

Development of a Modelling Framework for Core Data Integration using XRF Scanning

Menno Rudolf Bloemsma

Cover illustration: *Focus-stack* image of core "Kampen-1-S1", constituting alternation of coarse-grained, anhydrite-cemented (light colour) and fine-grained (red-brown colour) sand laminae. This image was acquired by Pieter van der Klugt (TNO), and was modified from Henares et al. (2014).

Development of a Modelling Framework for Core Data Integration using XRF Scanning

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.Ch.A.M. Luyben,
voorzitter van het College van Promoties,
in het openbaar te verdedigen op dinsdag 29 september 2015 om 15.00 uur
door

Menno Rudolf BLOEMSMA

ingenieur in de geomatica
geboren te Almelo, Nederland

This dissertation has been approved by the promotor:

Prof. dr. S.M. Luthi

Prof. dr. G.J. Weltje

Composition of the doctoral committee:

Rector Magnificus

Prof. dr. S.M. Luthi

Prof. dr. G.J. Weltje

Dr. R. Tjallingii

Delft University of Technology (CiTG), chairperson

Delft University of Technology (CiTG), promotor

University of Leuven, promotor

GFZ Potsdam

Independent members:

Prof. dr. P.M.J. van Espen

Prof. dr. J.J. Egozcue

Prof. dr. J. Dik

Dr. W.J.E. van de Graaff

Prof. dr. J. Bruining

University of Antwerp

Polytechnic University of Catalonia

Delft University of Technology (3ME)

VDG Geologische Diensten

Delft University of Technology (CiTG), reserve member

ISBN 978-94-6295-352-9

Keywords: compositional data, multivariate regression, quantitative spectroscopy, analytical uncertainty, maximum likelihood, facies prediction, geochemistry.

This research is conducted at the faculty of civil engineering and geosciences, department of geoscience & engineering, section of applied geology, and is financially supported by Wintershall Noordzee BV and TNO Geological Survey of The Netherlands.

Copyright © 2015 by M.R. Bloemsmma. All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without the prior permission of the copyright owner.

Contact: bloemsmamenno@gmail.com

Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress

Published by: Uitgeverij BOXPress, 's-Hertogenbosch

Contents

1	Introduction	1
1.1	General introduction	1
1.2	Core analysis and core-scanning technology	2
1.3	XRF and its application in geoscience	3
1.4	The multivariate approach	4
1.5	Objectives	5
1.6	Structure of this thesis	6
2	Multivariate logratio calibration of XRF-CS data	7
2.1	Introduction	7
2.2	Calibration in conventional XRF spectrometry	8
2.3	Log-ratio calibration	9
2.4	"Absolute" concentrations	11
2.5	MLC workflow	11
2.5.1	Multivariate calibration approach	12
2.5.2	Parameter estimation	12
2.5.3	Replicates and scaling	14
2.5.4	Summary statistics	15
2.5.5	Automatic selection of calibration samples	16
2.6	Comparative calibration exercise	17
2.6.1	Data set 1: GeoB7920	17
2.6.2	Data set 2: AU10v	17
2.7	Results	17
2.8	Discussion	26
2.8.1	Comparative performance of calibration models	26
2.8.2	Recommended measurement and sampling strategies	27
2.9	Conclusions	28

3	Modelling the joint variability of grain size and chemical composition in sediments	31
3.1	Introduction	31
3.2	Material and methods	32
3.2.1	Materials	32
3.2.2	Analytical methods	34
3.2.3	Data preprocessing	35
3.3	Modelling approach	37
3.3.1	Conceptual model	37
3.3.2	Statistical model	39
3.4	Synthetic examples	39
3.4.1	Synthetic example 1	40
3.4.2	Synthetic example 2	40
3.5	Results	45
3.6	Discussion	45
3.6.1	Correlation between bulk properties	45
3.6.2	Residuals analysis	49
3.6.3	Palaeo-climatological interpretation	50
3.6.4	General discussion	52
4	Reduced-rank approximations to compositional and spectroscopic data in log-ratio space	53
4.1	Introduction	53
4.2	Statistical basis	55
4.2.1	Notation	55
4.2.2	Fixed-size count data	56
4.2.3	Fixed-time count data	57
4.2.4	CLR space error model	58
4.2.5	ECM in the clr space	59
4.2.6	Limiting properties of the clr-space ECM	59
4.2.7	ECM of a count data set	61
4.3	Simulation experiment	62
4.3.1	Reduced-rank methods	62
4.3.2	Generating synthetic data	65
4.3.3	Example data set	67
4.3.4	Experimental design	68
4.3.5	Diagnostics	68
4.3.6	Computational load	70
4.4	Results	70
4.5	Discussion	71
4.5.1	Method performance	71
4.5.2	Method selection	74
4.5.3	Computational performance	75
4.5.4	Optimal Scale-Invariant Reduced-rank Approximation (OSIRA)	75
4.5.5	General remarks	76
4.6	Conclusions	76

5	Modelling the uncertainty of routine chemical analyses	79
5.1	Introduction	79
5.2	UFs and spectrochemical theory	80
5.2.1	UFs from PT data	80
5.2.2	Parametric forms	82
5.2.3	UF examples	82
5.2.4	Physical plausibility	83
5.2.5	Calibration	83
5.2.6	Propagation of counting errors	84
5.2.7	SDR at low concentrations	85
5.2.8	SDR at high concentrations	85
5.3	UFs and statistical theory	86
5.3.1	Properties of compositional data	86
5.3.2	Symmetry of uncertainty	87
5.3.3	Plausibility of concentration-uncertainty relations	88
5.3.4	Evaluation of existing UFs	89
5.4	Synthesis	90
5.4.1	Physical and statistical theory compared	90
5.4.2	BUCE and CUCE compared	91
5.4.3	Re-establish consistency by means of closure?	92
5.4.4	Mimicking closure	93
5.4.5	Symmetry in empirical uncertainty data	93
5.5	Modelling framework	94
5.5.1	Definition of UFs	94
5.5.2	Modelling stage 1: System characterization	95
5.5.3	Modelling stage 2: Pseudo concentrations to concentrations	95
5.5.4	Domain of the UF	98
5.5.5	Measuring performance	98
5.6	Application	98
5.6.1	GeoPT data set	98
5.6.2	GeoPT Results	99
5.6.3	Discussion of GeoPT results	99
5.7	Discussion and conclusions	100
6	Integrated core analysis using XRF-CS. Part I: Prediction of lithofacies	103
6.1	Introduction	103
6.2	Materials and Methods	104
6.2.1	Geological setting and sedimentological interpretation	104
6.2.2	Core analysis	104
6.2.3	XRF core scanning	105
6.2.4	Integrity of XRF-CS data	106
6.3	Modelling methods	107
6.3.1	Discriminant-function analysis	107
6.3.2	Predictive performance	108
6.3.3	Dendrogram and biplot analysis	108
6.4	Results	109
6.4.1	Core E10-3	109

6.4.2	Core B38D	109
6.5	Discussion	112
6.5.1	Dendrogram and biplot analysis	112
6.5.2	Relation between confusion matrix and dendrogram	114
6.5.3	Interpretation of element intensities	115
6.5.4	Nature of discrepancies	116
6.6	General discussion, conclusions and recommendations	116
6.6.1	General discussion and conclusions	116
6.6.2	Outlook	117
7	Integrated core analysis using XRF-CS. Part II: Prediction of chemical and petrophysical properties	121
7.1	Introduction	121
7.2	Materials and Methods	121
7.2.1	Core analysis	121
7.2.2	Calibration data integrity	122
7.3	Modelling methods	123
7.3.1	MLC method	123
7.3.2	GLC and data scaling	124
7.3.3	Summary statistics	124
7.3.4	Predictive performance (petrophysical properties)	125
7.3.5	Predictive performance (geochemical composition)	125
7.3.6	Repeatability of XRF-CS data	126
7.4	Results	127
7.4.1	Geochemical calibration	127
7.4.2	Petrophysical calibration	133
7.5	Discussion	133
7.5.1	Geochemical calibration	133
7.5.2	Element intensity importance	135
7.5.3	Extrapolation abilities (histogram analysis)	138
7.5.4	Extrapolation abilities (explicit poroperm relation)	139
7.6	General discussion, conclusions and recommendations	141
7.6.1	Quantitative performance evaluation	141
7.6.2	Possible improvements	141
7.6.3	Concluding remarks	142
8	Integrated core analysis using XRF-CS. Part III: Applications	145
8.1	Introduction	145
8.1.1	General introduction	145
8.1.2	Core E10-3: controls on reservoir quality	145
8.2	Application 1: Enhanced facies properties	146
8.2.1	Objective	146
8.2.2	Approach	146
8.2.3	Results and discussion	146
8.3	Application 2: Automatic reservoir quality prediction	149
8.3.1	Objective	149
8.3.2	Approach	149

8.3.3	Results and discussion	149
8.4	Application 3: Diagenetic controls on reservoir properties	151
8.4.1	Objective	151
8.4.2	Approach	151
8.4.3	Results and discussion	152
8.5	General discussion and conclusions	157
9	General discussion and outlook	159
9.1	Discussion	159
9.1.1	Objectives	159
9.1.2	Performance	160
9.1.3	Application of XRF-CS	161
9.2	Future developments	162
9.2.1	Line-scan imaging	162
9.2.2	Minipermeametry	164
9.2.3	Alternative technological improvements	165
9.3	Outlook	165
	Appendices	169
A	Appendix of Chapter 3	171
B	Appendix of Chapter 4	173
C	Appendix of Chapter 5	175
C.1	Modelling approach	175
C.1.1	UFs and closure	175
C.1.2	Derivation of proper UF	176
C.1.3	From pseudo to proper UF: Poissonian case	176
C.1.4	From pseudo to proper UF: General case	177
C.2	GeoPT unpublished references	177
D	Appendix of Chapters 6 and 7	181
D.1	RCA of core E10-3	181
D.1.1	Petrographical analysis	182
D.2	Lithofacies classification scheme (E10-3)	182
D.3	Features of the GLC	183
D.3.1	Extension 1: Model selection	183
D.3.2	Extension 2: Transformation selection	184
D.3.3	Extension 3: Removing uninformative X-variables	184
D.3.4	Extension 4: Removing unpredictable Y-variables	185
D.3.5	Extension 5: Modelling incomplete compositions	185
E	List of acronyms	187
	Bibliography	189
	Curriculum Vitae	211

Acknowledgements	213
List of publications	215
9.1 Journal papers	215
9.2 Conference proceedings and abstracts	215
9.3 Miscellaneous	217

Introduction

1.1 General introduction

According to Clarke and Washington (1924), 95% of the earth's crust consists of igneous and metamorphic rocks. The other 5% are sedimentary rocks, including 4% shale, 0.75% sandstone and 0.25% limestone. Despite their low crustal abundance, sedimentary rocks are of key economical and societal importance because they contain important natural resources such as hydrocarbons, rare earth elements and drinking water. In addition, sedimentary records are important archives of the Earth's history and deciphering them provides insight into ancient climatic, tectonic and eustatic conditions (Allen, 2009). Understanding the mechanisms which control the properties, morphology and spatial distribution of sedimentary deposits is therefore of utmost importance.

A conceptual approach to the formation of sedimentary deposits is the "source-to-sink" model whose genetic units are (1) an erosional area in which chemical and mechanical weathering cause rocks to break down into sediments (i.e., the source); (2) a sedimentary routing system in which sediments are transported by some medium (e.g., water, air) under the influence of a gradient; (3) a topographic low where the sediments are deposited (i.e., the sink) (Allen and Allen, 2005). The sedimentary record reflects the complex interplay of compositional and physical alterations imposed on the sediment while residing in each of these genetic units (Blatt et al., 1980). Unravelling information from sedimentary records therefore requires solving an inverse problem which, given that it generally does not have a unique solution, can be a challenging task (Johnsson, 1993; Weltje et al., 1998; Weltje and von Eynatten, 2004). Hence, it is of key importance to use all information and data at hand and analyse them in an integrated manner.

Characterisation of sedimentary deposits in the subsurface generally involves acquisition of different sets of data, including seismic, well-log and, ideally, core data. Only by employing these different analysis methods, it is possible to adequately capture the heterogeneity of the sedimentary deposits across many different scales. For instance, the resolution of seismic data is not sufficient to capture the small-scale variations in sediment prop-

erties which, nevertheless, may be of vital importance for the ability of the sediment body to transmit fluids. On the other hand, given that it covers the entire deposit, seismic data enable determination of the dimensions and geometry of large-scale structures. It is evident that to a certain extent these data sets contain redundant information. This redundancy enables the transfer of information across the scales. For instance, a first-order approximation of the physical properties throughout the deposit may be obtained by assigning physical properties derived from cores to stratigraphic units recognisable on seismic data.

Cores play a key role in integrating the different available data sets. Mobil attempted to quantify the value of core information in their operations and estimated that 10 to 15% of their oil production would not be possible without core analysis (Forbes, 1998). One reason for their importance is that cores are the only physical and intact sample of the sedimentary deposit, hence they enable calibration and validation of remotely-sensed data, such as porosity logs, seismic impedance (Serra, 1987) and more specialised logs such as geochemical logs (Hertzog et al., 1987). Secondly, cores enable analysis of the sediment on the smallest possible scale: surprisingly enough, variations on this smallest scale can uniquely provide information about the large scale basin evolution with regard to e.g., diagenetic phases, sediment source and the source-to-sink pathway. Unravelling such basin-scale variations from analysis of grains is known as provenance analysis (Basu, 2003; Weltje, 2012).

1.2 Core analysis and core-scanning technology

Although limited compared to the basin scale, sedimentological variability on the core scale also spans several orders of magnitude. In the order of meters, cores reflect the major variations in lithologies and sedimentary facies. In the order of centimeters, cores reflect sedimentary structures such as ripples and burrows, i.e. structures which may serve as telltales of the conditions under which the sediments were deposited. In the order of micrometers to millimeters, cores contain information about the pore network and the characteristics of individual grains (e.g., roundness).

A similar genetic ordering is reflected by the data typically acquired from cores, which includes: (1) high-resolution ($\leq 0.01\text{m}$) qualitative descriptions in terms of lithology, texture and sedimentary structures, (2) medium-resolution (0.1m-1m) point measurements of bulk properties (chemistry, mineralogy, porosity), and (3) low-resolution (1m-10m) point measurements acquired using the most sophisticated and costly analytical techniques like single-grain geochemistry and scanning-electron microscopy. Hence, whereas the most basic properties of the sediment are at hand at a virtually infinite resolution, there is generally only a handful of detailed analysis of sediment properties on the grain scale. These resolution gaps are the consequence of practical limitations: acquiring all these data on the same high resolution is not feasible due to time and financial constraints. It would also require taking numerous discrete samples from the core which is unwanted and in some cases not even allowed because it reduces the core's quality and can even destroy its integrity.

The fact that not all data are available at the same resolution complicates core analysis. In an attempt to overcome this problem, automated core scanning techniques have been developed. Core-scanning techniques are based on in-situ spectroscopic analysis, which are therefore time- and cost-efficient as well as non-destructive. This technology enables acquisition of bulk properties on a high-resolution (sub-centimeter), whereas it requires

minimal sample preparation (Jansen et al., 1998; Haschke et al., 2002; Haschke, 2006). One spectroscopic technique which is often employed in core scanners is X-ray Fluorescence (XRF) spectrometry (Jansen et al., 1998; Wien et al., 2005; Gé et al., 2005; Rothwell et al., 2006). The principle underlying XRF is that incident γ - or x-radiation causes electrons in the inner shells to be ejected (De Vries and Vrebos, 2002). Because this is an unstable situation, the vacant positions created are filled by electrons in higher shells, which emit their excess energy by means of secondary radiation. The energy of this radiation is specific to a particular combination of inner and outer shell energies which, in turn, is element-specific. Under standardised conditions, the intensity of the secondary radiation is proportional to the concentration of the associated element in the specimen.

1.3 XRF and its application in geoscience

XRF is a widely-used method for chemical analysis of rocks and sediments (Ramsey et al., 1995; Jenkins, 1999). Geochemical data have been used to study: (i) the composition of the source rock (Amorosi et al., 2002; McLennan et al., 1993; Argast and Donnely, 1987), (ii) the intensity of entrainment sorting (Garcia et al., 2004), (iii) variations in grain size (Whitmore et al., 2004; Kiminami and Fuji, 2007), (iv) the extent of chemical weathering (Nesbitt and Young, 1982; Ohta and Arai, 2007), (v) and sedimentary facies (Svendsen et al., 2007). Despite the potential of XRF core scanning for geological studies, however, XRF core scanning is mainly used as a qualitative tool which enables fast acquisition of high-resolution "wiggles". It is commonplace to analyse these "wiggles" in combination with other independently-acquired data and display them side-by-side, either to highlight consistency between the data sets or to illustrate patterns unique to a particular data set (e.g., Bakke et al., 2013). This approach is often referred to as "multi-proxy" analysis (see e.g., Cronin et al., 2005).

The underlying reason for XRF core scanners to be mainly used as a qualitative tool is that quantitative calibration is considered problematic (Jansen et al., 1998; Jaccard et al., 2005; Croudace et al., 2006; Kido et al., 2006; Böning et al., 2007; Tjallingii et al., 2007). The problem is that the XRF data are acquired in-situ, which means that both the properties of the matrix as well as the measurement geometry (e.g., distance between sample and detector) vary throughout the core. This, in turn, invalidates the proportionality between intensities and analyte concentrations, i.e., the fundamental assumption of quantitative spectroscopy (Jenkins, 1999; De Vries and Vrebos, 2002). An additional problem is that in the models commonly used to calibrate core-scanning data there is neither implicit nor explicit control on physical plausibility (e.g., concentrations could theoretically become negative and larger than 100%). This lack of control, in turn, frequently results in negative concentration estimates or concentration estimates exceeding 100%.

A major breakthrough in the calibration of XRF core-scanning data has been the study of Weltje and Tjallingii (2008). Firstly, they solved the lack of control on physical plausibility, common to many conventional calibration equations. Secondly, they proposed a definition of the calibration problem in such a way that problems associated with the variable measurement geometry and matrix were greatly reduced, if not eliminated.

Solving these problems was achieved by combining well-established spectroscopic theory with the statistical theory of compositional data analysis, as developed by John Aitchison in the 1980s (Aitchison, 1986) and extended ever since (see e.g., Egozcue et al., 2003;

Pawlowsky-Glahn and Egozcue, 2001). This theoretical framework is concerned with describing the statistical properties of compositional quantities, such as their sample space and the problems surrounding modelling and statistical inference. Furthermore, Aitchison (1986) proposed a mathematical framework with which these problems can be circumvented. This framework consists of invertible transformations (i.e. log-ratio transformations or simply "log-ratios") which ensure that the (from a statistical standpoint) problematic data are represented in a suitable metric space.

Key behind the calibration method of Weltje and Tjallingii (2008) was to define the calibration problem in terms of log-ratios, thereby treating the element intensities as well as the analyte concentrations as compositional quantities. This reduces the effects from the variable measurement geometry, it greatly simplifies modelling the highly non-linear matrix effects and it ensures that concentration estimates are always physically tractable. Weltje and Tjallingii (2008) conclude that for the soft-sediment core they analysed, the quantitative performance of the XRF core scanner was comparable to conventional geochemical analysis.

1.4 The multivariate approach

The fact that the quantitative performance of XRF core scanning has become comparable to conventional geochemical analysis means that it has great potential for studies governing mass balances. However, because the geochemical composition is typically strongly correlated with numerous other sediment properties, the potential value of geochemical core logs is certainly not limited to constraining mass-balance exercises. Hence, XRF core scanning technology potentially provides much more than "just" a way to quickly acquire geochemical data.

Using chemical data to predict other properties requires a multivariate approach. This approach is followed throughout this thesis. The rationale behind the multivariate approach is illustrated in Figure 1.1. The ellipsoids in Figure 1.1 represent data sets which reflect a certain variability. Correlation between the two data sets X and Y is represented as overlap, which is highlighted in grey. Calibration can be considered going from a not very meaningful data set (X) to a meaningful data set (Y) by using some parametric model to capture this overlap. In the light of calibration, $X \setminus Y$ may be interpreted as noise and $Y \setminus X$ reflects the unpredictable variability in Y which may be interpreted as a shortcoming of X . Since the goal is to predict Y from X and $Y \setminus X$ has a different significance than $X \setminus Y$, calibration is an *asymmetric* exercise. Also when both X and Y are meaningful measures, we may attempt to parameterise the overlap between X and Y to gain more insight into their relation. In this case, the unique information (i.e. $X \setminus Y$ and $Y \setminus X$) is of potential value because both data sets are meaningful. For instance, the chemical and mineralogical composition will overlap to a large extent. Variability unique to the mineralogical composition will be controlled by minerals with the same chemical composition. In contrast to calibration, multivariate analysis of sets of meaningful data sets may be considered a *symmetric* exercise. An important task is to turn the conceptual model behind the multivariate approach into algorithms with practical value.

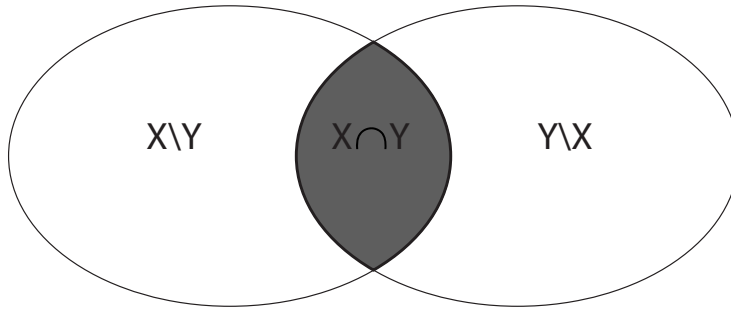


Figure 1.1: Graphical representation of the rationale behind the "multivariate approach". Details are provided in the text. Modified from Bloemsmma et al. (2012).

1.5 Objectives

Ever since the introduction of the first XRF core scanner, the performance and versatility of these instruments has been improved. Important milestones were the implementation of efficient (Si-drift) x-ray detectors, and the introduction of an adjustable slit opening. Although there is still room for improvement, core scanning technology is relatively mature. Data processing and handling methods, on the other hand, are lagging behind. The research presented in this thesis is therefore centred around developments in software, rather than hardware.

The first goal of this study is to build a mathematical-statistical framework for the construction of so-called *integrated core descriptions* using XRF core scanning. We define an integrated core description as a collection of lithofacies, chemical and petrophysical records which have the same, high (1 cm) resolution and quantified uncertainties. The interpretation of these descriptions, possibly in combination with other data sets, is referred to as *integrated core analysis*. The second goal of this study is development of methods that facilitate integrated core analysis.

To reach these goals, we define the following six objectives:

1. **Characterise statistically** the relevant data types.
2. **Formalise** interpretation of, and geological inference from geochemical data using mathematical-statistical methods.
3. **Improve** the logratio-based calibration method so that the misfit between composition predicted by the scanner and composition estimated using conventional analytical techniques may be reduced.
4. **Extend** the logratio-based calibration method by enabling prediction of properties which were not directly observed.
5. **Evaluate** the performance of newly-proposed calibration methods for different types of cores (soft-sediment cores, and consolidated-sediment cores).
6. **Explore** the added value of core-scanning technology within the routine core analysis workflow using real-life cases.

1.6 Structure of this thesis

Chapter 2 provides a mathematical description of the calibration problem and shows the rationale behind the logratio-based approach. In addition, a multivariate **extension** is presented which enables using the scanner as a fully quantitative geochemical tool (i.e., also prediction of the total concentration of unobserved analytes). The predictive performance of this method is compared to other methods using two core data sets.

In Chapter 3, we describe the problems surrounding interpretation of geochemical data, which have the tendency to reflect the combined response of different mechanisms. One important bulk-chemistry-controlled mechanism is variation in grain size. We **formalise** interpretation of bulk chemical composition and grain size data using a latent-variable technique.

Given that our ultimate goal is to compare the quantitative performance of the core scanner with that of conventional methods, it is important that we **statistically characterise** core-scanning and geochemical data. This is attempted in Chapters 4 and 5. In Chapter 4, we perform error propagation to study the properties of spectroscopic and compositional data in logratio space. Whereas the starting point in Chapter 4 is the theory, the starting point in Chapter 5 is the existing empirical evidence with regard to analytical uncertainty. It is investigated how this evidence compares to what is known about spectroscopic and compositional data which led to the identification of some inconsistencies. Finally, a new uncertainty modelling workflow is proposed.

In Chapters 6 and 7, we build an *integrated core description* of two cores: one heterogeneous soft-sediment core from the shallow subsurface (B38D) and one heterogeneous core of siliciclastic reservoir rock (E10-3). In Chapter 6, a framework is presented which enables prediction of categorical core data from scanning data. In Chapter 7, a new calibration method is proposed which enables prediction of chemical as well as petrophysical properties. The quality of the predictions is compared with the quality of conventional analytical methods.

Chapter 8 is centered around *integrated core analysis*. By means of three applications it will be illustrated how core-scanning data can facilitate in making a detailed reservoir quality characterisation which is not possible based on the data acquired in routine core analysis.

An **evaluation**, discussion and recommendations are given in Chapter 9.

Multivariate logratio calibration of XRF-CS data ¹

2.1 Introduction

It is common practice to convert XRF core scanner output to element or oxide concentrations using linear regression on an element-by-element basis (referred to as direct linear calibration; DLC) (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:

$$W_{ij} = a_j I_{ij} + b_j \quad (2.1)$$

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 Equation 2.1. 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), irregu-

¹Based on: Weltje, G.J., Bloemsmma, M.R., Tjallingii, R., Heslop, D., Röhl, U., Croudace, I.W., 2015. Prediction of geochemical composition from XRF-core-scanner data: A new multivariate approach including automatic selection of calibration samples and quantification of uncertainties. In: Croudace, I.W., Rothwell, G. (Eds.), *Micro-XRF Studies of Sediment Cores*. DPER special publication.

larities 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).

2.2 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):

$$W_{ij} = K_j I_{ij} M_{ij} S_i \quad (2.2)$$

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).

2.3 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 content and grain-size distribution of sediments, and the presence of burrows) 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 Equation 2.2 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 Pederson, 2007). Down-core patterns of element-intensity ratios have proved extremely useful for correlation (e.g. Pälke 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 bivariate log-ratio calibration (BLC) equation by combining two conventional calibration equations (Eqn. 2.2) 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:

$$\frac{W_{ij}}{W_{iD}} = \frac{K_j I_{ij} M_{ij} S_i}{K_D I_{iD} M_{iD} S_i} = \frac{K_j I_{ij} M_{ij}}{K_D I_{iD} M_{iD}} \quad (2.3)$$

Calibration in terms of ratios removes the measurement geometry from the problem to be solved, given that 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:

$$\frac{K_j}{K_D} = B_{jD} \quad (2.4)$$

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

$$\frac{M_{ij}}{M_{iD}} = C_{jD} \left(\frac{I_{ij}}{I_{iD}} \right)^{A_{jD}} \quad (2.5)$$

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 Equations 2.4 and 2.5 into Equation 2.2 gives:

$$\frac{W_{ij}}{W_{iD}} = B_{jD} C_{jD} \left(\frac{I_{ij}}{I_{iD}} \right)^{A_{jD}+1} \quad (2.6)$$

Equation 2.6 may be rewritten by defining the empirical coefficients:

$$\begin{aligned} \alpha_{jD} &= A_{jD} + 1 \\ \beta_{jD} &= \log(B_{jD} C_{jD}) \end{aligned} \quad (2.7)$$

If we substitute these into Equation 2.6 and take logarithms, we obtain the BLC equation:

$$\log \left(\frac{W_{ij}}{W_{iD}} \right) = \alpha_{jD} \log \left(\frac{I_{ij}}{I_{iD}} \right) + \beta_{jD} \quad (2.8)$$

Equation 2.8 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 (Eqn. 2.2), 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 BLC 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. Eqn. 2.1).

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:

$$x_{ij} = \log \left(\frac{\tilde{W}_{ij}}{\tilde{W}_{iD}} \right) \quad (2.9)$$

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:

$$\tilde{W}_{ij} = \frac{\exp(x_{ij})}{\sum_{j=1}^D \exp(x_{ij})} \quad (2.10)$$

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).

2.4 "Absolute" concentrations

Although BLC presents a major step forward compared to DLC, it does not allow "absolute" concentrations to be determined. This flaw is the result of the fact that XRF core scanners measure a limited range of elements only. Hence, the sum of all detectable elements is equal to a proportion 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:

$$\text{Undef} = W_{i,D+1} = 1 - \sum_{j=1}^D W_{ij} \quad (2.11)$$

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", however, requires a more flexible definition of the calibration problem permitting prediction of variables which were not directly observed. In contrast to the bivariate nature of Equation 2.8, this requires a multivariate approach.

2.5 MLC workflow

To enable prediction of "Undef" and potentially increase the predictive capabilities of calibrated XRF core scanning data, a multivariate extension of the BLC model is presented, i.e. the Multivariate Calibration (MLC) model. Following suggestions of Weltje and Tjallingii (2008), this multivariate calibration method approach is embedded in an integrated workflow which is shown in Figure 2.1. Its components are: (1) a multivariate calibration approach which permits unbiased prediction of "absolute" concentrations, (2) an algorithm which estimates the level of noise based on the measured spread among replicate intensities and concentrations, and (3) an algorithm for automatic selection of calibration samples, based on the multivariate geometry of the scaled intensities.

2.5.1 Multivariate calibration approach

The transformation in Equation 2.9 may be extended to materials comprising more than two components. This operation is known as the additive log-ratio (alr) transformation:

$$\mathbf{y} = \text{alr}_D(\mathbf{x}) = \left[\log\left(\frac{x_1}{x_D}\right), \log\left(\frac{x_2}{x_D}\right), \dots, \log\left(\frac{x_{D-1}}{x_D}\right) \right] \quad (2.12)$$

Using this operator, Equation 2.8 may be written in vector notation:

$$\text{alr}_D(\mathbf{w}) = \text{alr}_D(\mathbf{e})\mathbf{A}_D + \mathbf{b}_D \quad (2.13)$$

where \mathbf{A}_D is a diagonal matrix with all α 's associated with denominator D on its diagonal, \mathbf{b}_D is a vector containing all β 's, and \mathbf{e} and \mathbf{w} are vectors with the measured intensities and concentrations, respectively.

Additive logratio transformations employed in BLC belong to a larger family of logratio transformations (Aitchison, 1986; Aitchison and Greenacre, 2002; Egozcue et al., 2003). The transformation most suitable for multivariate analysis is the centred log-ratio (clr) transformation which is defined as:

$$\text{clr}(\mathbf{x}) = [\log(x_1) - \log(g(\mathbf{x})), \log(x_2) - \log(g(\mathbf{x})), \dots, \log(x_D) - \log(g(\mathbf{x}))] \quad (2.14)$$

In this transformation, $g(\mathbf{x})$ is the geometric mean over all elements. The major advantage of the centred log-ratio (clr) transformation is that transformed data may be analysed with least-squares methods. Furthermore, the clr-transformation enables calibration of all elements simultaneously (i.e., including "Undef"). The MLC is therefore based upon the following equation:

$$\text{matclr}(\mathbf{W}) = \text{matclr}(\mathbf{E})\mathbf{A} + \mathbf{b} \quad (2.15)$$

In this equation, $\text{matclr}(\dots)$ and $\text{matclr}^{-1}(\dots)$ reflect the row-wise application of the forward and inverse clr-transformation, respectively.

Equation 2.15 has essentially the same form as the BLC. The matrix \mathbf{A} and vector \mathbf{b} serve the same purpose as \mathbf{A}_D and \mathbf{b}_D in Equation 2.8. One fundamental difference, however, is that the MLC model is unique since it does not rely on selection of the denominator element. It also enables prediction of elements which have not been measured by the XRF core scanner using multivariate estimation, thereby exploiting the covariance among intensities and concentrations, so as to generate "absolute" concentrations. In practice, this means that \mathbf{A} will not be diagonal. Covariance among intensities may be due to absorption and enhancement. Covariance among analytes may be caused by the fact that certain elements may reside in the same mineral. By exploiting these covariances, the MLC model allows further minimisation of the prediction error relative to the BLC.

2.5.2 Parameter estimation

Parameter estimation in the MLC workflow 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 the cross-covariance matrix (de Jong, 1993). PLS is the most suitable regression method in case data matrices are not of full rank (Wold et al., 1983), which is the case for clr-transformed data sets. Another property of PLS, which it shares

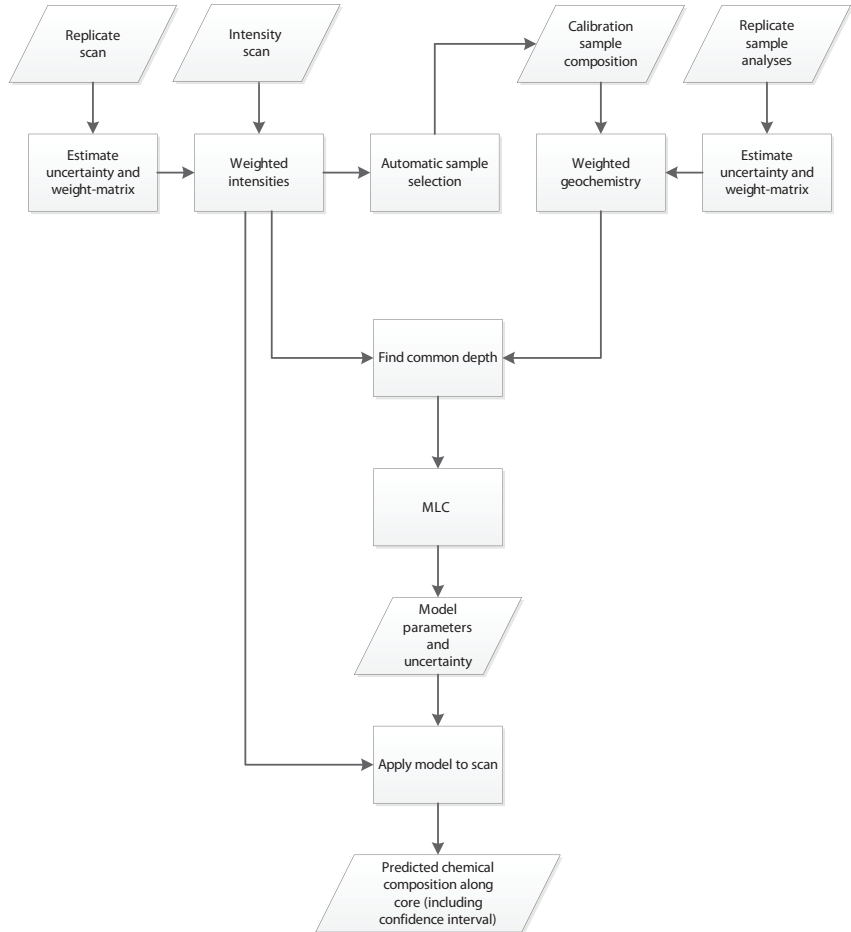


Figure 2.1: MLC workflow; see text for discussion.

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, 1993). 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 (we recommend a minimum of 30 specimens).

A straightforward goodness-of-fit measure of logratio-transformed compositional data is the mean squared Euclidian distance between predicted and measured composition:

$$\text{MSE} = \frac{1}{nD} \sum_{i=1}^n \sum_{j=1}^D [\text{clr}(\hat{w}_{ij}) - \text{clr}(w_{ij})]^2 \quad (2.16)$$

In the BLC 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 Equation 2.16, 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.

2.5.3 Replicates and scaling

In the absence of prior knowledge about the uncertainty of intensities and analyte concentrations, we may collect replicate measurements according to a systematic sampling scheme, e.g. collect at least three measurements at every tenth location down-core. Using replicate sets of intensities, we may estimate the level of noise from the average spread within each set of replicates in log-ratio space. Assuming that the noise follows a normal distribution in this space, these estimates of spread facilitate calculation of confidence limits around logratios of intensities for a user-defined significance level (e.g., 5% or 10%).

The spread among replicate intensities and concentrations may also be used to robustify the calibration model. Scaling of clr-transformed variables ensures that the resulting uncertainties have on average the same magnitude, which is the fundamental assumption underlying least-squares estimation methods. In case two variables reflect the same underlying signal, scaling ensures that the variable with the high precision will have a greater importance than the variable with the low precision in the calibration model. The calibration model in terms of scaled variables reads:

$$\text{matclr}(\mathbf{W})\mathbf{L}_w = \text{matclr}(\mathbf{E})\mathbf{L}_e\mathbf{A}_* + \mathbf{b}_* \quad (2.17)$$

where \mathbf{L}_e and \mathbf{L}_w are diagonal matrices with the inverse of the average standard deviation among clr-transformed replicate intensities and concentrations on their diagonals.

The MSPE is also calculated in this scaled clr space, which makes it a pseudo maximum likelihood criterion.

Whereas \mathbf{A}_* and \mathbf{b}_* are model parameters in terms of the scaled clr-transformed data, a regression model in terms of the unscaled data may be obtained by applying:

$$\mathbf{A} = \mathbf{L}_e \mathbf{A}_* \mathbf{L}_w^{-1} \quad (2.18)$$

and subsequently:

$$\mathbf{b} = \bar{\mathbf{w}} - \bar{\mathbf{e}} \mathbf{A} \quad (2.19)$$

where $\bar{\mathbf{e}}$ and $\bar{\mathbf{w}}$ are the column means of the unscaled clr-transformed intensities and concentrations, respectively. Given \mathbf{A} and \mathbf{b} , the $N \times D$ matrix with concentration estimates $\hat{\mathbf{W}}$ associated with the $N \times D$ matrix with measured intensities \mathbf{E} may be obtained as follows:

$$\hat{\mathbf{W}} = \text{matclr}^{-1}(\text{matclr}(\mathbf{E})\mathbf{A} + \mathbf{1}_n \mathbf{b}) \quad (2.20)$$

2.5.4 Summary statistics

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. To illustrate the nature of the misfit, we 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). Hence, these quantities are not being used to justify model selection, but are purely intended to summarise data-model discrepancies.

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 $u = [1, 1]$ (i.e. the line $y = x$ in vector form) and v_1 is the first major axis or eigenvector of the data as determined by means of Singular Value Decomposition, B is given by:

$$B_j = 100\% \times \frac{|0.5\pi - \cos^{-1}(u \cdot v_1)|}{0.5\pi} \quad (2.21)$$

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:

$$S_j = 100\% \times \frac{\lambda_2}{\lambda_1 + \lambda_2} \quad (2.22)$$

Hence, S and B are in principle independent and can therefore also be interpreted as such. Figure 2.2 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 \bar{S} and \bar{B} .

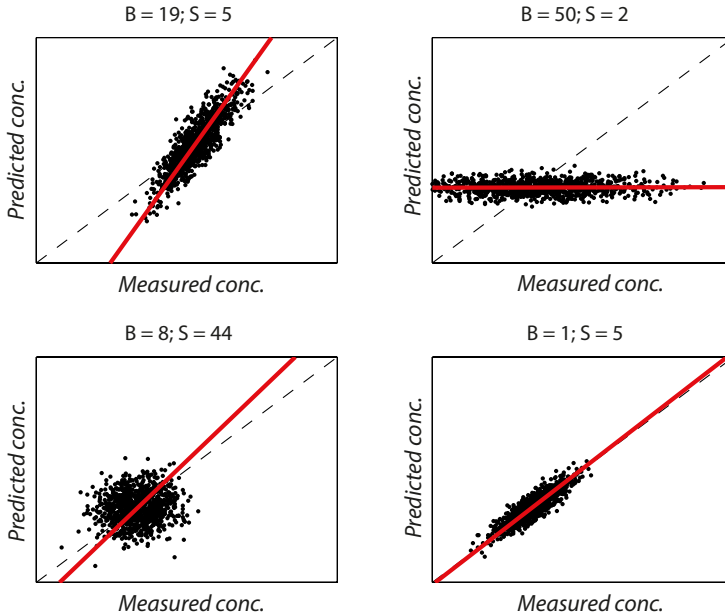


Figure 2.2: Definition of scatter (S) and bias (B) in compositional space (i.e. concentrations).

2.5.5 Automatic selection of calibration samples

Effective calibration requires that the compositional trends in the data are adequately sampled. The average uncertainty of the clr-transformed and scaled intensities is identical, which implies that Euclidian distances between intensity vectors may be directly interpreted in terms of compositional variability. This provides a solid basis for automatic selection of calibration samples using least-squares estimation techniques which forms an integral part of the MLC workflow (Fig. 2.1).

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 et al., 1999) 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. \ln -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. 2.1).

2.6 Comparative calibration exercise

We will compare the performance of four different calibration models (DLC-1, DLC-2, BLC, 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.

2.6.1 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 μ A) 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).

2.6.2 Data set 2: AU10v

Core AU10v was collected in July 2008 from Augusta Harbour, Sicily (see also Croudace et al., 2015). The core was split and the top 9.6 cm was scanned at 500 μ m 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.

2.7 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

Data set	Model	B [%]	S [%]	MSPE [$\times 10^{-4}$]
GeoB7920	DLC1	11.69	14.31	319
GeoB7920	DLC2	15.66	13.02	131
GeoB7920	BLC	3.46	8.59	124
GeoB7920	MLC	3.80	6.90	80
AU10v	DLC1	14.88	9.48	530
AU10v	DLC2	13.86	12.09	239
AU10v	BLC	7.40	8.80	218
AU10v	MLC	5.43	7.90	163

Table 2.1: Comparative performance of calibration techniques based on cross validation.

the calibration samples. Elements which could not be calibrated were added to "Undef" (except in the case of the BLC 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 2.1 gives the goodness-of-fit statistics of the four calibration models, whereas Figures 2.3 to 2.8 permit a visual appraisal of their predictive power. The MSPE values of both data sets show that the MLC model performs better than the BLC model, the BLC model performs better than the DLC-2 model, and the DLC-1 model is the least satisfactory (Table 2.1). 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 (BLC) performs much better than the most advanced direct linear calibration model (DLC-2).

Figure 2.9 provides a visual impression of core GeoB7920. In the upper half of this figure, the high-resolution RGB image (Fig. 2.9A) is shown alongside the records of $\ln(\text{Ca}/\text{Ti})$ intensities (Fig. 2.9B) and predicted concentrations (Fig. 2.9C, D). The gray bands in these four graphs represent the 95% confidence interval of measured intensities and predicted concentrations. Comparison of the raw (Fig. 2.9B) and calibrated (Fig. 2.9C, D) intensities highlights their close similarity. Furthermore, the MLC estimate (Fig. 2.9D) has a much smaller confidence interval than the BLC estimate (Fig. 2.9C). The lower half of the image (Fig. 2.9E to 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 2.10 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 20), the automatic selection method yields the lowest MSPE and thus provides better results than the other methods.

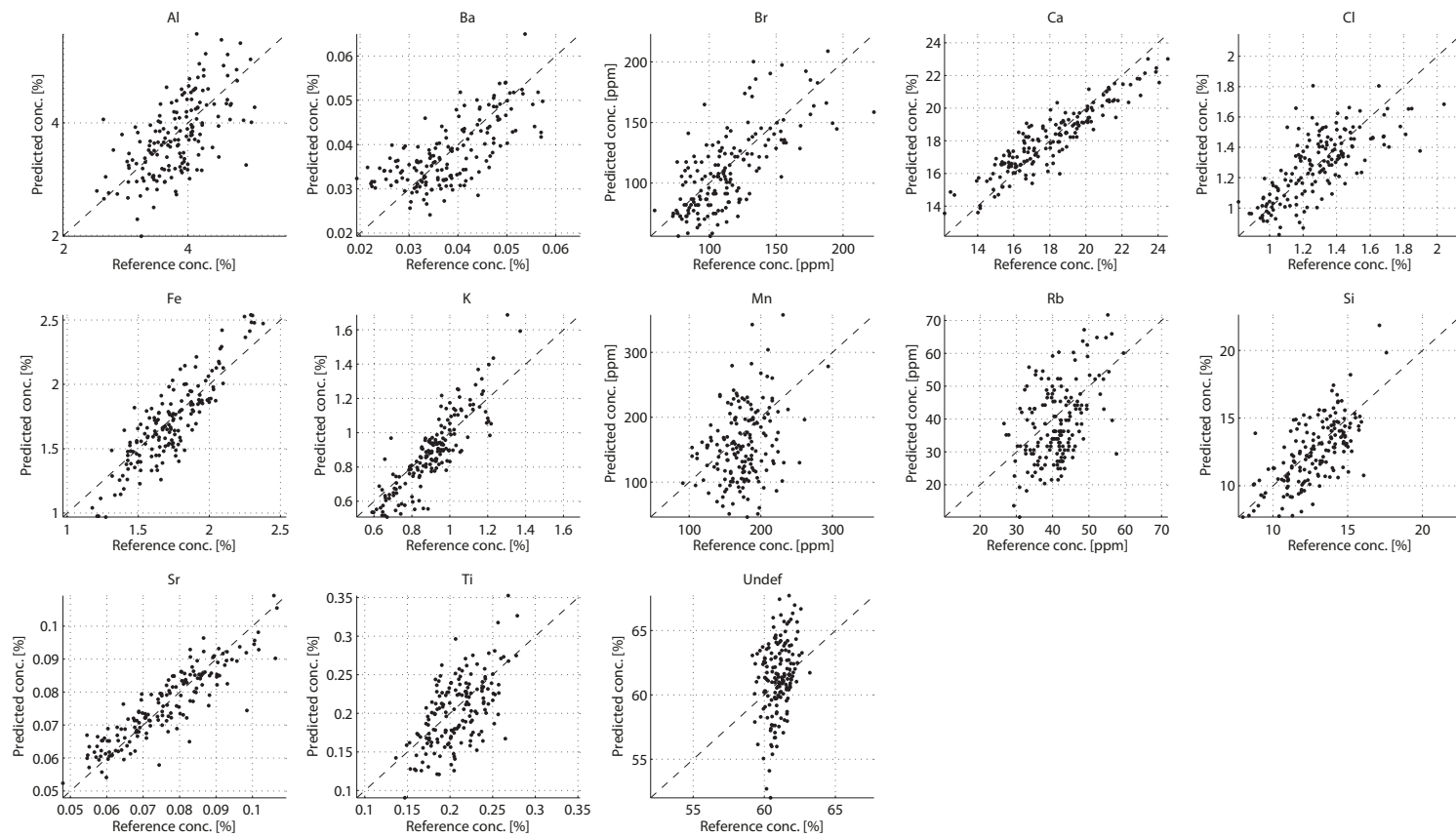


Figure 2.3: Measured versus predicted concentrations for DLC-1 model of GeoB7920.

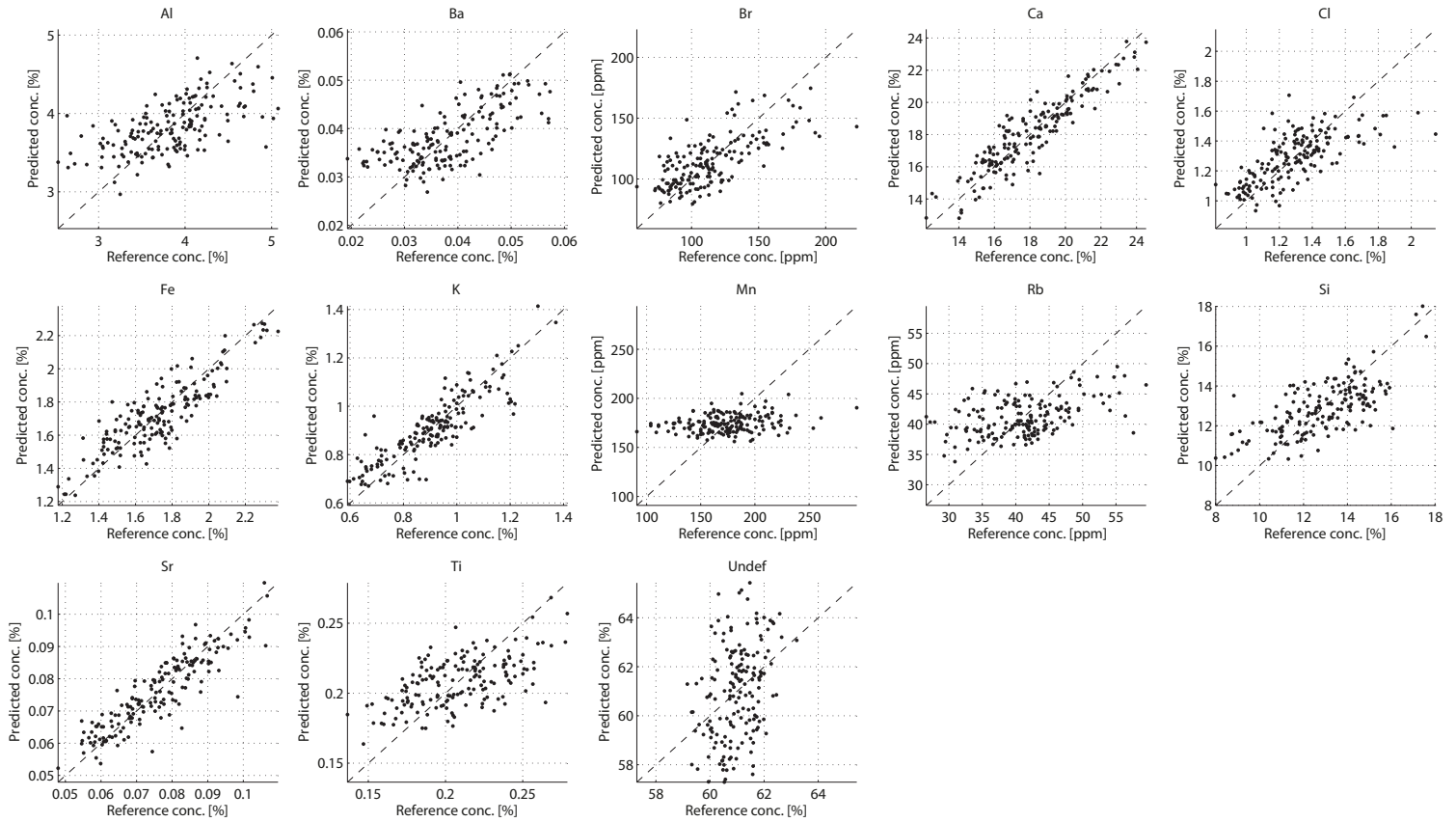


Figure 2.4: Measured versus predicted concentrations for DLC-2 model of GeoB7920.

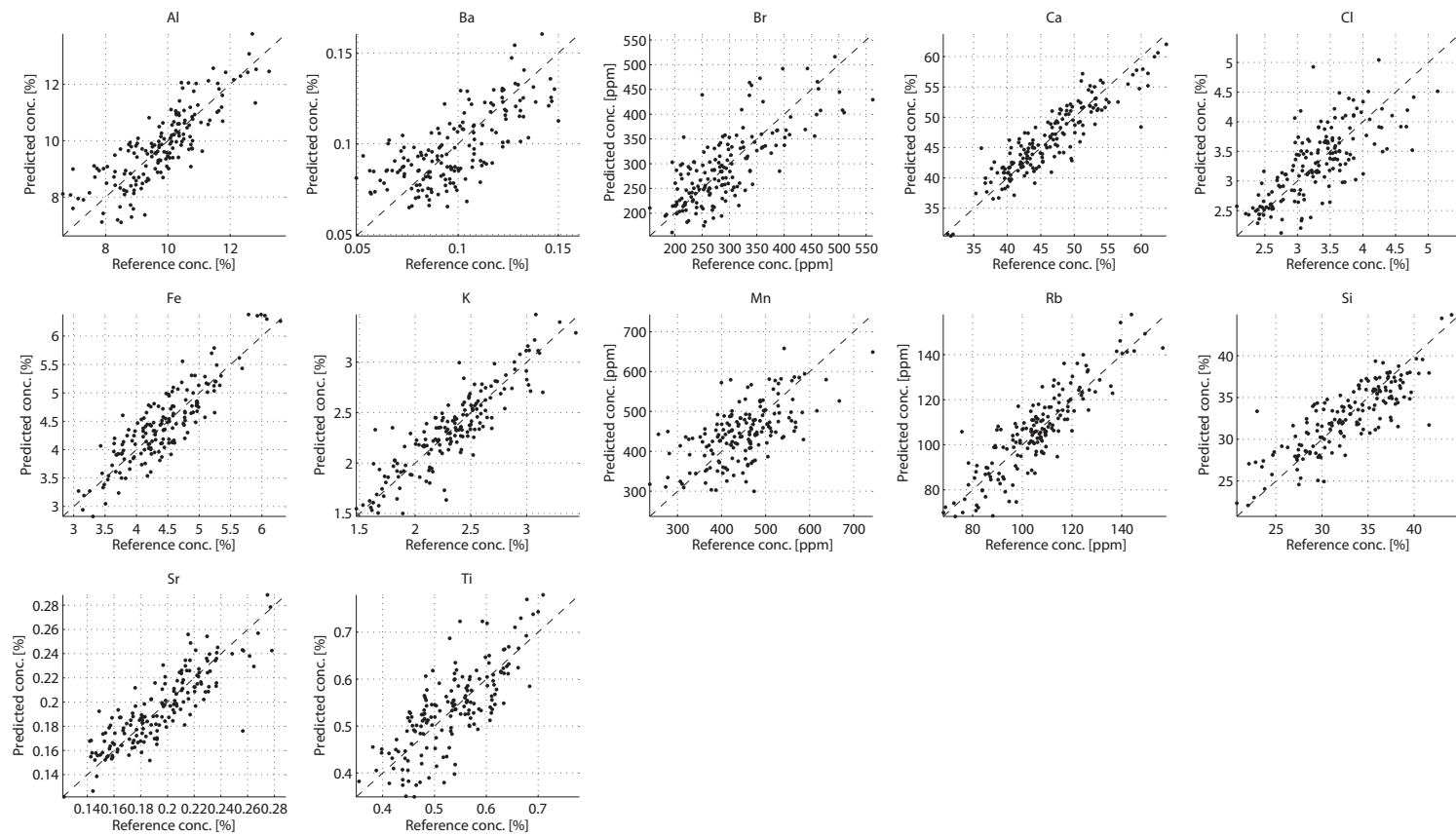


Figure 2.5: Measured versus predicted concentrations for BLC model of GeoB7920.

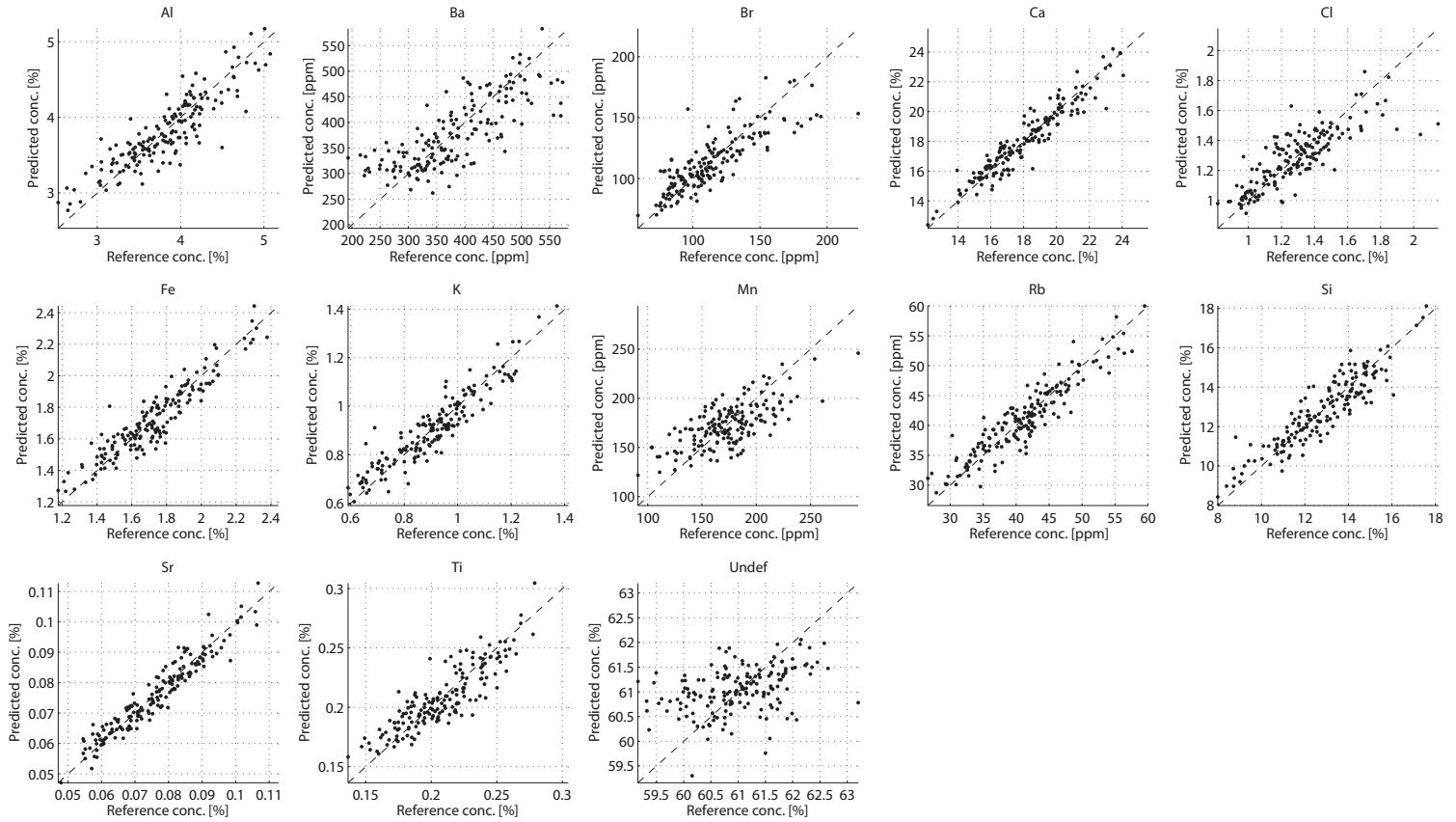


Figure 2.6: Measured versus predicted concentrations for MLC model of GeoB7920.

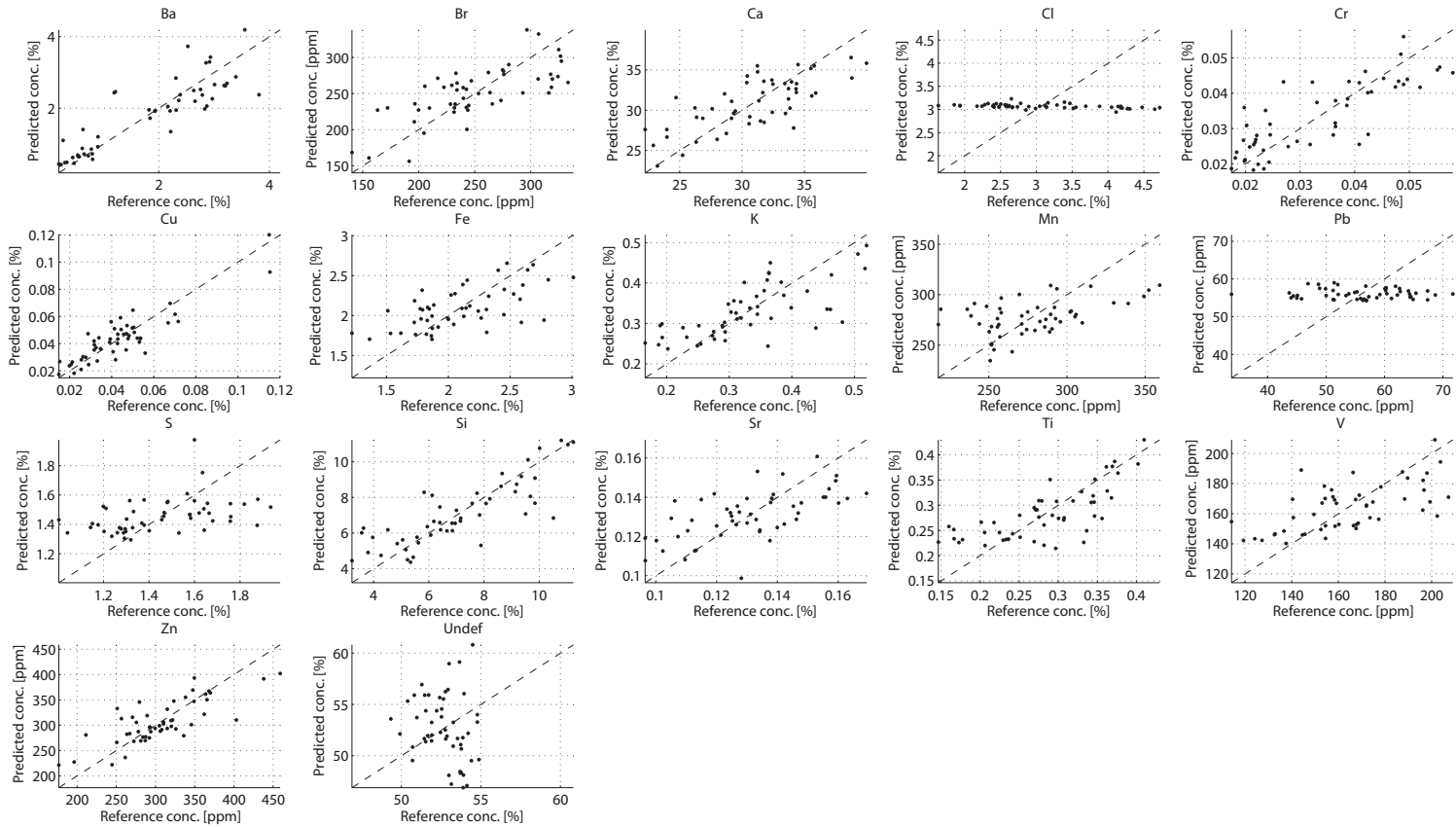


Figure 2.7: Measured versus predicted concentrations for DLC-2 model of AU10v.

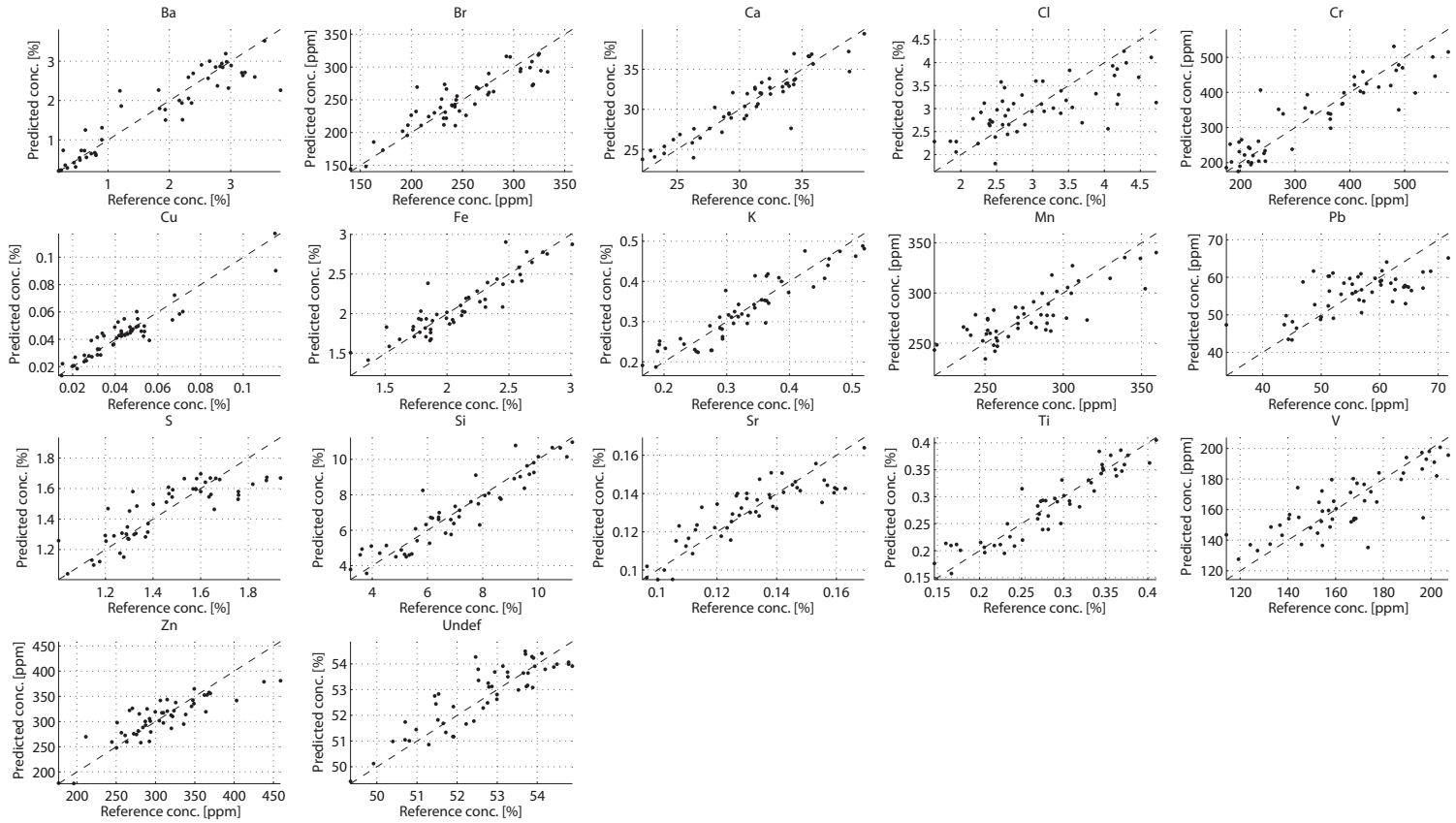


Figure 2.8: Measured versus predicted concentrations for MLC model of AU10v.

