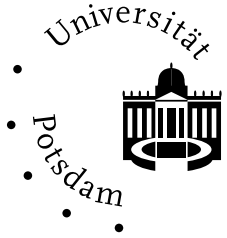
ADVANCED METHODS FOR ANALYSING AND MODELLING MULTIVARIATE PALAEOCLIMATIC TIME SERIES

DISSERTATION
ZUR ERLANGUNG DES AKADEMISCHEN GRADES
DOKTOR DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
IN DER WISSENSCHAFTSDISZIPLIN
THEORETISCHE PHYSIK / NICHTLINEARE DYNAMIK

vorgelegt von

DIPL.-PHYS. REIK DONNER



ARBEITSGRUPPE NICHTLINEARE DYNAMIK
INSTITUT FÜR PHYSIK
MATHEMATISCH-NATURWISSENSCHAFTLICHE FAKULTÄT
UNIVERSITÄT POTSDAM

Juni 2006

Kurzzusammenfassung

Die Separation natürlicher und anthropogen verursachter Klimaänderungen ist eine bedeutende Aufgabe der heutigen Klimaforschung. Hierzu ist eine detaillierte Kenntnis der natürlichen Klimavariabilität während Warmzeiten unerlässlich. Neben Modellsimulationen und historischen Aufzeichnungen spielt hierfür die Analyse von sogenannten Klima-Stellvertreterdaten eine besondere Rolle, die anhand von Archiven wie Baumringen oder Sediment- und Eisbohrkernen erhoben werden. Um solche Quellen paläoklimatischer Informationen vernünftig interpretieren zu können, werden geeignete statistische Modellierungsansätze sowie Methoden der Zeitreihenanalyse benötigt, die insbesondere auf kurze, verrauschte und instationäre uni- und multivariate Datensätze anwendbar sind.

Korrelationen zwischen verschiedenen Stellvertreterdaten eines oder mehrerer klimatologischer Archive enthalten wesentliche Informationen ber den Klimawandel auf großen Zeitskalen. Auf der Basis einer geeigneten Zerlegung solcher multivariater Zeitreihen lassen sich Dimensionen schätzen als die Zahl der signifikanten, linear unabhängigen Komponenten des Datensatzes. Ein entsprechender Ansatz wird in der vorliegenden Arbeit vorgestellt, kritisch diskutiert und im Hinblick auf die Analyse von paläoklimatischen Zeitreihen weiterentwickelt. Zeitliche Variationen der entsprechenden Maße erlauben Rückschlüsse auf klimatische Veränderungen. Am Beispiel von Elementhäufigkeiten und Korngrößenverteilungen des Cape-Roberts-Gebietes in der Ostantarktis wird gezeigt, dass die Variabilität der Dimension der untersuchten Datensätze klar mit dem Übergang vom Oligozän zum Miozän vor etwa 24 Millionen Jahren sowie regionalen Abschmelzereignissen korreliert.

Korngrößenverteilungen in Sedimenten erlauben Rückschlüsse auf die Dominanz verschiedener Transport- und Ablagerungsmechanismen. Mit Hilfe von Finite-Mixture-Modellen lassen sich gemessene Verteilungsfunktionen geeignet approximieren. Um die statistische Unsicherheit der Parameterschätzung in solchen Modellen umfassend zu beschreiben, wird das Konzept der asymptotischen Unsicherheitsverteilungen eingeführt. Der Zusammenhang mit dem Überlapp der einzelnen Komponenten sowie der aufgrund des Abschneidens und Binnens der gemessenen Daten verloren gehenden Informationen wird anhand eines geologischen Beispiels diskutiert.

Die Analyse einer Sequenz von Korngrößenverteilungen aus dem Baikalsee zeigt, dass bei der Anwendung von Finite-Mixture-Modellen bestimmte Probleme auftreten, die eine umfassende klimatische Interpretation der Ergebnisse verhindern. Statt dessen wird eine lineare Hauptkomponentenanalyse verwendet, um den Datensatz in geeignete Fraktionen zu zerlegen, deren zeitliche Variabilität stark mit den Schwankungen der mittleren Sonneneinstrahlung auf der Zeitskala von Jahrtausenden bis Jahrzehntausenden korreliert. Die Häufigkeit von grobkörnigem Material hängt offenbar mit der jährlichen Schneebedeckung zusammen, während feinkörniges Material möglicherweise zu einem bestimmten Anteil durch Frühjahrsstürme aus der Taklamakan-Wüste herantransportiert wird.

Abstract

The separation of natural and anthropogenically caused climatic changes is an important task of contemporary climate research. For this purpose, a detailed knowledge of the natural variability of the climate during warm stages is a necessary prerequisite. Beside model simulations and historical documents, this knowledge is mostly derived from analyses of so-called climatic proxy data like tree rings or sediment as well as ice cores. In order to be able to appropriately interpret such sources of palaeoclimatic information, suitable approaches of statistical modelling as well as methods of time series analysis are necessary, which are applicable to short, noisy, and non-stationary uni- and multivariate data sets.

Correlations between different climatic proxy data within one or more climatological archives contain significant information about the climatic change on longer time scales. Based on an appropriate statistical decomposition of such multivariate time series, one may estimate dimensions in terms of the number of significant, linear independent components of the considered data set. In the presented work, a corresponding approach is introduced, critically discussed, and extended with respect to the analysis of palaeoclimatic time series. Temporal variations of the resulting measures allow to derive information about climatic changes. For an example of trace element abundances and grain-size distributions obtained near the Cape Roberts (Eastern Antarctica), it is shown that the variability of the dimensions of the investigated data sets clearly correlates with the Oligocene/Miocene transition about 24 million years before present as well as regional deglaciation events.

Grain-size distributions in sediments give information about the predominance of different transportation as well as deposition mechanisms. Finite mixture models may be used to approximate the corresponding distribution functions appropriately. In order to give a complete description of the statistical uncertainty of the parameter estimates in such models, the concept of asymptotic uncertainty distributions is introduced. The relationship with the mutual component overlap as well as with the information missing due to grouping and truncation of the measured data is discussed for a particular geological example.

An analysis of a sequence of grain-size distributions obtained in Lake Baikal reveals that there are certain problems accompanying the application of finite mixture models, which cause an extended climatological interpretation of the results to fail. As an appropriate alternative, a linear principal component analysis is used to decompose the data set into suitable fractions whose temporal variability correlates well with the variations of the average solar insolation on millenial to multi-millenial time scales. The abundance of coarse-grained material is obviously related to the annual snow cover, whereas a significant fraction of fine-grained sediments is likely transported from the Taklamakan desert via dust storms in the spring season.

Contents

Introduction

The present global climate change has severe effects on the entire biosphere of the Earth. In addition to the successive environmental pollution due to the increasing human population and industrial activity, the climatic conditions controlling the growth of vegetation are changing. For example, in several areas of the world, the desertification has become significantly more intensive during the last centuries, which leads to a decrease of the area suitable for agriculture. In addition, a changing vegetation has an influence on the climate because the surface albedo is closely related to the global energy balance. Together with limited natural resources, the climate change is believed to be the most severe problem mankind is confronted with in the near future. Predictions based on intensive model studies suggest that under an industrial business-as-usual scenario, the global average temperature will increase by several degrees until the end of this century, which is underlined by recent studies coordinated by the intergovernmental panel of climate change (IPCC). This rise of temperatures (which will be nonhomogeneously distributed over the Earth) is likely to lead to a successive melting of polar ice caps followed by a dramatic change of the oceanic sea level, to shifts of atmospheric oscillation patterns, to a higher number and intensity of extreme weather events, and to a variety of other possible phenomena.

The question which contribution to this climate change is actually "man-made" (i.e., of anthropogenic origin) has been an intensive matter of debate during the last years. The knowledge of the corresponding answer is important in order to develop strategies for coping with the changing environmental conditions: whereas the anthropogenic part might be modified by sustainable environmental policies, the natural variability of environmental conditions can hardly be controlled. Hence, estimates of the ratio between anthropogenic and natural climate change are urgently required, which means that both contributions have to be separated in recent climate records as well as models. For this purpose, a detailed knowledge about the natural variability of the climate is necessary. Besides the simulation with suitable climate models, the study of palaeoclimatic proxy data from other historical periods with similar environmental conditions, but without anthropogenic influence may contribute essential information. Hence, there is particular interest in studying the climate variability in the early Holocene (the time period following the last glacial) and previous warm stages like the Eemian (about 120,000 years before present).

The Earth's climate varies on very different temporal scales related to different driving forces: The annual cycle is caused by the variations of the solar net radiation. Cycles on decadal scales as the El Niño/Southern Oscillation (ENSO) or the North-Atlantic Oscillation (NAO) are related to the internal variability of the athmosphere-ocean system. Solar cycles on the decadal scale (as the with a period around 11 years or the Gleisberg cycle with an average period length of 88 years) are known to drive climate variations. Centennial scale oscillations of the solar output have an effect on the ocean circulation. The orbital dynamics of the Sun/Earth system is responsible for climate variations on millenial scales and in particular for glaciation/deglaciation cycles. Plate tectonics and elevation of the continents and changes in their positions have a direct influence

on the atmospheric and oceanic circulation on very long time scales.

Palaeoclimatic proxy data representing variations of environmental conditions can be obtained on, e.g., sediments, ice cores, or rock deposits from locations distributed over the entire Earth. Therefore, the study of palaeoclimatic proxy data obtained from multiple sites is a major foundation of our present-day knowledge about the climate system and its natural and anthropogenically undisturbed behaviour on long time scales. To achieve an overall picture of the information included in a particular geological source under investigation, complementary measurements and analyses are thus performed. These analyses involve measurements of physical, chemical, and (in the case of sedimentary sequences) biological or sedimentological parameters as the corresponding observables are influenced by different climatic variables in different ways. Furthermore, age models (that quantify the age-depth relationship of a sediment or ice core) incorporating information of their uncertainties must be developed.

The intention of geological studies based on the observation of multiple palaeoclimatic proxy data is to identify and analyse signatures of climate change and attribute them to the variability of climatologically meaningful quantities. A direct interpretation of the individually measured time series in terms of meteorological parameters is often not possible. Therefore, it is a standard approach to derive variability patterns from multivariate geological time series which can be assigned to changes of meaningful climatic observables like temperature, moisture conditions (i.e., seasonal precipitation, snow volumes), vegetation cover, or the strength and location of different atmospheric oscillatory patterns. For this purpose, one frequently makes use of transfer functions. However, as such transfer functions are usually derived based on heuristic arguments and uncertain or incomplete data, this approach may be a potential matter of criticism (for examples, see, e.g., [Telford et al. 2004b, Telford and Birks 2005]).

This thesis contributes some new methodological ideas to the field of multivariate palaeoclimatic data analysis. As it is further discussed in the next chapter, there are many well-developed approaches of linear and non-linear time series analysis, which however can hardly be applied to palaeoclimatic data. This is due to the fact that time series obtained from geological sequences are characterised by features like an uneven sampling in the time domain, an uncertain chronology, a small number of observations with rather high noise levels, and instationarity including the potential presence of transitions between different states of the local climate system (e.g., glacial/interglacial variability).

The scientific achievements presented in this thesis are organised as follows: In Sect. 1, recent developments and remaining challenges in palaeoclimatic time series analysis are summarised. As a particular approach relevant for the statistical modelling of grain-size distributions (a palaeoclimatic proxy of increasing importance), finite mixture models are introduced in Sect. 2. New results about the statistical assessment of uncertainty in such models and open problems still to be solved on the way towards a standardised application of this approach to particle-size analysis in geology are discussed. In Sect. 3, dimension estimates based on the Karhunen-Loéve decomposition (KLD) of a record are introduced as a novel concept for quantifying the temporally variable content of information in general multivariate data sets. The potential power of this approach is presented by applying it to generic model systems and real-world palaeoclimatic data sets from Cape Roberts (East Antarctic). Possible directions of further research on this topic are outlined. An extensive discussion of the applicability of both, finite mixtures and KLD-based analysis, to a well-resolved sequence of grain-size distributions from Lake Baikal is presented in Sect. 4.

Chapter 1

Time Series Analysis in Palaeoclimatology

1.1 Motivation

The basic idea of time series analysis is to consider observational records as a realisation of a certain stochastic process. The features of this process can be described by suitable statistical characteristics, like probability distribution functions, correlations, Fourier spectra or other appropriate measures, which can be estimated from the measured time series. As most of these characteristics have originally been developed to describe the features of certain idealised stochastic models, their traditional estimators assume also ideal conditions, which are rarely present in case of real-world systems.

The special features of palaeoclimatic time series require some very specific modifications of these standard approaches of data analysis (which are discussed in the next sections). Typical difficulties include rapid transitions between states with different environmental conditions, a small number of observations, and a rather high degree of uncertainty of the respective measurements. In addition, standard estimators of measures in time series analysis use data with a constant sampling rate, i.e., measurements are carried out in constant time intervals. In palaeoclimatology, time series of proxy data are gained from geological sequences like sediments, ice cores, or rocks. Here, the accumulation rate of the deposits under investigation is a priori unknown and may significantly vary with time. Hence, even measurements of proxies with a constant sampling interval along the sequence do usually not guarantee that there is a constant sampling in the time domain as well. In general, this is clearly not the case, as in many situations, older material at the bottom of the sequence is more compressed than younger one at the top due to physical (higher pressure due to the mass of the upper layers) or (especially in the case of ice cores) chemical processes.

Fig. 1.1 shows three typical examples of palaeoclimatic records:

- The Deuterium record of an Antarctic ice core record near the Vostok station [Petit et al. 1999] covers the last about 420,000 years of climate history (see Sec. 1.5). δD describes the relative deviation of the deuterium content with respect to an international standard, SMOW (standard mean ocean sea water), and can be used as a proxy for palaeotemperature.

- The magnetic susceptibility record from a lacustrine sediment obtained in Lake Baikal, Eastern Siberia, covers the last about 20,000 years. Although this proxy is a parameter

strongly influenced by the palaeomagnetism, it is also very sensitive to palaeoclimatic conditions.

- The width of treen rings (averages over a suitably large ensemble of samples) and other records allow to reconstruct temperature fluctuations over the last centuries to millenia. As an example, an annual reconstruction of [Moberg et al. 2005] of northern hemisphere mean temperatures is shown.

In particular, the first example clearly illustrates the problem of uneven sampling, as there are only very few data in the older part of the sequence, although the measurement of the palaeotemperature proxy has been performed in regular intervals of 1 m along the ice core.

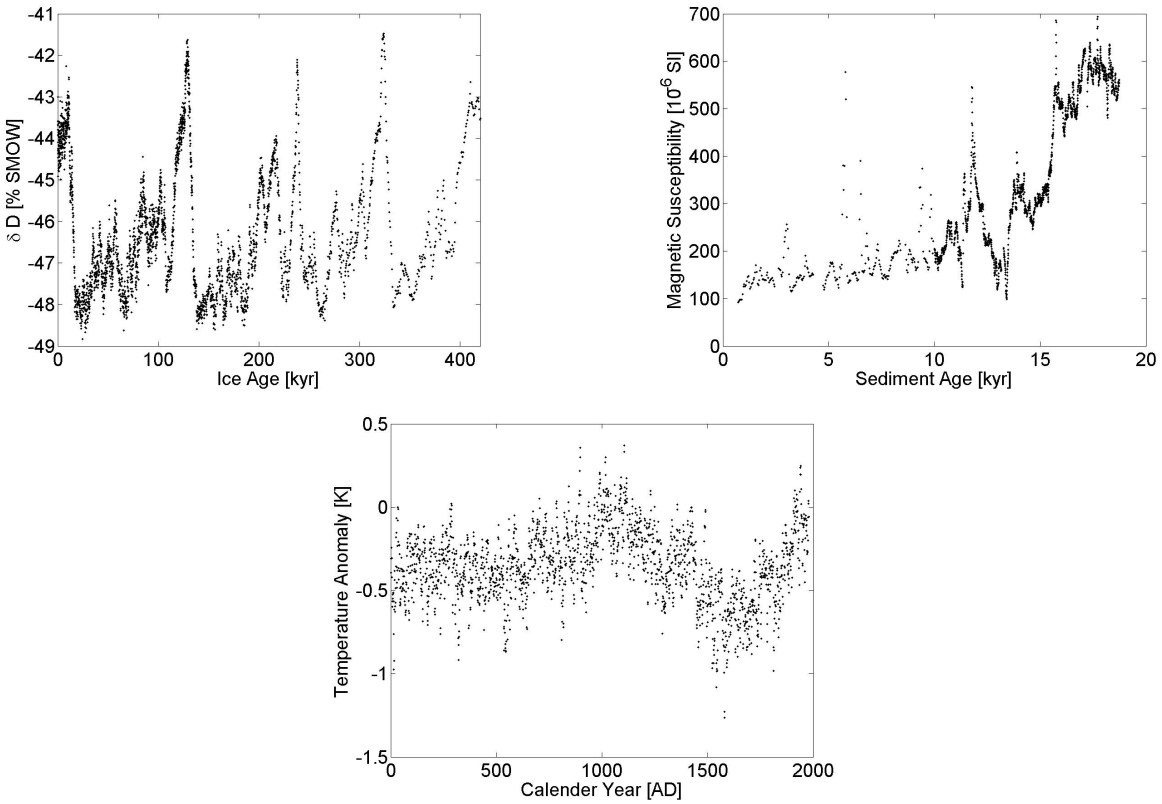


Figure 1.1: Three examples of palaeoclimate records. Upper left figure: Deuterium content from the Vostok ice core (Antarctica) as a measure of palaeotemperature over the last glacial cycles. Upper right figure: Magnetic susceptibility of a sediment record from Lake Baikal. Lower figure: Reconstruction of northern hemisphere annuan average temperatures during the last centuries (difference with respect to the mean value of the reference interval 1961-1990).

In this chapter, typical features of palaeoclimatic proxy data and the resulting challenges for time series analysis are summarised. The power and possible problems of wavelet analysis of geological records are studied as an example applicable to univariate data. Finally, potential approaches to the study of interrelationships of climatological time series and the corresponding relevant questions are briefly discussed.

1.2 Linear Time Series Analysis

The probably most traditional methods for the analysis of an univariate time series $X(t)$ with length T are Pearson's autocorrelation function

$$C_X(\tau) = \frac{\langle (X(t) - \langle X(t) \rangle) (X(t+\tau) - \langle X(t) \rangle) \rangle}{\left\langle (X(t) - \langle X(t) \rangle)^2 \right\rangle} \tag{1.1}$$

(where $\langle \cdot \rangle$ denotes the average value of the observable X, which is usually approximated by the sample mean taken from the time series of observations), and the power spectrum

$$S_X(k) = \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X(t) e^{2\pi i k t/T} \right|^2 . \tag{1.2}$$

Both characteristics are closely related via the Wiener-Khinchin theorem, stating that the Fourier transform of $C_X(\tau)$ equals the power spectrum. One can easily formulate appropriate generalisations to the case of bivariate data (i.e., two time series $X(t)$ and $Y(t)$) [Priestley 1981], leading to the cross-correlation function

$$C_{XY}(\tau) = \frac{\langle (X(t) - \langle X(t) \rangle) (Y(t+\tau) - \langle Y(t) \rangle) \rangle}{\sqrt{\left\langle (X(t) - \langle X(t) \rangle)^2 \right\rangle \left\langle (Y(t) - \langle Y(t) \rangle)^2 \right\rangle}} \tag{1.3}$$

and the cross-spectral density function

$$S_{XY}(k) = \mathcal{F} \{ C_{XY}(\tau) \} \tag{1.4}$$

where $\mathcal{F}(\cdot)$ denotes the Fourier transform operator. In a similar way, one can also find generalisations to multivariate time series consisting of $N > 2$ simultanaeosly recorded observables, however, in this case, correlations and spectral densities become $(N \times N)$ matrix-valued for any τ and k, resp.

Together with the probability density function, correlation functions and power spectra are referred to as methods of linear data analysis. This is due to the fact that their application bases on the assumption that the underlying statistical processes are linear-stochastic. The appropriateness of this assumption has to be validated by testing the corresponding null hypothesis using a suitable test statistics. The most conventional approach uses surrogate data with the same linear features like the original time series, which can be constructed by shuffling the phases of the Fourier-transformed time series or substituting them by random numbers [Theiler et al. 1992, Prichard and Theiler 1994]. In a similar way, one may also fit an autoregressive model of appropriate order to the data and consider a number of realisations of this model. From ensembles of such surrogate time series, one can compute suitable characteristics quantifying the deviation from a linear-stochastic process, e.g., the so-called Q-statistics measuring the skewness of a time series [Theiler et al. 1992]

$$Q(\tau) = \frac{\left\langle (X(t+\tau) - X(t))^3 \right\rangle}{\left\langle (X(t+\tau) - X(t))^2 \right\rangle}. \tag{1.5}$$

If the value of Q computed from the original time series is outside of a certain quantile of the distribution of the values from the surrogate data sets, the null hypothesis has to be rejected with the corresponding probability.

In addition to a linear-stochastic nature of the underlying process, the Gaussianity and stationarity of the time series are typical prerequisites for a success- and meaningful application of linear time series analysis. Here, Gaussanity means that the observed data are distributed according to a normal distribution, i.e., the statistical features of the time series are completely described by the first and second moments as all higher-order moments vanish. This condition is often implicitly assumed before applying linear methods in time series analysis. For example, it is convenient to standardise time series by substracting means and dividing by the standard deviations before calculating correlation functions. The above mentioned approach of Fourier surrogates for testing the hypothesis of a linear-stochastic process has usually to be improved in a similar way by an appropriate adjustment of the data to Gaussian distributions, leading to the (iterative) ampltude adjusted Fourier transform ((I)AAFT) surrogates [Schreiber and Schmitz 1996, Paluš and Novotna 1999, Schreiber and Schmitz 2000, Venema et al. accepted].

In comparison to the stationarity of a time serie, the Gaussianity condition is relatively mild. In particular, there are *non-parametric* statistical analoga of Pearson's correlation function for non-Gaussian time series based on a rank-ordering of the observed values. Examples from this class of rank-order correlation functions are Spearman's Rho (which converges to the Pearson correlation if $T \rightarrow \infty$) or Kendall's Tau [Lehmann 1975, Conover1980].

In contrast to a non-Gaussianity of a time series, instationarity may easily cause any linear method of data analysis to fail. Stationarity means that all statistical features of a time series (in particular, moments and correlations) are constant over the entire record. Depending on how the term "all statistical features" is interpreted, one may distinguish between different levels, from strong stationarity (all moments of arbitrary order are constant) to weak stationarity (the first and second moments are constant), where the latter one is again related to an implicit assumption of a Gaussian process. In general, a sophisticated proof of the stationarity of an observed process requires an appropriate statistical test [Witt et al. 1998, Rieke et al. 2002, Rieke et al. 2004]. However, even if a time-series is found to be non-stationary (which is the typical case for observational records, e.g., in geosciences), there are still possible generalisations of the above mentioned linear methods, for example, time-dependent (evolutionary) spectra [Priestley 1988] and probability density functions, or methods of time-frequency analysis like wavelets [Holschneider 1995].

1.3 Nonlinear Time Series Analysis

In real-world systems, completely linear processes are rather exceptional. Therefore, it has been necessary to develop appropriate concepts for the analysis of time series originated from nonlinear processes. In the following, the basic requirements of linearity and Gaussianity are skipped, however, stationarity of the time series is still assumed.

For time series from nonlinear stationary processes, there is a large variety of methods for quantifying the underlying dynamics [Abarbanel 1996, Kantz and Schreiber 1997]. These concepts include measures of predictability, complexity, instability, or fractality, which characterise the attractor properties (note that a basic assumption of time series analysis is that the observed system is in some equilibrium state, which is reflected by the stationarity of the record). Many of these approaches use an appropriate coarse-graining of the data, which loses information about the explicitly observed values, but preserves the topological features of the trajectory in phase space. Among the latter ones, fractal dimenensions, entropies, and measures derived from are some of the most prominent nonlinear methods of time series analysis.

For characterising the dimension of a dynamical system, there are different approaches, in-

cluding the Hausdorff, box counting, and several other dimensions [Falconer 1990]. However, these mathematically well-defined concepts can hardly be applied to real-world time series. Grassberger and Procaccia [Grassberger and Procaccia 1983] presented the concept of correlation dimension as a practically applicable alternative. For calculating this measure, one has to consider the correlation sum

$$C(\epsilon) = \frac{2}{N(N-1)} \sum_{t=1}^{T} \sum_{s=t+1}^{T} \Theta\left(\epsilon - ||X(t) - X(s)||\right) \tag{1.6}$$

where $\Theta(\cdot)$ denotes the Heaviside step function. Grassberger and Procaccia could show that this correlation sum has in the limit $T \to \infty$ and $\epsilon \to 0$ a characteristic scaling law, $C(\epsilon) \propto \epsilon^{D_2}$, such that the correlation dimension can be defined according to

$$D_2 = \lim_{\epsilon \to 0} \lim_{T \to \infty} \frac{\partial \ln C(\epsilon, T)}{\partial \ln \epsilon}. \tag{1.7}$$

To approximate the corresponding limit in an appropriate way, Grassberger and Procaccia proposed an algorithm using an m-dimensional embedding of the time series (i.e., considering the time series $Y(t) = (X(t), X(t+1), \ldots, X(t+m-1))$ instead of $X(t)$ itself) which estimates D_2 by the value of $D_2(m)$ where (as a function of m) a plateau is approached.

Apart from the Grassberger-Procaccia algorithm, there is a variety of other approaches for estimating the fractal dimension of a time series. The most traditional methods use the specific scaling behaviour of auto-correlation function and power spectrum of self-similar processes, however, they work well only in rather exceptional cases. Other approaches consider the curve length of the embedded time series in dependence of the choice of the sampling interval used for embedding [Burlaga and Klein 1986, Higuchi 1988]. I has been demonstrated that this concept works rather well for geoscientific and other "irregular" time series. A similar concept is the so-called *fluctuation analysis*, where the scaling of rms displacements

$$F(\tau) = \sqrt{\left\langle \left(X(t+\tau) - X(t)\right)^2 \right\rangle - \left\langle X(t+\tau) - X(t) \right\rangle^2} \tag{1.8}$$

with varying τ is considered. Peng and co-workers [Peng et al. 1994] proposed to improve this method by considering pieces of the time series which have been locally detrended, i.e., the original time series is replaced by its residual with respect to a low-order polynomial obtained by a least-square fitting of the observations in the considered time interval. The resulting *detrended fluctuation analysis* has found many applications in the geosciences, however, the resulting scaling behaviour is often interpreted as a long-range memory of the generating process which may be misleading in certain situations [Maraun et al. 2004].

In climatology, the concept of fractal dimensions has been used to find a characterisation of a supposed "climate attractor" in terms of fractal theory. In particular, the Grassberger-Procaccia algorithm has been applied to different univariate time series from meteorology [Sahay and Sreenivasan 1996] and palaeoclimatology [Maasch 1989, Schulz et al. 1994, Mudelsee and Stattegger 1994a, Mudelsee and Stattegger 1994b, Mudelsee 1995]. The latter articles clearly demonstrated that this approach has severe methodological problems when being applied to palaeoclimatic data, which is mainly related to the limited amount of data and the instationarity of the record, including certain transitions of the climate system [Mudelsee and Stattegger 1997].

A further obstacle might be even more fundamental: as the concept of fractal dimensions is based on the assumption of a self-similar process, the corresponding measures may be misleading

if this particular condition is not fulfilled. In turbulence, but also in a wide range of other complex systems, the underlying system has been found to violate the self-similarity condition, leading to more sophisticated concepts like extended self-similarity or multifractal analysis. For example, in geosciences, observables characterising the hydrological cycle (like precipitation or river runoff records) are rather frequently found to have a multifractal character. In this case, the record cannot be described by a single fractal dimension, but requires a complete spectrum of such dimensions. Hence, if an observational record is not self-similar, the estimation of D_2 or similar measures may not be appropriate for characterising the generating physical system.

The methods of nonlinear time series analysis discussed above have been originally introduced for applications to univariate time series. However, several of the measures can be appropriately extended to the multivariate case, including the Lyapunov spectrum [Bünner and Hegger 1999], the scaling of fractal dimensions [Politi and Witt 1999], or dimension densities based on a normalised Grassberger-Procaccia algorithm [Bauer et al. 1993]. As shown by [Olbrich et al. 1998], the dimensions calculated from data sets depend crucially on the resolution of the observations. To overcome the corresponding difficulties in practical applications, Raab and co-workers [Raab and Kurths 2001, Raab et al. 2005] have proposed a normalisation for approaching large-scale correlation dimension densities (LASDID). Although this method has been designed for applications to relatively short multivariate time series, it remains hardly possible to use measures based on the "traditional" correlation dimension for characterising palaeoclimatic data sets. Some of the reasons for this will be discussed in the next section.

1.4 Typical Problems in Palaeoclimatic Data Analysis

In the previous sections, some of the problems frequently occurring in the analysis of palaeoclimatic data have already been briefly mentioned. For example, measurements of a particular observable carried out on a part of a sediment or ice core reflect always information aggregated over some part of the climate history, whereas it is treated as if corresponding to one particular point in time. Hence, the sampling rate leads in general to a filtering of the data. Related problems are the uncertainty of measurement, possible perturbations of the deposited material (e.g., bioturbation, chemical reactions, diffusion etc.), and the sensitivity of palaeoclimatic proxy data to other influences not directly related to climate. In summary, all these effects cause geological time series to have (apart from their typical irregular pattern) a rather high noise level.

Another problem is that most methods of time series analysis are designed for the treatment of stationary data, i.e., the generating system is supposed to be in *one particular* equilibrium state or can at least be considered to be in a quasi-equilibrium where changes occur on timescales which are sufficiently large compared to the time interval under consideration. In contrast to this assumption, it is hardly possible to consider the climate system to be in an equilibrium, as it is subjected to variable external forcings (solar irradiation, variations of the Earth's orbital parameters, changes in the geomagnetic field, volcanic activity, etc.). As the components of the climate system interact with each other in a highly complicated way and obey a variety of complex internal feedback mechanisms, these variable forcings necessarily drive the system out of possible equilibrium states and may even generate dynamic transitions between states characterised by different climatic conditions and dynamic behaviour of the corresponding observables (e.g., a predominance of different atmospheric circulation patterns).

A very prominent example for externally induced transitions between states with different climatic conditions is the sequence of glacials (popularly called "ice ages") and interglacials. For the last about 1.8 million years (i.e., the Pleistocene and Holocene epochs), it is known

that intervals of cold and warm periods (corresponding to a significant glaciation and deglaciation of the high latitudes of both hemispheres) have been alternating in a more or less regular way, firstly with a period of about 41,000 years, later about 100,000 years. The dynamics of the corresponding mid-Pleistocene transition has been a subject of intensive research during the last decades [Mudelsee and Schulz 1997]. Following ideas already developed by Joseph Adhémar in 1842 and James Croll in the 1860s, Milutin Milankovich founded an astronomical theory to explain this switching behaviour in terms of variations of the Earth's orbital parameters and corresponding changes in the solar insolation acting as a pacemaker of the climate system [Paillard 2001, Berger and Loutre 2004, Bard 2004]. However, the so-called Milankovich theory is still subjected to intensive debates, as there are features which cannot be completely explained only by the varying irradiation [Raymo 1997, Rial and Anaclerio 2000, Elkibbi and Rial 2001, Wunsch 2003, Huybers and Wunsch 2003]. The observational fact of alternating glacial/interglacial variability motivated the development of the theory of stochastic resonance, stating the possibility of an amplification of a weak periodic forcing in a complex system due to the presence of noise [Benzi et al. 1982, Benzi et al. 1983].

Another example for alternations of the climate system on shorter time scales are the so-called *Dansgaard-Oeschger cycles* originally observed in ice-core records from Greenland covering the last glacial period [Dansgaard et al. 1993]. Surprisingly, the occurrence of the underlying characteristic pattern seems to follow a well-defined cycle with a priod of 1470 years [Bond et al. 1997, Mayewski et al. 1997, Grootes and Stuiver 1997]. Similar quasi-regular oscillatory patterns have also been observed in palaeoclimatic proxies from well-resolved marine sediments and southern-hemispheric ice cores, where the latter ones allow insights into the dynamics during earlier glacial periods. However, the southern-hemispheric millenial-scale oscillation seems to be out-of-phase with respect to the northern hemispheric one [Hinnov et al. 2002]. The appropriate determination of the corresponding leads or lags (being in the order of magnitude of the uncertainty of the age models assigned to the respective time series) is another intensively discussed problem which resembles the classical chicken-or-egg question. In addition, signatures of the 1470-year climatic cycle have also been found during the Holocene (i.e., the current warm stage) [Bianchi and McCave 1999, Bond et al. 2001]. During the last years, there has been an intensive debate whether the observed periodic signals are actually significant [Wunsch 2000, Schulz 2002a, Rahmstorf 2003, Witt and Schumann 2005].

Whereas the occurrence of Dansgaard-Oeschger events during glacial epochs is commonly believed to be related to abrupt freshwater discharges in the North Atlantic (leading to a change in the mode of the thermo-haline circulation), the actual mechanism leading to the pronounced periodicity is still unclear as there is no orbital counterpart acting on the corresponding timescale. A variety of possible explainations has been proposed, including internal instabilities of the ice shield [Schulz et al. 1999, Schulz 2002b] or ocean dynamics, tidal action, noise excitation, or an external forcing of unknown origin (see [Timmermann et al. 2003] for a review). The latter idea motivated a stochastic resonance approach [Alley et al. 2001], which may be triggered by suitable internal dynamics of the northern-hemispheric ice cover. Recently, the general possibility of such a mechanism has been demonstrated in climate models [Ganopolski and Rahmstorf 2002].

The two examples (glacial/interglacial and Dansgaard-Oeschger variability) illustrate that in the climate system, there are several possible states between whose transitions occur rather frequently. Hence, the climate system cannot be considered to be in an equilibrium state. As a consequence of the strong variability, palaeoclimatic records are usually far from being stationary, which causes problems for the statistically appropriate time series analysis. In addition to the dynamically meaningful transitions, there may be also strong outliers in the data caused by local (e.g., landslides, floods,...) or larger-scale (e.g., earthquakes) extreme climatic or non-climatic

events which increase the problems for data analysis. Concerning the high noise level, there are (in principle) techniques allowing to separate stochastic and deterministic contributions to the dynamics contained in a time series. However, as palaeoclimatic data are often characterised by a low amount of data (or, equivalently, a low sampling rate) due to the extreme costs for collecting and measuring samples, the number of available data is usually by far too low to apply such methods successfully.

Apart from these rather general features of palaeoclimatic time series, the major problem of data analysis in this field of research is related to the uneven sampling of the data in the time domain, which calls for highly specified methods of time series analysis to derive meaningful information about the dynamics of the underlying dynamics. In particular, the above mentioned periodicities of climatic variations cannot be studied using simple power spectra due to the instationarity of data, but require the application of time-dependent spectral estimators or time-frequency techniques. For stationary data, appropriate techniques have been developed to estimate the power spectrum [Schulz and Stattegger 1997, Heslop and Dekkers 2002], persistence [Mudelsee 2002], or correspondung red-noise spectrum [Schulz and Mudelsee 2002] of a unevenly sampled time series.

In the case of instationary data, one may use wavelets instead of the standard Fourier analysis. For transferring this time-frequency method to unevenly sampled data, different approaches have been proposed, including the weighted wavelet Z-transform [Foster 1996c, Andronov 1997, Andronov 1998, Andronov 1999, Schumann 2004, Witt and Schumann 2005, Brauer et al. accepted, Witt and Oberhänsli 2006] based on a projection method in Fourier space [Foster 1996a, Foster 1996b], application of gapped wavelets [Frick et al. 1997, Frick et al. 1998], or a generalised multiresolution analysis approach [Otazu et al. 2002, Otazu et al. 2004]. However, although these methods are available, there is a number of articles describing the direct application of *standard* wavelet analysis to geological time series using a certain interpolation to equal sampling intervals, which (beside the age uncertainty always present in palaeoclimatic studies) necessarily leads to additional errors [Bolton et al. 1995, Guyodo et al. 2000, Hargreaves and Abe-Ouchi 2003, Glushkov et al. 2005]. This particular issue will be discussed in some detail in the next section. To completely overcome the problem of uncertain age models [Telford et al. 2004a], research efforts are currently made to combine the particular analysis method with an appropriate statistical approach, e.g., using Monte Carlo age models or probability density distributions for the uncertain ages in a Bayesian framework.

1.5 Wavelet Analysis in Palaeoclimatology: A Univariate Example

Wavelet analysis is a suitable tool for investigating periodic components, which occur temporarily or with non-constant amplitudes. In particular, when studying the dynamics of the Earth's climate on large time scales, wavelet analysis of palaeoclimatic proxy data provides age intervals that are dominated by certain periodicities. In general, one has to distinguish between continuous and discrete wavelet transforms, which again may have different versions (for example, the non-decimated wavelet transform (NWT) utilised by [Glushkov et al. 2005] is a particular version of discrete wavelet transform (DWT)).

As already discussed in the previous section, the uncertainty of age models of palaeoclimatic records (see, e.g., [Telford et al. 2004b]) is the major source of problems and errors in spectral analysis. Moreover, this uncertainty is not constant along a geological sequence: With increasing

depth, there are usually less points of observation within a given time interval, such that the uncertainty of the corresponding age estimates in this part of the record is larger than for younger sediments. In addition, the uncertainty of every separate value increases with increasing age due to limits of the corresponding methods (like radiocarbon / AMS ^{14}C, luminescence dating, etc.).

Since observational evidence of Milankovitch theory has been found, records are frequently tuned to the variability curve of solar irradiation. This means that between isolated points with directly measured (but usually uncertain) age values, timescales are obtained by graphical adjustment of large-scale variability patterns in the data with respect to their apparent representations in the reference. It is questionable if such records can be used to study variations on Milankovitch scales because they are implicitly used to generate these records. Variations on Milankovitch scales of a record can be analysed if the corresponding age model is based on a suitably large amount (or equivalently, a small spacing) of directly measured age values, and (realistic) confidence intervals of all estimated ages that are significantly smaller than the considered period.

In the following, two exemplary time series originally studied by [Glushkov et al. 2005] are again investigated: the isotope record from a well-studied Antarctic ice core obtained at the Vostok station, and a composite record containing observations from three tropical sediment cores.

For the Vostok ice core, various age models based on different approaches have been published (see, e.g., [Ruddiman and Raymo 2003]). As [Glushkov et al. 2005] have examined the deuterium-based relative temperature data, it is likely that the GT4 timescale (also known as the extended glaciological timescale EGT 20) according to [Petit et al. 1999] has been applied. GT4 gives estimates for the age of the ice which is appropriate for analyzing the Deuterium signal. However, the estimated uncertainty of the corresponding age values ranges between 5 and 15 kyr where at least the latter value (for the older part of the record) leads to severe problems for reconstructing variability signals on the 20 kyr band. Alternatively, there are more recent age models that adjust the atmospheric δ^{18}O to a synthetic orbital signal ([Shackleton 2000]) or use CH_4 measurements ([Ruddiman and Raymo 2003]). Both models are based on the chemical characteristics of the gas that is confined in bubbles of the ice cores and, therefore, the gas age may differ from the ice age significantly.

Following a proposal of [Shackleton 1995], the records of the marine cores V19-30 (replacing the originally used SPECMAP stack for the upper 620 kyr of the composite sequence), ODP 677 and ODP 846 have been combined to one marine sequence (sometimes referred to as the S95 composite ([Lisiecki and Raymo 2005])) covering the last approx. 6 Myr of climate history. The corresponding timescale has been obtained by combining the age models of the respective components. Recently, several groups have considered benthic δ^{18}O records from various marine cores to construct more sophisticated composites with improved age models (e.g., [Karner et al. 2002]). Correlating data sets with certain age measurements yields more reliable timescales due to the larger amount of references. The resulting age models are based on different approaches. Very promising for this type of data analysis is the depth-derived, minimally tuned HW04 timescale of [Huybers and Wunsch 2004] spanning the last 780 kyr. In contrast, the (most recent) LR04 age model by [Lisiecki and Raymo 2005] is essentially aligned to the orbital forcing. By choosing either one or the other timescale, significant differences in the spectral domain may be expected. A fair analysis should involve different age models and discuss their effects on the spectral properties of the records.

A typical (but nonetheless problematic) approach to cope with unevenly sampled data is an appropriate interpolation of the observed time series to a given, uniformly spaced grid. For example, [Glushkov et al. 2005] have presented a variant of this method by interpolating with

cubic Hermite polynomials. However, as spline interpolation has been found to fail in their application, the appropriate choice of the local polynomials seems to play a crucial role for the data sets considered.

Observational records do not provide information about the behaviour on time scales shorter than the temporal resolution. The interpolation approach implicitely assumes that the observable is well-behaved in time. This assumption is problematic for palaeoclimatic proxy data which frequently show large differences (e.g., of temperatures) within small time windows. If a signal consisting of temporary averages of a particular parameter (being the typical case in palaeoclimatology) is interpolated again, the corresponding reconstruction of the variability may remarkably differ from the actual one which influences the results of the wavelet analysis. Even if the a particular wavelet approach performs well for analysing data with regular sampling, its outcome must be treated with special care in the discussed application.

There are several approaches to overcome the problem of being dependent on information distributed on a regular grid. The simplest idea is based on an application of the Haar wavelet, a piecewise continuous function ([Scargle 1997]). A more realistic decomposition is provided by using differentiable functions as mother wavelets (e.g., the Morlet wavelet). [Frick et al. 1998] proposed appropriate corrections of the wavelet transform for gapped data records (GWM). [Foster 1996c] introduced the weighted wavelet Z-transform (WWZ), a projection method reorthogonalizing the three basic functions (real and imaginary part of the Morlet wavelet and a constant) by rotating the matrix of their scalar products. He furthermore introduces appropriate statistical tests to distinguish between periodic components and a noisy background signal for unevenly sampled data. [Andronov 1998] suggested a further improvement of WWZ by introducing additional weighting factors. Recently, WWZ was successfully applied to palaeoclimatic records by [Witt and Schumann 2005]. Because GWN and WWZ consider only the information directly measured along the cores, they are not affected by additional uncertainties caused by interpolation. From this point of view, the corresponding approaches are more suitable for the analysis of unevenly sampled data.

Apart from problems with interpolating data, wavelet analysis cannot be applied as a black-box method because different choices for the mother wavelets, i.e., their shapes and spectral representations are possible. Moreover, in case of a continuous wavelet transform, the scale parameter that represents the spectral bandwidth must be fixed. It has to be statistically tested if the wavelet coefficients indicate a signal that is significantly different from a white or red noise climatic background. Only if such a test fails and the variability beyond the noisy background signal is related to Milankovitch cycles, the reconstruction of the original signal as a superposition of wavelet filtered components on Milankovitch scales is justified.

To illustrate the dependence on the wavelet basis, a continuous wavelet analysis (as in [Witt and Schumann 2005]) has been performed for the Vostok deuterium record (this data set is analysed rather than the temperature reconstruction for a better consistency to the marine δ^{18}O records studied by [Glushkov et al. 2005]). Considering the wavelet amplitude map shown in Fig. 1.2, it is found that this record has a rather complex variation structure. It is a very rough approximation to model the variations of the Vostok deuterium record by a superposition of variations on the three Milankovitch scales of about 20, 42, and 100 kyr, and a red noise background (see Fig. 1.3). In particular, an additional period of about 60 kys (eventually caused by a superposition of higher frequencies) occurs for the last 150 kys and has to be taken into consideration.

Even if the above mentioned problems are ignored for reconstructing the variations on Milankovitch scales, these reconstructions are far from being identical to those found by [Glushkov et al. 2005]. Remarkable differences are especially found in the signal correspond-

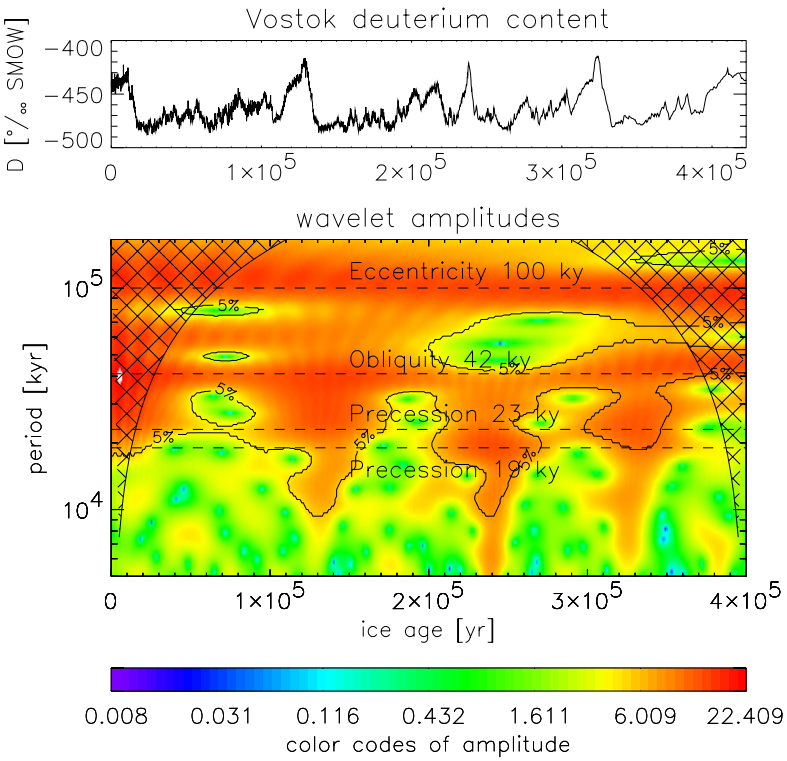


Figure 1.2: Original time series of the Vostok deuterium record (upper panel) and the corresponding wavelet amplitudes (lower panel) depending on the localization (age) and the period length. For calculations, the WWZ method and an abbreviated Morlet wavelet have been used. Cross hatched regions indicate the cone of influence, where the wavelet analysis is affected by edge effects. Contour lines mark periodic shares (red colors indicating the strongest periodicities) that are significantly different from a red noise background assuming an error of 5% (see [Witt and Schumann 2005] for details). The periods of the major Milankovitch cycles are displayed by dashed lines. The bottom panel displays the color codes.

ing to eccentricity. This difference is caused on the one hand by the broader spectral bandwidth of the Daubechies wavelet compared to the abbreviated Morlet wavelet. On the other hand, the discrete wavelet transform reconstructs the entire signal, whilst the continuous wavelet transform only its variability. This finding (which shall not claim if either of the reconstructions is correct) illustrates that the choices of the wavelet basis and the wavelet scale parameters matter and have to be discussed in detail.

1.6 Correlations in and between Palaeoclimate Records

In the previous sections, two particular questions have already been addressed related to the question of adjustment and causality of palaeoclimatological records: On the one hand, one frequently combines different observational data sets to one geological composite record. Similarly, age models are often artificially transferred from well-studied sequences to cores where no proper age estimates are available. In particular, dominating variability patterns are adjusted to variations of orbital parameters (astronomical tuning), e.g., by using the so-called SPECMAP

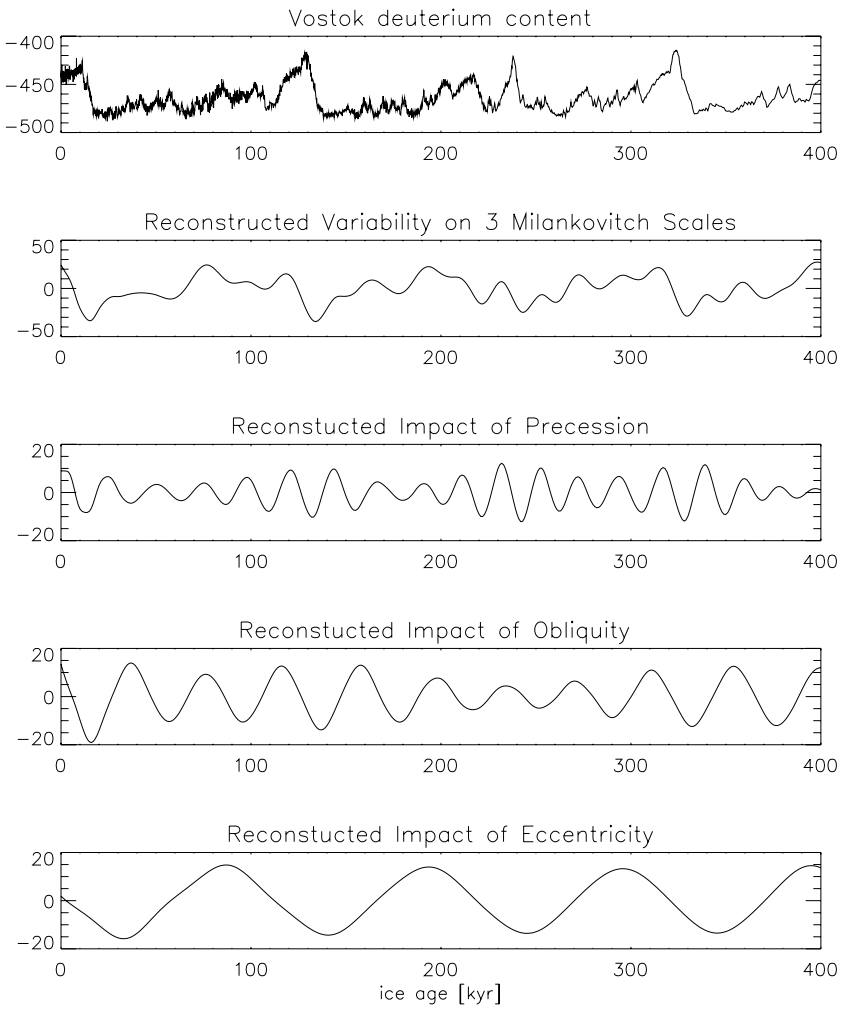


Figure 1.3: Reconstruction of wavelet amplitudes at the major Milankovitch scales of 100 (eccentricity), 42 (obliquity), and 20 kyr (precession), and the composition of these scales compared to the long-term variability of the original time series (from bottom to top).

curve. To validate all these approaches, one has to assure that the physical mechanisms and processes behind all considered data sets are actually comparable, which must not necessarily be the case. If this assumption cannot be proven, the entire approach may be misleading. On the other hand, the evaluation of leads and lags between climatic transitions on both hemispheres is an intensively discussed problem, which is essentially related to a proper mechanistic understanding of the underlying processes. The observation-based study of the corresponding causality requires very accurate age estimates, which are usually missing.

The first question has already been under investigation for several decades. The traditional approach in geology is the so-called *s*equence slotting (see [Thompson and Clark 1989] and references therein) which aims on identifying prominent "events" in different observational records with each other and (usually linearly) interploating the relative age model between these points. Beside a manual adjustment, application of dynamic programming techniques allows a wide-ranging automatisation of the process [Lisiecki and Lisiecki 2002].

A more sophisticated approach for deriving such relative age-depth models from the intercomparison of different (univariate) data sets may use non-parametric regression, for example, the method of maximum correlation and optimum transformations which can be estimated using the ACE algorithm [Breiman and Friedman 1985, Voss and Kurths 1997]. Witt and Oberhänsli have proposed to transfer the (unknown) relative age model $T(t)$ into a three-dimensional phase space spanned by the sediment depth t, the age model $T(t)$, and its derivative dT/dt, and compute a path in this phase space which has minimum length under an appropriate smoothness constraint [Witt and Oberhänsli 2003].

Finally, a third approach for synchronising the time scales of two geological records uses the so-called line of synchronisation in a cross-recurrence plot [Marwan et al. 2002a]. The concept of recurrence plots [Eckmann et al. 1987] has originally been designed as a tool to visualise the correlation pattern within a single time series by comparing the observed value at any given time t_i with the values at any other times t_j. A simple graphical representation is obtained by comparing the difference to a prescribed threshold value ϵ and encoding this difference in dependence on both times t_i and t_j according to the order relation with respect to ϵ. Mathematically, the corresponding recurrence matrix of a time series $X(t)$ is formulated using the Heavyside function as

$$R_X(t_i, t_j) = \Theta\left(\epsilon - ||X(t_i) - X(t_j)||\right) \tag{1.9}$$

(note that this matrix depends on the particular choice of ϵ) and may be used for defining a bunch of nonlinear characteristics, based on statistics of either the diagonal or the horizontal structures in terms of the so-called recurrence quantification analysis (RQA) [Zbilut and Webber 1992, Webber and Zbilut 1994, Marwan et al. 2002b]. The simplest idea is to consider the recurrence rate (i.e., the relative frequency of occurrence of the value $R_X(t_i, t_j) = 1$ on a diagonal defined by a fixed value of $\tau = t_j - t_i$) as a generalised correlation function sensitive also to nonlinear dependences. Moreover, the consideration of recurrence plots allows to estimate several dynamic invariants, including the second-order Renyi entropy K_2 (characterising the predictability of the system), the correlation dimension D_2, and the mutual information (see below) [Thiel et al. 2004, Asghari et al. 2004, von Bloh et al. 2005].

In a complex system, the question of a co-evolution of the dynamics of different observables may be of particular relevance. The recurrence plot approach can consequently be extended to an intercomparison between two time series $X(t)$ and $Y(t)$ by (i) substituting $X(t_j)$ by $Y(t_j)$ in the definition of the recurrence matrix in terms of cross recurrence plots $R_{XY}(t_i, t_j)$ [Zbilut et al. 1998], or (ii) point-wise multiplication of the component recurrence plots yielding joint recurrence plots [Romano et al. 2004].

Whereas in a standard recurrence matrix, the main diagonal only contains the value "1" by definition, this is not necessarily the case in a cross-recurrence plot. However, if both time series are strongly correlated, the value "1" will still have a very high probability. If now the time scale of one time series is transformed with respect to the second one, the predominant pattern of the value "1" will remain, but is shifted from the main diagonal. This pattern is called the line of synchronisation and may be used for adjusting the time scales of different geological records (at least if both reflect the same observable and/or have been obtained at neighboring locations. Reversing this argument, if the time scales of two records would be exactly known, a shift of the line of synchronisation would allow to estimate an eventual lead or lag between the palaeoclimate dynamics recorded in the considered sequences, giving some important information about the causality of the associated climate change. Unfortunately, as long as the uncertainties of the age estimates are in the order of the typically expected leads and lags, this issue remains to be a problem of mainly academic nature.

Apart from the consideration of correlations between different palaeoclimatic records as discussed above, the concept of correlation functions introduced in Sect. 1.2 can in general be adapted in a framework of nonlinear data analysis. One possibility is the generalised auto-correlation function derived from a recurrence plot (see above) which can be generalised to the bi- and multivariate case by considering cross- or joint-recurrence plots instead. Beside this approach, there is a variety of other nonlinear measures quantifying the statistical dependence between two time series. The probably best known of these measures is the (cross-)mutual information [Fraser and Swinney 1986]. To estimate this quantity, one traditionally considers an appropriate discretisation of the time series $X(t)$ and $Y(t)$ into symbols $\{x_i\}$ and $\{y_j\}$, resp. The probability of these symbols, p_i and p_j, as well as the joint probability $p_{ij}(\tau)$ that x_i and y_j occur simultaneously if the time series Y is lagged by τ time steps are empirically approximated by their frequencies of occurrence in the observational records. These probabilities are used to compute the corresponding Shannon entropies

$$H_X^{(1)} = -\sum_i p_i^{(X)} \log p_i^{(X)} \tag{1.10}$$

$$H_Y^{(1)} = -\sum_j p_j^{(Y)} \log p_j^{(Y)} \tag{1.11}$$

$$H_{XY}^{(1)}(\tau) = -\sum_{ij} p_{ij}(\tau) \log p_{ij}(\tau). \tag{1.12}$$

Finally, these Shannon entropies are combined yielding the following definition of the (cross-) mutual information function:

$$I_{XY}(\tau) = H_X^{(1)} + H_X^{(1)} - H_{XY}^{(1)}(\tau) = \sum_{ij} p_{ij}(\tau) \log \frac{p_{ij}(\tau)}{p_i p_j}. \tag{1.13}$$

$I_{XY}(\tau)$ is by definition restricted to non-negative values, but not normalised. To approach a corresponding standardisation to values in the unit interval, different proposals have been made (for an overview, see [Kojadinovic 2005]). One particular approach adapts the procedure used for the linear covariance function, i.e.,

$$J_{XY}(\tau) = \frac{I_{XY}(\tau)}{\sqrt{I_X(0)}\sqrt{I_Y(0)}}. \tag{1.14}$$

Here I_X and I_Y are the univariate mutual information of the time series $X(t)$ and $Y(t)$, resp. The concept of mutual information may also be further generalised by substituting the Shannon entropies $H^{(1)}$ by Rényi entropies of order q,

$$H_X^{(q)} = \frac{1}{1-q} \log \sum_i \left[p_i^{(X)}\right]^q \tag{1.15}$$

etc. The resulting generalised (cross-) mutual information functions may be useful for studying the dynamics of suitable model systems in some detail [Pompe 1993], however, their applicability to noisy and nonstationary (palaeo-)climatological data sets with a small amount of observations is rather doubtful. In addition, the generalised mutual information functions are not necessarily bounded from below by zero, which makes their interpretation more difficult. The concept of (generalised) mutual information can also be transferred to the analysis of multivariate data sets in terms of the so-called redundancies which can be applied to detect and test for nonlinearity in multivariate observational records [Paluš et al. 1993, Paluš and Novotna 1994, Paluš 1995, Paluš 1996].

1.7 Climate Records: Correlation or Synchronisation ?

In order to be able to interprete the results of palaeoclimatic data analysis in an appropriate way, it is necessary to have a good knowledge about present-day climatology. The climate system is a high-dimensional complex system which is subjected to different global and local forcings and the nonlinear action of internal feedback mechanisms. Therefore, its behaviour is highly chaotic and characterised by an extreme sensitivity which may lead to sudden changes in the dynamics of the entire system. Time series of variables recording the corresponding variability are therefore typically very irregular and have rather high noise levels. This holds in particular for the case of hydro-meteorological data obtained from direct measurements since the start of the instrumental period, reconstructions of earlier time intervals, and modelling studies. Moreover, the variability of meteorological parameters like temperature and precipitation in both, observations and climate models, is characterised by properties like non-Gaussian probability distribution functions, multifractality and long-term persistence.

Atmospheric patterns are characterised by spatio-temporal scales on which meteorological observables like temperature, air pressure or humidity vary only weakly. Measuring the temporal evolution of such parameters at different locations influenced by the same pattern, it is thus likely that the corresponding time series are more or less strongly correlated, with a maximum correlation occuring at a time lag which corresponds to the spatial distance between the sites and the typical drift velocity of the pattern. Due to the dynamic evolution of the observed patterns during their spatial motion, the correlations between meteorological and hydrological records decay with an increasing distance between the considered locations. This statement holds in general for very different spatial and temporal scales.

For complex systems consisting of several sub-systems, a suitable coupling of the components may lead to synchronisation, a very special kind of statistical dependence between the dynamics of the sub-systems [Pikovsky et al. 2001]. The concept of synchronisation originally describes a coupling-induced correlation between well-defined oscillatory dynamics in two systems, but may be generalised to systems in which a certain specific behaviour is observed from time to time (e.g., the so-called event synchronisation important in neurophysiology [Quian Quiroga et al. 2002]). In general, there are different types of synchronisation, including complete synchronisation, phase synchronisation, generalised synchronisation, or lag synchronisation. To test for synchronisation in real-world observational records, a sophisticated definition of a phase is required which is hardly possible for ver noisy, non-stationary or non-coherent data. Recently, [Romano et al. 2005] suggested to define synchronisation indices based on bivariate recurrence plots which are applicable to a broad class of data where standard methods of synchronisation analysis fail.

Although there are major conceptual differences between correlations and synchronisation (i.e., the presence of correlations does not necessarily mean synchronisation), both features are frequently identified with each other. In the case of time series from climatology, this identification is clearly wrong as observational records of different observables or at different stations reflect different aspects of the same spatially extended system, while it is misleading to understand the climate system to be composed of different well-defined and separated subsystems interacting which each other in a complex way. Nonetheless, [Rybski et al. 2003] proposed that for long-term daily temperature and precipitation records from different European stations, measures of phase synchronisation yield more significant results than the classical linear correlation function.

On a world-wide scale, the interrelationships between sea-level pressure records obtained from reanalysis data have been utilised to derive a network-like structure [Tsonis and Roebber 2004].

Similar features are likely to be found in simulations of climate models as well, however, the behaviour of such models is known to differ from reanalysis data not only in terms of absolute variabilities and correlations, but also with respect to their non-linear features like the local predictability [von Bloh et al. 2005]. On continental scales (i.e., several hundreds to thousands of kilometers), simple linear cross-correlation functions may (depending on the particular geographic situation) not necessarily be an optimal measure for describing the interrelationships and exactly detecting the delay corresponding to a maximum correlation between meteorological time series.

In order to test the results of [Rybski et al. 2003] for generality, the dynamics of daily minimum and maximum temperature data has been studied for records from the meteorological observatories at Armagh (Northern Ireland) [Butler et al. 2005] and Potsdam (Germany)[1] covering the time interval between 1900 and 1999. Some isolated missing data in the Armagh time series have been substituted by spline interpolates. The mean annual cycles over the considered time interval have been removed from both time series, which have been standardised to zero mean and unit variance afterwards to approach approximately stationary signals and allow a further comparison of the records.

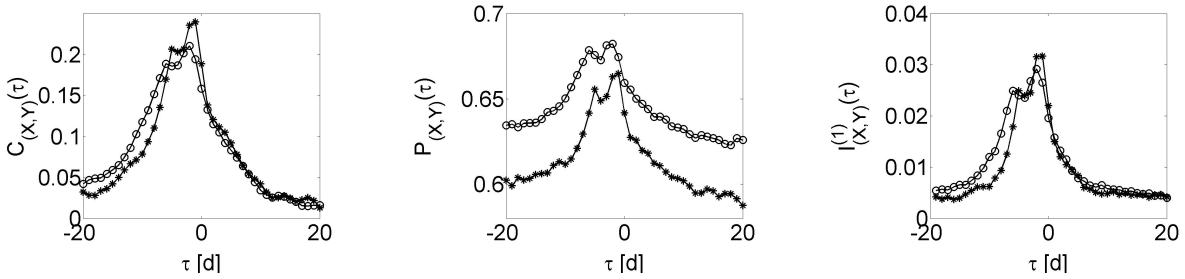


Figure 1.4: Linear cross-correlation (left), recurrence-based generalised cross-correlation (middle), and cross-mutual information (left) for 100-year records of daily maximum (circles) and minimum (asterisks) temperatures at Armagh and Potsdam.

Fig. 1.4 shows the results of correlation analysis with different methods described in the previous sections. In particular, linear cross-correlations, recurrence-based generalised cross-correlations and cross-mutual information give very clear evidence of existing correlations. There are, however, two interesting features that may be worth to be discussed further. Firstly, one observes that the maximum correlations occur at a time lag of two days in the case of daily maximum temperatures, whereas for the corresponding minimum values, a delay of only one day is found. This finding indicates a true time lag between one and two days, which however is hardly to be exactly detected even in records with a higher temporal resolution as the temperature fluctuations during one day are typically much larger than the differences between successive days. In addition, the different lags may have been supported by different observational strategies. As a second feature, one observes a double-peak structure of all considered functions indicating that there are significant correlations with an additional lag of four more days which possibly indicates a very pronounced time scale of atmospheric oscillations.

In order to test the applicability of phase synchronisation analysis to this kind of data, a phase variable has to be defined firstly for both time series considered. For a system with a well-

[1]These stations have been chosen because of the free availability of the corresponding time series from the websites of the Armagh meteorological observatory and the German Weather Service DWD, resp. In addition, both locations have a similar distance like Oxford and Vienna studied by [Rybski et al. 2003].

defined oscillatory dynamics, one may reconstruct an analytic signal [Rosenblum et al. 1996] by setting

$$S_X(t) = X(t) + i\tilde{X}(t) = A_X(t)e^{i\phi_X(t)} \tag{1.16}$$

where $\tilde{X}(t)$ is the Hilbert transform of the original time series defined as

$$\tilde{X}(t) = \frac{1}{\pi} P.V. \int_{-\infty}^{\infty} \frac{X(\tau)}{t-\tau} d\tau. \tag{1.17}$$

Hence, $\phi_X(t) = \arctan(\tilde{X}(t)/X(t))$ defines a proper phase for an oscillating system. However, as real-world data are rarely phase-coherent, an equivalent phase definition may be used based on time derivative of $S_X(t)$ instead [Osipov et al. 2003, Maraun and Kurths 2005].

Indices for phase synchronisation are based on the distribution of phase differences of the two time series lagged by a time shift τ, $\Delta\phi(t,\tau) = |\phi_Y(t+\tau) - \phi_X(t)|$ (in the case of $t > 0$), which are normalised to the interval $[0, 2\pi]$. If one considers a partition of this interval into M bins with $p_k(t)$ being the relative frequency of phase differences in the k-th bin, [Tass et al. 1998] to use the Shannon entropy computed from these probabilities,

$$S(\tau) = -\sum_k = 1^M p_k(\tau) \log p_k(\tau), \tag{1.18}$$

to define a synchronisation index as

$$\rho(\tau) = \frac{\log M - S(\tau)}{\log M}. \tag{1.19}$$

Alternative approaches consider the standard deviation of the normalised phase differences, $\sigma_{\delta\phi}(\tau)$, or the mean resultant length,

$$\lambda(\tau) = \frac{1}{N} \left| \sum_{t=1}^{N-|\tau|} e^{i\Delta\phi(t,\tau)} \right|. \tag{1.20}$$

[Paluš 1997] suggested to consider the mutual information between the two phases as a measure of statistical dependence indicating phase synchronisation in noisy systems. However, that all these approaches *depend crucially on the existence of well-defined phase variables* which is problematic in systems where a coherent oscillatory component is missing.

The results of the corresponding analysis for the temperature time series are shown in Fig. 1.5. One observes that for the standard phase definition of [Rosenblum et al. 1996], only the entropy-based index gives a very weak indication of phase coherence at a time lag of 1 to 2 days, whereas the remaining measures completely fail. However, the positive result is successively lost when a finer partition is applied to the normalised phase differences. Following the suggestions in [Pikovsky et al. 2001], the applied phase definition is only useful in the case of phase-coherent data with a narrow frequency spectrum which is not present for noisy meteorological data. Using the more generally applicable curvature method of [Osipov et al. 2003] instead, all three indices yield no evidence at all for phase synchronisation. These results do not mean that there is no synchronisation at all as the applied method is only sensitive to phase synchronisation as one particular kind of synchronisation. In contrast, there may still be intermittent epochs of enhanced phase coherence or even phase synchronisation which might be found by a separate analysis of different time intervals. For example, [Maraun and Kurths 2005] have recently reported evidence of such epochs in the relationship between the All-Indian Rainfall (AIR) and the

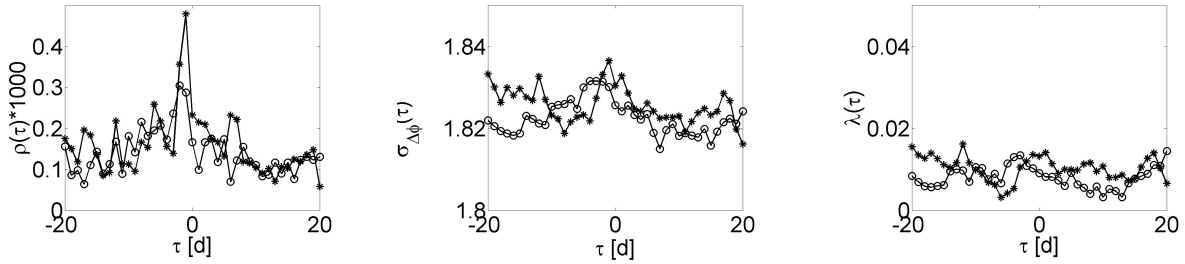


Figure 1.5: Phase synchronisation indices $\rho(\tau)$ (left, computed with $M = 20$ bins), σ (middle) and λ (right) for the daily maximum (circles) and minimum (asterisks) temperature time series from Armagh and Potsdam. All phases have been computed with the standard analytic signal approach of [Rosenblum et al. 1996].

NINO3 indices which manifest the known dependences between variations of the Indian monsoon system and the El Niño phenomenon [Webster and Yang 1992, Torrence and Webster 1999, Krishna Kumar et al. 1999, Sarkar et al. 2004].

