

Mathematical Statistics Stockholm University

Decomposition of time series of geological data into long- and short-timescale variations under derived time-dependent approximate Gaussian state space models.

Jelena Bojarova

Research Report 2008:1

ISSN 1650-0377

Postal address:

Mathematical Statistics Dept. of Mathematics Stockholm University SE-106 91 Stockholm Sweden

Internet:

 $\rm http://www.math.su.se/matstat$



DECOMPOSITION OF TIME SERIES OF GEOLOGICAL DATA INTO LONG- AND SHORT-TIMESCALE VARIATIONS UNDER DERIVED TIME-DEPENDENT APPROXIMATE GAUSSIAN STATE SPACE MODELS.

JELENA BOJAROVA*

January 2008

ABSTRACT. Time series of geological ice core data from Greenland are investigated by statistical modelling. The decomposition of the total variation into a long timescale variation (trend), a short timescale variation (fluctuations around the trend) and a noise component (including measurement errors) is proposed and discussed. A usually used time-invariant linear Gaussian model is compared with more advanced time-dependent Gaussian models, derived as approximations from the heavy-tailed distributional assumptions. In essence, the derived time-dependent Gaussian models result in a local smoothing in contrast to the global smoothing provided by the time-invariant model. To describe the mechanism of the local smoothing we introduce a concept of a local variance function. The local variance function is derived from a heavy-tailed density and is estimated from the observations. In the approximating Gaussian model, the time-dependent error variance works as a time-dependent measure of uncertainty about the dynamical development of the model state and it controls the influence of observations on the estimates of the model state components. The great advantage of the derived time-dependent Gaussian model is that the Kalman filter and the Kalman smoother can be used as an efficient computational tool performing the variation decomposition. Even though geological time series are investigated in this paper, the study is performed from the perspective of data assimilation. The main objective of the study is to investigate how different distributional assumptions on the stochastic drift (the model error) influence the estimate of the model state components. A few results on the interpretation of the data are given in addition.

Key words: Data assimilation; on-line estimation; Kalman filter and Kalman smoother; variation decomposition; local linear Gaussian model

^{*} Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: jelena@math.su.se.

1. INTRODUCTION

1.1. Data modelling strategy. An objective of statistical inference is to construct and interpret a statistical model, which is a simplistic representation of some data generating mechanism (Lindsey, 1996). The data generating mechanism itself is a complex process by which the data under investigation are produced. However, all complexities of the process are not of interest when the information contained in a particular random data set should be utilised. Rather, certain aspects of the process, determined by the goal of the study in the particular case, should be isolated and modelled.

In this paper we will analyse time series of measurements of some chemical substances performed on two ice cores drilled at different locations in Greenland. The ice cores taken from the arctic ice span a long period of time, and there is variation present on several different timescales. Here, analysing these time series, the specific interest is devoted to the decomposition of the variation into different timescale variation components, in accordance with the model structure

data = long timescale variation + short timescale variation

+**noise**

A typical model assumption in statistical analysis of time series is that Gaussian processes can describe variation. Convenience, simplicity and computational efficiency play a great role in choosing just a Gaussian process to describe variability. In this paper we will see that the standard Gaussian framework can easily be extended to non-Gaussian processes. Such an extension provides a much better description of the structure of ice core data, still preserving many attractive features of the Gaussian framework. Because the physical laws controlling the generation of the time series are complex and largely unknown, the assumptions to be made about model dynamics are quite general, with strong influence from the observations themselves on the model design.

The raw time series of observations are shown in Figure 1.

It is well known that the decomposition of the sample variability into different temporal and/or spatial scales is not unique (Chiles and Delfiner, 1999). Subjective prior assumptions about the amount of variability attributed to different components will influence the decomposition. At the same time, the choice of model describing the variability should fit the data under investigation.

In this paper the following model is assumed to describe the variability of the observations $\{y(t), t \ge 1\}$ with most variability being attributed to the short timescale variation component $\psi(t)$

(1)
$$y(t) = \mu(t) + \psi(t) + \epsilon(t), \quad t \ge 1$$

where:

 $\mu(t) \equiv$ is a non-stationary, smooth mean structure, which is called the long timescale variation (or the "trend").

 $\psi(t) \equiv$ is a stationary, zero mean random process, showing strong positive auto-correlation between neighbouring values. This is called the short timescale variation.

 $\epsilon(t) \equiv$ is a zero-mean white-noise process and it is called the observation error or the observation noise, even though other sources of variation may also contribute.

All three processes $\mu(t)$, $\psi(t)$ and $\epsilon(t)$ are assumed mutually independent. A basic assumption of the decomposition applied in this paper is that extraordinary



FIGURE 1. The raw time series of observations of four selected substances reconstructed from the measurements performed on the ice cores drilled at two different locations. The title of each diagram consists of the name of the corresponding ice-drilling project (NGRIP or GISP2) and the name of the selected substance: $\delta^{18}O$, $\log(Ca^{2+})$, $\log(SO_4^{2-})$ and MS^- .

observations in comparison with their neighbourhood should influence mainly the short timescale variation.

These eight time series, shown in Figure 1, will be modelled by linear state space models. The short timescale variation will be modelled through a stochastic cycle with a non-Gaussian (in most cases) assumption on the *model error* (the stochastic drift in the dynamics of the short timescale variation). To perform the total variation decomposition into different timescale variation components, and to estimate parameters entering in the dynamical and statistical model formulation, the Kalman filter and the Kalman smoother will be used as efficient computational tools. An approximate time-dependent Gaussian model will be derived from the original non-Gaussian linear state space model in order to create an environment where the Kalman filter and the Kalman smoother can work. Conditionally, given the observations, this model is able to capture the variability of essentially nonlinear time series of observations with regime of behaviour changing in time.

Even though the geological data are analysed in this paper, the our study is performed from the perspective of data assimilation and concerns important subjects of data assimilation. One of the most attractive properties of the Kalman filtering is the possibility to model the time evolution of the conditional moments (the two first moments) of the unobservable model state. A proper modelling of these statistics substantially improves the efficiency in assimilating observations within the framework of the mathematical model. Under the approximating time-dependent Gaussian model, the time evolution of the first two conditional moments of the model state, given observations, will partly be estimated from the observations.

When a heavy-tailed distribution is assumed under the original non-Gaussian model, the approximate time-dependent Gaussian model will provide a type of local smoothing, the mechanism of which is discussed in this paper. The impact of an observation on the different components of the variability decomposition (given in eqn. (1)) depends on its neighbourhood. The $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ series from the NGRIP and GISP2 ice cores under the local smoothing will be compared.

The local smoothing will be carried out and compared for the following four heavy-tailed alternatives to a Gaussian density: a mixture of two normal densities, a density of a general error distribution, a *Student t*-density and a Cauchy density.

1.2. Structure of the study. The observational data and some aspects of the data processing are discussed in Section 2. Section 3 concerns standard statistical tools used in this paper. A time-invariant linear Gaussian state space model is first fitted to each of the eight time series. Then the problem of the total variation decomposition into different timescales is formulated as the solution to a system of linear equations, which can be solved through Kalman filter and Kalman smoother recursions. Diagnostics to check the model are also considered in Section 3.

Based on the conclusions from the diagnostic check, a linear non-Gaussian state space model is proposed to model the variability of the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series from both ice cores, and this model is discussed in Section 4. A linear non-Gaussian state space model with heavy-tailed distributional assumptions on the stochastic drift of the short timescale variation is formulated. The Monte Carlo approach for computing the likelihood under this non-Gaussian model is discussed. An approximate time-dependent Gaussian state space model is derived from the original non-Gaussian model in order to perform the variation decomposition through an iterative application of a Kalman filter and a Kalman smoother. The diagnostics of the statistical fit of the approximate time-dependent Gaussian model for all six time series are considered as well.

The local variance function in the time-dependent Gaussian approximating model, which is estimated from the observations, provides a time-dependent measure of the uncertainty about the dynamical development of the model components. This local variance function is also discussed in Section 4. The mechanism of the local smoothing, when the impact of an observation depends on its neighbourhood in time, is finally also discussed.

Section 5 contains discussion on statistical modelling and some notes on interpretation of the time series model. Conclusions of the study are summurised at the end of this section.

2. A description of data

The two ice cores, referred to below by the names of the corresponding projects NGRIP (Northen Greenland Ice core Project, $72.10^{\circ}N, 45.32^{\circ}W$) and GISP2 (Greenland Ice Sheet Project 2, 320 km to the south of NGRIP), were drilled in early 1990's (GISP2) and early 2000's (NGRIP) on the Central (GISP2) and the Northern (NGRIP) Greenland Ice sheet. The time series represent high resolution measurements of $\delta^{18}O$, $\log(Ca^{2+})$, MS^{-} and $\log(SO_4^{2-})$ during the period 95,800 -11,400 B.P.("before present"), averaged over sections of 200 years, see Table 1 for details (**ppb** (parts per billion by weight) is a concentration measure).

TABLE 1. Investigated quantities -Meaning and Notations (**ppb** (parts per billion by weight) is a concentration measure).

| Notation | Meaning | Units |
|-------------------|---|-------------|
| $\delta^{18}O$ | $= \{ ({}^{18}O/{}^{16}O)_{sample} - ({}^{18}O/{}^{16}O)_{std} \} / ({}^{18}O/{}^{16}O)_{std} \}$ | |
| | where the std denotes the ratio in standard sea water | |
| | and <i>sample</i> denotes the ratio in the current sample and | |
| | ^{16}O and ^{18}O are isotopes of oxygen. | |
| | (Grootes et al., 1993) | |
| | The quantity is used as an air temperature proxy. | |
| $\log(Ca^{2+})$ | calcium $(2+)$ ion | $\log(ppb)$ |
| | The quantity is used as a terrestrial mineral dust proxy | (by weight) |
| | (Fuhrer <i>et al.</i> , 1999) | |
| $\log(SO_4^{2-})$ | sulfate (2-) ion | $\log(ppb)$ |
| | originates from volcanoes, mineral dust and | (by weight) |
| | marine biological activity | |
| MS^- | methanesulfonate (CH_3SO_3H) | ppb |
| | solely marine biological activity : | (by weight) |
| | as a result of oxidation of dimethylsulfide | |
| | (Saltzman, 1995) | |

A first brief insight into the data, supplementing Figure 1, can be obtained from the first-order directed scatter diagram of each time series, presented in Figure 2. These are diagrams of (y_{t-1}, y_t) with adjacent points, that is (y_{t-1}, y_t) and (y_t, y_{t+1}) , linked by a straight line. Even from a first glance it can be seen that the time series of observations of $\delta^{18}O$, $\log(Ca_{2+})$ and $\log(SO_4^{2-})$ exhibit some type of a cyclic behaviour superimposed on a small magnitude variation.

The raw time series are shown in Figure 1.

The $\delta^{18}O$ value is an air temperature proxy and shows the effect of climatic oscillations. The Ca^{2+} value is a terrestrial mineral dust proxy. The SO_4^{2-} and MS^- measure concentrations of sulfur ions. SO_4^{2-} originates from a mixture of sources such as volcanoes, mineral dust and marine biological activity as well as aerosols induced anthropogenically in modern time (Charlson *et al.*, 1992). Volcanic eruptions can be recognized as distinct SO_4^{2-} concentration peaks in the ice core records (Hammer, 1977). On the other hand, MS^- originates solely from marine biological activity and represents a result of oxidation of dimethylesulfide, which is produced by phytoplankton in the oceans. To study the time evolution of the global sulfur budget it is important to estimate the anthropogenic influence on the climate (e.g. Sugarawa, 1961; Delmas and Boutron, 1976).

climate (e.g. Sugarawa, 1961; Delmas and Boutron, 1976). The concentration data of the Ca^{2+} , SO_4^{2-} and MS^- from the NGRIP ice core used in this paper (on log-scale for Ca^{2+} and SO_4^{2-}) were obtained by ion chromatography on aliquots of decontaminated melt water from continuous flow analysis (Röthlisberger *et al.*, 2000a). Synchronisation of the GISP2 and the NGRIP time series was done through outstanding features (the *tie points*) identified in the SO_4^{2-} time series from the GISP2 ice core (Jonsell *et al.*, in review). (GISP2 data sources: SO_4^{2-} Mayewski *et al.*, 1997, available on the Greenland Summit ice core CD-rom published on Internet, MS^- , Saltzman *et al.*, 1997, $\delta^{18}O$, White, 2004; NGRIP data source: North Greenland Ice Core Project members, 2004). The outstanding features identified in the SO_4^{2-} time series from the NGRIP and GISP2 ice cores are assumed to be deposited simultaneously. The outstanding features are considered to represent Greenland Stadials (GS) (the abrupt climatic cold events) and Greenland Interstadials (GIS) (abrupt climatic warm events, also known as



FIGURE 2. The first-order directed scatter diagrams of four selected substances reconstructed from measurements performed on ice cores drilled at two different locations. The title of each diagram consists of the name of correspondent ice-drilling project (*NGRIP* or *GISP2*) and the name the selected substance in the following order: $\delta^{18}O$, $\log(Ca^{2+})$, $\log(SO_4^{2-})$ and MS^-

Dansgaard-Oeschger (DO) events) and they are used as tie points to synchronise the measurements from these two ice cores. The ages of the tie points from the NGRIP ice core, according to NGRIP(GRIP)-chronology (Johnsen *et al.*, 2001), were transferred to the corresponding SO_4^{2-} observations from the GISP2 ice core. A linear interpolation was used between adjacent tie points.

The six time series of $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ from the NGRIP and GISP2 ice cores look quite similar, whereas the two MS^- series look both mutually different and different from the others. For example, the log-concentrations $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ are nearly linear functions of the ratio $\delta^{18}O$ (the estimated zero-lag cross-correlations are $\rho_0(\log(Ca^{2+}), \delta^{18}O) = 0.94$ and $\log(SO_4^{2-}) \lesssim \log(Ca^{2+}) = 0.91$ for MCDED.

 $\rho_0(\log(SO_4^{2-}), \delta^{18}O) = 0.91$ for measurements from the NGRIP ice core).

As seen from the directed scatter diagrams of the time series shown in Figure 2 as well as from the raw time series in Figure 1, one can notice two regimes in the behaviour of the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series from both ice cores: a relatively smoothly varying *background signal* and a number of strong, *abrupt deviations* from this *background signal*.

A common approach to model such time series with two regimes of behaviour are the non-linear SETAR models (the self-exciting threshold autoregressive models,

Tong, 1990). Trials to capture the variability of the $\delta^{18}O$ series through a simple auto-regressive process are also known (e.g. Roe and Steig, 2004).

In this paper, a stationary stochastic cycle (or a stochastic wave process) with a heavy-tailed random source will be used to model the behaviour of the short timescale variation of these time series and to separate two regimes in the behaviour of observations. Other possibilities of a statistical separation of the peaks caused by volcanic eruption from the background signal have been studied by e.g. Fischer *et al.*, 1998 and Castellano *et al.*, 2004.

The model used in this paper has a property of local smoothing. Under this model the abrupt deviations from the background signal, which could originate from some external force (such as volcano eruptions), will influence the estimate of the short timescale variation but only to a minor extent affect the estimate of the long timescale variation. What belongs to the background signal is determined by neighbouring observations. The background signal will be modelled explicitly via a trend and a wave-like variation around the trend. The abrupt deviation from the background signal will be modelled implicitly via assumptions on a stochastic drift of the wave-like process.

The time series of methanesulfonate $(MS^- \text{ from both ice cores})$ have different patterns of behaviour. Investigation of Figure 2 and Figure 1 suggest a nearly stationary zero-mean process with near-Gaussian time increments as a model for the MS^- time series from the GISP2 ice core. No apparent cyclic pattern is present. Taking into account that the data investigated in this paper are averages of the original high-resolution data over 200 years periods, the Gaussian process can be sufficient to capture the variability of this time series. On other hand, the MS^- series from the NGRIP ice core indicates non-stationary fluctuations with a spread dependent on the mean value. The observations are noisy and the observation noise is especially strong for large values. This noise could originate from the observational process itself (the data-collecting procedure) or from some fluctuations un-resolvable on the time grid of 200 years.

3. Standard statistical tools

3.1. A Gaussian linear state space model. Structural time series models, where the large- and small- scale variations as well as the error term are modelled explicitly, have been used to describe time series of observations with great success, see Harvey (1989) for a detailed discussion. A linear Gaussian state space model provides a powerful tool in practical implementation of structural time series models. Under the state space model approach, the different, unobservable, components that make up the time series are modelled separately. The components define unobservable time series of model state variables representing the model development through time.

This development is formulated through a first-order vector autoregressive model which is called *the state equation*. The model state is related to observations through the so-called *observation equation*, which has the form of a linear regression (Durbin and Koopman, 2001). One of the standard ways in which the general linear Gaussian state space model can be written is

$$y(t) = Z_t \alpha(t) + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_{\epsilon t}^2),$$

$$\alpha(t) = T_t \alpha(t-1) + R_t \eta_{t-1}, \quad \eta_t \sim \mathcal{N}(0, \Omega_t), \quad t = 1, \dots, m$$

where y(t), t = 1, ..., n is a sequence of observations, $\alpha(t)$, t = 0, ..., n is an unobservable vector-sequence of model states and ϵ_t , t = 1, ..., n and η_t , t = 0, ..., n-1 are observation and model state errors respectively. Due to the

Markovian nature of the model, calculations necessary for the practical implementation can be performed in a very efficient recursive way, known as the Kalman filter and Kalman smoother recursions (Fahrmeir and Knorr-Held, 2000). This computational algorithm allows efficient estimation of unknown parameters in the model formulation. The same algorithm can treat even non-Gaussian and non-linear state space models. In this case, the original model is iteratively approximated by a sequence of linear Gaussian models.

