Spring 5-15-2015

A Hierarchical Bayesian Model for the Unmixing Analysis of Compositional Data subject to Unit-sum Constraints

Shiyong Yu
Department of Mathematics, syu@uno.edu

A Hierarchical Bayesian Model for the Unmixing Analysis of Compositional Data subject to Unit-sum Constraints

A Thesis

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Master of Science
in
Mathematics

by

Shiyong Yu

M.S. Nanjing University, 1997
Ph.D. Lund University, 2003

May, 2015

Acknowledgements

Life appears not simply to be a Bernoulli random variable, which has a binary outcome of either success or failure; rather, it is a voyage full of ups and downs, following a multimodal distribution bounded by birth and death with love in between. I cannot believe that I could have a chance to come back to graduate school ten-plus years after completing my Ph.D., and I had the courage to confront the unforeseen challenges that may pose on my life during the study.

Upon completion of this thesis, first and foremost, I would like to extend my special thanks to my supervisor Prof. Linxiong Li. His friendly sarcasm and gentle encouragement really served as the impetus for me to finish this work. He remained a supporter and provided insights and direction-right up from the beginning to the end of this thesis.

I cannot begin to express my gratitude to my wife Mrs. Minxia Yang for her continued support, both financially and spiritually over the years. Thank you Minxia for being persistent and encouraging, for believing in me, and for the many precious memories along the way. No acknowledgments would be complete without giving thanks to my beloved daughter Kelly Yu. Her beautiful violin music always cheers me up after a day of heavy course work.

Last, but certainly not least, I would like to thank the Chinese Academy of Sciences for providing me a grant (XDA05120401), which enabled me to enter inland Asia to collect lake sediments for testing the model. My deepest gratitude is also due to Dr. E. Dietze, who shared her codes with me so that I can validate the model using her method.

Table of Contents

List of Figures

List of Tables

Abstract

Modeling of compositional data is emerging as an active area in statistics. It is assumed that compositional data represent the convex linear mixing of definite numbers of independent sources usually referred to as end members. A generic problem in practice is to appropriately separate the end members and quantify their fractions from compositional data subject to nonnegative and unit-sum constraints. A number of methods essentially related to polytope expansion have been proposed. However, these deterministic methods have some potential problems.

In this study, a hierarchical Bayesian model was formulated, and the algorithms were coded in MATLAB®. A test run using both a synthetic and real-word dataset yields scientifically sound and mathematically optimal outputs broadly consistent with other non-Bayesian methods. Also, the sensitivity of this model to the choice of different priors and structure of the covariance matrix of error were discussed.

Compositional data, nonnegativity, constant sum, simplex, end member, unmixing, hierarchical Bayesian model, Markov chain, Monte Carlo, birth-death process

Introduction

Compositional data are common in many subjects of social and natural sciences ranging from archaeology to zoology such as the mineral composition of ceramics, household budget composition, and species assemblage of zooplankton in the ocean and so on (Aitchison, 1982). In statistics, compositional data are referred to as the quantitative descriptions of the parts of some whole, usually denoted as an *L*-tuple of nonnegative real numbers whose sum is a constant expressed as per unit (1), percent (100%), per mil (1000‰), parts per million (ppm), and parts per billion (ppb). Figure 1 is a pie chart illustrating the relative abundance of gases making up the Earth's atmosphere. Note that the abundance of all gases was expressed as percentage here. But usually ppm and ppb are used when expressing the concentration of some rare gases such as carbon dioxide (CO_2) and methane (CH_4). This diagram clearly shows that the atmosphere is dominated by nitrogen and oxygen.
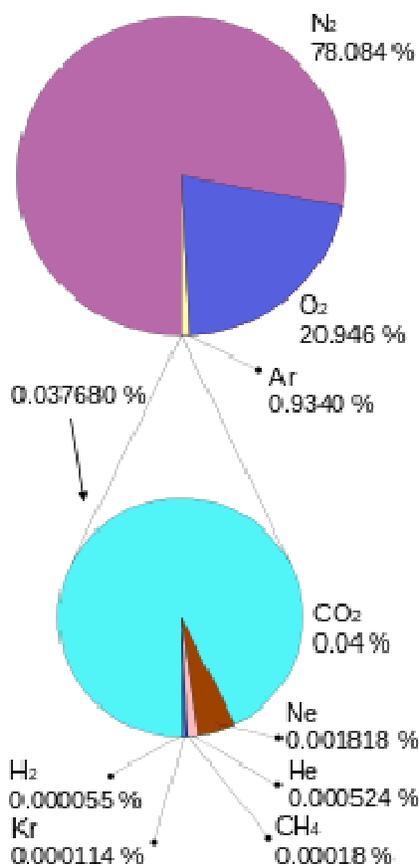
Figure 1. Major constituents of dry air by volume (http://en.wikipedia.org/wiki/Atmosphere_of_Earth)

Ternary plot is another frequently used method to graphically depict the fractions of three variables (more variables also can be grouped into three families based on similar or related attributes) as positions in an equilateral triangle. Figure 2 is such a plot showing the chemical and mineral composition of archaeological ceramics excavated from Shaanxi Province, China. We can see that the raw materials of the ceramics are mainly composed of SiO_2 and Al_2O_3, Fe_2O_3, and CaO, MgO, and SO_3 in total only account for 20% (Figure 2A). The mineralogical composition is dominated by quartz and amorphous phases, and the rest (20%) is made up with feldspars, illite, and calcite (Figure 2B). A comparison reveals that the chemical and mineral compositions of archaeological ceramics are similar to loess sediments, a material used for making fired clay bricks in ancient China some 5000 years ago (Yang et al., 2014).
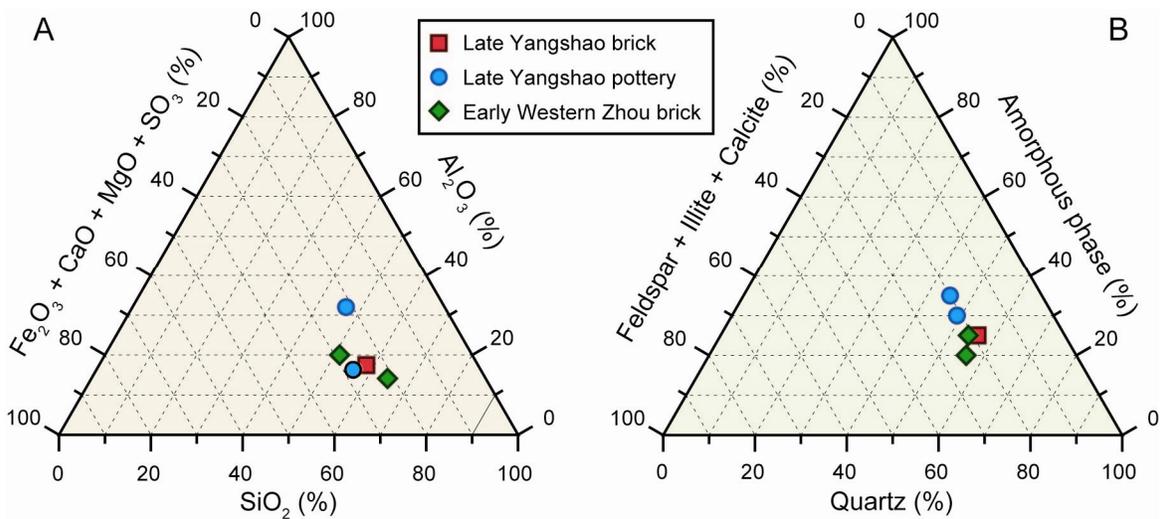
Figure 2. Ternary diagram showing the chemical (A) and mineral (B) composition of ceramics excavated from archaeological sites of late Yangshao and early Western Zhou Dynasty in Shaanxi Province, China

Essentially based on the series plot, area graph is also commonly used to display compositional data, particularly to illustrate the trend of changes of the composition over time. Figure 3 is stacked area graph showing the changes in the assemblage of fossil pollen and spores along a sediment core from Lake Gaotai, a small playa situated on the southern margin of the Badain Jaran Desert in NW China. A total of 36 taxa were identified in the laboratory, and their relative abundance is expressed as percentage, calculated based on the total counts of the taxa in each sample. Here these taxa are grouped as trees, shrubs, upland herbs, wetland herbs, and ferns and algae, and their percentages are potted against depth. As we can see, vegetation around the lake is dominated by shrubs and upland herbs, representing a desert-steppe landscape under a dry climate condition prevailing over recent geological time.
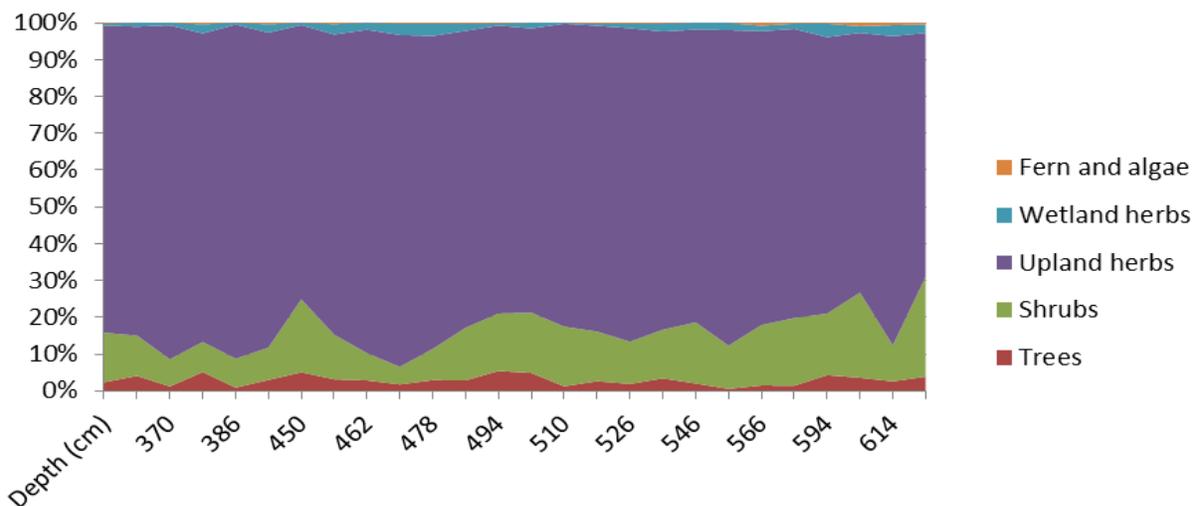
Figure 3. Relative abundance (%) of fossil pollen and spore in sediments of Gaotai Lake, NW China. Shrubs and upland herbs appear to be the dominant component of the local vegetation in this arid area

In some case, frequency distribution diagram is used to display compositional data, particularly when there are too many components. Figure 4 is grain-size distribution of dust trapped in Xinjiang, NW China during three storm events occurring in 2003, 2004, and 2005. The frequency (%) of particles passing through 93 bins having different sizes was determined with a granulometer using laser diffraction technique. This graph reveals a bimodal structure of the grain-size frequency distribution. The mode of 10 μm represents the suspension of particles with upward turbulent winds, while the mode of 100 μm represents the saltation transport by near-ground winds.
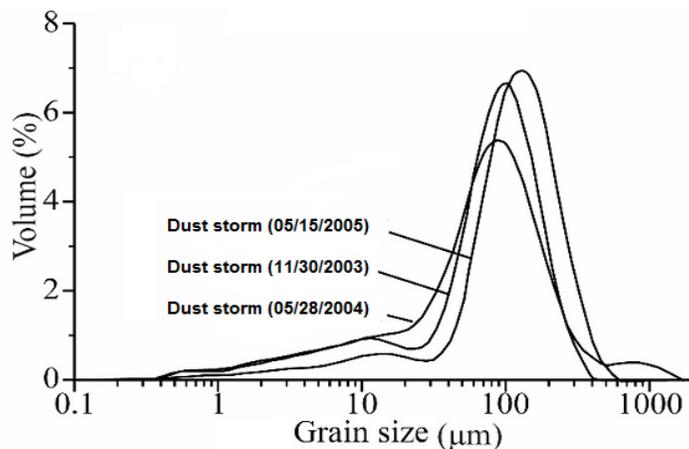
Figure 4. Grain-size distribution of dust trapped in Xinjiang, NW China during three storm events (Data courtesy: Dr. C.-B. An)

3

The above examples only serve as an intuitive exposition of the constrained nature of compositional data commonly known as nonnegativity and constant sum. Let $\mathbf{x} = [x_1, x_2, \ldots, x_L] \in {}^L$ be an L-part compositional data vector where all components are nonnegative real numbers that convey only relative information of each component for a subject, then $\mathbf{x}$ has the following properties (Aitchison, 1986).

(1) The sample space of $\mathbf{x}$ is an $L-1$ simplex defined as

$$\Delta^{L-1} = \left\{ \mathbf{x} = [x_1, x_2, \ldots, x_{L-1}] \in {}^{L-1} : 0 \le x_j \le \kappa, \sum_{j=1}^{L-1} x_j \le \kappa, \text{ and } x_L = \kappa - \sum_{j=1}^{L-1} x_j \right\}, \text{ where } j = 1, \ldots, L, \text{ and } \kappa$$

is a unity (e.g. 1, 100%, or 1000‰). This means that we can always omit one element while analyzing compositional data.

(2) Closure under affine transformation. Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M \in \Delta^{L-1}$. There exist scalars $\lambda_1, \lambda_2, \ldots, \lambda_M \left(0 \le \lambda_i \le 1 \text{ and } \sum_{i=1}^{M} \lambda_i = 1 \right)$ such that $\mathbf{x} = \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \ldots + \lambda_M \mathbf{x}_M \in \Delta^{L-1}$. This is a crucial property for the unmixing analysis of compositional data.

3) Scale invariance. For example, suppose $\mathbf{x}_1 = [0.2,\ 0.3,\ 0.5]$ and $\mathbf{x}_2 = [20\%,\ 30\%,\ 50\%]$, clearly $\mathbf{x}_1$ and $\mathbf{x}_2$ are equivalent.

(4) Permutation invariance. Changing the order of components in the L-tuple would not alter the sum.

(5) Linear dependence. A change of one element in an L-tuple would lead to the changes of the other elements accordingly.

Statistical analysis of compositional data should take into account these properties, particularly when constructing models, designing algorithm, and dealing with the covariance matrix. In this study, the above properties are fully acknowledged as revealed in the following sections.

Chapter 1   Statements of the End-member Unmixing Problem

Analysis and modeling of compositional data are emerging as an active area in statistics. A classical problem in practice is source identification known as the unmixing problem, which is illustrated in Figure 5. For example, a lake receives water from several sources each having a characteristic composition of pollutants. Given an observational dataset of pollutant composition of lake water measured from different sites of the lake, how to identify the sources of pollutants and quantify the relative contribution (fraction) of each source by unmixing the data? Here, we assume that each source has a unique composition. Therefore, they are mutually independent commonly referred to as end members.
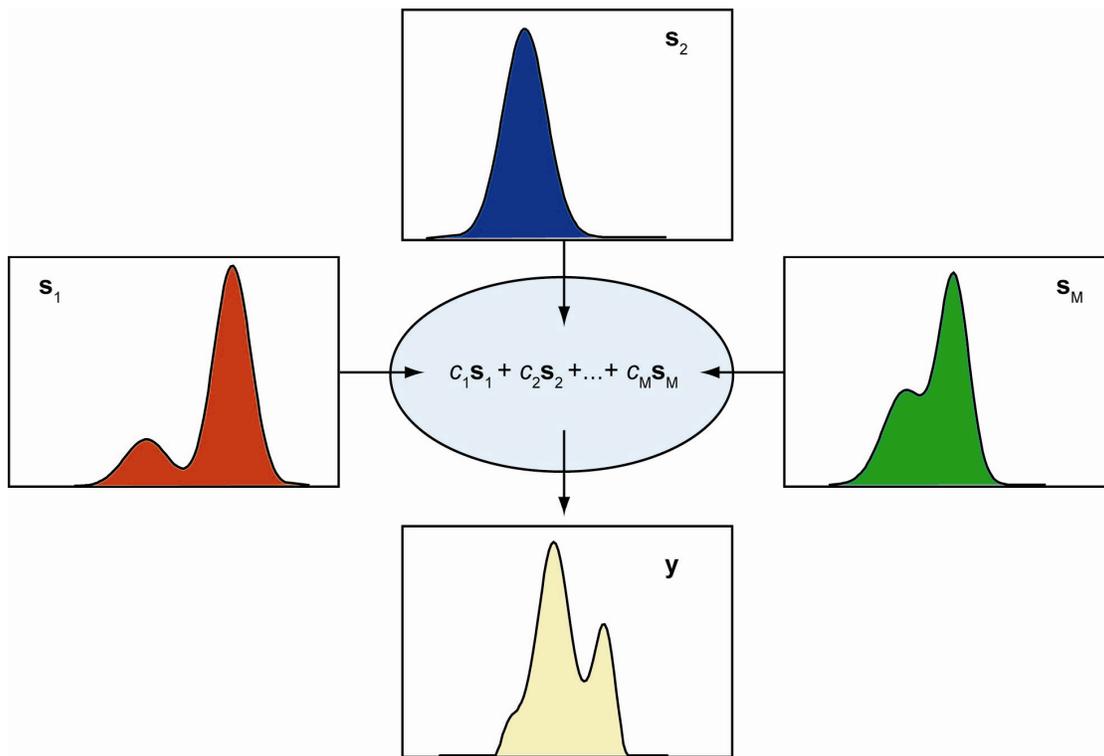
Figure 5. Schematic diagram illustrating the end-member unmixing problem in compositional data analysis. The vectors $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_M$ denote M mutually independent sources involved in the linear mixing during a process, which results in a complicated composition measured as $\mathbf{y}$. The scalars c_1 to c_M denote the relative contribution (weight or fraction) of each source

The following is a full account of this problem. Regarding the notation throughout the text, boldface lowercase letters refer to vectors, and boldface uppercase letters denote matrices. Italics lowercase letters refer to the entries, subscripts indicate the dimension, and superscript T denotes the

transpose operation of a vector or matrix. Unless otherwise stated elsewhere, $\mathbf{0}$ represents the zero vector, $\mathbf{u}$ denotes the unit vector, and $\mathbf{I}$ refers to the identity matrix. The operand $|\cdot|$ and $tr(\cdot)$ denotes the determinant and trace of a matrix, respectively. Scalars are written in italics.

