that attained by conventional XRF analysis of heterogenous materials under laboratory conditions. A solution to the long-standing problem of XRF core scanner calibration implies that high-resolution records of sediment composition with associated uncertainties can now be routinely established, which should increase the range of quantitative applications of XRF-core-scanning devices and strengthen inferences based on analysis of geochemical proxies.

Keywords Log-ratio transformation · Partial least squares regression · Compositional data · Chemometrics · Palaeoceanography · Micro-XRF core analysis

Introduction

X-ray fluorescence analysis (XRF) is a well-established technique for determining the composition of rocks and sediments (Ramsey et al. 1995; Jenkins 1999; De Vries and Vrebos 2002). Progress in XRF instrumentation has opened the way to on-line measurement of (soft) sediment cores aboard ship by means of XRF-core-scanning devices (Jansen et al. 1998; Wien et al. 2005; Ge et al. 2005; Rothwell and Rack 2006). Technical descriptions of the various XRF core scanners in current use are given by Jansen et al. (1998), Croudace et al. (2006), Haschke et al. (2002), Haschke (2006), and Richter et al. (2006) and summarised in this volume by Jarvis et al. The major advantage of XRF core scanning over conventional geochemical analysis of discrete specimens is that element intensities are obtained directly at the surface of a split sediment core, which implies up to two orders of magnitude increase in analytical speed. In addition, the spatial resolution of XRF core-scanning devices is much higher than that of conventional destructive methods, and allows the extraction of near-continuous records of element intensities from sediment cores. XRF core scanning is therefore recognised as a technique with great potential for palaeoenvironmental research (Calvert and Pedersen 2007; Weltje and Tjallingii 2008).

In order to realize the full potential of high-resolution XRF core scanning, a universally applicable, robust procedure must be developed for converting core scanner output to quantitative measures of sediment composition with associated

measures of uncertainty. Previous attempts to convert XRF core scanner output to element or oxide concentrations by means of linear regression on an element-by-element basis (referred to as direct linear calibration; DLC) have been only moderately successful (Jansen et al. 1998; Jaccard et al. 2005; Croudace et al. 2006; Kido et al. 2006; Böning et al. 2007; Tjallingii et al. 2007).

Direct linear calibration (DLC) models take the following form:

$$W_{ij} = a_j I_{ij} + b_j \tag{21.1}$$

where W_{ij} represents the concentration (weight proportion) of element j in specimen i. I_{ij} represents the net intensity of element j in specimen i, obtained by preprocessing of the raw spectrum by background subtraction, sum-peak and escape-peak correction, deconvolution and peak integration. Coefficients a_j and b_j are empirical constants specific to the data set and element under consideration.

The simplest DLC model is one in which direct proportionality is assumed, i.e. $b_j = 0$. In this model, referred to as DLC-1, concentrations and intensities are both constrained to be non-negative. A more flexible model, referred to as DLC-2, is obtained by allowing non-zero values of b_j in Eq 21.1. The DLC-2 model implies that the relation between W and I cannot be linear over their full range, because negative concentrations or intensities would occur. Therefore, the conventional interpretation of the DLC-2 model is that it represents the approximately linear relation between W and I over the limited range of intensities and concentrations covered by the data.

Practical problems associated with DLC models are apparent from considerable scatter and bias in cross-plots of intensity and concentration, which are attributable to inhomogeneity of the specimens (e.g. variable water content and grain-size distribution), irregularities of the split core surface, and in some setups, spatial variations in thickness of an adhesive pore-water film which forms directly below a protective foil covering the core surface. The conceptual problems of DLC are apparent from the fact that the seemingly straightforward (linear) calibration of intensities to concentrations is inconsistent with the parametric model used in XRF spectrometry (Weltje and Tjallingii 2008). The absence of a viable underlying physical model manifests itself in a number of ways: (1) None of the published regression equations passes through the origin, or alternatively, through the point representing the detection limits corresponding to the two measurements being compared; (2) Calibration on an element-by-element basis provides no guarantee that predicted concentrations are positive and sum to unity, which is a violation of fundamental physical constraints on compositional data.

In view of the problems associated with quantification of XRF core scanner output by DLC, the general view has been that such data should be regarded as semi-quantitative only (Croudace et al. 2006; Richter et al. 2006; Rothwell and Rack 2006). In a recent paper however, Weltje and Tjallingii (2008) showed that rigorous quantitative calibration of XRF core scanner output is possible. The key to this breakthrough was provided by a new mathematical formulation of the calibration equation in terms of log-ratios of intensities and concentrations, which is consistent with basic XRF-spectrometry theory, as well as the statistical theory of

compositional data analysis (Aitchison 1982, 1986). The present contribution documents the progress made since publication of the log-ratio calibration model of Weltje and Tjallingii (2008), and introduces a software package designed for efficient quantitative calibration of XRF core scanner output.

Calibration in Conventional XRF Spectrometry

In conventional (destructive) quantitative XRF analysis under well-constrained laboratory conditions, conversion of the net intensity of an element to a weight proportion is provided by the following general equation (Jenkins 1999; De Vries and Vrebos 2002):

$$W_{ij} = K_j I_{ij} M_{ij} S_i \tag{21.2}$$

Where W_{ij} and I_{ij} are defined as above, K_j represents a device-specific calibration constant for element j (the sensitivity or detection efficiency of the measurement device), and M_{ij} is the matrix effect which corrects for scattering, absorption and enhancement effects on I_{ij} caused by the presence of other elements in the specimen.

Note that for a series of specimens covering a range of compositions, the matrix effect is a non-linear function of the concentrations (or intensities) of the full range of elements present. S_i is the specimen effect which captures the measurement geometry and specimen homogeneity relative to the standard configuration.

Under laboratory conditions, K_j and S_i are fixed, and W_{ij} is estimated from I_{ij} with a correction factor given by M_{ij}. The matrix effect is commonly expressed as a function of the concentrations or intensities of the other elements present in the specimen under consideration. Various methods for estimating M_{ij} have been proposed, most of which are based on a combination of theory and empirical evidence (calibration specimens). Under ideal conditions, entirely theoretical methods for estimating M_{ij} (so-called fundamental parameter methods) may be utilized to predict net intensities based on known specimen compositions. Fundamental parameter methods are commonly implemented in the form of non-linear optimisation techniques, in which intensities calculated from an initial guess of a specimen's composition are compared to measured intensities, and the estimated composition of the sample is iteratively adjusted so as to minimise some measure of discrepancy between predicted and measured intensities (Jenkins 1999; De Vries and Vrebos 2002).

Log-Ratio Calibration

The fundamental problem in on-line XRF core scanning is that S_i is unconstrained, owing to inhomogeneity of the specimens (e.g. variable water, organic content, grain-size distribution of sediments, the presence of burrows, etc) and the irregular

surface of a split core. In some setups, spatial variations in thickness of an adhesive pore-water film which forms directly below a protective foil covering the split core surface also have to be considered (Kido et al. 2006; Tjallingii et al. 2007). Down-core variability of these factors implies that the measurement geometry S_i is not constant, contrary to XRF analysis under laboratory conditions, and Eq. 21.2 cannot be solved within reasonable limits of uncertainty, because every single measurement requires the solution of a set of calibration equations associated with a location-specific value of S_i. In other words, there is no unique relation between I and W, which implies that every core scanner measurement must be calibrated by means of destructive analysis, which would render the whole measurement strategy meaningless. This seemingly insurmountable problem may be solved by considering the problem in terms of log-ratios of element intensities and concentrations (Weltje and Tjallingii 2008).

Although log-ratio transformations (Aitchison 1982, 1986) have not been widely applied to geochemical compositions and XRF core scanner output, several authors have pointed to the advantage of using ratios of element intensities (or concentrations) instead of intensities (or concentrations) of single elements (Croudace et al. 2006; Richter et al. 2006; Rothwell et al. 2006; Calvert and Pedersen 2007; Löwemark et al. 2011). Down-core patterns of element-intensity ratios have proved extremely useful for correlation (e.g. Pälike et al. 2001; Vlag et al. 2004; Bahr et al. 2005), because they are unaffected by variations in the concentrations of other elements in a specimen (so-called dilution effects), which is especially relevant considering the fact that XRF core scanners do not measure the full range of elements.

Despite their usefulness, ratios do not permit rigorous statistical modelling of compositional data. Ratios have the undesirable property of asymmetry, i.e., conclusions based on evaluation of the ratio of two elements, say A/B, cannot be directly translated into equivalent statements about B/A. This implies that the results of statistical analysis of ratios depend on arbitrary decisions, since there is no Law of Nature to suggest which element should act as numerator or denominator. Fortunately, this problem was solved by Aitchison (1982), who discovered that rigorous statistical modelling of compositional data merely requires that compositions are expressed in terms of logarithms of ratios of component abundances (so-called log-ratios), in order to achieve the desired symmetry, and allow compositional data to be described with a unique set of statistics. Concepts and applications of compositional data analysis by means of log-ratios are covered by Aitchison (1986), Aitchison and Egozcue (2005), and in the monograph edited by Buccianti et al. (2006).