To describe the variation of the time series shown in Figure 1, we will try a structural time series model with a similar structure for the time series of each substance. As a first step in the model fitting, a time-invariant Gaussian linear model, in which all system matrices $(Z_t, T_t, R_t, \Omega_t \text{ and } \sigma_{\epsilon t}^2)$ are constant over time, will be tried to describe the variation.

The model state will have long $\mu(t)$ and short $\psi(t)$ timescale variation components. The observations will be related linearly to the model state, $y(t) = \mu(t) + \psi(t) + \epsilon_t$, where ϵ_t is an observational noise. Because the physical laws describing deposition of the selected substances in the ice and snow are very complicated, only general assumptions will be made about dynamical development of the model state, at the same time allowing observations to have strong impact on the dynamics.

The long timescale variation $\mu(t)$ is modelled through a Gaussian integrated random walk (Young *et al.*, 1991) $\Delta^2 \mu(t+1) = \xi_t$, where Δ^2 is the second-order difference operator defined by $\Delta^2 x(t) = x(t) - 2x(t-1) + x(t-2)$.

(2)
$$\mu(t+1) = 2\mu(t) - \mu(t-1) + \xi_t, \quad \xi_t \sim N(0, \sigma_{\xi}^2), \quad t \ge 1$$

with independent identically distributed (i.i.d.) ξ_t and with a diffuse prior on the starting values $\mu(0), \mu(1) \sim N(0, K), \quad K \to \infty$. For simplicity $\mu(0)$ and $\mu(1)$ are assumed mutually independent. Such a formulation of the dynamics of the long timescale variation component allows the non-stationary development of the time series to be absorbed by the trend component, at the same time as imposing strong requirements on the smoothness of the trend estimate.

The short timescale variation $\psi(t)$ is modelled as a Gaussian stationary cyclic process (a stochastic wave process $(\psi(t), \psi^*(t)))$ of a certain frequency λ_c and a variance σ_{ψ}^2 with a stochastic drift (Harvey and Streibel, 1998).

(3)
$$\begin{pmatrix} \psi(t+1) \\ \psi^*(t+1) \end{pmatrix} = \rho \begin{pmatrix} \cos(\lambda_c) & \sin(\lambda_c) \\ -\sin(\lambda_c) & \cos(\lambda_c) \end{pmatrix} \begin{pmatrix} \psi(t) \\ \psi^*(t) \end{pmatrix} + \begin{pmatrix} \mathcal{X}_t \\ \mathcal{X}_t^* \end{pmatrix}, \ t \ge 0,$$

where the stochastic drift \mathcal{X}_t and \mathcal{X}_t^* are i.i.d. stochastic variables \mathcal{X}_t , $\mathcal{X}_t^* \sim N(0, (1 - \rho^2)\sigma_{\psi}^2)$. Because of the stationarity of the process, initial values of the cyclic component have a distribution $(\psi(0), \psi^*(0))' \sim N(\mathbf{0}, \sigma_{\psi}^2 \mathbf{I}_2)$ where **0** and \mathbf{I}_2 are 2×2 zero and identity matrices, respectively.

A Gaussian stationary stochastic process can be considered as an extension of a second-order autoregressive process. The wave-like development with the constant wave length (3) will provide a simplistic representation of the dynamics of the short timescale variation of the background signal. The stochastic drift will take care of the abrupt deviations from the background signal.

The "damping factor" ρ ($\rho < 1$) helps to specify the extent of a wave-like development in the formulation of the dynamical propagation of the short timescale variation. If $\rho \rightarrow 1$, the cyclic component tends to a deterministic behaviour of the wave and the wave process is assumed to capture well the time development of the short timescale variation. The smaller the "damping factor", the more stochastic is the pattern in the behaviour of the wave and the stronger is the influence of

the model error term. With the presence of a certain percentage of a stochastic drift in the dynamics of the short timescale variation, the frequency parameter λ_c stands for the averaged (the first-lag) auto-correlation of the time series. Small values of λ_c , say $\lambda_c < 0.45$, indicate presence of a strong first-lag auto-correlation $\rho \cos(\lambda_c) > 0.9\rho$ (provided ρ is sufficiently large).

The observation equation which relates the model state to the observations $\mathcal{Y} = \{y(1), \dots, y(n)\}$ is

(4)
$$y(t) = \mu(t) + \psi(t) + \epsilon_t, \ \epsilon_t \sim N(0, \sigma_\epsilon), \ 1 \le t \le n$$

All model error components $\xi_t, \mathcal{X}_t, \mathcal{X}_t^*$ and $\epsilon_t, 1 \leq t \leq n$ are assumed to be mutually uncorrelated and uncorrelated with the initial model state values $\mu(0), \mu(1), \psi(0), \psi^*(0).$

Thus, the time-invariant linear Gaussian model in the state space form describing the variation of the time series represented in Figure 1 is

$$y(t) = Z\alpha(t) + \epsilon =$$

$$= \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \alpha(t) + \epsilon_t, \ 1 \le t \le n$$

$$\alpha(t+1) = T\alpha(t) + \mathcal{R}\eta_t =$$
(5)
$$= \begin{pmatrix} 2 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \rho \cos(\lambda_c) & \rho \sin(\lambda_c) \\ 0 & 0 & -\rho \sin(\lambda_c) & \rho \cos(\lambda_c) \end{pmatrix} \alpha(t) + \mathcal{R}\eta_t, \ t \ge 1$$

where the model state is

$$\alpha(t) = \begin{pmatrix} \mu(t) \\ \mu(t-1) \\ \psi(t) \\ \psi^*(t) \end{pmatrix} \quad t = 1, \dots, n;$$

and the model error is

$$\eta_t = \begin{pmatrix} \xi_t \\ \mathcal{X}_t \\ \mathcal{X}_t^* \end{pmatrix} = \begin{pmatrix} \mu(t+1) - 2\mu(t) + \mu(t-1) \\ \psi(t+1) - \rho \cos(\lambda_c)\psi(t) - \rho \sin(\lambda_c)\psi^*(t) \\ \psi^*(t+1) + \rho \sin(\lambda_c)\psi(t) - \rho \cos(\lambda_c)\psi^*(t) \end{pmatrix}$$
$$\sim N(0,\Omega), \ \Omega = \begin{pmatrix} \sigma_{\xi} & 0 \\ 0 & (1-\rho^2)\sigma_{\psi} & 0 \\ 0 & 0 & (1-\rho^2)\sigma_{\psi} \end{pmatrix}; \ t = 1, \dots, n-1.$$

The model error projection matrix \mathcal{R} takes the form

$$\mathcal{R} = \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right).$$

Model (5) is a Gaussian linear model since both the model state equations (2)and (3) and the observation equation (4) are linear in the model state components $\mu(\cdot), \psi(\cdot), \psi^*(\cdot)$, and all errors of the model $\epsilon, \xi, \mathcal{X}$ and \mathcal{X}^* are assumed to have Gaussian distributions.

The set of model parameters $\Theta = \{\sigma_{\xi}^2, \sigma_{\psi}^2, \sigma_{\epsilon}^2, \lambda_c, \rho\}$ is estimated univariately, for each time series $\mathcal{Y} \equiv \mathcal{Y}_n = \{y(t), 1 \leq t \leq n\}$ maximising the log-likelihood 9

TABLE 2. Direct parameter estimates and transformed parameter estimates with their *standard errors* (stderr), obtained under the assumption of a linear Gaussian state space model and valid for the $\delta^{18}O$ series on each ice core. Transform^{*} means log-transforms for the parameters σ_{ξ} , σ_{ϵ} , σ_{ψ} and λ_c and a *logit*-transform for the parameter ρ .

| | core | parameter | estimate | $transform^*$ | stderr |
|----------------|-------|-------------------|----------|---------------|--------|
| | NGRIP | σ_{ξ} | 0.010 | -4.609 | 0.242 |
| | GISP2 | | 0.008 | -4.844 | 0.335 |
| | NGRIP | σ_ϵ | 0.081 | -2.518 | - |
| | GISP2 | | 0.001 | -7.073 | - |
| $\delta^{18}O$ | NGRIP | σ_ψ | 1.489 | 0.398 | 0.078 |
| | GISP2 | | 1.307 | 0.267 | 0.068 |
| | NGRIP | λ_c | 0.233 | -1.499 | 0.394 |
| | GISP2 | | 0.182 | -1.704 | 0.586 |
| | NGRIP | ρ | 0.768 | 1.198 | 0.369 |
| | GISP2 | | 0.727 | 0.980 | 0.179 |

 $l(\mathcal{Y}; \Theta) = \log(p(\mathcal{Y})) = \sum_{1}^{n} \log(p(y(t) | \mathcal{Y}_{t-1})))$. A multivariate treatment of these eight time series is a outside the scope of the present work.

Under the assumption of a linear Gaussian model the one step forecast errors, often called innovations, in a linear case $v(t) = y(t) - E(y(t) | \mathcal{Y}_{t-1})$ ($p(v(t)) = p(y(t) | \mathcal{Y}_{t-1})$ in general), are Gaussian and independently distributed (see section Diagnostics). In this case, the exact log-likelihood (Schweppe, 1965, also known from Harvey, 1989, as the prediction error decomposition) can efficiently be calculated through Kalman filter recursions

(6)
$$l(\mathcal{Y};\Theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{1}^{n}(\log(F(t)) + \frac{v(t)^2}{F(t)})$$

where F(t) is the variance of v(t). Maximisation can be performed by a numerical iterative maximum search routine, for example the BFGS (Broyden-Fletcher-Goldfarb-Shannon) implementation of Newton's method (Fletcher, 1987). In Table 2, estimated parameters for $\delta^{18}O$ from the NGRIP and GISP2 ice cores are shown. In practice it is more efficient to maximise the likelihood with respect to the log-/logit-transformed parameters, where $logit(\rho) = log(\frac{\rho}{1-\rho}), \quad 0 < \rho < 1,$ $\Theta^* = \{log(\sigma_{\xi}), log(\sigma_{\psi}), log(\sigma_{\epsilon}), log(\lambda_c), logit(\rho)\}$. The estimates of the original parameters and the estimates of the transformed parameters together with their standard errors are given in Table 2.

3.2. Maximum posterior estimate of the variation decomposition under the assumption of a Gaussian state space model. When the unknown parameters Θ have been estimated, the decomposition of the variation into the long and short timescale variations and observational noise can be obtained as a maximum posterior estimate of the model state vector $\alpha \equiv \{\alpha(t), t = 1, ..., n\}$, given the whole set of observations \mathcal{Y} . According to Bayes theorem, the posterior density for α is

(7)
$$p(\alpha \mid \mathcal{Y}; \Theta) = \frac{p(\mathcal{Y} \mid \alpha; \Theta)p(\alpha; \Theta)}{p(\mathcal{Y}; \Theta)} \propto p(\mathcal{Y} \mid \alpha; \Theta)p(\alpha; \Theta)$$

where $p(\alpha; \Theta)$ is determined by the prior on the starting values and the state equations (2) and (3) through the distributional assumption on the *model error* vector $\eta \equiv \{\xi_t, \mathcal{X}_t, \mathcal{X}_t^*, t = 0, ..., n-1\}$, and $p(\mathcal{Y} \mid \alpha; \Theta)$ is determined by the observation equation (4) through the distributional assumptions on the observation error vector $\epsilon \equiv \{\epsilon_t, t = 1, ..., n\}$.

Due to the mutual independence of the errors, the log-posterior density of α under the assumption of a linear Gaussian model (5) differs by an additive constant from

$$\log(p(\alpha \mid \mathcal{Y}; \Theta)) = constant - \frac{1}{2} \sum_{1}^{n} \frac{(y(t) - \mu(t) - \psi(t))^{2}}{\sigma_{\epsilon}^{2}}$$

$$= \frac{\mu(0)^{2}}{2K} - \frac{\mu(1)^{2}}{2K} - \frac{1}{2} \sum_{2}^{n} \frac{(\mu(t) - 2\mu(t-1) + \mu(t-2))^{2}}{\sigma_{\xi}^{2}}$$

$$= \frac{\psi(0)^{2}}{2\sigma_{\psi}^{2}} - \frac{1}{2} \sum_{1}^{n} \frac{(\psi(t) - \rho \cos(\lambda_{c})\psi(t-1) - \rho \sin(\lambda_{c})\psi^{*}(t-1))^{2}}{\sigma_{\psi}^{2}(1-\rho^{2})}$$

$$= \frac{\psi^{*}(0)^{2}}{2\sigma_{\psi}^{2}} - \frac{1}{2} \sum_{1}^{n} \frac{(\psi^{*}(t) + \rho \sin(\lambda_{c})\psi(t-1) - \rho \cos(\lambda_{c})\psi^{*}(t-1))^{2}}{\sigma_{\psi}^{2}(1-\rho^{2})}$$

where K is a predefined large number (in the actual calculations $K = 10^6$ was chosen). The method used here follows the technique discussed by Shephard and Pitt (1997) and in more detail by Durbin and Koopman (1992, 2001).

Under the assumption of a linear Gaussian model, the log-posterior density $\log(p(\alpha \mid \mathcal{Y}; \Theta))$ is a quadratic function in the model state components $\alpha(t) = (\mu(t), \mu(t-1), \psi(t), \psi^*(t))'$ or, respectively, in the model error components $\eta_t = (\xi_t, \mathcal{X}_t, \mathcal{X}_t^*)'$ and ϵ_t . Therefore, the maximum of the posterior density, provided that standard regularity conditions hold, can be obtained by solving the system of linear equations

$$\nabla_{\alpha} \log(p(\alpha \mid \mathcal{Y}; \Theta)) = 0.$$

In the case of a model specified by equation (5), the system of linear equations will have the following form

$$\begin{aligned} \frac{\partial p(\alpha \mid \mathcal{Y}; \Theta)}{\partial \mu(t)} &= \frac{\epsilon_t}{\sigma_\epsilon^2} - \frac{\xi(t-1)}{\sigma_\xi^2} + 2\frac{\xi_t}{\sigma_\xi^2} - \frac{\xi(t+1)}{\sigma_\xi^2}, \\ \frac{\partial p(\alpha \mid \mathcal{Y}; \Theta)}{\partial \psi(t)} &= \frac{\epsilon_t}{\sigma_\epsilon^2} - \frac{\mathcal{X}_{t-1}}{\sigma_\psi^2(1-\rho^2)} + \rho \cos(\lambda_c) \frac{\mathcal{X}_t}{\sigma_\psi^2(1-\rho^2)} - \rho \sin(\lambda_c) \frac{\mathcal{X}_t^*}{\sigma_\psi^2(1-\rho^2)}, \\ \frac{\partial p(\alpha \mid \mathcal{Y}; \Theta)}{\partial \psi^*(t)} &= -\frac{\mathcal{X}_{t-1}^*}{\sigma_\psi^2(1-\rho^2)} + \rho \sin(\lambda_c) \frac{\mathcal{X}_t}{\sigma_\psi^2(1-\rho^2)} + \rho \cos(\lambda_c) \frac{\mathcal{X}_t^*}{\sigma_\psi^2(1-\rho^2)}, \end{aligned}$$

for t = 2, ..., n - 2 together with the equations for the boundary observations at t = 0, 1 and t = n - 1, n.

Because of Gaussianity, the maximum posterior estimator of the model state vector $\hat{\alpha}$, that is the most probable estimator of α given the whole set of observations \mathcal{Y} , is identical to the conditional expectation $E(\alpha \mid \mathcal{Y})$ and can be efficiently computed through the classical Kalman filter forward recursion and smoother backward recursion (e.g. Kalman, 1960; Kohn and Ansley, 1989; de Jong, 1989; Anderson and Moore, 1979; Koopman, 1993). Notice that the time evolution of the smoothed trend component $\hat{\mu}(t)$ and of the smoothed cyclic component $\hat{\psi}(t), t = 1, \ldots, n$, can be obtained from either the Kalman model state smoother $E(\alpha \mid \mathcal{Y})$ or the Kalman disturbance smoother $\hat{\eta} = E(\eta \mid \mathcal{Y}), \hat{\epsilon} = E(\epsilon \mid \mathcal{Y})$. The Kalman filter and the smoother algorithms can be regarded as computational tools solving the system of equations (9) without necessity for a Bayesian interpretation. All computations necessary for the variation decomposition were performed using the statistical package **SsfPack 2.2**, developed by Koopman *et al.* (1999). The package works under the **Ox** computing environment developed by J.A. Doornik (1998). Both packages **Ox** and **SsfPack 2.2** are freely available on the Internet.

One example of the total variation decomposition into the large timescale $(\hat{\mu}(t))$, the short timescale $(\hat{\psi}(t))$ and the noise $(\hat{\epsilon}_t)$ components, obtained by the timeinvariant linear Gaussian state space model (5), is shown in Figure 3 for the $\delta^{18}O$ time series from the *NGRIP* ice core.