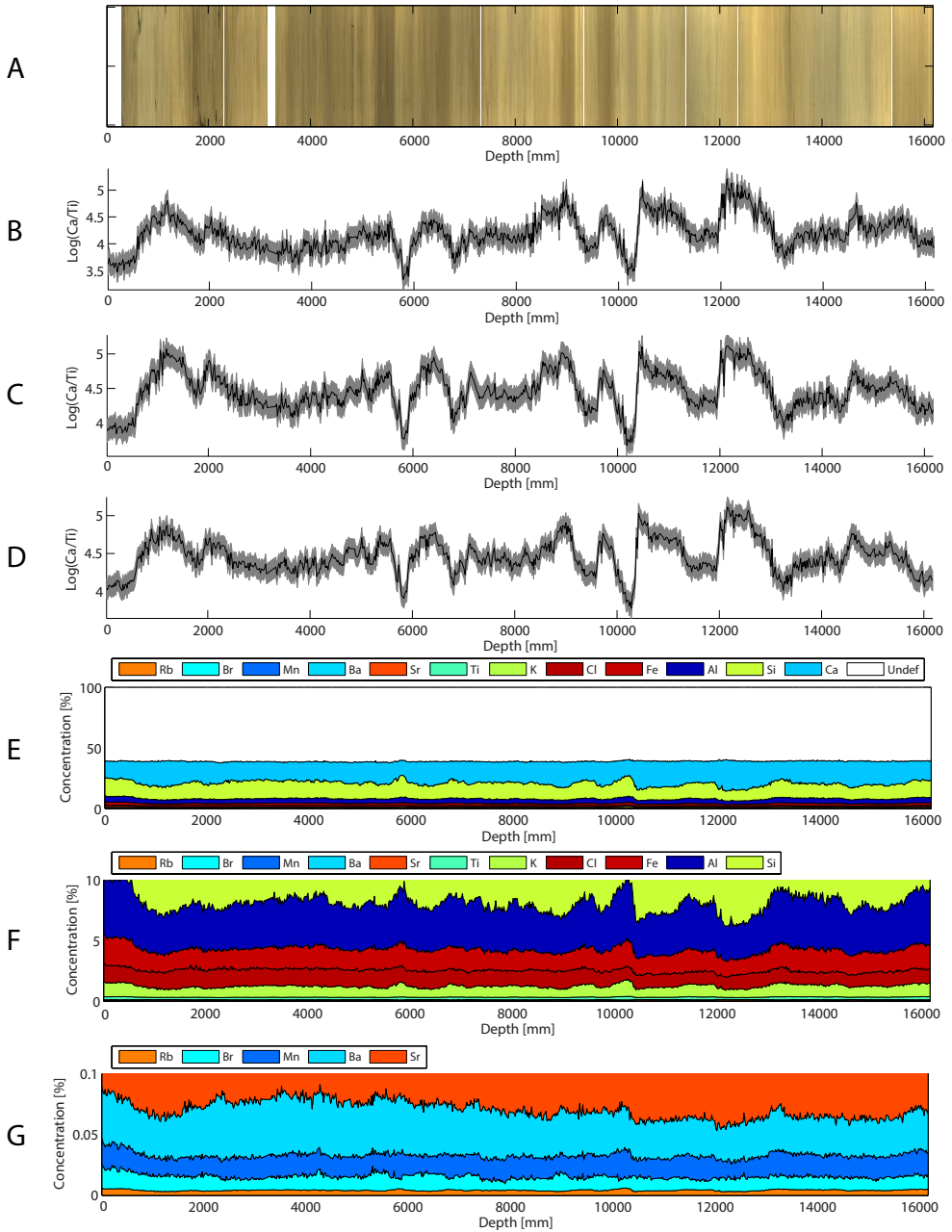


Figure 2.9: Overview of core GeoB7920. A: high-resolution RGB image; B: raw $\ln(\text{Ca}/\text{Ti})$ intensities with 95% confidence interval; C: BLC prediction of $\ln(\text{Ca}/\text{Ti})$ concentrations with 95% confidence interval; D: MLC prediction of $\ln(\text{Ca}/\text{Ti})$ concentrations with 95% confidence interval; E: Concentrations from 0% to 100%; F: Concentrations from 0% to 10%; G: Concentrations from 0% to 0.1%. Elements are sorted in descending order of average concentration.

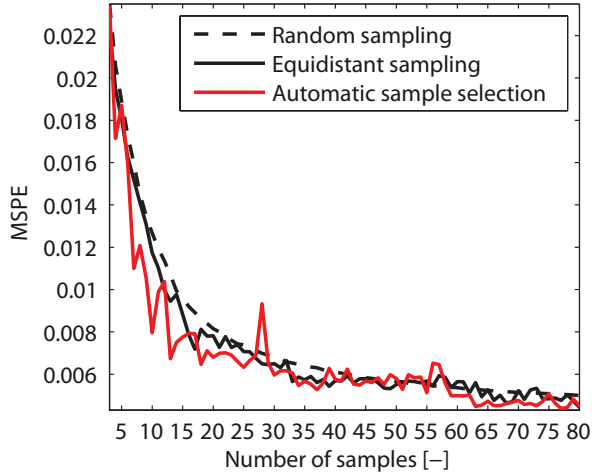


Figure 2.10: Comparative performance of sampling strategies in the MLC model for core GeoB7920.

2.8 Discussion

2.8.1 Comparative performance of calibration models

Because all statistics in Table 2.1 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 (Eqn. 2.1) bears no relation to the calibration equation used in XRF spectrometry (Eqn. 2.2); (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 (Eqn. 2.2), 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 BLC and MLC models indicates that the latter provides a more comprehensive description of the calibration problem at hand. The BLC 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 BLC predictions depends on which element is selected as 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 BLC model.

The BLC model (Eqn. 2.8) 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. 2.9B to 2.9C). Moreover, the linear transformation embodied in Equation 2.8 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 Bloemsmas 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.

2.8.2 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. 2.10) 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:

$$N_{I,s} \geq \text{ceil}(n^{1/2}) \quad (2.23a)$$

$$\Delta N_{I,s} \geq \text{floor}(n/N_{I,s}) \quad (2.23b)$$

$$N_{I,r} \geq 3 \quad (2.23c)$$

$$N_{I,tot} = n + N_{I,s}(N_{I,r} - 1) \quad (2.23d)$$

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 equations 2.23a-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:

$$N_c \geq 3D \quad (2.24a)$$

$$N_{W,s} \geq \text{ceil}(D^{1/2}) \quad (2.24b)$$

$$N_{W,r} \geq 3 \quad (2.24c)$$

$$N_{W,tot} = N_c + N_{W,s}(N_{W,r} - 1) \quad (2.24d)$$

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.

2.9 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 heterogenous materials under laboratory conditions. It represents a vast improvement

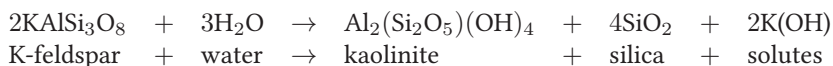
over previous attempts at direct calibration of intensities using concentrations, and is significantly better than the bivariate BLC model proposed by Weltje and Tjallingii (2008). The main advantages of multivariate log-ratio calibration (MLC) over bivariate log-ratio calibration (BLC) are (1) elimination 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; Bloemsmas et al., 2012).

Modelling the joint variability of grain size and chemical composition in sediments ¹

3.1 Introduction

Geochemical analysis is a powerful method for quantifying bulk sediment properties. It may be used to characterise the composition of the parent rock or the climatic-physiographic conditions under which the sediment was formed (Pettijohn et al., 1987; Johnsson, 1993; Basu, 2003; Weltje and von Eynatten, 2004). Alternatively, it may be used to assess compositional modifications caused by weathering, sorting, mixing and diagenesis (McLennan et al., 1993; Nesbitt and Young, 1996; Amorosi et al., 2002; Dinelli et al., 2007; Pe-Piper et al., 2008). Because geochemical variation typically reflects the superposition of these factors, independent information is required to resolve them individually. This applies specifically to analysis of clastic sediments, where multiple equally plausible hypotheses may be postulated, given the observed compositional data.

It is well known that grain size and bulk geochemical composition of clastic sediments are strongly correlated. The tight connection between grain size and composition is a consequence of the processes which govern the generation of sediments from crystalline rocks. Chemical weathering leads to release of unstable elements as solutes, while stable elements such as Al remain in the solid phase (e.g., Nesbitt and Young (1984)). Consider for instance chemical weathering of K-feldspar:



Since kaolinite particles are relatively fine grained, whereas the K-feldspar grains are of sand size, chemical alteration is accompanied by textural modification of the sediment. If the K-feldspar and kaolinite are transported away from their source area and deposited in

¹Based on: M.R. Bloemsmma, M. Zabel, J.B.W. Stuut, R. Tjallingii, J.A. Collins, G.J. Weltje (2012). Modelling the joint variability of grain size and chemical composition in sediments. *Sedimentary Geology* 280. pp. 135-148.

a sedimentary basin, the resulting product is a sediment with a spatially variable grain-size distribution. If we sample this material at different localities, we expect Si/Al of the bulk sediment to correlate positively with mean grain size. Quite often, the objective of geochemical analysis is not to approximate grain size, but to provide a unique compositional fingerprint of the sediment for the purpose of palaeoclimate or provenance reconstruction. In such cases, we would like to eliminate the size-dependency from our data.

A method to eliminate variation of bulk grain-size distribution from sediment geochemical analysis is to analyse a narrow size fraction (Weltje and Brommer, 2011). In the above example, the composition of a narrow size fraction would be constant across the basin. However, reality is more complicated. If we assume that the extent of weathering in the source area of the sediment has varied over time, we expect that the composition of a narrow size fraction extracted from samples taken at different stratigraphic levels records this variation. For example, the Si/Al ratio within the sand fraction will correlate positively with the extent of weathering, whereas the Si/Al ratio of the clay fraction will correlate negatively with the extent of weathering. Because separation of samples into narrow size fractions is time consuming and expensive, we should try to find more practical ways of eliminating the grain-size bias from geochemical data.

In marine geosciences, it is commonplace to use a so-called multi-proxy approach to infer palaeo-climatological signals. In this approach, high-resolution records of physical and chemical properties, for example grain-size distribution and bulk chemical composition, are measured on the same sediment core. Subsequently, these data are displayed side by side to highlight consistency between two data sets, or to illustrate patterns of variability unique to each data set. A widely accepted method to select the relevant signals (i.e. the relevant element ratios and/or grain-size parameters) from multivariate data sets is not yet in place, and the same applies to generic procedures to examine the internal consistency of selected proxies. In this contribution, we propose a fully quantitative and formal method for simultaneous analysis of multiple data sets. Our analysis of the relation between the geochemical composition and grain-size distribution of sediment will serve to illustrate the implications of the proposed statistical framework for multi-proxy analysis.

3.2 Material and methods

3.2.1 Materials

We use three marine soft-sediment cores, recovered from the margin off West Africa and Chile, to investigate geochemical variation with grain size of lithogenic sediments. These cores are well suited for testing this variation since both West Africa and Chile receive a considerable amount of sediment from the continent transported by wind and rivers (e.g., Sarnthein et al., 1981; Stuut et al., 2007). In addition to terrestrial input, both areas are characterised by strong year-round upwelling, resulting in a considerable biogenic sediment input (Haslett and Smart, 2006; De Pol-Holz et al., 2007).

Cores GeoB9508-5 and GeoB7920-2 were retrieved from the continental margin off West Africa, and core GeoB7139-2 was retrieved from the margin off Chile (Fig. 3.1). Core GeoB9508-5 (15°30N/17°57W, 2384 m water depth) is approximately 9.5 m long and was obtained during RV Meteor cruise M65/1 (Mulitza and cruise participants, 2006). Core GeoB7920-2 (20°45N/ 18°35W 2278 m water depth) is approximately 16 m long and was retrieved during RV Meteor cruise M53/1 (Meggers and cruise participants, 2002). The

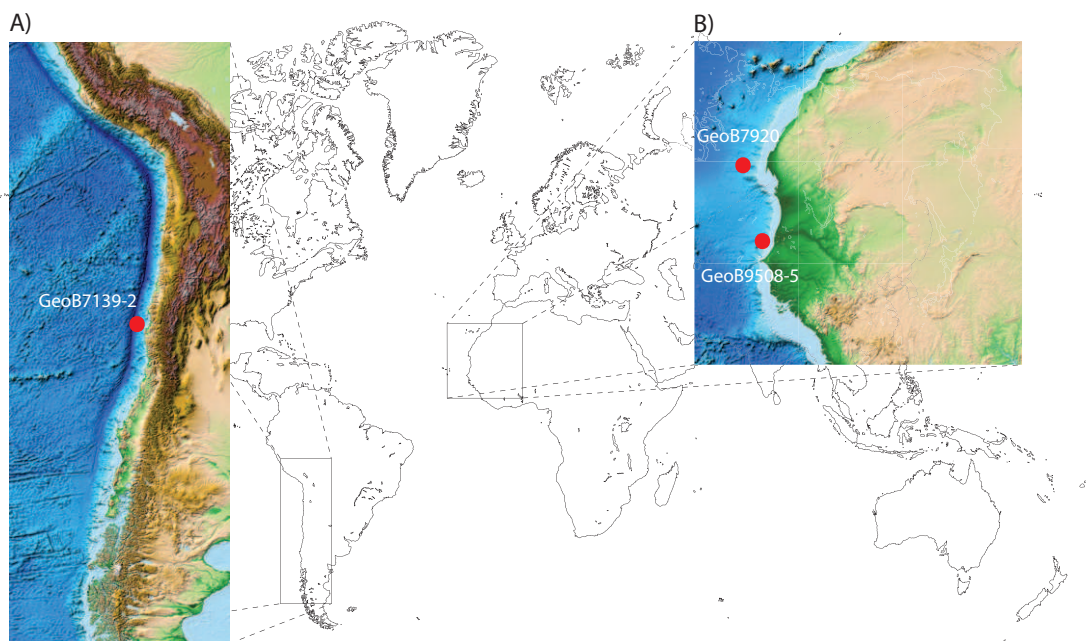


Figure 3.1: Location of the cores used in this study. Core GeoB7139-2 was retrieved off the Chilean coast (inset A), whereas cores GeoB9508-5 and GeoB7920-2 were retrieved off the West-African coast (inset B). The insets are derived from the ETOPO1 topographical map (Amante and Eakins, 2008).

Chilean core GeoB7139-2 (30°12S/71°59W) was retrieved during the PUCK expedition on RV Sonne in 2001 (Hebbeln and cruise participants, 2001) and is approximately 8 m long. In all three cores, the sediments were dated using stable oxygen isotopes and radiocarbon dates, which show that they date back to 57 kyr BP (GeoB9508-5; Mulitza et al. (2008)), 118 kyr BP (GeoB7920; Tjallingii et al. (2008)) and 65 kyr BP (GeoB7139-2; De Pol-Holz et al. (2007)). Marine sediment cores have provided information necessary to infer Late Quaternary palaeoclimatic variations in South America (e.g., Lamy et al., 1998) and West Africa (e.g., Tjallingii et al., 2008). Climatic conditions have been inferred from grain-size records (e.g., Holz et al., 2004; Stuut and Lamy, 2004; Tjallingii et al., 2008), from geochemical records (e.g., Mulitza et al., 2008; Haslett and Davies, 2006) and various combinations of different sediment properties (e.g., Romero et al., 2008; Kaiser et al., 2008).

Marine sedimentation in West Africa is characterised by aeolian and fluvial input, where the former sediments are considerably coarser grained than the latter (Gac and Kane, 1986; Stuut et al., 2005). Proximal to the source, grain-size variations in the lithogenic fraction of marine cores are therefore considered to reflect the transporting mechanism (Koopmann, 1981; Sarnthein, 1978; Holz et al., 2004; Stuut et al., 2007). In West Africa, sediment with a grain size below $6\mu\text{m}$ is typically assumed to be of fluvial origin, whereas sediment coarser than $6\mu\text{m}$ is assumed to be transported by wind (Koopmann, 1981; Lamy et al., 1998; Holz et al., 2004). If transport distances are relatively large, the initial grain-size of aeolian dust may be reduced by proximal to distal fining (e.g., Weltje and Prins, 2003; Stuut et al., 2005). The same information (i.e. transporting medium, energy and transport distance) may be contained in geochemical signals; Boyle (1983) proposed Al/Ti as a generic grain-size proxy because it is controlled by the concentration of heavy minerals which are transported along with the coarse fraction. Downcore variations in Al/Ti, in turn, were therefore considered a proxy for aridity and/or wind strength (Yarincik et al., 2000).

Apart from information about grain size, geochemical records are considered to contain information about the source area. Sediment that is derived from chemically-weathered terrain typically has high Al and Fe concentrations (Moreno et al., 2006; Mulitza et al., 2008). Elemental ratios with either Al or Fe in the numerator may therefore serve as a proxy for humidity (Sarnthein, 1978; deMenocal et al., 1993). Schneider et al. (1997) related an increase in Al/K towards the tropics to the relative enrichment in kaolinite, at the expense of K-feldspar. Yarincik et al. (2000) used the same ratio as a proxy for the amount of illite relative to kaolinite which is considered to be controlled by the intensity of chemical weathering. Compared to West Africa, the effect of chemical weathering on the geochemical composition of sediments from continental South America is low, due to short transport distances and arid climate conditions (Lamy et al., 2000). Because this implies that the sediments retain their initial composition, geochemical records may serve as a tracer for parent rock lithology (Lamy et al., 1998, 2000; Klump et al., 2000).

3.2.2 Analytical methods

Samples for grain-size analysis were acquired every 5 cm for cores GeoB9508-5 and GeoB7920, and every 2 cm in core GeoB7139-2. The siliciclastic sediment fraction was isolated by dissolving carbonate, organic matter and biogenic opal in HCl, H₂O₂ and NaOH, respectively. The samples were heated with about 300 mg of Na₄P₂O₇ · 10H₂O directly before measuring to avoid the formation of aggregates in the fine-grained fraction. The grain-size distribution was determined with a Coulter LS200 laser particle sizer, which detects grains in the

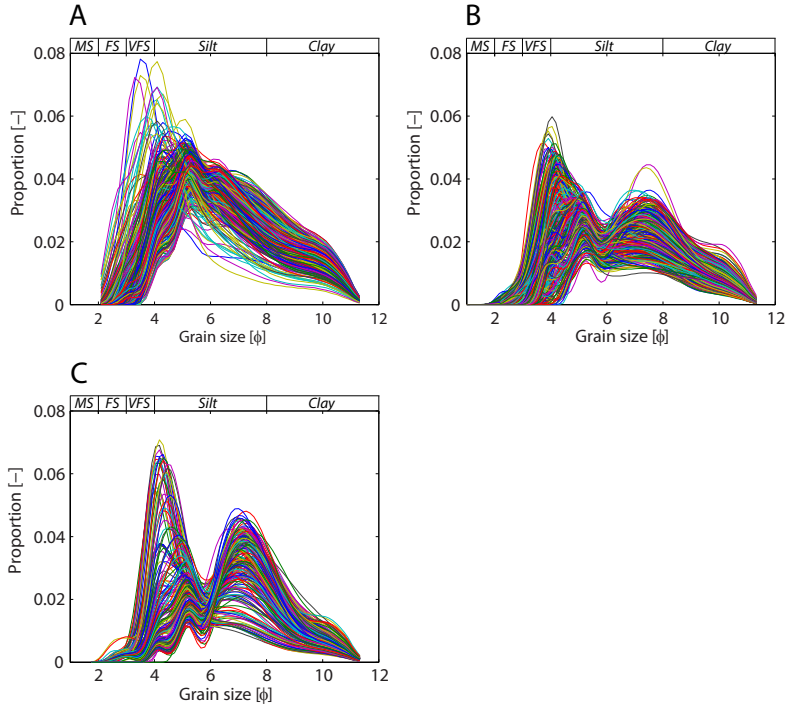


Figure 3.2: Grain-size distributions of cores GeoB7139-2 (A), GeoB7920-2 (B) and GeoB9508-5 (C). Distributions are displayed in phi-scale, and in Udden Wentworth scale (i.e., clay, silt, very fine sand, fine sand, medium sand).

range from 0.4 to 2000 μm (Fig. 3.2).

All three cores have been scanned with an Avaatech core scanner (e.g., Richter et al., 2006) at a 2 cm resolution, a source current of 10kV and a measurement time of 30s. The measured XRF spectra were converted to a record of net element intensities using the WinAxil software (Swerts and Van Espen, 1993). In order to convert the XRF scanner elemental intensities to concentrations, quantitative geochemical analysis of powdered sediment samples was carried out with a portable Spectro Xepos XRF Analyzer (Wien et al., 2005). For core GeoB9508-05, 229 samples were analyzed in total, whereas the geochemical reference data set of GeoB7920-2 and GeoB7139-2 comprises 165 and 20 samples, respectively.

3.2.3 Data preprocessing

Quantitative calibration of the XRF core scanning records was performed using the log-ratio calibration equation (LRCE) (Weltje and Tjallingii, 2008). In this study, the predictive power of the calibration model is empirically quantified by taking the median squared discrepancy between the predicted and the reference geochemical composition at a particular downcore position. We calculate this discrepancy using leave-one-out cross-validation (LOOCV) (Geisser, 1993). As a result, the empirically derived residual variance accounts for all factors influencing the deviation between XRF core scanning records and destruct-

GeoB7139-2	Al	Fe	K	Si	Ti
Al	-	0.002	0.0003	0.0014	0.0019
Fe	0.0784	-	0.0032	0.0009	0.0035
K	0.0322	0.0995	-	0.0016	1.74
Si	0.0660	0.0504	0.0695	-	0.0012
Ti	0.0759	0.1039	2.3280	0.0621	-
GeoB7920-2	Al	Fe	K	Si	Ti
Al	-	0.0024	0.0039	0.0072	0.0039
Fe	0.0792	-	0.0068	0.0157	0.0056
K	0.1030	0.1319	-	0.0049	0.0041
Si	0.1383	0.2034	0.11139	-	0.0154
Ti	0.1025	0.1211	0.1049	0.1990	-
GeoB9508-5	Al	Fe	K	Si	Ti
Al	-	0.0046	0.0032	0.0015	0.0021
Fe	0.1119	-	0.0033	0.0056	0.0033
K	0.0930	0.0954	-	0.0019	0.0032
Si	0.0642	0.1239	0.0716	-	0.0055
Ti	0.0750	0.0949	0.0931	0.1223	-

Table 3.1: The upper diagonal shows the median variances between the observed geochemistry and the geochemistry predicted on the basis of the XRF core scan. The lower diagonal shows the 95% confidence limits corresponding to these residual variances.

ive analysis (e.g. difference in analysed sample volume). A detailed description of these factors is provided by Tjallingii et al. (2007).

Based on the average discrepancy between predicted and reference composition, the LRCE calibration model is derived by selecting the optimal denominator element for each core, which turned out to be Si (GeoB7139-2) and K (GeoB7920-2 and GeoB9508-5). The LOOCV-variances of the different log-ratios are shown in the upper diagonal of Table 3.1. The lower diagonal shows the half-width of 95% confidence limits (cf. Weltje and Tjallingii, 2008). To analyse the relation between grain size and chemical composition, we use samples of which both grain size and geochemical composition were measured. Because the resolution of the grain-size data is different from that of the core scan, a subset of the grain-size and bulk geochemical data meets these requirements, namely 184 (GeoB7139-2), 162 (GeoB7920) and 92 (GeoB9508-5) samples.

To facilitate the application of logratio analysis to the grain size data, zero-valued channels have to be accounted for. In this study, all channels of the grain-size distribution which contain a zero in any of the observations are amalgamated with an adjacent channel that contains only positive values. Generally this leaves more than 80% of the channels unchanged. Given the high level of redundancy in grain-size data (Weltje and Prins, 2003), we assume that this step has a negligible effect on the information content of the grain-size data.

3.3 Modelling approach

3.3.1 Conceptual model

From a theoretical perspective, the information provided by textural and geochemical data may be subdivided into (1) information shared by the two records, (2) information unique to the geochemical record, and (3) information unique to the grain-size record. The partitioning of the geochemical and grain-size variability into these three independent components is illustrated in Figure 3.3A. The significance of these three parts is discussed below.

If a series of sediment samples has been derived from a single source area in which the extent of chemical weathering did not vary significantly over time, most variation in bulk chemical composition may be attributed to fractionation during entrainment, transport, and deposition. Hence, if it were possible to apply a "grain-size correction" to the bulk chemical composition of this hypothetical data set, the residual geochemical composition would not show any significant downcore variation. Such a one-to-one correspondence between grain-size and geochemistry implies that the geochemical record carries no unique information and the bulk chemical composition can be accurately predicted from the grain-size distribution and vice versa.

In practice, this one-to-one correspondence between composition and grain size will be an exception, and a single grain-size distribution may correspond to different geochemical signatures. Various mechanisms for generating size-independent geochemical variability may be envisaged (see Figure 3.3):

1. Chemical weathering: If the degree of chemical weathering varies over time at a single locality, the residual Si/Al ratio will correlate positively with the extent of weathering. Similar trends will be present in other element ratios (Xiong et al., 2010).
2. Hydraulic/aerodynamic sorting: Sediment-forming minerals in the silt to sand fraction span a wide range of densities (from 2.6 to about 5.0 g/cm³) and shapes (spherical to platy). Mineral grains which are susceptible to size-independent fractionation (i.e., depletion and enrichment not mirrored in the bulk grain-size distribution) must have a combination of density and shape which deviates strongly from the bulk, and be present in small proportions. Chemical elements which exclusively reside in such minerals (e.g. Ti and Zr) are therefore excellent tracers of fractionation.
3. Mixing: If different source areas shed sediments with distinct geochemical signatures within a given grain-size range, the residual geochemical signal represents the variation of mixing coefficients of these compositionally-distinct sediment types.
4. Diagenesis: If the degree of diagenetic modification varies across a series of samples with similar grain-size distributions, a size-independent signal will be generated. Examples include dissolution and precipitation at different stratigraphic levels within a core.

For the sake of completeness, one should also consider occurrence of grain-size variations which do not correspond to geochemical variability (Fig. 3.3), which might be the

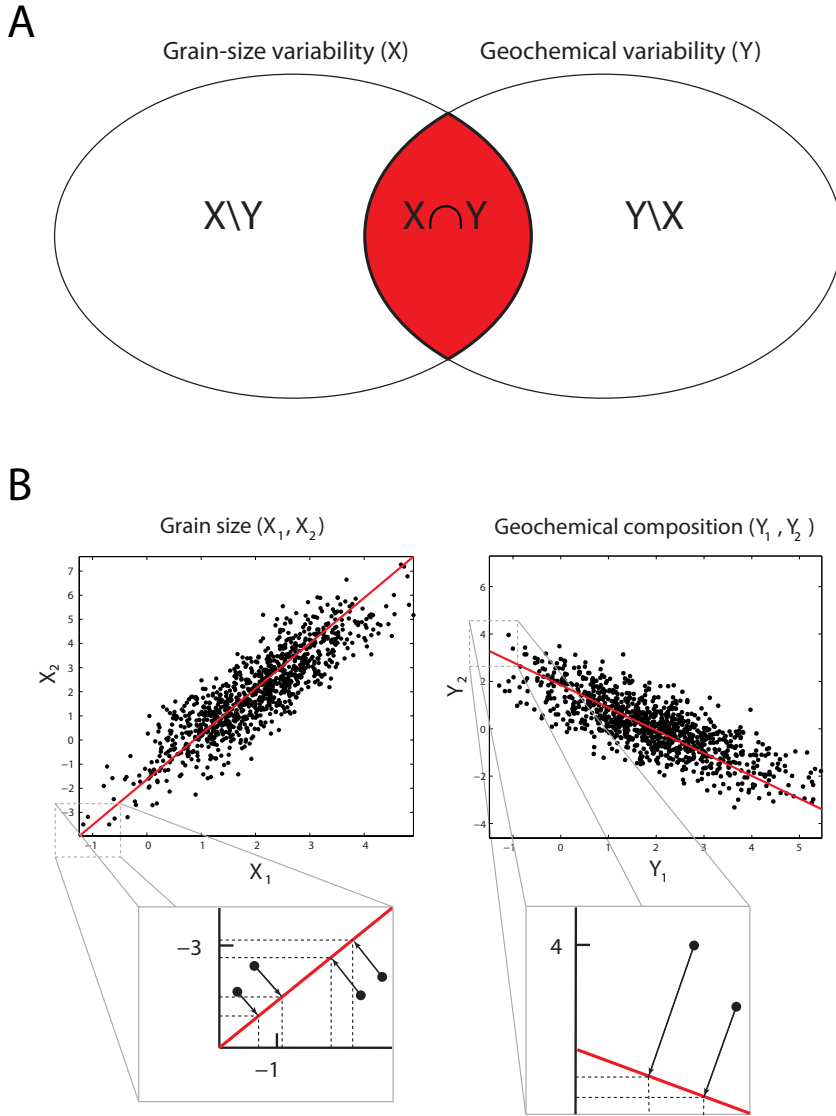


Figure 3.3: In A, we illustrate geochemical and grain-size variabilities in the form of a Venn-diagram. The variability shared by both data sets is indicated by $X \cap Y$, which is highlighted in red. In contrast to this shared component, the variability unique to the geochemical data is the portion that potentially holds relevant signals (e.g., provenance). The unique geochemical variability is indicated by $Y \setminus X$, representing the residual geochemical variability. As shown in the inserts in B, the shared signals are derived by an orthogonal projection of each data set on a set of basis vectors. The direction of these basis vectors is such that they jointly maximise the common covariance in the two data sets. The residual signal may then be obtained by subtracting the shared signals from their corresponding raw data set.

case if rocks have been mechanically ground by glaciers, or mechanically weathered by wind-action. However, even under extremely arid conditions, geochemical variability was found to be accompanied by grain-size variations (Nesbitt and Young, 1996; Solano-Acosta and Dutta, 2005). We therefore consider this a theoretical possibility rather than something which merits further investigation.

3.3.2 Statistical model

If we were able to remove the shared geochemical variability (the red part in Figure 3.3A), the residual record may reveal patterns which cannot be distinguished in the original record. We propose a generic approach to determine the shared variability, based on the assumption that a monotonic relation exists between grain size and bulk chemical composition.

Grain-size distributions and geochemical compositions are compositional in nature (values are non-negative by definition, and their sum over all grain size classes, or all chemical constituents should equal unity), which requires that we model them as log-ratios (Aitchison, 1986). More specifically, we use a centered log-ratio (clr) transformation in order to facilitate application of multivariate methods (Aitchison and Greenacre, 2002). Linear models of log-ratio-transformed variables are compatible with trends of compositional variability of sediments (e.g., von Eynatten et al., 2003; Tolosana-Delgado and von Eynatten, 2008). Exceptions stressed by Tolosana-Delgado and von Eynatten (2010) refer to the 'intrinsic' relation between chemistry and composition, derived from chemical analyses of narrow size fractions, rather than the relation between bulk grain size and bulk chemical composition. Since the latter typically will be more smooth, we consider the logratio-linear approach a suitable parametric model for the relation between bulk grain size and chemical composition.

The model we propose is based on orthogonal projections. A graphical illustration of the model is shown in Figure 3.3B. The model establishes two orthogonal bases (coordinate systems) which maximise the common variability of grain size and geochemical composition. We derive these bases using a method based on Partial Least Squares (PLS) (Martens and Naes, 1989). By projecting both data sets onto their basis vectors, we obtain the scores on each vector. Using the scores and the bases, we can calculate the shared signals. The variability unique to both data sets may be derived by subtracting the shared signals from their corresponding data set.

This process may be repeated until either the variance of the chemical or the grain size data set equals zero. The number of orthogonal components that are removed is a hyper-parameter and may take on any value between zero (the raw geochemical data) and $N-1$, with N being the minimum number of variables (either grain-size classes or chemical elements). We employ a χ^2 -test to determine whether the residual variance of a log-ratio of two elements deviates significantly from the error variance of the geochemical data set for every order k . A more detailed description of the model is given in Appendix A.

3.4 Synthetic examples

To illustrate the working principle of the model, we will apply the model to two synthetic data sets. The numerical experiments are intended to demonstrate the working principle

of the proposed model. We construct two different data sets: data set 1 only has variability induced by grain-size variations, whereas data set 2 reflects grain-size independent compositional variability.

3.4.1 Synthetic example 1

The first step in generating data set 1 is stochastic simulation of grain-size distributions. These distributions are constructed to be log-normal, with a random mean and a standard deviation that is a function of the simulated mean. This function yields a larger standard deviation for fine-grained sediments than for coarse-grained sediments. The fact that these two parameters are correlated implies that the grain size data occupy only one dimension in clr space.

To generate bulk chemical compositions, a logratio-linear relation between composition and grain size is constructed (Fig. 3.4A). Subsequently, this "fingerprint" is transformed back into proportions using the inverse clr-transformation in order to obtain the function giving element concentrations as a function of grain size (Fig. 3.4B). The bulk chemical composition is obtained by multiplying this fingerprint with the grain-size distribution. This provides a unique bulk geochemical composition for any given grain size distribution. Finally, these bulk chemical composition are clr-transformed and some normally-distributed noise is added.

Figure 3.5A shows the simulated grain-size record and Figure 3.5B shows the bulk geochemical record expressed in proportions. The model output is illustrated in Figures 3.5C, D and E. Since the rank of the grain size data is one, and the geochemistry is directly inferred from the grain size, the common signal carried by the first vectors in both solution spaces explains all systematic variation. As a consequence, Figure 3.5D reflects only the superimposed noise centered around the mean bulk chemical composition. We derive the strongest signal in the residual and the shared geochemical variability using Principal Component Analysis (PCA). The principal components (PCs) are obtained by means of a Singular Value Decomposition (SVD) (Press et al., 1994). Figures 3.5E and 3.5F show the scores on PC1 of the shared and the residual signal, respectively. A gray bar represents 95% confidence limits derived from the superimposed noise. Note that for the PC1 scores of the residual geochemical record, the proportion of observations plotting outside of these limits not much larger than 5%. The scores of the shared geochemical signal show significantly more exceedences, indicating a strong correlation between bulk chemical composition and grain size. Using the proposed model, this common trend could be eliminated from the data without any prior knowledge, leaving nothing but noise.

3.4.2 Synthetic example 2

In example 2 we increase the complexity of the data by simulating random mixing of two distinct sediment sources. The size-composition function of the second source is shown in Figures 3.4C and 3.4D. The grain size distributions were simulated according to the procedure outlined in example 1, and contain therefore no information about provenance (Fig. 3.6B). A real-world analogue of this example is a river with two tributaries, draining different parent-rock types. We assume that the proportional contributions of the tributaries has varied over time and we repeatedly sampled the river-mouth sediments. The grain-size distributions of these sediments reflect the hydrodynamic conditions under which the

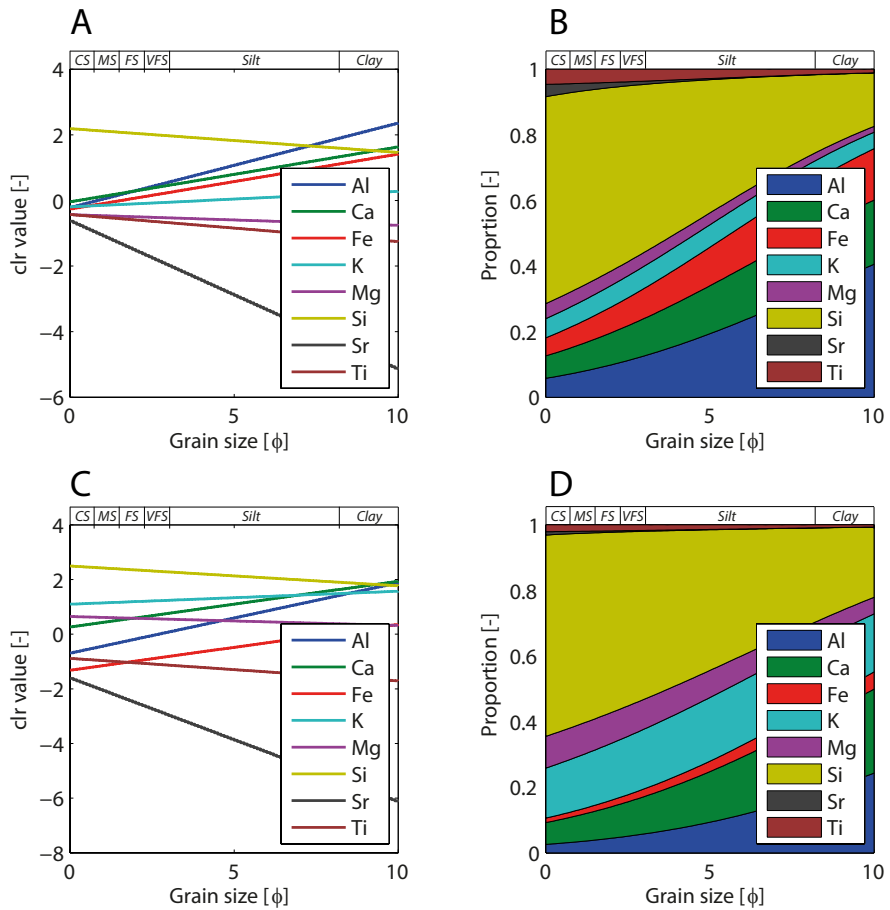


Figure 3.4: Fingerprints of sediment sources 1 and sources 2 (B and D, respectively), which are parameterised by a log-ratio linear function between grain size and composition (A and C).

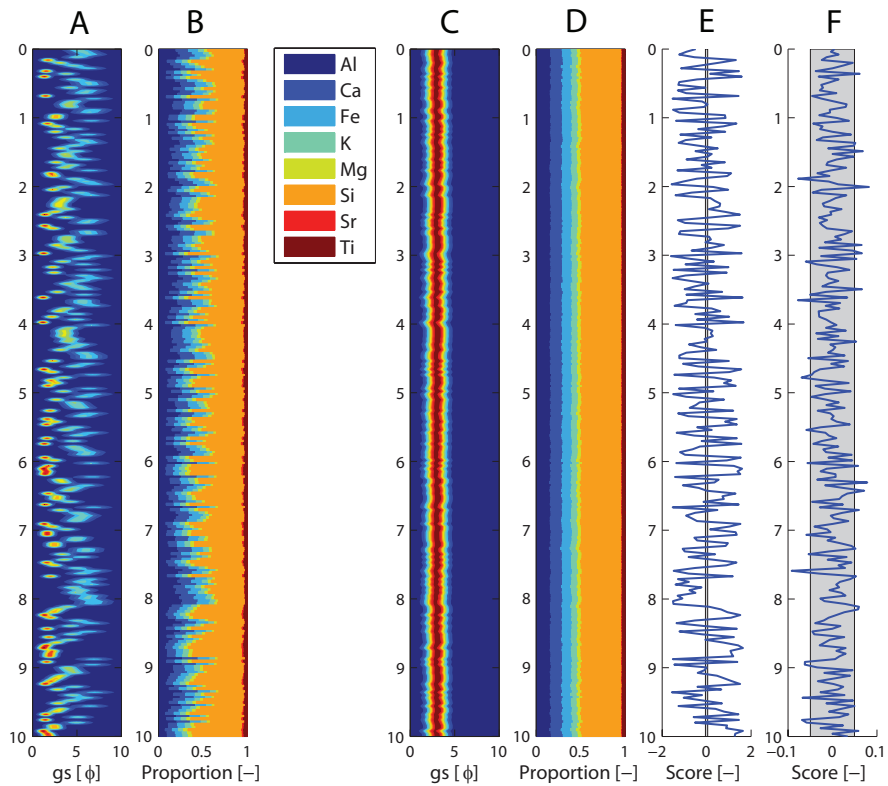


Figure 3.5: Results of synthetic data set 1 ($n=250$), reflecting one provenance. Plots A and B show the grain size distributions and geochemical record, respectively. The residual grain size and geochemical signals are shown in C and D, respectively. The first principal component scores of the shared geochemical signal is shown in E, whereas plot F shows the first principal component scores of the residual geochemical signal. In E and F, a gray bar represents a 95% confidence interval derived from the imposed noise. Note that approximately 5% of the residual PC1 scores (F) are located outside the 95% confidence interval (grey).

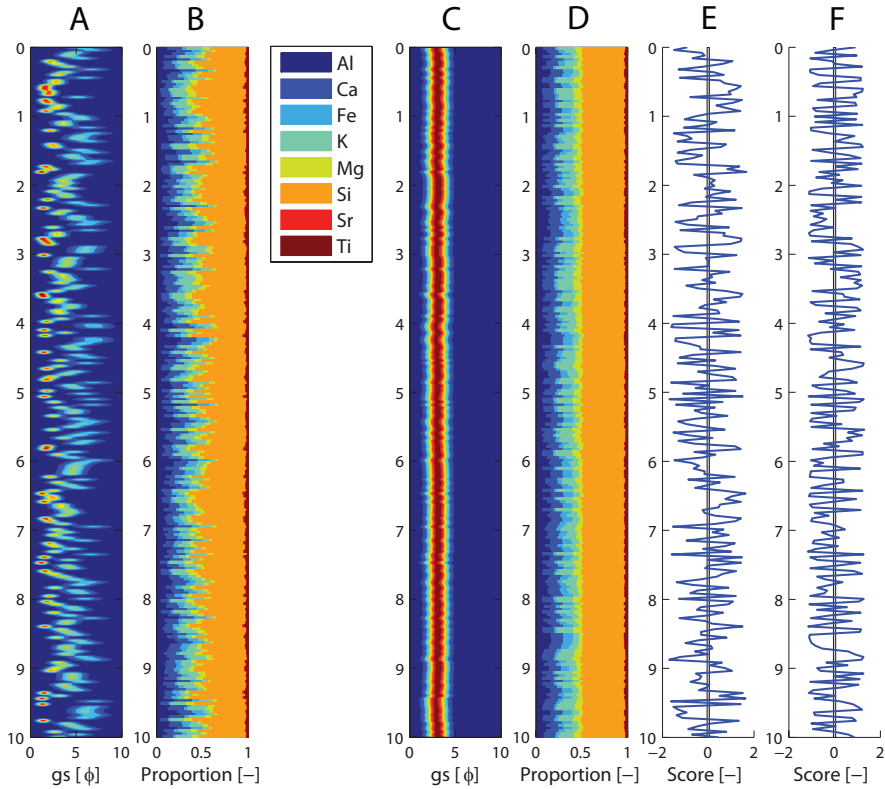


Figure 3.6: Results of synthetic data set 2 ($n=250$), which is constructed by mixing of two different provenances. Plot A shows the grain-size distributions and plot B shows geochemical record. The residual grain size and geochemical record are shown in C and D, respectively. The first principal component of the common size-composition trend is shown in E, whereas plot F shows the first principal component scores of the residuals. In E and F, a gray bar represents a 95% confidence interval derived from the imposed noise. Note that residuals scores on PC1 exceed the 95% confidence interval at almost every stratigraphic level.

sediments were deposited, and thus contain no information regarding provenance.

Data set 2 exhibits a geochemical record which does not appear to be very different from that of data set 1 (Fig. 3.6B). This is not surprising since the fingerprints of the two sources are very similar (see Figure 3.4). As a result, differences induced by variations in the source are overshadowed by variability induced by grain-size variations. In contrast to example 1, however, the residual geochemical record shows significant variability independent from the grain size (Fig. 3.6D). The associated PC1 scores exceed the 95% confidence interval of the superimposed noise (Fig. 3.6F). Hence, we would conclude that there is evidence for grain-size independent variability in the bulk-chemical composition.

If our aim was to infer either grain size or provenance from the raw geochemical data, a straightforward approach would be to analyse the scores on PC1 (i.e. the main direction of variance). Figure 3.7A shows the relation between these scores and the mean grain size, whereas Figure 3.7B shows the relation between these scores and the provenance (i.e. the contribution from source 1). In both cases there is a weak correlation.

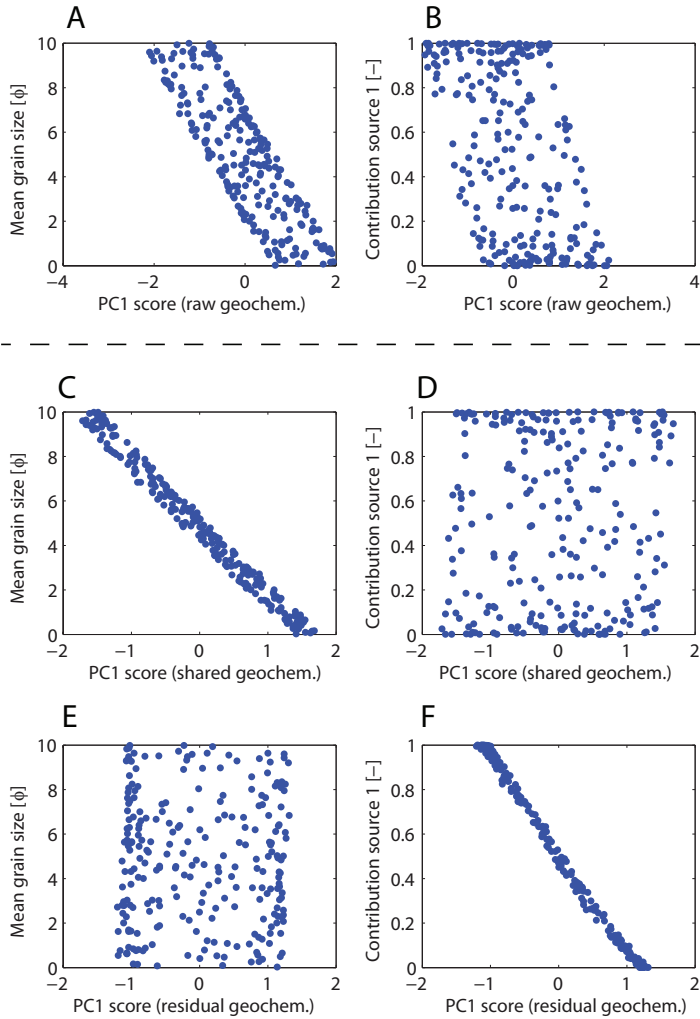


Figure 3.7: Cross plots in plots A and B show the PC1 scores of the raw geochemical data against the mean grain size, and the contribution from source 1, respectively. Note that the raw PC1 scores correlate poorly with both provenance and grain size. Plot C, on the other hand, shows that the correlation between the shared geochemical signal scores and the mean grain size is strong. Similarly, plot F shows that residual scores correlate strongly with the contribution from source 1.

Instead of using the raw data, we may also use the model outcome to derive proxies for both grain size and provenance; PC1 scores of the shared geochemical record represent the grain-size proxy, whereas the residual geochemical record may serve as a provenance proxy. The relation between the grain-size proxy, and the grain size and (known) source contributions, respectively, are shown in Fig. 3.7 (C and D). Fig. 3.7C demonstrates that the grain-size proxy correlates strongly with the mean grain size and does not contain any provenance information. The provenance proxy (Fig. 3.7, E and F), on the other hand, varies independently from the mean grain size, but records the known changes in provenance perfectly.

3.5 Results

We applied the model to the data sets of the three Quaternary marine sediment cores. Before studying the model outcome, however, we analyse the correlation structure between composition and the mean grain size in each of the three cores (see Fig. 3.8). The steepness of the fitted log-ratio linear models of cores GeoB7920-2 and GeoB9508-5 in particular, demonstrate that the geochemical composition strongly depends on the mean grain size. Based on these plots, we also conclude that our data shows no systematic departures from the compositional linear trend, which justifies the use of a log-ratio linear approach.

Figure 3.8D shows Pearson's correlation coefficients between clr transformed elements concentrations and mean grain size. In addition, the correlation coefficients between the mean grain size and the PC1 scores of the clr transformed element concentrations are shown. The high correlation (0.9) between PC1 scores and mean grain size in core GeoB9508-5 demonstrates the correspondence between geochemical and grain-size data, respectively. In core GeoB7139-2, on the other hand, the grain-size control on the geochemical variability is relatively low. In this core, only Ti shows a fairly strong correlation with the mean grain size.

After applying the proposed model to the three cores, significance tests on the residual records ($\alpha = 5\%$) allow us to identify additive log-ratio pairs which show significant residual variability in bulk composition (see Table 3.2). The strongest residual signal in the three cores are shown in Figures 3.9 and 3.10. In contrast to the synthetic example, the residuals are plotted without their mean added and reflect the difference between modelled and observed signals. An exception is made for $k = 0$: since these patterns are by definition only a centered version of the input data, the raw data are shown instead.

The results demonstrate that the variance of the residuals is largely controlled by the number of removed components (i.e., the order of the residuals). In addition, the cores show a marked difference, in the sense that the variance of the third-order residuals ranges from low in core GeoB7920, to high in core GeoB9508-5. Note that the limited number of chemical elements ($D=5$) used in our study limits the maximum number of removed shared signals to four.

3.6 Discussion

3.6.1 Correlation between bulk properties

The correlation coefficients in Figure 3.8D show that in cores GeoB7920-2 and GeoB9508-5, Al and Fe correlate positively with the mean grain size whereas Si and K correlate negat-

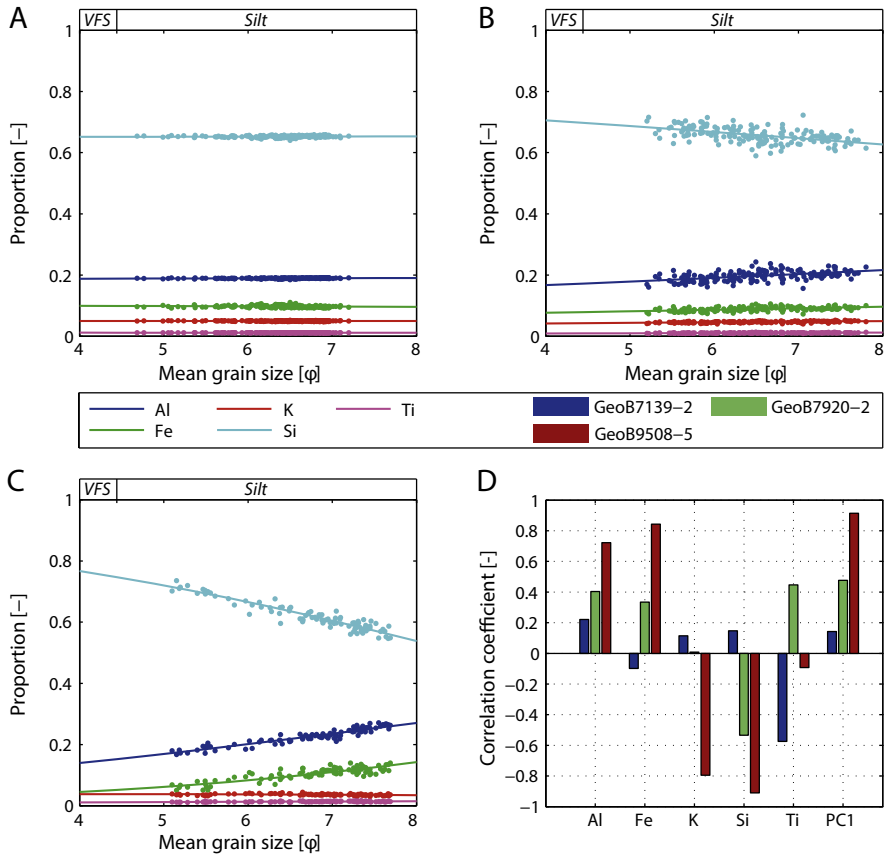


Figure 3.8: The bulk chemical composition as function of the mean grain size in cores GeoB7139-2 (A), GeoB7920-2 (B) and GeoB9508-5 (C). The line passing through the data reflects the least-squares solution to a logratio-linear model. The correlation between the different clr-transformed chemical components and the mean grain size is shown in D. In addition, the correlation coefficients between the mean grain size and the first principal component (PC1) of the raw clr-transformed geochemical data are shown.

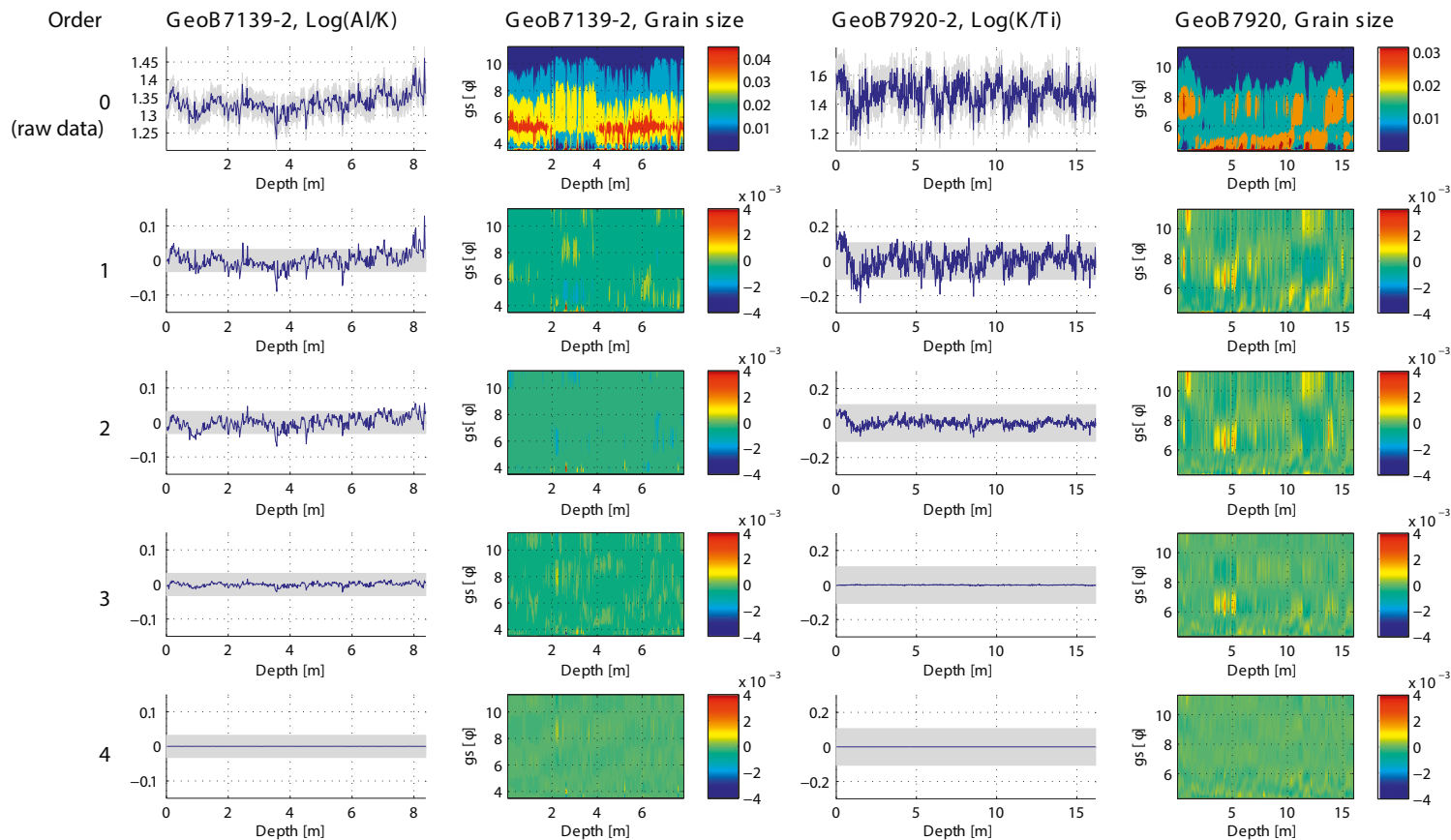


Figure 3.9: The strongest downcore residual signals in terms of signal-to-noise ratio in cores GeoB7139-2 and GeoB7920-2, with the model order ranging between zero (i.e., the raw data) and four. The residual geochemical signals are presented as variations around the mean, with a gray bar indicating a 95% confidence limit. The residual grain-size distributions are presented as the difference between the raw and modeled clr-transformed grain size distributions.

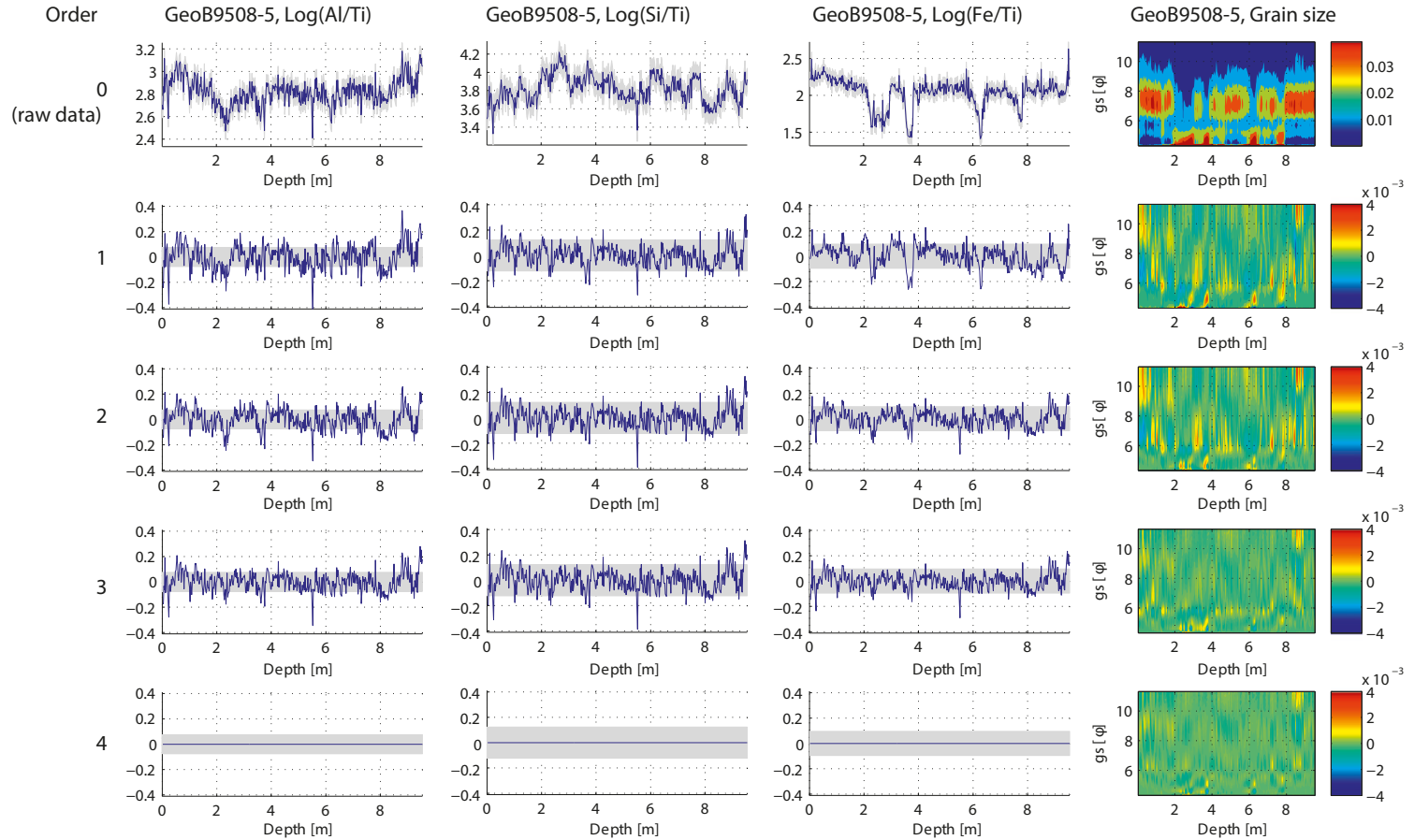


Figure 3.10: The three strongest downcore residual signals in terms of signal-to-noise ratio in core GeoB9508-5, with the model order ranging between zero (i.e., the raw data) and four. The residual geochemical signals are presented as variations around the mean, with a gray bar indicating a 95% confidence limit. The residual grain size distributions are presented as the difference between the raw and modeled clr-transformed grain size distributions.

Core	Order	Significant residuals
GeoB7139-2	1	Al/K
GeoB7139-2	2	Al/K
GeoB7139-2	3	none
GeoB7139-2	4	none
GeoB7139-2	5	none
GeoB7920-2	1	Al/Fe
GeoB7920-2	2	Al/Fe
GeoB7920-2	3	none
GeoB7920-2	4	none
GeoB7920-2	5	none
GeoB9508-5	1	Ti/*,Al/*, K/Si
GeoB9508-5	2	Ti/*, K/Si
GeoB9508-5	3	Al/Ti, Fe/Ti, Si/Ti, K/Si
GeoB9508-5	4	none
GeoB9508-5	5	none

Table 3.2: The significant residuals of the different cores ($\alpha=5\%$). A "*" indicates that all elements (except the numerator) exhibit significant residual variability.

ively with the mean grain size. These trends are in line with typical weathering trends, leading to relative enrichment of clay-minerals (reflected by Al and Fe) in the fine fraction and enrichment of quartz (reflected by Si) and feldspar (reflected by K) in the coarse fraction.

In core GeoB7139-2 the correlation structure is different, given that Si shows a weak positive correlation with the mean grain size in phi units. Moreover, Ti appears to be the only element closely linked to grain size, reflected by a negative correlation with the mean grain size in phi units. The weak grain-size control on the geochemical composition in this core may be due to continental climate conditions. Core GeoB7139-2 is situated close to the Atacama desert, which is characterised by extreme aridity (Stuut and Lamy, 2004; Kaiser et al., 2008). The absence of chemical weathering under arid conditions implies that the size-composition trend primarily reflects mechanical weathering. The weak but positive correlation between Si and the mean grain size may be attributable to additional input of biogenic silica, which obviously disturbs the size-composition trend.

Based on these results we may select a proxy for grain size by choosing a ratio of an element showing a high and positive correlation coefficient, and an element showing a negative correlation coefficient with the mean grain size. This yields Al/Ti (GeoB7139-2), Ti/Si (GeoB7920-2) and Fe/Si (GeoB9508-5). Differences among these proxies illustrate the empirical nature of grain-size proxies.

3.6.2 Residuals analysis

Significance tests yielded that, depending on the order, the cores reflect residual patterns of Ti (GeoB9508-5), Al/Fe (GeoB7920) and Al/K (GeoB7139-2). In contrast to the synthetic examples, however, we do not know the number of dimensions of the grain size data. It is obvious that the more shared signals we remove from the data, the smaller the variance of

the residual records. The limitations of our data are reached at $D - 1$, which equals four in these data set (we measured only five elements).

Following the principle of parsimony, we should stop subtracting signals when the shared grain size variability is within acceptable limits of the measured grain variability, or when the removed grain-size signals do not correlate with the removed bulk chemical signals. This means that quantitative control on the model outcome does not only require uncertainties of the geochemical data, but also of the grain size data. Unfortunately, we can only infer the uncertainties associated with the geochemical data, meaning that we are not able to identify the 'true' order of the residuals. However, we may interpret the results for the analysed cores relative to each other.

The downcore records in Figures 3.9 and 3.10 suggest that from these three cores, core GeoB9508-5 contains the strongest geochemical residual signals. The same is reflected by the large number of logratios that are identified as significant (Table 3.2). The raw Al/Ti, Fe/Ti, Si/Ti, as well as their first and second order residuals (Fig. 3.10) are positively correlated, suggesting that the significance is attributable to relative changes in Ti. Apparently, sediments with a similar grain-size distribution have different Ti concentrations. Variable Ti concentrations independent from the grain size have also been observed in present-day dust samples collected off northwest Africa; dust samples with the same mean grain size showed different Al/Ti values ($\text{Log}(\text{Al}/\text{Ti}) \sim [2.65, 2.9]$; Stuut et al. (2005)). Moreover, the variability of these Al/Ti values is in the same range as the residuals observed in this study.

Residual grain size signals in cores GeoB7139-2 and GeoB7920-2 (Fig. 3.9) suggest that at least three common patterns should be removed to obtain a grain-size invariant geochemical record. At this point, there are no significant residual geochemical signals present in the data (Table 3.1). Hence, we conclude that these cores contain no significant residual geochemical variability.

3.6.3 Palaeo-climatological interpretation

In order to relate the residuals unambiguously to the proposed size-independent mechanisms, mineralogical information is indispensable. In core GeoB9508-5, K most likely, but not exclusively, resides in K-feldspar, whereas Si resides in detrital quartz and possibly in opal. Because (i) we cannot unambiguously associate these elements to minerals, and (ii) the magnitude of their associated residual signal (i.e. K/Si) is approximately 50% smaller than that of the Ti-residuals, we do not attempt to interpret this signal.

Compared to the other elements, Ti can be more confidently associated with mineral phases: it is considered to be of exclusively terrigenous origin (Murray and Leinen, 1996) and mostly resides in heavy minerals such as sphene, rutile and anatase (Spears and Kanaris-Sotiriou, 1976). These minerals have a high density relative to quartz, feldspars, and clays, which make up the bulk of siliciclastic sediments. Although Ti-rich heavy minerals are considered essentially inert components (Young and Nesbitt, 1998), diagenetic alterations have been described under highly specific conditions (Goldberg and Arrhenius, 1958; Pe-Piper et al., 2011). However, given the fact that the observed Al/Ti are approximately equal to Al/Ti values found in other marine sediment cores located in the Atlantic (Zabel et al., 1999), we consider diagenetic modification a highly unlikely cause. This implies that the residuals may be explained by any of the postulated grain-size independent processes that take place before deposition, such as weathering, density/shape sorting and mixing.

In the area of core GeoB9508-5, contrasting patterns of deposition have been linked to

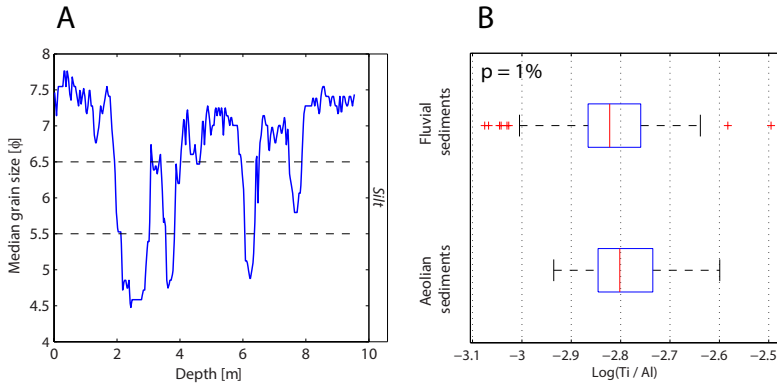


Figure 3.11: As indicated in A, we partitioned the data into fluvial, aeolian and mixed provenance, based on the median grain size. For both the fluvial and aeolian sediment, a boxplot associated with the residual Ti/Al is shown in B.

climate changes (Mulitza et al., 2008). Dry and cold periods were characterised by strong winds and low fluvial sediment input, whereas the opposite holds for relatively warm and humid intervals (e.g. Sarnthein, 1978; deMenocal et al., 1993). Hence, we are dealing with a situation in which the grain size of the sediments, and therefore their transport mechanisms, are likely to correlate with provenance. We make use of this knowledge by separating the residual geochemical data into aeolian-dominated, fluvial-dominated, and mixed sediments, based on the median grain size (P50 of the cumulative grain size distributions) with cutoffs set at 5.5ϕ and 6.5ϕ (Fig. 3.11A). The relation between transport mechanism and residual Ti/Al is shown in Figure 3.11B. The probability that the residual Ti/Al values of the two have identical distributions was estimated using a non-parametric Wilcoxon test (Hollander and Wolfe, 1973). Based on the p-value of 1% and a significance level of 5%, we conclude that the third-order Ti/Al residuals correlate with transport mechanism (and therefore also the Al/Ti residuals, because logratios are symmetrical).

If the sediment deposited in the sink has been exclusively derived from one source, and the conditions under which the sediment was transported were constant in time, the residual Al/Ti signal reflects varying climatic conditions in the source area(s). An alternative scenario is that the Al/Ti residuals reflect the different areas that have served as a source, rather than varying conditions within the source area(s). In both cases, the underlying assumption is that the sediment is transport-invariant, i.e., that it has been transported under the same hydro- and aerodynamic conditions (Weltje, 2004). Another scenario is that the residual Al/Ti variability is induced by mixing of fluvial and aeolian sediment and/or sorting processes during either river- or wind-transport. This scenario is characterised by the assumption of "provenance-invariance".