Summarising, the application of phase synchronisation analysis is not appropriate in the case of noisy, non-coherent data which are typical in climatology. Whereas there is no evidence for long-term phase synchronisation in the considered temperature records, strong linear and nonlinear correlations occur with a time lag between one and two days. The strength and lag of correlations is explained by the fact that both locations are essentially influenced by the same major atmospheric circulation patterns (for example, both stations are situated in the preferred direction of North Atlantic storm tracks). In particular, the correlations are much stronger than between Oxford and Vienna (the examples studied by [Rybski et al. 2003]) because the latter station is additionally influenced by continental and mediterranean oscillation patterns.

Chapter 2

Statistical Modelling of Finite Mixture Distributions

2.1 Motivation

An important characteristic of a data set is its probability distribution function (PDF). This function can be estimated from the frequencies of possible values in a sufficiently long series of independent measurements of the associated observable. Hence, estimating the PDF of a given data set is a frequent task in data analysis which can be approached either nonparametrically [Silverman 1986] or parametrically, i.e., by estimating the parameters of a statistical model prescribed in terms of an appropriate distribution function.

In general, an exact determination of the distribution function from data requires an infinite amount of data and is practically not possible. On the one hand, data sets of observations typically contain a (possibly rather low) finite number of measurement points or observations due to restricted observation time and data storage capacity and thus yield only imperfect information. On the other hand, for a huge number of observations or individual measurements with very uncertain values, it is a typical approach to group the data into different classes and consider the resulting histogram of group frequencies for the PDF estimation as methods based on the explicit data become very inefficient.

In many situations, data from natural or industrial processes show a multimodal structure as components with different statistical properties contribute to the observed sample and are represented by subpopulations displayed as such more or less distinct modes in the PDF. A particular example which is extensively discussed in this thesis are object-size distributions which are frequently studied in many areas of research (some important applications are summarised in Sect. A.5). To assign the different modes in the PDF to the different subpopulations or subprocesses, it is an important problem to statistically decompose these components from observations in an appropriate way. In particular, one is interested in the statistical weights of the components as well as the shapes of the components themselves. For the latter purpose, different parametric as well as non-parametric modelling approaches can be performed.

In the case of parametric models, one has to simultaneously estimate the parameters and statistical weights of the probability distributions of the respective components if their total number and type is given. This approach has the advantage that the parameters may be closely related to certain physical models describing the generating process. In the following, a *K-finite mixture model* is defined by its PDF according to [Everitt and Hand 1981, Titterington et al. 1985,

McLachlan and Peel 2000] as

$$f(x;\vec{\pi},\mathbf{\Theta}) = \sum_{i=1}^{K} \pi_i f_i(x;\mathbf{\Theta}). \tag{2.1}$$

Here, $x \in X$ is the independent variable while $\mathbf{\Theta}$ is the vector of parameters of all subpopulations f_i, $i = 1,\ldots,K$. This vector is usually assumed to be unmixed, i.e., model parameters always influence the shape of only one particular component i auch that $f_i(x;\mathbf{\Theta}) = f_i(x;\mathbf{\Theta}_i)$ and $\mathbf{\Theta} = (\mathbf{\Theta}_1,\ldots,\mathbf{\Theta}_K)$. The respective types of subpopulations (i.e., the general forms of the functions $f_i(x;\mathbf{\Theta})$) are assumed to be known. For convenience, the statistical weights π_i with $\sum_{i=1}^{K} \pi_i = 1$ are combined in a statistical weight vector $\vec{\pi} = (\pi_1,\ldots,\pi_K)$ such that all unknown parameters of the total distribution can be written as $\mathbf{\Psi} = (\vec{\pi},\mathbf{\Theta})$. f and f_i are probability functions provided that

$$\int_X f(x;\mathbf{\Psi})dx = \int_X f_i(x;\mathbf{\Psi})dx = 1, \quad i = 1,\ldots,K. \tag{2.2}$$

As it is intensively discussed in Sect. 4.3, K-finite mixture models are important candidates for describing the multimodal shape of grain-size distributions (see, e.g., [Sun et al. 2002]), which motivates to study parameter estimation for such models in some more detail. A second important geoscientific application for this kind of statistical model is the analysis of isothermal remanent magnetisation (IRM) acquisition curves, which are studied to extract palaeomagnetic information from sedimentary sequences, but also yields secondary information about climatic influences on the magnetic properties of the deposits (in particular, different minerals have different magnetic properties and different grain-sizes, which sometimes relates both quantities [Potter et al. 2004]). IRM acquisition curves, describing the IRM in dependence on the amplitude of an applied external magnetic field, are structurally related to cumulative distribution functions, and their derivatives (being the equivalents of the PDFs) frequently show multimodality. To quantify and explain this multimodality in a sophisticated way, it has been proposed to model the curves by a finite mixture of lognormal distributions, which has been demonstrated to give a reasonable approximation of the observed data [Robertson and France 1994, Stockhausen 1998, Kruiver et al. 2001, Heslop et al. 2002].

The special importance of this type of statistical models in the analysis of geoscientific data motivates a more detailed study of finite mixtures in the following. For practical reasons and due to its special importance for the mentioned applications, only the case of (log-)normal components will be discussed, although many results are likely to be generalised to more complicated distribution functions. As for the case of grain-size distributions being a specific subject of palaeoclimatic studies, data are given in terms of group frequencies, a statistical framework is required which allows parameter estimation from grouped and possibly truncated data. Such a framework is described in the following in terms of the expectation-maximisation (EM) algorithm. In particular, the goodness-of-fit, the uncertainties of the estimated models, and possible problems relevant for applications in geoscientific research are intensively discussed.

2.2 The Expectation-Maximisation (EM) Algorithm

The expectation-maximisation algorithm [Dempster et al. 1977, McLachlan and Krishnan 1997] is a robust and relatively efficient method for parameter estimation of distribution functions (including the statistical separation of components in finite-mixture distributions) which provides an iterative computation of a maximum likelihood estimate of the considered statistical model.

In this chapter, the key features of this algorithm and its mechanism for explicit as well as grouped and truncated data are summarised. As a particular example, the case of Gaussian components is discussed in App. A.

2.2.1 Likelihood Functions and Maximum Likelihood Principle

Consider an ensemble of measured data $\vec{x} = (x_1, \ldots, x_J)$ given as single values $x_j \in X$ with J being the total number of observations in a given sample. Under the assumption of statistical independence of the respective observations, the joint probability of the given sample with respect to a prescribed parameter vector $\mathbf{\Psi}$ is given as

$$p(\vec{x}; \mathbf{\Psi}) = \prod_{j=1}^{J} f(x_j, \mathbf{\Psi}). \tag{2.3}$$

In the following, the symbol $p(\cdot; \vec{\Psi})$ is used for the joint probability of an ensemble of observations, whereas $f(\cdot; \vec{\Psi})$ refers to the underlying probability density function which may have been evaluated at certain discrete values of the respective observable.

It is possible to consider this quantity from a different point of view. Consider the parameter vector $\mathbf{\Psi}$ as the independent quantity. Then, the probability of any $\mathbf{\Psi}$ with respect to a given random vector $\vec{x}$ of observations can be established in terms of the likelihood function $L(\mathbf{\Psi}) = L(\mathbf{\Psi}; \vec{x})$ with respect to the given data sample. In general, a likelihood function is a function of $\mathbf{\Psi}$ calculated with respect to $\vec{x}$ that is (up to a constant factor independent of $\mathbf{\Psi}$ depending on the respective convention) equivalent to the joint probability of the single data x_j. In contrast to the latter one, $L(\mathbf{\Psi})$ is a function of the parameters for a fixed data sample (while for a probability density function, the parameters are considered fixed) and is therefore a random quantity like $\vec{x}$. Typically, equality of likelihood and probability function is assumed. Given statistical independence of the $\{x_j\}$, this joint probability factorises to the product of the respective single-data probabilities, i.e.,

$$L(\mathbf{\Psi}) = p(\vec{x}; \mathbf{\Psi}) = \prod_{j=1}^{J} f(x_j, \mathbf{\Psi}), \tag{2.4}$$

or, equivalently, its logarithm

$$\log L = \sum_{j=1}^{J} \log f(x_j, \mathbf{\Psi}) = \sum_{j=1}^{J} \log \sum_{i=1}^{K} \pi_i f_i(x_j, \mathbf{\Theta}) \tag{2.5}$$

in the case of a mixture distribution.

For the principle of ML estimation, constant prefactors are not important. The idea is to find the parameters which maximise L. This can be established by calculating the partial derivatives of $\log L$ with respect to all parameters and setting them to zero as

$$\frac{\partial \log L(\mathbf{\Psi})}{\partial \mathbf{\Psi}} = 0. \tag{2.6}$$

This formula is usually referred to as the likelihood equation. Its solution corresponds to the maximum likelihood estimate which is consistent, efficient and normal for $J \to \infty$ [Cramer 1946].

2.2.2 Expectation-Maximisation Algorithm: The Basic Idea

The EM algorithm allows an iterative computation of conditional probabilities leading to a maximum likelihood (ML) estimate of $\mathbf{\Psi}$. The basic procedure is best explained by considering samples of measured data. As such samples are incomplete, additional independent measurements of x would allow to improve the knowledge about the data distribution and therefore more reasonable estimates of the PDF. An EM algorithm implicitly considers these unknown data by formally defining complete data vectors $\vec{y} = (\vec{x}, \vec{x}')$ where $\vec{x}'$ contains the (in total, J') unknown data, referred to as the unobserved or missing data (in the following, primes will always indicate unobserved quantities).

The first step of the EM algorithm requires the estimation of some reasonable initial estimates for both parameters $\mathbf{\Theta}^{(0)}$ and statistical weights $\vec{\pi}^{(0)}$ from the given sample. Depending on the respective types of subpopulations, there is no general procedure for approaching this first rough estimation. The idea of the EM algorithm is then to iteratively calculate maximum likelihood estimates of the unknown parameters $\mathbf{\Psi}$. For this purpose, the complete-data likelihood function $L_c(\vec{y}, \mathbf{\Psi})$ is considered.

In the **expectation (E) step**, the expectation value of $\log L_c$ with respect to the initial parameter vector $\mathbf{\Psi}^{(0)}$ is calculated given the observed data $\vec{x}$ reading (for a continuous random variable $x \in X$) as follows:

$$\begin{aligned} Q(\mathbf{\Psi}; \mathbf{\Psi}^{(0)}) &= E_{\mathbf{\Psi}^{(0)}} \{\log L_c(\mathbf{\Psi}) | \vec{x}\} = \left(\int_X dx \; f(x; \mathbf{\Psi}^{(0)}) \; \log L_c(\mathbf{\Psi}) \middle| \vec{x} \right) \\ &= \sum_{j=1}^{J} \int_X dx \; f(x, \mathbf{\Psi}^{(0)}) \; \log f(x_j; \mathbf{\Psi}) = \sum_{j=1}^{J} \log f(x_j; \mathbf{\Psi}) \int_X dx \; f(x, \mathbf{\Psi}^{(0)}) \\ &= \sum_{j=1}^{J} \log f(x_j; \mathbf{\Psi}) = \log L(\mathbf{\Psi}). \end{aligned} \tag{2.7}$$

In the following **maximisation (M) step**, this expectation value is maximised with respect to the parameter vector $\mathbf{\Psi}$ which means chosing $\mathbf{\Psi}^{(1)}$ such that $Q(\mathbf{\Psi}^{(1)}, \mathbf{\Psi}^{(0)}) \geq Q(\mathbf{\Psi}, \mathbf{\Psi}^{(0)})$. The procedure of successive E- and M-steps is iteratively repeated (where the above superscripts $^{(0)}$ and $^{(1)}$ are replaced by $^{(l)}$ and $^{(l+1)}$, resp.) until the difference $L(\mathbf{\Psi}^{(l+1)}) - L(\mathbf{\Psi}^{(l)})$ changes only by values which are smaller than a certain predefined threshold value (which means a convergence of the likelihood function to its maximum). Under rather general conditions, it can be shown that for the observed data vector $\vec{x}$, the likelihood does not decrease after an EM iteration, i.e., $L(\mathbf{\Psi}^{(l+1)}) \geq L(\mathbf{\Psi}^{(l)})$ [Wu 1983].

2.2.3 Parameter Estimation in Finite Mixture Models

To practically implement the EM algorithm, one needs an explicit formulation of the respective E- and M-steps considering the initial information. A simple way of writing the respective iteration steps is the following one [Arcidiacono and Jones 2003]:

E-step: Given initial parameter values $\mathbf{\Psi}^{(l)}$, one computes the total probability of any observation x_j as

$$f(x_j; \mathbf{\Psi}^{(l)}) = \sum_{i=1}^{K} \pi_i^{(l)} f_i(x_j; \mathbf{\Theta}^{(l)}). \tag{2.8}$$

and the conditional probability of any observation x_j to belong to the subpopulation k as

$$t_k(x_j;\mathbf{\Psi}^{(l)}) = Pr(k|x_j;\mathbf{\Psi}^{(l)}) = \frac{\pi_k^{(l)} f_k(x_j;\mathbf{\Theta}^{(l)})}{f(x_j;\mathbf{\Psi}^{(l)})} = \frac{\pi_k^{(l)} f_k(x_j;\mathbf{\Theta}^{(l)})}{\sum_{i=1}^{K}\pi_i^{(l)} f_i(x_j;\mathbf{\Theta}^{(l)})}. \tag{2.9}$$

Note that, in this formulation, neither the complete-data likelihood, $L_c(\mathbf{\Psi})$, nor the observed-data likelihood, $L(\mathbf{\Psi})$, have to be computed explicitly.

M-step: The problem of calculating the maximum likelihood estimates of the statistical weights, $\vec{\pi}$, under the additional condition $\sum_{i=1}^{K}\pi_i = 1$ can be solved by applying the standard solution method for constrained extrema. For this purpose, one constructs a first-order Lagrange-type function

$$\mathcal{L} = \log L(\mathbf{\Psi}^{(l)}) + \lambda\left(1 - \sum_{i=1}^{K}\pi_i^{(l)}\right) \tag{2.10}$$

where λ is a Lagrangian multiplier combining the constraint with the log-likelihood function to be maximised. To approach the constrained maximum, the derivatives of $\mathcal{L}$ with respect to both π_k and λ have to be considered while the latter one is identical with the constraint. The derivative with respect to a particular π_k then reads:

$$\begin{aligned} 0 = \frac{\partial\mathcal{L}}{\partial\pi_k} &\propto \frac{\partial}{\partial\pi_k}\sum_{j=1}^{J}\log\sum_{i=1}^{K}\pi_i^{(l)} f_i(x_j,\mathbf{\Theta}^{(l)}) - \lambda = \sum_{j=1}^{J}\frac{\sum_{i=1}^{K}\frac{\partial\pi_i^{(l)}}{\partial\pi_k} f_i(x_j,\mathbf{\Theta}^{(l)})}{\sum_{i=1}^{K}\pi_i^{(l)} f_i(x_j;\mathbf{\Theta}^{(l)})} - \lambda \\ &= \sum_{j=1}^{J}\frac{f_k(x_j;\mathbf{\Theta}^{(l)})}{f(x_j,\mathbf{\Psi}^{(l)})} - \lambda = \frac{1}{\pi_k}\sum_{j=1}^{J} t_k(x_j;\mathbf{\Psi}^{(l)}) - \lambda. \end{aligned} \tag{2.11}$$

or

$$\sum_{j=1}^{J} t_k(x_j;\mathbf{\Psi}^{(l)}) = \lambda\pi_k^{(l+1)}. \tag{2.12}$$

Taking the sum over all $k = 1,\ldots,K$ in (2.12) and considering that for every x_j, the sum over all $t_k(x_j;\mathbf{\Psi})$ must be one due to the probability character of $t_i(x_j;\mathbf{\Psi})$, it follows that $\sum_{i=1}^{K}\sum_{j=1}^{J} t_i(x_j;\mathbf{\Psi}) = \lambda = J$. Finally, one ends up with

$$\boxed{\pi_k^{(l+1)} = \frac{1}{J}\sum_{j=1}^{J} t_k(x_j;\mathbf{\Psi}^{(l)}).} \tag{2.13}$$

Inserting this result into the derivative of $\log L$ with respect to θ_k, one finds the following condition for the maximum likelihood estimates:

$$\begin{aligned} 0 = \frac{\partial\log L}{\partial\theta_k} &\propto \sum_{j=1}^{J}\frac{\partial}{\partial\theta_k}\log\sum_{i=1}^{K}\left(\pi_i^{(l)} f_i(x_j;\mathbf{\Theta}^{(l)})\right) = \sum_{j=1}^{J}\frac{\sum_{i=1}^{K}\pi_i^{(l)}\frac{\partial}{\partial\theta_k} f_i(x_j,\mathbf{\Theta}^{(l)})}{\sum_{i=1}^{K}\pi_i^{(l)} f_i(x_j,\mathbf{\Theta}^{(l)})} \\ &= \sum_{j=1}^{J}\sum_{i=1}^{K}\frac{\pi_i^{(l)} f_i(x_j;\mathbf{\Theta}^{(l)})}{f(x_j;\mathbf{\Psi}^{(l)})}\frac{\partial}{\partial\theta_k}\log f_i(x_j;\mathbf{\Theta}^{(l)}) = \sum_{j=1}^{J}\sum_{i=1}^{K} t_i(x_j;\mathbf{\Psi}^{(l)})\frac{\partial\log f_i(x_j;\mathbf{\Theta}^{(l)})}{\partial\theta_k}. \end{aligned} \tag{2.14}$$

Hence, one can express the ML solution for the next iteration as

$$\boldsymbol{\Theta}^{(l+1)} = \arg\max_{\boldsymbol{\Theta}} \sum_{j=1}^{J} \sum_{i=1}^{K} t_i(x_j; \boldsymbol{\Psi}^{(l)}) \log f_i(x_j, \boldsymbol{\Theta}^{(l)}). \tag{2.15}$$

By iteratively solving of the set of equations (2.8,2.9,2.13,2.15), optimum ML estimates for the statistical weights and distribution parameters of the different mixture components can be calculated. Note that $Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(l)})$ has not to be evaluated explicitly in this formulation. Here, the E-step corresponds to the calculation of the conditional probabilities while the M-step includes (as usual) the recalculations of all parameters. The explicit form of the maximisation step in the case of Gaussian components is derived in App. A.1.

2.3 Parameter Estimation for Grouped and Truncated Data

The EM algorithm can be applied to a large variety of missing-data problems. In particular, it is not limited to data given in explicit form. For example, due to the particular strategy of measurement, the uncertainty of single measurements, or a large number of observations, data are often given in grouped form, i.e., in terms of histograms giving frequencies of observations falling into certain mutually exclusive classes (for an example, see Fig. 2.1). Already in [Dempster et al. 1977], the possibility of using the EM framework for parameter estimation in such situations has been intensively discussed. The details of the corresponding algorithm have been explicitly worked out by [McLachlan and Jones 1988, McLachlan and Krishnan 1997].

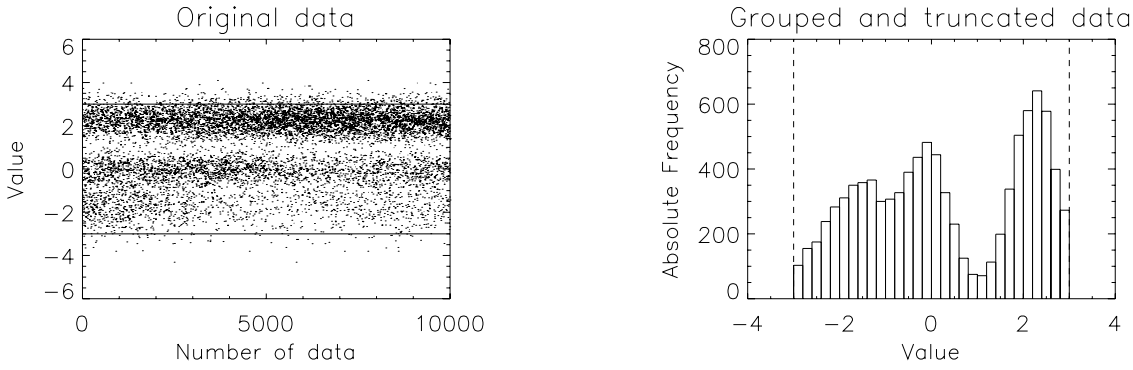


Figure 2.1: An example for a set of grouped and truncated data from a superposition of three Gaussian component distributions. Left: Original data (horizontal lines indicate the truncation values). Right: Data subjected to grouping and additional truncation of the lower- and uppermost values.

To consider grouped data, let the range of possible observations X (i.e., $X = \mathrm{supp} f(x; \boldsymbol{\Psi})$) contain M mutually excluding subsets X_m. The available information about the system is restricted to counts (absolute or relative frequencies) $\vec{n} = (n_1, \ldots, n_M)$ of events $x_j \in X$, $j = 1, \ldots, J$, where n_m is the number of counts with $x_j \in X_m$. The total number of observed counts is $n = \sum_{m=1}^{M} n_m$. For the derivation of the general expressions, one may first restrict to a single-component distribution function. The extension to finite mixtures is discussed later. For simplicity, one may assume a univariate distribution and intervals $X_m = [a_m, b_m]$ with $b_m = a_{m+1}$.

2.3.1 The Problem of Truncation

In practical applications, one is often confronted with the problem of truncated data. For example, values exceeding a certain threshold may not be detected by measurement devices, or the corresponding results are extremely uncertain or are influenced by problems during the measurement process such that the respective observations have to be removed. Simply neglecting these data in the estimation of the PDF may lead to serious falsifications of the expected parameters.

As an example, in the following the bias of the corresponding parameter values is explicitly discussed for the case of explicit observations following a Gaussian distribution. For $N \to \infty$, the empirical parameters μ and σ^2 calculated from the discrete data converge towards the integral expression of the corresponding expectation values of x and $(x-\mu)^2$, resp., as

$$\hat{\mu} = \int_{-\infty}^{\infty} dx\ x\ f(x;\mathbf{\Psi}) = \mu \int_{-\infty}^{\infty} dx\ f(x;\mathbf{\Psi}) - \sigma^2 \int_{-\infty}^{\infty} dx\ \frac{df}{dx}(x;\mathbf{\Psi}) \tag{2.16}$$

$$\hat{\sigma}^2 = \int_{-\infty}^{\infty} dx\ (x-\mu)^2\ f(x;\mathbf{\Psi}) = \sigma^4 \int_{-\infty}^{\infty} dx\ \frac{d^2 f}{dx^2}(x;\mathbf{\Psi}) + \sigma^2 \int_{-\infty}^{\infty} dx\ f(x;\mathbf{\Psi}) \tag{2.17}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ correspond to the parameters estimated from the data while μ and σ^2 reflect the exact parameters of the underlying distribution $f(x;\mathbf{\Psi})$. In the case of truncation, the integration ranges are restricted to the respective intervals. It is convenient to express the corresponding remaining integrals in the grouped-data formulation to approach the following equations:

$$\hat{\mu} = \mu \sum_{m=1}^{M} \int_{a_m}^{b_m} dx\ f(x;\mathbf{\Psi}) - \sigma^2 \sum_{m=1}^{M} \int_{a_m}^{b_m} dx\ \frac{df}{dx}(x;\mathbf{\Psi}) \tag{2.18}$$

$$\hat{\sigma}^2 = \sum_{m=1}^{M} \int_{a_m}^{b_m} dx\ f(x;\mathbf{\Psi}) + \sigma^4 \sum_{m=1}^{M} \int_{a_m}^{b_m} dx\ \frac{d^2 f}{dx^2}(x;\mathbf{\Psi}). \tag{2.19}$$

For explicitly expressing the occuring integrals, one may again use the respective identities for the normal distributions as

$$\sum_{m=1}^{M} \int_{a_m}^{b_m} dx\ f(x;\mathbf{\Psi}) = \sum_{m=1}^{M} P_m(\mathbf{\Psi}) = P(\mathbf{\Psi}) = 1 - P_{tr}(\mathbf{\Psi}) \tag{2.20}$$

$$\sum_{m=1}^{M} \int_{a_m}^{b_m} dx\ \frac{df}{dx}(x;\mathbf{\Psi}) = \sum_{m=1}^{M} \Delta_m f \tag{2.21}$$

$$\begin{aligned} \sum_{m=1}^{M} \int_{a_m}^{b_m} dx\ \frac{d^2 f}{dx^2}(x;\mathbf{\Psi}) &= \sum_{m=1}^{M} \frac{df}{dx}\bigg|_{a_m}^{b_m} = -\sum_{m=1}^{M} \frac{x-\mu}{\sigma^2} f\bigg|_{a_m}^{b_m} \\ &= \frac{\mu}{\sigma^2} \sum_{m=1}^{M} \Delta_m f - \frac{1}{\sigma^2} \sum_{m=1}^{M} \Delta_m \phi \end{aligned} \tag{2.22}$$

with

$$\Delta_m f^{(l)} = f(b_m;\mathbf{\Psi}^{(l)}) - f(a_m;\mathbf{\Psi}^{(l)}) \tag{2.23}$$

$$\Delta_m \phi^{(l)} = b_m f(b_m;\mathbf{\Psi}^{(l)}) - a_m f(a_m;\mathbf{\Psi}^{(l)}) \tag{2.24}$$

to finally end up with the following estimates:

$$\hat{\mu} = \mu\left(1 - P_{tr}(\mu,\sigma^2)\right) - \sigma^2 \sum_{m=1}^{M} \Delta_m f(\mu,\sigma^2) \tag{2.25}$$

$$\hat{\sigma}^2 = \sigma^2\left(1 - P_{tr}(\mu,\sigma^2)\right) + \sigma^2\mu \sum_{m=1}^{M} \Delta_m f(\mu,\sigma^2) - \sigma^2 \sum_{m=1}^{M} \Delta_m \phi(\mu,\sigma^2). \tag{2.26}$$

In the special case of a symmetrically double-side truncated standard $\mathcal{N}(0,1)$ distribution, these equations simplify to

$$\hat{\mu} = 0 \tag{2.27}$$

$$\hat{\sigma}^2 = 1 - P_{tr} - 2x_c f(x_c). \tag{2.28}$$

For such a distribution, the behaviour of the estimated variance, $\hat{\sigma^2}$, in dependence on the truncated probability, P_{tr}, or, equivalently, the point of truncation, x_c, is shown in Fig. 2.2. In this example, the $\mathcal{N}(0,1)$ was realised by a sample of only $n = 1000$ explicit data which explains the significant deviations from the derived expression for large truncations as the uncertainty of the estimated parameters increases with $1/\sqrt{n}$ (see Sect. 2.4). Moreover, in the limit of a small number of observations, n, the discrete summation for the calculation of σ^2 may not be replaced by an integration any more. In general, the parameters estimated from a short realisation may not sufficiently resolve the actual distribution in the presence of large truncations.

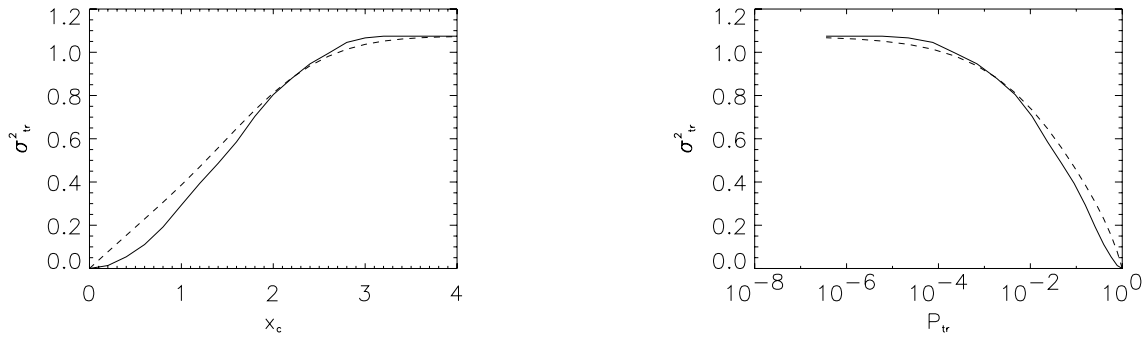


Figure 2.2: Values of the expected variance σ^2 estimated from a symmetrically truncated $\mathcal{N}(0,1)$ distribution by the explicit approximate equation (2.26) (dashed line) and direct calculation from the explicit data (solid line). The variance is shown depending on the truncation cutoff x_c (left panel) and the corresponding truncated probability P_{tr} (right panel).

2.3.2 Likelihood Functions for Grouped Non-Truncated Data

To formulate the EM algorithm for grouped and truncated data, the appropriate formulation of the corresponding likelihood function has to be considered firstly. For this purpose, it is convenient to start with the non-truncated case before discussing the modifications due to the presence of truncation.

Probability and Likelihood of the Observed Data. Under the assumption that the single events x_j are statistically independent, one may consider the observed group frequencies $\vec{n}$ to be taken from a multinomial distribution gained from n draws out of M categories. The relative probability of any category X_m is given by

$$P_m(\mathbf{\Psi}) = \int_{X_m} dx \; f(x; \mathbf{\Psi}). \tag{2.29}$$

For a multinomial distribution with the above relative frequencies, the joint probability of the observed data $\vec{n}$ is given by

$$p(\vec{n}; \mathbf{\Psi}) = \left(\frac{n!}{\prod_{m=1}^{M} n_m!} \right) \prod_{m=1}^{M} [P_m(\mathbf{\Psi})]^{n_m}. \tag{2.30}$$

This function actually defines a probability because $\sum_{m=1}^{M} P_m(\mathbf{\Psi}) = 1$. Identifying this probability with the observed-data likelihood function yields [Hartley 1971]

$$\log L = \sum_{m=1}^{M} n_m \log P_m(\mathbf{\Psi}) + \log \frac{n!}{\prod_{m=1}^{M} n_m!}. \tag{2.31}$$

Probability and Likelihood of the Complete Data. Unlike the grouped data combined in the observed-data vector $\vec{n}$, the explicit values of observations are unknown. To formulate the estimation problem based on grouped data within the EM framework, these data are considered as a vector $\vec{x}'_m = (x'_{m1}, \ldots, x'_{mn_m})$ for $m = 1, \ldots, M$. Here, each $\vec{x}'_m$ contains n_m independent observations of $x \in X_m$ with the probability density

$$p_m(x; \mathbf{\Psi}) = \begin{cases} \frac{f(x;\mathbf{\Psi})}{P_m(\mathbf{\Psi})}, & x \in X_m \\ 0, & \text{else.} \end{cases} \tag{2.32}$$

Explicitly including the unknown values $\{x_{mj}\}$ of single observations, the complete-data vector has the form $\vec{y} = (\vec{n}^T, \vec{n}'^T, \vec{x}_1'^T, \ldots, \vec{x}_M'^T)$. It follows that the corresponding complete-data log-likelihood has the form

$$\log L_c(\mathbf{\Psi}) \propto \sum_{m=1}^{M} \sum_{j=1}^{n_m} \log f(x'_{mj}; \mathbf{\Psi}). \tag{2.33}$$

In addition, one has to consider the total probability of the explicit data $p(\{x'_{mj}\}; \mathbf{\Psi})$ as being conditional with respect to $\vec{n}$ such that

$$p(\vec{y}; \mathbf{\Psi}) = p(\vec{n}, \vec{x}'_1, \ldots, \vec{x}'_M; \mathbf{\Psi}) = p(\{x'_{mj}\} | \vec{n}; \mathbf{\Psi}) \cdot p(\vec{n}; \mathbf{\Psi}). \tag{2.34}$$

Given the respective measurements $\{x_{mj}\}$ being statistically independent, the first factor can be further evaluated such that

$$p(\{x'_{mj}\} | \vec{n}; \mathbf{\Psi}) = \prod_{m=1}^{M} \prod_{j=1}^{n_m} p_m(x'_{mj}; \mathbf{\Psi}) = \prod_{m=1}^{M} \prod_{j=1}^{n_m} \frac{f(x'_{mj}; \mathbf{\Psi})}{P_m(\mathbf{\Psi})} \tag{2.35}$$

which implies

$$L_c(\mathbf{\Psi}) = p(\vec{y}; \mathbf{\Psi}) = \prod_{m=1}^{M} \prod_{j=1}^{n_m} \frac{f(x'_{mj}; \mathbf{\Psi})}{P_m(\mathbf{\Psi})} \cdot p(\vec{n}; \mathbf{\Psi}). \tag{2.36}$$

2.3.3 Likelihood Functions for Grouped Truncated Data

Probability and Likelihood of the Observed Data. For the observed data, one may proceed similar to the non-truncated case. However, as $P(\mathbf{\Psi}) := \sum_{m=1}^{M} P_m(\mathbf{\Psi}) < 1$, one has to substitute $P_m(\mathbf{\Psi})$ by $P_m(\mathbf{\Psi})/P(\mathbf{\Psi})$ in all corresponding considerations (in particular, in Eqs. (2.30) and (2.33)) to approach a proper probability distribution function. This leads to the following expression:

$$\begin{aligned} \log L(\mathbf{\Psi}) = p(\vec{n};\mathbf{\Psi}) &= \sum_{m=1}^{M} n_m \log \frac{P_m(\mathbf{\Psi})}{P(\mathbf{\Psi})} + \log \frac{n!}{\prod_{m=1}^{M} n_m!} \\ &= \sum_{m=1}^{M} n_m \log P_m(\mathbf{\Psi}) - n \log P(\mathbf{\Psi}) + \log \frac{n!}{\prod_{m=1}^{M} n_m!}. \end{aligned} \tag{2.37}$$

To illustrate the correctness of this approach, Fig. 2.3 shows the values of the log-likelihood function for a $\mathcal{N}(0,1)$ distribution in dependence on the two parameters μ and σ for grouped, non-truncated data according to Eq. (2.33), and for grouped truncated data according to Eq. (2.37). One clearly observes that both functions take their maximum values at the desired point in the parameter space. However, as it can be inferred from the figure, it is rather likely to overestimate the variance σ in the case of truncated distributions.

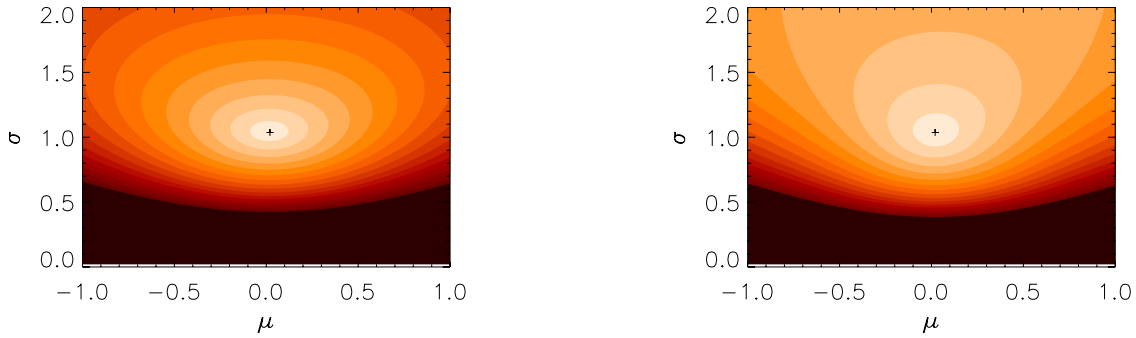


Figure 2.3: Color-coded representation of the observed-data log-likelihood function from completely grouped (left panel) and grouped truncated data (right panel) from a $\mathcal{N}(0,1)$ distribution using $n = 1000$ observations. The data have been truncated at $x_c = \pm 2.0$ and grouped into 20 equally sized bins within this interval and the tail intervals. The + corresponds to the sample mean and standard deviation estimated from the explicit data.

Probability and Likelihood of the Complete Data. As the neglection of missing groups leads to a systematic and significant misestimation of the distribution parameters for both direct and EM estimates, it is possible to explicitly consider these unknown groups included in the complete-data set within the EM framework. For this purpose, one may consider the possible range of observations X to be *completely* covered by $M + M'$ mutually exclusive subsets where M is again the number of subsets with observed group frequencies while M' is the number of intervals with unknown counts. The corresponding additional group frequencies $\vec{n}' = (n_{M+1}, \ldots, n_{M+M'})$ sum up to a total number of $n' = \sum_{m=M+1}^{M+M'} n_m$ events which are not observed. In an expectation-maximisation algorithm, the complete-data vector contains the exact values belonging to all observed as well as truncated counts explicitly included in data vectors

$\vec{x}'_m = (x'_{m1}, \ldots, x'_{mn_m})$ for $m = 1, \ldots, M + M'$ where $\vec{x}'_m$ contains n_m independent observations of $x \in X_m$ with a probability density according to Eq. (2.32) such that the complete data read $\vec{y} = (\vec{n}^T, \vec{n}'^T, \vec{x}'^T_1, \ldots, \vec{x}'^T_{M+M'})$. It follows that the corresponding complete-data log-likelihood has the form

$$\log L_c(\mathbf{\Psi}) = \sum_{m=1}^{M+M'} \sum_{j=1}^{n_m} \log f(x'_{mj}; \mathbf{\Psi}). \tag{2.38}$$

Considering again the total probability of the formal single-data observations $p(\{x'_{mj}\}; \mathbf{\Psi})$ as being conditional with respect to $\vec{n}$ and $\vec{n}'$ yields

$$p(\vec{y}; \mathbf{\Psi}) = p(\vec{n}, \vec{n}', \vec{x}'_1, \ldots, \vec{x}'_{M+M'}; \mathbf{\Psi}) = p(\{x'_{mj}\} | \vec{n}, \vec{n}'; \mathbf{\Psi}) \cdot p(\vec{n}' | \vec{n}; \mathbf{\Psi}) \cdot p(\vec{n}; \mathbf{\Psi}). \tag{2.39}$$

Under the usual assumption of statistical independence of the $\{x_{mj}\}$, the first factor becomes

$$p(\{x'_{mj}\} | \vec{n}, \vec{n}'; \mathbf{\Psi}) = \prod_{m=1}^{M+M'} \prod_{j=1}^{n_m} p_m(x'_{mj}; \mathbf{\Psi}) = \prod_{m=1}^{M+M'} \prod_{j=1}^{n_m} \frac{f(x'_{mj}; \mathbf{\Psi})}{P_m(\mathbf{\Psi})} \tag{2.40}$$

such that

$$L_c(\mathbf{\Psi}) = p(\vec{y}; \mathbf{\Psi}) = \prod_{m=1}^{M+M'} \prod_{j=1}^{n_m} \frac{f(x'_{mj}; \mathbf{\Psi})}{P_m(\mathbf{\Psi})} \cdot p(\vec{n}' | \vec{n}; \mathbf{\Psi}) \cdot p(\vec{n}; \mathbf{\Psi}). \tag{2.41}$$

Up to here, the conditional probability $p(\vec{n}' | \vec{n}; \mathbf{\Psi})$ is still unknown but crucial for the explicit formulation of the EM algorithm. To further evaluate this conditional probability, one may consider

$$\begin{aligned}
\log L_c(\mathbf{\Psi}) &= \log p(\vec{n}; \mathbf{\Psi}) + \log p(\vec{n}' | \vec{n}; \mathbf{\Psi}) + \sum_{m=1}^{M+M'} \sum_{j=1}^{n_m} \log \frac{f(x'_{mj}; \mathbf{\Psi})}{P_m(\mathbf{\Psi})} \\
&= \sum_{m=1}^{M} n_m \log \frac{P_m(\mathbf{\Psi})}{P(\mathbf{\Psi})} + \log \frac{n!}{\prod_{m=1}^{M} n_m!} \\
&\quad + \sum_{m=1}^{M+M'} \sum_{j=1}^{n_m} \log \frac{f(x'_{mj}; \mathbf{\Psi})}{P_m(\mathbf{\Psi})} + \log p(\vec{n}' | \vec{n}; \mathbf{\Psi}) \\
&= \sum_{m=1}^{M+M'} \sum_{j=1}^{n_m} \log f(x'_{mj}; \mathbf{\Psi}) + \log p(\vec{n}' | \vec{n}; \mathbf{\Psi}) + \log \frac{n!}{\prod_{m=1}^{M} n_m!} \\
&\quad + \sum_{m=1}^{M} n_m \log P_m(\mathbf{\Psi}) - \sum_{m=1}^{M} n_m \log P(\mathbf{\Psi}) \\
&\quad - \sum_{m=1}^{M} n_m \log P_m(\mathbf{\Psi}) - \sum_{m=M+1}^{M+M'} n_m \log P_m(\mathbf{\Psi}).
\end{aligned} \tag{2.42}$$

Comparing this expression with Eq. (2.38), it follows that

$$p(\vec{n}' | \vec{n}; \mathbf{\Psi}) \propto \frac{\prod_{m=1}^{M} n_m!}{n!} [P(\mathbf{\Psi})]^n \prod_{m=M+1}^{M+M'} [P_m(\mathbf{\Psi})]^{n_m} \tag{2.43}$$

To approach a proper probability distribution function, that this expression has to be renormalised as follows [McLachlan and Krishnan 1997]:

$$p(\vec{n}'|\vec{n};\mathbf{\Psi}) = \frac{(n'+n-1)!}{(n-1)!}[P(\mathbf{\Psi})]^n \prod_{m=M+1}^{M+M'} \frac{[P_m(\mathbf{\Psi})]^{n_m}}{n_m!}. \tag{2.44}$$

Comparing this result with Eq. (4.2.3) in [Dempster et al. 1977] (where all truncated intervals have been summarised in one group), this functional form may be identified as the result of a negative multinomial density.

2.3.4 The EM Algorithm for Grouped Truncated Data

In general, one has to firstly consider the expectation of the complete-data log-likelihood function which is given for grouped data by

$$\begin{aligned} Q(\mathbf{\Psi};\mathbf{\Psi}^{(l)}) &= E_{\mathbf{\Psi}^{(l)}}\left\{\log L_c(\mathbf{\Psi})|\vec{n}\right\} = E_{\mathbf{\Psi}^{(l)}}\left\{ \sum_{m=1}^{M+M'} \sum_{j=1}^{n_m} \log f(x'_{mj};\mathbf{\Psi}) \middle| \vec{n} \right\} \\ &= \sum_{m=1}^{M+M'} E_{\mathbf{\Psi}^{(l)}}\left\{n_m(\mathbf{\Psi})|\,\vec{n}\right\} E_{\mathbf{\Psi}^{(l)}}\left\{\log f(x;\mathbf{\Psi})|\,x\in X_m\right\}. \end{aligned} \tag{2.45}$$

If the observations are truncated, the expectation step splits up into separate calculations of the expectations of the unknown group frequencies and of the log-likelihood function itself.

Expected Group Frequencies. In the case of non-truncated data, the expected probability in any group is given by $P_m = n_m/n$, i.e., $n_m = nP_m$. Transferring this idea to the case of truncated data, it follows that the expected unknown group frequencies (given a parameter vector $\mathbf{\Psi}$ and observations $\vec{n}$) must have the corresponding form where P_m has to be replaced by $P_m(\mathbf{\Psi})/P(\mathbf{\Psi})$:

$$n_m(\mathbf{\Psi}^{(l)}) = E_{\mathbf{\Psi}^{(l)}}\left\{n_m(\mathbf{\Psi})|\,\vec{n}\right\} = \begin{cases} n_m & , \quad m = 1,\ldots,M \\ n\frac{P_m(\mathbf{\Psi}^{(l)})}{P(\mathbf{\Psi}^{(l)})} & , \quad m = M+1,\ldots,M+M' \end{cases}. \tag{2.46}$$

Expected Complete Data Log-Likelihood. Eq. (2.45) is additive with respect to the different groups and may therefore be written as

$$Q(\mathbf{\Psi};\mathbf{\Psi}^{(l)}) = \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)})\; Q_m(\mathbf{\Psi};\mathbf{\Psi}^{(l)}) \tag{2.47}$$

with

$$Q_m(\mathbf{\Psi};\mathbf{\Psi}^{(l)}) = E_{\mathbf{\Psi}^{(l)}}\left\{\log f(x;\mathbf{\Psi})|x\in X_m\right\}. \tag{2.48}$$

The expectation values of any function $g(x)$ of the random variable x with respect to a given parameter estimate $\mathbf{\Psi}^{(l)}$ for a given interval X_m can be calculated as follows:

$$\begin{aligned} E_{\mathbf{\Psi}^{(l)}}\left\{g(x)|x\in X_m\right\} &= \int_{X_m} dx\; p_m(x;\mathbf{\Psi}^{(l)})\; g(x) = \frac{1}{P_m(\mathbf{\Psi}^{(l)})}\int_{X_m} dx\; f(x;\mathbf{\Psi}^{(l)})\; g(x) \\ &= \frac{\int_{X_m} dx\; f(x;\mathbf{\Psi}^{(l)})\; g(x)}{\int_{X_m} dx\; f(x;\mathbf{\Psi}^{(l)})}. \end{aligned} \tag{2.49}$$

This yields the following explicit expression for the expectation step:

$$Q(\mathbf{\Psi};\mathbf{\Psi}^{(l)}) = \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) \frac{\int_{X_m} dx\ f(x;\mathbf{\Psi}^{(l)})\ \log f(x;\mathbf{\Psi})}{\int_{X_m} dx\ f(x;\mathbf{\Psi}^{(l)})}. \tag{2.50}$$

Maximisation Step. With the above results, for the $(l+1)$th iteration, the solution $\mathbf{\Psi}^{(l+1)}$ is obtained by computing the solution of

$$0 = \frac{\partial Q(\mathbf{\Psi};\mathbf{\Psi}^{(l)})}{\partial \mathbf{\Psi}} = \sum_{m=1}^{M} n_m(\mathbf{\Psi}^{(l)}) \frac{\partial Q_m(\mathbf{\Psi};\mathbf{\Psi}^{(l)})}{\partial \mathbf{\Psi}} = \sum_{m=1}^{M} n_m(\mathbf{\Psi}^{(l)}) \frac{\int_{X_m} dx\ f(x;\mathbf{\Psi}^{(l)}) \frac{\partial}{\partial \mathbf{\Psi}} \log f(x;\mathbf{\Psi})}{\int_{X_m} dx\ f(x;\mathbf{\Psi}^{(l)})}. \tag{2.51}$$

This expression has to be evaluated separately for each type of distribution. In App. A.2, the explicit equations for E- and M-step of a Gaussian distribution are derived analytically. Note that, for general distribution function, no analytical expression for the parameter estimates exist such that numerical iteration procedures have to be used in applications.

2.3.5 Parameter Estimation in Finite Mixture Models

As in the standard EM algorithm for explicitly given data, the theoretical framework of the calculus for grouped data can be extended to the case of finite mixture distributions. This offers a broad range of applications from clustering to parameter estimation in models based on binned data.

Expectation Step. The implementation of the EM algorithm for grouped data from finite mixture distributions requires again the calculation of conditional probabilities for any (unknown) observation x'_{mj} to belong to the subpopulation i. Because these relative probabilities cannot be calculated explicitly, one introduces zero-one component indicator variables $\vec{z}'_{mj} = (z'_{1mj}, \ldots, z'_{Kmj})$ with $m = 1, \ldots, M$ and $j = 1, \ldots, n_m$ with the properties

$$z'_{imj} = \begin{cases} 1 & \text{if } x'_{mj} \text{ belongs to } f_i \\ 0 & \text{else} \end{cases} \tag{2.52}$$

and $\sum_{i=1}^{K} z'_{imj} = 1\ \forall m, j$ [McLachlan and Jones 1988, McLachlan and Peel 2000]. Given the x'_{mj}, the $\vec{z}'_{mj}$ are conditionally independent with conditional probabilities

$$t_i(x'_{mj};\mathbf{\Psi}) = Pr(z'_{imj} = 1 | x'_{mj}) = \frac{\pi_i f_i(x'_{mj};\mathbf{\Theta})}{f(x'_{mj},\mathbf{\Psi})}. \tag{2.53}$$

In this case, it follows that the log-likelihood of the complete data can be reformulated as

$$\begin{aligned} \log L_c(\mathbf{\Psi}) &= \sum_{m=1}^{M+M'} \sum_{j=1}^{n_m} \log f(x'_{mj};\mathbf{\Psi}) = \sum_{m=1}^{M+M'} \sum_{j=1}^{n_m} \log \sum_{i=1}^{K} z'_{imj} \pi_i f_i(x'_{mj};\mathbf{\Theta}) \\ &= \sum_{i=1}^{K} \sum_{m=1}^{M+M'} \sum_{j=1}^{n_m} z'_{imj} \left[\log \pi_i + \log f_i(x'_{mj};\mathbf{\Theta})\right], \end{aligned} \tag{2.54}$$

where the last step follows due to the zero-one character of the z'_{imj}. Including the indicator variables into the equation for the expectation values, the expression for the $Q_m(\mathbf{\Psi};\mathbf{\Psi}^{(l)})$ reads

as follows:

$$Q_m(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(l)}) = \sum_{i=1}^{K} E_{\boldsymbol{\Psi}^{(l)}}\left\{ t_i(\{x'_{mj}\},\boldsymbol{\Psi}^{(l)}) \left(\log f_i(\{x'_{mj}\};\boldsymbol{\Theta}) + \log\pi_i\right) \middle| x'_{mj} \in X_m \right\}. \quad (2.55)$$

Maximisation Step. In the case of a mixture distribution, the particular M-step for the iterative estimation of the statistical weights is independent of the actual type of subpopulations. Again, there is a constraint that the sum of all weights must be equal to one. Considering this fact, one has to evaluate the derivative of the complete log-likelihood function's expectational value extended by the constraint

$$\mathcal{L} = Q(\boldsymbol{\Psi},\boldsymbol{\Psi}^{(l)}) + \lambda\left(1 - \sum_{i=1}^{K}\pi_i\right) \quad (2.56)$$

with respect to the different π_k as follows:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\pi_k} = \frac{\partial}{\partial\pi_k}\sum_{i=1}^{K}\sum_{m=1}^{M+M'} n_m(\boldsymbol{\Psi}^{(l)}) E_m^{(l)}\left\{ t_i(\{x'_{mj}\};\boldsymbol{\Psi}^{(l)})\left[\log f_i(\{x'_{mj}\};\boldsymbol{\Theta}) + \log\pi_i\right]\right\} \\ + \frac{\partial}{\partial\pi_k}\left[\lambda\left(1-\sum_{i=1}^{K}\pi_i\right)\right] = \sum_{m=1}^{M+M'} n_m(\boldsymbol{\Psi}^{(l)}) E_m^{(l)}\left\{ t_k(\{x'_{mj}\};\boldsymbol{\Psi}^{(l)})\frac{1}{\pi_k}\right\} - \lambda = 0\end{aligned} \quad (2.57)$$

with $E_m^{(l)}\{\cdot\} = E_{\boldsymbol{\Psi}^{(l)}}\{\cdot | x \in X_m\}$ such that

$$\lambda\pi_i^{(l+1)} = \sum_{m=1}^{M+M'} n_m(\boldsymbol{\Psi}^{(l)}) E_m^{(l)}\left\{ t_i(\{x'_{mj}\};\boldsymbol{\Psi}^{(l)})\right\}. \quad (2.58)$$

Summing up over all $i = 1,\ldots,K$ gives

$$\lambda = \lambda\sum_{i=1}^{K}\pi_i^{(l+1)} = \sum_{m=1}^{M+M'} n_m(\boldsymbol{\Psi}^{(l)}) E_m^{(l)}\left\{\sum_{i=1}^{K} t_i(\{x'_{mj}\};\boldsymbol{\Psi}^{(l)})\right\} = \sum_{m=1}^{M+M'} n_m(\boldsymbol{\Psi}^{(l)}) = n + n'(\boldsymbol{\Psi}^{(l)}). \quad (2.59)$$

Hence, the next-step ML estimate for the statistical weights of the mixture is given by

$$\pi_i^{(l+1)} = \frac{1}{n}\sum_{m=1}^{M+M'} n_m(\boldsymbol{\Psi}^{(l)}) E_m^{(l)}\left\{ t_i(\{x'_{mj}\};\boldsymbol{\Psi}^{(l)})\right\}. \quad (2.60)$$

Note that this particular result is actually valid independent of the type of mixture components. In contrast to this finding, the M-step for the parameter estimates must obviously depend on the subpopulation structure in terms of an explicit solution of

$$0 = \frac{\partial Q(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(l)})}{\partial\boldsymbol{\Theta}} = \sum_{i=1}^{K}\sum_{m=1}^{M+M'} n_m(\boldsymbol{\Psi}^{(l)}) E_{\boldsymbol{\Psi}^{(l)}}\left\{ t_i(\{x'_{mj}\};\boldsymbol{\Psi}^{(l)})\frac{\partial}{\partial\boldsymbol{\Theta}}\log f_i(\{x'_{mj}\};\boldsymbol{\Theta}) | x'_{mj}\in X_m\right\} \quad (2.61)$$

Eqs. (2.59) and (2.61) may be combined in the following general equation for the maximisation step:

$$0 = \frac{\partial Q(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(0)})}{\partial\Psi_k} = \sum_{m=1}^{M+M'} n_m(\boldsymbol{\Psi}^{(l)}) \int_{X_m} dx\ f(x;\boldsymbol{\Psi}^{(l)})\ t_k(x;\boldsymbol{\Psi}^{(l)})\ \frac{\partial}{\partial\Psi_k}\log f(x;\boldsymbol{\Psi}) \quad (2.62)$$

In App. A.3, the explicit solution of this equation is derived for Gaussian mixture models.

2.3.6 Related and Concurring Approaches

While the standard EM algorithm offers a robust method to simultaneously cluster given data and estimate the parameters of the corresponding probability density distributions, its relatively slow convergence triggered research to derive alternative approaches for both particular problems. While there is a large number of results concerning the problem with explicitly given data, the estimation based on grouped data has attracted attention very recently. As examples, some particular approaches concurring to the standard EM-based method may be mentioned:

[Wu and Perloff 2003] consider the application of a maximum entropy method to successively estimate the moments of an unspecified distribution function based on grouped data resulting in a nonlinear least squares estimator. However, by construction this approach is not necessarily appropriate for simultaneous clustering of data. Consequently, [Wu and Perloff 2003] used the method for estimating single-component densities only, in particular, for income distributions [Wu 2003]. The combination with a separate clustering algorithm might allow the estimating of finite mixture distributions as well but probably requires larger computational efforts than in the EM case. As an advantage with respect to the EM algorithm, [Wu and Perloff 2003] show that due to its intrinsically non-parametric nature, the maximum entropy method may be applied as well to the case of grouped data with unknown interval limits when combining it with a sequential EM-type algorithm [Arcidiacono and Jones 2003].

[Lam and Ip 2003] used standard and residual maximum-likelihood approaches to clustering and parameter estimation in logistic and grouped proportional regression models based on binned survival time data.

[Samé and Govaert 2002, Samé et al. 2003] extended the framework of the EM algorithm applied to grouped data by combining it with a classification EM (CEM) algorithm for clustering of data. The power of the resulting EM-CEM algorithm has been demonstrated for some numerical examples of mixtures of two bi-dimensional Gaussian distributions.

2.4 Estimation of Parameter Uncertainty

If the distribution function of observations is described by means of a parametric model, the knowledge of not only of the parameters themselves, but also of their corresponding statistical uncertainties and their relationship to model validity is of particular interest. Usually, the uncertainty is quantified in terms of so-called standard errors which can be estimated either by methods based on the information matrix of the estimator or by application of resampling techniques. In the following, the realisation of both approaches and their possible disadvantages are discussed for the EM algorithm.

2.4.1 Information-based Standard Errors

To calculate standard errors for the estimated parameters, a wide spread approach is approximating the corresponding covariance matrix by the inverse of either the observed or the expected information matrix which may be additionally used to improve the convergence of the algorithm [Louis 1982, Meilijson 1989, Meng and Rubin 1991, Jones and McLachlan 1992, Lange 1995, Jamshidian and Jennrich 1997]. The appropriate choice and the approximation or estimation of these information matrices have been intensively studied [Berndt et al. 1974, Efron and Hinkley 1978, Redner and Walker 1984, Griffiths et al. 1987, Dolan and Molenaar 1991, Baker 1992, Oakes 1999, Jamshidian and Jennrich 2000]. In particular, applications to grouped data have been discussed [Hartley 1971,

Jones and McLachlan 1992]. The information matrix method has been applied to different estimation problems, e.g., for quantifying the uncertainty of component means [Behboodian 1972, Basford et al. 1997] and weights [Hill 1963] in Gaussian mixture models.

Following [McLachlan and Krishnan 1997], one defines the score vector $\vec{s}(\mathbf{\Psi})$ as the gradient, and the information matrix $I(\mathbf{\Psi})$ as the negative Hessian of the log-likelihood function

$$\vec{s}(\mathbf{\Psi}) = \frac{\partial \log L(\mathbf{\Psi})}{\partial \mathbf{\Psi}} \tag{2.63}$$

$$I(\mathbf{\Psi}) = -\frac{\partial^2 \log L(\mathbf{\Psi})}{\partial \mathbf{\Psi} \partial \mathbf{\Psi}^T}. \tag{2.64}$$

When evaluated at the maximum likelihood solution (which is approximated - in case of its existence - by the parameter set $\hat{\mathbf{\Psi}}$ asymptotically approached by the EM algorithm iterations), $I(\hat{\mathbf{\Psi}})$ is the Fisher information matrix. The information-based approach to parameter uncertainty quantification in this framework is then given by

$$SE(\hat{\Psi}_i) = \left[(I^{-1}(\hat{\mathbf{\Psi}}))_{ii} \right]^{1/2}. \tag{2.65}$$

Computing the observed and expected information matrices involves the explicit calculation of the Hessian which is often not possible analytically. For suitably approximating the observed information matrix, numerical differentiation of the Fisher score vector as well as of the EM maximization operator have been proposed [Meng and Rubin 1991, Jamshidian and Jennrich 2000]. Alternatively, the expectation of the complete-data information matrix can be estimated consistently and bias-free by the observed-data score covariance matrix [Behboodian 1972, Berndt et al. 1974, Redner and Walker 1984, Griffiths et al. 1987], or, alternatively, by the corresponding empirical covariance matrix [Meilijson 1989] which allows an analytic approach to the calculation of $I(\mathbf{\Psi})$.

In the case of explicitly given data, the score vector and the corresponding score and empirical covariance matrices read:

$$\vec{s}(\mathbf{\Psi}) = \sum_{j=1}^{J} \frac{\partial \log f(x_j; \mathbf{\Psi})}{\partial \mathbf{\Psi}} = \sum_{j=1}^{J} \vec{s}_j(\mathbf{\Psi}) \tag{2.66}$$

$$I_s(\mathbf{\Psi}) = \sum_{j=1}^{J} \vec{s}_j(\mathbf{\Psi}) \vec{s}_j^T(\mathbf{\Psi}) \tag{2.67}$$

$$I_e(\mathbf{\Psi}) = I_s(\mathbf{\Psi}) - \frac{1}{J} \left(\sum_{j=1}^{J} \vec{s}_j(\mathbf{\Psi}) \right) \left(\sum_{j=1}^{J} \vec{s}_j(\mathbf{\Psi}) \right) \tag{2.68}$$

For transferring these expressions to grouped and truncated data, one may introduce the abbreviations

$$\vec{h}_m(\mathbf{\Psi}) = \frac{\partial \log P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}} \tag{2.69}$$

$$\bar{\vec{h}}(\mathbf{\Psi}) = \sum_{m=1}^{M} \frac{P_m(\mathbf{\Psi})}{P(\mathbf{\Psi})} \vec{h}_m(\mathbf{\Psi}) = \frac{\partial \log P(\mathbf{\Psi})}{\partial \mathbf{\Psi}} \tag{2.70}$$

$$\tilde{\vec{h}}(\mathbf{\Psi}) = \sum_{m=1}^{M} \frac{n_m}{n} \vec{h}_m(\mathbf{\Psi}) \tag{2.71}$$

(with $\bar{\vec{h}}(\mathbf{\Psi}) = \tilde{\vec{h}}(\mathbf{\Psi})$ at the maximum likelihood solution) to achieve [Jones and McLachlan 1992]:

$$\vec{s}(\mathbf{\Psi}) = \sum_{m=1}^{M} n_m \vec{h}_m(\mathbf{\Psi}) - n\bar{\vec{h}}(\mathbf{\Psi}) \tag{2.72}$$

$$I_s(\mathbf{\Psi}) = \sum_{m=1}^{M} n_m \vec{h}_m(\mathbf{\Psi})\vec{h}_m^T(\mathbf{\Psi}) - n\bar{\vec{h}}(\mathbf{\Psi})\bar{\vec{h}}(\mathbf{\Psi}) \tag{2.73}$$

$$I_e(\mathbf{\Psi}) = \sum_{m=1}^{M} n_m \vec{h}_m(\mathbf{\Psi})\vec{h}_m^T(\mathbf{\Psi}) - n\tilde{\vec{h}}(\mathbf{\Psi})\tilde{\vec{h}}(\mathbf{\Psi}). \tag{2.74}$$

Consider now the parameter vector $\mathbf{\Psi} = (\pi_1, \ldots, \pi_K, \mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_K)$ for a K-component mixture. As any π_i $(i = 1, ..., K)$ may be expressed by the remaining π_k $(k \neq i)$, the information matrix with respect to this complete parameter vector has the rank $\dim(I) - 1$ which means that I is non-invertable. To overcome this problem, one may consider a reduced parameter vector $\mathbf{\Psi}^{[i]}$ equaling $\mathbf{\Psi}$ with the π_i component left out. In this case, one has to explicitly express this component by all π_j with $j \neq i$, leading to additional terms when evaluating the corresponding score vector contributions $\vec{h}_m(\mathbf{\Psi})$ (see App. B.8). The resulting estimates of the information matrix $I(\mathbf{\Psi})$ have full rank and may therefore be inverted without further problems.

2.4.2 Resampling-based Standard Errors

An alternative approach for calculating standard errors is the application of resampling (also known as bootstrapping) techniques [Efron and Tibshirani 1993] to generate artificial samples with a distribution corresponding either to the original data or to the estimated distribution functions. For EM parameter estimates of normal component means, it has been demonstrated that although large sample sizes may be necessary, bootstrap estimates of standard errors (i.e., the standard deviations of the resampled parameters) are favourable as the information-based standard errors do not always provide realistic and stable results [Basford et al. 1997]. Hence, resampling-based standard errors are often well-suited to provide uncertainty estimates for all parameters in a mixture.

The practical realisation of the resampling may be as follows:
(i) Generate an artificial set of data $\{x_j^*\} = (x_1, \ldots, x_{N_d})$ consistent with either the observed data themselves or the distribution function $f(x; \hat{\mathbf{\Psi}})$ estimated thereof.
(ii) Recalculate the parameters $\mathbf{\Psi}^*$ of this model with respect to the bootstrap sample $\{x_j^*\}$.
(iii) Repeat step (i) and (ii) a sufficient number of times N_s.
(iv) Calculate the bootstrap covariance matrix

$$\mathrm{cov}(\mathbf{\Psi}^*) = E\left\{\left[\mathbf{\Psi}^* - E\left\{\mathbf{\Psi}^*\right\}\right]\left[\mathbf{\Psi}^* - E\left\{\mathbf{\Psi}^*\right\}\right]^T\right\} \tag{2.75}$$

with the empirical approximations

$$E\left\{\Psi_i^*\right\} \approx \bar{\Psi}_i^* = \frac{1}{N_s}\sum_{\nu=1}^{N_s} \Psi_{i,\nu}^* \tag{2.76}$$

$$\mathrm{cov}(\mathbf{\Psi}^*)_{ij} \approx \frac{1}{N_s - 1}\sum_{\nu=1}^{N_s}\left(\Psi_{i,\nu} - \bar{\Psi}_i\right)\left(\Psi_{j,\nu} - \bar{\Psi}_j^*\right). \tag{2.77}$$

(v) Define bootstrap standard errors as $SE(\Psi_i^*) = [\mathrm{cov}(\mathbf{\Psi}^*)_{ii}]^{1/2}$.

The simplest and most general idea for generating the bootstrap sample (step (i)) is to apply the inverse transform method (see [Ross 2002]). In this case, the only prerequisite is a proper representation of the cumulative distribution function $F(x)$ of the observed data. Depending on this representation, the resampling approach may be performed both nonparametrically and parametrically.

In the nonparametric version, the observations themselves are taken as the fundamental reference for resampling by considering the cumulative distribution $F_0(x) = \int_{x_{min}}^{x} f_0(\xi)d\xi$ derived from their empirical distribution function $f_0(x)$. Typically, in mathematical statistics $F_0(x)$ is defined as a function that is piecewise constant between any observed value. For applying the inverse transform method, this point of view is not helpful. In the contrary, it is appropriate to approximate $F_0(x)$ for any type of data by a suitable continuous function.

In the case of data consisting of J explicit observations $\{x_j\}$, one may decompose the interval $[0, 1]$ (the range of $F_0(x)$) into $J + 1$ subintervals of equal size. If the observed values are increasingly ordered, one may define $F_0(x_j) = j/(J + 1)$. For grouped data, the unit interval may be decomposed similarly according to the respective group frequencies leading to $F_0(a_1) = 0$, $F_0(b_{m'}) = \sum_{m=1}^{m'} n_m/n$. Between these prescribed points, one can interpolate $F_0(x)$ piecewise linearly or by any monotonously increasing function. To overcome numerical problems due to very large classes with $P_m \to 0$, it is recommended to introduce a suitable (artifical) trunction even if the original data themselves are non-truncated.

For a parametric bootstrap, the cumulative distribution function provided by the estimated model $F(x; \hat{\mathbf{\Psi}}) = \int_{x_{min}}^{x} f(\xi; \hat{\mathbf{\Psi}})d\xi$ is used instead of $F_0(x)$. Thus, this approach allows to use non-truncated bootstrap samples even in the case of truncated observations which is in general not the case for the nonparametric resampling.

Applying the inverse transform method, the generation of the bootstrap samples may then be performed by first simulating an ensemble of data s_j $(j = 1, \ldots, N_d)$ with uniform distribution in $(0, 1)$ and then using the inverse cumulative distribution function (which uniquely exists in any subset of X where $f_0(x)$ (or $f(x; \hat{\mathbf{\Psi}})$, resp.) is non-zero) to calculate the surrogates x_j with $F_0(x_j) = s_j$ $(F(x_j; \hat{\mathbf{\Psi}}) = s_j)$. If in the case of a nonparametric bootstrap, $F_0(x)$ is approximated by a simple analytic (e.g., a piecewise linear) function, the generation of nonparametric surrogates typically requires significantly less computational power than the parametric approach.

2.4.3 A Numerical Example

To illustrate the performance of standard error estimates, the following rather simple example is considered: Let a sample of N =10,000 data points be constructed by simulating 4,000 points with a $N(-1.5, 0.81)$ distribution, 2,000 points with a $N(0, 0.16)$ distribution, and 4,000 points with a $N(2.25, 0.25)$ distribution (where $N(\mu, \sigma^2)$ corresponds to a normal distribution with mean μ and variance σ^2). These data are then grouped into bins of size 0.2 covering the interval $[-3.0, 3.0]$ (see Fig. 2.1 for a graphical representation). The resulting group frequencies are then used to run the EM algorithm for a 3-component normal mixture model yielding the results

$\hat{\pi}_1 = 0.375$ (0.4)	$\hat{\pi}_2 = 0.226$ (0.2)	$\hat{\pi}_3 = 0.399$ (0.4)
$\hat{\mu}_1 = -1.563$ (-1.5)	$\hat{\mu}_2 = -0.014$ (0.0)	$\hat{\mu}_3 = 2.250$ (2.25)
$\hat{\sigma}_1 = 0.850$ (0.9)	$\hat{\sigma}_2 = 0.445$ (0.4)	$\hat{\sigma}_3 = 0.498$ (0.5)

(the values given in brackets correspond to the parameters expected by construction of the data set).

Apparently, the estimates for the first two components do not match the prescribed values very well. In contrast, the parameters of the third component are much closer to the expected ones. The observed bias thus cannot be caused by the finite number of data, grouping coarseness, or even an insufficiency of the random number generator used for creating the data sample. A more reasonable explanation is that the first two components are not as well separated from each other as any of both from the third one.