For the univariate case, let $\mathbf{y} = [y_1, y_2, \ldots, y_L]^T \in \mathbb{R}^L$ be an L-part compositional data vector whose sample space is an $L-1$ simplex. Suppose that $\mathbf{y}$ is generated from a linear combination of M independent sources such that

$$\mathbf{y} = \mathbf{Sc} + \boldsymbol{\varepsilon}, \tag{1.1}$$

where $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_M]$ is the source matrix containing M vectors whose attribute is the same as the data vector (i.e. an $L-1$ simplex), $\mathbf{c} = [c_1, c_2, \ldots, c_M]^T$ is the fraction vector whose sample space is an $M-1$ simplex, and $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_L]^T$ is an L-part vector denoting the random errors.

In practice, we usually need to deal with a large dataset. Let's consider the multivariate case. Suppose we have a dataset containing a large number of, say N, subjects (samples) each being an L-part compositional data vector. Let $\mathbf{Y}$ denote this $L \times N$ data matrix

$$\mathbf{Y}_{L \times N} = [\mathbf{y}_1, \ldots, \mathbf{y}_N] = \begin{bmatrix} y_{11} & \cdots & y_{1N} \\ \vdots & \ddots & \vdots \\ y_{L1} & \cdots & y_{LN} \end{bmatrix}, \tag{1.2}$$

$\mathbf{S}$ denote the $L \times M$ source matrix

$$\mathbf{S}_{L \times M} = [\mathbf{s}_1, \ldots, \mathbf{s}_M] = \begin{bmatrix} s_{11} & \cdots & s_{1M} \\ \vdots & \ddots & \vdots \\ s_{L1} & \cdots & s_{LM} \end{bmatrix}, \tag{1.3}$$

$\mathbf{C}$ denote the $M \times N$ fraction matrix

$$\mathbf{C}_{M \times N} = [\mathbf{c}_1, \ldots, \mathbf{c}_N] = \begin{bmatrix} c_{11} & \cdots & c_{1N} \\ \vdots & \ddots & \vdots \\ c_{M1} & \cdots & c_{MN} \end{bmatrix}, \tag{1.4}$$

and $\mathbf{E}$ denote the $L \times N$ error matrix

$$\mathbf{E}_{L \times N} = [\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_N] = \begin{bmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1N} \\ \vdots & \ddots & \vdots \\ \varepsilon_{L1} & \cdots & \varepsilon_{LN} \end{bmatrix}. \tag{1.5}$$

Then we can express the linear mixing problem in terms of matrices

$$\mathbf{Y} = \mathbf{SC} + \mathbf{E}. \tag{1.6}$$

With the assumption of mutual independence, this kind of question has three levels of complexity *per se* (Renner, 1991): (1) both the composition of the source, $\mathbf{S}$, and the number of sources, M, are known. In

this case, we only need to solve for $\mathbf{C}$; (2) only the composition of the sources, $\mathbf{S}$, are known, but the number of sources, M, and the fraction of the sources, $\mathbf{C}$, are unknown; and (3) the composition of the sources, $\mathbf{S}$, the number of sources, M, and the fraction of the sources, $\mathbf{C}$, are all unknown.

The end-member unmixing problem appears to be ill-posed in terms of model identifiability. This implies that the solution is non-unique. Specifically, there always exists a positive definite matrix, $\mathbf{X}$, which can slightly rotate $\mathbf{S}$ and $\mathbf{C}$ but makes $\mathbf{Y}$ be invariant such that $\mathbf{Y} = \mathbf{SX}^{-1}\mathbf{XC} + \mathbf{E}$. However, this issue can be overcome and optimal solutions may be obtained by implementing the full unit-sum constraints through a weighted least-square regression (Weltje, 1997; Weltje and Prins, 2007). Alternatively, this issue can be tackled through a probabilistic approach as proposed in this work.

In this work, the end-member unmixing problem with a complexity of level three will be tackled. Note that in this problem, both $\mathbf{S}$ and $\mathbf{C}$ are subject to the unit-sum constraint, adding another level of complexity. Usually, the source vector, $\mathbf{s}$, is sum to 100%, and the fraction vector, $\mathbf{c}$, is sum to 1. A hierarchical Bayesian model will be presented in the following sections along with a full exposition of the reversible jump Markov chain Monte Carlo algorithm. Non-informative prior distributions for the source and fraction vectors are used in order to real the lack of knowledge about these random variables. By making use of the conjugacy to the multivariate Gaussian likelihood function, an analytical expression of the posterior distribution of the model parameters was derived. The posterior distribution of the number of end members was modeled using the birth-death process, a generalization of Poisson process commonly used to model the changes in the number of components in a system. The composition and fraction of the end members were updated from their posterior distribution using a Gibbs sampler. These procedures were coded in MATLAB® by making use of some built-in functions for generating random numbers and visualizing the results. Codes are given in appendices. The model was validated using a synthetic and a real-world dataset, yielding satisfactory outputs broadly consistent with other non-Bayesian methods.

Chapter 2   Literature Review

2.1 Deterministic Approaches

Compositional data arise naturally in both social and natural sciences, typically in earth science where the relative abundance of chemical elements is usually expressed relative to a total in terms of parts per one, percentage, ppm, ppb, molar concentrations and so on (Buccianti et al., 2006). This constrained nature makes compositional data difficult to handle statistically, because the variables involved in the analysis are restricted in part of the real space (i.e. the simplex). The concept of compositional data can be traced back to the work of Ferrers (1866). Pearson (1896) was the first to realize the complexity of analyzing compositional data, and he pointed out that a spurious correlation may arise when analyzing ratio data whose numerators and denominators contain common parts. Aitchison (Aitchison, 1982, 1994) and several others (Piepel, 1988; Reyment, 1989; Pawlowsky-Glahn and Egozcue, 2006) have formulated the concept of compositional data analysis during recent years, pointing out the pitfalls of traditional statistical methods used to treat compositional data.

A generic problem in compositional data analysis is to extract the underlying independent sources from a large set of observational data. This linear mixing problem is common in audio, radio, and hyperspectral image processing (Wölfel and McDonough). A number of methods essentially about polytopic vector analysis have been proposed heuriscically over the years. Prevailing algorithms include the N-FINDR method (Winter, 1999), independent component analysis (Hyvärinen et al., 2004), and vertex component analysis (Nascimento and Bioucas Dias, 2005). For this kind of problem, the unit-sum constraint for the source vector is not essential. However, in some applications, the components of source vector must sum to unity. For example, it has become known that sediment transport is a sorting process (Visher, 1969; Ashley, 1978; McLaren and Bowles, 1985; Le Roux and Rojas, 2007), which may be approximated as a continuous process of fractionation both mechanically (grains may break down into smaller sizes) and dynamically (only grains with specific combination of sizes, shapes, and densities will be transported together), thereby resulting in characteristic spectra of grain-size distributions that is sum to 100%. Therefore, the above-mentioned methods are not readily applicable to the source vector subject to the unit-sum constraint, and methods are needed ad hoc to address this issue.

Numerical separation of independent sources (end members) subject to the unit-sum constraint in earth science was initially proposed by Klovan (1966), and subsequently improved by Renner (1993) and Weltje (1997). Several other methods such as factor analysis (Klovan and Imbrie, 1971; Clark, 1976; Miesch, 1976; Dietze et al., 2012), linear programming (Banks, 1979; Braun, 1986; Gordon and Dipple, 1999) have also been proposed, but the weighted least-square regression (Renner, 1993; Weltje, 1997)

represents a conceptual innovation in the analysis of compositional data subject to unit-sum constraints. Curve fitting using a prescribed kernel distribution function, for example, log-hyperbolic (Sutherland and Lee, 1994), log-Laplace (Flenley et al., 1987; Fieller et al., 1992; Purkait, 2002; Parker and Bloemendal, 2005), log-normal (Sun et al., 2001; Leys et al., 2005), and Weibull (Kondolf and Adhikari, 2000; Sun et al., 2002) was also widely used. However, this method appears to be misleading, because the frequency distribution of the end members is not necessarily monomodal.

Given that each independent source (end member) represents a vertex of a simplex, an approach through polytope expansion has been proposed (Weltje and Prins, 2003, 2007). The prevailing algorithm based on constrained weighted least-square approximation and eigenvector decomposition collectively known as the end-member modeling analysis (EMMA) originally proposed by Weltje (1997) and improved recently by Dietze et al. (2012) represents a robust and flexible method. However, the major problem of this method is the optimal estimate of the number of end members, although it can be roughly determined by examining the Scree plot (Dietze et al., 2012). Another issue is the method used for data transformation. Usually there are so many leading and trailing zeros for some samples in the dataset. The normalization of the data using an inter-percentile range may cause numerical problems when it happens to be zero.

2.2 Bayesian Approach

Since Green's (1995) innovative work in analyzing compositional data by introducing the reversible jump Markov chain Monte Carlo (MCMC) method (see Chapter 4 for technical details), the end-member unmixing problem of compositional data tends to be casted into a Bayesian paradigm. The Bayesian approach is advantageous over the deterministic approach in several aspects. A significant advantage is its flexibility in dealing with prior information, because the Bayesian approach treats the unknown model parameter as a random variable. As such, the initial belief of the unknown model parameter can be incorporated in the model as prior information, which will be updated and the objective belief (posterior) will be obtained when observational data are applied. Also, the Bayesian approach can efficiently deal with the uncertainty of the model parameter in terms of probability distribution function. Bayesian models have been increasingly used to solve the end-member unmixing problem in the areas of environmental hydrology and geochemistry (Billheimer, 2001; Soulsby et al., 2003; Palmer and Douglas, 2008; Brewer et al., 2011; Tolosana-Delgado et al., 2011; Parnell et al., 2013). However, there are still several issues remain unsolved: (1) How to estimate the number of sources optimally within the context of a Bayesian model; (2) How to appropriately handle the nonnegative and unit-sum constraints of the source vector by choosing a proper prior? (3) How to properly deal with the covariance matrix of the error in terms of linear dependence?

Chapter 3   Theoretical Framework

3.1 Likelihood Function

In this study, we assume that the *L*-part error sequence, $\boldsymbol{\varepsilon}$, is a random vector having a multivariate normal distribution with mean vector $\mathbf{0}_L$ and *L×L* positive definite covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1L} \\ \vdots & \ddots & \vdots \\ \sigma_{L1} & \cdots & \sigma_{LL} \end{bmatrix}. \tag{3.1}$$

As compositional data are linearly dependent, the covariance matrix of error appears not to have a pattered form. Due to lack of knowledge about the actual structure of the covariance matrix of the error, a fully unstructured covariance matrix of error is used in this study. Also, for the sake of simplicity, it is reasonable to assume that the covariance matrix of error for each data vector is the same. By virtue of definition of multivariate normal distribution as well as the above assumptions, the likelihood function for the linear mixing model turns out to be

$$\ell\left(\mathbf{Y}|\mathbf{S},\mathbf{C},M,\boldsymbol{\Sigma}\right) = \left(2\pi\right)^{-\frac{NL}{2}} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{N}{2}} \prod_{i=1}^{N} \exp\left[-\tfrac{1}{2}\left(\mathbf{y}_i - \mathbf{S}\mathbf{c}_i\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{S}\mathbf{c}_i\right)\right]$$

$$= \left(2\pi\right)^{-\frac{NL}{2}} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{N}{2}} \exp\left[-\tfrac{1}{2}\sum_{i=1}^{N}\left(\mathbf{y}_i - \mathbf{S}\mathbf{c}_i\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{S}\mathbf{c}_i\right)\right], \tag{3.2}$$

$$= \left(2\pi\right)^{-\frac{NL}{2}} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{N}{2}} \exp\left\{-\tfrac{1}{2}tr\left[\left(\mathbf{Y} - \mathbf{SC}\right)\left(\mathbf{Y} - \mathbf{SC}\right)^T \boldsymbol{\Sigma}^{-1}\right]\right\}$$

where $\mathbf{Y}$ is the compositional data matrix, $\mathbf{y}$ the compositional data vector, $\mathbf{S}$ the source matrix, $\mathbf{C}$ the fraction matrix, $\mathbf{c}$ the fraction vector, M the number of sources, N the number of compositional data vectors, and L the length of the compositional data vector as defined before.

3.2 Prior Distribution of Model Parameters and Hyperparameters

Given the likelihood function, Bayes' theorem states that the initial beliefs (prior information) of the unknown model parameters can be updated from the observational data. The prior distribution of the unknown model parameters are defined as below.

The discrete uniform distribution supported on $M \in \left[1, \cdots, M_{max}\right]$ was chosen for the prior of the number of end members. The probability mass function is defined as

$$f\left(M\right) = 1/M_{max}, \tag{3.3}$$

where $M_{max} \leq L - 1$ is a positive integer that represents the maximum number of end members possibly involved in the unmixing process.

The symmetric Dirichlet distribution with a concentration parameter being one was assigned to the prior of the source vector. It is equivalent to the uniform distribution supported on a $L-1$ simplex, which is appropriate for modeling the prior of compositional data where knowledge favoring one component over another is lacking (Dobigeon et al., 2008). The probability distribution function is given by

$$f\left(\mathbf{s}|M\right) = \Delta^{L-1}\left(\mathbf{s}\right)$$
$$= \left\{\left[s_1, \cdots, s_{L-1}\right]^T \in ^{L-1} : 0 \le s_j \le \kappa, \ \sum_{j=1}^{L-1} s_j \le \kappa, \ \text{and} \ s_L = \kappa - \sum_{j=1}^{L-1} s_j\right\}, \tag{3.4}$$

where κ is a constant (e.g. 1, 100%, or 1000‰). The assumption of mutual independence of the source vectors yields the prior distribution for the source matrix

$$f\left(\mathbf{S}|M\right) = \prod_{k=1}^{M} f\left(\mathbf{s}_k|M\right). \tag{3.5}$$

Similar to the source vector, the symmetric Dirichlet distribution was chosen for the prior of the fraction vector. The probability distribution function is defined as

$$f\left(\mathbf{c}|M\right) = \Delta^{M-1}\left(\mathbf{c}\right)$$
$$= \left\{\left[c_1, \cdots, c_{M-1}\right]^T \in ^{M-1} : 0 \le c_k \le 1, \ \sum_{k=1}^{M-1} c_k \le 1, \ \text{and} \ c_M = 1 - \sum_{k=1}^{M-1} c_k\right\}. \tag{3.6}$$

Assuming the samples are mutually independent, the prior of fraction matrix $\mathbf{C}$ can be expressed as

$$f\left(\mathbf{C}|M\right) = \prod_{i=1}^{N}\left(\mathbf{c}_i|M\right). \tag{3.7}$$

The inverse Wishart distribution was chosen for the prior of the covariance matrix of the error vector. The probability distribution function of inverse Wishart distribution is written as

$$f\left(\mathbf{\Sigma}|\mathbf{\Psi},\nu\right) = \frac{|\mathbf{\Psi}|^{\frac{\nu}{2}}}{2^{\frac{\nu L}{2}} \Gamma_L\left(\frac{\nu}{2}\right)} \left|\mathbf{\Sigma}^{-1}\right|^{\frac{\nu+L+1}{2}} \exp\left[-\tfrac{1}{2}tr\left(\mathbf{\Psi}\mathbf{\Sigma}^{-1}\right)\right], \tag{3.8}$$

where $\mathbf{\Psi}$ is a $L{\times}L$ positive definite scale matrix, $\nu > L-1$ is the degrees of freedom, and $\Gamma_L\left(\cdot\right)$ is the multivariate gamma function. In practice, ν remains fixed, while $\mathbf{\Psi}$ changes so as to make the prior vague. Here, $\mathbf{\Psi}$ is assumed to follow the Wishart distribution with a probability distribution function defined by

$$f\left(\mathbf{\Psi}|\mathbf{\Phi},\vartheta\right) = \frac{|\mathbf{\Psi}|^{\frac{\vartheta-L-1}{2}}}{2^{\frac{\vartheta L}{2}} \Gamma_L\left(\frac{\vartheta}{2}\right)} \left|\mathbf{\Phi}^{-1}\right|^{\frac{\vartheta}{2}} \exp\left[-\tfrac{1}{2}tr\left(\mathbf{\Phi}^{-1}\mathbf{\Psi}\right)\right], \tag{3.9}$$

where $\mathbf{\Phi}$ is a $L{\times}L$ positive definite matrix and ϑ the degrees of freedom. In practice, a less informative prior can be obtained by setting hyperparameters $\vartheta = L$ and $\mathbf{\Phi} = \vartheta\mathbf{I}_L$.