It is evident that, without additional constraints, there is no unique solution to this problem. Even when adopting the (strong) assumption that the source area and the climatic conditions were constant over time, we cannot ascribe the residuals to either selective transport or mixing. Preferential entrainment of Ti-bearing minerals in the windblown sediment due to increased wind strength (Sarnthein, 1978; Rea, 1994) will have the same effect on Ti residuals as mixing of Ti-rich windblown sediment, and Ti-poor riverine sedi-

ment of constant geochemical composition. Moreover, Stuut et al. (2005) showed that even on a short time-scale (days), both source area and pathway length (which controls selective transport) substantially vary, which means that neither the provenance-invariance nor the transport-invariance assumption is defensible. More data is necessary to unambiguously interpret the residuals in terms of the proposed grain-size independent mechanisms.

3.6.4 General discussion

In the analysed cores, the elements showing the strongest correlation with the mean grain size were Ti (GeoB7139-2) and Si (GeoB7920-2 and GeoB9508-5). In addition, Ti was designated as the main residual component in GeoB9508-5. This leads to the conclusion that geochemistry-based proxies for mean grain size depend on the specific setting, which automatically implies that the same is true for geochemistry-based proxies considered to be associated with other processes. We showed that, using the model and a set of overlapping grain-size and geochemical observations, these proxies can be identified in a robust way. However, whether or not signals associated with the size-independent processes can be inferred from the data depends on the magnitude of this signal, relative to the uncertainty of both data sets. Hence, more significant signals may appear if the uncertainty of the bulk chemical data will be reduced. Furthermore, applying the model to a more comprehensive set of chemical element concentrations is more likely to reveal informative residual patterns. As for the interpretation of bulk geochemical records in general, interpretation of the residuals is difficult because they may reflect numerous different processes. In addition, chemical elements cannot always be uniquely associated with a single mineral. We can thus not provide a 'recipe' for the interpretation of the residuals. However, by correcting the data for grain size, at least one hypothesis can be tested, i.e., that the chemical composition only reflects grain-size variations.

Reduced-rank approximations to compositional and spectroscopic data in log-ratio space ¹

4.1 Introduction

The ability to properly account for uncertainty is essential to optimal multivariate modeling (Leger et al., 2005; Reis and Saraiva, 2006a,b; Wentzell and Lohnes, 1999). For spectroscopic data, one of the most fundamental sources of uncertainty is the spread caused by the stochastic nature of a counting process. However, counting processes are by no means limited to spectroscopy: also data acquisition in the fields of geology, social sciences and ecology frequently involves counting different kinds of discrete objects falling into an exhaustive set of mutually exclusive classes. Examples of such data sets are point counts, opinion polls and species-richness data.

In addition to these obvious examples of count data, there are also data types whose enumerative nature is not evident because:

1. The data are not intrinsically categorical but represent discretised continuous variables;
2. The counting process is not directly reflected in the output of a measurement device or protocol.

We use a benchtop XRF analyser to illustrate these two cases. A benchtop XRF analyser will give element concentrations as output. However, the underlying data are fluorescence spectra reflecting the number of counted photons per wavelength or energy bin and are therefore an example of case 1. The element concentrations are derived from the area under the corresponding peak in the discretised spectrum and therefore associated with a finite number of counts. Hence, the compositional data provided by the instrument are

¹Published as: Bloemsmma, M.R. and G.J. Weltje (2015). Reduced-rank approximations to spectroscopic and compositional data: A universal framework based on log-ratios and counting statistics. *Chemometrics and Intelligent Laboratory Systems*, Vol. 142, pp. 206-218.

an example of case 2. Finally, there are also types of data which are technically not count data, although the way they have been acquired is very similar and therefore the errors are likely to exhibit similar properties. For example, a particle-size distribution acquired by sieving is derived from masses in different size classes, and is thus not acquired by means of counting. However, the data may be converted to numbers of particles by adopting a particle density and the data are therefore likely to possess statistical properties similar to those of "true" count data.

We distinguish two categories of count data applications: the relative and absolute information content are either considered meaningful, or the count data are considered to carry relative information only. For instance, if we are counting the numbers and types of birds spotted during one hour at a single locality, the sample size does matter which makes it an example of the first category. However, the size of the sample is of no importance when analysing the silica content of two differently-sized rocks: it is much more meaningful to compare their silica content in weight percent. All compositional data therefore fall in the second category. Spectroscopic applications may fall into both categories. The sample size (i.e. the total number of counts) of spectroscopic data is largely determined by instrument settings and measurement geometry (e.g. counting time, distance between sample and sensor). In case the data are acquired under standardised conditions, the scale will be a meaningful quantity and the data fall into the first category. However, in case instrument settings or measurement geometry vary within one data set, one should consider focusing only on the relative information content (Weltje and Tjallingii, 2008).

This contribution is dedicated to the second category of count data applications. Separation of relative from absolute variability can be achieved by making the following decomposition:

$$\mathbf{X} = \text{diag}(\mathbf{t})\mathbf{P} \quad (4.1)$$

where \mathbf{P} is a matrix with count proportions on its rows and \mathbf{t} is a vector containing the number of counted objects in each row. An important property of \mathbf{P} is that its elements are discrete because \mathbf{X} and \mathbf{t} are discrete. However, as the sample size goes to infinity, \mathbf{P} becomes a continuous variable. Another important property of \mathbf{P} is that the variables are not independent, which stems from the set of constraints that apply to each row \mathbf{p} spanning D variables (Chayes, 1960):

$$\mathbf{p} : \sum_{i=1}^D p_i = 1, \quad 0 \leq p_i \leq 1 \quad (4.2)$$

Data characterised by these constraints are referred to as compositional data and the sample space in which these data reside is referred to as the simplex (Aitchison, 1986). It is well-known that the simplex is problematic for multivariate analysis and modelling: the variables have upper and lower limits and the sum-constraint imposes dependency (they have effective rank $D - 1$).

Instead of modelling the data in this problematic simplex, the commonly accepted procedure is to model the data in an alternative space (Aitchison, 1986). Several of these spaces have been proposed, but they all rely on representing variables as logarithms of ratios, i.e., log-ratios. The advantages of analysing and modelling log-ratio transformed count data are that (1) log-ratios are unconstrained, (2) they are insensitive to multiplicative operations

and therefore to scale variations, (3) the log-ratio space is a metric space (Pawlowsky-Glahn and Egozcue, 2002). Because log-ratio transformations are invertible, the modelling result may be transformed back to the original units after multivariate analysis. Furthermore, modelling of log-ratios is numerically convenient because many non-linear trends observed in nature tend to become linearised upon log-ratio transformation (Aitchison, 1999; Weltje, 2012; von Eynatten et al., 2003).

Multivariate analysis of compositional data is frequently performed in the so-called centred log-ratio (clr) space (Aitchison and Greenacre, 2002). The transformation which maps data from the simplex to the clr space is given by:

$$\mathbf{z} = \text{clr}(\mathbf{x}) \rightarrow z_j = \log(x_j) - \frac{1}{D} \sum_{i=1}^D \log(x_i) \quad (4.3)$$

for $j = 1, 2, \dots, D$. The inverse transformation reads:

$$\mathbf{x} = \text{clr}^{-1}(\mathbf{z}) = \mathcal{C} [\exp(x_1), \exp(x_2), \dots, \exp(x_D)] \quad (4.4)$$

where $\mathcal{C}[\dots]$ is the so-called closure operator, i.e., the operator which divides all vector elements by their sum to impose the sum constraint. Equation 4.3 shows that the clr-transformation is a row-wise centring operation in the log-space. The geometric mean is used because it is the BLU-estimator of the center of a composition with respect to the geometry of the simplex (Pawlowsky-Glahn and Egozcue, 2002).

Given that models governing compositional data in log-ratio space are generally based upon minimising a least-squares criterion, it is implicitly assumed that the errors in this space are independent and identically distributed (*iid*). However, we will demonstrate that, under the assumption that counting errors make up the dominant source of uncertainty, errors are not *iid*. The general aim of this paper is to show that accommodating for counting errors is essential to improving multivariate modelling of count data in general, and spectroscopic and compositional data in particular. More specifically, we investigate methods for making Reduced Rank Approximations (RRAs) to count data in clr space. The methods we investigate range from straightforward RRA based on least-squares principles to RRA by means of a maximum likelihood method. Since many different types of data originate from counting or an analogous process, this contribution is of potential importance to a wide audience. Furthermore, the results are directly applicable to other methods which rely on subspace estimation such as Principal Component Analysis (PCA).

4.2 Statistical basis

4.2.1 Notation

We use the notation conventions common to chemometrics and compositional data analysis. Hence, matrices and vectors are written in bold-face upper- and lower-case letters, respectively. In addition, the convention is used that a set of count data is represented by the matrix \mathbf{X} (N rows and D columns). The log-transformed data matrix is given by \mathbf{Y} and the clr-transformed data matrix is given by \mathbf{Z} . Furthermore, we use the forward and

inverse clr-transformation operators which act upon matrices so that the following holds:

$$\begin{aligned}\mathbf{Z} &= \text{matclr}(\mathbf{X}) \\ \mathbf{X} &= \text{diag}(\mathbf{t}) \text{matclr}^{-1}(\mathbf{Z})\end{aligned}\quad (4.5)$$

where \mathbf{t} contains the number of counted objects.

The index $j = 1, 2, \dots, D$ is used for columns and the index $k = 1, 2, \dots, N$ is used for rows. If two indices for the D variables are required, i is also used as a row index. Upper-case letters X and T refer to the counts and sample size as random variables, respectively. Hence, T is the random variable defined as:

$$T = \sum_{j=1}^D X_j \quad (4.6)$$

Short-hand notations for the theoretical standard deviation and variance of X are δX and $\delta^2 X$, respectively.

We also use the notation that a symmetric matrix is summarized by showing the explicit expressions for the diagonal and off-diagonal entries. For example, the D -part identity matrix \mathbf{I}_D reads:

$$\begin{aligned}\mathbf{I}_D(j, j) &= 1, \quad j \in \{1 : D\} \\ \mathbf{I}_D(i, j) &= 0, \quad i, j \in \{1 : D\} \mid i \neq j\end{aligned}\quad (4.7)$$

However, for the sake of simplicity we will not specify the index ranges and simply write $\mathbf{I}_D(j, j) = 1$ and $\mathbf{I}_D(i, j) = 0$ for the diagonal and off-diagonal entries, respectively.

Greek letters represent distribution parameters, hence they are not observed (e.g. p_{kj} is the estimate of the true probability π_{kj} of finding an object of class j in specimen k). We use $\text{var}(\mathbf{x})$ as a short-hand notation for the sample variance over the values in the vector \mathbf{x} . Likewise, the total variance of a data matrix is written as $\text{var}(\mathbf{X})$ which essentially means $\text{var}(\text{vec}(\mathbf{X}))$. Finally, ' \circ ' denotes the Hadamard product (i.e. element-wise matrix multiplication) and $\text{diag}()$ reflects the operator which puts the elements of the argument on the diagonal of a matrix, i.e.:

$$\text{diag}(\mathbf{1}_D) = \mathbf{I}_D \quad (4.8)$$

where $\mathbf{1}_D$ is the D -part unit vector.

4.2.2 Fixed-size count data

First, we consider count data which have a fixed sample size. For these data, acquisition stops as soon as the number of counted objects reaches some user-specified value τ , as in point counting. Given that the objects are randomly drawn from the population, the count data are multinomial distributed and therefore referred to as Multinomial Distributed Data (MDD).

The distribution of a multinomial-distributed variable X_j is conditional on τ and not independent from the other variables. Therefore, we cannot express the distribution of one variable without considering all others and we define MDD as follows:

$$[X_1, X_2, \dots, X_D] \sim \text{Mult}(\tau, \boldsymbol{\pi}) \quad (4.9)$$

where $\boldsymbol{\pi}$ is the vector with probabilities π_j , i.e. the probability of drawing an object of class j from the population. For MDD, the sample size T does not vary so that $T = \tau$.

For the standard deviations associated with a multinomial distributed random vector the following holds:

$$\mathbf{d}_{mdd} = [\tau\pi_1(1 - \pi_1), \tau\pi_2(1 - \pi_2), \dots, \tau\pi_D(1 - \pi_D)]^{1/2} \quad (4.10)$$

For the diagonal and off-diagonal elements of the correlation matrix \mathbf{R} the following holds:

$$\begin{aligned} \mathbf{R}_{mdd}(i, i) &= 1 \\ \mathbf{R}_{mdd}(i, j) &= -\sqrt{\pi_i\pi_j(1 - \pi_i)^{-1}(1 - \pi_j)^{-1}} \end{aligned} \quad (4.11)$$

The Error Covariance Matrix (ECM) $\boldsymbol{\Gamma}_{mdd}$ of multinomial data in the count space is given by:

$$\boldsymbol{\Gamma}_{mdd} = \mathbf{R}_{mdd} \circ (\mathbf{d}'_{mdd}\mathbf{d}_{mdd}) = \tau \begin{bmatrix} \pi_1(1 - \pi_1) & -\pi_1\pi_2 & \dots & -\pi_1\pi_D \\ -\pi_2\pi_1 & \pi_2(1 - \pi_2) & \dots & -\pi_2\pi_D \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_D\pi_1 & -\pi_D\pi_2 & \dots & \pi_D(1 - \pi_D) \end{bmatrix} \quad (4.12)$$

4.2.3 Fixed-time count data

Next, we will treat count data whose acquisition protocol is characterised by setting a fixed time interval of length Δt in which objects are counted. For these "fixed-time" data, the probability that an object is counted is random in time and independent from all other objects. Hence, we rather refer to "events" than objects and the samples are not characterised in terms of count probabilities π (as for MDD) but in terms of count rates ξ (in counts per unit of time). Peak areas in spectroscopic data are an example of this type of data.

If the events are random in time, the expected number of observed counts within time interval Δt is Poisson distributed and we refer to these data as Poisson Distributed Data (PDD). The distribution of such a variable is given by:

$$X_j \sim Poiss(\Delta t \xi_j) \quad (4.13)$$

In contrast to MDD, the sample size of PDD is not constant:

$$T \sim Poiss\left(\sum_{j=1}^D \Delta t \xi_j\right) = Poiss(\tau) \quad (4.14)$$

and only controlled indirectly by Δt .

Because the data are uncorrelated, the following holds for the error correlation matrix:

$$\mathbf{R}_{pdd} = \mathbf{I}_D \quad (4.15)$$

The Poisson distribution also has the following property:

$$E(X_j) = \delta^2 X_j = \Delta t \xi_j \quad (4.16)$$

and therefore the following holds for the vector with standard deviations:

$$\mathbf{d}_{pdd} = [\Delta t \xi_1, \dots, \Delta t \xi_D]^{1/2} \quad (4.17)$$

Hence, the ECM of PDD in the count space is given by:

$$\Gamma_{pdd} = \mathbf{R}_{pdd} \circ (\mathbf{d}'_{pdd} \mathbf{d}_{pdd}) = \Delta t \text{diag}([\xi_1, \dots, \xi_D]) \quad (4.18)$$

4.2.4 CLR space error model

Our next goal is to propagate the errors of MDD and PDD to the clr space. First, an error model in the log-space is derived. Given that $Y = \log(X)$, applying the rules of error propagation yields that $\delta Y = \frac{dY}{dX} \delta X$. For the derivative term $\frac{dy}{dx}$ holds that $\frac{dy}{dx} = \frac{d \ln(X)}{dx} = X^{-1}$ so that $\delta Y = \frac{\delta X}{X}$. The relative error in X translating to an absolute error in Y is a fundamental property of the logarithmic transformation.

For MDD it holds that $\delta X = \sqrt{\tau\pi(1-\pi)}$ (Eq. 4.10). For the associated standard deviation in the log-space δY we can write:

$$\delta Y = \frac{\delta X}{X} \simeq \frac{\sqrt{\tau\pi(1-\pi)}}{\tau\pi} = \sqrt{\frac{1-\pi}{\tau\pi}} \quad (4.19)$$

where we used the substitution $X = E(X) = \tau\pi$. For one sample originating from a D -component multinomial distribution, these D standard deviations may be stored in a vector \mathbf{q}_{mdd} :

$$\mathbf{q}_{mdd} = \left[\frac{1-\pi_1}{\tau\pi_1}, \frac{1-\pi_2}{\tau\pi_2}, \dots, \frac{1-\pi_D}{\tau\pi_D} \right]^{1/2} \quad (4.20)$$

For PDD we can write:

$$\delta Y = \frac{\delta X}{X} = \frac{\sqrt{\Delta t \xi}}{\Delta t \xi} = (\Delta t \xi)^{-1/2} \quad (4.21)$$

where we used the substitution $X = E(X) = \Delta t \xi$.

The vector containing the standard deviations is given by:

$$\mathbf{q}_{pdd} = [\Delta t \xi_1, \Delta t \xi_2, \dots, \Delta t \xi_D]^{-1/2} \quad (4.22)$$

To derive the ECM in the clr space, we specify the clr transformation in terms of a linear transformation matrix \mathbf{A}_D acting upon the log-transformed data matrix \mathbf{Y} so that:

$$\mathbf{Z} = \mathbf{Y} \mathbf{A}_D = \begin{bmatrix} y_{11} & \cdots & y_{1D} \\ \vdots & \ddots & \vdots \\ y_{N1} & \cdots & y_{ND} \end{bmatrix} \begin{bmatrix} (1-D^{-1}) & -D^{-1} & \cdots & -D^{-1} \\ -D^{-1} & (1-D^{-1}) & \cdots & -D^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ -D^{-1} & -D^{-1} & \cdots & (1-D^{-1}) \end{bmatrix} \quad (4.23)$$

The ECM then follows from error propagation:

$$\Psi = \mathbf{A}_D (\mathbf{R} \circ (\mathbf{q}' \mathbf{q})) \mathbf{A}'_D \quad (4.24)$$

where \mathbf{q} and \mathbf{R} are the standard deviation vector and error correlation matrix associated with MDD or PDD, respectively. For clarity we post-multiply by the transpose of \mathbf{A}_D , however, note that $\mathbf{A}'_D = \mathbf{A}_D$.

4.2.5 ECM in the clr space

To study the differences among the error properties of PDD and MDD, we construct a random composition and analyse the corresponding ECMs. Within the multinomial framework this composition represents the count probabilities π_j and within the Poissonian framework this composition represents the count rates ξ_j . The composition vector consists of 5 components (i.e., $D = 5$) and reads $[0.1, 0.15, 0.3, 0.4, 0.05]$. The total number of counted objects τ and the time interval Δt are set at 500 for the MDD and PDD, respectively. The ECMs are shown in Figure 4.1.

Figures 4.1, A and D, show that in the count space the ECM of PDD is diagonal whereas the ECM of MDD has nonzero off-diagonals. This structure is, however, not maintained upon log-transformation: the off-diagonals of MDD become equal (Fig. 4.1E). Moreover, upon clr-transformation, the ECMs of the two data types become identical (Fig. 4.1, C and F).

This is proven by deriving the ECM in the log-space Φ for MDD:

$$\begin{aligned}\Phi_{mdd}(j, j) &= \frac{1-\pi_j}{\tau\pi_j} \\ \Phi_{mdd}(i, j) &= -\tau^{-1}\end{aligned}\quad (4.25)$$

and PDD:

$$\begin{aligned}\Phi_{pdd}(j, j) &= (\Delta t \xi_j)^{-1} \\ \Phi_{pdd}(i, j) &= 0\end{aligned}\quad (4.26)$$

The proof of equivalence of the clr-space ECMs is given in Appendix B. The structure of this universal ECM is:

$$\begin{aligned}\Psi(j, j) &= a_d^2 E(X_j)^{-1} + a_{od}^2 \sum_{m \neq j}^D E(X_m)^{-1} \\ \Psi(i, j) &= a_{od} a_d (E(X_i)^{-1} + E(X_j)^{-1}) + a_{od}^2 \sum_{m \neq \{i, j\}}^D E(X_m)^{-1}\end{aligned}\quad (4.27)$$

where $E(X_j)$ is the expectancy of a multinomial or Poisson random variable and a_d and a_{od} are the diagonal and off-diagonal entries of the clr-transformation matrix \mathbf{A}_D (i.e. $1 - D^{-1}$ and $-D^{-1}$), respectively.

4.2.6 Limiting properties of the clr-space ECM

If the count probabilities (MDD) or count rates (PDD) are identical among the variables, the count probability vector or count rate vector reads $[D^{-1}, \dots, D^{-1}]$. For the corresponding ECM the following holds:

$$\begin{aligned}\lim_{var(\boldsymbol{\xi}) \rightarrow 0} \Psi_{pdd}(j, j) &= (1 - D^{-1})(\Delta t \xi)^{-1} \\ \lim_{var(\boldsymbol{\xi}) \rightarrow 0} \Psi_{pdd}(i, j) &= -(\Delta t \xi)^{-1} [2a_d a_{od} + a_{od}^2 (D - 2)] = -(D \Delta t \xi)^{-1}\end{aligned}\quad (4.28)$$

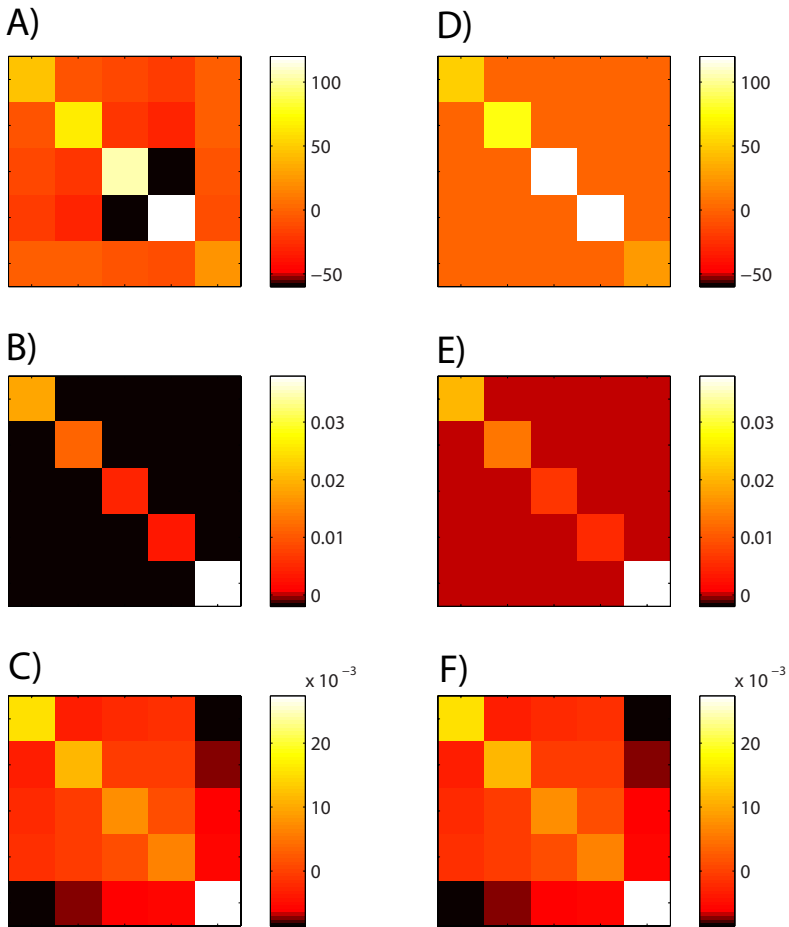


Figure 4.1: MDD and PDD ECMs in the clr space associated with the composition $[0.1, 0.15, 0.3, 0.4, 0.05]$. A, B and C represent the ECMs of this composition for MDD in the count space, log space and clr space, respectively. D, E, and F show ECMs for PDD in the count, log and clr space, respectively.

Another limiting property of the universal ECM is:

$$\begin{aligned}\lim_{D \rightarrow \infty} \Psi_{pdd}(j, j) &= (\Delta t \xi)^{-1} \\ \lim_{D \rightarrow \infty} \Psi_{pdd}(i, j) &= 0\end{aligned}\quad (4.29)$$

For convenience we assume that the data are PDD, which means that $E(X_j)^{-1} = (\Delta t \xi_j)^{-1} = \frac{D}{\Delta t}$. By replacing ξ by π and Δt by τ the equivalent result for MDD is obtained.

4.2.7 ECM of a count data set

Until now we have analysed the error properties of count data in terms of their distribution (i.e. multinomial or Poisson). In reality, however, we do not observe the population parameters but we observe a random sample from this underlying distribution. Furthermore, the aim in multivariate analysis is to characterise a set of count data rather than individual realisations.

An important aspect when analysing noisy sets of count data is whether or not it is reasonable to assume that there is one ECM which applies to the set as a whole. This can be predicted on the basis of the above limiting cases. Therefore we consider a set of N MDD contained in \mathbf{X} , which are drawn from N unique populations. Their count probabilities are contained in matrix $\mathbf{\Pi}$ and their sample sizes in the vector $\boldsymbol{\tau}$. The matrix \mathbf{X} is decomposed by first calculating the vector with observed samples sizes \mathbf{t} :

$$\mathbf{t} = \mathbf{X} \cdot \mathbf{1}_D \quad (4.30)$$

Because we consider MDD, $t_k = \tau_k$ for any $k = 1, 2, \dots, N$. For the count probability estimates \mathbf{P} the following holds:

$$\mathbf{P} = \text{diag}(\mathbf{t}^{-1}) \mathbf{X} \quad (4.31)$$

Next we consider the error properties of \mathbf{P} after clr-transformation. Given the proportionality between mean and counting error variance of PDD and MDD, counting errors become identically-distributed only if the data set reflects repeated and identically-sized samples from the same population. Hence, the rows of $\text{diag}(\boldsymbol{\tau}) \mathbf{\Pi}$ must be identical. If, in addition, the dimensionality approaches infinity, the errors become independent. In this limiting case, the sample size may be factored out so that the following holds for the theoretical ECM Ψ of sample k :

$$\lim_{D \rightarrow \infty, \text{var}(\boldsymbol{\pi}_j) \rightarrow 0} \Psi_k = \tau_k^{-1} \text{diag}([\bar{\pi}_1, \dots, \bar{\pi}_D]^{-1}), \quad j \in 1 : D \quad (4.32)$$

where $\bar{\pi}_j$ is the mean probability in the j -th column of $\mathbf{\Pi}$ and $\text{var}(\boldsymbol{\pi}_j) \rightarrow 0$ means that the variance in the j -th column of $\mathbf{\Pi}$ approaches zero.

In case $\bar{\pi}_1 = \bar{\pi}_2, \dots = \bar{\pi}$, this reduces to:

$$\lim_{D \rightarrow \infty, \text{var}(\mathbf{\Pi}) \rightarrow 0} \Psi_k = (\tau_k \bar{\pi})^{-1} \mathbf{I}_D \quad (4.33)$$

which implies *iid* errors.

From these limiting cases we draw two conclusions. Firstly, considering the fact that a data set generally reflects samples drawn from different populations with finite D , counting errors in a clr-transformed data set are not identically distributed and not independent.

Secondly, whereas sample size causes an inflation/deflation of the ECM, the structure of the ECM is fully controlled by the dimensionality in combination with the count probability (MDD) or count rate (PDD) matrix.

4.3 Simulation experiment

A simulation experiment was conducted to investigate whether accommodating for counting errors can help to produce more robust RRAs. In the experiment count data are modelled as compositions that are either subject to multinomial or Poisson random errors. In case of Poisson random errors, a variable scale is imposed.

4.3.1 Reduced-rank methods

Making a RRA is based on estimating a linear basis in \mathbb{R}^D . Subsequently this coordinate system is partitioned into two disjoint subspaces: an included and an excluded subspace. One purpose of making a RRA is to allocate as much of the underlying signal to the included subspace while simultaneously allocating noise to the excluded subspace (i.e., denoising).

The benchmark in RRA is to estimate the included subspace based on a least-squares criterion. In that case, RRA is numerically equivalent to Principal Component Analysis (PCA) (Jolliffe, 2002). The underlying assumption is that errors are *iid*. To investigate the impact of this assumption and successively less restrictive assumptions on the resulting RRAs a simulation experiment is conducted. Various RRA methods are compared which accommodate for counting errors to a different extent. These alternative methods use either scaling or Maximum Likelihood estimation and they are discussed in more detail in the next sections.

u-RRA (method 0)

Unweighted RRA (u-RRA) is based on estimating the data covariance matrix (DCM) Σ using the zero-mean clr transformed count data matrix \mathbf{Z}_0 :

$$\Sigma = \left(\frac{1}{N-1} \right) \mathbf{Z}_0' \mathbf{Z}_0 \quad (4.34)$$

From this DCM Σ , a Singular Value Decomposition is constructed:

$$\Sigma = \mathbf{U} \mathbf{S} \mathbf{V}' \quad (4.35)$$

The loadings are contained in \mathbf{V} and the corresponding scores matrix \mathbf{T} is given by:

$$\mathbf{T} = \mathbf{Z}_0 \mathbf{V} \quad (4.36)$$

A RRA of rank R is obtained by using only the scores and loadings of the largest R eigenvalues:

$$\hat{\mathbf{Z}}_0 = \mathbf{T}_{1:R} \mathbf{V}'_{1:R} \quad (4.37)$$

In this unweighted approach, it is assumed that $\Psi_k \simeq c \mathbf{I}_D$ for any $c > 0$ and $k = 1, 2, \dots, N$.

w-RRA (method 1)

A straightforward way to compensate for variable errors is to apply SVD to a scaled data matrix (Martens et al., 2003). We refer to this approach as weighted RRA (w-RRA). In case the columns as well as the rows exhibit variable errors, the data matrix should be pre- and post-multiplied by scaling matrices which have the inverse row and column standard deviations on their diagonals, respectively. This methodology was used for PCA of spectral images (Keenan and Kotula, 2004) in which the squared row and column mean of the count-data matrix were placed on the diagonal of the left and right scaling matrix, respectively. In case the matrix with errors has unit rank, applying SVD to a scaled matrix is optimal in a maximum likelihood sense (Cochran and Horne, 1977).

Our implementation of w-RRA is based on parameterisation of the errors in the same manner as the count data, i.e., the matrix with clr space error variances \mathbf{F} is decomposed as the product of a row vector $\tilde{\mathbf{t}}$ containing the scale and a matrix $\tilde{\mathbf{P}}$ having compositions as rows:

$$\mathbf{F} = \text{diag}(\tilde{\mathbf{t}})\tilde{\mathbf{P}} \quad (4.38)$$

From $\tilde{\mathbf{t}}$, a vector \mathbf{w} is derived containing row weights:

$$\mathbf{w} = N\mathcal{C}[\tilde{\mathbf{t}}]^{-1/2} \quad (4.39)$$

Next, the mean error vector is calculated in a weighted manner:

$$\tilde{\mathbf{p}} = \text{clr}^{-1} \left(\frac{1}{N} \sum^N \text{matclr}(\tilde{\mathbf{P}}) \right) \quad (4.40)$$

i.e., the errors are treated as if they were compositions. The corresponding scaling matrix is given by:

$$\mathbf{W} = \text{diag}(\tilde{\mathbf{p}})^{-1/2} \quad (4.41)$$

The w-RRA method includes multiplication of the clr-transformed data by the scaling matrix \mathbf{W} , after which the weighted mean is calculated whereby the elements of \mathbf{w} are used as weights. This mean is subtracted from \mathbf{Z} to form \mathbf{Z}_0 . Next, the DCM Σ is calculated in a weighted manner:

$$\Sigma = \left(\frac{1}{N-1} \right) \mathbf{Z}'_0 \text{diag}(\mathbf{w}) \mathbf{Z}_0 \quad (4.42)$$

Then, the same steps are followed as in u-RRA: (1) SVD of the DCM, (2) extraction of scores, and (3) constructing the RRA to the scaled data for a certain rank. To find the RRA in terms of the original units, they should be multiplied by \mathbf{W}^{-1} and the mean should be added.

The w-RRA method implicitly makes the following approximation to the ECM:

$$\Psi_k \simeq \text{diag}(\tilde{t}_k \tilde{\mathbf{p}})^{-1} \quad (4.43)$$

for any $k = 1, 2, \dots, N$.

d-MLRRA (method 2)

The second and third alternative RRA methods are fully based upon Maximum Likelihood PCA (MLPCA) (Wentzell and Lohnes, 1999), and they are therefore referred to as Maximum Likelihood RRA (MLRRA) methods. MLPCA was first introduced by Wentzell et al. (1997). In that same contribution, an algorithm was proposed based on Alternating Least Squares (ALS). This algorithm is believed to always converge and it is also employed here.

The rationale behind MLRRA is that, given a rank R , ML approximations to the data can be made in both the column- and row-space using SVD (i.e. the "original" and "alternate" space, respectively). By consecutively using the basis that was determined in the alternate space as input for the original space and vice versa, the objective function values in the two spaces will converge which marks the ML-solution. Although it is technically not truly ML, we estimate and subtract the weighted mean before applying the method (cf. Wentzell, 2009).

In contrast to the previously introduced methods, in MLRRA the approximation is not obtained by means of orthogonal projection. Instead, the MLRRA is given by:

$$\hat{\mathbf{z}}_k = \hat{\mathbf{V}}(\hat{\mathbf{V}}'\Psi_k^{-1}\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}'\Psi_k^{-1}\mathbf{x} \quad (4.44)$$

In diagonal MLRRA (d-MLRRA), it is implicitly assumed that Ψ_k is diagonal.

f-MLRRA (method 3)

In full MLRRA (f-MLRRA), use is made of the full ECM associated with every observation. The ML solution can no longer be found by consecutively estimating the ML-solution in the row and column space because there is no straightforward relation between the two spaces. To solve this we employ the approach proposed by Wentzell et al. (1997): instead of solving the ML problem for N points in D -dimensional space, the problem may be defined as finding the ML solution given 1 point in $N \times D$ dimensional space. Otherwise, f-MLRRA is based on the same numerical scheme as d-MLRRA.

The f-MLRRA algorithm requires Ω^{-1} as input, which is a block-diagonal matrix with the inverse of all Ψ_k on its diagonal. The ECM estimate associated with every row in the clr-transformed data matrix may be obtained using Equation 4.24. However, in a numerical scheme this would require the use of a for-loop. It is more efficient to calculate the ECMs all at once by matrix multiplication. We therefore propose calculating a $NN \times ND$ block-diagonal matrix with all ECMs on its diagonal referred to as Ω :

$$\Omega = (\mathbf{I}_N \otimes \mathbf{A}_D) \text{diag}(\text{vec}(\mathbf{X}')^{-1}) (\mathbf{I}_N \otimes \mathbf{A}_D) \quad (4.45)$$

The operation $\mathbf{I}_N \otimes \mathbf{A}_D$ is the construction of a matrix with N repetitions of \mathbf{A}_D on its diagonal.

Because Ω is block-diagonal, inversion may be carried out on a block-by-block basis for speed. The inverse of the ECM in the alternate space is denoted as Ξ^{-1} and is derived from Ω^{-1} using the commutation matrix \mathbf{K} :

$$\Xi^{-1} = \mathbf{K}\Omega^{-1}\mathbf{K}' \quad (4.46)$$

Following Wentzell et al. (1997), the approximations in the two spaces are derived as:

$$\begin{aligned} \text{vec}(\hat{\mathbf{X}})_{orig} &= \hat{\mathbf{V}}(\hat{\mathbf{V}}'\boldsymbol{\Omega}^{-1}\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}'\boldsymbol{\Omega}^{-1}\text{vec}(\mathbf{X})_{orig} \\ \text{vec}(\hat{\mathbf{X}})_{alt} &= \hat{\mathbf{V}}(\hat{\mathbf{V}}'\boldsymbol{\Xi}^{-1}\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}'\boldsymbol{\Xi}^{-1}\text{vec}(\mathbf{X})_{alt} \end{aligned} \quad (4.47)$$

Since f-MLRRA is based on the same numerical scheme as d-MLRRA, it is assumed that convergence marks the ML solution. This method does not make any assumptions regarding the error covariance structure.

4.3.2 Generating synthetic data

To study the performance of the different algorithms, synthetic data sets were generated. This allows us to compare the RRAs to noisy data with the "truth", i.e. the noise-free data. The process of synthetic data generation is shown in Figure 4.2.

An expected number of counts per variable f_k was simulated so that for τ_k holds that $\tau_k = f_k \times D$. The value for f_k was drawn from a uniform distribution ranging between f_{min} and f_{max} . A vector \mathbf{m} of column means of the clr-transformed count probability matrix was generated. The D values contained in \mathbf{m} were drawn from a uniform distribution ranging between $-\mu_{max}$ and μ_{max} . The larger the μ_{max} , the more spread among the column means of the proportions matrix \mathbf{P} . We also generated a D -part vector with variable-specific variances \mathbf{v} whose entries were randomly selected from a uniform distribution ranging between 0 and σ_{max}^2 , which were subsequently turned into a vector of standard deviations \mathbf{s} .

Data correlation matrices \mathbf{R} were generated using the MATLAB function "randcorr" in the "matrix gallery" (MATLAB, 2011). The produced correlation matrices were, however, not a suitable matrix for clr-transformed data because clr-transformed data are rank deficient. Hence, the correlation matrices were deflated so as to constrain them to the clr space. This, in combination with the introduction of additional rank deficiencies was achieved by first calculating the SVD of \mathbf{R} :

$$\mathbf{R} = \mathbf{U}\boldsymbol{\Theta}\mathbf{V}' \quad (4.48)$$

where $\mathbf{U} = \mathbf{V}$ because \mathbf{R} is symmetric. The basis vector \mathbf{h} orthogonal to the clr space reads:

$$\mathbf{h} = [D^{-1/2}, D^{-1/2}, \dots, D^{-1/2}] = D^{-1/2} \cdot \mathbf{1}_D \quad (4.49)$$

where $\mathbf{1}_D$ is D -part unit column vector. The clr space covariance matrix was then obtained by deflating \mathbf{V} in the following manner:

$$\check{\mathbf{V}} = \mathbf{V} - \mathbf{h}(\mathbf{h}\mathbf{V}) \quad (4.50)$$

Next, additional rank deficiencies were introduced. First, an integer between 1 and $D-2$ was selected randomly, reflecting the desired effective rank of the covariance matrix. The additional rank deficiency was achieved by setting $D-1-R$ randomly selected diagonal elements of $\boldsymbol{\Theta}$ to zero. The result is referred to as $\boldsymbol{\Theta}_{rd}$ and the clr-deflated and rank deficient correlation matrix could now be derived as follows:

$$\mathbf{R} = (\check{\mathbf{V}}\boldsymbol{\Theta}_{rd}^{1/2}\check{\mathbf{V}}')'(\check{\mathbf{V}}\boldsymbol{\Theta}_{rd}^{1/2}\check{\mathbf{V}}') \quad (4.51)$$

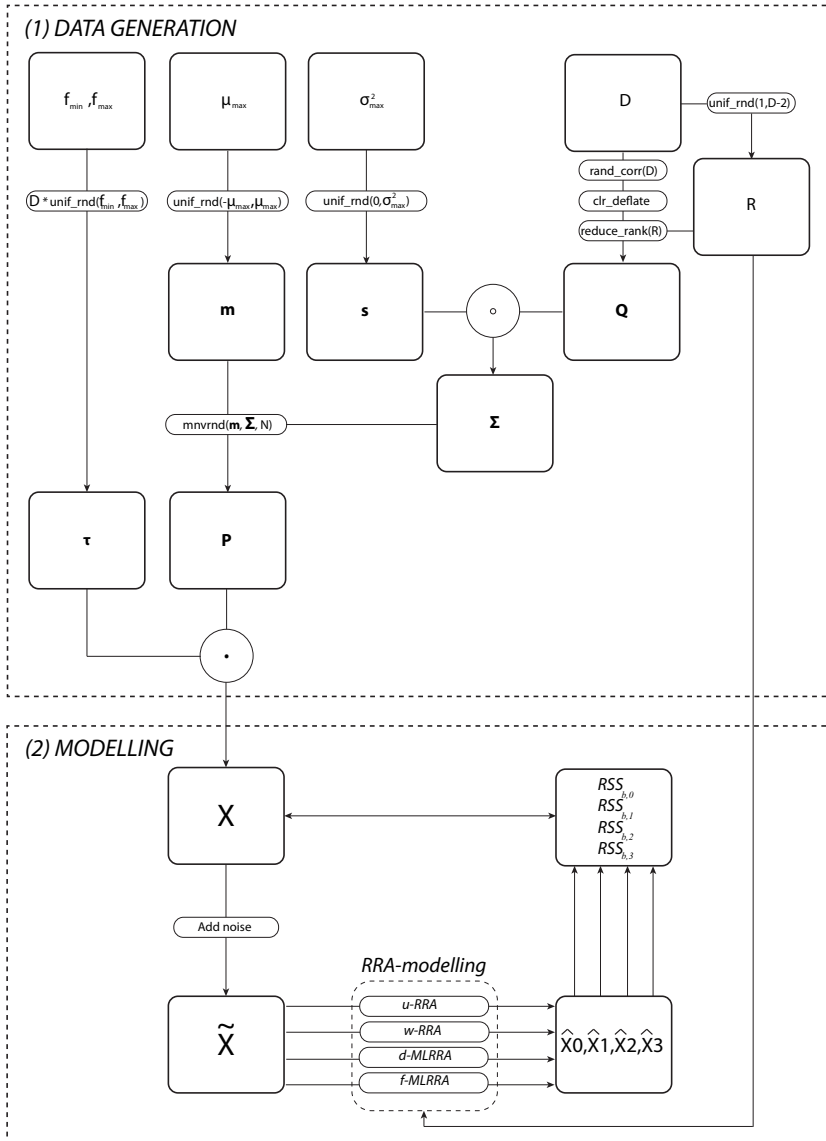


Figure 4.2: Setup of the simulation experiment, which consists of two parts: a data generation part (1) and a modelling part (2).

so as to ensure positive-definiteness. Finally, the DCM Σ was obtained in the following manner:

$$\Sigma = \mathbf{R} \circ (\mathbf{s}'\mathbf{s}) \quad (4.52)$$

Then, N unique clr-transformed compositions were generated from multivariate normal distributions, characterised by the mean vector \mathbf{m} and covariance-matrix Σ . These clr-transformed compositions were transformed back to proportions to form $\mathbf{\Pi}$. Next, samples were drawn from the distributions characterised by the distribution parameters contained in $\mathbf{\Pi}$ and τ . These $k = 1, \dots, N$ realizations were either drawn from a Poisson distribution or from a multinomial distribution. For PDD this implies:

$$x_{kj} \sim \text{Poiss}(\lambda_{kj}) = \text{Poiss}(\tau_k \pi_{kj}) \quad (4.53)$$

whereas for MDD:

$$\mathbf{x}_k \sim \text{Mult}(\tau_k, [\pi_1, \dots, \pi_D]) \quad (4.54)$$

The occurrence of zeros jeopardises the use of log-ratio analysis. The probability matrix $\mathbf{\Pi}$ contains no zeros because it originates from a logarithmic (clr) space, however, the sample size may still cause zeros to occur in \mathbf{X} . In reality, a significant number of zeros in \mathbf{X} indicates that the sample is too small to properly characterise the population. On the other hand, a sample size which is too large is uneconomical. Designing a measurement strategy is aimed at finding this trade-off between data quality and costs. The same was done in this simulation experiment: the sample sizes were set sufficiently large in relation to the simulated count probabilities. Nevertheless, \mathbf{X} sporadically contained zeros which were replaced by a small positive number. Therefore we employed the multiplicative replacement method proposed by Martín-Fernández et al. (2003), which is considered to have a minimal influence on the data covariance structure. Given that \mathbf{x} is a composition where the j -th entry equals zero, this replacement method is given by:

$$x_j = \begin{cases} \delta_j, & \text{if } x_j = 0, \\ 1 - \frac{\sum_{k|x_k=0} \delta_k}{\kappa} x_j & \text{if } x_j > 0, \end{cases} \quad (4.55)$$

Following suggestions by Sandford et al. (1993) we adopt $\delta = 0.55$.

4.3.3 Example data set

All proposed RRA methods were applied in clr space. Their ability to reconstruct the noise-free data was also evaluated in clr space because the aim of this study is to reconstruct relative (i.e. proportional) variations in counts. However, the clr-transformation is invertible and the RRAs may be transformed back to the count space by applying the following inverse transformation to $\hat{\mathbf{Z}}$:

$$\hat{\mathbf{X}} = \text{diag}(\mathbf{t}) \text{matchr}^{-1}(\hat{\mathbf{Z}}) \quad (4.56)$$

In case a RRA produced by the benchmark method (i.e., u-RRA) is inferior to an RRA produced by any of the alternative methods, the same is likely to be the case in the count space.

Parameter	Values
N	100
D	3, 5, 10, 15, 20, 25
f_{min}/f_{max}	300/300 (MDD), 100/500 (PDD)
μ_{max}	0, 0.5, 1
σ_{max}^2	0.1, 0.2, 0.5, 1

Table 4.1: Parameter values adopted in the simulation experiment.

In Figure 4.3, the RRA methodology is illustrated using an example data set of PDD ($D = 10$, $N = 100$ and $f_{min} = 100$, $f_{max} = 1000$). Figures 4.3A and 4.3B show the clr-transformed count data and raw count data, respectively. Because $f_{min} > f_{max}$, much variability in the raw counts is associated with variable sample size. However, as a result of the scale-invariance of clr-transformed data, this variability is not reflected by the clr transformed count data (4.3A). Figures 4.3C and 4.3E show the RRA obtained with u-RRA, whereas Figures 4.3D and 4.3F show the f-MLRRA result as clr-transformed counts and raw counts, respectively. It is immediately apparent that the cross plot associated with u-RRA shows more scatter than the one associated with f-MLRRA.

4.3.4 Experimental design

The ability of the RRA methods to reconstruct noise-free count data was investigated for data sets with different properties. In the simulation experiment, seven parameters were varied (Table 4.1): the number of observations N , the dimensionality D , the sample size parameters f_{min} and f_{max} , the maximum variance of the clr-transformed variables σ_{max}^2 , the maximum mean μ_{max} of the clr-transformed variables, and the distribution from which the data are drawn (i.e. MDD or PDD). MDD have a constant sample size and therefore $f_{min} = f_{max}$. For PDD, $f_{min} < f_{max}$. Note that $f_{min} + f_{max}$ is constant. Also N is held constant ($N = 100$) in all experiments. All combinations of parameter values were tested.

In the experiment, one configuration serves a special goal: a data set with a maximum variance of 0.1 represents a workable example of a data set containing multiple realizations from the same specimen. Because reduced-rank methods rely on covariance, it is impossible to analyse the performance of RRA methods for data sets with zero variance (i.e. with $\sigma_{max}^2 = 0$).

4.3.5 Diagnostics

The ability of the various RRA methods to reconstruct the noise-free data is tested according to the flow chart in Figure 4.2. Each experiment consists of $l = 1, 2, \dots, B$ data sets. Because each method was applied to every data set, their outcomes can be evaluated in a paired manner. First, the residual sum of squares is calculated for every data set b and method q :

$$RSS_{bq} = \sum_{k=1}^N \sum_{j=1}^D (\hat{z}_{kj} - z_{kj})^2 \quad (4.57)$$

where \hat{z} is the RRA in clr space and z is the noise-free data in clr space. Then, the

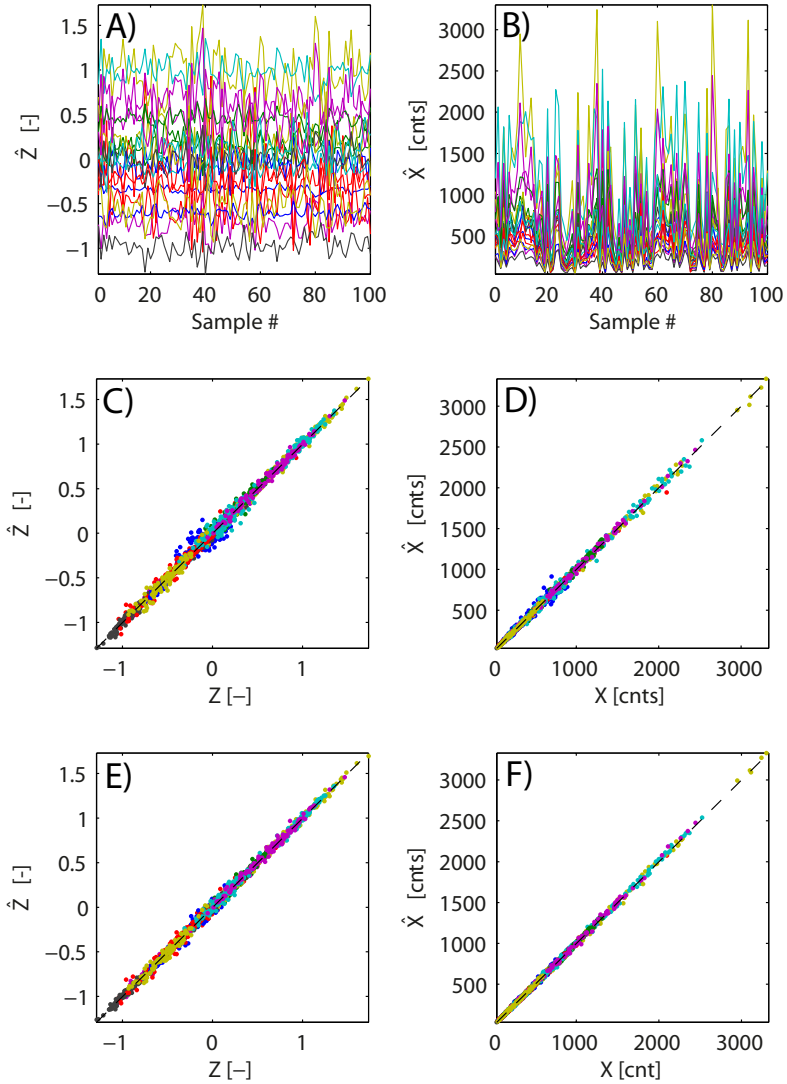


Figure 4.3: Modelling result for the example data set. In A and B, the RRA is shown, both as clr-transformed counts and as counts, respectively. The cross plots in C-F show the rank-deficient noise-free data against the u-RRA result (C and D) and the f-MLRRA result (E and F). C and E show the results in clr-transformed counts whereas D and F show the results in counts. In all plots, different colors reflect different variables.

performance gain parameter PG is calculated with respect to the benchmark (i.e. u-RRA):

$$PG_q = \exp \left(\frac{1}{B} \sum_{l=1}^B \log \left(\frac{RSS_{b0}}{RSS_{bq}} \right) \right) \quad (4.58)$$

where B is the number of data sets generated within each experiment. The value of PG_q reflects the average proportional decrease in residual variance obtained when using the method q as alternative for the null method.

When comparing the results associated with different methods and different data sets, it is important that the degrees of freedom are taken into account. All RRA methods comprise the same number of parameters and likewise consume the same number of degrees of freedom. However, the degrees of freedom of the residuals vary among data sets with different N and D . Furthermore, the residuals analysed in this study represent residuals between the RRA and the noise-free data whose degrees of freedom are difficult to predict.

We circumvent the necessity of defining the degrees of freedom by evaluating the resulting PG -values also in a relative sense so as to be able to compare experiments comprising different D . This measure is referred to as the relative performance gain RPG , reflecting the decrease in residual variance as a result of choosing a particular method in favour of another. For instance, RPG_{32} is the relative performance gain when using method 3 (f-MLRRA) instead of method 2 (d-MLRRA) which is given by:

$$RPG_{32} = \frac{PG_3 - PG_2}{PG_2} \quad (4.59)$$

RPG reflects a proportional performance gain, i.e. $RPG_{32} = 0.10$ implies that method 3 performs 10% better than method 2.

4.3.6 Computational load

In addition to the potential performance increase obtained by the three alternative methods, also the computational load was monitored. We monitored the time it took for each method to produce an answer. Furthermore, for the ML-based methods which rely on an iterative scheme, we adopted a maximum number of iterations (500). If convergence was not reached within this number of iterations, the routine was terminated. We also exited the algorithm when the SVD algorithm of MATLAB did not converge when inverting a covariance matrix. All code was written in MATLAB and executed on a PC with an Intel Core i5 (3.1 GHz) processor.

4.4 Results

In Figure 4.4, the PG -values are shown for both MDD and PDD. We see that the alternative methods always perform better than our null method (u-RRA). Only when $\mu_{max} = 0$, w-RRA displays the same performance as u-RRA. MLPCA-based methods are not better than u-RRA in case both σ_{max} and μ_{max} are close to zero. The results also show that generally, PDD benefit more from the alternative methods than MDD.

Figure 4.5 shows the effect of σ_{max}^2 on RPG_{21} . It turns out that as σ_{max}^2 decreases, so does RPG_{21} . In Figure 4.6, the effect of dimensionality is shown on RPG_{32} . We observe that as D increases, RPG_{32} approaches zero. This applies to both MDD and PDD: in both

cases, the performance gain has decreased to less than 10% for $D = 15$. Finally, Figure 4.7 shows the median convergence time of the different methods as function of the size of the data matrix.

4.5 Discussion

4.5.1 Method performance

The results in Figure 4.4 show that all alternative methods generally perform better than u-RRA. The fact that MLRRA methods give better results than u-RRA is consistent with earlier work by Wentzell and Lohnes (1999). Furthermore, we observe that this effect is the strongest for PDD. This can be explained from the fact that in MDD there are only variations among the errors in the columns whereas PDD is characterised by variable row and column error variances.

For $\mu_{max} = 0$ and for all σ_{max}^2 , PG_1 is practically zero (Fig. 4.4) which indicates that for MDD, w-RRA gives virtually the same result as u-RRA. This stems from the fact that both row and column variances are equal for MDD. The row variances are equal because the number of counted objects is held constant and the column variances are equal because $\mu_{max} = 0$. In this hypothetical case, w-RRA and u-RRA perform equally well.

In the left-most column of Figure 4.4 we observe that as both μ_{max} and σ_{max}^2 approach zero, none of the alternative methods perform better than u-RRA. This can be explained from the fact that when both μ_{max} and σ_{max}^2 are zero, error variances associated with every entry in the data set become equal. Likewise, compensating for errors will not contribute to a better RRA.

From the results in Figure 4.5 it follows that RPG_{21} is largely controlled by σ_{max}^2 . This is caused by the fact that, as σ_{max}^2 approaches zero, the rank of the data matrix approaches unity. Because the error estimates are equal to the observed counts, a data matrix with unit rank implies that the matrix with error variances also has unit rank: in that case the w-RRA solution is equal to the f-MLRRA solution.

The results in Figure 4.6 show that RPG_{32} is a function of the dimensionality. This can be explained from the fact that off-diagonals of the ECM approach zero as the dimensionality goes to infinity. Likewise, as D becomes large, d-MLRRA gives approximately the same result as f-MLRRA.

Equal column means yield on average equal off-diagonals, whereas large differences between the column means result in dispersion among the off-diagonals: intuitively, we expected that ignoring equally-valued off-diagonals would have a smaller impact on the result than ignoring dispersed off-diagonals with the same mean magnitude. However, the results of the experiment are inconclusive with regard to this hypothesis (see Fig. 4.6, A and B): between $D = 5$ and $D = 15$, the $\mu_{max} = 0$ curve is lower than that of $\mu_{max} = 1$ suggesting that the d-MLRRA solution is more close to the f-MLRRA solution when column means are equal. For $D = 3$ and $D \geq 20$, however, it seems to be the other way around.

The three conclusions we can draw are that (1) the impact of the dispersion among column means on the quality of the d-MLRRA is much smaller than that of the dimensionality, (2) the relative performance gain when using f-MLRRA instead of d-MLRRA can be predicted on the basis of D and (3) the relative performance gain as a result of using d-MLRRA instead of w-RRA can be predicted on the basis of σ_{max}^2 .

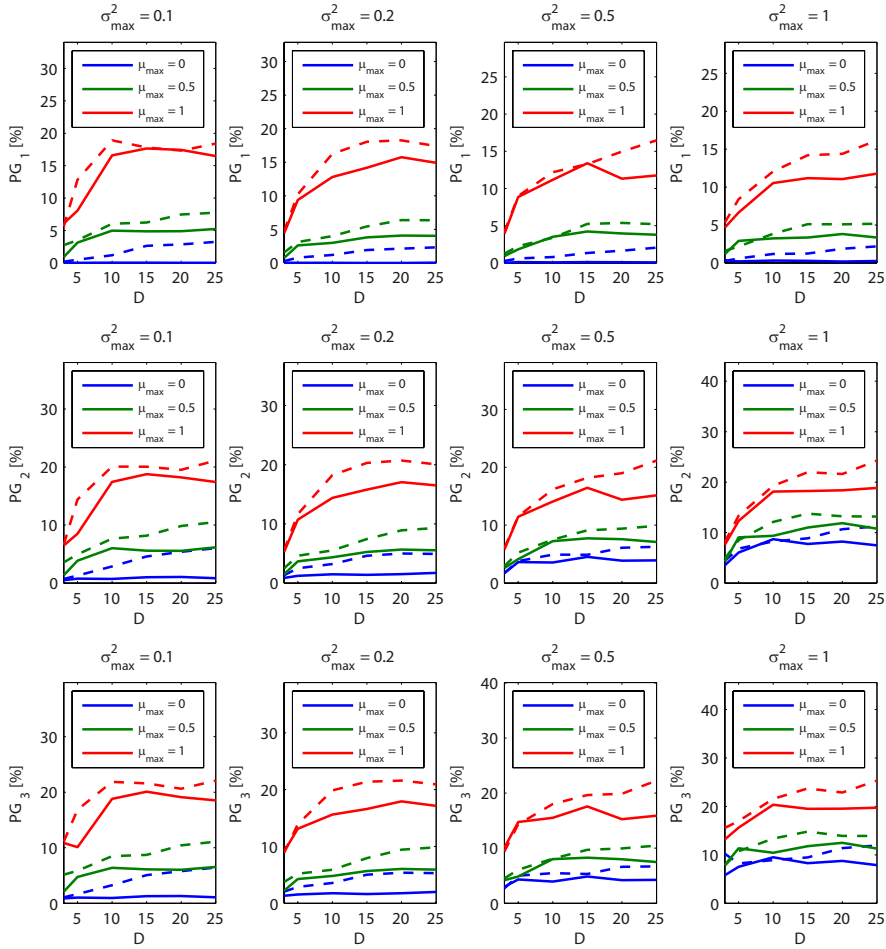


Figure 4.4: Results of the simulation experiment. The plots represent the performance gain PG of the three alternative methods relative to u-PCA (rows) as function of the number of variables D and for different σ_{max}^2 (columns). The solid and dashed lines show the MDD and PDD results, respectively .

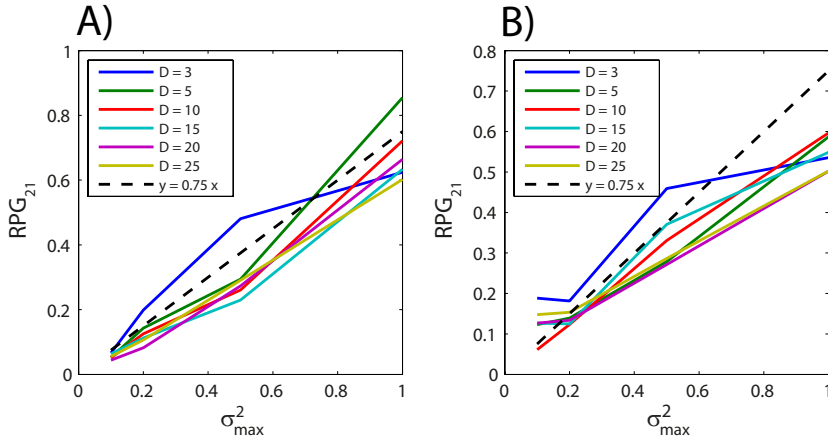


Figure 4.5: RPG_{21} as function of σ_{max}^2 for $\mu_{max} = 1$. A and B show the MDD and PDD results, respectively.

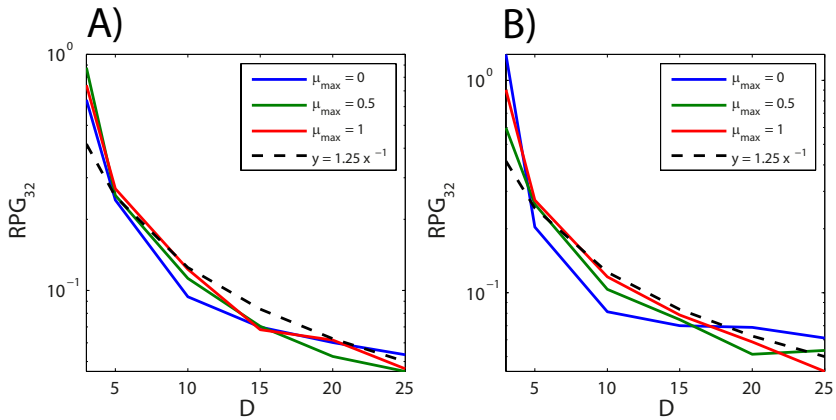


Figure 4.6: RPG_{32} as function of the dimensionality D for the various values of μ_{max} and $\sigma_{max}^2 = 1$. A and B show the MDD and PDD results, respectively.

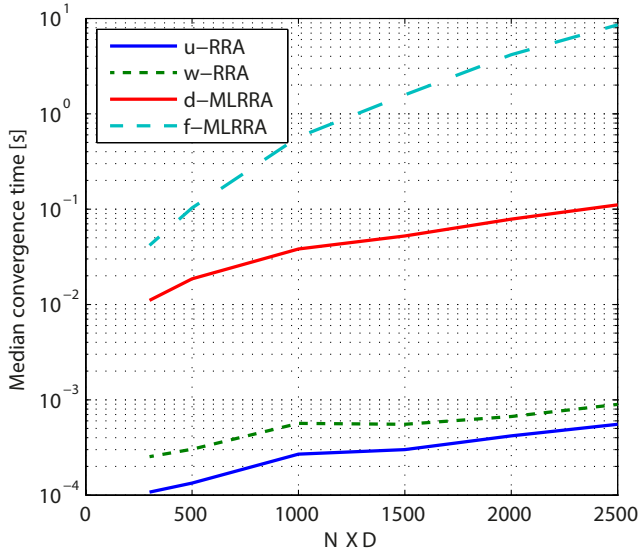


Figure 4.7: Median convergence times observed in the simulation experiment.

4.5.2 Method selection

Our findings regarding the different methods can be summarised in terms of limiting cases. Firstly, we can state that:

$$\lim_{\text{var}(\mathbf{X}) \rightarrow \bar{x}} \left(\hat{\mathbf{Z}}_{\text{u-RRA}} \right) = \hat{\mathbf{Z}}_{\text{d-MLRRA}} \quad (4.60a)$$

$$\lim_{\text{var}(\mathbf{p}_j) \rightarrow \bar{p}_j} \left(\hat{\mathbf{Z}}_{\text{w-RRA}} \right) = \hat{\mathbf{Z}}_{\text{d-MLRRA}} \quad (4.60b)$$

which implies that as the total variance of a count data set can be completely explained by random counting errors, u-RRA gives the same result as d-MLRRA. This was illustrated by the case where $\sigma_{\text{max}}^2 \rightarrow 0$ and $\mu_{\text{max}} = 0$. The fact that u-RRA finds the same solution as d-MLRRA only in case there is no variability in the data set is an interesting result, given that it is generally assumed that the errors in compositional data sets are *iid*. Secondly, Equations 4.60a and 4.60b may be extended because:

$$\lim_{D \rightarrow \infty} \left(\hat{\mathbf{Z}}_{\text{d-MLRRA}} \right) = \hat{\mathbf{Z}}_{\text{f-MLRRA}} \quad (4.61)$$

so that $\hat{\mathbf{Z}}_{\text{d-MLRRA}}$ is interchangeable with $\hat{\mathbf{Z}}_{\text{f-MLRRA}}$ for large D .

Because real data sets generally reflect samples with different properties and/or different sample-sizes we conclude that the simplest method suitable for practical applications is w-RRA. Practical applications in which w-RRA will prove useful are spectroscopic applications where the variability in peak height is small and the number of channels is large. For spectroscopic applications with larger column variability, the d-MLRRA approach is the method of choice. Data sets with small D always require f-MLRRA for accurate RRAs.

4.5.3 Computational performance

The fact that d-MLRRA can serve as an alternative for data sets with large D is convenient because, as shown in Figure 4.7, f-MLRRA is computationally more expensive. In general, f-MLRRA is between one and two orders of magnitude slower than d-MLRRA. Furthermore, the computational load of f-MLRRA increases much faster with the size of the data matrix than that of the other methods. Although the MLPCA code we employed was not specifically designed for speed, these results show that f-MLRRA becomes impractical when combined with resampling techniques such as leave-one-out cross-validation, or when it is applied to very large data sets. Apart from its high computational load, f-MLRRA may be jeopardised by the fact that the matrices become too large for the available storage capacity.

In this study, both d- and f-MLRRA were initialized with the u-RRA result. To increase efficiency we investigated whether computational load may be reduced by means of a different initialization strategy. In some simulation experiments (not shown here) it was tested whether computational load decreased by initializing (1) d-MLRRA with the w-RRA solution and (2) f-MLRRA with the d-MLRRA solution.

It turned out that initialization of d-MLRRA with the w-RRA basis does not necessarily save computation time. In case $\sigma_{max}^2 = 0.1$ we found that computation time is reduced, whereas for $\sigma_{max}^2 = 1$ the computation time was found to have increased. Apparently, d-MLRRA only benefits from w-RRA initialization in case the two solutions are almost equivalent. In that case, however, the d-MLRRA solution has no added value.

We also found that d-MLRRA and f-MLRRA require approximately the same number of iterations to reach convergence, irrespective of the properties of the data set. Initializing f-MLRRA with the d-MLRRA solution reduces the number of f-MLRRA iterations by approximately 50%. This translates to a decrease in convergence time: a time reduction of 40% was observed for $N = 100$ and $D = 25$.

4.5.4 Optimal Scale-Invariant Reduced-rank Approximation (OSIRA)

The results of the simulation experiment may be used to define an algorithm for optimal RRA of noisy count data. This algorithm is based on selecting the optimal RRA method in terms of efficiency and performance based on the properties of a data set.

Although f-MLRRA displayed the best performance, it is unnecessary and not efficient to use f-MLRRA if any of the computationally less demanding methods perform equally satisfactory. Therefore, we propose to calculate a priori the expected RPG -values and use these to select the most suitable method. Selection, in turn, is performed based on a user-specified threshold value r . This value controls when a less efficient method is used in order to get a more reliable RRA, e.g. $r = 0.05$ implies that the user wishes to use a more expensive method if that method performs at least 5% better in terms of residual variance.