Weltje and Tjallingii (2008) derived a univariate log-ratio calibration (ULC) equation by combining two conventional calibration equations (Eq. 21.2) in the form of a ratio. They assumed that all values of W and I are positive (i.e., greater than or equal to the detection limits of the measurement devices used). Elements are indicated by subscripts j and D, whereas subscript i denotes the measurement location and the corresponding specimen:

$$\frac{W_{ij}}{W_{iD}} = \frac{K_j I_{ij} M_{ij} S_i}{K_D I_{iD} M_{iD} S_i} = \frac{K_j I_{ij} M_{ij}}{K_D I_{iD} M_{iD}}$$

Calibration in terms of ratios removes the measurement geometry from the problem to be solved, because S_i drops out of the equation. For each pair of elements, the ratio of detection efficiencies is a constant, the so-called relative detection efficiency:

$$\frac{K_j}{K_D} = B_{jD} \tag{21.3}$$

Further simplification is achieved by assuming that the ratio of matrix effects is a non-linear function of the ratio of measured element intensities:

$$\frac{M_{ij}}{M_{iD}} = C_{jD}\left(\frac{I_{ij}}{I_{iD}}\right)^{A_{jD}} \tag{21.4}$$

In the above expression, A_{jD} and C_{jD} are empirical coefficients specific to the pair of elements in the data set under consideration. Substitution of Eqs. 21.3 and 21.4 into Eq. 21.2 gives:

$$\frac{W_{ij}}{W_{iD}} = B_{jD}C_{jD}\left(\frac{I_{ij}}{I_{iD}}\right)^{A_{jD}+1} \tag{21.5}$$

Equation (21.5) may be rewritten by defining the empirical coefficients:

$$\alpha_{jD} = A_{jD} + 1 \tag{21.6a}$$

$$\beta_{jD} = \ln(B_{jD}C_{jD}) \tag{21.6b}$$

If we substitute Eqs. 21.6a and 21.6b into Eq. 21.5 and take logarithms, we obtain the ULC equation:

$$\ln\left(\frac{W_{ij}}{W_{iD}}\right) = \alpha_{jD}\ln\left(\frac{I_{ij}}{I_{iD}}\right) + \beta_{jD} \tag{21.7}$$

Equation 21.7 is a non-dimensional unconstrained linear equation expressed in terms of log-ratios of element intensities and concentrations. The coefficients α and β are the log-ratio equivalents of the matrix effect and detection efficiency in single-element XRF-spectrometry (Eq. 21.2), respectively. Although the variability of the specimen effect has been eliminated, the average values of variables such as grain-size, water content, and any matrix effects attributable to the presence of elements which have not been measured, will be reflected in the empirical model parameters α and β. The ULC equation is largely insensitive to down-core variability of these

quantities, which implies that the scatter of *I*-*W* log-ratio plots will be much smaller than the scatter observed in conventional ("raw") *I*–*W* plots under conditions of variable measurement geometry. An additional advantage of the above approach is that the relation between log-ratios of *I* and *W* is expected to be linear, unlike the relation between raw *I* and *W* data (cf. Eq. 21.1).

Solution of the log-ratio calibration equation for α and β yields positive concentration estimates $\tilde{W}_{ij}$ which sum to unity. This is easily shown by letting the log-ratios of concentrations predicted from log-ratios of measured intensities be:

$$x_{ij} = \ln\left(\frac{\tilde{W}_{ij}}{\tilde{W}_{iD}}\right) \tag{21.8a}$$

The above expression accounts for all elements but one: the element acting as common denominator, whose log-ratio value is by definition equal to zero: $x_{iD} = 0$. The predicted concentrations $\tilde{W}_{ij}$ of all elements, including the one used as common denominator ($j = 1, D$) are obtained by applying the inverse log-ratio transformation:

$$\tilde{W}_{ij} = \frac{e^{x_{ij}}}{\sum_{j=1}^{D} e^{x_{ij}}} \tag{21.8b}$$

The above method was used by Weltje and Tjallingii (2008) to obtain unbiased predictions of "relative" concentrations of elements which were measured by the XRF core scanner ("relative" means that the sum of all elements measured is automatically constrained to unity). Although ULC presents a major step forward compared to DLC, it does not allow "absolute" concentrations to be determined.

MLC Workflow

Ongoing research into the underlying mathematical form of the log-ratio calibration equation has resulted in a further generalisation of the estimation procedure proposed by Weltje and Tjallingii (2008). The second-generation multivariate calibration algorithm introduced in this contribution (MLC) makes effective use of the covariance between and among measured element intensities and concentrations, which reflect absorption and enhancement of intensities, as well as the fact that several elements may reside in the same mineral. The MLC model allows us to further minimize the uncertainty of predicted log-ratio quantities and extend predictions to elements which have not been measured by the XRF core scanner, so as to generate "absolute" concentrations. The MLC workflow also includes more rigorous estimation of prediction uncertainties and automatic selection of calibration samples. The automated MLC workflow is schematically shown in Fig. 21.1.

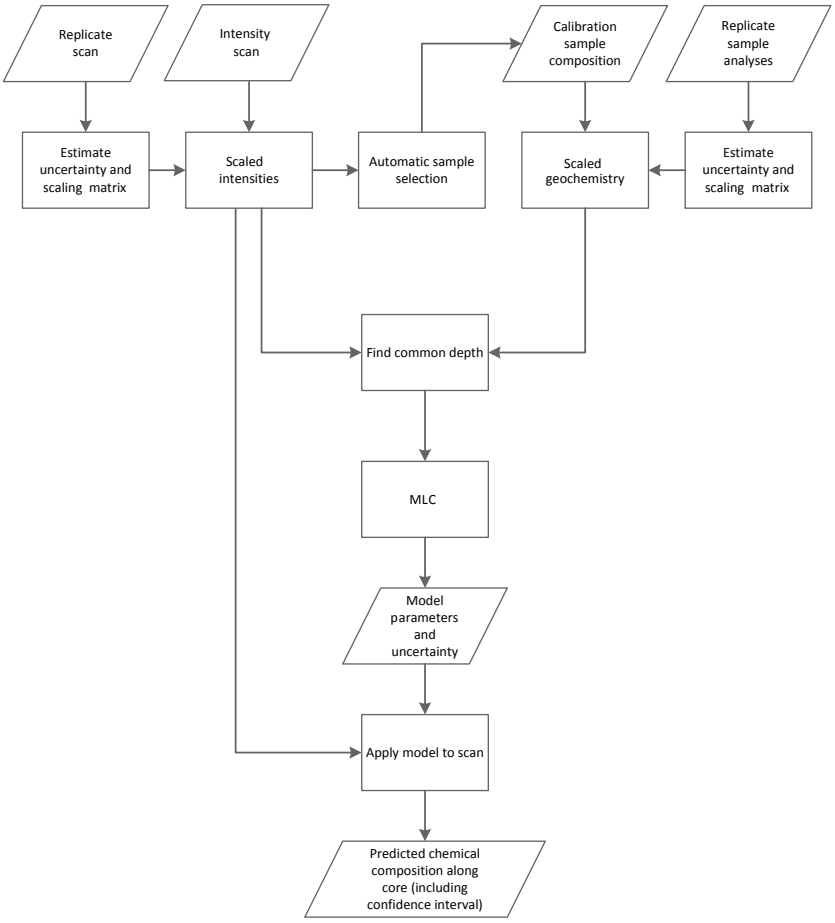

Fig. 21.1 MLC workflow; see text for discussion

Prediction of "Absolute" Element Concentrations

XRF core scanners measure a limited range of elements only. Hence, the sum of all detectable elements is equal to a proportion, p, of the total mass. In order to express log-ratio-calibrated quantities in terms of actual concentrations (and not just as mass fractions of the total number of elements actually measured), we must account for the total mass of the material analysed. We therefore introduce one additional unknown to be calibrated, which is simply the mass fraction of the material which cannot be assigned to specific elements:

$$\text{Undef} = W_{i,D+1} = 1 - \sum_{j=1}^{D} W_{ij} \tag{21.9}$$

In practice, "Undef" represents the union of all elements which have not been detected by the core scanner and by the chemical technique used to provide the reference concentrations for calibration. Prediction of "Undef" requires that calibration is formulated as a multivariate problem. The MLC workflow is therefore based on a slightly different log-ratio representation of compositional data relative to the ULC model of Weltje and Tjallingii (2008). Instead of using additive log-ratios to perform pairwise element calibration, we have adopted centred log-ratios which enable calibration of all elements at the same time (including "Undef"). The centred log-ratio transformation of a composition vector **x** comprising *D* parts is defined as:

$$\mathrm{clr}(\mathbf{x}) = [\ln x_1 - \ln g, \ldots, \ln x_j - \ln g, \ldots, \ln x_D - \ln g] \tag{21.10a}$$

Where:

$$\ln g = \frac{1}{D}\sum_{j=1}^{D} \ln x_j \tag{21.10b}$$

Centred log-ratio-transformed compositional data may be analysed with standard multivariate statistical methods, which implies that model parameters may be determined with unconstrained least-squares methods. Therefore, we will implicitly assume throughout the remainder of this text that all data have been clr-transformed. It is important to note that clr-transformed variables can always be retransformed to concentrations by using Eq. 21.8b.