FIGURE 3. The decomposition of the total variation of the time series of $\delta^{18}O$ from the NGRIP ice core into the large-scale $\hat{\mu}(t)$ (upper right diagram), the small-scale $\hat{\psi}(t)$ (lower left diagram) and the observation noise $\hat{\epsilon}$ (lower right diagram) components under the time-invariant linear Gaussian model specified by equation (5). The fit of the "trend" + "cyclic" (the systematic part of observations $Z\hat{\alpha}(t)$) against raw observations, denoted by "stars", is shown on the upper left diagram.*Note*: The noise component is so small that the difference between the curve and the stars in the upper left diagram is impossible to distinguish.

The upper left diagram of Figure 3 contains the fit of observations cleaned from observational noise (called as the *systematic part* of observations hereafter, equal $Z\hat{\alpha}(t)$) together with raw observations marked by "stars". The estimated long timescale variation itself is shown on the upper right diagram. The estimated short timescale variation is shown on the lower left diagram. The lower right diagram illustrates the estimated observational noise. The raw observations are the $\delta^{18}O$ series from the *NGRIP* ice core already shown in Figure 1. The fit of observations

NGRIP GISP2 $\log Ca^{2+}$ $\delta^{18}O$ $\delta^{18}O$ $M\overline{S}$ $\log Ca^{2+}$ $\log SO_A^2$ $\log SO_4^2$ -606.466-247.804-176.806-572.508-301.279-134.700-248.250 $\hat{\sigma}_{\epsilon}$ -606.468-247.805-176.808-572.508-301.280-134.701-248.251 $\sigma_{\epsilon} \equiv 0$

TABLE 3. The log-likelihood of corresponding time series under the general Gaussian model (first line) and under the more parsimonious model with a perfect fit to observations (second line).

is nearly perfect under the model. The estimated observational noise is negligible compared with the total variation of the time series: $\hat{\sigma}_{\epsilon} = 0.081$ compared to $\hat{\sigma}_{\psi} = 1.489$. According to the Akaike information criterion, which is a log-likelihood function with a maximum likelihood estimate of the parameters penalised by the dimension of the vector of parameters (Akaike, 1974), a more parsimonious model with perfect observation fit $(y(t) = Z\alpha(t)$ instead of the original model defined by equation(4)) should be used to model the variation of the time series of $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ from each ice core. The summary of the observed values of the log-likehoods of the corresponding time series under the general and under the more parsimonious model is given Table 3. The only observational error of a similar order as the total variation of the time series was indicated for the $MS^$ observations from the NGRIP ice core with $\hat{\sigma}_{\epsilon} = 0.380$ compared to $\hat{\sigma}_{\psi} = 0.354$. This supports the visual impression based on a brief investigation of Figure 1 and Figure 2.

Even if the fit of the observations is nearly perfect by the time-invariant linear Gaussian model, the large sample diagnostic tests reject normality of the underlying state space model for all time series except the MS^- from GISP2.



FIGURE 4. Empirical cumulative distribution function from 1000 simulations of Box-Ljung statistics Q(15) based on the first 15 autocorrelations under the hypothesis of mutually i.i.d. $\mathcal{N}(0, 1)$ -distributed time series of N (N = 418) observations. The dashed lines mark quantiles (as indicated) of the distribution of Q(15).

The test of normality is performed on the standardised one-step forecast errors known also as standardised innovations

$$\tilde{v}_t = \frac{v(t)}{\sqrt{F(t)}} \quad t = 2, \dots, n$$

When a linear Gaussian model underlies the time series, the innovations are Gaussian and orthogonal to the past observations. Thus the standardised innovations are i.i.d. $\mathcal{N}(0,1)$ -distributed (e.g. Tong, 1990, Tests for linearity). The Box-Ljung portmanteau statistic (Ljung and Box, 1978) was calculated to check the serial correlation of the standardised innovations. This statistic is given by

(10)
$$Q(k) = (n-1)(n+1)\sum_{j=1}^{k} \frac{c_j^2}{n-1-j}$$

for some preset positive value k where c_j is the *j*th correlogram value, $j = 1, \ldots, k$,

(11)
$$c_j = \frac{1}{(n-1)m_2} \sum_{t=j+2}^n (\tilde{v}_t - m_1)(\tilde{v}_{t-j} - m_1)$$

Here m_1 and m_2 are usual the first- and the second-order moments of \tilde{v}_t .

$$m_1 = \frac{1}{n-1} \sum_{t=2}^n \tilde{v}_t, \ m_2 = \frac{1}{n-1} \sum_{t=2}^n (\tilde{v}_t - m_1)^2.$$

Figure 4 shows a cumulative empirical distribution function $F_{Q(15)}$, constructed from 1000 simulations of the Box-Ljung statistic Q(15) under the hypothesis of an i.i.d. $\mathcal{N}(0, 1)$ -distributed time series of N = 418 observations (N = 418 is the length of each of the time series $y(t), t = 1, \ldots, N$ investigated in this paper). Large values of the statistics Q(k) >> k indicate presence of strong auto-correlations. Figure 4 will be used below as a reference for checking serial correlation.

Figure 5 shows diagnostic plots for the standardised innovations \tilde{v}_t of the MS^- series from the GISP2 ice core. The figure contains four diagnostic diagrams: a plot of the standardised innovations (upper left), their histogram together with the estimated density (Parzen, 1962; Rosenblatt, 1956; Sarda and Vieu, 2000) Estimated densities shown in this paper are of the type Parzen-Rosenblatt kernel density estimate with Epanechnikov kernel and over-smoothed bandwidth $h = 1.5h_{opt}$, where the h_{opt} is the asymptotically optimal bandwidth in the case of the standard Gaussian density $h_{opt} = 1.62 * n^{-0.2}$. In our case the bandwidth h = 1.5 * 1.62 * 0.299 = 0.73. (upper right), the QQ-plot of the standardised innovations against the quantiles of their theoretical density N(0,1), in case the hypothesis about normality of the underlying model holds (lower left), and the correlogram (equation (11)) of the standardised innovations (lower right). The estimated density of the standardised innovations corresponding to the MS^- series from the GISP2 ice core looks much like a Gaussian one. The QQ-plot also shows quite a good correspondence between the ordered standardised innovations and the N(0,1)-quantiles with presence only of some positive outliers and a minor discrepancy in the left tail. It is in fact only 6-% of data (25 of total 416) which exceed the two-standard deviation level. The standardised innovations are nearly uncorrelated as can be seen from the plot of the correlogram. The corresponding value of Q(15) statistics is 16.69. The empirical cumulative distribution function $F_{Q(15)}(16.69) = 0.681$. Based on the diagnostic plots presented in Figure 5, the time-invariant linear Gaussian model (5) can be considered acceptable to describe the variation of the MS^- series from the GISP2 ice core.

The results of the variation decomposition based on model (5) for the $MS^$ time series from the GISP2 ice core are shown in Figure 6. The smoothed estimate



FIGURE 5. Diagnostics plots for standardised innovations of the time series of methanesulfonate MS^- observations from the GISP2 ice core. The upper left diagram shows the standardised innovations. The upper right diagram shows their histogram together with the estimated density (solid curve). The Gaussian density is shown as a reference (dashed curve). The lower left diagram shows the QQ-plot of the ordered standardised innovations against the N(0,1) quantiles (innovations are plotted along the y-axes and normal quantiles along the x-axes). The lower right diagram shows the correlogram c_j of the standardised innovations (lag j is along x-axes.).

of the trend $\hat{\mu}(t)$ has the simple form of a decreasing linear function in time, $\mu(t + 1) - \mu(t) = \mu(t) - \mu(t-1)$. Under the Akaike information criterion, a simpler state space model with no stochastic drift in the dynamical trend development ($\hat{\sigma^2}_{\xi} = 0$) and no observation error ($\hat{\sigma^2}_{\epsilon} = 0$) adequately describes the MS^- series from the GISP2 ice core. The estimated value of the damping factor ρ is relatively small (see Table 4) and 54% of the short-range variation comes from a stochastic drift.

When a heavy-tailed *Student* t-distribution was tried for modelling the stochastic drift of the cyclic component, the estimated degree of freedom ν was relatively high ($\nu = 19$). This fact supports again that the Gaussian assumption on the distribution is reasonable. Only minor differences in the variation decomposition were noticed under these two different underlying models. The estimated parameters (the original and the transformed) together with the standard error of the transformed parameters are shown in Table 4 (see section 3.1 for an explanation of the meaning of the parameters).



FIGURE 6. The decomposition of the total variation of the $MS^$ series from the GISP2 ice core into the large-scale $(\hat{\mu}(t))$ (the upper right diagram), the small -scale $(\hat{\mathcal{X}}_t)$ (the lower left diagram) variations and the observation noise (the lower right diagram; note scale) under the time-invariant linear Gaussian model specified by equation (5). The systematic part $\hat{y}(t) = Z\hat{\alpha}(t) = \hat{\mu}(t) + \hat{\psi}(t)$ against the raw observations y(t) (stars) is shown on the upper left diagram.

TABLE 4. Parameter estimates for the MS^- series from the GISP2 ice core under the time-invariant linear Gaussian state space model. The notation "transform*" means the log-transform of parameters σ_{ψ} and λ_c and logit transform of parameter ρ . The standard error corresponds to the transformed parameters.

| parameter | estimate | $transform^*$ | stderr (Gauss) |
|---------------------|----------|---------------|----------------|
| $\hat{\sigma}_\psi$ | 0.589 | -0.529 | 0.057 |
| $\hat{\lambda}_c$ | 0.175 | -1.751 | 0.755 |
| $\hat{ ho}$ | 0.697 | 0.832 | 0.166 |

The results of the diagnostic check for normality of the standardised innovations \tilde{v}_t , the standardised smoothed trend residuals $\tilde{\xi}_t$ ($\tilde{\xi}_t = \hat{\xi}_t/stderr(\hat{\xi}_t)$) and the standardised smoothed cyclic component residuals $\tilde{\mathcal{X}}_t$ ($\tilde{\mathcal{X}}_t = \hat{\mathcal{X}}_t/stderr(\hat{\mathcal{X}}_t)$) corresponding to the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series are presented in Figure 7 in the form of QQ-plots. The standardised smoothed model state residuals $\tilde{\eta}(t) = (\tilde{\xi}_t, \tilde{\mathcal{X}}_t, \tilde{\mathcal{X}}_t^*)'$ are often called *auxiliary residuals* and are discussed in



FIGURE 7. The QQ-plots of the standardised innovations \tilde{v}_t (the first column of diagrams), of the *standardised* smoothed trend residuals $\tilde{\xi}_t$ (the second column of the diagrams) and of the standardised smoothed cyclic component residuals $\tilde{\mathcal{X}}_t$ (the third column of the diagrams) against their theoretical quantiles $\mathcal{N}(0, 1)$ of the $\delta^{18}O$ series (the first row of the diagrams), of the $\log(Ca^{2+})$ series (the second row of the diagrams) and of the $\log(SO_4^{2-})$ series (the third row of diagrams) from the NGRIP ice core (the solid line) and from the GISP2 ice core (the dashed line). The time-invariant linear Gaussian model (5) is applied.

detail in Harvey and Koopman (1992). In general auxiliary residuals are cross- and auto-correlated.

In the QQ-plots, the NGRIP and GISP2 data are represented by solid lines and dashed lines, respectively. The QQ-plots from the two sites are remarkably similar. The QQ-plots of the standardised innovations clearly indicate that the hypothesis about an underlying time-invariant Gaussian model should be rejected for the $\delta^{18}O$, log(Ca^{2+}) and log(SO_4^{2-}) time series from both ice cores. The \tilde{v}_t are heavy-tailed: the ones corresponding to the $\delta^{18}O$ series have a heavy right tail and the ones corresponding to log(Ca_{2+}) and log(SO_4^{2-} have heavy left tails. The non-Gaussianity noted in the distribution of the innovations \tilde{v}_t comes basically from the non-Gaussian behaviour of the cyclic component residuals \tilde{X}_t , as seen from comparison of columns 1 and 3 in Figure 7.

The standardised trend component residuals ξ_t for the time series are symmetric and coincide well with Gaussian ones except in the tails which are too light. The light tails in the distribution of the ξ_t are caused by a deficiency in the dynamics of the trend model, which provides an over-smoothed trend estimate. An alternative dynamical model of the trend time development with a weaker assumption on the smoothness $(\mu(t+1) = \mu(t) + \xi_t \text{ instead of } (2))$ improves the behaviour of the smoothed trend residuals in the tails. On the other hand, it is usually expected that the curve representing a trend should be very smooth.

The non-Gaussian character of the standardised innovations \tilde{v}_t is even more clearly illustrated in Figure 8. Here the same plots as for MS^- from the GISP2ice core, presented in Figure 5, are shown for the \tilde{v}_t of the $\log(SO_4^{2-})$ time series from the GISP2 ice core. The estimated density is clearly non-Gaussian with a very heavy left tail.



FIGURE 8. Diagnostics plots for the standardised innovations \tilde{v}_t of the log(SO_4^{2-}) time series from the GISP2 ice core. The upper left diagram shows the standardised innovations. The upper right diagram shows their histogram together with the estimated density. The lower left diagram shows the QQ-plot of the ordered standardised innovations against the N(0,1)-quantiles. The lower right diagram shows the correlogram of the standardised innovations. $(Q(15) = 12.64, F_{Q(15)}(12.64) = 0.37)$

It is interesting to notice that \tilde{v}_t corresponding to the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series all seem to be nearly mutually uncorrelated within the series. This means that the underlying state space model is close to a linear one. Table 5 presents a summary of the Box-Ljung statistics Q(15) (eqn. 10) based on the first 15 autocorrelations c(j), $j = 1, \ldots, 15$, for the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series from the NGRIP and the GISP2 ice cores. Included are also the corresponding P-values, $F_{Q(15)} = \mathcal{P}\{Q(15) \leq Q(15)^{obs}\}$, and the maximal absolute value of the autocorrelations $c(j)_{max} = max_{1 \leq j \leq 15} |c(j)|$. An empirical cumulative

| substance | ice core | Q(15) | $F_{Q(15)}$ | $c(j)_{max}$ |
|-------------------|----------|-------|-------------|--------------|
| $d^{18}O$ | NGRIP | 8.63 | 0.111 | -0.07 |
| | GISP2 | 10.81 | 0.239 | -0.09 |
| $\log(Ca^{2+})$ | NGRIP | 16.05 | 0.634 | -0.09 |
| | GISP2 | 15.77 | 0.610 | -0.10 |
| $\log(SO_4^{2-})$ | NGRIP | 12.30 | 0.347 | -0.09 |
| | GISP2 | 12.64 | 0.370 | -0.11 |

TABLE 5. The Box-Ljung statistics based on the first 15 autocorrelations for the standardised innovations corresponding to the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series from both ice cores.

distribution function, constructed from 1000 simulated values of the Box-Ljung statistics Q(15) under the hypothesis of the i.i.d. N(0, 1)-distributed time series with N = 416 observations, is shown Figure 4.



FIGURE 9. The decomposition of the total variation of the $MS^$ time series from the NGRIP ice core under the time-invariant linear Gaussian state space model (5) and the diagnostics plots for the standardised innovations and auxiliary residuals testing the fit of the model. The four diagrams to the left present the total variation decomposition into different timescales: the systematic part of the MS^- series (solid curve) $\hat{y}(t) = \hat{\mu}(t) + \hat{\psi}(t)$ together with the raw observations (dots) y(t) (upper left), the smoothed estimate of the long-range variation $\hat{\mu}(t)$ (upper 2^{nd} from the left), the smoothed estimate of the short-range variation $\hat{\psi}(t)$ (lower left) and the smoothed estimate of the observation noise $\hat{\epsilon}(t)$ (lower 2^{nd} from the left). The four diagrams to the right present the diagnostic plots of the model fit : the QQ-plots of the standardised innovations \tilde{v}_t (upper 2nd from the right), the standardised smoothed trend component residuals $\tilde{\xi}_t$ (upper right), the standardised smoothed cyclic component residuals $\tilde{\mathcal{X}}_t$ (lower 2nd from the right) and standardised smoothed observations residuals $\tilde{\epsilon}(t)$ (lower right) against their theoretical $\mathcal{N}(0,1)$ quantiles.

The diagnostics performed on the MS^- time series from the NGRIP ice core indicate that the underlying structure is not Gaussian. The decomposition of the total variation into different timescales and the diagnostic plots to check normality of the standardised innovations are shown in Figure 9. It is interesting to notice that the trend component $\hat{\mu}(t)$ (the long timescale variation) absorbs the largest amount of variation. The non-stationary process dominates the modelling of the variation of the time series. The standardised trend component residuals ξ_t are heavy-tailed in this case. The substantial observation noise with a number of outliers contributes additionally to a non-Gaussian behaviour of the state space model based on this MS^- time series. As it was already mentioned, the observational noise could originate from the measurement process itself, or it may represent processes, unresolvable on the current time grid, where 1 time step is equal to 200 years.