3.3 Posterior Distribution of Model Parameters and Hyperparameters

According to Bayes' theorem, the joint posterior distribution of the model parameters and hyperparameters can be expressed as a hierarchical structure

$$p(\mathbf{S},\mathbf{C},M,\boldsymbol{\Sigma},\boldsymbol{\Psi}|\mathbf{Y}) \propto \ell(\mathbf{Y}|\mathbf{S},\mathbf{C},M,\boldsymbol{\Sigma})f(\mathbf{S}|M)f(\mathbf{C}|M)f(M)f(\boldsymbol{\Sigma}|\boldsymbol{\Psi},v)f(\boldsymbol{\Psi}|\boldsymbol{\Phi},\vartheta), \quad (3.10)$$

where $\propto$ denotes "proportional to." Integrating out the covariance matrix of error by making use of the definition of the multivariate gamma function lead to a simplified expression of the above structure

$$p(\mathbf{S},\mathbf{C},M,\boldsymbol{\Sigma}|\mathbf{Y}) \propto \int \ell(\mathbf{Y}|\mathbf{S},\mathbf{C},M,\boldsymbol{\Sigma})f(\mathbf{S}|M)f(\mathbf{C}|M)f(M)f(\boldsymbol{\Sigma}|\boldsymbol{\Psi},v)d\boldsymbol{\Sigma}$$

$$= f(\mathbf{S}|M)f(\mathbf{C}|M)f(M)\frac{|\boldsymbol{\Psi}|^{\frac{v}{2}}}{2^{\frac{(v+N)L}{2}}\pi^{\frac{NL}{2}}\Gamma_L\left(\frac{v}{2}\right)}|\boldsymbol{\Sigma}^{-1}|^{\frac{v+N+L+1}{2}}\int \exp\left[-\tfrac{1}{2}tr(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})\right]d\boldsymbol{\Sigma}$$

$$= \frac{f(\mathbf{S}|M)f(\mathbf{C}|M)f(M)|\boldsymbol{\Psi}|^{\frac{v}{2}}}{\pi^{\frac{NL}{2}}\Gamma_L\left(\frac{v}{2}\right)|\boldsymbol{\Lambda}|^{\frac{v+N}{2}}}\int \exp\left[-\tfrac{1}{2}tr(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})\right]|\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}|^{\frac{v+N}{2}-\frac{L+1}{2}}d(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}) \quad , \quad (3.11)$$

$$= f(\mathbf{S}|M)f(\mathbf{C}|M)f(M)\frac{\Gamma_L\left(\frac{v+N}{2}\right)|\boldsymbol{\Psi}|^{\frac{v}{2}}}{\pi^{\frac{NL}{2}}\Gamma_L\left(\frac{v}{2}\right)|\boldsymbol{\Lambda}|^{\frac{v+N}{2}}}$$

where $\boldsymbol{\Lambda} = \left[(\mathbf{Y}-\mathbf{SC})(\mathbf{Y}-\mathbf{SC})^T + \boldsymbol{\Psi}\right]$.

This joint posterior distribution is still too complicated to handle. Here, the Markov chain Monte Carlo (MCMC) method was used to simulate the conditional posterior distribution of individual model parameter, which takes an analytical form in terms of the Gibbs sampler as given below.

Let $\mathbf{S}_{\bar{k}}$ be the source matrix with the kth ($k = 1,\ldots,M$) source vector being removed, and $\mathbf{c}_{\bar{k},i}$ be the fraction vector with the corresponding kth element being removed for the ith ($i = 1,\ldots,N$) sample. According to Bayes' theorem, tedious algebraic operations yield the conditional posterior of the source matrix, which follows a multivariate normal distribution supported on a $L-1$ simplex

$$p(\mathbf{S}|\mathbf{Y},\mathbf{C},M,\boldsymbol{\Sigma}) \propto \ell(\mathbf{Y}|\mathbf{S},\mathbf{C},M,\boldsymbol{\Sigma})f(\mathbf{S}|M)$$

$$= (2\pi)^{-\frac{NL}{2}}|\boldsymbol{\Sigma}^{-1}|^{\frac{N}{2}}\prod_{i=1}^{N}\exp\left[-\tfrac{1}{2}(\mathbf{y}_i-\mathbf{Sc}_i)^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i-\mathbf{Sc}_i)\right]\prod_{k=1}^{M}\Delta^{L-1}(\mathbf{s}_k)$$

$$= (2\pi)^{-\frac{NL}{2}}|\boldsymbol{\Sigma}^{-1}|^{\frac{N}{2}}\prod_{i=1}^{N}\exp\left[-\tfrac{1}{2}\left(\mathbf{y}_i-\mathbf{S}_{\bar{k}}\mathbf{c}_{\bar{k},i}-\mathbf{s}_k c_{k,i}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}_i-\mathbf{S}_{\bar{k}}\mathbf{c}_{\bar{k},i}-\mathbf{s}_k c_{k,i}\right)\right]\prod_{k=1}^{M}\Delta^{L-1}(\mathbf{s}_k)$$

$$= (2\pi)^{-\frac{NL}{2}}|\boldsymbol{\Sigma}^{-1}|^{\frac{N}{2}}\prod_{i=1}^{N}\exp\left\{-\tfrac{1}{2}\left[\mathbf{s}_k c_{k,i}-\left(\mathbf{y}_i-\mathbf{S}_{\bar{k}}\mathbf{c}_{\bar{k},i}\right)\right]^T\boldsymbol{\Sigma}^{-1}\left[\mathbf{s}_k c_{k,i}-\left(\mathbf{y}_i-\mathbf{S}_{\bar{k}}\mathbf{c}_{\bar{k},i}\right)\right]\right\}\prod_{k=1}^{M}\Delta^{L-1}(\mathbf{s}_k) \quad , \quad (3.12)$$

$$= \prod_{k=1}^{M}(2\pi)^{-\frac{NL}{2}}|\boldsymbol{\Sigma}^{-1}|^{\frac{N}{2}}\exp\left\{-\tfrac{1}{2}\sum_{i=1}^{N}\left[\mathbf{s}_k c_{k,i}-\left(\mathbf{y}_i-\mathbf{S}_{\bar{k}}\mathbf{c}_{\bar{k},i}\right)\right]^T\boldsymbol{\Sigma}^{-1}\left[\mathbf{s}_k c_{k,i}-\left(\mathbf{y}_i-\mathbf{S}_{\bar{k}}\mathbf{c}_{\bar{k},i}\right)\right]\right\}\Delta^{L-1}(\mathbf{s}_k)$$

$$= \prod_{k=1}^{M}(2\pi)^{\frac{NL}{2}}|\boldsymbol{\Sigma}^{-1}|^{\frac{N}{2}}\exp\left\{-\tfrac{1}{2}\left[\mathbf{s}_k-\sum_{i=1}^{N}\left(\mathbf{y}_i-\mathbf{S}_{\bar{k}}\mathbf{c}_{\bar{k},i}\right)c_{k,i}\Big/\sum_{i=1}^{N}c_{k,i}^2\right]^T\left(\boldsymbol{\Sigma}\Big/\sum_{i=1}^{N}c_{k,i}^2\right)^{-1}\right.$$

$$\left[\mathbf{s}_k - \sum_{i=1}^{N}\left(\mathbf{y}_i - \mathbf{S}_{\bar{k}}\mathbf{c}_{\bar{k},i}\right)c_{k,i} \Big/ \sum_{i=1}^{N} c_{k,i}^2 \right] \Delta^{L-1}\left(\mathbf{s}_k\right)$$

$$= \prod_{k=1}^{M}\left(2\pi\right)^{\frac{NL}{2}}\left|\mathbf{\Sigma}^{-1}\right|^{\frac{N}{2}}\exp\left\{-\tfrac{1}{2}\left[\mathbf{s}_k - \mathbf{\mu}_k\right]^{T}\mathbf{\Xi}^{-1}\left[\mathbf{s}_k - \mathbf{\mu}_k\right]\right\}\Delta^{L-1}\left(\mathbf{s}_k\right)$$

$$= \prod_{k=1}^{M} N\left(\mathbf{s}_k \mid \mathbf{\mu}_k, \mathbf{\Xi}\right)\Delta^{L-1}\left(\mathbf{s}_k\right)$$

where $\mathbf{\mu}_k = \sum_{i=1}^{N}\left(\mathbf{y}_i - \mathbf{S}_{\bar{k}}\mathbf{c}_{\bar{k},i}\right)c_{k,i} \Big/ \sum_{i=1}^{N} c_{k,i}^2$ is the mean vector and $\mathbf{\Xi} = \mathbf{\Sigma}\Big/\sum_{i=1}^{N} c_{k,i}^2$ is the covariance matrix.

Let $\mathbf{S} = \left[\mathbf{S}_{\bar{M}}, \mathbf{s}_M\right]$ and $\mathbf{c}_i = \left[\mathbf{c}_{\bar{M},i}, 1 - \mathbf{uc}_{\bar{M},i}\right]^{T}$ be a partition of the source matrix, $\mathbf{S}$, and the ith

$(i = 1,\ldots,N)$ fraction vector, $\mathbf{c}$, where $\mathbf{S}_{\bar{M}}$ is the source matrix except the last column, say $\mathbf{s}_M$, $\mathbf{c}_{\bar{M},i}$ is the

fraction vector except the last element, and $\mathbf{u} = \left[1, \cdots, 1\right]$ is a unit vector of size $M-1$. By making use of

Bayes' theorem, elaborative algebraic operations yield the conditional posterior of the fraction matrix,

which follows a multivariate normal distribution supported on a $M-1$ simplex

$$p\left(\mathbf{C} \mid \mathbf{Y}, \mathbf{S}, M, \mathbf{\Sigma}\right) \propto \ell\left(\mathbf{Y} \mid \mathbf{S}, \mathbf{C}, M, \mathbf{\Sigma}\right) f\left(\mathbf{C} \mid M\right)$$

$$= \left(2\pi\right)^{-\frac{NL}{2}}\left|\mathbf{\Sigma}^{-1}\right|^{\frac{N}{2}}\prod_{i=1}^{N}\exp\left[-\tfrac{1}{2}\left(\mathbf{y}_i - \mathbf{Sc}_i\right)^{T}\mathbf{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{Sc}_i\right)\right]\Delta^{M-1}\left(\mathbf{c}_i\right)$$

$$= \frac{\left|\mathbf{\Sigma}^{-1}\right|^{\frac{N}{2}}}{\left(2\pi\right)^{\frac{NL}{2}}}\prod_{i=1}^{N}\exp\left\{-\tfrac{1}{2}\left[\mathbf{y}_i - \mathbf{S}_{\bar{M}}\mathbf{c}_{\bar{M},i} - \mathbf{s}_M\left(1 - \mathbf{uc}_{\bar{M},i}\right)\right]^{T}\mathbf{\Sigma}^{-1}\left[\mathbf{y}_i - \mathbf{S}_{\bar{M}}\mathbf{c}_{\bar{M},i} - \mathbf{s}_M\left(1 - \mathbf{uc}_{\bar{M},i}\right)\right]\right\}\Delta^{M-1}\left(\mathbf{c}_i\right)$$

$$= \frac{\left|\mathbf{\Sigma}^{-1}\right|^{\frac{N}{2}}}{\left(2\pi\right)^{\frac{NL}{2}}}\prod_{i=1}^{N}\exp\left\{-\tfrac{1}{2}\mathbf{c}_{\bar{M},i}^{T}\left[\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)^{T}\mathbf{\Sigma}^{-1}\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)\right]\mathbf{c}_{\bar{M},i} - \mathbf{c}_{\bar{M},i}^{T}\left[\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)^{T}\mathbf{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{s}_M\right)\right] - \right.$$

$$\left.\left[\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)^{T}\mathbf{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{s}_M\right)\right]^{T}\mathbf{c}_{\bar{M},i}\right\}\Delta^{M-1}\left(\mathbf{c}_i\right)$$

$$= \frac{\left|\mathbf{\Sigma}^{-1}\right|^{\frac{N}{2}}}{\left(2\pi\right)^{\frac{NL}{2}}}\prod_{i=1}^{N}\exp\left\langle -\tfrac{1}{2}\left\{\mathbf{c}_{\bar{M},i} - \left[\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)^{T}\mathbf{\Sigma}^{-1}\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)\right]^{-1}\left[\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)^{T}\mathbf{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{s}_M\right)\right]\right\}^{T} \right.$$

$$\left[\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)^{T}\mathbf{\Sigma}^{-1}\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)\right]\left\{\mathbf{c}_{\bar{M},i} - \left[\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)^{T}\mathbf{\Sigma}^{-1}\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)\right]^{-1}\right.$$

$$\left.\left.\left[\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M\mathbf{u}\right)^{T}\mathbf{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{s}_M\right)\right]\right\}\right\rangle\Delta^{M-1}\left(\mathbf{c}_i\right)$$

$$= \frac{\left|\mathbf{\Sigma}^{-1}\right|^{\frac{N}{2}}}{\left(2\pi\right)^{\frac{NL}{2}}}\prod_{i=1}^{N}\exp\left\{-\tfrac{1}{2}\left[\mathbf{c}_{\bar{M},i} - \mathbf{\theta}_i\right]^{T}\mathbf{\Omega}_i^{-1}\left[\mathbf{c}_{\bar{M},i} - \mathbf{\theta}_i\right]\right\}\Delta^{M-1}\left(\mathbf{c}_i\right) \qquad , (3.13)$$

$$= \prod_{i=1}^{N} N\left(\mathbf{c}_{\bar{M},i} \mid \mathbf{\theta}_i, \mathbf{\Omega}_i\right)\Delta^{M-1}\left(\mathbf{c}_i\right)$$

where $\boldsymbol{\theta}_i = \left[\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M \mathbf{u}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M \mathbf{u}\right)\right]^{-1} \left[\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M \mathbf{u}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{y}_i - \mathbf{s}_M\right)\right]$ is the mean vector, and

$\boldsymbol{\Omega}_i = \left[\left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M \mathbf{u}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{S}_{\bar{M}} - \mathbf{s}_M \mathbf{u}\right)\right]^{-1}$ is the covariance matrix.

Applying Bayes' theorem results in the conditional posterior for the covariance matrix of error, $\boldsymbol{\Sigma}$, which follows the inverse Wishart distribution with updated parameters

$$p\left(\boldsymbol{\Sigma}|\mathbf{Y},\mathbf{S},\mathbf{C},M\right) \propto \ell\left(\mathbf{Y}|\mathbf{S},\mathbf{C},M,\boldsymbol{\Sigma}\right) f\left(\boldsymbol{\Sigma}|\boldsymbol{\Psi},v\right)$$

$$= \left(2\pi\right)^{-\frac{NL}{2}} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{N}{2}} \exp\left\{-\tfrac{1}{2}tr\left[\left(\mathbf{Y}-\mathbf{SC}\right)\left(\mathbf{Y}-\mathbf{SC}\right)^T \boldsymbol{\Sigma}^{-1}\right]\right\} \frac{\left|\boldsymbol{\Psi}\right|^{\frac{v}{2}}}{2^{\frac{vL}{2}} \Gamma_L\left(\tfrac{v}{2}\right)} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{v+L+1}{2}} \exp\left[-\tfrac{1}{2}tr\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\right)\right]$$

$$= \frac{\left|\boldsymbol{\Psi}\right|^{\frac{v}{2}}}{2^{\frac{(v+N)L}{2}} \pi^{\frac{NL}{2}} \Gamma_L\left(\tfrac{v}{2}\right)} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{v+N+L+1}{2}} \exp\left\langle-\tfrac{1}{2}tr\left\{\left[\left(\mathbf{Y}-\mathbf{SC}\right)\left(\mathbf{Y}-\mathbf{SC}\right)^T + \boldsymbol{\Psi}\right]\boldsymbol{\Sigma}^{-1}\right\}\right\rangle$$

$$= W^{-1}\left(\left[\mathbf{Y}-\mathbf{SC}\right]\left[\mathbf{Y}-\mathbf{SC}\right]^T + \boldsymbol{\Psi}, v+N\right)$$

(3.14)

A straightforward application of Bayes' theorem yields the conditional posterior of hyperparameter $\boldsymbol{\Psi}$, which follows the Wishart distribution as well but with updated parameters

$$p\left(\boldsymbol{\Psi}|\boldsymbol{\Sigma},\boldsymbol{\Phi},v,\vartheta\right) \propto f\left(\boldsymbol{\Sigma}|\boldsymbol{\Psi},v\right) f\left(\boldsymbol{\Psi}|\boldsymbol{\Phi},\vartheta\right)$$

$$= \frac{\left|\boldsymbol{\Psi}\right|^{\frac{v}{2}}}{2^{\frac{vL}{2}} \Gamma_L\left(\tfrac{v}{2}\right)} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{v+L+1}{2}} \exp\left[-\tfrac{1}{2}tr\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\right)\right] \frac{\left|\boldsymbol{\Psi}\right|^{\frac{\vartheta-L-1}{2}}}{2^{\frac{\vartheta L}{2}} \Gamma_L\left(\tfrac{\vartheta}{2}\right)} \left|\boldsymbol{\Phi}^{-1}\right|^{\frac{\vartheta}{2}} \exp\left[-\tfrac{1}{2}tr\left(\boldsymbol{\Phi}^{-1}\boldsymbol{\Psi}\right)\right]$$

$$= \frac{\left|\boldsymbol{\Psi}\right|^{\frac{v+\vartheta-L-1}{2}}}{2^{\frac{(v+\vartheta)L}{2}} \Gamma_L\left(\tfrac{v}{2}\right)\Gamma_L\left(\tfrac{\vartheta}{2}\right)} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{v+L+1}{2}} \left|\boldsymbol{\Phi}^{-1}\right|^{\frac{\vartheta}{2}} \exp\left\{-\tfrac{1}{2}tr\left[\left(\boldsymbol{\Phi}^{-1} + \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\Psi}\right]\right\}$$

$$= W\left(\boldsymbol{\Phi}^{-1} + \boldsymbol{\Sigma}^{-1}, v+\vartheta\right)$$

(3.15)

Chapter 4 Algorithms

The unknown model parameters were inferred from their conditional posterior using the reversible jump Markov chain Monte Carlo (MCMC) method. For the source and the fraction matrices, a reversible-jump scheme (Green, 1995; Richardson and Green, 1997) in conjunction with a Gibbs sampler (Dobigeon et al., 2009) was employed to propose the move of the Markov chains and to update the state of the chains sequentially, while for the number of sources, a birth-death process was used to browse the parameter space, say $[1, M_{max}]$. The algorithm was outlined in Table 1.