Automatic Selection of Calibration Samples

Effective calibration requires that the compositional trends in the data are adequately sampled. Improvements to the workflow proposed by Weltje and Tjallingii (2008) have been achieved by introducing two processing steps in the MLC workflow: (1) an algorithm which estimates the level of noise based on the measured spread among replicate intensities and scales the data accordingly, and (2) an algorithm for automatic selection of calibration samples, based on the multivariate geometry of the scaled intensities.

The MLC workflow (Fig. 21.1) relies on replicate intensity measurements for empirical uncertainty quantification. In the absence of prior knowledge we may collect replicate intensity measurements according to a systematic sampling scheme, e.g. collect at least three measurements at every tenth location down-core. We estimate the level of noise from the average spread within each set of replicates for each element separately.

For the purpose of automatic sample selection, the clr-transformed intensities are scaled with the inverse of their empirically determined uncertainties in order to meet the requirements for rigorous application of least-squares estimation techniques. The average uncertainty is now identical for all intensities, which implies that Euclidian dis-

tances between intensity vectors may be directly interpreted in terms of compositional variability. This provides a solid basis for automatic selection of calibration samples.

In order to automatically select the locations at which calibration samples should be collected, the desired number of unique calibration sites (N_c) must be specified by the analyst. The most suitable locations are selected by means of hierarchical cluster analysis of the scaled clr-transformed intensities. We use Ward's method (Ward 1963) in conjunction with a Euclidian distance measure to subdivide the clr-transformed scaled intensities into N_c clusters, and select one data point from each cluster (the one closest to the cluster centroid). The set of N_c intensities thus selected is subjected to another round of cluster analysis with the objective to further subdivide the set of calibration locations into the desired number of replicate sets, $N_{W,s}$. From each of these $N_{W,s}$ clusters, one data point is selected to be sampled and analysed in replicate (three times or more). This two-step sampling strategy ensures that calibration and uncertainty estimation make use of the full range of variation in the intensity data. Because this procedure does not rely on stratigraphic information, additional sampling may be deemed desirable if one wishes to achieve a more uniform stratigraphic/spatial coverage of the core.

The chemical composition of the set of calibration samples should be determined by conventional destructive methods such as XRF and ICP-AES. The compositional data are subjected to the same treatment as the intensities, i.e. clr-transformation and scaling with the inverse of the estimated uncertainties. The resulting data sets provide the optimal starting point for multivariate calibration with the MLC algorithm (Fig. 21.1).

MLC Algorithm

Parameter estimation for the purpose of predicting element concentrations with the MLC model is carried out by means of Partial Least Squares (PLS) regression, a generalized multivariate linear regression model which maps one data set onto another by making use of a common covariance matrix (De Jong 1993). One of the properties of PLS, which it shares with all other regression models, is that the regression error is inversely proportional to the number of model coefficients (model complexity). When the number of model parameters equals the size of the data set, each data point can be reproduced exactly. Because our data are not noise-free and we should allow for the possibility that the model may not be fully correct, we use a more objective and robust measure of model performance than the regression error: leaving-out-one cross validation (Geisser 1975). The idea behind cross validation is to estimate the parameters of the calibration equation by leaving one measurement out of the data set, and predict the value of this missing data point from the equation fitted to the other data points. By repeating this procedure for all data points, the discrepancies between predicted and measured values may be used to estimate the actual prediction uncertainty. This approach works well if the calibration data set is sufficiently large (the bare minimum would be ~30 specimens).

A straightforward goodness-of-fit measure of logratio-transformed compositional data is the mean squared Euclidian distance between $\hat{\mathbf{x}}$ and $\mathbf{x}$ (predicted and measured composition, respectively):

$$\mathrm{MSE} = \frac{1}{nD}\sum_{i=1}^{n}\sum_{j=1}^{D}\left\{\mathrm{clr}(x_j) - \mathrm{clr}(\hat{x}_j)\right\}^2 \tag{21.11}$$

In the ULC workflow, the set of D models with different common denominators is evaluated to find the one which corresponds to the minimum MSE (cf. Weltje and Tjallingii 2008). Model selection in the MLC workflow is more advanced, because use is made of cross validation to estimate prediction uncertainties. In cases where cross validation has been used, the value of MSE as defined by Eq. 21.11 will be referred to as the mean squared prediction error (MSPE). A similar approach is used to construct error bars around predicted log-ratios of concentrations of specific element pairs. In cases where log-ratio calibration is not required, error bars around log-ratios of measured intensities may be derived from the spread among replicate intensity measurements (Weltje and Tjallingii 2008).

The misfit between the reference and predicted concentrations varies among the different elements and the different calibration models. This misfit is quantified by calculating the residual variance (MSPE), which is a statistically meaningful measure that can be used to construct confidence intervals. However, the MSPE does not tell us anything about the nature of the misfit. We therefore introduce two pseudo-statistical quantities which serve as a measure of the bias and the scatter of estimated concentrations, respectively. Although these measures provide insight into model-data discrepancies, it should be kept in mind that concentrations cannot be rigorously evaluated owing to compositional constraints (Aitchison 1986; Weltje and Tjallingii 2008).

The bias-indicator B is defined as the angle between the major axis of the point-cloud and the line $y = x$. When the two are exactly equal, the bias-indicator B equals zero. If, however, the major axis is perpendicular to the line $y = x$, B will have a value of 100 (it does not matter how the major axis is oriented with respect to the line $y = x$). B is calculated by subtracting the mean from the point cloud and then estimating the major axis by means of a Singular Value Decomposition (Press et al. 1994). Given that $u = \begin{bmatrix}1 & 1\end{bmatrix}$ (i.e. the line $y = x$ in vector form) and v_1 is the major axis vector:

$$B_j = 100 \times \frac{\left|0.5\pi - \cos^{-1}(u \cdot v_1)\right|}{0.5\pi} \tag{21.12}$$

The scatter-indicator S, which also ranges from zero to 100, is based on the percentage of variance along the minor axis, i.e., the axis perpendicular to the earlier calculated major axis. Given that λ_1 is the eigenvalue of the major axis and λ_2 the eigenvalue of the minor axis:

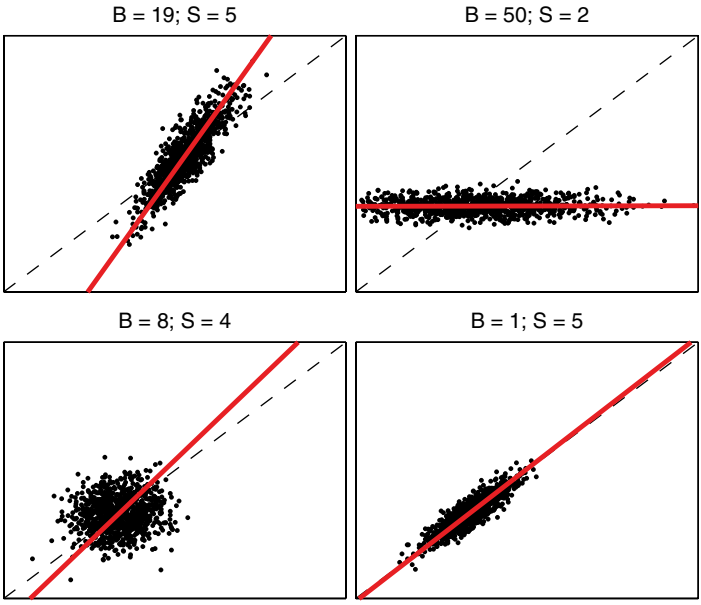

Fig. 21.2 Definition of scatter and bias in compositional space

$$S_j = 100 \times \frac{\lambda_2}{\lambda_1 + \lambda_2} \tag{21.13}$$

Hence, S and B are in principle independent and can therefore also be interpreted as such. Figure 21.2 gives an impression of S and B in a number of different situations. For the purpose of comparing calibration models, we derive their median values across all elements, which we refer to as $\overline{S}$ and $\overline{B}$.

Comparative Calibration Exercise

We will compare the performance of four different calibration models (DLC-1, DLC-2, ULC, and MLC), using two data sets: Core GeoB7920 acquired with an Avaatech scanner, and Core AU10v acquired with an Itrax scanner. Each of these instruments is equipped with a software package which extracts intensities of a range of chemical elements from the spectra recorded by the detector. For the purpose of the comparative calibration exercise, we will simply assume that these intensities are correct. Moreover, no attempts will be made to evaluate the results of our calibrations in the light of geological knowledge.

Data Set 1: GeoB7920

This legacy data set of core GeoB7920, taken offshore West Africa, consists of 168 intensity measurements at 2-cm resolution measured at the University of Bremen with a second-generation Avaatech core scanner (ca. 2004) using 10 kV (250 μA) and 50 kV (1 mA) tube settings, at 30 s per measurement. Automatic processing of the XRF spectra provided intensities of 13 elements. Five replicate sets of intensity measurements (measured five or six times) are available for estimation of uncertainties. A set of 168 table-top ED-XRF measurements (Wien et al. 2005) of samples taken at each down-core location measured with the core scanner is used for calibration. No replicate ED-XRF measurements are available. Detailed information about the sample preparation and the geological setting of the core may be found in Weltje and Tjallingii (2008), Tjallingii et al. (2008), and Bloemsma et al. (2012).