The outcome of the EM algorithm is used to calculate the standard errors

$$\begin{array}{lll} SE(\hat{\pi}_1) = 0.021 & SE(\hat{\pi}_2) = 0.020 & SE(\hat{\pi}_3) = 0.006 \\ SE(\hat{\mu}_1) = 0.050 & SE(\hat{\mu}_2) = 0.025 & SE(\hat{\mu}_3) = 0.011 \\ SE(\hat{\sigma}_1) = 0.086 & SE(\hat{\sigma}_2) = 0.018 & SE(\hat{\sigma}_3) = 0.010 \end{array}$$

of all parameters from the information matrix $I(\mathbf{\Psi})$ with respect to all three possible reduced parameter vectors. The results are independent of the choice of the left-out component π_i in the reduced parameter vector. Concerning the effect of separation between the respective components, it is found that the parameters of the third component have actually considerably lower standard errors than for the first and second ones which underlines the relationship between the component overlap on the one hand and the parameter uncertainty and bias of the estimated mixture model on the other hand.

Using the considered example, one may demonstrate the consistency of the resampled parameters with the original estimates, as well as of the standard errors obtained with the different bootstrap approaches and the information matrix method. For the first purpose, the bias and mean squared error of the bootstrap parameter estimates are considered which are defined as (see [Nityasuddhi and Böhning 2003])

$$BIAS(\Psi_i^*) = \frac{1}{N_s}\sum_{\nu=1}^{N_s}\Psi_{i,\nu}^* - \hat{\Psi}_i = \bar{\Psi}_i - \hat{\Psi}_i \tag{2.78}$$

$$MSE(\Psi_i^*) = \frac{1}{N_s}\sum_{\nu=1}^{N_s}\left(\Psi_{i,\nu}^* - \hat{\Psi}_i\right)^2. \tag{2.79}$$

Note that for a vanishing bias, the mean squared error corresponds to the squared value of the corresponding bootstrap standard error.

The results are shown in Tab. 2.1-2.3 and Fig. 2.4. For a better comparability, the same grouping (and eventual truncation) has been applied to both, the original and the resampled data (of course, one may also perform the EM algorithm directly with the explicit surrogate data). Moreover, the size of the bootstrap samples was restricted to $N_d = N$ =10,000 points, i.e., the size of the prescribed (non-truncated) original data set. Three different types of resampled data have been generated: truncated nonparametric and parametric as well as non-truncated parametric bootstrap surrogates (by definition, nonparametric surrogates obey the same truncation as the original data).

From the calculated parameters, it may be inferred that there is actually a high degree of consistency between the results of all three approaches. Note, however, that as the nonparametric bootstrap is closer to the original data, this approach may be more reliable than the parametric resampling (which involves a priori unknown information about the truncated part of the sampling space and assumes a zero residual between the orginal data and the model distribution estimated thereof) but on the cost of larger bias and dispersion of the calculated parameters due to grouping coarseness. Moreover, in the example considered here, the standard errors given by the information matrix method and the different resampling approaches yield similar values.

The bias of the resampled parameters is sufficiently small such that $\left|SE(\Psi_i^*) - \sqrt{MSE(\Psi_i^*)}\right|$ is of the order of 10^{-4} to 10^{-6}. Of course, bootstrapping requires significantly more computational efforts than the information matrix method. Nonetheless, as it will be demonstrated next, the resampling approach to parameter uncertainty has serious advantages as it yields more detailed information about the distribution of uncertainty than the standard errors themselves.

Ψ_i	$\bar{\Psi}_i^*$	$SE(\Psi_i^*)$	$BIAS(\Psi_i^*)$	$MSE(\Psi_i^*)$
π_1	0.3765	0.0219	0.0013	0.0005
π_2	0.2252	0.0209	−0.0010	0.0004
π_3	0.3983	0.0057	−0.0002	< 0.0001
μ_1	−1.5598	0.0595	0.0033	0.0035
μ_2	−0.0143	0.0253	−0.0005	0.0006
μ_3	2.2505	0.0109	0.0004	0.0001
σ_1	0.8519	0.0531	0.0021	0.0028
σ_2	0.4430	0.0218	−0.0017	0.0005
σ_3	0.4970	0.0105	−0.0005	0.0001

Table 2.1: Estimated parameters and results of the parametric bootstrap resampling without truncation with 10,000 surrogate data sets: expected parameter values, standard errors, bias, and mean squared errors.

Ψ_i	$\bar{\Psi}_i^*$	$SE(\Psi_i^*)$	$BIAS(\Psi_i^*)$	$MSE(\Psi_i^*)$
π_1	0.3767	0.0218	0.0014	0.0005
π_2	0.2251	0.0209	−0.0011	0.0004
π_3	0.3982	0.0056	−0.0003	< 0.0001
μ_1	−1.5597	0.0590	0.0034	0.0035
μ_2	−0.0145	0.0253	−0.0007	0.0006
μ_3	2.2505	0.0110	0.0003	0.0001
σ_1	0.8524	0.0530	0.0026	0.0028
σ_2	0.4433	0.0219	−0.0014	0.0005
σ_3	0.4970	0.0106	−0.0005	0.0001

Table 2.2: Estimated parameters and results of the parametric bootstrap resampling with truncation with 10,000 surrogate data sets: expected parameter values, standard errors, bias, and mean squared errors.

2.4.4 Uncertainty Distributions and their Asymptotic Behaviour

Although the concept is frequently applied, the description of parameter uncertainty by standard errors is intrinsically faced with problems.

On the one hand, if certain parameters are subjected to constraints (like variances which are bounded from below by zero, or component weights in a mixture satisfying $\sum_{i=1}^{K} \pi_i = 1$), the likelihood is forced to be asymmetrically distributed around the estimated values. In such cases, the description of the resulting asymmetric parameter uncertainty by a (per definition) symmetric measure like the standard error may be not sufficient.

Ψ_i	$\bar{\Psi}_i^*$	$SE(\Psi_i^*)$	$BIAS(\Psi_i^*)$	$MSE(\Psi_i^*)$
π_1	0.3783	0.0245	0.0030	0.0006
π_2	0.2236	0.0227	-0.0027	0.0005
π_3	0.3982	0.0058	-0.0003	< 0.0001
μ_1	-1.5559	0.0648	0.0072	0.0043
μ_2	-0.0143	0.0227	-0.0005	0.0005
μ_3	2.2506	0.0105	0.0004	0.0001
σ_1	0.8566	0.0609	0.0068	0.0038
σ_2	0.4415	0.0234	-0.0032	0.0006
σ_3	0.4970	0.0108	-0.0006	0.0001

Table 2.3: Estimated parameters and results of the nonparametric bootstrap resampling with 10,000 surrogate data sets: expected parameter values, standard errors, bias, and mean squared errors.

On the other hand, for both information- and resampling-based methods, the approach of standard errors as estimates of parameter uncertainty implies that the distribution of uncertainty for any parameter is completely described by its first and second moments, i.e., by Gaussian distributions. In the case of bootstrap samples, this assumption can be easily checked by calculating higher-order characteristics of the distributions of resampled parameters. For the example from the previous section, the evolution of the standard deviation $Var^{1/2}$, the skewness γ_1, and the kurtosis γ_2 of these distributions are displayed in Fig. 2.5. As $\gamma_1 = \gamma_2 = 0$ for normal distributions, the latter two measures are well suited to quantify deviations from a Gaussian.

For the bootstrap parameter distributions, it is clearly observed that the calculated values of γ_1 and γ_2 are more or less clearly separated from zero such that a normal distribution is not a valid assumption for the resampled parameters. This is already underlined by a detailed inspection of Fig. 2.4 where a remarkable asymmetry of the resampling-based parameter distributions is visible for almost all model parameters. In particular, for component weights π_i and standard deviations σ_i whose values are bounded from below (for the π_i additionally from above), this asymmetry is a natural consequence of the respective constraint and causes a non-Gaussianity of the distribution of parameter "uncertainties" around the expected values. Moreover, one observes that for the third component, the absolute values of both skewness and kurtosis are remarkably smaller than those of the first and second ones (this is particularly true for π_3 and σ_3 compared to π_1, π_2 and σ_1, σ_2, resp.). This result indicates that in a mixture, the deviation of the bootstrap parameter distributions from a Gaussian is closely linked to the component overlap.

As an additional outcome, it is found that in the case of grouped truncated data, more bootstrap samples are necessary for stable standard errors than for explicit observations where about $50-100$ samples are typically found to be sufficient in earlier studies [Efron and Tibshirani 1993]. Depending on the respective parameters, in the considered example, typical values for the number of samples range up to several 1000 (see Fig. 2.5), depending on the considered parameter and the demands on accuracy.

As the distribution of parameters calculated from the bootstrap sample is non-Gaussian even for our Gaussian mixture example, one may consider the additional information for uncertainty assessment. While the dependence of the uncertainties on the number of bootstrap samples N_s has been described above, the natural question of the corresponding dependence on the size of the surrogate data sets N_d does still remain. In the case of the information-based standard

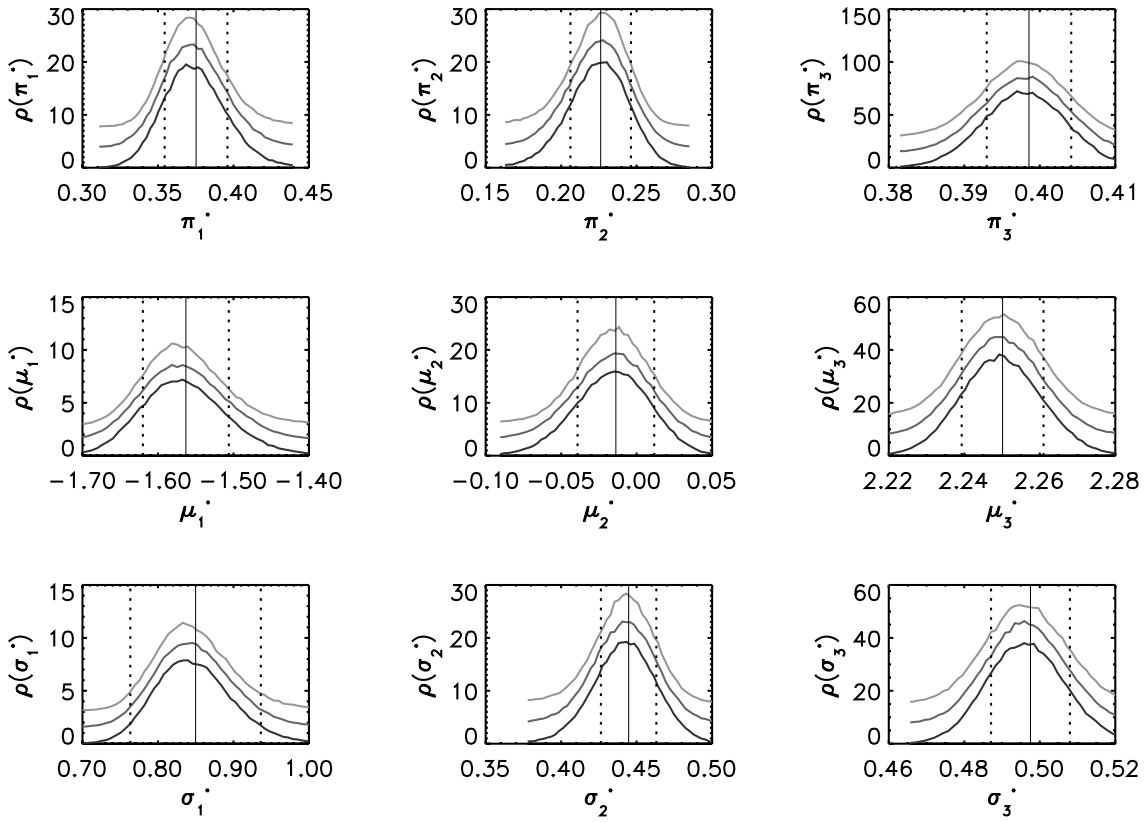


Figure 2.4: Frequency distributions of estimated parameters for the resampled data sets ($N_d = 10,000$, $N_s = 100,000$) with the nonparametric (truncated) bootstrap, and the parametric bootstrap without and with truncation (lines from top to bottom). For the latter one, the graph corresponds to the normalised density obtained from the set, whereas the other lines are vertically shifted for better identifiability. Vertical lines indicate the original estimates and the confidence levels of $\pm SE(\Psi_i)$ obtained with the information matrix method.

errors, it can be easily inferred that $SE(\hat{\Psi}_i) \sim 1/\sqrt{n}$. As the standard errors obtained by the resampling approach are consistent with those of the information matrix method, they have the same scaling behaviour. Following this idea, the bootstrap uncertainty distributions $\rho(\Psi_i^*(N_d))$ rescaled by a factor of $\sqrt{N_d}$ should approach *asymptotic parameter distributions* which may be defined as

$$\bar{\rho}(\Psi_i^*) = \lim_{N_d \to \infty} \rho\left(\sqrt{N_d}\left(\Psi_i^*(N_d) - \bar{\Psi}_i^*(N_d)\right)\right) \tag{2.80}$$

(alternatively, one may use $\hat{\Psi}_i$ instead of Ψ_i^* to approach an appropriate normalised uncertainty function). For the Gaussian mixture example, the correctness of this consideration is demonstrated in Fig. 2.6. This allows to calculate measures of parameter uncertainties even for the case of data given in terms of relative group frequencies only where the appropriate choice of N_d is not clear as n itself is unknown. Moreover, this result allows to calculate uncertainty distributions with larger bootstrap samples and to rescale them afterwards to the appropriate number of observations according to

$$\rho(\Psi_i^*(n)) \approx \rho\left(\sqrt{N_d/n}\left(\Psi_i^*(N_d) - \bar{\Psi}_i^*(N_d)\right) + \bar{\Psi}_i^*(N_d)\right). \tag{2.81}$$

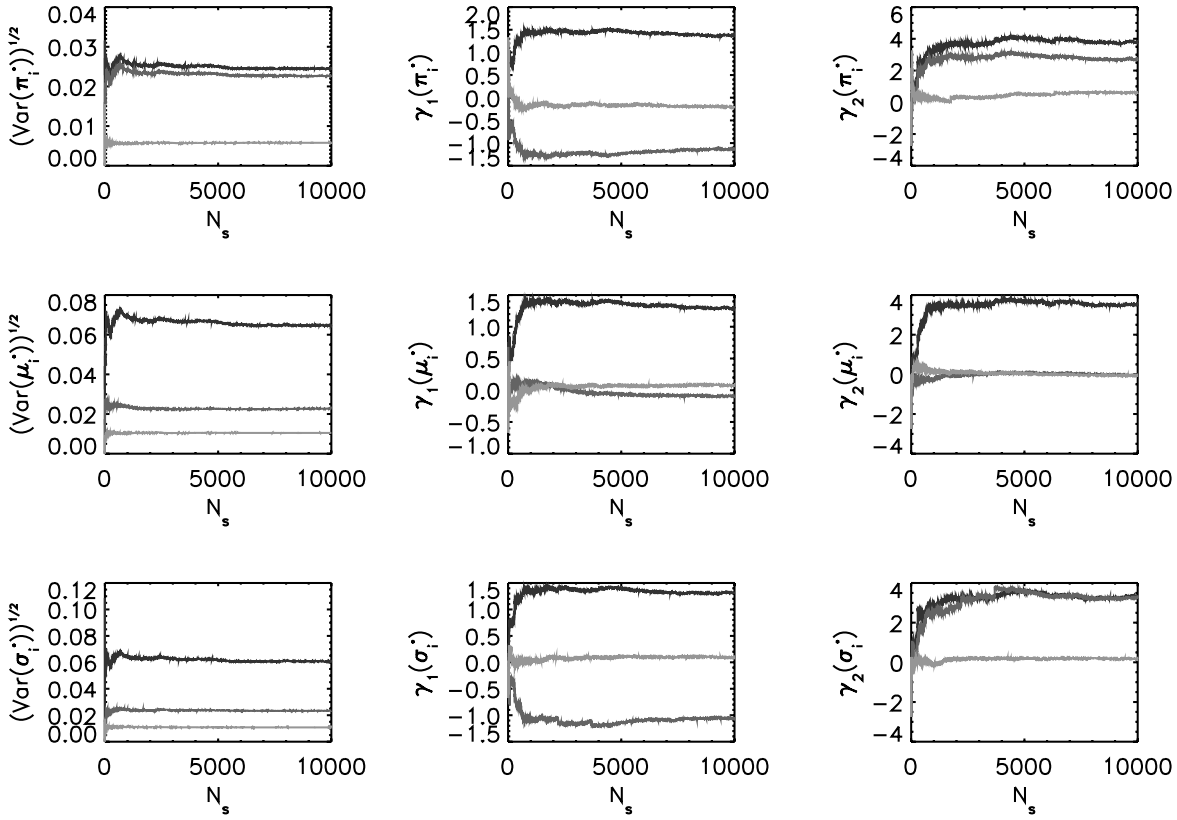


Figure 2.5: Standard deviations (first row), skewness $\gamma_1(\Psi_i)$ (second row), and kurtosis $\gamma_2(\Psi_i)$ (third row) of the bootstrap parameter distributions for the component weights (first column), mean values (second column), and standard deviations (third column) of the first (dark lines), second and third (bright lines) mixture component. The results have been obtained with non-parametric bootstrap samples containing $N_d = 10,000$ data points each for the first N_s respective parameter estimates from one continuous sequence of bootstrap realisations.

For $N_d \to \infty$, this measure yields all relevant information about the parameter uncertainty due to grouping and truncation of the data and a potential insufficiency of the model considered. Note that the asymptotic approach to parameter distributions is useful only if both original and resampled data are of the same type (i.e., either both are explicit ($n = N$) or grouped and truncated in the same way (possibly $N \neq n$)) as otherwise, a sufficient comparability of measurement and bootstrap samples may be missing.

2.4.5 Application: Grain-Size Distributions from Lake Baikal Sediments

The estimation of model parameters and the statistical assessment of their corresponding uncertainties has yet been discussed only for artificial data. In real-world examples, the grouped data may be much more problematic, involving for example more significant component overlap or a remaining uncertainty concerning the model itself (i.e., the number and shape of component functions).

The corresponding potential problems may be underlined by one short example: Consider the present-day aeolian dust grain-size record from Lake Baikal shown in Fig. 2.7. Fitting a

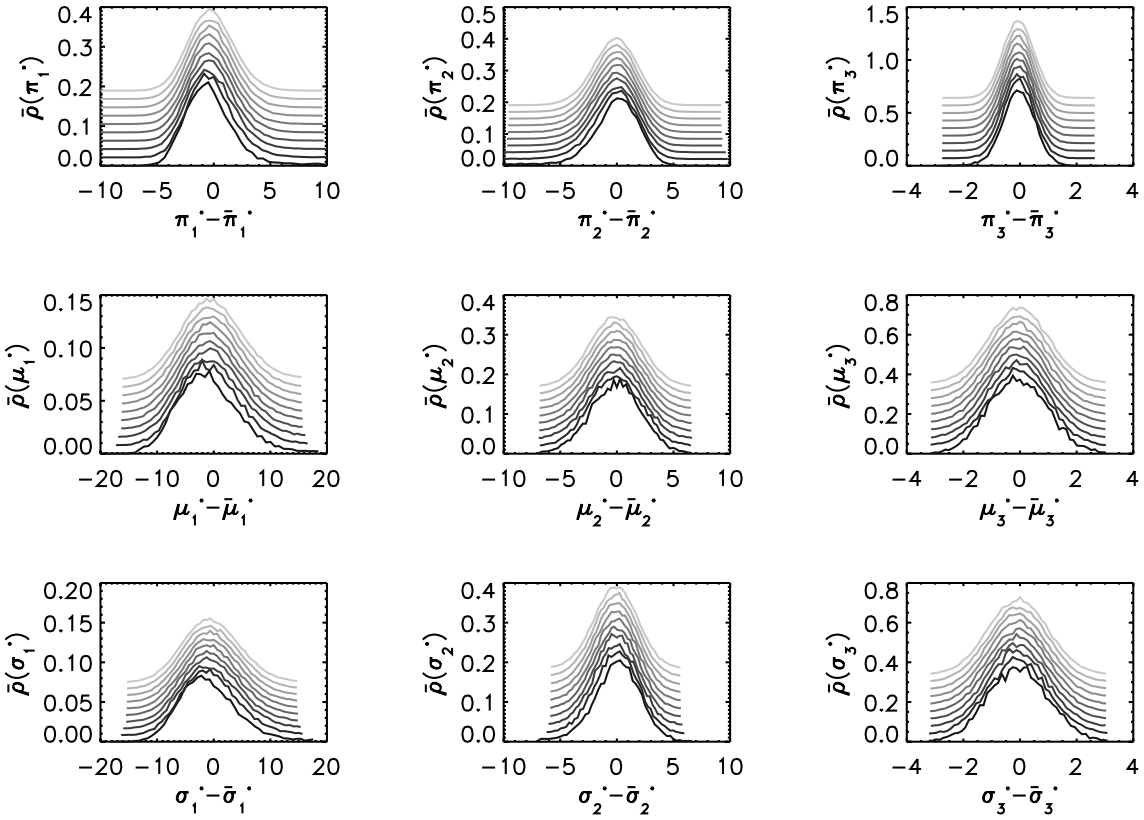


Figure 2.6: Asymptotic parameter distributions $\bar{\rho}(\Psi_i^*)$ estimated from $N_s = 100,000$ nonparametric bootstrap samples for the 3-component Gaussian mixture model. The size of the bootstrap samples ranges from $N_d = 10,000$ $(= n)$ (lowermost line) in steps of $10,000$ up to $100,000$ (uppermost line). For $N_d = 10,000$, the graph corresponds to the normalised density obtained from the set, whereas the other lines are vertically shifted for better identifiability. For a better comparability, we have used the alternative definition for $\bar{\rho}(\Psi_i^*)$ involving the original EM estimates $\hat{\Psi}_i$ instead of the sample means $\bar{\Psi}_i^*(N_d)$ which may slightly differ between the respective simulations.

finite mixture of lognormally distributed components to this data set requires three components to capture the essential shape of the observed distribution (see Fig. 2.7). To see this, one has to switch from the bar-chart representation (bar heights correspond to abundances) usually used in geosciences to a histogram representation (bar areas correspond to abundances). The minor "dusty" peak below 2μm appears not significant in the bar chart, but is a clearly evident indicator of a minor component (about 5% of the entire sediment) whereas the bulk distribution is composed by two strongly overlapping major components.

From Fig. 2.7, one may infer problems with the reliability and significance of the considered mixture models. Although obviously, three components are necessary to describe the data set sufficiently, the penalised-likelihood criteria AIC, BIC, and HIC favour less complex models with up to two components only (BIC even suggests that a single lognormal distribution would be the optimum choice, however, this suggestion may be a particular result of the bin dimension [Biernacki 2004c]. In the special case considered here, this problem may be related to the loga-

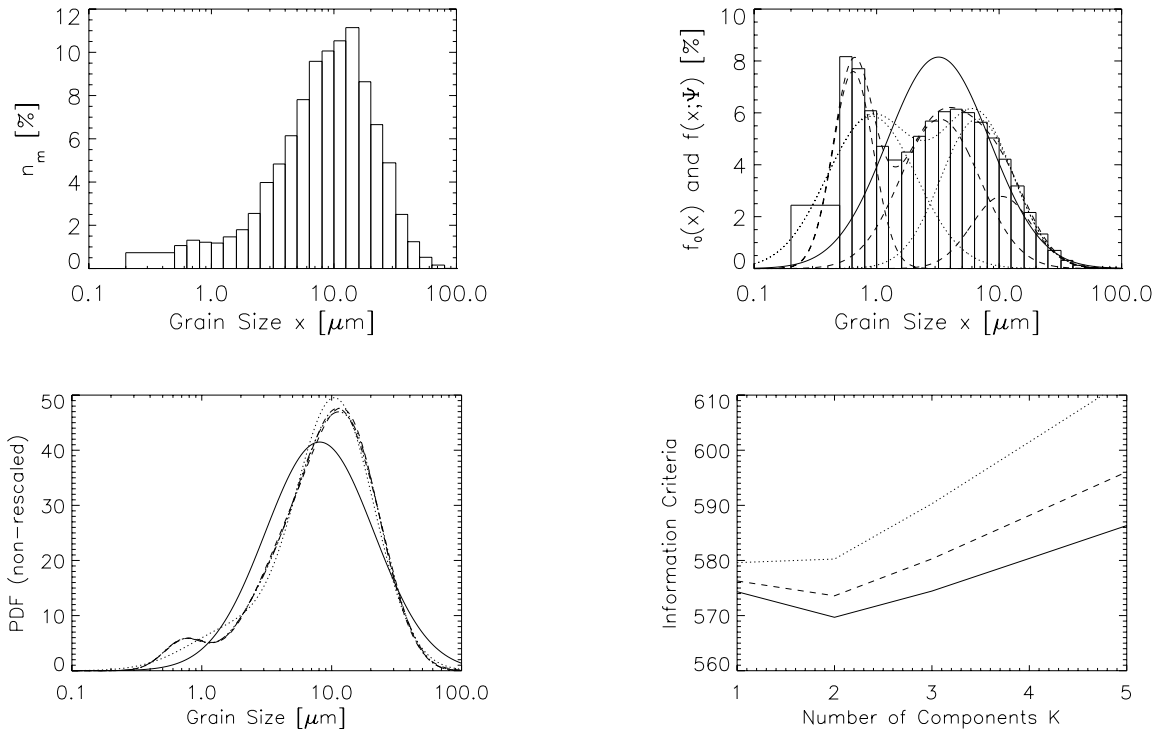


Figure 2.7: Upper panels: Bar chart (left) and histogram (right) representation of the present-day aeolian dust record from Lake Baikal. In the histogram, the (rescaled) probability distribution functions corresponding to finite mixture models with one (solid), two (dotted), and three (dashed) lognormal components. Lower left panel: The associated non-rescaled probability distribution functions associated to the finite mixture models. Lower right panel: Behaviour of the penalised-likelihood criteria AIC (solid), BIC (dotted), and HIC (dashed) in dependence on the number of component functions.

rithmic spacing of the size intervals, which is underlined by the fact that also the improvement of the likelihood or the associated χ^2 statistics of the estimated model distributions is only very weak when comparing two- and three-component mixtures.

The example underlines typical problems of the finite mixture approach for modelling real-world data sets. Although a particular model appears promising for deriving informations about sedimentation and transport mechanisms, a generally applicable test for an optimum model type and order is missing as standard criteria like AIC fail, hence, there is no unified approach for model validation. In addition, there is the potential problem of overfitting: On the one hand, with increasing model complexity, the undertainty of the estimated model parameters increases as well. On the other hand, there are distributions which cannot be appropriately described by a model involving only few components.

To investigate how these problems involve the uncerteinty assessment, the three-component lognormal mixture model is further analysed. In particular, the standard errors of all model parameters have been computed with both, the information matrix method and via resampling with (truncated) non-parametric as well as parametric bootstrap surrogates. The results displayed in Tab. 2.4 show that the standard errors derived from information matrix estimates and

different bootstrapping approaches are (in contrast to the numerical example studied in Sect. 2.4.3 and 2.4.4) not anymore consistent with respect to each other and are of the typical order of (in some cases even larger than) the estimated parameter values themselves. This result is obviously related to the fact that the two major components have a very strong overlap. As a consequence, whereas parametric bootstrapping allows to approximately recover the originally estimated parameter values, nonparametric resampling may lead to completely different results which underlines that in the example considered, the original outcome of the EM algorithm for the corresponding model can hardly be used for qualitative and quantitative interpretations of the record as the location of the computed maximum likelihood solution in parameter space changes crucially if the data are only slightly modified.

Ψ_i	$\hat{\Psi}_i$	$SE(\hat{\Psi}_i)$	$\bar{\Psi}^*_{i,np}$	$SE(\Psi^*_{i,np})$	$\bar{\Psi}^*_{i,p}$	$SE(\Psi^*_{i,p})$
π_1	0.0499	0.0936	0.1492	0.1490	0.0504	0.0613
π_2	0.4646	4.3073	0.4845	0.7862	0.4555	2.0173
π_3	0.4855	4.2261	0.3663	0.8388	0.4940	1.9783
μ_1	-0.2968	0.6139	0.4038	0.9361	-0.2933	0.4710
μ_2	1.7426	6.4788	1.7560	1.0415	1.7240	3.1052
μ_3	2.6685	1.5995	2.5662	0.6706	2.6620	0.9144
σ_1	0.3794	0.2995	1.0639	0.2806	0.3811	0.3169
σ_2	0.7441	3.9626	0.6446	0.3299	0.7337	1.3216
σ_3	0.5651	0.8171	0.5251	0.2149	0.5647	0.3918

Table 2.4: Estimated parameters $\hat{\Psi}_i$, information-matrix based standard errors $SE(\hat{\Psi}_i)$, and mean parameters $\bar{\Psi}^*_i$ and standard errors $SE(\Psi^*_i)$(rescaled by a factor of $\sqrt{N_d/n}$ with $n = 100$) from the non-parametric (subscript np) and parametric (subscript p) resampling with $N_s = 100,000$ surrogate data sets with $N_d = 100,000$ data each. The grain-size values (in μm) have been logarithmised to approach normal components whose weights π_i, means μ_i, and standard deviations σ_i are given.

Although the resulting distributions of resampled parameters show again significant deviations from a Gaussian and a particularly large asymmetry, the qualitative behaviour of the standard errors is reproduced by the asymptotic bootstrap parameter distributions displayed in Figs. 2.8 and 2.9. In particular, the distributions associated to the minor component are still rather symmetric, whereas those for the strongly overlapping major components are much more asymetric (for example, the parameters of the second component show a clear preference of smaller values). Comparing the results of non-parametric and parametric bootstrap, it is found that the non-parametric uncertainty distributions are wider than the parametric ones for the minor component, but larger for the major components. This behaviour indicates that the parametric approach is unusually sensitive with respect to the minor component, i.e., this component is better resolved than the original data would actually allow. In contrast, for the major components, the parametric approach seems to lose information.

2.5 Open Problems

Although the EM algorithm is a robust and efficient tool for parameter estimation and is thus frequently applied, there are several properties of the algorithm which are problematic for certain practical applications. In the following, the most prominent of these problems will be described.

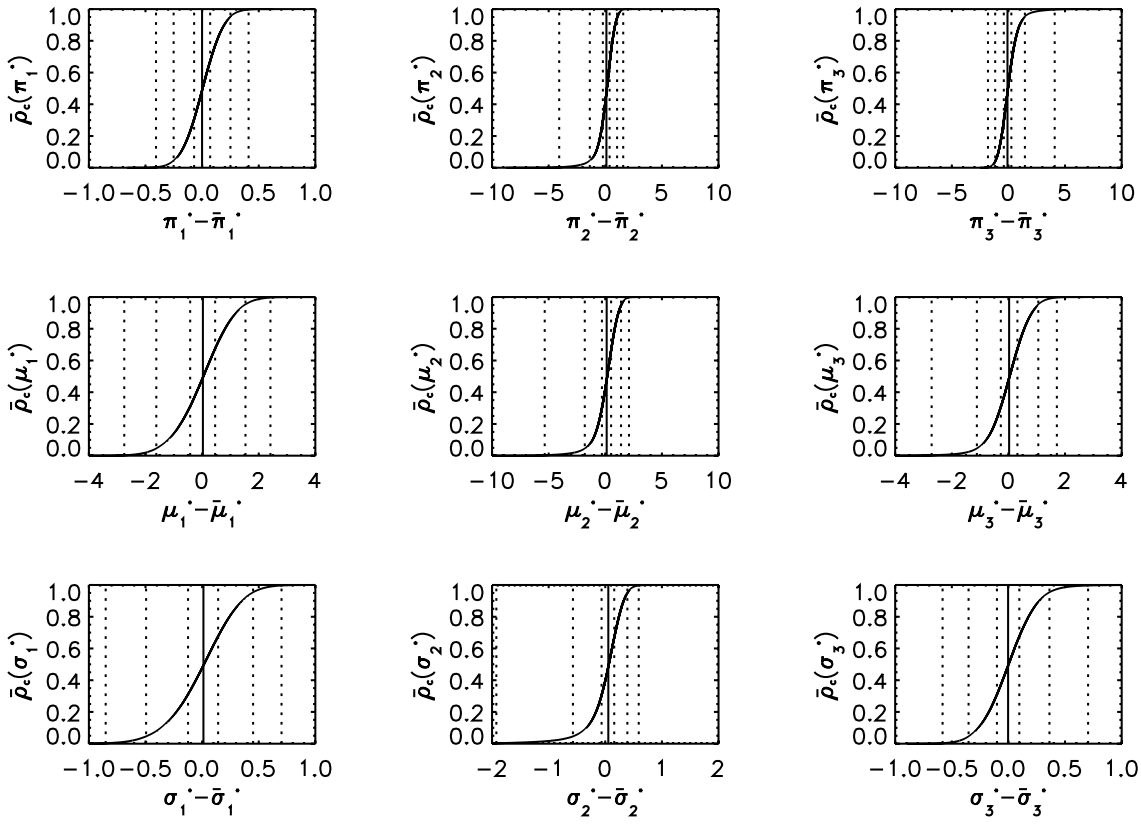


Figure 2.8: Cumulative asymptotic parameter distributions $\bar{\rho}_c(\Psi_i^*) = \int_{-\infty}^{\Psi_i^*} \bar{\rho}(\psi)\, d\psi$ for all model parameters obtained with the non-parametric bootstrap. The vertical lines correspond to the sample median (solid) and the 1σ, 2σ, and 3σ probability levels of a Gaussian distribution (dotted).

Recent solution approaches are given, however, there is still almost no general concept which helps to avoid the discussed difficulties for any particular example. Therefore, it is appropriate to refer to the following points to as open or at least not yet completely solved problems.

2.5.1 Uniqueness and Convergence

The question of general convergence of the EM algorithms was firstly addressed by [Wu 1983] who particularly proved that an unimodal likelihood function with a certain differentiability condition is a sufficient criterion for the convergence of an EM algorithm and the uniqueness of its solution. However, these prerequisites are rarely fulfilled (in particular, for the case of mixture models discussed in this chapter) which may lead to a multiplicity of local maxima of the likelihood function which are approached by the EM algorithm depending on the choice of starting values [McLachlan 1988]. As a consequence, the appropriate choice of initial parameter values may be an important problem in practical realisations of the EM algorithm (for an overview about some possible strategies and a corresponding comparison of their performance, see [Seidel et al. 2000b, Karlis and Xekalaki 2003]). [Biernacki et al. 2003] pointed out that instead of considering a random initialisation which is frequently applied, the strategy of initiating

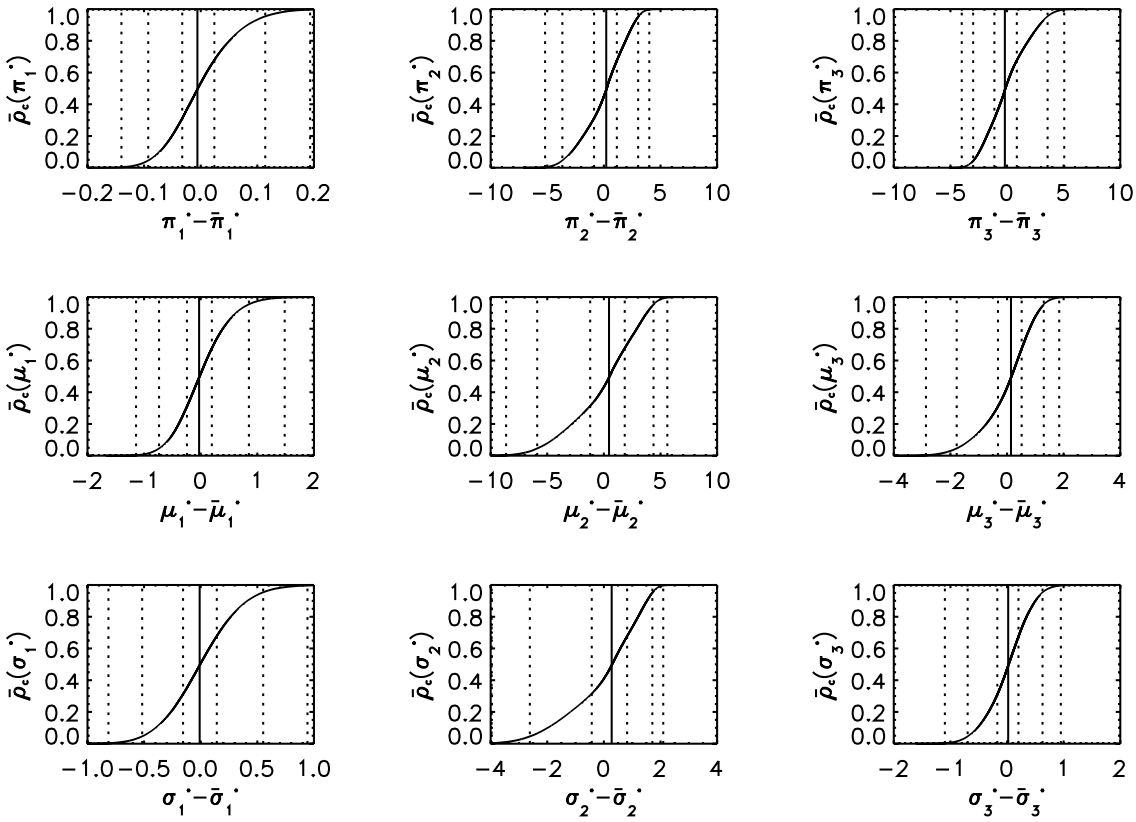


Figure 2.9: Cumulative asymptotic parameter distributions $\bar{\rho}_c(\Psi_i^*)$ for all model parameters obtained with the parametric bootstrap. The vertical lines correspond to the sample median (solid) and the 1σ, 2σ, and 3σ probability levels of a Gaussian distribution (dotted).

the EM algorithm based on short runs of the EM is recommended. [Seidel and Sevcikova 2004] have demonstrated that in two-component mixtures of exponential distributions, different strategies for starting the likelihood maximisation algorithm converge to different types of maxima, hence, there is a need for criteria and strategies which may identify the statistically meaningful maxima, e.g., based on sophisticated resampling approaches [Seidel et al. 2003]. Further improvement of the corresponding behaviour of the EM algorithm may be achieved by gradient function updates [Böhning 2002, Seidel and Sevcikova 2002a].

Beside its dependence on the initial values, the EM algorithm usually converges rather slowly. Thus, methods for improving its convergence have been extensively discussed in the literature [McLachlan 1996, Karlis and Xekalaki 1999], including quasi-Newton acceleration steps [Lange 1995, Jamshidian and Jennrich 1997], alternations of EM iterations with Gauss-Newton iterations [Aitkin and Aitkin 1996], conjugate-gradient methods [Jamshidian and Jennrich 1993], or the so-called ECME algorithm [Liu and Rubin 1994]. A number of these approaches involves appropriate estimates of the information matrix of the considered estimator [Louis 1982, Meilijson 1989, Meng and Rubin 1991, Jones and McLachlan 1992, Lange 1995, Jamshidian and Jennrich 1997].

A third problem is the detection of convergence being a crucial point in practical im-

plementations. Standard stopping rules are based on the relative change of parameters and/or the log-likelihood function (e.g., [Agha and Ibrahim 1984]), the so-called gradient function or directional derivative [Lindsey 1983, Böhning et al. 1992, Pilla and Lindsay 2001], or the Aitken acceleration [Böhning et al. 1994]. Although the first approach is mostly applied, the corresponding criteria indicate a lack of progress rather than actual convergence [Lindstrom and Bates 1988, Seidel et al. 2000a, Karlis 2001]. However, also the concurring approaches have some serious disadvantages in practical applications as pointed out by [Karlis 2001].

Unlike for example quasi-Newton type algorithms, the EM algorithm converges to a (local) maximum likelihood solution under rather general conditions [Davenport et al. 1988]. However, in Gaussian mixtures, in the case of explicit data, the likelihood function is unbounded for any parameters [Kiefer and Wolfowitz 1956, Day 1969] which can be seen by placing a Dirac at any observed sample point [Biernacki and Chrétien 2003]. In the case of grouped data, the global maximum of the likelihood function is approached in a similar way when placing a Dirac in any non-empty interval [Biernacki 2004b]. Around such degenerate solutions, the EM algorithm iterates move very slowly [Biernacki and Chrétien 2003, Biernacki 2004b] which may falsely indicate convergence of the algorithm when applying the function-based stopping criteria as described above. Moreover, there is a domain of attraction around degeneracy where convergence to the degenerate solution occurs very fast [Biernacki and Chrétien 2003, Biernacki 2004b]. To avoid such degenerate solutions, [Biernacki 2004a] theoretically derived an asymptotic upper bound for the likelihood function of Gaussian mixtures and demonstrated its performance for both, artificial and real-world data.

2.5.2 Model Validation

In most practical applications of finite mixture models, the number and shape of component functions is a priori unknown. Thus, the assessment of model validity is a crucial point, in particular, for comparing a particularly chosen model structure with a possible alternative. The simplest idea is considering the appropriately quantified goodness-of-fit of the respective models with respect to the observed data, e.g., in terms of the χ^2 statistics in the case of grouped data. However, this approach may be problematic regarding an unknown number of components to be tested for: for real-world data sets, the consideration of additional components usually leads to a further improvement of the *value* of the test statistics, whereas this does not necessarily hold for the associated *probability* of rejecting the null hypothesis of the simpler model (i.e., the model with a smaller number of unknown components).

As a promising alternative, penalised-likelihood criteria are often used. Instead of considering the computed maximum value of the log-likelihood function itself as a criterion for the goodness-of-fit, this value is here penalised by an additional additive factor including the number of unknown parameters. For deriving these penalty terms, the expected log-likelihood must be appropriately approximated based on sophisticated assumptions on the underlying likelihood ratio (LR) test statistics (for a summary of some appropriate validity functionals, see [Miloslavsky and van der Laan 2003]). As an alternative, the likelihood ratio between the models corresponding to the alternative and null hypothesis may be directly considered for model validation itself.

Practical realisations of both, penalised-likelihood criteria and the likelihood ratio statistic, are often problematic as the distribution of the LR statistics is itself unknown. To give appropriate probability estimates for rejecting a given null hypothesis, this distribution has to be approximated itself which is frequently done in terms of resampling approaches

[McLachlan 1987, McLachlan et al. 1995, McLachlan and Peel 1997, McLachlan and Peel 2000, Schlattmann 2005] as the availability of explicit expressions is an exception rather than the general case (see, e.g., [Böhning et al. 1994, Seidel et al. 1997, Liu and Shao 2004, Hall and Stewart 2005]). In some cases, the likelihood ratio statistics can be efficiently described by its asymptotic distribution [Ruck 2001, Ruck 2002, Azais et al. 2004]. As this distribution may be very complex and difficult to use in practice, [Chen et al. 2001] proposed a modified LR test with a χ^2-type null distribution. Independent of the particular approximation, the appropriate maximisation of the likelihood function (see Sec. 2.5.1) remains a fundamental prerequisite for any model validation [Seidel et al. 2000a, Seidel et al. 2000b, Seidel and Sevcikova 2002b, Seidel et al. 2003, Seidel and Sevcikova 2003a, Seidel and Sevcikova 2003b, Seidel and Sevcikova 2004].

For a more detailed overview on the different approaches for assessing the number of components in mixture models, one may refer to Chapt. 6 of [McLachlan and Peel 2000] where the above mentioned methods are discussed and illustrated by some insightful examples.

2.5.3 Maximisation Step for Non-Gaussian Components

In the case of explicit data, the maximisation step of the EM algorithm can usually be explicitly solved for the unknown model parameters $\mathbf{\Psi}$. However, this is not the case any more for grouped and possibly truncated data where the summation over the different observations is replaced by integrations. For homogeneous Gaussian models and mixture models with Gaussian component functions, the corresponding derivations are discussed in App. A. Further examples where the maximisation step can be performed without further secondary iterations include negative binomial distributions [Schader and Schmid 1985, Adamidis 1999] and Poisson mixtures in binomial proportions [Adamidis and Loukas 1993].

As in practical applications, other components functions may be of interest (see, e.g., Sec. 4.3), it is an important question to avoid or at least improve the computational efficiency of secondary iterations in the maximsation step of the algorithm. This problem is rather model-specific and shall not be further discussed here.

Chapter 3

Dimension Estimates of Multivariate Time Series

3.1 Motivation

As it was already discussed in the introductory chapter of this thesis, palaeoclimatic time series usually contain simultaneous measurements of a variety of different parameters which can serve as climatic proxies. Although both, the campaigns for recovering samples from a geological source (e.g., ice or sediment cores) and the subsequent measurement processes are not only very time-consuming, but also rather expensive with respect to the required manpower and technique, there is typically no discussion about which complementary information can actually be extracted from such multivariate records.

In this chapter, this natural and important question is addressed quantitatively. For this purpose, appropriate measures are developed and discussed that allow to quantify the amount of information about the variability of the underlying system contained in general observed multivariate time series. This quantification is performed in terms of the number of statistically meaningful components that can be derived from the multivariate record, which may be considered as a measure for the strength of ensemble correlations or, more general, mutual interdependences. In addition, a time-dependent calculation of these measures (i.e., a separate computation for different parts of the climate history) allows to consider the variability of the appropriately quantified information content as a novel measure for long-term climate change at the location under investigation.

In univariate time series analysis, a number of nonlinear measures is widely applied [Abarbanel 1996, Kantz and Schreiber 1997] for characterising the complexity of the data (see Sec. 1.3). Several of them can be appropriately extended to the multivariate case, however, in the case of geological records, the length and resolution of the corresponding time series is highly restricted such that none of these methods may provide sufficient results any longer. To quantify the content of information about the variability of the underlying system contained in even extremely small data sets, one may consider the complexity of interrelationships between the respective component time series in terms of the number of variability patterns that can be derived from such data and show statistically meaningful amplitudes. The corresponding method is thus based on an appropriate estimation of this number after a suitable statistical decomposition of the data and will be discussed in this chapter.

It is worth to be mentioned that the statistical techniques developed in the following are of a broad interest in various fields of research dealing with multivariate data analysis. In particular,

in geosciences, there are possible applications in terms of spatio-temporal observational records obtained, for example, in seismology or (hydro-) meteorology, with variations of the complexity of linear (or, more general, nonlinear) correlations yielding indicators of an underlying change of the system's dynamics, which may be a predecessor of an extreme event (e.g., an earthquake, flooding, or heat wave). Similar examples may be taken from other areas of science where multivariate time series play an important role. As a corresponding potential field of application, the analysis of the dynamics in neuro-physiological, social, economical, or biological networks should be mentioned.

3.2 KLD-Based Dimension Estimates

3.2.1 Statistical Decomposition of Multivariate Data Sets

The characterisation of ensemble correlations by a single statistical parameter requires an appropriate statistical decomposition of the corresponding multivariate time series. In principle, this decomposition can be performed by a variety of different approaches, including purely linear methods like Karhunen-Loéve decomposition (KLD) (which is often referred to as principal component analysis (PCA) or empirical orthogonal function (EOF) method) [Joliffe 1986, Preisendorfer 1988], multi-dimensional scaling (MDS) [Cox and Cox 2000], or, referring to a separate consideration of patterns in the frequency domain, multi-channel singular spectrum (or system) analysis (MSSA) [Plaut and Vautard 1994], a combination of the "standard" singular spectrum analysis (SSA) [Broomhead and King 1986] with PCA. All these methods have the common concept that some matrix which is suitably constructed from the observational data is subjected to a singular value decomposition (SVD), i.e., it is decomposed into its eigenvalues and the corresponding eigenvectors. In the case of KLD, one makes use of the correlation (or scatter) matrix $S = A^T A$ of the observed data set A (whose components have to be transformed to zero means). For MDS, a transformed matrix of the squared linear inter-point distances is used, whereas MSSA is based on a Toeplitz-type lag-covariance matrix obtained from every univariate components time series.

Whereas the SVD step of all these methods may be easily and computationally efficiently performed, there are also different nonlinear generalisations. One possible way to obtain such generalisations is replacing the Euclidean metric by one defined by the local neighborhood, e.g., in terms of isometric feature mapping (ISOMAP) [Tenenbaum et al. 2000] or locally linear embedding (LLE) [Roweis and Saul 2000]. An alternative is realising the decomposition in terms of neural networks, including methods like nonlinear principal component analysis (NLPCA) [Kramer 1991] or independent component analysis (ICA) [Hyvärinen et al. 2001]. However, these nonlinear variants require a much larger amount of data for computation, while the linear methods can be applied to rather short time series as well. In addition, the methods based on neural networks do not lead to well-defined component variances such that the approach described in the following is not applicable in such cases.

For a temporally localised characterisation of the components of multivariate data sets, it is thus recommended to focus on the linear methods only. For simplicity, Karhunen-Loéve decomposition is considered as a particular example, as the components derived by this method have probably the most intuitive interpretation. As its principal idea has been introduced about 100 years ago (see, e.g., [Preisendorfer 1988] for some historical remarks), KLD is today frequently applied as a standard method for compressing spatiotemporal data by finding the largest linear subspace that contains substantial statistical variations of the data. In the case of observations with N simultaneously measured variables and M points in time, the $M \times N$-

dimensional data matrix A (rescaled to zero means for any component time series) is used to define a $N \times N$-dimensional symmetric and positive semidefinite scatter matrix $S = A^T A$. The matrix S can be completely described by its non-negative eigenvalues σ_i^2 $(i = 1, \ldots, N)$ and their corresponding eigenvectors (which are in geosciences usually referred to as the empirical orthogonal functions (EOF)). Without loss of generalisation, one may consider the σ_i^2 of S in decreasing order $\sigma_1^2 \geq \cdots \geq \sigma_N^2 \geq 0$ in the following. Moreover, the eigenvalues will be normalised to unit sum $\sum_{i=1}^{N} \sigma_i^2 = 1$ wherever appropriate.

In order to examine the dynamic features of the data, it is common to additionally study the time-dependence of the amplitudes corresponding to the respective EOF (if thus considered dynamically, KLD is usually referred to as principal component analysis (PCA)). However, this approach still reflects only the linear properties of the observations, but does not allow a nonlinear characterisation of the record in terms of quantitative measures.

3.2.2 KLD Dimension

The idea of using Karhunen-Loève decomposition for estimating the number of degrees of freedom in spatially extended systems is already presented in [Ciliberto and Nicolaenko 1991]. As in the case of weakly turbulent systems, the same quantity may be represented with methods based on the fractal dimension [Pomeau 1985] or Lyapunov exponents [Kaneko 1989, Mayer-Kress and Kaneko 1989], it is convenient to refer to this number to as "the" dimension of the considered system. Following this line of argumentation, one may extend the application of Karhunen-Loève decomposition with respect to the purely linear point of view described above.

To determine the number of degrees of freedom in spatially extended systems, [Zoldi and Greenside 1997] introduced the concept of KLD dimension for a quantitative characterisation of spatio-temporal chaos [Zoldi et al. 1998, Meixner et al. 2000, Varela et al. 2005]. The KLD dimension may be defined as the number of eigenvalues required to capture some specified fraction $0 \leq f \leq 1$ of the total variance $\sum_{i=1}^{N} \sigma_i^2$ of the data, i.e.,

$$D_{KLD}(f) = \min \left\{ p : \sum_{i=1}^{p} \sigma_i^2 \Big/ \sum_{i=1}^{N} \sigma_i^2 \geq f \right\} \tag{3.1}$$

with the limiting cases $D_{KLD}(0) = 0$ and $D_{KLD}(1) = N$. It should be noted that this definition is modified with respect to the original one introduced by [Zoldi and Greenside 1997] and [Meixner et al. 2000] who considered $D_{KLD}(f)$ being the maximum number of eigenmodes describing less than a fraction of f of the total variance. This modification is motivated by the fact that for applications in data analysis, for a given f the minimum number of modes that explains a given amount of total variance is usually the quantity of interest. Moreover, this redefinition leads to a more natural behaviour of the KLD dimension at the limiting cases $f = 0$ and $f = 1$ as described above.

In the case of simulations of spatio-temporally chaotic systems, Zoldi and coworkers observed (for any f) a linear scaling of D_{KLD} with the system size N. Whereas the KLD dimension is otherwise restricted to integer values, this finding suggested to study a normalised version, the KLD dimension density $\delta_{KLD} = D_{KLD}/N$ [Meixner et al. 2000], whose values are bounded to the unit interval.

The KLD dimension has mainly been used to characterise the dynamics of spatially extended model systems in the extensive chaotic state [Zoldi and Greenside 1997], spiral-defect chaos [Zoldi et al. 1998], and reaction-diffusion systems [Meixner et al. 2000]. Recently,

[Varela et al. 2005] applied D_{KLD} for an investigation of spatiotemporal data from electrochemical oscillator experiments (with $M \geq 6000$ and $N = 50$). It has been demonstrated that this measure is well suited for quantifying differences between regular and turbulent states.

To adapt the concept of KLD dimension within a more general framework of multivariate data analysis (in particular for geoscientific applications), one may in addition explicitly consider the temporal variability of the observations for a temporally localised characterisation of the dynamics. While the consideration of S for the complete data set loses any temporal information about the variations in the complexity of interrelationships between the different components (which may be significant especially if $M \gg N$), a separate computation of the KLD dimension for sliding windows in time [Meixner et al. 2000] allows the resolution of the varying complexity down to the scale of N points in time or even below.

3.2.3 LVD Dimension

Whereas the KLD dimension density can be widely applied to characterise large data sets from spatio-temporally chaotic systems, its direct use for the characterisation of an observational record is problematic in the case of small data sets (i.e., small N) or time windows (small M) due to different reasons:

Firstly, δ_{KLD} has a possible range of only $N + 1$ different, equally spaces values. Thus, the number of possible values becomes very small for the considered data. As a consequence, small changes of the structure of interrelationships between the component time series are not detected by this measure, whereas it changes discontinuously (with a step size of $1/N$) when these modifications of the data increase over a certain threshold. Thus, if N is rather small, only rather strong changes within the data are detected by a dramatic change of δ_{KLD}.

Secondly, there is no natural choice of the cutoff parameter f which has to be specified separately for each application. Thus, it is not appropriate to consider δ_{KLD} as an *absolute,* but rather as an *relative* dimension density. However, for applications where only a qualitative detection and description of changes of the complexity of interrelationships within multivariate data is requested, this subtile difference is no major problem.

Thirdly, due to the small amount of observational data in time, certain finite-size effects have to be expected which may cause any quantitative interpretation of δ_{KLD} to fail.

These arguments call for a definition of more general estimates for relative dimension densities which can be applied also to small multivariate data. As a possibility, one may consider the scaling of δ_{KLD} with the cutoff parameter f and fitting a suitable parametric function to the respective curve. For this purpose, one firstly observes that for a given value of $\delta_{KLD}(f) = p/N$ $(p = 0, \ldots, N)$, $1 - f$ plays the role of the remaining variances defined as $V_r(p/N) = 1 - \sum_{i=1}^{p} \sigma_i^2$ for $p = 1, \ldots, N$ $(V_r(0) = 1.0)$, where p/N is the relative number of components considered. For the component variances (i.e., the eigenvalues of the equal-time covariance matrix), the scaling behaviour has been investigated in some detail for random matrices [Farmer 1971, Probert-Jones 1973] as well as real-world geoscientific data [Craddock and Flood 1969] in terms of the logarithmic eigenvalue (LEV) curves (for an overview, see [Preisendorfer 1988]). The LEV curve is usually used as a simple possibility to graphically check whether the component variances decay sufficiently smooth which is an important prerequisite for a meaningful interpretation of KLD-based dimension estimates.

In contrast to the component variances, there are no studies analysing the scaling of the remaining variances in some detail. However, a rough inspection of the corresponding values for both, random matrices as well as observational data, shows that the decay corresponding to the major components (i.e., the consideration of the components with the highest variances) is in

general more or less well described by an exponential decay law (see Fig. 3.1). As a consequence, one can make the following ansatz:

$$V_r(p/N) = e^{-\frac{p}{N}/\delta} \quad \text{for} \quad p \leq p_{max} < N. \tag{3.2}$$

The corresponding value for δ may be computed by a simple linear least square approach. However, if N is rather small, there are only few points to interpolate the respective model function. Moreover, there are again only N possible choices of the threshold p_{max} for fitting this function (as $V_r(N) = 0.0$ by definition, an exponential decay law must be subjected to a certain cutoff at $p_{max} < N$). To overcome this difficulty and define the model function with respect to a continuously distributed cutoff parameter f, one can make use of the relationship between $V_r(p)$ and $1 - f$ which is illustrated in Figs. 3.1 and 3.2: reversing the axes in Fig. 3.2 and multiplying δ_{KLD} by $N(=32)$, one approaches a continuously defined equivalent of the right panel in Fig. 3.1 (where the illustrated function is defined only for integer values of p). A scaling law of the KLD dimension density corresponding to that of the remaining variances then looks as follows:

$$\delta_{KLD}(\phi) = -\delta(f)\ln(1-\phi) \quad \text{for} \quad \phi \in [0, f]. \tag{3.3}$$

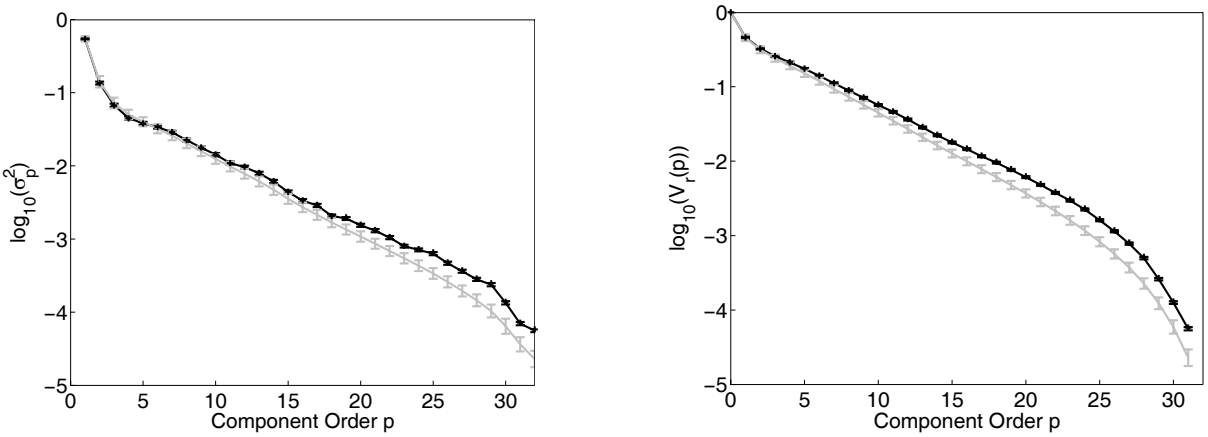


Figure 3.1: Scaling of the component variances σ_p^2 (left panel) and the corresponding remaining variances $V_r(p)$ (right panel) for the normalised trace element abundance data ($M = 60$ and $N = 32$, black lines) discussed in Sect. 3.5. For comparing the results with those of finite-size random matrices, we additionally computed $V_r(p)$ for ensembles of 1000 multivariate ($N = 32$) surrogate data sets consisting of normally distributed data (with prescribed component variances equal to those of the original data) with length $M = 60$ (gray) and $M = 1000$ (black) points in time. The displayed error bars correspond to the standard deviations of the values from the respective surrogates. The deviation between the black and the gray curve is mainly explained by the small size and non-Gaussian distribution of the observed time series values.

Instead of using the natural logarithm $\ln(\cdot)$, in the following, the decadal logarithm $\log_{10}(\cdot)$ will be applied for convenience. The resulting decay scale δ_{LVD} is modified with respect to δ by a constant factor of $(\log_{10} e)^{-1}$ (due to the remaining dependence on p_{max} or f, resp., one is still interested in a relative measure only such that the corresponding modification is not critical). The particular choice of the decadal logarithm, i.e., an explained fraction of 90% of the total variance of the data, is motivated by the fact that this threshold gives a reasonable number for

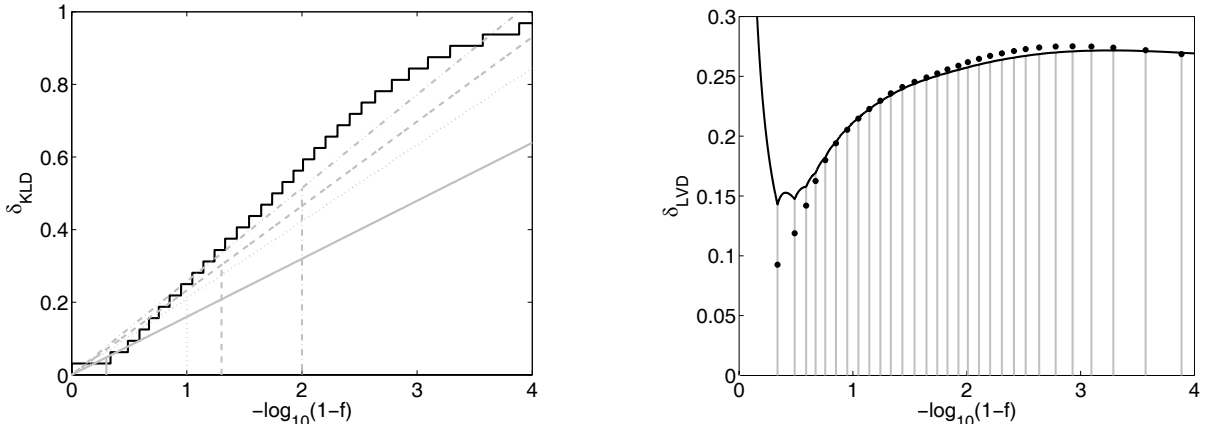


Figure 3.2: Left panel: Scaling of the KLD dimension density $\delta_{KLD}(f)$ with $\log(1-f)$ (black line) for the trace element abundance data discussed in Sect. 3.5. Vertical gray lines indicate the cutoff values of $f = 0.5$ (solid), 0.9 (dotted), 0.95 (dashed), and 0.99 (dash-dotted), whereas the slope of the associated gray diagonal lines correspond to the respective values of $\delta_{LVD}(f)$. Right panel: Scaling of the corresponding LVD dimension density $\delta_{LVD}(f)$ with $\log(1-f)$ (black line). Vertical gray lines indicate the total variances $\sum_{i=1}^{p} \sigma_i^2$ corresponding to a fixed number of components $p = 1, \ldots, N$ (from right to left). Asterisks indicate the estimates of $\delta_{LVD}(p/N)$ using the discrete least-square approach.

the effective degree of freedom in spatially extended systems, cf. [Ciliberto and Nicolaenko 1991] (93%), [Zoldi and Greenside 1997] (between 81% and 95%), or [Bayly et al. 1998] (90%). In particular, the values of $\delta_{LVD} \cdot N$ are more closely related to the degrees of freedom than those of $\delta \cdot N$.

As it quantifies the decay of remaining variances of the linear principal components of a multivariate data set, $\delta_{LVD}(f)$ is called the *linear variance decay* (LVD) dimension density of the respective data. As $\delta_{KLD}(f)$ is well-defined for $f \in [0, 1]$, this expression allows to calculate $\delta_{LVD}(f)$ for any $f \in (0, 1)$. For this purpose, it is recommended to apply a continuous least-square approach by minimising the functional

$$F_\alpha(f) = \int_{\log(1-f)}^{0} (\delta_{KLD}(x) + \alpha x)^2 e^x dx \tag{3.4}$$

with respect to α (here, the transformation $x = \ln(1-\phi)$ has been used). One easily convinces oneself that $F_\alpha(f)$ has (for any value of f) a unique global minimum at

$$\alpha_{min}(f) = -\frac{\int_{\log(1-f)}^{0} \delta_{KLD}(x) x e^x dx}{\int_{\log(1-f)}^{0} x^2 e^x dx} \tag{3.5}$$

which is easily computed by separately evaluating the integrals over all ranges in x where $\delta_{KLD}(x)$ has a constant value. This minimum, $\alpha_{min}(f)$, is then a reasonable estimate of the exponential decay scale δ such that $\delta_{LVD} = \alpha_{min} / \log_{10} e$.

Alternatively to this continuous least-square approach, δ_{LVD} may also be estimated using the discrete values of the remaining variances only (with a corresponding uncertainty), giving rise to a pointwise defined measure in dependence on $f = p/N$. Using an approach equivalent

to the one described above minimising

$$G_\beta(p_{max}) = \sum_{i=1}^{p_{max}} \left(\log_{10} V_r(i/N) + \frac{i}{N\beta} \right)^2 , \tag{3.6}$$

one finds that

$$\beta_{min}(p_{max}) = -\frac{1}{N} \frac{\sum_{i=1}^{p_{max}} i^2}{\sum_{i=1}^{p_{max}} i \log_{10} V_r(i/N)} \tag{3.7}$$

which may be identified with $\delta_{LVD}(p_{max}/N)$. However, when considering δ_{LVD} *dynamically*, the advantage of the continuous least-square estimate is obvious as a fixation of the explained fraction of variance f yields more comparable results than the restriction to a particular p_{max} where the assigned number of leading eigenmodes may cover a very different amount of information.

By considering its above definition, it is immediately clear that the LVD dimension density still depends on the cutoff parameter f, i.e., gives only a relative dimension density estimate again. In contrast, the appealing alternative of obtaining a parameter-free measure by, e.g., taking the minimum or maximum of δ_{LVD} over all values of f has severe disadvantages. In particular, as it is visualised in Fig. 3.2, there is only a local minimum and maximum of $\delta_{LVD}(f)$ for f within the open interval $(0,1)$ as $\log(1-f) \to 0$ for $f \to 0$ ($\delta_{LVD} \to +\infty$) and $\log(1-f) \to -\infty$ for $f \to 1$ ($\delta_{LVD} \to 0$). Moreover, the local minimum of δ_{LVD} taken over all $f \in (0, f_{max})$ occurs always at $f = 1-V_r(p)$ for a suitable $p \in \{1, \ldots, N-1\}$. Thus, a dynamic characterisation of the record by this local minimum LVD dimension density is not suitable as it may occur at completely different values of f (possibly even changing discontinuously if the associated value of p changes with time).

As a consequence of the features discussed above, one should consider δ_{LVD} always as a relative dimension estimate corresponding to a particularly chosen, fixed value of $f \in (0,1)$, which is clearly bounded from 1 (otherwise, the exponential decay model for the remaining variances would immediately lose its meaning). However, although it still shares this disadvantage with the KLD dimension density, δ_{LVD} is much more sensitive with respect to minor changes in the correlations of the component time series and simultaneously applicable to very small data sets. In the following, the corresponding features and limits of this approach are demonstrated in some detail.

3.3 Application to Stochastic Component Time Series

To study the performance of the KLD-based dimension estimates, the KLD and LVD dimension densities are firstly applied to different artificial data sets. In particular, the behaviour of both measures is studied in the limit of small data sets (i.e., either N or M is comparable to typical geological time series).

3.3.1 Independent Standardised Gaussian Components

The behaviour of the eigenvalues of random covariance matrices in the Gaussian case with a limited amount of data has yet been extensively studied both analytically and numerically (for an overview and further references, see chapter 5 of [Preisendorfer 1988]). In particular, there are analytic expressions for the probability of eigenvalues of such matrices. The resulting logarithmic eigenvector curves show a quasi-exponential decay of values steepening towards the major components as well as towards the components with the smallest variances. This fact is

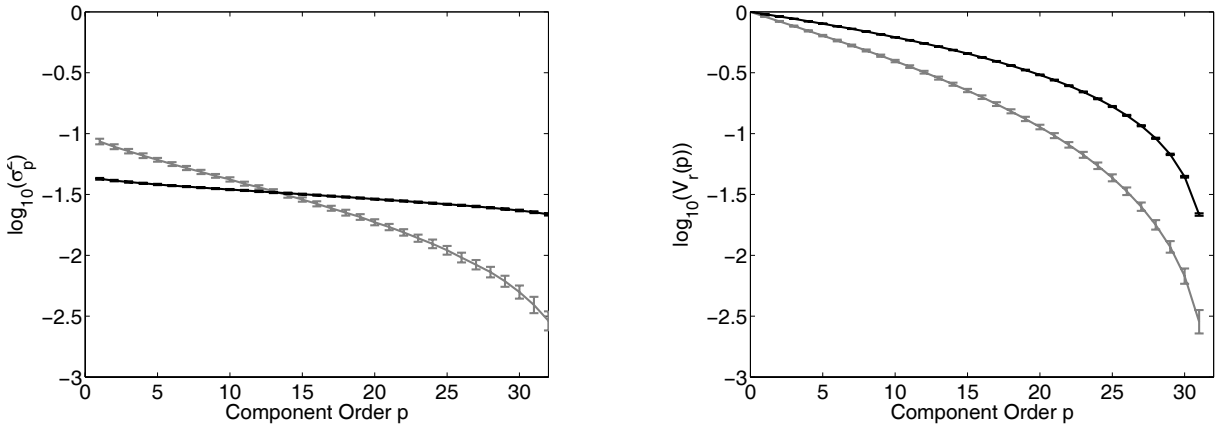


Figure 3.3: Scaling of the component variances σ_p^2 (left panel) (LEV curve) and the corresponding remaining variances $V_r(p)$ (right panel) for standardised Gaussian random data with $N = 32$ and $M = 60$ (gray lines) as well as $M = 1000$ (black lines) points in time, resp. The displayed error bars correspond to the standard deviations of the values from 100 realisations.

represented by numerical calculations on random matrices with standardised Gaussian components (i.e., components with unit variances) displayed in Fig. 3.3 which resemble the results of [Preisendorfer 1988], p. 240. The corresponding decay curves of the remaining variances $V_r(p)$ which are additionally shown in the figure start to significantly deviate from an exponential decay law much earlier (i.e., at lower numbers of components considered) than the corresponding LEV curve. Nonetheless, for the major components, the exponential model appears to be a reasonable approximation.