The presence of strong outliers among the MS^- observations from the NGRIPice core can be modelled assuming a heavy tailed distribution for the observation error. This is a commonly used technique, which results in a reduced impact of outliers on the smoothed estimate of the model state $\hat{\alpha}$. An observation y(t) influences the smoothed estimates of a model state at the particular moment t through a one-step-forecast error (or innovation) v(t) scaled by the inverse of its estimated variance F(t) = var(v(t)). The estimated variance F(t), which becomes large for the outliers if a heavy-tailed distribution is assumed for the observation error, weakens the influence of the outlier y(t) on the systematic part of the variation, $Z\hat{\alpha}(t)$, and increases the impact of the outlier on the estimate of the observation noise $\hat{\epsilon}_t$.

On the other hand, the fitted models for $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series from both ice cores contain essentially only the systematic part of the variation, $Z\hat{\alpha}(t)$, and are nearly free from the observation error ($\hat{\sigma}_{\epsilon}^2 \approx 0$). The non-Gaussian character of the variation of the time series will be modelled through a time-dependent linear Gaussian model derived from a heavy-tailed distributional assumption on the cyclic model error component $\mathcal{X}_t, \mathcal{X}_t^*$. This modelling approach has some nice properties of local smoothing and will be discussed in the next section.

There exists a large number of various approximate smoothing and filtering algorithms which deal with robustified state-space modelling (e.g. Masreliez, 1975; Masreliez and Martin, 1977; Kitagawa, 1987). We follow the methodology presented in Koopman *et al.*, 1998.

4. A LINEAR NON-GAUSSIAN STATE SPACE MODEL.

4.1. Heavy-tailed distributions for the cyclic model error component. Heavy-tailed distributions are assumed for the cyclic model error component to describe the variation of the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series from both ice cores.

(12)

$$y(t) = Z\alpha(t) + [\epsilon(t)], \quad \epsilon(t) \sim N(0, \sigma_{\epsilon}^{2}), \quad \sigma_{\epsilon} \equiv 0, \quad 1 \le t \le n$$

$$\alpha(t+1) = T\alpha(t) + \begin{pmatrix} \xi_{t} \\ \mathcal{X}_{t} \\ \mathcal{X}_{t}^{*} \end{pmatrix}, \quad \begin{cases} \xi_{t} \sim N(0, \sigma_{\xi}^{2}) \\ \mathcal{X}_{t}, \mathcal{X}_{t}^{*} \sim \mathcal{H}_{k} \end{cases}, \quad t \ge 0$$

where the stochastic drift \mathcal{X}_t and \mathcal{X}_t^* are assumed to have one of four zero-mean symmetric heavy-tailed densities $\mathcal{X}, \mathcal{X}^* \sim \mathcal{H}_k, \ k = \{1, 2, 3, 4\}.$

• \mathcal{H}_1 denotes a t – *Student*-distribution with ν degree of freedom and the scale parameter $\sigma_{\mathcal{X}}$ (variance $\sigma_{\mathcal{X}}^2$)

$$h_1(u;\nu,\sigma_{\mathcal{X}}^2) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(0.5(\nu+1))}{\Gamma(0.5\nu)} \frac{1}{\sqrt{(\nu-2)\sigma_{\mathcal{X}}^2}} \left(1 + \frac{u^2}{(\nu-2)\sigma_{\mathcal{X}}^2}\right)^{-0.5(\nu+1)}, \ \nu \ge 3$$

TABLE 6. The estimated values of the parameters involved into the state space model formulations describing the variation of the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series.

| substance | ice core | distribution | $\hat{\sigma}_{\mathcal{X}}$ | $\hat{\sigma}_{\xi}$ | $\hat{\sigma}_{\psi}$ | $\hat{ ho}$ | $\hat{\lambda}_c$ |
|----------------|----------|--|------------------------------|----------------------|-----------------------|-------------|-------------------|
| | NGRIP | $\mathcal{N}(0, \sigma_{\mathcal{X}}^2)$ | 0.953 | 0.010 | 1.489 | 0.768 | 0.233 |
| | | $\mathcal{H}_1(t\text{-dist}) \ \nu = 4$ | 0.695 | 0.010 | 1.349 | 0.877 | 0.135 |
| | | $\mathcal{H}_2(MN) \ \lambda = 0.9, \ \chi = 20$ | 0.507 | 0.007 | 1.011 | 0.865 | 0.014 |
| $\delta^{18}O$ | GISP2 | $\mathcal{N}(0, \sigma_{\mathcal{X}}^2)$ | 0.879 | 0.008 | 1.307 | 0.727 | 0.182 |
| | | $\mathcal{H}_1(t\text{-dist}) \ \nu = 4$ | 0.701 | 0.010 | 1.197 | 0.810 | 0.075 |
| | | \mathcal{H}_2 (MN) $\lambda = 0.9, \ \chi = 20$ | 0.465 | 0.008 | 0.993 | 0.843 | 0.026 |
| | NGRIP | $\mathcal{N}(0, \sigma_{\mathcal{X}}^2)$ | 0.395 | 0.005 | 0.722 | 0.837 | 0.252 |
| | GISP2 | $\mathcal{H}_1(t\text{-dist}) \ \nu = 4$ | 0.194 | 0.006 | 0.693 | 0.898 | 0.116 |
| | | \mathcal{H}_2 (MN) $\lambda = 0.85, \ \chi = 20$ | 0.164 | 0.003 | 0.496 | 0.944 | 0.008 |
| $\log(Ca_2+)$ | | $\mathcal{N}(0, \sigma_{\mathcal{X}}^2)$ | 0.456 | 0.005 | 0.817 | 0.830 | 0.226 |
| | | $\mathcal{H}_1(t\text{-dist}) \ \nu = 4$ | 0.222 | 0.006 | 0.784 | 0.897 | 0.104 |
| | | \mathcal{H}_2 (MN) $\lambda = 0.85, \ \chi = 20$ | 0.210 | 0.003 | 0.691 | 0.953 | 0.029 |
| | NGRIP | $\mathcal{N}(0, \sigma_{\mathcal{X}}^2)$ | 0.347 | 0.003 | 0.543 | 0.769 | 0.172 |
| $\log(SO_4-)$ | | $\mathcal{H}_1(t\text{-dist}) \ \nu = 4$ | 0.278 | 0.003 | 0.529 | 0.851 | 0.134 |
| | | \mathcal{H}_2 (MN) $\lambda = 0.85, \ \chi = 10$ | 0.193 | 0.002 | 0.384 | 0.862 | 0.026 |
| | GISP2 | $\mathcal{N}(0, \sigma_{\mathcal{X}}^2)$ | 0.317 | 0.002 | 0.459 | 0.759 | 0.142 |
| | | $\mathcal{H}_1(t\text{-dist}) \ \nu = 4$ | 0.256 | 0.003 | 0.448 | 0.820 | 0.117 |
| | | \mathcal{H}_2 (MN) $\lambda = 0.85, \ \chi = 10$ | 0.218 | 0.003 | 0.403 | 0.841 | 0.039 |

The larger the degree of freedom ν the closer the density $h_1(u;\nu,\sigma^{2x})$ is to the Gaussian $N(0,\sigma_{\chi}^2)$.

For each of the six time series, considered here, the same value of the degree of freedom $\nu = 4$ was found to provide a good model fit to the data. Such a low value of the estimated degree of freedom indeed indicates much heavier tails than for a Gaussian one.

• \mathcal{H}_2 denotes a mixture of two normal densities with parameters λ, χ and a basic variance σ_{χ}^2

$$h_2(u;\lambda,\chi,\sigma_{\mathcal{X}}^2) = \frac{\lambda}{\sqrt{2\pi\sigma_{\mathcal{X}}^2}} \exp\left(\frac{u^2}{2\sigma_{\mathcal{X}}^2}\right) + \frac{1-\lambda}{\sqrt{2\pi\chi\sigma_{\mathcal{X}}^2}} \exp\left(\frac{u^2}{2\chi\sigma_{\mathcal{X}}^2}\right)$$

where $(1-\lambda)$ is a percentage of outliers among the data, λ is near 1, say $\lambda \ge 0.85$ and χ is large, say from 10 to 100, and determines the variance of the outliers.

• \mathcal{H}_3 denotes a *Cauchy distribution* with the parameter $\sigma_{\mathcal{X}}^2$ (Feller, 1966)

$$h_3(u;\sigma_{\mathcal{X}}^2) = \frac{1}{\pi\sqrt{\sigma_{\mathcal{X}}^2}} \left((1 + \frac{u^2}{\sigma_{\mathcal{X}}^2}) \right)^{-1}$$

Note that the Cauchy density is very heavy-tailed and has an infinite variance. The Cauchy density corresponds to a limiting case of the *t*-densities with the degree of freedom $\nu \rightarrow 2$.

• \mathcal{H}_4 denotes a general error distribution with a parameter q and a variance $\sigma_{\mathcal{X}}^2$ (Box and Tiao, 1973)

$$h_4(u;q,\sigma_{\mathcal{X}}^2) = \frac{2\sqrt{\Gamma(0.75q)}}{l\sqrt{(\Gamma(0.25q))^3}} \frac{1}{\sqrt{\sigma_{\mathcal{X}}^2}} \exp\left(-\left(\frac{\Gamma(0.75q)}{\Gamma(0.25q)}\right)^{0.5q} \left(\frac{u^2}{\sigma_{\mathcal{X}}^2}\right)^{0.5q}\right), \ 1 < q < 2$$

The closer q is to 2, the closer is the density to the Gaussian $N(0, \sigma_{\chi}^2)$. A Gaussian density $\mathcal{N}(0, \sigma_{\chi}^2)$ can be considered as a special case of a general error distribution \mathcal{H}_4 with parameters q = 2 and σ_{χ}^2 .



FIGURE 10. Plot of the different estimates of the large-scale variation $\hat{\mu}(t)$ for the $\delta^{18}O$ time series from the NGRIP ice core under the Gaussian (solid curve) and the non-Gaussian distributional assumptions on the cyclic model error component $\mathcal{X}_t, \mathcal{X}_t^* \sim \mathcal{H}_k, k = \{1, 2, 3\}$: the *t*-distribution with $\nu = 4$ degree of freedom (dashed curve), the mixture of two normal densities with parameters $\lambda = 0.9$ and $\chi = 20$ (dash-dotted curve) and the Cauchy distribution (dotted curve). The observations are denoted by "*".

In order to preserve stationarity of the cyclic process the variance of the heavytailed distributions σ_{χ}^2 is parametrised as $\sigma_{\chi}^2 = (1 - \rho^2)\sigma_{\psi}^2$. The same definition of σ_{χ}^2 is used even in the case of the Cauchy distribution. Notice that in this case the cyclic process is not stationary anymore because of the infinite variance of the Cauchy distribution.

The ML (Maximum Likelihood) estimates of the parameters involved into the state space model formulations under the different assumptions on the distribution for the cyclic model error component, the Gaussian one, the *t*-distribution (\mathcal{H}_1) and the mixture of two normal densities (\mathcal{H}_2), are presented in Table 6. The parameters are estimated by maximising the corresponding likelihood function. The weaker the restrictions of the variability expressed in the distributional assumptions on the cyclic model error component, the larger part of the variability is attributed to the stochastic part of the dynamic development of the cyclic component. The damping factor $\hat{\rho}$ increases and the wave frequency $\hat{\lambda}_c$ decreases at the same time as the estimated variance $\hat{\sigma}_{\psi}$ of the cyclic component becomes more robust to the exceptional observations in the comparison with their surrounding values in time.

For all six time series, the observations seem to originate from two different regimes of the underlying model: *the background signal* with a smooth pattern

of the behaviour (originates from some continuous process in time), to which the majority of the observations belongs, and *the "alternative" regime*, which contains the severe deviations from the background signal, which could be caused by a momentary influence of some external forcing (volcanic eruptions, as one example). From the model perspective the splitting of the observations into

regime dominated by the background signal & the alternative regime corresponds to the spliting

those observations that the model is able to predict &

those observations that the model is not able to predict.

Under the non-Gaussian assumptions on the cyclic model error component, the observations, which bring up the background signal, are explicitly handled by the model dynamical propagator T (eqn 12). The observations which belongs to the alternative regime are handled implicitly via the assumptions on $\mathcal{X}_t, \mathcal{X}_t^*$. Different distributional assumptions on the cyclic model error result in different splitting of the observations into the regime dominated by the background signal and the alternative regime.

The different estimates of the long timescale variation $\hat{\mu}(t)$ for $\delta^{18}O$ from the NGRIP ice core under the Gaussian (solid curve) and the heavy-tailed assumptions on the distribution for the cyclic model error (the t-distribution with $\nu = 4$ (dashed curve), the mixture of two normal densities with parameters $\lambda = 0.9$ and $\chi = 20$ (dash-dotted curve) and the Cauchy distribution (dotted curve)) are presented in Figure 10. The estimates of the $\hat{\mu}(t)$ for $\delta^{18}O$ under Gaussian and the general error distribution \mathcal{H}_4 are given in Figure 11 and discussed in next section. Under the Cauchy distributional assumptions, almost all variation of the time series was attributed to the short-range variation, as it could be expected under so generous assumptions on the size of the allowable variation (the infinite variance) as is in the case with the Cauchy distribution. Due to the same reason different estimates of the long timescale variation $\hat{\mu}(t)$ are obtained assuming the t-distribution or the mixture of two normal densities as the distribution for $\hat{\mathcal{X}}_t, \hat{\mathcal{X}}_t^*$. The mixture of the two normal densities, with the parameters $\hat{\lambda} = 0.9$ and $\chi = 20$ fitted to the data, is less restrictive on the size of the variability for the cyclic model error than the t-distribution with $\nu = 4$ (see details in section 4.4.).

However, despite these different estimates of the long timescale variations, the systematic part of the variation $\hat{y}(t) = Z\hat{\alpha}(t) = \hat{\mu}(t) + \hat{\psi}(t)$ is nearly identical under the different model assumptions because of the nearly error-free observations: $\hat{y}(t) = y(t) - \hat{\epsilon}(t)$ with $\hat{\sigma}_{\epsilon}(t) \approx 0$.

4.2. The computation of a likelihood under the linear non-Gaussian state space model. The set of unknown parameters Θ^* (the specific ones ν under \mathcal{H}_1 , λ, χ under \mathcal{H}_2 , q under \mathcal{H}_4 and the general ones $\sigma_{\xi}, \sigma_{\eta}, \sigma_{\psi}, \lambda_c$ and ρ) are estimated by maximising the corresponding log-likelihood function $l(\Theta^*) = \log p(\mathcal{Y}; \Theta^*)$.

(13)
$$p(\mathcal{Y};\Theta^*) = \int p(\mathcal{Y} \mid \alpha;\Theta^*)p(\alpha;\Theta^*)d\alpha = \int \prod_{t=1}^n g(\epsilon(t);\Theta^*) \prod_{t=0}^{n-1} g(\eta_t;\Theta^*)h(\mathcal{X}_t;\Theta^*)h(\mathcal{X}_t;\Theta^*)d\eta_t d\mathcal{X}_t d\mathcal{X}_t^*$$

where $p(\cdot)$ denotes any probability density function, $g(\cdot)$ denotes a Gaussian density function and $h(\cdot)$ denotes one of the heavy-tailed density functions mentioned above. It is impossible to compute the exact value of log-likelihood $l(\Theta^*)$



FIGURE 11. The estimate of the long timescale variation $\hat{\mu}()t$ of $\delta^{18}O$ observations from the *NGRIP* ice core when the Gaussian distribution (the solid line) and the general error distribution \mathcal{H}_4 with q = 1.8 (the dashed line) are assumed for the cyclic model error. The observations are denoted by "*".

under a linear non-Gaussian state space model as it was done under a linear Gaussian one (see (6)). Still a close approximate value of the likelihood (13) can be computed using importance sampling (Ripley, 1987).

The idea is to use that

$$p(\mathcal{Y}; \Theta^*) = \int \frac{p(\mathcal{Y}, \alpha; \Theta^*)}{g(\alpha \mid \mathcal{Y})} g(\alpha \mid \mathcal{Y}) d\alpha$$

for any convenient sampling density $g(\alpha \mid \mathcal{Y})$ of the model state α , for example the one obtained under the time-dependent Gaussian approximate state space model to be described below (19).