Data Set 2: AU10v

Core AU10v was collected in July 2008 from Augusta Harbour, Sicily (see also Croudace et al., this volume). The core was split and the top 9.6 cm was scanned at 500 μm resolution using a high-resolution Itrax instrument (Croudace et al. 2006) at the University of Southampton. Tube settings used were 30 kV (30 mA) at 30 s per measurement. The XRF spectra were automatically processed and provided intensities of 20 elements. After Itrax scanning, the core was incrementally sub-sampled, pelleted and analysed to obtain quantitative element data using conventional WD-XRF analysis. A total of 51 calibration samples was collected (no replicates). Intensity replicates (126 repeat measurements each) were acquired on two homogenised and powdered pellets taken from the core for the purpose of calibration.

Results

Calibration of the intensities to concentrations of chemical elements involved a selection step, in which elements which could not be predicted (because their residual variance was almost equal to their total variance) were removed from the data sets. This category also included elements whose concentrations were below detection limits in more than half of the calibration samples. Elements which could not be calibrated were added to “Undef” (except in the case of the ULC model, which predicts “relative” concentrations only). The final models contain 12 elements for GeoB7920 (Al, Ba, Br, Ca, Cl, Fe, K, Mn, Rb, Si, Sr, Ti) and 16 elements for AU10v (Ba, Br, Ca, Cl, Cr, Cu, Fe, K, Mn, Pb, S, Si, Sr, Ti, V, Zn), respectively. Table 21.1 gives the goodness-of-fit statistics of the four calibration models, whereas Figs. 21.3, 21.4, 21.5, 21.6, 21.7 and 21.8 permit a visual appraisal of their predictive power. The