3.3.2 Independent Non-Standard Gaussian Components

Generalising the above results about the signatures of finite realisations of Gaussian processes in the eigenvalues and remaining variances of the associated covariance matrix, one may consider the more natural case of non-standard components, i.e., components with variances deviating from unity. Fig. 3.4 shows three different examples which may approximate real-world scenarios with component variances decaying exponentially ($\sigma_i^2 = \exp(-\frac{i}{N}/\delta)$), algebraically ($\sigma_i^2 = 1/(\frac{i}{N}/\delta)$), and linearly ($\sigma_i^2 = (N + 1 - i)/N$). In all these cases, it is observed that the corresponding "true" component variances are well approximated even in the case of rather short realisations ($M = 60$). In addition, it is shown that for the major part of the total variance ($f \gtrsim 90\%$), the remaining variances are mainly described by an exponential decay model as supposed in the previous section. In the case of component variances which follow an exponential distribution, the exponential decay law seems to hold exactly in the limit of large realisations of the corresponding processes ($M \to \infty$).

3.3.3 Behaviour of Variances in the Presence of Additive Noise

Next, the influence of additive Gaussian white noise on the eigenvalues and remaining variances of the covariance matrices is studied. As an example, the component variances are prescribed to fixed, exponentially decaying values $\sigma_p^2 = \exp(-\frac{p}{N}/\delta)$. In this case, additive noise dominates the decay of the eigenvalues only on scales where σ_p^2 is reasonably smaller than the noise

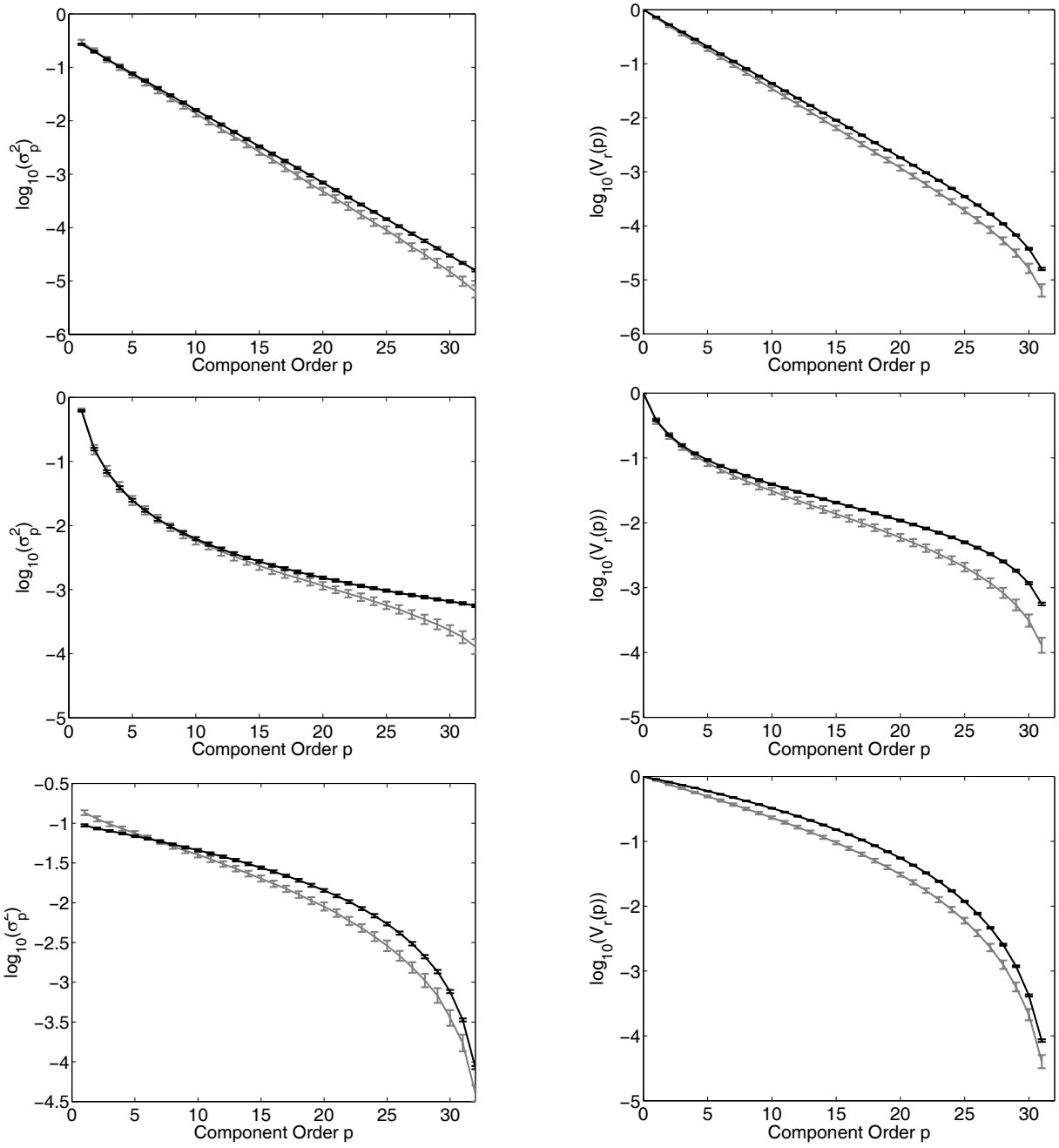


Figure 3.4: Scaling of the component variances σ_p^2 (left panels) (LEV curve) and the corresponding remaining variances $V_r(p)$ (right panels) for independent Gaussian random data with $N = 32$ and $M = 60$ (gray lines) as well as $M = 1000$ (black lines) points in time, resp., with an exponential (upper panels), algebraic, and linear (lower panels) decay of the component variances σ_i^2. The displayed error bars correspond to the standard deviations of the values from 100 realisations.

variance σ^2 (which is in this example related to the fact that both, signal and noise, are the same kind of process). In contrast to the eigenvalues themselves, the remaining variances are much more sensitive to the noise and show remarkable deviations already for $(V_r(p)/\sigma)^2 \sim 1$. For component orders p where the eigenvalues and remaining variances are smaller than these respective thresholds (whose values are closely related to the specific setting), the noise leads to a significant change of the slope of the corresponding decay curves shown in Fig. 3.5. Hence, the decay at these minor components is mainly described by the noise.

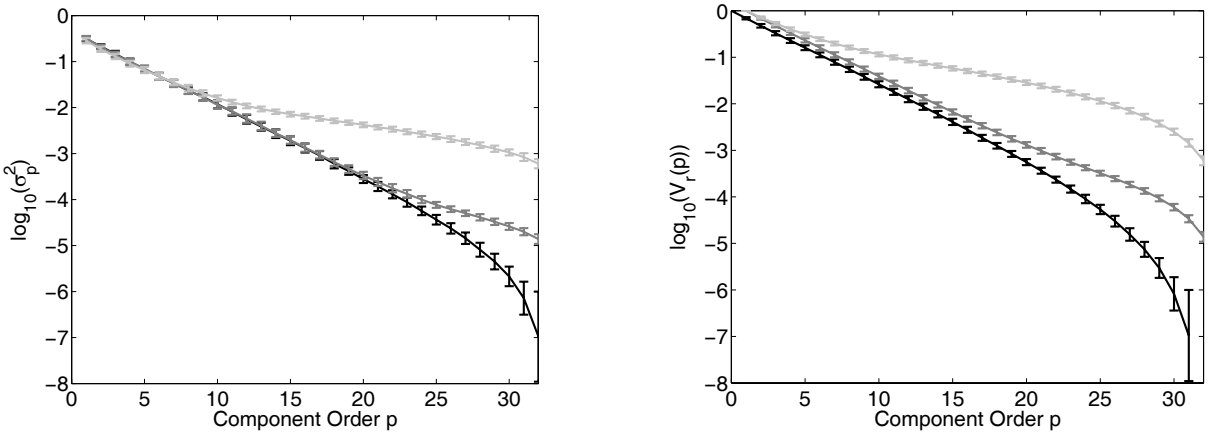


Figure 3.5: Scaling of the component variances σ_p^2 (left panel) (LEV curve) and the corresponding remaining variances $V_r(p)$ (right panel) for Gaussian random data with $N = 32$ and $M = 60$ with component variances $\sigma_p^2 = \exp(\frac{p}{N}/\delta)$ ($\delta = 0.2$) without noise (black lines) and subjected to additive Gaussian white noise with $\sigma^2 = 0.01$ (dark gray lines) and $\sigma^2 = 0.1$ (light gray lines), resp. The displayed error bars correspond to the standard deviations of the values from 100 realisations.

The different sensitivity of both, the eigenvalues of the covariance matrix and the corresponding remaining variances, is reflected by a larger sensitivity of the LVD dimension density against additive noise compared to that of the KLD dimension which is - as a coarse-grained estimate - much more robust against reasonably small changes of the covariance structure of the data. In Fig. 3.6, the behaviour of both measures and their respective uncertainties is systematically studied as a function of both, the cutoff variance fraction f and the noise amplitude σ^2. In particular, one observes that δ_{KLD} *in general* increases with f, whereas its values are relatively slowly increasing when increasing the noise with f kept fixed. In contrast to this behaviour, δ_{LVD} changes (for sufficiently large f) only slowly when increasing the cutoff f, but is still sensitive to changes of the amplitude of the applied noise. Note, however, that in the case of δ_{LVD}, the cutoff fraction f has to be chosen sufficiently high to avoid the strong and unbounded increase in the values of this measure for $f \to 0$ (cf. Fig. 3.2). Concerning the uncertainty of both dimension estimates, it is found that these are of similar orders of magnitude with maximum values at parameters where the corresponding measures have a large gradient.

3.3.4 Non-Gaussian Components

For completing the study of matrices with independent stochastic components, one may additionally consider different generalisations of the above settings. In particular, the effect of instationarities in the data (for example, trends) on the computed dimension densities has to

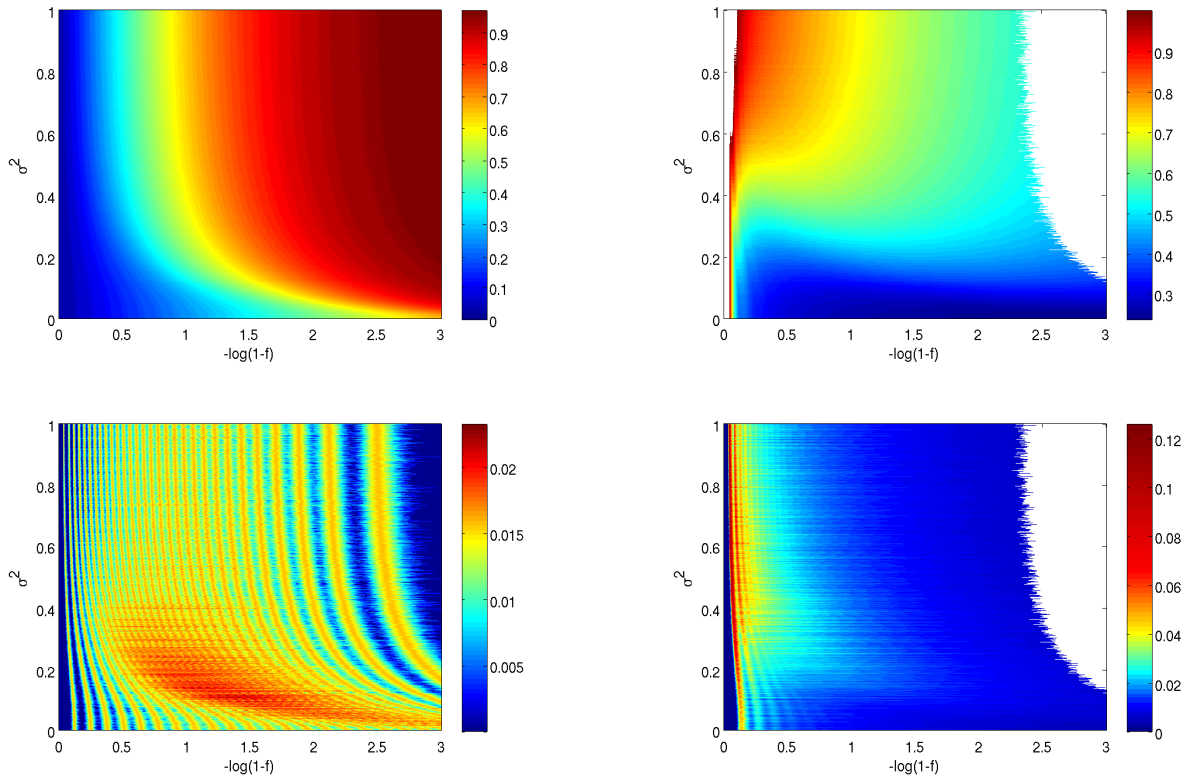


Figure 3.6: Upper panels: Color-coded representations of δ_{KLD} (left) and δ_{LVD} (right) for Gaussian random data with $N = 32$ and $M = 60$ with component variances $\sigma_p^2 = \exp(\frac{p}{N}/\delta)$ ($\delta = 0.2$) as a function of the cutoff level f for different additive noise amplitudes σ^2. Lower panels: The respective uncertainties of both dimension estimates (left: δ_{KLD}, right: δ_{LVD}), expressed by the standard deviations of the computed values from 100 realisations of the respective system for each parameter. White areas correspond to parameters where δ_{LVD} either could not be computed (very large f) or gave artificially high values > 1 (very small f).

be considered. Of course, for real-world data sets as in geosciences, non-stationarity will have an effect on the number values of the measures introduced above. However, as there are reasonable and simple methods for an appropriate trend removal, it is recommended to use these methods before computation. In the case of a dynamic characterisation of the data (i.e., a calculation of dimension estimates for reasonably short windows in time), one may typically assume stationarity and the dynamics on these small scales being statistically relevant.

In the above considerations, only the case of Gaussian components has been considered. However, when studying standardised data (i.e., component time series with unit variances), the actual distribution of the data is less relevant. As a particular example shown in Fig. 3.7, the behaviour of data sets with uniformly distributed components has been studied, which cannot be statistically distinguished from that of Gaussian components with respect to the error bars originated from the finite size of the data sets considered.

In real-world data sets, component time series may have different physical dimensions or values with different magnitudes. In such cases, an application of KLD-based dimension estimates without an appropriate rescaling has to fail as the components with the largest magnitudes of

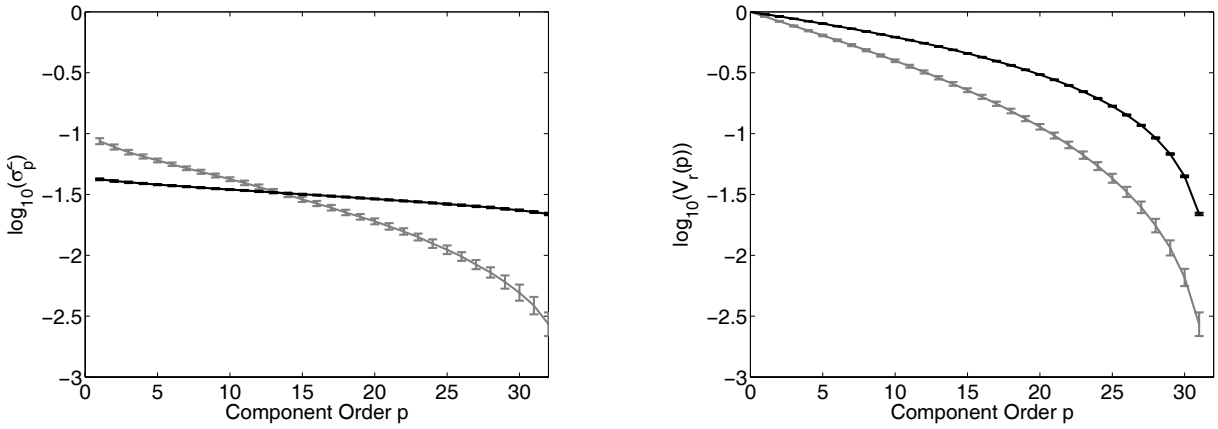


Figure 3.7: Scaling of the component variances σ_p^2 (left panel) (LEV curve) and the corresponding remaining variances $V_r(p)$ (right panel) for stochastic data taken from uniform distribution on $[0, 1]$ with $N = 32$ and $M = 60$ (gray lines) as well as $M = 1000$ (black lines) points in time, resp. The displayed error bars correspond to the standard deviations of the values from 100 realisations.

the recorded values dominate the covariance structure. Applying a normalisation to unit variances (the zero means are a general prerequisite of our method) in the case of Gaussian random matrices with distributed component variances necessarily leads to a shift and deformation of the decay curves of both eigenvalues and remaining variances, resulting in a slower decay and a corresponding increase of the computed dimension density estimates. However, if one wishes to use δ_{KLD} and δ_{LVD} as *relative* dimension densities only, this point is not crucial as long as we are only interested in the *qualitative* change of their corresponding values when the covariance structure of the underlying data is modified.

3.4 Application to Subsets of Large-Scale Systems

The case of completely stochastic component time series discussed so far is rather generic, whereas observational data from geoscientific systems are likely to have some deterministic, but eventually high-dimensional chaotic components. To demonstrate the power of KLD-based dimension estimates for such data sets, one may study their performances for systems modelling the behaviour of spatio-temporal chaos. A particular (linear) approach to construct such a system with a prescribed dimension density $d \in [0, 1]$ has been introduced by [Politi and Witt 1999]. For this purpose, the basis $\{F_1, \ldots, F_n\}$ of a sufficiently high-dimensional Fourier space (i.e., n is large, here $n = 1000$) is considered which may be expressed as

$$F_{kj} = \begin{cases} 1/\sqrt{n}, & \text{if } k = 1, \\ \sqrt{2/n}\cos\left(\frac{2\pi}{n}\left[\frac{k}{2}\right] j\right), & \text{if } k > 1 \text{ and odd}, \\ \sqrt{2/n}\sin\left(\frac{2\pi}{n}\left[\frac{k}{2}\right] j\right), & \text{if } k \text{ even}, \end{cases} \tag{3.8}$$

where $[\cdot]$ denotes the integer part. $j = 1, \ldots, N \leq n$ gives the "spatial" position on a regular one-dimensional lattice which is associated to each respective component time series of the resulting

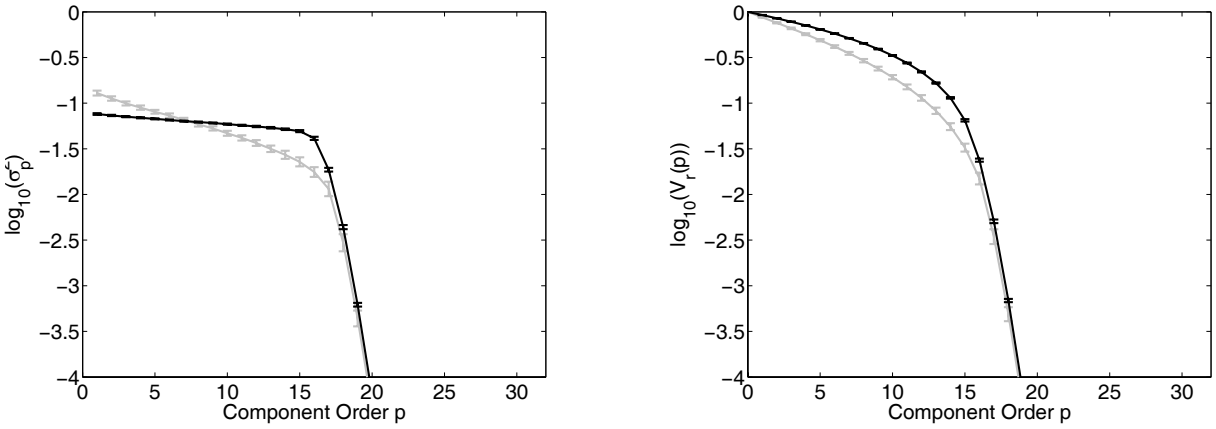


Figure 3.8: Scaling of the component variances σ_p^2 (left panel) (LEV curve) and the corresponding remaining variances $V_r(p)$ (right panel) for the model system for spatio-temporal chaos ($d = 0.5$) with $N = 32$ and $M = 60$ (gray lines) as well as $M = 1000$ (black lines) points in time, resp. The displayed error bars correspond to the standard deviations of the values from 100 realisations.

multivariate data set constructed as

$$x_{ij} = \sum_{k=1}^{dn} \xi_{ik} F_{kj}. \tag{3.9}$$

Here, ξ_{ik} (with $i = 1, \ldots, M$ corresponding to the position in time) is a set of random numbers taken from an appropriate distribution. If $|\xi_{ik}| < 1$, the set of values x_{ij} is contained in a dn-dimensional hypercube and forms a $M \times N$-dimensional data matrix. If M is sufficiently large, the eigenvalues σ_p^2 of the associated covariance matrix (which has a Toeplitz structure) show an abrupt decay at the component index dn, corresponding to the dimension of the underlying hypercube (see Fig. 3.8).

As an example, in the following the ξ_{ik} are taken from a uniform distribution on $[-3^{1/3}, 3^{1/3}]$. This setting corresponds to the system originally studied by [Politi and Witt 1999] both analytically and numerically. Fig. 3.9 shows the computed values of δ_{KLD} and δ_{LVD} for realisations containing $M = 100$ data. One firstly observes that δ_{KLD} fits the true dimension of the system better than δ_{LVD}, whereas the latter one is preferable for detecting small changes within the system. Moreover, for the considered system, the LVD dimension density shows a different behaviour compared to the case of random matrices: Whereas δ_{KLD} increases with increasing f due to its definition, δ_{LVD} is found to decrease in this example which is caused by the particular distribution of eigenvalues in the considered model system.

Concerning the dependence on the length M of the component time series shown in Fig. 3.10, it is found that for suitably large values of M ($M \leq n = 1000$), the KLD dimension density approaches constant values. However, for short time series, the elements of the correlation matrix are rather uncertain. Consequently, as δ_{KLD} approaches only discrete values, this can lead to changes of the computed values or their corresponding uncertainties with varying M. Unlike δ_{KLD}, the LVD dimension density δ_{LVD} is much more sensitive to even small changes of the length of the time series and indicates that an increasing size of the record leads to a gradually increasing dimension estimate, i.e., more components are found to be significant. Only for very long time series $M \geq n$, the computed values slowly saturate. In general, for

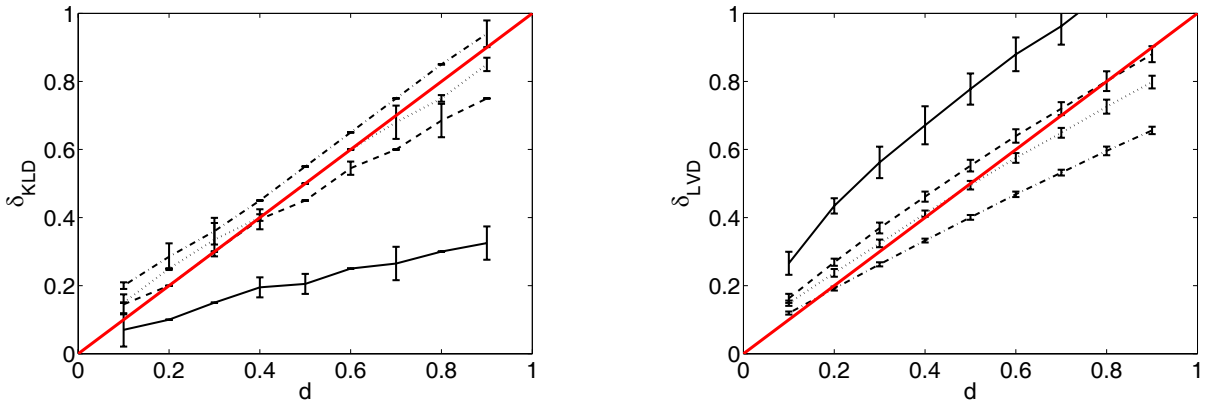


Figure 3.9: δ_{KLD} (left panel) and δ_{LVD} (right panel) with their corresponding 95% confidence levels from 50 realisations of the surrogate data with $N = 20$ and $M = 100$ in dependence of the predescribed dimension density d (note that for some parameters, all realisations gave the same δ_{KLD} such that the corresponding error bars are missing). Truncation levels have been chosen as $f = 0.5$ (solid), 0.9 (dashed), 0.95 (dotted), and 0.99 (dash-dotted). The red diagonal lines represent the true values of d.

both dimension estimates it is found that the saturation occurs for shorter time series if f is increased. Moreover, if f and M are both chosen sufficiently large, both δ_{KLD} and δ_{LVD} recover the prescribed dimension density d of the system.

The latter result may have important implications for real-world observational data sets: to be able to compare the correlation structure of two records by means of KLD-based dimension estimates, these must contain the same number of components and the same number of observations is time. This restricts the applicability of our approach to geoscientific data sets as for example, due to a different sampling in the time domain, for many geological records one cannot consider fixed windows in time as these may contain different numbers of data. Similar problems are expected to occur if there are missing data in a record. Thus, it is recommended to consider always data (sub)sets with an equal number of observations when applying dimension estimates to characterise temporal variations of the correlation structure of observational records.

When considering variations of the true system dimension d, the "optimum" truncation level f to recover d by the respective dimension estimates increases with increasing d in the case of the KLD dimension density δ_{KLD}, whereas for δ_{LVD}, the opposite behaviour is found (see Fig. 3.11). Note that the number values of both dimension density estimates depend on the specific setting, i.e., the choice of M and N.

The discrete values of δ_{KLD} lead to oscillations of the optimum truncation level f with varying d, whereas δ_{LVD} (having continuously distributed values) changes with d is a much smoother way, but yields (for our specific setting and "typical" values of f) a clearly worse quantitative estimate of d than δ_{KLD}. In general, for fixed f both dimension density estimates detect changes in the true dimension of the system. As the associated changes of their respective values are discrete in the case of δ_{KLD}, but continuous for δ_{LVD}, it is suggested that actually the latter measure is better suited for qualitatively detecting and describing changes of the correlation structure in multivariate data sets.

Concerning the respective uncertainties of both measures also shown in Fig. 3.11, a qualitatively similar behaviour is observed where the largest uncertainties occur at high values of d

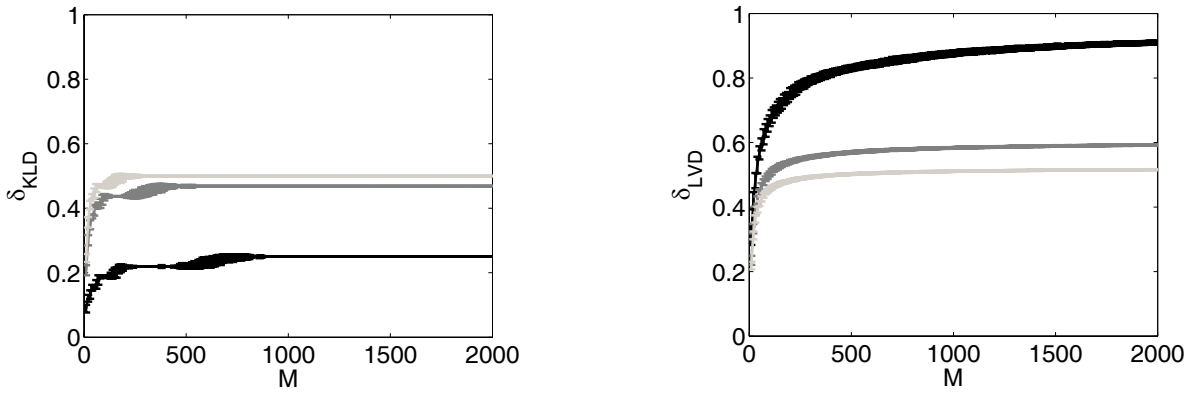


Figure 3.10: Scaling of the dimension estimates δ_{KLD} (left panel) and δ_{LVD} (right panel) for the model system for spatio-temporal chaos ($d = 0.5$) with $N = 32$ components as a function of the length M of the record for $f = 0.5$ (black), $f = 0.9$ (dark gray) and $f = 0.95$ (light gray). The displayed error bars correspond to the standard deviations of the values from 100 realisations.

and moderate cutoff values f (in our specific setting, $f \leq 0.5$ for δ_{LVD} and $f \approx 0.9$ for δ_{KLD}), whereas they decrease with increasing cutoff level f and decreasing dimension density d of the system considered. In addition, one finds that the uncertainty of the KLD dimension density shows small-sized patterns clearly related to the discrete values of this measure, whereas for the LVD dimension density, the uncertainty behaves again much smoother, but is at least for moderate values of f significantly higher than that of δ_{KLD} caused by sensitivity of δ_{LVD} for small f (see Fig. 3.2).

The equal choice of the distribution of the ξ_{ik} for the different components i of the model may appear rather artificial. One may easily modify the model by allowing different distributions for the Fourier coefficients belonging to any component. Fig. 3.12 shows how the resulting dimension densities change when the ξ_{ik} from the standard setting are additionally weighted by a factor of $(dn + 1 - k)/dn$, corresponding to linearly decreasing component variances. With respect to Fig. 3.10 where the corresponding results for the standard setting are displayed, one observes that the values of both, δ_{KLD} and δ_{LVD}, are significantly smaller for any choice of the cutoff variance fraction f, while their values already saturate for shorter realisations of the system with $M \ll n$.

3.5 Sedimentology of the Cape Roberts Project

As a particular example illustrating the power of the KLD-based dimension estimates for the analysis of multivariate geoscientific data, the corresponding measures have been applied to multivariate geological records from marine sediments obtained within the framework of the Cape Roberts project offshore the East Antarctic coast. The main objective of this campaign (consisting of three scientific drillings at slightly different locations) was a detailed study of glaciation and deglaciation intervals in the antarctic region in a time interval between about 17 to 34 million years before present. During this time interval, the global mean temperature was significantly higher than today with long-term temperature fluctuations caused by orbital cycles (for more details and a list of references, see [Naish et al. 2001]). The superimposed successive climate change is believed to be essentially controlled by the opening and closure of

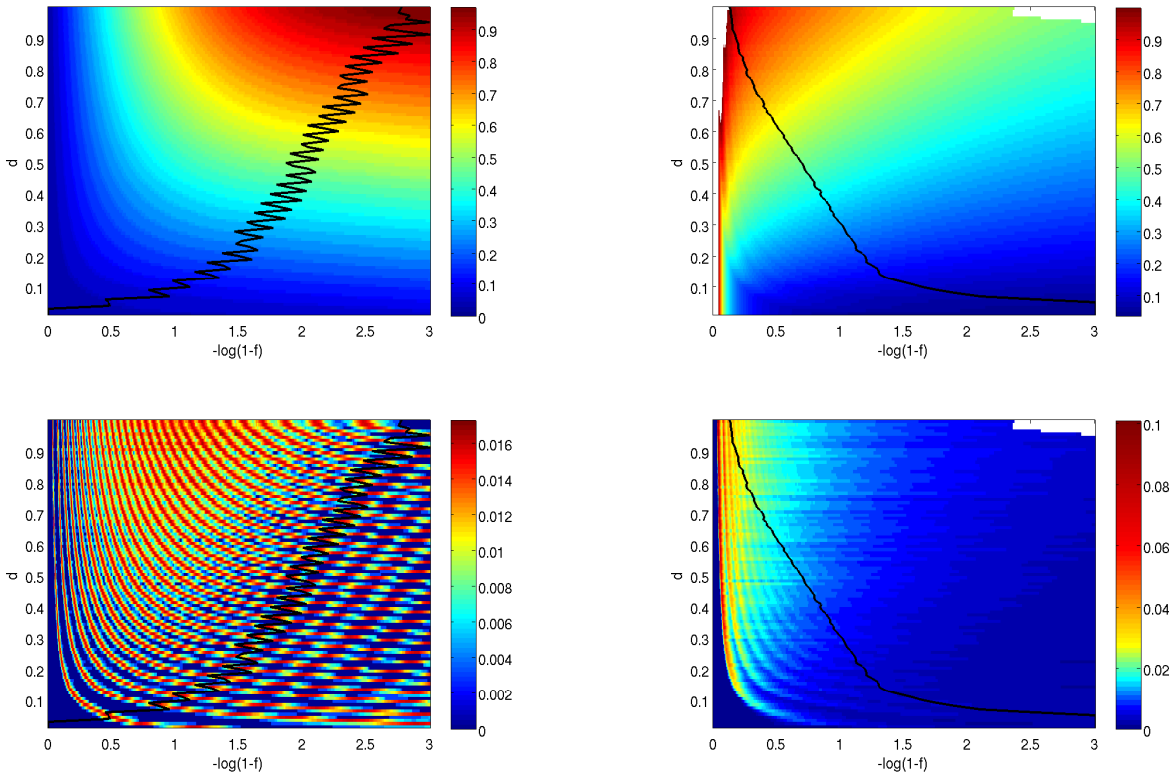


Figure 3.11: Upper panels: Color-coded representations of δ_{KLD} (left) and δ_{LVD} (right) for data from the space-time chaos model with $N = 32$ and $M = 60$ as a function of the cutoff level f for different prescribed dimension densities d of the system. Lower panels: The respective uncertainties of both dimension estimates (left: δ_{KLD}, right: δ_{LVD}), expressed by the standard deviations of the computed values from 100 realisations of the respective system for each parameter. Black lines correspond to cutoff levels f for which the prescribed dimension density d is recovered by the respective dimension estimates. White areas correspond to parameters where δ_{LVD} either could not be computed (very large f) or gave artificially high values > 1 (very small f).

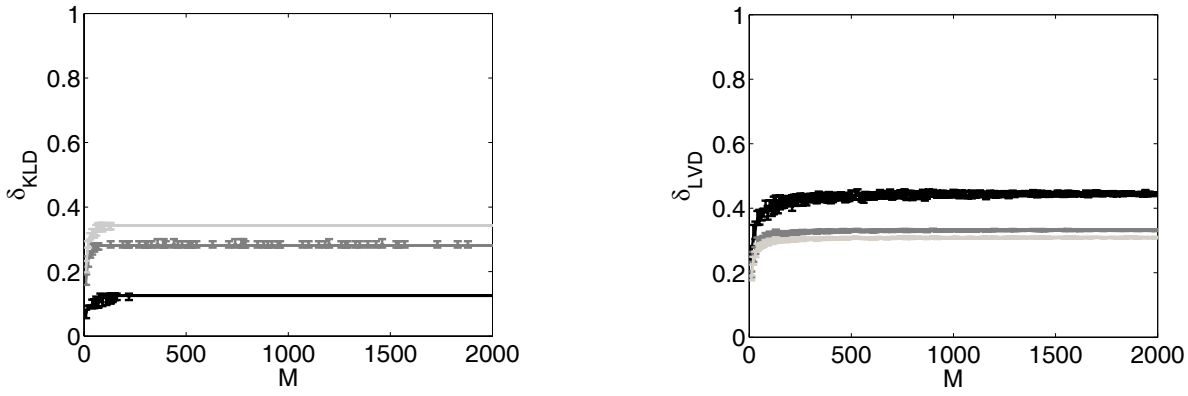


Figure 3.12: Scaling of the dimension estimates δ_{KLD} (left panel) and δ_{LVD} (right panel) for the model system for spatio-temporal chaos ($d = 0.5$) with $N = 32$ components as a function of the length M of the record for $f = 0.5$ (black), $f = 0.9$ (dark gray) and $f = 0.95$ (light gray). The displayed error bars correspond to the standard deviations of the values from 100 realisations.

ocean passages as a consequence of tectonic activity, which have lead to dramatic changes in the oceanic circulation.

3.5.1 Cenozoic Climate Evolution

The geologic history of the Earth is devided on different levels into eons, eras, periods, epochs, and stages. Eons mainly reflect the different parts of the planetary evolution on a long time scale, hence, the current eon, the Phanerozoic, covers the time interval from about 542 Myr BP (million years before present) which is characterised by the significant development of "macroscopic" life (after the so-called "Cambrian explosion"). The Phanerozoic is devided into the eras Palaeozoic, Mesozoic, and Cenozoic, the latter one starting at about 65.5 Myr BP.

Historically, the Cenozoic is devided into the Tertiary and Quaternary, however, these periods are today mainly referred to as Palaeogene and Neogene, which follows an approach towards a unified nomenclature for the entire geological time scale. The Cenozoic contains the Palaeogene epochs of Palaeocene (65.5 - 55.8 Myr BP), Eocene (55.8 - 33.9 Myr BP), and Oligocene (33.9 - 23.03 Myr BP) as well as the Neogene epochs of Miocene (23.03 - 5.33 Myr BP), Pliocene (5.33 - 1.81 Myr BP), Pleistocene (1.81 Myr to 11.4 kyr BP), and Holocene (reaching until present day)[1].

The Cenozoic climate is characterised by a successive transition from rather warm, tropical conditions towards the colder climate of present day. The corresponding transitions are essentially governed by tectonic activity, i.e., the shifting of the different continental plates relative to each other and the resulting opening and closing of oceanic gateways. As the oceanic water masses are a strong heat capacitor, changes in the ocean circulation are today believed to have major influence on the global heat budget and thus are a major mechanism for controlling the large-scale climate. In particular, the opening of the Tasmanian gateway and the Drake Passage between South America and the Antarctic peninsula are closely related to the strong enhancement of glacial activity in the Southern hemisphere, i.e., the occurrence of antarctic glaciations,

[1]The values given here have been taken from http://en.wikipedia.org/wiki/Geologic_time_scale and may differ according to the respective reference considered.

whereas the closing of the Central American seaway during the Pliocene is believed to be one important ingredient in the climatic puzzle leading to the glaciation of the high latitudes of the northern hemisphere.

About 30 years ago, [Kennett 1977] established the hypothesis of a link between the opening of the Drake Passage on the one hand and the initiation of the Antarctic Circumpolar Current (ACC) and the development of ice sheets on Antarctica on the other hand. During the Palaeocene, Australia and Antarctica were joined. In the early Eocene (about 55 Myr BP), Australia started to drift northward, while a circum-antarctic flow was still blocked by the continental South Tasman Rise and Tasmania. During the Eocene, the southern ocean was relatively warm and the Antarctic continent largely nonglaciated. During the late Eocene (about 39 Myr BP), a shallow water connection developed between the southern Indian and Pacific over the South Tasman Rise.

Although the first significant continental glaciations and eventually sea ice formations date to the late Eocene and early Oligocene, the successive development of a substantial ice cap required the thermic isolation of the continent due to the development of the ACC *after* an opening of deep-water passages though both, the Tasmanian gateway and the Drake Passage. The timing of the opening of Drake Passage to deep water flow as the last contribution to the ACC formation has been a long-standing debate with estimates ranging from around the Oligocene/Miocene boundary [Barker and Burrell 1977, Barker and Burrell 1982, Barker 2001] to the early Oligocene [Lawver and Gahagan 1998, Lawver and Gahagan 2003]. Numerous attempts have been made to constrain the opening by dating the onset of the ACC with different palaeoceanographic proxies (see [Barker and Thomas 2004] and references therein), the debate has endured [Scher and Martin 2004].

The changes in large-scale oceanic circulation patterns are believed to be the major reason for the long-term gradual cooling of the Earth since the beginning of the Cenozoic. Apart from this, different other influences have been discussed, like biogeochemical feedback mechanisms, the contribution of volcanic forcing to the atmospheric greenhouse gas contents, long-term variations of the solar insolation as well as the variability of the orbital parameters on shorter time scales. The latter ones are known as the most promising candidates for triggering climate variability on glacial and sub-glacial time scales such that it is a reasonable hypothesis to link them also to shifts of the climate regime during, for example, the Eocene, Oligocene, and Miocene. Indeed, there are numerous studies which report observations from high-resolution geologic sequences showing the dominant variability with comparable frequencies (e.g., [Flower et al. 1997a, Zachos et al. 1997, Paul et al. 2000, Naish et al. 2001, Zachos et al. 2001, Mallinson et al. 2003]) and discuss the corresponding potential mechanisms for orbit-climate feedbacks.

3.5.2 Location and Objectives

In the sediment core CRP-2/2A (see Fig. 3.13), the probably most remarkable climatic transition in the considered time interval, the Oligocene-Miocene transition (OMT), is well resolved within a long sequence due to high sedimentation rates. The origin and mechanism of the climate change associated to this transition will be discussed later. However, although there is much sediment available for analysis, the data of major palaeoclimatic proxies like trace element abundances or grain-size distributions have only been obtained for very few distinct time slices. Thus, the actual mechanism of the transition is not well resolved in the corresponding component time series. In this contribution, we aim to improve the resolution on the basis of the existing short, but multivariate records in order to get a better understanding about the transitional behaviour. In particular, we focus on the question whether the different climatic

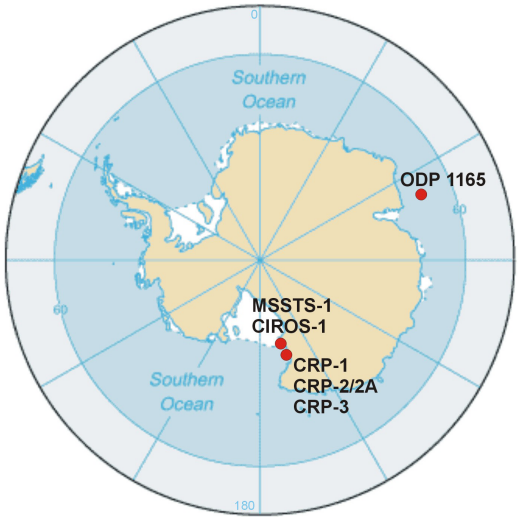


Figure 3.13: Map of the antarctic continent including the locations of the Cape Roberts project drill sites and the preceeding MSSTS-1 and CIROS-1 drillings. In addition, the location of the probably best studied marine core offshore the Antarctic coast (ODP site 1165) [Williams and Handwerger 2005] is displayed.

conditions in the older and the younger part of these records [Naish et al. 2001] are reflected by a varying strength and pattern of interrelationships between the different palaeoclimatic observables.

As a first example, a record of trace element abundances from the CRP-2A core [Krissek and Kyle 2000, Krissek 2004] is studied. The analysed data set consists of 46 parameters measured within 104 slices of the sedimentary core. Due to gaps in the measurement series, the presented analysis is restricted to a homogeneous subset of records of 32 trace elements from 60 time slices. The trace elements abundances were measured by X-ray fluorescence (XRF, 19 elements) and instrumental neutron activation analysis (INAA, 13 elements)[2]. As the absolute abundances of these elements (given in parts per million (ppm)) cover several orders of magnitude, all component time series are firstly standardised to unit variance before subjecting the data set to further analysis. Moreover, when applying the KLD or calculating the corresponding dimension estimates for a subset of measurements, the respective data are similarly standardised to zero means and unit variances of all component time series. In the following, this procedure is always applied when considering observational data.

To get complimentary information, an independent second palaeoclimatic proxy is considered. During the last decades, the statistical analysis of grain-size distributions has become an important tool in palaeoclimatic studies (see Chapt. 4). In particular, the use of KLD (in particular, the amplitude of the dominating components) for describing the information content of the data has been proposed rather early [Davis 1970, Chambers and Upchurch 1979]. Grain-size distributions can be obtained by sieving (mass-frequency distributions) or more advanced optical measurement techniques applied to the suspended material (number-frequency distributions). For the CRP-2/2A core, the corresponding data have been obtained by a SEDIGRAPH 5100 which measures the absorption of x-rays during the sedimentation of the material. The

[2]Among the 32 parameters, the abundances of Arsenic, Thorium and Uranium have been measured twice with both methods.

data [Barrett and Anderson 2000, Barrett and Anderson 2003] consist of relative frequencies of particles in 23 different size classes, which are equally spaced on a logarithmic scale (phi-scale). In total, measurements have been performed on 119 different time slices.

Explicit tables of data sets and an introductory discussion of their palaeoclimatic relevance are given in the references cited above. In addition, the data are freely available from the PANGAEA database.

3.5.3 Analysis of Trace Element Abundances

As a first example, the trace element record of the CRP-2/2A sediment core is investigated. The standardised version of the record is displayed in Fig. 3.14. Considering the eigenvalues of the covariance matrix S of the complete data set shown in Fig. 3.1, one observes a relatively smooth decay resembling that of random matrices with a larger scale at the major components and a more moderate decay at higher-order modes [Preisendorfer 1988]. However, as it is shown in Fig. 3.1, the decay of the remaining variances cannot be related to a superposition of Gaussian white noise processes with the corresponding variances. This behaviour is mainly caused by trends visible even in the original data, and a non-Gaussian probability distribution of the component time series. The behaviour of the associated dimension estimates with varying explained variance fraction f was already shown in Fig. 3.2.

Considering the abundances of the respective trace elements only separately, an exact determination of the starting and end points of transitions recorded in the record (corresponding to different colors in Fig. 3.14) is problematic as different parameters yield a slightly different variability pattern, i.e., have different sensitivity with respect to changing environmental conditions. In the case of trace element abundances, the source region of the sediment is encoded in the entire record where some elements vary only slightly between different sources whereas others show dramatic changes. In Fig. 3.14, the example of Niobium is shown where a remarkable concentration peak is observed at about 130 mbsf probably corresponding to volcanic detritus from the McMurdo Volcanic Group [Krissek and Kyle 2000]. The height of this peak also dominates the associated variability amplitude derived from subsamples of data. Both, mean values and variances show a remarkable trend for the corresponding element abundance corresponding to a (successive) change in sediment provenance between about 400 and 130 mbsf.

A similar behaviour may be observed when considering other particular elements. However, there are elements not showing the corresponding variability. Therefore, one may be interested in whether the variability of *all* measured element concentrations still gives rise to an aggregated variability pattern which is described by the local values of the KLD-based dimension estimates. Hence, the entire multivariate record is considered in terms of its temporary dimension estimates to infer a hopefully better signal of long-term climate change (an approach which is somehow counter-intuitive as the corresponding procedure applies a further strong coarse-graining to the data). In Fig. 3.14, the eigenvalues of the covariance matrix for subsets containing 45 (of 60) successive time slices are shown, giving rise to a decrease of the variances of the major components when considering the younger part of the sediment similar to the trends in the observed abundance of different elements.

Fig. 3.15 shows that the KLD dimension density is not sufficiently sensitive to reflect changes in the strength of interrelationships between the variations of the different chemical components in an appropriate way, whereas δ_{LVD} shows a much more pronounced variability whose significance will be proven next. For this purpose, the LVD dimension density has been computed for different age intervals, showing whether or not varying climate conditions are actually reflected in variations of the number of relevant components. The corresponding results shown in Fig.

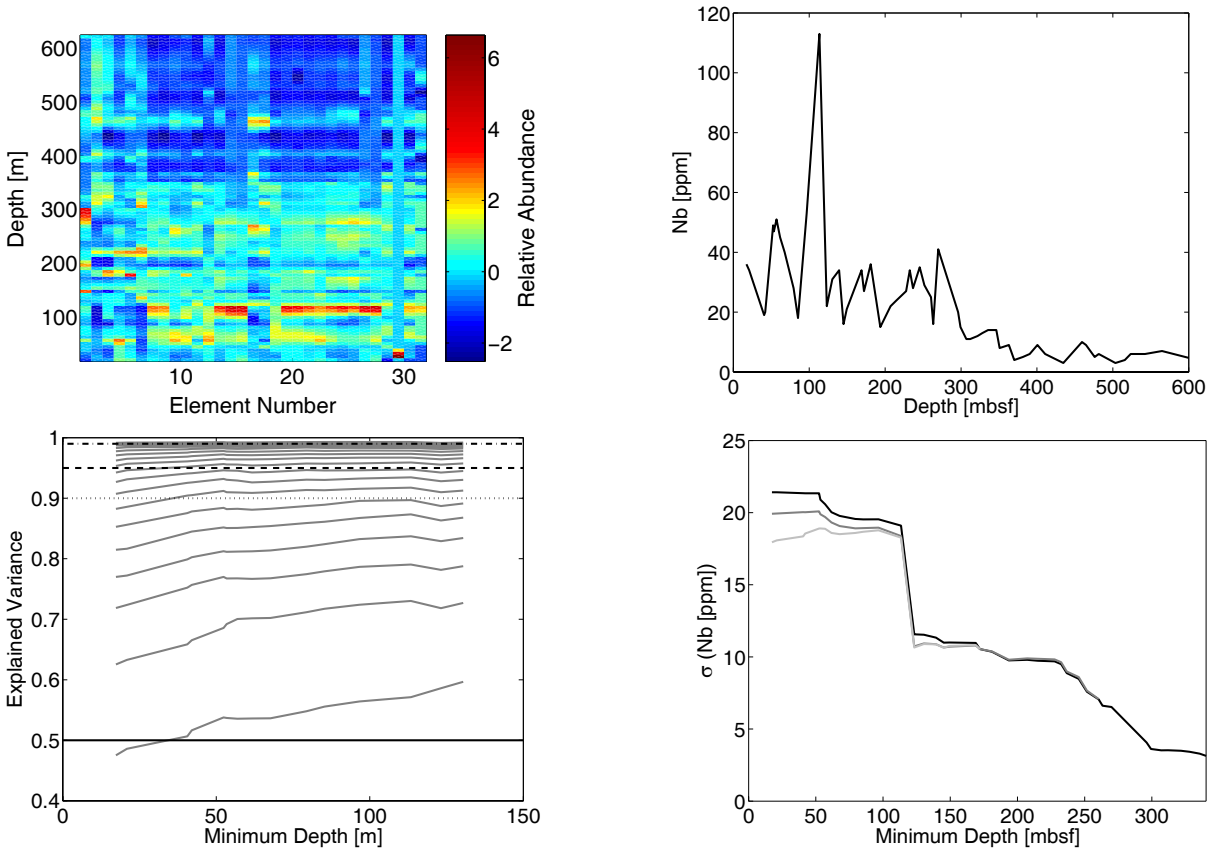


Figure 3.14: Upper panels: Color-coded representation of the normalised abundances of the 32 trace elements (left) and absolute abundance (in parts per million ([ppm])) of Niobium (Nb, element number 15) as a function of depth below seafloor. Lower panels: Left: Explained variances with increasing maximum component number (gray lines) for subsets of $M = 45$ observed samples as a function of the associated minimum depth in the sedimentary sequence. Horizontal black lines correspond explained variances of $f = 0.5$ (solid), $f = 0.9$ (dotted), $f = 0.95$ (dashed), and $f = 0.99$ (dash-dotted). Right: Local variability (standard deviation) of the absolute Nb abundance within time slices of $M = 20$ (black), 30 (dark gray), and 40 (light gray) points.

3.16 underline that the general variability pattern is displayed independent on the particular choice on the width of the considered time windows, M, which indicates significance of the corresponding results.

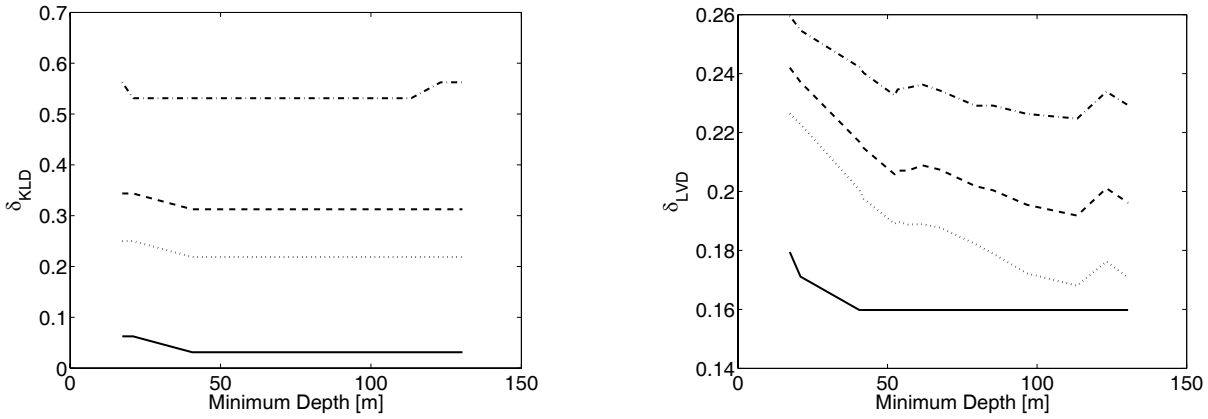


Figure 3.15: δ_{KLD} (left) and δ_{LVD} (right) for subsets of 45 observed samples from the CRP-2A trace element record as a function of the associated minimum depth in the sedimentary sequence for $f = 0.5$ (solid), $f = 0.9$ (dotted), $f = 0.95$ (dashed), and $f = 0.99$ (dash-dotted).

Significant changes of δ_{LVD} occur when sediment from below 400 mbsf (meters beyond sea floor) or above 130 mbsf (see Fig. 3.16) is considered for the analysis, which can be seen in the graphical representation of this measure as a function of the maximum and minimum depth associated to the considered data window, resp. The interval between these two horizons covers a relatively small time window between 24.3 and 23.8 Myr BP (depending on the age estimates based on different measurements), which includes the Oligocene-Miocene transition (OMT). It has to be noted there are rather different age estimates for the OMT, depending on either the alignment to certain reference time scales based on geomagnetic polarity [Cande and Kent 1992, Cande and Kent 1995], orbital tuning [Shackleton et al. 1999, Shackleton et al. 2000] to astronomical cycles [Laskar et al. 1993, Pälike and Shackleton 2000, Laskar et al. 2004], or a combination of both [Billups et al. 2004], or the establishment of a local chronology basing on measurement of appropriate quantities on a substantial amount of samples [Wilson et al. 2002, Roberts et al. 2003, Pfuhl and McCave 2003], see [Flower et al. 1997a, Flower et al. 1997b, Zachos et al. 1997, Paul et al. 2000, Zachos et al. 2001, Naish et al. 2001, Williams and Handwerger 2005].

Following the arguments presented in Sect. 3.5.1, the OMT is characterised by an opening of the Drake passage between South America and the Antarctic continent, which has led to an intensification of the Antarctic circumpolar current and a successive thermic isolation of the continent. These effects caused finally an enhanced glacial variability in the high latitudes of the Southern hemisphere. The qualitative changes recorded in the trace element data are associated with a change of the provenance of the material [Krissek and Kyle 2000] and an enhanced variability of the sedimentation. Both together result in a decrease of the interrelationships between the variability of different trace elements and, consequently, an increase of the dimension of the record.

To further prove the significance of the variation of δ_{LVD} with the age of the considered sediment, the uncertainty of this measure may be estimated. For this purpose, δ_{LVD} was computed for ensembles of slightly perturbed data constructed from the original data set by substituting the data of a single, but randomly chosen time slice with a N-dimensional Gaussian random

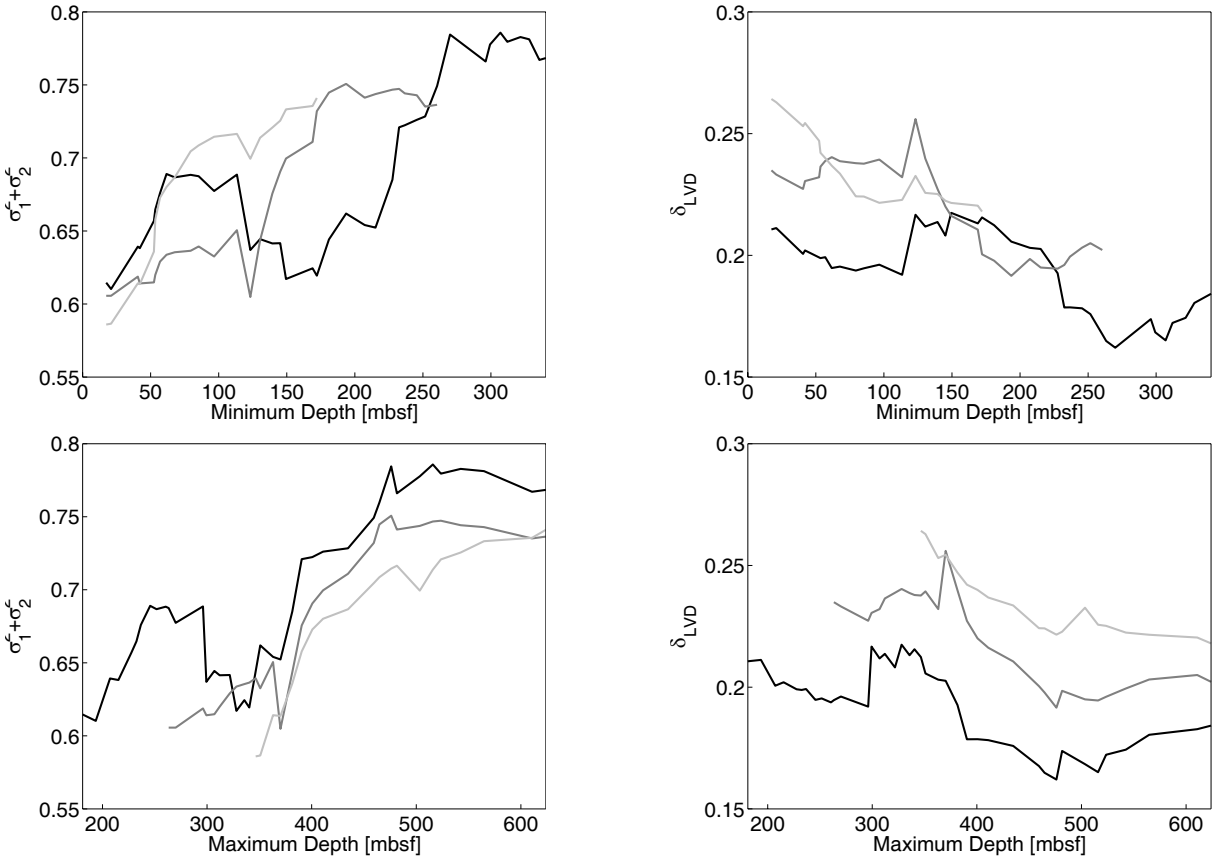


Figure 3.16: Sum of the first two eigenvalues $\sigma_1^2 + \sigma_2^2$ (left panels) and δ_{LVD} (right panels) for the normalised trace element abundances from the CRP-2A core as a function of the minimum (upper panels) and maximum (lower panels) depth (mbsf) of the sediment layer associated to the respective data subsets for sliding window of $M = 20$ (black), $M = 30$ (dark gray) and $M = 40$ (light gray) points in time. Vertical dotted lines indicate the major transitions recorded in the data.

variable. To study the influence of such distortions to δ_{LVD} systematically, a suitably large ensemble of these perturbed data have to be analysed. Fig. 3.17 presents the result for windows containing only 20 time slices: While the expectation of δ_{LVD} is shifted towards higher values with respect to the original data (i.e., the surrogate data are more stochastic), the qualitative behavior remains unchanged. The corresponding confidence levels indicate clearly that the variations in the dimension are significant even for such small numbers of data points, which demonstrates the qualitative robustness of the considered approach.

Although a significant number of components is required to explain a certain fraction of total variance, the temporal variation of the first two leading eigenvalues shows already a pattern similar to the LVD dimension density. For time windows containing $M = 20$ points, Fig. 3.18 shows these eigenvalues and the corresponding eigenvectors. One observes that the behaviour of the first two eigenvalues is completely different. For the older part of the record, the first eigenmode clearly dominates the record, whereas the second one becomes increasingly important when considering data resulting from the time interval associated to the OMT. The corresponding changes in the first eigenvector are mainly reflected by the components associated

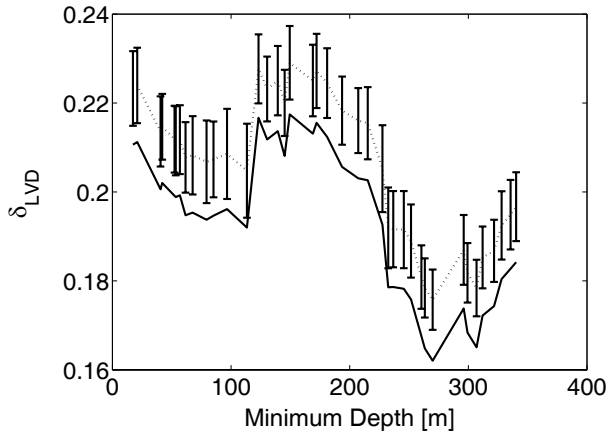


Figure 3.17: Expectation (dotted line) and 95% confidence levels of δ_{LVD} for $M = 20$ points in time calculated from 1000 random substitutions of one complete time slice each compared to the values for the original data (solid line).

to the trace elements Scandium (element number 2), Vanadium (3), Strontium (12), and Barium (18). The latter two ones are also the main recorders of the increase of the first eigenvalue when younger sediment ($\leq$ 23.8 Myr BP) from above 130 mbsf is considered. The onset of the increase of the second eigenvalue is reflected by the eigenvector component associated to Sulfur (element number 1), whereas a number of other components start to change later. These results apparently indicate that the climate change associated to the OMT is particularly pronounced by three elements of the record (S, Sr, Ba), which are also the trace elements with the highest absolut abundances in the record. This result is particularly remarkable as all component time series have been standardised to unit variance *before* our analysis.

One should briefly discuss one issue related to the term "transition": In geosciences, this term is maily used to describe an *abrupt* change of the behaviour of the system. In contrast to this, the physical meaning would rather be that of a process where a (multistable) dynamical system leaves a certain (equilibrium) state and relaxes to another stable solution. Apart from the stochastic components always present in the high-dimensional climate system, the latter interpretation appears to be more useful in the context of the OMT: The opening of a deep-water connection through the Drake Passage was likely not an abrupt, but a gradual process. This is also indicated by the smoothness of the transition recorded in the trace element data and the resulting dimension estimate, giving rise to the assumption of a successive change of provenance between 24.3 and 23.8 Myr BP (the corresponding layer is referred to as the *upper Oligocene* in [Krissek and Kyle 2000]), i.e., during a time interval of several 100,000 years. Considering the entire process as the Oligocene-Miocene *transition* however contradicts the usual point of view in geosciences where the OMT is assigned to a specific age rather than to an age interval, leading to a variety of different numbers (e.g., for the CRP-2/2A core, [Krissek and Kyle 2000] propose the OMT at 130 mbsf whereas [Naish et al. 2001] give a value of 183.7 mbsf).

Apart from the climate change associated to the OMT, the eigenvector analysis seems to indicate further transitions in the climate system. For example, a qualitative change of the second eigenvector in the youngest part of the sediment is found which cannot be assigned to a known climatic transition. To gain a deeper insight into the age interval, one has to consider data from other sources, e.g., the CRP-1 and CIROS-1 cores or, for the older part of climate history, the CRP-3 core. For these locations, however, there are no comparable

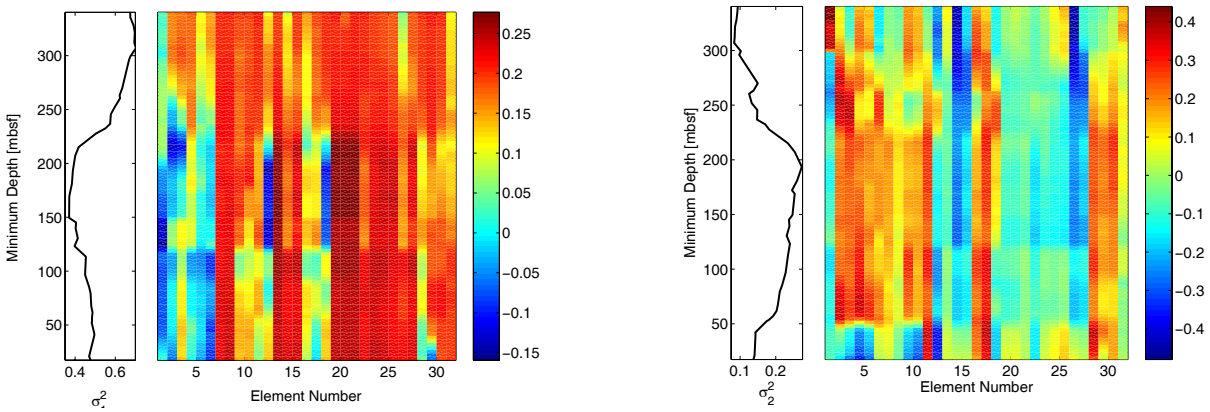


Figure 3.18: The first two eigenvalues σ_i^2 $(i = 1, 2)$ (left panels) and their associated eigenvectors (color-coded representation in right panels) for the normalised trace element abundances from the CRP-2A core as a function of the minimum core depth for sliding windows of $M = 20$ points in time.

records of trace element abundances available: On the one hand, available measurements of other sedimentological fractions (XRF analysis, see [Krissek and Kyle 1998]) or on sand and sandstone (energy dispersive analysis of x-rays (EDAX), see [Armienti et al. 1998]) from the CRP-1 and CIROS-1 cores include either only very few continuously recorded trace elements or too few samples in time. On the other hand, additional continuous geochemical data are available in terms of major element abundances, which do cover only about 10 to 12 different parameters. The latter ones are given in terms of percentages of the corresponding oxide abundances and thus form sets of compositional data which require a special statistical treatment. Similar data are also available for the CRP-2/2A and CRP-3 cores but not considered here.

3.5.4 Analysis of Grain-Size Distributions

Grain-size distributions are another example for compositional data which is used as a complementary source of information. For an arbitrary multivariate data set, a transformation dividing the original data by their respective sum at every point in time leads to a set of compositional data. This situation is present in the case of oxide abundances and grain-size distributions: As there are no absolute, but relative values, the statistically relevant quantities are no longer the component data themselves, but appropriate ratios thereof as the latter ones are invariant under the respective transformation. Many geological data (like grain-size distributions and major element abundances) belong to this class of data constrained to a constant sum in each time slice.

[Aitchison 1986] has demonstrated that there are three equivalent ways of considering either pairwise or centred ratios within a compositional vector. Among these, for a data vector $(x_1, \ldots, x_N)$, the N centred ratios are defined as $x_i^* = x_i / g(x_1, \ldots, x_N)$, where $g(x_1, \ldots, x_N)$ is the geometric mean of the vector. For the analysis of grain-size distributions, it is recommended to consider these centred ratios as they do not give particular weight to any fixed component of the original data set. Typically, one uses the corresponding log-ratio transformed data instead of the centred ratios themselves (i.e., $\log x_i^*$). However, as in the case of grain-size distributions, there are frequent zero "counts" occuring in the data, the consideration of logarithms leads to

numerical instabilities, such that a restriction to the non-logarithmised data is more appropriate.