Then

(14)
$$p(\mathcal{Y};\Theta^*) = g(\mathcal{Y};\Theta^*) \int \frac{p(\mathcal{Y},\alpha;\Theta^*)}{g(\mathcal{Y},\alpha;\Theta^*)} g(\alpha; | \mathcal{Y};\Theta^*) d\alpha = L_g(\Theta^*) E_g(\frac{p(\mathcal{Y},\alpha;\Theta^*)}{g(\mathcal{Y},\alpha;\Theta^*)}),$$

where $L_g(\Theta^*)$ is an exact likelihood under the time-dependent Gaussian approximating state space model computed by (6) and $E_g(p(\mathcal{Y}, \alpha; \Theta^*)/g(\mathcal{Y}, \alpha; \Theta^*))$ is the mathematical expectation of the functional $p(\mathcal{Y}, \alpha; \Theta^*)/g(\mathcal{Y}, \alpha; \Theta^*)$ with respect to the sampling density $g(\alpha \mid \mathcal{Y}; \Theta^*)$. Thus the likelihood $L(\hat{\Theta}^*)$ under the

TABLE 7. Estimated values of statistical and dynamical parameters describing the variation of the $\delta^{18}O$ time series from the *NGRIP* ice core under the Gaussian and under the heavy-tailed distributions assumed for the cyclic model error.

| $\mathcal{X}_t, \mathcal{X}_t^* \sim$ | $\hat{\sigma}_{\xi}$ | $\hat{\sigma}_{\psi}$ | $\hat{ ho}$ | $\hat{\rho}\cos(\hat{\lambda}_c)$ | $\hat{\rho}\sin(\hat{\lambda}_c)$ | $\sqrt{(1-\hat{ ho}^2)\hat{\sigma}_\psi^2}$ |
|---|----------------------|-----------------------|-------------|-----------------------------------|-----------------------------------|---|
| $N(0,\sigma_{\mathcal{X}}^2)$ | 0.010 | 1.489 | 0.768 | 0.749 | 0.170 | 0.953 |
| $\mathcal{H}_1, \ \nu = 4, \sigma_{\mathcal{X}}^2$ | 0.009 | 1.349 | 0.877 | 0.870 | 0.108 | 0.649 |
| $\mathcal{H}_2, \ \lambda = 0.9, \chi = 20, \sigma_{\mathcal{X}}^2$ | 0.007 | 1.011 | 0.865 | 0.865 | 0.012 | 0.507 |
| $\mathcal{H}_3, \sigma^2_{\mathcal{X}}$ | 0.006 | 1.195 | 0.906 | 0.906 | 0.024 | 0.506 |
| $\mathcal{H}_4, q = 1.8, \sigma_{\mathcal{X}}^2$ | 0.009 | 1.265 | 0.783 | 0.767 | 0.203 | 0.786 |

linear non-Gaussian model (12) can be approximated by

$$L(\Theta^*) = L_g(\Theta^*)\bar{w}$$

, where

$$\bar{w} = \frac{1}{N} \sum_{i=1}^{N} \frac{p(\mathcal{Y}; \alpha^i)}{g(\mathcal{Y}, \alpha^i)}$$

and where $\alpha^1, \alpha^2, \ldots, \alpha^N$ is the simulation sample generated by the sampling density $g(\alpha \mid \mathcal{Y}; \Theta^*)$. The current methodology originates from Durbin and Koopman, 2001, and Shephard and Pitt, 1997. In order to improve the precision of the $L(\Theta^*)$ estimation, the dependent sample of α^i , $i = 1, \ldots, 4 * N'$, with four antithetic variables was generated (Ripley, 1987; Durbin and Koopman, 1997). The sample was balanced for scale and for location.

As an example, Table 7 shows a number of important statistical and dynamical parameter estimates for the $\delta^{18}O$ time series from the NGRIP ice core. The parameters were estimated both under the linear Gaussian state space model (12) and under the heavy-tailed distributional assumptions on the cyclic error component $\mathcal{X}_t, \mathcal{X}_t^* \sim \mathcal{H}_k, \ k = \{1, 2, 3, 4\}$. To obtain the estimates both the exact likelihood (6) under the linear Gaussian state space model and the approximate likelihood (14) under the "heavy-tailed" state space model were maximised numerically. The approximate likelihood was computed using the simulation sample $\alpha^i, \ i = 1, \ldots, 4*N$ with N = 200. Under all distributional assumptions $\sigma_{\mathcal{X}}^2 = (1 - \rho^2)\sigma_{\psi}^2$.

One important aspect should be mentioned in connection with Table 7. As will be discussed later, the heavy-tailed assumption on the cyclic model error component distribution (\mathcal{H}_k , $k = \{1, 2, 3, 4\}$) results in a kind of local model fitting. The Cauchy distribution(\mathcal{H}_3) has the strongest local model fitting effect and the general error distribution with parameter q = 1.8 (\mathcal{H}_4) has the weakest such effect among the heavy-tailed distributions discussed in this paper. The Gaussian assumption on the distribution provides a global model fitting, for which all available observations influence the estimate of parameters.

It was mentioned in section 3.1 that the value of the parameter $\hat{\rho}$ contains important information about the dynamical model fitting. The closer the "damping factor" ρ is to 1, the better the wave-like process captures the time development of the short timescale variation of the background signal. As it can be seen from the Table 7, the estimate of $\hat{\rho}$ has a larder value under each of the heavy-tailed distributional assumptions on the cyclic model error than under the Gaussian assumption. The largest value of $\hat{\rho}$ is obtained assuming the Cauchy distribution for the cyclic model error. Indeed, under the Cauchy distribution nearly all variability is attributed to stochastic drift that does not have any restrictions on the amplitude. This, of course, does not guarantee the proper statistical model fitting.

Figure 11 illustrates the smoothed trend estimate $\hat{\mu}(t)$ of the $d^{18}O$ series from the **NGRIP** ice core under the Gaussian linear model (5) (the solid line) and under model (12) with the general error distribution assumed for the cyclic model error $\mathcal{X}_t, \mathcal{X}_t^* \sim \mathcal{H}_4, q = 1.8$ (the dashed line). The value q = 1.8 was estimated from the data by numerical maximisation of the non-Gaussian likelihood (13). In fact the same value of the parameter q = 1.8 was estimated for all six time series. This value makes the distribution very close to the Gaussian one that corresponds to q = 2. This model gives the total variation decomposition into the long and the short timescale variations very similar to the one under the time-invariant linear Gaussian model. The distributional assumption \mathcal{H}_4 was excluded from further investigations in this paper. We should mention, that the non-Gaussian log-likelihood seems to be very flat in q and small deviations from the chosen value q does not influence the evaluation of the non-Gaussian log-likelihood.

4.3. A time-dependent Gaussian state space model approximation. Given the maximum likelihood estimate of the model parameteres Θ , the decomposition of the total variation into the long timescale and the short timescale variation components is performed. In the case of the linear non-Gaussian models (12), the maximum posterior estimate of the trend and the cyclic development $\hat{\mu}(t)$ and $\hat{\psi}(t)$, $t = 1, \ldots, n$, are not identical to the posterior mean. Still the Kalman filter and the Kalman smoother recursions can be used as efficient computational tools for estimating the maximum posterior density of the non-Gaussian model. The posterior density for model (12) is defined by (7) in a similar way as in a linear Gaussian model (e.g. Fahrmeir and Tutz, 1994). In order to use the Kalman filter and the smoother recursions as computational tools to maximise the posterior density, the non-Gaussian heavy-tailed densities should be approximated by Gaussian ones in such a way that the system of equations which determines the posterior maximum

(15)
$$\nabla_{\alpha} \log(p(\alpha \mid \mathcal{Y})) = \nabla_{\alpha} \log(p(\alpha)) + \nabla_{\alpha} \log(p(\mathcal{Y} \mid \alpha)) = 0$$

has a form similar to the linear system of equation (9). We can say that instead of treating the original non-Gaussian model (12), we construct an approximate Gaussian model which will have the same posterior mode $\hat{\alpha}$ as the original non-Gaussian model. Then we will use the Kalman filter and Kalman smoother recursions to compute the posterior mode, which is identical to the posterior mean of the approximate Gaussian model. The methodology of solving equation (15), used in this paper, follows the methodology proposed by Durbin and Koopman (1997). Alternative possibilities to solve non-Gaussian state space models are given by Fahrmeir and Kaufmann (1991) and Fahrmeir and Wagenpfeil (1997).

It is useful to notice that all densities \mathcal{H}_k , like the Gaussian density

$$g(u; \sigma^2) = \frac{1}{\sqrt{(2\pi)}} \exp\left(-\frac{u^2}{2\sigma^2}\right),$$

are unimodal around 0.

The first-order derivative of the Gaussian log-density of the stochastic variable u with respect to the variable u is a linear function of variable u itself

$$\frac{d\log(g(u;\sigma^2))}{du} = -\frac{u}{\sigma^2}$$

One can express the first-order derivative of a heavy-tailed log-density of $h_k(u)$, in a similar form,

(16)
$$\frac{d\log(h_k(u))}{du} = -\frac{u}{\sigma_k^2(u)}$$

where $\sigma_k^2(u)$ now is not a constant but a certain function of the stochastic variable u.

Under the heavy-tailed distributional assumptions on the cyclic model error \mathcal{X}_t , $\mathcal{X}_t^* \sim \mathcal{H}_k$, and assuming that the model error vector components are mutually independent, the log-posterior density of the model state α will differ by an additive constant from

(17) $\log(p(\alpha \mid \mathcal{Y})) = constant$

$$-\frac{\mu(0)^2}{2K} - \frac{\mu(1)^2}{2K} - \frac{1}{2} \sum_{t=2}^n \frac{(\mu(t) - 2\mu(t-1) + \mu(t-2))^2}{\sigma_\xi} + \log h_k(\psi(0)) + \sum_{t=1}^n \log h_k(\psi(t) - \rho \cos(\lambda_c)\psi(t-1) - \rho \sin(\lambda_c)\psi^*(t-1)) + \log h_k(\psi^*(0)) + \sum_{t=1}^n \log h_k(\psi^*(t) + \rho \sin(\lambda_c)\psi(t-1) - \rho \cos(\lambda_c)\psi^*(t-1))$$

For simplicity, the dependency of the posterior density $p(\alpha \mid \mathcal{Y})$ on the observation error vector ϵ has been omitted, because a more parsimonious model with the perfect observation fit was suggested for the $\delta^{18}O$, $\log Ca^{2+}$ and $\log SO_4^{2-}$, using the Akaike information criterion, also motivated by the small contribution to model from the observation error component, for $\delta^{18}O$, $\log Ca^{2+}$ and $\log SO_4^{2-}$, as it is summarised in Table 3.

Taking into account expression (16) for the first-order derivative of the heavytailed log-densities $\log h_k(u)$ with respect to the stochastic variable u, the non-linear system of equations (15) can be written in a form similar to the system of linear equations (9),

(18)
$$\frac{\partial p(\alpha \mid \mathcal{Y})}{\partial \mu(t)} = -\frac{\xi_{t-1}}{\sigma_{\xi}^{2}} + 2\frac{\xi_{t}}{\sigma_{\xi}^{2}} - \frac{\xi_{t+1}}{\sigma_{\xi}^{2}},$$
$$\frac{\partial p(\alpha \mid \mathcal{Y})}{\partial \psi(t)} = -\frac{\mathcal{X}_{t-1}}{\sigma_{k}^{2}(\mathcal{X}_{t-1})} + \rho \cos(\lambda_{c})\frac{\mathcal{X}_{t}}{\sigma_{k}^{2}(\mathcal{X}_{t})} - \rho \sin(\lambda_{c})\frac{\mathcal{X}_{t}^{*}}{\sigma_{k}^{2}(\mathcal{X}_{t}^{*})},$$
$$\frac{\partial p(\alpha \mid \mathcal{Y})}{\partial \psi^{*}(t)} = -\frac{\mathcal{X}_{t-1}^{*}}{\sigma_{k}^{2}(\mathcal{X}_{t-1}^{*})} + \rho \sin(\lambda_{c})\frac{\mathcal{X}_{t}}{\sigma_{k}^{2}(\mathcal{X}_{t})} + \rho \cos(\lambda_{c})\frac{\mathcal{X}_{t}^{*}}{\sigma_{k}^{2}(\mathcal{X}_{t}^{*})}$$

for t = 2, ..., n - 2, together with the equations for the boundary points t = 0, 1and t = n - 1, n.

The log-likelihood for the non-Gaussian model (12) yields approximately the same equation system as if we had a Gaussian model with the time-varying variance $\sigma_k^2(\hat{X}_t), \sigma_k^2(\hat{X}_t^*)$ for the cyclic model error.

(19)

$$y(t) = Z\alpha(t), \quad 1 \le t \le n$$

$$\alpha(t+1) = T\alpha(t) + \begin{pmatrix} \xi_t \\ \mathcal{X}_t \\ \mathcal{X}_t^* \end{pmatrix}, \quad \begin{array}{l} \xi_t \sim N(0, \sigma_k^2) \\ \mathcal{X}_t \sim N(0, \sigma_k^2(\hat{\mathcal{X}}_t)) \\ \mathcal{X}_t^* \sim N(0, \sigma_k^2(\hat{\mathcal{X}}_t)) \end{array}; \quad t > 0$$

where $\sigma_k^2(\hat{\mathcal{X}}_t)$ $k = \{1, 2, 3\}$ are time-dependent functions of the maximum likelihood estimate of the cyclic model error (also called *the smoothed cyclic model residuals*) and they are determined by the form of the heavy-tailed distributions \mathcal{H}_1 - \mathcal{H}_3 , defined above.

Indeed, the Kalman filter and smoother recursions can be used to solve iteratively the non-linear equation system (18) and to compute the posterior mode of the approximating Gaussian model (eqn. 19) with the time-dependent variance of the cyclic model error. Here we just sketch the idea of this method. During *i*th iteration, given a currently available estimate of the posterior mode of the model state vector α^{i-1} or , equivalently, given a currently available estimate of the posterior mode of the model error vector $\eta^{i-1} = \{\xi_t^{i-1}, \mathcal{X}_t^{i-1}, \mathcal{X}_t^{*,i-1}, t = 0, \ldots, n-1\}$, the model (eqn. 19) is approximated by an analogous model with the time-varying variance $\sigma_k^2(\mathcal{X}_t^{i-1}), \sigma_k^2(\mathcal{X}_t^{*,i-1})$. The Kalman filter and the Kalman smoother are used to obtain new estimate of the posterior mode of the model state α^i or the model error η^i vector. The iterations are repeated until convergence, i.e. $\eta^{i-1} = \eta^i := \hat{\eta}$ ($\alpha^i = \alpha^{i-1} := \hat{\alpha}$). The solution of the iterative procedure of the equation system (eqn. 18) is the posterior mode of both the approximate time-dependent Gaussian model (eqn. 19) and the original non-Gaussian model (eqn. 12). For the detailed description of the iterative method see Durbin and Koopman, 1997.

Figures 12, 13 and 14 show diagnostic plots of the model fit for the approximate time-dependent linear Gaussian state space model (19) derived from the original non-Gaussian state space model: \mathcal{H}_1 (Figure 12), \mathcal{H}_2 (Figure 13) and \mathcal{H}_3 (Figure 14). Solid curves represents the time series obtained from the *NGRIP* ice core and dashed curves the ones from the *GISP2* ice core. Similar plots under the time-invariant linear state space model are shown in Figure 7 (see section 3.1 where the diagnostics are introduced and discussed).

It should be mentioned, that the values of the parameters Θ^* in the definition of the actual non-Gaussian linear state space model (both the specific and the general ones) were determined by numerically maximising the corresponding non-Gaussian likelihood. The solution of the maximisation procedure is not unique. The solution with the largest possible "damping factor" $\hat{\rho}$ was chosen among the "equally-likely" ones. The estimated values of the parameters Θ^* involved in the definition of the models are summarised in Table 7.

As it can be seen from Figures 12 and 13, especially comparing with Figure 7, the statistical model fit is improved under the time-dependent linear Gaussian model derived from the original non-Gaussian model, given by equation (12) with \mathcal{H}_1 or \mathcal{H}_2 . The time-dependent Gaussian models with the local variance functions $\sigma_1^2(\hat{\mathcal{X}}_t)$ and $\sigma_2^2(\hat{\mathcal{X}}_t)$ provide a more proper normalisation of the smoothed cyclic residuals, such that the distribution of the standardised $\hat{\mathcal{X}}$ becomes close to a Gaussian one. The proper weighting of the smoothed cyclic residuals (the third column of the diagrams), especially for its large values, improves the fit of the statistical model. This can be seen from the QQ-plot of the standardised innovations (the first column of the diagrams) for all six time series. The time-dependent linear Gaussian model derived from (12) with \mathcal{H}_3 (the Cauchy distribution) assumed for the cyclic mode error component cannot adequately describe the data. The down-weighting of large values of the smoothed cyclic residuals by the local variance function $\sigma_3^2(\hat{\mathcal{X}}_t)$ is so strong that the distributions of the standardised $\hat{\mathcal{X}}$ and the standardised innovations become too light-tailed (Figure 14).