First we postulate functions that predict RPG_{32} and RPG_{21} . We found that RPG_{21} may be predicted from σ_{max}^2 , whereas RPG_{32} can be predicted from D . Approximate fits to the simulation results are:

$$\begin{aligned}\widehat{RPG}_{21} &= 0.75\sigma_{max}^2 \\ \widehat{RPG}_{32} &= 1.25D^{-1}\end{aligned}\tag{4.62}$$

The two functions are shown in Figures 4.5 and 4.6. Both fits are acceptable for all $D \geq 5$: in that range they (1) fit reasonably well, and (2) they are conservative: the expected

performance gain is underestimated rather than overestimated.

Next, we use these predicted RPG -values to select the optimal method. In case \widehat{RPG}_{32} is large, f-MLRRA is used. In case \widehat{RPG}_{32} is small, the choice between w-RRA and d-MLRRA is made based on \widehat{RPG}_{21} . Combining these two selection rules defines an algorithm for RRA of count data, which we refer to as OSIRA:

$$\text{OSIRA} = \begin{cases} \widehat{RPG}_{32} > r & \text{d-MLRRA} \xrightarrow{\text{init}} \text{f-MLRRA} \\ \widehat{RPG}_{32} \leq r & \begin{cases} \widehat{RPG}_{21} > r & \longrightarrow \text{d-MLRRA} \\ \widehat{RPG}_{21} \leq r & \longrightarrow \text{w-RRA} \end{cases} \end{cases} \quad (4.63)$$

where $r < 0.25$. By imposing this restriction on r , we foresee that for $D < 5$, f-MLRRA will be the selected method: in this range the predicted RPG -values tend to underestimate the true RPG -values found in the simulation experiment (see Figs. 4.5 and 4.6). Code for OSIRA may be downloaded at <http://www.ascar.nl>.

4.5.5 General remarks

An aspect which we did not consider to solve in a ML sense is mean centering. When errors are uncorrelated among the rows and columns, the ML-estimate of the mean coincides with the weighted arithmetic mean, either along the rows or the columns. However, in case of clr-transformed count data this is only the case in the limit and the weighted arithmetic mean is only an approximation to the truly ML mean. In the future, it may be attempted to include ML-estimation of the mean in the algorithm, however, we expect that for data sets with a considerable number of observations ML-like mean centering will not give a significant improvement.

A way to potentially increase the signal-to-noise ratio of RRAs to spectroscopic data would be to amalgamate variables (i.e. channels). Since the sum of Poisson-distributed variables is again Poisson-distributed for which holds that $\lambda = \lambda_1 + \lambda_2$, the counting error in clr space associated with the newly-formed variable is smaller than those of the amalgamated variables individually. However, to prevent losing signal, only variables reflecting the same signal should be amalgamated. Combining RRA with smart amalgamation strategies is a subject of future research. Another logical extension of the work presented here is to generalize the methodology so that it is applicable to all log-ratio spaces (e.g., the isometric log-ratio space (Egozcue et al., 2003)).

4.6 Conclusions

1. All spectroscopic and compositional data either directly or indirectly originate from count data, or they can be considered to have similar error characteristics.
2. Count data may be decomposed into a scale vector and a proportions matrix \mathbf{P} . Because the rows of \mathbf{P} are compositions, data analysis and modelling should be performed in a suitable metric space such as the clr space. Although this space is scale invariant, the scale and the composition together determine the counting error associated with the clr-transformed variables: the error covariance structure is completely determined by \mathbf{P} where the scale t only causes an inflation/deflation of the error covariance matrix.

3. The error-covariance structure in clr space is identical for data originating from a multinomial or Poisson distribution. This implies that for count data in this space, there is a universal error model.
4. The counting errors in clr space become independent as the number of variables D goes to infinity. The errors become *iid* only in the hypothetical case where a data set comprises repeated realisations with the same sample size from the same population and expressed in terms of an infinite number of object classes. In non-trivial cases, the error covariances of a set of count data will never be *iid*.
5. In a simulation experiment, various RRA methods were tested including unweighted RRA (u-RRA), weighted RRA (w-RRA), diagonal MLRRA (d-MLRRA) and full MLRRA (f-MLRRA). Synthetic rank-deficient data sets were generated and, given that the true rank was known, the ability of the methods to approximate the noise-free data was evaluated. As expected, the f-MLRRA solution was closest to the noise-free data.
6. Under certain conditions, the results obtained with the less advanced methods are nearly identical to those obtained with f-MLRRA. Based on the simulation experiment, an algorithm was proposed which consists of a set of method selection rules employing a user-specified threshold value r whereby $r \leq 0.25$. This algorithm is referred to as Optimal Scale-Invariant Reduced-rank Approximation (OSIRA). Combining OSIRA with an intelligent amalgamation strategy is the subject of future research.

Modelling the uncertainty of routine chemical analyses ¹

5.1 Introduction

Estimating the concentration of chemical components in natural and artificial materials is the fundamental task of analytical chemistry. An equally important task is to estimate or predict the uncertainty of chemical analyses for statistical inference and data processing. An example of inference is process monitoring: if the objective is to approve or reject a batch of an industrial product on the basis of chemical analyses, one must incorporate the natural variability introduced in the analytical process to prevent inadvertent rejection of batches with acceptable quality. Examples of data-processing techniques which benefit from an appropriate model of uncertainty are principal components analysis (Wentzell et al., 1997), multivariate regression (Wentzell and Andrews, 1997; Schreyer et al., 2002) and reduced-rank approximation (Bloemsmas and Weltje, 2015).

The analytical uncertainty of spectro-chemical analysis is controlled by many factors, including instrument resolution and method/operator bias. A so-called bottom-up approach to predict the analytical uncertainty is described in the "Guide to the Expression of Uncertainty in Measurement", i.e. the "GUM" (JCGM, 2008). This approach relies on a thorough understanding of the data-generating process, and it is therefore mainly of practical value for laboratories and instrument manufacturers (Luo et al., 2006; EUROLAB, 2007). Moreover, this approach is risky because an inadequate or incomplete description of the mathematical/physical process may lead to underestimation of the "true" analytical uncertainty (Thompson and Ellison, 2011). The alternative is to use a top-down approach, i.e. use empirical estimates of variability associated with one or a combination of sources of uncertainty (e.g., counting) to predict the "true" uncertainty (Williams et al., 2000). A sensible estimate of the uncertainty is the so-called reproducibility standard deviation, typically denoted by "SDR". The SDR can be estimated empirically by letting similar specimens be analysed by different laboratories using different analytical techniques. Conducting such

¹Submitted as: Bloemsmas, M.R. and G.J Weltje. Predicting the uncertainty of routine chemical analyses: a new modelling approach based on evaluation of proficiency tests. *Analyst*.

an experiment for every sample, however, is very costly and impractical.

For an end user, the most practical approach would be to employ an Uncertainty Function (UF) which predicts the analytical uncertainty based on the measurements itself. The primary ingredient required to construct an UF is a suitable data set to constrain the function. End users are likely to (i) have no particular reason to choose one competent laboratory over the other, and (ii) do not prefer one particular analytical method over another. Suitable data for constructing UFs are therefore provided by Proficiency Tests (PTs) (see Horwitz and Albert, 2006; Hund et al., 2000, and references therein).

Construction of empirical UFs from the outcome of PTs with the objective to predict the SDR is not new. Most analytical chemists will be familiar with the Horwitz function which relates the concentration to an expected SDR by means of a power law (see e.g. Horwitz, 1982). The Horwitz function and the associated HorRat represent the current standard in uncertainty specification and performance evaluation, respectively (Horwitz and Albert, 2006). Despite their world-wide use, the extent to which these and other empirical UFs behave within limits of physical and statistical plausibility have been given little attention. The aim of this contribution is to examine current UFs in the light of physical and statistical theory. We show that existing UFs are broadly consistent with spectrochemical calibration equations. Fundamental problems arise, however, when we try to reconcile these findings with the definition of concentrations as mass fractions. We show that this apparent contradiction may be resolved by employing an appropriate UF. Finally, a new performance measure is proposed to replace the HorRat.

5.2 UFs and spectrochemical theory

5.2.1 UFs from PT data

The rationale behind this study is illustrated in Figure 5.1. In A, the theoretical Probability Density Function (PDF) is shown which characterises all existing chemical laboratories and analytical techniques, i.e. the population. The adopted assumption is that all these techniques and laboratories on average are unbiased, i.e., in the hypothetical case in which we could sample an infinite number of laboratories we would obtain the "true" concentration. The blue line represents this unknown "true" concentration ($=0.4$). In B, the results of a PT are demonstrated: the "population" was sampled, yielding a consensus value (~ 0.4) and an SDR shown in green (i.e. the mean/median analyte concentration and the spread). In C, the two are linked by an UF shown in red. In D, this UF is used to predict the analytical uncertainty associated with a "new" concentration estimate (dashed line). As suggested by the arrows, this workflow is dynamic because UFs may need to be updated whenever new PTs have been conducted.

Given that analytical uncertainty consists of random and systematic effects, it may seem as if using SDR as a proxy for the analytical uncertainty ignores the contribution of systematic effects. Key behind this approach is that although a single laboratory or group of laboratories may be biased, we assume that all laboratories together are unbiased. This is a reasonable assumption, given the general absence of a "knowable truth". Furthermore, bias behaves randomly among the laboratories and methods, which means that random and systematic effects are in effect indistinguishable.

Another assumption which underlies this approach is that the change in predicted SDR as function of concentration is small in relation to the SDR itself. In that case, the UF val-

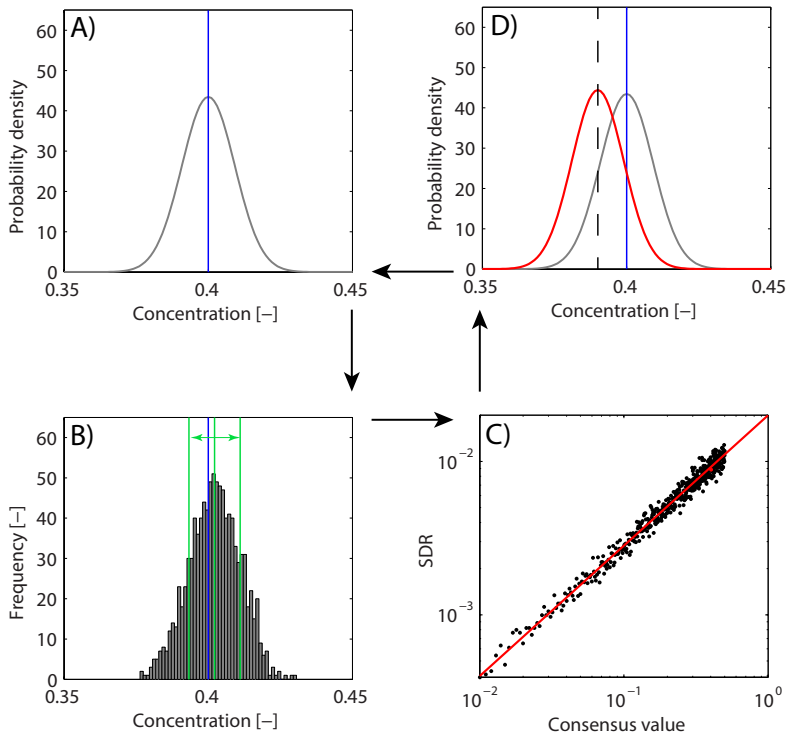


Figure 5.1: Illustration of the PT-based uncertainty modelling approach employed in this study. (A) The true concentration is represented by a blue line, together with a distribution characterising the population of laboratory-instrument combinations. (B) The PT experiment which yields a consensus value and an associated SDR (green lines). (C) An UF is fitted which links consensus value to SDR, and (D) this UF is used to estimate the analytical uncertainty (red) for the "new" concentration estimate, indicated by the dashed vertical line.

ues associated with the "true" and estimated concentration are very similar. Given that the concentration estimate lies within the 95% confidence interval of the distribution of laboratories, the true value thus lies within a 95% confidence interval around the concentration estimate. Despite being an approximation, this justifies statistically the approach presented in this contribution.

5.2.2 Parametric forms

Several types of UFs have been discussed in literature (e.g., Jiménez-Chacón and Alvarez-Prieto, 2006). The most basic UF has a linear form (Zitter and God, 1971):

$$f(c) : \sigma = a_L + b_L c \quad (5.1)$$

which implies a constant rate of change of the standard deviation with concentration, and a positive standard deviation at zero concentration. Zitter and God (1971) proposed a slightly different parametric form given by:

$$f(c) : \sigma = \sqrt{a_L^2 + (b_L c)^2} \quad (5.2)$$

The quadratic form of Eqn. 5.2 is more in line with statistical theory than that of Eqn. 5.1, because it is based on additivity of variances rather than standard deviations.

A more flexible parametric form with a variable exponential term was proposed by Thompson (2011):

$$f(c) : \sigma = \sqrt{a_C^2 + (b_C c^{d_C})^2} \quad (5.3)$$

It was referred to as the "Characteristic Function" (CF) because it is considered to be a function suitable for capturing the behaviour of a specific combination of analytical setup and analyte. Two examples of the CF are shown in Figure 5.2 which both have $b_C = 0.02$ and $d_C = 0.8495$ whereas their intercept terms a_C differ. This parametric form turns out to be adequate for predicting the standard deviation on different levels of precision: it was used to model the instrumental precision (Thompson and Cole, 2011), repeatability (Thompson and Cole, 2009) and reproducibility (Thompson et al., 2008).

5.2.3 UF examples

The best-known example of an UF is the Horwitz function (see e.g. Horwitz et al., 1980; Horwitz, 1982):

$$f(c) : \sigma = 0.02 c^{0.8495} \quad (5.4)$$

It has the same form as the CF with $a_C = 0$, $b_C = 0.02$, $d_C = 0.8495$ (see Fig. 5.2). The Horwitz function is based on an inter-laboratory experiment involving 50 laboratories and a wide range of commodities and analytes. Although based on collaborative trial data, the function is also considered to predict the SDR associated with PTs with acceptable accuracy (Thompson and Lowthian, 1997; Thompson, 2000).

Later studies revealed that the Horwitz function is biased: it overestimates the analytical uncertainty for concentrations exceeding 10% and it underestimates the analytical uncertainty for concentrations below 10^{-8} (Horwitz and Albert, 2006). To correct this bias,

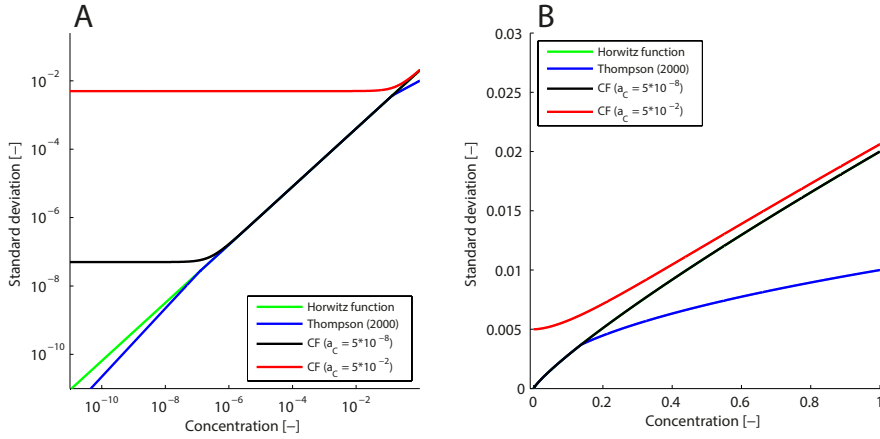


Figure 5.2: Relation between concentration and standard deviation of different UFs on logarithmic (A) and linear axes (B).

Thompson (2000) proposed an UF which subdivides the domain into three parts in which there is a different relation between concentration and uncertainty:

$$f(c) : \sigma = \begin{cases} 0.22c, & \text{if } c < 1.2 \times 10^{-7} \\ 0.02c^{0.8495}, & \text{if } 1.2 \times 10^{-7} \leq c < 0.138 \\ 0.01c^{0.5}, & \text{if } c \geq 0.138 \end{cases} \quad (5.5)$$

The function is shown in Figure 5.2. In the middle part, the function is identical to the Horwitz function. At the high-concentration end, the function suggests that the uncertainty increases more slowly with concentration than specified by the Horwitz function whereas the converse applies to the low-concentration end.

5.2.4 Physical plausibility

All UFs discussed thus far were based upon empirical evidence and only loosely based on statistical considerations. A few studies have attempted to provide a theoretical basis for the parametric form of these UFs post-hoc. They tried to explain the exponential relation between the concentration and the error from conceptual models. Hall and Selinger (1989), for instance explained this proportionality from the principle of least effort (i.e. Pareto's law). Albert and Horwitz (1997) explained proportionality between concentration and error from fractal theory, i.e., they stated that smaller concentrations are measured with a higher (concentration) resolution. Although intuitively these conceptual models make sense, they are not a proof in the strictly mathematical or statistical sense.

5.2.5 Calibration

In an attempt to break away from the empirical nature of UFs, we evaluate the physical tractability of these functions in relation to the nature of the data. A starting point for this

evaluation is to consider the basis of spectroscopic quantification methods (e.g., Jenkins, 1999; De Vries and Vrebos, 2002):

$$C \propto I \quad (5.6)$$

i.e. proportionality between absorption or emission peak intensities I (in, e.g., counts per second) and analyte concentrations C (in, e.g., proportions). It is assumed that I relates to a particular analyte peak and may be written as:

$$I = \sum_{n=1}^N S_n - B_n \quad (5.7)$$

whereby B_k reflects the background in the k -th channel and S_k is the spectral count in this same channel. If we write the proportionality in terms of an equality we obtain the basic univariate calibration equation (BUCE):

$$C = KI \quad (5.8)$$

where K is a constant whose units must be the inverse of the units of I because concentrations are dimensionless by definition.

5.2.6 Propagation of counting errors

A reasonable assumption with regard to spectroscopic quantification is that it behaves as a Poisson counting process. Based on Poisson counting statistics it can be shown that the background subtraction step in the calculation of I could explain the necessity of an intercept term. The counting error in I as a result of finite count times reads:

$$\delta^2 I = \sum_{n=1}^N (\delta^2 S_n + \delta^2 B_n) \quad (5.9)$$

and because the counting errors in S_n and B_n are given by Poisson counting statistics:

$$\delta^2 S = E(S) = E(B + I), \quad \delta^2 B = E(B) \quad (5.10)$$

the following expression is obtained for the counting error in I :

$$\delta^2 I = \sum_{n=1}^N (E(U_n) + E(2B_n)) \quad (5.11)$$

If we also assume that the background is flat (i.e. $B_n \approx B$ for all n), the following holds:

$$\delta^2 I \approx 2NB + I \quad (5.12)$$

Given BUCE the following may be written for the Poisson variance of C :

$$\delta^2 C \approx a + bI \quad (5.13)$$

where a and b have the same units as K . Taking the square root yields a function with the same form as the CF (Eqn. 5.3) with $a_C = a$, $b_C = b$ and $d_C = 1/2$.

5.2.7 SDR at low concentrations

Equation 5.13 shows that background subtraction, which is necessary to determine intensities, is a mechanism that can explain the necessity of an UF with intercept term.

Studies at various levels of precision showed that the intercept is generally greater than zero (see Thompson et al., 2008; Thompson and Cole, 2009, 2011). The Horwitz function, on the other hand, does not have an intercept term. A possible explanation for this difference is the fact that the intercept is hardware-, software- and analyte-specific. The study by Thompson et al. (2008) supports this hypothesis: analysis of PT data involving food stuffs yielded that intercepts varied between 0.015 (Nitrogen) and 4 (Acrylamide). If no distinction is made between analytical setups and analytes, which was the case for the Horwitz function, this non-linearity does not manifest itself as a trend that can be captured by one function. Moreover, this non-linearity may not always stand out from the noise.

An intercept in an UF causes a departure from a linear trend in log-log space which sets in at low concentrations, and manifests itself as a relative flattening (see Figure 5.2). The opposite behaviour is implied by the UF proposed by Thompson (2000) (Eqn. 5.5). This steepening trend has been ascribed to progress in data acquisition and data processing technology (Thompson and Lowthian, 1997). However, instead of progress in technology a simpler explanation is that censoring has affected the distribution of the data. Since negative values are by definition not physically meaningful, laboratories and/or software manufacturers may be tempted to replace these negative values by zeros or very small positive values (e.g. detection limits). Replacing negative realizations by zero or a small constants will result in the data having an asymmetric distribution whose spread may be smaller than expected on the basis of the Horwitz function.

To illustrate this phenomenon, we performed a simulation experiment. For a given concentration, 10,000 random realizations were generated from a normal distribution whose standard deviation conforms to the Horwitz function. All realisations which were smaller than zero were set at zero. Then, the experimental mean and standard deviations were calculated. This was performed for different concentrations: the concentrations are given by 10^x where x ranges between -15 and -5 with increments of 0.25. The results are shown in Figure 5.3 where the relation between the experimentally derived standard deviation and the theoretical standard deviation is shown in (A), and the ratio between theoretical and experimental standard deviation is shown in (B). Note that the influence of censoring becomes negligible for concentrations above 10^{-9} whereas the ratio between the theoretical and experimental standard deviation is greater than two for concentration below 10^{-14} .

5.2.8 SDR at high concentrations

Conversion of Equation 5.13 to a CF-like function yields an exponential term equal to 1/2. Hence only an UF with an exponential term of 1/2 can be explained by counting statistics.

Studies describing a concentration-uncertainty relation characterized by an exponential term approximately equal to 1/2 are rare (see e.g. Hughes and Hurley, 1987). Also the Horwitz function is incompatible with counting statistics given that its exponential term equals 0.85. The UF proposed by Thompson (2000) consists of three functions and the function in the high-concentration regime has an exponential term equal to 1/2, consistent with Poisson counting statistics. However, the fact that this trend is found only in the high-concentration regime contradicts expectations from spectrochemical analyses (Luo et al.,

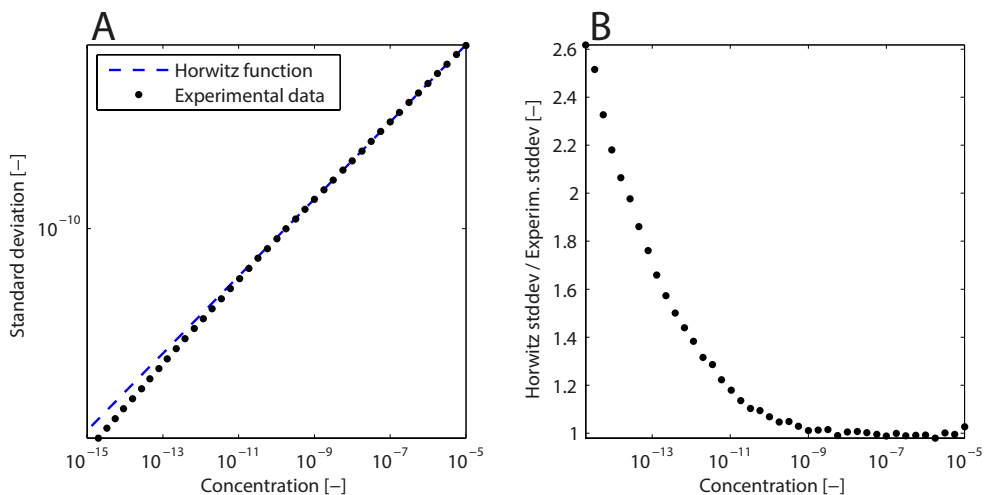


Figure 5.3: Results of the simulation experiment showing the effect of censoring. In A, the dashed blue line represents the theoretical concentration-uncertainty relation (i.e. the Horwitz function) whereas the black dots represent the estimated concentrations and standard deviations. In B, the ratio between the theoretical and estimated standard deviation is shown as function of the concentration.

2006): high concentrations correspond to high intensities for which counting errors are generally negligible. We would expect this exponent in the low-concentration regime.

Thompson and Lowthian (1997) attributed the flattening of the SDR at high concentrations to a relative decrease in analytical performance. This would imply that chemical and spectroscopical research should focus more on improving quantification of high-concentration analytes rather than analytes present in trace amounts. Unfortunately, however, no evidence has been presented for the root cause of this relative flattening towards the high-concentration end.

5.3 UFs and statistical theory

In the previous section we have reviewed the general parametric form of most UFs and we tried to relate them to basic spectrochemical theory. The starting point of this section is to consider the properties of the data we intend to acquire, i.e. concentrations.

5.3.1 Properties of compositional data

Concentrations are compositional quantities. The compositional nature of concentration data leads to a number of properties regarding their uncertainty. To study the implications of this compositional nature for uncertainty modelling, we consider a hypothetical sample which is a mixture of only two components: component *A* and component *B*. Given that a_A and a_B are the amounts of mass in arbitrary units (e.g. mg), the total mass of the sample

is given by $a_A + a_B$. The concentrations of A and B in proportions are given by:

$$\begin{aligned} c_A &= \frac{a_A}{a_A + a_B} \\ c_B &= \frac{a_B}{a_A + a_B} \end{aligned} \rightarrow c_A + c_B = 1 \quad (5.14)$$

This operation is commonly referred to as closure (Aitchison, 1982; Chayes, 1960).

Because amounts of mass are non-negative, the following holds for the D -part vector containing the amounts divided by their sum:

$$\mathbf{c} : \sum_{i=1}^D c_i = 1, \quad 0 \leq c_i \leq 1 \quad (5.15)$$

Hence, concentrations have an upper limit, a lower limit and a sum-constraint. Furthermore, they are intrinsically correlated and not statistically independent (Chayes, 1960; Aitchison, 1982).

5.3.2 Symmetry of uncertainty

The proposition that spectroscopic techniques are capable of measuring concentrations automatically leads to the desired properties of the analytical uncertainty, and the corresponding UFs. To illustrate this, we consider a substance which consists of two analytes (i.e., analyte A and B):

$$\begin{aligned} \hat{c}_A &= c_A \pm \delta c_A \\ \hat{c}_B &= c_B \pm \delta c_B \end{aligned} \quad (5.16)$$

We know that c_A and c_B are compositional quantities and if we assume that the concentration estimates \hat{c}_A and \hat{c}_B are also compositional quantities we obtain:

$$(c_A \pm \delta c_A) + (c_B \pm \delta c_B) = 1; \quad (5.17)$$

If we also consider that $c_A + c_B = 1$, the above expression reduces to:

$$\delta c_A = \delta c_B \quad (5.18)$$

Hence, the uncertainty associated with one component must be equal to the uncertainty associated with the other component. We derived this requirement for a mixture of two components. However, using the notion of *analyte* and *matrix*, any mixture can be expressed in terms of D two-component mixtures defined as c_A and $c_M = 1 - c_A$ with c_A the analyte concentration and c_M the matrix concentration. For an UF to be consistent with theory the following must hold:

$$f(c_A) = f(c_M) \rightarrow f(c) = f(1 - c) \quad c \in [0, 1] \quad (5.19)$$

i.e., the UF must be symmetric around $c = 0.5$.

Because amounts and therefore concentrations are by definition non-negative their associated UF predicts zero standard deviation at zero concentration:

$$f(0) = 0 \quad (5.20)$$

Hence, an empirically-derived UF is consistent with statistical theory if it does not have an intercept term. Combining this requirement with the symmetry requirement (Eqn. 5.19) implies that the predicted uncertainty at the other physical limit must also be zero:

$$f(1) = 0 \quad (5.21)$$

5.3.3 Plausibility of concentration-uncertainty relations

Not only the UF's behaviour at the physical boundaries determines whether concentrations potentially violate physical limits. The concentration-dependent part of the UF controls the domain in which physical plausibility is ensured. This domain is defined as the range in which the probability of violating the physical limits (Eqn. 5.15) is acceptably small. In order to specify this domain, we adopt the convention that uncertainties follow a normal distribution (cf. Thompson and Howarth, 1976).

The probability of violating the physical limits can be calculated for any combination of concentration and uncertainty. By defining a maximum probability of encountering physically implausible results, we are able to derive the Limits Of Physical Plausibility (LOPPs). The lower LOPP is given by:

$$\sigma_{lL} = \frac{c}{\Phi^{-1}(1 - \alpha)} \quad (5.22)$$

where $\Phi^{-1}(1 - \alpha)$ is the inverse cumulative distribution function of the standard normal distribution and $c \in [0, 1]$. Consistent with the principle of symmetry (Eqn. 5.19), the upper LOPP is defined analogously:

$$\sigma_{uL} = 1 - \frac{c}{\Phi^{-1}(1 - \alpha)} \quad (5.23)$$

Note that the lower and upper LOPP are characterised by proportionality between concentration and standard deviation. Together they define a region in concentration-uncertainty space. Any UF which aims at describing the relation between concentration and uncertainty is only valid within this region, which is therefore called the "permissible region" bounded by c_{lL} and c_{uL} . This region and the different UFs are indicated in Figure 5.4. Given that concentration estimates behave according to a particular UF, these estimates cannot be trusted outside the permissible region. Inside this region, the plausibility of an UF can be evaluated based on the symmetry criterion (Eqn. 5.19). The limiting concentrations are obtained by solving the following equations:

$$\begin{aligned} \text{Lower LOPP : } \sigma_{lL} &= f(c_{lL}) \\ \text{Upper LOPP : } \sigma_{uL} &= f(c_{uL}) \end{aligned} \quad (5.24)$$

which, depending on the UF, may be obtained analytically or numerically. For the Horwitz function and the piece-wise alternative of Thompson (2000) the domains read $[10^{-8.88}, 0.9552]$ and $[0, 0.9770]$, respectively.

These results show a fundamental difference between the Horwitz function and the function proposed by Thompson (2000): in contrast to the former, the latter implies that analytical methods are not affected by violations of the physical limits. It is impossible to calculate c_{lL} and c_{uL} for the CF because its function parameters are unspecified. However, a positive intercept term implies that the CF will always cross the lower LOPP.

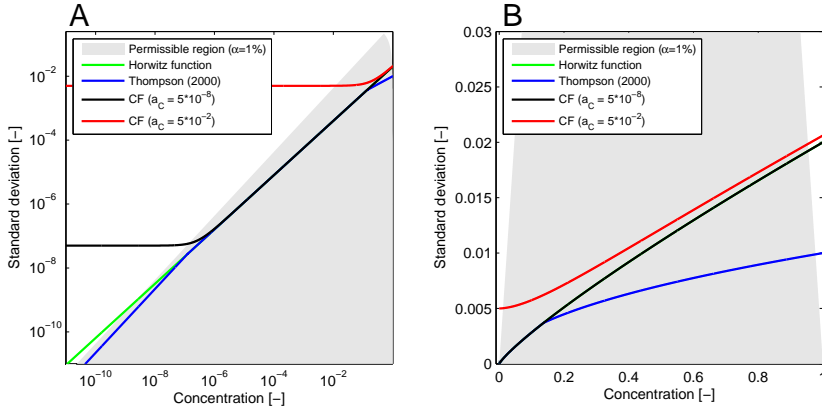


Figure 5.4: Relation between concentration and standard deviation of different UFs on logarithmic (A) and linear axes (B). Both plot also show the permissible region for $\alpha = 1\%$.

5.3.4 Evaluation of existing UFs

From the parametric forms of the proposed UFs it is immediately apparent that they all violate the symmetry criterion and are therefore at odds with the definition of concentrations as mass fractions (see Eqn. 5.14). This inconsistency can have important consequences for statistical inference.

A specific example of statistical inference is performance quantification. A well-known performance measure is the Horwitz Ratio or 'HorRat' (see Horwitz and Albert, 2006, and references therein):

$$HorRat = \frac{\sigma_{emp}}{0.02c^{0.8495}} \quad (5.25)$$

Consider that we wish to quantify the performance of an analytical instrument by analysing a substance which consists of two components, i.e. a low-concentration analyte and a high-concentration matrix. According to theory, the error in $\delta c_A = \delta c_M = \delta c$. Depending on whether we fill in c_A or c_M in Equation 5.25, two different HorRat values are obtained. Only in case $c_A = c_M$ the two values are equal. Hence, the HorRat performance measure is non-unique and relies on a subjective decision about what is considered analyte and what is considered matrix which, in turn, influences the result. This points towards a crucial error in this approach.

Problems related to the asymmetrical nature of UFs have been recognised by Thompson et al. (2008), who observed an inverse relation between moisture concentration and uncertainty which they ascribed to the fact that moisture was indirectly quantified by determination of the amount of dry residue. Horwitz and Albert (2006) also stressed the problems surrounding asymmetry and stated that overestimation of uncertainties by the Horwitz function at high concentrations is due to a "mathematical idiosyncrasy applied to paired values". They also proposed an ad-hoc solution being that "high variability for moisture in foods in the 5-20% region can be transformed to low variability if reported as solids in the 80-95% region". It is evident, however, that this is not a rigorous solution to a fundamental problem.

5.4 Synthesis

We have shown that asymmetrical UFs are inconsistent with the definition of concentrations as mass fractions. Remarkably, a symmetrical nature of UFs was not implied by BUCE, and likewise it did not follow from propagation of Poisson counting errors. This apparent contradiction merits a closer look at the nature of spectroscopic calibration.

5.4.1 Physical and statistical theory compared

The fundamental assumption of BUCE is a unique relation between intensity and concentration (Eqn. 5.8). Given that concentrations are dimensionless by definition, this would require that the units of the proportionality constant K are in units of concentration per units of intensity: something which can be adopted without further consequences because it is a constant with no other purpose than relating I to W . But, a problem arises when considering the sample spaces of I and W : whereas $W \in [0, 1]$, for I it holds that $I \in [0, \infty)$. This is the root cause for the inconsistency noted above.

It is possible to achieve consistency of the spectroscopic calibration model in terms of units and sample spaces. The simplest way to ensure that the sample spaces coincide is to build the model on the following proportionality:

$$A \propto I \quad (5.26)$$

where A_j is the amount of mass of component j . The sample space of these two quantities is identical, namely $[0, \infty)$. In addition, proportionality between the amount and intensity seems more physically tractable given that radiation is emitted or absorbed by matter. This proportionality may be written as an equality by introducing the constant G :

$$A = GI \quad (5.27)$$

G is very similar to K , however, it is expressed in units of mass per units of I (e.g. [cps/g]).

Concentrations are mass fractions so that $C = A/T$, where T is the total amount of mass given by:

$$T = \sum_{j=1}^D A_j \quad (5.28)$$

Using this definition, Equation 5.26 may be expressed in terms of concentrations:

$$\frac{A}{T} = \frac{GI}{T} = \frac{G}{T}I \quad (5.29)$$

The associated Consistent Univariate Calibration Equation (CUCE) reads:

$$C = \frac{G}{T}I \quad (5.30)$$

Note that T is explicitly resolved and therefore a function of all amount estimates A_j , rather than a constant as in BUCE.

5.4.2 BUCE and CUCE compared

BUCE and CUCE are very similar: they only differ with respect to the multiplicative term. The assumption implicitly made in BUCE is that T is constant. Hence the shortcut that:

$$K \approx \frac{G}{T} \quad (5.31)$$

Given that T is indirectly controlled by, e.g., the strength of the radiation source (active or passive) or the size of the irradiated area, this is a reasonable assumption given that K is applied to one analytical setup with constant settings. The above short-cut, however, has important consequences with regard to the uncertainty.

The statistical consequences will be illustrated by means of two examples. In these examples we consider a jar filled with black and white marbles whose composition needs to be determined. The jar contains 90 black marbles and 10 white marbles (i.e. $c_b = 0.9$ and $c_w = 0.1$). It is analyzed with a certain spectroscopic technique. In example 1 we adopt the assumption that T is known ($T = 100$) and constant, i.e. the analogue of BUCE. In example 2, we use the CUCE with $G = 1$. In both cases it is assumed that the dominant source of uncertainty is the counting process. Hence, the intensities are independent Poisson-distributed variables with Poisson parameters $\lambda_b = 90$ and $\lambda_w = 10$, i.e. they behave according to the CF with $b_C = 1$ and $d_C = 1/2$ (see Eqn. 5.3). This corresponds to a situation in which peak areas can be derived by summation over a pre-defined set of channels (i.e. no background and a negligible contribution from fitting to the error budget).

Example 1: T is constant (BUCE)

First, we will discuss the case in which proportionality is assumed between C and I . Since $G = 1$, the estimated amounts are also Poisson-distributed. Their variances are given by:

$$\begin{aligned} \delta^2 \hat{A}_b &= A_b = 90[\text{marbles}] \\ \delta^2 \hat{A}_w &= A_w = 10[\text{marbles}] \end{aligned} \quad (5.32)$$

It is evident that dividing the expected amounts by T (which equals 100) yields the expected concentrations:

$$\begin{aligned} E(\hat{C}_b) &= E\left(\frac{G\hat{A}_b}{T}\right) = 0.9 \\ E(\hat{C}_w) &= E\left(\frac{G\hat{A}_w}{T}\right) = 0.1 \end{aligned} \quad (5.33)$$

Because T is a constant, error propagation yields that for the variance of \hat{C} the following holds:

$$\begin{aligned} \delta^2 \hat{C}_b &= \frac{G\lambda_b}{T^2} = T^{-1}C_b = 9 \times 10^{-3} \\ \delta^2 \hat{C}_w &= \frac{G\lambda_w}{T^2} = T^{-1}C_w = 1 \times 10^{-3} \end{aligned} \quad (5.34)$$

Example 2: T is random (CUCE)

Our best estimate for T is defined as:

$$\hat{T} = \hat{A}_b + \hat{A}_w \quad (5.35)$$

Because the amounts estimated are random variables, so is \hat{T} . The concentrations of the black and white marbles are given by:

$$\begin{aligned} E(\hat{C}_b) &= E\left(\frac{G\hat{A}_b}{\hat{T}}\right) = 0.9 \\ E(\hat{C}_w) &= E\left(\frac{G\hat{A}_w}{\hat{T}}\right) = 0.1 \end{aligned} \quad (5.36)$$

Because the sum of Poisson-distributed variables is also Poisson distributed, \hat{T} is distributed according to $\lambda_T = \lambda_b + \lambda_w$. Error propagation then yields the following variances associated with the concentration estimates:

$$\begin{aligned} \delta^2 \hat{C}_b &= C_b^2 [\lambda_b^{-1} - \lambda_T^{-1}] = 9 \times 10^{-4} \\ \delta^2 \hat{C}_w &= C_w^2 [\lambda_w^{-1} - \lambda_T^{-1}] = 9 \times 10^{-4} \end{aligned} \quad (5.37)$$

These expressions are equivalent to the expression for the multinomial variance:

$$\begin{aligned} \delta^2 \hat{C}_b &= \lambda_T^{-1} C_b (1 - C_b) = 9 \times 10^{-4} \\ \delta^2 \hat{C}_w &= \lambda_T^{-1} C_w (1 - C_w) = 9 \times 10^{-4} \end{aligned} \quad (5.38)$$

Hence, in contrast to example 1, the variances of the concentration estimators are equal for both the black and white marbles.

Discussion of examples

In both examples we obtained a quantity whose expected value corresponds to the true concentration. However, the two cases differ fundamentally with respect to the obtained uncertainties:

$$\begin{aligned} \text{Example 1 : } \delta^2 \hat{C} &= T^{-1} C \\ \text{Example 2 : } \delta^2 \hat{C} &= \hat{T}^{-1} C (1 - C) \end{aligned} \quad (5.39)$$

In example 1, the uncertainty is asymmetric with respect to the concentration. The uncertainty in example 2, however, behaves in a symmetric manner because the expression contains the structural element $C(1 - C)$.

Provided that T has been properly estimated, these examples demonstrate that conventional approaches to spectroscopic calibration can potentially yield unbiased estimates of concentrations. Although unbiased, example 1 showed that BUCE yields counting errors which are incompatible with the nature of compositional quantities.

5.4.3 Re-establish consistency by means of closure?

In the calibration mode, BUCE (example 1) would require estimation of the analyte-specific constants K associated with the instrument and settings. This can be achieved using data from "pure" substances (i.e. one bowl with only black and one bowl with only white marbles) and set K_W and K_B such that $K_B I_B = K_W I_W = 1$. In the prediction mode, BUCE now allows for independent prediction of marble concentrations whereby T is implicitly contained in the constants K_W and K_B : using a larger bowl would require different K -values.

For CUCE (example 2), calibration would again involve analysis of "pure" substances but now the total amount of spectroscopically-active mass must be known (i.e. density

times the irradiated area times the radiation penetration depth). In a similar manner as BUCE, this then allows estimation of G_W and G_B . Prediction of future samples yields amounts which turn into concentrations after dividing the estimated amounts by their sum. In compositional data analysis theory this last step is referred to as *closure* (Aitchison, 1986).

Although CUCE is mathematically and physically more tractable than BUCE, CUCE has limited practical value because it is impossible to estimate T . Closure would impose the desired (symmetric) UF on concentration estimates. Using BUCE and subsequently applying closure therefore seems a potentially powerful and simple solution. However, the problem is that it relies on complete quantification, i.e. that all components in the mixture could be quantified with the analytical method used. For incomplete quantification, closure will induce bias, because the mass fractions of components which could not be identified is eliminated from the composition. Closure of incompletely quantified substances thus results in a subcomposition (Aitchison, 1986).

5.4.4 Mimicking closure

Concentrations are intrinsically correlated because they have an upper bound and it is this correlation which results in symmetric uncertainties. The fact that the dependency causes symmetry means that in principle any kind of data processing operation which introduces dependency can make the errors symmetric.

One process which may introduce dependency is multivariate calibration whereby an optimal estimate of the full composition is postulated on the basis of all measured intensities. Another relevant and closely related operation is matrix correction. Because matrix corrections induce changes to the concentration of an analyte based on the concentration of the matrix (i.e. the other components), they introduce correlation by definition. Hence, we hypothesize that carefully executed matrix corrections and/or multivariate calibration methods acknowledging the compositional nature of concentrations are indistinguishable from closure (Eqn. 5.15). In a strictly statistical context, BUCE does not yield concentrations but quantities whose first moment may be indistinguishable from that of concentrations, whereas their second moment is not representative for the "true" uncertainty. We will refer to such quantities as *pseudo concentrations*.

The extent to which these processes resemble closure is likely to be method- and analyte-specific. This means that uncertainties associated with pseudo concentrations can range from fully asymmetric to fully symmetric and that for every analytical setup, empirical evidence is indispensable.

5.4.5 Symmetry in empirical uncertainty data

The asymmetric nature of existing UFs appears to be related to their lack of fit, e.g., the overestimation of the analytical uncertainty by the Horwitz function for $c > 0.1$. The function proposed by Thompson (2000) aims to correct this bias by using a piece-wise exponential function. We interpret this as an attempt to capture the symmetrical behaviour of the analytical uncertainty which became apparent through evaluation of high-concentration data.

In order to understand the origin of this bias for $c > 0.1$ we consider the limiting property of the structural element $c(1 - c)$ which marks the difference between BUCE and

CUCE:

$$\lim_{c \rightarrow 0} c(1 - c) = c \quad (5.40)$$

The bias introduced by approximating $c(1 - c)$ by c is equal to c (i.e. a 10% error for $c = 0.1$). Hence, the symmetrical nature of the analytical uncertainty only becomes apparent for concentrations higher than 10% which manifests itself as a relative overestimation. In the past decades, however, analytical efforts have mostly focused on quantification of low-concentration analytes. Only on rare occasions concentrations exceed 10% (e.g. sugar analysis; see Horwitz and Albert (2006)). Likewise, the symmetrical nature of the analytical uncertainty was either considered a problem of minor importance Horwitz and Albert (e.g., 2006) or it was simply not recognised. The common way of plotting concentrations and errors on a log-log scale has certainly not contributed to the recognition of symmetry because the concentrations higher than 10% are squeezed in the right corner of an concentration-error plot (see Fig. 5.5).

These practical explanations are symptoms of a more fundamental problem, which is the "analyte-matrix perspective" on chemical analysis. With this perspective we refer to the univariate perspective on the problem of determining the concentration of low-concentration analytes in a high-concentration matrix. In the light of BUCE and counting statistics, determining low-concentration analytes seems indeed to be the most challenging because the number of counted objects decreases with concentrations. In the light of CUCE, however, determining the analyte is as challenging as determining the total matrix. We refer to this as the "compositional perspective": the corresponding task is not to determine individual concentrations, but to determine the composition of the substance as a whole. In that light, we encourage the analytical chemistry community to test the performance of analytical setups on both sides of the concentration spectrum.

5.5 Modelling framework

5.5.1 Definition of UFs

In this section we propose a modelling framework for assessing system performance and turning pseudo concentrations into proper concentrations by specifying statistically and physically meaningful UFs for every possible setup.

In this modelling framework we distinguish (1) the function which best captures the behaviour of the analytical system (i.e. the uncertainty associated with pseudo concentrations) and (2) the corresponding symmetrical UF suitable for statistical inference. The former is referred to as the *pseudo* UF whereas the latter is called the *proper* UF.

We propose a three-stage modelling approach. The first stage consists of system characterization, i.e. fitting a *pseudo* UF. The second stage consists of transforming the *pseudo* UF into a *proper* UF. In the third stage, the plausible range of the function is determined.

5.5.2 Modelling stage 1: System characterization

Pseudo UF candidates

From physics we derived that the following function is suitable for modelling the relation between concentration and analytical uncertainty:

$$\text{PUF}(a_P, b_P, d_P; \tilde{c}) \quad : \quad \delta c = \sqrt{a_P^2 + [b_P \tilde{c}^{d_P}]^2} \quad (5.41)$$

which we refer to as the Poisson-like UF (PUF). Under the influence of closure, matrix corrections or joint calibration, the following function is applicable:

$$\text{BUF}(a_B, b_B, d_B; \tilde{c}) \quad : \quad \delta c = \sqrt{a_B^2 + [b_B (\tilde{c}(1 - \tilde{c}))^{d_B}]^2} \quad (5.42)$$

which we refer to as the Binomial-like UF (BUF). If the analytical setup is characterised by a BUF, the observed pseudo concentrations are indistinguishable from proper concentrations.

Model selection

Model selection involves fitting PUF and BUF to experimentally-derived uncertainties. We demonstrated that for concentrations less than a few percent, BUF and PUF overlap (Eqn. 5.40). For higher concentrations, the two functions deviate significantly. We make use of this fact by partitioning the data into two subsets (set 1: $c < 0.1$, set 2: $c \geq 0.1$). Set 1 is used for function fitting whereas set 2 is used for distinguishing between BUF and PUF.

PUF and BUF are fitted to set 1 using non-linear regression based on the following criterion:

$$\text{argmin}_{a,b,d} \sum_{i=1}^N [\log(\sigma_i) - \log(f(a, b, d; \tilde{c}_i))]^2 \quad (5.43)$$

Because BUF and PUF have the same number of parameters, the UF which gives the smallest residual sum of squares for set 2 defines the optimal pseudo UF. Subsequently, this function is fitted to the data set as a whole which yields function parameter estimates with associated confidence intervals. These confidence intervals allow us to determine whether the intercept a is effectively zero. In that case, the pseudo UF without an intercept is refitted in order to yield a more parsimonious function with essentially the same goodness-of-fit. We perform this test because an intercept will not contribute to a better fit in case the data shows a relative steepening instead of flattening at the low-concentration end.

5.5.3 Modelling stage 2: Pseudo concentrations to concentrations

For inference we require a symmetric UF. In case BUF was regarded as the pseudo UF, this function can be readily used for inference. Hence, the fitted BUF is both the pseudo and the proper UF:

$$\begin{aligned} \text{Pseudo UF} & \quad \rightarrow \quad \text{Proper UF} \\ \text{BUF}(a_B, b_B, d_B; \tilde{c}) & \quad \rightarrow \quad \text{BUF}(a_B, b_B, d_B; \tilde{c}) \end{aligned} \quad (5.44)$$

Otherwise, the proper UF needs to be postulated from the pseudo UF. Given that closure defines concentrations, the proper UF is defined as the UF that would be obtained if the data were closed.

Explicit expression for the proper UF

In Appendix C we derived an expression for the uncertainty after closure (Eqn. C.1.2). This expression shows that upon closure, an intercept term in PUF becomes a concentration-dependent term. However, data processing in spectroscopy (i.e. background subtraction) gives rise to an additive component in the UF that is independent of the concentration of the other elements in the material analyzed.

In order to find a compromise between the physical reality of the data acquisition process and statistical theory, we propose the following expression for the proper analytical uncertainty $\delta^2 c_j$:

$$\delta^2 c_j = a_P^2 + \left[(1 - 2\tilde{c}_j)(b_P \tilde{c}_j^{d_P})^2 + \tilde{c}_j^2 \sum_{k=1}^D (b_P \tilde{c}_k^{d_P})^2 \right] \quad (5.45)$$

where D is the number of measured components in the mixture and the terms enclosed in brackets follow from error propagation. It rightfully shows that the uncertainties of proper concentrations are correlated among the variables.

A compact approximation to the proper UF

Our objective is to find a compact and univariate approximation to Equation 5.45. Hereby we try to stay as close as possible to current practice which is to predict the analytical uncertainty in a univariate manner. Our strategy is to construct a univariate modelling framework which enables the transfer from PUF into BUF.

In case the UF is given by a PUF with $a_P = 0$, $d_P = 1/2$ and arbitrary b_P , the UF corresponding to the data after closure is given by BUF with exactly the same function parameters, i.e. $\text{BUF}(0, b_P, 1/2; \tilde{c})$. In all other cases, this approach yields inexact and biased predictions of uncertainty. A proof is given in Appendix C.

Using the principle of error propagation, we developed a workflow to approximate the expression shown in Equation 5.45 by a BUF. It is an approximation because BUF is a univariate function whereas the exact solution shown in Equation 5.45 requires the concentration of all analytes. We found that irrespective of d_P , the intercept a_P and exponential term d_P can be transferred directly from PUF into BUF without introducing bias. For the multiplicative term (b_P) a correction factor k is required so that the relation between the fitted PUF and the proper BUF reads:

$$\begin{aligned} \text{Pseudo UF} & \rightarrow \text{Proper UF} \\ \text{PUF}(a_P, b_P, d_P; \tilde{c}) & \rightarrow \text{BUF}(a_P, kb_P, d_P; \tilde{c}) \end{aligned} \quad (5.46)$$

The details of, and mathematical justification for this approach are given in Appendix C.

Correction factor k

An explicit expression for k is obtained by adopting some "standard composition". An obvious choice for this composition is:

$$c_j = D^{-1} \quad \forall \quad j = 1, 2, \dots, D \quad (5.47)$$

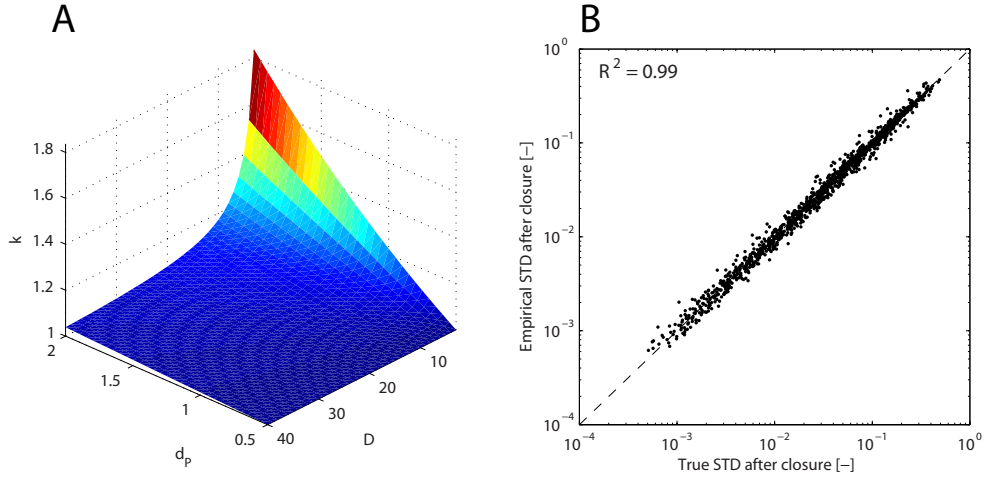


Figure 5.5: (A) Multiplicative correction factor k corresponding to the proposed "standard composition" as function of D (i.e. the number of components in the mixture) and d_P (i.e. the exponential term in PUF). (B) The true error in PUF data after closure against the univariate approximation proposed in this contribution. The data points reflect the errors of concentrations for which d_P ranges between $1/2$ and 2 and D ranges between 4 and 40 .

i.e. all analytes are present in the same concentration. The corresponding correction factor reads:

$$k = \sqrt{\frac{D^{-2d_P} - 2D^{-2d_P-1} + D^{-2d_P+2}}{[D^{-1}(1 - D^{-1})]^{2d_P}}} \quad (5.48)$$

As demonstrated by the equation, k is a function of the number of measured components D and the exponential term d_P . This function correctly yields $k = 1$ for $d_P = 1/2$, i.e. for an UF with $d_P = 1/2$ $b_P = b_B$. The behaviour of k as function of D and d_P is shown in Figure 5.5A. Clearly, k is largest for small D and large d_P . Furthermore, for $d_P = 1/2$, k is equal to unity for all D : this reflects the Poissonian case. For large D and relatively low spread among the concentrations it is therefore reasonable to take $k = 1$. As a guideline we propose to take $k = 1$ in case $D \geq 30$.

In Figure 5.5B, the relation between true and predicted standard deviation are plotted for data behaving according to BUF. The concentration were drawn from a uniform distribution after which the concentrations were closed to unit sum. The displayed data reflect realisations for which the exponential term d_P ranges between $1/2$ and 2 and D ranges between 3 and 40 . Note the unbiasedness of the prediction and the high R^2 (0.99). It is evident that in specific applications where the compositional variability is small (e.g. process monitoring applications) a different "standard composition" may be adopted to increase the quality of predictions.

5.5.4 Domain of the UF

A complete definition of a proper UF requires specification of its limits of application. The domain of the proper UF follows by solving the equality shown in Equation 5.24. This yields c_{LL} and because the proper UF is symmetric, it holds that $c_{uL} = 1 - c_{LL}$. After this final step, the proper UF has been fully specified for a given significance level α (e.g. 5%).

5.5.5 Measuring performance

The proper UF may be employed to measure the performance of an analytical setup. The Horwitz function is only valid for low-concentration domains and it can therefore not serve as a basis for quantifying the performance of analytical systems designed to measure major components.

We may take the Horwitz function and turn it into a proper UF:

$$BHF(c) : \sigma = 0.02 [c(1 - c)]^{0.8495} \quad 10^{-8.88} \leq c \leq 1 - 10^{-8.88} \quad (5.49)$$

where "BHF" stands for the Binomial Horwitz Function (BHF). It has a much larger domain than the Horwitz function (i.e., $[10^{-8.88}, 0.9552]$) which means that concentrations well above 0.955 can be trusted. Based on this proper UF, we derive the following alternative to the HorRat, referred to as the Binomial HorRat (BHR):

$$BHR = \text{med}_{ij} \left\{ \frac{|\hat{c}_{ij} - c_{ij}|}{BHF(c_{ij})} \right\} \quad 10^{-8.88} \leq c \leq 1 - 10^{-8.88} \quad (5.50)$$

In this expression, \hat{c} is the estimated concentration of analyte j in specimen i by a particular laboratory and c is the assigned value (e.g. the consensus value or the concentration associated with a particular Certified Reference Material). The operator $\text{med}_{ij} \{ \dots \}$ defines the median over all specimens i and all analytes j . A *BHR* close to unity means that the laboratory-instrument combination performs as expected. Values much larger than unity would merit further investigation of the data integrity.

5.6 Application

5.6.1 GeoPT data set

Our final objective is to apply the proposed modelling framework to real data. A data set suitable for this purpose should have a wide range of concentration values. Data from the GeoPT proficiency testing programme fulfill this requirement. This programme was initiated and executed by the International Association of Geoanalysts (IAG) between 2001 and 2011 with the purpose to allow geochemical laboratories to evaluate their performance in routine analysis of silicate rocks (Thompson et al., 1997; Thompson, 2002). In addition to a range of trace elements, the analyzed silicate rocks contain high concentrations of major elements (e.g. Si, Ca, Fe) (Potts et al., 2013).

In every GeoPT round, identical samples were sent to numerous laboratories for chemical analysis. Every GeoPT round thus yields standard errors and consensus values of a series of analytes present in particular specimen. The consensus value and spread were calculated using different measures of central tendency or spread (Thompson, 2006). In rounds 8 to 22, the consensus values reflected the robust mean (Lawn et al., 1993) and from

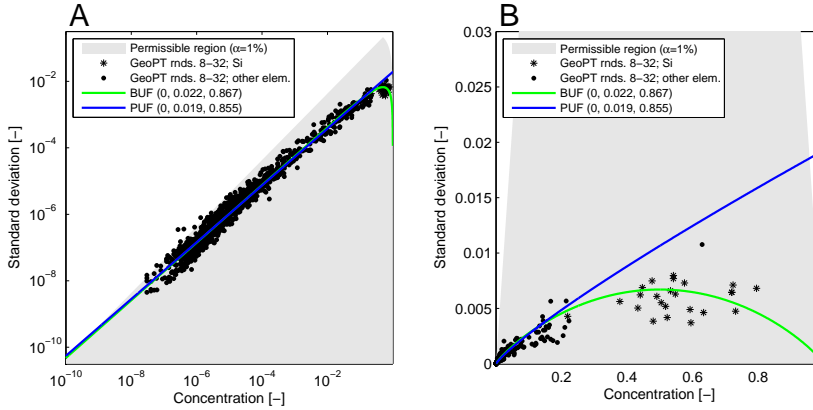


Figure 5.6: Modelling results for the GeoPT data displayed on (A) logarithmic and (B) linear scales.

round 23 onwards the consensus values were derived from the robust mean, the median or the mode (Thompson, 2006). The choice for the respective method was made by the Steering Committee (Thompson, 2002).

We included the consensus values and empirical standard deviations of 25 rounds in our analysis (i.e. GeoPT rounds 8 to 32). The raw data of rounds 1 to 7 could not be recovered. References to unpublished reports are provided in Appendix C2.

5.6.2 GeoPT Results

It was found that BUF shows the best fit to the GeoPT data. It was also found that the intercept term a_B was effectively zero. The pseudo UF of this data set is given by $\text{BUF}(0, 0.022, 0.867)$, which is displayed in Figure 5.6. Figure 5.6 also displays the fitted PUF for comparison. The calculated 95% confidence intervals around the BUF parameters read: $0.021 \leq b_B \leq 0.024$ and $0.860 \leq b_B \leq 0.875$. The same confidence intervals around PUF read: $0.018 \leq b_B \leq 0.021$ and $0.847 \leq b_B \leq 0.862$.

Because BUF is symmetric, it is also the proper UF. The domain of this function is given by $[10^{-9.68}, 1 - 10^{-9.68}]$. Given that the GeoPT data are representative for geochemical analysis, this would lead to the following performance measure for geochemical laboratories worldwide:

$$GPR = \text{med}_{ij} \left\{ \frac{|\hat{c}_{ij} - \bar{c}_{ij}|}{0.022[\bar{c}_{ij}(1 - \bar{c}_{ij})]^{0.867}} \right\} \quad 10^{-9.68} \leq c \leq 1 - 10^{-9.68} \quad (5.51)$$

where 'GPR' stands for Geochemical Performance Ratio. Analogous to the HorRat, $GPR = 1$ means that a laboratory performs as expected, whereas $GPR \geq 2$ means that the laboratory performance is unacceptably low.

5.6.3 Discussion of GeoPT results

Because the intercept is effectively zero and BUF is the pseudo UF, the GeoPT pseudo concentrations are indistinguishable from proper concentrations. This may be due to closure

or successful application of matrix corrections.

Note that the function parameters of BUF and PUF are approximately equal whereas the functions themselves have a very different appearance. Also note that for PUF, the 95% confidence intervals of the function parameter include those of the Horwitz function (i.e. $b_P = 0.02$ and $d_P = 0.8495$). Hence, the quality of geochemical analyses is comparable to analyses carried out in other fields of application. The BUF parameters, on the other hand do not fall inside the 95% confidence interval which means that the fitted parameters are significantly different from those of the Horwitz function. This shows that in cases where the data may display symmetric behaviour that went unnoticed, re-evaluation of data is necessary in order to obtain accurate pseudo UFs.

Practically all high-concentration data are associated with Si. As suggested by the deviation of the single data point around 60% concentration from the fitted BUF, matrix corrections for Ca seem to be less successful. The fact that the single data point which deviates from BUF is the only high-concentration data point which is not associated with Si suggests that the extent to which matrix corrections impose symmetry differs between analytes. Hence, analyte- and method-specific UFs may be deemed necessary.

5.7 Discussion and conclusions

We have identified and diagnosed inconsistencies between the parametric form of UFs proposed in literature and statistical theory. Derivation of the behaviour of the analytical uncertainty yielded that for strictly univariate calibration without matrix corrections, the analytical uncertainty may behave in a fully asymmetrical manner which is unrealistic. Ideally, matrix corrections or appropriate multivariate calibration will compensate for this flaw and make the concentration estimates produced by an analytical instrument fully symmetric and therefore indistinguishable from proper concentrations. In practical applications, however, the analytical uncertainty can behave anywhere between fully asymmetric and fully symmetric.

The covariance structure of the uncertainties is a potentially important indicator for the extent to which a certain matrix correction or multivariate calibration process is able to resemble the effect of closure. Hence, we encourage manufacturers of calibration software to analyse the nature of the empirically-derived standard deviation under various levels of precision and over a wide range of concentrations, possibly stratified by analyte. Given that compositional variability in natural samples is generally limited, the only way to achieve this would be to study synthetic mixtures tailored to this application. Furthermore, we recommend users to be careful when employing uncertainties reported by software packages in their analysis. Only when it is empirically demonstrated that these calculated uncertainties behave in a symmetric manner can they be readily used for inference.

Application of the modelling framework is not limited to the analytical uncertainty. Given that the common objective of an end user is to characterize a population by analysing a finite sample, the total uncertainty is given by the sum of analytical uncertainty and sampling uncertainty. It is well-known that the latter behaves in a fully symmetric manner (e.g., Gy, 1979). This means that in case the modelling framework is employed to produce a proper symmetrical UF predicting the analytical uncertainty, the parametric form of this function will be similar to the one that predicts the sampling uncertainty. In that sense, the proposed modelling approach does not only establish the link between physical and

statistical theory, but also between analytical and sampling uncertainty.

The results presented in this contribution have important consequences for performance quantification. Particularly the replacement of the HorRat by its symmetric equivalent is vital because this yields a unique and unambiguous criterion. Hence, we urge regulatory bodies and professional organizations (e.g. ISO, AOAC, EU) to consider adopting the proposed modelling framework in their guidelines/legislation.

Integrated core analysis using XRF scanning. Part I: Prediction of lithofacies

6.1 Introduction

Routine Core Analysis (RCA) includes measurements of porosity, grain density and horizontal permeability at constant intervals (typically 1 ft.), as well as a lithological description. Generally, these data are complemented with analyses obtained with more specialised techniques. In the hydrocarbon industry, for instance, it is commonplace to perform thin-section analysis to gain insight into controls on reservoir quality such as diagenetic history and provenance. Ideally, all core data are available at the same, high resolution. In reality, however, they have different coverage and resolution which severely complicates quantitative core characterisation.

In this study, we investigate whether X-ray Fluorescence Core Scanning (XRF-CS) may be used to obtain *integrated core descriptions*, i.e., a collection of lithofacies, chemical and petrophysical records which have the same, high (1 cm) resolution. XRF is a chemical analysis technique. It is well-known that the chemical composition of sediments and sedimentary rocks is strongly correlated with their lithology and petrophysical properties. Integrated core descriptions using XRF-CS rely on exploitation of the correlation between geochemical composition and other sediment properties by multivariate statistical methods to link the high resolution XRF-CS data to calibration data (e.g., assigned lithofacies, porosity, permeability) and estimate the associated prediction uncertainty using cross validation.

This chapter is the first of three parts. In this chapter we use the XRF-CS data to predict lithofacies. In Chapter 7 we study the ability to predict the bulk chemical composition and the petrophysical properties from XRF-CS data which together represent the *integrated core descriptions*. In Chapter 8, we demonstrate the added value of integrated core descriptions for unraveling the controls on reservoir quality. We refer to this new way of core characterisation as *integrated core analysis*.

6.2 Materials and Methods

6.2.1 Geological setting and sedimentological interpretation

Well E10-3 (Fig. 6.1, A) is a vertical well which was drilled in the Southern North Sea from an offshore platform in 2003. Its development was commissioned by Wintershall Noordzee BV and it targeted the Westphalian deposits. In Northern and Central Europe, Westphalian deposits are an important source of natural gas (Lokhorst, 1997). In the south, the Westphalian basin is bounded by the Variscan front. In the north, the Westphalian basin is limited by the Mid North Sea High. Sediments were deposited in the basin by meandering and braided river systems (Lokhorst, 1997).

The core retrieved from well E10-3 covers the interval between 3648.00 and 3720.38 m depth. The most conspicuous features are three sandstone intervals of around 10 m thickness. These channel deposits represent the most important reservoir units in the cored interval. The intervals were interpreted as Braided Channel Units (BCUs) and they are embedded in thick mudstone deposits interpreted as inter-distributary fines. Another feature of this core is that it contains numerous coal-bearing layers. It is this wide variety of lithologies which makes this core suitable for studying the quantitative performance and added value of XRF core scanning technology for routine core analysis.

Well B38D4316 (Fig. 6.1, B) is located in the Dutch province of South Holland; a core was retrieved in 2010 using geoprobing. The core comprises sediments from the Quaternary. The well forms part of a larger project aimed at characterisation of the Dutch shallow subsurface, conducted by the Geological Survey of the Netherlands. Modern-day sedimentation in the study area is strongly influenced by the Waal, Lek and Meuse river transporting coarse-grained material. In this area there are also extensive marshlands where the Waal and Meuse river flow into the "Hollands Diep" estuary. Hence, the core retrieved from well B38D reflects a wide variety of sediment types: it contains clay, coarse-sand and loam intervals as well as peat. The core has a total length of 35 m and its dominant features are two peat layers which bound an extensive interval of coarse-grained sand. In terms of lithology, this core can be considered a modern analogue of core E10-3.

6.2.2 Core analysis

Core E10-3 has been analysed by Panterra Geoconsultants in September 2002 (Bakker and Pruno, 2004). In a later study also conducted by Panterra Geoconsultants (Boels, 2003), the sedimentology, petrography and reservoir quality of core E10-3 were analysed. Both reports are publicly available online at the Netherlands Oil and Gas Portal (<http://www.nlog.nl>). Routine core analysis of core B38D has been conducted by TNO, the Geological Survey of The Netherlands.