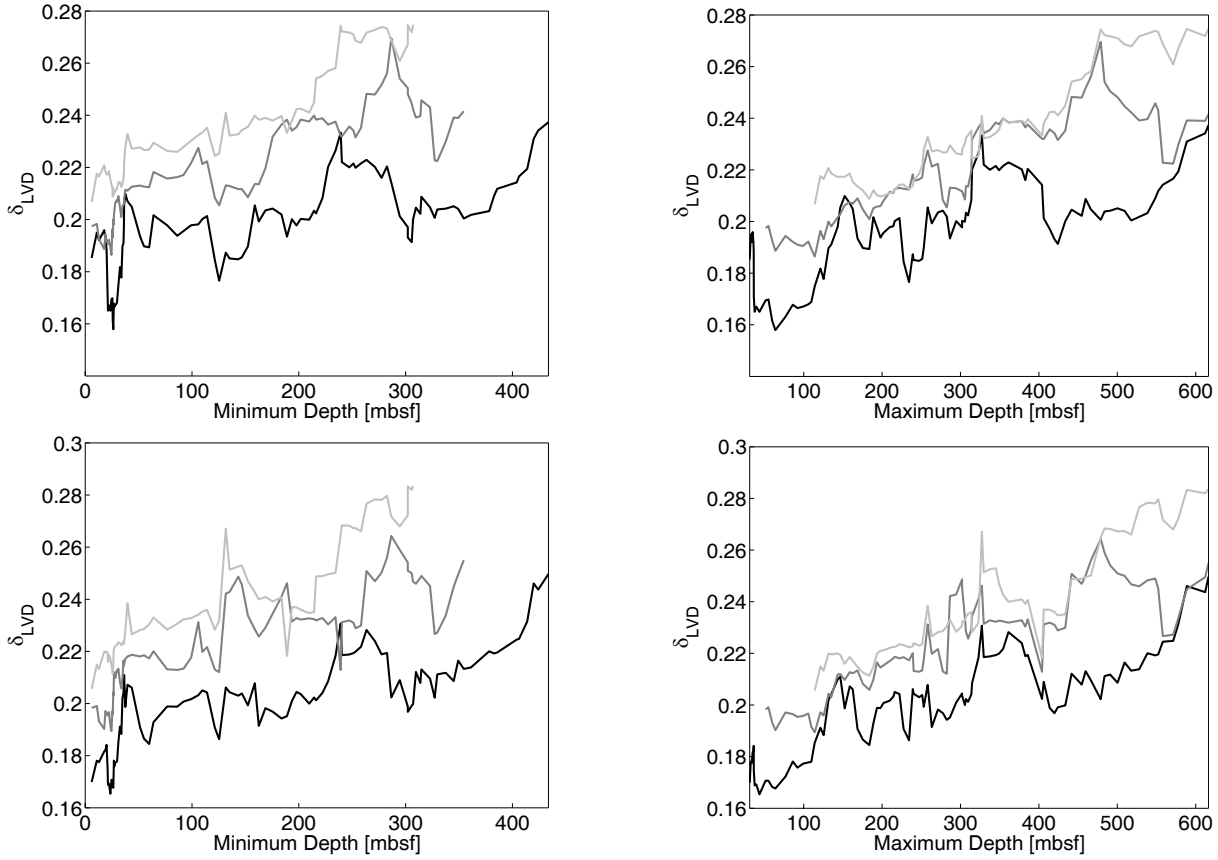


Figure 3.19: δ_{LVD} computed for grain-size distributions from the CRP-2/2A core without (upper panel) and with (lower panel) a transformation of the data to centred ratios. The results are displayed as a function of both, the minimum (left) and maximum (right) depth of the sediment layer considered for sliding windows of $M = 20$ (black), $M = 30$ (dark gray) and $M = 40$ (light gray) points in time. Vertical dotted lines correspond to common features of the three settings representing major climatic transitions recorded in the data.

Considering the results displayed in Fig. 3.19, one firstly observes a much more detailed variability pattern compared to the trace element data discussed in the previous section, which is an effect of the higher total number of time slices where observations have been available. Secondly, the non-transformed data give a more diffuse pattern compared to the transformed ones which underlines the necessity of a transformation for obtaining statistically meaningful results on compositional data. Thirdly, the recorded transitions in the climate system associated to the OMT are consistent with the results trace element record, but are better resolved due to the larger number of time slices. In particular, one observes that the pattern corresponding to the Oligocene part of the sediment is reflected by a strong successive decrease of the LVD dimension, which is followed by an increase when sediment from between 350 and 400 mbsf is considered, which is probably related to the onset of the OMT. Furthermore, the decrease of dimension above 130 mbsf is resolved as in the case of the trace elements, which probably determines the end of the transition with a full development of the antarctic circumpolar current and a resulting provenance change.

As an additional feature, the further decrease of the complexity of interrelationships due to a consideration from sediment from above 50 mbsf is well resolved by the grain-size distributions, underlining the results from the geochemical data and proving their actual relevance. The age associated to the corresponding layer is about $20.3 \ldots 20.4$ Myr BP. In another study from Prydz Bay (ODP Site 1165) [Williams and Handwerger 2005], this age interval was found correspond to the probably most pronounced layer of ice-rafted debris during the Early miocene, which indicates that the transition found in the Cape Roberts data is probably related to a major deglaciation event on the Antarctic continent.

3.6 Related Work

During the last years, several authors have adapted techniques from random matrix theory (RMT) [Mehta 1990] for analysing the equal-time correlation matrix S of empirically obtained multivariate data sets, including financial [Laloux et al. 1999, Plerou et al. 1999, Plerou et al. 2000, Drozdz et al. 2000, Drozdz et al. 2001, Maslov 2001, Plerou et al. 2002, Kwapien et al. 2003], neurophysiological [Kwapien et al. 2000, Seba 2003], and atmospheric time series [Santhanam and Patra 2001]. The corresponding approach offers the possibility to compare the fluctuation properties of the spectrum of S with the analytical results obtained from random matrix ensembles. It is proposed that the part of the spectrum which can be described by random matrices corresponds to random correlations or noise and thus does not reflect any relevant information, while only the eigenvalues which show significant deviations from the RMT predictions (in particular, the largest eigenvalues) represent the "true" correlation structure of the system.

Another approach which is very similar to the one applied in this chapter has recently been proposed by [Müller et al. 2005] who considered also the temporal evolution of the eigenvalues and corresponding eigenvectors of the equal-time covariance matrix S obtained from short windows from multivariate data sets standardised in the same way like in the above presented method, i.e., considering component time series with zero means and unit variances in the considered windows. Instead of considering the decay of remaining variances, the authors of this work proposed two alternative measures: the participation ratio or number of principal components of the i-th eigenvector,

$$N_i^p = \frac{1}{N \sum_{j=1}^{N} |a_{ij}|^4} \tag{3.10}$$

(where the a_{ij} are the expansion coefficients of the i-th eigenvector of S), and the so-called symmetry parameter

$$S_i = \left| \sum_{j=1}^{N} sgn(a_{ij}) |a_{ij}|^2 \right| \tag{3.11}$$

which measures to which extent an eigenvector is generated by constructive or destructive interference of the basis states.

The considered measures may have a sophsiticated meaning, however, they refer only to one (typically, the largest) eigenvalue and thus lose information about the additional information contained in the remaining modes. However, it might be of interest to study the whole spectrum of participation ratios and symmetry parameters, resp., for a given data set. Moreover, [Müller et al. 2005] apply their measures for short time windows, but do not consider the dependence of their behaviour on the particular choice of M, which may be (as it has been shown for the method used in this work) rather crucial for even a qualitative interpretation of the results.

3.7 Open Problems

Considering the examples discussed in this chapter, it turns out that in the case of standardised component time series from model systems without additive noise, the remaining variance is a convex function of the maximum component order p, i.e., the local slope of $V_r(p)$ shows an increasing absolute value when increasing p (see Figs. 3.3, 3.7 and 3.8). In contrast to this, if the data are contaminated by additive stochasticity, this characteristic shape is changed towards a function steepening at both, small and large values of p. The same type of decay is also found in the case if standardised measurements in the geological example as displayed in Fig. 3.1. These observations give rise to two different questions which are not yet answered:

Firstly: Is it possible to give a more sophisticated functional form for the decay of standardised model data? For the rather specific case of independent standardised Gaussian components, an analytic expression for the corresponding component variances has been derived (see [Preisendorfer 1988] and references therein), such that it might be possible to give a similar expression also for the remaining variances at least for this specific case. If a corresponding modified decay law for $V_r(p)$ could be described by only one parameter, this could be used to define a more sophisticated dimension estimate for the multivariate data set which is (in contrast to δ_{KLD} and δ_{LVD}) parameter-free, i.e., independent of the choice of a particular cutoff f or p_{max}, resp.

Secondly: In a similar way, one may ask whether the specific signature of additive noise could be extracted in a similar way at least if these can be independent realisations of a Gaussian process. In particular, if this would be possible, one might use this approach as a new tool for noise-level estimation in multivariate time series. Moreover, if the signature of Gaussian white noise is actually separable in this way, one might modify the potential decay law of the remaining variances in an appropriate way. Ideally, the resulting function would be described by two parameters only, one yielding an estimate of the dimension of the unperturbed underlying system and one describing the noise level. Following the results of Sect. 3.3, it seems likely that in the case of noisy "data", an exponential decay model for the remaining variances fitted only in an interval $[f_{min}, f_{max}]$ would asymptotically yield a constant decay scale (related to the noise) within a certain range of values of f_{min} and f_{max}.

It has to be underlined that the above considerations are currently rather speculative, but may be used as a possible outline for further methodological research on KLD-based dimension estimates. Moreover, there are different other questions to be answered, including the signature of multiplicative stochasticity, auto- or even cross-correlated noise etc. Generalisations of the KLD-based approach to other appropriate methods of statistical decomposition are to be investigated as well, in particular for quantifying the dimensionality of sufficiently long, stationary time series from nonlinear, e.g., spatially extended systems. For such systems, one should also perform a fair comparison to "fully nonlinear" dimension estimates of multivariate time series in terms of both, correctness and computational efficiency.

Chapter 4

Analysis of Grain-Size Distributions from Lake Baikal

4.1 Motivation

During the last years, grain-size distributions have gained increasing interest as a recorder of changes of environmental conditions. The typical size of particles in the sediment and, more general, the distribution of grain sizes depends on the origin of the material and the mechanisms of transport to the final deposition. Records can be obtained from lacustrine and marine cores, but also from continental sequences distributed all around the world. As rather different size classes may be involved, corresponding studies may yield information about the geomorphology as well as the palaeoclimatic or, more general, palaeoenvironmental conditions, and record both, long-term changes and short-term extreme "events".

Minerogenic dust plays an important role in climate forcing as it affects the radiation balance in the atmosphere [Harrison et al. 2001, Kohfeld and Harrison 2001, Prospero et al. 2002]. During the past, dust fluxes are known to have changed significantly. From polar ice sheets, [Thompson et al. 1989] reported that dust accumulation has increased by more than one order of magnitude from interglacial to glacial times. The high accumulation of minerogenic aerosols during glacial time is (besides changes in atmospheric dynamics) mainly controlled by increased aridity in mid, non-glaciered latitudes. As result of reduced moisture availability, forest and prairie vegetations covering and stabilising soils retreated drastically in the Northern hemisphere. Hence, dust is an important proxy for tracing atmospheric dynamics and, in combination with pollen data, to infer on aridity changes in the past. Furthermore, grain size distributions may be used to highlight changes in the the position of relevant wind trajectories and to discriminate between local and far distant atmospheric transport.

This chapter presents the results of an analysis of data from a sediment core obtained in Lake Baikal, Eastern Siberia. These data have been obtained by a laser-assisted grain size analysis using the detrital fraction $> 2\mu$m. Furthermore, the different fractions (like opal, clays, silts and sand) from the sediment of the CON01-603-2 core (which has been retrieved from the Continent Ridge in the Northern Basin of Lake Baikal, see Fig. 4.1) are quantified to evaluate the processes which control the formation of the different detrital particles distribution. Studies of the detrital input, contrarily to biogenic proxy studies (whose availability is highly dependent on the temperatures) are able provide continuous paleoclimatic records including cold periods. Temperature and moisture indices reconstructed from pollen assemblages in the Late Glacial and the Holocene [Demske et al. 2005, Granoszewski et al. 2005, Tarasov et al. 2005] are therefore

compared with grain size data to highlight the processes driving the detrital input into Lake Baikal during the last 150 kyr.

4.2 Measurement of Grain-Size Distributions

Grain-size distributions can be obtained by a variety of different procedures, ranging from the traditional sieving, microscopic measurements in thin-sections (in modern days usually combined with some sophisticated pattern recognition algorithm for automatic processing of the microscopic images) to sedimentation methods. A rather new approach realised in many present-day device is based on the single-particle scattering of a laser beam in a suspension including the probe material. The raw data are given in terms of relative frequences of occurrence in different pre-defined size classes. Depending on the goal of the project and the measurement strategy, the corresponding grouped (compositional) data are given as either weight (mass) or particle number percentages. For converting one type of distribution to the other, different approaches have been discussed in the literature [Greenman 1951, Friedman 1958], however, all these approaches depend on certain assumptions (e.g., about the shape and density of the particles) which are often violated in reality.

[McBride 1971] lists five typical goals of grain-size analysis:

1. To describe samples in terms of statistical measures.
2. To correlate samples from similar depositional environments or stratigraphic units.
3. To determine the agent (wind, river, turbidity current, etc.) of transportation and deposition.
4. To determine the process (suspension, traction, saltation, etc.) of final deposition.
5. To determine the environment of deposition (channel, plain flow, beach, dune, neritic marine, etc.).

Performing grain-size analysis for samples of a sedimentary sequence belonging to different time intervals of deposition thus allows to derive information about the variations of the (palaeo)environmental conditions.

Traditionally, grain-size distributions are classified using a logarithmic size scale. Today, the so-called ϕ scale introduced by [Krumbein 1934, Krumbein 1936, Krumbein 1938] is commonly used which is defined as $d[\phi] = -\log_2 d[mm]$. Size classes are thus often defined in equi-distant units of ϕ, $\phi/2$, or $\phi/4$, depending on the respective measurement device. The grain-size spectrum is roughly devided into clay ($< 2\mu$m), silt (2-64μm), and sand ($> 64\mu$m), however, when relating the terms clay and sand to certain mineralogical fractions, slightly different classifications are possible.

4.3 Statistical Approaches to Grain-Size Analysis

The probably most used approach of extracting climatically relevant information from grain-size distributions is considering simple statistical parameters of the entire distribution function, for example, the mean (i.e., the average particle size), mode (the location of the maximum, i.e., the most probable particle size), or median (the 50% quantile, i.e., the size of the mean ranked observation) of the entire distribution. This approach may be useful if the observed distributions are

rather "smooth", in particular, unimodal. In addition, higher order moments (ore characteristics based on these) like the variance or (in the typical case of asymmetric distributions) skewness and kurtosis may accompany the corresponding analysis. The corresponding parameters have the advantage that they are automatically computed by many measurement devices or suitable software packages [Blott and Pye 2001, Poppe et al. 2004] and are thus easy to use for further analysis.

In many cases, the particle-size distribution is however not unimodal. In this case, the consideration of simple global statistical parameters is not useful and may yield variability patterns which cannot be interpreted climatologically. To overcome the corresponding difficulties, different heuristically defined parameters have been proposed, like the mean/mode and median/mode ratios [Demory 2004] or the so-called U-ratio, which is defined as the ratio between the relative abundances in the size classes 16 to 44 μm and 5.5 to 16 μm [Vandenberghe et al. 1993, Vandenberghe et al. 1997, Nugteren et al. 2004]. The U-ratio has mainly been used to characterise grain-size distributions from Chinese Loess deposits and has the advantage that it disregards secondary formed minerals in the clay fraction ($\ll 5.5\mu$m) and sand-sized particles ($\gg 44\mu$m) probably deposited by saltation [Vandenberghe et al. 1997]. However, this advantage comes on the cost of an overemphasis of the coarse fraction in glacials, since a small increase in coarse material results in a much higher U-ratio.

A possible alternative for the statistical analysis of grain-size distributions is the appropriate decomposition of the entire (multi-layer) data set. The simplest approach for such a decomposition is principal component analysis (which has been referred to as Karhunen-Loéve decomposition (KLD) in the previous chapter) [Davis 1970, Chambers and Upchurch 1979], eventually combined with an appropriate (log-ratio) transformation of the data (see, e.g., [Aitchison 1982, Aitchison 1983, Aitchison 1986, Aitchison 2002]). Basing on similar approaches, a whole theory of nonparametric geostatistical "unmixing" of such data sets has been developed, with the so-called *endmember modelling* as the most prominent and probably most used technique [Renner 1991, Weltje 1997, Prins and Weltje 1999].

Endmember modelling and related approaches consider the entire time series of grain-size distributions to infer typical patterns superposed in all time slices in an appropriate way. However, this approach has also some disadvantages: Adding a new time slice with measurements will typically change the shape of the inferred endmembers and, consequently, also their statistical weights. Hence, a sufficient amount of samples is required to give a reasonable low statistical uncertainty of the derived patterns (for the determination of uncertainty, the application of resampling techniques considering "slightly" perturbed data is outlined). In addition, the components resulting from endmember modelling may have a complicated (sometimes even multimodal) shape which can hardly be assigned to a particular physical generation mechanism.

Complimentary to the global characterisation with an endmember modelling approach, a separate description of the distributions obtained in any time slice may give important information about the composition of the sample of sediment with different origin and/or transport history. For this purpose, a successive parametric modelling may be applied for all time slices. For this prupose, a suitable and meaningful statistical model has to be described first. Starting already in the 1930's, there has been a long debate on suitable model functions suggesting either log-normal [Kolmogorov 1941, Inman 1952, Folk and Ward 1957, Sheridan et al. 1987, Gorokhovski and Saveliev 2003, Gorokhovski 2003] or Weibull (Rosin-Rammler) [Rosin et al. 1933, Rosin and Rammler 1933, Rosin and Rammler 1934, Krumbein and Tisdel 1940, Kittleman Jr. 1964, Braun 1975, Zobeck et al. 1999] distributions as the most promising candidates depending on the mechanisms responsible for generation, transport and deposition of the particles [Ibbeken 1983, Schleyer 1987, Hartmann 1988,

Schleyer 1988, Kondolf and Adhikari 2000, Lu et al. 2002]. A possible unification of both approaches is given in terms of the sequential fragmentation an transport (SFT) theory [Wohletz et al. 1989, Lirer et al. 1996].

The above mentioned approaches may typically be used to describe homogeneous sediments, i.e., grain-size distributions originating from only one well-distinguished source. For the more general case of distributions with a complicated shape, the use of log-hyperbolic or log-skew Laplace distributions [Bagnold and Barndorff-Nielsen 1980, Fieller et al. 1984, Wyrwoll and Smyth 1985, Christiansen and Hartmann 1988, Wyrwoll and Smyth 1988, Hartmann and Christiansen 1992, Hartmann and Bowman 1993, Sutherland and Lee 1994, Hill and McLaren 2001, Knight et al. 2002, Hartmann and Flemming 2002, Hill and McLaren 2003, Scalon et al. 2003] or other model functions [Passe 1997] has been discussed, however, the corresponding results are usually rather poor. For multi-modal grain-size distributions, it is in contrast an appealing alternative to interprete the observations as a superposition [Doeglas 1946] of "standard" component functions (log-normal, Weibull)[1] [Sun et al. 2002] describing contributioons with different origin (i.e., mineralogical composition), and erosional history (related to the transport by wind, water, etc.). Hence, finite mixture models (see Chapt. 2) are promising candidates for a parametric description. The appropriate methodology for the analysis of such models has been discussed by several authors (e.g., [Clark 1976]). Note, however, that the significance of the corresponding genetic interpretation may be rather poor [Weltje and Prins accepted].

4.4 Mechanisms of Detrital Input into Lake Baikal

In the region of Lake Baikal, the detrital input originates today from aerosol transport or from river discharge. The balances between aeolian and fluviatile detrital inputs is seasonally controlled. During late spring and summer, dust may settle from large-scale dust plumes, which generate sporadically in the vast plains of the Angara River north of Irkutsk or come from distant desert regions. Well documented is e.g., a local dust storm occurring during summer 1890, when a 5 mm thick lid of dust settled around Lake Baikal during a single event. Recently such events can be pined-down with the Total Ozone Mapping Spectrometers (TOMS) (NASA) from the source area to the location of final deposition. For 19 through 23 May 1989 a regional storm event loading in the Takla Makan Desert has been documented. Such events may carry as much as 109 tons of dust. These exceptional strong convective turbulences are due to extreme heating of the surface during late spring and summer. During dry winters coarser silt and fine sand particles are carried on the ice cover from the shore by saltation to the centre of the lake [Karabanov et al. 1998].

Besides the aeolian dust particles, detrital input is also controlled by the regional hydrology. Today the summer precipitation exceeds winter precipitation by a factor of 10. The lake is feed by many rivers which spread mostly eastward and south eastward of the lake. With the 3 main tributaries - Selenga, Barguzin and Upper Angara Rivers - they cover covering a huge large catchment area (as large as 560,000 km^3). At present the seasonally varying moisture distribution, which regulates the precipitation in the catchment area of Lake Baikal, is mostly driven by the Westerly winds which have its biggest imprint on the western shore of the lake. South-eastern winds which are deviated at the eastern face of the Sayan mountains contribute to the water balance of Selenga River, the biggest tributary of Lake Baikal. These moisture

[1] In addition, [Jones and McLachlan 1989, Fieller et al. 1990] suggest mixtures of log-hyperbolic and log-skew Laplace components.

loaded winds occasionally cause strong summer flood events, which significantly increase the suspension discharge at the Selenga Delta [Heim et al. 2005].

In-situ reworking is a third process, which must be addressed though its quantification is difficult. As earthquakes are quite frequent and slopes along the rift flanks are steep, turbidites are common in the centre of basins. Based on high resolution seismic studies, redeposition of reworked detrital components due to strong currents within the water body and near the lake bottom according to [Ceramicola et al. 2002, Charlet et al. 2005] can mostly be ruled at water depths beyong 600 metres.

To date, the reconstruction of climate variability in Lake Baikal sediments is mostly based on pollen and diatoms assemblage studies [Mackay et al. 1997, Khursevich et al. 2001, Demske et al. 2002, Demske et al. 2005, Granoszewski et al. 2005, Rioual and Mackay 2005]. Only few paleoclimatic studies from Lake Baikal deal with detrital proxies, like clay mineralogy [Yuretich et al. 1999] and grain size analyses, for the first time on turbiditic sediments using the image analysis technique [Francus 1998, Francus and Karabanov 2000]. A subsequent study by [Ochiai and Kashiwaya 2003] showed with the laser counter technique, that abundance of fine particles increased indeed during cold periods. [Fagel et al. 2003] notably confirmed that cristallinity of clay minerals and their weathering are related to temperature and moisture conditions. From rock magnetic studies [Peck et al. 1994, Demory et al. 2005b], it turned out that a parameter estimating the abundance of hematite is the best marker of detrital input.

4.5 Description of the Data

Within the framework of the CONTINENT campaign, several lacustrine sediment cores have been obtained at three different parts of Lake Baikal, Eastern Siberia. For palaeoclimatic studies, this site is of particular importance because of the relatively continuous sedimentation for up to several million years [Kashiwaya et al. 1998, Kashiwaya et al. 2001]. The corresponding sediments thus give important information for a more detailed understanding of long-term climate variability within Central Eurasia as there are rather few comparable sites within this relatively large area. Lake Baikal is of particular interest as it is influenced by westerlies, the Siberian High, and particularly by the East Asian monsoon circulation differently for different epochs in climate history. Thus, detailed multi-proxy studies at this location allow to study changes of the dynamics of any of these atmospheric patterns.

The grain-size data set studied in the following has been obtained from a sequence build of the pilot and piston cores CON01-602-2a and CON01-602-2 retrieved in 2001 from the Continent Ridge (Fig. 4.1). The 11 metres long composite section consists of hemipelagic sediments deposited on the geomorphological high located at the prolongation of the Academician Ridge. The sedimentary record of the core is nearly continuous and ranges from the Holocene to through the interstadial equivalent to the marine isotope stage[2] MIS 7 [Demory et al. 2005a]. The cold periods (glacials and stadials) are characterised by clay-rich sedimentation whereas the warm periods (interglacials and interstadials) are characterised by diatomaceous-rich sediments [Charlet et al. 2005]. The uppermost section was dated using the AMS ^{14}C method

[2]The marine isotope stages (MIS) have been introduced to classify climatic periods with colder and warmer conditions, which can be classified according to the isotopic composition of different chemical elements firstly described for marine sediments. The current interglacial (Holocene) is referred to as MIS 1, the last interglacial (mainly referred to as the Eemian) corresponds to MIS 5e. In between, the isotope stages 2 (last glacial maximum), 3, and especially 4, 5a, 5b, 5c, and 5d represent alternating time intervals with colder and warmer average temperatures, the so-called stadials (cold) and interstadials (warm), resp. In general, higher numbers of an isotopic stage indicate older time periods.

for the last 15 kyr [Piotrowska et al. 2004]. Beyond 15 kyr, palaeomagnetic data have been used to construct the age model [Demory et al. 2005a]. The reference curve established by [Channell 1999] at ODP site 984 was used to date the so-called Kazantsevo (the time equivalent of the European Eemian), i.e., the duration of the last interglacial in Eastern Siberia. For dating the time window from 15 to 110 kyr, the palaeomagnetic intensity record of the MD 95-2024 core from the Labrador Sea [Stoner et al. 2000] has been used. The corresponding record covers only 110 kyr but seems to be better dated than the record from [Channell 1999] since the former is tuned to the high-resolution chronology of the GISP-2 ice core [Grootes et al. 1993] while the latter is tuned to the low-resolution chronology of the orbitally tuned SPECMAP curve [Martinson et al. 1987]. The sequence has yet been subjected to studies of different other proxies to infer geochemical, palaeolimnological, and sedimentological information (for a recent summary of these studies, see [Oberhänsli and Mackay 2005]).

From 448 samples (continuously sampled at 2 cm intervals) covering the entire sedimentary sequence, the organic carbon has firstly been removed by soaking repeatedly 1 g of freeze dried sediment in a H_2O_2 solution (5%) until all the organic carbon was dissolved. Then, the clay fraction was separated by centrifugation (1 min at 1000 rev. min^{-1}). In a next step, the biogenic silica was dissolved at 90^oC, while shaking the sample for 5 hours in a 2M solution of Na_2CO_3. As expected from other studies [Olivarez Lyle and Lyle 2002], the opal dissolution was not always completed after this procedure. For 56 samples, which still contained the most resistant diatoms and spicules of sponge, the remaining opal was separated using a sodium poly-tungstate solution with a density of 2.32 g/cm^3. The opal content was measured with ICP-OES. For removing carbonates and authigenic hydroxides, the remaining sediment was soaked in a HCl solution (1.1N) and treated with ultrasonic for one hour. Despite this procedure, few samples were still containing aggregates which had partly formed during freeze drying [Rajaram and Erbach 1999]. Some of the aggregates are faecal pellets, which can be abundant in Lake Baikal sediments [Tani et al. 2002]. Subsequently, the clay fraction as well as the detrital fraction > 2 μm were dried and weighted.

The grain size distribution of the detrital fraction $> 2\mu$m was measured using the Malvern Mastersizer 2000 equipment at the University of Lille (see Fig. 4.2). Separation of the clay and silt fractions was difficult to complete. After multiple separation steps an irreducible fraction of clays still adhered to particles of the size fraction $> 2\mu$m. However, the remaining amount of clay is small (less than 5%) and remains relatively constant. Using tests of reproducibility after laboratory treatment [Demory 2004], estimated error bars of an order of $\pm 1\mu$m for the mode of the particle size distribution have been found to be realistic. The error bar for the size-frequency data reported by the manufacturer (Malvern) of the grain size measurement device is very low ($< 1\%$) but increases with decreasing size of measured particles and reaches 2% between 900 μm and 5μm and 6% for the finer fractions.

4.6 Statistical Analysis of the Lake Baikal Record

The grain-size record shown in Fig. 4.2 already underlined variations of the particle size distributions which are visible in terms of shifts of the maxima or tails towards large particle sizes. Fig. 4.3 presents grain-size distributions of the detrital fraction $> 2\mu$m for selected periods including clay-rich and diatomaceous muds. Fine silts (2-32 μm) generally dominate the grain size distribution of the detrital fraction $> 2\mu$m. The fraction is fine (mode $\sim 9\mu$m) during glacial and interglacial optimums while the silt fraction coarsens (mode $\sim 10.5\mu$m) during climatic transitions. In addition, the figure displays the grain-size distribution of dust trapped in fresh

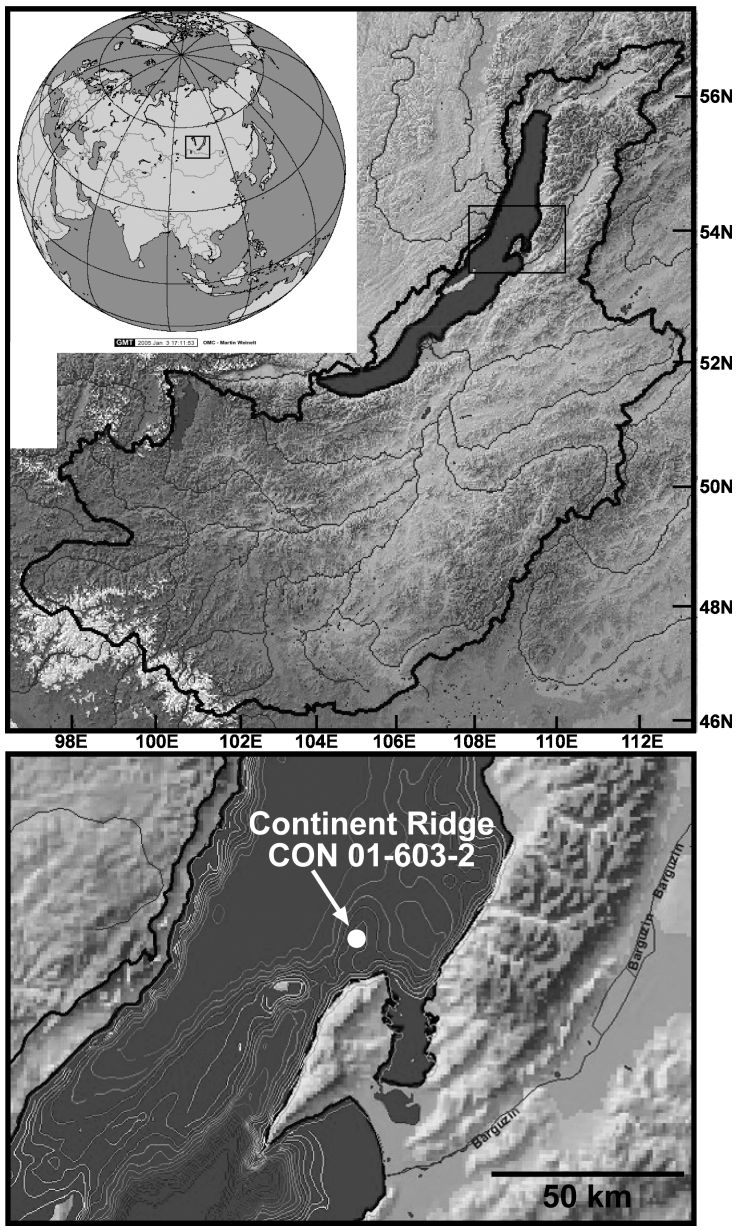


Figure 4.1: Maps showing the location of Lake Baikal and the coring station for core CON 01-603-2. The middle scale map shows the relief around Lake Baikal (Landsat TM-Mosaic UTM48, source: Baikal Online-GIS, http://dc108.gfz-potsdam.de/website), and the black line shows limits of the catchment area of the lake. The focused map shows in addition some bathymetric lines.

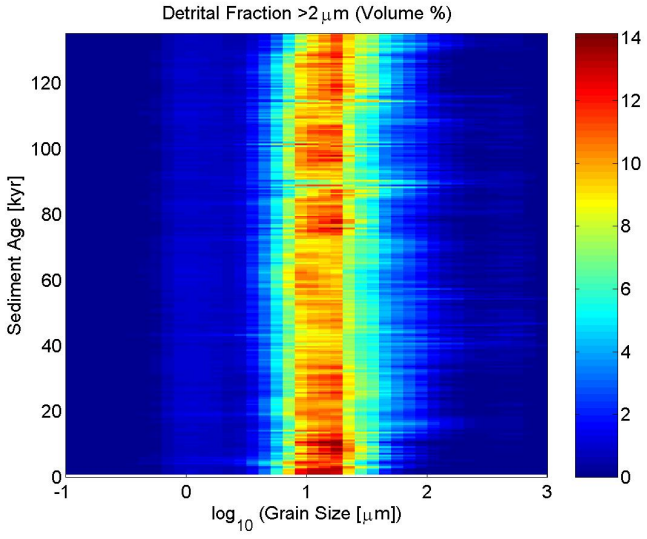


Figure 4.2: Color-coded representation of the relative abundance (in %) of particles in the different size intervals as a function of the estimated sediment age.

snow accumulating on Lake Baikal ice cover. This dust, which is of aeolian origin, is dominated by fine silt particles like the detrital fraction $> 2\mu$m extracted from Lake Baikal sediments. The clay content is low although no treatment (i.e. no grain size separation) was applied to this recent aeolian sediment. In the core, coarse grains $> 100\mu$m are restricted to cold and transitional climate periods.

Clay and opal contents in the sediment mimic the well-known clay-rich/diatom-rich alternation, which reflects the cold/warm cycles [Demory 2004]. Cold periods are characterised by high amounts of clay (up to 80 weight percentage) and by an absence of opal (i.e., organic material) whereas warm periods are characterised by low amounts of clay (as low as 40 wt %) and high amounts of opal (up to 25 wt %) (except in the marine isotope stage 5b equivalent which seems to be warmer in Eastern Siberia than one would expect from other records). The cold/warm pattern is generally well reflected in the quantity of detrital particles $> 2\mu$m. The corresponding percentage is high ($\sim$ 30%) in sediments representing warm periods and low ($\sim$ 15%) during colder periods, except (i) during the Late Holocene and in the Kazantsevo when silt abundance is low and (ii) during the late MIS 6 when silt contents are high. As to the mode of this fraction, Fig. 4.4 shows more detailed variations. The silt fraction has a low mode average of approx. 9-10 μm at the end of MIS 1, 2, 4, 5d, and at the beginning of MIS 5e while the silt fraction coarsens to a maximum mode average value of 11 up to 14 μm during the end of MIS 6, 5e, 5c-5a, 3, at the beginning of MIS 3 and during most of the Holocene (Fig. 4.4).

4.6.1 Global Statistical Parameters

Beside the mode giving only the position of the maximum, other parameters describing the "location" of the grain-size distribution on the size axis have been computed (see Fig. 4.4). As with respect to a linear size scale, the particles sizes are strongly assymetrically distributed, the consideration of additional parameters may allow to infer additional information about the shape of the distribution function. Whereas median and mode (and their corresponding ratio) have been directly computed by the measurement device, the mean value has been estimated

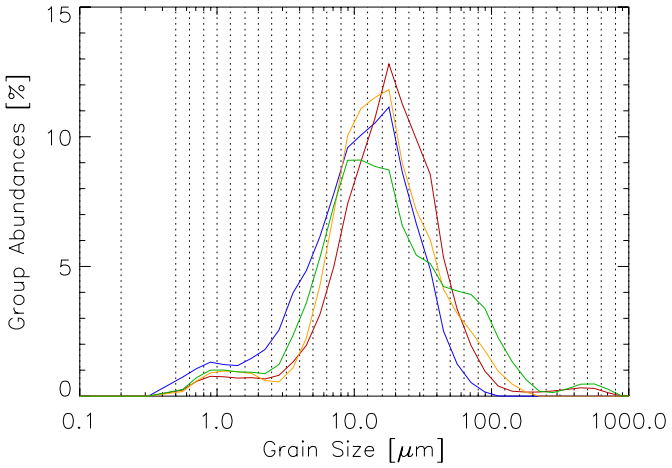


Figure 4.3: Grain size distribution of the detrital fraction $> 2\mu$m for 4 representative samples: present-day dust sample from the fresh snow cover on the lake (blue), Early Holocene (approx. 10.2 ka BP, red), Last Glacial Maximum (approx. 30.1 ka BP, orange), and Kazantsevo / Eemian interglacial (approx. 125.1 ka BP, green). The vertical lines indicate the respective grain size intervals.

using the observed group frequencies n_m and the (linear) size interval midpoints $\bar{x}_m$ according to $\hat{\mu} = \sum_{m=1}^{M} n_m x_m / \sum_{m=1}^{M} n_m$ (with $\sum_{m=1}^{M} n_m = 100$ as the relative frequencies of occurrence have been given in percentages).

The temporal variability of median and mode does not directly correspond to known temperature or precipitation patterns as have been inferred from other proxies. In contrast, the mean/mode ratio shows rather pronounced minima during the globally warmer periods equivalent to the MIS 1 (Holocene, about 12 kyr BP - present), 5a (about 75-85 kyr BP), 5c (about 95-105 kyr BP), and 5e (Eemian/Kazantsevo, about 120-130 kyr BP) [Demory 2004]. Similar minima are also found when considering the mean grain size, however, the mean shows an additional minimum during MIS 2 which cannot be linked to the global temperature signal. Hence, when relating the statistical properties of the particle size distributions from the detrital fraction in Lake Baikal to known climatic variability patterns, the median/mode ratio (yielding information about the asymmetry of the entire distribution function) gives by far the most convincing signal. To conclude about the reason for this behaviour, more detailed statistical analyses are required.

4.6.2 Statistical Modelling

As it was discussed in Sect. 4.3, the statistical modelling of the grain-size distribution functions in any time slice is an appealing alternative to the simple consideration of global parameters. In particular, as the asymmetry of the distribution flanks (which can be observed in Fig. 4.3) suggests the application of finite mixture models with strongly overlapping components, however, due to the overlap, large undertainties (and eventually some bias) are to be expected.

To determine the applicability of different model functions, all observed distributions have been firstly analysed using a software package developed by Ken Wohletz in the framework of the sequential fragmentation and transport (SFT) theory. This software is freely available and

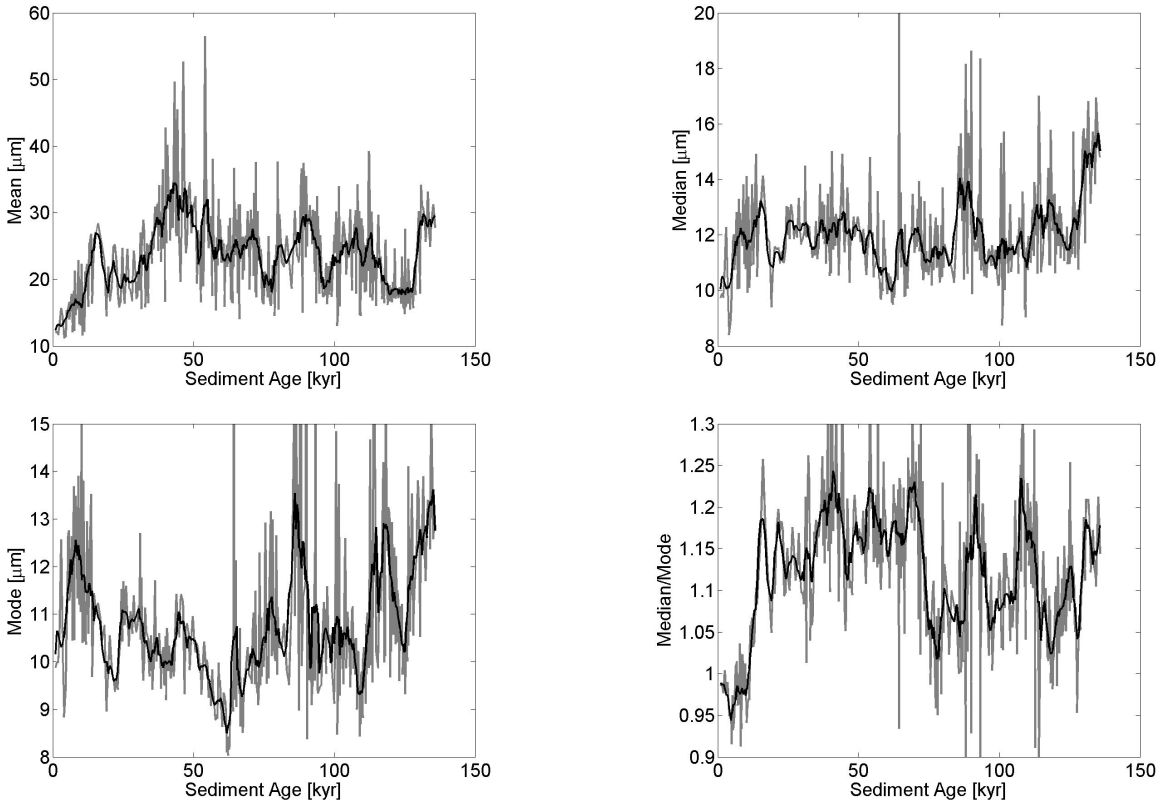


Figure 4.4: Global statistical parameters describing the "typical" size of particles in the sediment: Mean (upper left panel), median (upper right panel), mode (lower left panel), and median/mode ratio (lower right panel) of the grain-size distributions as a function of the estimated sediment age. Gray lines give indicate the separate values computed for all 448 time slices, whereas the black lines correspond to the values computed with a 2,000 year (non-weighted) moving average filter.

allows to fit finite mixture models with up to four log-normal or SFT-type (Weibull) components to observed grain-size data. In general, it has been found that models with four lognormal components minimise the residuals in most cases, hence, the corresponding model was successively applied to all samples. The temporal variability of the corresponding model parameters gives, however, mainly results which cannot be interpreted palaeoclimatically.

The rather irregular behaviour of the estimated model parameters is to some extent related to the features of the software package used[34]: Firstly, the starting values have to be prescribed manuelly. Secondly, the algorithm does not work using a particular statistical estimator like the EM algorithm described in Chapt. 2, but performs optimisation with respect to an interpolating spline function. Thirdly, the data (given with respect to size classes predefined by the measurement device) have to be interpolated in advance to half-ϕ or full-ϕ units (see Sect. 4.2), which leads to additional large uncertainties[5]. The only interpretable and robust pattern is given by

[3]In addition, the potential variability of the "optimum" model contributes here as well.

[4]For comparison, the EM algorithm was applied as well to the complete data set, yielding similar results.

[5]The corresponding problem has been tested by comparing data locally interpolated by linear and quadratic polynoms, resp., yielding significantly different results of most parameters.

the statistical weight of the coarse-grained component, which pretty well follows the variability of known northern hemispheric temperature signals, however, a corresponding pattern may be much simpler inferred by considering the abundance of the sand fraction $> 64\mu$m shown in Fig. 4.5.

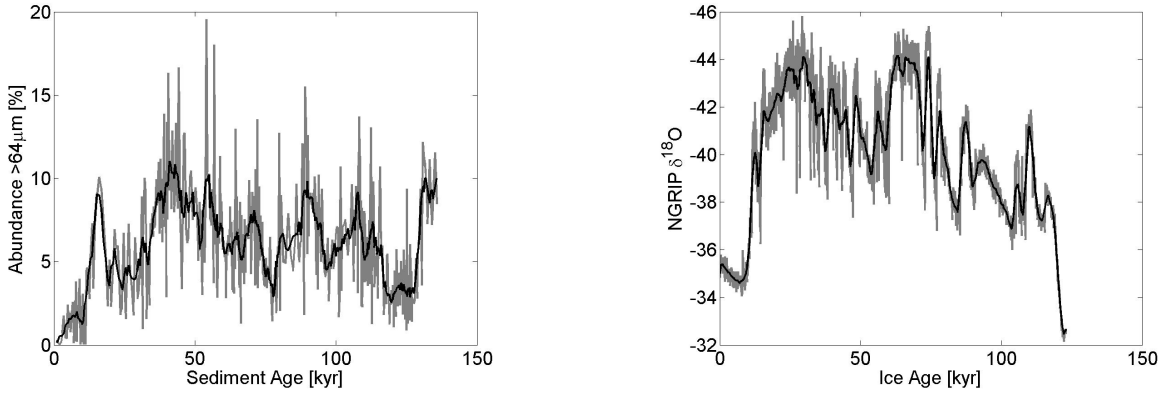


Figure 4.5: Relative abundance of sand ($> 64\mu$m) in the detrital fraction $< 2\mu$m of the Lake Baikal sediment, and the δ^{18}O palaeotemperature proxy record from the Greenland NRIP ice core [NGRIP members 2004]. Gray lines give indicate the separate values computed for all time slices, whereas the black lines correspond to the values computed with a 2,000 year (non-weighted) moving average filter.

Considering the general applicability of finite mixture models for describing sequences of grain-size distributions, one has to firstly recall the problems with parameter estimation in the presence of an unknown model structure, relevant residuals, or a significant component overlap already discussed in Sect. 2.4.5. In addition, the presented approach implicitly assumes not only the correctness of the predefined model structure, but also its constancy over the entire sequence, which is likely a wrong assumption if sediments from periods with very different environmental conditions are considered. Hence, one has to conclude that for the Lake Baikal grain-size data, the application of the finite mixture model approach does not lead to appropriate results.

4.6.3 Principal Component Analysis

As it has been demonstrated above that statistical modelling with a fixed, formal model structure gives only very poor palaeoclimatically interpretable information, there is the question of more robust statistical methods for the analysis of grain-size distributions. A particular approach would be the non-parametric endmember modelling [Weltje 1997], which has widely been applied in a variety of geological studies. However, the components inferred by this approach have usually a rather complicated shape which can often not be assigned to any particular sedimentological component. In addition, the statistical significance of the results should be discussed, in particular, in the case of small sample sizes. For this purpose, a resampling approach similar to that used in the case of the LVD dimension in Chapt. 3 might be useful which could work as follows:

1. Take the complete grain-size data set as it is.

2. Choose a particular percentage of the measured time slices by chance.

3. Replace the data within these time slices by random data.

4. Apply the statistical analysis and/or modelling approach, e.g., endmember modelling.

5. Repeat this procedure a sufficient number of times.

6. Calculate a suitable statistics over the results obtained from all randomised realisations (for example, mean and dispersion give information about bias and uncertainty of the original approach).

In the following, an approach is used which is even simpler than endmember modelling. The principal component analysis was already introduced in Chapt. 3 of this thesis as *Karhunen-Loève decomposition* (KLD). In the following, the information about the dynamics gained from this approach is considered: as KLD is based on a transformation of the covariance matrix to diagonal form, the entire data set can be recovered by an appropriate superposition of the corresponding eigenvectors. As the latter ones are orthonormal by definition, the corresponding (time-dependent) expansion coefficients may be easily computed as the scalar product of the data in a particular time slice with the respective eigenvector.

In Sect. 3.5.4, it has been argued that an appropriate principal component analysis requires a certain statistical treatment of the data in terms of a (log-)ratio transformation [Aitchison 1983]. Going yet another step back, this recommendation is not considered in the following, hence, the statistical decomposition is applied with respect to the original (non-transformed) compositional data set.

Applied this way, principal component analysis gives as a first-hand information components which are described by (linearly in a certain way optimised) weights associated to each size class. Fig. 4.6 shows the corresponding shapes of the first major principal components. One recognised that the first component shows one zero crossing, the second one two, etc. This behaviour has also been found for grain-size distributions from other locations with rather different shapes, which indicated that it is intrinsically related to the particular decomposition approach.

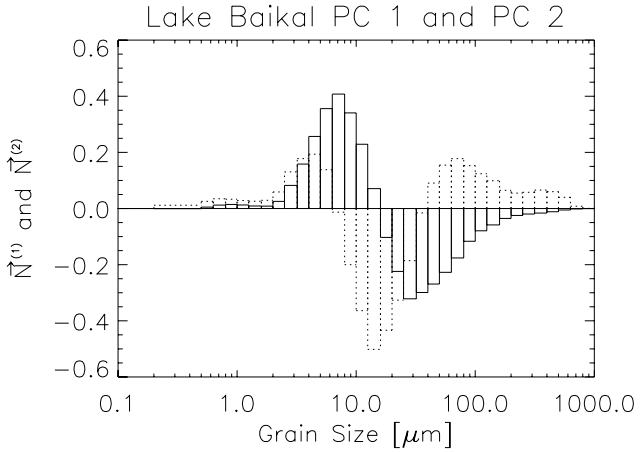


Figure 4.6: Shape of the first two major principal components (weights for every size class, indicated by the height of the displayed bars). Solid lines correspond to the PC 1, dotted ones to the PC 2.

While the principal components (i.e., the eigenvectors of the covariance matrix) may be interpreted in terms of the shape of the grain size distributions, the evolution coefficients represent the whole information about the temporal variability of this shape, including the complete information on the variability of environmental conditions from the original data set. PC 1, explaining 50.8% of the total variability, has a positive (negative) amplitude if more (less) sediment with smaller than with larger grains (compared with an average distribution from the data set) occurs. The boundary between smaller and larger grains is approximately fixed by the position of the zero crossing at about 10.5μm. While thus PC 1 mainly quantifies the asymmetry between both flanks of the total distribution (i.e., has its highest absolute values at about 6 and 25μm, resp.), PC 2 (explaining another 38.5% of variance) represents the asymmetry between the bulk region (determined by its two zero crossings at about 6 and 30μm) and both flanks such that a wider distribution with more fine and/or large-grained sediment (and therefore more pronounced tails) results in more positive coefficients while negative values correspond to a very narrow distribution with only few sediment with grains outside this size interval.

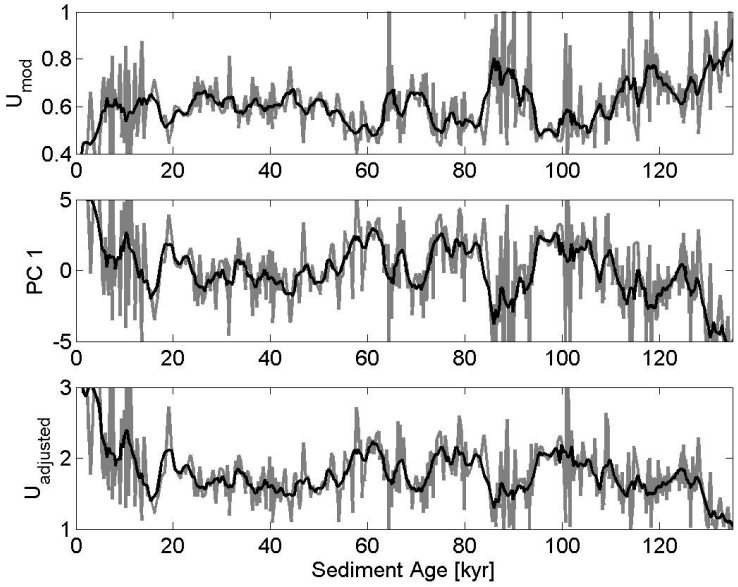


Figure 4.7: From top to bottom as a function of the sediment age: U-ratio (modified according to the size classes prescribed by the measurement device), amplitude of the first principal component, and modified U-ratio defined as the ratio of sediment abundance above and below 16μm (the position of the zero-crossing of the PC 1).

As a next step, the temporal evolution of the principal component amplitudes is studied. The corresponding variability describes the most prominent temporal changes in the observed grain-size distribution and may therefore be used to assess information on the variability of environmental conditions. Considering the first component, Fig. 4.7 shows that the corresponding amplitude is closely related to the U-ratio of [Vandenberghe et al. 1993][6]. Equivalently, one may use the ratio of sediment above and below the zero crossing of the first principal component at

[6]As the size classes have been predefined by the measurement device, the definition of the U-ratio has been slightly modified according to these classes as the ratio of sedimented material with grain sizes between 16 and 50μm and between 5 and 16μm, resp.

16μm, which is referred to as the *adjusted U-ratio*. Obviously, all three parameters show almost the same variability pattern.

The temporal evolution of the first two principal components shows variations on time scales related to the well-known alternation of marine isotope stages (see above). Using wavelet analysis as described in Sect. 1.5, clear evidence is found for significant variations on frequencies corresponding to the precession and obliquity oscillations of the Earth's orbit. Fig. 4.8 shows that the corresponding oscillatory components are relevant over almost the entire record, i.e., the complete last glacial/interglacial cycle. Although the temporal resolution is extraordinary well for a grain-size record, it is still not sufficient to resolve millenial-scale oscillations. Hence, it is not possible to derive information about climate oscillations on these smaller time scales which would be an indicator for large-scale northern hemispheric teleconnections, i.e., interrelationships of the atmospheric circulation patterns over large spatial distances.

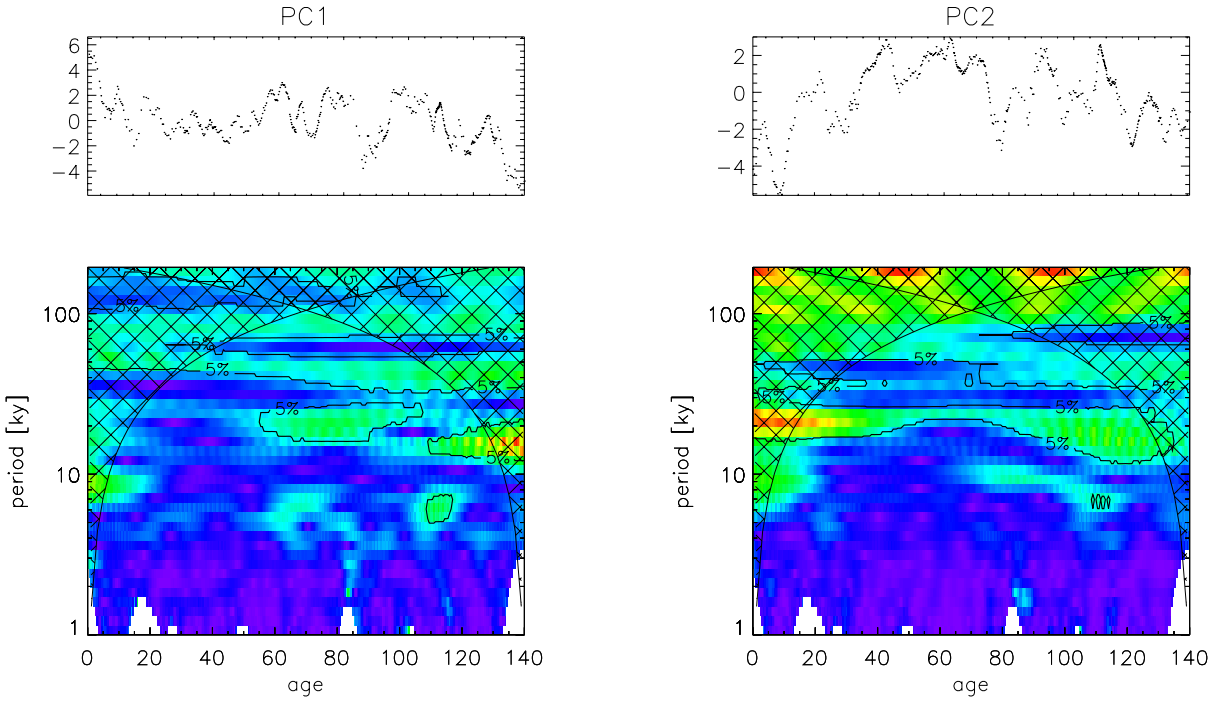


Figure 4.8: Temporal evolution of the amplitudes (averaged over 2,000 years) of the first two principal components (upper panels), and the corresponding wavelet amplitude maps including the estimated 95% significance levels from a test against white noise.

4.7 Interpretation

As an alternative to the information provided by the amplitudes of the major principal components, it may be sufficient to consider the abundance of grains within size classes which are defined by the zero crossings of these components. For the Lake Baikal data set, these roots are at 16 μm (PC 1), 6 μm and 40 μm (PC 2). Hence, principal component analysis allows to define four disjoint size classes which are further evaluated. The finest fraction with grain sizes below 6 μm may still be contaminated by some remaining clay and will therefore not be discussed. The medium silt fraction with grain sizes between 16 and 40 μm gives no climatologically inter-

pretable information. Hence, it is appropriate to focus on the fractions of fine silts (6 to 16 μm) and coarse silts and sand (above 40 μm).

The variations of the abundance of coarse material have already been described above. It is likely that coarse material cannot be transported over large distances, hence, there must be a local source and transportation mechanism. As the corresponding abundance (compare Fig. 4.5) is clearly increased during colder periods, it is evident that the efficiency of transport is higher in the case of glacial conditions. A possible explanation is that coarse material is mainly transported from the nearby shore by saltation on the ice. In this case, the efficiency depends on the snow cover on both, the ice cap on the lake and the shore, which suggests that the coarse detrital input might be a proxy for the snow cover. This hypothesis is underlined by a strong coincidence of variations of the corresponding abundance of this fraction and the late winter insolation (which has a direct influence on the duration of the annual frost period) illustrated in Fig. 4.9. In particular, this insolation has been mainly controlled by precession cycles over the last glacial/interglacial cycle.

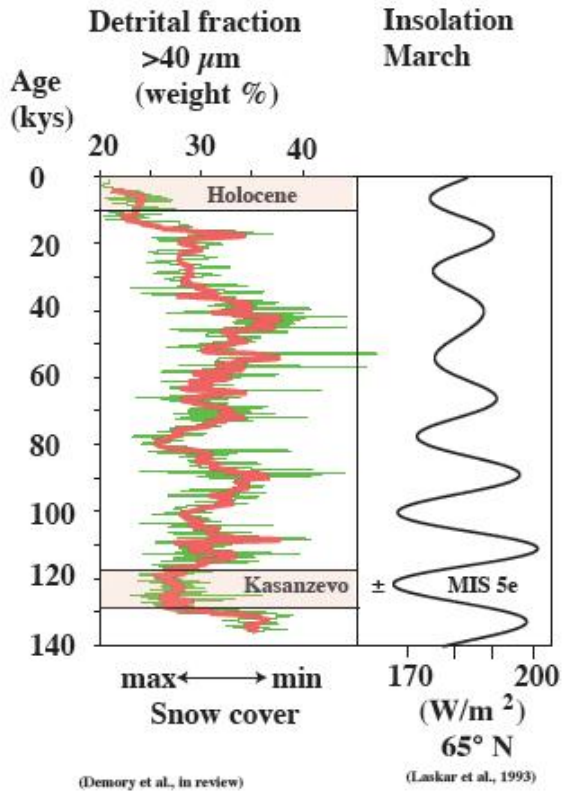


Figure 4.9: Variations in the detrital fraction $> 40\mu$m and the march insolation at 65^o North according to [Laskar et al. 1993].

The explanation for the variability of the fine silt fraction is slightly more subtile. In contrast to the coarse material, fine grains may be efficiently transported by winds, i.e., it is likely that this part of the sediment is of aeolian origin. Fig. 4.10 shows that the corresponding abundance varies similar to the late spring / early summer insolation over Central Eurasia. A more detailed analysis suggests that there is also a dependence of the global ice volume on this insolation: a

minimum summer insolation leads a maximum global ice volume by several 1,000 years over the last glacial cycle. As a maximum of the global ice volume corresponds to a high continentality of the climate, the fine silt fraction can be interpreted as a proxy for continentality.

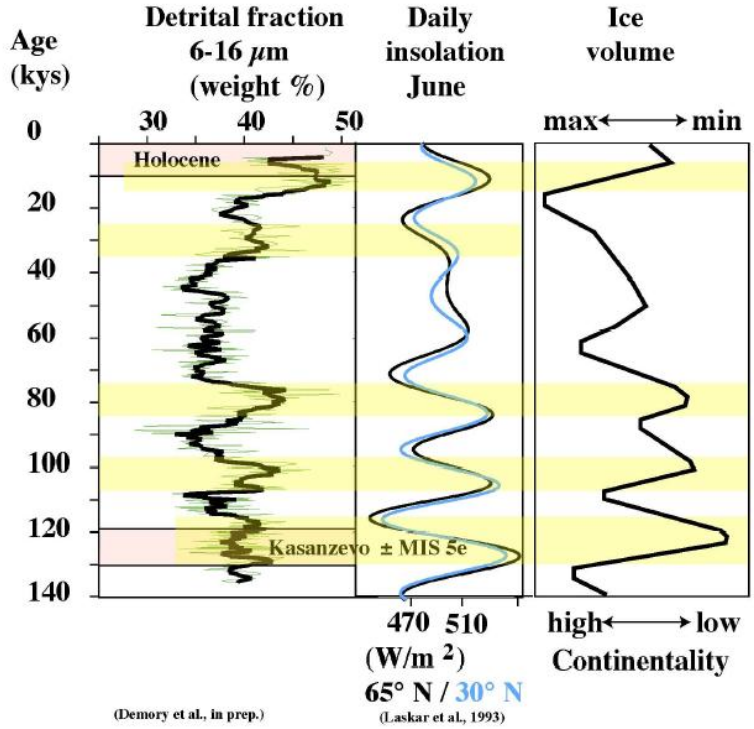


Figure 4.10: Variations in the detrital fraction 6-16μm, the june insolation at 65^o North according to [Laskar et al. 1993], and a surrogate curve of the global ice volume.

A possible source of the fine silt material are dust storms which today occur rather frequently during last spring. The most likely origin of a large amount of aeolian material is thus located in the Taklamakan desert in Northeastern China, which is underlined by satellite measurements of atmospheric aerosols as shown in Fig. 4.11. The typical direction of transport under present-day climatic conditions support the hypothesis that Lake Baikal may be essentially supplied by this source.

The presented attempt for interpreting the climatic conditions around Lake Baikal based on grain-size distributions explains the main features by the varying efficiency of different types of aeolian material transport. However, for the geological site under investigation, it is likely that the sediment support is mainly fluvially controlled, i.e., the largest part of the material is transported by numerous small rivers entering Lake Baikal near the coring site. To defend the explanations proposed above, it would be necessary to collect additional information about the grain-size distributions of the fluvial material. In addition, one has to investigate whether the fluvial input mechanisms are less influenced by changing environmental conditions than aeolian ones. As a working hypothesis, it seems possible that fine aeolian material as well as coarse-grained sediments are less abundant in fluvial sediments such that changes of the corresponding aeolian supply are recorded by smaller "tails" of the grain-size distributions. At least in the case of the coarse grains, this explanation is strongly suggested by the results presented above.

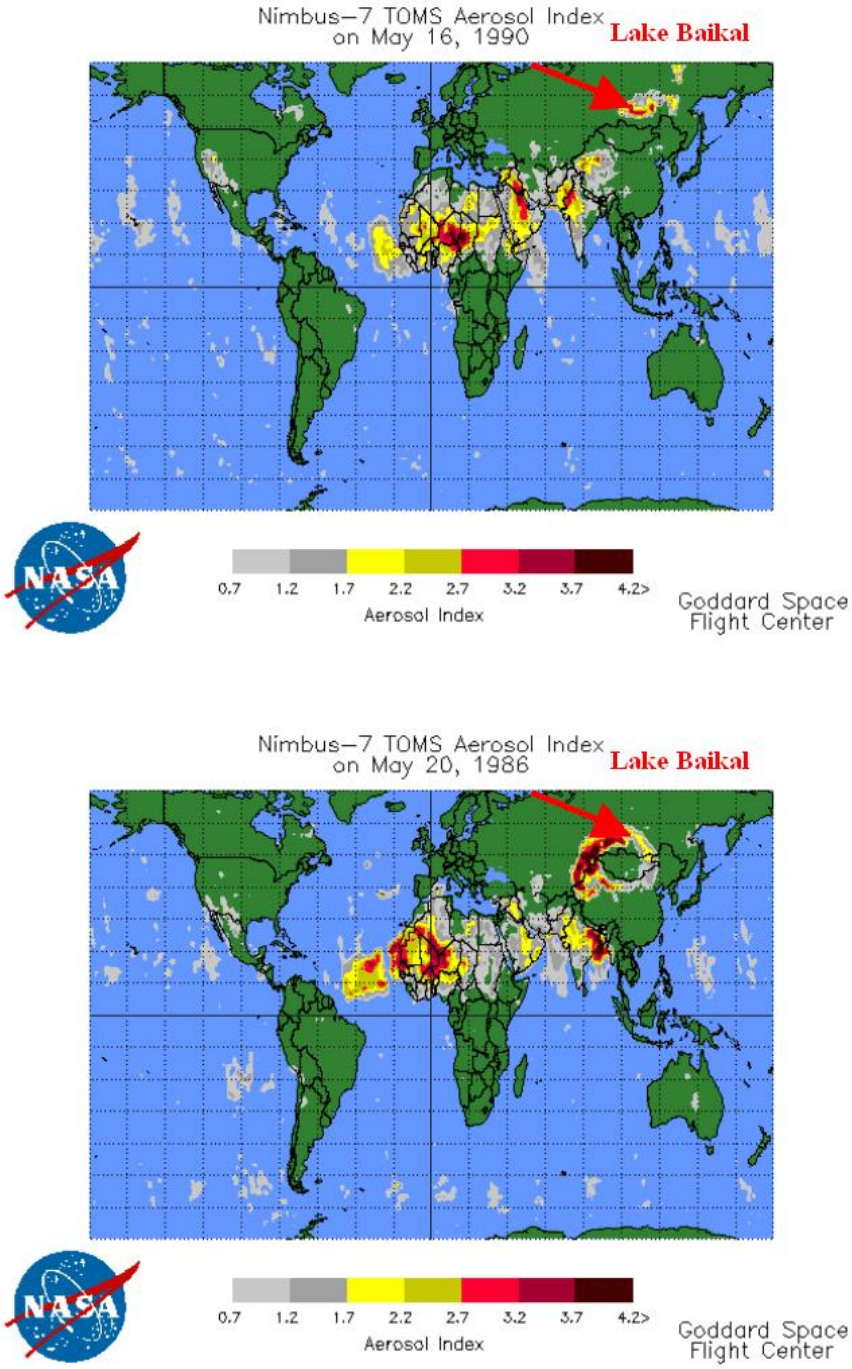


Figure 4.11: Satellite-based measurement of present-day atmospheric aerosols in late spring. One clearly observes a dust cloud moving from the Taklamakan desert towards the Lake Baikal region.

Chapter 5

Summary

1. An important task of modern studies in climatology is the separation of natural and anthropogenic contributions to the observed climate change. For this purpose, a detailed knowledge about the natural variability of the climate system during warm stages is essential. Besides model simulations and the study of historical observations spanning the last centuries, the corresponding information is mainly taken from the study of palaeoclimatic proxies. In order to appropriately interpret the corresponding records, sophisticated statistical modelling approaches and methods of linear and nonlinear time series analysis are required which must be (in contrast to most conventional methods) applicable to very short, noisy, and instationary univariate as well as multivariate time series.

2. Correlations of palaeoclimatic proxies observed at different locations or of multiple proxies from one specific site include important information about climate change. It has been demonstrated that the diagonalisation of the covariance matrix of a multivariate time series in terms of Karhunen-Loéve decomposition (KLD) allows to define estimates of the dimensions, i.e., the number of statistically relevant components of such data sets. To improve the performance of earlier approaches, the linear variance decay (LVD) dimension density is introduced. It has been demonstrated that this method is sensitive with respect to small changes in the correlation structure. Modifications or extensions of the presented method are possible by substituting the KLD by other statistical decomposition methods.

3. The LVD dimension density has been used to analyse the temporal variations of the information content in multivariate palaeoclimatic time series. As an example, trace element abundances and grain-size distributions from Cape Roberts, East Antarctica, have been studied. The corresponding investigation has demonstrated a significant change of the LVD dimension density across the Oligocene/Miocene boundary about 24 Myr BP, which is related to a rather abrupt change of sediment provenance and an intensification of glacial activity. In addition, the sensitivity of this measure to outliers in the multivariate data allows the identification of short-time events caused by certain climatic or non-climatic extreme events influencing the geological source.

4. The instationarity of palaeoclimatic records is often related to the fact that external influences forcing the climate system as well as internal oscillations are amplified temporarily. In order to study the resulting varying influence of different time-periodic contributions in univariate climate records, wavelet analysis is a promising tool. As palaeoclimatic time series are characterised by an uneven sampling in the time domain, the traditional approach has to be modified, for example, in terms of the weighted wavelet Z transform.

The resulting time-frequency pattern depends crucially on the applied age model and its uncertainty. To improve the validity of the results of a wavelet analysis in palaeoclimatology, a combination with stochastic techniques like Monte Carlo resampling or Bayesian statistics is outlined.

5. Grain-size distributions are an important palaeoclimate proxy which allows to identify and quantify different mechanisms of sediment transport and deposition caused by varying palaeoenvironmental conditions. To extract the corresponding information from the observed data sets, special techniques of multivariate statistical analysis and modelling are required. In particular, finite mixture models are a sophisticated approach for the statistical modelling of grain-size distributions. The number of components corresponds to different origins of the deposited material, whereas the shape of the components is related to the dominating transport mechanisms.