The result of the variation decomposition obtained by assuming the mixture of two normal densities as a distribution for the $\mathcal{X}_t, \mathcal{X}_t^*$ should be interpreted with great care (Figure 13). The trend component residuals seem to have lighter tails under \mathcal{H}_2 than under \mathcal{H}_1 in comparison with their theoretical quantiles. The estimate of the trend under \mathcal{H}_2 is over-smoothed for the log(Ca^{2+}) and the log(SO_4^{2-}) time series because as much as 15% of the observations ($\lambda = 0.85$) are assumed to belong to the "alternative regime" with the strong deviations from the background signal



FIGURE 12. The QQ-plots of the standardised innovations \tilde{v}_t (the first column of diagrams), the standardised smoothed trend residuals $\tilde{\xi}_t$ (the second column of the diagrams) and the standardised smoothed cyclic component residuals $\tilde{\mathcal{X}}_t$ (the third column of the diagrams) against their theoretical quantiles $\mathcal{N}(0, 1)$, under the approximate time-dependent Gaussian model (19) with \mathcal{H}_1 (the *Student* t) distributional assumptions on the cyclic model error. The diagnostics plots for the $\delta^{18}O$ time series are given in the first row of diagrams, for the $\log(Ca^{2+})$ time series in the second row of diagrams and for the $\log(SO_4^{2-})$ time series in the third row of the diagrams. Solid curves respresent time series from the *NGRIP* ice core and dashed curves ones from the *GISP2* ice core. The standardised innovations or standardised residuals are plotted along the y-axes and quantiles of $\mathcal{N}(0, 1)$ are plotted along the x-axes.

and enter through the stochastic drift $\mathcal{X}_t, \mathcal{X}_t^*$ only. This destroys the variation decomposition into the long and the short timescale variation as it can be seen from the estimated time-averaged autocorrelation function ρ_{τ} for the short cyclic model component $\psi(t)$ with a very long timescale variation being present. Figure 15 shows the time-averaged auto-correlation functions $\rho(\tau)$ of the short timescale variation component $\hat{\psi}(t)$ of the $\delta^{18}O$ (solid curve), the $\log(Ca^{2+})$ (dashed curve) and the $\log(SO_4^{2-})$ time series (dash-dotted curve) estimated under the time-invariant linear Gaussian model (the diagram to the left), under the time-dependent Gaussian models derived from the \mathcal{H}_1 (the middle diagram) and from the \mathcal{H}_2 (the diagram to the right) distributional assumptions on the $\mathcal{X}_t, \mathcal{X}_t^*$. The estimated auto-correlation functions for all three time series oscillate around non-zero value when the mixture



FIGURE 13. The QQ-plots of the standardised innovations \tilde{v}_t (the first column of diagrams), the standardised smoothed trend residuals $\tilde{\xi}_t$ (the second column of the diagrams) and the standardised smoothed cyclic component residuals $\tilde{\mathcal{X}}_t$ (the third column of the diagrams) against their theoretical quantiles $\mathcal{N}(0,1)$ under the time-dependent Gaussian model (19) with \mathcal{H}_2 (the mixture of two normal densities) distributional assumptions on $\mathcal{X}_t, \mathcal{X}_t^*$. The diagrams, for the $\log(Ca^{2+})$ time series are given in the first row of diagrams and for the $\log(SO_4^{2-})$ time series in the third row of the diagrams. Solid curves represent time series from the NGRIP ice core and dashed curves the ones from the GISP2 ice core.

of two normal densities (\mathcal{H}_2) was assumed as the distribution for $\mathcal{X}_t, \mathcal{X}_t^*$. This indicates the presence of a long timescale variation in the estimate of the short timescale variation.

In Table 8 we present a summary of the Box-Ljung statistics Q(15) for the check of the serial correlation of the standardised innovations \tilde{v}_t under the time-invariant linear Gaussian model (\mathcal{N}) and the time-dependent Gaussian models derived from the non-Gaussian assumptions on the $\mathcal{X}_t, \mathcal{X}_t^* \colon \mathcal{H}_1$ (Student t-distribution), \mathcal{H}_2 (the mixture of two normal densities) and \mathcal{H}_3 (the Cauchy distribution). One can notice the influence of the iterative solution of the system of the non-linear equations (18), which determine the mode of the posterior density. The observed value of the Box-Ljung statistics Q(15), based on the first 15 autocorrelations of \tilde{v}_t , is larger under all the time-dependent Gaussian models than under the time-invariant Gaussian model. Still the observed values of Q(15) presented in Table 8, are not large enough to indicate the presence of strong autocorrelations among the corresponding series



FIGURE 14. The QQ-plots of the standardised innovations \tilde{v}_t (the first column of diagrams), the standardised smoothed trend residuals $\tilde{\xi}_t$ (the second column of the diagrams) and the standardised smoothed cyclic component residuals $\tilde{\mathcal{X}}_t$ (the third column of the diagrams) against their theoretical quantiles $\mathcal{N}(0, 1)$ under the time-dependent linear Gaussian model (19) with \mathcal{H}_3 (the Cauchy) distributional assumptions on $\mathcal{X}_t, \mathcal{X}_t^*$. The diagnostics plots for the $\delta^{18}O$ time series are given in the first row of diagrams, for the $\log(Ca^{2+})$ time series in the second row of diagrams. Solid curves represent time series from the NGRIP ice core and dashed curves the ones from the GISP2 ice core.

of the \tilde{v}_t (compare with Figure 4), although weakly positive correlations appear due to the non-linearity of the system (eqn. 9). The highest value of Q(15) as well as the largest absolute values of the autocorrelation are observed among the innovations \tilde{v}_t corresponding to the $\log(Ca^{2+})$ time series, which have the worst fit to the linear Gaussian model (see diagnostic diagrams in Figures 12, 13 and 7). Notice that the Ca^{2+} series were not used explicitly in the synchronisation process (see section 2).

The diagnostics of the statistical model fit discussed above concerns the approximate time-dependent Gaussian model, given by equation (19), only. To investigate if the original non-Gaussian model, given by equation (12), fits the time series, a more advanced and comprehensive analysis using simulation techniques must be performed. On the other hand, it is the approximate time-dependent Gaussian model that determines the mechanism of the total variation decomposition into different timescale variation components.



FIGURE 15. The autocorrelation functions $\rho(\tau)$ of the cyclic model component $\psi(t)$ of the time series of $\delta^{18}O$ (solid curve), $\log(Ca^2)$ (dashed curve) and $\log(SO_4^{2-})$ (dash-dotted curve) from the GISP2 ice core plotted as functions of lag τ under different assumptions on the distribution of the cyclic model error: a Gaussian one (the left diagram), a *t-Student* distribution (the middle diagram) and a mixture of two normal densities (the diagram to the right).

TABLE 8. The Box-Ljung statistics based on the first 15 autocorrelation of \tilde{v}_t corresponding to the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series from the NGRIP ice core under the time-invariant linear Gaussian model and under the time-dependent Gaussian models derived from \mathcal{H}_1 , \mathcal{H}_2 and \mathcal{H}_3 assumed as a distribution for $\mathcal{X}_t, \mathcal{X}_t^*.$

| distribution | $d^{18}O$ | | | $\log(Ca^{2+})$ | | | $\log(SO_4^{2-})$ | | |
|-----------------|-----------|-------------|--------------|-----------------|-------------|--------------|-------------------|-------------|--------------|
| distribution | Q(15) | $F_{Q(15)}$ | $c(j_{max})$ | Q(15) | $F_{Q(15)}$ | $c(j_{max})$ | Q(15) | $F_{Q(15)}$ | $c(j_{max})$ |
| \mathcal{N} | 8.627 | 0.111 | -0.067 | 16.050 | 0.634 | -0.094 | 12.305 | 0.347 | -0.089 |
| \mathcal{H}_1 | 13.566 | 0.444 | 0.075 | 26.753 | 0.967 | 0.155 | 12.782 | 0.380 | 0.095 |
| \mathcal{H}_2 | 11.356 | 0.291 | 0.087 | 20.935 | 0.858 | 0.142 | 10.813 | 0.239 | 0.091 |
| \mathcal{H}_3 | 13.264 | 0.421 | 0.103 | 21.216 | 0.867 | 0.146 | 18.849 | 0.778 | 0.127 |

4.4. The local variance function. The local variance functions $\sigma_k^2(\hat{\mathcal{X}}_t)$, in the formulation of the time-dependent Gaussian approximating model (equation (19)), depend on the chosen distributions for $\mathcal{X}_t, \mathcal{X}_t^* \sim \mathcal{H}_k, \ k = \{1, 2, 3\}$ and are given by

• $\sigma_1^2(u)$ is corresponding to a *t*-density \mathcal{H}_1 with ν degrees of freedom and a variance (a scaling parameter) $\sigma_{\mathcal{X}}^2$

$$\sigma_1^2(u) = \frac{1}{\nu+1}u^2 + \frac{\nu-2}{\nu+1}\sigma_{\mathcal{X}}^2$$

• $\sigma_2^2(u)$ is corresponding to a mixture of two normal densities \mathcal{H}_2 with parameters λ , χ and a basic variance $\sigma_{\mathcal{X}}^2$

$$\sigma_2^2(u) = \frac{\sigma_{\mathcal{X}}^2 \lambda \chi \sqrt{\chi} \exp(-u^{*2}) + \chi \sigma_{\mathcal{X}}^2 (1-\lambda) \exp(-u^{*2}/\chi)}{\lambda \chi \sqrt{\chi} \exp(-u^{*2}) + (1-\lambda) \exp(-u^{*2}/\chi)}$$

where $u^{*2} = \frac{u^2}{2\sigma_{\chi}^2}$. • $\sigma_3^2(u)$ is corresponding to a Cauchy distribution \mathcal{H}_3 with a scaling parameter σ^2_{χ}



 $\sigma_3^2(u) = 0.5(u^2 + \sigma_\mathcal{X}^2)$

FIGURE 16. Diagram to the left: Plot of the local variance functions $\sigma_k^2(\hat{\mathcal{X}}_t)$ (on the y-axes) of the time-dependent Gaussian model derived from the non-Gaussian assumptions on the $\mathcal{X}_t, \mathcal{X}_t^* \sim \mathcal{H}_k, \ k = \{1, 2, 3\}$ plotted as functions of the smoothed cyclic residuals $\hat{\mathcal{X}}_t$ (on the *x*-axes): $\sigma_1^2(\hat{\mathcal{X}}_t)$ ("+"), $\sigma_2^2(\hat{\mathcal{X}}_t)$ ("*") and $\sigma_3^2(\hat{\mathcal{X}}_t)$ ("o"). The plot corresponds to the log(Ca^{2+}) time series from the NGRIP ice core. The global estimate of the variance of the cyclic model error $\sigma_{\mathcal{X}}^2 = 0.141$ (from a Gaussian assumption on the error distribution, solid line). Diagram to the right: Plot of the smoothed estimates of the long timescale variation $\hat{\mu}(t)$ for the $\log(Ca^{2+})$ time series from the NGRIP ice core under the different distributional assumptions on the $\mathcal{X}_t, \mathcal{X}_t^*$: the Gaussian distribution (solid curve), the *Student* t-distribution \mathcal{H}_1 with $\nu = 4$ degrees of freedom (dashed curve), the mixture of two normal densities \mathcal{H}_2 with $\lambda = 0.85$ and $\chi = 20$ (dash-dotted curve) and the Cauchy distribution \mathcal{H}_3 (dotted curve). The raw observations are denoted by "x"

As one can see, all local variance functions $\sigma_k^2(\hat{\chi}_t)$ are some kind of a compromise between the global estimate of the variance $\sigma_{\mathcal{X}}^2$ of the cyclic model error component \mathcal{X}_t and the momentary estimates of the variance being simply $(\hat{\mathcal{X}}_t)^2$. One example of the local variance functions corresponding to the $\log(Ca^{2+})$ time series from the NGRIP ice core is shown in Figure 16 (the diagram to the left). The corresponding variation decomposition is shown in Figure 16 (the diagram to the right).

The local variance functions corresponding to a *t*-distribution $\sigma_1^2(\hat{\mathcal{X}}_t)$ and to a Cauchy distribution $\sigma_3^2(\hat{\mathcal{X}}_t)$ are additive combinations of the *momentary* and the global variance estimates, with increasing weight of the *momentary* variance estimate for the large values of $\hat{\mathcal{X}}_t$, (see Figure 16). The local variance function $\sigma_3^2(\hat{\mathcal{X}}_t)$ is almost completely dominated by $(\hat{\mathcal{X}}_t)^2$ for large values of $\hat{\mathcal{X}}_t$.

A very interesting pattern has the local variance function corresponding to the mixture of two normal densities $\sigma_2^2(\hat{\mathcal{X}}_t)$, which splits the smoothed cyclic residuals by assigning to them the basic variance $\sigma_{\mathcal{X}}^2$ (for the small values of $\hat{\mathcal{X}}_t$) or the outlier variance $\chi \sigma_{\mathcal{X}}^2$ (for the large values of $\hat{\mathcal{X}}$) or a linear combination of both of them

with weights exponentially depending on the momentary variance estimate $(\hat{\mathcal{X}}_t)^2$ (see Figure 16). The global estimate of the variance $\hat{\sigma}^2_{\mathcal{X}}$ is equal 0.141 and is obtained under the time-invariant linear Gaussian model.

It is important to notice that even if the same parameter σ_{χ}^2 enters in the formulation of all local variance functions σ_k^2 , $k = \{1, 2, 3\}$, the estimates of the parameter are different under the different distributional assumptions on the cyclic model error. The estimate of the parameter $\hat{\sigma}_{\chi}^2$ is smaller under each of the investigated heavy-tailed assumptions on the cyclic model error component than under the Gaussian one. This is exactly as it should be under a more robust measure of the error, for which the eventually observed large values have a smaller impact on the variance estimate.



FIGURE 17. **Diagram to the left:** Plots of the local variance functions $\sigma_k^2(\hat{X}_t)$, k = 1 (upper plot), k = 2 (middle plot), k = 3(lower plot), corresponding to the $\log(SO_4^{2-})$ time series from the *NGRIP* ice core. The global estimate of the variance of the cyclic model error is $\hat{\sigma}_{\chi}^2 = 0.144$ under the time-invariant linear Gaussian model. **Diagram to the right:** Plot of the different smoothed estimates of the long timescale variation $\hat{\mu}(t)$ for the $\log(SO_4^{2-})$ time series from the *NGRIP* ice core under different distributional assumptions on the $\mathcal{X}_t, \mathcal{X}_t^*$: the Gaussain distribution (solid curve), the *Student t*-distribution \mathcal{H}_1 with $\nu = 4$ degrees of freedom (dashed curve), the mixture of two normal densities \mathcal{H}_2 with $\lambda = 0.85$ and $\chi = 10$ (dash-dotted curve) and the Cauchy distribution \mathcal{H}_3 (dotted line). Raw observations are denoted by "x".

A local variance function can be considered as a temporal (time-dependent) measure of the uncertainty about the dynamical development of $\psi(t)$. Figure 17 (the diagram to the left) shows the local variance functions in the time-dependent linear Gaussian models, derived from the different distribution assumptions, as functions of time. Figure 17 corresponds to the $\log(SO_4^{2-})$ time series from the NGRIP ice core. The corresponding variation decomposition is shown in Figure 17 (the diagram to the right).

The local variance function plays an essential role in the total variation decomposition into the long $(\hat{\mu}(t))$ and short timescale $(\hat{\psi}(t))$ variations. It is important to notice that, under this time-dependent measure of the uncertainty about the dynamical development of the $\psi(t)$, not only do the large values of the smoothed cyclic residuals $\hat{\mathcal{X}}_t$ indicate larger (temporal) uncertainty about the development of the cyclic component $\hat{\psi}(t)$ in comparison to the global estimation of the uncertainty. At the same time do the small values of the smoothed cyclic residuals indicate larger relative uncertainty about the development of the long timescale variation $\mu(t)$.

The simple dynamical model used in this paper exaggerates the influence of the model error specification on the variation decomposition. At the same time it helps to outline one of the basic properties of Kalman filtering and Kalman smoothing. The larger the relative uncertainty about a particular model state component is in comparison to other model state components, the larger will the impact of newly in-coming observations be on just that component. In other words, the data assimilation by a linear Gaussian model will always work for a reduction of the uncertainty.

When a Gaussian distribution is assumed for $\mathcal{X}_t, \mathcal{X}_t^*$, a global smoothing is performed, under which all observations y(t) of the time series have the same impact on $\hat{\mu}(t)$ and $\hat{\psi}(t)$, as soon as the steady state solution (see next section) is achieved. Under a time-dependent Gaussian model, derived from the non-Gaussian assumption on the distribution of the cyclic model error, a local smoothing is performed. The impact of each observation y(t) on the estimated long timescale $\hat{\mu}(t)$ and short timescale $\hat{\psi}(t)$ variations depends on its similarity with the surrounding values in time, expressed through the local variance function in the model state space. A scaled innovation, the one-step-forecast error $v(t) = y(t) - \hat{y}(t)$ scaled by its variance F(t), is the most natural measure, taken in observational space, of how exceptional the observation y(t) is in comparison with its surrounding. As shown later, the smoothed cyclic residual $\hat{\mathcal{X}}_{t-1}$ is in a nearly deterministic linear relationship with the innovation v(t). This is why the local variance function $\sigma_k^2(\hat{\mathcal{X}}_{t-1})$, being a symmetric function of $\hat{\mathcal{X}}_{t-1}$, is a measure of dissimilarity of the observation y(t) to its neighbourhood, taken in model state space.