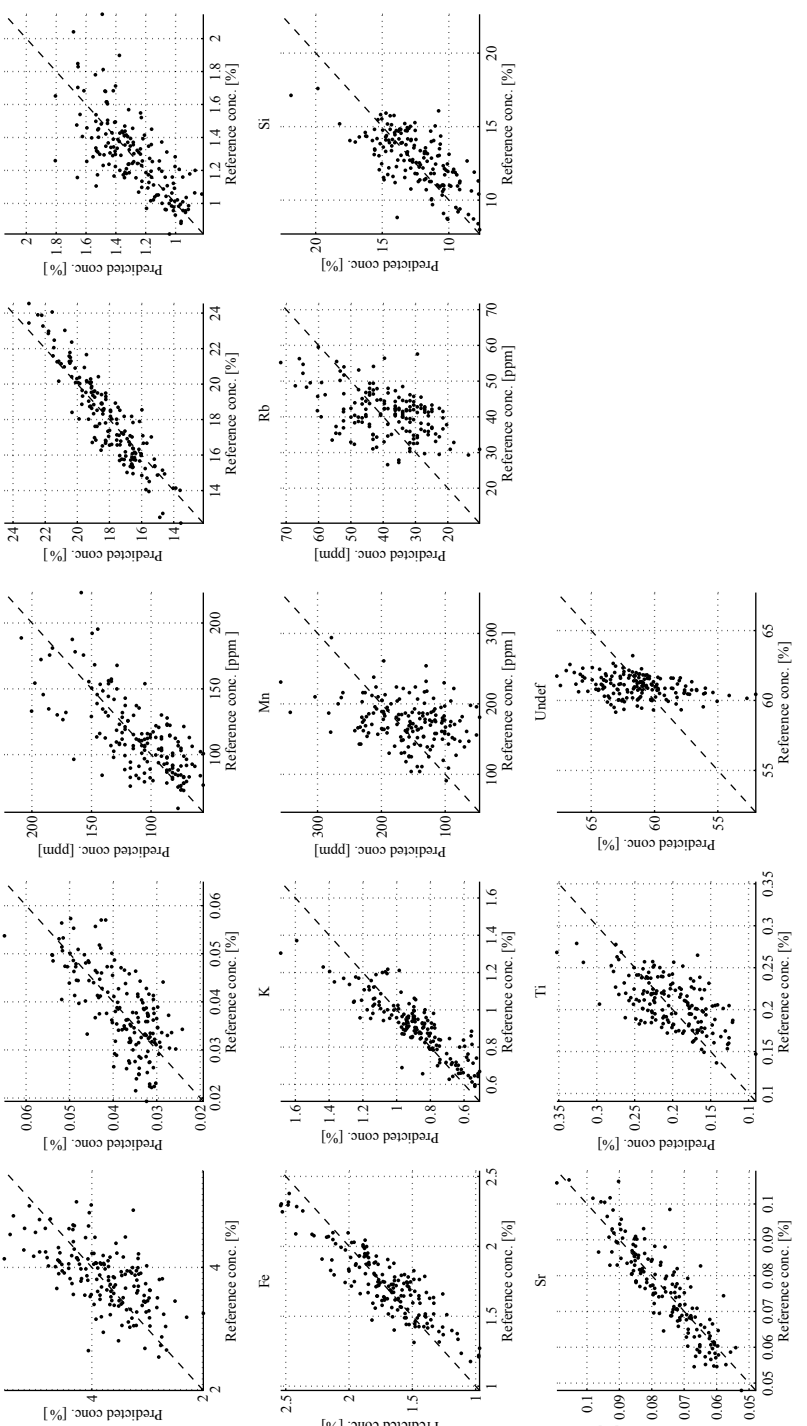

Fig. 21.3 Measured versus predicted concentrations for DLC-1 model of GeoB7920

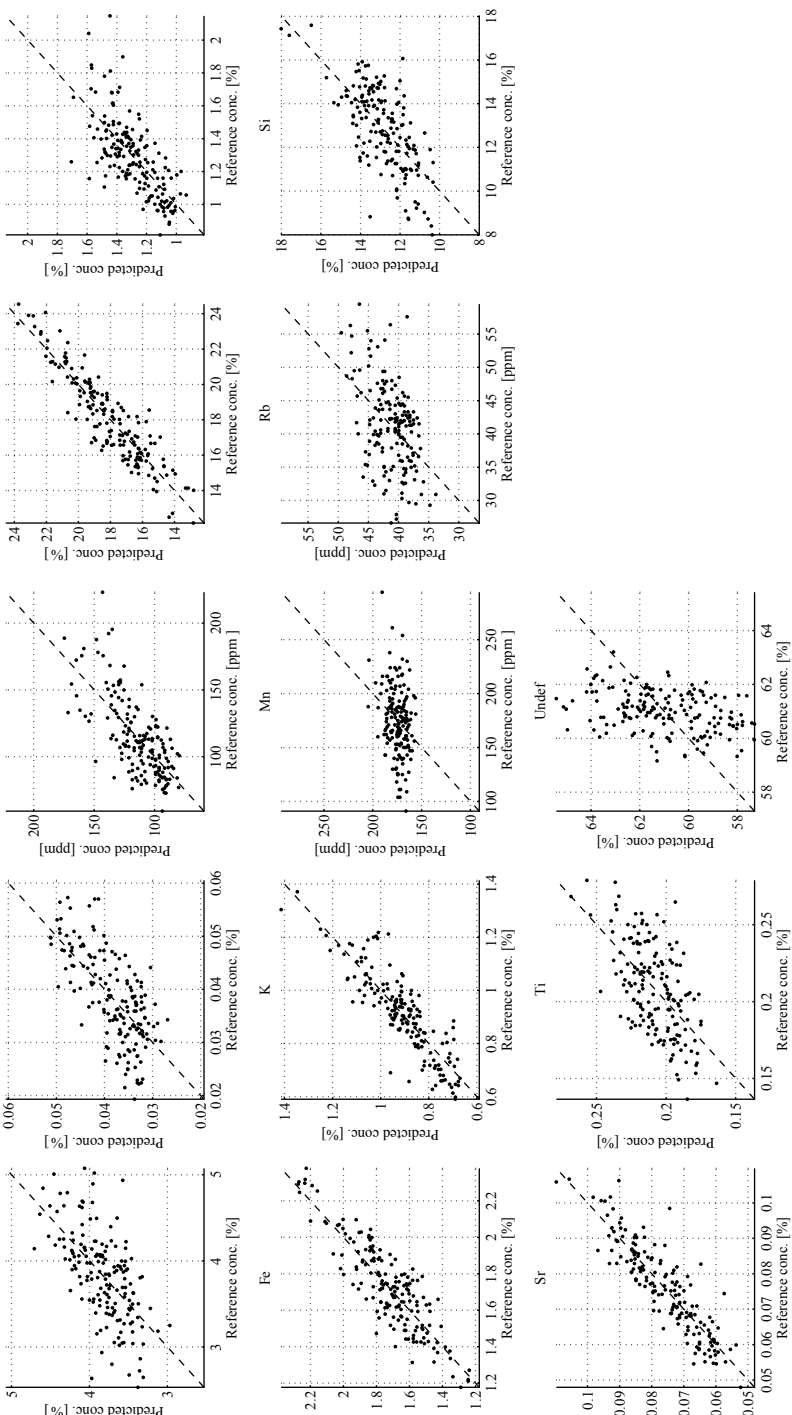

Fig. 21.4 Measured versus predicted concentrations for DLC-2 model of GeoB7920

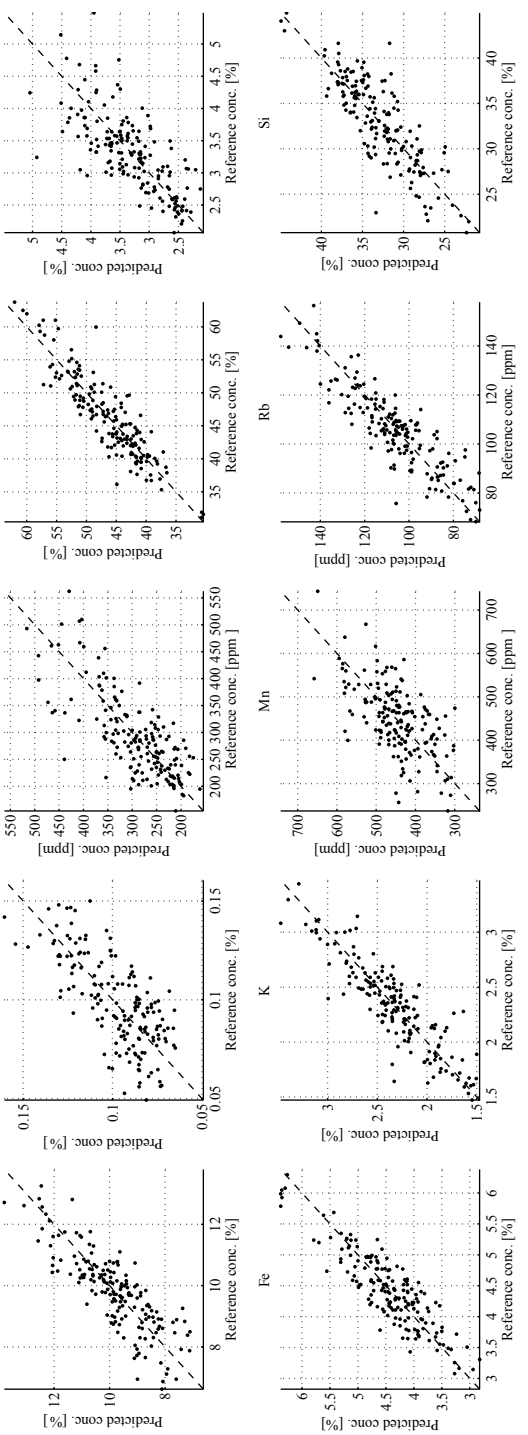

Fig. 21.5 Measured versus predicted concentrations for ULC model of GeoB7920

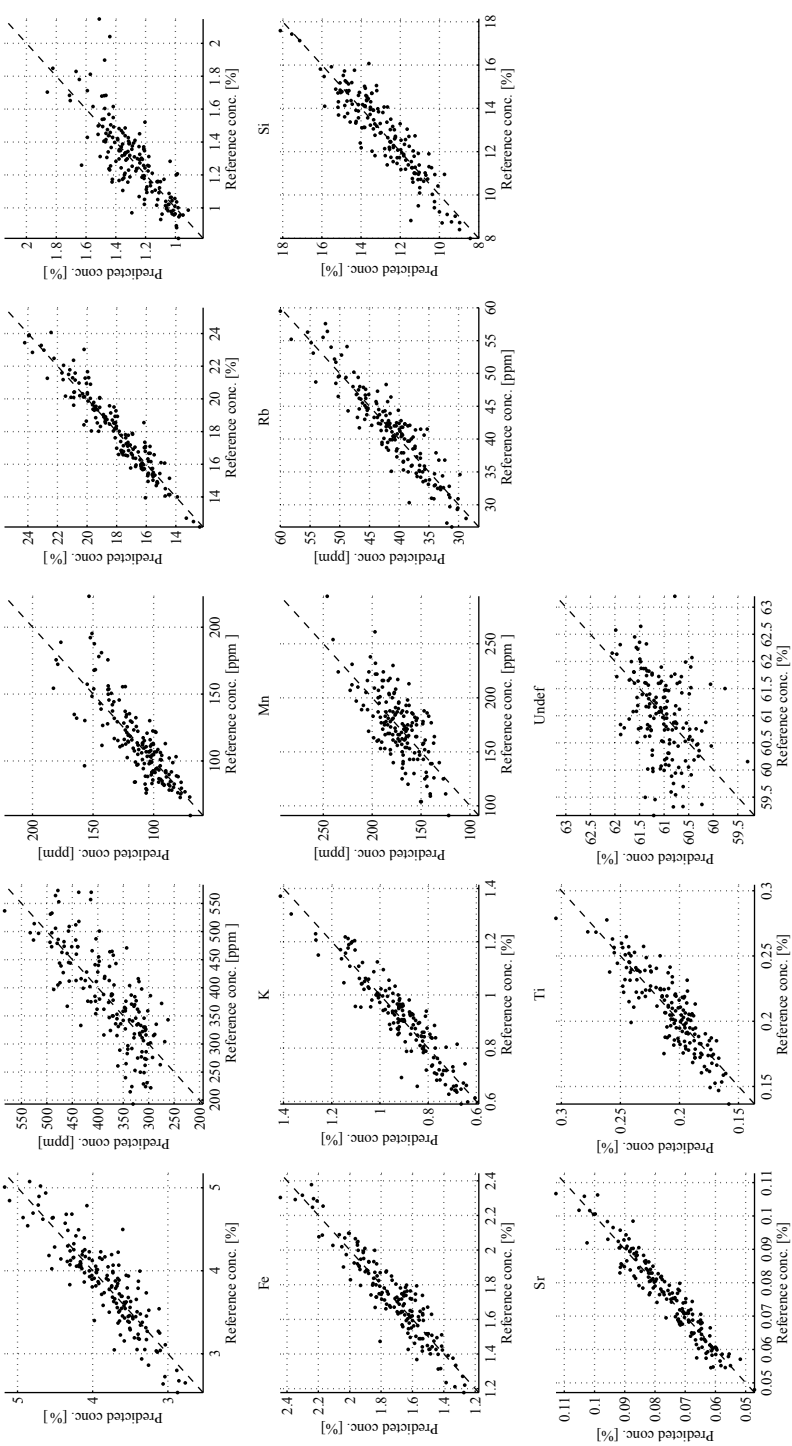

Fig. 21.6 Measured versus predicted concentrations for MLC model of GeoB7920

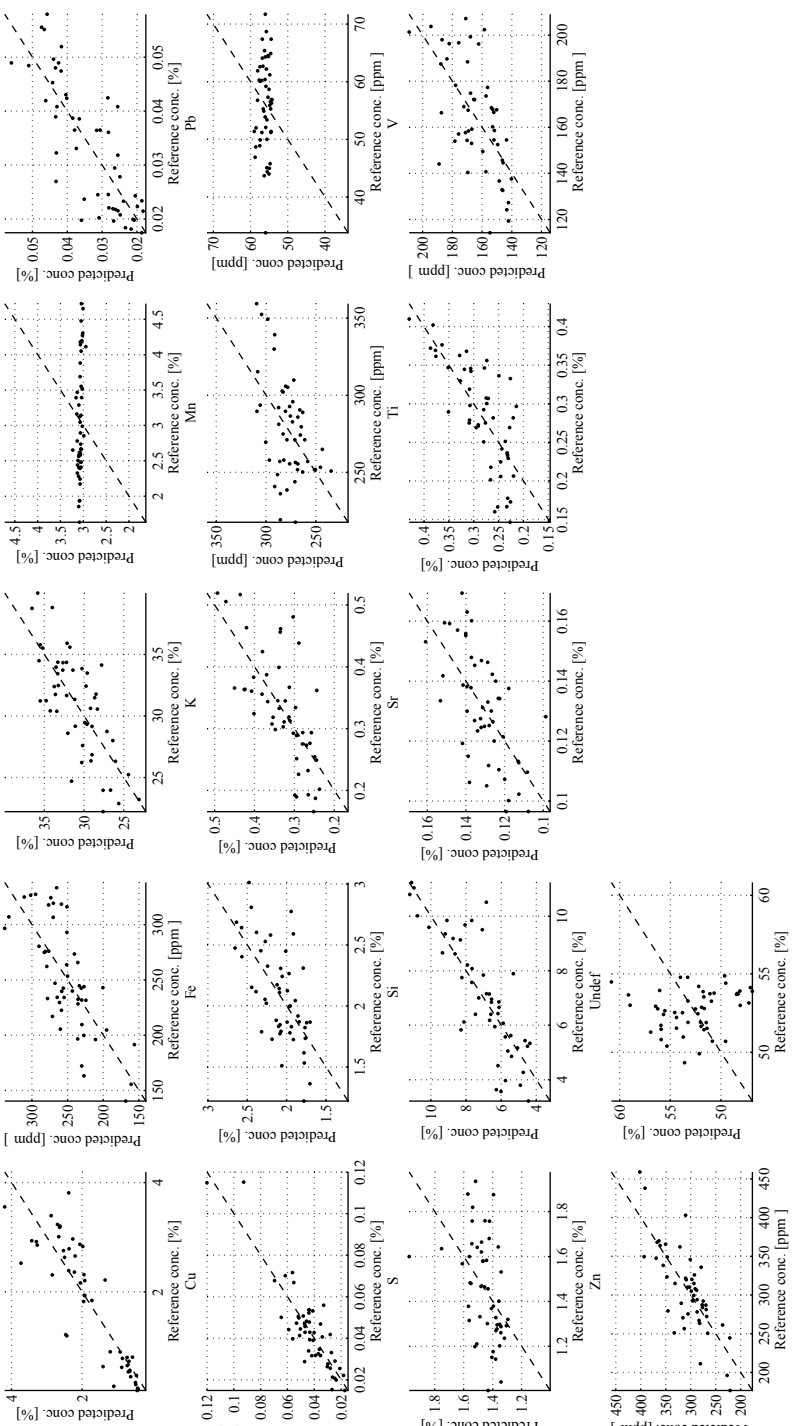

Fig. 21.7 Measured versus predicted concentrations for DLC-2 model of AU10v

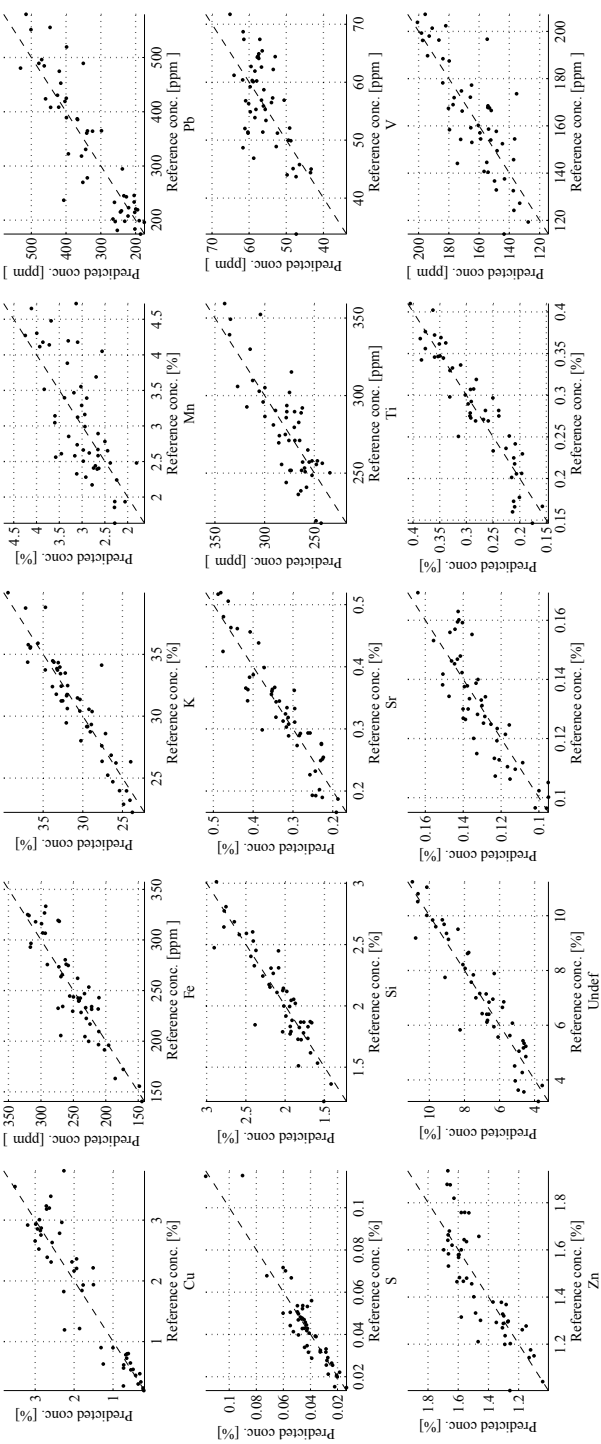

Fig. 21.8 Measured versus predicted concentrations for MLC model of AU10v

Table 21.1 Comparative performance of calibration techniques based on cross validation

Data set	Model	B (%)	S (%)	MSPE
GeoB7920	DLC1	11.69	14.31	0.0319
GeoB7920	DLC2	15.66	13.02	0.0131
GeoB7920	ULC	3.46	8.59	0.0124
GeoB7920	MLC	3.80	6.90	0.008
AU10v	DLC1	14.88	9.48	0.0530
AU10v	DLC2	13.86	12.09	0.0239
AU10v	ULC	7.40	8.80	0.0218
AU10v	MLC	5.43	7.90	0.0163

MSPE values of both data sets show that the MLC model performs better than the ULC model, the ULC model performs better than the DLC-2 model, and the DLC-1 model is the least satisfactory (Table 21.1). The median values of scatter and bias for each model confirm this ranking, and draw attention to the fact that the least advanced log-ratio-based model (ULC) performs much better than the most advanced direct linear calibration model (DLC-2).

Figure 21.9 provides a visual impression of core GeoB7920. In the upper half of this figure, the high-resolution RGB image (Fig. 21.9a) is shown alongside the records of ln(Ca/Ti) intensities (Fig. 21.9b) and predicted concentrations (Fig. 21.9c, d). The gray bands in these four graphs represent the 95% confidence interval of measured intensities and predicted