Table 1. Outline of the Metropolis-Hastings-Green-with-Gibbs algorithm

---

Initialization
    Set $\mathbf{S}^{[0]}, \mathbf{C}^{[0]}, M^{[0]}, \mathbf{\Psi}^{[0]}, \mathbf{\Sigma}^{[0]}$;

Iterations and Updates
    for $i = 1$: the number of MCMC runs **do**
        %The Metropolis-Hastings-Green procedure
        Draw m from a discrete uniform distribution supported on $[-1,1]$;
        if $m = 1$, propose a birth move (Table 2);
        else propose a death move (Table 2);
        Calculate the acceptance probability, A;
        Draw ρ from a uniform distribution $U(0,1)$;
        if $\rho < \min(1, A)$, accept the proposed move;
        else stay in current state;

        %The Gibbs procedure
        Update the source matrix, $\mathbf{S}$, according to Eq. 3.12;
        Update the fraction matrix, $\mathbf{C}$, according to Eq. 3.13;
        Update the covariance matrix of error, $\mathbf{\Sigma}$, according to Eq. 3.14;
        Update the hyperparameter matrix, $\mathbf{\Psi}$, according to Eq. 3.15;
    end

Post processing
    Sample the Markov chain of $\mathbf{S}$, $\mathbf{C}$, M, and Σ;
    Calculate the descriptive statistics of $\mathbf{S}$, $\mathbf{C}$, and M;
    Visualize $\mathbf{S}$, $\mathbf{C}$, and M;

Specifically, given an initial guess of the model parameters, a death or birth move is proposed first by drawing a random number from a population of $[-1,\ 1]$ with equal probability (birth is defined by 1, while death is defined by -1), and then a new state of the model parameters is proposed and the acceptance probability is calculated. If the move is accepted, then the model parameters and their hyperparameters will be updated accordingly by sampling their conditional posterior distributions using the Gibbs samplers as given explicitly in Eqs. 12 to 15.

Note that for the birth process, a move when M is equal to one is mandatory, while a move when M is equal to M_{max} is prohibited. Therefore, the transit probability is defined as $b = [1, 0.5, ..., 0.5, 0]$. Conversely, for the death process, a move when M is equal to one is prohibited, while a move when M is equal to M_{max} is mandatory. Therefore, the transit probability turns out to be $d = [0, 0.5, ..., 0.5, 1]$. Each birth or death move would lead to a change in the number of end members by one (Figure 6). Therefore, a rescaling of the fraction vector is required (Eches et al., 2010). The algorithm was given in Table 2.
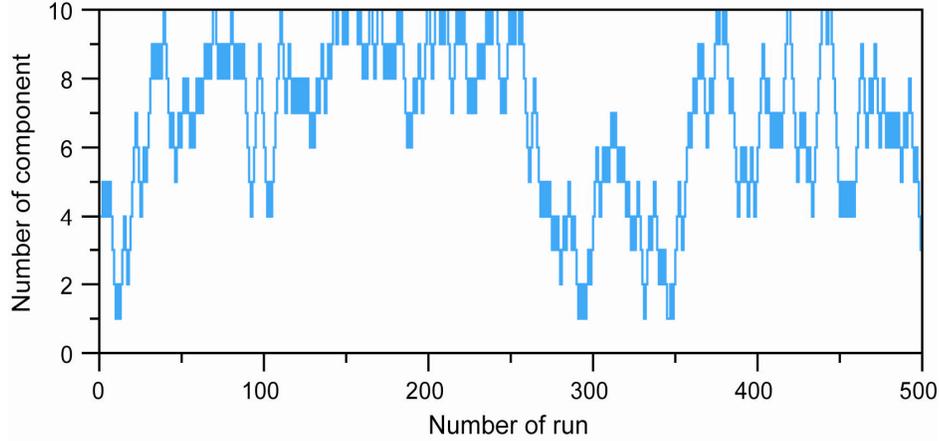
Figure 6. Diagram demonstrating the birth-death process. It models the changes in the number of component for a system with sample space of [1, 2, …, 10]

Table 2. Algorithm for the birth and death move

Birth Move

Set $M^* = M + 1$;

Draw $\mathbf{s}_p$ from the symmetric Dirichlet distribution supported on a $L-1$ simplex;

Add $\mathbf{s}_p$ to the source matrix such that $\mathbf{S}^* = [\mathbf{s}_1, \cdots, \mathbf{s}_M, \mathbf{s}_p]$;

Draw $\boldsymbol{\omega} = [\omega_1, \cdots, \omega_N]$ from the Beta distribution $B(1, M)$;

Add $\boldsymbol{\omega}$ to the fraction matrix $\mathbf{C}$ and rescale it to satisfy the unit-sum constraint such that

$\mathbf{C}^* = [\mathbf{C}diag(1 - \boldsymbol{\omega}) \quad \boldsymbol{\omega}]^T$;

Calculate the acceptance probability of the birth move, A_b, according to Eq. 4.1;
Set the acceptance probability $A = A_b$;

Death Move

Set $M^* = M - 1$;

Draw p from the discrete uniform distribution $U(1, \cdots, M)$;

Set $\boldsymbol{\omega} = [c_{p,1}, \cdots, c_{p,N}]$;

Remove the pth column from the source matrix such that $\mathbf{S}^* = [\mathbf{s}_1, \cdots, \mathbf{s}_{p-1}, \mathbf{s}_{p+1}, \cdots, \mathbf{s}_M]$;

Remove the pth row (i.e. $\boldsymbol{\omega}$) from the fraction matrix and rescale it to satisfy the unit-sum constraint such that $\mathbf{C}^* = \mathbf{C}_{\bar{\omega}}diag\left(\frac{1}{1-\omega}\right)$;

Calculate the acceptance probability of the death move, A_d, according to Eq. 4.2;
Set the acceptance probability $A = A_d$;

16

The acceptance probability for a birth move is

$$A_b = \frac{\ell\left(\mathbf{Y}\big|\mathbf{S}^*,\mathbf{C}^*,M^*,\mathbf{\Sigma}\right)}{\ell\left(\mathbf{Y}\big|\mathbf{S},\mathbf{C},M,\mathbf{\Sigma}\right)}\frac{f\left(\mathbf{S}^*\mid M^*\right)}{f\left(\mathbf{S}\mid M\right)}\frac{f\left(\mathbf{C}^*\mid M^*\right)}{f\left(\mathbf{C}\mid M\right)}\frac{f\left(M^*\right)}{f\left(M\right)}\frac{q\left(M,\mathbf{S},\mathbf{C}\mid M^*,\mathbf{S}^*,\mathbf{C}^*\right)}{q\left(M^*,\mathbf{S}^*,\mathbf{C}^*\mid M,\mathbf{S},\mathbf{C}\right)}|J|$$

$$= \frac{\exp\left\{-\tfrac{1}{2}tr\left[\left(\mathbf{Y}-\mathbf{S}^*\mathbf{C}^*\right)\left(\mathbf{Y}-\mathbf{S}^*\mathbf{C}^*\right)^T\mathbf{\Sigma}^{-1}\right]\right\}}{\exp\left\{-\tfrac{1}{2}tr\left[\left(\mathbf{Y}-\mathbf{S}\mathbf{C}\right)\left(\mathbf{Y}-\mathbf{S}\mathbf{C}\right)^T\mathbf{\Sigma}^{-1}\right]\right\}}\frac{\prod_{k=1}^{M+1}\Gamma_k\left(L\right)}{\prod_{k=1}^{M}\Gamma_k\left(L\right)}\frac{\prod_{i=1}^{N}\Gamma_i\left(M+1\right)}{\prod_{i=1}^{N}\Gamma_i\left(M\right)}\times \qquad , \qquad (4.1)$$

$$\frac{d\left(M+1\right)}{b\left(M\right)}\frac{\left(\frac{1}{M+1}\right)^{N+1}}{\Gamma\left(L\right)\prod_{i=1}^{N}B\left(\omega_i\mid 1,M\right)}\prod_{i=1}^{N}\left(1-\omega_i\right)^{M}$$

and the acceptance probability for a death move is

$$A_d = \frac{\ell\left(\mathbf{Y}\big|\mathbf{S}^*,\mathbf{C}^*,M^*,\mathbf{\Sigma}\right)}{\ell\left(\mathbf{Y}\big|\mathbf{S},\mathbf{C},M,\mathbf{\Sigma}\right)}\frac{f\left(\mathbf{S}^*\mid M^*\right)}{f\left(\mathbf{S}\mid M\right)}\frac{f\left(\mathbf{C}^*\mid M^*\right)}{f\left(\mathbf{C}\mid M\right)}\frac{f\left(M^*\right)}{f\left(M\right)}\frac{q\left(M,\mathbf{S},\mathbf{C}\mid M^*,\mathbf{S}^*,\mathbf{C}^*\right)}{q\left(M^*,\mathbf{S}^*,\mathbf{C}^*\mid M,\mathbf{S},\mathbf{C}\right)}|J|$$

$$= \frac{\exp\left\{-\tfrac{1}{2}tr\left[\left(\mathbf{Y}-\mathbf{S}^*\mathbf{C}^*\right)\left(\mathbf{Y}-\mathbf{S}^*\mathbf{C}^*\right)^T\mathbf{\Sigma}^{-1}\right]\right\}}{\exp\left\{-\tfrac{1}{2}tr\left[\left(\mathbf{Y}-\mathbf{S}\mathbf{C}\right)\left(\mathbf{Y}-\mathbf{S}\mathbf{C}\right)^T\mathbf{\Sigma}^{-1}\right]\right\}}\frac{\prod_{k=1}^{M-1}\Gamma_k\left(L\right)}{\prod_{k=1}^{M}\Gamma_k\left(L\right)}\frac{\prod_{i=1}^{N}\Gamma_i\left(M-1\right)}{\prod_{i=1}^{N}\Gamma_i\left(M\right)}\times \qquad . \qquad (4.2)$$

$$\frac{b\left(M-1\right)}{d\left(M\right)}\frac{\Gamma\left(L\right)\prod_{i=1}^{N}B\left(\omega_i\mid 1,M-1\right)}{\left(\frac{1}{M}\right)^{N+1}}\prod_{i=1}^{N}\left(\frac{1}{1-\omega_i}\right)^{M-1}$$

Chapter 5   Model Validation

5.1 Synthetic Data

The performance of the model was validated using a small-size synthetic dataset, which contains 20 samples each was mixed through a superposition of two independent sources whose composition is known. These independent sources (end members) were generated by sampling a lognormal distribution with a structured covariance matrix of variance components. The components of each source were rescaled so as they are sum to 100%. The fractions (weight) of each source were constructed by sampling a Dirichlet distribution (numbers generated from this distribution are always sum to 1).

The Markov chain of the number of end members was plotted in Figure 7A. After a short burn-in period beginning with an initial value of M equal to four, the chain converges at M equal to two, revealing the presence of two end members in the dataset as known *a priori*. The unmixed end-member spectra were plotted in Figure 7B along with the actual end members used for the synthesis of the dataset for comparison. As we can see that the actual end members can be successfully recovered using this model.
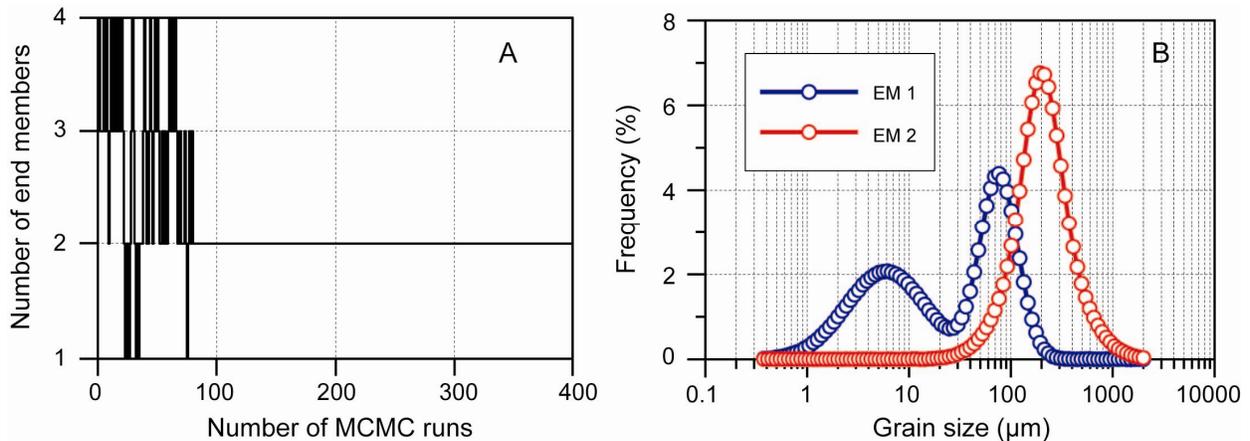
Figure 7. (A) Markov chain of the number of end members (sources) modeled using a birth-death process for the synthetic dataset. The chain moves from an initial state of M equal to four and converges to M equal to two after the burn-in period (ca. 90 iterations); (B) Posterior distribution of the composition of two independent sources identified from the synthetic dataset. Solid lines show the results from model run, and open dotted lines represents the actual composition of two independent sources used to generate the synthetic data

5.2 Real-world Data

A test run of the model was also conducted using a relatively large dataset obtained from Gaotai Lake, a small playa located on the southern margin of the Badain Jaran Desert, NW China (Yu et al.,

2014). Climatic conditions in this area are primarily controlled by the westerlies. The sediment sequence studied here was recovered by digging a trench up to 4 m deep on the north shore of the lake (39°46′42″ N, 99′12′38″ E).
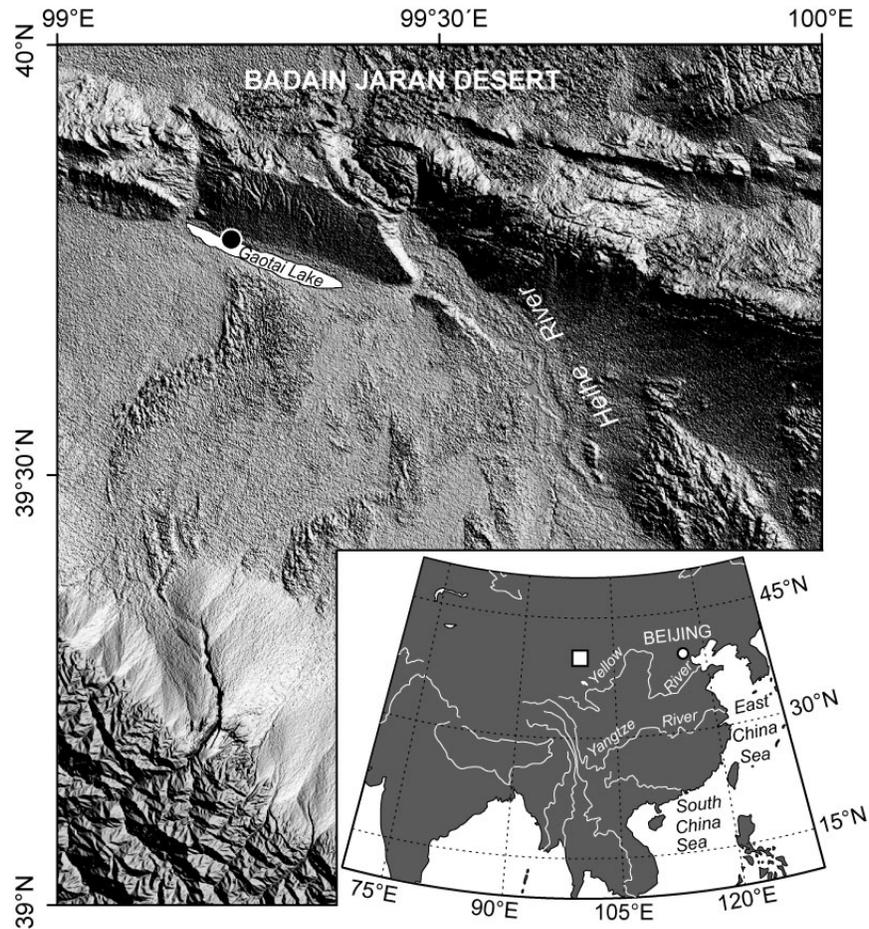
Figure 8. Map showing the location and topographical features of the Gaotai Lake and surrounding area, NW China. Filled circle indicates the location of the stratigraphical section studied here