Core E10-3 comprises Carboniferous sedimentary rock, and was described in terms of a lithofacies scheme proposed by Reijers et al. (1993). The facies architecture of the cored interval in Well E10-3 comprises the following classes: poorly drained floodplain (*IFL*), floodplain (*F*), swamp (*SW*), crevasse splay (*CS*), inter-distributary bay (*IB*) and braided channel (*BC3*) deposits. Bedboundaries in *BC3* sporadically occur as erosive or scoured, with lags of pebbly sandstones. The grain size in this core ranges from clay (i.e., mudstones and coal beds) to gravel. The dominant grain size (D50) along-core was determined in terms of a 1/2 phi unit scheme common in siliciclastic reservoir characterisation: "*S1I*" reflects

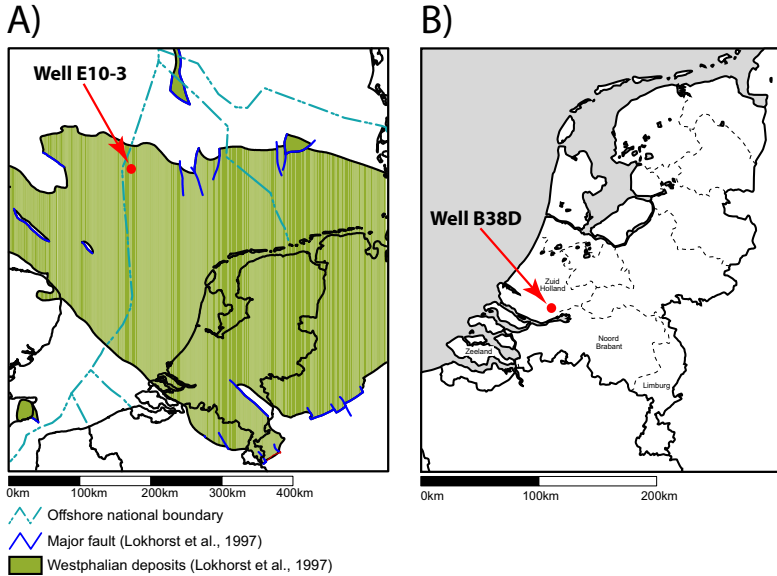


Figure 6.1: The location of well E10-3 (A) and well B38D (B). The map in (A) is modified from Lokhorst (1997).

very fine sandstone ($\phi = 3.75$), and "S5l" reflects very coarse sandstone ($\phi = -0.25$). In addition to sandstone, core E10-3 contains coal beds ("O") and siltstone intervals ("J"). More details about core E10-3 may be found in Appendix D.

In core B38D, eight different lithofacies classes were recognised including 3 types of swamp/peat, 3 types of sand-dominated facies, 2 types of clay/loam. These classes are: eutrophic swamp (FOVEU), oligotrophic swamp (FOVOL), swamp (FOVAM), bank and crevasse (FFL), channel (FFG and FFZ), floodplain (FFKMA), abandoned channel (FFLKA).

6.2.3 XRF core scanning

Both cores have been scanned using an Avaatech X-ray Fluorescence core scanner at the Royal Netherlands Institute for Sea Research (NIOZ). The downcore step size was for both cores set at 1 cm. Core B38D was scanned along the central axis with a cross-core slit size of 1 cm. Because core E10-3 was plugged prior to slabbing, the scan track of core E10-3 was positioned 23 mm from the central axis to avoid the plug holes. Furthermore, the cross-core slit size was set to 8 mm to prevent the scanner from measuring air and ensure reliable measurements. To cover a large suite of elements, both cores were scanned with different instrument settings. The core-scanner settings are summarised in Table 6.1.

To facilitate estimation of the precision, replicate analyses were carried out. These replicates were acquired at fixed intervals. In core E10-3, the average interval equals 30 cm with a total number of 247 replicate sets each consisting of three consecutive measurements. In core B38D, 275 replicate sets were acquired: the average interval in this core equals 10 cm. After data acquisition, the XRF spectra have been processed using the WinAxil processing software to obtain element intensities (see e.g., Vekemans et al., 1994). The integrity of the

Core Run	E10-3	E10-3	B38D	B38D	B38D
	1	2	1	2	3
Measurement time [s]	12	12	10	20	60
Tube voltage [kV]	10	30	10	30	50
Tube current [μ A]	750	750	1000	1000	1000
Filter	none	Pd-thin	none	Pd-thin	Cu
Elements	Al,Ca,Cr,Fe K,Mn,S,Si,Ti	Br,Cu,Ga,Ni Rb,Sr,Y,Zn,Zr	Al,Ca,Cl,Fe,K Mn,Rh,S,Si,Ti	Br,Rb, Sr,Zr	Ag, Ba

Table 6.1: Core-scanner settings used in this study.

element intensities was verified by visual inspection of the peak fit from spectra associated with the lithological extremes.

In addition to XRF measurements, line-scan images were acquired. For E10-3, this was done with a JAI-107 line scan camera which yielded images with an effective pixel size of 0.07×0.07 mm. The exposure time and aperture were set at 143 ms and f11, respectively. Core B38D have been analysed with a JAI-105 camera with an effective pixel size of 0.09×0.09 mm and an exposure time and aperture of 154 ms and f8, respectively.

6.2.4 Integrity of XRF-CS data

Fluorescent radiation from light elements interacts with air. Hence, in case the fluorescent radiation traveled through a thick layer of air, the data are likely to exhibit systematic deviations. Air was measured in case the core was fractured or in case a data point was acquired at the edge of a core section. Therefore, these data points were removed prior to further analysis.

Core B38D constitutes a continuous sedimentary record which was scanned prior to sampling. Hence, we only had to remove data points acquired at the margins of the core sections, which was done manually. The core scanning data set before and after quality checking consists of 3650 and 3462 measurements per run. The total number of unique core positions in the quality-checked data set is 2875.

Core E10-3, however, contains a considerable number of fractures which formed during (plug) drilling. Furthermore, core E10-3 was plugged prior to core scanning: although the scan track was narrow and avoided the plug holes, some measurements at the plug positions were unreliable because of suboptimal alignment between scanner and core slab. In core E10-3, detection of these potential outliers was based on the total count rate associated with run 2 (i.e. the 30kV run): in case the total count rate at a certain scan position was below a threshold value, all data acquired at this position were omitted. A threshold value of 10^5 counts per second gave the best results: it resulted in the removal of data points at fractures and slab edges only. Figure 6.2A shows the total count rate of run 2 for the entire core with the data points whose total count rate is below this threshold value indicated in red. The core scanning data set before and after quality checking consists of 7200 and 7005 measurements, respectively. The total number of unique core positions in the quality-controlled data set is 6511.

Figure 6.2B shows the total count rate for six example slabs of core E10-3: these slabs were selected because they contain all lithofacies (sandstone, siltstone, mudstone, coal)

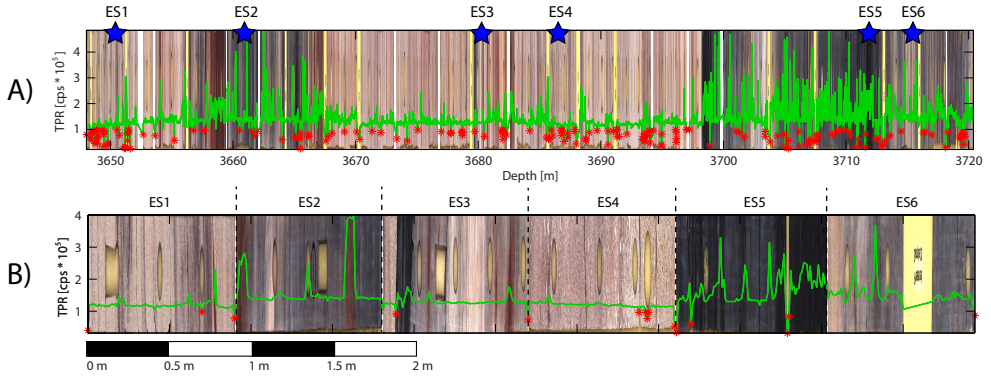


Figure 6.2: Total count rate associated with run 2 (i.e. 30kV scan) of the complete core E10-3 (A) and for ES1-ES6 (B). Data points removed in the quality-checking process are indicated in red.

present in this core. The six example slabs (ES1-ES6) include three slabs from core 1 (boxes 3, 14 and 38) and three slabs from core 2 (boxes 3, 33 and 38) and they are used throughout this study to illustrate detailed analysis of the modelling results. In this case, the example slabs clearly demonstrate that most excluded scan positions reflect edges of the slab or an abrupt lithological transition which was often associated with a small gap.

6.3 Modelling methods

6.3.1 Discriminant-function analysis

Predicting class membership from numerical data is known as Discriminant Function Analysis (DFA). To perform DFA, the core descriptions, which are given by an assigned class between two bed boundaries, are transformed into a vector containing the assigned class at every scanned core position. Next, a prior probability is assigned to each class. In this study an empirical prior is adopted, i.e., the prior probability is given by the proportion of each class in the calibration data. After determining the prior, the discriminant function must be derived. We employ Quadratic Discriminant Function Analysis (QDFA) which is based on stratified mean and covariance estimation, i.e., the mean XRF-CS signature and associated covariance matrix are calculated for every class individually. Combining these means and covariances with the prior yields the quadratic discriminant function given by:

$$\delta_k(\mathbf{x}_i) = \frac{1}{2} \log \|\Sigma_k\| - \frac{1}{2} (\mathbf{x}_i - \mathbf{m}_k)' \Sigma_k^{-1} (\mathbf{x}_i - \mathbf{m}_k) + \log(\pi_k) \quad (6.1)$$

where \mathbf{x} is a vector with the properly transformed element intensities and \mathbf{m}_k are the mean transformed element intensities for class k . The prior associated with class k is denoted by π_k . The predicted class associated with any (new) XRF-CS signature is given by:

$$\hat{C}(\mathbf{x}) = \arg \max_k \delta_k(\mathbf{x}) \quad (6.2)$$

i.e., the class for which the discriminant function value is largest.

6.3.2 Predictive performance

Leave-one-out cross-validation (LOOCV) is used to make quasi-independent predictions for every position. These LOOCV predictions enable determination of the model's predictive power, i.e., its ability to predict the proper class for "new" data (Geisser, 1993). We quantify this predictive power by calculating the True Positive Ratio TPR for every class k :

$$\text{TPR}_k = 100\% \times \frac{\text{count}(\hat{C} == k)}{\text{count}(C == k)} \quad (6.3)$$

where \hat{C} is the predicted class by LOOCV, and C is the class given by the core description. The operator $\text{count}(\hat{C} == k)$ yields the integer number of occurrences that \hat{C} equals k . A estimate for the global prediction error of the model is given by the Mean Relative Prediction Error (MRPE):

$$\text{MRPE}(\%) = 100\% \times \sum_{k=1}^K (1 - \text{TPR}_k) \frac{\text{count}(C == k)}{\text{count}(C)} \quad (6.4)$$

where K is the total number of classes. Note that the MRPE is essentially the mean prediction error over all classes weighted by their prior. We also calculate a confusion matrix which summarizes the relation between "true" class (rows) and the predicted class by LOOCV (columns).

6.3.3 Dendrogram and biplot analysis

To gain insight into the reason behind the (in)ability to predict the "true" class, use is made of two statistical tools.

The first tool is the dendrogram. Making a dendrogram includes calculating the pairwise Euclidean distances between the different class means in terms of the clr-transformed and properly scaled intensities. From these distances a linkage tree may be constructed by adopting a linkage function. In this study that is single linkage, i.e., the two (composite) classes with the smallest distance are merged. Finally, we display this linkage tree using a dendrogram. Classes which are contained in the same branch in the dendrogram are relatively similar whereas classes which are in different branches at the same level are relatively dissimilar.

The second tool we employ is the (compositional) biplot. A biplot represents a visual representation of the outcome of principal components analysis (PCA), i.e., scores and loadings of the covariance matrix eigenvectors. The horizontal and vertical axes of a biplot represent the first and second principal component and therefore the principal directions of variability. The variable vectors show their importance relative to the principal components, and relative to each other. In this study we employ biplots of both geochemical compositions as well as element intensities. In biplots derived from compositional data such as geochemical compositions, the covariance matrix of the clr-transformed data may be used (Aitchison and Greenacre, 2002). Intensities, on the other hand, are not compositional quantities because they still require calibration. According to Weltje and Tjallingii (2008), intensities are a linear transformation of concentrations in the additive logratio space. In this study, a similar assumption is adopted for the centered logratio space. If

	<i>BC3</i>	<i>CS</i>	<i>F</i>	<i>IB</i>	<i>IFL</i>	<i>SW</i>	TPR [%]
<i>BC3</i>	3364	71	0	25	22	1	96.6
<i>CS</i>	14	514	12	35	104	0	75.7
<i>F</i>	0	6	215	3	5	0	93.9
<i>IB</i>	8	10	0	475	176	3	70.7
<i>IFL</i>	7	66	63	54	936	28	81.1
<i>SW</i>	0	0	0	1	50	167	76.6

Table 6.2: Confusion matrix associated with the true and predicted (LOOCV) lithofacies (core E10-3).

this assumption holds, this means that the correlation matrix of the clr-transformed intensities is identical to that of the clr-transformed concentrations. Hence, we construct biplots of uncalibrated intensities from their correlation matrix (i.e. the variables are given unit-variance prior to calculation of the eigenvectors).

6.4 Results

6.4.1 Core E10-3

QDFA was employed to predict the lithofacies and grain size class based on the core scanning data. The MRPE associated with the lithofacies and the grain size class were 11.9% and 27%, respectively. The average MRPE therefore equals 20%.

Figure 6.3 and 6.4 show the predictions for the entire core and for the example slabs, respectively. These figures clearly demonstrate that the discrepancies between assigned and predicted core description are concentrated in specific intervals. In ES6, for instance, the crevasse splay (*CS*) interval in the assigned core description is considerably thicker than in the predicted core description (Fig. 6.4). In addition, there are many small-scale discrepancies, i.e., places where the true and predicted core descriptions are broadly consistent apart from thin beds mainly associated with individual core-scanning measurements. Another kind of discrepancy is observed in ES4 (Fig. 6.4): the assigned core description suggests a single boundary from *S4* to *S3* whereas the predicted core description suggests that this transition is not discrete but takes place over an interval of 20 cm.

Tables 6.2 and 6.3 show the lithofacies and grain-size group confusion matrices, respectively. For the former, the highest TPR is found for *BC3* (96.6%) whereas the lowest TPR is *IB* (70.7%). Furthermore, primary confusion is between *IB* and *IFL*. Table 6.3 shows that for the grain-size group, the highest TPR is found for *M* whereas the lowest TPR values are found for *S5* and *O* (0% and 21.4%, respectively). Furthermore, *S5* is most often erroneously classified as *S4* and *O* is mostly confused with *M*. The dendrograms and biplots are shown in Figures 6.5 and 6.6. In the biplot the predicted class membership are indicated using colors. More than 50% of the variation in intensities can be explained by only two principal components.

6.4.2 Core B38D

Applying QDFA to the facies classification scheme of core B38D yields a MRPE of 10.9%. The confusion matrix given in Table 6.4 shows that *FOVOL* has the lowest TPR and is mainly

	J (sst)	M (clst)	O (org)	$S1$ (VFSst)	$S2$ (FSst)	$S3$ (MSst)	$S4$ (CSst)	$S5$ (VCSst)	TPR [%]
J (sst)	368	32	0	60	9	0	0	0	78.5
M (clst)	143	1798	3	45	3	1	0	0	90.2
O (org)	0	11	3	0	0	0	0	0	21.4
$S1$ (VFSst)	67	75	0	576	181	4	0	0	63.8
$S2$ (FSst)	6	6	0	88	1018	433	32	0	64.3
$S3$ (MSst)	1	1	0	5	243	770	202	0	63
$S4$ (CSst)	0	0	0	0	10	72	136	0	62.4
$S5$ (VCSst)	0	0	0	0	0	4	11	0	0

Table 6.3: Confusion matrix associated with the true and predicted (LOOCV) grain-size group (core E10-3). Key to the grain sizes: siltstone (sst), claystone (clst), organic (org.), very fine sandstone (VFSst), fine sandstone (FSst), medium sandstone (MSst), coarse sandstone (CSst) and very coarse sandstone (VCSst)

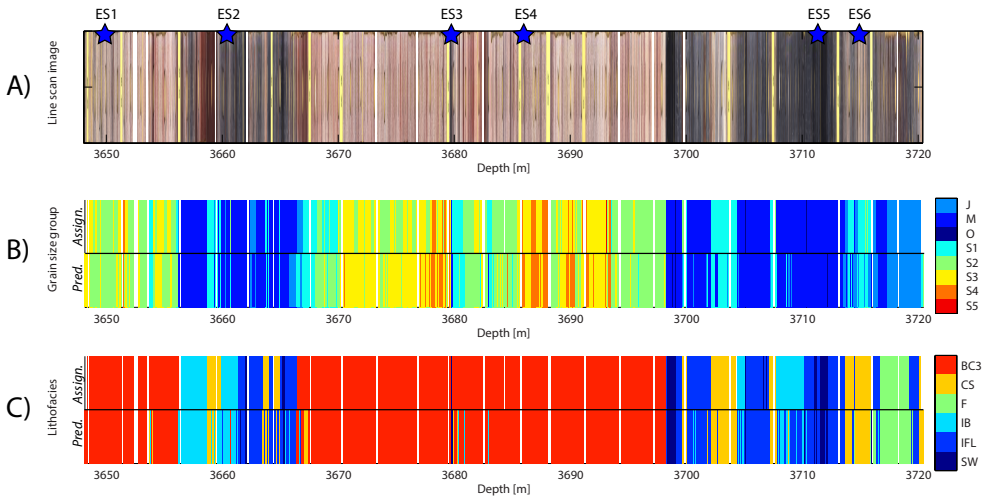


Figure 6.3: Results of the proposed calibration method for core E10-3: line scan image (A), grain-size group (B), lithofacies (C). In A, the position of the example slabs are indicated. In B and C, the assigned and predicted core descriptions are shown side by side.

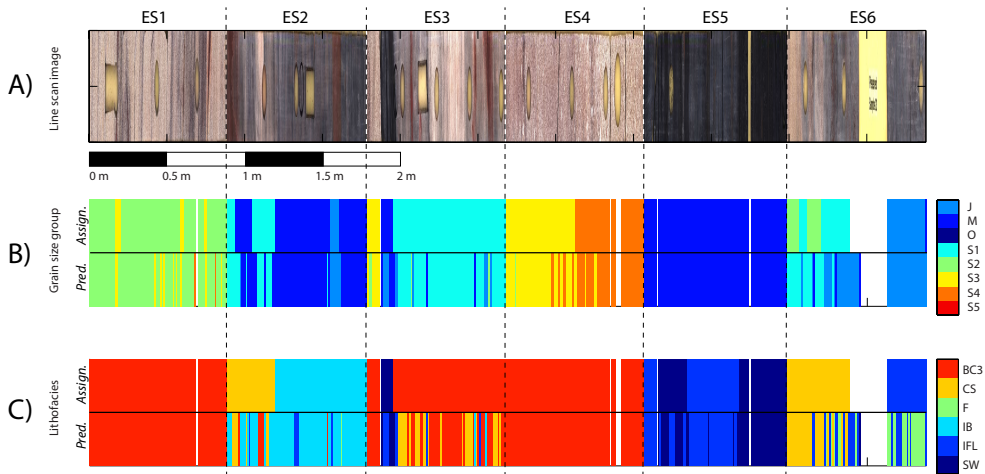


Figure 6.4: The calibration results of the example slabs, displayed in a similar manner as Figure 6.3.

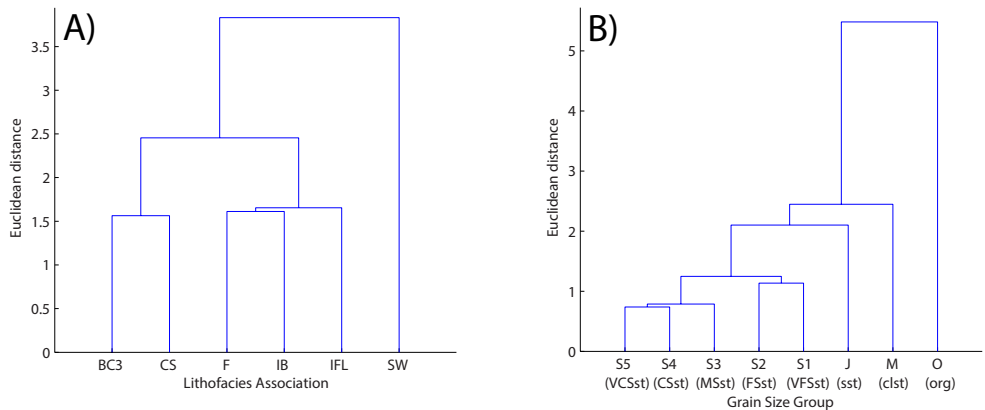


Figure 6.5: Dendrograms associated with the lithofacies (A) and the grain-size group (B) of core E10-3.

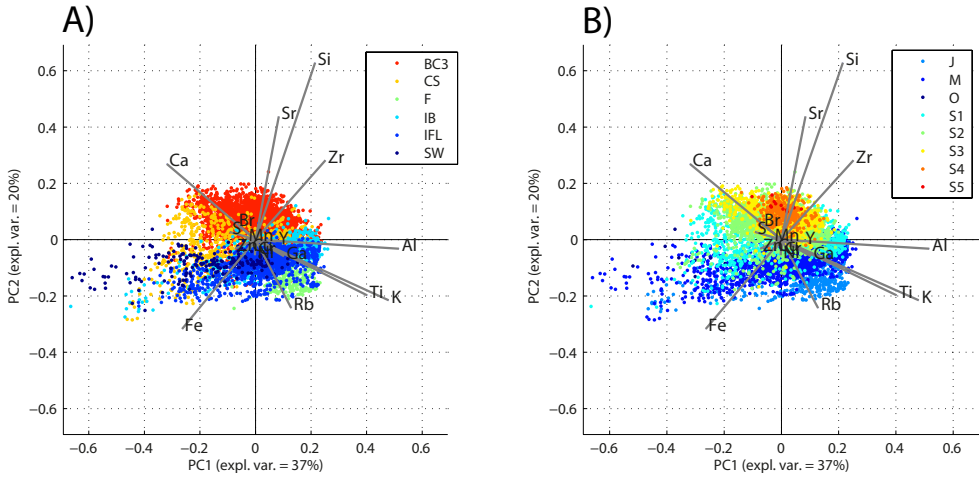


Figure 6.6: Biplots of the clr transformed and scaled intensities of core E10-3, with the predicted lithofacies (A) and grain-size group (B) indicated using colors .

	<i>FFG</i>	<i>FFKLA</i>	<i>FFKMA</i>	<i>FFL</i>	<i>FFZ</i>	<i>FOVAM</i>	<i>FOVEU</i>	<i>FOVOL</i>	TPR [%]
<i>FFG</i>	472	0	0	19	53	0	0	0	86.8
<i>FFKLA</i>	0	601	1	1	0	0	6	0	98.7
<i>FFKMA</i>	0	1	555	22	0	1	3	0	95.4
<i>FFL</i>	88	0	26	622	29	0	0	1	81.2
<i>FFZ</i>	39	0	0	11	173	0	0	0	77.6
<i>FOVAM</i>	0	0	0	0	0	23	0	0	100
<i>FOVEU</i>	0	2	2	0	0	0	99	0	96.1
<i>FOVOL</i>	0	0	5	1	0	0	0	2	25

Table 6.4: Confusion matrix associated with the true and predicted (LOOCV) lithofacies (Core B38D).

confused with *FFKMA*. *FOVAM*, on the other hand is predicted with a 100% TPR. The assigned and predicted facies descriptions are shown in Figure 6.7. Differences between assigned and predicted core description are found almost exclusively between 10 and 20 m. The most noticeable difference is the assignment of *FFG* to intervals which, according to the core description, are *FFL* sediments. In addition, both *FFZ* and *FFL* intervals around 20 m depth are classified as *FFG* based on the core-scanning data. These discrepancies in particular, have translated to the relatively low TPR of *FFZ* (TPR=77.6%) and *FFL* (TPR=81.2%)

6.5 Discussion

6.5.1 Dendrogram and biplot analysis

Cutting the lithofacies dendrogram of E10-3 (Fig. 6.5A) at a Euclidean distance of 2 yields three branches: branch 1 containing *BC3* and *CS*, branch 2 consisting of *F*, *IB* and *IFL*, and

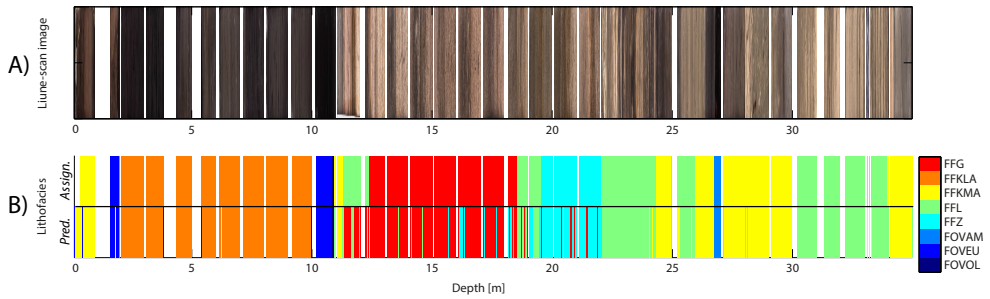


Figure 6.7: Results of the proposed calibration method for core B38D: line scan image (A), lithofacies (B). B shows the assigned and predicted core descriptions side by side.

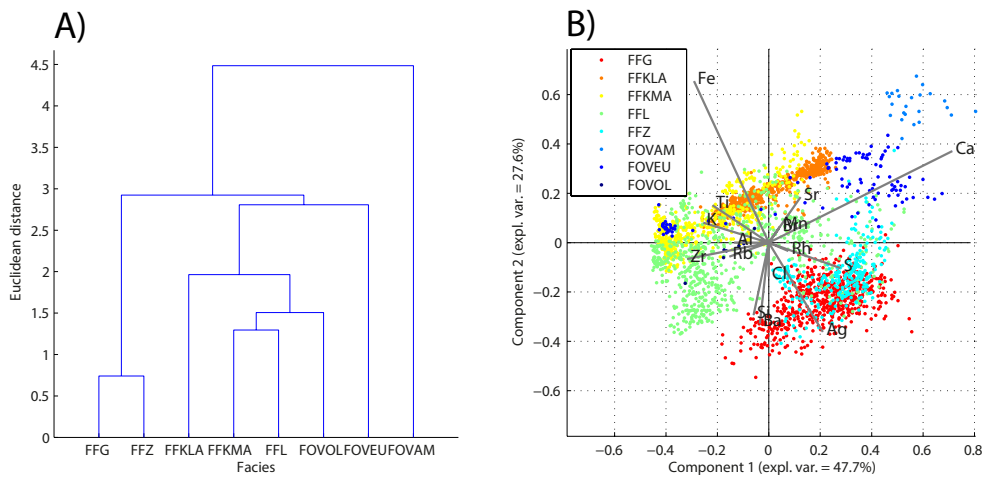


Figure 6.8: Dendrogram (A) and biplot (B) of the clr-transformed and scaled intensities with the predicted lithofacies of core B38D indicated using colors.

branch 3 containing only *SW*. The branches have a lithological significance: branch 1 reflects the siltstone and sandstone, branch 2 reflects the mudstone deposits and branch 3 reflects the coal deposits. This shows that sedimentologically-similar classes also have a relatively similar XRF-CS signature. In addition, the dendrogram of the grain-size classes (Fig. 6.5B) directly reflects the genetic relation between the grain-size classes, i.e., first *S5* and *S4* are grouped, then *S5* and *S4* are merged with *S3*, etc. The dendrogram of the facies in core B38D gives a similar result: we observe that *FFG* and *FFZ* have the most similar XRF-CS signature whereas *FOVAM*, *FOVEU*, *FOVOL* (i.e., the organic-rich sediments) have a relatively dissimilar XRF-CS signature compared to the other classes.

The dendrogram of core E10-3 suggests that the composition of the sandstone converges as the grain size increases. This, in turn, explains why the MRPE associated with the grain-size classification in core E10-3 is higher than that associated with the lithofacies association: whereas the lithofacies scheme is primarily aimed at discriminating between different types of mudstone units in terms of their depositional environment (i.e., it lumps all braided-channel sands into one group), the grain size classification scheme consists mainly of sandstone classes with a relatively similar XRF-CS signature.

The fact that *IFL* and *IB* are in the same branch (Fig. 6.5A) implies that their XRF-CS signatures are similar. This similarity explains the confusion between *IFL* and *IB* (see Table 6.2). Why they nevertheless represent different lithofacies groups follows from their definitions: whereas *IB* is defined as parallel-laminated and undisturbed claystone, *IFL* is defined as claystone with abundant rootlets (Boels, 2003). Hence, the main difference between *IFL* and *IB* is the presence or absence of sedimentary structures which is not reflected by the XRF-CS signature.

A similar situation applies to *FFG* and *FFZ* in core B38D, which are also in the same branch of the dendrogram (see Fig. 6.8) and are also frequently confused (see Table 6.4). Both are channel sands and the difference between the classes is the formation in which they are found, i.e. the Kreftenheye and Sterksel fm. for *FFG* and *FFZ*, respectively. This suggests that there is little to no compositional difference between the sands in these two formations.

The dendrogram of the grain-size group (Fig. 6.5B) shows that the XRF-CS signature of both *SW* and *O* are dissimilar from those of the other classes. The dissimilarity of the coal-bearing sediments (i.e., *SW* and *O*) from the other classes is not reflected by the lithofacies biplot, given the overlap between *SW* with both *CS*, *IFL* and *IB*. For *O* it is virtually impossible to derive its structure in the biplot because of the small number of data points. The compositional dissimilarity of the coal-bearing mudstones represented by *SW* is most likely reflected by the low-concentration elements (e.g. Ga, Zn) whose variations are more subtle and therefore mainly reflected by higher-order PCs.

6.5.2 Relation between confusion matrix and dendrogram

In core E10-3, beds classified as *O* are exclusively found in the *SW* lithofacies. Although the dendrograms of the grain size and lithofacies results both suggest the XRF-CS signatures of these classes to be dissimilar from the other classes, their TPR values are markedly different: whereas the TPR associated with *SW* is 77%, the TPR associated with *O* is 21%. A similar situation applies to the peat classes in core B38D: whereas the TPR of *FOVAM* is 100%, that of *FOVOL* is only 25%.

There are numerous intervals in core E10-3 where the sedimentological description

mentions "organic claystone interbedded by cm-scale coal beds" and where we predict *SW* but no *O*. This may be observed, for instance, in ES5 (Fig. 6.4, A and B). To explain why the coal laminae were not classified as *O* we consider the class definitions (see Appendix D). *SW* is defined as parallel laminated organic-rich claystone with cm-thick coal laminae. Hence, the description is consistent with the assignment of *SW*. Unfortunately, the grain-size class *O* is not clearly defined in the core report. Therefore, we are unable to provide a conclusive explanation for the coal laminae not being classified as *O*. *O* was assigned in case the concentration of organic material exceeds 50%. Determination of the concentration of organic carbon macroscopically, however, is questionable.

Another factor which controlled the low TPR of *O* (E10-3) is that for this class, there were only 14 data points to constrain the XRF-CS signature whereas for *SW* there were 218. Relative to the high number of intensity variables ($D = 18$), the limited number of *O* data points probably jeopardised the ability to constrain the XRF-CS signature. The same applies to *FOVOL* in core B38D which has a TPR of 25% whereas the other peat lithofacies have a TPR higher than 96%: in contrast to *FOVEU* and *FOVAM* there were only 8 core-scanning measurements to constrain the XRF-CS signature of *FOVOL*.

6.5.3 Interpretation of element intensities

Biplots show the correlation structure of the clr-transformed element intensities. By indicating the predicted lithofacies for every data point, they permit linking of lithofacies classes to particular elements.

In core E10-3, the grain-size biplot (Fig. 6.6B) shows that mainly Sr, Si and Zr are associated with *S4* and *S5*, i.e. the most coarse-grained sandstones. It also shows that K, Ti, Al, Fe and Rb are associated with mudstone and siltstone. The former may be explained geochemically from the fact that Si typically reflects the quartz content, whereas Zr is primarily contained in zircon and other high-density minerals: both are deposited under high-energy conditions. Rb and Al, on the other hand, are generally concentrated in feldspar and mica as well as their weathering products. In this core, Rb and Al probably reflect the concentration of clay minerals (kaolinite, illite). *M* and *J* show a profound elongated trend in Figure 6.6B which seems to be associated with variations in Al, Fe, K and Ti intensities. The structure suggests that a relatively low Fe intensity is compensated by an increase in K, Al and Ti and vice versa. In class *M*, the core description mentions siderite concretions which are also considered to be present in ES5. Also within class *J*, red colourations (see ES2) suggest local iron enrichments.

The biplot of the lithofacies in core B38D, shows that sandy sediments are primarily associated with elevated Ag, S, Ba and Si. Peat and organic-rich intervals, on the other hand, are characterised by elevated Ca intensities. The mixed clay and sand units such as *FFL* (clay-dominated) and *FFKMA* (sand-dominated) are mainly reflected by Fe, K, Ti, Fe, Zr. Like in core E10-3, we find that K, Ti and Al are closely related with clay-rich facies, probably because they reflect clay-mineral enrichment. Furthermore, the sandy intervals are characterised by elevated Si intensities because they are rich in quartz. The enrichment of peat layers with Ca is consistent with the work of Mol et al. (2012). In the Dutch shallow subsurface, an important carrier of Ca is CaCO_3 (Vermooten et al., 2011).

The strong association between sandy intervals and Ag intensities is difficult to explain. Ag is contained in the beam collimator and the Ag intensities should therefore be constant. It is very unlikely that Ag is contained in the sediment in considerable concen-

tration, hence these intensities cannot be explained mineralogically. Possibly, incoming X-rays are scattered in the sample, both elastically and inelastically, which influence the Ag intensities. This scatter, in turn, may be largely controlled by grain size. Similar conclusions were drawn by Kuhn et al. (2008).

6.5.4 Nature of discrepancies

A confusion matrix treats all discrepancies equal. However, the downcore records show that the nature of the discrepancies varies. We introduce the following two types of discrepancies: "stratigraphically-random" and "stratigraphically-systematic". The difference between the two is the thickness of the interval over which the two descriptions differ.

The discrepancies in the sandstone intervals in core B38D around 20 m (i.e., the confusion between *FFG*, *FFL* and *FFZ*) are an example of "stratigraphically-random" differences: although the majority of the data are correctly classified as *FFZ*, there are numerous individual XRF-CS measurements which are classified as *FFG*. Stratigraphically-random differences may be caused strictly by the measurement-by-measurement basis with which the predictions were made and the higher intrinsic resolution of the core-scanning data. In addition, they may be caused by the fact that when postulating the core description from the XRF-CS data, no use is made of prior knowledge.

A specific example of an interval where the assigned core description relies on prior knowledge is found in ES4 of core E10-3 (see Fig. 6.4), which contains a fining-upward trend. The core description captures this trend by one interface between *S3* and *S4*. The predictions, on the other hand, go from 100% *S4* at the base to 100% *S3* at the top, with *S4* and *S3* beds in between. Both are in principle correct. However, whereas the predicted core description potentially resembles reality more closely, the true core description may be qualified as a reasonable sedimentological interpretation.

The discrepancies in core B38D around 11 m and 12 m depth are an example of "stratigraphically-systematic" differences. Stratigraphically-systematic differences suggest that the assigned core description is at odds with the XRF-CS signature of the sediments. These intervals therefore merit a closer investigation of the core description's integrity.

6.6 General discussion, conclusions and recommendations

6.6.1 General discussion and conclusions

We applied QDFA to predict lithofacies in an automatic and objective manner. The average MRPE associated with the prediction of lithofacies is 16.6%. We acknowledge a number of factors which controlled the discrepancies between true and predicted core description. Firstly, the core description is based on assigning properties to sedimentary beds or bed sets using prior knowledge and the stratigraphic context, whereas the predictions are made on a measurement-by-measurement basis. Given that the classes are compositionally distinct, the predictions represent the "truth" more accurately because at the scale of the XRF-CS data they are strictly objective. Secondly, a core description is not only aimed at distinguishing lithologies: classes may also differ in terms of their appearance (i.e., color, sedimentary structures). This information is not contained in the XRF-CS signature which introduces discrepancies. Thirdly, in case of a sharp bed boundary, the core scan may not always be optimally vertically referenced with respect to the core description. Apart from being a

		To						#
		<i>BC3</i>	<i>CS</i>	<i>F</i>	<i>IB</i>	<i>IFL</i>	<i>SW</i>	
From	<i>BC3</i>	98.3	1.1	0	0.4	0.3	0	3420
	<i>CS</i>	4.7	79.9	1.4	2.7	11.3	0	701
	<i>F</i>	0	2.7	88.5	0.3	8.1	0.3	296
	<i>IB</i>	2.9	3.6	0.2	78.2	14.7	0.5	591
	<i>IFL</i>	0.7	5.8	1.8	7.2	81.8	2.8	1297
	<i>SW</i>	0.5	0	0	1	18	80.5	205

Table 6.5: Upward transition probability matrix of the predicted lithofacies in core E10-3.

potential source of prediction errors, however, the vertical position of a bed boundary implied by a core description is under some circumstances also highly subjective, for instance when the transition is smooth. In that case, the predictions based on core-scanning data are at least made in an objective manner. Taking those factors into account, we consider the proposed methodology of great potential value for automatic prediction of lithofacies. Moreover, in the future it may be attempted to use the same methodology to control the quality of conventional core description and to harmonise cores which were described in terms of different classification schemes.

6.6.2 Outlook

Further improvements to the proposed methodology include the application of intelligent post-processing steps and complementing the XRF-CS data with other data sets in order to accommodate for the above-mentioned factors.

An example of a potentially fruitful post-processing step is "stratigraphic smoothing", i.e., in case the predicted lithofacies for a particular XRF-CS measurement is different from those of the neighbouring classes which have the same lithofacies, all three measurements are assigned the lithofacies of the neighbouring measurements. This is likely to reduce the discrepancies between assigned and predicted core description.

In certain cases, the cm-scale predictions are not meaningful geologically. In case of a fining-upward sequence, for instance, the cm-scale predictions of lithofacies will include all grain-size classes. In that case, it is important that we can automatically recognise the genetic ordering of the lithofacies and turn the sequence into one super class. The genetic ordering in a categorical data set is often analysed using a transition probability matrix. In addition to the recognition of sequences and the improved understanding of the mechanisms which controlled the stratigraphic record, exploitation of transition probability matrices in the prediction of lithofacies by modifying the priors is likely to increase the predictive capabilities.

The upward TPM of the lithofacies in core E10-3 is shown in Table 6.5. We observe that, in case of a change in lithofacies, a *SW* bed is mostly overlain by *IFL* (TP = 18%) and rarely by *BC3* (TP = 0.5%) and *IB* (TP = 1%). The association between *SW* and *IFL* is in line with the core report. As expected, Table 6.5 also shows that *CS* is mostly associated with *BC3*: it is evident that crevasse splays (*CS*) are mostly found in braided channel intervals (*BC3*).

Given that a geologist makes a core description also on the basis of the core's appearance, exploiting core photographs data is likely to increase performance of lithofacies pre-

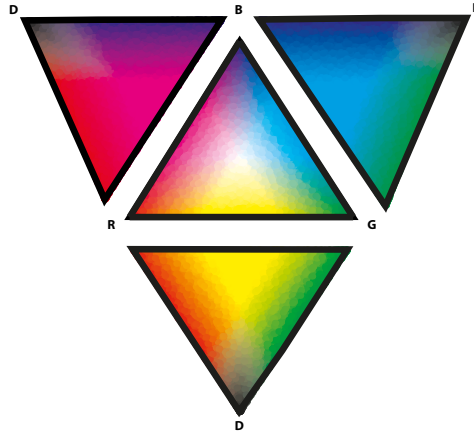


Figure 6.9: Unfolded tetrahedron whose faces are ternary diagrams associated with combinations of R, G, B and D.

diction. One of the most basic properties which can be retrieved from line-scan images is the color of the sediment. It is evident that distinguishing between sandstone and mudstone on the basis of color is relatively easy. However, it may also help in distinguishing between lithofacies in a more detailed manner. Therefore we consider the variability in the color composition of the sediment. Given that the colors data represent 8-bit R(red), G(green) and B(blue), we may turn the color data into a composition by defining the shadow variable D (i.e., darkness):

$$D = 3 \times (2^8 - 1) - (R + G + B) \quad (6.5)$$

The structure of the RGBD space is shown by means of an unfolded tetrahedron in Figure 6.9.

Because R, G, B and D define a 4-part composition, we may analyse them by means of a compositional biplot. Figure 6.10 shows the RGBD biplot of core E10-3 with the predicted lithofacies indicated. It clearly demonstrates that the sandstone units (i.e., *BC3*) are characterised by redness (R), whereas the coal-bearing swamp (*SW*) deposits are mainly characterised by darkness (D). Both results are in line with what may be expected.

Given that color composition, in contrast to the geologists's judgment, is a quantitative measure, complementing XRF records with color data is a potentially powerful way to formalise the construction of core descriptions. Apart from color, other relevant information which can potentially be retrieved from core images is the grain-size distribution, the bedding direction and the presence and nature of sedimentary structures. All these properties are vital for interpretation of depositional environment, but require a substantial effort in terms of image analysis before they can be used to make integrated core descriptions on a routine basis.

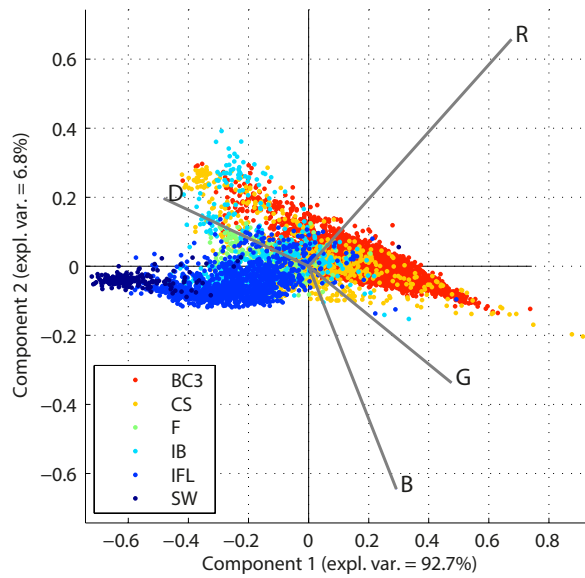


Figure 6.10: Biplot of the clr-transformed average RGBD values at the core-scan positions (core E10-3). Colors indicate the predicted lithofacies on the basis of the XRF core scanning data.

Integrated core analysis using XRF scanning. Part II: Prediction of chemical and petrophysical properties

7.1 Introduction

In Chapter 6 we analysed to what extent it is possible to predict lithofacies from XRF-CS records. The goal of this second part of the construction of an *integrated core description* is to predict chemical and petrophysical core properties at the centimeter scale. Therefore the same quality-controlled XRF-CS data sets are used as outlined in Chapter 6.

7.2 Materials and Methods

7.2.1 Core analysis

Routine core analysis of core E10-3 yielded between 230 and 244 plug analyses including grain density, porosity and horizontal permeability (Bakker and Pruno, 2004). The grain density varies between 2.6 and 3.2 g/cm³ and the porosity ranges from 0 to 20%. The horizontal permeabilities in this core range roughly up to 400 mD. To enable geochemical calibration of the core scan, another 35 core positions (both at and between plug locations) were sampled and analysed using pressed-pellet XRF. These analysis were conducted by Chemostrat Ltd. The XRF analyses yielded elemental concentration estimates for Al, Si, Ti, Fe, Mn, Mg, Ca, Na, K, P, S, Cl, Ba, Br, Ce, Co, Cr, Cs, Cu, Ga, La, Mo, Nb, Ni, Pb, Rb, Sb, Sc, Sn, Sr, Ta, Th, U, V, W, Y, Zn, Zr. We only included the elements which were also measured by the core scanner, i.e., Al, Si, Ti, Fe, Mn, Ca, K, S, Br, Cr, Cu, Ga, Ni, Rb, Sr, Y, Zn, Zr.

Analysis of core B38D was conducted by TNO. Vertical permeability analyses (n=30) were acquired using a constant-head conductivity test. For the sand and clay samples, the bulk geochemical composition was determined using pressed-pellet XRF. The composition of the samples rich in organic carbon was determined using both glass-bead XRF as well as ICP-OES. We combined these results such that a bulk composition was obtained in terms of the following selected elements: Al, Si, Ti, Fe, Mn, Ca, K, Cr, Cu, Ga, Ni, Rb, Sr, Ba, Y, Zn,

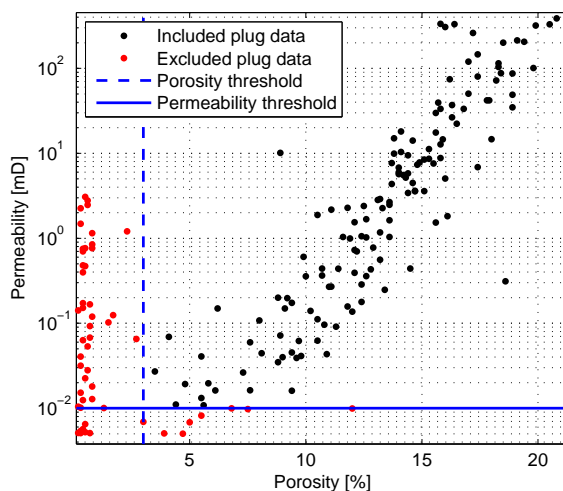


Figure 7.1: Plug porosities and permeabilities of core E10-3 included (black) and excluded (red) in the calibration process. Inclusion was based on whether their porosity and permeability exceeded the threshold values of 3% and 0.01 mD (shown in blue).

Zr. In case the analyte concentration was determined by both glass-bead XRF and ICP-OES, we selected the XRF analyses to maximise method consistency.

7.2.2 Calibration data integrity

For core B38D, we only used the data which passed the quality-control protocol of TNO. The data from core E10-3 have also been subjected to a quality-control procedure by Panterra Geoconsultants. We subjected the E10-3 data to another round of quality checking because some analyses were suspect. This suspicion originated from the porosity-permeability relation, as shown in Figure 7.1. As also mentioned in the core report, the increase in permeability as the porosity reaches 0% is probably artificial. According to the report, these analyses are probably influenced by leakage (along fractures or along the Hassler sleeve). Apart from these physical constraints, the increasing uncertainty with decreasing magnitude at the low end, is a common feature of non-negative petrophysical data, because of finite instrument sensitivities. Hence, we omitted all measured permeabilities smaller than 0.01 mD from the calibration procedures and we adopted a porosity threshold of 3% based on the poroperm relation (Fig. 7.1).

Geochemical data are also subject to detection or quantitation limits. For the selected set of elements, there were only a handful data points for which the concentration was reported as below the level of detection (LOD). Because the modelling methods used in this study rely on positive concentrations, we had to define a strategy for coping with zeros in the geochemical data set. We applied the multiplicative zero-replacement technique proposed by Martín-Fernández et al. (2003), which is based upon replacing zeros by a small value δ and simultaneously deflating the concentration of all other components by the

same amount. The analyte-specific replacement value δ_j was set at the minimum positive concentration in the data set. Apart from being positive, we also had to define a way to cope with compositions whose total concentration exceeds 100%. We decided to close these data to the maximum total concentration in the data set which does not exceed the physical limit of 100%.

7.3 Modelling methods

7.3.1 MLC method

The Multivariate Logratio Calibration (MLC) method (see Weltje et al. (2015) and Chapter 2) forms the basis for all calibration exercises undertaken in this study. The rationale behind the MLC method is that both core-scanning and chemical composition data are modelled as logratios (Weltje and Tjallingii, 2008). To establish a framework suitable for calibrating all considered data types with high predictive capability, a further generalisation of the MLC method is presented here. We refer to this method as the Generalised Logratio Calibration (GLC) and, compared to the MLC method, it possesses a number of new features. The different features are given below. They are described in more detail in Appendix D.

1. **Flexibility.** The GLC method is more flexible than the MLC method because it encompasses a larger set of possible invertible transformations that can be applied to the Y-data in order to make calibration of non-compositional quantities possible. Transformations are selected on the basis of the nature of the Y-data: we employ the invertible transformation which makes the Y-data unconstrained and ensures physically tractable predictions.
2. **Robustness.** Weltje et al. (2015) defined the optimal number of PLS components as the number for which the MSPE is minimal. In contrast to their definition, the MSPE in the GLC is based on a robust measure of scale. This will not only give a more reliable estimate of the optimal number of PLS components but it will also give a more reliable estimate of the prediction uncertainty compared to the MLC method.
3. **X-variable selection.** The GLC is based on X-variable selection using the Variable Importance in Projection (VIP) (see Eriksson et al., 2001). Removal of uninformative X-variables may be beneficial, given that counting statistics prescribe that low-intensity variables have a higher uncertainty than high-intensity variables, and errors in clr-space are not independent which means that low-intensity variables potentially "pollute" all other variables Bloemsma and Weltje (2015). Hence, variables must be included in the model only in case they contribute to a better prediction.
4. **Y-variable selection.** Given that clr-transformed variables are intrinsically correlated, removal of unpredictable Y-variables may also increase the inferred predictive performance. In this study, the only multivariate Y data are geochemical data. Y-variable selection in the GLC is based on the removal of all elements whose R^2 is smaller than some threshold value.
5. **Independent prediction of scale and composition.** The MLC approach to construct complete from incomplete geochemical compositions (i.e. that the sum over

all variables is smaller than 100%) by adding the "Undef" to the composition implies that all Y-variables must be included in the model (see Weltje et al., 2015). To enable Y-variable selection in the GLC, we follow a two-step approach whereby scale (scale = 1 - undef) and composition are modelled separately. Apart from the fact that this two-step approach ensures that the predictions are unbiased relative to the "absolute" concentrations, it is also expected to give better predictions because composition and scale are also estimated in a fundamentally different manner by conventional spectrochemical methods (Chapter 5).

7.3.2 GLC and data scaling

The MLC method is based on scaling the X- and Y-data matrices prior to calibration, in order to reduce the importance of imprecise variables: in case two variables contain the same information, scaling ensures that the one with a high precision is implicitly made more important than the one with a low precision. The scaling factors may be derived from replicate analyses (see Weltje et al. (2015), and Chapter 2, this thesis). Progress in uncertainty modelling of geochemical data, however, showed that the uncertainty associated with chemical analyses can be predicted on the basis of the concentration c by means of the Binomial Horwitz Function (Bloemsmas and Weltje (submitted), and Chapter 5):

$$\delta c = 0.02[c(1 - c)]^{0.8495} \quad 10^{-8.88} \leq c \leq 1 - 10^{-8.88} \quad (7.1)$$

where δc is the expected standard deviation of c . Given that $q = \log(c)$ and assuming that the errors δc are independent among the analytes, the error in q reads:

$$\delta q = 0.02[c(1 - c)]^{0.8495-1} \approx 0.02[c(1 - c)]^{-0.15} \quad 10^{-8.88} \leq c \leq 1 - 10^{-8.88} \quad (7.2)$$

The fact that the exponential term in the above equation is relatively small means that the errors associated with the log-transformed, and therefore also of the clr-transformed variables, become concentration-independent. If the number of X-variables is reasonably large (> 20), these errors are also independent (Bloemsmas and Weltje (2015)). From the standpoint of least-squares estimation, scaling the element concentrations according to their uncertainty is therefore an unnecessary operation because for all practical purposes, the PLS model will be indistinguishable from the Maximum Likelihood solution. Scaling the petrophysical data will have no effect on the associated calibration models because all petrophysical data included in this study are univariate.

7.3.3 Summary statistics

Model selection in the GLC is executed on the basis of the MSPE (see Appendix D). In case the data are of type 1 (i.e., unconstrained data such as the mean grain size in phi-units), the MSPE may be readily used as an estimate of the prediction uncertainty. For all bounded data types (types 2, 3 and 4), however, the MSPE is defined in terms of transformed data which makes the MSPE difficult to interpret in terms of the original units of measurement. Moreover, the uncertainty of these data types is much more likely to behave proportionally. For these data types, it is therefore more sensible to evaluate a Mean Relative Prediction Error:

$$\text{MRPE}(\%) = 100\% \times \text{std} \left(\log \left(\frac{\mathbf{y}}{\hat{\mathbf{y}}} \right) \right) \quad (7.3)$$

For multivariate Y , one summary statistic is obtained by vectorising the Y -data matrix prior to MRPE calculation.

Whereas the MSPE and the MRPE are "absolute" error estimates, another relevant summary statistic is the signal strength of the predictions, i.e. the variance of the residuals relative to the variance of the predictions. We refer to this metric as the R_{eff}^2 . For data types 1, 2 and 3, $R_{\text{eff}}^2 = R^2$ reflects the squared Pearson correlation coefficient associated with the data and predictions in the modelling space, i.e.:

$$R_{\text{eff}}^2 = 1 - \frac{\text{var}(\mathbf{y}^* - \hat{\mathbf{y}}^*)}{\text{var}(\mathbf{y}^*)} \quad (7.4)$$

For compositional data, R_{eff}^2 is derived in a slightly different manner. Although the same equation is used, \mathbf{y}^* is substituted by \mathbf{z}^* whose elements are given by:

$$z_j^* = \log \left(\frac{w_j}{1 - w_j} \right) \quad (7.5)$$

The rationale behind this transformation is discussed in more detail in Appendix D.

7.3.4 Predictive performance (petrophysical properties)

Our goal is to calculate the ratio between mean prediction error and uncertainty of conventional analyses, which we refer to as the Predictive Performance Ratio (PPR).

All petrophysical properties analysed in this study are strictly positive. Hence, it is reasonable to assume a constant relative error for both the prediction uncertainty and the uncertainty of laboratory data so that the PPR may be defined as:

$$\text{PPR}_{\text{petro}} = \text{MRPE} / \text{MRE}_{IL} \quad (7.6)$$

where the MRPE is the mean prediction error inferred using LOOCV. An inter-laboratory experiment conducted by Thomas and Pugh (1989) involving sedimentary rock yielded values for the MRE_{IL} for both the grain density (0.13%), porosity (0.67%) and permeability (4.96%). The minimum porosity and permeabilities values in that study were considerably larger than found in this study. Under the influence of the decreasing sensitivity of the analytical methods, it is very likely that the data used in this study have a larger associated error. To compensate for this, we adopt the maximum values reported by Thomas and Pugh (1989) as the MRE_{IL} , i.e. 0.20%, 1.5% and 20% for the grain density, porosity and permeability, respectively. Although the petrophysical data of B38D were derived using different analytical techniques, the same values are applied to both cores.

7.3.5 Predictive performance (geochemical composition)

The absolute and relative inter-laboratory error associated with geochemical data is concentration-dependent (Horwitz et al., 1980). Hence, the structure of the PPR used for petrophysical data (Eqn. 7.6) is unsuitable.

An often used performance criterion for analytical instruments is the HorRat (see Horwitz and Albert, 2006, and references therein):

$$\text{HorRat} = \frac{\sigma_{\text{emp}}}{0.02c^{0.8495}} \quad (7.7)$$

In this equation, σ_{emp} reflects the spread between "true" analyte concentrations and the concentration c determined with some analytical instrument. In Chapter 5, we discussed the problems surrounding the *denominator* (i.e., the Horwitz function) of the HorRat which reflects the expected inter-laboratory spread. We concluded that the asymmetry of the Horwitz function is inconsistent with the definition of analyte concentrations as mass fractions. Applying a rigorous modelling framework to a database with inter-laboratory spreads associated with routine analysis of geological samples yielded the following alternative to the HorRat:

$$GPR = \text{med}_{ij} \left\{ \frac{|\hat{c}_{ij} - \bar{c}_{ij}|}{0.022[\bar{c}_{ij}(1 - \bar{c}_{ij})]^{0.867}} \right\} \quad 10^{-9.68} \leq c \leq 1 - 10^{-9.68} \quad (7.8)$$

In this equation, med_{ij} reflects the median over all calibration samples i and analytes j . In Chapter 5 it was proposed that a GPR smaller or equal to two means that the analytical setup performs satisfactory.

What was not considered is the fact that the empirical standard deviation estimate (i.e., the *numerator* of Eqn. 7.7) is estimated in the problematic concentration space. In this chapter, we go one step further and estimate this spread also in a binomial-like space. Therefore we propose the following transformation:

$$z = \log(c/(1 - c)) = \log(c) - \log(1 - c) \quad (7.9)$$

Previously (Eqn. 7.2) we showed that, given $q = \log(c)$, it holds that $\delta q = 0.02[c(1 - c)]^{-0.15}$. It also holds that:

$$\delta z = \sqrt{(0.02[c(1 - c)]^{-0.15})^2 + (0.02[c(1 - c)]^{-0.15})^2} = 2^{0.5}0.02[c(1 - c)]^{-0.15} \quad (7.10)$$

Based on this expression we propose the following definition for the PPR associated with the geochemical data:

$$\text{PPR}_{chem} = \text{med}_{ij} \left\{ \left| \log \left(\frac{\hat{w}_{ij}}{1 - \hat{w}_{ij}} \right) - \log \left(\frac{w_{ij}}{1 - w_{ij}} \right) \right| \right\} \quad (7.11)$$

In this equation, \hat{w}_{ij} is the LOOCV predicted concentration. The fundamental difference between Equation 7.11 and the GPR is that, in contrast to the spread around concentration estimates \hat{w}_{ij} , the spread around $\log \left(\frac{\hat{w}_{ij}}{1 - \hat{w}_{ij}} \right)$ is by definition symmetric. This makes the PPR_{chem} a robust and unbiased statistical measure. We will use the PPR_{chem} in a similar manner as the GPR , i.e., in case PPR_{chem} is smaller than two, the core scanner performs as good as conventional analytical methods.

7.3.6 Repeatability of XRF-CS data

Replicate scans enable quantification of the contribution of repeatability errors to the inferred prediction errors. Since we have a well-defined UF only for the geochemical uncertainty, this investigation is limited to the geochemical data. We quantify the contribution of repeatability errors to the geochemical prediction errors by means of the X-data Repeatability Ratio:

$$\text{XRR} = \text{med}_{kj} \left\{ \frac{\text{std} \left[\log \left(\frac{\hat{w}_{kj}}{1 - \hat{w}_{kj}} \right) \right]}{2^{0.5}0.02[\hat{w}_{kj}(1 - \hat{w}_{kj})]^{-0.1505}} \right\} \quad (7.12)$$

	Comp.	Scale	Geochem.	Comp.	Scale	Geochem.
X-vars out	Cr	Y	none	Cr, Cu, Ni, Y	Cr, Zn	-
Y-vars out	Br, Cr, S, Zr	-	-	Br, Cr, S, Zr (set)	-	-
No. PLS comp.	12	12	-	9	5	-
R^2	0.65	0.53	0.70 (R_{eff}^2)	0.63	0.22	0.70 (R_{eff}^2)
MSPE	0.22	0.30	-	0.18	0.17	-
MSI%	57.93	55.98	-	73.66	75.37	-
MRPE%	-	-	54.46	-	-	51.75
PPR _{chem}	-	-	3.12	-	-	3.13
XRR	-	-	0.74	-	-	0.28

Table 7.1: Summary of the geochemical calibration results of core E10-3, with (left columns) and without (right columns) the coal sample.

where the vector \check{w}_{kj} contains the replicate intensities of the k -th replicates set and analyte j , transformed to concentrations. The mean predicted concentration at this core position is given by \hat{w} . A XRR equal to unity implies that the repeatability errors have the same magnitude as the expected inter-laboratory errors. Comparison of the XRR and the PPR provides insight into the contribution of repeatability errors to the prediction error (MSPE).

7.4 Results

7.4.1 Geochemical calibration

The core scanning data were geochemically calibrated using the GLC method with a VIP threshold of 0.25 and a R^2 threshold of 0.5. To obtain one XRF-CS measurement for each calibration point, all XRF-CS measurements within 1 cm from the calibration point were averaged in logratio space.

The summary statistics of the geochemical calibration exercise are shown in Table 7.1 (E10-3) and 7.2 (B38D). For core E10-3, the predictions along the entire core and along the six example slabs are given in Figures 7.2B and 7.3B, respectively. The predictions for core B38D are shown in Figure 7.4. Cross plots of data and predictions are shown in Figures 7.5 and 7.6 for cores E10-3 and B38D, respectively.

For both cores, a total of 15 elements could be adequately calibrated. In core E10-3, S, Zr, Br and Cr were consecutively removed because their R^2 value was below the threshold value. From the set of calibrated elements in core E10-3, Ga and SiO₂ display the highest signal strength ($R_{\text{eff}}^2 = 0.86$ and $R_{\text{eff}}^2 = 0.81$, respectively) whereas the lowest R_{eff}^2 values were found for Sr ($R_{\text{eff}}^2 = 0.49$) and TiO₂ ($R_{\text{eff}}^2 = 0.52$) (Fig. 7.5). In core B38D, Cu, Ga and Rb were removed on the basis of their R^2 and as shown in Figure 7.6, SiO₂ and Fe₂O₃ demonstrate the highest signal strengths of 0.92 and 0.90, respectively. For this core, the elements with the lowest signal strength include Y ($R_{\text{eff}}^2 = 0.39$) and Ni ($R_{\text{eff}}^2 = 0.47$).

Cross plots shown in Figures 7.5 and 7.6 suggest that the ability to predict the concentration is concentration-dependent: particularly for major elements such as SiO₂, the scatter seems to increase as the concentration decreases.

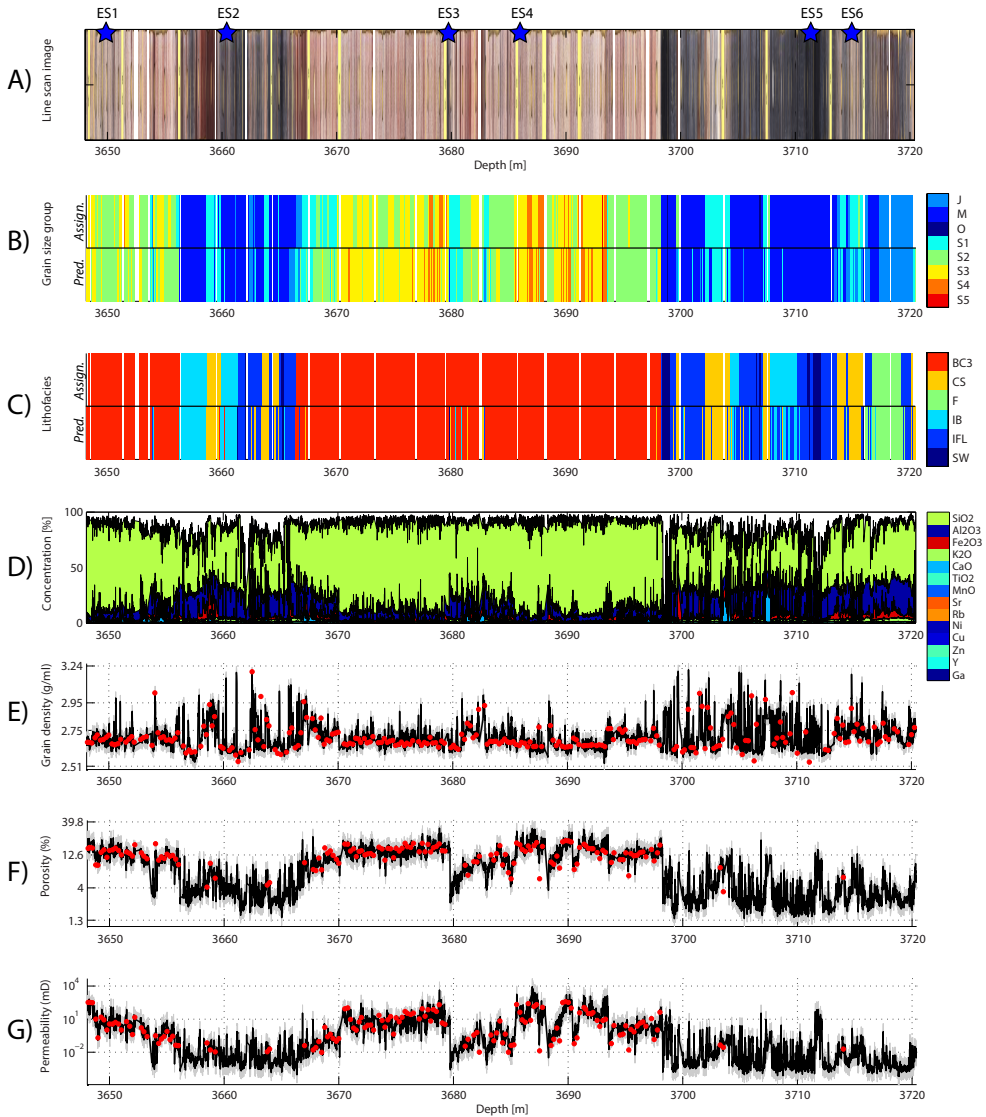


Figure 7.2: Results of the proposed calibration method for core E10-3: line scan image (A), grain size group (B), lithofacies association (C), bulk geochemical composition (D), grain density (E), porosity (F), and permeability (G). In A, the position of the example slabs are indicated. E, F and G show the plug data (red), the core scan predictions (black) and the associated 90% confidence interval (gray). In B and C, the assigned and predicted core descriptions are shown side by side.

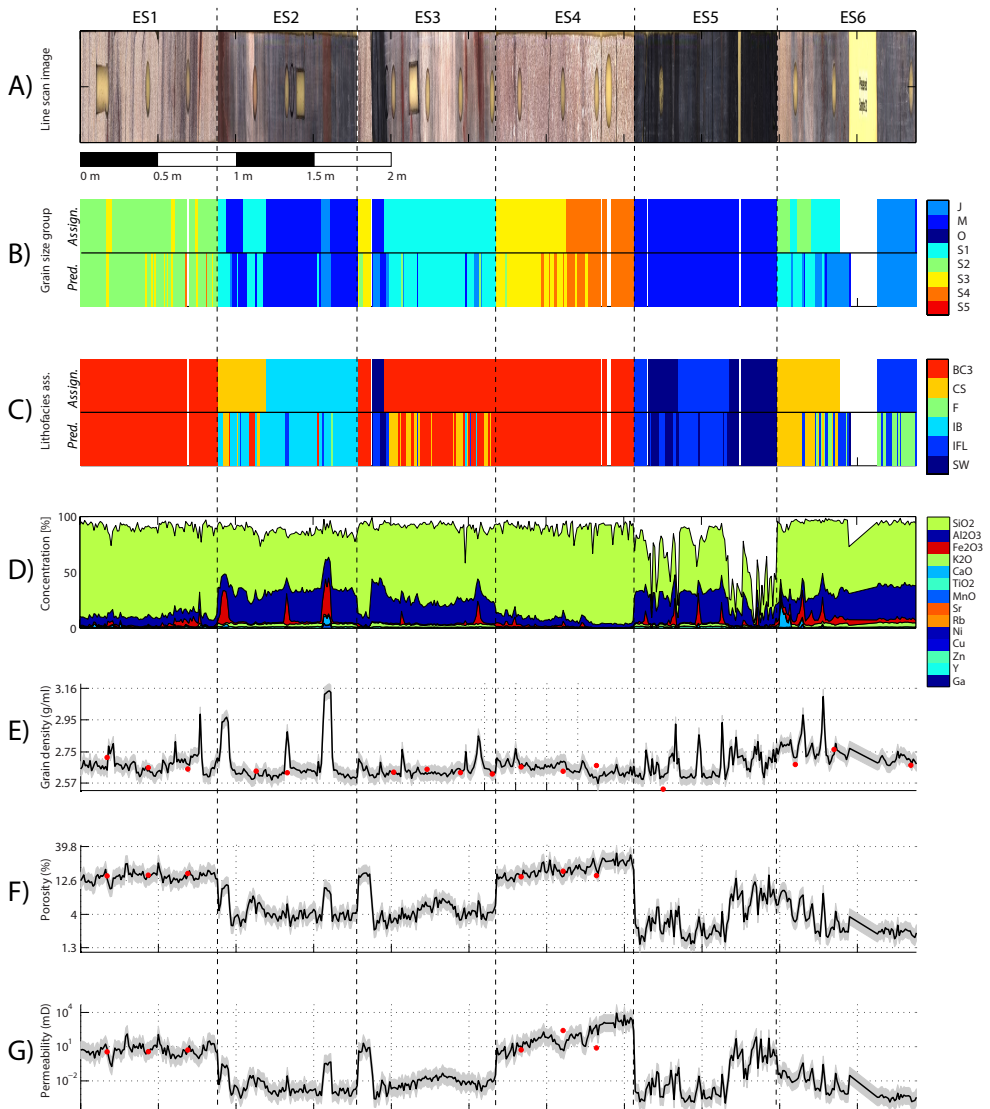


Figure 7.3: The calibration results of the example slabs, displayed in a similar manner as Figure 7.2.