concentrations. Comparison of the raw (Fig. 21.9b) and calibrated (Fig. 21.9c, d) intensities highlights their close similarity. Furthermore, the MLC estimate (Fig. 21.9d) has a smaller confidence interval than the ULC estimate (Fig. 21.9c). The lower half of the image (Fig. 21.9e, f, g) illustrate the chemical composition predicted by the MLC model. Note the strong correlations between element concentrations arising from the compositional constraints. In this particular case, "Undef" is nearly constant down-core and does not contribute much to the variability of "absolute" concentrations.

Figure 21.10 shows the relation between the number of randomly selected calibration samples and the median values of the MSPE for MLC models of core GeoB7920 using different methods of sample selection. The performance of random sampling is represented by the median value of MSPE over a series of 1000 simulations. The performance of systematic sampling (using a fixed sampling interval) is represented by the median MSPE over a series of 83 simulations, and is therefore more "spiky." Our automatic sampling strategy produced only one MSPE for every number of calibration samples, and is therefore even more "spiky". For small numbers of calibration samples (less than 40), the automatic selection method yields the lowest MSPE and thus provides better results than the other methods.

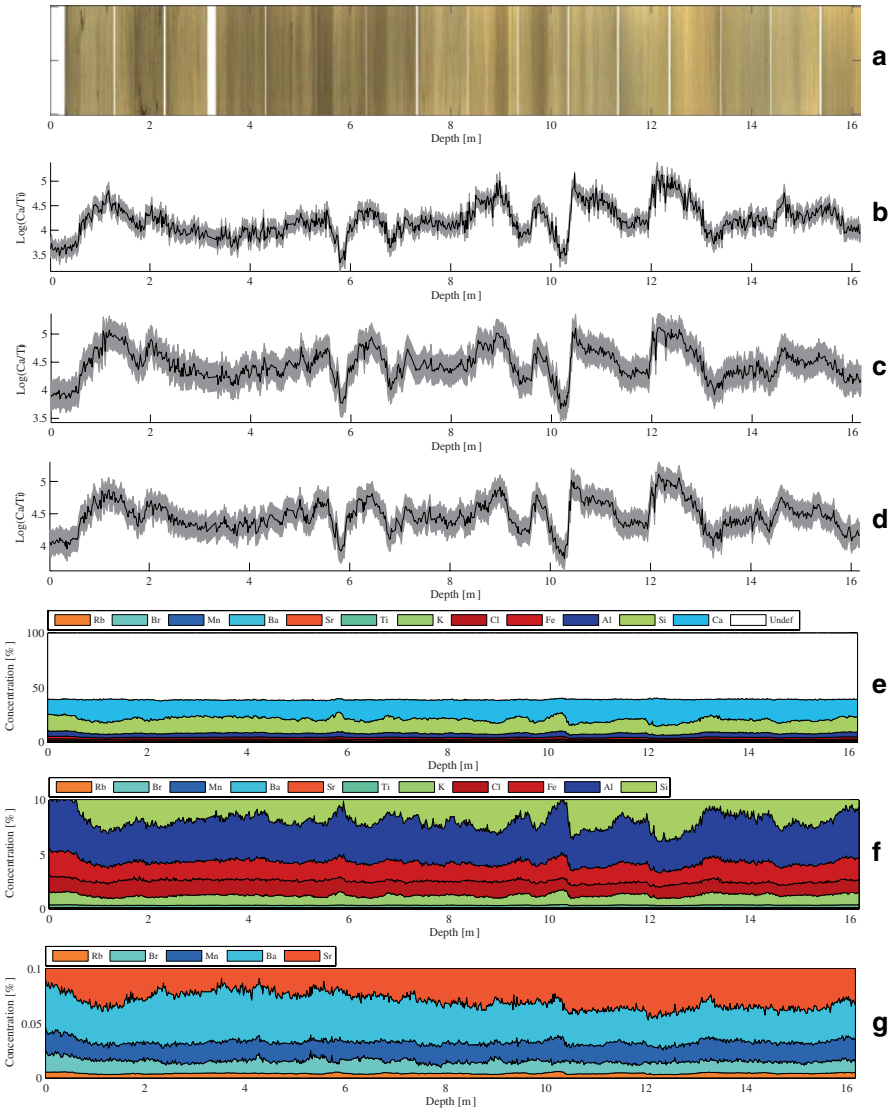

Fig. 21.9 Overview of core GeoB7920. **a** High-resolution RGB image; **b** raw ln(Ca/Ti) intensities with 95 % confidence interval; **c** ULC prediction of ln(Ca/Ti) concentrations with 95 % confidence interval; **d** MLC prediction of ln(Ca/Ti) concentrations with 95 % confidence interval; **e** concentrations from 0 to 100 %; **f** concentrations from 0 to 5 %; **g** concentrations from 0 to 0.5 %. Elements are sorted in descending order of average concentration