Three local variance functions $\sigma_1^2(\hat{\mathcal{X}}_t)$, $\sigma_2^2(\hat{\mathcal{X}}_t)$ and $\sigma_3^2(\hat{\mathcal{X}}_t)$, which are different measures of dissimilarity of the $\log(SO_4^{2-})$ observations to their neighbourhoods, are illustrated in Figure 17 (the diagram to the left). All three time series are observed on the NGRIP ice core. Each of these measures recognizes both regimes from which the observations originate. The majority of observations represent the "background signal". They are modelled explicitly through the dynamical propagator T, are well predicted and they induce small (scaled) innovations. As a result, a small temporal uncertainty will be assigned to the dynamical development of the cyclic model component just before the observation comes, and thus a large relative uncertainty will be assigned to the dynamical development of the trend component at that moment. That observation will thus be an influential source of information for the long timescale variation $\hat{\mu}(t)$. The observations from the alternative regime, extraordinary observations in comparison with their neighbourhoods, are not modelled explicitly, and they induce large scaled innovations, and thus a large temporal uncertainty will be assigned to the dynamical development of the cyclic model component just before the observation comes. The extraordinary observations will therefore have a strong impact on the short timescale variation $\psi(t).$

When the *Student* t-distribution is assumed for the cyclic model error component, the local variance function $\sigma_1^2(\hat{\mathcal{X}}_t)$ increases slowly with increasing v(t+1). Thus a few large values of the local variance function are induced by the most extraordinary observations, at the same time as the well predictable majority of observations causes a lot of small values of the local variance function. As a result, the estimated trend under the time-dependent Gaussian model derived from $\mathcal{X}_t, \mathcal{X}_t^* \sim \mathcal{H}_1$ has more energy on shorter scales due to the extra strong influence of the well predictable observations than under the Gaussian assumption on the



FIGURE 18. The spectral density functions $f(\omega)$ of the autocorrelations of the trend component residuals $\hat{\xi}_t$ of the time series of $\delta^{18}O$ (the solid line), $\log(Ca^{2+})$ (the dashed line) and $\log(SO_4^{2-})$ (the dash-dotted line) from NGRIP ice core plotted as functions of wave number ω under different assumptions on the distribution of the cyclic model error: a Gaussian one (the left diagram), a *t-Student* distribution (the middle diagram) and a mixture of two normal densities (the diagram to the right.)

distribution. If the mixture of two normal densities is assumed for the $\mathcal{X}_t, \mathcal{X}_t^*$, the $\sigma_2^2(\hat{\mathcal{X}}_t)$ sharply increases in magnitude as soon as the observation y(t+1) from the *alternative* regime is waiting to come. So all observations from the *alternative* regime are prohibited to influence the estimate of the long timescale variation and they enter directly in the estimate of the short timescale variation. The estimate of the long timescale variation becomes over-smoothed, as can be seen from the spectral densities of the autocorrelation function of the trend component residuals $\overline{\rho_{\tau}(\hat{\xi}_t, \hat{\xi}_{t+\tau})}$, shown in Figure 18. Even more, the process of total variation decomposition fails. As we already have mentioned above, the estimate of the short timescale variation clearly contains long timescale variation as can be seen from the time-averaged estimated autocorrelations $\rho(\tau)$ of the short timescale variation $\hat{\psi}(t)$, which oscillate around a non-zero value for all three time series, (see Figure 15, the diagram to the right).

Even worse results of the total variation decomposition into different timescales are obtained with the Cauchy distribution for which the $\sigma_3^2(\hat{\mathcal{X}}_{t-1})$ increases dramatically with v(t). The estimation of the trend becomes un-acceptably over-smoothed and clearly too much variation is attributed to the short timescale variation, see Figure 17.

4.5. The local smoothing mechanism. As already mentioned above, a linear Gaussian model (5) provides a global treatment of the time series. The reason is that all parameters, which define the model, are constant in time. Linear Gaussian models have an exceptional property that innovations (or the one-step-ahead forecast errors $v(t) = y_t - ZE(\alpha(t) | \mathcal{Y}_{t-1}))$ are mutually independent (Harvey, 1989). This very strong property has the consequence that the conditional variance of the innovations, $F(t) = Var(v(t) | \mathcal{Y}_{t-1}) = Var(v(t))$, is free from dependence on data \mathcal{Y}_{t-1} .

Even conditional variances of the model state and model error, given observations, are not dependent on the observations (except through the parameter estimates). The time development of the conditional mean and variance of the model state, given observations, is governed by the Kalman filter and the smoother recursions. The forward Kalman filter recursions provide a sequential update of the knowledge about the model state each time t a new observation y_t is available and express the time evolution of the two first conditional moments of the model state given the set of observations available up to time t (Durbin and Koopman, 2001).

(20)

$$E(\alpha(t+1) \mid \mathcal{Y}_{t}) := a_{t+1} = Ta(t) + K_{t}v(t),$$

$$Var(\alpha(t+1) \mid \mathcal{Y}_{t}) := P_{t+1} = TP_{t}T' - TP_{t}Z'(ZP_{t}Z' + \sigma_{\epsilon}^{2})^{-1}ZP_{t}T' + \Omega$$

$$K_{t}^{*} = P_{t}Z'(ZP_{t}Z' + \sigma_{\epsilon}^{2})^{-1},$$

$$K_{t} = TK_{t}^{*},$$

$$v(t) = y_{t} - Za(t),$$

$$Var(v(t)) := F(t) = ZP_{t}Z' + \sigma_{\epsilon}^{2}, \ t = 1, \dots, n$$

The backward Kalman smoothing recursions finalise the estimation of the model state when the whole set of observations \mathcal{Y} is given. For example, the smoothed model state estimate $\hat{\alpha}(t)$ is equal to the conditional mean of the model state, given the whole set of observations under the assumption of a linear Gaussian model. The smoothed model state estimate $\hat{\alpha}(t)$ consists of the filtered value and the linear influence of future innovations. Notice that the filtered value itself is a certain linear combination of the history (previously occurred) innovations.

$$E(\alpha(t) \mid \mathcal{Y}) := \hat{\alpha}(t) = a_t + \sum_{j=0}^{n-t} c_j^t v(t+j)$$

= $a_s + \sum_{j=1}^{t-s} c_{-j}^t v(t-j) + c_0^t v(t) + \sum_{j=1}^{n-t} c_j^t v(t+j), \quad 0 \le s < t \le n$

The influence from the innovations can be computed from the output of the Kalman filter recursions (20)

$$c_0^t = K_t^* = P_t Z'(F(t))^{-1}$$

$$c_j^t = P_t L'_t \dots L'_{t-1+j} Z'(F(t+j))^{-1}, \quad j = 1, \dots, n-t$$

$$c_{-j}^t = (T \cdot)_{j-1} K_{t-j}, \qquad j = 1, \dots, t-s$$

where $L_s = (T - K_s Z)$.

The notation $(T \cdot)_j$ stands for the *j*-times repeated application of the operator T. With the current definition of the dynamical propagator T, equation (5), we will have

$$T(\cdot)_j = \begin{pmatrix} j+1 & -(j) & 0 & 0\\ j & -(j-1) & 0 & 0\\ 0 & 0 & \rho^j \cos(j * \lambda_c) & \rho^j \sin(j * \lambda_c)\\ 0 & 0 & -\rho^j \sin(j * \lambda_c) & \rho^j \cos(j * \lambda_c) \end{pmatrix}$$

Under such a time-invariant linear state space model (5), the Kalman filter recursions (20) converge to a steady state solution after a certain number of updates. This steady-state solution is determined by a matrix equation for the conditional variance of the model state $P_{t+1} = Var(\alpha(t+1) | \mathcal{Y}_t) \equiv \tilde{P}, t \geq t_{conv}$, where t_{conv} is a time when the convergence of P_t to \tilde{P} (Durbin and Koopman, 2001; Harvey, 1989) has occured.

(22)
$$\tilde{P} = T\tilde{P}T' - T\tilde{P}Z'(Z\tilde{P}Z' + \sigma_{\epsilon}^2)^{-1}Z\tilde{P}T' + \Omega$$

As soon as the steady state solution is achieved (for $t \ge t_{conv}$), the output of the Kalman filter becomes constant.

Smoothing weights c_j^t , $-100 \le j \le 100$, corresponding to the steady state solution of the Kalman filter for the $\log(Ca^{2+})$ time series from the NGRIP ice



FIGURE 19. The smoothing weights of innovations c_j^t , $-100 \leq j \leq 100$ for the estimate of the smoothed model state for the $\log(Ca^{2+})$ time series from the *NGRIP* ice core under the timeinvariant linear Gaussian model. The smoothing weights of the innovations, lagging j time steps from the moment t, represent the influence of the innovation v(t + j) on the different components of the model state at time t, are shown as functions of the lag. The weights for the estimate of the trend component $\hat{\mu}(t)$ at moment t are denoted by a solid curve and the weights for the estimate on the short timescale variation $\hat{\psi}(t)$ by a dashed curve. The weights correspond to time moments when the steady state solution has been achieved.

core are given in Figure 19. The impact of innovations v(t + j), lagging j time steps from the moment t, on the smoothed estimate of the different components of the model state at time moment t, are shown as functions of the lag j. The solid line is corresponding to the influence of the standardised innovations on the long timescale variation $\hat{\mu}(t)$ and the dashed line is corresponding to the influence on the standardised innovations on the short timescale variation $\hat{\psi}(t)$.

As soon as the steady state solution (22) has been achieved $(t_{conv} \approx 30 \text{ to} achieve convergence with 3 decimal digits and <math>t_{conv} \approx 15$ to achieve convergence with 2 decimal digits), the influence of the innovation v(t + j) on the smoothed estimate of the model state components at moment t is dependent only on the time lag j and is completely independent of the size of the innovation itself. This property puts a strong requirement on the allowable size of the innovations. The historical innovations $(a(s) + \sum_{j=1}^{t-s} c_{-j}^t v(t-j) + c_0^t v(t), s < t)$ determine, in principle, the value of the estimate at moment t and the future innovations $\sum_{j=1}^{n-t} c_j^t v(t+j)$ provide the smoothness of the estimate as function of t. The presence of a number of large innovations, which correspond to extraordinary observations in comparison with the surrounding, can strongly bias the estimate of the trend toward the extraordinary observations. At the same time, the effect of the smoothing of the surrounding innovations in time.

A local smoothing, with the impact of the innovations depending on their size, is a better alternative to represent variations of such time series where extraordinary observations in comparison to their surrounding in time are present. A non-Gaussian linear model with a heavy-tailed assumption on the model error distribution, such as the one used in this paper to represent the variation of the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series from both ice cores, has the desirable property.



FIGURE 20. The Kalman gain of the trend $\mu(t)$ (the first row of the diagrams) and of the cyclic component $\psi(t)$ (the second row of the diagrams) under the time-dependent Gaussian model, derived from the non-Gaussian assumptions on the $\mathcal{X}, \mathcal{X}^* \sim \mathcal{H}_k, \quad k = 1, 2, 3$ (shown in the corresponding columns of the diagrams) as a function of the innovations v(t). The Kalman gain corresponds to the $\log(Ca^{2+})$ time series from the NGRIP ice core. Under a time-invariant linear Gaussian model $K_t^{trend} = 0.291$ and $K_t^{cyclic} = 0.684$.

The linear dependency of the $\hat{\mu}(t)$ and $\hat{\psi}(t)$ on the innovations (21) holds even under the time-dependent Gaussian approximating model (19). However, the dependency of $\hat{\mu}(t)$ and $\hat{\psi}(t)$ on the observations themselves is not linear anymore, but expressed via the system of non-linear equations (18).

Under the time-dependent Gaussian approximating model the Kalman filter does not converge to a steady state solution due to the local variance function $\sigma_k^2(\hat{\mathcal{X}}_t)$, k = 1, 2, 3. The dynamics of the conditional variance of $\hat{\psi}(t)$ is dominated by the time fluctuations of the local variance function, which is a function of the cyclic component residuals. One example of the time fluctuations of the local variance function $\sigma_k^2(\hat{\mathcal{X}}_t)$, k = 1, 2, 3 is shown on Figure 17 (the diagram to the left). The dynamics of the conditional variance of $\hat{\mu}(t)$ is mainly determined by the deterministic time propagator (σ_{ξ}^2 is quite small in comparison to other sources of variation).

$$P_{t+1} = TP_t'L_t + (\sigma_{\xi}^2, 0, \sigma_k^2(\hat{\mathcal{X}}_t), \sigma_k^2(\hat{\mathcal{X}}_t^*))'$$

The conditional variance P_t enters non-linearly into the Kalman gain and makes it dependent on the size of the cyclic component residual $\hat{\mathcal{X}}_t$.

(23)
$$(K*)_{t+1}^{trend} = \frac{d_1}{d_1 + d_2 + \sigma_{\epsilon}^2}$$
$$(K*)_{t+1}^{cyclic} = \frac{d_2}{d_1 + d_2 + \sigma_{\epsilon}^2}$$

where

$$d_1 = (p_{11})_t + (p_{13})_t + \sigma_{\xi}^2$$

$$d_2 = (p_{13})_t + (p_{33})_t + \sigma_k^2(\hat{\mathcal{X}}_t) \quad (k = 1, 2, 3).$$

Here $(p_{ij})_t$ is the corresponding element of the one-step-forward deterministically developed analysis variance from time moment t, i e. $TP_tL'_t$. Notice that $F_{t+1} = d_1 + d_2 + \sigma_{\epsilon}^2$. For the large values of the cyclic component residuals d_2 will dominate F_{t+1} and will thus force $(K*)_{t+1}^{cyclic}$ to come close to 1.

For comparison, we may note that in the case of the Gaussian linear model, the dynamics of the conditional variance of the model state does not depend on the amplitude of residuals.

$$\tilde{P} = T\tilde{P}'L_t + (\sigma_{\xi}^2, 0, (1-\rho^2)\sigma_{\psi}^2, (1-\rho^2)\sigma_{\psi}^2)'.$$

After the steady state solution is attained, the Kalman gain converges to a constant value, where $d_1 = \tilde{p}_{11} + \tilde{p}_{13} + \sigma_{\xi}^2$ and $d_2 = \tilde{p}_{13} + \tilde{p}_{33} + (1 - \rho^2)\sigma_{\psi}^2$. Here (\tilde{p}_{ij}) is a corresponding element of the steady state matrix \tilde{P} .

The total impact of the new observation y_{t+1} , attributed to the model state is dependent on the assumptions about the observation error

$$(K*)_{t+1}^{trend} + (K*)_{t+1}^{cyclic} = 1 - \frac{\sigma_{\epsilon}^2}{d_1 + d_2 + \sigma_{\epsilon}^2}.$$

For error-free observations (as the time series of $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ observations could be considered to be) we will have

$$(K*)_{t+1}^{trend} + (K*)_{t+1}^{cyclic} = 1.$$

One example of the Kalman gain $K_t = TK_t^*$ under the time-dependent Gaussian approximating model is shown in Figure 20. The Kalman gain is derived for the $\log(Ca^{2+})$ time series from the NGRIP ice core. The estimates of the unknown parameters in the model specification are obtained by maximising the corresponding log-likelihood and are given in Table 6. Performing the global smoothing, the Kalman gain of the trend component $K_t^{trend} = 0.291$ and for the cyclic component $K_t^{cyclic} = 0.684$ under the steady state solution. One can notice a clear dependency of the Kalman gain on the magnitude of the innovations. The large magnitude innovations will have a much smaller impact on the estimate of the $\hat{\mu}(t)$ and much a larger impact on the $\hat{\psi}(t)$ than they would have under the time-invariant linear Gaussian model. How small the innovations are and thus how large influence they will have on the estimate of the long and the short timescale variation components will be determined from the surrounding observations in time and depend on the particular choice of the local variance function $\sigma_k^2(\hat{\chi}_t)$.



FIGURE 21. The inverse innovation variance as a function of the innovations v(t) of the $\log(Ca^{2+})$ time series from the NGRIP ice core under the time-invariant linear Gaussian model (".") and under three different non-Gaussian distributional assumptions on the \hat{X}_t, \hat{X}_t^* : the t-distribution ("+"), the mixture of two normal densities ("*") and the Cauchy distribution ("o".)

Under the time-dependent approximate Gaussian model the components of the residuals $\hat{\xi}_t, \hat{\mathcal{X}}_t, \hat{\mathcal{X}}_t^*$ are certain additive combinations of the future innovations $\{v(t+1), \ldots, v(n)\}$.