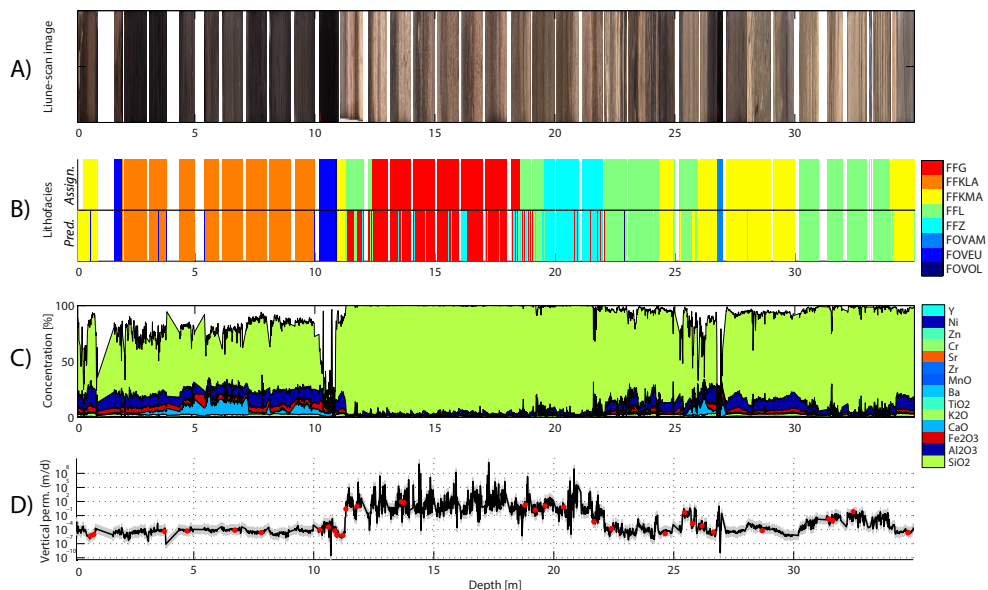


Figure 7.4: Results of the proposed calibration method for core E10-3: line scan image (A), lithofacies (B), bulk chemical composition (C), and permeability (D). The lithofacies predictions shown in B were derived in Chapter 6. D shows the plug data (red points) and the predictions (black line) with a 90% confidence interval (gray).

	Comp.	Scale	Geochem.	Comp.	Scale	Geochem.
X-vars out	none	none	-	none	Rh	-
Y-vars out	Cu, Ga, Rb	-	-	Cu, Ga, Rb (set)	-	-
No. PLS comp.	8	8	-	10	3	-
R ²	0.78	0.77	0.77 (R _{eff} ²)	0.75	0.74	0.89 (R _{eff} ²)
MSPE	0.07	0.36	-	0.04	0.38	-
MSI%	77.59	64.28	-	74.47	84.15	-
MRPE	-	-	40.23	-	-	28.17
PPR _{chem}	-	-	2.76	-	-	1.57
XRR	-	-	0.35	-	-	0.46

Table 7.2: Summary of the geochemical calibration results of core B38D, with (left columns) and without (right columns) the samples with a high TOC content.

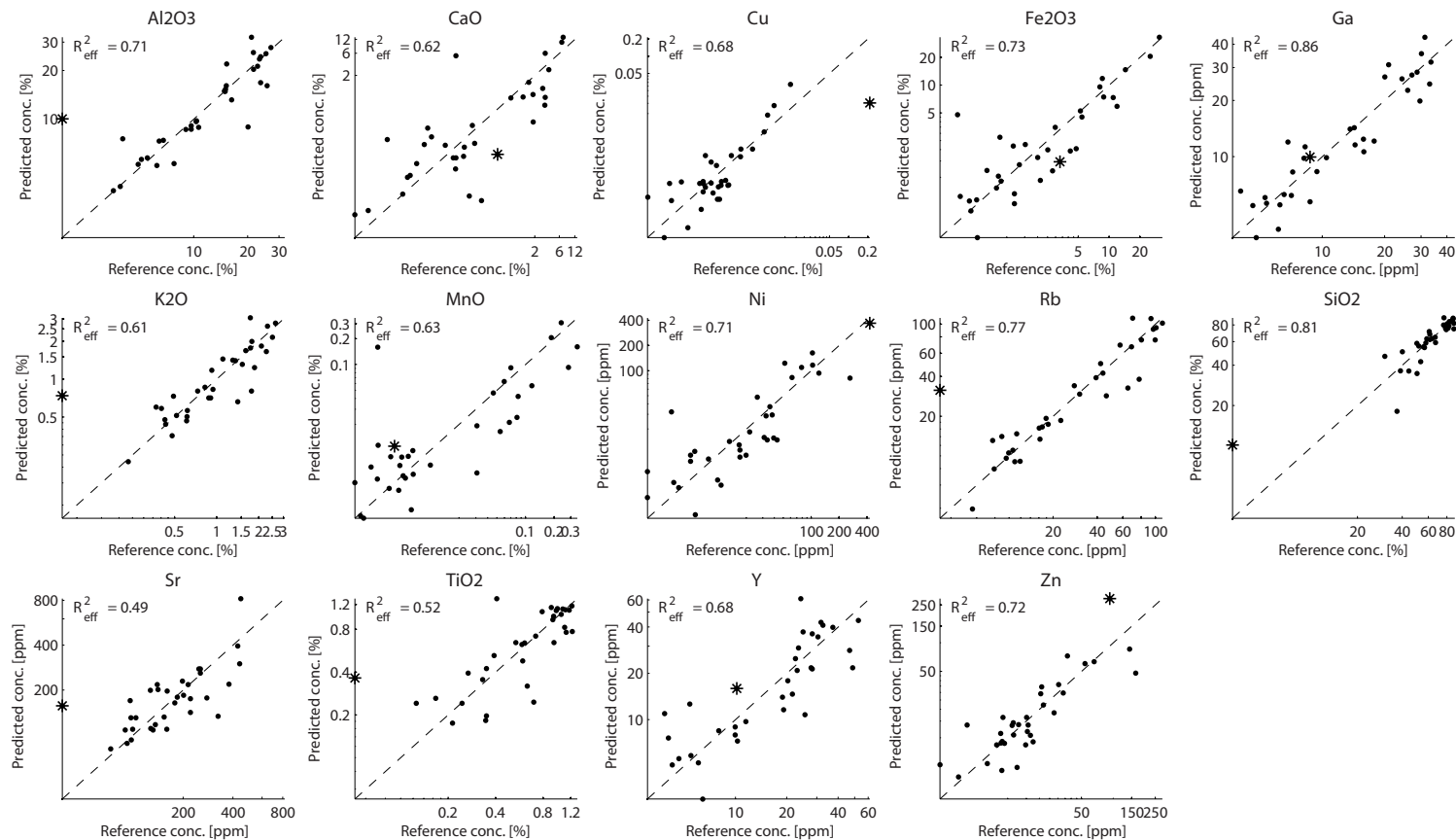


Figure 7.5: Cross plots of the analyte concentration in the calibration data against the predicted (LOOCV) analyte concentration in core E10-3. Each cross plot shows the R_{eff}^2 and the line $y=x$. The data point indicated by a '*' represents the coal sample.

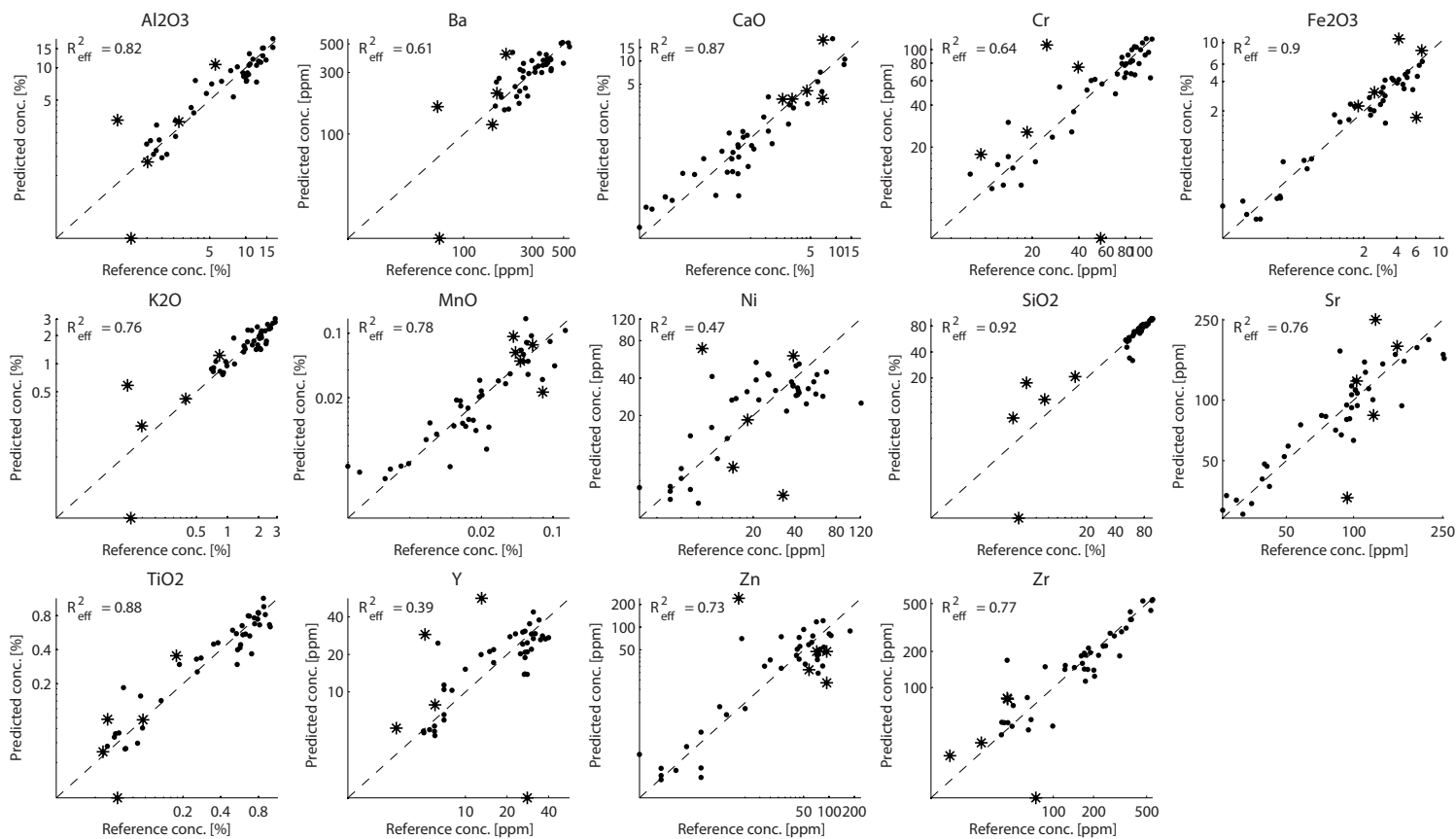


Figure 7.6: Cross plots of the analyte concentration in the calibration data against the predicted (LOOCV) analyte concentration in core B38D. Each cross plot shows the R_{eff}^2 and the line $y=x$. The data points indicated by a "*" represent the organic-rich samples.

Geochemical calibration (only siliciclastics-dominated samples)

The results we obtained by geochemical calibration illustrate the ability of the GLC model to predict the geochemical composition of the cores as a whole. However, the fact that the uncertainty seems to increase as the concentration of major elements decreases suggests that excluding samples with a small compositional scale may have a positive effect on the inferred predictive performance. Hence, we perform another geochemical calibration exercise, but now without the samples with a scale smaller than 40%, i.e., samples with a high carbon content. The resulting summary statistics are shown in Tables 7.1 (E10-3) and 7.2 (B38D).

The associated summary statistics illustrate the ability of the GLC model to predict the analyte concentrations in all but the coal and peat layers. There are a number of noticeable differences with the initial calibration results. Firstly, the PPR remained constant for core E10-3, and it decreased for core B38D from 2.76 to 1.57. The same applies to the MSPE values associated with both composition and scale. We also observe that for core E10-3, the number of excluded intensity variables is smaller in the first model: in contrast to the initial model, Cu, Ni and Y were not included in the final calibration model. Stability generally increased substantially from the first to the second model: only for the compositional calibration model in B38D the MSI was effectively unchanged. We get mixed results for the XRR: it decreased in core E10-3 from 0.70 to 0.27, whereas in core B38D it increased from 0.36 to 0.46.

7.4.2 Petrophysical calibration

We applied the GLC model to the petrophysical data. The same thresholds and tolerance were used as adopted for the geochemical calibration. The results are summarised in Table 7.3 and cross plots between plug data and LOOCV predictions are shown in Figure 7.7.

Table 7.3 shows that the signal strength R_{eff}^2 does not vary much among the properties and is comparable to the values associated with the geochemical calibration. The only exception is the permeability of core B38D which demonstrates a relatively high signal strength of 0.90. The PPR varies between 6.0 for the permeability and 12.6 for the porosity (both in core E10-3), suggesting that the quality of the predictions is in all cases lower than that of conventional petrophysical analysis. Compared to conventional analyses, the core scanner has the highest predictive performance for the permeability in core E10-3 whereas performance for the porosity in core E10-3 is lowest. The permeability in the soft-sediment core B38D scores roughly in between these two extremes with a PPR of 8.05.

The MSI values we found for core E10-3 vary between 81% and 94%, suggesting that the set of calibration data is sufficient to constrain the calibration models. The stability associated with the permeability model of core B38D is only 26%. For this core, more calibration data are therefore likely to increase the predictive performance.

7.5 Discussion

7.5.1 Geochemical calibration

For cores comprising effectively all different lithologies or sediment types, the core scanner apparently performs up to three times worse than conventional analysis. The summary

	E10-3 Grain density [gr/cm ³]	E10-3 Porosity [%]	E10-3 Permeability [mD]	B38D Permeability [m/D]
X-vars excluded	Cr, Cu, Mn, Ni	Cr, Cu, Y	Cr, Cu, Ga, Mn, Y	none
PLS comp.	12	8	9	10
R_{eff}^2	0.75	0.72	0.82	0.91
MSPE	9.11e-5	2.73e-5	1.07	1.61
MSI%	94.2907	81.39	86.90	26.22
MRPE%	1.35	16.49	116.72	161.02
PPR _{petro}	6.41	10.99	5.84	8.05

Table 7.3: Summary of the petrophysical calibration results.

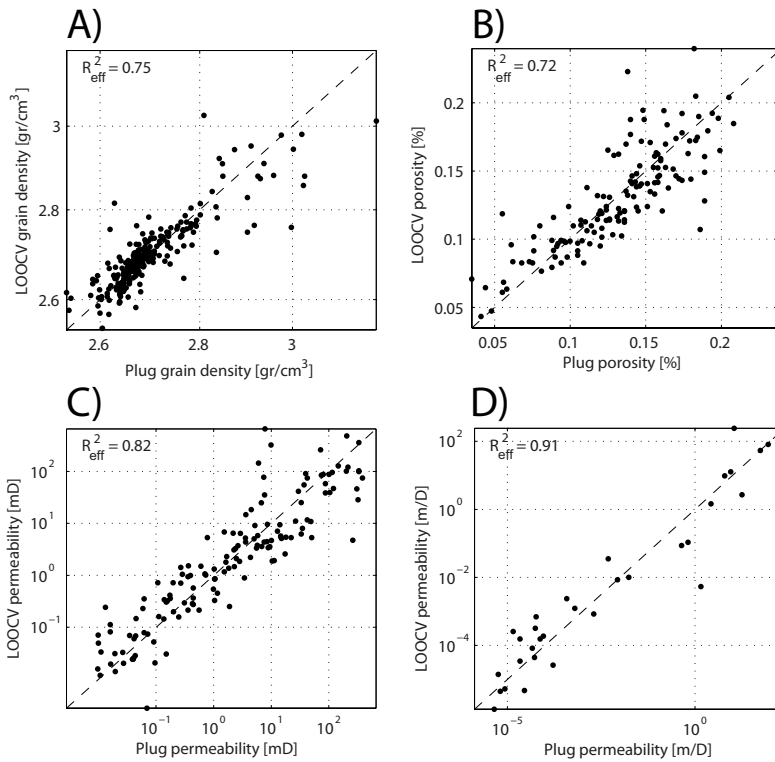


Figure 7.7: Cross plots between the plug data and the predictions (LOOCV) for the grain density (A), porosity (B) and permeability (C) of core E10-3, and the permeability of core B38D (D). Each cross plot shows the R_{eff}^2 and the line $y=x$.

statistics associated with the reduced set of calibration samples of core B38D show that limiting the model's application range to siliciclastics has a positive effect on the quantitative performance. The PPR_{chem} of core B38D without the ICP-OES and glass-bead XRF analyses was as small as 1.57, indicating that the performance of the core-scanning methodology is comparable to that of conventional analysis methods. An explanation for the fact that the PPR_{chem} remained constant in core E10-3 is that there was only one coal sample excluded in combination with the robust definition of PPR_{chem} (Eqn. 7.11). Supported by the fact that the lowest MSI values were around 50%, we want to stress that if more calibration data would have been available, the PPR value is likely to decrease.

Comparing the first and second calibration results yields that the XRR in core E10-3 decreased whereas that in core B38D slightly increased. An explanation for the reduced XRR in core E10-3 is that many low-intensity elements were excluded from the model (e.g., Cr, Cu, Ni and Y) probably because they primarily reside in the coal-bearing intervals (see Fig. 7.2). In core B38D, on the other hand, the set of included intensity variables was practically equal. For the complete set of calibration data, the XRR values were 0.74 (E10-3) and 0.35 (B38D). This implies that the magnitude of repeatability errors is slightly smaller than the inter-laboratory error associated with conventional analyses. Comparing the XRR to the PPR values, however, shows that repeatability errors have a marginal contribution to the total prediction error.

An important aspect is whether the inability to predict the analyte concentrations in samples with a low concentration of siliciclastics is due to incorrect predictions of scale (i.e., the total concentration of quantifiable analytes) or of composition. Therefore we analyse the predictions of samples that were excluded in the second calibration exercise and compare them to the measured concentrations (i.e. the calibration data). The results are shown in Figure 7.8. In core E10-3 we observe that the predicted scale for the coal sample is 63% whereas the scale determined using destructive XRF is 7.5%. Apparently, scale cannot be accurately predicted by extrapolation for these cores. For core B38D, the core scanner seems to be better at predicting the composition than the scale. We only observe appreciable bias for Zn and CaO, but it is beyond the scope of this contribution to find an explanation for this bias.

Carbon does not produce fluorescent energy and is therefore not explicitly resolved in the intensity data set. A high carbon content will cause the intensities of all "active components" to decrease (i.e., dilution effect). However, the proposed modelling approach is based on clr-transforming the intensities which means that information about the total intensity is removed. The reason for doing so is to obtain a model which is less sensitive to variations in the measurement geometry and which is therefore more robust. The consequence of this approach, however, is that prediction of compositional scale in coal- and peat-bearing intervals is possible only in case the scale correlates with any of the measurable components (e.g., S intensities). The success of this approach is highly case-specific.

7.5.2 Element intensity importance

The VIP values associated with all initial calibration models (i.e. before X-variable removal) as well as the associated cross-correlation coefficients are shown in Figure 7.9. Their high VIP values indicate that Ca, Fe, Si were important in all calibration models. Elements which were relatively unimportant for the prediction of the various core properties are Cl (B38D), Cr (E10-3), Y (E10-3) and Rh (B38D).

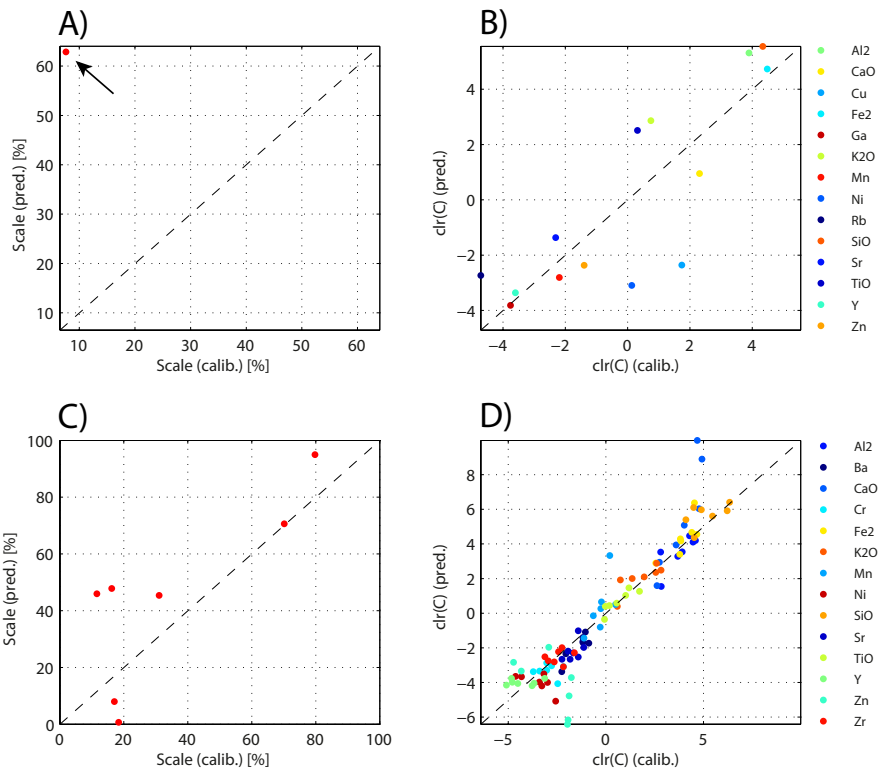


Figure 7.8: Cross plots between the plug and predicted (LOOCV) chemical scale (A, C) and clr-transformed composition (B, D) of cores E10-3 (A, B) and B38D (C, D). All cross plots show the line $y=x$.

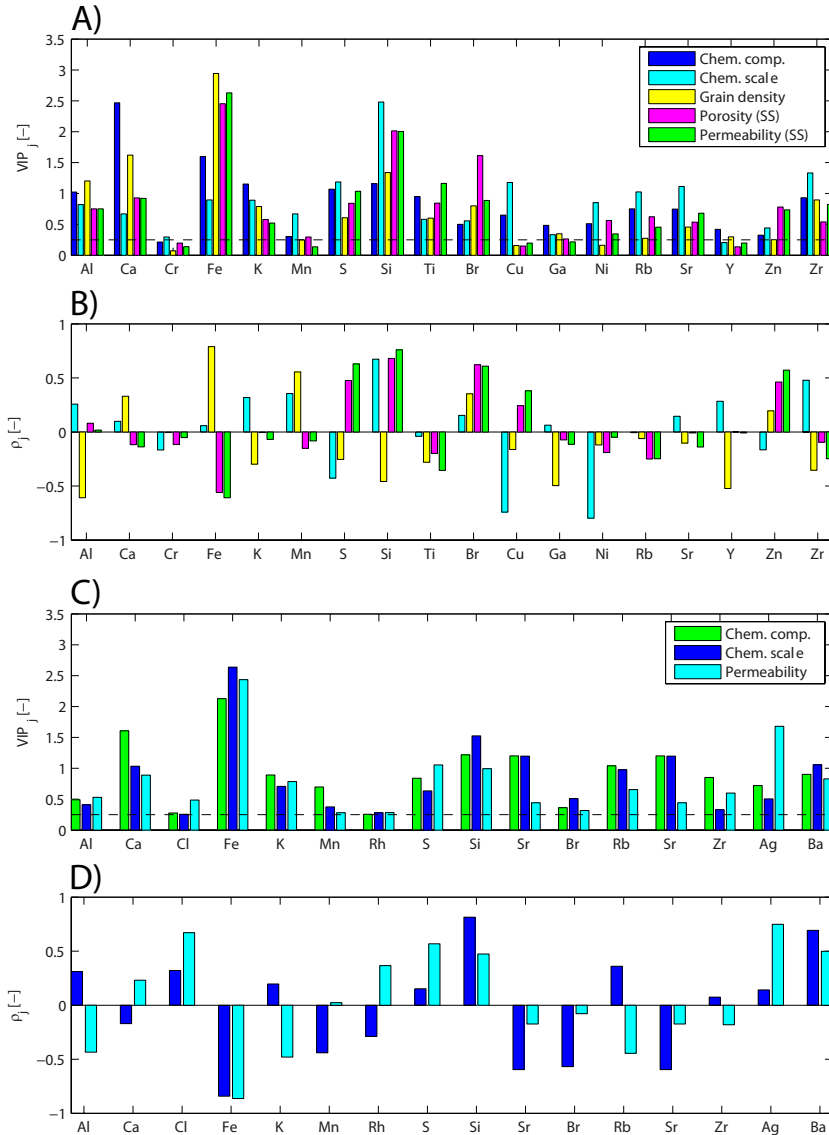


Figure 7.9: Prior VIP (A,C) and cross-correlation coefficients (B,D) associated with the calibration models of core E10-3 (A,B) and core B38D (C,D).

Typically, a high VIP was accompanied by a high cross-correlation coefficient (see Fig. 7.9), suggesting that variables were included in the model *because* they demonstrate a monotonic relation with that particular property. For instance, Si has a high VIP values and a strongly positive correlation coefficient with the chemical scale in both cores. This can be explained geochemically by considering that the most prominent unobservable chemical component is carbon which, in turn, is found exclusively in Si-poor intervals. Another example is Fe in core E10-3 which displays the highest VIP value for the grain density and also the strongest (positive) correlation coefficient. Fe is important for the prediction of grain density because (i) this core contains high concentrations of Fe-bearing minerals, and (ii) Fe-bearing minerals have a high density. In contrast to core E10-3, however, Fe is strongly negatively correlated with chemical scale in core B38D. In this core, Fe turns out to be enriched in intervals with a high concentration of organic material. With Fe₂O₃ concentrations not exceeding 10%, Fe intensities most likely reflect variations in pyrite content (Griffioen et al., 2012).

The majority of the element intensities with a high VIP were acquired in the 10kV run. For core E10-3, the calibration models which benefited most from intensities acquired in 30 or 50kV runs are the chemical scale (Cu, Ni, Rb, Sr, Zr) and the porosity (Br). The former set of elements is important for prediction of chemical scale probably because these elements are primarily enriched in coal layers characterised by a small compositional scale. Given that the formation water contained substantial amounts of Br ions (> 7 mg/L), the Br intensities may reflect the amount of precipitated formation water, which is presumably proportional to porosity. However, this interpretation is by no means conclusive because we were unable to properly calibrate Br.

Core B38D was also scanned at 50kV yielding Ba and Ag intensities. The VIP values show that in this core, Ba was moderately important for the prediction of the petrophysical properties. Biplot analysis (Chapter 6, Fig. 6.8) showed that Ba is primarily a proxy for sandy sediments. Including Ag intensities in the calibration model was only beneficial for the prediction of permeability. As we already discussed, Ag is only present in the instrument and probably reflects the scattering properties of the sediment which will be influenced by e.g., water content and porosity. This shows that including intensities which have little geochemical significance may increase prediction performance.

7.5.3 Extrapolation abilities (histogram analysis)

By inferring the predictive capabilities using LOOCV it is implicitly assumed that predictions are reasonably close to the calibration data. We can gain more insight into the limitations of the proposed methodology by analysing histograms of measured and predicted permeabilities (Fig. 7.10). These histograms demonstrate that the predictions of core B38D (Fig. 7.10, B and D) resemble the two populations in the calibration data set, reflecting the sandy and peat/loam intervals. However, the predictions exceed the maximum measured permeability in quite a few cases up to three orders of magnitude. High-end exceedances are less common in core E10-3 (Fig. 7.10, A and C). In this core, however, there are a considerable number of permeability predictions which are much smaller than the lowest measured permeability.

Inspection of core B38D reveals that high-end exceedances occur exclusively in coarse-grained sediment intervals. The grain size in these intervals range up to gravel. The most permeable intervals are characterised by a high quartz content. In case the core-

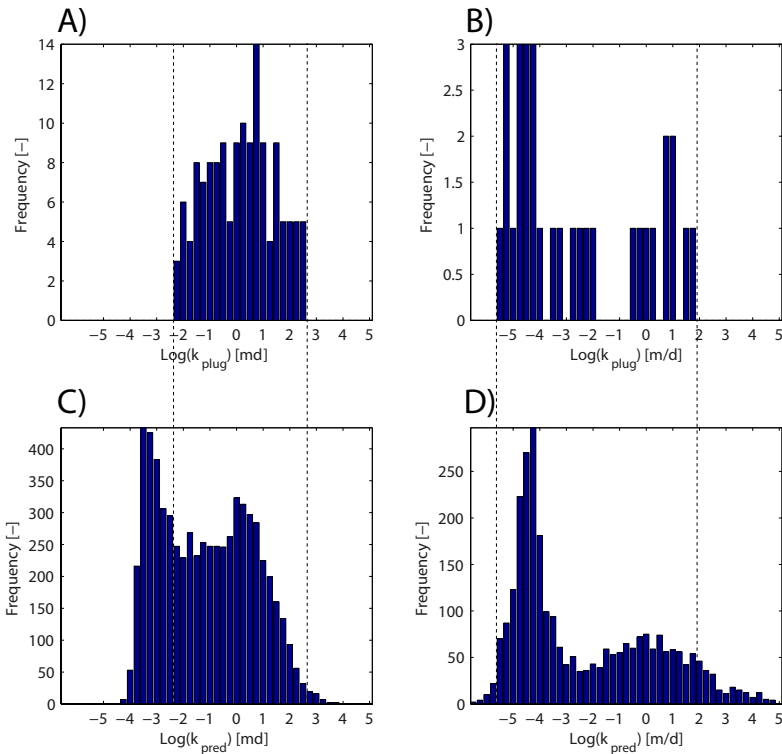


Figure 7.10: Histograms of the permeability values in the calibration data set (A,B) and the core-scanning predictions (C,D) for core E10-3 (A,C) and B38D (B,D).

scanning measurement corresponds to a single pebble which consists almost exclusively of quartz, these scanning measurements were therefore erroneously "interpreted" as an ultra-permeable sand (i.e., more permeable than the maximum measured permeability). The fact that these exceedances are less common in core E10-3 suggests that also the stability of the calibration model plays a role. Hence, we conclude that the integrity of the predictions is controlled by the representativity of the scan measurement for the sediment bed, and the size of the calibration data set.

7.5.4 Extrapolation abilities (explicit poro-perm relation)

To further investigate the integrity of the core-scanning-based predictions, we study the poro-perm relation of the independently calibrated predictions of core E10-3 and compare this relation to that of the plug data. This "explicit" poro-perm relation is shown in Figure 7.11A only for the plug positions. In this figure, also the plug data are shown which were not included in the calibration procedure. The relation associated with all scan positions is shown in Figure 7.11B. To facilitate interpretation, the assigned and predicted lithofacies class are indicated for the plug and core-scanning data, respectively.

Figure 7.11A shows that the poro-perm relation of the plug data is very similar to that

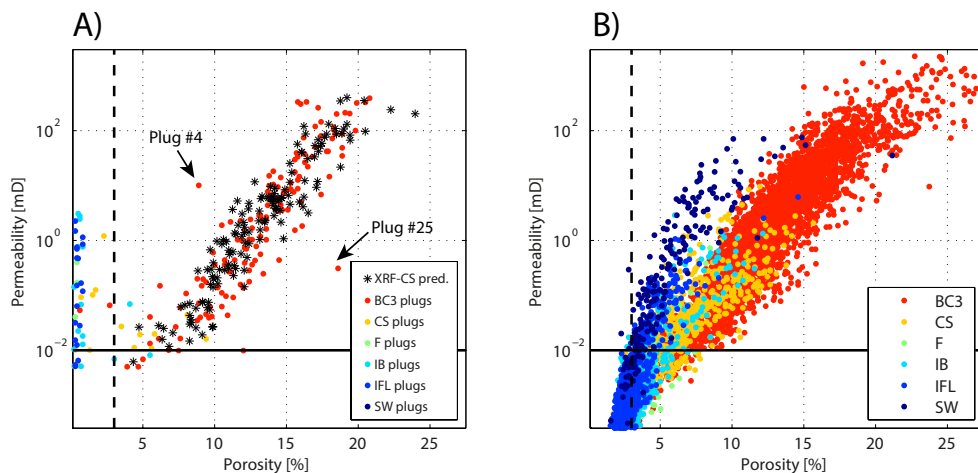


Figure 7.11: (A) Plug porosity versus plug permeability as well as the core scanning predictions at the plug positions (black stars). For the plug data, the color indicates the lithofacies association of the core description. (B) The predicted (LOOCV) porosities and permeabilities for the whole core. Colors indicate the predicted lithofacies association.

of the associated predictions. This underlines the integrity of the core-scanning-based predictions within the calibrated range. Analysing the predictions outside the calibrated range (Fig. 7.11B), however, yields that the predictions tend to deviate from the main poro-perm relation both at the high and the low end. Outside the range of calibration values, the poro-perm relation of the predictions shows a curvilinear relation. On the low end, we suspected the plug permeability data to be incorrect due to leakage along fractures and the Hassler sleeve. The predicted poro-perm relation is more in line with what we expect, namely positive correlation between porosity and permeability along the entire range. With regard to the deviation at the high end we cannot draw any firm conclusions, because, as we discussed in Section 7.5.3, the representativity of the XRF-CS measurements deteriorates as the grain size increases. We suspect that the curvilinear appearance of the extrapolations is to some extent controlled by the logarithmic transformation that is applied to the porosity before linking it to the intensities.

Two plug measurements which do not fall within the general poro-perm trend (indicated with arrows in Figure 7.11A) are not reflected by the predictions at the plug locations. For a given porosity, plug #25 has a lower than expected permeability whereas plug #4 has a higher than expected permeability. The sedimentological description at the position of the former mentions cm-scale mudstone and siltstone intraclasts whereas at the position of the latter large concretions of anhydrite were found. The presence of a large intraclast or concretion inside a plug may have caused their deviation from the main poro-perm trend, which would justify labelling them as "outliers". In contrast to plug #25, however, the poro-perm predictions for the whole core (Fig. 7.11B) suggest that the combination of porosity and permeability of plug #4 is quite common in core E10-3. This would imply that this plug measurement is not an outlier but a sample from an important subpopulation characterised by a different poro-perm relation. As demonstrated by Figure 7.11B, how-

ever, this subpopulation includes all coal-bearing *SW* intervals. Poro-perm predictions are unreliable because they are not supported by plug data. An example of such predictions may be observed in ES5 (Fig. 7.3). Furthermore, plug #4 is not coal-bearing but a sandstone plug. These two factors suggest that plug #4 is indeed not representative for the core.

7.6 General discussion, conclusions and recommendations

7.6.1 Quantitative performance evaluation

The quantitative performance of the XRF core scanner is generally lower than that of conventional geochemical and petrophysical analyses. Based on this study and a PPR threshold of 2, the proposed methodology can be considered a full-fledged alternative *only* for conventional geochemical analysis of unconsolidated-sediment cores and *only* for siliciclastics-dominated sediments. Ranking the calculated predictive performances in decreasing order yields: geochemistry, permeability, grain density, porosity. The fact that the highest PPR is that of the geochemical composition is not surprising, given that XRF is a chemical technique.

We derived that repeatability errors in the intensities are relative small compared to the prediction error. Hence, longer measurement times and more stable spectrum interpretation software are not the key to smaller prediction errors. The MSI values suggested that for most properties (in particular the bulk geochemical composition) the PPR will increase when more calibration data become available. The results also showed that extrapolations in the Y-data range are of questionable integrity. In addition, the explicit poro-perm relation showed that the integrity of the predictions is also determined by the X-data range, i.e., predictions associated with XRF-CS signatures which are very different from the calibration data are questionable. For core E10-3, a pragmatic solution would have been to use only the predictions which (i) fall within the range of the calibration data, and (ii) which were assigned the *BC3* lithofacies class.

The performance criterion PPR we defined is based on the premise that all calibration data were internally consistent. Given the professionalism of the laboratories involved, this is a reasonable assumption. For the geochemical data, however, this assumption was possibly violated because determining the geochemical composition of sediments with a carbon content of more than 90% is also pushing the limits of conventional XRF analyses. The inability to predict the compositional scale is therefore probably influenced by the limited number of calibration samples, as well as systematic errors in the calibration data.

7.6.2 Possible improvements

Possible improvements are stratified modelling (i), flagging of extrapolations (ii) and reduction of sampling errors (iii).

The proposed modelling framework is founded upon mathematical and statistical considerations. Modelling intensities by means of logratios was assumed to remove the variable measurement geometry and linearise matrix effects (Weltje and Tjallingii, 2008). As these authors showed, these assumptions hold for cores with little lithological variability such as marine soft-sediment cores (Chapter 2). It is very likely, however, that this assumption does not hold for cores comprising more lithological variability, such as cores E10-3 and

B38D. Instead of setting up one calibration model, setting up different calibration models stratified by lithology is therefore likely to improve predictive capabilities.

Extrapolations made by the model in both the X and the Y space may be unreliable. Extrapolation may be prevented by carefully selecting calibration samples (see Weltje et al. (2015) and Chapter 2, this thesis). In the hydrocarbon industry, however, cores are plugged prior to slabbing which implies that sample selection based on XRF-CS data is not possible and prediction by extrapolation cannot be avoided. For these cores it is important to have a safety mechanism which can 'flag' extrapolation by the model in both the X and the Y space in an automated manner, for instance by determining the convex hull in both multi-dimensional spaces. Predictions would then only be marked as reliable in case both X-data and Y-predictions fall within the two convex hulls.

Sampling errors are introduced because the scanner did not measure the same sample volume as the conventional analytical instrument: the core scanner analysed the core slab right next to the plug location whereas the petrophysical properties were derived from the plug sample as a whole. In the future, it may be attempted to mitigate sampling errors by scanning the plugs after trimming and before they have been cleaned.

7.6.3 Concluding remarks

It is evident that mitigating sampling errors may reduce the prediction error. The question is, however, whether reducing this prediction error is *necessary* and even advisable. In order to answer that question, we should consider the objective of core scanning and core analysis in general: the objective is not to accurately characterise the uppermost millimeters of a core slab or the core slab as a whole, but to characterise the properties of a stratigraphic interval in the subsurface. Reducing the inferred prediction error by mitigating sampling errors may result in a false sense of certainty because the inferred prediction error may become smaller than the geological uncertainty. In that case, we ignore the intrinsic uncertainty associated with cores as samples of the subsurface. On the other hand, mitigating sampling errors will give better calibration models because they require fewer data to estimate their parameters.

To make use of the improved calibration models by mitigating sampling errors requires a hybrid approach. Firstly, establish calibration models from closely-linked core scan and calibration data, yielding a calibration model with a minimal associated prediction error. Secondly, quantify the intrinsic variability of these predictions in a particular application, for instance by determining the variance in sediment properties parallel to the bedding direction. In the light of the objective to characterise the subsurface, a realistic uncertainty of the core-scanning data is then given by the sum of the prediction error associated with the calibration model, and the sedimentological error associated with the bed-parallel compositional variability.

Although the results showed that the XRF-CS predictions are inferior to conventional geochemical and petrophysical analyses, what should be kept in mind is that XRF-CS and conventional methods are not directly comparable. In practical applications it is simply impossible to acquire conventional geochemical or petrophysical analyses on a 1-cm resolution due to financial and time constraints, but also due to the fact that conventional methods are destructive. Generally it is unwanted or even not allowed to sample a core so intensively that its integrity is jeopardised. Hence, in cases where the core is heterogeneous and the objective is to make a detailed description of the core, there is simply no

conventional alternative to in-situ core-scanning techniques. We will illustrate the value of the integrated core description of E10-3 in Chapter 8.

Integrated core analysis using XRF scanning. Part III: Applications

8.1 Introduction

8.1.1 General introduction

In Chapter 6 and 7, new methods were introduced for the prediction of lithofacies, chemical and petrophysical properties from XRF-CS data. The high resolution predictions of lithofacies, chemical and petrophysical properties together constitute the "integrated core description". Two cores were analysed: core E10-3 constituting consolidated sediments and core B38D containing unconsolidated sediments. In this chapter, the added value of the "integrated core description" is investigated for core E10-3.

The added value of the integrated core description will be explored by means of three applications. In these applications, we combine the integrated core description with all other available core data, i.e., also with those which were not linked to the XRF-CS data in Chapters 6 and 7, such as thin-section and XRD analyses. Ultimately, this leads to a far more detailed description of core E10-3 in terms of reservoir quality and its controlling mechanisms: a detailed description which could not have been made based on the data acquired in routine core analysis.

8.1.2 Core E10-3: controls on reservoir quality

Hereafter we summarise the conclusions drawn with regard to the factors controlling reservoir quality in core E10-3, as given in the core report (see Boels, 2003). The sediments of core E10-3 are dominated by three sandstone intervals, which were interpreted as braided-channel deposits (Boels, 2003). The reservoir quality of the sandstones is mainly controlled by variations in grain size, and diagenetic overprint. Early-diagenetic kaolinite and hematite reduces reservoir quality. Pore-filling kaolinite formed as a result of grain dissolution and hematite replaced matrix grains. Its occurrence could not be distinguished from grain-

size variations, because matrix grains are more abundant in the fine-grained sandstones. Quartz precipitation during intermediate diagenesis enhances reservoir quality. Ferroan dolomite is late diagenetic and destroys porosity. According to Boels (2003), however, its occurrence is too local and little to have a large effect on the reservoir quality. Furthermore, diagenesis was qualified as heterogeneous in core E10-3, and it was not attempted to reconstruct the relative importance of these mechanisms: the resolution of the data at hand (i.e., 10 thin section petrographical and 10 XRD analyses) was not sufficient for this task.

8.2 Application 1: Enhanced facies properties

8.2.1 Objective

One of the most important goals of core analysis is to obtain reliable estimates of petrophysical properties which, in turn, serve as input for reservoir models. The standard approach is to drill plugs at regular intervals (e.g., 1 ft or 25 cm). The rationale behind this approach is that the plugs are a representative sample of the core and that the plugs are sufficient to constrain the average poro-perm values associated with each lithofacies. The core scan, however, covers the entire core on a virtually continuous scale which means that, given that the predictions are unbiased, the nearly continuous poro-perm predictions are a representative sample by definition. Our objective is to investigate the impact of using the core-scan predictions instead of plug data to calculate the average porosity and permeability of a particular grain-size group or lithofacies association.

8.2.2 Approach

In addition to the conventional approach whereby the average class properties are calculated by taking the mean value over all plug data acquired from that particular class (i.e., method 0: see Figure 8.1A), we propose two alternative methods which rely increasingly more on the XRF-CS predictions. Method 1 is based on calculation of the average porosity and permeability as predicted from the core scan, however, using the original core description (see Figure 8.1B). Comparison of the outcome of method 1 with the conventional approach gives insight into the effect of inadequate sampling on the lithofacies properties. Method 2 is similar to the first, but the predicted core description (i.e., predicted on the basis of XRF-CS data) is used to label porosity and permeability predictions (see Fig. 8.1C). Comparing the results of method 1 and 2 gives insight into the relative importance of using a predicted core description instead of the original core description for prediction of lithofacies properties.

8.2.3 Results and discussion

In our analysis, we only include *BC3* (lithofacies) and *S1-S5* (grain-size group), and only the porosities between 3% and 20.8% and permeabilities between 0.01 and 387 mD (i.e. the calibrated ranges). We derive average poro-perm values by employing the arithmetic mean for the porosity, and the geometric mean for the permeability. The average porosities and permeabilities of each class, derived using the different methods, are shown in Figure 8.2 (A and B).

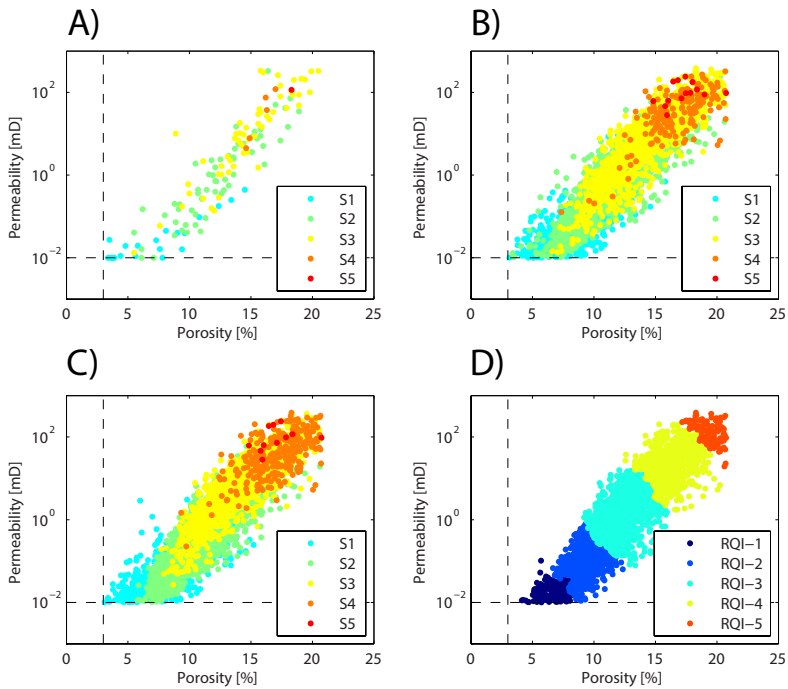


Figure 8.1: A, B and C illustrate the various methods (method 0, 1 and 2, respectively) that were used in application 1 to derive average grain size group properties: average plug data (A), average predicted poro-perm using existing core description (B) and using updated core description (C). D shows the poro-perm predictions labeled with their assigned RQI class, as derived in application 2 (Sect. 8.3).

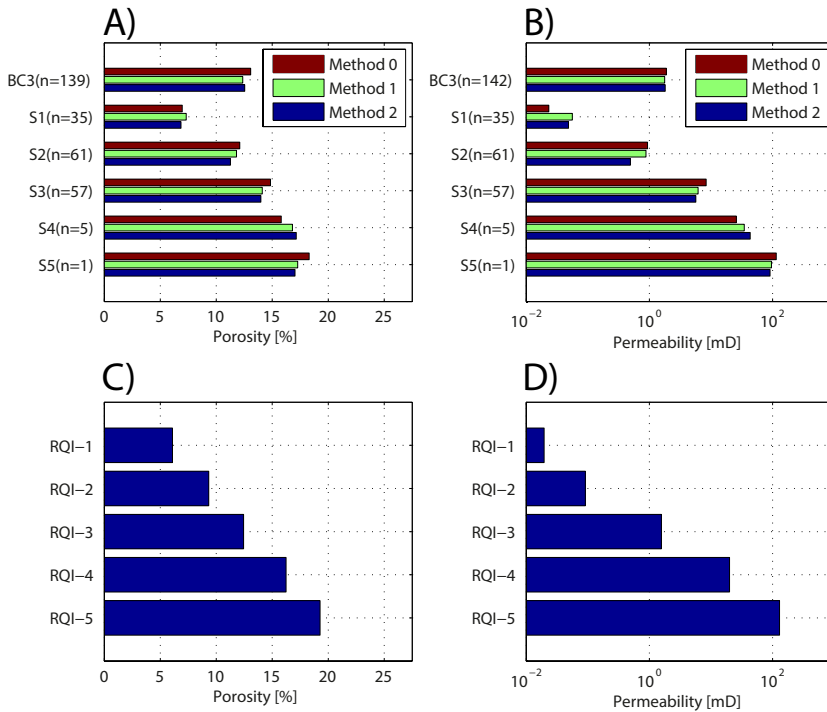


Figure 8.2: A and B show the average poro-perm values, as estimated with the three methods. C and D show the average poro-perm values associated with the RQIs (See Sect. 8.3 for discussion).

We observe that the main differences are among the conventional method and the two alternatives. Hence, it is not so much the change in the classification scheme but merely the increase in resolution which influences the average poro-perm per class.

We also observe that the methods yield similar poro-perm values for *BC3*. The reason why the differences are so small for *BC3* is that it was already frequently plugged, which means that the average poro-perm values are well-constrained. For most other classes, however, the alternative methods yield noticeably different average poro-perm values. According to the conventional method (method 0), for instance, grain-size group *S5* has higher poro-perm values than those derived using methods 1 and 2, whereas the opposite applies to grain-size group *S4* (Fig. 8.2, A and B).

In terms of magnitude, the differences in porosity estimates derived using the methods is small (1%). For the permeability, the spread among the methods ranges from 67% for *S1* and 25% for *S5*. Although the relative spread is smallest for *S5*, this spread may have the largest impact on associated reservoir models. The 25% spread of *S5* translates to 20 mD difference between method 0 (114.8 mD) and method 2 (91.2 mD).

8.3 Application 2: Automatic reservoir quality prediction

8.3.1 Objective

Core descriptions serve different purposes, i.e., they may reflect the lithology of the core or they may reflect an interpretation of the environment in which the sediments were deposited. Under particular circumstances, lithology in combination with sedimentary environment may be a good indicator for the porosity and permeability and therefore of the reservoir quality. In other cases, it is impossible to obtain a continuous description of the core in terms of reservoir quality without having to employ specialised measurement techniques such as mini-permeametry. Our next goal is to use the core-scanning predictions to automatically describe the core in terms of a set of reservoir quality classes with unique properties in an objective manner.

8.3.2 Approach

One way to automatically subdivide the core into classes with distinct poro-perm values is to apply some sort of binning to porosity and permeability values. A more rigorous approach is to design a new classification scheme with non-overlapping classes in the desired space (in this case the porosity-permeability space) using cluster analysis. We use hierarchical clustering based on Ward's distance measure (Ward, 1963) to construct such a classification scheme. The methodology presented here is suitable for developing any kind of classification scheme by using a different set of variables (e.g., geochemical facies, grain-density facies). The only parameter which has to be set when employing cluster analysis is the desired number of clusters.

Because the poro-perm relation is generally modelled in a linear-logarithmic space, clustering is applied to the raw porosities and to logarithmically-transformed permeabilities. To ensure that the variables are given equal weight, they are normalised to unit variance. After cluster analysis, the mean poro-perm values are calculated for every cluster. The classes are subsequently ranked according to their mean permeability. We refer to this classification as the Reservoir Quality Index (RQI) scheme.

8.3.3 Results and discussion

In core E10-3 only the Braided Channel Units (BCUs) have significant reservoir potential. Therefore, only those poro-perm data were included whose predicted lithofacies is *BC3*, and all other poro-perm data were labeled as *RQI-0*. The included data were divided into five clusters, i.e., the same number of sandstone grain-size groups as used in the original core description.

The poro-perm relation labeled with the RQI classes displayed in Figure 8.1D shows that, as expected, RQI classes are always non-overlapping. Assuming that the aim of a classification scheme is to continuously describe the core in terms of classes with unique reservoir qualities, this scheme is superior to the original grain size and lithofacies classification schemes. The poro-perm relation also shows that for this particular core, binning the porosity and permeability along their major axes would have given a similar result, which is due to the fact that the poro-perm relation is monotonic. This, however, may not always be the case.

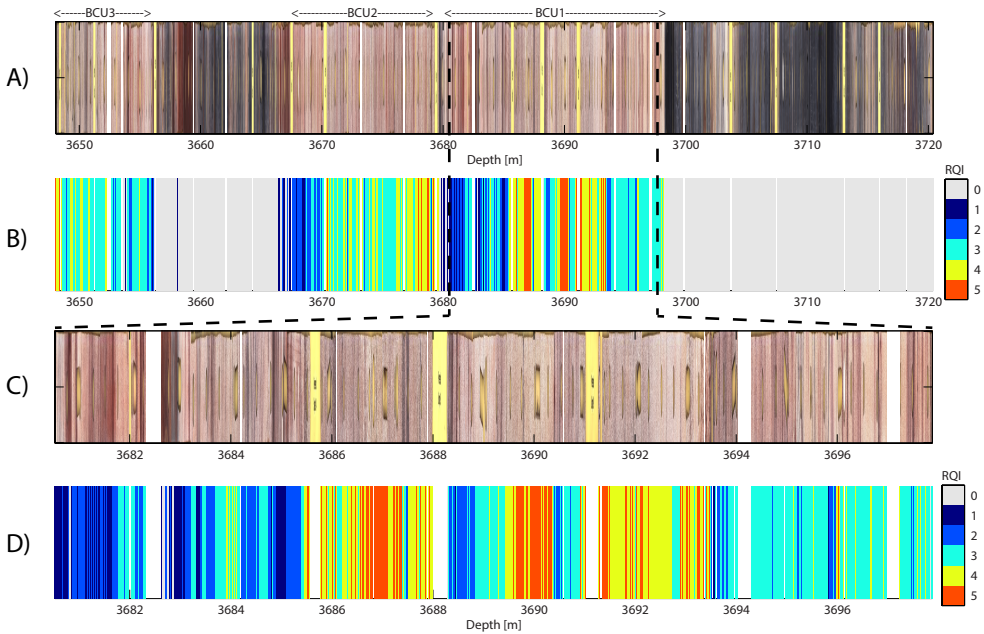


Figure 8.3: RQI classification along for the entire core (A and B), and for BCU1 only (C and D).

Comparing the average poro-perm values of the RQI classification with the grain-size scheme (Figure 8.2) indicates that the properties of *RQI-5* are more favourable than those of *S5* whereas the predicted abundance of *S5* is much lower than that of *RQI-5*. Despite the RQI classification scheme being based upon the same data set, it suggests that core E10-3 has intervals with better reservoir quality than implied by the lithofacies and grain size descriptions. The RQI scheme is more specific, i.e., the average poro-perm values of each class are a better approximation of the cm-scale poro-perm predictions.

The RQI classification along core (Fig. 8.3) readily shows that the reservoir quality is heterogeneous. Moreover, there are some noticeable difference between the BCUs. Intervals with high (RQI-3 and higher) reservoir quality are found mainly in BCU1: it is the only BCU which contains RQI-5 intervals with appreciable thickness. A difference between BCU1 and BCU2 is that whereas in the former the highest quality sandstones are found in the middle, in the latter the highest reservoir quality is found at the base. The highest quality sandstones in BCU3 are found at the top.

These results are illustrative for the disadvantages of the RQI scheme relative to conventional lithofacies descriptions. RQI classes are difficult to interpret geologically because the reservoir quality is controlled by different, possibly independent mechanisms. Our next challenge is to unravel the control on this heterogeneity by integrating the RQI description with the other available data sets.

8.4 Application 3: Diagenetic controls on reservoir properties

8.4.1 Objective

In Chapter 7, we discussed the importance of grain-size variations for the reservoir quality of core E10-3. Thin-section analysis yielded that reservoir quality is controlled by different mechanisms whose importance varies strongly throughout the cored interval (Boels, 2003). In the core study, however, it was not attempted to reconstruct the relative importance of these mechanisms due to the limited data that were at hand. Our objective is to investigate to what extent core-scanning data can help to make a more detailed description of the diagenetic overprints in the stratigraphic record.

8.4.2 Approach

Using the classification of the core in terms of RQI and the geochemical record, we may distinguish two types of mechanisms: those that control the reservoir quality (positively and negatively), and those that have no effect on reservoir quality. Under the assumption that all relevant mechanisms have a geochemical response, our goal is to subdivide the signals of the different chemical elements into three groups, and associate them with phenomena that were distinguished by thin section analysis. We use a methodology which we refer to as *crosscorrelation-variance analysis*. This method is based on classifying all chemical elements based on crosscorrelation coefficients between the clr-transformed concentrations and the properly-transformed poro-perm. Pearson's (cross)correlation coefficients are sensitive to linear relations of, in this case, elemental abundance and poro-perm values. This approach can be considered a univariate implementation of the method proposed by Bloemsma et al. (2012) (Chapter 3, this thesis).

The elements that correlate significantly (positively and negatively) with porosity and permeability are placed in group 1. The other elements are subdivided into those that have a high variance (group 2), and the elements which have a low variance (group 3). The elements in group 1 are considered to reflect the geochemical response of reservoir-quality-controlling mechanisms and may be readily interpreted. The elements in group 2 reflect the response of an important mechanism, whose response is not monotonically related to the poro-perm values. Because this relation may also be non-monotonic, it is important to study the nature of the relation between the abundance of group 2 elements and poro-perm values graphically by means of cross plots. The elements in group 3 do not deserve further investigation because they probably reflect noise.

After this classification step, the correlation *among* the elements is examined using biplot analysis. The results of the biplot analysis are combined with XRD analysis, petrographical analysis and general sedimentological and geochemical principles.

Generally, the significance of correlation coefficients is determined using a statistical criterion which is based upon the null hypothesis that two properties are uncorrelated. Evaluation of the t-statistic-based criterion (Eqn. A.7 in App. A), for instance, yields for $\alpha = 0.05$ and $N = 142$ (i.e. the number of plugs used to constrain the permeability calibration), a critical value of 0.16. In this case, however, testing against the null hypothesis of uncorrelatedness is pointless, given that the predicted concentrations and the poro-perm predictions are by definition not independent: they were both generated from the same set of intensities which disqualifies this strictly statistical approach. We therefore adopt a

relatively high (but arbitrary) correlation-coefficient threshold of 0.25. It is also difficult to define a variance threshold, which determines whether elements are placed in group 2 or group 3. We choose to place the two elements with the highest variance in group 2.

8.4.3 Results and discussion

The crosscorrelation-variance analysis is limited to the viable reservoir units by considering only RQI-1 up to RQI-5. The results are shown in Figure 8.4 and the biplot of the geochemical composition is shown in Figure 8.5. Figure 8.4 shows that for both porosity and permeability, group 1 contains SiO_2 and Cu (positive correlation), and Y and Fe_2O_3 (negative correlation). For both porosity and permeability, group 2 contains CaO, MnO and Ni.

Group-1 elements

The simplest mineralogical explanation for the positive correlation between permeability and SiO is that SiO reflects the quartz content. The quartz content is generally positively correlated with grain size, and the grain size controls the pore size and therefore the permeability. Fe_2O_3 has replaced matrix grains. Because matrix grains are primarily abundant in fine-grained sandstone, this diagenetic signal cannot be distinguished from variations in detrital grain size.

Relating Cu and Y uniquely to a particular mineral would require single-grain analysis. However, analysis of the biplot (Fig. 8.5) shows that SiO_2 and Cu are highly positively correlated, suggesting that Cu is positively correlated with grain size. Possibly, Cu is contained in heavy minerals which are deposited under the same conditions as large quartz grains.

Y is a relatively immobile accessory element found in biotite, pyroxene, feldspar and several high-density minerals (e.g., garnet, zircon). The compositional biplot (Fig. 8.5) shows that Y is strongly linked to a suite of other elements including Rb and K_2O . These elements are generally associated with clay minerals. Following hydrodynamic principles, the concentration of clay minerals is determined by the mean grain size and the sorting which, in turn, influence porosity and permeability. Y is probably enriched in the weathering product of the above-mentioned minerals (i.e., clay minerals), which would make Y a proxy for the clay content.

Our interpretation of the crosscorrelation-variance analysis results is that reservoir quality is controlled by various mechanisms which, however, cannot be distinguished from grain size. Given that variations in geochemistry mainly reflect variations in grain size (Bloemsma et al., 2012), one carefully selected geochemical proxy may therefore give a good indication of reservoir quality. A proxy suitable for this purpose would be the log-ratio between the element with the most positive and the most negative correlation coefficient. For both the porosity and permeability this yields $\text{Log}(\text{SiO}_2/\text{Y})$ which is displayed against porosity and permeability in Figures 8.4C and 8.4D, respectively.

Group-2 elements

CaO and MnO are not (linearly) correlated with reservoir quality and they constitute dominant geochemical signals. Hence, they merit closer investigation. The biplot showed that CaO and MnO are positively correlated and therefore probably associated with the same

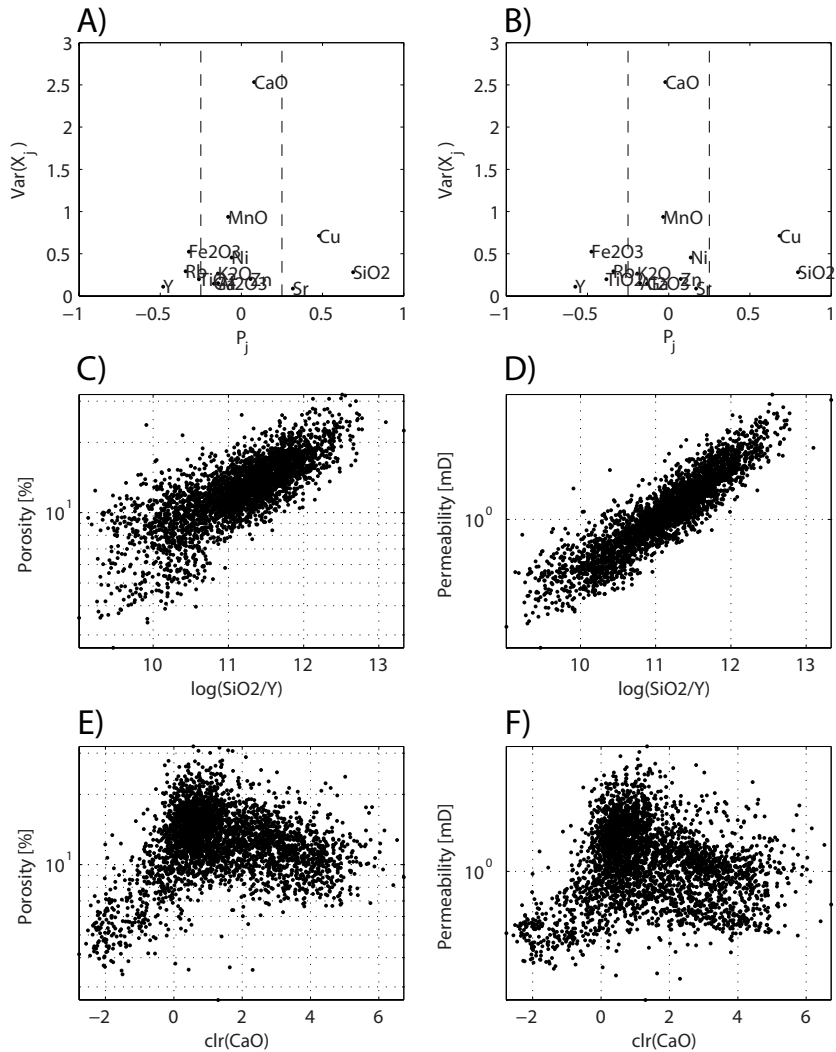


Figure 8.4: Results of the correlation-variance analysis conducted in application 3. A and B show the correlation coefficient against the variance of the different element concentrations for the porosity and permeability, respectively. C and D show the selected geochemical proxies whereas E and F show the non-correlating element with the highest variance.

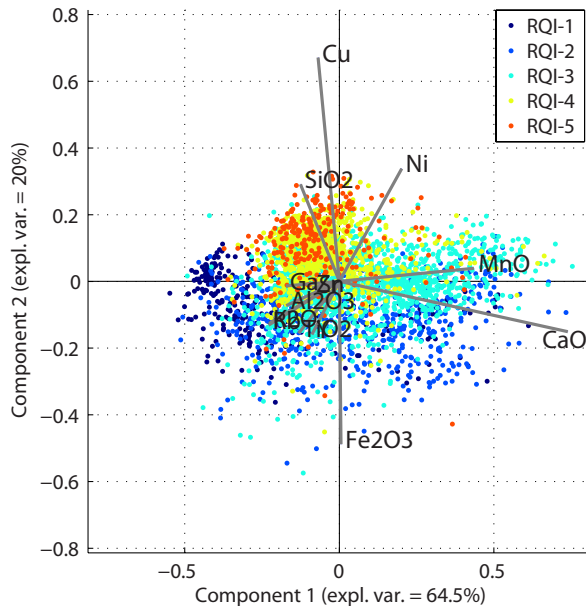


Figure 8.5: Biplot of the predicted geochemical composition with the RQI classification scheme indicated. The two principal components together explain 85% of the geochemical variability.

mechanism. Cross plots between CaO and both porosity and permeability (Fig. 8.4, E and F) show that poro-perm is not independent from CaO, but not related in a monotonic (linear) manner.

For clr -values below unity, Ca is positively correlated with porosity and permeability. For clr -values higher than unity, CaO is negatively correlated with porosity and permeability. A clr -value of one corresponds roughly to a concentration of 0.1%. Although not shown here, similar results were obtained for the MnO concentrations. The fact that we observe two different regimes suggest that there are two different mechanism affecting the CaO concentration. To better understand this behaviour we need to consider mineralogical information.

A conclusive explanation for positive correlation between CaO abundance and reservoir quality in the low-concentration regime cannot be given: because of its low concentration, the Ca-bearing minerals which may have caused this trend may not even be identifiable from XRD or thin-section analysis. Given the strong grain-size control on reservoir quality, the simplest explanation for this trend is that it is (again) caused by variations in grain size. Probably, CaO is contained in small amounts in a detrital component which is either relatively large or has a relatively high density, such as schist fragments or anhydrite grains (Henares et al., 2014).