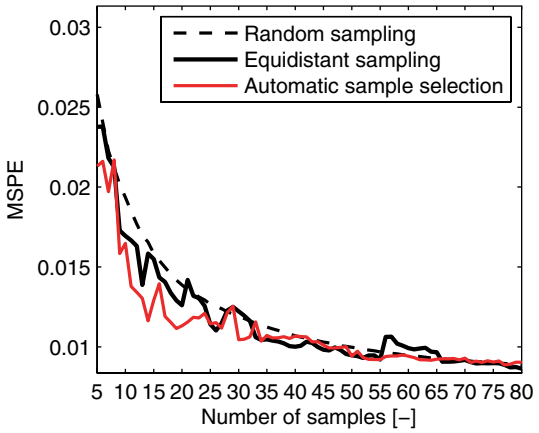

Fig. 21.10 Comparative performance of sampling strategies in the MLC model for core GeoB7920

Discussion

Comparative Performance of Calibration Models

Because all statistics in Table 21.1 are based on cross validation, they tell us something about the actual predictive power of the four calibration models. The comparative evaluation clearly brings out the lack of fit of the DLC models, which stems from a combination of deficiencies: (1) the parametric form of the DLC model (Eq. 21.1) bears no relation to the calibration equation used in XRF spectrometry (Eq. 21.2); (2) there is no unique relation between intensities and concentrations if the measurement geometry is unconstrained; (3) the "single element" calibration does not take into account that concentrations are compositional data. It should be noted that the third deficiency is also present in the "single element" calibration equation used in XRF spectroscopy (Eq. 21.2), but its implications are far less severe, because prediction uncertainties are very small owing to standardization of measurement geometry, which implies that ignoring this fundamental problem will not usually lead to unrealistic results (although these are by no means precluded). It follows directly from the parametric form of the DLC-2 model that its predictive power outside the range of concentrations with which it has been calibrated is extremely small. Hence, extrapolation of DLC-2 models may lead to erroneous results even under laboratory conditions.

Comparison of the ULC and MLC models indicates that the latter provides a more comprehensive description of the calibration problem at hand. The ULC approach is based on the simplifying assumption that the problem to be solved can be reduced to estimation of relative concentrations of two elements simultaneously. This is clearly an oversimplification, because the quality of ULC predictions depends on which element is selected as the common log-ratio denominator. The

MLC approach takes the covariances of all log-ratios into account and allows for direct fitting of matrix effects (absorption and enhancement), which cannot be adequately represented in the ULC model.

The ULC model (Eq. 21.7) implies that log-ratios of concentrations are linear functions of log-ratios of intensities. Hence, if the interest of a researcher lies with down-core changes in a specific proxy, such as the ratio of calcium to iron—commonly taken as a robust marine-to-terrestrial flux ratio indicator—there is no need to perform calibration. In cases where researchers are merely interested in the relative fluxes of marine and terrestrial material, converting intensities to concentrations does not provide any additional information, and XRF-core scanning may be carried out in fully non-destructive mode (compare for instance Fig. 21.9b, c). Moreover, the linear transformation embodied in Eq. 21.7 implies that the correlation matrix of log-ratio intensities is an excellent predictor of the correlation matrix of log-ratio concentrations. This fact may be exploited by researchers who are interested in studying the correlation between log-ratios of element intensities on the one hand, and other properties measured on the same core (or time series obtained from other locations) on the other hand. Hence, multi-proxy analysis may be formalised using exploratory statistical analyses for the purpose of identifying specific geochemical log-ratios as proxies of palaeo-environmental conditions in a specific basin. The usefulness of this empirical approach was demonstrated by Bloemsma et al. (2012), who showed that log-ratios of element intensities or concentrations obtained from bulk measurements do not provide universal shortcuts to palaeo-environmental interpretation, but are site-specific and should thus be validated before they may be applied with some confidence. Calibration of XRF output to obtain estimates of element concentrations is only required if the objective of the research is to carry out quantitative analyses of fluxes or mass-balance calculations.

The ULC model implies that log-ratios of concentrations are linear functions of log-ratios of intensities. Hence, if the interest of a researcher lies with down-core changes in a specific proxy, such as the ratio of calcium to iron—commonly taken as a robust marine-to-terrestrial flux ratio indicator—there is no need to perform calibration.

Calibration of XRF output to obtain estimates of element concentrations is only required if the objective of the research is to carry out quantitative analyses of fluxes or mass-balance calculations.

Recommended Measurement and Sampling Strategies

Automatic selection of calibration samples appears to be a promising method for minimizing the damage to sediment cores in cases where calibration to

concentrations is required. Stochastic simulation experiments carried out with core GeoB7920 (Fig. 21.10) suggest that the prediction error of the MLC model stabilizes at ~60 calibration samples. Although such knowledge is useful if we intend to measure many more cores from the same basin, it is not possible to make broad generalizations about the number of calibration samples in relation to the quality of the calibration model, because this depends on many different factors, such as the type of material analysed, the performance of the core scanner and the device used to measure the compositions of calibration samples, and the practical limitations of time and money.

A generic (i.e., hardware-independent) approach to calibration of XRF core scanner data requires that data-model discrepancies are interpreted in the light of the measurement errors associated with the input data (intensities and concentrations). Hence, replicate measurements of intensities and concentrations are required. Empirical estimates of measurement errors may be derived from the variability of repeated intensity measurements at the same spot and from replicate geochemical analyses. The latter should be obtained by splitting samples into subsamples and analysing them separately. Although the two data sets used for the comparative analysis of calibration models fulfil some of these requirements, they are by no means ideal for effective use of the MLC model. Below we list some guidelines for compiling high-quality core scanner data which are to be converted to quantitative estimates of bulk composition.

As a rule of thumb, adequate coverage of replicate intensities along a core requires the following measurement strategy:

$$N_{I,s} \geq ceil(n^{1/2}) \tag{21.14a}$$

$$\Delta N_{I,s} \leq floor(n / N_{I,s}) \tag{21.14b}$$

$$N_{I,r} \geq 3 \tag{21.14c}$$

$$N_{I,tot} = n + N_{I,s}(N_{I,r} - 1) \tag{21.14d}$$

The number of replicate intensity measurements is coupled to the length of the record, which is defined by the number of locations at which unique intensity measurements are collected, n. For example, scanning of a 1-m long segment of core at 1-cm resolution gives $n = 100$. $N_{I,s}$ represents the number of replicate sets, i.e. the number of locations at which replicate intensity measurements are to be collected. The spacing between these locations is defined as $\Delta N_{I,s}$, and the number of replicate measurements collected at every location as $N_{I,r}$. The total number of intensity measurements on the core segment is given by $N_{I,tot}$. The terms *ceil* and *floor* refer

to the method of rounding to adjacent integer values (up and down, respectively). In the above case where $n=100$, we obtain $N_{I,s}=10$, $\Delta N_{I,s}=10$, $N_{I,r}=3$. The measurement strategy is complete if we specify where the first set of replicates is to be collected, for instance at location 5. The other sets of replicates are then collected at locations 15, 25, …, 95. The total number of measurements needed, $N_{I,tot}=120$, indicates that the overhead associated with this strategy equals 20%. The use of inequalities in Eqs. 21.11a–c implies that the above guidelines should be regarded as reasonable minimum values. Collecting more replicates may contribute to improving the uncertainty estimation of measured intensities, but it will also increase the overhead.

Based on our current experience, we recommend the following strategy for acquisition of calibration samples:

$$N_c \geq 3D \tag{21.15a}$$

$$N_{W,s} \geq ceil(D^{1/2}) \tag{21.15b}$$

$$N_{W,r} \geq 3 \tag{21.15c}$$

$$N_{W,tot} = N_c + N_{W,s}(N_{W,r} - 1) \tag{21.15d}$$

The number of unique sites at which calibration samples should be taken, N_c, is coupled to the number of elements to be calibrated, D. Replicates of some of the calibration samples are needed for the purpose of uncertainty quantification. $N_{W,s}$ represents the number of replicate sets, and $N_{W,r}$ the number of replicates in each set. The total number of calibration samples to be analysed thus equals $N_{W,tot}$. For example, if the number of elements to be calibrated equals 10, $N_c=30$, $N_{W,s}=4$, $N_{W,r}=3$, and $N_{W,tot}=38$. The overhead associated with this strategy equals 27%. Again, more samples may be analysed if deemed necessary.