Four lithological units can be identified (Figure 9). The lowermost unit (Unit I, 285 to 350 cm) is light-brown very fine sand deposited in shallow waters; Unit II (85 to 285 cm) is a thick layer of light-gray clay interbedded with dark-gray silt, indicating an open water environment; Unit III (15 to 85 cm) is dark-brown crossbedded fine sand deposited in a littoral environment; and the uppermost unit (Unit IV, 0 to15 cm) is modern sand dune composed of light-brown fine sand.
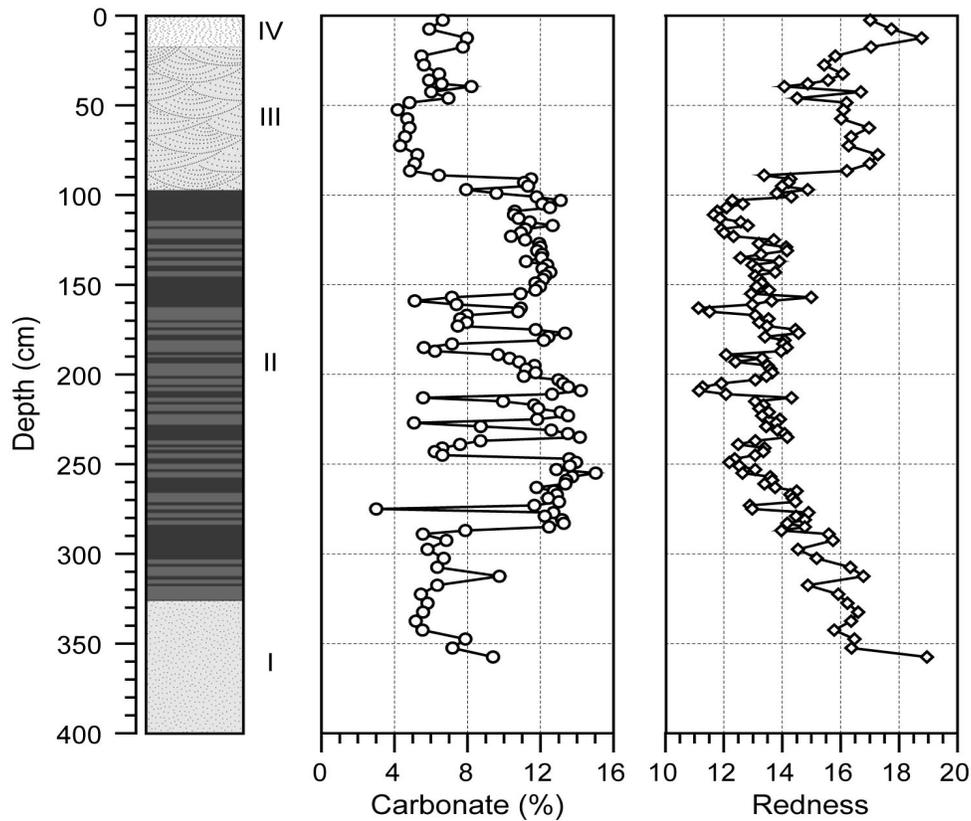
Figure 9. Diagram showing changes in lithology, carbonate content, and redness along with depth in the Gaotai Lake, NW China

A total of 136 samples were collected at 2 to 3 cm intervals along the exposure. The concentration of carbonate was measured using the gas balance method described by Yu (2007), and color analyses were conducted using a MINOLTA CM-508i spectrophotometer. Bulk samples of ca. 1 cm^3 were pretreated with 10% HCl to remove carbonate minerals. After repeated rinsing with deionized water, the samples were further treated with 30% H_2O_2 consecutively for seven days to allow the complete removal of organic matter. Grain-size distributions of the samples were measured using a Malven Mastersizer 2000 laser diffraction grain-size analyzer with 100 bins ranging from 0.02 to 2000 μm. All of the above analyses were conducted in the State Key Laboratory of Loess and Quaternary Geology, CAS. Table 1 shows part of the entire data set, which is too large to display here. The data are summarized by grouping all of the grain sizes into three grades such as clay (very fine, fine, medium, and coarse), silt (very fine, fine, medium, and coarse), and sand (fine, medium, and coarse).

Table 3. Frequency (%) of different grain sizes in sediments of Gaotai Lake, NW China

Sample no.	Clay				Silt				Sand		
	1.46μm	1.95μm	3.91μm	7.81μm	15.63μm	31.25μm	62.50μm	125μm	250μm	500μm	1000μm
1	1.79	2.31	4.57	6.64	6.62	4.25	5.87	25.74	31.65	10.56	0.00
2	1.39	1.83	3.49	4.60	4.83	3.85	7.55	29.30	32.81	10.35	0.00
3	0.83	1.08	2.16	2.81	2.82	2.13	3.91	25.83	39.83	18.42	0.19
4	1.50	1.87	3.91	6.38	8.26	10.02	17.08	26.80	18.79	5.34	0.06
5	1.57	2.15	4.72	8.14	10.67	10.38	12.96	22.78	19.78	6.79	0.07
6	0.86	1.35	2.93	4.44	5.00	3.91	3.84	19.90	36.06	20.84	0.87
7	2.90	4.17	8.03	10.38	10.22	10.55	15.66	21.14	12.62	4.29	0.06
8	2.58	3.41	7.12	10.87	13.45	17.11	20.64	17.06	6.16	1.53	0.06
9	6.91	9.50	18.66	22.67	15.76	9.10	7.97	6.38	2.21	0.70	0.14
10	8.51	10.23	18.62	21.81	14.58	8.85	6.76	3.96	1.69	4.10	0.90
11	1.09	1.54	3.22	4.37	4.47	3.32	5.92	26.23	33.51	15.22	1.10
12	3.24	4.14	8.59	13.02	12.00	7.30	7.61	16.74	16.61	9.94	0.82
13	1.68	2.35	4.89	7.34	8.55	8.71	10.93	19.33	22.03	13.27	0.91
14	1.96	2.69	5.32	7.91	9.14	8.05	9.06	19.30	23.19	12.83	0.54
15	0.88	1.39	3.09	4.67	5.29	5.03	5.16	19.29	35.70	19.29	0.21
16	1.85	2.57	5.32	7.93	9.66	10.83	14.69	22.11	17.57	7.35	0.12
17	0.74	1.17	2.74	4.41	5.92	5.56	8.87	29.06	31.24	10.27	0.01
18	1.19	1.42	3.35	6.68	11.23	13.05	17.74	27.18	16.82	1.34	0.00
19	1.71	2.22	4.69	7.81	11.39	14.58	21.78	25.33	9.58	0.91	0.00
20	2.92	4.55	9.64	13.86	14.46	13.22	17.25	18.81	5.30	0.00	0.00
21	3.85	5.41	13.49	22.00	20.89	13.34	10.60	8.56	1.87	0.00	0.00
22	5.14	7.30	15.40	22.03	22.78	16.89	7.63	2.42	0.42	0.00	0.00
23	6.50	8.11	15.53	22.56	23.46	15.38	6.05	2.05	0.36	0.00	0.00
24	7.49	7.79	14.22	24.35	27.89	14.78	3.04	0.42	0.00	0.00	0.00
25	7.27	8.06	16.12	28.17	26.37	10.48	2.61	0.90	0.02	0.00	0.00

The model begins to run with an initial guess of five end members (Figure 10). After about 200 iterations, the chain of the number of end members tends to converge to three, suggesting that, in this example, essentially three end members can be decomposed. After the burn-in period, the posterior distributions of the grain-size end-member spectra were sampled from the Markov chains (Figure 11).
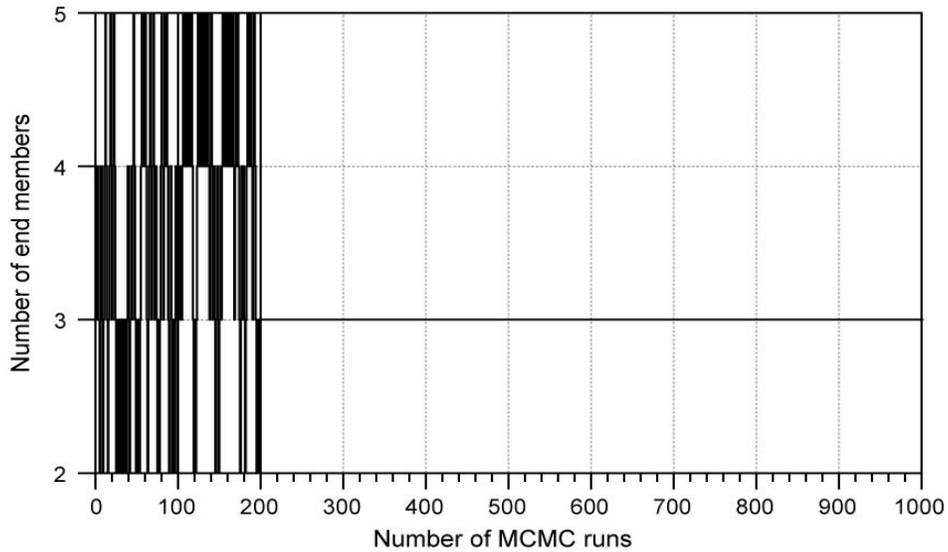
Figure 10. Diagram showing the burn-in period and convergence of the Markov chain of the number of end members of grain-size distributions for Gaotai Lake, NW China
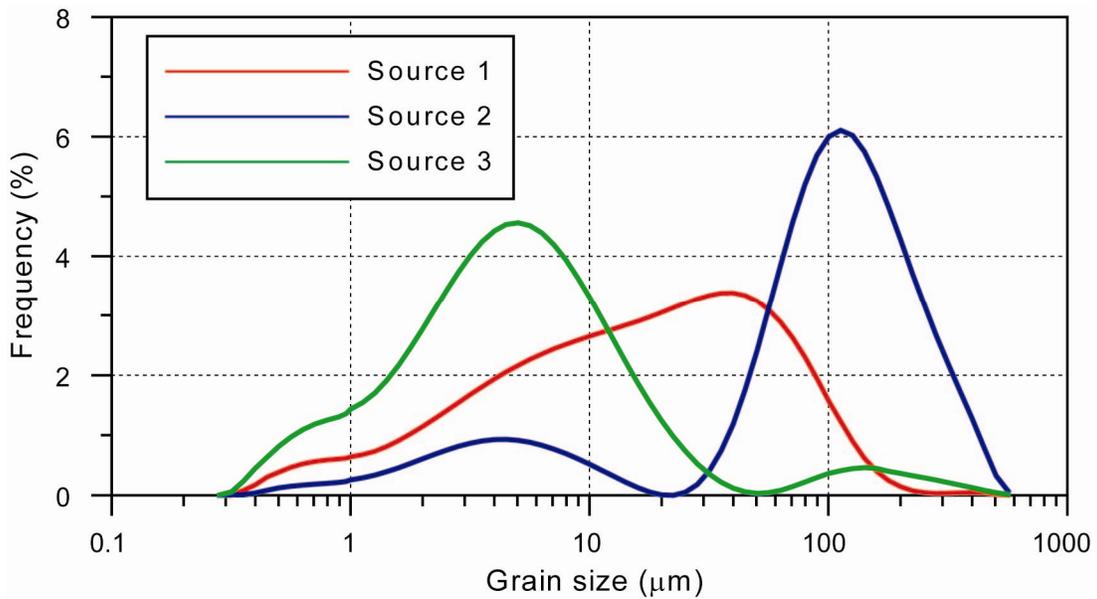
Figure 11. Posterior distribution of the grain-size source (end-member) spectra for Gaotai Lake, NW China obtained using this model

The corresponding grain-size source (end-member) spectra obtained from the weighted least-square regression method (Dietze et al., 2012) were also plotted here for comparison (Figure 12). The results from these two models are broadly consistent. Both models reveal the multimodal and asymmetrical structure of the grain-size end-member spectra. The slight deviation may have resulted from the different scheme of normalization and rescaling employed by the weighted least-square regression method. For example, the Bayesian method does not normalize the data, while the weighted least-square regression method uses the inter-percentile range (i.e. between the 99th and the first percentile) for normalization of the data. Also, rescaling is not needed in this method to satisfy the unit-sum constraint, as the simplex was sampled to simulate the posterior distribution of the source and fraction matrices. However, the weighted least-square regression method imposes a "brutal force" (i.e. rescaling) to the intermediate results in order to meet the unit-sum constraints of the source vector.
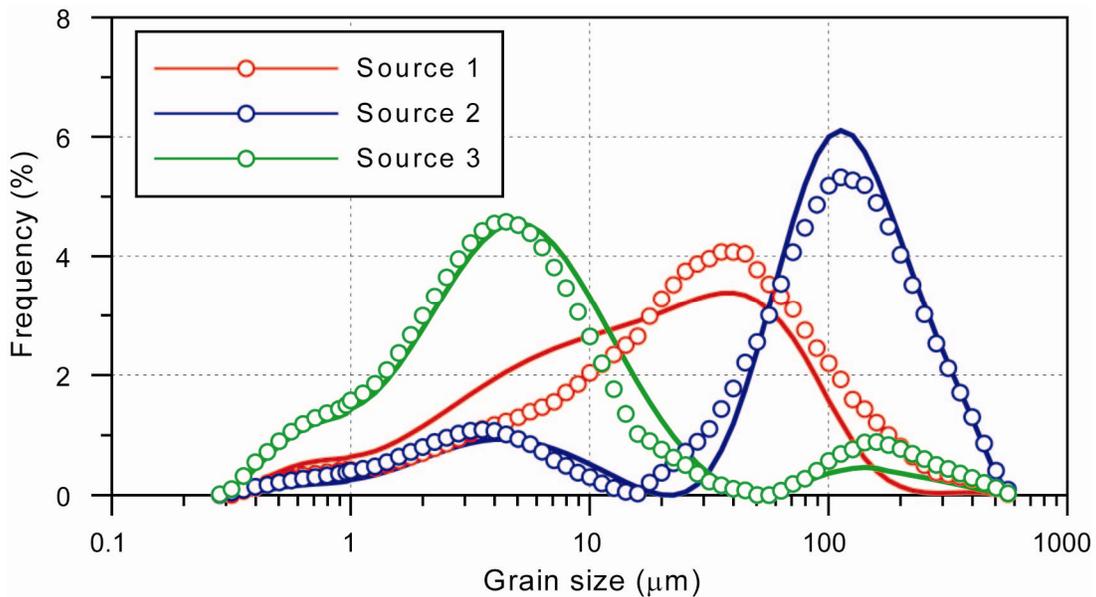
Figure 12. Comparison of the grain-size source (end-member) spectra for Gaotai Lake, NW China obtained using this method (solid lines) with those obtained using the weighted least-square regression method (open dotted lines)

The results are also completed with those obtained from the vertex component analysis (Nascimento and Bioucas Dias, 2005), another deterministic method in the spirit of polytopic vector analysis. Both methods yield a multimodal and asymmetrical structure of the grain-size end-member spectra. Again, the minor difference is due likely to the rescaling employed by the vertex component analysis (Figure 13).