Conventional XRF analysis showed that CaO concentration ranges up to 7% and in the XRF-CS predictions the CaO concentration even exceeds 30%. Because CaO concentrations are fairly high, the associated mineral(s) must have been detected by XRD and thin-section analysis. Both data sets revealed that there are two main sources of CaO in this core which

are anhydrite (CaSO_4) and dolomite ($\text{CaMg}(\text{CO}_3)_2$). The concentration of dolomite is in the same range as the CaO concentrations: whereas the anhydrite concentration does not exceed 1%, the dolomite concentration ranges up to 18%. Given that the transition from positive to negative correlation was around 0.1%, the results suggest that both anhydrite and dolomite may be the potential cause for reduced reservoir quality. Thin section analysis confirms that both dolomite and anhydrite act as a reservoir-quality reducing mineral, because both minerals were found to fill secondary pores. Hence we interpret the negative correlation between CaO and MnO in the high-concentration regime as being caused by dolomite and/or anhydrite cementation, which reduced reservoir quality.

Synthesis

By combining the mineralogical interpretation with the core-scanning data we will try to reconstruct the importance of the various reservoir-quality controlling mechanisms in detail.

Grain size and anhydrite/dolomite cementation have been identified as the dominant reservoir-quality controlling mechanisms. Grain size controls the detrital composition and is therefore an important factor throughout the core. However, as thin section analysis revealed, the abundance of anhydrite and dolomite cement varies strongly throughout the cored interval. To determine their spatial distribution we make use of their physical and chemical properties. Compared to quartz, anhydrite and dolomite have a relatively high density (average densities are 2.97 and 2.85 g/cm^3 , respectively). Hence, their abundance will be reflected by the high-resolution grain density record. Dolomite and anhydrite may be distinguished based on the concentration of S and Mg. Unfortunately, Mg nor S could be predicted with enough certainty from the XRF-CS data. To permit this distinction to be made, we limit ourselves to the Mg and S concentrations, as determined using destructive XRF analysis.

The core photograph, RQI classification, predicted grain density, predicted CaO concentrations, measured MgO concentrations and measured S concentrations are shown in Figure 8.6 (A-F). Firstly, we observe that in the top of BCU3 and BCU2, grain densities are positively correlated with the concentration of CaO. Secondly, we observe that elevated CaO concentrations are common in BCU3, and relatively rare in BCU1 and BCU2. At the top of BCU1 and BCU2, high CaO concentrations are associated with elevated S. At the base of BCU1, the elevated CaO concentrations are accompanied by high S nor Mg concentrations. Only in BCU3, elevated CaO concentrations are accompanied with high S concentrations.

Elevated S concentrations suggest the presence of anhydrite. Anhydrite was indeed recognised based on visual inspection of the uppermost core section: according to the core report, spotty, brown-yellow concretions are visible with the naked eye. Also the declining anhydrite abundance with depth suggested by the S concentration (Fig. 8.6 D) is confirmed by the fact that the size and frequency of these concretions reduce with depth (they are not visible anymore below section 2). At the base of BCU1 we observe neither elevated S nor Mg, indicating that there is possibly a third source of Ca (possibly calcite). Based on these results we propose the following interpretation for the controls on the reservoir quality and their relative importance (Fig. 8.6 F):

1. Grain size (BCU3, BCU2, BCU1),

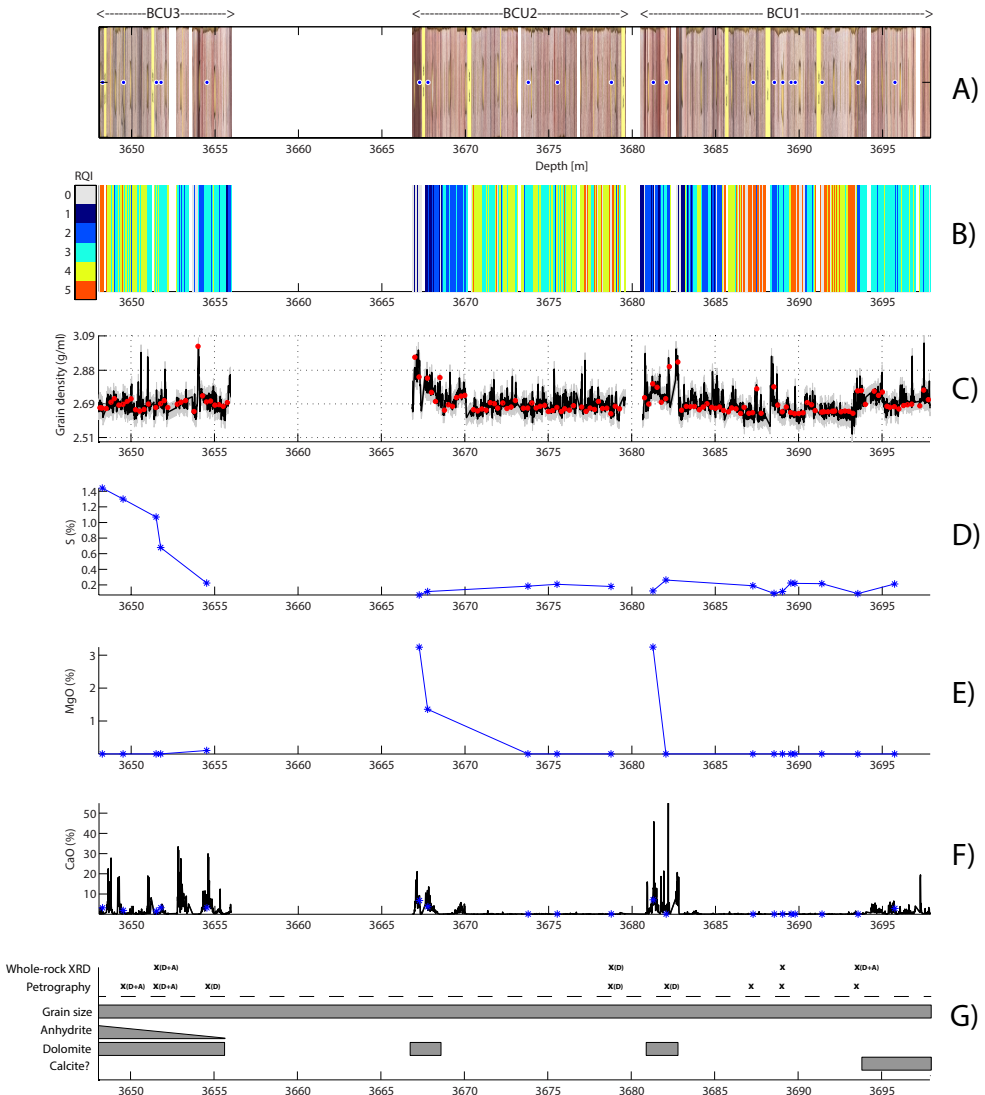


Figure 8.6: Reconciliation of the controls on reservoir quality in the BCUs (application 3). Displayed are the line scan image (A), the plug and predicted grain density (B), the concentration of S (C) and Mg (D) derived using pressed-pellet XRF, the plug and predicted CaO concentration (E), RQI classification scheme (F) and our interpretation of the mechanisms controlling reservoir quality in this core (G). The interpretation is based on all of the above sources of information in addition to the XRD analysis and petrographical analysis: an "X" indicates a measurement whereas "D" and "A" imply the presence of dolomite and anhydrite.

2. Dolomite cementation (BCU3, BCU2, BCU1),
3. Anhydrite cementation (BCU3),
4. Calcite cementation (BCU1).

These findings are consistent with the core report, which states that diagenetically-controlled reservoir quality varies throughout this core.

8.5 General discussion and conclusions

We demonstrated that upscaling the lithofacies properties with the high-resolution poro-perm predictions can result in different average values. This, in turn, can have a substantial impact on reservoir models. It was also shown that by complementing petrophysical analysis with core scanning data, a core may be described in terms of its reservoir quality on a centimeter resolution (potentially even on a sub-centimeter scale) in a fully automated and objective manner. Finally, we demonstrated that core scanning data enable estimation of the thickness of diagenetically altered intervals, and unfolding conclusions drawn from thin sections to the entire core. These conclusions could not be drawn from data acquired in routine core analysis because their resolution and coverage are insufficient.

Comparison between the RQI (Fig. 8.6 B) and our interpretation of diagenetic and grain-size controls (Fig. 8.6 B) confirms that reduced reservoir quality is accompanied by elevated CaO concentrations. At the top of BCU1 and BCU2, these elevated CaO concentrations are assigned to dolomite cement. The thickness of these cemented intervals is roughly between two and five meter. The conclusion in the core report that the occurrence of dolomite "is too local and little to have a large effect on reservoir properties" is therefore untrue. Furthermore, we identified a possible third reservoir-quality-reducing mechanism, probably associated with calcite cement, which reduced the quality in an even thicker interval at the base of BCU1.

The RQI classification in combination with the average geochemical composition facilitates rigorous selection of intervals or plugs which require more in-depth analysis with XRD, SEM, etc. In this core, for instance, the geochemical and RQI records immediately show that sampling the lowermost section of BCU1 may be deemed necessary for a better description of this core. When considering the link between core and reservoir characterisation, the high resolution geochemical records facilitate more detailed characterisation of diagenetically-altered intervals. Although the lateral extent of such diagenetic processes having a geochemical response may be limited, the high resolution of the scanning data ensures that a horizon which has a distinct geochemical response will be picked up. Such high resolution records potentially lead to better well-to-well correlations. Considering the extra investment that has to be made negligible compared to the costs of core recovery and routine core analysis, we consider XRF-CS a necessary addition to the core-analysis workflow.

General discussion and outlook

9.1 Discussion

9.1.1 Objectives

One goal of this study was to build a mathematical-statistical framework for the construction of so-called *integrated core descriptions* using XRF core scanning. The other goal of this study was the development of methods that facilitate integrated core analysis. To reach those goals, six objectives were defined which we will evaluate below:

1. **Characterise statistically** the relevant data types. We found that counting errors in the clr space are not *iid*, as requested by conventional least-squares methods (Chapt. 4). We analysed the behaviour of the inter-laboratory error of geochemical data in Chapter 5, and concluded that its behaviour could not be explained by counting. We also found that the structure of commonly-used uncertainty functions is inconsistent with theoretical considerations. A new formulation and modelling approach was proposed which accommodates for this flaw. The implications of the statistical evaluations presented in Chapters 4 and 5 for calibration of XRF-CS data were that: (i) the errors of element intensities can often be adequately "whitened" using appropriate scaling, (ii) it is reasonable to calibrate core-scanning data using a method based on a least-squares criterion, (iii) uncertain element intensities "pollute" the other element intensities when expressed in clr-space, (iv) the total analyte concentration in quantitative spectroscopy is established in a fundamentally different manner than the analyte concentration itself.
2. **Formalise** interpretation and geological inference from geochemical data using mathematical and statistical methods. In Chapter 3 we demonstrated that a model inspired by PLS may be used to formalise geological interpretation of geochemical data. PLS enables symmetric decomposition of data sets into unique signals and signals shared with other data sets. These two types of signals have geological significance when applied to geochemical composition and grain size. Applying this methodology to

three soft-sediment cores yielded that employing "textbook" proxies like Al/Ti for grain size can be risky: for the three analysed cores, Ti concentrations once showed no correlation, once positive correlation and once negative correlation with grain size.

3. **Improve** the existing logratio-based calibration method so that the misfit between composition predicted by the scanner and composition estimated using conventional analytical techniques may be reduced. We proposed a multivariate calibration method in Chapter 2 which was further refined in Chapter 7 by introducing X- and Y-variable selection routines and a robust model-selection criterion. We found that this new method has considerably higher predictive power than the univariate and bivariate calibration methods.
4. **Extend** the existing logratio-based calibration method by enabling prediction of properties which were not directly observed. The multivariate calibration approach proposed in Chapter 2 was an **extension** of conventional calibration methods, because it enables estimation of "true" analyte concentrations. This was achieved by padding the "undef" so as to ensure that the geochemical data become "complete" compositions. The other important extension proposed in Chapter 7 is selection of the appropriate Y-data transformation based on the sample space. This extension enables estimation of chemical as well as non-chemical properties, such as porosity and permeability.
5. **Evaluate** the performance of newly-proposed calibration methods for different types of cores (marine soft-sediment cores, terrestrial soft-sediment cores and consolidated-sediment cores). We **evaluated** the quantitative performance of XRF-CS for prediction of lithofacies, chemical composition and petrophysical properties for consolidated and unconsolidated sediments. On the basis of the analysed data, we concluded that the proposed methodology is a full-fledged alternative *only* for conventional geochemical analysis of unconsolidated-sediment cores and *only* for siliciclastics-dominated sediments.
6. **Explore** the added value of core-scanning technology within the conventional routine core analysis workflow using real-world cases. We have thoroughly explored the added value and new applications of XRF core scanning analysis within the core-analysis workflow. Firstly, we showed that the use of core scanning-based predictions of the petrophysical properties can result in different average facies properties which, in turn, can have important consequences for reservoir models. Secondly, we showed that by complementing petrophysical analysis with core scanning data, a core may be automatically classified in terms of its reservoir quality on a centimeter scale. Thirdly, we showed XRF-CS enables extrapolation of thin-section analysis and associated interpretations to the entire core on a high resolution.

9.1.2 Performance

We may turn the performance evaluation conducted in this thesis into a set of guidelines. In Chapter 7 we obtained that, depending on the type of core and its lithological range (i.e. including or excluding organic-rich sediments), the PPR value ranges between 1.6 and 3.2.

Re-evaluation of the data sets analysed in Chapter 2 (i.e., a marine soft-sediment core) and application of the state-of-the-art calibration method proposed in Chapter 7 yields $PPR_{chem} = 0.82$ (core GeoB7920) and $PPR_{chem} = 1.00$ (core AU10v). Drawing firm conclusions about the quantitative performance would require the analysis of more cores, however, the results suggest that the quantitative performance is related to the type of sediment and/or the associated compositional variability. Assuming that the results are representative for their respective class of cores, we postulate the following guidelines:

1. Homogeneous soft-sediment cores (i.e., only siliciclastics): $PPR \leq 2$
2. Heterogeneous soft-sediment cores (i.e., siliciclastics as well as organic material): $1 \leq PPR \leq 3$
3. Heterogeneous cores of sedimentary rock (organic and diagenetically-altered sediments): $3 \leq PPR \leq 4$

If we adopt a PPR-fence value of 2, we conclude that the core scanner can serve as a fully-fledged alternative only to homogeneous soft-sediment cores. Nevertheless, more calibration data and stratified calibration models (e.g., by lithofacies) are likely to give lower PPR values.

9.1.3 Application of XRF-CS

Choosing the optimal data acquisition strategy is not only a matter of choosing the most accurate analytical method. Defining a data-acquisition strategy is a matter of finding an optimum between time and cost constraints on the one hand, and required data quality and information content on the other hand. Hence, in applications where time is more important than quality, core scanning data with a lower predictive performance may still be a favourable option. In that case, the question is whether the quality of the core scanning data is *sufficient* for finding an answer to the respective scientific question. In practice, defining the requirements which the data should fulfill is considered very difficult: frequently, the purpose of the data is not specified on forehand, hence restricting its versatility is unwanted. Furthermore, it requires an extensive knowledge about mathematical statistics which is not always present/available. Given the modest additional costs of a full core scan, however, we recommend considering to complement conventional techniques such as laser-particle analysis and pressed-pellet XRF with a core scan.

In addition to efficiency and accuracy, also the non-destructiveness of core scanning can be an important aspect when setting up a data-acquisition strategy. Obviously, it is possible to retrieve samples every centimeter and analyse them with a conventional analytical technique. In certain applications, however, the fact that intensive sampling disturbs the integrity of the core is unwanted or not even allowed, for instance when the core material is state-owned or property of a third party. In those cases, core scanning may be the only way to retrieve high resolution data of sediment properties.

Core scanning can play a key role in the analysis of hydrocarbon reservoir rocks. For these cores, there are clearly-defined objectives and data requirements. Moreover, hydrocarbon exploration focuses increasingly more on smaller and more heterogeneous (thin-bedded) reservoirs. Considering the modest additional investment necessary to run a full core scan, high resolution core scanning can contribute to increasing the understanding of the control on reservoir quality, and reducing statistical uncertainty and financial risk.

9.2 Future developments

9.2.1 Line-scan imaging

When it comes to future developments of core-scanning technology, we think that it is key to combine XRF-CS with other analytical methods which fulfill the following requirements: data acquisition is time- and cost-efficient, data can be acquired in-situ and non-destructively, the data have a high spatial resolution. Line-scan imaging fulfills these requirements.

In this thesis, line-scan imaging has been used only in a qualitative sense. Given that a geologist makes a core description on the basis of the core's appearance, exploiting image data in a quantitative sense is likely to increase predictive performance. Apart from the average color (see Chapter 6), relevant and unique information which can potentially be retrieved from core images is the grain-size distribution. However, the pixel size of 70 microns of the cameras mounted on the Avaatech scanners is insufficient for accurate prediction of grain size in the silt to sand range. Hence, our colleagues at the NIOZ have attempted to increase the resolution of the line-scan camera system. It turned out, however, that increased resolution (20 micron) translated to a increase in the amount of data, thereby introducing new problems concerning data transfer and storage.

Apart from technical challenges, grain-size extraction from line-scan images is certainly not a straightforward operation. Firstly, extraction of grain size is hampered by the presence of sedimentary structures. Secondly, grain size is not a feature of a particular pixel, but merely a feature of a sedimentary bed. Retrieving grain-size information from core images therefore requires estimation of the bedding direction at every vertical position. It was found that an effective, though complicated method to estimate this direction is to find maximum cross-correlation between two short "traces" along the axis of the core (MSc thesis work of Idtz Wieling (2013)). In an attempt to develop a simpler method, an alternative approach based on 2D autocorrelation functions (ACFs) of small segments of the core image was explored. Apart from being less complicated, an additional argument behind the application of ACFs was that they have been used quite effectively for determination of the grain size from images of isotropic sediments (Rubin, 2004). Employing ACFs for both estimation of bedding direction as well as grain size would therefore yield a universal approach.

The first step in this process was to obtain one unique data set from the three individual color channels (i.e., RGB). In order to stay within the context of compositional data analysis, our approach was to derive the first principal component from the clr-transformed RGBD values. In case the image is anisotropic, the ACF of the first principal component has the appearance of a 2D multivariate Gaussian probability density function having a long axis and a short axis. The long axis is the direction of maximum autocorrelation which most likely corresponds to the bedding direction. Determination of this direction may be done by Singular Value Decomposition of the covariance matrix of the x- and y-direction gradient of this ACF. Moreover, the ratio between largest and smallest eigenvalue gives an indication of the anisotropy of the image. This approach proved to be quite powerful for the determination of clearly linear features such as fractures (see Fig. 9.1), particularly. The ability to retrieve more subtle trends turned out to be more tricky. More research is necessary to turn this basic idea into a robust method with practical value.

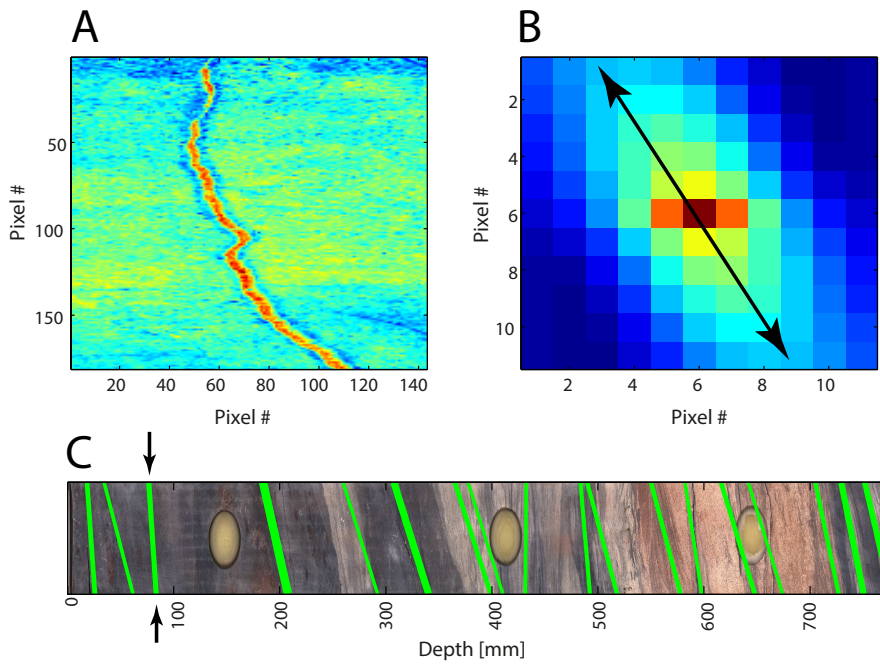


Figure 9.1: (A) small slice of the RGB image, represented by means of the first principal component of the clr-transformed RGBD values, (B) 2D ACF, (C) core image with the green lines showing the determined bedding direction. The thickness of the green lines represents the ratio between the first and second eigenvalue of the 2D ACF.

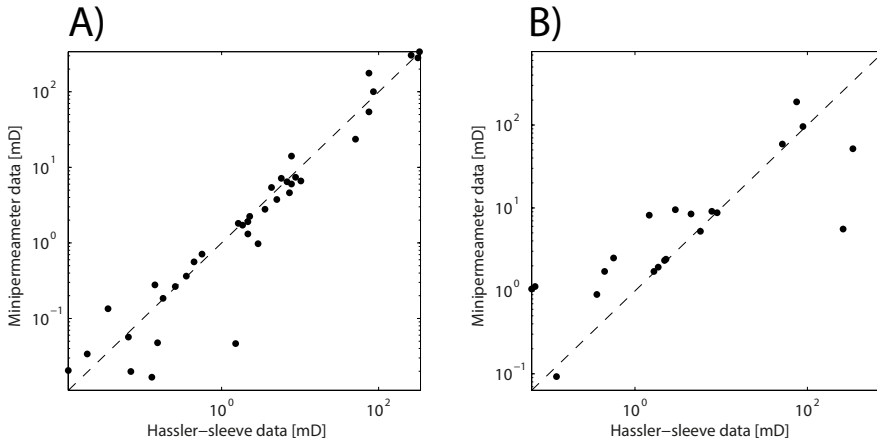


Figure 9.2: Comparison between the plug permeabilities acquired using Hassler-sleeve experiments and the minipermeametry data acquired directly from the plugs (A) and from the resinated slabs (B).

9.2.2 Minipermeametry

Another technique which is time-effective, cost-effective and non-destructive is minipermeametry. Minipermeametry is based on placing a small, air-tight nozzle on the rock and putting on an over-pressure. By carefully processing the pressure-decay curve, it is possible to estimate the horizontal gas permeability (Jones, 1994). Minipermeametry can be of great value because after slabbing the core, it is no longer possible to retrieve permeability data with the conventional techniques (Hassler-sleeve experiment). Minipermeametry and XRF core scanning data are therefore a potentially powerful combination. The fact that minipermeametry can be conducted nearly continuous means that the XRF-CS data can be better calibrated.

To investigate the accuracy of minipermeametry data on slabs we conducted an experiment. Firstly, we compared minipermeameter data derived from plugs with Hassler-sleeve data. Secondly, minipermeameter data acquired from the resinated slab at the plug positions were compared to permeability measurements of the plugs. For this experiment we used the slabs and plugs of core E10-3. The minipermeameter data were acquired with a PDPK-400 of Corelab. The results are shown in Figure 9.2.

Under the assumption that the Hassler-sleeve data are correct, we found $MRE=82\%$ for the minipermeametry measurements of the plugs, and $MRE=138\%$ for the minipermeametry measurements of the resinated slabs. The MRE associated with the plug minipermeametry data is considerably smaller than that associated with the core-scanning-based predictions ($MRE=117\%$). On the other hand, the MRE associated with the minipermeametry measurements of the resinated slabs is larger. We also observe that the minipermeametry measurements on the resinated slab overestimate the Hassler-sleeve permeabilities on the low end. Possibly, there was air leaking between the annulus and the rock. Another possibility is that air could escape laterally: at the plug positions there was typically only a 1-cm wide piece of rock left.

The fact that this systematic overestimation was not apparent in the minipermeametry

data of the plugs may be caused by the fact that in contrast to the slab, the plugs are relatively smooth and flat. Also at the high end there are data points for which the Hassler-sleeve data and minipermeametry data of the slabs are inconsistent. Here, however, the most extreme deviations are characterised by the minipermeameter yielding a considerably lower permeability than the Hassler-sleeve experiment. This underestimation could be explained by the resin having penetrated the rock too close to the surface of the slab.

The results demonstrate that more research is necessary to further optimise the data-acquisition workflow. Moreover, it needs to be investigated whether a minipermeameter can be integrated within the XRF-CS instrument for efficiency. The importance of permeability in hydrocarbon exploration means that we think that it is worth investigating whether XRF-CS and minipermeametry may be integrated, both in one machine as well as in the core-analysis workflow.

9.2.3 Alternative technological improvements

Apart from further exploitation of line-scan images and minipermeametry, there are other sensors which are already mounted on core-scanning instruments which may be used more effectively. For example, Itrax core scanners (by Cox) are already equipped with lasers to determine the exact vertical position of the core relative to the x-ray source and detector. In the future, it may be attempted to use laser-backscattering information to retrieve information about surface roughness. Roughness, in turn, will be correlated with grain and/or pore size.

Another sensor which may be mounted on an XRF-CS instrument is a digital microscope. Microscopic imagery at the scan positions is a way to circumvent the problems surrounding data storage and data handling and yet have sufficient resolution. Ideally, these images enable direct estimation of grain size and bedding direction at the position of the scan data. Another measurement technique which potentially provides relevant information not contained in geochemical records is hyperspectral imaging. Hyperspectral imaging includes all imaging techniques which gather reflectance data across a wide spectral range. Relative to the visible-light spectrum, both short and long wavelengths can provide relevant information about mineralogical composition and the content and nature of organic compounds (Trachsel et al., 2010). Combining XRF core scanning with hyperspectral core scanning therefore potentially enables integration of geochemical and mineralogical composition. This is of major importance because, unlike chemical components, minerals have physical properties such as density and size: these properties are indispensable to make the connection between composition and models of sediment transport (e.g., Komar, 2007).

9.3 Outlook

In this thesis, the predictive performance of the core scanner was tested on the basis of individual cores. However, the core as a genetic unit is an arbitrary choice when a project involves numerous cores from the same rock or sediment body, which is the case for a mature field or province (e.g., the gas-bearing Rotliegend reservoir rocks in The Netherlands). Ideally, the core scanner facilitates prediction of sediment properties in a fully non-destructive manner at a certain stage, i.e., after enough measurements have been acquired. This stage may be achieved by following a completely formalised and computer-driven ap-

proach which, in the context of core analysis, would constitute three phases: a training phase, optimization phase and a development phase.

In the *training phase*, calibration/prediction models associated with the core-scanning data are not yet established which means that any new core should be sampled with the same sampling intensity. In contrast to conventional equal-interval sampling, sampling should rely on optimizing the stratigraphic coverage as well as the chemical coverage. This may be achieved in an automated manner by applying a clustering-based technique to the joint stratigraphic and chemical spaces. In principle, this phase is therefore equivalent to what has been outlined in Chapter 2.

As more calibration data become available, calibration models tend to stabilise and adding more calibration data will not change the model parameters. This marks the *optimization phase*. In this phase the goal should be to maximise predictive performance of the calibration models, for instance by introducing stratified modelling. For instance, it may be advantageous to set up a separate calibration/prediction model for grain density within the sand and the clay, or for Rotliegend and Carboniferous cores. This may be done manually, but there are also unsupervised techniques to establish optimal model stratification such as Multivariate Adaptive Regression Splines (MARS). (Friedman, 1991). The rationale behind manual and automatic stratification is demonstrated in Figure 9.3.

When optimal calibration models have been established, the core scanner may be used in predictive mode. This marks the start of the *development phase*. Key element of this phase is to model the discrepancy between available calibration data and predictions, i.e., a second-order model (as presented in Chapter 5 of this thesis). By doing so, we also quantitatively model the limits of the core scanner. The ability to predict the prediction error enables optimal sample selection in a strictly automated manner: we sample only if the expected prediction error is outside the acceptable limits. To what kind of facies this applies cannot be said beforehand, however, it could be that it is simply impossible to accurately predict the permeability of the most coarse-grained facies because for these facies, their chemical compositions are all close to 100% SiO₂.

In this last phase, core scanning and conventional methods may be treated "symmetrically". With a symmetric approach we mean that the core-scanning predictions and calibration data are treated as two independently-acquired estimates of sediment properties, although possibly with a different uncertainty. The advantage of a symmetric approach is that it enables using the core-scanning data for the purpose of data-quality control: the core-scanning data in combination with a calibration model allow us to estimate chemical and petrophysical properties prior to conventional analysis. In case these estimates turn out to be at odds with the conventional analysis, we have reason to believe that (i) either our core scanning calibration model is inadequate or incomplete, (ii) the integrity of the direct measurements is questionable. Either way, this "flagged" data point deserves further investigation. In this study, we experienced such discrepancies for the pressed-pellet XRF analysis of core E10-3 which were sporadically at odds with both the core-scanning predictions, the XRD data and the petrographical analysis. In principle, this approach is similar to that applied in Chapter 3 of this thesis.

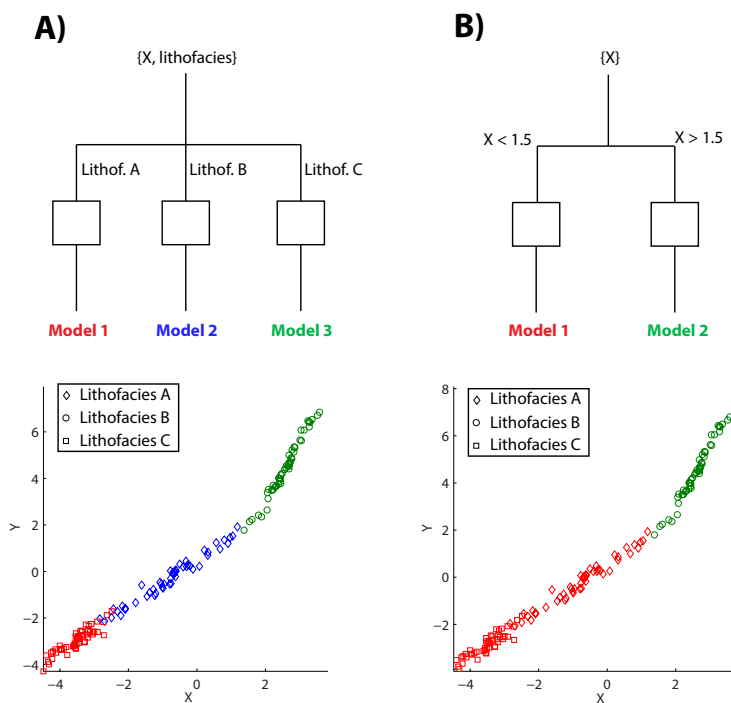


Figure 9.3: Conceptual representation of the rationale behind stratified modelling, aimed at relating X (e.g., XRF-CS signature) to a variable of interest Y (e.g., total-organic carbon). (A) Manual stratification by employing different models for different lithofacies, and (B) automatic stratification: classes showing similar behaviour are lumped based on X and modelled accordingly.

Appendices

Appendix of Chapter 3

Our goal is to decompose the geochemical record into a part which is correlated with grain size, and a part which varies independently from the grain size. More specifically, our aim is to obtain a basis Q in \mathbb{R}^D (i.e., the clr-transformed geochemical solution space) that maximizes the geochemical variance explained by the grain size, and vice versa. A multivariate method to maximize the common covariance is Partial Least Squares, or PLS (Wold et al., 1983). The model employed in this study is closely related to PLS.

Bulk chemistry and grain-size distribution are both compositional data. From a modelling perspective, the most important implication of their compositional nature is that standard statistical methods cannot be applied. Instead, processing of compositional data should be done using log-ratio transformations, or logratios. For multivariate statistical analysis, the data should be centred log-ratio (clr) transformed.

Given that we have measured the geochemical composition (D variables) and the grain-size distributions (L grain-size classes) of a set of m specimens, we first derive the matrices \mathbf{X} and \mathbf{Y} containing the clr-transformed grain-size distributions and chemical compositions of the same set of samples, respectively. Subtracting the mean yields \mathbf{X}_0 and \mathbf{Y}_0 . Next, we calculate the cross-covariance matrix \mathbf{C} :

$$\mathbf{C} = \mathbf{X}_0^T \mathbf{Y}_0 \quad (\text{A.1})$$

from which we derive the Singular Value decomposition:

$$\mathbf{C} = \mathbf{U} \mathbf{S} \mathbf{V}' \quad (\text{A.2})$$

The first column of \mathbf{U} (\mathbf{u}_1) is a vector in \mathbb{R}^D whereas the first column of \mathbf{V} spans a vector in \mathbb{R}^L . They jointly maximize the cross-covariance among \mathbf{X} and \mathbf{Y} . We store \mathbf{u}_1 and \mathbf{v}_1 in the first column of the matrices \mathbf{P} and \mathbf{Q} , respectively. Next, we perform direct deflation of the cross-covariance matrix \mathbf{C} (i.e. the SIMPLS method):

$$\mathbf{C} = \mathbf{C} - \mathbf{u}_1 (\mathbf{u}_1' \mathbf{C}) \quad (\text{A.3})$$

and return to equation A.2 to calculate the scores and loadings of the second PLS component. We proceed until the number of calculated components equals the minimum rank of X and Y . The final product of this algorithm are the matrices \mathbf{P} and \mathbf{Q} .

Using \mathbf{P} and \mathbf{Q} we can apply the model to the grain-size and geochemical data set as a whole, i.e., also to the observations related to samples contained in either the grain-size or the geochemical data set. Because $(\mathbf{P}^T)^{-1} =$ and \mathbf{Q}^T are both orthonormal, we may calculate the scores matrices T and U in the following manner:

$$\begin{aligned}\mathbf{T} &= \mathbf{X}_0 \mathbf{P} \\ \mathbf{U} &= \mathbf{Y}_0 \mathbf{Q}\end{aligned}\tag{A.4}$$

Subsequently, the k -component PLS approximations are given by:

$$\begin{aligned}(\hat{\mathbf{X}}_0)_k &= \mathbf{T}_{1:k} \mathbf{P}'_{1:k} \\ (\hat{\mathbf{Y}}_0)_k &= \mathbf{U}_{1:k} \mathbf{Q}'_{1:k}\end{aligned}\tag{A.5}$$

whereby $\mathbf{T}_{1:k}$ stands for the first k columns of \mathbf{T} . The 'residuals' are obtained by subtracting the common variability in both data sets from the input data:

$$\begin{aligned}(\check{\mathbf{X}}_0)_k &= \mathbf{X}_0 - (\hat{\mathbf{X}}_0)_k \\ (\check{\mathbf{Y}}_0)_k &= \mathbf{Y}_0 - (\hat{\mathbf{Y}}_0)_k\end{aligned}\tag{A.6}$$

Alternatively, we may add the mean again to form $(\check{\check{\mathbf{X}}}_0)_k$ and $(\check{\check{\mathbf{Y}}}_0)_k$, so that the residual variability centers around the mean of their parent data.

To determine the maximum number of components that may be removed, we perform a significance test on the correlation between the scores \mathbf{t}_k and \mathbf{u}_k for every k . Significance of correlation on the k -th basis vector may be tested using the following criterion (Kendall and Stuart, 1973):

$$|r| \sqrt{\frac{m-2}{1-r^2}} > t^{-1}(m-2, p)\tag{A.7}$$

where r is Pearson's correlation coefficient between \mathbf{t}_k and \mathbf{u}_k (i.e. the projection of the data on the k -th basis vector). The right-hand side is the inverse t -distribution with probability p and degrees of freedom $m-2$. For any r and number of observations m that were used to fit the model, we can calculate the probability of getting the observed r in the absence of correlation. We remove the common component if the criterion in Equation A.7 with $p = 1 - \alpha/2$ is met.

Appendix of Chapter 4

For MDD we derived that the ECM in the log-space reads:

$$\begin{aligned}\Phi_{mdd}(j, j) &= (\tau\pi_j)^{-1}(1 - \pi_j) \\ \Phi_{mdd}(i, j) &= r(i, j)q_{mdd}(i)q_{mdd}(j) = -\sqrt{\frac{\pi_i\pi_j}{(1-\pi_i)(1-\pi_j)}}\sqrt{\frac{\pi_i}{(1-\pi_i)\tau}}\sqrt{\frac{\pi_j}{(1-\pi_j)\tau}} = -\tau^{-1}\end{aligned}\tag{B.1}$$

After some algebra we obtain the following result for the product of the clr-transformation matrix \mathbf{A}_D and Φ_{mdd} :

$$\begin{aligned}\mathbf{A}_D\Phi_{mdd}(j, j) &= a_d(\tau\pi_j)^{-1} \\ \mathbf{A}_D\Phi_{mdd}(i, j) &= a_{od}(\tau\pi_j)^{-1}\end{aligned}\tag{B.2}$$

where a_d and a_{od} are the diagonal and off-diagonal entries of \mathbf{A}_D , respectively.

Using the result of this first matrix product we may write for the full matrix product $\mathbf{A}_D\Phi\mathbf{A}_D$:

$$\begin{aligned}\Psi_{mdd}(j, j) &= \mathbf{A}_D\Phi_{mdd}\mathbf{A}_D(j, j) = a_d^2(\tau\pi_j)^{-1} + a_{od}^2\sum_{m\neq j}^D(\tau\pi_m)^{-1} \\ \Psi_{mdd}(i, j) &= \mathbf{A}_D\Phi_{mdd}\mathbf{A}_D(i, j) = a_{od}a_d(\tau(\pi_i + \pi_j))^{-1} + a_{od}^2\sum_{m\neq\{i, j\}}^D(\tau\pi_m)^{-1}\end{aligned}\tag{B.3}$$

For PDD the following holds for the ECM in the log-space:

$$\begin{aligned}\Phi_{pdd}(j, j) &= (\Delta t\xi_i)^{-1} \\ \Phi_{pdd}(i, j) &= 0\end{aligned}\tag{B.4}$$

Because the log-space ECM is diagonal, it is not difficult to derive that the following

holds for the clr-space ECM:

$$\begin{aligned}\Psi_{pdd}(j, j) &= \mathbf{A}_D \Phi_{pdd} \mathbf{A}_D(j, j) = a_d^2 (\Delta t \xi_j)^{-1} + a_{od}^2 \sum_{m \neq j}^D (\Delta t \xi_m)^{-1} \\ \Psi_{pdd}(i, j) &= \mathbf{A}_D \Phi_{pdd} \mathbf{A}_D(i, j) = a_{od} a_d (\Delta t (\xi_i + \xi_j))^{-1} + a_{od}^2 \sum_{m \neq \{i, j\}}^D (\Delta t \xi_m)^{-1}\end{aligned}\tag{B.5}$$

The proof is complete if we consider the following properties of MDD and PDD:

$$[X_1, \dots, X_D] \sim Mult(\tau, \boldsymbol{\pi}) \rightarrow E(X_j) = \tau \pi_i \tag{B.6a}$$

$$X_j \sim Poiss(\Delta t \xi_j) \rightarrow E(X_j) = \Delta t \xi_i \tag{B.6b}$$

Substitution yields that:

$$\begin{aligned}\Psi(j, j) &= a_d^2 E(X_j)^{-1} + a_{od}^2 \sum_{m \neq j}^D E(X_m)^{-1} \\ \Psi(i, j) &= a_{od} a_d (E(X_i) + E(X_j))^{-1} + a_{od}^2 \sum_{m \neq \{i, j\}}^D E(X_m)^{-1}\end{aligned}\tag{B.7}$$

QED.

Appendix of Chapter 5

C.1 Modelling approach

C.1.1 UFs and closure

Following Aitchison (1982), closure is given by:

$$\mathcal{C}[\mathbf{c}] = \left[\frac{c_1}{\sum_k c_k}, \frac{c_2}{\sum_k c_k}, \dots, \frac{c_D}{\sum_k c_k} \right] \quad (\text{C.1.1})$$

Because closure defines concentrations, we propose to derive the proper UF by propagating the associated uncertainty through closure.

Given that the UF of pseudo concentrations is given by $f(\tilde{c}_k)$, the uncertainty of the closed pseudo concentrations (i.e. the proper concentrations c) is given by:

$$\delta^2 c_j = \left(\frac{\tilde{c}_j}{\sum_k \tilde{c}_k} \right)^2 \left[\left(\frac{f(\tilde{c}_j)}{\tilde{c}_j} \right)^2 + \left(\frac{\sqrt{\sum_k f^2(\tilde{c}_k)}}{\sum_k \tilde{c}_k} \right)^2 - 2 \frac{f^2(\tilde{c}_j)}{\tilde{c}_j \sum_k \tilde{c}_k} \right] \quad (\text{C.1.2})$$

Under the assumption that pseudo concentrations are unbiased estimates of the true concentration, we can assume that $\sum_k \tilde{c}_k = 1$ and apply the substitution $c_j = \tilde{c}_j / \sum_k \tilde{c}_k$ so that Equation C.1.2 reduces to:

$$\delta^2 c_j = (1 - 2\tilde{c}_j) f^2(\tilde{c}_j) + \tilde{c}_j^2 \sum_k f^2(\tilde{c}_k) \quad (\text{C.1.3})$$

Equation C.1.3 allows us to calculate the uncertainty associated with the closed pseudo concentrations. Note that, consistent with the definition of concentrations, the uncertainty becomes correlated among the variables. Given that the first and second moment of closed pseudo concentrations are indistinguishable from proper concentrations, we treat them as such and refer to them as c .

C.1.2 Derivation of proper UF

If we would strictly follow the approach to derive the proper UF from the pseudo UF by error propagation upon closure, the proper UF would always be concentration-dependent. This can be shown by substituting $f^2(\tilde{c}) = a$ in Equation C.1.3, which leads to the following expression for $\delta^2\tilde{c}_j$:

$$\delta^2\tilde{c}_j = a - 2a\tilde{c}_j + aD\tilde{c}_j^2 \quad (\text{C.1.4})$$

To facilitate a constant proper UF, we approximate Equation C.1.3 by the following function:

$$\delta^2c_j = a_P^2 + \left[(1 - 2\tilde{c}_j)(b_P\tilde{c}_j^{d_P})^2 + \tilde{c}_j^2 \sum_{k=1}^D (b_P\tilde{c}_k^{d_P})^2 \right] \quad (\text{C.1.5})$$

i.e. we omit the concentration-dependent part in Equation C.1.4. Still, the expression shown in Equation C.1.5 is not compact and since it is a function of all analyte concentrations in the mixture, it is multivariate by nature. Our next goal is to approximate this expression with a BUF, i.e. an univariate function.

C.1.3 From pseudo to proper UF: Poissonian case

To study the ability to approximate the proper uncertainty (Eq. C.1.5) with a BUF we first consider the Poissonian case, i.e., data whose UF is given by:

$$\delta^2c_j = b_P^2\tilde{c} \quad (\text{C.1.6})$$

Substituting Equation C.1.6 into Equation C.1.5 yields:

$$\delta^2c_j = (1 - 2\tilde{c}_j)(b_P^2\tilde{c}_j) + \tilde{c}_j^2 \sum_k b_P^2\tilde{c}_k \quad (\text{C.1.7})$$

If we assume that $\sum_k \tilde{c}_k = 1$, this can be further reduced to:

$$\delta^2c_j = b_P^2c_j(1 - c_j) \quad (\text{C.1.8})$$

Hence, in case of proportionality between concentration and variance (cf. the overdispersed Poisson distribution) b_P and d_P from PUF can be readily plugged into BUF in order to obtain an exact proper UF (i.e. the UF associated with the closed pseudo concentrations). The reason for the portability of function parameters in case $d_P = 1/2$ is that only then the following holds:

$$\sum_k f^2(\tilde{c}_k) = f^2\left(\sum_k \tilde{c}_k\right) = f^2(1) = b_P \quad (\text{C.1.9})$$

C.1.4 From pseudo to proper UF: General case

After the Poissonian case, we will now analyse the more general case. As we let the concentration approach zero, the following holds:

$$\begin{aligned}\lim_{c \rightarrow 0} \delta^2 \tilde{c}_j &= a_P^2 \\ \lim_{c \rightarrow 0} \delta^2 c_j &= a_P^2\end{aligned}\quad (\text{C.1.10})$$

Hence, the intercept term is directly portable from PUF to BUF. For the other terms we adopt the "standard" composition, i.e.:

$$c_j = 1/D \quad (\text{C.1.11})$$

for any $j \in 1, 2, \dots, D$. Based on this standard composition it can be shown that in the following limiting cases the exponential term d_P is directly transferable between PUF and BUF:

$$\begin{aligned}\lim_{D \rightarrow 0} b^2 [D^{-2d} - 2D^{-2d-1} + D^{-2d-2}] &= b^2 D^{-2d} \\ \lim_{D \rightarrow 0} b^2 [D^{-1}(1 - D^{-1})]^{2d} &= b^2 D^{-2d}\end{aligned}\quad (\text{C.1.12})$$

By numerical simulation (not shown here) it was verified that the exponential terms before and after closure are indeed equal, irrespective of D , a_P and b_P .

In case $d_P \neq 1/2$, plugging b_P into BUF yields a biased prediction of the proper uncertainty. Given that $d_P > 1/2$ and \tilde{c} is smaller than unity, this follows from:

$$\sum_k f^2(\tilde{c}_k) > f^2\left(\sum_k \tilde{c}_k\right) \quad (\text{C.1.13})$$

Approximating the expression for the proper uncertainty shown in Equation C.1.5 by directly transferring the PUF parameters into BUF will introduce bias. This bias can be mitigated by again assuming the "standard" composition. Its corresponding multiplicative correction factor reads:

$$k = \sqrt{\frac{f^2(D^{-1})}{(1 - 2D^{-1})f^2(D^{-1}) + D^{-2} \sum_k f^2(D^{-1})}} \quad (\text{C.1.14})$$

where the numerator is the UF before closure and the denominator is the UF after closure. The multiplicative factor b_B which has to be plugged into BUF in order to obtain the proper UF is given by $b_B = kb_P$. This leads to the following general methodology for deriving a proper BUF from a pseudo PUF:

$$\begin{aligned}\text{Pseudo UF} &\rightarrow \text{Proper UF} \\ \text{PUF}(a_P, b_P, d_P; \tilde{c}) &\rightarrow \text{BUF}(a_P, kb_P, d_P; \tilde{c})\end{aligned}\quad (\text{C.1.15})$$

C.2 GeoPT unpublished references

Thompson, M., 2002. Protocol for the operation of GeoPT proficiency testing scheme, International Association of Geoanalysts.

Potts P.J., Thompson M., Kane J.S., Webb, P.C. and J.S. Watson (2000). GeoPT8 - an international proficiency test for analytical geochemistry laboratories - report on round 8 / February 2001 (OU-4 Penmaenmawr microdiorite). Unpublished IAG report.

Potts P.J., Thompson M., Webb, P.C. and J.S. Watson (2001). GeoPT9 - an international proficiency test for analytical geochemistry laboratories - report on round 9 / July 2001 (OU-6 Penrhyn slate). Unpublished IAG report.

Potts P.J., Thompson M., Webb, P.C., Watson, J.S. and Y. Wang (2001). GeoPT10 - an international proficiency test for analytical geochemistry laboratories - report on round 10 / December 2001 (CH-1 Marine sediment). Unpublished IAG report.

Potts P.J., Thompson M., Chenery S.R., Webb, P.C. and J.S. Watson (2002). GeoPT11 - an international proficiency test for analytical geochemistry laboratories - report on round 11 / July 2002 (OU-5 Leaton dolerite). Unpublished IAG report.

Potts P.J., Thompson M., Chenery S.R., Webb, P.C. and B. Batjargal (2003). GeoPT12 - an international proficiency test for analytical geochemistry laboratories - report on round 12 / January 2003 (GAS Serpentinite). Unpublished IAG report.

Potts P.J., Thompson M., Chenery S.R., Webb, P.C. and H.U. Kaspar (2003). GeoPT13 - an international proficiency test for analytical geochemistry laboratories - report on round 13 / July 2003 (Köln Loess). Unpublished IAG report.

Potts P.J., Thompson M., Chenery S.R., Webb, P.C. and B. Batjargal (2004). GeoPT14 - an international proficiency test for analytical geochemistry laboratories - report on round 14 / January 2004 (OShBO - alkaline granite). Unpublished IAG report.

Potts P.J., Thompson M., Chenery S.R., Webb, P.C. and Y. Wang (2004). GeoPT15 - an international proficiency test for analytical geochemistry laboratories - report on round 15 / June 2004 (Ocean floor sediment MSAN). Unpublished IAG report.

Potts P.J., Thompson M., Webb, P.C. and S. Wilson (2005). GeoPT16 - an international proficiency test for analytical geochemistry laboratories - report on round 16 / February 2005 (Nevada basalt, BNV-1). Unpublished IAG report.

Potts P.J., Thompson M., Webb, P.C. and J.N. Walsh (2005). GeoPT17 - an international proficiency test for analytical geochemistry laboratories - report on round 17 / July 2005 (Calcareous sandstone, OU-8). Unpublished IAG report.

Webb, P.C., Thompson M., Potts P.J. and L.P. Bedard (2006). GeoPT18 - an international proficiency test for analytical geochemistry laboratories - report on round 18 / Jan 2006 (Quartz Diorite, KPT-1). Unpublished IAG report.

Webb, P.C., Thompson M., Potts P.J. and B. Batjargal (2006). GeoPT19 - an international proficiency test for analytical geochemistry laboratories - report on round 19 / July 2006

(Gabbro, MGR-N). Unpublished IAG report.

Webb, P.C., Thompson M., Potts P.J. and M. Burnham (2007). GeoPT20 - an international proficiency test for analytical geochemistry laboratories - report on round 20 / Jan 2007 (Ultramafic rock, OPY-1). Unpublished IAG report.

Webb, P.C., Thompson M., Potts P.J. and B. Batjargal (2007). GeoPT21 - an international proficiency test for analytical geochemistry laboratories - report on round 21 / July 2007 (Granite, MGT-1). Unpublished IAG report.

Webb, P.C., Thompson, M., Potts, P.J. and B. Batjargal (2008). GeoPT22 - an international proficiency test for analytical geochemistry laboratories - report on round 22 / January 2008 (Basalt, MBL-1). Unpublished IAG report.

Webb, P.C., Thompson, M., Potts, P.J., Watson, J.S. and C. Kriete (2008). GeoPT23 - an international proficiency test for analytical geochemistry laboratories - report on round 23 / September 2008 (Separation Lake pegmatite, OU-9) and 23A (Manganese nodule, FeMn-1). Unpublished IAG report.

Webb, P.C., Thompson, M., Potts, P.J. and J.S. Watson (2009). GeoPT24 - an international proficiency test for analytical geochemistry laboratories - report on round 24 / January 2009 (Longmyndian greywacke, OU-10). Unpublished IAG report.

Webb, P.C., Thompson, M., Potts, P.J. and J. Enzweiler (2009). GeoPT25 - an international proficiency test for analytical geochemistry laboratories - report on round 25 / July 2009 (Basalt, HTP-1). Unpublished IAG report.

Webb, P.C., Thompson, M., Potts, P.J. and M. Loubser (2010). GeoPT26 - an international proficiency test for analytical geochemistry laboratories - report on round 26 / January 2010 (Ordinary Portland cement, OPC-1). Unpublished IAG report.

Webb, P.C., Thompson, M., Potts, P.J. and B. Batjargal (2010). GeoPT27 - an international proficiency test for analytical geochemistry laboratories - report on round 27 / July 2010 (Andesite, MGL-AND). Unpublished IAG report.

Webb, P.C., Thompson, M., Potts, P.J. and S. Wilson (2011). GeoPT28 - an international proficiency test for analytical geochemistry laboratories - report on round 28 / January 2011 (Shale, SBC-1). Unpublished IAG report.

Webb, P.C., Thompson, M., Potts, P.J. and S. Wilson (2011). GeoPT29 - an international proficiency test for analytical geochemistry laboratories - report on round 29 / July 2011 (Nephelinite, NKT-1). Unpublished IAG report.

Webb, P.C., Thompson, M., Potts, P.J., Long, D. and B. Batjargal (2012). GeoPT30 - an international proficiency test for analytical geochemistry laboratories - report on round 30 / January 2012 (Syenite, CG-2) and 30A (Limestone, ML-2). Unpublished IAG report.

Webb, P.C., Thompson, M., Potts, P.J and S. Wilson (2012). GeoPT31 - an international proficiency test for analytical geochemistry laboratories - report on round 31 / July 2012 (Modified river sediment, SdAR-1). Unpublished IAG report.

Webb, P.C., Thompson, M., Potts, P.J and E. Webber (2013). GeoPT32 - an international proficiency test for analytical geochemistry laboratories - report on round 32 / January 2013 (Woodstock Basalt, WG-1). Unpublished IAG report.

Appendix of Chapters 6 and 7

D.1 RCA of core E10-3

Petrophysical analysis

As part of RCA, 4-cm long core plug samples (n=246) with a 1 inch diameter were retrieved at 25 centimeter intervals and parallel to the bedding direction (i.e. horizontal plugs). The plugs were cleaned with an azeotropic mixture of methanol, chloroform and water to remove any hydrocarbons and salts. Subsequently, the plugs were dried in an oven of 60 °C until their weight remained constant.

The porosity and grain density (n=244) were determined by direct measurement of grain volume and bulk volume. The grain volume was measured by helium expansion in a Boyle's Law Porosimeter whereas the bulk volume was determined by submerging the sample in mercury. The grain volume together with the bulk volume allows calculation of the porosity. The grain density was calculated from the weight of the sample in combination with the grain volume. Information about uncertainty or detection limits were unavailable.

The horizontal permeability was determined using nitrogen gas (n=230). The plugs were mounted in a "Hassler" type core holder at a confining pressure of 400 psig and a steady state nitrogen gas flow was established through the sample. The flow rate, pressure differential, gas temperature and ambient pressure were recorded. These parameters were used in conjunction with the callipered length and diameter of the plugs to compute the permeability from Darcy's equation. The permeabilities were corrected for gas-slippage using an empirical correlation. The permeability data set is slightly smaller than the porosity and grain density data sets because it was not possible to determine the permeability of fractured plug. These plugs were not naturally fractured but fractured during plug drilling.

Core descriptions

The core was continuously described in terms of grain-size classes: organic (*O*), mudstone (*M*), siltstone (*f*), very fine sandstone (*S1*), fine sandstone (*S2*), medium sandstone (*S3*),

coarse sandstone (*S4*), and very coarse sandstone (*S5*). Hence, the sandstone was classified in terms of classes comprising 1/2 phi unit intervals.

Core E10-3 was also described in terms of lithofacies associations which are a description of a sedimentary bed in terms of grain size, sedimentary structures and relevant diagenetic characteristics. The following lithofacies associations were distinguished: poorly drained floodplain (*IFL*), floodplain (*F*), swamp (*SW*), crevasse splays (*CS*), inter-distributary bay (*IB*) and braided channel (*BC*) deposits. Both the grain-size description as well as the lithofacies description are continuous data sets in the sense that they reflect bed boundaries rather than point measurements.

D.1.1 Petrographical analysis

Detailed petrographical analysis including thin section petrography, whole rock as well as clay fraction X-Ray Diffraction (XRD) and Scanning Electron Microscopy (SEM) were conducted on the trim ends of 10 plugs. These samples were taken from the braided channel complex (n=8), interdistributary bay deposit (n=1) and the poorly-drained floodplain (n=1) deposits. Point counts yielded that all sandstones samples are sublitharenites with an average quartz content of 53%. Based on thin section analysis, several diagenetic phases were recognised with a different mineralogical response. This includes (i) early-diagenetic precipitative hematite and replacive kaolinite, (ii) quartz cement, and (iii) late-diagenetic dolomite. The nature and intensity of diagenesis turned out to be heterogeneous in this core. However, compared to other reservoirs in this area of similar age, Boels (2003) concluded that the general importance of diagenesis for the reservoir quality in this well is limited and that most variation in reservoir quality is controlled by grain size. However, not all diagenetic mechanisms could be distinguished from variations in mean grain size, e.g., hematite was found to be replacing pseudo matrix whose abundance was correlated with grain size.

D.2 Lithofacies classification scheme (E10-3)

The facies architecture of the cored Carboniferous interval in Well E10-was classified in terms of following lithofacies associations: poorly drained floodplain (*IFL*), floodplain (*F*), swamp (*SW*), crevasse splays (*CS*), inter-distributary bay (*IB*) and braided channel (*BC3*) deposits.

The dominant lithofacies association in core E10-3 is the braided channel sandstone lithofacies (*BC3*): approximately 55% of the core is classified as such. *BC3* is assigned to parts of example slabs 1, 3 and 4 (Fig. 6.4). Sedimentological features of the braided channel sandstone include cross-bedding (both low- and high-angle), horizontal and ripple lamination. The sandstone units are mainly composed of moderately sorted fine- to medium-grained sands that have a vague fining upward trend and an erosive base. Depositional style varies between the Channel Units: where Channel Unit 1 (lowermost) has internal small-scaled fining upward cycles related to local cut-off, Channel Unit 2 has a gradual fining upward sequence. The sandstones of the braided channel lithofacies association are interpreted as bedload deposits of a high-energy braided channel complex system.

About 11% of the cored interval was assigned to the Crevasse Splay (*CS*) deposits. *CS* is assigned to parts of example slab 2 (Fig. 6.4). *CS* deposits consist of parallel, low-angle, and ripple laminated very fine-grained sandstone beds with moderate amounts of carbonaceous

matter and clay (5-35%) occurring as intraclasts and laminations. Also, these deposits show rootlets at the top. These deposits are interpreted as crevasse splay deposits on the basis of the fine-grained nature of the sandstone beds, the relatively thin beds, the clay occurring as laminae and the occurrence within floodplain sediments.

Poorly drained floodplain deposits (*IFL*) comprise nearly 18% of the cored interval. *IFL* is assigned to parts of ES5 (see Fig. 6.4). The association consists of dark grey to black claystones with regular coal laminae, abundant rootlets, regular coal laminae, and frequent siderite nodules. The poorly drained floodplain is recognised as such by the dark grey colour, the abundant rootletting, the siderite nodules, and good preservation of organic matter. This all indicates a waterlogged or slightly submerged, reducing environment with intense plant growth.

Nearly 12% of the cored interval consists of Interdistributary Bay (*IB*) deposits. *IB* is assigned to parts of ES2 (see Fig. 6.4). *IB* deposits consists of dark grey, relatively undisturbed laminated claystones with a few silt or sand laminae. Some beds have a yellowish colouration due to siderite cementation. Distortion through soft sediment deformation occurs sporadically. The well-preserved parallel lamination of the claystones, and the absence of rootletting imply deposition within a permanently standing body of water, hence their classification as interdistributary bay deposits.

Swamps (*SW*) comprise a small percentage of the cored interval (ca. 3%). *SW* is assigned to parts of ES5 (see Fig. 6.4). The deposits consist of parallel laminated organic rich claystones with cm-thick coal laminae. The coal beds, and the abundance of organic material suggest a waterlogged, anoxic environment with abundant floral inhabitation and no clastic input.

Approximately 3% of the cored interval consist of Well-drained Floodplain (*F*), which only occurs in the deepest sections of core E10-3. *F* was assigned to parts of ES6 (see Fig. 6.4). It consists of thoroughly bioturbated sandy claystones with a mottled appearance, and an overall reddish colouration. Dark fragments occur, which may represent the former host sediments, giving the rock its brecciated appearance. These sediments are termed Well-drained Floodplain deposits on basis of the following arguments. The strong disturbance of the sediment and the presence of pisoids suggest slow sedimentation rates and pedogenesis. The primary red colouration indicated that at time of deposition the ground water table, Well E10-3 was periodically below the sediment surface at the time of deposition. For example, the environment could have been subject to seasonal flooding.

D.3 Features of the GLC

We propose a number of extensions to the MLC in order make it suitable for prediction of petrophysical properties and to obtain improved calibration performance. We refer to this generalization of the MLC method as the Generalized Logratio Calibration (GLC) method.

D.3.1 Extension 1: Model selection

The first extension of is selection of the optimal number of components on the basis of the Median Squared Prediction Error (MSPE) given by:

$$\text{MSPE} = \text{med}_j \left\{ \frac{\sum_{j=1}^D (y_{ij} - \hat{y}_{ij})^2}{D \left(1 - \frac{2}{9D}\right)^3} \right\} \quad (\text{D.3.1})$$

where y reflects the clr variables of W . We define the optimal PLS model as the one that gives the smallest MSPE. Because the MSPE is based upon a median, it is a robust measure of scale. By doing so, we reduce the influence of outliers. Furthermore the denominator term is introduced in order to ensure that the MSPE is a consistent estimator for the variance: the denominator is an approximation to the median of a χ^2 -distributed random variable (Abrahamowitz and Stegun, 1964).

D.3.2 Extension 2: Transformation selection

The second extension is the transformation applied to the Y-data. Because petrophysical data have different statistical properties, they also require different transformations. The GLC is based on selecting the appropriate transformation for a particular data type based on the sample space. The goal is to ensure that the data become unbound. By doing so, we ensure that the predictions are physically tractable.

We distinguish between four different data types with the following transformations and associated inverse transformations, respectively:

$$\begin{array}{ll|ll}
 \text{Type 1} & y \in [-\infty, \infty] & | & y_* = y \\
 \text{Type 2} & y \in [0, 1] & | & y_* = \ln\left(\frac{y}{1-y}\right), \quad y = \frac{\exp(y_*)}{\exp(y_*)+1} \\
 \text{Type 3} & y \in [0, \infty] & | & y_* = \ln(y), \quad y = \exp(y_*) \\
 \text{Type 4} & \mathbf{y} \in S^D & | & \mathbf{y}_* = \text{clr}(\mathbf{y}), \quad \mathbf{y} = \text{clr}^{-1}(\mathbf{y}_*)
 \end{array} \tag{D.3.2}$$

When expressed in terms of the logarithmic phi-scale, mean grain size is an example of type-1 data. Porosity and permeability are examples of type-2 and type-3 data, respectively. Geochemistry is data of type 4. Since compositions are intrinsically correlated and must therefore be modelled simultaneously, type 4 data are the only type where the Y-data are explicitly multivariate (type-2 data are compositional and therefore bivariate, however the information can be captured by just one variable). Note that geochemical calibration is now a special case of the GLC whereby the data are of type 4.

D.3.3 Extension 3: Removing uninformative X-variables

Another extension is removal of uninformative X-variables. Removal of uninformative X-variables is important because clr-transformed variables are not independent and likewise all variables are "polluted" when noisy, uninformative variables are included in the model.

X-variable removal is performed based on the SIMPLS weights matrix \mathbf{R} . Investigating the weights associated with the first PLS component (i.e. the first column of \mathbf{R}) gives insight into the added value of the X-variables. However, for a more complete overview over the variable importance in prediction all PLS components need to be assessed. Hence, we perform this step on the basis of the Variable Importance in Projection (Eriksson et al., 2001):

$$\text{VIP}_j = \sqrt{D_x \sum_{p=1}^{D_x} \frac{SS_p}{\sum_{p=1}^P (SS_p)} (w_{jp}/|w_j|)^2} \tag{D.3.3}$$

In this Equation, SS_p represents the variance explained by the p th component which is given by:

$$SS_p = \mathbf{t}'_p \mathbf{t}_p \mathbf{q}'_p \mathbf{q}_p \tag{D.3.4}$$

The VIP has been used extensively in process chemometrics for variable selection in PLS modelling. A threshold of unity is most commonly used (Mehmood et al., 2012). In this contribution, however, we use a VIP threshold of 0.25. The reason for choosing a relatively low VIP threshold is that clr-transformed data are by definition statistically dependent which means that too rigorous variable selection may result in the removal of informative variables. Moreover, because clr-transformed variables are correlated simply removing a variable is not possible without forming new clr-transformed data. Hence, after variable exclusion new X- and Y-variables need to be formed.

Apart from variable selection, the VIP is also used as a qualitative tool to gain insight into the significance of chemical analytes for the petrophysical properties of the sediment or sedimentary rock. Another measure which serves the same purpose is Pearson's cross-correlation coefficients between the element intensities and the univariate petrophysical properties:

$$\rho_j = \text{corr}(\mathbf{x}_j^*, \mathbf{y}^*) \quad (\text{D.3.5})$$

In this equation, \mathbf{x}_j^* is a transformed and appropriately scaled intensity variable j , and \mathbf{y}^* is the transformed property of interest (e.g., the log-transformed permeability). For multivariate Y-data, more sophisticated methods need to be used to analyse the cross-correlation structure (see Bloemsma et al. (2012) and Chapter 3 of this thesis).

D.3.4 Extension 4: Removing unpredictable Y-variables

The fourth extension is the identification and removal of unpredictable Y-variables. Given that the multivariate data are compositional (i.e., data of type 4), we propose a methodology tailored to compositional data.

Y-variable removal and is done in a sequential manner. Identification of unpredictable Y-variables is carried out by calculating the following intermediate variable:

$$\mathbf{z}_j = \log \left(\frac{\mathbf{c}_j}{\mathbf{1} - \mathbf{c}_j} \right) \quad (\text{D.3.6})$$

Subsequently, the signal strength R^2 is calculated for every analyte j :

$$R_j^2 = 1 - \frac{\text{var}(\mathbf{z}_j - \hat{\mathbf{z}}_j)}{\text{var}(\mathbf{z}_j)} \quad (\text{D.3.7})$$

The variable with the smallest R_j^2 is removed, unless it is larger than the user-defined threshold value. The calibration model is then re-executed using the reduced set of variables. The process is terminated in case no variables are removed.

The reason for calculating the signal strength in this manner is that treating the D -part composition as D 2-part compositions consisting of the respective element and its complement, allows us to analyse the signal strength of the predicted element concentrations in a pseudo-independent manner. In addition, notice that by defining it in this manner it reduces to the same form as the R_{eff}^2 in case the number of components equals two.

D.3.5 Extension 5: Modelling incomplete compositions

Type-4 data are compositional data which means that their sum equals unity. Geochemical data, however, are generally not proper compositional data sets because most analytical methods are unable to quantify the concentration of all components in the mixture.

For instance, carbon and oxygen cannot be quantified using XRF or ICP. General log-ratio methods are not appropriate for modelling these incomplete compositional data because "absolute" information, i.e. information about the proportion of quantified components is lost.

In Chapter 2, this problem was circumvented by calculating the "undef" vector whose elements are given by:

$$u_i = 1 - \sum_{j=1}^D w_{ij} \quad (\text{D.3.8})$$

It represents the proportion of mass that could not be adequately quantified. Adding this "undef" column vector to the matrix with measured concentrations yields a proper compositional matrix \mathbf{C} :

$$\mathbf{C} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_D, \mathbf{u}] \quad (\text{D.3.9})$$

As a result, predictions can be made in an unbiased manner. This approach proved to be quite effective for data sets with limited variations in this "undef" (see Weltje et al. (2015) and Chapter 2). In the light of the extensions proposed here, mixing up scale and composition can be problematic in case "undef" is difficult to predict: the correlation among clr-variables considerably decreases the predictive power of the model as a whole and jeopardises Y-variable removal. Moreover, removal of the "undef" variable would imply that "absolute" information is lost which is undesirable.