6. The EM algorithm is discussed as an efficient method for parameter estimation in such models basing on grouped and eventually truncated observational data. For Gaussian mixture models, it has been shown that the performance of this approach is closely related to the component overlaps and weights, the grouping coarseness, and the truncation of the observed data. In order to estimate the statistical uncertainty of the resulting parameter estimates, methods based on approximations of the information matrix as well as the resampling of either the original grouped data or the estimated probability distribution functions have been compared. Consistency of both approaches is observed only under rather idealised conditions like an optimal knowledge of the model structure and a reasonably small component overlap.

7. For real-world observations of grain-size distributions, the optimum number and shape of the usually strongly overlapping component functions is not a priori known. Consequently, information- and resampling based methods may yield inconsistent results. An additional problem of this kind of data is that the absolute uncertainty depends on the number of observations which is not known in the case of grain-size distributions which are described in terms of relative frequencies. For the statistical assessment of parameter uncertainty important for palaeoclimatic interpretations of the results of modelling, asymptotic uncertainty distributions have been proposed as a new resampling-based concept applicable in this situation. This concept uses the complete information about the probability of parameter estimates in appropriately simulated data sets and is (in contrast to traditional uncertainty estimates) also applicable in the case of relative group frequencies.

8. For grain-size distributions from a sedimentary sequence obtained in Lake Baikal, Eastern Siberia, it has been shown that the finite mixture approach does not give results that can be well interpreted palaeoclimatically. As an alternative, linear principal component analysis (PCA) allows to define different size classes which are closely related to heuristically defined parameters frequently used in geological studies. The abundance of material $> 40\mu$m is found to be a useful proxy for the snow cover at Lake Baikal, whereas the abundance of fine silts ($6-16\mu$m) is closely related to the strength of late spring / early summer heating over Central Eurasia and may be related to aeolian dust from the Taklamakan desert. Wavelet analysis demonstrates that the two major principal components show significant oscillatory contributions with the typical frequencies of the precession and obliquity variations of the Earth.

Danksagung

An dieser Stelle möchte ich noch einmal all jenen danken, die mich in meiner Arbeit in den vergangenen Jahren und speziell bei der Anfertigung dieser Dissertation fachlich wie moralisch unterstützt haben. Mein Dank geht in erster Linie an Dr. Annette Witt und Prof. Jürgen Kurths, die mir die Möglichkeit gegeben haben, in der Arbeitsgruppe Nichtlineare Dynamik nicht nur über die Analyse und Modellierung von Paläoklima-Zeitreihen zu forschen, sondern darüber hinaus auch an vielen anderen spannenden Fragestellungen zum Verhalten komplexer Systeme in Natur, Wirtschaft und Gesellschaft mitzuarbeiten. In diesem Zusammenhang möchte ich auch all jene Institutionen nicht unerwähnt lassen, die mir durch die Gewährung finanzieller Unterstützung die Möglichkeit gegeben haben, meine Arbeit in den vergangenen Jahren kontinuierlich fortzuführen und mich bei Workshops und Tagungen mit internationalen Kollegen auszutauschen und weiterzubilden. Hierfür möchte ich der Volkswagen-Stiftung, der Deutschen Forschungsgemeinschaft (SFB 555), der Europäischen Union (NEST-Projekt E2C2) sowie der Universität Potsdam und den Veranstaltern der Grand Combin Summer School 2003 im italienischen Aostatal sehr herzlich danken.

Wissenschaft lebt nicht allein von den vielen einsamen Stunden vor dem Rechner, sondern auch vom lebendigen, intensiven und knstruktiven Austausch von Ideen und Ergebnissen mit anderen Kollegen. An dieser Stelle möchte ich daher ganz besonders Dr. Annette Witt, Dr. Maria Carmen Romano, Dr. Marco Thiel und Dr. Udo Schwarz danken, die bei Fragen immer sofort ein offenes Ohr für mich hatten und mich bei der Lösung von Problemen zu jeder Zeit tatkräftig unterstützt haben. Danken möchte ich auch den Kollegen, die es mir durch intensive Diskussionen im kleinen Kreis wie auch auf Konferenzen und Workshops ermöglicht haben, meine Kenntnisse im Bereich der Geologie und Paläoklimatologie immer weiter zu erweitern, und die mich stets zur Weiterführung meiner entsprechenden Arbeiten zur Zeitreihenanalyse von Beobachtungsdaten aus diesem Bereich angespornt haben. Besonders nennen möchte ich an dieser Stelle meine Kooperationspartner vom Geoforschungszentrum Potsdam und den mit diesem kooperierenden Einrichtungen, Dr. Hedi Oberhänsli, Dr. François Demory und Hans von Suchodoletz. Die beiden erstgenannten Kollegen haben darüber hinaus auch einige der in Kapitel 4 dieser Arbeit verwendeten Grafiken zur Verfügung gestellt. Herrn Dr. Maarten Prins danke ich für die Einladung zur Präsentation meiner Ergebnisse zur statistischen Modellierung von Korngrößenverteilungen in Sedimenten auf der letzten Jahrestagung der European Geosciences Union und die dortigen intensiven und produktiven Diskussionen. Prof. Martin Claussen und Prof. Gerald Haug sowie den Dozenten der Grand Combin Summer School 2003 möchte ich schließlich dafür meinen Dank aussprechen, dass sie mir durch ihre ausgezeichneten Vorlesungen über Paläoklimatologie bzw. Quartärgeologie die entsprechenden fachlichen Grundlagen zur Anfertigung dieser Dissertation vermittelt haben.

Da Zeitreihenanalyse von dem Vorhandensein guter Mess- und Beobachtungsdaten abhängt, gebührt ein spezieller Dank all jenen Kolleginnen und Kollegen, die in langen Stunden in Feld und Labor diese Daten erhoben haben und mir die Möglichkeit gegeben haben, mit diesen zu

arbeiten und sie auch zum Teil in dieser Dissertation zu verwenden. Insbesondere danke ich den Mitarbeitern des CONTINENT-Projekts am GFZ Potsdam und den anderen beteiligten Einrichtungen, speziell Dr. Hedi Oberhänsli, Dr. François Demory und Matthias Zopperitsch, für die Bereitstellung der Korngrößendaten aus dem Baikalsee. Hans von Suchodoletz hat mir darüber hinaus weitere Daten von verschiedenen Sedimentproben aus dem Bereich der Kanarischen Inseln zur Verfügung gestellt, die im Rahmen dieser Dissertation leider nicht mit präsentiert werden konnten. Für die Erhebung und die Möglichkeit des weltweiten Abrufs der Daten aus dem Cape-Roberts-Projekt über die PANGAEA-Datenbank danke ich allen an diesem Projekt beteiligten Kolleginnen und Kollegen. Die im ersten Kapitel dieser Arbeit als Beispiel heran gezogenen meteorologischen Zeitreihen wurden freundlichenweise vom Deutschen Wetterdienst und den Wissenschaftlichen Mitarbeitern des Armagh Meteorological Observatory zur Verfügung gestellt. Keinen Platz in dieser Arbeit gefunden haben schließlich hydro-meteorologische Zeitreihen (insbesondere Niederschläge und Abflüsse), die von mir in Zuammenarbeit mit Malaak Kallache, Henning Rust und Dr. Jürgen Kropp vom Potsdam-Institut für Klimafolgenforschung analysiert wurden. Für die entsprechenden intensiven Diskussionen möchte ich den genannten Kolleginnen und Kollegen an dieser Stelle noch einmal herzlich danken.

Last but not least braucht gute Wissenschaft nicht nur eine solide Datengrundlagen und freundliche und fachkundige Kollegen, sondern auch ein entsprechendes Umfeld, in welchem man Verwaltungsfragen und sonstige technische Dinge in guten Händen weiß. Ich möchte daher zwei Menschen einen besonderen Dank aussprechen: unserem Systemadministrator Jörg-Uwe Tessmer alias Tessi für den hervorragenden technischen Support und das Immer-Ansprechbar-Sein, und unserer Sekretärin Birgit Voigt für ihre Mithilfe und Unterstützung bei allen Fragen rund um Arbeitsverträge und andere Verwaltungsangelegenheiten. Herrn Siegfried Riedel danke ich für die Überlassung eines Labtops, auf dem der grösste Teil dieser Arbeit entstanden ist.

Schließlich und endlich gebührt der größte Dank jedoch meiner Frau und meinem Sohn, die mich in schwierigen Zeiten immer wieder aufgefangen haben, aber auch (fast) immer Verständnis dafür hatten, wenn mein Arbeitspensum in der Woche (wie meistens) die 40 Stunden wieder einmal deutlich überschritten hat. Ohne Eure Hilfe und Unterstützung wäre es sicher nicht möglich gewesen, die Ergebnisse zu erzielen, die im Rahmen dieser Dissertation und darüber hinaus in den vergangenen Jahren entstanden sind.

List of Publications

The results presented in this thesis have been published in several scientific articles, which are listed in the following (in the order of occurrence of the corresponding topic in this work):

R. Donner, A. Witt: *Interactive Comment on "Orbital forcings of the Earth's climate in wavelet domain" by A.V. Glushkov et al.* Climate of the Past Discussions **1**, S127-S137 (2005).

R. Donner: *Interdependences between Daily European Temperature Records: Correlation or Phase Synchronisation ?* In: P. Marquié (ed.): *Nonlinear Dynamics of Electronic Systems (NDES 2006).* Dijon, France, 26-29 (2006).

R. Donner: *Uncertainty Assessment in Parameter Estimation of Finite Mixture Distributions with Grouped Truncated Data.* Computational Statistics & Data Analysis, submitted.

R. Donner, A. Witt: *Temporary dimensions of multivariate data from paleoclimate records - A novel measure for dynamic characterization of long-term climate change.* International Journal of Bifurcation and Chaos, in press.

R. Donner, A. Witt: *Characterisation of Long-Term Climate Change by Dimension Estimates of Multivariate Palaeoclimatic Proxy Data.* Nonlinear Processes in Geophysics, accepted.

R. Donner: *Spatial Correlations of Hydro-Meteorological Records in a River Catchment.* In: J. Kropp, H.-J. Schellnhuber (eds.): *Correlations and Extremes in Hydrology and Climate.* (Springer, Berlin, in preparation), submitted.

F. Demory, R. Donner, H. Oberhänsli, M. Zopperitsch, A. Trenteseaux: *Dynamics of the detrital sedimentation in Lake Baikal (Siberia) during the last 140 ka: paleoclimatic implications.* In preparation.

Further results, which are of particular interest with respect to further problems in geosciences or applications of several methods discussed in this thesis, are discussed in the following articles:

A. Cser, R. Donner, U. Schwarz, A. Otto, M. Geiger, U. Feudel: *Towards a better understanding of laser beam melt ablation using methods of statistical analysis.* In: R. Teti (eds.): *Intelligent Computation in Manufacturing Engineering - 3.* CIRP, Paris, 203-208 (2002).

R. Donner, A. Cser, U. Schwarz, A. Otto, U. Feudel: *An Approach to a Process Model of Laser Beam Melt Ablation Using Methods of Linear and Non-linear Data Analysis.* In: Proceedings of the 4th International Symposium on Investigations of Non-Linear Dynamic Effects in Production Systems, Chemnitz (Germany), April 8-9, Art.-No. 31 (2003).

R. Donner, A. Cser, U. Schwarz, A. Otto, U. Feudel: *An Approach to a Process Model of Laser Beam Melt Ablation Using Methods of Linear and Non-linear Data Analysis.* In: G. Radons, R. Neugebauer (eds.): *Nonlinear Dynamics of Production Systems.* Wiley Europe, Weinheim, 443-458 (2004).

K. Bube, C. Rodrigues Neto, R. Donner, U. Schwarz, U. Feudel: *Linear and nonlinear characterization of surfaces from a laser beam melt ablation process.* Journal of Physics D: Applied Physics, **39**(7), 1405-1412 (2006).

B. Scholz-Reiter, U. Hinrichs, R. Donner, A. Witt: *Modelling of Networks of Production and Logistics and Analysis of Their Nonlinear Dynamics.* In: Wamkeue, R. (ed.): *Modelling and Simulation.* IASTED, Montreal, Quebec, Canada, 178-183 (2006).

R. Donner, U. Hinrichs, B. Scholz-Reiter, A. Witt: *Nonlinear Dynamics and Control of Small-Scale Manufacturing Networks.* In: P. Marquié (ed.): *Nonlinear Dynamics of Electronic Systems (NDES 2006).* Dijon, France, 22-25 (2006).

R. Donner, F. Feudel, N. Seehafer, M.A.F. Sanjuan: *Hierarchical modelling of a forced Roberts dynamo.* International Journal of Bifurcation and Chaos, in press.

R. Donner, N. Seehafer, M.A.F. Sanjuan, F. Feudel: *Low-dimensional dynamo modelling and symmetry-breaking bifurcations.* Physica D, submitted.

Bibliography

[Abarbanel 1996] Abarbanel, H.D.I.. *Analysis of Observed Chaotic Data:* Springer, New York (1996).

[Adamidis and Loukas 1993] Adamidis, K.; Loukas, S.: *ML estimation in the Poisson binomial distribution with grouped data via the EM algorithm.* Journal of Statistical Computation and Simulation **45**, 33-39 (1993).

[Adamidis 1999] Adamidis, K.: *An EM algorithm for estimating negative binomial parameters.* Australian and New Zealand Journal of Statistics **41**(2), 213-221 (1999).

[Adams 1969] Adams, A.G.: *Areas Under the Normal Curve (Algorithm 39).* Computer Journal **12**, 197-198 (1969).

[Agha and Ibrahim 1984] Agha, M.; Ibrahim, M.T.: *Algorithm AS 203: Maximum Likelihood Estimation of Mixtures of Distributions.* Applied Statistics 33, 327-332 (1984).

[Aitchison 1982] Aitchison, J.: *The Statistical Analysis of Compositional Data.* Journal of the Royal Statistical Society B **44**(2), 139-177 (1982).

[Aitchison 1983] Aitchison, J.: *Principal Component Analysis of Compositional Data.* Biometrika **70**(1), 57-65 (1983).

[Aitchison 1986] Aitchison, J.: *The Statistical Analysis of Compositional Data.* Chapman and Hall, London (1986).

[Aitchison 2002] Aitchison, J.: *Biplots of Compositional Data.* Applied Statistics **51**(4), 375-392 (2002).

[Aitkin and Aitkin 1996] Aitkin, M,; Aitkin, I.: *A hybrid EM/Gauss-Newton algorithm for maximum likelihood in mixture distributions.* Statistics and Computing **6**, 127-130 (1996).

[Alley et al. 2001] Alley, R.B.; Anandakrishnan, S.; Jung, P.: *Stochastic Resonance in the North Atlantic.* Paleoceanography **16**(2), 190-198 (2001).

[Andronov 1997] Andronov, I.L.: *Method of running parabolae: Spectral and statistical properties of the smoothing function.* Astronomy and Astrophysics Supplement Series **125**, 207-217 (1997).

[Andronov 1998] Andronov, I.L.: *Wavelet Analysis of Time Series by the Least-Squares Method with Supplementary Weights.* Kinematics and Physics of Celestial Bodies **14**(6), 37-392 (1998).

[Andronov 1999] Andronov, I.L.: *Wavelet analysis of the irregularly spaced time series.* In: Priezzhev, V.B.; Spiridonov, V.P. (eds.): *Self-Similar Systems.* Joint Institute for Nuclear Research, Dubna, 57-70 (1999).

[Arcidiacono and Jones 2003] Arcidiacono, P.; Jones, J.B.: *Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm.* Econometrica **71** (3), 933-946 (2003).

[Armienti et al. 1998] Armienti, P.; Messiga, B.; Vannucci, R.: *Sand provenance from major and trace element analyses of bulk rock and sand grains.* Terra Antartica **5**(3), 589-599 (1998).

[Asghari et al. 2004] Asghari, N.; Broeg, C.; Carone, L.; Casas-Miranda, R.; Castro Palacio, J.C.; Csillik, I.; Dvorak, R.; Freistetter, F.; Hadjivantsides, G.; Hussmann, H.; Khramova, A.; Khristoforova, M.; Khromova, I.; Kitiashivilli, I.; Kozlowski, S.; Laakso, T.; Laczkowski, T.; Lytvinenko, D.; Miloni, O.; Morishima, R.; Moro-Martin, A.; Paksyutov, V.; Pal, A.; Patidar, V.; Pečnik, B.; Peles, O.; Pyo, J.; Quinn, T.; Rodriguez, A.; Romano, M.C.; Saikia, E.; Stadel, J.; Thiel, M.; Todorovic, N.; Veras, D.; Vieira Neto, E.; Vilagi, J.; von Bloh, W.; Zechner, R.; Zhuchkova, E.: *Stability of Terrestrial Planets in the Habitable Zone of GI 777 A, HD 72659, GI 614, 47 Uma and HD 4208.* Astronomy and Astrophysics **426**, 353-365 (2004).

[Azais et al. 2004] Azais, J.M.; Gassiat, E.; Mercadier, C.: *Asymptotic distribution and power of the likelihood ratio test for mixtures: bounded and unbounded cases.* Preprint (2004).

[Bagnold and Barndorff-Nielsen 1980] Bagnold, R.A.; Barndorff-Nielsen, O.: *The pattern of natural size distributions.* Sedimentology **27**, 199-207 (1980).

[Baker 1992] Baker, S.G.: *A Simple Method for Computing the Observed Information Matrix When Using the EM Algorithm With Categorical Data.* Journal of Computational and Graphical Statistics **1**, 63-76 (1992).

[Bard 2004] Bard, E.: *Greenhouse effect and ice ages: historical perspective.* Comptes Rendus Geoscience **336**, 603-638 (2004).

[Barker and Burrell 1977] Barker, P.F.; Burrell, J.: *The Opening of Drake Passage.* Marine Geology **25**, 15-34 (1977).

[Barker and Burrell 1982] Barker, P.F.; Burrell, J.: *The Influence upon Southern Ocean Circulation, Sedimentation, and Climate of the Opening of Drake Passage.* In: Craddock, C. (ed.): *Antarctic Geoscience* University of Wisconsin Press, Madison, WI, 377-385 (1982).

[Barker 2001] Barker, P.F.: *Scotia Sea regional tectonic evolution: implications for mantle flow and palaeocirculation.* Earth-Science Reviews **55**, 1-39 (2001).

[Barker and Thomas 2004] Barker, P.F.; Thomas, E.: *Origin, signature and palaeoclimatic influence of the Antarctic Circumpolar Current.* Earth-Science Reviews **66**, 143-162.

[Barrett and Anderson 2000] Barrett, P.; Anderson, J.B.: *Grain-size amalysis of samples from CRP-2/2A, Victoria Land Basin, Antarctica.* Terra Antartica **7**(3), 373-378 (2000).

[Barrett and Anderson 2003] Barrett, P.; Anderson, J.B.: *Frequency percent in each size class for grain size analysis of sediment core CRP-2/2A (Tab. 1).* PANGAEA database, doi:10.1594/PANGAEA.133964 (2003).

[Basford et al. 1997] Basford, K.E.; Greenway, D.R.; McLachlan, G.J.; Peel, D.: *Standard Errors of Fitted Component Means of Normal Mixtures.* Computational Statistics **12**, 1-17 (1997).

[Bauer et al. 1993] Bauer, M.; Heng, H.; Martienssen, W.: *Characterization of Spatiotemporal Chaos from Time Series.* Physical Review Letters **71**(4), 521-524 (1993).

[Bayly et al. 1998] Baily, P.V.; KenKnight, B.H.; Rogers, J.M.; Johnson, E.E.; Ideker, R.E.; Smith, W.M.: *Spatial organization, predictability, and determinism in ventricular fibrillation.* Chaos **8**(1), 103-115 (1998).

[Behboodian 1972] Behboodian, J.: *Information Matrix for a Mixture of Two Normal Distributions.* Journal of Statistical Computation and Simulation **1**, 295-314 (1972).

[Benzi et al. 1982] Benzi, R.; Parisi, G.; Sutera, A.; Vulpiani, A.: *Stochastic resonance in climatic change.* Tellus **34**, 10-16 (1982).

[Benzi et al. 1983] Benzi, R.; Parisi, G.; Sutera, A.; Vulpiani, A.: *A theory of stochastic resonance in climatic change.* SIAM Journal of Applied Mathematics **43**(3), 565-578 (1983).

[Berger 1978] Berger, A.: *Long-term variations of daily insolation and Quaternary climatic changes.* Journal of Atmospheric Sciences **35**, 2362–2367 (1978).

[Berger and Loutre 2004] Berger, A.; Loutre, M.F.: *Astronomical theory of climate change.* Journal de Physique IV France **121**, 1-35 (2004).

[Berndt et al. 1974] Berndt, E.K.; Hall, B.H.; Hall, R.E.; Hausman, J.A.: *Estimation and Inference in Nonlinear Structural Models.* Annals of Economic and Social Measurement **3/4**, 653-665 (1974).

[Bianchi and McCave 1999] Bianchi, G.G.; McCave, N.: *Holocene periodicity in North Atlantic climate and deep-ocean flow south of Iceland.* Nature **397**, 515-517 (1999).

[Biernacki et al. 2003] Biernacki, C.; Celeux, G.; Govaert, G.: *Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models.* Computational Statistics & Data Analysis **41**, 561-575 (2003).

[Biernacki and Chrétien 2003] Biernacki, C.; Chrétien, S.: *Degeneracy in the maximum likelihood estimation of univariate Gaussian mixtures with EM.* Statistics & Probability Letters **61**, 373-382 (2003).

[Biernacki 2004a] Biernacki, C.: *An Asymptotic Upper Bound of the Likelihood to Prevent Gaussian Mixtures from Degenerating.* Preprint, Université de Franche-Comté, Besançon (2004).

[Biernacki 2004b] Biernacki, C.: *Degeneracy in the Maximum Likelihood Estimation of Univariate Gaussian Mixtures for Grouped Data and Behaviour of the EM Algorithm.* Preprint, Université de Franche-Comté, Besançon (2004).

[Biernacki 2004c] Biernacki, C.: *Influence of the Bin Dimension on Selecting a Model by the BIC criterion in Gaussian mixtures with Grouped Data.* Preprint, Université de Franche-Comté, Besançon (2004).

[Billups et al. 2004] Billups, K.; Pälike, H.; Channell, J.E.T.; Zachos, J.C.; Shackleton, N.J.: *Astronomic calibration of the late Oligocene through early Miocene geomagnetic polarity time scale.* Earth and Planetary Science Letters **224**, 33-44 (2004).

[Blott and Pye 2001] Blott, S.J.; Pye, K.: *GRADISTAT: A Grain Size Distribution and Statistics Package for the Analysis of Unconsolidated Sediments.* Earth Surface Processes and Landforms **26**, 1237-1248 (2001).

[Böhning et al. 1992] Böhning, D.; Schlattmann, P.; Lindsay, B.: *Computer-Assisted Analysis of Mixtures (C.A.MAN): Statistical Algorithms.* Biometrics **48**, 283-303 (1992).

[Böhning et al. 1994] Böhning, S.; Dietz, E.; Schaub, R.; Schlattmann, P.; Lindsay, B.G.: *The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family.* Annals of the Institute of Statistical Mathematics **46**(2), 373-388 (1994).

[Böhning 2002] Böhning, D.: *The EM Algorithm with Gradient Function Update for Discrete Mixtures with Known (Fixed) Number of Components.* Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 101, Universität der Bundeswehr Hamburg (2002).

[Bolton et al. 1995] Bolton, E.W.; Maasch, K.A.; Lilly, J.M.: *A wavelet analysis of Plio-Pleistocene climate indicators: A new view of periodicity evolution.* Geophysical Research Letters **22**(20), 2753-2756 (1995).

[Bond et al. 1997] Bond, G.; Showers, W.; Cheseby, M.; Lotti, R.; Almasi, P.; de Menocal, P.; Priore, P.; Cullen, H.; Hajdas, I.; Bonani, G.: *A Pervasive Millenial-Scale Cycle in North Atlantic Holocene and Glacial Climates.* Science **278**(5341), 1257-1266 (1997).

[Bond et al. 2001] Bond, G.; Kromer, B.; Beer, J.; Muscheler, R.; Evans, M.N.; Showers, W.; Hoffmann, S.; Lotti-Bond, R.; Hajdas, I.; Bonani, G.: *Persistent Solar Influence on North Atlantic Climate During the Holocene.* Science **294**, 2130-2136 (2001).

[Brauer et al. accepted] Brauer, A.; Mangili, C.; Moscariello, A.; Witt, A.: *Palaeoclimatic implications from micro-facies data of a 5,900 varve time series from the Pianico interglacial sediment record, Southern Alps.* Paleogeography Paleoclimatology Paleoecology, accepted.

[Braun 1975] Braun, A.F.: *Die genetische Deutung natürlicher Haufwerke mit Hilfe des doppeltlogarithmischen Körnungsnetzes nach Rosin, Rammler und Sperling (DIN 4190).* Zeitschrift der deutschen geologischen Gesellschaft **126**, 199-205 (1975).

[Breiman and Friedman 1985] Breiman, L.; Friedman, J.H.: *Estimating Optimal Transformations for Multiple regression and Correlation.* Journal of the American Statistical Society **80**(391), 580-595 (1985).

[Broomhead and King 1986] Broomhead, D.S.; King, G.P.: *Extracting Qualitative Dynamics from Experimental Data.* Physica D **20**(2-3), 217-236 (1986).

[Bünner and Hegger 1999] Bünner, M.J.; Hegger, R.: *Estimation of Lyapunov spectra from space-time data.* Physics Letters A **258**(1), 25-30 (1999).

[Burlaga and Klein 1986] Burlaga, L.F.; Klein, L.W.: *Fractal structure of the interplanetary magnetic field.* Journal of Geophysical Research **91**(A1), 347-350 (1986).

[Butler et al. 2005] Butler, C.J.; Garca Surez, A.M.; Coughlin, A.D.S.; Morrell, C.: *Air temperatures at Armagh Observatory, Northern Ireland, from 1796 to 2002.* International Journal of Climatology **25** 1055-1079 (2005).

[Cadez et al. 1999] Cadez, I.V.; McLaren, C.E.; Smyth, P.; McLachlan, G.J.: *Hierarchical Models for Screening of Iron Deficiency Anemia.* In: Bratko, I.; Džeroski, S. (eds.): Proceedings of the Sixteenth International Conference on Machine Learning (ICML '99), 77-86 (1999).

[Cadez et al. 2002] Cadez, I.V.; Smyth, P.; McLachlan, G.J.; McLaren, C.E.: *Maximum Likelihood Estimation of Mixture Densities for Binned and Truncated Multivariate Data.* Machine Learning **47**, 7-34 (2002).

[Cande and Kent 1992] Cande, S.C.; Kent, D.V.: *A new geomagnetic polarity time scale for the Late Cretaceous and Cenozoic.* Journal of Geophysical Research **97**, 13917-13951 (1992).

[Ceramicola et al. 2002] Ceramicola, S.; Rebesco, M.; De Batist, M.; Khlystov, O.: *Seismic evidence of small-scale lacustrine drifts in Lake Baikal (Russia).* Marine Geophysical Researches **22**(5-6), 445-464 (2002).

[Cande and Kent 1995] Cande, S.C.; Kent, D.V.: *Revised calibration of the geomagnetic polarity time scale for the Late Cretaceous and Cenozoic.* Journal of Geophysical Research **100**, 6093-6095 (1995).

[Chambers and Upchurch 1979] Chambers, R.L.; Upchurch, S.B.: *Multivariate Analysis of Sedimentary Environments Using Grain-Size Frequency Distributions.* Mathematical Geology **11**(1), 27-43 (1979).

[Channell 1999] Channell, J.E.T.: *Geomagnetic paleointensity and directional secular variation at Ocean Drilling Program (ODP) site 984 (Bjorn Drift) since 500 ka: Comparisons with ODP site 983 (Gardar drift).* Journal of Geophysical Research B: Solid Earth **104**(10), 22,937-22,951 (1999).

[Charlet et al. 2005] Charlet, F.; Fagel, N.; De Batist, M.; Hauregard, F.;, Minnebo, B.; Meischner, D.; Team, T.S.: *Sedimentary dynamics on isolated highs in Lake Baikal: evidence from detailed high-resolution geophysical data and sediment cores.* Global and Planetary Change **46**(1-4), 125-144 (2005).

[Chen et al. 2001] Chen, H.; Chen, J.; Kalbfleisch, J.D.: *A modified likelihood ratio test for homogeneity in finite mixture models.* Journal of the Royal Statistical Society B **63**(1), 19-29 (2001).

[Christiansen and Hartmann 1988] Christiansen, C.; Hartmann, D.: *On Using the Log-Hyperbolic Distribution to Describe the Textural Characteristics of Eolian Sediments - Discussion.* Journal of Sedimentary Petrology **58**(1), 159-160 (1988).

[Ciliberto and Nicolaenko 1991] Ciliberto, S.; Nicolaenko, B.: *Estimating the Number of Degrees of Freedom in Spatially Extended Systems.* Europhysics Letters **14**(4), 303-308 (1991).

[Clark 1976] Clark, M.W.: *Some Methods for Statistical Analysis of Multimodal Distributions and Their Applications to Grain-Size Data.* Mathematical Geology **8**(3), 267-281 (1976).

[Conover1980] Conover, W.J.: *Practical Nonparametric Statistics.* 2nd edition. Wiley, New York (1980).

[Cox and Cox 2000] Cox, T.F.; Cox, M.A.A.: *Multidimensional Scaling.* 2nd edition. Chapman and Hall, London (2000).

[Craddock and Flood 1969] Craddock, J.M.; Flood, C.R.: *Eigenvectors for representing the 500 mb geopotential surface over the Northern Hemisphere.* Quarterly Journal of the Royal Meteorological Society **95**, 576-593 (1969).

[Cramer 1946] Cramer, H.: *Mathematical Methods of Statistics.* Princeton University Press, Princeton (1946).

[Dansgaard et al. 1993] Dansgaard, W.; Johnsen, S.J.; Clausen, H.B.; Dahl-Jensen, D.; Gundestrup, N.S.; Hammer, C.U.; Hvidberg, C.S.; Steffensen, J.P.; Sveinbjörnsdottir, A.E.; Jouzel, J., Bond, G.: *Evidence for general instability of past climate from a 250-kyr ice-core record.* Nature **364**, 218-220 (1993).

[Davenport et al. 1988] Davenport, J.W.; Pierce, M.A.; Hathaway, R.J.: *A Numerical Comparison of EM and Quasi-Newton Type Algorithms for Computing MLE's for a Mixture of Normal Distributions.* Computing Science and Statistics - Proceedings of the 20th Symposium on the Interface. American Statistical Association, Alexandria, Virginia, 410-415 (1988).

[Davis 1970] Davis, J.C.: *Information Contained in Sediment-Size Analyses.* Mathematical geology **2**(2), 105-112 (1970).

[Day 1969] Day, N.E.: *Estimating the components of a mixture of normal distributions.* Biometrika **56**(3), 463-474 (1969).

[Demory 2004] Demory, F.: *Paleomagnetic dating of climatic events in late Quaternary sediments of Lake Baikal (Siberia).* PhD Thesis, University of Potsdam (2004).

[Demory et al. 2005a] Demory, F.; Nowaczyk, N.R.; Witt, A.; Oberhänsli, H.: *High-resolution magnetostratigraphy of late Quaternary sediments from Lake Baikal, Siberia: timing of intracontinental paleoclimatic responses.* Global and Planetary Change **46**(1-4), 167-186 (2005).

[Demory et al. 2005b] Demory, F.; Oberhänsli, H.; Nowaczyk, N.R.; Gottschalk, M.; Wirth, R.; Naumann, R.: *Detrital input and early diagenesis in sediments from Lake Baikal revealed by rock magnetism.* Global and Planetary Change **46**(1-4), 145-166 (2005).

[Dempster et al. 1977] Dempster, A.P.; Laird, N.M.; Rubin, D.B.: *Maximum Likelihood from Incomplete Data via the EM Algorithm.* Journal of the Royal Statistical Society B **39** (1), 1-38 (1977).

[Demske et al. 2002] Demske, D.; Mohr, B.; Oberhänsli, H.: *Late Pliocene vegetation and climate of the Lake Baikal region, southern East Siberia reconstructed from palynological data.* Palaeogeography, Palaeoclimatology, Palaeoecology **184**(1-2), 107-129 (2002).

[Demske et al. 2005] Demske, D.; Heumann, G.; Granoszewski, W.; Nita, M.; Mamakowa, K.; Tarasov, P.E.; Oberhänsli, H.: *Late glacial and Holocene vegetation and regional climate variability evidenced in high-resolution pollen records from Lake Baikal.* Global and Planetary Change **46**(1-4), 255-279 (2005).

[Doeglas 1946] Doeglas, D.J.: *Interpretation of the Results of Mechanical Analyses.* Journal of Sedimentary Petrology **16**(1), 19-40 (1946).

[Dolan and Molenaar 1991] Dolan, C.V.; Molenaar, P.C.M.: *A comparison of four methods of calculating standard errors of maximum-likelihood estimates in the analysis of covariance structure.* British Journal of Mathematical and Statistical Psychology **44**, 359-368 (1991).

[Drozdz et al. 2000] Drozdz, S.; Grümmer, F.; Ruf, F.; Speth, J.: *Dynamics of competition between collectivity and noise in the stock market.* Physica A **287**(3-4), 440-449 (2000).

[Drozdz et al. 2001] Drozdz, S.; Kwapień, J.; Grümmer, F.; Ruf, F.; Speth, J.: *Quantifying dynamics of financial correlations.* Physica A **299**(1-2), 144-153 (2001).

[Eckmann et al. 1987] Eckmann, J.-P.; Oliffson Kamphorst, S.; Ruelle, D.: *Recurrence plots of dynamic systems.* Europhysics Letters **4**(9), 973-977 (1987).

[Efron and Hinkley 1978] Efron, B.; Hinkley, D.V.: *Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information.* Biometrika **65** (3), 457-487 (1978).

[Efron and Tibshirani 1993] Efron, B.; Tibshirani, R.J.: *An Introduction to the Bootstrap.* Chapman and Hall, Boca Raton (1993).

[Elkibbi and Rial 2001] Elkibbi, M.; Rial, J.A.: *An outsider's review of the astronomical theory of the climate: is the eccentricity-driven insolation the main driver of ice ages ?* Eart-Science Reviews **56**, 161-177 (2001).

[Everitt and Hand 1981] Everitt, B.S.; Hand, D.J.: *Finite Mixture Distributions.* Chapman and Hall, London (1981).

[Everitt 1984] Everitt, B.S.: *Maximum Likelihood Estimation of the Parameters in a Mixture of Two Univariate Normal Distributions; a Comparison of Different Algorithms.* The Statistician, Journal of the Royal Statistical Society D **33**, 205-215 (1984).

[Fagel et al. 2003] Fagel, N.; Boski, T.; Likhoshway, L.; Oberhänsli, H.: *Late Quaternary clay mineral record in Central Lake Baikal (Academician Ridge, Siberia).* Palaeogeography, Palaeoclimatology, Palaeoecology **193**, 159-179 (2003).

[Falconer 1990] Falconer, K.J.: *Fractal Geometry. Mathematical Foundations and Applications.* Wiley, Chichester (1990).

[Farmer 1971] Farmer, S.A.: *An Investigation into the Results of Principal Component Analysis of Data derived from Random Numbers.* The Statistician **20** (4), 63-72 (1971).

[Fieller et al. 1984] Fieller, N.R.J.; Gilbertson, D.D.; Olbricht, W.: *A new method for environmental analysis of particle size distribution data from shoreline sediments.* Nature **311**, 648-651 (1984).

[Fieller et al. 1990] Fieller, N.R.J.; Flenley, E.C.; Gilbertsin, D.D.; Thomas, D.S.G.: *Dumbbells: a plotting convention for "mixed" grain size populations.* Sedimentary Geology **69**, 7-12 (1990).

[Fieller et al. 1992] Fieller, N.R.J.; Flenley, E.C.; Olbricht, W.: *Statistics of Particle Size Data.* Applied Statistics, Journal of the Royal Statistical Society C **41** (1), 127-146 (1992).

[Fisher 1925] Fisher, R.A.: *Theory of statistical estimation.* Proceedings of the Cambridge Philosophical Society **22**, 700-725 (1925).

[Flower et al. 1997a] Flower, B.P.; Zachos, J.C.; Paul, H.: *Milankovitch-scale climate variability recorded near the Oligocene/Miocene boundary.* In: Shackleton, N.J.; Curry, W.B.; Richter, C.; Bralower, T.J. (eds.): Proceedings of the Ocean Drilling program, Scientific Results **154**, 433-439 (1997).

[Flower et al. 1997b] Flower, B.P.; Zachos, J.C.; Martin, E.: *Latest Oligocene thorugh early Miocene isotopic stratigraphy and deep-water paleoceanography of the Western Equatorial Atlantic: Sites 926 and 929.* In: Shackleton, N.J.; Curry, W.B.; Richter, C.; Bralower, T.J. (eds.): Proceedings of the Ocean Drilling program, Scientific Results **154**, 451-461 (1997).

[Folk and Ward 1957] Folk, R.L.; Ward, W.C.: *Brazes river bar: A study in the significance of grain size parameters.* Journal of Sedimentary Petrology **27**, 3-26 (1957).

[Foster 1996a] Foster, G.: *Time Series Analysis by Projection. I. Statistical Properties of Fourier Analysis.* The Astronomical Journal **111**(1), 541-554 (1996).

[Foster 1996b] Foster, G.: *Time Series Analysis by Projection. I. Tensor Methods for Time Series Analysis.* The Astronomical Journal **111**(1), 555-566 (1996).

[Foster 1996c] Foster, G.: *Wavelets for Period Analysis of Unevenly Sampled Time Series.* The Astronomical Journal **112**(4), 1709-1729 (1996).

[Francus 1998] Francus, P.: *An image-analysis technique to measure grain-size variation in thin sections of soft clastic sediments.* Sedimentary Geology **121**(3-4), 289-298 (1998).

[Francus and Karabanov 2000] Francus, P.; Karabanov, E.: *A computer-assisted thin-section study of lake Baikal sediments : a tool for understanding sedimentary processes and deciphering their climate signal.* International Journal of Earth Sciences **89**(2), 260-267 (2000).

[Fraser and Swinney 1986] Fraser, A.M.; Swinney, H.L.: *Independent coordinates for strange attractors from mutual information.* Physical Review A **33**(2), 1134-1140 (1986).

[Frick et al. 1997] Frick, P.; Baliunas, S.L.; Galyagin, D.; Sokoloff, D.; Soon, W.: *Wavelet Analysis of Stellar Chromospheric Activity Variations.* The Astrophysical Journal **483**, 426-434 (1997).

[Frick et al. 1998] Frick, P.; Grossmann, A.; Tchamitchian, P.: *Wavelet analysis of signals with gaps.* Journal of Mathematical Physics **39**(8), 4091-4107 (1998).

[Friedman 1958] Friedman, G.M.: *Determination of sieve-size distribution from thin-section data for sedimentary petrological studies.* Journal of Geology **66**, 394-416 (1958).

[Ganopolski and Rahmstorf 2002] Ganopolski, A.; Rahmstorf, S.: *Abrupt Glacial Climate Changes due to Stochastic Resonance.* Physical Review Letters **88**(3), 038501 (2002).

[Glushkov et al. 2005] Glushkov, A.V.; Khokhlov, V.N.; Loboda, N.S.; Rusov, V.D.; Vaschenko, V.N.: *Orbital forcings of the Earth's climate in wavelet domain.* Climate of the Past Discussions **1**, 193-214 (2005).

[Gorokhovski and Saveliev 2003] Gorokhovski, M.A.; Saveliev, V.E.: *Analyses of Kolmogorov's model of breakup and its application into Lagrangian computation of liquid sprays under air-blast conditions.* Physics of Fluids **15**(1), 184-192 (2003).

[Gorokhovski 2003] Gorokhovski, M.A.: *Fragmentation under the scaling symmetry and turbulent cascade with intermittency.* Stanford Center for Turbulence Research Annuals Briefs, 197-203 (2003).

[Granoszewski et al. 2005] Granoszewski, W.; Demske, D.; Nita, M.; Heumann, G.; Andreev, A.A.: *Vegetation and climate variability during the Kazantsevo (Eemian) Interglacial evidenced in a pollen record from Lake Baikal.* Global and Planetary Change **46**(1-4), 187-198 (2005).

[Grassberger and Procaccia 1983] Grassberger, P.; Procaccia, I.: *Characterization of Strange Attractors.* Physical Review Letters **50**(5), 346-349 (1983).

[Greenman 1951] Greenman, N.M.: *The mechanical analysis of sediments from thin-section data.* Journal of Geology **59**, 447-462 (1951).

[Griffiths et al. 1987] Griffiths, W.E.; Hill, R.C.; Pope, P.J.: *Small Sample Properties of Probit Model Estimators.* Journal of the American Statistical Association **82** (399), 929-937 (1987).

[Grootes et al. 1993] Grootes, P.M.; Stuiver, M.; White, J.W.C.; Johnsen, S.; Jouzel, J.: *Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores.* Nature **366**(6455), 552-554 (1993).

[Grootes and Stuiver 1997] Grootes, P.M.; Stuiver, M.: *Oxygen 18/16 variability in Greenland snow and ice with 10^{-3}- to 10^{5}-year time resolution.* Journal of Geophysical Research (Oceans and Atmospheres) **102**(C12), 26,455-26,470 (1997).

[Guyodo et al. 2000] Guyodo, Y.; Gaillot, P.; Channell, J.E.T.: *Wavelet analysis of relative geomagnetic paleointensity at ODP Site 983.* Earth and Planetary Science Letters **184**, 109-123 (2000).

[Hald 1952] Hald, A.: *Statistical Theory with Engineering Applications.* Wiley, New York (1952).

[Hall and Stewart 2005] Hall, P.; Stewart, M.: *Theoretical analysis of power in a two-component normal mixture model.* Journal of Statistical Planning and Inference **134**, 158-179 (2005).

[Hargreaves and Abe-Ouchi 2003] Hargreaves, J.C.; Abe-Ouchi, J.: *Timing of ice-age terminations determined by wavelet methods.* Paleoceanography **18**(2), 000825 (2003).

[Harrison et al. 2001] Harrison, S.P.; Kohfeld, K.E.; Roelandt, C.; Claquin, T.: *The role of dust in climate changes today, at the last glacial maximum and in the future.* Earth-Science Reviews **54**(1-3), 43-80 (2001).

[Hart et al. 1968] Hart, J.F.; Cheney, E.W.; Lawson, C.L.; Maehly, H.J.; Mesztenyi, C.K.; Rice, J.R.; Thacher Jr., H.G.; Witzgall, C.: *Computer Approximations.* Wiley, New York, 136-140 (1968).

[Hartley 1971] Hartley, H.O.; Hocking, R.R.: *The Analysis of Incomplete Data.* Biometrics **27**, 783-823 (1971).

[Hartmann 1988] Hartmann, D..: *The goodness-of-fit to ideal Gauss and Rosin distributions: a new grain-size parameter - Discussion.* Journal of Sedimentary Petrology **58**(5), 913-917 (1988).

[Hartmann and Christiansen 1992] Hartmann, D.; Christiansen, C.: *The hyperbolic shape triangle as a tool for discriminating populations of sediment samples of closely connencted origin.* Sedimentology **39**, 697-708 (1992).

[Hartmann and Bowman 1993] Hartmann, D.; Bowman, D.: *Efficiency of the Log-Hyperbolic Distribution - A Case Study: Pattern of Sediment Sorting in a Small Tidal-Inlet - Het Zwin, The Netherlands.* Journal of Coastal Research **9**(4), 1044-1053 (1993).

[Hartmann and Flemming 2002] Hartmann, D.; Flemming, B.: *Discussion of: A comparison between log-hyperbolic and model-independent grain size distributions in sediment trend analysis (STA (R)).* Journal of Coastal Research **18**(3), 592-595 (2002).

[Hasselblad 1966] Hasselblad, V.: *Estimation of Parameters for a Mixture of Normal Distributions.* Technometrics **8**(3), 431-444 (1966).

[Hasselblad 1969] Hasselblad, V.: *Estimation of Finite Mixtures of Distributions from the Exponential Family.* Journal of the American Statistical Association **64**, 1459-1471 (1969).

[Hasselblad et al. 1980] Hasselblad, V.; Stead, A.G.; Galke, W.: *Analysis of Coarsely Grouped Data From the Lognormal Distribution.* Technometrics **75**, 771-778 (1980).

[Heim et al. 2005] Heim, B.; Oberhänsli, H.; Fietz, S.; Kaufmann, H.: *Variation in Lake Baikal's phytoplankton distribution and fluvial input assessed by SeaWiFS satellite data.* Global and Planetary Change **46**(1-4), 9-27 (2005).

[Heslop et al. 2002] Heslop, D.; Dekkers, M.J.; Kruiver, P.P.; van Oorschot, I.H.M.: *Analysis of isothermal remanent magnetization acquisition curves using the expectation-maximization algorithm.* Geophysical Journal International **148**, 58-64 (2002).

[Heslop and Dekkers 2002] Heslop, D.; Dekkers, M.J.: *Spectral analysis of unevenly spaced climatic time series using CLEAN: signal recovery and derivation of significance levels using a Monte Carlo simulation.* Physics of the Earth and Planetary Interior **130**, 103-116 (2002).

[Higuchi 1988] Higuchi, T.: *Approach to an irregular time series on the basis of the fractal theory.* Physica D **31**, 277-283 (1988).

[Hill 1963] Hill, B.G.: *Information for estimating the proportions in mixtures of exponential and normal distributions.* Journal of the American Statistical Association **58**, 918-932 (1963).

[Hill 1973] Hill, I.D.: *The Normal Integral (Algorithm AS 66).* Applied Statistics, Journal of the Royal Statistical Society C **22** (3), 424-427 (1973).

[Hill and McLaren 2001] Hill, S.; McLaren, P.: *A comparison between log-hyperbolic and model-independent grain size distributions in sediment trend analysis (STA (R)).* Journal of Coastal Research **17**(4), 931-935 (2001).

[Hill and McLaren 2003] Hill, S.; McLaren, P.: *Response to: A comparison between log-hyperbolic and model-independent grain size distributions in sediment trend analysis (STA (R)).* Journal of Coastal Research **19**(1), 218-220 (2003).

[Hinnov et al. 2002] Hinnov, L.A.; Schulz, M.; Yiou, P.: *Interhemispheric space-time attributes of the Dansgaard-Oeschger oscillations between 100 and 0 ka.* Quaternary Science Reviews **21**, 1213-1228 (2002).

[Holschneider 1995] Holschneider, M.: *Wavelets: An Analysis Tool.* Oxford University Press, Oxford (1995).

[Huybers and Wunsch 2003] Huybers, P.; Wunsch, C.: *Rectification and precession signals in the climate system.* Geophysical Research Letters **30**(19), 017875 (2003).

[Huybers and Wunsch 2004] Huybers, P.; Wunsch, C.: *A depth-derived Pleistocene age model: Uncertainty estimates, sedimentation variability, and nonlinear climate change.* Paleoceanography **19**, 000857 (2004).

[Hyvärinen et al. 2001] Hyvärinen, A.; Karhunen, J.; Oja, E.: *Independent Component Analysis.* Wiley, New York (2001).

[Ibbeken 1983] Ibbeken, H.: *Jointed source rock and fluvial gravels controlled by Rosin's law: a grain size study in Calabria, South Italy.* Journal of Sedimentary Petrology **53**, 1213-1231 (1983).

[Inman 1952] Inman, D.L.: *Measures for describing the size distribution of sediments.* Journal of Sedimentary Petrology **22**, 125-145 (1952).

[Jamshidian and Jennrich 1993] Jamshidian, M.; Jennrich, R.I.: *Conjugate gradient acceleration of the EM algorithm.* Journal of the American Statistical Association **88**, 221-228 (1993).

[Jamshidian and Jennrich 1997] Jamshidian, M.; Jennrich, R.I.: *Acceleration of the EM Algorithm by using Quasi-Newton Methods.* Journal of the Royal Statistical Society B **59**, 569-587 (1997).

[Jamshidian and Jennrich 2000] Jamshidian, M.; Jennrich, R.I.: *Standard errors for EM estimation.* Journal of the Royal Statistical Society B **62**, 257-270 (2000).

[Joliffe 1986] Joliffe, I.T.: *Principal Component Analysis* Springer, New York (1986).

[Jones and McLachlan 1989] Jones, P.N.; McLachlan, G.J.: *Modelling Mass-Size Particle Data by Finite Mixtures.* Communications in Statistics - Theory and Methods **18**(7), 2629-2646 (1989).

[Jones and McLachlan 1990] Jones, P.N.; McLachlan, G.J.: *Maximum Likelihood Estimation from Grouped and Truncated Data with Finite Normal Mixture Models (Algorithm AS 254).* Applied Statistics, Journal of the Royal Statistical Society C **39** (2), 273-282 (1990).

[Jones 1991] Jones, P.N.: *On collagen fibril diameter distributions.* Connective Tissue Research **26**, 11-21 (1991).

[Jones and McLachlan 1991] Jones, P.N.; McLachlan, G.J.: *Fitting Mixture Distributions to Phenylthiocarbamide (PTC) Sensitivity.* American Journal of Human Genetics **48**, 117-120 (1991).

[Jones and McLachlan 1992] Jones, P.N.; McLachlan, G.J.: *Improving the Convergence Rate of the EM Algorithm for a Mixture Model Fitted to Grouped Truncated Data.* Journal of Statistical Computation and Simulation **43**, 31-44 (1992).

[Kaneko 1989] Kaneko, K.: *Spatiotemporal chaos in one-dimensional and two-dimensional coupled map lattices.* Physica D **37**(1-3), 60-82 (1989).

[Kantz and Schreiber 1997] Kantz, H.; Schreiber, T.: *Nonlinear Time Series Analysis.* Cambridge University Press, Cambridge (1997).

[Karabanov et al. 1998] Karabanov, E.B.; Prokopenko, A.A.; Williams, D.F.; Colman, S.M.: *Evidence from Lake Baikal for Siberian glaciation during oxygen-isotope substage 5d.* Quaternary Research **50**(1), 46-55 (1998).

[Karlis 2001] Karlis, D.: *A cautionary note about the EM algorithm for finite exponential mixtures.* Technical Report 150, Department of Statistics, Athens University of Economics and Business (2001).

[Karlis and Xekalaki 1999] Karlis, D.; Xekalaki, E.: *Improving the EM algorithm for mixtures.* Statistics and Computing **9**, 303-307 (1999).

[Karlis and Xekalaki 2003] Karlis, D.; Xekalaki, E.: *Choosing initial values for the EM algorithm for finite mixtures.* Computational Statistics & Data Analysis **41**, 577-590 (2003).

[Karner et al. 2002] Karner, D.B.; Levine, J.; Medeiros, B.P.; Muller, R.A.: *Constructing a stacked benthic $\delta^{18}O$ record.* Paleoceanography **17**, 000667 (2002).

[Kashiwaya et al. 1998] Kashiwaya, K.; Ryugo, M.; Sakai, H.; Kawai, T.: *Long-term climato-limnological oscillation during the past 2.5 million years printed in Lake Baikal sediments.* Geophysical Research Letters **25**, 659-663 (1998).

[Kashiwaya et al. 2001] Kashiwaya, K.; Ochiai, S.; Sakai, H.; Kawai, T.: *Orbit-related long-term climate cycles revealed in a 12-Myr continental record from Lake Baikal.* Nature **410**, 71-74 (2001).

[Kennett 1977] Kennett, J.P.: *Cenozoic Evolution of Antarctic Glaciation, the Circum-Antarctic Ocean, and Their Impact on Global Paleoceanography.* Journal of Geophysical Research **82**(27), 3843-3860 (1977).

[Khursevich et al. 2001] Khursevich, G.K.; Karabanov, E.B.; Prokopenko, A.A.; Williams, D.F.; Kuzmin, M.I.; Fedenya, S.A.; Gvozdkov, A.A.: *Insolation regime in Siberia as a major factor controlling diatom production in Lake Baikal during the past 800,000 years.* Quarternary International **80-81**, 47-58 (2001).

[Kiefer and Wolfowitz 1956] Kiefer, J.; Wolfowitz, J.: *Consistency of the maximum likelihood estimates in the presence of infinitely many incidental parameters.* Annals of Mathematical Statistics **27**, 887-906 (1956).

[Kittleman Jr. 1964] Kittleman Jr., L.R.: *Application of Rosin's distribution in size-frequency analysis of clastic rocks.* Journal of Sedimentary Petrology **34**, 483-502 (1964).

[Knight et al. 2002] Knight, J.; Orford, J.D.; Wilson, P.; Braley, S.M.: *Assessment of temporal changes in coastal sand dune environments using the log-hyperbolic grain-size method.* Sedimentology **49**(6), 1229-1252 (2002).

[Kohfeld and Harrison 2001] Kohfeld, K.E.; Harrison, S.P.: *DIRTMAP: the geological record of dust.* Earth-Science Reviews **54**(1-3), 81-114 (2001).

[Kojadinovic 2005] Kojadinovic, I.: *On the use of mutual information in data analysis: an overview.* Proceedings of Applied Statistical Models and Data Analysis 2005, VII.16 (2005).

[Kolmogorov 1941] Kolmogorov, A.N.: *On the log-normal distribution of particles sizes during break-up processes.* Doklady Akademii Nauk SSSR **31**(2), 99-101 (1941).

[Kondolf and Adhikari 2000] Kondolf, G.M.; Adhikari, A.: *Weibull vs. Lognormal Distributions for Fluvial Gravels.* Journal of Sedimentary Research **70**, 456-460 (2000).

[Kramer 1991] Kramer, M.A.: *Nonlinear Principal Component Analysis Using Autoassociative Neural Networks.* American Institute for Chemical Engineering Journal **37** (2), 233-243 (1991).

[Krishna Kumar et al. 1999] Krishna Kumar, K.; Rajagopalan, B.; Cane, M.A.: *On the Weakening Relationship Between the Indian Monsoon and ENSO.* Science **284**, 2156-2159 (1999).

[Krissek 2004] Krissek, L.A.: *Element abundances, loss on ignition, total analyzed abundances, CIA values, and Al-oxide/Ti-oxide ratios of core CRP-2A.* PANGAEA database, doi:10.1594/PANGAEA.144432 (2004).

[Krissek and Kyle 1998] Krissek, L.A.; Kyle, P.R.: *Geochemical indicators of weathering and Cenozoic palaeoclimates in sediments from CRP-1 and CIROS-1, McMurdo Sound, Antarctica.* Terra Antartica **5**(3), 673-680 (1998).

[Krissek and Kyle 2000] Krissek, L.A.; Kyle, P.R.: *Geochemical indicators of weathering, Cenozoic palaeoclimates, and provenance from fine-grained sediments in CRP-2/2A.* Terra Antartica **7**(4/5), 589-597 (2000).

[Kruiver et al. 2001] Kruiver, P.P.; Dekkers, M.J.; Heslop, D.: *Quantification of magnetic coercivity components by the analysis of acquisition curves of isothermal remanent magnetisation.* Earth and Planetary Science Letters **189**(3-4), 269-276 (2001).

[Krumbein 1934] Krumbein, W.C.: *Size Frequency Distributions of Sediments.* Journal of Sedimentary Petrology **4**(2), 65-77 (1934).

[Krumbein 1936] Krumbein, W.C.: *Application of logarithmic moments to size frequency distributions of sediments.* Journal of Sedimentary Petrology **6**(1), 35-47 (1936).

[Krumbein 1938] Krumbein, W.C.: *Size Frequency Distributions of Sediments and the Normal Phi Curve.* Journal of Sedimentary Petrology **8**(3), 84-90 (1938).

[Krumbein and Tisdel 1940] Krumbein, W.C.; Tisdel, F.W.: *Size Distribution of Source Rocks of Sediments.* American Journal of Science **238**(4), 296-305 (1940).

[Krumbein 1941] Krumbein, W.C.: *Measurement and geological significance of shape and roundness of sedimentary particles.* Journal of Sedimentary Petrology **11**, 64-72 (1941).

[Kwapien et al. 2000] Kwapień, J.; Drozdz, S.; Ioannides, A.A.: *Temporal correlations versus noise in the correlation matrix formalism: An example of the brain auditory response.* Physical Review E **62**, 5557-5564 (2000).

[Kwapien et al. 2003] Kwapień, J.; Drozdz, S.; Speth, J.: *Alternation of different fluctuation regimes in the stock market dynamics.* Physica A **330**(3-4), 605-621 (2003).

[Laloux et al. 1999] Laloux, L.; Cizeau, P.; Bouchaud, J.-P.; Potters, M.: *Noise Dressing of Financial Correlation Matrices.* Physical Review Letters **83**(7), 1467-1470 (1999).

[Lam and Ip 2003] Lam, K.F.; Ip, D.: *REML and ML estimation for clustered grouped survival data.* Statistics in Medicine **22**, 2025-2034 (2003).

[Lange 1995] Lange, K.: *A Quasi-Newton Acceleration of the EM Algorithm.* Statistica Sinica **5**, 1-18 (1995).

[Laskar et al. 1993] Laskar, J.; Joutel, F.; Boudin, F.: *Orbital, precessional, and insolation quantities for the Earth from -20 Myr to +10 Myr.* Astronomy & Astrophysics **270**, 522-533 (1993).

[Laskar et al. 2004] Laskar, J.; Robutel, P.; Joutel, F.; Gastineau, M.; Correia, A.C.M.; Levrard, B.: *A long term numerical solution for the insolation quantities of the Earth.* Astronomy & Astrophysics **428**, 261-285 (2004).

[Lawver and Gahagan 1998] Lawver, L.A.; Gahagan, L.M.: *Opening of Drake Passage and its impact on Cenozoic ocean circulation.* In: Crowley, T.J.; Burke, K.C. (eds.): *Tectonic Boundary Conditions for Climate Reconstructions.* Oxford University Press, Oxford, 212-223 (1998).

[Lawver and Gahagan 2003] Lawver, L.A.; Gahagan, L.M.: *Evolution of Cenozoic seaways in the Circum-Antarctic region; Antarctic Cenozoic palaeoenvironments; geologic records and models.* Palaeogeography Palaeoclimatology Palaeoecology **198**, 11-37 (2003).

[Lehmann 1975] Lehmann, E.L.: *Nonparametrics. Statistical Methods Based on Ranks.* Holden-Day, San Francisco (1975).

[Lin and Chao 1998] Lin, H.S.; Chao, B.F.: *Wavelet Spectral Analysis of the Earth's Orbital Variations and Paleoclimatic Cycles.* Journal of Atmospheric Sciences **55**, 227-236 (1998).

[Lindsey 1983] Lindsey, B.G.: *The Geometry of Mixture Likelihoods: A General Theory.* The Annals of Statistics **11**(1), 86-94 (1983).

[Lindstrom and Bates 1988] Lindstrom, M.J.; Bates, D.M.: *Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data.* Journal of the American Statistical Association **83**, 1014-1022 (1988).

[Lirer et al. 1996] Lirer, L.; Sheridan, M.; Vinci, A.: *Deconvolution of pyroclastic grain-size spectra for interpretation of transport mechanisms: an application to the AD 79 Vesuvio deposits.* Sedimentology **43**, 913-926 (1996).

[Lisiecki and Lisiecki 2002] Lisiecki, L.E.; Lisiecki, P.A.: *Application of dynamic programming to the correlation of paleoclimate records.* Paleoceanography **17**(4), 000733 (2002).

[Lisiecki and Raymo 2005] Lisiecki, L.E.; Raymo, M.E.: *A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records.* Paleoceanography **20**, 001071 (2005).

[Liu and Rubin 1994] Liu, C.; Rubin, D.B.: *The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence.* Biometrika **81**(4), 633-648 (1994).

[Liu and Shao 2004] Liu, X.; Shao, Y.: *Asymptotics for the likelihood ratio test in a two-component normal mixture model.* Journal of Statistical Planning and Inference **123**(1), 61-81 (2004).

[Louis 1982] Louis, T.A.: *Finding the Observed Information Matrix when Using the EM algorithm.* Journal of the Royal Statistical Society B **44** (2), 226-233 (1982).

[Lu et al. 2002] Lu, C.; Danzer, R.; Fischer, F.D.: *Fracture statistics of brittle materials: Weibull or normal distribution.* Physical Review E **65**, 067102 (2002).

[Maasch 1989] Maasch, K.A.: *Calculating climate attractor dimension from $\delta^{18}O$ records by the Grassberger-Procaccia algorithm.* Climate Dynamics **4**(1), 45-55 (1989).

[Mackay et al. 1997] Mackay, A.W.; Flower, R.J.; Kuzmina, A.E.; Granina, L.Z.; Rose, N.L.; Appelby, P.G.; Boyle, J.F.; Battarbee, R.W.: *Diatom succession trends in recent sediments from Lake Baikal and their relation to atmospheric pollution and to climate change.* Philosophical Transactions of the Royal Society London, Series B **353**, 1011-1055 (1997).

[Mallinson et al. 2003] Mallinson, D.J.; Flower, B.; Hine, A.; Brooks, G.; Molina Garza, R.: *Paleoclimate implications of high latitude precession-scale mineralogic fluctuations during early Oligocene Antarctic glaciation: the Great Australian Bight record.* Global and Planetary Change **39**, 257-269 (2003).

[Maraun et al. 2004] Maraun, D.; Rust, H.W.; Timmer, J.: *Tempting long-memory - on the interpretation of DFA results.* Nonlinear Processes in Geophysics **11**, 495-503 (2004).

[Maraun and Kurths 2005] Maraun, D.; Kurths, J.: *Epochs of phase coherence between El Nino/Southern Oscillation and Indian Monsoon.* Geophysical Research Letters **32**, L15709 (2005).

[Martinson et al. 1987] Martinson, D.G.; Pisias, N.G.; Hays, J.D.; Imbrie, J.; Moore Jr., T.C.; Shackelton, N.J.: *Age dating and the orbital theory of the ice ages: development of a high-resolution 0 to 300 000-year chronostratigraphy.* Quaternary Research **27**(1), 1-29 (1987).

[Marwan et al. 2002a] Marwan, N.; Thiel, M.; Nowaczyk, N.R.: *Cross Recurrence Plot Based Synchronization of Time Series.* Nonlinear Processes in Geophysics **9**, 325-331 (2002).

[Marwan et al. 2002b] Marwan, N.; Wessel, N.; Meyerfeldt, U.; Schirdewan, A.; Kurths, J.: *Recurrence Plot Based Measures of Complexity and its Application to Heart Rate Variability Data.* Physical Review E **66**(2), 026702 (2002).

[Maslov 2001] Maslov, S.: *Measures of globalization based on cross-correlations of world financial indices.* Physica A **301**(1-4), 397-406 (2001).

[Mayer-Kress and Kaneko 1989] Mayer-Kress, G.; Kaneko, K.: *Spatiotemporal Chaos and Noise.* Journal of Statistical Physics **54**(5-6), 1489-1508 (1989).

[Mayewski et al. 1997] Mayewski, P.A.; Meeker, L.D.; Twickler, M.S.; Whitlow, S.; Yang, Q.; Lyons, W.B.; Prentice, M.: *Major features and forcing of high latitude northern hemisphere atmospheric oscillation over the last 110,000 years.* Journal of Geophysical Research (Oceans and Atmospheres) **102**(C12), 26,345-26,366 (1997).

[McBride 1971] McBride, E.F.: *Mathematical Treatment of Size Distribution Data.* In: Carver, R.E. (ed.): *Procedures in sedimentary petrology.* Wiley-Interscience, New York, 109-127 (1971).

[McLachlan 1987] McLachlan, G.J.: *On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture.* Applied Statistics **36**(3), 318-324 (1987).

[McLachlan 1988] McLachlan, G.J.: *On the choice of starting values for the EM algorithm in fitting mixture models.* The Statistician **37**, 417-425 (1988).

[McLachlan and Basford 1988] McLachlan, G.J.; Basford, K.E.: *Mixture Models - Inference and Applications to Clustering.* Marcel Dekker, New York (1988).

[McLachlan and Jones 1988] McLachlan, G.J.; Jones, P.N.: *Fitting Mixture Models to Grouped and Truncated Data via the EM algorithm.* Biometrics **44**, 571-578 (1988).

[McLachlan et al. 1995] McLachlan, G.J.; McLaren, C.E.; Matthews, D.: *An algorithm for the likelihood ratio test of one versus two components in a normal mixture model fitted to grouped and truncated data.* Communications in Statistics - Simulation and Computation **24** (4), 965-985 (1995).

[McLachlan 1996] McLachlan, G.J.: *On Aitken's method and other approaches for accelerting convergence of the EM algorithm.* Proceedings of the A.C. Aitkin Centenary Conference, University of Otago, August 1995. University of Otago Press, Dunedin, 201-209 (1996).

[McLachlan and Krishnan 1997] McLachlan, G.J.; Krishnan, T.: *The EM Algorithm and Extensions.* Wiley, New York (1997).

[McLachlan and Peel 1997] McLachlan, G.J.; Peel, D.: *On a Resampling Approach to Choosing the Number of Components in Normal Mixture Models.* In: Billard, L.; Fisher, N.I. (Eds.): *Computing Science and Statistics* **28**, 260-266. Interface Foundation of North America, Fairfax Station, Virginia (1997).

[McLachlan and Peel 2000] McLachlan, G.J.; Peel, D.: *Finite Mixture Models.* Wiley, New York (2000).

[McLaren et al. 1986a] McLaren, C.E.; Brittenham, G.M.; Hasselblad, V.: *Analysis of the Volume of Red Blood Cells: Application of the Expectation-Maximization Algorithm to Grouped Data from the Doubly-Truncated Lognormal Distribution.* Biometrics **42**, 143-158 (1986).

[McLaren et al. 1986b] McLaren, C.E.; Brittenham, G.M.; Gordeuk, V.R.; Hughes, M.A.; Keating, L.J.: *Statistical modelling of the distribution of red blood cell volumes in iron deficiency anemia using the expectation-maximization algorithm.* The Statistician, Journal of the Royal Statistical Society D **35**, 135-142 (1986).

[McLaren et al. 1987] McLaren, C.E.; Brittenham, G.M.; Hasselblad, V.: *Statistical and graphical evaluation of erythrocyte volume distributions.* American Journal of Physiology - Heart and Circulatory Physiology **252**(4), 857-866 (1987).

[McLaren et al. 1991] McLaren, C.E.; Wagstaff, M.; Brittenham, G.M.; Jacobs, A.: *Detection of Two-Component Mixtures of Lognormal Distributions in Grouped, Doubly-Truncated Data: Analysis of Red Blood Cell Volume Distributions.* Biometrics **47**, 607-622 (1991).

[McLaren et al. 1993] McLaren, C.E.; Houwen, B.; Koepke, J.A.; Rowan, R.M.; McKay, P.J.; Ortner, B.R.; Bishop, M.L.: *Analysis of red blood cell volume distributions using the ICSH reference method: detection of sequential changes in distributions by hydrodynamic focusing.* Clinical and Laboratory Haematology **15**, 173-184 (1993).

[McLaren 1996] McLaren, C.E.: *Mixture models in haematology: a series of case studies.* Statistical Methods in Medical Research **5**, 129-153 (1996).

[McLaren et al. 1998] McLaren, C.E.; McLachlan, G.J.; Halliday, J.W.; Webb, S.I.; Leggett, B.A.; Jazwinska, E.C.; Crawford, D.H.; Gordeuk, V.R.; McLaren, G.D.; Powell, L.W.: *Distributions of Transferrin Saturation in an Australian Population: Relevance to the Early Diagnosis of Hemochromatosis.* Gastroenterology **114**, 543-549 (1998).

[McLaren et al. 2000] McLaren, C.E.; Kambour, E.L.; McLachlan, G.J.; Lukaski, H.C.; Li, X.; Brittenham, G.M.; McLaren, G.D.: *Patient-specific analysis of sequential haematological data by multiple regression and mixture distribution modelling.* Statistics in Medicine **19**, 83-98 (2000).

[McLaren et al. 2001a] McLaren. C.E.; Cadez, I.V.; Smyth, P.; McLachlan, G.J.: *Classification of disorders of anemia on the basis of mixture model parameters.* Technical Report No. 01-56, Information and Computer Science Department, University of California, Irvine (2001).

[McLaren et al. 2001b] McLaren, C.E.; Li, K.-T.; Gordeuk, V.R.; Hasselblad, V.; McLaren, G.D.: *Relationship between transferrin saturation and iron stores in the African American and US Caucasian populations: analysis of data from the third National Health and Nutrition Examination Survey.* Blood **98**(8), 2345-2351 (2001).