23
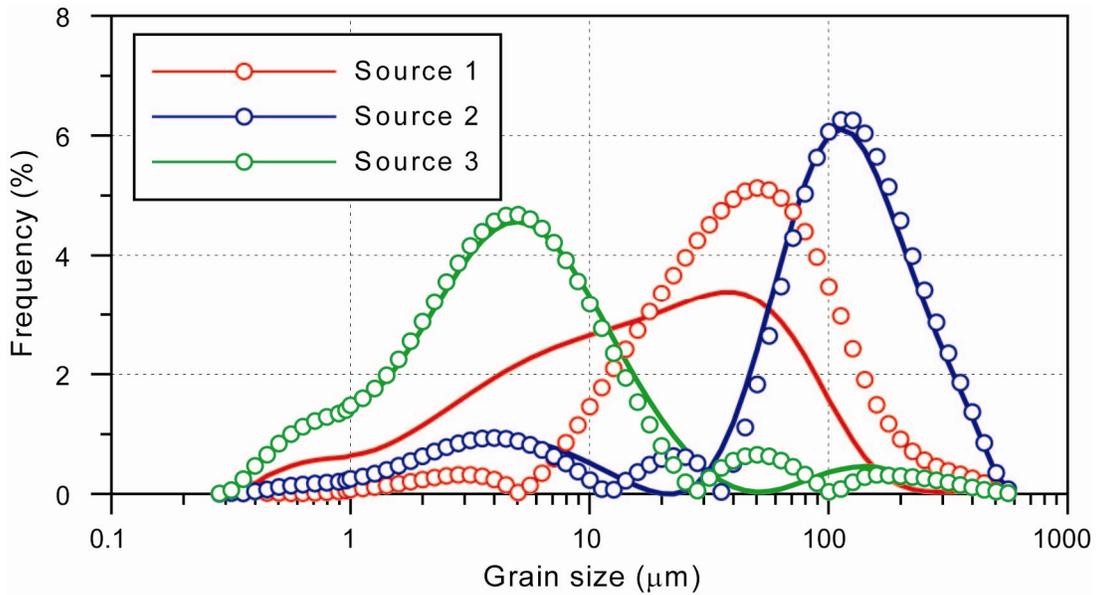
Figure 13. Comparison of the grain-size source (end-member) spectra for Gaotai Lake, NW China obtained using this method (solid lines) with those obtained using the vertex component analysis (open dotted lines)

The grain-size end members identified here represent three distinct transport processes under certain climatic and hydrological conditions. For example, the first end member is characterized by an asymmetrical monomodal structure with a mode at about 40 µm (coarse dust), This grain-size distribution may represent the poorly sorted (reworked dune) sediments, suggesting the near-source transport and thus the short-term suspension and saltation processes by local storms breaking out in the winter and spring when near-ground turbulent airflow prevails (Qin et al., 2005; Qiang et al., 2007). The second end member of grain-size distributions is marked by a bimodal structure with a dominant mode at about 125 µm (very fine sand) and a minor mode at about 4 µm (coarse dust). Given the location of the study site, this end member may represent the suspended sediments from the fluvial/alluvial process or prolonged sediment reshuffling in the littoral zone of the lake while lake-level was lowering. The third end member also exhibits a bimodal structure but with a dominant mode at about 4 µm (coarse dust) and a minor mode at about 150 µm (fine sand). This end member represents the well-sorted sediments, indicating remote dust input associated with the upper-level westerly wind transport and localized sediment trapping dynamics related to the tropospheric turbulence structure (Dietze et al., 2013a).

Changes in the fraction of these end-member spectra with depth were plotted in Figure 14. The first end member dominated the period of higher lake level corresponding to lithological unit II. This period is marked by frequent storm activities, which delivered massive coarse dusts (>50%) to the lake from the surrounding dune fields. The second end member dominated the early and late phases of lower lake level.

It contributed ca. 80% of detrital materials from the neighboring deserts and alluvial fans, indicating enhanced erosion due to seasonal flooding of the Heihe River. The third end member contributed >50% clay to the lake. Superimposed on a weakening trend, the westerlies exhibited a remarkable variability, which exhibits a nearly anti-phase variability with the first one, revealing a dynamic connection of local storminess with the westerlies.
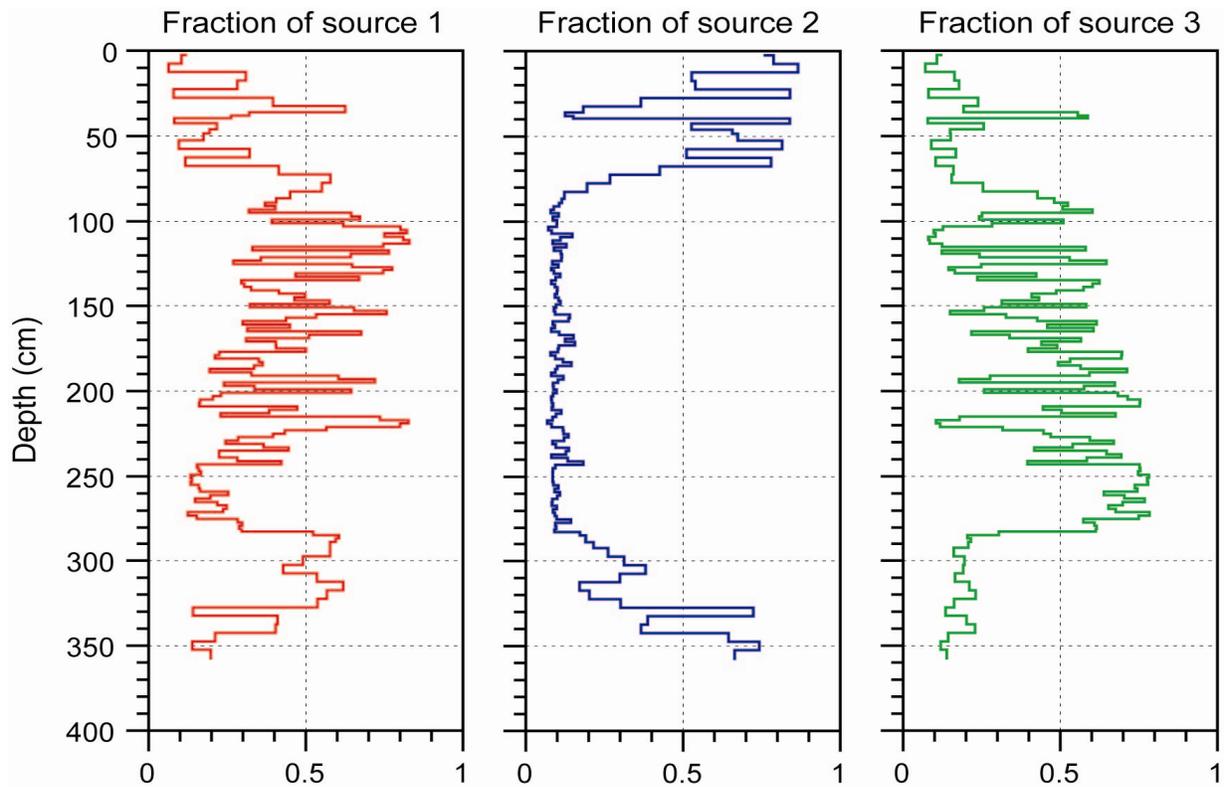
Figure 14. Changes in the fraction of the grain-size end members alone with depth for Gaotai Lake, NW China

Chapter 6   Discussion

6.1 Data Transformation

In order to eliminate the spurious correlations between the components of compositional data, Aitchison (1986) proposed the logratio transformation. Let $y = [y_1, \ldots, y_L]$ be an L-part compositional data vector. The mapping $T : \mathbf{x} \to \mathbf{y}$ is defined as

$$\mathbf{x} = \left[\log\left(\frac{y_1}{y_L}\right), \ldots, \log\left(\frac{y_{L-1}}{y_L}\right) \right].$$ (6.1)

The one-to-one inverse mapping $T^{-1} : \mathbf{y} \to \mathbf{x}$ is given as

$$\mathbf{y} = \left[\frac{e^{x_1}}{1 + \sum_{j=1}^{L-1} e^{x_j}}, \ldots, \frac{e^{x_{L-1}}}{1 + \sum_{j=1}^{L-1} e^{x_j}}, \frac{1}{1 + \sum_{j=1}^{L-1} e^{x_j}} \right].$$ (6.2)

The major advantages of logratio transform are: (1) as the value of variable $\mathbf{x}$ varies in the interval $(-\infty, +\infty)$, the unit-sum constraint was relaxed, and it is easier to choose a function fitting for the model in the real space; and (2) it has been proved that if compositional data vector $\mathbf{y}$ follows the additive logistic normal distribution, then the transformed vector $\mathbf{x}$ will follow the multivariate normal distribution (Aitchison and Bacon-Shone, 1999). However, it was called question by Tauber (1999) when applying this transformation to compositional data with zero components. Actually, numerical problem arises immediately when the last component, say y_L happens to be zero. This situation is common in practice. Padding with tiny positive numbers would violate the unit-sum constraint.

A transformation based on inter-percentile range has also been proposed (Dietze et al., 2012). However, this method also has potential numerical problem. If a compositional dataset contains many leading and trailing zero components, a numerical problem of diving by zero would occur. The method presented here relies on random sampling of the $L-1$ simplex and no transformation is applied on the raw data, thereby circumventing this problem. Nevertheless, the logratio transformation of Aitchison (1986) provides a promising approach to the relaxation of the unit-sum constraint so that one can sample the multivariate normal distribution in the $^{L-1}$ real space. Actually, this method can be improved using the permutation invariance property. A random permutation of the L-tuple can be conducted until the last component is non-zero. Alternatively, one can simply pick any non-zero component in the data vector as a reference, with respect to which the logratio transformation can be performed. In the future, a transformation of the data using this improved method will be considered.

6.2 Choice of Prior Distribution of Model Parameters

It has become known that Bayesian inference is sensitive to the choice of prior (Berger, 1990; Lavine, 1991). In this model, the unit-sum constraint for both the source and fraction vectors was considered as prior information. Therefore, a least-informative prior (i.e. the uniform distribution on a simplex) was chosen for these model parameters. To explore how sensitive is this model to the choice of prior, two cases were compared: uniform vs. multivariate normal distribution for the source vector (Figure 15). The difference appears to be negligible. As no one knows the actual distribution of the source vector other than this constraining information, the uniform prior is preferable. If the data were transformed using the improved method proposed above, the informative logistic-normal distribution (Aitchison and Bacon-Shone, 1999) could be an alternative.
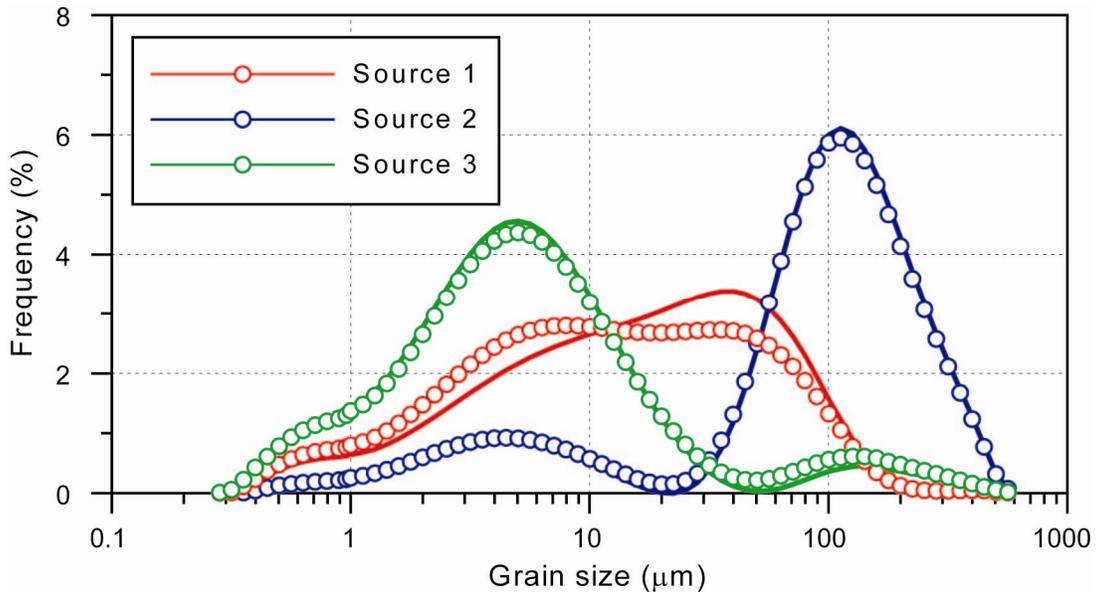
Figure 15. Comparison between models with different prior distribution of the source vector. Solid lines denote the uniform prior and dotted lines indicate the multivariate normal prior on a simplex

6.3 Structure of Covariance Matrix of Error

Compositional data are linearly dependent, which in turn would affect the structure of the covariance matrix of error (Ledoit and Wolf, 2004). In this model, a fully unstructured covariance matrix of error was specified, adding $L(L+1)/2$ to the degrees of freedom of the model. To explore the sensitivity of the model to the structure of the covariance matrix of error, a comparative study was conducted with two

extreme scenarios: unstructured vs. simple structure (i.e. constant multiple of an $\mathbf{I}_{L\times L}$ matrix). The result was presented in Figure 16. The difference is large for the first source, but trivial for the other two sources, implying that this model could be simplified by using a reduced covariance structure (e.g. compound symmetry or Toeplitz).
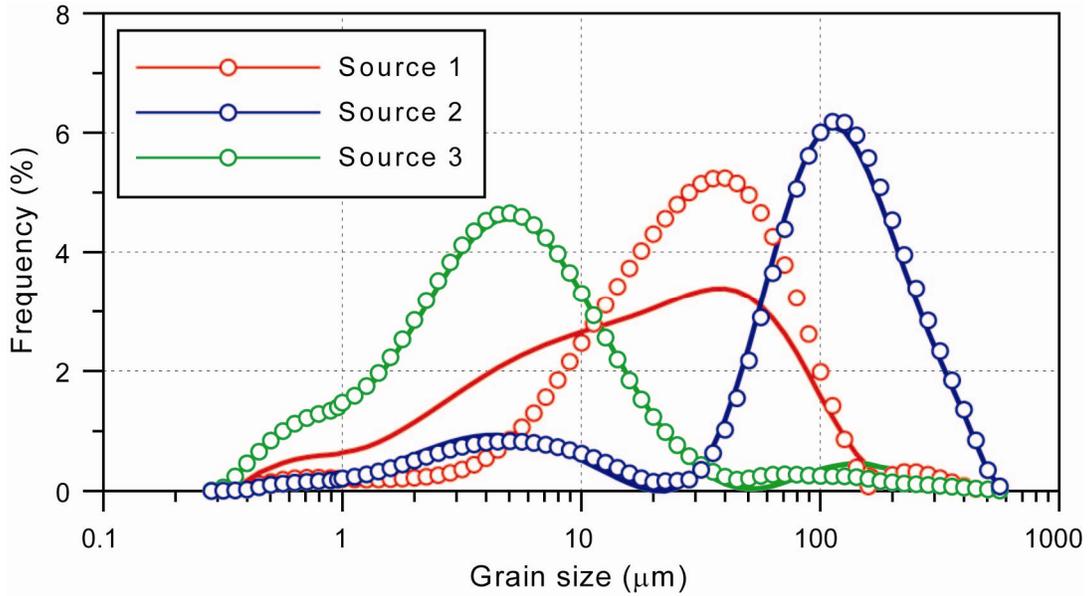
Figure 16. Comparison between models with different structure of the covariance matrix of error. Solid lines denote results from model with an unstructured covariance matrix of error and dotted lines indicate results from model with a simple covariance matrix of error

Chapter 7   Concluding Remarks

It is assumed that compositional data represent the convex linear mixing of definite numbers of independent ("pure") sources commonly referred to as end members, which has a unique spectrum that characterizes the composition of their source. Therefore, the overarching objective of end-member unmixing of compositional data is to separate these sources. Within the framework of Bayesian inference, a hierarchical Bayesian model for end-member unmixing of compositional data was formulated. By making use of the reversible jump MCMC method in conjunction with the Gibbs samplers, this model could not only provide an optimal estimate of the number of end members, but also produce the posterior distribution of the end-member spectra and their fractions completely satisfying the non-negativity and unit-sum constraints. A test run using both a synthetic and real-word dataset yields satisfactory outputs consistent with other non-Bayesian methods.

It is noteworthy that the end-member modeling of compositional data is not mechanistic, as no physical processes are included. This may undermine the scientific soundness of the results. To develop a full inversion towards the informative and reliable inferences of the sources from observational data, physical modeling should be implemented along with end-member modeling for a specific problem. An alternate and probably feasible approach is to build up a library that includes the spectra of all possible sources. As such, the end-member unmixing problem would degenerate to a semi-linear unmixing problem (e.g. complexity of level 2). This is a trade-off between scientific soundness and mathematical feasibility, but it works particularly for the fingerprinting of sources.

References

Aitchison, J., 1982. The statistical analysis of compositional data. Journal of the Royal Statistical Society. Series B - Methodological, 139-177.

Aitchison, J., 1986. The statistical analysis of compositional data.

Aitchison, J., 1994. Principles of compositional data analysis. Lecture Notes-Monograph Series, 73-81.

Aitchison, J., Bacon-Shone, J., 1999. Convex linear combinations of compositions. Biometrika 86, 351-364.

Ashley, G.M., 1978. Interpretation of polymodal sediments. Journal of Geology 86, 411-421.

Banks, R., 1979. Use of linear programming in the analysis of petrological mixing problems. Contributions to Mineralogy and Petrology 70, 237-244.

Berger, J.O., 1990. Robust Bayesian analysis: sensitivity to the prior. Journal of Statistical Planning and Inference 25, 303-328.

Billheimer, D., 2001. Compositional receptor modeling. Environmetrics 12, 451-467.

Braun, G.E., 1986. Quantitative analysis of mineral mixtures using linear programming. Clays and Clay Minerals 34, 330-337.