To circumvent these problems we make the following decomposition of the matrix \mathbf{W} containing incomplete compositions:

$$\mathbf{W} = \text{diag}(\mathbf{t})\mathbf{C} \quad (\text{D.3.10})$$

The matrix \mathbf{C} contains compositional data and the vector \mathbf{t} contains the row-specific scale. The advantage of decomposing an incomplete composition in this manner is that the composition and the scale can be separately modelled. The first model is associated with the subcompositional variability in the observed components (i.e. data type 4), the second model is associated with the scale which is a compositional quantity of data type 2.

List of acronyms

alr	additive log-ratio (transformation)
ALS	Alternating Least Squares
BCU	Braided Channel Unit
BHF	Binomial Horwitz Function
BHR	Binomial Horwitz Ratio
BLC	Bivariate Logratio Calibration
blr	binary log-ratio (transformation)
BUCE	Basic Univariate Calibration Equation
BUF	Binomial-like Uncertainty Function
CF	Characteristic Function
clr	centred log-ratio (transformation)
CUCE	Consistent Univariate Calibration Equation
DCM	Data Covariance Matrix
DFA	Discriminant Function Analysis
DLC	Direct Linear Calibration
ECM	Error Covariance Matrix
GLC	Generalised Logratio Calibration
GPR	Geochemical Performance Ratio
iid	independent and identically distributed
LOOCV	Leave-one-out cross-validation
LOPP	Limit of Physical Plausibility
MSI	Model Stability Index
MSE	Mean Squared Error
MRE	Mean Relative Error
MSPE	Mean Squared Prediction Error
MRPE	Mean Relative Prediction Error
MDD	Multinomial-Distributed Data
MLC	Multivariate Logratio Calibration

MLRRA	Maximum Likelihood Reduced-Rank Approximation
MLPCA	Maximum Likelihood Principal Components Analysis
OSIRA	Optimal Scale-Invariant Reduced-rank Approximation
PCA	Principal Components Analysis
PDD	Poisson-Distributed Data
PG	Performance Gain
PLS(R)	Partial Least Squares (Regression)
PT	Proficiency Test
PPR	Predictive Performance Ratio
PUF	Poisson-like Uncertainty Function
QDFA	Quadratic Discriminant Function Analysis
RPG	Relative Performance Gain
RQI	Reservoir Quality Index
RRA	Reduced-Rank Approximation
SDR	Inter-laboratory standard deviation
SVD	Singular Value Decomposition
TPR	True Positive Ratio
TPM	Transition Probability Matrix
UF	Uncertainty Function
ULC	Univariate Logratio Calibration
VIP	Variable Importance in Projection
XRR	X-data Repeatability Ratio
XRF	X-Ray Fluorescence
XRF-CS	X-Ray Fluorescence Core Scanning
XRD	X-Ray Diffraction

Bibliography

- Abrahamowitz, M., Stegun, I. (Eds.), 1964. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 1st Edition. Dover Publications.
- Aitchison, J., 1982. The statistical analysis of compositional data. *Journal of the Royal Statistical Society* 44, 139–177.
- Aitchison, J., 1986. *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.
- Aitchison, J., 1999. Logratios and natural laws in compositional data analysis. *Math. Geol.* 31, 563–580.
- Aitchison, J., Egozcue, J. J., 2005. Compositional data analysis: Where are we and where should we be heading? *Math. Geol.* 37, 829–850.
- Aitchison, J., Greenacre, M., 2002. Biplots for compositional data. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 51, 375–392.
- Albert, R., Horwitz, W., 1997. A heuristic derivation of the Horwitz curve. *Anal. Chem.* 69, 780–790.
- Allen, P. A., 2009. *Earth Surface Processes*. Blackwell Science Ltd.
- Allen, P. A., Allen, J. R., 2005. *Basin Analysis: Principles and Applications*, 2nd Edition. Blackwell Publishing.
- Amante, C., Eakins, B. W., 2008. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. National Geophysical Data Center, NESDIS, NOAA, U.S. Department of Commerce, (publicly available at: <http://www.ngdc.noaa.gov/mgg/global/global.html>).
- Amorosi, A., Centineo, M. C., Dinelli, E., Lucchini, F., Tateo, F., 2002. Geochemical and mineralogical variations as indicators of provenance changes in Late Quaternary deposits of SE Po Plain. *Sediment. Geol.* 151, 273–292.

- Argast, S., Donnelly, T. W., 1987. The chemical discrimination of clastic sedimentary components. *J. Sediment. Petrol.* 57, 813–823.
- 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.
- Bakke, J., Trachsel, M., Kvisvik, B. C., Nesje, A., Lysa, A., 2013. Numerical analyses of a multi-proxy data set from a distal glacier-fed lake, Sørsendalsvatn, western Norway. *Quatern. Sci. Rev.* 73, 182–195.
- Bakker, E., Pruno, S., 2004. Conventional and special core analysis on cores from well E10-3XX. Tech. Rep. No. c0222, Panterra Nederland BV.
- Basu, A., 2003. A perspective on quantitative provenance analysis. In: Valloni, R., Basu, A. (Eds.), *Quantitative Provenance Studies in Italy*, *Memorie Descrittive della Carta Geologica dell'Italia*. Vol. 64. pp. 11–22.
- Blatt, H., Middleton, G. V., Murray, R. C., 1980. *Origin of Sedimentary Rocks*. Prentice-Hall, Englewood Cliffs, NJ.
- Bloemsma, M. R., Weltje, G. J., 2015. Reduced-rank approximations to spectroscopic and compositional data: A universal framework based on log-ratios and counting statistics. *Chemom. Intell. Lab. Syst.* 142, 206–218.
- Bloemsma, M. R., Weltje, G. J., submitted. Predicting the uncertainty of routine chemical analyses: a new modelling approach based on evaluation of proficiency tests. *Analyst*.
- Bloemsma, M. R., Zabel, M., Stuut, J. B., Tjallingii, R., Collins, J., Weltje, G. J., 2012. Modelling the joint variability of grain size and chemical composition in sediments. *Sediment. Geol.* Corrigendum in 2013 (doi: 10.1016/j.sedgeo.2012.12.001).
- Boels, J., 2003. Sedimentology, petrography and reservoir quality of the Upper Carboniferous in well E10-3. Tech. Rep. G453, Panterra Nederland BV.
- 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, doi:10.1029/2006GC001480.
- Boyle, E. A., 1983. Chemical accumulation variations under the Peru Current during the past 130,000 years. *J. Geophys. Res.* 88(C12), 7667–7680, doi:10.1029/JC088iC12p07667.
- Buccianti, A., Mateu-Figueiras, G., Pawlowsky-Glahn, V. (Eds.), 2006. *Compositional Data Analysis in the Geosciences*. Geological Society, London, Special Publication 264.
- Calvert, S. E., Pederson, T. F., 2007. Elemental proxies for palaeoclimatic and palaeoceanographic variability in marine sediments: Interpretation and Application. In: Hillaire-Marcel, C., de Vernal, A. (Eds.), *Proxies in Late Cenozoic Paleoclimatology*. Developments in Quaternary Research, vol. 1. Elsevier Science, Amsterdam, pp. 567–644.
- Chayes, F., 1960. On correlation between variables of constant sum. *J. Geophys. Res.* 65, 4185–4193.

- Clarke, F. W., Washington, H. S., 1924. The composition of the Earth's crust. United States Geological Survey, Professional Paper 127, 117.
- Cochran, R., Horne, F. H., 1977. Statistically weighted principal component analysis of rapid scanning wavelength kinetics experiments. *Anal. Chem.* 49 (6), 846–853.
- Cronin, T. M., Thunell, R., Dwyer, G. S., Saenger, C., Mann, M. E., Vann, C., Seal, R. R., 2005. Multiproxy evidence of Holocene climate variability from estuarine sediments, eastern North America. *Paleoceanography* 20, 135–174, pA4006, doi:10.1029/2005PA001145.
- Croudace, I. W., Bull, M. J., Cotteril, C. J., Dix, J. K., Taylor, R. N., 2015. X-ray core scanners as an environmental forensic tool: a case study of polluted harbour sediment (Augusta Bay, Sicily). *Developments in Paleoenvironmental Research: Micro-XRF Studies of Sediment Cores*.
- Croudace, I. W., Rindby, A., Rothwell, R. G., 2006. Itrax: description and evaluation of a new multi-function x-ray core scanner. In: Rothwell, R. G. (Ed.), *New Techniques in Sediment Core Analysis*. Geological Society, London, Special Publication 267, pp. 79–98.
- de Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. *Chemometr. Intell. Lab. Syst.* 18, 251–263.
- De Pol-Holz, R., Ulloa, O., Lamy, F., Dezileau, L., P., S., Hebbeln, D., 2007. Late Quaternary variability of sedimentary nitrogen isotopes in the Eastern South Pacific Ocean. *Paleoceanography* 22, pA2207.
- De Vries, J. L., Vrebos, B. A. R., 2002. Quantification of infinitely thick specimens by XRF analysis. In: van Grieken, R. E., Markovicz, A. A. (Eds.), *Handbook of X-ray Spectrometry*, 2nd Edition. Marcel Dekker, New York, pp. 314–405.
- deMenocal, P. B., Ruddiman, W. F., Pokras, E. M., 1993. Influences of high- and low- latitude processes on African terrestrial climate: Pleistocene eolian records from equatorial Atlantic ocean drilling program Site 663. *Paleoceanography* 8, 209–242.
- Dinelli, E., Tateo, F., Summa, V., 2007. Geochemical and mineralogical proxies for grain size in mudstones and siltstones from the Pleistocene and Holocene of the Po River alluvial plain, Italy. In: Arribas, J., Critelli, S., Johnsson, M. (Eds.), *Sedimentary Provenance and Petrogenesis: Perspectives from Petrography and Geochemistry*. Geological Society of America Special Papers, pp. 25–36.
- Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. *Math. Geol.* 35, 279–300.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S., 2001. *Multi- and Megavariate Data Analysis: Principles and Applications*. Umetrics Academy, Umeå.
- EUROLAB, 2007. Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation, eurolab technical report no. 1/2007. <http://www.eurolab.org>.
- Forbes, P., 1998. The status of core analysis. *J. Petrol. Sci. Eng.* 19, 1–6.
- Friedman, J. H., 1991. Multivariate adaptive regression splines. *Annals of Statistics* 19, 167.

- Gac, J. Y., Kane, A., 1986. Le fleuve Senegal: I. Bilan hydrologique et flux continentaux de matières particulaire à l'embouchure. *Sci. Geol. Bull.* 39, 99–130.
- Garcia, D., Ravenec, C., Marechal, B., Mouttea, J., 2004. Geochemical variability induced by entrainment sorting: quantified signals for provenance analysis. *Sediment. Geol.* 171, 113–128.
- Gé, L. Q., Lai, W. C., Lin, Y. C., 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., 1993. *Predictive inference: An introduction*. Chapman and Hall, New York.
- Goldberg, E. D., Arrhenius, G. O. S., 1958. Chemistry of Pacific pelagic sediments. *Geochim. Cosmochim. Acta* 13, 153–212.
- Griffioen, J., Klein, J., van Gaans, P. F. M., 2012. Reaction capacity characterization of shallow sedimentary deposits in geologically different regions of the Netherlands. *J. Cont. Hydr.* 127, 30–46.
- Gy, P., 1979. *Sampling of Particulate Materials. Theory and Practice*. Elsevier, Amsterdam.
- Hall, P., Selinger, B., 1989. A statistical justification relating interlaboratory coefficients of variation with concentration levels. *Anal. Chem.* 61, 1465–1466.
- Haschke, M., 2006. The Eagle III BKA system, a novel sediment core X-ray fluorescence analyzer with very high spatial resolution. In: Rothwell, R. (Ed.), *New Techniques in Sediment Core Analysis*. Geological Society, London, Special Publication 267, pp. 31–37.
- Haschke, M., Scholz, W., Theis, U., Nicolosi, J., Scruggs, B., Herzceg, L., 2002. Description of a new micro-x-ray spectrometer. *Journal de Physique IV France* 12, 6–83.
- Haslett, S. K., Davies, C. F. C., 2006. Late Quaternary climate-ocean changes in western North Africa: offshore geochemical evidence. *Transactions of the Institute of British Geographers* 31, 34–52.
- Haslett, S. K., Smart, C. W., 2006. Late Quaternary upwelling off tropical NW Africa: new micropalaeontological evidence from ODP Hole 658C. *Journal of Quaternary Science* 21, 259–269.
- Hebbeln, D., cruise participants, 2001. PUCK, report and preliminary results of R/V Sonne cruise 156, Valparaiso (Chile)-Talcahuano (Chile), March 29 - May 14, 2001. Univ. Bremen, Bremen, Germany.
- Henares, S., Bloemsmma, M. R., Donselaar, M. E., Mijnlief, H. F., Redjosentono, A. E., Veldkamp, H. G., Weltje, G. J., 2014. The role of detrital anhydrite in diagenesis of aeolian sandstones (Upper Rotliegend, The Netherlands): Implications for reservoir-quality prediction. *Sediment. Geol.* 314, 60–74.
- Hertzog, R., Colson, L., Seeman, B., O'Brien, M., Scott, H., McKeon, D., Wraight, P., Garu, J., Ellis, D., Schweitzer, J., Herron, M., 1987. Geochemical logging with spectrometry tools. SPE Paper 16792.

- Hollander, M., Wolfe, D. A., 1973. *Nonparametric Statistical Methods*. Wiley.
- Holz, C., Stuut, J.-B. W., Henrich, R., 2004. Terrigenous sedimentation processes along the continental margin off NW Africa: implications from grain-size analysis of seabed sediments. *Sedimentology* 51, 1145–1154.
- Horwitz, W., 1982. Evaluation of analytical methods used for regulation of foods and drugs. *Anal. Chem.* 54, 67–76.
- Horwitz, W., Albert, R., 2006. The Horwitz Ratio (HorRat): A useful index of method performance with respect to precision. *J. Assoc. Off. Anal. Chem.* 89, 1095–1109.
- Horwitz, W., Kamps, L. R., Boyer, K. W., 1980. Quality assurance in the analysis of foods and trace constituents. *J. Assoc. Off. Anal. Chem.* 63, 1344–1354.
- Hughes, H., Hurley, P. W., 1987. Precision and accuracy of test methods and the concept of k-factors in chemical analysis. *Analyst* 112, 1445–1449.
- Hund, E., Massart, D., Smeyers-Verbeke, J., 2000. Inter-laboratory studies in analytical chemistry. *Anal. Chim. Acta.* 423, 145–165.
- Jaccard, S. L., Haug, G. H., Sigman, D. M., Pedersen, T. F., Thierstein, H. R., Röhl, U., 2005. Glacial/interglacial changes in Subarctic North Pacific stratification. *Science* 308, 1003–1006.
- Jansen, J. H. F., van der Gaast, S. J., Koster, B., Vaars, A., 1998. Cortex, a shipboard XRF-scanner for element analyses in split sediment cores. *Mar. Geol.* 151, 143–153.
- JCGM, 2008. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. Tech. rep., Joint Committee for Guides in Metrology (JCGM).
- Jenkins, R., 1999. *X-Ray Fluorescence Spectroscopy*. Wiley & Sons, New York.
- Jiménez-Chacón, J., Alvarez-Prieto, M., 2006. Modelling uncertainty in a concentration range. *Accred. Qual. Assur.* 14, 15–27.
- Johnsson, M. J., 1993. The system controlling the composition of clastic sediments. In: Johnsson, M. J., Basu, A. (Eds.), *Processes Controlling the Composition of Clastic Sediments*. Geological Society of America Special Paper 284, pp. 1–19.
- Jolliffe, D. E., 2002. *Principal Component Analysis*. Springer-Verlag, New York.
- Jones, S. C., 1994. A new, fast, accurate pressure-decay probe permeameter. *SPE Formation Evaluation* 9, 1993–199.
- Kaiser, J., Schefuss, E., Lamy, F., Mohtadi, M., Hebbeln, D., 2008. Glacial to Holocene changes in sea surface temperature and coastal vegetation in north central Chile: high versus low latitude forcing. *Quaternary Science Reviews* 27, 2064–2075.
- Keenan, M. R., Kotula, P. G., 2004. Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images. *Surf. Interface Anal.* 36, 203–212.

- Kendall, M. G., Stuart, A., 1973. *The Advanced Theory of Statistics, Volume 2: Inference and Relationships*. Griffin, London.
- Kido, Y., Koshikawa, T., Tada, R., 2006. Rapid and quantitative major element analysis method for wet fine-grained sediments using an XRF microscanner. *Marine Geol.* 229, 209–225.
- Kiminami, K., Fuji, K., 2007. The relationship between major element concentration and grain size. *Sediment. Geol.* 195, 203–215.
- Klump, J., Hebbeln, D., Wefer, G., 2000. The impact of sediment provenance on barium-based productivity estimates. *Mar. Geol.* 169, 259–271.
- Komar, P. D., 2007. The entrainment, transport, and sorting of heavy minerals by waves and currents. In: Mange, M., Wright, D. (Eds.), *Heavy Minerals in Use: Developments in Sedimentology*. Amsterdam, Elsevier, pp. 3–48.
- Koopmann, B., 1981. Sedimentation von Saharastaub im subtropischen Nordatlantik während der letzten 25.000 Jahre. *Meteor Forschungsergeb., Reihe C* 35, 23–59.
- Kuhn, G., Helling, D., von Eynatten, H., Niessen, F., Magens, D., 2008. Past warmer climate periods at the Antarctic margin detected from proxies and measurements of biogenic opal in the ABD-1B core: the XRF spectral silver (Ag) peak used as a new tool for biogenic opal quantification. In: *Proceedings of the AGU Fall Meeting, December 14-19, San Fransisco, USA*. Hdl:10013/epic.32435.
- Lamy, F., Hebbeln, D., Wefer, G., 1998. Terrigenous sediment supply along the Chilean continental margin: modern regional patterns of texture and composition. *Geologische Rundschau* 87, 477–494.
- Lamy, F., Klump, J., Hebbeln, D., Wefer, G., 2000. Late Quaternary rapid climate change in northern Chile. *Terra Nova* 12, 8–13.
- Lawn, R. E., Thompson, M., Walker, R. F., 1993. *Proficiency testing in analytical chemistry*. The Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge, UK.
- Leger, M. N., Vega-Montoto, L., Wentzell, P. D., 2005. Methods for systematic investigation of measurement error covariance matrices. *Chemometr. Intell. Lab. Syst.* 77, 181–205.
- Lokhorst, A. (Ed.), 1997. *NW European Gas Atlas*. Haarlem (NITG-TNO), The Netherlands.
- Luo, Y., Geo, S., Longerich, H. P., Günther, D., Wunderli, S., Yuan, H.-. L., Liu, X.-. M., 2006. The uncertainty budget of the multi-element analysis of glasses using LA-ICP-MS. *J. Anal. Atom. Spectrom.* 22, 122–130.
- Martens, H., Hoy, M., Wise, B. M., Bro, R., Brockhoff, P. B., 2003. Pre-whitening of data by covariance-weighted pre-processing. *J. Chemometr.* 17, 153–165.
- Martens, H., Naes, T., 1989. *Multivariate Calibration*. Wiley, Chicester.

- Martín-Fernández, J. A., Barceló-Vidal, C., Pawlowsky-Glahn, V., 2003. Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Math. Geol.* 35, 253–278.
- MATLAB, 2011. Version 7.11.0 (R2010b). The MathWorks Inc., Natick, Massachusetts, United States.
- McLennan, S. M., Hemming, S., McDaniel, D. K., Hanson, G. N., 1993. Geochemical approaches to sedimentation, provenance and tectonics. In: Johnsson, M., Basu, A. (Eds.), *Processes Controlling the Composition of Clastic Sediments*. Geological Society of America, Special Papers 285, pp. 21–40.
- Meggers, H., cruise participants, 2002. Report and preliminary results of R/V Meteor cruise M 53/1, Limasol - Las Palmas - Mindelo, 30.03 - 03.05.2002. Univ. Bremen, Bremen, Germany.
- Mehmood, T., Liland, K. H., Snipen, L., Sæbø, S., 2012. A review of variable selection methods in Partial Least Squares Regression. *Chemometr. Intell. Lab. Syst.* 118, 62–69.
- Mol, G., Spijker, J. H., Van Gaans, P. F. M., Römkens, P. F. A. M. (Eds.), 2012. *Geochemische bodematlas van Nederland*. Wageningen: Wageningen Academic Publishers.
- Moreno, T., Querol, X., Castillo, S., Alastuey, A., Cuevas, E., Herrmann, L., Mounkaila, M., Elvira, J., Gibbons, W., 2006. Geochemical variations in aeolian mineral particles from the Sahara-Sahel Dust Corridor. *Chemosphere* 65, 261–270.
- Mulitza, S., cruise participants, 2006. Report and preliminary results of R/V METEOR Cruise M65/1, Dakar-Dakar, 11.06.2005-01.07.2005. Univ. Bremen, Bremen, Germany.
- Mulitza, S., Prange, M., Stuut, J. B., Zabel, M., von Dobeneck, T., Itambi, A. C., Nizou, J., Schulz, M., Wefer, G., 2008. Sahel megadroughts triggered by glacial slowdowns of Atlantic meridional overturning. *Paleoceanography* 23, PA4206, doi:10.1029/2008PA001637.
- Murray, R. W., Leinen, M., 1996. Scavenged excess aluminum and its relationship to bulk titanium in biogenic sediment from the central equatorial Pacific Ocean. *Geochim. Cosmochim. Acta* 60, 3869–3878.
- Nesbitt, H. W., Young, G. M., 1982. Early Proterozoic climates and plate motions inferred from major element chemistry of lutites. *Nature* 299, 715–717.
- Nesbitt, H. W., Young, G. M., 1984. Prediction of some weathering trends of plutonic and volcanic rocks based on thermodynamic and kinetic considerations. *Geochim. Cosmochim. Acta* 48, 1523–1548.
- Nesbitt, H. W., Young, G. M., 1996. Petrogenesis of sediments in the absence of chemical weathering: effects of abrasion and sorting on bulk composition and mineralogy. *Sedimentology* 43, 341–358.
- Ohta, T., Arai, H., 2007. Statistical empirical index of chemical weathering in igneous rocks. *Chem. Geol.* 240, 280–297.

- Pälike, H., Shackleton, N. J., Röhl, U., 2001. Astronomical forcing on late Eocene marine sediments. *Earth Planet. Sci. Lett.* 193, 589–602.
- Pawlowsky-Glahn, V., Egozcue, J. J., 2001. Geometric approach to statistical analysis on the simplex. *Stochastic Environ. Res. Risk Assess.* 15, 384–398.
- Pawlowsky-Glahn, V., Egozcue, J. J., 2002. BLU Estimators and compositional data. *Math. Geol.* 34, 259–274.
- Pe-Piper, G., Karim, A., Piper, D. J. W., 2011. Authigenesis of titania minerals and the mobility of Ti, new evidence from pro-deltaic sandstones, Cretaceous Scotian Basin, Canada. *J. Sediment. Res.* 81, 762–773.
- Pe-Piper, G., Triantafyllidis, S., Piper, D. J. W., 2008. Geochemical identification of clastic sediment provenance from known sources of similar geology: the Cretaceous Scotian basin, Canada. *J. Sediment. Res.* 78, 595–607.
- Pettijohn, F. J., Potter, P. E., Siever, R., 1987. *Sand and Sandstone*, Second Edition. Springer, New York.
- Potts, P. J., Webb, P. C., Thompson, M., 2013. An assessment of performance in the routine analysis of silicate rocks based on an analysis of data submitted to the GeoPT proficiency testing program for geochemical laboratories (2001–2011). *Geostandards and Geoanalytical Research* 37, 403–416.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., 1994. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, Second Edition. Cambridge University Press, Cambridge.
- Ramsey, M. H., Potts, P. J., Webb, P. C., Watkins, P., Watson, J. S., Coles, B. J., 1995. Objective assessment of analytical method precision: comparison of ICP-AES and XRF for the analysis of silicate rocks. *Chem. Geol.* 124, 1–19.
- Rea, D. K., 1994. The paleoclimatic record provided by eolian deposition in the deep sea: the geologic history of wind. *Geophysics* 32, 159–195.
- Reijers, T. J. A., Mijnlief, H. F., Pestman, P. J., Kouwe, W. F. P., 1993. *Lithofacies and their interpretation: a guide to standardised description of sedimentary deposits*. Tech. rep., Rijks Geologische Dienst.
- Reis, M. S., Saraiva, P. M., 2006a. Heteroscedastic latent variable modelling with applications to multivariate statistical process control. *Chemometr. Intell. Lab. Syst.* 80, 57–66.
- Reis, M. S., Saraiva, P. M., 2006b. Integration of data uncertainty in linear regression and process optimization. *AIChE J.* 51, 3007–3019.
- Richter, T. O., Van der Gaast, S., Koster, B., Vaars, A., Gieles, R., de Stigter, H., de Haas, H., van Weering, T. C. E., 2006. The Avaatech XRF core scanner: technical description and applications to NE Atlantic sediments. In: Rothwell, R. E. (Ed.), *New Techniques in Sediment Core Analysis*. Special Publication. Vol. 267. Geological Society of London, pp. 39–50.

- Romero, O. E., Kim, J. H., Donner, B., 2008. Submillennial-to-millennial variability of diatom production off Mauritania, NW Africa during the last glacial cycle. *Paleoceanography* 23, pA3218, doi:10.1029/2008PA001601.
- Rothwell, R. G., Hoogakker, B., Thomson, J., Croudace, I. W., 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, R. G. (Ed.), *New Techniques in Sediment Core Analysis*. Geological Society, London, Special Publication 267, pp. 79–98.
- Rothwell, R. G., Rack, F. R., 2006. New techniques in sediment core analysis: an introduction. In: Rothwell, R. G. (Ed.), *New Techniques in Sediment Core Analysis*. Geological Society, London, Special Publication 267, pp. 1–29.
- Rubin, D. M., 2004. A simple autocorrelation algorithm for determining grain size from digital images of sediment. *J. Sediment. Res.* 74, 160–165.
- Sandford, R. F., Pierson, C. T., Crovelli, R. A., 1993. An objective zero-replacement method for censored geochemical data. *Math. Geol.* 25:1, 59–80.
- Sarnthein, M., 1978. Sand deserts during glacial maximum and climatic optimum. *Nature* 272, 43–45.
- Sarnthein, M., Tetzlaff, G., Koopmann, B., Wolter, K., Pflaumann, U., 1981. Glacial and interglacial wind regimes over the eastern subtropical Atlantic and North-West Africa. *Nature* 293, 193–196, doi: 10.1038/293193a0.
- Schneider, R. R., Price, B., Müller, P. J., Kroon, D., Alexander, I., 1997. Monsoon related variations in Zaire (Congo) sediment load and influence of fluvial silicate supply on marine productivity in the east equatorial Atlantic during the last 200,000 years. *Paleoceanography* 12, 463–481.
- Schreyer, S. K., Bidinosti, M., Wentzell, P. D., 2002. Application of maximum likelihood principal components regression to fluorescence emission spectra. *App. Spectr.* 56 (6), 789–796.
- Serra, O., 1987. *Fundamentals of Well-Log Interpretation. II: The Interpretation of Logging Data*. Elsevier, Amsterdam.
- Solano-Acosta, W., Dutta, P. K., 2005. Unexpected trend in the compositional maturity of second-cycle sand. *Sedimen. Geol.* 178, 275–283.
- Spears, D. A., Kanaris-Sotiriou, R., 1976. Titanium in some Carboniferous sediments from Great Britain. *Geochim. Cosmochim. Acta* 40, 345–351.
- Stuut, J.-B., Kasten, S., Lamy, F., Hebbeln, D., 2007. Sources and modes of terrigenous sediment input to the Chilean continental slope. *Quaternary International* 161, 67–76.
- Stuut, J. B., Zabel, M., Ratmeyer, V., Helmke, P., Schefuss, E., Lavik, G., Schneider, R., 2005. Provenance of present-day eolian dust collected off North-Africa. *J. Geophys. Res.* 110, doi:10.1029/2004JD005161.

- Stuut, J.-B. W., Lamy, F., 2004. Climate variability at the southern boundaries of the Namib (southwestern Africa) and Atacama (northern Chile) coastal deserts during the last 120,000 yr. *Quaternary Research* 62, 301–309.
- Svendsen, J., Friis, H., Stollhofen, H., Hartley, N., 2007. Facies discrimination in a mixed fluvio-eolian setting using elemental whole-rock geochemistry applications for reservoir characterization. *J. Sediment. Res.* 77, 23–33.
- Swerts, J., Van Espen, P., 1993. Partial Least Squares techniques in the Energy-Dispersive X-ray Fluorescence determination of sulfur-graphite mixtures. *Analytical Chemistry* 65, 1181–1185.
- Thomas, D. C., Pugh, V. J., 1989. A statistical analysis of the accuracy and reproducibility of standard core analysis. *Log Anal.*, 71–77(journal version of SCA 8701).
- Thompson, M., 2000. Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing. *Analyst* 125, 385–386.
- Thompson, M., 2002. Protocol for the operation of GeoPT proficiency testing scheme, international Association of Geoanalysts. <http://www.geoanalyst.org/documents/GeoPT-protocol.pdf>.
- Thompson, M., 2006. Using mixture models for bump-hunting in the results of proficiency tests. *Accred. Qual. Assur.* 10, 501–505.
- Thompson, M., 2011. Uncertainty functions, a compact way of summarising or specifying the behaviour of analytical systems. *TrAC* 30, 1168–1175.
- Thompson, M., Cole, B. J., 2009. Examples of the characteristic function applied to instrumental precision in chemical measurement. *Accred. Qual. Assur.* 14, 147–150.
- Thompson, M., Cole, B. J., 2011. Use of the characteristic function for modelling repeatability precision. *Accred. Qual. Assur.* 16, 13–19.
- Thompson, M., Ellison, S. L. R., 2011. Dark uncertainty. *Accred. Qual. Assur.* 16, 483–487.
- Thompson, M., Howarth, R. J., 1976. Duplicate analysis in Geochemical Practice: Part I. Theoretical Approach and Estimation of Analytical Reproducibility. *Analyst* 101, 690–698.
- Thompson, M., Lowthian, P. J., 1997. The Horwitz function revisited. *J. Assoc. Off. Anal. Chem.* 80, 676–679.
- Thompson, M., Mathieson, K., Damant, A. P., Wood, R., 2008. A general model for interlaboratory precision accounts for statistics from proficiency testing in food analysis. *Accred. Qual. Assur.* 13, 223–230.
- Thompson, M., Potts, P. J., Webb, P. C., Kane, J. S., 1997. GeoPT - A proficiency test for geoanalysis. *Analyst* 122, 1249–1254.

- Tjallingii, R., Claussen, M., Stuut, J.-B., Fohlmeister, J., Jahn, A., Bickert, T., Lamy, F., Röhl, U., 2008. Coherent high- and low-latitude control of the northwest African hydrological balance. *Nature Geoscience* 1, 670–675.
- 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, doi:10.1029/2006GC001393.
- Tolosana-Delgado, R., von Eynatten, H., 2008. Grain-size control on petrographic composition of sediments: Compositional regression and rounded zeros. *Math. Geosci.* 41, 869–886.
- Tolosana-Delgado, R., von Eynatten, H., 2010. Simplifying compositional multiple regression: Application to grain size controls on sediment geochemistry. *Comput. Geosci.* 36, 577–589.
- Trachsel, M., Grosjean, M., Schnyder, D., Kamenik, C., Rein, B., 2010. Scanning reflectance spectroscopy (380–730 nm): a novel method for quantitative high-resolution climate reconstructions from minerogenic lake sediments. *J. Paleolimnol.* 44, 979–994.
- Vekemans, B., Janssen, K., Vincze, L., Adams, F., van Espen, P., 1994. Analysis of X-ray Spectra by Iterative Least Squares (AXIL): New developments. *X-ray Spectrometry* 23, 278–285.
- Vermooten, S., Heerdink, R., Griffioen, J., 2011. Regionale geologische variatie in samenstelling van de grond. *Bodem* 5, 19–21.
- Vlag, P. A., Kruiver, P. P., Dekkers, M. J., 2004. Evaluating climate change by multivariate statistical techniques on magnetic and chemical properties of marine sediments (Azores region). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 212, 23–44.
- von Eynatten, H., Barceló-Vidal, C., Pawlowsky Glahn, V., 2003. Modelling compositional change: The example of chemical weathering of granitoid rocks. *Math. Geol.* 35, 231–251.
- Ward, C. R., Taylor, J. C., Cohen, D. R., 1999. Quantitative mineralogy of sandstones by X-ray diffractometry and normative analysis. *J. Sediment. Res.* 69, 1050–1062.
- Ward, J. H., 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58, 236–244.
- Weltje, G. J., 2004. A quantitative approach to capturing the compositional variability of modern sands. *Sediment. Geol.* 171, 59–77.
- Weltje, G. J., 2012. Quantitative models of sediment generation and provenance: State of the art and future developments. *Sediment. Geol.* 280, 4–20.
- Weltje, G. J., Bloemsmas, M. R., Tjallingii, R., Heslop, D., Röhl, U., Croudace, I. W., 2015. Prediction of geochemical composition from XRF-core-scanner data: A new multivariate approach including filtering, sample selection, and quantification of uncertainties. *Developments in Paleoenvironmental Research: Micro-XRF Studies of Sediment Cores.*

- Weltje, G. J., Brommer, M. B., 2011. Sediment-budget modelling of multi-sourced basin fills: application to recent deposits of the western Adriatic mud wedge (Italy). *Basin Res.* 23, 291–308.
- Weltje, G. J., Meijer, X. D., de Boer, P. L., 1998. Stratigraphic inversion of siliciclastic basin fills: a note on the distinction between supply signals resulting from tectonic and climatic forcing. *Basin Res.* 10, 129–153.
- Weltje, G. J., Prins, M. A., 2003. Muddled or mixed? inferring palaeoclimate from size distributions of deep-sea clastics. *Sediment. Geol.* 162, 39–62.
- Weltje, G. J., 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.
- Weltje, G. J., von Eynatten, H., 2004. Quantitative provenance analysis of sediments review and outlook. *Sediment. Geol.* 171, 1–11.
- Wentzell, P., Andrews, D. T., Hamilton, D. C., Faber, K., Kowalski, B. R., 1997. Maximum likelihood principal component analysis. *J. Chemometr.* 11, 339–366.
- Wentzell, P. D., 2009. Other topics in soft-modeling: Maximum likelihood-based soft-modeling methods. In: Brown, S. D., Tauler, R., Walczak, B. (Eds.), *Comprehensive Chemometrics*. Elsevier, Amsterdam, pp. 507–558, vol. 2.
- Wentzell, P. D., Andrews, D. T., 1997. Maximum likelihood multivariate calibration. *Anal. Chem.* 69, 2299–2311.
- Wentzell, P. D., Lohnes, M. T., 1999. Maximum likelihood principal component analysis with correlated measurement errors. *Chemometr. Intell. Lab. Syst.* 45, 65–85.
- Whitmore, G. P., Crook, K. A. W., Johnson, D. P., 2004. Grain size control of mineralogy and geochemistry in modern river sediment. *Sediment. Geol.* 195, 203–215.
- Wieling, I. S., 2013. Facies and permeability prediction based on analysis of core images. Master's thesis, Delft University of Technology, the Netherlands.
- Wien, K., Wissmann, D., Külling, M., Schulz, H. D., 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.
- Williams, A., Ellison, S. L. R., Roesslein, M. (Eds.), 2000. Quantifying uncertainty in analytical measurement. LGC Ltd, London, 2nd edn. LGC Ltd, London (ISBN: 0-948926-15-5).
- Wold, S., Martens, H., Wold, H., 1983. The multivariate calibration problem in chemistry solved by the PLS method. In: Ruhe, A., Kagstrom, B. (Eds.), *Proc. Conf. Matrix Pencils*. Springer Verlag, Heidelberg, pp. 286–293.
- Xiong, S., Ding, Z., Zhu, Y., Lu, H., 2010. A ~ 6 Ma chemical weathering history, the grain size dependence of chemical weathering intensity, and its implications for provenance change of Chinese loess-red clay deposit. *Quaternary Science Reviews* 29, 1911–1922.

- Yarincik, K. M., Murray, R. W., Peterson, L. C., 2000. Climatically sensitive eolian and hemipelagic deposition in the Cariaco Basin, Venezuela, over the past 578,000 years: Results from Al/Ti and K/Al. *Paleoceanography* 15, 210–228.
- Young, G. M., Nesbitt, H. W., 1998. Processes controlling the distribution of Ti and Al in weathering profiles, siliciclastic sediments and sedimentary rocks. *J. Sediment. Res.* 68, 448–455.
- Zabel, M., Bickert, T., Dittert, L., Häse, R. R., 1999. Significance of the sedimentary Al:Ti ratio as an indicator for variations in the circulation patterns of the equatorial North Atlantic. *Paleoceanography* 14, 789–799.
- Zitter, H., God, C., 1971. Investigation, evaluation and sources of error of plant analyses. *Fresenius Z. Anal. Chem.* 255, 1.

Summary

Sedimentary deposits are important archives of the Earth's history. In addition, they are of key economical and societal importance because they contain natural resources (e.g., hydrocarbons and drinking water). Hence, it is of utmost importance that we understand the mechanisms controlling the heterogeneity, morphology and spatial distribution of sedimentary deposits. Tools for characterisation of sediment bodies in the subsurface include seismic data, well-log data and core data. Since cores reflect the only continuous physical sample of the rock body under investigation, they provide the input necessary to construct detailed reservoir models. Despite their high information content, financial constraints, time constraints and the desire to keep cores intact limit the amount and resolution of quantitative data that can be retrieved from cores. In an attempt to overcome these problems, spectroscopic core scanners have been developed which facilitate fast, inexpensive, high resolution and non-destructive acquisition of quantitative core data *in situ*. This research is centered around one such method, which is X-ray fluorescence core scanning (XRF-CS).

The first goal of this study is to build a mathematical-statistical framework for the construction of so-called *integrated core descriptions* using XRF-CS. We define an integrated core description as a collection of lithofacies, chemical and petrophysical records on the same, high (1 cm) resolution, and with quantified uncertainties. The interpretation of these descriptions, possibly in combination with other data sets, is referred to as *integrated integrated core analysis*. Developing methods which facilitate integrated core analysis is the second goal of this study. To reach these goals, six objectives have been defined: (1) to **characterise statistically** the relevant data types, (2) to **formalise** interpretation of bulk chemical data, to **improve** (3) and **extend** (4) the existing state-of-the-art calibration framework, (5) to **evaluate** the performance of the new modelling frameworks for different sediment properties and different cores, and (6) to **explore** the added value of XRF-CS in the core analysis framework.

Prior to this study, the state-of-the-art calibration method for XRF-CS was a bivariate method based on logratios (i.e., the BLC method). In an attempt to improve and extend this method, a multivariate alternative (i.e., the MLC method) is proposed which uses Partial Least Squares (PLS) (Chapt. 2). The MLC is an **extension** of the BLC because it facilitates

prediction of "absolute" concentrations. The quality of the MLC method is compared to that of a variety of alternative models, including the BLC method. Results show that (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) method, which is consistent with the compositional nature of the calibration problem but does not fully incorporate absorption and enhancement effects on intensities nor does it permit estimation of "relative" concentrations, is markedly better, and (3) the MLC model 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 **improved** predictive capabilities of the MLC method compared to the other methods are fully exploited by employing automatic sample selection based on the multivariate geometry of intensity measurements in log-ratio space.

In Chapter 3, PLS is used to **formalise** interpretation of geochemical data (Chapter 3). The rationale behind PLS is decomposition of two data sets into unique signals, and signals that are shared among the two data sets. When applied to geochemical composition and grain size, these two types of signals have geological significance: whereas the unique chemical signals are likely to be the response of provenance, the shared signal mainly reflects the conditions under which the sediment was deposited. Applying this methodology to three marine soft-sediment cores yields that employing "textbook" proxies for grain size, such as Al/Ti, can be risky: for the three analysed cores, Ti concentrations once showed no correlation, once positive correlation and once negative correlation with grain size. As for XRF-CS data, chemical proxies therefore require "calibration" for which the PLS-based model provides a framework.

Uncertainty estimates are indispensable for statistically rigorous inference, and for quantification of the predictive performance of XRF-CS. In an attempt to **characterise statistically** spectroscopic and compositional data, Chapters 4 and 5 deal with theoretical and empirical error models associated with these data. We propose that, given their counting or similar statistical nature, all spectroscopic and compositional data are prone to errors caused by counting a finite-sized sample (i.e. counting errors). Given that it removes the effect of scale and transforms the data to a suitable metric space, analysing these count data in terms of centered logratios is a potentially powerful approach. Error propagation shows that in this space, the error-correlation structure of multinomial (e.g., point counts) and Poisson-distributed (e.g., element intensities) data become identical (Chapt. 4). Furthermore, counting errors in clr space are generally not *iid*: in clr-space, the errors are correlated among the variables and the error associated with low counts is higher than that associated with high counts. To make reduced-rank approximations to count data in clr space in a maximum likelihood manner, an algorithm is proposed, which is referred to as Optimal Scale-invariant Reduced-Rank Approximation (OSIRA).

In Chapter 5 we review the structure of widely-used functions for predicting the uncertainty of chemical analyses. The structure of these functions suggests that counting is not the dominant mechanism controlling chemical uncertainty. Moreover, their structure is inconsistent with the definition of concentrations as mass fractions. This inconsistency reflected by the asymmetric nature of these UFs (i.e., $f(c) \neq f(1-c)$ where $c \in [0, 1]$). Derivations using physical-spectroscopic theory yield that for strictly univariate calibration

without matrix corrections, the analytical uncertainty may indeed behave in a fully asymmetrical manner. Ideally, matrix corrections or appropriate multivariate calibration compensate for this flaw, making the concentration estimates indistinguishable from proper concentrations. In practical applications, however, the analytical uncertainty can behave anywhere between fully asymmetric and fully symmetric. Irrespective of its behaviour, however, statistical theory prescribes that UFs used for inference must be symmetric. A new modelling framework is developed to cope with this inconsistency and a fundamentally different definition of the well-known Horwitz function (i.e., the Binomial Horwitz Function (BHF)) and the associated performance criterion 'HorRat' is proposed.

In Chapter 6 we investigate the ability to predict lithofacies (i.e. categorical data) from XRF-CS data. Given that the associated mean prediction error turned out to be 16%, we conclude that XRF-CS can be of great value for automatic lithofacies prediction. Next, we tried to further **improve** the prediction of chemical composition and petrophysical properties. An extension of the modelling framework presented in Chapter 2 is proposed in Chapter 7, which is tested on two cores: one unconsolidated- and one consolidated-sediment core, both comprising a large compositional and sedimentological variability. The quality of the XRF-CS predictions is found to be significantly lower than conventional geochemical and petrophysical analysis. Only for the geochemical composition of relatively homogeneous sediments, the core scanner performs as good as destructive analysis. **Evaluation** of the results yields a set of guidelines for the expected ratio between prediction uncertainty associated with XRF-CS data and the uncertainty of conventional destructive analysis (i.e., the PPR): homogeneous soft-sediment cores ($PPR_{chem} \leq 2$), heterogeneous soft-sediment cores ($1 \leq PPR_{chem} \leq 3$), heterogeneous cores of sedimentary rock ($3 \leq PPR_{chem} \leq 4$).

Chapter 8 **explores** the added value of *integrated* core analysis. We show that incorporating XRF-CS predictions in the workflow can yield different average facies properties which, in turn, can have important consequences for reservoir models. Secondly, by complementing standard petrophysical analysis with XRF-CS data, it is shown how a core may be described in terms of its reservoir quality in a fully automated manner. Thirdly, the use of high-resolution predictions of reservoir quality in combination with petrographical analyses facilitate detailed identification of the control on reservoir quality in a detailed manner: in addition to grain size and dolomite cement, it turns out that also calcite cement controlled reservoir quality. Furthermore, dolomite cement plays a larger role for reservoir quality than suggested by the petrographical analyses. This illustrates the importance of high resolution geochemistry-controlled extrapolation of thin-section analysis.

Future improvements to calibration and processing of XRF-CS data may be achieved by establishing separate calibration models for different lithologies (i.e., stratified calibration). Exploitation of the core images will also contribute to more detailed descriptions. Complementing the XRF data with other data sets such as hyperspectral images, in-situ microscopic images and minipermeametry will further support integrated core analysis. Hyperspectral imaging hopefully leads to a better understanding of the relation between chemical and mineralogical composition, which is important to make the connection with sediment-transport equations, whereas microscopic image analysis potentially facilitates direct estimation of grain size. Another subject of future research is the application of XRF-CS in a project involving numerous cores (e.g., re-evaluation of the structure of a mature gas field).

Samenvatting

Sedimentaire afzettingen bevatten een schat aan informatie over de geschiedenis van de aarde. Daarnaast zijn ze van groot economisch en maatschappelijk belang, omdat ze natuurlijke grondstoffen bevatten zoals olie en drinkwater. Het is daarom belangrijk om te begrijpen welke mechanismen hun voorkomen, heterogeniteit en morfologie bepalen. Beschikbare technieken voor het karakteriseren van sedimentlichamen in de ondergrond zijn seismische surveys, metingen aan boorgaten en de analyse van kernmateriaal. Gezien het feit dat kernen gedetailleerde en directe metingen aan het gesteente mogelijk maken, spelen zij een belangrijke rol bij het integreren en kalibreren van seismische en andere "indirecte" metingen. Ook leveren ze de benodigde input voor gedetailleerde reservoir modellen. Ondanks hun schat aan kwantitatieve informatie, zorgen financiële en tijds-beperkingen, alsmede de wens om kernen intact te houden ervoor dat het niet mogelijk is om kernen optimaal te benutten. Om deze beperkingen het hoofd te bieden zijn er spectroscopische scanners ontwikkeld die het mogelijk maken om snel, goedkoop, op een hoge resolutie én op een non-destructieve manier kwantitatieve kerndata *in situ* te vergaren. Dit onderzoek richt zich op één van deze technieken, te weten X-straling Fluorescentie kernscanning (XRF-CS).

Eén doel van deze studie is om een mathematisch-statistisch raamwerk te bouwen waarmee zogenaamde *geïntegreerde kernbeschrijvingen* kunnen worden gemaakt met behulp van XRF-CS. Wij definiëren een geïntegreerde kernbeschrijving als een verzameling van lithofacies, chemische samenstelling en petrofysische data op dezelfde, hoge (1 cm) resolutie en met bijbehorende onzekerheid. De interpretatie van een geïntegreerde kernbeschrijving, al dan niet gecombineerd met andere data types, definiëren wij als *geïntegreerde kernanalyse*. Ontwikkeling van methodes voor geïntegreerde kernanalyse is het tweede doel van deze studie. Om deze twee doelen te bereiken definiëren we de volgende concrete subdoelen: (1) **statistisch karakteriseren** van relevante data types, (2) **formaliseren** van interpretatie van bulk chemische data, **verbeteren** (3) en **uitbreiden** (4) van de meest recente kalibratie methode, (5) **evalueren** van de prestaties van nieuw ontwikkelde methodes, (6) **exploreren** van de toegevoegde waarde van geïntegreerde kern analyse.

Voordat deze studie begon was de meest geavanceerde kalibratie methode voor XRF-CS een bivariate methode (BLC) gebaseerd op additieve log-ratio's. In een poging om kalibratie van core scanning data te **verbeteren** en **uit te breiden** hebben we een multivariate

methode (MLC) ontwikkeld dat is gebaseerd op gecentreerde log-ratio's en Partial Least Squares (PLS) (Hoofdstuk 2). Deze methode is een uitbreiding op de BLC methode, omdat hiermee voorspelling van "absolute concentraties" mogelijk is. We hebben de prestatie van dit en andere modellen vergeleken. De resultaten laten zien dat de gebruikelijke directe lineaire kalibratie methodes die zijn gestoeld op de aanname dat er een unieke relatie bestaat tussen element intensiteiten en concentraties en daarom dus geen recht doen aan de compositionele aard van deze data, slecht presteren. Aanzienlijk beter presteert het bivariate log-ratio kalibratie model. Dit model is consistent met de compositionele aard van het probleem, maar incorporeert matrix effecten niet optimaal en maakt geen voorspelling van "absolute" concentraties mogelijk. Het voorgestelde multivariate model dat wel rekening houdt met laatstgenoemde zaken presteert het beste. Ook hebben we laten zien dat de voorspellende waarde van het model toeneemt, door gebruik te maken van automatische selectie van kalibratie monsters op basis van element intensiteiten in de logratio ruimte.

We laten in Hoofdstuk 3 zien dat PLS ook worden gebruikt om interpretatie van geochemische data te **formaliseren**. We laten zien dat een op PLS gebaseerd model het mogelijk maakt om de variatie in een data set te partitioneren in unieke signalen, en signalen die de data set gemeenschappelijk heeft met een andere data set. Als we dit toepassen op geochemische en korrelgrootte data, dan hebben deze type signalen een geologische relevantie. De gedeelde geochemische signalen reflecteren de omstandigheden waaronder het sediment is afgezet, terwijl de unieke signalen mogelijk informatie over provenance bevatten. Toepassing van dit model op drie mariene sediment kernen liet zien dat het gebruiken van chemische proxies voor korrelgrootte uit "tekstboeken" risicovol kan zijn: Ti was respectievelijk positief gecorreleerd, negatief gecorreleerd en niet gecorreleerd met korrelgrootte in elk van de drie kernen. Net als XRF-CS data, moeten chemische korrelgrootte proxies dus "gekalibreerd", waarvoor het gepresenteerde model kan worden gebruikt.

Schattingen van de onzekerheid van zowel de kalibratie data als de XRF-CS voorspellingen zijn onontbeerlijk. In een poging om spectroscopische en compositionele data **statistisch te karakteriseren** hebben we in Hoofdstukken 4 en 5 theoretische en empirische onzekerheidsmodellen geanalyseerd en ontwikkeld. Wij stellen dat, aangezien alle spectroscopische en compositionele data tot stand komen door tellen of een soortgelijk proces, zij onderhevig zijn aan zogenaamde telfouten die het resultaat zijn van de eindige grootte van de monsters. Aangezien de gecentreerde logratio (clr) transformatie schaal effecten wegneemt en de data representeert in een metrische ruimte is het analyseren van deze soorten data in de clr ruimte een potentieel krachtige aanpak. Foutenpropagatie liet zien dat in deze ruimte, de covariantiestructuur van de fout geassocieerd met multinomiale en Poisson-gedistribueerde data identiek wordt (Hoofdstuk 4). Een eigenschap van deze universele covariantiestructuur is dat de fouten niet *iid* zijn: ze zijn gecorreleerd en de fout geassocieerd met een klein aantal getelde objecten is relatief groter dan die geassocieerd met een groot aantal getelde objecten. We hebben een nieuw algoritme ontwikkeld (OSIRA) om *reduced-rank approximations* te maken van teldata, met volledige inachtneming van deze eigenschappen.

In Hoofdstuk 5 hebben we de structuur geanalyseerd van bestaande functies die de inter-laboratorium spreiding van chemische analyses voorspellen. Hun structuur suggereert dat de onzekerheid van chemische analyses niet wordt gedomineerd door telfouten. Ook hebben we aangetoond dat hun structuur niet consistent is met de definitie van concentraties als massafractionen. Eén van deze inconsistenties is hun asymmetrische vorm (i.e., $f(c)$)

$\neq f(1-c)$ waarbij $c \in [0, 1]$). Afleiding van deze vorm uit standaard fysisch-spectroscopische theorie leverde op dat strikt univariate kalibratie zonder matrix correcties inderdaad kan resulteren in een asymmetrische onzekerheid. In het ideale geval zorgen matrix correcties of multivariate kalibratie ervoor dat de concentratie schattingen niet te onderscheiden zijn van "echte" concentraties in de strikt statistische zin. In praktische applicaties zal het gedrag van deze onzekerheid variëren tussen perfect symmetrisch en asymmetrisch. Onafhankelijk van dit gedrag schrijft statistische theorie echter voor dat als het doel is om deze voorspelde onzekerheid te gebruiken voor statistische gevolgtrekking, de onzekerheid symmetrisch moet zijn. We hebben een methode voorgesteld om met deze tegenstrijdigheid om te gaan. Dit leidde tot een wijziging in de parametrische vorm van de veelgebruikte Horwitz functie, i.e., de Binomiale Horwitz Functie (BHF).

In Hoofdstuk 6 onderzoeken we of het mogelijk is om lithofacies te voorspellen o.b.v. XRF-CS data. Een methode is ontwikkeld die is toegepast op twee kernen met veel lithologische variatie: een ongeconsolideerde en een geconsolideerde sediment-kern. De gemiddelde fout in de voorspelling van lithofacies is 16%. Op basis hiervan concluderen we dat XRF-CS data van grote waarde kunnen zijn voor het genereren en controleren van de kwaliteit van kernbeschrijvingen. Op basis van onze bevindingen in Hoofdstukken 4 en 5, zijn we tot een **verbetering** en een **uitbreiding** van de MLC methode voor chemische en petrofysische data gekomen (i.e., de GLC methode). De kwaliteit van de kwantitatieve voorspellingen is in de meeste gevallen lager dan die van conventionele technieken. Alleen voor de geochemische compositie van relatief homogene sedimentkernen presteerde de core scanner even goed als conventionele technieken. **Evaluatie** van deze resultaten, gecombineerd met andere resultaten uit deze studie resulteert in een richtlijn voor de verwachte ratio (i.e., de PPR) tussen de fout in XRF-CS voorspelling en gemeten bulk geochemische samenstelling: homogene sediment kernen ($PPR \leq 2$), heterogene sediment kernen ($1 \leq PPR \leq 3$), heterogene sedimentaire gesteente kernen ($3 \leq PPR \leq 4$).

In Hoofdstuk 8 **exploreren** we de toegevoegde waarde van *geïntegreerde kern analyse* aan de hand van drie voorbeelden. Allereerst laten we zien dat het gebruik van hoge-resolutie voorspellingen van petrofysische eigenschappen kan leiden tot significant verschillende facies eigenschappen, wat weer van grote invloed kan zijn op reservoir modellen (8). Vervolgens laten we zien dat door petrofysische analyses te combineren met XRF-CS data het mogelijk is om op een automatische manier een voorspelling van de reservoirkwaliteit te krijgen op de centimeter-schaal. Tenslotte demonstreren we zien dat gebruik van geïntegreerde kernbeschrijvingen het mogelijk maakt om op een meer gedetailleerde manier te analyseren welke mechanismen van invloed zijn geweest op de reservoirkwaliteit. Naast korrelgrootte en dolomietcement, suggereren de resultaten dat ook calcietcement de reservoirkwaliteit beïnvloedt. Daarnaast is de rol van dolomietcement groter dan gesuggereerd door het beperkte aantal petrografische analyses. Dit betekent dat hoge-resolutie geochemische data het mogelijk maken om op een betrouwbare manier bevindingen van petrografische analyses te extrapoleren naar de gehele kern.

Een aanbeveling ten aanzien van toekomstig onderzoek is het opstellen van aparte modellen voor verschillende lithologiën. Waarschijnlijk zal dit bijdragen aan het reduceren van onzekerheid. Daarnaast is het aan te raden om meer informatie te halen uit kernfoto's. In meer algemene zin is het complementeren van XRF-CS data met andere data bronnen een kansrijke vervolgstap: wij denken hierbij aan hyperspectrale metingen, in-situ microscopische foto's en minipermeametrie. De eerste uitbreiding zal hopelijk leiden tot een

betere koppeling tussen chemische en mineralogische kompositie, welke onontbeerlijk is om de link te kunnen leggen met de fysica van sediment transport. Analyse van in-situ microscopische beelden zal continue schattingen van korrelgrootte mogelijk moeten maken, welke nodig zijn voor het toepassen van het PLS model (Hoofdstuk 3). Tenslotte zal moeten worden onderzocht hoe de scanner optimaal gebruikt kan worden in een project dat meerdere kernen behelst (e.g., herevaluatie van de structuur van een ontwikkeld gasveld).

Curriculum Vitae



Menno Bloemsma was born in Almelo on June 16th, 1984. After obtaining his VWO at Erasmus College in Almelo (2002), his academic education started in Leiden with the joint Leiden/Delft BSc program in Life Sciences and Technology. In 2003, he made the transition to Applied Earth Sciences at TU Delft. Subsequently, he started the MSc program in Geomatics at the same university. After an internship with Fugro-Jason, he executed his MSc thesis project under supervision of Gert Jan Weltje, which was aimed at investigating the potential of XRF core scanning, in particular for the hydrocarbon industry (graduation in August 2010). His thesis was nominated for the 'best thesis award' of the department, and he presented his results at the illustrious Working Group on Sediment Generation (WGSG) in Calabria, Italy. Being inspired by this scientific gathering and fascinated by the topic, Menno continued at the TU Delft with the task to develop a software package for analysis of XRF core scanning data. Hereafter, Wintershall Noordzee BV and TNO gave him the opportunity to pursue a PhD on the same topic, which started in January 2011. Menno communicated his results by means of several journal papers and conference abstracts. Another important product of his research is a software package which is far more advanced than the package developed prior to his PhD research, and which is used by several universities. Menno continued his career at Tata Steel as a process and product technologist.

Acknowledgements

Finishing your PhD is by no means a one-man job: many persons have contributed to this thesis, whom I would like to thank hereafter.

I would like to thank my mentor, Gert Jan Weltje, for his endless support. I still remember vividly how we discussed our fascination for latent-variable models during the BSc field work in Vesc which, at the time of writing, is more than six years ago. Who would have thought that this discussion was the onset of a fruitful collaboration, which led to my PhD degree. Although technically you were my supervisor, I sincerely appreciate the fact that our collaboration was based on equality and teamwork. Especially when taking into consideration that this style of supervision can be nerve wrecking sometimes. I am also very grateful for the fact that you have taught me the skill (and art of) scientific writing: a valuable skill which I will employ throughout my career, I'm sure. Thanks for the numerous trips we undertook together, e.g., to Italy, Germany and Austria. I will certainly miss our vivid discussions about a variety of topics, e.g., vectors in compositional space, fall-velocity corrections and the nature of compositional data. And although there currently lies an end product on your desk, I am confident about the fact that this thesis will not be the end of our collaboration.

I also want to thank Rik Tjallingii. I always enjoyed our discussions, the last year mostly by phone, about the link between numerical models and geological objectives. Apart from being pleasant, these discussions taught me to select appropriate starting points for mathematically-oriented papers for a geologist audience. I am also very grateful for the time and effort you put into the various scanning projects. Not all projects made it to this thesis, notwithstanding a paper, which I regret. However, I certainly hope that this thesis is not the end of our collaboration. And if not about XRF scanners, I am sure we will stay in touch. I also want like to thank my promotor, Stefan Luthi. Although you followed my scientific progressions from a distance, I am grateful for your confidence in Gert Jan and myself to successfully execute this project. Also my other committee members are thanked for taking the time to read and assess my dissertation, and come to Delft.

Thirdly, I am deeply grateful for the sponsors and partners of this project. I acknowledge financial support of Wintershall Noordzee BV, and thank Bert de Wijn for the freedom he gave us in determining the direction of research. I also thank TNO, the Geological Survey of the Netherlands, for sponsoring. Of the Survey, I have worked with Michiel van der Meulen, Ronald Vernes, Ronald Harting and Eppie de Heer, whom I thank for their time and efforts. I also want to acknowledge the support I received from the NIOZ: they facilitated

my research to a great extent. The support of Geert-Jan Brummer and Rik Tjallingii, as well as the assistance of Rineke Gieles have been valuable. I also thank Greg van de Bilt, Sergio Fernandez and Mike Burns of Panterra Geoconsultants BV for their support, and for letting me and some of our MSc students use their lab facilities.

At TUD, I felt myself surrounded by helpful and sincere people with creative minds. Rik Donselaar, I am very grateful for the fact that you got me involved in the "anhydrite enigma" project, and I admire the way you and Gert Jan paired up and set up new research projects, some of them with a key role for XRF core scanning. I also want to mention my colleagues of the applied geology department: Adriaan, Andrea, Cees, Geertje, Helena, Ilja, Jiaguang, Koen, Kevin, Liang, Navid, Nico, Pantelis, Remi, Rodolfo, Siddarth, Thais, Wieske, Xiaoxi. You are thanked for the numerous coffee breaks, drinks, discussions and "think and drink" sessions. I also want to thank Rik Noorland for the inspirational discussions about numerical problems, the latest nerdy gadgets, about the beauty of programming. I thank the IAG and the members of the GeoPT steering committee (Peter Webb, Michael Thompson and Phil Potts) for providing the GeoPT data set used in Chapter 5. During my research, I enjoyed working with two MSc students: Idtz Wieling (MSc student of Delft Univ. of Techn.) who investigated core scanning and image analysis and who acquired the core-scanning data, and Martijn van de Boor for his contributions to gamma-ray logging.

Wie ik graag wil noemen zijn mijn oud-huisgenoten, studiegenoten en clubgenoten. Ook al heb ik ze de laatste jaren minder vaak gesproken of gezien dan ik wenste, jullie hebben de mogelijkheid geboden om werk zo af en toe even helemaal te vergeten. PLM, nu ik in de buurt kom wonen moeten we zeker weer eens samen op de wielrenfiets stappen. Kokki en Geertrui, wanneer gaan we die alpiene skills die we hebben opgedaan in Zwitserland in de praktijk brengen? Kwak, je bent een waardevolle vriend en ik waardeer het dat je mijn paranimf wil zijn. En clubje Tool, vergeet niet dat we cool zijn, althans, soms dan...

I want to extend my deepest gratitude to my family. All in all, they were the ones who established the boundary conditions which allowed me to develop myself as an independent researcher, and as a person. And last but certainly not least, I want to thank the love of my life; Annelous, I am very grateful for your unconditional support. This thesis is partly your accomplishment.

Menno Rudolf Bloemsma
Barendrecht, August 2015

List of Publications

9.1 Journal papers

Bloemsma, M.R., Weltje, G.J., submitted. Predicting the uncertainty of routine chemical analyses: a new modelling approach based on evaluation of proficiency tests. *Analyst*.

Bloemsma, M.R., Weltje, G.J., 2015. Reduced-rank approximations to spectroscopic and compositional data: A universal framework based on log-ratios and counting statistics. *Chemometrics and Intelligent Laboratory Systems*, 142, pp. 206-218.

Bloemsma, M.R., Zabel, M., Stuut, J.B., Tjallingii, R., Collins, J., Weltje, G.J., 2012. Modelling the joint variability of grain size and chemical composition in sediments. *Sedimentary Geology*. doi:10.1016/j.sedgeo.2012.04.009.

Weltje, G.J., **Bloemsma, M.R.**, Tjallingii, R., Heslop, D., Röhl, U., Croudace, I.W. (in press). Prediction of geochemical composition from XRF-core-scanner data: A new multivariate approach including automatic selection of calibration samples and quantification of uncertainties. In: Croudace, I.W., Rothwell, G. (Eds.), *Micro-XRF Studies of Sediment Cores*. Developments in Paleoenvironmental Research (DPER) special publication.

Henares, S., **Bloemsma, M.R.**, Donselaar, M.E., Mijnlief, H.F., Redjosentono, A.E., Veldkamp, H.G., Weltje, G.J., 2014. The role of detrital anhydrite in diagenesis of aeolian sandstones (Upper Rotliegend, The Netherlands): Implications for reservoir-quality prediction. *Sedimentary Geology*, 314, 60-74.

Li, J., Luthi, S.M., Donselaar, M.E., Weltje, G.J., Prins, M.A., **Bloemsma, M.R.**, 2014. An ephemeral meandering river system: Sediment dispersal processes in the Río Colorado, Southern Altiplano Plateau, Bolivia. *Zeitschrift für Geomorphologie*.

9.2 Conference proceedings and abstracts

Bloemsma, M.R., Weltje, G.J., 2014. Partial least squares in a compositional framework: a powerful tool for process-based statistical analysis of sediment properties. Oral presentation at the 2nd workshop of the Working Group On Sediment Generation (WGSG). June

25-27, 2014. Geoscience Center of Univ. Göttingen, Germany.

Weltje, G.J., **Bloemsma, M.R.**, 2014. Process-based compositional analysis of clastic sediments: provenance fingerprinting using bulk properties. Keynote at the 2nd workshop of the Working Group On Sediment Generation (WGSG). June 25-27, 2014. Geoscience Center of Univ. Göttingen, Germany.

Barrett, S., Tjallingii, R., **Bloemsma, M.R.**, Brauer, A., Starnberger, R., Spötl, C., Dulski, P., 2015. Compositional classification and sedimentological interpretation of the laminated lacustrine sediments at Baumkriechen (Western Austria) using XRF core scanning data. EGU General Assembly 2015, Vienna, Austria.

Martin-Puertas, C., Tjallingii, R., **Bloemsma, M.R.**, Brauer, A., 2015. Developments in the use of high-resolution X-Ray fluorescence core scanning data of varved sediments for paleoclimate studies: an example of Lake Meerfelder Maar, Germany. EGU General Assembly 2015, Vienna, Austria.

Bloemsma, M.R., Tjallingii, R., 2013. Developments in calibration and data processing of X-Ray fluorescence core scanning data: theory behind, and a demonstration of the AvaaXel-erate software package. Invited lecture centered around the software package developed by MRB in close collaboration with RT.

Bloemsma, M.R., Weltje, G.J., 2012. Advances in core scanning techniques. Invited lecture at the Dutch IODP symposium. Utrecht, The Netherlands.

Bloemsma, M.R., Weltje, G.J., Stuut, J.B.W., Zabel, M., Tjallingii, R., Collins, J.A., 2012. Grain-size independent fingerprinting using Partial Least Squares: a new tool for chemostratigraphic correlation. Proceedings of the 29th IAS Meeting of Sedimentology. Schlading, Austria.

Weltje, G.J., **Bloemsma, M.R.**, Garzanti, E., 2012. Provenance analysis of heavy-mineral assemblages: an old problem revisited. Proceedings of the 29th IAS Meeting of Sedimentology. Schlading, Austria.

Bloemsma, M.R., Weltje, G.J., Stuut, J.B.W., Zabel, M., 2011. Modelling the relation between grain-size distribution and geochemical composition: application to a marine record of mixed fluvial-aeolian provenance (Late Quaternary, offshore Northern Senegal). EGU General Assembly 2011, Vienna, Austria.

Bloemsma, M.R., Weltje, G.J., Stuut, J.B.W., Zabel, M., 2010. Characterisation of sediments by linear combinations of grain-size-dependent source compositions. Oral presentation at the 1st workshop of the Working Group On Sediment Generation (WGSG). June 29th - July 1st, 2010.

9.3 Miscellaneous

Tjallingii, R., **Bloemsma, M.R.**, 2013. *Getting started with AvaaXelerate*. User manual for the Xelerate software package, which was developed as part of this thesis.

Bloemsma, M.R., Weltje, G.J., 2012. *Proefproject core scanner: Het gebruik van XRF scanning binnen een modelgedreven kernanalyse workflow*. Technical report prepared for Geological Survey of the Netherlands (TNO). 123 p.