[Mehta 1990] Mehta, M.L.: *Random Matrices.* Academic Press, New York (1990).

[Meilijson 1989] Meilijson, I.: *A Fast Improvement to the EM Algorithm on its Own Terms.* Journal of the Royal Statistical Society B **51**(1), 127-138 (1989).

[Meixner et al. 2000] Meixner, M.; Zoldi, S.M.; Bose, S.; Schöll, E.: *Karhunen-Loève local characterization of spatiotemporal chaos in a reaction-diffusion system.* Physical Review E **61** (2), 1382-1385 (2000).

[Mélice et al. 2001] Mélice, J.L.; Coron, A.; Berger, A.: *Amplitude and Frequency Modulations of the Earth's Obliquity for the Last Million Years.* Journal of Climate **14**, 1043-1054 (2001).

[Meng and Rubin 1991] Meng, X.-L.; Rubin, D.B.: *Using EM to Obtain Asymptotic Variance-Covariance Matrices: The SEM Algorithm.* Journal of the American Statistical Association **86** (416), 899-909 (1991).

[Miloslavsky and van der Laan 2003] Miloslavsky, M.; van der Laan, M.J.: *Fitting of mixtures with unspecified number of components using cross validation distance estimate.* Computational Statistics & Data Analysis **41**, 413-428 (2003).

[Moberg et al. 2005] Moberg, A.; Sonechkin, D.M.; Holmgren, K.; Datsenko, N.M.; Karl, W.: *Highly variable Northern Hemisphere temperatures reconstructed from low-and high-resolution proxy data.* Nature **433**(7026), 613-617 (2005).

[Mudelsee and Stattegger 1994a] Mudelsee, M.; Stattegger, K.: *Application of the Grassberger-Procaccia Algorithm to the $\delta^{18}O$ Record from ODP Site 659: Selected Methodological Aspects.* In: Kruhl, J.H. (ed.): *Fractals and Dynamic Systems in Geoscience.* Springer, Berlin, 399-413 (1994).

[Mudelsee and Stattegger 1994b] Mudelsee, M.; Stattegger, K.: *Plio-/Pleistocene Climate Modeling Based on Oxygen Isotope Time Series from Deep-Sea Sediment Cores: The Grassberger-Procaccia Algorithm and Chaotic Climate Systems.* Mathematical Geology **26**(7), 799-815 (1994).

[Mudelsee 1995] Mudelsee, M.: *Entwicklung neuer statistischer Analysemethoden für Zeitreihen mariner, stabiler Isotopen: die Evolution des globalen Plio-/Pleistozänen Klimas.* PhD Thesis, University of Kiel (1995).

[Mudelsee and Schulz 1997] Mudelsee, M.; Schulz, M.: *The Mid-Pleistocene climate transition: onset of 100 ka cycle lags ice volume build-up by 280 ka.* Earth and Planetary Science Letters **151**, 117-123 (1997).

[Mudelsee and Stattegger 1997] Mudelsee, M.; Stattegger, K.: *Exploring the structure of the mid-Pleistocene revolution with advanced methods of time-series analysis.* Geologische Rundschau **86**, 499-511 (1997).

[Mudelsee 2002] Mudelsee, M.: *TAUEST: a computer program for estimating persistence in unevenly spaced weather/climate time series.* Computers & Geosciences **28**, 69-72 (2002).

[Müller et al. 2005] Müller, M.; Baier, G.; Galka, A.; Stephani, U.; Muhle, H.: *Detection and characterization of changes of the correlation structure in multivariate time series.* Physical Review E **71**, 046116 (2005).

[Naish et al. 2001] Naish, T.R.; Woolfe, K.J.; Barrett, P.J.; Wilson, G.S.; Atkins, C.; Bohaty, S.M.; Bücker, C.J.; Claps, M.; Davey, F.J.; Dunbar, G.B.; Dunn, A.G.; Fielding, C.R.; Florindo, F.; Hannah, M.J.; Harwood, D.M.; Henrys, S.A.; Krissek, L.A.; Lavelle, M.; van der Meer, J.; McIntosh, W.C.; Niessen, F.; Passchier, S.; Powell, R.D.; Roberts, A.P.; Sagnotti, L.; Scherer, R.P.; Strong, C.P.; Talarico, F.; Verosub, K.L.; Villa, G.; Watkins, D.K.; Webb, P.N.; Wonik, T.: *Orbitally induced oscillations in the East Antarctic ice sheet at the Oligocene/Miocene boundary.* Nature **413**, 719-723 (2001).

[NGRIP members 2004] North Greenland Ice Core Project members: *High resolution Climate Record of the Northern Hemisphere back into the last Interglacial Period.* Nature **431**, 147-151 (2004).

[Nityasuddhi and Böhning 2003] Nidyasuddhi, D.; Böhning, D.: *Asymptotic properties of the EM algorithm estimate for normal mixture models with component specific variances.* Computational Statistics and Data Analysis **41**, 591-601 (2003).

[Nugteren et al. 2004] Nugteren, G.; Vandenberghe, J.; van Huissteden, J.K.; Zhisheng, A.: *A Quaternary climate record based on grain size analysis from the Luochuan loess section on the Central Loess Plateau, China.* Global and Planetary Change **41**(3-4), 167-183 (2004).

[Oakes 1999] Oakes, D.: *Direct calculation of the information matrix via the EM algorithm.* Journal of the Royal Statistical Society B **61**, 479-482 (1999).

[Oberhänsli and Mackay 2005] Oberhänsli, H.; Mackay, A.W. (eds.): *Progress towards reconstructing past climate in Central Eurasia, with special emphasis on Lake Baikal.* Global and Planetary Change **46**, 1-383 (2005).

[Ochiai and Kashiwaya 2003] Ochiai, S.; Kashiwaya, K.: *A conceptual model of sedimentation processes for a hydrogeomorphological study in Lake Baikal.* In: Kashiwaya, K. (ed.): *Long Continental Records from Lake Baikal.* Springer, Tokyo, 297-312 (2003).

[Olbrich et al. 1998] Olbrich, E.; Hegger, R.; Kantz, H.: *Analysing local observations of weakly coupled maps.* Physics Letters A **244**(6), 538-544 (1998).

[Olivarez Lyle and Lyle 2002] Olivarez Lyle, A.; Lyle, M.W.: *Determination of biogenic opal in pelagic marine sediments: a simple method revisited.* Proceedings of Ocean Drilling Program, Initial Reports **199**, 1-21 (2002).

[Orchard and Woodbury 1972] Orchard, T.; Woodbury, M.A.: *A Missing Information Principle: Theory and Applications.* In: Le Cam, L.M.; Neyman, J. (eds.): Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1 (Theory of Statistics), Berkeley, California: University of California Press, 697-715 (1972).

[Osipov et al. 2003] Osipov, G.V.; Hu, B.; Zhou, C.S.; Ivanchenko, M.V.; Kurths, J.: *Three Types of Transitions to Phase Synchronization in Coupled Chaotic Oscillators.* Physical Review Letters **91**, 024101 (2003).

[Otazu et al. 2002] Otazu, X.; Ribó, M.; Peracaula, M.; Paredes, J.M.; Núñez, J.: *Detection of superimposed periodic signals using wavelets.* Monthly Notices of the Royal Astronomical Society **333**, 365-372 (2002).

[Otazu et al. 2004] Otazu, X.; Ribó, M.; Paredes, J.M.; Peracaula, M.; Núñez, J.: *Multiresolution approach for period determination on unevenly sampled data.* Monthly Notices of the Royal Astronomical Society **351**(1), 215-219 (2004).

[Paillard 2001] Paillard, D.: *Glacial Cycles: Toward a New Paradigm.* Reviews of Geophysics **39**(3), 325-346 (2001).

[Pälike and Shackleton 2000] Pälike, H.; Shackleton, N.J.: *Constraints on astronomical parameters from the geological record for the last 25 Myr.* Earth and Planetary Science Letters **182**(1), 1-14 (2000).

[Paluš et al. 1993] Paluš, M.; Albrecht, V.; Dvorak, I.: *Information theoretic test for nonlinearity in time-series.* Physics Letters A **175**(3-4), 203-209 (1993).

[Paluš and Novotna 1994] Paluš, M.; Novotna, D.: *Testing for Nonlinearity in Weather Records.* Physics Letters A **193**(1), 67-74 (1994).

[Paluš 1995] Paluš, M.: *Testing for nonlinearity using redundancies - quantitative and qualitative aspects.* Physica D **80**(1-2), 186-205 (1995).

[Paluš 1996] Paluš, M.: *Detecting nonlinearity in multivariate time series.* Physics Letters A **213**, 138-147 (1996).

[Paluš 1997] Paluš, M.: *Detecting phase synchronization in noisy systems.* Physics Letters A **235**, 341-351 (1997).

[Paluš and Novotna 1999] Paluš, M.; Novotna, D.: *Sunspot cycle: A driven nonlinear oscillator?* Physical Review Letters **83**(17), 3406-3409 (1999).

[Passe 1997] Passe, T.: *Grain size distribution expressed as* tanh*-functions.* Sedimentology **44**, 1011-1014 (1997).

[Paul et al. 2000] Paul, H.; Zachos, J.C.; Flower, B.P.; Tripati, A.: *Orbitally induced climate and geochemical variability across the Oligocene/Miocene boundary.* Paleoceanography **15**(5), 471-485 (2000).

[Pearson 1914] Pearson, K.: *A study of Trypanosome Strains.* Biometrika **10**, 85-143 (1914).

[Peck et al. 1994] Peck, J.A.; King, J.W.; Colman, S.M.; Kravchinsky, V.A.: *A rock-magnetic record from Lake Baikal, Siberia: Evidence for Late Quarternary climate change.* Earth and Planetary Science Letters **122**, 221-238 (1994).

[Peng et al. 1994] Peng, C.-K.; Buldyrev, S.V.; Havlin, S.; Simons, M.; Stanley, H.E.; Goldberger, A.L.: *Mosaic organization of DNA sequences.* Physical Review E **49**(2), 1685-1689 (1994).

[Petit et al. 1999] Petit, J.R.; Jouzel, J.; Raynaud, D.; Barkov, N.I.; Barnola, J.-M.; Basile, I.; Bender, M.; Chappellaz, J.; Davis, M.; Dalaygue, G.; Delmotte, M.; Kotlyakov, V.M.; Legrand, M.; Lipenkov, V.Y.; Lorius, C.; Pépin, L.; Ritz, C.; Saltzman, E.; Stievenard, M.: *Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica.* Nature **399**, 429-436 (1999).

[Pilla and Lindsay 2001] Pilla, R.S.; Lindsay, B.G.: *Alternative EM methods for nonparametric finite mixture models.* Biometrika **88**(2), 535-550 (2001).

[Pfuhl and McCave 2003] Pfuhl, H.A.; McCavem N.: *Integrated age models for the early Oligocene - early Miocene, Sites 1168 and 1170-1172.* In: Exon, N.F.; Kennett, J.P.; Malone, M.J.: Proceedings of the Ocean Drilling Program, Scientific Results **189**, 1-21 (2003).

[Pikovsky et al. 2001] Pikovsky, A.; Rosenblum, M.; Kurths, J.: *Synchronization - A Universal Concept in Nonlinear Sciences.* Cambridge University Press, Cambridge (2001).

[Piotrowska et al. 2004] Piotrowska, N.; Bluszcz, A.; Demske, D.; Granoszewski, W.; Heumann, G.: *Extraction and AMS radiocarbon dating of pollen from Lake Baikal sediments.* Radiocarbon **46**(1), 181-188 (2004).

[Plaut and Vautard 1994] Plaut, G.; Vautard, R.: *Spells of Low-Frequency Oscillations and Weather Regimes in the Northern Hemisphere.* Journal of the Atmospheric Sciences **51**(2), 210-236 (1994).

[Plerou et al. 1999] Plerou, V.; Gopikrishnan, P.; Rosenow, B.; Nunes Amaral, L.A.; Stanley, H.E.: *Universal and Nonuniversal properties of Cross Correlations in Financial Time Series.* Physical Review Letters **83**, 1471-1474 (1999).

[Plerou et al. 2000] Plerou, V.; Gopikrishnan, P.; Rosenow, B.; Nunes Amaral, L.A.; Guhr, T.; Stanley, H.E.: *A Random Matrix Approach to Cross-Correlations in Financial Data.* Physica A **287**(3-4), 374-382 (2000).

[Plerou et al. 2002] Plerou, V.; Gopikrishnan, P.; Rosenow, B.; Nunes Amaral, L.A.; Guhr, T.; Stanley, H.E.: *Random matrix approach to cross correlations in financial data.* Physical Review E **65**, 066126 (2002).

[Politi and Witt 1999] Politi, A.; Witt, A.: *Fractal Dimension of Space-Time Chaos.* Physical Review Letters **82**(15), 3034-3037 (1999).

[Pomeau 1985] Pomeau, Y.: *Measurement of the Information Density in Turbulence.* Comptes Rendus de l'Academie des Sciences Serie II **300**(7), 239-241 (1985).

[Pompe 1993] Pompe, B.: *Measuring Statistical Dependences in a Time Series.* Journal of Statistical Physics **73**, 587-610 (1993).

[Poppe et al. 2004] Poppe, L.J.; Eliason, A.H.; Hastings, M.E.: *A Visual Basic Program to Generate Sediment Grain-Size Statistics and to Extrapolate Particle Distributions.* Computers & Geosciences **30**(7), 791-795 (2004).

[Portier 2001] Portier, K.M.: *Statistical Issues in Assessing Anthropogenic Background for Arsenic.* Environmental Forensics **2**, 155-160 (2002).

[Potter et al. 2004] Potter, D.K.; Corbett, P.W.M.; Barcklay, S.A.; Haszeldine, R.S.: *Quantification of Illite Content in Sedimentary Rocks Using Magnetic Susceptibility - A Rapid Complement or Alternative to X-Ray Diffraction.* Journal of Sedimentary Research **74**(5), 730-735 (2004).

[Preisendorfer 1988] Preisendorfer, R.W.: *Principal Component Analysis in Meteorology and Oceanography.* Elsevier, Amsterdam (1988).

[Press 1999] Press, W.H.: *Numerical Recipes in C: The Art of Scientific Computing.* 2nd edition. Cambridge University Press, Cambridge (1999).

[Prichard and Theiler 1994] Prichard, D.; Theiler, J.: *Generating surrogate data for time series with several simultaneously measured variables.* Physical Review Letters **73**, 951-954 (1994).

[Priestley 1981] Priestley, M.B.: *Spectral Analysis and Time Series.* Academic Press, London (1981).

[Priestley 1988] Priestley, M.B.: *Non-Linear and Non-Stationary Time Series Analysis.* Academic Press, London (1988).

[Prins and Weltje 1999] Prins, M.A.; Weltje, G.J.: *End-member modeling of siliclastic grain-size distributions: The late Quaternary record of eolian and fluvial sediment supply to the Arabian Sea and its paleoclimatic significance.* In: *Numerical Experiments in Stratigraphy: Recent Advances in Stratigraphic and Sedimentologic Computer Simulations.* SEPM Special Publications **63** (1999).

[Probert-Jones 1973] Probert-Jones, J.R.: *Orthogonal Pattern (Eigenvector) Analysis of Random and Partly Random Fields.* Conference on Probability and Statistics in Atmospheric Sciences **3**, 187-192 (1973).

[Prospero et al. 2002] Prospero, J.M.; Ginoux, P.; Torres, O.; Nicholson, S.E.; Gill, T.E.: *Environmental characterization of global sources of atmospheric soil dust identified with the Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) absorbing aerosol product.* Reviews of Geophysics **40**(1), 1002 (2002).

[Protassov 2004] Protassov, R.S.: *EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ.* Statistics and Computing **14**, 67-77 (2004).

[Quian Quiroga et al. 2002] Quian Quiroga, R.; Kreuz, T.; Grassberger, P.: *Event synchronization: A simple and fast method to measure synchronicity and time delay patterns.* Physical Review E **66**, 041904 (2002).

[Raab and Kurths 2001] Raab, C.; Kurths, J.: *Estimation of large-scale dimension densities.* Physical Review E **64**, 016216 (2001).

[Raab et al. 2005] Raab, C.; Wessel, N.; Schirdewan, A.; Kurths, J.: *Large-Scale Dimension Densities for Heart rate Variability Analysis.* Computers in Cardiology **32**, 985-988 (2005).

[Rahmstorf 2003] Rahmstorf, S.: *Timing of abrupt climate change: A precise clock.* Geophysical Research Letters **30**(10), 017115 (2003).

[Rajaram and Erbach 1999] Rajaram, G.; Erbach, D.C.: *Effect of wetting and drying on soil physical properties.* Journal of Terramechanics **36**(1), 39-49 (1999).

[Raymo 1997] Raymo, M.E.: *The timing of major climate transitions.* Paleoceanography **12**(4), 577-585 (1997).

[Redner and Walker 1984] Redner, R.A.; Walker, H.F.: *Maxture Densities, Maximum Likelihood and the EM Algorithm.* SIAM Review **26** (2), 195-239 (1984).

[Renner 1991] Renner, R.M.: *An Examination of the Use of the Logratio Transformation for the Testing of Endmember Hypotheses.* Mathematical Geology **23**(4), 549-563 (1991).

[Rial and Anaclerio 2000] Rial, J.A.; Anaclerio, C.A.: *Understanding nonlinear responses of the climate system to orbital forcing.* Quaternary Science Reviews **19**, 1709-1722 (2000).

[Rieke et al. 2002] Rieke, C.; Sternickel, K.; Andrzejak, R.G.; Elger, C.E.; David, P.; Lehnertz, K.: *Measuring Nonstationarity by Analyzing the Loss of Recurrence in Dynamical Systems.* Physical Review Letters **88**(24), 244102 (2002).

[Rieke et al. 2004] Rieke, C.; Andrzejak, R.G.; Mormann, F.; Lehnertz, K.: *Improved statistical test for nonstationarity using recurrence time statistic.* Physical Review E **69**, 046111 (2004).

[Rioual and Mackay 2005] Rioual, P.; Mackay, A.W.: *A diatom record of centennial resolution for the Kazantsevo Interglacial stage in Lake Baikal (Siberia).* Global and Planetary Change **46**(1-4), 199-219 (2005).

[Roberts et al. 2003] Roberts, A.P.; Wilson, G.S.; Harwood, D.M.; Verosub, K.L.: *Glaciation across the Oligocene-Miocene boundary in southern McMurdo Sound, Antarctica: new chronology from the CIROS-1 drill hole.* Palaeogeography, Palaeoclimatology, Palaeoecology **198**, 113-130 (2003).

[Robertson and France 1994] Robertson, D.J.; France, D.E.: *Discrimination of remanence-carrying minerals in mixtures, using isothermal remanent magnetisation acquisition curves.* Physics of the Earth and Planetary Interiors **82**, 223-234 (1994).

[Romano et al. 2004] Romano, M.C.; Thiel, M.; von Bloh, W.: *Multivariate Recurrence Plots.* Physics Letters A **330**(3-4), 214-223 (2004).

[Romano et al. 2005] Romano, M.C.; Thiel, M.; Kurths, J.; Kiss, I.Z.; Hudson, J.L.: *Detection of synchronization for non-phase-coherent and non-stationary data.* Europhysics Letters **71**(3), 466-472 (2005).

[Rosenblum et al. 1996] Rosenblum, M.G.; Pikovsky, A.S.; Kurths, J.: *Phase Synchronization of Chaotic Oscillators.* Physical Review Letters **76**, 1804-1807 (1996).

[Rosin et al. 1933] Rosin, P.; Rammler, E.; Sperling, K.: *Korngrössenprobleme des Kohlenstaubes und ihre Bedeutung für die Vermahlung.* Berichte der Technisch-Wirtschaftlichen Sachverständigenausschüsse des Reichskohlenrates **C52**. Reichskohlenrat, Berlin (1933).

[Rosin and Rammler 1933] Rosin, P.R.; Rammler, E.: *The laws governing the fineness of powdered coal.* Journal of the Institute of Fuel **7**, 29-36 (1933).

[Rosin and Rammler 1934] Rosin, P.R.; Rammler, E.: *Die Kornzusammensetzung des Mahlgutes im Lichte der Wahrscheinlichkeitslehre.* Kolloid-Zeitschrift **67**, 16-26 (1934).

[Ross 2002] Ross, S.M.: *Simulation.* 3rd edition, Academic Press, San Diego (2002).

[Roweis and Saul 2000] Roweis, S.; Saul, L.: *Nonlinear dimensionality reduction by locally linear embedding.* Science **290** (5500), 2323-2326 (2000).

[Ruck 2001] Ruck, A.: *Calculating the (Asymptotic) Distribution of the Log-LRT Statistic in a Contamination Mixture Model.* Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 93, Universität der Bundeswehr Hamburg (2001).

[Ruck 2002] Ruck, A.: *Calculating the Asymptotic Distribution of the Log-LRT Statistic for Testing One Against two Populations in Normal Mixtures.* Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 98, Universität der Bundeswehr Hamburg (2002).

[Ruddiman and Raymo 2003] Ruddiman, W.F.; Raymo, M.E.: *A methane-based time scale for Vostok ice.* Quaternary Science Reviews **22**, 141-155 (2003).

[Rybski et al. 2003] Rybski, D.; Havlin, S.; Bunde, A.: *Phase synchronization in temperature and precipitation records.* Physica A **320**, 601-610 (2003).

[Sahay and Sreenivasan 1996] Sahay, A.; Sreenivasan, K.R.: *The search for a low-dimensional characterization of a local climate system.* Philosophical Transactions of the Royal Society London Series A **354**, 1715-1750 (1996).

[Samé and Govaert 2002] Samé, A.B.; Govaert, G.: *Classification de données discrétisées* (in French). In: Proceedings of XXXIVèmes Journées de Statistique, Bruxelles / Louvain-la-Neuve, Belgium, http://www.stat.ucl.ac.be/jsbl2002/articles.html (2002).

[Samé et al. 2003] Samé, A.; Ambroise, C.; Govaert, G.: *A mixture model approach for binned data clustering.* In: Berthold, M.R.; Lenz, H.-J.; Bradley, E.; Kruse, R.; Borgelt, C. (eds.): Advances in Intelligent Data Analysis V. Proceedings of the 5th International Symposium on Intelligent Data Analysis (IDA 2003). Lecture Notes in Computer Science (LNCS) **2810**. Springer, Berlin, 265-274 (2003).

[Sansom and Thomson 1998] Sansom, J.; Thomson, P.J.: *Detecting Components in Censored and Truncated Meteorological Data.* Environmetrics **9**, 673-688 (1998).

[Santhanam and Patra 2001] Santhanam, M.S.; Patra, P.K.: *Statistics of atmospheric correlations.* Physical Review E **64**, 016102 (2001).

[Sarkar et al. 2004] Sarkar, S.; Singh, R.P.; Kafatos, M.: *Further evidences for the weakening relationship of Indian rainfall and ENSO over India.* Geophysical Research Letters **31**, L13209 (2004).

[Scalon et al. 2003] Scalon, J.; Fieller, N.R.J.; Stillman, E.C.; Atkinson, H.V.: *A model-based analysis of particle size distributions in composite materials.* Acta Materialia **51**, 997-1006 (2003).

[Scargle 1997] Scargle, J.D.: *Wavelet methods in astronomical time series analysis.* In: Subba Rao, T.; Priestley, M.B.; Lessi, O. (eds.): *Applications of Time Series Analysis in Astronomy and Meteorology.* Chapman & Hall, London, 226-248 (1997).

[Schader and Schmid 1985] Schader, M.; Schmid, F.: *Computation of ML estimates for the parameters of a negative binomial distribution from grouped data. A comparison of the scoring, Newton-Raphson and EM algorithms.* Applied Stochastic Models and Data Analysis **1**, 11-23 (1985).

[Scher and Martin 2004] Scher, H.D.; Martin, E.E.: *Circulation in the Southern Ocean during the Paleogene inferred from neodymium isotopes.* Earth and Planetary Science Letters **228**, 391-405 (2004).

[Schlattmann 2005] Schlattmann, P.: *On bootstrapping the number of components in finite mixtures of Poisson distributions.* Statistics and Computing **15**, 179-188 (2005).

[Schleyer 1987] Schleyer, R.: *The goodness-of-fit to ideal Gauss and Rosin distributions: a new grain-size parameter.* Journal of Sedimentary Petrology **57**, 871-880 (1987).

[Schleyer 1988] Schleyer, R.: *The goodness-of-fit to ideal Gauss and Rosin distributions: a new grain-size parameter - Reply.* Journal of Sedimentary Petrology **58**(5), 917-918 (1988).

[Schreiber and Schmitz 1996] Schreiber, T.; Schmitz, A.: *Improved surrogate data for nonlinearity tests.* Physical Review Letters **77**, 635-638 (1996).

[Schreiber and Schmitz 2000] Schreiber, T.; Schmitz, A.: *Surrogate time series.* Physica D **142**, 346-382 (2000).

[Schulz et al. 1994] Schulz, M.; Mudelsee, M.; Wolf-Welling, T.C.W.: *Fractal Analyses of Pleistocene Marine Oxygen Isotope Records.* In: Kruhl, J.H. (ed.): *Fractals and Dynamic Systems in Geoscience.* Springer, Berlin, 377-387 (1994).

[Schulz and Stattegger 1997] Schulz, M.; Stattegger, K.: *SPECTRUM: Spectral Analysis of Unevenly Spaced Paleoclimatic Time Series.* Computers & Geosciences **23**(9), 929-945 (1997).

[Schulz et al. 1999] Schulz, M.; Berger, W.H.; Sarnthein, M.; Grootes, P.M.: *Amplitude variations of 1470-year climate oscillations during the last 100,000 years linked to fluctuations of continental ice mass.* Geophysical Research Letters **26**(22), 3385-3388 (1999).

[Schulz 2002a] Schulz, M.: *On the 1470-year pacing of Dansgaard-Oeschger warm events.* Paleoceanography **17**(2), 000571 (2002).

[Schulz 2002b] Schulz, M.: *The tempo of climate change during Dansgaard-Oeschger interstadials and its potential to affect the manifestation of the 1470-year climate cycle.* Geophysical Research Letters **29**(1), 013277 (2002).

[Schulz and Mudelsee 2002] Schulz, M.; Mudelsee, M.: *REDFIT: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series.* Computers & Geosciences **28**, 421-426 (2002).

[Schumann 2004] Schumann, A.Y.: *Waveletanalyse von Sedimentdaten unter Einbeziehung von Alters-Tiefen-Modellen.* Diploma Thesis, University of Potsdam (2004).

[Seba 2003] Seba, P.: *Random Matrix Analysis of Human EEG Data.* Physical Review Letters **91**, 198104 (2003).

[Seidel et al. 1997] Seidel, W.; Mosler, K.; Alker, M.; Ruck, A.: *Size and Power of Likelihood Ratio Tests in Exponential Mixture Models Based on Different Implementations of the EM Algorithm.* Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 79, Universität der Bundeswehr Hamburg (1997).

[Seidel et al. 2000a] Seidel, W.; Mosler, K.; Alker, M.: *A cautionary note on likelihood ratio tests in mixture models.* Annals of the Institute of Statistical Mathematics **52**(3), 481-487 (2000).

[Seidel et al. 2000b] Seidel, W.; Sevcikova, H.; Alker, M.: *On the Power of Different Versions of the Likelihood Ratio test for Homogeneity in an Exponential Mixture Model.* Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 92, Universität der Bundeswehr Hamburg (2000).

[Seidel and Sevcikova 2002a] Seidel, W.; Sevcikova, H.: *On EM Versions with Gradient Function Update for Finite Mixtures.* Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 103, Universität der Bundeswehr Hamburg (2002).

[Seidel and Sevcikova 2002b] Seidel, W.; Sevcikova, H.: *Tools for Analyzing and Maximizing Likelihood Functions in Mixture Models.* Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 104, Universität der Bundeswehr Hamburg (2002).

[Seidel et al. 2003] Seidel, W.; Sevcikova, H.; Ali, S.M.Y.: *On Resampling Approaches for Identifying Statistically Meaningful Maxima of Likelihood Functions in Mixture Models.* Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 105, Universität der Bundeswehr Hamburg (2003).

[Seidel and Sevcikova 2003a] Seidel, W.; Sevcikova, H.: *A Detailed Investigation of Likelihood Maxima in Two-Component Exponential Mixture Models and their Implication on LR Tests.* Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 106, Universität der Bundeswehr Hamburg (2003).

[Seidel and Sevcikova 2003b] Seidel, W.; Sevcikova, H.: *An Analysis of Tests Against Nonparametric Alternatives in Exponential Mixture Models.* Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 109, Universität der Bundeswehr Hamburg (2003).

[Seidel and Sevcikova 2004] Seidel, W.; Sevcikova, H.: *Types of likelihood maxima in mixture models and their implication on the performance of tests.* Annals of the Institute of Statistical Mathematics **56**(4), 631-654 (2004).

[Shackleton 1995] Shackleton, N.: *New Data on the evolution of Pliocene climatic variability.* In: Vrba, E.S.; Denton, G.H.; Partridge, T.C.; Burckle, L.H. (eds): *Paleoclimate and evolution with emphasis on human origins.* Yale University Press, Yale, 242-248 (1995).

[Shackleton et al. 1999] Shackleton, N.J.; Crowhurst, S.J.; Weedon, G.P.; Laskar, J.: *Astronomical calibration of Oligocene-Miocene time.* Philosophical Transactions of the Royal Society London, Series A **357**, 1907-1929 (1999).

[Shackleton 2000] Shackleton, N.: *The 100,000-Year Ice-Age Cycle Identified and Found to Lag Temperature, Carbon Dioxide, and Orbital Eccentricity.* Science **289**, 1897-1902 (2000).

[Shackleton et al. 2000] Shackleton, N.J.; Hall, M.A.; Raffi, I.; Tauxe, L.; Zachos, J.C.: *Astronomical calibration age for the Oligocene-Miocene boundary.* Geology **28**, 447-450 (2000).

[Sheridan et al. 1987] Sheridan, M.F.; Wohletz, K.H.; Dehn, J.: *Discrimination of grain-size subpopulations in pyroclastic deposits.* Geology **15**, 367-370 (1987).

[Silverman 1986] Silverman, B.W.: *Density Estimation for Statistics and Data Analysis.* Chapman and Hill, London (1986).

[Smith 1995] Smith, M.G.: *Survival of E. Coli and Salmonella after Chilling and Freezing in Liquid Media.* Journal of Food Science **60**(3), 509-512 (1995).

[StatLib] StatLib - Applied Statistics algorithms, *http://lib.stat.cmu.edu/apstat* .

[Stockhausen 1998] Stockhausen, H.: *Some new aspects for the modelling of isothermal remanent magnetisation acquisition curves by cumulative log-Gaussian functions.* Geophysical Research Letters **25**, 2217-2220 (1998).

[Stoner et al. 2000] Stoner, J.S.; Channell, J.E.T.; Hillaire-Marcel, C.; Kissel, C.: *Geomagnetic paleointensity and environmental record from Labrador Sea core MD 95-2024 : global marine sediment and ice core chronostratigraphy for the last 110kyr.* Earth and Planetary Science Letters **183**(1-2), 161-177 (2000).

[Sun et al. 2002] Sun, D.; Bloemendal, J.; Rea, D.K.; Vandenberghe, J.; Jiang, F.; Zhisheng, A.; Su, R.: *Grain-size distribution function of polymodal sediments in hydraulic and aeolian environments, and numerical partitioning of the sedimentary components.* Sedimentary Geology **152**, 263-277 (2002).

[Sutherland and Lee 1994] Sutherland, R.A.; Lee, C.T.: *Application of the log-hyperbolic distribution to Hawaiian beach sands.* Journal of Coastal Research **10**(2), 251-262 (1994).

[Tani et al. 2002] Tani, Y.; Kurihara, K. ; Nara, F.; Itoh, N.; Soma, M.; Soma, Y.; Tanaka, A.; Yoneda, M.; Hirota, M.; Shibata, Y.: *Temporal changes in the phytoplankton community of the southern basin of Lake Baikal over the last 24,000 years recorded by photosynthetic pigments in a sediment core.* Organic Geochemistry **33**(12), 1621-1634 (2002).

[Tarasov et al. 2005] Tarasov, P.; Granoszewski, W.; Bezrukova, E.; Brewer, S.; Nita, M.; Abzaeva, A.; Oberhänsli, H.: *Quantitative reconstruction of the last interglacial vegetation and climate based on the pollen reocrd from Lake Baikal, Russia.* Climate Dynamics **25**(6), 625-637 (2005).

[Tass et al. 1998] Tass, P.; Rosenblum, M.G.; Weule, J.; Kurths, J.; Pikovsky, A.; Volkmann, J.; Schnitzler, A.; Freund, H.-J.: *Detection of n:m Phase Locking from Noisy Data: Application to Magnetoencephalography.* Physical Review Letters **81** 3291-3294 (1998).

[Telford et al. 2004a] Telford, R.J.; Heegard, E.; Birks, H.J.B.: *All age-depth models are wrong: but how badly ?* Quaternary Science Reviews **23**, 1-5 (2004).

[Telford et al. 2004b] Telford, R.J.; Andersson, C.; Birks, H.J.B.; Juggins, S.: *Biases in the estimation of transfer function prediction errors.* Paleoceanography **19**, PA4014 (2004).

[Telford and Birks 2005] Telford, R.J.; Birks, H.J.B.: *The secret assumption of transfer functions: problems with spatial autocorrelation in evaluating model performance.* Quaternary Science Reviews **24**, 2173-2179 (2005).

[Tenenbaum et al. 2000] Tenenbaum, J.; de Silva, V.; Langford, J.C.: *A Global Geometric Framework for Nonlinear Dimensionality Reduction.* Science **290** (5500), 2319-2323 (2000).

[Theiler et al. 1992] Theiler, J.; Eubank, S.; Longtin, A.; Galdrikian, B.; Farmer, J.D.: *Testing for nonlinearity in time series: the method of surrogate data.* Physica D **58**, 77-94 (1992).

[Thiel et al. 2004] Thiel, M.; Romano, M.C.; Kurths, J.: *Estimation of Dynamical Invariants without Embedding by Recurrence Plots.* Chaos **14**(2), 234-243 (2004).

[Thompson and Clark 1989] Thompson,R.; Clark, R.M.: *Sequence slotting for stratigraphic correlation between cores: theory and practice.* Journal of Paleolimnology **2**, 173-184 (1989).

[Thompson et al. 1989] Thompson, L.G.; Mosley-Thompson, E.; Davis, M.E.; Bolzan, J.F.; Dai, J.; Yao, T.; Gundestrup, N.; Wu, X.; Klein, L.; Xie, Z.: *Holocene Late Pleistocene Climatic Ice Core Records from Qinghai-Tibetan Plateau.* Science **246**(4929), 474-477 (1989).

[Timmermann et al. 2003] Timmermann, A.; Gildor, H.; Schulz, M.; Tziperman, E.: *Coherent Resonant Millenial-Scale Climate Oscillations Triggered by Massive Meltwater Pulses.* Journal of Climate **16**, 2569-2585 (2003).

[Titterington et al. 1985] Titterington, D.M.; Smith, A.F.M.; Makov, U.E.: *Statistical Analysis of Finite Mixture Distributions.* Wiley, Chichester (1985).

[Torrence and Webster 1999] Torrence, C.; Webster, P.J.: *Inzterdecadal Changes in the ENSO-Monsoon System.* Journal of Climate **12**, 2679-2690 (1999).

[Tsonis and Roebber 2004] Tsonis, A.A.; Roebber, P.J.: *The architecture of the climate network.* Physica A **333**, 497-504 (2004).

[Vandenberghe et al. 1993] Vandenberghe, J.; Mommersteeg, H.; Edelman, D.: *Lithogenesis and geomorphological processes of the Pleistocene deposits at Belvédère.* Mededelingen-Rijks Geologische Dienst **47**, 7-18 (1993).

[Vandenberghe et al. 1997] Vandenberghe, J.; Zhisheng, A.; Nugteren, G.; Lu, H.; van Huissteden, J.: *A new absolute timescale for the Quaternary climate in the Chinese loess region based on grain size analysis.* Geology **25**, 35-38 (1997).

[Varela et al. 2005] Varela, H.; Beta, C.; Bonnefort, A.; Krischer, K.: *Transitions to Electrochemical Turbulence.* Physical Review Letters **94**, 174104 (2005).

[Venema et al. accepted] Venema, V.; Ament, F.; Simmer, C.: *A Stochastic Iterative Amplitude Adapted Fourier Transform Algorithm with Improved Accuracy.* Nonlinear Processes in Geophysics, accepted for publication.

[von Bloh et al. 2005] von Bloh, W.; Romano, M.C.; Thiel, M.: *Long-term predictability of mean daily temperature data.* Nonlinear Processes in Geophysics **12**, 471-479 (2005).

[Voss and Kurths 1997] Voss, H.; Kurths, J.: *Reconstruction of nonlinear time delay models from data by the use of optimal transformations.* Physics Letters A **234**, 336-344 (1997).

[Webber and Zbilut 1994] Webber Jr., C.L.; Zbilut, J.P.: *Dynamical assessment of physiological systems and states using recurrence plot strategies.* Journal of Applied Physiology **76**(2), 965-973 (1994).

[Webster and Yang 1992] Webster, P.J.; Yang, S.: *Monsoon and ENSO: Selectively interacting systems.* Quarterly Journal of the Royal Meteorological Society **118**, 887-926 (1992).

[Weltje 1997] Weltje, G.J.: *End-Member Modeling of Compositional Data: Numerical-Statistical Algorithms for Solving the Explicit Mixing Problem.* Mathematical Geology **29**(4), 503-549 (1997).

[Weltje and Prins accepted] Weltje, G.J.; Prins, M.A.: *Genetically meaningful decomposition of grain-size distributions.* Sedimentary Geology, accepted for publication.

[Williams and Handwerger 2005] Williams, T.; Handwerger, D.: *A high-resolution record of early Miocene Antarctic glacial history from ODP Site 1165, Prydz Bay.* Paleoceanography **20**, PA2017 (2005).

[Wilson et al. 2002] Wilson, G.S.; Lavelle, M.; McIntosh, W.C.; Roberts, A.P.; Harwood, D.M.; Watkins, D.K.; Villa, G.; Bohaty, S.M.; Fielding, C.R.; Florindo, F.; Sagnotti, L.; Naish, T.R.; Scherer, R.P.; Verosub, K.L.: *Integrated chronostratigraphic calibration of the Oligocene-Miocene boundary at 24.0 ± 0.1 Ma from the CRP-2A drill core, Ross Sea, Antarctica.* Geology **30**(11), 1043-1046 (2002).

[Witt et al. 1998] Witt, A.; Kurths, J.; Pikovsky, A.: *Testing stationarity in time series.* Physical Review E **58**(2), 1800-1810 (1998).

[Witt and Oberhänsli 2003] Witt, A.; Oberhänsli, H.: *Relative Zeitmodelle aus univariaten Messdaten.* Unpublished preprint (2003).

[Witt and Schumann 2005] Witt, A.; Schumann, A.Y.: *Holocene climate variability on millennial scales recorded in greenland ice cores.* Nonlinear Processes in Geophysics **12**, 345-352 (2005).

[Witt and Oberhänsli 2006] Witt, A.; Oberhänsli, H.: *Millenial scale variability of Northern hemisphere atmospheric dust loading.* Preprint (2005).

[Wohletz et al. 1989] Wohletz, K.H.; Sheridan, M.F.; Brown, W.K.: *Particle Size Distributions and the Sequential Fragmentation/Transport Theory Applied to Volcanic Ash.* Journal of Geophysical Research B **94**(11), 15,703-15,721 (1989).

[Wu 1983] Wu, C.F.J.: *On the Convergence Properties of the EM Algorithm.* The Annals of Statistics **11**(1), 95-103 (1983).

[Wu 2003] Wu, X.: *Calculation of maximum entropy densities with application to income distribution.* Journal of Econometrics **115**, 347-354 (2003).

[Wu and Perloff 2003] Wu, X.; Perloff, J.M.: *Maximum Entropy Density Estimation with Grouped Data.* Working Paper (2003).

[Wunsch 2000] Wunsch, C.: *On sharp spectral lines in the climate record and the millenial peak.* Paleoceanography **15**(4), 417-424 (2000).

[Wunsch 2003] Wunsch, C.: *The spectral description of climate change including the 100 ky energy.* Climate Dynamics **20**, 353-363 (2003).

[Wyrwoll and Smyth 1985] Wyrwoll, K.-H.; Smyth, G.K.: *On Using the Log-Hyperbolic Distribution to Describe the Textural Characteristics of Eolian Sediments.* Journal of Sedimentary Petrology **55**(4), 471-478 (1985).

[Wyrwoll and Smyth 1988] Wyrwoll, K.-H.; Smyth, G.: *On Using the Log-Hyperbolic Distribution to Describe the Textural Characteristics of Eolian Sediments - Reply.* Journal of Sedimentary Petrology **58**(1), 161-162 (1988).

[Yuretich et al. 1999] Yuretich, R.; Melles, M.; Sarata, B.; Grobe, H.: *Clay minerals in the sediments of Lake Baikal: a useful climate proxy.* Journal of Sedimentary Research **69**(3), 588-596 (1999).

[Zachos et al. 1997] Zachos, J.C.; Flower, B.P.; Paul, H.: *Orbitally paced climate oscillations across the Oligocene/Miocene boundary.* Nature **388**, 567-570 (1997).

[Zachos et al. 2001] Zachos, J.C.; Shackleton, N.J.; Revenaugh, J.S.; Pälike, H.; Flower, B.P.: *Climate Response to Orbital Forcing Across the Oligocene-Miocene Boundary.* Science **292**, 274-278 (2001).

[Zbilut and Webber 1992] Zbilut, J.P.; Webber Jr., C.L.: *Embeddings and delays as derived from quantification of recurrence plots.* Physics Letters A **171**(3-4), 199-203 (1992).

[Zbilut et al. 1998] Zbilut, J.P.; Giuliani, A.; Webber Jr., C.L.: *Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification.* Physics Letters A **246**(1-2), 122-128 (1998).

[Zhang 2002] Zhang, J.: *Powerful goodness-of-fit tests based on the likelihood ratio.* J. R. Statist. Soc. B **64** (2), 281-294 (2002).

[Zobeck et al. 1999] Zobeck, T.M.; Gill, T.E.; Popham, T.W.: *A Two-Parameter Weibull Function to Describe Airborne Dust Particle Size Distributions.* Earth Surface Processes and Landforms **24**, 943-955 (1999).

[Zoldi and Greenside 1997] Zoldi, S.M.; Greenside, H.M.: *Karhunen-Loève Decomposition of Extensive Chaos.* Physical Review Letters **78**(9), 1687-1690 (1997).

[Zoldi et al. 1998] Zoldi, S.M.; Liu, J.; Bajaj, K.M.S.; Greenside, H.S.; Ahlers, G.: *Extensive scaling and nonuniformity of the Karhunen-Loève decomposition for the spiral-defect chaos state.* Physical Review E **58**(6), 6903-6906 (1998).

Appendix A

EM Algorithm for Gaussian Mixture Models

A.1 Maximisation Step for Gaussian Components using Explicit Observations

As a particular application, the parameter estimation in mixtures of normal distributions has attracted much interest during the last century (for a historical review, see [Everitt and Hand 1981]). For a one-component normal distribution

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \tag{A.1}$$

one has that

$$Q(\mathbf{\Psi};\mathbf{\Psi}^{(l)}) = \sum_{j=1}^{J} \log f(x_j;\mathbf{\Psi}^{(l)}) = -\sum_{j=1}^{J}\left(\frac{\log(2\pi)}{2} + \frac{\log(\sigma^2)^{(l)}}{2} + \frac{\left(x-\mu^{(l)}\right)^2}{2(\sigma^2)^{(l)}}\right) \tag{A.2}$$

and therefore

$$0 = \frac{\partial Q(\mathbf{\Psi};\mathbf{\Psi}^{(l)})}{\partial \mu} = \sum_{j=1}^{J} \frac{x_j-\mu}{(\sigma^2)^{(l)}} \tag{A.3}$$

$$0 = \frac{\partial Q(\mathbf{\Psi};\mathbf{\Psi}^{(l)})}{\partial (\sigma^2)^{(l)}} = -\sum_{j=1}^{J}\left(\frac{1}{2(\sigma^2)^{(l)}} - \frac{(x_j-\mu)^2}{2\left((\sigma^2)^{(l)}\right)^2}\right) \tag{A.4}$$

such that solving for the respective parameters yields

$$\mu^{(l+1)} = \frac{1}{J}\sum_{j=1}^{J} x_j, \tag{A.5}$$

$$(\sigma^2)^{(l+1)} = \frac{1}{J}\sum_{j=1}^{J}\left(x_j - \mu^{(l+1)}\right)^2. \tag{A.6}$$

which corresponds to the standard estimates for the first and second moments. As both results do not depend on the previous-step parameter estimates, the ML solution is ultimately reached

with only one step corresponding to the empirical mean and variance of the observed data as expected (note, however, that the variance estimator is biased in this case).

For superpositions of normal distributions, the EM calculus actually leads to algorithmic estimates as

$$0 = \frac{\partial Q(\mathbf{\Psi}; \mathbf{\Psi}^{(l)})}{\partial \mu_k} = \sum_{j=1}^{J} \frac{\pi_k f_k(x_j; \mathbf{\Theta}^{(l)})}{\sum_{i=1}^{K} \pi_i^{(l)} f_i(x_j; \mathbf{\Theta}^{(l)})} \frac{x_j - \mu_k}{2\sigma_k^2} \tag{A.7}$$

$$0 = \frac{\partial Q(\mathbf{\Psi}; \mathbf{\Psi}^{(l)})}{\partial (\sigma_k^2)^{(l)}} = \sum_{j=1}^{J} \frac{\pi_k f_k(x_j; \mathbf{\Theta}^{(l)})}{\sum_{i=1}^{K} \pi_i^{(l)} f_i(x_j; \mathbf{\Theta}^{(l)})} \left(\frac{1}{2\sigma_k^2} - \frac{(x_j - \mu_k)^2}{2\left(\sigma_k^2\right)^2} \right) \tag{A.8}$$

such that

$$\pi_k^{(l+1)} = \pi_k^{(l)} \frac{1}{J} \sum_{j=1}^{J} \frac{f_k(x_j; \mathbf{\Theta}^{(l)})}{\sum_{i=1}^{K} \pi_i^{(l)} f_i(x_j; \mathbf{\Theta}^{(l)})} \tag{A.9}$$

$$\mu_k^{(l+1)} = \frac{\sum_{j=1}^{J} x_j \frac{f_k(x_j; \mathbf{\Theta}^{(l)})}{\sum_{i=1}^{K} \pi_i^{(l)} f_i(x_j; \mathbf{\Theta}^{(l)})}}{\sum_{j=1}^{J} \frac{f_k(x_j; \mathbf{\Theta}^{(l)})}{\sum_{i=1}^{K} \pi_i^{(l)} f_i(x_j; \mathbf{\Theta}^{(l)})}} \tag{A.10}$$

$$(\sigma_k^2)^{(l+1)} = \frac{\sum_{j=1}^{J} \left(x_j - \mu_k^{(l+1)} \right)^2 \frac{f_k(x_j; \mathbf{\Theta}^{(l)})}{\sum_{i=1}^{K} \pi_i^{(l)} f_i(x_j; \mathbf{\Theta}^{(l)})}}{\sum_{j=1}^{J} \frac{f_k(x_j; \mathbf{\Theta}^{(l)})}{\sum_{i=1}^{K} \pi_i^{(l)} f_i(x_j; \mathbf{\Theta}^{(l)})}} \tag{A.11}$$

have to be iterated until convergence is approached.

In [Hasselblad 1966], the estimation of parameters in a normal mixture was considered using different approximations to the ML solution. [Hasselblad 1969] used the corresponding results as initial estimates for an iterative procedure corresponding to the above EM algorithm whose power was demonstrated for Poisson, binomial, and exponential mixtures. [Everitt and Hand 1981] give the explicit equations for the case of normal distributions as shown above and points the connection to the EM algorithm. In [Everitt 1984], it is shown that the EM algorithm is one of the most efficient parameter estimation strategies for Gaussian mixtures.

A.2 Maximisation Step for Grouped Normal Data

For grouped and possibly truncated data, there is in general no explicit formulation of the maximisation step. However, in the case of observations following a Gaussian distribution, i.e.,

$$\log f(x; \mu, \sigma^2) = -\frac{1}{2} \left[\log(2\pi) + \log \sigma^2 \right] - \frac{(x - \mu)^2}{2\sigma^2}, \tag{A.12}$$

the equivalent of equation 2.15 may be explicitly solved for the unknown parameters μ and σ as discussed in [McLachlan and Krishnan 1997] (the explicit equations for the expectation values to compute here have already been given in [Hasselblad et al. 1980]). Evaluating the expectation

value of $\log L_c(\mathbf{\Psi})$ with respect to the unknown explicit observations $\{x'_{mj}\}$ yields

$$\begin{aligned} Q(\mathbf{\Psi}, \mathbf{\Psi}^{(l)}) &= E_{\mathbf{\Psi}^{(l)}} \{\log L_c(\mathbf{\Psi})|\vec{n}\} = \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)})\ E_{\mathbf{\Psi}^{(l)}} \left\{\log f(x;\mu,\sigma^2)|x \in X_m\right\} \\ &= -\frac{1}{2}\left(\log(2\pi) + \log\sigma^2\right) \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) - \frac{1}{2\sigma^2} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_{\mathbf{\Psi}^{(l)}} \left\{(x-\mu)^2|x \in X_m\right\} \end{aligned} \tag{A.13}$$

Note that the estimates from the previous iteration step enter only in the calculation of the expectation values.

In the maximisation step, the maxima of $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(l)})$ with respect to the different distribution parameters are evaluated. For this purpose, one firstly calculates the derivative with respect to μ as

$$\begin{aligned} 0 = \frac{\partial Q(\mathbf{\Psi}, \mathbf{\Psi}^{(l)})}{\partial\mu} &= -\frac{1}{2\sigma^2} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_{\mathbf{\Psi}^{(l)}} \left\{\frac{\partial}{\partial\mu}(x-\mu)^2|x \in X_m\right\} \\ = \frac{1}{\sigma^2} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_{\mathbf{\Psi}^{(l)}} \{(x-\mu)|x \in X_m\} &= \frac{1}{\sigma^2} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) \left[E_{\mathbf{\Psi}^{(l)}} \{x|x \in X_m\} - \mu\right], \end{aligned} \tag{A.14}$$

which is solved by

$$\mu^{(l+1)} = \frac{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_{\mathbf{\Psi}^{(l)}}\{x|x \in X_m\}}{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)})} = \frac{1}{n + n'(\mathbf{\Psi}^{(l)})} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) \frac{\int_{X_m} dx\ x\ f(x;\mathbf{\Psi}^{(l)})}{\int_{X_m} dx\ f(x;\mathbf{\Psi}^{(l)})}. \tag{A.15}$$

For further computations, one may set $X_m = (a_m, b_m)$ with $a_{m+1} = b_m$ for $m = 1, \ldots, M-1$. To evaluate the integrals, it is useful to consider of the following identity for a normal distribution:

$$\frac{df(x;\mathbf{\Psi}^{(l)})}{dx} = -\frac{x-\mu^{(l)}}{(\sigma^{(l)})^2} f(x;\mathbf{\Psi}^{(l)}) \tag{A.16}$$

which may be transformed into the form

$$x f(x;\mathbf{\Psi}^{(l)}) = \mu^{(l)} f(x;\mathbf{\Psi}^{(l)}) - (\sigma^2)^{(l)} \frac{df(x;\mathbf{\Psi}^{(l)})}{dx} \tag{A.17}$$

Inserting this equation into the integral leads to the final equation for the computation of $\mu^{(l+1)}$:

$$\mu^{(l+1)} = \frac{1}{n + n'(\mathbf{\Psi}^{(l)})} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) \left[\mu^{(l)} - (\sigma^2)^{(l)} \frac{f(b_m;\mathbf{\Psi}^{(l)}) - f(a_m;\mathbf{\Psi}^{(l)})}{F(b_m;\mathbf{\Psi}^{(l)}) - F(a_m;\mathbf{\Psi}^{(l)})}\right] \tag{A.18}$$

where

$$F(x;\mathbf{\Psi}) = \int_{-\infty}^{x} dy\ f(y;\mathbf{\Psi}) \tag{A.19}$$

is the corresponding cumulative distribution function, in case of a normal distribution the famous error function.

For the computation of $(\sigma^2)^{(l+1)}$, one has to solve

$$0 = \frac{\partial Q(\mathbf{\Psi}, \mathbf{\Psi}^{(l)})}{\partial \sigma^2} = -\frac{1}{2\sigma^2} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) \left(1 - \frac{1}{\sigma^2} E_{\mathbf{\Psi}^{(l)}} \left\{(x-\mu)^2 | x \in X_m\right\}\right) \tag{A.20}$$

which leads to the next-step ML estimate for σ^2 as

$$\begin{aligned}(\sigma^2)^{(l+1)} &= \frac{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_{\mathbf{\Psi}^{(l)}} \left\{(x-\mu)^2 | x \in X_m\right\}}{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)})} \\ &= \frac{1}{n + n'(\mathbf{\Psi}^{(l)})} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) \frac{\int_{a_m}^{b_m} dx\ (x-\mu)^2\ f(x; \mathbf{\Psi}^{(l)})}{\int_{a_m}^{b_m} dx\ f(x; \mathbf{\Psi}^{(l)})}.\end{aligned} \tag{A.21}$$

Here, for μ the next-step estimate $\mu^{(l+1)}$ has to be inserted. As $\mu^{(l+1)}$ itself does not depend on $(\sigma^2)^{(l+1)}$ itself, one may firstly compute μ and then σ^2. Note that, for arbitrary distribution functions, the parameters may depend on each other in a much more complicated way.

The integral in the denominator consists of three parts:

$$\begin{aligned}\int_{a_m}^{b_m} dx\ (x-\mu^{(l+1)})^2\ f(x; \mathbf{\Psi}^{(l)}) = \int_{a_m}^{b_m} dx\ x^2\ f(x; \mathbf{\Psi}^{(l)}) - 2\mu^{(l+1)} \int_{a_m}^{b_m} dx\ x\ f(x; \mathbf{\Psi}^{(l)}) \\ + (\mu^{(l+1)})^2 \int_{a_m}^{b_m} dx\ f(x; \mathbf{\Psi}^{(l)})\end{aligned} \tag{A.22}$$

While the expressions in the second and third terms have yet been solved above, the first integral still needs to be computed. For this purpose, one may apply (A.17) twice to obtain

$$\begin{aligned}\frac{d^2 f(x; \mathbf{\Psi}^{(l)})}{dx^2} &= \frac{d}{dx}\left(-\frac{x-\mu^{(l)}}{(\sigma^2)^{(l)}} f(x; \mathbf{\Psi}^{(l)})\right) = -\left(\frac{d}{dx}\frac{x-\mu^{(l)}}{(\sigma^2)^{(l)}}\right) - \frac{x-\mu^{(l)}}{(\sigma^2)^{(l)}} \frac{df(x; \mathbf{\Psi}^{(l)})}{dx} \\ &= -\frac{1}{(\sigma^2)^{(l)}} f(x; \mathbf{\Psi}^{(l)}) + \frac{(x-\mu^{(l)})^2}{((\sigma^2)^{(l)})^2} f(x; \mathbf{\Psi}^{(l)})\end{aligned} \tag{A.23}$$

or

$$\begin{aligned}x^2 f(x; \mathbf{\Psi}^{(l)}) &= ((\sigma^2)^{(l)})^2 \frac{d^2 f(x; \mathbf{\Psi}^{(l)})}{dx^2} + \left((\sigma^2)^{(l)} - (\mu^{(l)})^2\right) f(x; \mathbf{\Psi}^{(l)}) + 2\mu^{(l)} x f(x; \mathbf{\Psi}^{(l)}) \\ &= ((\sigma^2)^{(l)})^2 \frac{d^2 f(x; \mathbf{\Psi}^{(l)})}{dx^2} + \left((\sigma^2)^{(l)} - (\mu^{(l)})^2\right) f(x; \mathbf{\Psi}^{(l)}) + \\ &\qquad 2\mu^{(l)} \left(\mu^{(l)} f(x; \mathbf{\Psi}^{(l)}) - (\sigma^2)^{(l)} \frac{df(x; \mathbf{\Psi}^{(l)})}{dx}\right) \\ &= ((\sigma^2)^{(l)})^2 \frac{d^2 f(x; \mathbf{\Psi}^{(l)})}{dx^2} + \left((\sigma^2)^{(l)} + (\mu^{(l)})^2\right) f(x; \mathbf{\Psi}^{(l)}) - 2\mu^{(l)} (\sigma^2)^{(l)} \frac{df(x; \mathbf{\Psi}^{(l)})}{dx}.\end{aligned} \tag{A.24}$$

The integral over the second derivative of f becomes

$$\int_{a_m}^{b_m} dx \frac{d^2 f(x; \mathbf{\Psi}^{(l)})}{dx^2} = \left.\frac{df(x; \mathbf{\Psi}^{(l)})}{dx}\right|_{a_m}^{b_m} = \left.\frac{\mu^{(l)}}{(\sigma^2)^{(l)}} f(x; \mathbf{\Psi}^{(l)})\right|_{a_m}^{b_m} - \left.\frac{x}{(\sigma^2)^{(l)}} f(x; \mathbf{\Psi}^{(l)})\right|_{a_m}^{b_m} \tag{A.25}$$

such that

$$\begin{aligned}\int_{a_m}^{b_m} dx\ x^2\ f(x;\mathbf{\Psi}^{(l)}) &= ((\sigma^2)^{(l)})^2 \int_{a_m}^{b_m} dx \frac{d^2 f(x;\mathbf{\Psi}^{(l)})}{dx^2} \\ &\quad + \left((\sigma^2)^{(l)} + (\mu^{(l)})^2\right) \int_{a_m}^{b_m} dx\ f(x;\mathbf{\Psi}^{(l)}) - 2\mu^{(l)}(\sigma^2)^{(l)} \int_{a_m}^{b_m} dx \frac{df(x;\mathbf{\Psi}^{(l)})}{dx} \\ &= (\sigma^2)^{(l)}\mu^{(l)} \left(f(b_m;\mathbf{\Psi}^{(l)} - f(a_m;\mathbf{\Psi}^{(l)})\right) \\ &\quad - (\sigma^2)^{(l)} \left(b_m f(b_m;\mathbf{\Psi}^{(l)} - a_m f(a_m;\mathbf{\Psi}^{(l)})\right) \\ &\quad + \left((\sigma^2)^{(l)} + (\mu^{(l)})^2\right) \left(F(b_m;\mathbf{\Psi}^{(l)} - F(a_m;\mathbf{\Psi}^{(l)})\right) \\ &\quad - 2(\sigma^2)^{(l)}\mu^{(l)} \left(f(b_m;\mathbf{\Psi}^{(l)} - f(a_m;\mathbf{\Psi}^{(l)})\right)\end{aligned} \tag{A.26}$$

and therefore

$$\begin{aligned}\int_{a_m}^{b_m} dx\ (x-\mu^{(l+1)})^2\ f(x;\mathbf{\Psi}^{(l)}) &= \left((\sigma^2)^{(l)} + (\mu^{(l)} - \mu^{(l+1)})^2\right) \left(F(b_m;\mathbf{\Psi}^{(l)} - F(a_m;\mathbf{\Psi}^{(l)})\right) \\ &\quad + \left(2\mu^{(l+1)} - \mu^{(l)}\right) (\sigma^2)^{(l)} \left(f(b_m;\mathbf{\Psi}^{(l)} - f(a_m;\mathbf{\Psi}^{(l)})\right) \\ &\quad - (\sigma^2)^{(l)} \left(b_m f(b_m;\mathbf{\Psi}^{(l)} - a_m f(a_m;\mathbf{\Psi}^{(l)})\right)\end{aligned} \tag{A.27}$$

To finally give a short expression for the M-step parameter estimates for grouped normal data, it is useful to use a brief notation for expressing the final integrals, e.g.

$$\Delta_m F^{(l)} = F(b_m;\mathbf{\Psi}^{(l)}) - F(a_m;\mathbf{\Psi}^{(l)}) = P_m(\mathbf{\Psi}^{(l)}) \tag{A.28}$$

$$\Delta_m f^{(l)} = f(b_m;\mathbf{\Psi}^{(l)}) - f(a_m;\mathbf{\Psi}^{(l)}) \tag{A.29}$$

$$\Delta_m \phi^{(l)} = b_m f(b_m;\mathbf{\Psi}^{(l)}) - a_m f(a_m;\mathbf{\Psi}^{(l)}) \tag{A.30}$$

to formulate the following set of equations

$$\mu^{(l+1)} = \mu^{(l)} - \frac{(\sigma^2)^{(l)}}{n + n'(\mathbf{\Psi}^{(l)})} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) \frac{\Delta_m f^{(l)}}{\Delta_m F^{(l)}} \tag{A.31}$$

$$\begin{aligned}(\sigma^2)^{(l+1)} &= (\sigma^2)^{(l)} + (\mu^{(l+1)} - \mu^{(l)})^2 + \frac{(\sigma^2)^{(l)}}{n + n'(\mathbf{\Psi}^{(l)})} \left(2\mu^{(l+1)} - \mu^{(l)}\right) \times \\ &\quad \times \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) \frac{\Delta_m f^{(l)}}{\Delta_m F^{(l)}} - \frac{(\sigma^2)^{(l)}}{n + n'(\mathbf{\Psi}^{(l)})} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) \frac{\Delta_m \phi^{(l)}}{\Delta_m F^{(l)}}\end{aligned} \tag{A.32}$$

A.3 Finite Normal Mixtures

For grouped data from normal mixtures, there are again certain complications with respect to the single-component distribution's parameter estimation problem. In particular, the introduction

of indicator variables and their corresponding probabilities to belong to the respective mixture components makes the explicit calculation of the ML estimates for the different parameters becoming significantly more difficult.

The derivation of the next-step estimate for the statistical weights has yet been described. For the recalculation of the distribution parameters in terms of the ML estimation, one has to compute again the corresponding derivatives of the log-likelihood's expectational value with respect to the different parameters involved.

First, the recalculation of the means of the different component densities shall be discussed here. The derivative of $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(l)})$ with respect to the μ_k reads:

$$\begin{aligned} 0 = \frac{\partial}{\partial \mu_k} Q(\mathbf{\Psi}, \mathbf{\Psi}^{(l)}) &= \frac{\partial}{\partial \mu_k} \sum_{i=1}^{K} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_i(x; \mathbf{\Psi}^{(l)}) \left(\log f_i(x; \mathbf{\Theta}) + \log \pi_i \right) \right\} \\ &= \sum_{i=1}^{K} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_i(x; \mathbf{\Psi}^{(l)}) \frac{\partial}{\partial \mu_k} \log f_i(x; \mathbf{\Theta}) \right\} \\ &= \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_k(x; \mathbf{\Psi}^{(l)}) \frac{\partial}{\partial \mu_k} \log f_k(x; \mathbf{\Theta}) \right\} \\ &= \frac{1}{(\sigma_k^2)^{(l)}} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_k(x; \mathbf{\Psi}^{(l)}) (x - \mu_k) \right\} \end{aligned} \tag{A.33}$$

which immediately leads to the expression for $\mu_i^{(l+1)}$ as follows:

$$\mu_i^{(l+1)} = \frac{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_i(x; \mathbf{\Psi}^{(l)}) x \right\}}{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_i(x; \mathbf{\Psi}^{(l)}) \right\}} \tag{A.34}$$

Using this result, one computes the derivative with respect to σ_k^2

$$\begin{aligned} 0 = \frac{\partial}{\partial \sigma_k^2} Q(\mathbf{\Psi}, \mathbf{\Psi}^{(l)}) &= \frac{\partial}{\partial \sigma_k^2} \sum_{i=1}^{K} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_i(x; \mathbf{\Psi}^{(l)}) \left(\log f_i(x; \mathbf{\Theta}) + \log \pi_i \right) \right\} \\ &= \sum_{i=1}^{K} \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_i(x; \mathbf{\Psi}^{(l)}) \frac{\partial}{\partial \sigma_k^2} \log f_i(x; \mathbf{\Theta}) \right\} \\ &= \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_k(x; \mathbf{\Psi}^{(l)}) \frac{\partial}{\partial \sigma_k^2} \log f_k(x; \mathbf{\Theta}) \right\} \\ &= \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_k(x; \mathbf{\Psi}^{(l)}) \frac{1}{2(\sigma_k^2)} \left(\frac{1}{(\sigma_k^2)} (x - \mu_k)^2 - 1 \right) \right\} \end{aligned} \tag{A.35}$$

leading to the following equation for $(\sigma_i^2)^{(l+1)}$:

$$(\sigma_i^2)^{(l+1)} = \frac{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_i(x; \mathbf{\Psi}^{(l)}) (x - \mu_i^{(l+1)})^2 \right\}}{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_m^{(l)} \left\{ t_i(x; \mathbf{\Psi}^{(l)}) \right\}} \tag{A.36}$$

where the μ_i have been replaced by their new estimates $\mu_i^{(l+1)}$.

For the required expectation values, one may again use the general expression

$$E_{\mathbf{\Psi}^{(l)}}\left\{g(x)|x\in X_m\right\} = E_m^{(l)}\left\{g(x)\right\} = \frac{\int_{a_m}^{b_m} dx\ f(x;\mathbf{\Psi}^{(l)})\ g(x)}{\int_{a_m}^{b_m} dx\ f(x;\mathbf{\Psi}^{(l)})}. \tag{A.37}$$

with $g(x)$ appearing here in any case as a sum over terms of the form $t_i(x;\mathbf{\Psi}^{(l)})x^s$ with $s = 0,1,2$. Because taking the expectation value is a linear operation, one may first compute the expectations of these particular terms and then evaluate the required expressions.