Brewer, M., Tetzlaff, D., Malcolm, I., Soulsby, C., 2011. Source distribution modelling for end-member mixing in hydrology. Environmetrics 22, 921-932.

Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V., 2006. Compositional data analysis in the geosciences: from theory to practice. Geological Society of London.

Clark, M.W., 1976. Some methods for statistical analysis of multimodal distributions and their application to grain-size data. Journal of the International Association for Mathematical Geology 8, 267-282.

Dietze, E., Hartmann, K., Diekmann, B., IJmker, J., Lehmkuhl, F., Opitz, S., Stauch, G., Wünnemann, B., Borchers, A., 2012. An end-member algorithm for deciphering modern detrital processes from lake sediments of Lake Donggi Cona, NE Tibetan Plateau, China. Sedimentary Geology 243, 169-180.

Dietze, E., Maussion, F., Ahlborn, M., Diekmann, B., Hartmann, K., Henkel, K., Kasper, T., Lockot, G., Opitz, S., Haberzettl, T., 2013a. Sediment transport processes across the Tibetan Plateau inferred from robust grain size end-members in lake sediments. Climate of the Past 9, 4855-4892.

Dobigeon, N., Moussaoui, S., Tourneret, J.-Y., Carteret, C., 2009. Bayesian separation of spectral sources under non-negativity and full additivity constraints. Signal Processing 89, 2657-2669.

Dobigeon, N., Tourneret, J.Y., Chang, C.I., 2008. Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery. IEEE Transactions on Signal Processing 56, 2684-2695.

Eches, O., Dobigeon, N., Tourneret, J.Y., 2010. Estimating the number of endmembers in hyperspectral images using the normal compositional model and a hierarchical Bayesian algorithm. IEEE Journal of Selected Topics in Signal Processing 4, 582-591.

Ferrers, N.M., 1866. An Elementary Treatise on Trilinear Coordinates. Macmillan, London

Fieller, N.R.J., Gilbertson, D.D., Griffin, C.M., Briggs, D.J., Jenkinson, R.D.S., 1992. The statistical modeling of the grain-size distributions of cave sediments using log skew Laplace distributions - Creswell Crags, Near Sheffield, England. Journal of Archaeological Science 19, 129-150.

Flenley, E.C., Fieller, N.R.J., Gilbertson, D.D., 1987. The statistical analysis of mixed grain size distributions from aeolian sands in the Libyan Pre-Desert using log skew Laplace models. Geological Society, London, Special Publications 35, 271-280.

Gordon, T.M., Dipple, G.M., 1999. Measuring mineral abundance in skarn. II. A new linear programming formulation and comparison with projection and Rietveld methods. Canadian Mineralogist 37, 17-26.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711-732.

Hyvärinen, A., Karhunen, J., Oja, E., 2004. Independent Component Analysis. John Wiley & Sons.

Klovan, J., 1966. The use of factor analysis in determining depositional environments from grain-size distributions. Journal of Sedimentary Research 36, 115-125.

Klovan, J., Imbrie, J., 1971. An algorithm and Fortran-iv program for large-scale Q-mode factor analysis and calculation of factor scores. Journal of the International Association for Mathematical Geology 3, 61-77.

Kondolf, G.M., Adhikari, A., 2000. Weibull vs. lognormal distributions for fluvial gravels. Journal of Sedimentary Research 70, 456-460.

Lavine, M., 1991. Sensitivity in Bayesian statistics: the prior and the likelihood. Journal of the American Statistical Association 86, 396-399.

Le Roux, J.P., Rojas, E.M., 2007. Sediment transport patterns determined from grain size parameters: Overview and state of the art. Sedimentary Geology 202, 473-488.

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 88, 365-411.

Leys, J., McTainsh, G., Koen, T., Mooney, B., Strong, C., 2005. Testing a statistical curve-fitting procedure for quantifying sediment populations within multi-modal particle-size distributions. Earth Surface Processes and Landforms 30, 579-590.

McLaren, P., Bowles, D., 1985. The effects of sediment transport on grain-size distributions. Journal of Sedimentary Petrology 55, 457-470.

Miesch, A.T., 1976. Q-mode factor analysis of compositional data. Computers and Geosciences 1, 147-159.

Nascimento, J.M., Bioucas Dias, J.M., 2005. Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 43, 898-910.

Palmer, M.J., Douglas, G.B., 2008. A Bayesian statistical model for end member analysis of sediment geochemistry, incorporating spatial dependences. Journal of the Royal Statistical Society Series C, Applied Statistics 57, 313-327.

Parker, E.J., Bloemendal, J., 2005. Aeolian process and pedogenesis under the influence of the East Asian monsoon: A statistical approach to particle-size distribution variability. Sedimentary Geology 181, 195-206.

Parnell, A.C., Phillips, D.L., Bearhop, S., Semmens, B.X., Ward, E.J., Moore, J.W., Jackson, A.L., Grey, J., Kelly, D.J., Inger, R., 2013. Bayesian stable isotope mixing models. Environmetrics 24, 387-399.

Pawlowsky-Glahn, V., Egozcue, J., 2006. Compositional data and their analysis: an introduction. Geological Society, London, Special Publications 264, 1-10.

Pearson, K., 1896. Mathematical contributions to the theory of evolution - On a form of spurious correlation which may arise when indices are used in the measurement of organs. Proceedings of the Royal Society of London 60, 489-498.

Piepel, G.F., 1988. The statistical analysis of compositional data. Technometrics 30, 120-121.

Purkait, B., 2002. Patterns of grain-size distribution in some point bars of the Usri River, India. Journal of Sedimentary Research 72, 367-375.

Qiang, M., Chen, F., Zhang, J., Zu, R., Jin, M., Zhou, A., Xiao, S., 2007. Grain size in sediments from Lake Sugan: a possible linkage to dust storm events at the northern margin of the Qinghai-Tibetan Plateau. Environmental Geology 51, 1229-1238.

Qin, X.G., Cai, B.G., Liu, T.S., 2005. Loess record of the aerodynamic environment in the east Asia monsoon area since 60,000 years before present. Journal of Geophysical Research-Solid Earth 110, B01204.

Renner, R.M., 1991. An examination of the use of the logratio transformation for the testing of endmember hypotheses. Mathematical Geology 23, 549-563.

Renner, R.M., 1993. The resolution of a compositional data set into mixtures of fixed source compositions. Journal of the Royal Statistical Society. Series C - Applied Statistics 42, 615-631.

Reyment, R.A., 1989. Compositional data analysis. Terra Nova 1, 29-34.

Richardson, S., Green, P.J., 1997. On Bayesian analysis of mixtures with an unknown number of components. Journal of the Royal Statistical Society. Series B - Statistical Methodology 59, 731-758.

Soulsby, C., Petry, J., Brewer, M., Dunn, S., Ott, B., Malcolm, I., 2003. Identifying and assessing uncertainty in hydrological pathways: a novel approach to end member mixing in a Scottish agricultural catchment. Journal of Hydrology 274, 109-128.

Sun, D.H., An, Z.S., Su, R.X., Wu, X.H., Wang, S.M., Sun, Q.L., Rea, D.K., Vandenberghe, J., 2001. Mathematical approach to sedimentary component partitioning of polymodal sediments and its applications. Progress in Natural Science 11, 374-382.

Sun, D.H., Bloemendal, J., Rea, D.K., Vandenberghe, J., Jiang, F.C., An, Z.S., Su, R.X., 2002. Grain-size distribution function of polymodal sediments in hydraulic and aeolian environments, and numerical partitioning of the sedimentary components. Sedimentary Geology 152, 263-277.

Sutherland, R.A., Lee, C.T., 1994. Application of the log-hyperbolic distribution to Hawaiian beach sands. Journal of Coastal Research 10, 251-262.

Tauber, F., 1999. Spurious clusters in granulometric data caused by logratio transformation. Mathematical Geology 31, 491-504.

Tolosana-Delgado, R., von Eynatten, H., Karius, V., 2011. Constructing modal mineralogy from geochemical composition: a geometric-Bayesian approach. Computers & Geosciences 37, 677-691.

Visher, G.S., 1969. Grain size distributions and depositional processes. Journal of Sedimentary Petrology 39, 1074-1106.

Wölfel, M., McDonough, J., Blind source separation. Distant Speech Recognition, 387-408.

Weltje, G.J., 1997. End-member modeling of compositional data: Numerical-statistical algorithms for solving the explicit mixing problem. Mathematical Geology 29, 503-549.

Weltje, G.J., Prins, M.A., 2003. Muddled or mixed? Inferring palaeoclimate from size distributions of deep-sea clastics. Sedimentary Geology 162, 39-62.

Weltje, G.J., Prins, M.A., 2007. Genetically meaningful decomposition of grain-size distributions. Sedimentary Geology 202, 409-424.

Winter, M.E., 1999. N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data, SPIE's International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics, pp. 266-275.

Yang, Y., Yu, S.Y., Zhu, Y., Shao, J., 2014. The making of fired clay bricks in China some 5000 years ago. Archaeometry 56, 220-227.

Yu, F.H., 2007. Situation and comparison of carbonate detection methods. Marine Geology Letters 23, 35-39.

Yu, S.-Y., Cheng, P., Hou, Z., 2014. A caveat on radiocarbon dating of organic-poor bulk lacustrine sediments in arid China. Radiocarbon 56, 127-141.

Appendices

MATLAB® Codes

```
clc;
clear;
load ..//data/GTlake.txt;
Y = GTlake(2:end,2:end);
depth = GTlake(2:end,1);
binsize = GTlake(1,2:end);
nMCMC = 500;
Y = Y';
[NM,NC,NS] = bunmixing(Y,nMCMC);
[mode_M,mean_C,mean_S] = postprocess(NM,NC,NS,Y);
visualization(NM,mode_M,mean_S,mean_C,binsize,depth,Y);
EM_spct = [binsize' mean_S];
EM_abun = [depth mean_C'];
save ..//outputs/GTlake_S.dat EM_spct -ascii -tabs;
save ..//outputs/GTlake_C.dat EM_abun -ascii -tabs;



function [NM,NC,NS] = bunmixing(Y,nMCMC)
%Main program for Bayesian end-member unmixing
%----------------------------------------------------------------------
%INPUT
%       Y          : grain-size observation dataset
%       nMCMC      : number of MCMC runs
%OUTPUT
%       NM         : posterior distribution of the number of end members
%       NC         : Posterior distribution of the abundance matrix
%       NS         : posterior distribution of the end-member matrix
%----------------------------------------------------------------------
[L,N] = size(Y);
Mmax = maxem(Y);
M = randsample((2:Mmax),1,true,ones(1,Mmax-1)/(Mmax-1));
burnin = floor(nMCMC/5);
nv = L + 1;
theta = L;
phi = theta*eye(L);
psi = wishrnd(phi,theta);
sigma = iwishrnd(psi,nv);
C = randfixedsum(M,N,1,0,1);
S = randfixedsum(L,M,100,0,100);
proba_b = [1 1 ones(1,Mmax-3)/2 0];
proba_d = [0 0 ones(1,Mmax-3)/2 1];
NC = zeros(Mmax,N,nMCMC+1);
NS = zeros(L,Mmax,nMCMC+1);
NM = zeros(1,nMCMC+1);
NC(1:M,:,1) = C;
NS(:,1:M,1) = S;
NM(1,1) = M;
rho = log(rand(1,nMCMC));
h = waitbar(0,'Please wait...','CreateCancelBtn','closereq', 'name',...
```

```matlab
    'BAYESIAN END-MEMBER UNMIXING');
for comp = 1:nMCMC
    waitbar(comp/nMCMC,h)
    rhoi = rho(:,comp);
    [C,S,M] = move_bd(C,S,M,sigma,Y,proba_d,proba_b,rhoi,comp,burnin);
    C = update_C(C,S,sigma,Y);
    S = update_S(S,C,sigma,Y);
    sigma = update_sigma(Y,S,C,psi);
    phi = update_psi(sigma,phi);
    NM(1,comp+1) = M;
    NC(1:M,:,comp+1) = C;
    NS(:,1:M,comp+1) = S;
end
close(h)
end


function Max = maxem(Y)
%-------------------------------------------------------------------------
% Determine the maximum number of end members based on cumulative pca load.
%-------------------------------------------------------------------------
[L,N] = size(Y);
Y_bar = mean(Y,2);
Yc = Y - Y_bar*ones(1,N);
covMatrix = Yc*Yc';
 [E, D] = eig(covMatrix);
Ld = diag(D);
Ln = Ld / sum(Ld);
Lv = cumsum(flipud(Ln));
for i=1:length(Lv)
    if Lv(i)>0.995
        Max=i;
        break;
    end
end
end


function [x,v] = randfixedsum(n,m,s,a,b)
%-------------------------------------------------------------------------
% Generate an n*m array x, each of whose m columns contains n random values
% lying in the interval [a,b], but sum to s.
% -------------------------------------------------------------------------
if (m~=round(m))|(n~=round(n))|(m<0)|(n<1)
 error('n must be a whole number and m a non-negative integer.')
elseif (s<n*a)|(s>n*b)|(a>=b)
 error('Inequalities n*a <= s <= n*b and a < b must hold.')
end
s = (s-n*a)/(b-a);
k = max(min(floor(s),n-1),0);
s = max(min(s,k+1),k);
s1 = s - [k:-1:k-n+1];
s2 = [k+n:-1:k+1] - s;
w = zeros(n,n+1); w(1,2) = realmax;
t = zeros(n-1,n);
tiny = 2^(-1074);
```