ASCAR: Advanced Sediment and Core Analysis Research www.ascar.nl

To boost developments in XRF core scanning technology and interpretation, a software package was developed. On this website one finds a version of this software package XELERATE, which is a software package to evaluate, process and visualize XRF core scanning data. A download of this software package is available when you are logged in, at the section ‘downloads’

Conclusions

The second-generation multivariate log-ratio calibration (MLC) algorithm illustrated in this contribution allows unbiased prediction of geochemical compositions from XRF core scanner output with a degree of precision that is comparable to conventional XRF analysis of heterogeneous materials under laboratory conditions. It represents a vast improvement over previous attempts at direct calibration of intensities using concentrations, and is significantly better than the univariate ULC model proposed by Weltje and Tjallingii (2008). The main advantages of multivariate log-ratio calibration (MLC) over univariate log-ratio calibration (ULC) are (1) elimination of the need to select the best model from the set of D alternative models; (2) effective use of the covariances of intensity and concentration measurements, which reflect absorption and enhancement of intensities, as well as the fact that certain elements reside in the same minerals, and (3) the possibility to estimate the mass fractions of samples which could not be attributed to specific elements ("Undef"), allowing prediction of "absolute" concentrations. Solution of the long-standing problem of XRF core scanner calibration implies that high-resolution records of quantitative sediment composition with associated uncertainties can now be routinely established, which should increase the usefulness of XRF-core-scanning devices and pave the way for quantitative evaluation of geochemical proxies (Weltje and Tjallingii 2008; Bloemsma et al. 2012).

Acknowledgments GJW and MRB wish to thank Wintershall Noordzee BV, TNO, Avaatech, and Panterra for their support of the research project "Advanced Sediment Characterisation, Analysis and Research." RT was supported by the Netherlands Organisation for Scientific Research (NWO) as part of INATEX-B (grant number 839.08.430) and the SCAN2 program on advanced instrumentation. The data set of cores GeoB7920 and Augusta Harbour core AU10v (see Croudace et al., this volume), as well as executables of the program and a user guide may be downloaded from web site "www.ascar.nl" Please note that the current version of the program has been designed to read output of the Avaatech and Itrax core scanners only. As this is research code and not a commercial product, we do not provide support.

References

Aitchison J (1982) The statistical analysis of compositional data (with discussion). J R Stat Soc Ser B 44:139–177

Aitchison J (1986) The statistical analysis of compositional data. Chapman and Hall, London, p 416

Aitchison J, Egozcue JJ (2005) Compositional data analysis: where are we and where should we be heading? Math Geol 37:829–850

Bahr A, Lamy F, Arz H, Kuhlmann H, Wefer G (2005) Late glacial to Holocene climate and sedimentation history in the NW Black Sea. Mar Geol 214:309–322

Bloemsma MR, Zabel M, Stuut JBW, Tjallingii R, Collins JA, Weltje GJ (2012) Modelling the joint variability of grain-size and chemical composition in sediments. Sediment Geol 280: 135–148 (+ Erratum: 284–285, 214)

Böning P, Bard E, Rose J (2007) Toward direct, micron-scale XRF elemental maps and quantitative profiles of wet marine sediments. Geochem Geophys Geosyst 8

Buccianti A, Mateu-Figueiras G, Pawlowsky-Glahn V (eds) (2006) Compositional data analysis in the geosciences (Special Publication). Geological Society, London, p 264

Calvert SE, Pedersen TF (2007) Elemental proxies for palaeoclimatic and palaeoceanographic variability in marine sediments: interpretation and application. In: Hillaire C, de Vernal A (eds) Proxies in Late Cenozoic paleoceanography. Developments in marine geology, vol 1. Elsevier Science, Amsterdam, pp 567–644

Croudace IW, Rindby A, Rothwell RG (2006) ITRAX: description and evaluation of a new multi-function X-ray core scanner. In: Rothwell RG (ed) New techniques in sediment core analysis, vol 267 (Special Publication). Geological Society, London, pp 51–63

De Jong S (1993) SIMPLS: an alternative approach to partial least squares regression. Chemom Intell Lab Syst 18:251–263

De Vries JL, Vrebos BAR (2002) Quantification of infinitely thick specimens by XRF analysis. In: van Grieken RE, Markovicz AA (eds) Handbook of X-ray spectrometry, 2nd edn. Marcel Dekker, New York, pp 341–405

Ge LQ, Lai WC, Lin YC (2005) Influence of and correction for moisture in rocks, soils and sediments on in situ XRF analysis. X-Ray Spectrom 34:28–34

Geisser S (1975) The predictive sample reuse method with applications. J Am Stat Assoc 70:320–328

Haschke M (2006) The Eagle III BKA system, a novel sediment core X-ray fluorescence analyzer with very high spatial resolution. In: Rothwell RG (ed) New techniques in sediment core analysis, vol 267 (Special Publication). Geological Society, London, pp 31–37

Haschke M, Scholz W, Theis U, Nicolosi J, Scruggs B, Herzceg L (2002) Description of a new micro-X-ray spectrometer. J Phys IV Fr 12:6–83

Jaccard SL, Haug GH, Sigman DM, Pedersen TF, Thierstein HR, Röhl U (2005) Glacial/interglacial changes in subarctic North Pacific stratification. Science 308:1003–1006

Jansen JHF, Van der Gaast SJ, Koster B, Vaars AJ (1998) CORTEX, a shipboard XRF-scanner for element analyses in split sediment cores. Mar Geol 151:143–153

Jenkins R (1999) X-ray fluorescence spectroscopy, 2nd edn. Wiley, New York, 207 pp

Kido Y, Koshikawa T, Tada R (2006) Rapid and quantitative major element analysis method for wet fine-grained sediments using an XRF microscanner. Mar Geol 229:209–225

Löwemark L, Chen HF, Yang TN, Kylander M, Yu EF, Hsu YW, Lee TQ, Song SR, Jarvis S (2011) Normalizing XRF-scanner data: a cautionary note on the interpretation of high-resolution records from organic-rich lakes. J Asian Earth Sci 40:1250–1256

Pälike H, Shackleton NJ, Röhl U (2001) Astronomical forcing on late Eocene marine sediments. Earth Planet Sci Lett 193:589–602

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1994) Numerical recipes in FORTRAN: the art of scientific computing, 2nd edn. University Press, Cambridge, 963 p

Ramsey MH, Potts PJ, Webb PC, Watkins P, Watson JS, Coles BJ (1995) An objective assessment of analytical method precision: comparison of ICP-AES and XRF for the analysis of silicate rocks. Chem Geol 124:1–19

Richter TO, Van der Gaast SJ, Koster B, Vaars A, Gieles R, De Stigter H, De Haas H, van Weering TCE (2006) The Avaatech XRF core scanner: technical description and applications to NE Atlantic sediments. In: Rothwell RG (ed) New techniques in sediment core analysis, vol 267 (Special Publication). Geological Society, London, pp 39–50

Rothwell RG, Rack FR (2006) New techniques in sediment core analysis: an introduction. In: Rothwell RG (ed) New techniques in sediment core analysis, vol 267 (Special Publication). Geological Society, London, pp 1–29

Rothwell RG, Hoogakker B, Thomson J, Croudace IW, Frenz M (2006) Turbidite emplacement on the southern Balearic Abyssal Plain (western Mediterranean Sea) during Marine Isotope Stages 1–3: an application of ITRAX XRF scanning of sediment cores to lithostratigraphic analysis. In: Rothwell RG (ed) New techniques in sediment core analysis, vol 267 (Special Publication). Geological Society, London, pp 79–98

Tjallingii R, Röhl U, Kölling M, Bickert T (2007) Influence of the water content on X-ray fluorescence core-scanning measurements in soft marine sediments. Geochem Geophys Geosyst 8

Tjallingii R, Claussen M, Stuut JB, Fohlmeister J, Jahn A, Bickert T, Lamy F, Röhl U (2008) Coherent high- and low-latitude control of the northwest African hydrological balance. Nat Geosci 1:670–675

Vlag PA, Kruiver PP, Dekkers MJ (2004) Evaluating climate change by multivariate statistical techniques on magnetic and chemical properties of marine sediments (Azores region). Palaeogeogr Palaeoclimatol Palaeoecol 212:23–44

Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–244

Weltje GJ, Tjallingii R (2008) Calibration of XRF core scanners for quantitative geochemical logging of sediment cores: theory and application. Earth Planet Sci Lett 274:423–438

Wien K, Wissmann D, Kölling M, Schulz HD (2005) Fast application of X-ray fluorescence spectrometry aboard ship: how good is the new portable Spectro Xepos analyzer? Geo-Mar Lett 25:248–264