For $s = 0$, the expectation value is easily established:

$$E_m^{(l)}\left\{t_i(x;\mathbf{\Psi}^{(l)})\right\} = \frac{\int_{a_m}^{b_m} dx\ \pi_i^{(l)} f_i(x,\mathbf{\Theta}^{(l)})}{\int_{a_m}^{b_m} dx\ f(x;\mathbf{\Psi}^{(l)})} = \pi_i^{(l)}\frac{F_i(b_m;\mathbf{\Theta}^{(l)}) - F_i(a_m;\mathbf{\Theta}^{(l)})}{F(b_m;\mathbf{\Theta}^{(l)}) - F(a_m;\mathbf{\Theta}^{(l)})} \tag{A.38}$$

with

$$F_i(x;\mathbf{\Theta}^{(l)}) = \int_{-\infty}^{x} dy\ f_i(y;\mathbf{\Theta}^{(l)}) \tag{A.39}$$

$$F(x;\mathbf{\Psi}^{(l)}) = \int_{-\infty}^{x} dy\ f(y;\mathbf{\Psi}^{(l)}) = \sum_{i=1}^{K}\pi_i F_i(x;\mathbf{\Theta}^{(l)}) \tag{A.40}$$

For $s > 0$, one needs again the identities of the normal distribution (A.17) and (A.24) for computation applied to any component i of the mixture. Then, the expectation value for $t_i(x;\mathbf{\Psi}^{(l)})x$ reads as follows:

$$\begin{aligned}
E_m^{(l)}\left\{t_i(x;\mathbf{\Psi}^{(l)})x\right\} &= \frac{\int_{a_m}^{b_m} dx\ \pi_i^{(l)}\ x\ f_i(x,\mathbf{\Theta}^{(l)})}{\int_{a_m}^{b_m} dx\ f(x;\mathbf{\Psi}^{(l)})} = \pi_i^{(l)}\frac{\int_{a_m}^{b_m} dx\ x\ f_i(x,\mathbf{\Theta}^{(l)})}{F(b_m;\mathbf{\Theta}^{(l)}) - F(a_m;\mathbf{\Theta}^{(l)})} \\
&= \frac{\pi_i^{(l)}}{F(b_m;\mathbf{\Theta}^{(l)}) - F(a_m;\mathbf{\Theta}^{(l)})}\left[\mu_i^{(l)}\int_{a_m}^{b_m} dx\ f_i(x,\mathbf{\Theta}^{(l)}) - (\sigma_i^2)^{(l)}\int_{a_m}^{b_m} dx\ \frac{df_i(x,\mathbf{\Theta}^{(l)})}{dx}\right] \\
&= \pi_i^{(l)}\mu_i^{(l)}\frac{F_i(b_m;\mathbf{\Theta}^{(l)}) - F_i(a_m;\mathbf{\Theta}^{(l)})}{F(b_m;\mathbf{\Theta}^{(l)}) - F(a_m;\mathbf{\Theta}^{(l)})} - \pi_i^{(l)}(\sigma_i^2)^{(l)}\frac{f_i(b_m;\mathbf{\Theta}^{(l)}) - f_i(a_m;\mathbf{\Theta}^{(l)})}{F(b_m;\mathbf{\Theta}^{(l)}) - F(a_m;\mathbf{\Theta}^{(l)})}
\end{aligned} \tag{A.41}$$

For $s = 2$, the expectation value is computed analogously to a one-component Gaussian distribution by replacing f, μ, σ in (A.26) by $\pi_i f_i$, μ_i, σ_i, resp., yielding

$$\begin{aligned}
E_m^{(l)}\left\{t_i(x;\mathbf{\Psi}^{(l)})x^2\right\} &= \frac{\int_{a_m}^{b_m} dx\ \pi_i^{(l)}\ x^2\ f_i(x;\mathbf{\Theta}^{(l)})}{\int_{a_m}^{b_m} dx\ f(x;\mathbf{\Psi}^{(l)})} = \pi_i^{(l)}\frac{\int_{a_m}^{b_m} dx\ x^2\ f_i(x;\mathbf{\Theta}^{(l)})}{F(b_m;\mathbf{\Psi}^{(l)}) - F(a_m;\mathbf{\Psi}^{(l)})} \\
&= \pi_i^{(l)}((\mu_i^{(l)})^2 + (\sigma_i^2)^{(l)})\frac{F_i(b_m;\mathbf{\Theta}^{(l)}) - F_i(a_m;\mathbf{\Theta}^{(l)})}{F(b_m;\mathbf{\Theta}^{(l)}) - F(a_m;\mathbf{\Theta}^{(l)})} \\
&- \pi_i^{(l)}\mu_i^{(l)}(\sigma_i^2)^{(l)}\frac{f_i(b_m;\mathbf{\Theta}^{(l)}) - f_i(a_m;\mathbf{\Theta}^{(l)})}{F(b_m;\mathbf{\Theta}^{(l)}) - F(a_m;\mathbf{\Theta}^{(l)})} - \pi_i^{(l)}(\sigma_i^2)^{(l)}\frac{b_m f_i(b_m;\mathbf{\Theta}^{(l)}) - a_m f_i(a_m;\mathbf{\Theta}^{(l)})}{F(b_m;\mathbf{\Theta}^{(l)}) - F(a_m;\mathbf{\Theta}^{(l)})}.
\end{aligned} \tag{A.42}$$

Finally, one may combine the expectation values for $s = 0,1,2$ to approach the corresponding

value for $t_i(x;\mathbf{\Psi}^{(l)})(x-\mu_i^{(l+1)})^2$ as follows:

$$\begin{aligned}
&E_m^{(l)}\left\{t_i(x;\mathbf{\Psi}^{(l)})\left(x-\mu_i^{(l+1)}\right)^2\right\}\\
&\quad= E_m^{(l)}\left\{t_i(x;\mathbf{\Psi}^{(l)})x^2\right\} - 2\mu_i^{(l+1)}E_m^{(l)}\left\{t_i(x;\mathbf{\Psi}^{(l)})x\right\} + (\mu_i^{(l+1)})^2 E_m^{(l)}\left\{t_i(x;\mathbf{\Psi}^{(l)})\right\}\\
&\quad= \pi_i^{(l)}\left[\left((\sigma_i^2)^{(l)} + (\mu_i^{(l)})^2 + (\mu_i^{(l+1)})^2 - 2\mu_i^{(l)}\mu_i^{(l+1)}\right)\frac{F_i(b_m;\mathbf{\Theta}^{(l)}) - F_i(a_m;\mathbf{\Theta}^{(l)})}{F(b_m;\mathbf{\Theta}^{(l)}) - F(a_m;\mathbf{\Theta}^{(l)})}\right.\\
&\qquad+\left(2\mu_i^{(l+1)}(\sigma_i^2)^{(l)} - \mu_i^{(l)}(\sigma_i^2)^{(l)}\right)\frac{f_i(b_m;\mathbf{\Theta}^{(l)}) - f_i(a_m;\mathbf{\Theta}^{(l)})}{F(b_m;\mathbf{\Theta}^{(l)}) - F(a_m;\mathbf{\Theta}^{(l)})}\\
&\qquad\left.-(\sigma_i^2)^{(l)}\frac{b_m f_i(b_m;\mathbf{\Theta}^{(l)}) - a_m f_i(a_m;\mathbf{\Theta}^{(l)})}{F(b_m;\mathbf{\Theta}^{(l)}) - F(a_m;\mathbf{\Theta}^{(l)})}\right]
\end{aligned} \tag{A.43}$$

Adapting the abbreviations for the occuring differences from the case of a one-component normal distribution as

$$\Delta_m F_i^{(l)} = F_i(b_m;\mathbf{\Psi}^{(l)}) - F_i(a_m;\mathbf{\Psi}^{(l)}) \tag{A.44}$$

$$\Delta_m f_i^{(l)} = f_i(b_m;\mathbf{\Psi}^{(l)}) - f_i(a_m;\mathbf{\Psi}^{(l)}) \tag{A.45}$$

$$\Delta_m \phi_i^{(l)} = b_m f_i(b_m;\mathbf{\Psi}^{(l)}) - a_m f_i(a_m;\mathbf{\Psi}^{(l)}) \tag{A.46}$$

such that

$$\Delta_m F^{(l)} = \sum_{i=1}^{K}\pi_i^{(l)}\Delta_m F_i^{(l)} \tag{A.47}$$

$$\Delta_m f^{(l)} = \sum_{i=1}^{K}\pi_i^{(l)}\Delta_m f_i^{(l)} \tag{A.48}$$

$$\Delta_m \phi^{(l)} = \sum_{i=1}^{K}\pi_i^{(l)}\Delta_m \phi_i^{(l)} \tag{A.49}$$

allows to explicitly write the equations for the next-step estimates in rather compact form:

$$\pi_i^{(l+1)} = \pi_i^{(l)}\,\frac{1}{n+n'(\mathbf{\Psi}^{(l)})}\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)})\frac{\Delta_m F_i^{(l)}}{\Delta_m F^{(l)}} \tag{A.50}$$

$$\mu_i^{(l+1)} = \mu_i^{(l)} - (\sigma_i^2)^{(l)}\frac{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)})\frac{\Delta_m f_i^{(l)}}{\Delta_m F^{(l)}}}{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)})\frac{\Delta_m F_i^{(l)}}{\Delta_m F^{(l)}}} \tag{A.51}$$

$$\begin{aligned}
(\sigma_i^2)^{(l+1)} &= (\sigma_i^2)^{(l)} + (\mu_i^{(l+1)} - \mu_i^{(l)})^2 + (\sigma_i^2)^{(l)}\left(2\mu_i^{(l+1)} - \mu_i^{(l)}\right)\times\\
&\quad\times\frac{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)})\frac{\Delta_m f_i^{(l)}}{\Delta_m F^{(l)}}}{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)})\frac{\Delta_m F_i^{(l)}}{\Delta_m F^{(l)}}} - (\sigma_i^2)^{(l)}\frac{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)})\frac{\Delta_m \phi_i^{(l)}}{\Delta_m F^{(l)}}}{\sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)})\frac{\Delta_m F_i^{(l)}}{\Delta_m F^{(l)}}}
\end{aligned} \tag{A.52}$$

A.4 Numerical Approximation of the Error Function

In contrast to the case of explicit data where the assumed distribution function is explicitly given and therefore the corresponding probability of a given value may be computed without any problem, for grouped data it is necessary to calculate the cumulative distribution function of the model at the bin boundaries as well. In case of normal distributions, this is the well-known error function which cannot be expressed analytically and therefore needs numerical approximation. In this work, three different approaches have been considered using the programs included in [StatLib] and [Press 1999]. In particular, the final calculations in the examples discussed in this thesis have been performed using an algorithm based on the approximation as given by [Hart et al. 1968]. Other approximations, as that of [Adams 1969, Hill 1973], yield similar results while the implementation of [Press 1999] was found to work less stable.

A.5 Recent Applications

[Hasselblad et al. 1980] was probably the first to use the expectation-maximisation algorithm for fitting one-component lognormal distributions to blood lead data from a large-scale screening program in the 1970's in New York city. His method was adapted by [McLaren et al. 1986a] for modelling doubly-truncated lognormal distributions in the analysis of red blood cell volume distributions relevant for detection of different forms of anemia [McLaren et al. 1986b, McLaren et al. 1987]. Basing on the algorithms proposed in this work, [McLachlan and Jones 1988] introduced the general theoretical framework to the analysis of finite-mixture distributions with particular respect to normal-type components (see also [Jones and McLachlan 1990]) which in the following lead to a number of applications in different fields of science. An efficient algorithm for parameter estimation of grouped normal mixtures based on the above results was published by [Jones and McLachlan 1990] in terms of a Fortran 77 program and is included in the StatLib library [StatLib].

In further extensions of the haematological studies, [McLaren et al. 1991] fitted two-component mixtures of lognormal distributions and examined the conditions under which a superposition may be detected sufficiently. The results were used to distinguish between blood data from healthy persons and patients with different forms of anemia [McLaren et al. 1991, McLaren et al. 1993], and for early diagnoses of hemochromatosis by modelling Transferrin saturation distributions [McLaren 1996, McLaren et al. 1998]. To statistically test the hypotheses of one and two components against each other (particularly relevant for the detection of anemia), [McLachlan et al. 1995] proposed an algorithm basing on a likelihood ratio test used by [McLaren et al. 2001b] for describing the relationship between Transferrin saturation and iron storage by testing mixtures of two and three components against each other for samples of male and female African American and US Caucasian populations.

Further applications of the EM algorithm for grouped data with particular respect to normal-type mixture models included the analysis of collagen fibril diameter distributions [Jones 1991], Phenylthiocarbamide (PTC) sensitivity [Jones and McLachlan 1991], and the survival of different bacteria after chilling and freezing in liquid environments [Smith 1995].

For the parallel analysis of red blood cell data including respective cell volume and haemoglobin concentration distributions given in terms of bivariate histograms, the method was extended to multivariate distributions [McLaren et al. 2001a, Cadez et al. 2002] whose parameter determination is (combined with different other statistical analysis and modelling approaches) relevant for hierarchical screenings for iron deficiency anemia [Cadez et al. 1999] and, more general, for patient-specific analysis of haematological data [McLaren et al. 2000].

In [Jones and McLachlan 1989], the modelling of mass-size particle data using the EM algorithm based on grouped data was discussed including mixture models consisting of lognormal, log-hyperbolic and log-skew Laplace components. This analysis is particularly relevant to geology as well as materials science.

Similar to the case of grouped data, the occurrence of data subjected to certain kinds of censoring may be relevant in application. [Dempster et al. 1977] pointed out that this problem may be handled in a very similar way as the grouping and truncating of data. Particular applications basing on the results for binned data include the detection of components in meteorological data [Sansom and Thomson 1998] and the assessment of anthropogenic arsenic background in nature [Portier 2001]. For the case of hyperbolic distributions, an extension of the method applied to censored multivariate data was recently proposed by [Protassov 2004].

Appendix B

Standard Errors Based on the Information Matrix

In Sect. 2.4.1, the equations for standard error estimates based on the information matrix of the EM estimator have been given. In the following, the details of this approach will be presented. For a proper estimation of the information matrices in applications, the so-called score statistics of the estimator is of particual importance.

B.1 The Score Statistics

If $\vec{x}$ and $\vec{y}$ are the observed and complete data of a given estimation problem, the gradient vectors of the observed and complete-data log-likelihood function,

$$\vec{s}(\vec{x};\mathbf{\Psi}) = \frac{\partial \log L(\mathbf{\Psi})}{\partial \mathbf{\Psi}} \quad \text{and} \quad \vec{s}_c(\vec{y};\mathbf{\Psi}) = \frac{\partial \log L_c(\mathbf{\Psi})}{\partial \mathbf{\Psi}}, \tag{B.1}$$

are referred to as the (incomplete- and complete-data) score statistics [McLachlan and Krishnan 1997]. The interrelationship between observed and complete-data likelihood directly transfers to the score statistics as

$$\begin{aligned}
\vec{s}(\vec{x};\mathbf{\Psi}) &= \frac{\partial \log p(\vec{x};\mathbf{\Psi})}{\partial \mathbf{\Psi}} = \frac{1}{p(\vec{x};\mathbf{\Psi})}\frac{\partial p(\vec{x};\mathbf{\Psi})}{\partial \mathbf{\Psi}} \\
&= \frac{1}{p(\vec{x};\mathbf{\Psi})}\int_{\{x_j\}} d\vec{y}\; \frac{\partial p_c(\vec{y};\mathbf{\Psi})}{\partial \mathbf{\Psi}} = \int_{\{x_j\}} d\vec{y}\; \frac{\partial \log p_c(\vec{y};\mathbf{\Psi})}{\partial \mathbf{\Psi}}\frac{p_c(\vec{y};\mathbf{\Psi})}{p(\vec{x};\mathbf{\Psi})} \\
&= \int_{\{x_j\}} d\vec{y}\; \frac{\partial \log L_c(\mathbf{\Psi})}{\partial \mathbf{\Psi}} f(\vec{y}|\vec{x};\mathbf{\Psi}) = E_{\mathbf{\Psi}}\left\{ \frac{\partial \log L_c(\mathbf{\Psi})}{\partial \mathbf{\Psi}} \middle| \vec{x} \right\} = E_{\mathbf{\Psi}}\left\{ \vec{s}_c(\vec{y};\mathbf{\Psi}) \middle| \vec{x} \right\}
\end{aligned} \tag{B.2}$$

where $p_c(\vec{y};\mathbf{\Psi})$ is the joint probability of the complete data.

In the case of explicit data, the likelihood function factorises such that the incomplete-data score statistics has the form

$$\vec{s}(\vec{x};\mathbf{\Psi}) = \sum_{j=1}^{J} \frac{\partial \log f(x_j;\mathbf{\Psi})}{\partial \mathbf{\Psi}} = \sum_{j=1}^{J} \vec{s}_j(\mathbf{\Psi}). \tag{B.3}$$

Note that the incomplete-data score statistics approaches zero when being evaluated at the maximum likelihood solution $\hat{\mathbf{\Psi}}$ by definition.

Consider now the case of grouped and truncated data. The calculation of derivatives of the log-likelihood function with respect to the different parameters corresponds to an evaluation of the corresponding derivatives of the expectation values of $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(l)})$ at $\mathbf{\Psi}^{(l)} = \mathbf{\Psi}$, i.e.,

$$\begin{aligned}
\left.\frac{\partial Q(\mathbf{\Psi}, \mathbf{\Psi}^{(l)})}{\partial \mathbf{\Psi}}\right|_{\mathbf{\Psi}^{(l)}=\mathbf{\Psi}} &= \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) E_{\mathbf{\Psi}^{(l)}} \left\{ \left. \frac{\partial \log f(x;\mathbf{\Psi})}{\partial \mathbf{\Psi}} \right| x \in X_m \right\}_{\mathbf{\Psi}^{(l)}=\mathbf{\Psi}} \\
&= \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}^{(l)}) \int_{X_m} dx \; \frac{f(x;\mathbf{\Psi}^{(l)})}{P_m(\mathbf{\Psi}^{(l)})} \left. \frac{\partial \log f(x;\mathbf{\Psi})}{\partial \mathbf{\Psi}} \right|_{\mathbf{\Psi}^{(l)}=\mathbf{\Psi}} \\
&= \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}) \frac{1}{P_m(\mathbf{\Psi})} \frac{\partial}{\partial \mathbf{\Psi}} \int_{X_m} dx \; f(x;\mathbf{\Psi}) \\
&= \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}) \frac{1}{P_m(\mathbf{\Psi})} \frac{\partial P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}} = \sum_{m=1}^{M+M'} n_m(\mathbf{\Psi}) \frac{\partial}{\partial \mathbf{\Psi}} \log P_m(\mathbf{\Psi})
\end{aligned} \tag{B.4}$$

To approach an expression including only the observed group frequencies, one may have a detailed look onto the part of the sum involving the truncated intervals

$$\begin{aligned}
\sum_{m=M+1}^{M+M'} n_m(\mathbf{\Psi}) \frac{\partial}{\partial \mathbf{\Psi}} \log P_m(\mathbf{\Psi}) &= \sum_{m=M+1}^{M+M'} n \frac{P_m(\mathbf{\Psi})}{P(\mathbf{\Psi})} \frac{1}{P_m(\mathbf{\Psi})} \frac{\partial P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}} \\
&= \frac{n}{P(\mathbf{\Psi})} \frac{\partial}{\partial \mathbf{\Psi}} \sum_{m=M+1}^{M+M'} P_m(\mathbf{\Psi}) = \frac{n}{P(\mathbf{\Psi})} \frac{\partial}{\partial \mathbf{\Psi}} \left(1 - P(\mathbf{\Psi})\right) \\
&= -\frac{n}{P(\mathbf{\Psi})} \frac{\partial P(\mathbf{\Psi})}{\partial \mathbf{\Psi}} = -n \frac{\partial \log P(\mathbf{\Psi})}{\partial \mathbf{\Psi}}
\end{aligned} \tag{B.5}$$

such that

$$\vec{s}(\vec{n};\mathbf{\Psi}) = \left. \frac{\partial Q(\mathbf{\Psi};\mathbf{\Psi}^{(l)})}{\partial \mathbf{\Psi}} \right|_{\mathbf{\Psi}^{(l)}=\mathbf{\Psi}} = \sum_{m=1}^{M} n_m \frac{\partial \log P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}} - n \frac{\partial \log P(\mathbf{\Psi})}{\partial \mathbf{\Psi}} \equiv \frac{\partial \log L(\mathbf{\Psi})}{\partial \mathbf{\Psi}}. \tag{B.6}$$

Furthermore, one has

$$n \frac{\partial \log P(\mathbf{\Psi})}{\partial \mathbf{\Psi}} = \frac{n}{P(\mathbf{\Psi})} \frac{\partial P(\mathbf{\Psi})}{\partial \mathbf{\Psi}} = \sum_{m=1}^{M} \frac{n}{P(\mathbf{\Psi})} \frac{\partial P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}} = \sum_{m=1}^{M} n \frac{P_m(\mathbf{\Psi})}{P(\mathbf{\Psi})} \frac{\partial \log P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}} \tag{B.7}$$

and therefore

$$\vec{s}(\vec{n};\mathbf{\Psi}) = \sum_{m=1}^{M} \left(n_m - n \frac{P_m(\mathbf{\Psi})}{P(\mathbf{\Psi})} \right) \frac{\partial \log P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}}. \tag{B.8}$$

It is useful to reformulate the latter result by rewriting the derivative of the $\log P_m(\mathbf{\Psi})$ as follows:

$$\begin{aligned}
\vec{h}_m(\mathbf{\Psi}) := \frac{\partial \log P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}} &= \frac{1}{P_m(\mathbf{\Psi})} \frac{\partial P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}} = \frac{1}{P_m(\mathbf{\Psi})} \int_{X_m} dx \; \frac{\partial f(x;\mathbf{\Psi})}{\partial \mathbf{\Psi}} \\
&= \int_{X_m} dx \; \frac{f(x;\mathbf{\Psi})}{P_m(\mathbf{\Psi})} \frac{\partial \log f(x;\mathbf{\Psi})}{\partial \mathbf{\Psi}} = E_{\mathbf{\Psi}} \left\{ \left. \frac{\partial \log f(x;\mathbf{\Psi})}{\partial \mathbf{\Psi}} \right| x \in X_m \right\}.
\end{aligned} \tag{B.9}$$

In case of a K-finite mixture distribution, one may evaluate the involved distribution function explicitly in terms of the component densities by using the linearity of differentiation, integration, and expectation and taking the conditional probabilities into account yielding

$$\begin{aligned}\vec{h}_m(\mathbf{\Psi}) &= \frac{1}{P_m(\mathbf{\Psi})}\sum_{i=1}^{K}\int_{X_m} dx\ \frac{\partial}{\partial\mathbf{\Psi}}\left(\pi_i f_i(x;\mathbf{\Psi})\right)\\ &= \sum_{i=1}^{K}\int_{X_m} dx\ \frac{f(x;\mathbf{\Psi})}{P_m(\mathbf{\Psi})}\frac{\pi_i f_i(x;\mathbf{\Psi})}{f(x;\mathbf{\Psi})}\frac{\partial}{\partial\mathbf{\Psi}}\log\left(\pi_i f_i(x;\mathbf{\Psi})\right)\\ &= \sum_{i=1}^{K}\int_{X_m} dx\ \frac{f(x;\mathbf{\Psi})}{P_m(\mathbf{\Psi})}t_i(x;\mathbf{\Psi})\frac{\partial}{\partial\mathbf{\Psi}}\log\left(\pi_i f_i(x;\mathbf{\Psi})\right)\\ &= \sum_{i=1}^{K} E_{\mathbf{\Psi}}\left\{ t_i(x;\mathbf{\Psi})\frac{\partial}{\partial\mathbf{\Psi}}\log\left(\pi_i f_i(x;\mathbf{\Psi})\right)\middle| x\in X_m\right\}\end{aligned} \tag{B.10}$$

As it is shown below, this result is important for the calculation of information-based standard parameter errors.

B.2 Conditional Information Matrix

The negative Hessian (i.e., the matrix of the second partial derivatives) of the observed-data log-likelihood function is usually referred to as the observed information matrix. For the general case, an explicit expression for this matrix has been given by [Louis 1982] derived as follows: Firstly, one evaluates

$$\begin{aligned}I(\mathbf{\Psi};\vec{x}) &= -\frac{\partial^2\log L(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} = -\frac{\partial}{\partial\mathbf{\Psi}}\left(\frac{1}{p(\vec{x};\mathbf{\Psi})}\frac{\partial p(\vec{x};\mathbf{\Psi})}{\partial\mathbf{\Psi}^T}\right)\\ &= -\frac{1}{p(\vec{x};\mathbf{\Psi})}\frac{\partial^2 p(\vec{x};\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} + \frac{1}{p(\vec{x};\mathbf{\Psi})}\frac{\partial p(\vec{x};\mathbf{\Psi})}{\partial\mathbf{\Psi}}\cdot\frac{1}{p(\vec{x};\mathbf{\Psi})}\frac{\partial p(\vec{x};\mathbf{\Psi})}{\partial\mathbf{\Psi}^T}.\end{aligned} \tag{B.11}$$

While the second term can be easily expressed in terms of the incomplete-data score statistic as

$$\frac{1}{p(\vec{x};\mathbf{\Psi})}\frac{\partial p(\vec{x};\mathbf{\Psi})}{\partial\mathbf{\Psi}}\cdot\frac{1}{p(\vec{x};\mathbf{\Psi})}\frac{\partial p(\vec{x};\mathbf{\Psi})}{\partial\mathbf{\Psi}^T} = \vec{s}(\vec{x};\mathbf{\Psi})\cdot\vec{s}^T(\vec{x};\mathbf{\Psi}), \tag{B.12}$$

the first contribution needs further reformulation in terms of the complete-data likelihood similar to the above considerations for the corresponding score statistic as follows:

$$\begin{aligned}\frac{1}{p(\vec{x};\mathbf{\Psi})}\frac{\partial^2 p(\vec{x};\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} &= \frac{1}{p(\vec{x};\mathbf{\Psi})}\int_{\{x_j\}} d\vec{y}\ \frac{\partial^2 p_c(\vec{y};\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} = \int_{\{x_j\}} d\vec{y}\ \frac{p_c(\vec{y};\mathbf{\Psi})}{p(\vec{x};\mathbf{\Psi})}\frac{1}{p_c(\vec{y};\mathbf{\Psi})}\frac{\partial^2 p_c(\vec{y};\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T}\\ &= \int_{\{x_j\}} d\vec{y}\ f(\vec{y}|\vec{x};\mathbf{\Psi})\frac{1}{L_c(\mathbf{\Psi})}\frac{\partial^2 L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} = E_{\mathbf{\Psi}}\left\{\frac{1}{L_c(\mathbf{\Psi})}\frac{\partial^2 L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T}\middle|\vec{x}\right\}\\ &= E_{\mathbf{\Psi}}\left\{\frac{\partial^2\log L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} + \frac{\partial\log L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\cdot\frac{\partial\log L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T}\middle|\vec{x}\right\}\end{aligned} \tag{B.13}$$

where the last identity follows from

$$\frac{\partial^2\log L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} = \frac{\partial}{\partial\mathbf{\Psi}}\frac{1}{L_c(\mathbf{\Psi})}\frac{\partial L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T} = \frac{1}{L_c(\mathbf{\Psi})}\frac{\partial^2 L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} - \frac{1}{L_c(\mathbf{\Psi})}\frac{\partial L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\cdot\frac{1}{L_c(\mathbf{\Psi})}\frac{\partial L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T}. \tag{B.14}$$

Again, one identifies the complete-data score statistic and the complete-data information matrix

$$I_c(\mathbf{\Psi}; \vec{y}) = -\frac{\partial^2 \log L_c(\mathbf{\Psi})}{\partial \mathbf{\Psi} \partial \mathbf{\Psi}^T}. \tag{B.15}$$

In brief notation, it follows that

$$I(\mathbf{\Psi}; \vec{x}) = E_{\mathbf{\Psi}} \left\{ I_c(\mathbf{\Psi}; \vec{y}) \middle| \vec{x} \right\} - E_{\mathbf{\Psi}} \left\{ \vec{s}_c(\vec{y}; \mathbf{\Psi}) \cdot \vec{s}_c^T(\vec{y}; \mathbf{\Psi}) \middle| \vec{x} \right\} + \vec{s}(\vec{x}; \mathbf{\Psi}) \cdot \vec{s}^T(\vec{x}; \mathbf{\Psi}). \tag{B.16}$$

Defining the conditional expected complete-data information matrix,

$$\mathcal{I}_c(\mathbf{\Psi}; \vec{x}) = E_{\mathbf{\Psi}} \left\{ I_c(\mathbf{\Psi}; \vec{y}) \middle| \vec{x} \right\}, \tag{B.17}$$

and the expected information matrix for the distribution of $\vec{y}$ conditional on $\vec{x}$ (usually referred to as the missing information matrix which corresponds to the covariance matrix of the complete-data score statistics),

$$\begin{aligned} \mathcal{I}_m(\mathbf{\Psi}; \vec{x}) &= E_{\mathbf{\Psi}} \left\{ \vec{s}_c(\vec{y}; \mathbf{\Psi}) \cdot \vec{s}_c^T(\vec{y}; \mathbf{\Psi}) \middle| \vec{x} \right\} - \vec{s}(\vec{x}; \mathbf{\Psi}) \cdot \vec{s}^T(\vec{x}; \mathbf{\Psi}) \\ &= E_{\mathbf{\Psi}} \left\{ \vec{s}_c(\vec{y}; \mathbf{\Psi}) \cdot \vec{s}_c^T(\vec{y}; \mathbf{\Psi}) \middle| \vec{x} \right\} - E_{\mathbf{\Psi}} \left\{ \vec{s}_c(\vec{y}; \mathbf{\Psi}) \middle| \vec{x} \right\} \cdot E_{\mathbf{\Psi}} \left\{ \vec{s}_c(\vec{y}; \mathbf{\Psi}) \middle| \vec{x} \right\}^T, \end{aligned} \tag{B.18}$$

the expression may be rewritten as

$$I(\mathbf{\Psi}; \vec{x}) = \mathcal{I}_c(\mathbf{\Psi}; \vec{x}) - \mathcal{I}_m(\mathbf{\Psi}; \vec{x}) \tag{B.19}$$

yielding the famous missing information principle of [Orchard and Woodbury 1972].

B.3 Unconditional Information Matrix

Until this point, all information matrices considered have been conditional with respect to the observed-data vector $\vec{x}$. In practical applications, one is interested in more global expressions not involving the particular observations. For this purpose, one may define the expected incomplete-data information matrix

$$\mathcal{I}(\mathbf{\Psi}) = E_{\mathbf{\Psi}} \left\{ I(\mathbf{\Psi}; x) \right\} \tag{B.20}$$

and the expected complete-data information matrix

$$\mathcal{I}_c(\mathbf{\Psi}) = E_{\mathbf{\Psi}} \left\{ I_c(\mathbf{\Psi}; y) \right\} \tag{B.21}$$

[McLachlan and Basford 1988, McLachlan and Peel 2000]. Note that here, x and y occur as independent variables and not as particular representations thereof anymore. From

$$p(\vec{x}; \mathbf{\Psi}) = \frac{p_c(\vec{y}; \mathbf{\Psi})}{f(\vec{y}|\vec{x}; \mathbf{\Psi})}, \tag{B.22}$$

it follows that

$$\log L(\mathbf{\Psi}) = \log L_c(\mathbf{\Psi}) - \log f(\vec{y}|\vec{x}; \mathbf{\Psi}) \tag{B.23}$$

and therefore, by computing the second derivatives,

$$I(\mathbf{\Psi}; \vec{x}) = I_c(\mathbf{\Psi}; \vec{y}) + \frac{\partial^2 \log f(\vec{y}|\vec{x}; \mathbf{\Psi})}{\partial \mathbf{\Psi} \partial \mathbf{\Psi}^T}. \tag{B.24}$$

Taking the conditional expectations with respect to the observed data $\vec{x}$ then yields

$$I(\mathbf{\Psi};\vec{x}) = \mathcal{I}_c(\mathbf{\Psi};\vec{x}) - \mathcal{I}_m(\mathbf{\Psi};\vec{x}) \quad \text{with} \quad \mathcal{I}_m(\mathbf{\Psi};\vec{x}) = -E_{\mathbf{\Psi}}\left\{ \frac{\partial^2 \log f(\vec{y}|\vec{x};\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T}\middle| \vec{x}\right\} \tag{B.25}$$

Finally, by taking the expectation over the distribution of $\{x\}$, one approaches

$$\mathcal{I}(\mathbf{\Psi}) = \mathcal{I}_c(\mathbf{\Psi}) - E_{\mathbf{\Psi}}\left\{\mathcal{I}_m(\mathbf{\Psi};\vec{x})\right\}. \tag{B.26}$$

Under proper regularity conditions,

$$E_{\mathbf{\Psi}}\left\{\mathcal{I}_c(\mathbf{\Psi};\vec{x}) - E_{\mathbf{\Psi}}\left\{\vec{s}_c(\vec{y};\mathbf{\Psi})\cdot\vec{s}_c^{\,T}(\vec{y};\mathbf{\Psi})\middle|\vec{x}\right\}\right\} = 0. \tag{B.27}$$

Hence, $E_{\mathbf{\Psi}}\left\{\vec{s}_c(\vec{y};\mathbf{\Psi})\cdot\vec{s}_c^{\,T}(\vec{y};\mathbf{\Psi})\middle|\vec{x}\right\}$ is a bias-free estimator for the conditional expected complete-data information matrix, and

$$\mathcal{I}(\mathbf{\Psi}) = E_{\mathbf{\Psi}}\left\{\vec{s}(\vec{x};\mathbf{\Psi})\cdot\vec{s}^{\,T}(\vec{x};\mathbf{\Psi})\right\} \tag{B.28}$$

such that $\vec{s}(\vec{x};\mathbf{\Psi})\cdot\vec{s}^{\,T}(\vec{x};\mathbf{\Psi})$ is a bias-free estimator of $I(\mathbf{\Psi};\vec{x})$ as well.

B.4 Score Covariance Matrix

The values of $I(\mathbf{\Psi};\vec{x})$ and $\mathcal{I}(\mathbf{\Psi})$ at the maximum likelihood solution $\hat{\mathbf{\Psi}}$ are usually referred to as the observed and expected Fisher information matrices [Fisher 1925, Efron and Hinkley 1978]. Evaluating (B.16) at $\hat{\mathbf{\Psi}}$, the term directly involving the incomplete-data score statistics vanishes such that

$$I(\hat{\mathbf{\Psi}};\vec{x}) = E_{\mathbf{\Psi}}\left\{ I_c(\hat{\mathbf{\Psi}};\vec{y})\middle|\vec{x}\right\} - E_{\mathbf{\Psi}}\left\{\vec{s}_c(\vec{y};\hat{\mathbf{\Psi}})\cdot\vec{s}_c^{\,T}(\vec{y};\hat{\mathbf{\Psi}})\cdot\middle|\vec{x}\right\}. \tag{B.29}$$

Hence, the observed Fisher information matrix may be calculated only in terms of the gradient and curvature of the complete-data log-likelihood function as

$$I(\hat{\mathbf{\Psi}};\vec{x}) = -E_{\mathbf{\Psi}}\left\{ \frac{\partial^2 \log L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} + \frac{\partial \log L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\frac{\partial \log L_c(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T}\middle|\vec{x}\right\}\Bigg|_{\mathbf{\Psi}=\hat{\mathbf{\Psi}}}. \tag{B.30}$$

As for pairwise independent explicitly given data, the both score statistics and information matrices separate with respect to the particular observations, it is a common approximation to estimate the observed Fisher information by the observed-data score covariance matrix ([McLachlan and Basford 1988])

$$I_s(\hat{\mathbf{\Psi}};\vec{x}) = \sum_{j=1}^{J}\vec{s}_j(\hat{\mathbf{\Psi}})\vec{s}_j^{\,T}(\hat{\mathbf{\Psi}}). \tag{B.31}$$

This approach is additionally motivated by the fact that

$$\begin{aligned} I(\mathbf{\Psi};\vec{x}) &= -\frac{\partial^2 \log L(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} = -\sum_{j=1}^{J}\frac{\partial^2 \log f(x_j;\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} \\ &= \sum_{j=1}^{J}\left(\frac{\partial \log f(x_j;\mathbf{\Psi})}{\partial\mathbf{\Psi}}\frac{\partial \log f(x_j;\mathbf{\Psi})}{\partial\mathbf{\Psi}^T} - \frac{1}{f(x_j;\mathbf{\Psi})}\frac{\partial^2 f(x_j;\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T}\right) \end{aligned} \tag{B.32}$$

where the second term has zero expectation such that

$$\mathcal{I}(\mathbf{\Psi}) = E_{\mathbf{\Psi}} \left\{ \sum_{j=1}^{J} \frac{\partial \log f(x_j; \mathbf{\Psi})}{\partial \mathbf{\Psi}} \frac{\partial \log f(x_j; \mathbf{\Psi})}{\partial \mathbf{\Psi}^T} \right\} = E_{\mathbf{\Psi}} \left\{ \sum_{j=1}^{J} \vec{s}_j(\mathbf{\Psi}) \cdot \vec{s}_j^T(\mathbf{\Psi}) \right\}. \tag{B.33}$$

yields a valid expression of the expected Fisher information matrix (see [Behboodian 1972, Berndt et al. 1974, Redner and Walker 1984]). From the last identity, it follows that (B.31) is a bias-free estimator of $I(\mathbf{\Psi}; \vec{x})$.

B.5 Empirical Covariance Matrix

For the case of independent identically distributed data, [Meilijson 1989] pointed out that it may be preferable to use the empirical covariance matrix of the data for describing the variance of the estimated parameters rather than the standard Fisher information. In this configuration, the score statistics are sums over contributions from the different single data, and the expected information matrix may be written in similar form as

$$\mathcal{I}(\mathbf{\Psi}) = J i(\mathbf{\Psi}) \tag{B.34}$$

where

$$i(\mathbf{\Psi}) = E_{\mathbf{\Psi}} \left\{ \vec{s}(x; \mathbf{\Psi}) \cdot \vec{s}^T(x; \mathbf{\Psi}) \right\} \tag{B.35}$$

is the covariance matrix of a single observation. As an empirical representation of this covariance matrix in terms of the observed data (the empirical covariance matrix), one may calculate

$$\bar{i}(\mathbf{\Psi}) = \frac{1}{J} \sum_{j=1}^{J} \vec{s}_j(\mathbf{\Psi}) \cdot \vec{s}_j^T(\mathbf{\Psi}) - \bar{\vec{s}}(\mathbf{\Psi}) \cdot \bar{\vec{s}}^T(\mathbf{\Psi}) \tag{B.36}$$

with

$$\bar{\vec{s}}(\mathbf{\Psi}) = \frac{1}{J} \sum_{j=1}^{J} \vec{s}_j(\mathbf{\Psi}) \tag{B.37}$$

such that

$$I_e(\mathbf{\Psi}; \vec{x}) = J\bar{i}(\mathbf{\Psi}) = \sum_{j=1}^{J} \vec{s}_j(\mathbf{\Psi}) \cdot \vec{s}_j^T(\mathbf{\Psi}) - \frac{1}{J} \vec{s}(\vec{x}; \mathbf{\Psi}) \cdot \vec{s}^T(\vec{x}; \mathbf{\Psi}). \tag{B.38}$$

As this consistent estimator of the information matrix is computed rather easily only by calculating sums over the observed-data score statistics, it may be efficiently used in terms of the EM algorithm. In particular, when being evaluated at the maximum likelihood solution, $\bar{I}(\mathbf{\Psi}; \vec{x})$ reduces to $\hat{I}(\hat{\mathbf{\Psi}})$ as defined in (B.31).

B.6 Covariance Matrices for Grouped and Truncated Data

The formulation of the information matrix in terms of grouped and truncated data has been explicitly considered by [Jones and McLachlan 1992]. From

$$\log L(\mathbf{\Psi}) = \sum_{m=1}^{M} n_m \log P_m(\mathbf{\Psi}) - n \log P(\mathbf{\Psi}), \tag{B.39}$$

the computation of the second partial derivatives immediately yields

$$I(\mathbf{\Psi};\vec{n}) = -\sum_{m=1}^{M} n_m \frac{\partial^2 \log P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} + n\frac{\partial^2 \log P(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T}. \tag{B.40}$$

The explicit evaluation of the derivatives leads to the following expression:

$$\begin{aligned}
I(\mathbf{\Psi};\vec{n}) =& -\sum_{m=1}^{M} n_m \frac{\partial}{\partial\mathbf{\Psi}}\left(\frac{1}{P_m(\mathbf{\Psi})}\frac{\partial P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T}\right) + n\frac{\partial}{\partial\mathbf{\Psi}}\left(\frac{1}{P(\mathbf{\Psi})}\frac{\partial P(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T}\right)\\
=& -\sum_{m=1}^{M} n_m \left(-\frac{1}{P_m^2(\mathbf{\Psi})}\frac{\partial P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\frac{\partial P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T} + \frac{1}{P_m(\mathbf{\Psi})}\frac{\partial^2 P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T}\right)\\
&+ n\left(-\frac{1}{P^2(\mathbf{\Psi})}\frac{\partial P(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\frac{\partial P(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T} + \frac{1}{P(\mathbf{\Psi})}\frac{\partial^2 P(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T}\right)\\
=& \left(\sum_{m=1}^{M}\frac{n_m}{P_m^2(\mathbf{\Psi})}\frac{\partial P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\frac{\partial P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T} - \frac{n}{P^2(\mathbf{\Psi})}\frac{\partial P(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\frac{\partial P(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T}\right)\\
&- \left(\sum_{m=1}^{M}\frac{n_m}{P_m(\mathbf{\Psi})}\frac{\partial^2 P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T} - \frac{n}{P(\mathbf{\Psi})}\frac{\partial^2 P(\mathbf{\Psi})}{\partial\mathbf{\Psi}\partial\mathbf{\Psi}^T}\right)\\
=& I_s(\mathbf{\Psi};\vec{n}) - R(\mathbf{\Psi};\vec{n})
\end{aligned} \tag{B.41}$$

As the second term has zero expectation, $I_s(\mathbf{\Psi};\vec{n})$ is a bias-free estimator for the observed-data information matrix, and $I_s(\mathbf{\Psi};\vec{n})/n$ is also consistent [Jones and McLachlan 1992].
For further calculations, the explicit structure of the information matrix estimator has to be computed. This may be done analogously to the above considerations concerning the corresponding score statistics as

$$I_s(\mathbf{\Psi};\vec{n}) = \sum_{m=1}^{M} n_m \frac{\partial \log P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\frac{\partial \log P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T} - n\frac{\partial \log P(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\frac{\partial \log P(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T} \tag{B.42}$$

and

$$\begin{aligned}
\frac{n}{P^2(\mathbf{\Psi})}\frac{\partial P(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\frac{\partial P(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T} &= n\left(\sum_{m=1}^{M}\frac{1}{P(\mathbf{\Psi})}\frac{\partial P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\right)\left(\sum_{m=1}^{M}\frac{1}{P(\mathbf{\Psi})}\frac{\partial P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T}\right)\\
&= n\left(\sum_{m=1}^{M}\frac{P_m(\mathbf{\Psi})}{P(\mathbf{\Psi})}\frac{\partial \log P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}}\right)\left(\sum_{m=1}^{M}\frac{P_m(\mathbf{\Psi})}{P(\mathbf{\Psi})}\frac{\partial \log P_m(\mathbf{\Psi})}{\partial\mathbf{\Psi}^T}\right)\\
&= n\left(\sum_{m=1}^{M}\frac{P_m(\mathbf{\Psi})}{P(\mathbf{\Psi})}\vec{h}_m(\mathbf{\Psi})\right)\left(\sum_{m=1}^{M}\frac{P_m(\mathbf{\Psi})}{P(\mathbf{\Psi})}\vec{h}_m^T(\mathbf{\Psi})\right) =: n\bar{\vec{h}}(\mathbf{\Psi})\bar{\vec{h}}^T(\mathbf{\Psi})
\end{aligned} \tag{B.43}$$

such that

$$I_s(\mathbf{\Psi};\vec{n}) = \sum_{m=1}^{M} n_m\vec{h}_m(\mathbf{\Psi})\vec{h}_m^T(\mathbf{\Psi}) - n\bar{\vec{h}}(\mathbf{\Psi})\bar{\vec{h}}^T(\mathbf{\Psi}). \tag{B.44}$$

To find an analogue for (B.31) in terms of grouped truncated data, one may use the identity

$$0 = \vec{s}(\vec{n};\hat{\mathbf{\Psi}}) = \sum_{m=1}^{M} n_m\vec{h}_m(\hat{\mathbf{\Psi}}) - n\bar{\vec{h}}(\hat{\mathbf{\Psi}}) \tag{B.45}$$

to proove that

$$\begin{aligned} I_e(\mathbf{\Psi};\vec{n}) &= \sum_{m=1}^{M} n_m \left(\vec{h}_m(\hat{\mathbf{\Psi}}) - \bar{\vec{h}}(\hat{\mathbf{\Psi}})\right)\left(\vec{h}_m(\hat{\mathbf{\Psi}}) - \bar{\vec{h}}(\hat{\mathbf{\Psi}})\right)^T \\ &= \sum_{m=1}^{M} n_m \vec{h}_m(\hat{\mathbf{\Psi}})\vec{h}_m^T(\hat{\mathbf{\Psi}}) - n\bar{\vec{h}}(\hat{\mathbf{\Psi}})\bar{\vec{h}}^T(\hat{\mathbf{\Psi}}) = K(\mathbf{\Psi};\vec{n})|_{\mathbf{\Psi}=\hat{\mathbf{\Psi}}}\,. \end{aligned} \tag{B.46}$$

To calculate this quantity without explicitly considering $P(\hat{\mathbf{\Psi}})$, one may make use of

$$\bar{\vec{h}}(\hat{\mathbf{\Psi}}) = \sum_{m=1}^{M} \frac{P_m(\hat{\mathbf{\Psi}})}{P(\hat{\mathbf{\Psi}})}\vec{h}_m(\hat{\mathbf{\Psi}}) = \sum_{m=1}^{M} \frac{n_m}{n}\vec{h}_m(\hat{\mathbf{\Psi}}) \tag{B.47}$$

which follows directly from (B.45).

As for grouped data, it is implicitly assumed that the single events are independent and identically distributed, the formalism of empirical covariance matrix may be adopted to this case as well. The corresponding expression then reads as follows [Jones and McLachlan 1992]:

$$I_e(\mathbf{\Psi};\vec{n}) = \sum_{m=1}^{M} n_m \frac{\partial \log P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}} \frac{\partial \log P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}^T} - n\left(\sum_{m=1}^{M} \frac{n_m}{n} \frac{\partial \log P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}}\right)\left(\sum_{m=1}^{M} \frac{n_m}{n} \frac{\partial \log P_m(\mathbf{\Psi})}{\partial \mathbf{\Psi}^T}\right) \tag{B.48}$$

This equation is completely equal to (B.44) when evaluated at the maximum likelihood solution which follows directly from (B.47).

B.7 Information-based Standard Errors

For the analytical computation of information-based standard errors, one should first note that the squared standard errors (corresponding to the variances of the respective parameters) are computed as diagonal elements of the asymptotic covariance matrix of the parameters computed at the ML solution. This matrix is under proper regularity conditions sufficiently approximated by the inverse of the expected Fisher complete-data information matrix. [Efron and Hinkley 1978] noted however that, at least in the case of one-parameter families of distributions, it is more useful to approximate the asymptotic covariance matrix by the inverse of the observed-data Fisher matrix rather than than the expected one. In this case, the *standard error* of a single parameter Ψ_i is given by

$$SE(\hat{\Psi}_i) \approx \sqrt{(I^{-1}(\hat{\mathbf{\Psi}};\vec{x}))_{ii}}. \tag{B.49}$$

From the above considerations, it follows that there are different levels of approximation to the observed Fisher matrix. The direct evaluation of the matrix by computing the second derivatives of $\log L(\mathbf{\Psi})$ with respect to the parameters gives the exact result, but may be computationally inefficient because it might be hard to give analytical expressions for this purpose. Problems may arise as well when expanding the Fisher information matrix in terms of gradient and curvature of the complete-data log-likelihood function as (B.30). In contrast, a very simple bias-free estimate of the Fisher information is given by the variance of the incomplete-data score statistics which involves only the likelihood gradient of the observations (B.31) without significant loss of accuracy ([Griffiths et al. 1987]) and may therefore be efficiently computed. The latter estimator may be equivalently replaced by the corresponding empirical covariance

matrix. In App. B.8, the equations for the score covariance matrices for grouped truncated data from Gaussian mixtures are explicitly derived. The results for some test data briefly discussed in App. C.

B.8 Grouped Truncated Data from Gaussian Mixtures

In general, the number of unknown model parameters in a K-finite mixture model is

$$N_m(\mathbf{\Psi};\vec{n}) = K + \sum_{i=1}^{K} N_i \tag{B.50}$$

where N_i is the number of unknown parameters of the i-th component function $f_i(x;\theta_i)$ (i.e., the dimension of θ_i). The first contribution K occurs due to the unknown statistical weight of each component. However, remember that the statistical weights are not independent as $\sum_{i=1}^{K}\pi_i = 1$. An information matrix involving all N_m parameters has a rank of only $N_m - 1$, which would mean non-invertability in this case.

To avoid the corresponding problems, one has to consider a reduced parameter vector $\mathbf{\Psi}^{[i]}$ equaling $\mathbf{\Psi}$ with the π_i component left out. In this case, this component is expressed by all π_j with $j \neq i$ according to

$$\begin{aligned} f(x;\mathbf{\Psi}) &= \sum_{i=1}^{K} \pi_i f_i(x;\mathbf{\Theta}_i) \\ &= \sum_{i\neq j} \pi_i f_i(x;\mathbf{\Theta}_i) + \left(1 - \sum_{i\neq j}\pi_i\right) f_j(x;\mathbf{\Theta}_j), \end{aligned} \tag{B.51}$$

which leads to additional terms in the corresponding score vector

$$\begin{aligned} \vec{h}_m^{\pi_k} &= E_{\vec{\Psi}}\left\{ t_k(x;\vec{\Psi})\frac{1}{\pi_k} - t_j(x;\vec{\Psi})\frac{1}{\pi_j}\middle| x \in X_m\right\} \\ &= \frac{1}{\pi_k}\frac{\int_{a_m}^{b_m} dx\ \pi_k f_k(x;\vec{\theta}_k)}{\int_{a_m}^{b_m} dx\ f(x;\vec{\Psi})} - \frac{1}{\pi_j}\frac{\int_{a_m}^{b_m} dx\ \pi_j f_j(x;\vec{\theta}_k)}{\int_{a_m}^{b_m} dx\ f(x;\vec{\Psi})} = \frac{\Delta_m F_k - \Delta_m F_j}{\Delta_m F} \end{aligned} \tag{B.52}$$

for $k \neq j$.

The contributions with respect to the particular component parameters do not explicitly involve the constraint and are thus computed straightforward as

$$\begin{aligned} h_m^{(\mu_k)} &= E_{\vec{\Psi}}\left\{ t_k(x;\vec{\Psi})\frac{(x-\mu_k)}{\sigma_k^2}\middle| x \in X_m\right\} \\ &= \frac{1}{\sigma_k^2} E_{\vec{\Psi}}\left\{ x\ t_k(x;\vec{\Psi})\middle| x \in X_m\right\} - \frac{\mu_k}{\sigma_k^2} E_{\vec{\Psi}}\left\{ t_k(x;\vec{\Psi})\middle| x \in X_m\right\} \end{aligned} \tag{B.53}$$

and

$$\begin{aligned} h_m^{(\sigma_k^2)} &= E_{\vec{\Psi}}\left\{ t_k(x;\vec{\Psi})\left(\frac{(x-\mu_k)^2}{2\sigma_k^4} - \frac{1}{2\sigma_k^2}\right)\middle| x \in X_m\right\} \\ &= \frac{1}{2\sigma_k^4} E_{\vec{\Psi}}\left\{ x^2\ t_k(x;\vec{\Psi})\middle| x \in X_m\right\} - \frac{\mu_k}{\sigma_k^4} E_{\vec{\Psi}}\left\{ x\ t_k(x;\vec{\Psi})\middle| x \in X_m\right\} \\ &\quad + \frac{1}{2\sigma_k^2}\left(\frac{\mu_k^2}{\sigma_k^2} - 1\right) E_{\vec{\Psi}}\left\{ t_k(x;\vec{\Psi})\middle| x \in X_m\right\}. \end{aligned} \tag{B.54}$$

Using the expressions for the respective expectation values as derived above, after some algebraic simplifications one ends up with

$$h_m^{(\mu_k)} = \pi_k \frac{\Delta_m f_k}{\Delta_m F} \tag{B.55}$$

$$h_m^{(\sigma_k^2)} = \frac{\pi_k \mu_k}{2\sigma_k^2} \frac{\Delta_m f_k}{\Delta_m F} - \frac{\pi_k}{2\sigma_k^2} \frac{\Delta_m \phi_k}{\Delta_m F}. \tag{B.56}$$

which allows the computation of $I_s(\vec{\Psi}; \vec{n})$ by using (B.46).

Appendix C

Real-World Examples of Grouped Data

Before applying all codes to the geoscientific problem discussed in this thesis, the performance of the EM algorithm for grouped and truncated data from Gaussian mixture models as described above has been tested on different real-world data sets which have been subjected to corresponding analyses in the statistical literature yet.

C.1 Mixtures of Normal Distributions

As a first example, one may recall the length distribution of trypanosome, a parasitic protozoon. The corresponding data set (see Tab. C.1) was originally studied by [Pearson 1914] and later discussed in the text book of [Everitt and Hand 1981] on finite mixture distributions. The results of the implementation of the EM algorithm used in this thesis are shown in Fig. C.1 and listed below:

$$\begin{aligned}
\pi_1 &= 0.67 \pm 0.19 \ (0.65) \\
\pi_2 &= 0.33 \pm 0.10 \ (0.35) \\
\mu_1 &= 19.99 \pm 0.25 \ (19.96) \\
\mu_2 &= 26.31 \pm 0.67 \ (26.16) \\
\sigma_1 &= 2.25 \pm 0.72 \ (2.15) \\
\sigma_2 &= 2.78 \pm 2.08 \ (2.76)
\end{aligned}$$

(the $\pm$ values give the standard errors estimated with the information matrix method). In comparison to [Everitt and Hand 1981] whose values are given in brackets, the parameter values for the minor component are slightly shifted which is related to the fact that in the presented implementation, the improved adaptation of [McLachlan and Jones 1988] to grouped and truncated data was used resulting in a remarkable lower χ^2 value (tails are not taken into account here) of 7.71 (8.96). This improvement is particularly seen at the point where the predominance of the components changes. Here, the new estimate fits the data histogram clearly better than the old pdf.

Another typical example are the ash content data (see Table C.2) of [Hald 1952] which have been analysed by [Hasselblad 1966] using the steepest descent approximation of the maximum likelihood solution. In this example, application of the EM algorithm does not lead to an significant improvement of the fit quality with respect to Hasselbald's approximation as the corresponding differences of the χ^2 values are within the range of numerical errors (EM: 5.47,

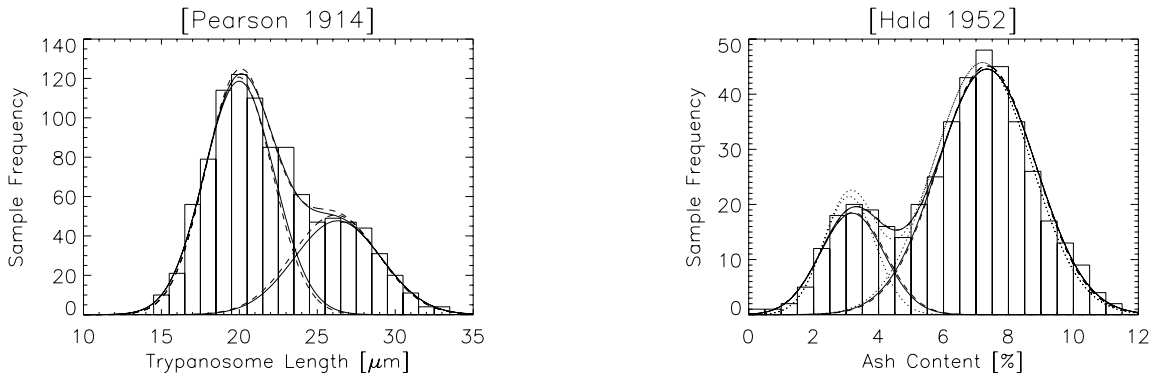


Figure C.1: Left panel: Trypanosome length data (histogram) of [Pearson 1914], the probability distribution function fitted by [Everitt and Hand 1981] (dotted line), and the PDF obtained with the new parameter estimates (solid line). Right panel: Ash content data (histogram) [Hald 1952], and the PDF resulting from the original estimate (dotted line), the steepest descent ML method (dashed line) [Hasselblad 1966], and the EM algorithm for grouped truncated data (solid line).

l_{min}	l_{max}	Observation	EH1981	EM
14.5μm	15.5μm	10	8.77	10.35
15.5μm	16.5μm	21	22.65	24.88
16.5μm	17.5μm	56	47.34	49.27
17.5μm	18.5μm	79	80.14	80.41
18.5μm	19.5μm	114	110.19	108.39
19.5μm	20.5μm	122	123.79	121.34
20.5μm	21.5μm	110	115.49	114.36
21.5μm	22.5μm	85	93.18	93.85
22.5μm	23.5μm	85	71.09	72.20
23.5μm	24.5μm	61	58.27	58.25
24.5μm	25.5μm	47	54.15	52.66
25.5μm	26.5μm	49	52.71	50.60
26.5μm	27.5μm	47	48.68	46.99
27.5μm	28.5μm	44	40.51	39.76
28.5μm	29.5μm	31	29.82	29.94
29.5μm	30.5μm	20	19.32	19.91
30.5μm	31.5μm	11	10.99	11.66
31.5μm	32.5μm	4	5.50	6.01
32.5μm	33.5μm	4	2.41	2.73

Table C.1: Trypanosome length data of [Pearson 1914], group frequencies estimated by [Everitt and Hand 1981] (EH1981), and the result of the EM algorithm for grouped truncated data.

Ash Content [%]	Ash Content [%]	Observation	H1952	H1966	EM
0.0	0.5	1	0.05	0.25	0.23
0.5	1.0	1	0.33	0.95	0.91
1.0	1.5	2	1.60	2.81	2.76
1.5	2.0	5	5.38	6.50	6.48
2.0	2.5	12	12.42	11.83	11.87
2.5	3.0	18	19.81	16.96	17.03
3.0	3.5	20	22.21	19.41	19.41
3.5	4.0	19	18.64	18.45	18.31
4.0	4.5	16	14.46	16.15	16.01
4.5	5.0	14	14.83	15.81	15.84
5.0	5.5	20	20.35	19.35	19.53
5.5	6.0	25	28.77	26.35	26.49
6.0	6.5	35	37.34	34.67	34.58
6.5	7.0	43	43.55	41.60	41.23
7.0	7.5	48	45.51	44.84	44.28
7.5	8.0	45	42.60	43.25	42.74
8.0	8.5	35	35.72	37.32	37.06
8.5	9.0	26	26.83	28.80	28.82
9.0	9.5	17	18.05	19.88	20.18
9.5	10.0	13	10.88	12.27	12.67
10.0	10.5	9	5.87	6.78	7.15
10.5	11.0	4	2.84	3.35	3.62
11.0	11.5	2	1.23	1.48	1.65

Table C.2: Ash content data of [Hald 1952] and group frequencies estimated by [Hald 1952] (H1952), [Hasselblad 1966] (H1966) and the EM algorithm for grouped truncated data.

estimate of [Hasselblad 1966]: 5.74, rough fit of [Hald 1952]: 25.91). Compared to Hald's estimate, the particular underestimation of the variance of the minor component is significantly improved. The estimated distribution parameters read as follows:

$$\begin{aligned}
\pi_1 &= 0.2113 \pm 0.0409 \text{ (Hasselblad: 0.2162, Hald: 0.20)} \\
\pi_2 &= 0.7887 \pm 0.1528 \text{ (Hasselblad: 0.7838, Hald: 0.80)} \\
\mu_1 &= 3.1860 \pm 0.2018 \text{ (Hasselblad: 3.210, Hald: 3.1)} \\
\mu_2 &= 7.3357 \pm 0.1257 \text{ (Hasselblad: 7.339, Hald: 7.2)} \\
\sigma_1 &= 0.9842 \pm 0.2299 \text{ (Hasselblad: 1.000, Hald: 0.8)} \\
\sigma_2 &= 1.5186 \pm 0.3018 \text{ (Hasselblad: 1.490, Hald: 1.5)}
\end{aligned}$$

C.2 Lognormal Distributions

A lognormal distribution can be derived from a normal distribution by replacing the independent variable x by its logarithm. Hence, the corresponding probability distribution function reads:

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}x}\exp\left[-\frac{(\log x-\mu)^2}{2\sigma^2}\right]. \tag{C.1}$$

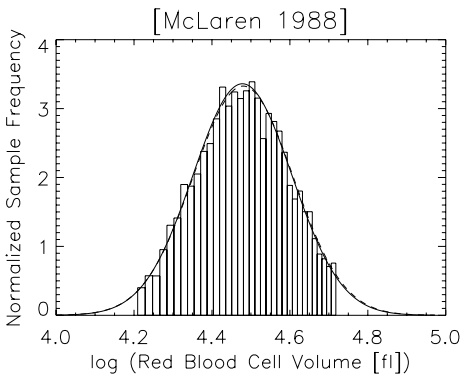


Figure C.2: Red blood cell volume data (histogram) of [McLaren et al. 1986a] and the corresponding probability distribution function estimated by the original authors (dashed line) and the EM algorithm for grouped and truncated normal data (solid line) on logarithmic scale.

Because the additional factor $1/x$ is separated from the parameters μ and σ^2 in the logarithmic representation, it plays no role in the derivatives of the log-likelihood function. Therefore, the M-step iterations for the parameters of a lognormal distribution can be obtained by taking those of a normal distribution with the replacement $\{x'_{mj}\} \to \log\{x'_{mj}\}$. The result corresponds to the findings of [McLaren et al. 1986a].

As an exercise for the implemented algorithm, the original data set (see Tab. C.3) of [McLaren et al. 1986a] is re-analysed. The result is shown in Fig. C.2. The corresponding χ^2 value is found to be 24.54 (for the parameters from the original publication 27.51) with the estimated parameters $\mu = 4.478 \pm 0.002$ (4.48) and $\sigma = 0.128 \pm 0.001$ (0.128).

C.3 Mixtures of Lognormal Distributions

As in the case of a one-component Gaussian-type density, the step from normal to lognormal populations of a finite mixture distribution is rather trivial because of the symmetry between the respective functions. In [McLaren et al. 1991], the EM procedure for normal data was adapted to be applied to a two-component mixture of lognormal functions in grouped, doubly-truncated data of red blood cell volume distributions. Applying the expressions derived for the case of normal component densities under the substitution $x \to \log x$ to the special case of two-component doubly-truncated mixture density data yields the corresponding results.

As a typical example for a mixture of two lognormal distributions, one may consider the red blood cell volume data sets for cows originally studied by [McLachlan and Jones 1988]. Again, the explicit data may be found in the corresponding Tab. C.4. Plotting the distributions estimated in the original work, one recognises that the location parameters are clearly shifted, which may be due to the bin locations not having been considered appropriately in the corresponding calculations (see Fig. C.3). This is recovered by χ^2 values of 260.66 (set 1) and 486.57 (set 2) which are out of discussion. However, re-analysis lead to more suitable parameters (χ^2 of 7.24 and 11.34, resp.) as follows:

$$\pi_1 = 0.4756 \pm 0.1051 \ (0.45) \text{ and } 0.1833 \pm 0.0643 \ (0.17)$$
$$\pi_2 = 0.5244 \pm 0.1159 \ (0.55) \text{ and } 0.8167 \pm 0.2865 \ (0.83)$$
$$\mu_1 = 3.9599 \pm 0.0472 \ (4.07) \text{ and } 3.7052 \pm 0.0625 \ (3.86)$$

log (Min. Vol. [fl])	log (Max. Vol. [fl])	Observation	M1986	EM
4.20935	4.22866	32	31.2581	31.1667
4.22866	4.24760	45	41.1017	41.1376
4.24760	4.26619	44	52.6178	52.8375
4.26619	4.28445	72	65.6719	66.1315
4.28445	4.30237	97	80.0127	80.7621
4.30237	4.31998	103	95.2782	96.3549
4.31998	4.33729	136	111.013	112.437
4.33729	4.35430	132	126.697	128.463
4.35430	4.37103	142	141.774	143.856
4.37103	4.38748	162	155.695	158.042
4.38748	4.40367	167	167.953	170.492
4.40367	4.41959	188	178.112	180.754
4.41959	4.43527	215	185.838	188.484
4.43527	4.45071	194	190.912	193.459
4.45071	4.46591	204	193.238	195.589
4.46591	4.48088	195	192.841	194.909
4.48088	4.49563	199	189.857	191.571
4.49563	4.51017	204	184.515	185.822
4.51017	4.52450	187	177.116	177.986
4.52450	4.53863	150	168.010	168.435
4.53863	4.55256	169	157.574	157.564
4.55256	4.56630	160	146.189	145.771
4.56630	4.57985	150	134.221	133.439
4.57985	4.59322	131	122.010	120.914
4.59322	4.60642	103	109.853	108.501
4.60642	4.61944	91	98.0032	96.4575
4.61944	4.63230	96	86.6656	84.9848
4.63230	4.64499	79	75.9946	74.2349
4.64499	4.65753	78	66.0992	64.3113
4.65753	4.66990	57	57.0462	55.2740
4.66990	4.68213	45	48.8661	47.1461
4.68213	4.69421	41	41.5592	39.9201
4.69421	4.70615	35	35.1017	33.5646
4.70615	4.71794	37	29.4513	28.0308

Table C.3: Red blood cell volume data of [McLaren et al. 1986a] and group frequencies estimated by [McLaren et al. 1986a] (M1986) and the EM algorithm for grouped truncated data.

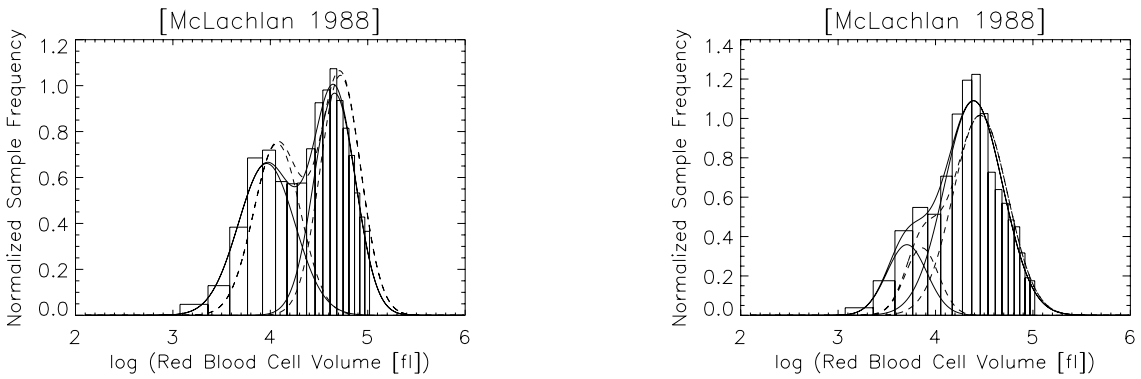


Figure C.3: Two different sets of red blood cell volume data [McLachlan and Jones 1988] for cows (histogram), the probability distribution functions fitted using the parameters from the original work (dashed line) and the outcome of the EM algorithm for grouped truncated data (solid line).

$$\mu_2 = 4.6598 \pm 0.0258 \ (4.72) \text{ and } 4.3903 \pm 0.0300 \ (4.46)$$
$$\sigma_1 = 0.2885 \pm 0.0154 \ (0.24) \text{ and } 0.2050 \pm 0.0136 \ (0.17)$$
$$\sigma_2 = 0.2176 \pm 0.0107 \ (0.21) \text{ and } 0.2997 \pm 0.0146 \ (0.28).$$

The values refer to both data sets analysed, brackets show the values given in [McLachlan and Jones 1988]. However, calculations show that while the estimates of the new implementation match the expected group frequencies given in the cited reference rather well, the recalculation using the parameters from the reference lead to clearly differing results underlying the above finding.

Min. Vol. [fl]	Max. Vol. [fl]	Observation	M1988	EM
21.6000	28.8000	10	0.504597	6.21018
28.8000	36.0000	21	6.48772	26.7743
36.0000	43.2000	51	26.6346	53.5962
43.2000	50.4000	77	53.6482	67.6374
50.4000	57.6000	70	68.0681	64.3377
57.6000	64.8000	50	64.1215	52.8130
64.8000	72.0000	44	51.7141	43.4683
72.0000	79.2000	40	42.2382	41.1096
79.2000	86.4000	46	40.2759	44.5327
86.4000	93.6000	54	43.9898	49.4815
93.6000	100.800	53	48.9454	52.0276
100.800	108.000	54	51.4385	50.4025
108.000	115.200	44	49.9237	44.9861
115.200	122.400	36	44.8060	37.3138
122.400	129.600	29	37.5051	29.0601
129.600	136.800	21	29.5651	21.4562
136.800	144.000	16	22.1460	15.1455
144.000	151.200	13	15.8859	10.2940

Min. Vol. [fl]	Max. Vol. [fl]	Observation	M1988	EM
21.6000	28.8000	9	0.229607	6.97041
28.8000	36.0000	32	6.73479	36.6318
36.0000	43.2000	64	32.9382	60.9851
43.2000	50.4000	69	55.4240	61.8875
50.4000	57.6000	56	57.1506	63.5373
57.6000	64.8000	68	58.8183	72.9124
64.8000	72.0000	88	66.8367	81.3142
72.0000	79.2000	93	73.5178	82.7472
79.2000	86.4000	87	73.9155	76.9760
86.4000	93.6000	67	68.1333	66.5468
93.6000	100.800	44	58.4892	54.3154
100.800	108.000	36	47.4512	42.3855
108.000	115.200	30	36.8089	31.9324
115.200	122.400	24	27.5517	23.4005
122.400	129.600	21	20.0415	16.7778
129.600	136.800	14	14.2477	11.8238
136.800	144.000	8	9.94374	8.22009
144.000	151.200	7	6.83776	5.65408

Table C.4: Red blood cell volume data for cows of [McLachlan and Jones 1988] and group frequencies estimated using the parameters given in [McLachlan and Jones 1988] (M1988) and this paper's implementation of the EM algorithm. Note that [McLachlan and Jones 1988] give different frequencies which are not consistent with the estimated parameters as published in the reference.