34

```matlab
for i = 2:n
 tmp1 = w(i-1,2:i+1).*s1(1:i)/i;
 tmp2 = w(i-1,1:i).*s2(n-i+1:n)/i;
 w(i,2:i+1) = tmp1 + tmp2;
 tmp3 = w(i,2:i+1) + tiny;
 tmp4 = (s2(n-i+1:n) > s1(1:i));
 t(i-1,1:i) = (tmp2./tmp3).*tmp4 + (1-tmp1./tmp3).*(~tmp4);
end
v = n^(3/2)*(w(n,k+2)/realmax)*(b-a)^(n-1);
x = zeros(n,m);
if m == 0, return, end
rt = rand(n-1,m);
rs = rand(n-1,m);
s = repmat(s,1,m);
j = repmat(k+1,1,m);
sm = zeros(1,m); pr = ones(1,m);
for i = n-1:-1:1
 e = (rt(n-i,:)<=t(i,j));
 sx = rs(n-i,:).^(1/i);
 sm = sm + (1-sx).*pr.*s/(i+1);
 pr = sx.*pr;
 x(n-i,:) = sm + pr.*e;
 s = s - e; j = j - e;
end
x(n,:) = sm + pr.*s;
rp = rand(n,m);
[ig,p] = sort(rp);
x = (b-a)*x(p+repmat([0:n:n*(m-1)],n,1))+a;
end


function Ae = gen_em(R,varargin)
%-------------------------------------------------------------------------
% Generate the mean of end-member spectra
%-------------------------------------------------------------------------
        verbose = 'on';
        snr_input = 0;
        dim_in_par = length(varargin);
        if (nargin - dim_in_par)~=1
           error('Wrong parameters');
        else if rem(dim_in_par,2) == 1
           error('Optional parameters should always go by pairs');
        else
           for i = 1 : 2 : (dim_in_par-1)
               switch lower(varargin{i})
                 case 'verbose'
                    verbose = varargin{i+1};
                 case 'endmembers'
                    p = varargin{i+1};
                 case 'snr'
                    SNR = varargin{i+1};
                    snr_input = 1;
                 otherwise
                    fprintf(1,'Unrecognized parameter:%s\n', varargin{i});
               end
           end
```

```
            end
        if isempty(R)
            error('there is no data');
        else
            [L N]=size(R);
        end
        if (p<0 || p>L || rem(p,1)~=0),
            error('ENDMEMBER parameter must be integer between 1 and L');
        end
        if snr_input==0,
            r_m = mean(R,2);
            R_m = repmat(r_m,[1 N]);
            R_o = R - R_m;
            [Ud,Sd,Vd] = svds(R_o*R_o'/N,p);
            x_p =  Ud' * R_o;
            SNR = estimate_snr(R,r_m,x_p);
            if strcmp (verbose, 'on')
                    fprintf(1,'SNR estimated = %g[dB]\n',SNR);
            end
        else if strcmp (verbose, 'on')
            fprintf(1,'input    SNR = %g[dB]\t',SNR);
        end
        end
        SNR_th = 15 + 10*log10(p);
        if SNR < SNR_th,
                if strcmp (verbose, 'on'),
                    fprintf(1,'... Select the projective proj.\n',SNR);
                end
                d = p-1;
                if snr_input==0,
                    Ud= Ud(:,1:d);
                else
                    r_m = mean(R,2);
                    R_m = repmat(r_m,[1 N]);
                    R_o = R - R_m;
                    [Ud,Sd,Vd] = svds(R_o*R_o'/N,d);
                x_p =  Ud' * R_o;
                end
                Rp =  Ud * x_p(1:d,:) + repmat(r_m,[1 N]);
                x = x_p(1:d,:);
                c = max(sum(x.^2,1))^0.5;
                y = [x ; c*ones(1,N)] ;
        else
                if strcmp (verbose, 'on'),
                    fprintf(1,'... Select proj. to p-1\n',SNR);
                end
                d = p;
                [Ud,Sd,Vd] = svds(R*R'/N,d
                x_p = Ud'*R;
                Rp =  Ud * x_p(1:d,:);
                x =  Ud' * R;
                u = mean(x,2);
                y =  x./ repmat( sum( x .* repmat(u,[1 N]) ) ,[d 1]);
        end
indice = zeros(1,p);
A = zeros(p,p);
```

```matlab
A(p,1) = 1;
for i=1:p
    w = rand(p,1);
    f = w - A*pinv(A)*w;
    f = f / sqrt(sum(f.^2));
    v = f'*y;
    [v_max indice(i)] = max(abs(v));
    A(:,i) = y(:,indice(i));
end
Ae = Rp(:,indice);
Ae = abs(Ae);
Ae = 100*Ae./(ones(size(Ae,1),1)*sum(Ae,1));
return;
%%
function snr_est = estimate_snr(R,r_m,x)
        [L N]=size(R);
        [p N]=size(x);
        P_y = sum(R(:).^2)/N;
        P_x = sum(x(:).^2)/N + r_m'*r_m;
        snr_est = 10*log10( (P_x - p/L*P_y)/(P_y- P_x) );
return;



function [C_out,S_out,M_out]=
move_bd(C,S,M,sigma,Y,proba_d,proba_b,rhoi,comp,burnin)
%--------------------------------------------------------------------------
% Propose a birth/death move
%--------------------------------------------------------------------------
[L N] = size(Y);
move = randsample([-1 1],1,true,[proba_d(M) proba_b(M)]);
M_star = M + move;
S_star = gen_em(Y,'Endmembers',M_star,'verbose','off');
if move == 1
    w = betarnd(1,M,1,N);
    C_star = [C*diag(1-w); w];
else
    indm = unidrnd(M);
    C_new = C;
    w = C_new(indm,:);
    C_new(indm,:) = [];
    C_star = C_new*diag(1./(1-w));
end
accept=proba_bd(S_star,S,C_star,C,M_star,M,sigma,Y,proba_d,proba_b,w,rhoi,comp,burnin);
if accept == 1
    M_out = M_star;
    C_out = C_star;
    S_out = S_star;
else
    M_out = M;
    C_out = C;
    S_out = S;
end
end
```

```matlab
function accept =
proba_bd(S_star,S,C_star,C,M_star,M,sigma,Y,proba_d,proba_b,w,rhoi,comp,burni
n)
%-------------------------------------------------------------------------
%Compute the acceptance ratio of birth/death move
%-------------------------------------------------------------------------
[L,N] = size(Y);
a1 = (Y-S_star*C_star)*(Y-S_star*C_star)'*inv(sigma);
a2 = (Y-S*C)*(Y-S*C)'*inv(sigma);
A = 0.5*(trace(a2)-trace(a1));
if M_star-M == 1
    A = A + gammaln(L);
else
    A = A - gammaln(L);
end
if M_star-M == 1
    A = A + N*(gammaln(M+1)-gammaln(M));
else
    A = A - N*(gammaln(M)-gammaln(M-1));
end
if M_star-M == 1
    A = A + log(proba_d(M+1)) - log(proba_b(M)) ;
else
    A = A + log(proba_b(M-1)) - log(proba_d(M)) ;
end
if M_star-M == 1
    A = A - sum(log(betapdf(w,1,M)),2) - (N+1)*log(M+1);
else
    A = A + sum(log(betapdf(w,1,M-1)),2) + (N+1)*log(M);
end
if M_star-M == 1
    A = A + sum(log((1-w).^M),2);
else
    A = A - sum(log((1-w).^(M-1)),2);
end
if comp < burnin
    accept = rhoi < max([0,A]);
else
    accept = rhoi < min([0,A]);
end
end


function S_out = update_S(S,C,sigma,Y)
%-------------------------------------------------------------------------
% Sample the posterior distribution of the end-member matrix
% INPUT
%   S : current state of end-member matrix
%   C : current state of fraction matrix
%   sigma : covariance matrix of error
%   Y : observational data
% OUTPUT
%   S_out : updated end-member matrix
%-------------------------------------------------------------------------
[L,N] = size(Y);
M = size(S,2);
S_out = S;
```

```matlab
[matP,matU,Y_bar] = whitenmat(Y,1,M-1);
T_out = matP*(S - Y_bar*ones(1,M));
em_proj = matP*(S - Y_bar*ones(1,M));
K = size(T_out, 1);
for r = randperm(M)
    comp_r = setdiff(1:M,r);
    alpha_r = C(comp_r,:);
    alphar(1:N) = C(r,:);
    invSigma_r = sum(C(r,:).^2)*(matU'*inv(sigma)*matU);
    Sigma_r = inv(invSigma_r);
    er = em_proj(:,r);
    for k = randperm(K);
        tr = T_out(:,r);
        comp_k = setdiff((1:K),k);
        S_r = S_out(:,comp_r);
        Delta_r = (Y-S_r*alpha_r-Y_bar*alphar);
        mu = Sigma_r*matU'*(sum(inv(sigma)*(Delta_r.*(ones(L,1)*alphar)),2));
        skr = Sigma_r(comp_k,k);
        Sigma_r_k = Sigma_r(comp_k,comp_k);
        inv_Sigma_r_k = inv(Sigma_r_k);
        muk = mu(k) + skr'*inv_Sigma_r_k*(tr(comp_k,1)-er(comp_k,1));
        s2k = Sigma_r(k,k) - skr'*inv_Sigma_r_k*skr;
        vect_e = (-Y_bar - matU(:,comp_k)*tr(comp_k,1))./matU(:,k);
        setUp = (matU(:,k)>0);
        setUm = (matU(:,k)<0);
        mup = max([-1/eps max(vect_e(setUp))]);
        mum = min([ 1/eps min(vect_e(setUm))]);
        T_out(k,r) = dtrandn_MH(T_out(k,r),muk,sqrt(s2k),mum,mup);
        S_out(:,r) = abs(matU*T_out(:,r) + Y_bar);
    end
end


function [wmat, dwmat, Y_bar] = whitenmat(Y, firstEig, lastEig)
%-------------------------------------------------------------------------
% Whiten a matrix using the eigenvectors between firstEig and lastEig
% INPUT
%   Y : observational data
%   firstEig : first eigenvalue to keep
%   lastEig : last eigenvalue to keep
% OUTPUT
%   wmat : whiten matrix
%   dwmat : unwhiten matrix
%   Y_bar : columnwise mean of Y
%-------------------------------------------------------------------------
[L,N] = size(Y);
Y_bar = mean(Y,2);
Yc = Y - Y_bar*ones(1,N);
covMatrix = Yc*Yc';
[E, D] = eig(covMatrix);
rankTolerance = 1e-7;
maxLastEig = sum (diag (D) > rankTolerance);
if maxLastEig == 0,
  fprintf (['Eigenvalues of the covariance matrix are' ...
        ' all smaller than tolerance [ %g ].\n' ...
        'Please make sure that your data matrix contains' ...
        ' nonzero values.\nIf the values are very small,' ...
```

```matlab
          ' try rescaling the data matrix.\n'], rankTolerance);
  error ('Unable to continue, aborting.');
end
eigenvalues = sort(diag(D), 'descend');
oldDimension = size (Y, 1);
if lastEig < oldDimension
  lowerLimitValue = (eigenvalues(lastEig) + eigenvalues(lastEig + 1))/2;
else
  lowerLimitValue = eigenvalues(oldDimension) - 1;
end
lowerColumns = diag(D) > lowerLimitValue;
if firstEig > 1
  higherLimitValue = (eigenvalues(firstEig - 1) + eigenvalues(firstEig))/2;
else
  higherLimitValue = eigenvalues(1) + 1;
end
higherColumns = diag(D) < higherLimitValue;
selectedColumns = lowerColumns & higherColumns;
E = selcol(E, selectedColumns);
D = selcol(selcol(D, selectedColumns)', selectedColumns);
E = rotatefactors(E(:,firstEig:lastEig), 'Normalize', 'off');
wmat = sqrt(D)\E';
dwmat = E * sqrt (D);
return;
%%
function newMatrix = selcol(oldMatrix, maskVector)
if size(maskVector, 1) ~= size(oldMatrix, 2),
  error ('The mask vector and matrix are of uncompatible size.');
end
numTaken = 0;
for i = 1 : size (maskVector, 1),
  if maskVector(i, 1) == 1,
    takingMask(1, numTaken + 1) = i;
    numTaken = numTaken + 1;
  end
end
newMatrix = oldMatrix(:, takingMask);
return;

function C_out = update_C(C,S,sigma,Y)
%-------------------------------------------------------------------------
%Sample the posterior distribution of the fraction matrix
% INPUT
%       C : Current state of fraction matrix
%       S : current state of end-member matrix
%       Y : observational data
%       sigma : current state of the covariance of error
% OUTPUT
%       C_out : updated state of abundance matrix
%-------------------------------------------------------------------------
[M,N] = size(C);
ord = randperm(M);
k = ord(M);
comp_k = ord(1:(M-1));
alpha  = C(comp_k,:);
u = ones(1,M-1);
```

```matlab
SM = S(:,k);
SM_u = SM*u;
S_M = S(:,comp_k);
T = (S_M-SM_u)'*inv(sigma)*(S_M-SM_u);
for j=1:N
    Sigma = inv(T);
    Mu = Sigma*((S_M-SM_u)'*inv(sigma)*(Y(:,j)-SM));
    alpha(:,j) = dtrandnmult(alpha(:,j),Mu,Sigma,1);
end
C_out(ord(1:(M-1)),:) = alpha;
C_out(ord(M),:) = max(1-sum(alpha,1),0);
end


function S = dtrandnmult(S,Mu,Re,unit)
%-------------------------------------------------------------------------
% Sample a multivariate normal distribution with mean Mu and
% covariance Re on a simplex bounded between 0 and unit
%-------------------------------------------------------------------------
S = S(:);
Mu = Mu(:);
R = length(S);
if R==1
    S = trandn(Mu,sqrt(Re));
else
    for r=1:R
        Rm = Re;
        Rm(r,:) = [];
        Rv = Rm(:,r);
        Rm(:,r) = [];
        Sigma_mat{r} = inv(Rm);
        Sigma_vect{r} = Rv;
    end
    for iter=1:10
        for k=randperm(R)
            Sk = S;
            Sk(k) = [];
            Muk = Mu;
            Muk(k) = [];
            Moy_Sv(k) = Mu(k) + Sigma_vect{k}'*Sigma_mat{k}*(Sk-Muk);
            Var_Sv(k) = Re(k,k) - Sigma_vect{k}'*Sigma_mat{k}*Sigma_vect{k};
            Std_Sv(k) = sqrt(abs(Var_Sv(k)));
            S(k) = dtrandn_MH(S(k),Moy_Sv(k),Std_Sv(k),0,(unit-sum(S)+S(k)));
        end
    end
end
return;
%%
function X = trandn(Mu,Sigma)
%-------------------------------------------------------------------------
% Sample a positive normal distribution with mean Mu and standard
% deviation Sigma
%-------------------------------------------------------------------------
Mu = Mu(:);
Sigma = Sigma(:);
T = length(Mu);
U = rand(T,1);
```

```matlab
V = erf(- Mu./(sqrt(2)*max(Sigma,eps)));
X = Mu + sqrt(2*Sigma.^2) .* erfinv(-(((1-V).*U + V)==1)*eps+(1-V).*U + V);
X = max(X,eps);
return;



function X = dtrandn_MH(X,Mu,Sigma,Mum,Mup)
%-------------------------------------------------------------------------
% Sample a truncated normal distribtuion bounded between Mum and Mup with
% mean Mu and standard deviation Sigma
%-------------------------------------------------------------------------
Mu_new = Mu - Mum;
Mup_new = Mup -Mum;
if Mu<Mup
    Z= randnt(Mu_new,Sigma,1);
else
    delta = Mu_new - Mup_new;
    Mu_new = -delta;
    Z= randnt(Mu_new,Sigma,1);
    Z = -(Z-Mup_new );
end
    Z = Z+Mum;
    cond = (Z<=Mup) && (Z>=Mum);
    X = (Z.*cond + X.*(~cond));
return;
%%
function x = randnt(m,s,N)
%-------------------------------------------------------------------------
% Generate N random numbers from a positive normal distribution with
% mean M and standard deviation S
%-------------------------------------------------------------------------
if s<0,  error('Standard deviation must be positive.'); end;
if N<=0, error('N is wrong.'); end;
Tindcand = [];
x = [];
NN = N;
A  = 1.136717791056118;
mA = (1-A^2)/A*s;
mC = s * sqrt(pi/2);
while length(x)<NN,
    if m < mA,
        a = (-m + sqrt(m^2+4*s^2)) / 2 / s^2;
        z = -log(1-rand(N,1))/a;
        rho = exp( -(z-m).^2/2/s^2 - a*(m-z+a*s^2/2) );
    else if m <= 0,
        z = abs(randn(N,1))*s + m;
        rho = (z>=0);
    else if m < mC,
        r = (rand(N,1) < m/(m+sqrt(pi/2)*s));
        u = rand(N,1)*m;
        g = abs(randn(N,1)*s) + m;
        z = r.*u + (1-r).*g;
        rho = r.*exp(-(z-m).^2/2/s^2) + (1-r).*ones(N,1);
    else
        z = randn(N,1)*s + m;
        rho = (z>=0);
```

```matlab
    end;
    reject = (rand(N,1) > rho);
    z(reject) = [];
    if ~isempty(z), x = [x ; z]; end;
    N = N-length(z);
end
return;


function sigma_post = update_sigma(Y,S,C,psi)
%-------------------------------------------------------------------------
%sample the posterior distribution of the covariance matrix of error
%-------------------------------------------------------------------------
[L,N] = size(Y);
nv = L + 1;
a = (Y-S*C)*(Y-S*C)'+ psi;
b = nv + N;
sigma_post = iwishrnd(a,b);
end


function psi_post = update_psi(sigma,phi)
%-------------------------------------------------------------------------
%sample the posterior distribution of hyperparameter matrix psi
%-------------------------------------------------------------------------
L = length(sigma);
nv = L + 1;
theta = L;
a = inv(phi)+inv(sigma);
b = nv+theta;
psi_post = wishrnd(a,b);
end


function [mode_M,mean_C,mean_S] = postprocess(NM,NC,NS,Y)
%-------------------------------------------------------------------------
% Find the modal value of the number of end members
% INPUT
%       NM: Markov chain of the number of end members
%       NC: Markov chain of the abundance of the end members
%       NS: Markov chiain of the spectra of the end members
% OUTPUT
%       mode_M: modal value of the number of end members
%       mean_C: mean of the abundance of the end members
%       mean_S: mean of the spectra of the end members
%-------------------------------------------------------------------------
Mmax = maxem(Y);
h = hist(NM,1:Mmax)/length(NM);
mode_M = find((h == max(h)));
indM = [];
for i = 1:length(NM)
    if NM(i) == mode_M
        im = i;
        indM = [indM im];
    end
end
C = NC(1:mode_M,:,indM);
```

```matlab
S = NS(:,1:mode_M,indM);
mean_S = 0;
mean_C = 0;
for i = 1:length(indM)
    mean_S = mean_S + S(:,:,i);
    mean_C = mean_C + C(:,:,i);
end
mean_S = mean_S/length(indM);
mean_C = mean_C/length(indM);
end


function visualization(NM,mode_M,mean_S,mean_C,binsize,depth,Y)
%-----------------------------------------------------------------------
% Plot the results
%-----------------------------------------------------------------------
Mmax = maxem(Y);
figure (1)
subplot(1,2,1);stairs(NM(1:end-1));
set(gca,'yLim',[1 Mmax]);
set(gca,'YTick',1:1:Mmax);
set(gca,'fontsize',12);
xlabel('Number of MCMC run');
ylabel('Number of end members');
grid on;
subplot(1,2,2);bar(1:Mmax,hist(NM,1:Mmax)/length(NM),'b');
axis([0.5 Mmax+0.5 0 1]);
set(gca,'fontsize',12);
xlabel('Number of end members');
ylabel('Posterior probability');
grid on;
figure (2)
LEG = [];
for i = 1:mode_M
  LEG = [LEG; 'End member ', num2str(i)];
end
semilogx(binsize,mean_S,'-');
set(gca,'fontsize',12);
xlabel('Grain size (\mum)');
ylabel('Frequency (%)');
legend(LEG);
grid on;
figure (3)
for i=1:mode_M
    subplot(1,mode_M,i);plot(mean_C(i,:),depth,'-');
    set(gca,'fontsize',12);
    xlabel('Abundance');
    ylabel('Depth');
    set(gca, 'YDir', 'reverse');
    title(['End member ', num2str(i)]);
    grid on;
end
end
```

Vita

The author was born in Jiaozhou, Shandong Province, China. He obtained his Master's degree in Quaternary Geology and Geomorphology from Nanjing University in 1997 and Doctor's degree in Quaternary Geology from Lund University in 2003. After 10 years dedication to the research into climatic changes at University of Minnesota Duluth and Tulane University, he joined the University of New Orleans mathematics graduate program to pursue a M.S. in statistics in 2013.