(24)

$$\hat{\xi}_{t} = \frac{\sigma_{\xi}^{2}}{F_{t+1}}v(t+1) + \sum_{j=t+1}^{n-1}\sigma_{\xi}^{2}[a_{t+1,j}]v(j+1)$$

$$\hat{\chi}_{t} = \frac{\sigma_{k}^{2}(\hat{\chi}_{t})}{F(t+1)}v(t+1) + \sum_{j=t+1}^{n-1}\sigma_{k}^{2}(\hat{\chi}_{t})[a_{t+1,j}]v(j+1)$$

$$\hat{\chi}_{t}^{*} = \sum_{j=t+1}^{n-1}\sigma_{k}^{2}(\hat{\chi}_{t}^{*})[a_{t+1,j}]v(j+1)$$

where $[a_{t+1,j}]_l$ is the *l*-th element of the vector $a_{t+1,j} = (\prod_{i=t+1}^j L'_i)Z'(F(j+1))^{-1}$. The relationship between the components of the residuals and the innovations is non-linear because the coefficients in front of the innovations are functions of the residuals themselves. As it was already mentioned, the local variance function $\sigma_k^2(\hat{\chi}_t)$ is the dominating part in the expression of the innovation variance F(t+1). As a result, the smoothed cyclic component residual $\hat{\chi}_t$ is in a very good agreement with the innovation v(t+1) (the coefficient in front of v(t+1)) in the expression of $\hat{\chi}_t$

TABLE 9. The Box-Ljung statistics Q(15) based on the first 15 autocorrelations of $\{\tilde{\xi}_t, 0 < t < n\}$ and $\{\tilde{\mathcal{X}}_t, 0 < t < n\}$ residuals corresponding to the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series from the NGRIP ice core under the time-invariant linear Gaussian model (\mathcal{N}) and under the time-dependent Gaussian models derived from the heavy-tailed distributions \mathcal{H}_1 , \mathcal{H}_2 and \mathcal{H}_3 , assumed for the cyclic model error component.

| substance | disturbance | distribution | Q(15) | $F_{Q(15)}$ | $c(j_{max})$ |
|-------------------------|----------------------|-----------------|---------|-------------|--------------|
| | | \mathcal{N} | 5104.50 | - | 0.997 |
| δ ¹⁸ Ω | $\widetilde{\xi}$ | \mathcal{H}_1 | 5129.00 | - | 0.994 |
| 0 0 | | \mathcal{H}_2 | 5267.11 | - | 0.995 |
| | | \mathcal{H}_3 | 5587.10 | - | 0.997 |
| | | \mathcal{N} | 14.57 | 0.528 | -0.079 |
| | $	ilde{\mathcal{X}}$ | \mathcal{H}_1 | 18.85 | 0.778 | -0.089 |
| | | \mathcal{H}_2 | 16.24 | 0.645 | -0.082 |
| | | \mathcal{H}_3 | 17.36 | 0.696 | -0.089 |
| | | \mathcal{N} | 4862.30 | - | 0.995 |
| $\log(Ca^{2+})$ | $\widetilde{\xi}$ | \mathcal{H}_1 | 4613.20 | - | 0.989 |
| log(Cu) | | \mathcal{H}_2 | 5179.60 | - | 0.992 |
| | | \mathcal{H}_3 | 5530.50 | - | 0.995 |
| | | \mathcal{N} | 22.34 | 0.891 | -0.135 |
| | $	ilde{\mathcal{X}}$ | \mathcal{H}_1 | 29.43 | 0.987 | -0.165 |
| | | \mathcal{H}_2 | 29.83 | 0.988 | -0.159 |
| | | \mathcal{H}_3 | 30.40 | 0.989 | -0.166 |
| | ~ | \mathcal{N} | 5128.10 | - | 0.997 |
| $\log(SO_{\star}^{2-})$ | ξ | \mathcal{H}_1 | 4996.80 | - | 0.995 |
| 108(004) | | \mathcal{H}_2 | 5455.90 | - | 0.997 |
| | | \mathcal{H}_3 | 5650.10 | - | 0.995 |
| | ~ | \mathcal{N} | 16.44 | 0.564 | -0.106 |
| | X | \mathcal{H}_1 | 22.33 | 0.891 | -0.123 |
| | | \mathcal{H}_2 | 19.25 | 0.800 | -0.118 |
| | | \mathcal{H}_3 | 22.65 | 0.900 | -0.128 |

is close to 1). At the same time the influence of the innovations $\{v(t+2), \ldots, v(n)\}$ on $\hat{\mathcal{X}}_t$ is negligible in comparison with the influence from v(t+1) due to the nearly noise-free observations. This is why the dependency of the covariance function P_{t+1} on the cyclic model component residual $\hat{\mathcal{X}}_t$ makes the Kalman gain K_{t+1} and the innovation variance F(t+1) to be dependent on the innovation v(t+1).

One example of the inverse innovation variance $(F(t))^{-1}$ as a function of the innovations v(t) is presented in Figure 21. Under the time-dependent Gaussian approximating model, the innovation variance increases with the magnitude of innovations and is not constant anymore, as in the case of the time-invariant linear Gaussian model. As it can be seen from the diagnostics plots (Figure 12 and Figure 13), the dependency of the innovation variance on the innovation magnitude improves the fit of the statistical model.

The cyclic component residuals $\{\hat{\mathcal{X}}_t, 0 < t < n\}$, all of them being functions of the future innovations, are nearly serially uncorrelated (see Table 9 and compare with Figure 4) due to the particular property of the cyclic component residual $\hat{\mathcal{X}}_t \approx v(t+1)$. Therefore, the structure of the short timescale variation is captured

TABLE 10. The averaged zero-lag cross-correlation between the trend and the cyclic component residuals $\rho_0(\hat{\xi}_t, \hat{\mathcal{X}}_t)$ under the different distributional assumptions on the cyclic model error. The results correspond to the data from the *NGRIP* ice core. The cross-correlations with maximal magnitude $\rho_{\tau*}^{max}$ are given as well.

| distribution | | $\delta^{18}O$ | $\log(C)$ | $Ca^{2+})$ | $\log(SO_4^{2-})$ | | |
|-----------------|---------|---|-----------|---------------------|-------------------|---------------------|--|
| distribution | $ ho_0$ | $\rho_{\tau*}^{max}: max \rho_{\tau} $ | $ ho_0$ | $\rho_{	au*}^{max}$ | $ ho_0$ | $\rho_{	au*}^{max}$ | |
| \mathcal{N} | -0.089 | -0.150 | -0.071 | -0.149 | -0.110 | -0.167 | |
| \mathcal{H}_1 | 0.009 | 0.073 | 0.013 | -0.090 | -0.018 | -0.084 | |
| \mathcal{H}_2 | -0.012 | 0.077 | -0.014 | -0.082 | -0.038 | -0.081 | |
| \mathcal{H}_3 | -0.013 | 0.079 | -0.024 | -0.071 | 0.032 | 0.096 | |

through the dynamical propagator (3) and is summarised in the parameters ρ and λ_c estimated from data.

On the other hand, the estimate of the trend component residual ξ is just a sum of a large number of small impacts from future innovations without any dominant member in the sum. This is why the Gaussian assumption on the distribution of $\hat{\xi}_t$ is reasonable. The same dynamical model is used for describing the time development of the large timescale variation (equation (2)) for all of the time series in this paper, which essentially expresses a requirement on the smoothness of the estimate. Therefore the trend component residuals themselves $\{\hat{\xi}_t, 0 < t < n\}$ contain the particular structure of the data in the form of their autocorrelations.

Table 9 contains the summary of the Box-Ljung statistics Q(15) based on the first 15 auto-correlations of the standardised trend component and cyclic component residuals, corresponding to the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series from the NGRIP ice core under the time-invariant linear Gaussian model (\mathcal{N}) and under the time-dependent Gaussian model derived from the heavy-tailed assumptions on the distribution for the cyclic model error. Besides the observed values of Q(15) statistics, included are the corresponding P-values $F_{Q(15)}$ (were appropriate) and the maximal absolute value of the autocorrelations $c(j)_{max} = max_{1 \leq j \leq 15} |c(j)|$. Some problems can be noticed in the reconstruction of the structure of the $\log(Ca^{2+})$ time series using the time-dependent Gaussian approximating model. A very weak serial autocorrelation exists among the cyclic component residuals { \hat{X}_t , 0 < t < n} on the expense of a minor reduction of the autocorrelation of the trend component residuals { $\hat{\xi}_t$, 0 < t < n}. Notice, the series of Ca^{2+} records were not used explicitly in the synchronisation procedure.

The time-dependent Gaussian approximating model provides a nearly orthogonal decomposition of the total variation on the long and on the short timescale variations in a similar way as hierarchical ANOVA type models do. The timeaverages of the zero-lag cross-correlations $\rho(\hat{\xi}(t), \hat{X}_t)$ are presented in Table 10 for all combinations of time series ($\delta^{18}O$, log Ca^{2+} and log SO_4^{2-}) from the NGRIP ice core and of distributions: Gaussian and heavy-tailed \mathcal{H}_1 , \mathcal{H}_2 , \mathcal{H}_3 . They all are small, and the decomposition is particulary successful in this respect under the heavy-tailed distributional assumptions.

The reason for the orthogonality is a sharper scale separation under the timedependent Gaussian approximating models for which the $\hat{\mathcal{X}}_t$ coincide to a large extent with the innovations v(t+1) (equation (24)) at the same time as the $\hat{\xi}_t$ contains impact from a large number of the future innovations $\{v(s), t < s < n\}$. Notice that the innovations $\{v(t), 0 < t < n\}$ are nearly serially uncorrelated also under the time-dependent Gaussian approximating model, (see Table 8). It is important to notice that such a perfect separation of scales, with nearly orthogonal estimates for the long and the short timescale variations, is possible mainly for time series nearly free from observational noise ($\sigma_{\epsilon}^2 \approx 0$), such as the $\delta^{18}O$, log(Ca^{2+}) and log(SO_4^{2-}) time series. In case stronger observational noise, comparable with other sources of the variation, would be present, the averaging over several future innovations would be necessary for the estimation of the cyclic model component residuals in order to reduce impact of the noise. It should be mentioned that a linear non-Gaussian model with a heavy-tailed assumption on the distribution for the model error component should be used with great care in the treatment of noisy observations, such as, for example, the MS^- series from the NGRIP ice core. Weaker restrictions on the variations of the model component residual may allow the observational noise to influence the estimate of the model components and to make the predictions loss reliability.

5. Discussion on statistical modelling and Notes on the Interpretation of time series model.

5.1. Discussion on the statistical modelling. A large collection of computationally efficient techniques for treatment of the time series models has been developed for linear Gaussian models. However geophysical time series often do not fit linear Gaussian models. One may transform a non-Gaussian stochastic variable to become more Gaussian-like and in this way one preserves the possibility to apply the efficient computational algorithms. A scaling of a non-Gaussian variable by a variance dependent on the magnitude of the variable is one approach that has been applied in meteorology. The development of data assimilation schemes for humidity is one example of research in this direction (Elias Holm, ECMWF, personal communication).

A similar method has been applied in this paper treating some geological series of measurements from the ice cores drilled at two different locations (NGRIP and GISP2) in Greenland: $\delta^{18}O$ (commonly used as a temperature proxy), $\log(Ca^{2+})$ (Ca^{2+} is commonly used as a terrestrial dust proxy) and $\log(SO_4^{2-})$ (the sulfate depositions). A linear Gaussian model with a time-dependent evolution of the covariance of the model state conditional on observations is derived in such a way that the variance of innovations (a one-step-forward prediction error) becomes dependent on the innovation size. A heavy-tailed distribution is used as a framework to create the dependency.

After a diagnostic check and an investigation of the variation decomposition had been carried out, it could be concluded that the *t*-distribution was the only one from the four initially tried heavy-tailed distributions which can be used as a reference for deriving the time-dependent Gaussian approximating model to capture the variability of the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series from each ice core. The mixture of two normal densities fails in the variation decomposition, allowing the long timescale variation to influence the estimate of the short timescale variation. The Cauchy distribution fails already in the fitting of the statistical model. It scales the innovations too severely and makes initially heavy-tailed, non-Gaussian, innovations become light-tailed, non-Gaussian, ones. The general error distribution provides a variation decomposition which does not differ much from the one obtained under the Gaussian assumptions.

Figure 22 shows the estimated long timescale variation $\hat{\mu}(t)$ together with the systematic part $Z\hat{\alpha}(t) = \hat{\mu}(t) + \hat{\psi}(t)$ for the $\delta^{18}O$, the $\log(Ca^{2+})$ and the $\log(SO_4^{2-})$ series from the NGRIP and GISP2 ice cores when the Student- t distribution was used to derive the time-dependent Gaussian model. One should take into account that the estimated variation decomposition is biased in the beginning and at the



FIGURE 22. The estimated long $(\hat{\mu}(t))$ and short $(\hat{\psi}(t))$ timescale variation of the $\delta^{18}O$ (the temperature proxy, upper plot), log (Ca^2) (the terrestrial dust proxy, middle plot) and log (SO_4^{2-})) (the nonsea salt sulfate, lower plot) time series from the NGRIP (solid thick curves) and from the GISP2 (dashed thick curve fot the trend and solid thin curve for the short timescale variation) ice core under the time-dependent Gaussian model, derived from the Student-t distributional assumptions (\mathcal{H}_1 with $\nu = 4$) on $\mathcal{X}_t, \mathcal{X}_t^*$.

end of the time period. Both future and past innovations, lagging up to ≈ 30 time steps (1 time step = 200 years), influence the estimation of the trend.

One of the main contributions of this paper is that we show that the linear Gaussian model with a time-varying variance can be used for treatment of essentially non-linear time series. All six time series represent systems which has two different regimes of behaviour: the relatively smooth background signal, originating from continuous processes on Earth, and the abrupt deviations from the background signal, which can originate from some momentary external or internal forcing, such as volcanoe eruptions. The advantage of analysing the time series via a Gaussian linear model is that the Kalman filter and the Kalman smoother recursions can be used as efficient computational tools to perform the variation decomposition and to estimate parameters.

Even though geological series are modelled in our paper, the investigations are carried out from the perspective of data assimilation and they touch essential subjects of data assimilation. From the model perspective the splitting of observations into these two different regimes corresponds to the splitting of observations into those that the model is able to predict (the background signal) and those that the model is not able to predict (the "alternative" regime).



FIGURE 23. Upper part: The estimated long timescale variation $\hat{\mu}(t)$ for $\delta^{18}O$ (solid curve), $\log(Ca^{2+})$ (dashed curve) and $\log(SO_4^{2-})$ (dash-dotted curve) from NGRIP (to the left) and GISP2 (to the right). Lower part: The time-averaged estimated cross-correlations $\hat{\psi}(t)$ for $\hat{\rho}_{\tau}(\delta^{18}O), \log(Ca^{2+}))$ (solid curve), $\hat{\rho}_{\tau}(\delta^{18}O, \log(SO_4^{2-}))$ (dashed curve) and $\hat{\rho}_{\tau}(\log(Ca^{2+}), \log(SO_4^{2-}))$ (dash-dotted curve) from NGRIP(to the left) and GISP2 (to the right). The short timescale variation is estimated under the \mathcal{H}_1 distributional assumptions on $\mathcal{X}_t, \mathcal{X}_t^*$.

The background signal is modelled through a time-invariant dynamical propagator. A non-stationary random walk process with second-order requirements on smoothness is used to model the trend, and a stationary cyclic (wave) process is used for modelling the short timescale variation of the background signal. The abrupt deviations from the background signal are modelled via a stochastic forcing of the dynamical model for the short timescale variation. The stochastic forcing is Gaussian with zero mean and a time-dependent variance. To describe the dependency of the variance on time we introduce a concept of a local variance function. The local variance function is iteratively estimated from the data and is derived from the density of the t-distribution. The on-line estimation of the conditional variance of the model state, given observations, to improve efficiency of the data assimilation procedure is a well-known approach in the meteorological community (Dee, 1995).

Another important contribution of this paper is that we discuss the properties and the mechanism of the local smoothing, which lies behind the total variation

TABLE 11. The parameter estimates, the transformed parameter (the log-transform for σ_{ξ} , σ_{ψ} and λ_c and logit transform for ρ) estimates and the standard errors (stderr) of the transformed parameters estimates under the time-dependent Gaussian model derived from the \mathcal{H}_1 distributional assumptions on the $\mathcal{X}_t, \mathcal{X}_t^*$ for the $\delta^{18}O$, the $\log(Ca^{2+})$ and the $\log(SO_4^{2-})$ series from each ice core. The estimated degree of freedom $\nu = 4$ for each 6 series of observations. In the horizontal subsections of this table upper rows correspond to the NGRIP ice core and lower rows correspond to the GISP2 ice core.

| substance | $\hat{\sigma}_{m{\xi}}$ | | $\hat{\sigma}_\psi$ | | ρ | | $\hat{\lambda}_c$ | |
|-----------------|-------------------------|-------------|---------------------|-------------|------|-----------------|-------------------|-------------|
| $\delta^{18}O$ | 0.009 | -4.69(0.44) | 1.35 | 0.30(0.49) | 0.88 | 1.96(0.20) | 0.13 | -2.10(0.27) |
| | 0.010 | -4.59(0.38) | 1.20 | 0.18(0.71) | 0.81 | 1.45(0.17) | 0.08 | -2.60(0.44) |
| $\log(Ca^{2+})$ | 0.006 | -5.11(0.29) | 0.69 | -0.37(0.74) | 0.90 | 2.18(0.17) | 0.12 | -2.16(0.33) |
| $\log(Ca^{-+})$ | 0.006 | -5.19(0.36) | 0.78 | -0.24(0.69) | 0.90 | 2.16(0.19) | 0.10 | -2.26(0.39) |
| $\log(SO^{2-})$ | 0.003 | -5.72(0.41) | 0.53 | -0.64(1.02) | 0.85 | 1.74(0.25) | 0.13 | -2.01(0.51) |
| $\log(SO_4)$ | 0.003 | -5.93(0.81) | 0.45 | -0.80(0.48) | 0.82 | $1.52 \ (0.27)$ | 0.12 | -2.14(1.35) |

decomposition and which originates from the time-dependent Gaussian approximation of the non-Gaussian state space model. Under the approximating model, the time fluctuations of the local variance function dominate the dynamics of the conditional variance of the cyclic model component. The local variance function works as a local, time-dependent, measure of the uncertainty about the dynamical development of the model and it controls the influence of newly coming observations on the estimates of the different model state components. These estimates are given an impact from coming observations relative to the local (in time) estimate of the uncertainty about their dynamical development. The larger the relative uncertainty of the particular component about its dynamical development, the stronger will the impact be of the new coming observation just on this component. Discussing the mechanism of the local smoothing we stress the core properties of data assimilation. The data assimilation by linear Gaussian models always contributes to reduction of the uncertainty and the data assimilation is based on the relative (but not absolute) measure of the uncertainty.

The local variance function which we use in our paper as the way to assign the time-dependent (as opposite to the time-invariant) conditional variance of the model state, given observations, has a number of advantages. The local variance function is derived in such a way that the nearly deterministic dependency between innovations and the cyclic component residuals is established. This dependency allows us to transform the inability of model to predict observations, expressed in the space of observations through innovations, into the uncertainty about the dynamical development of the short and long timescale variations, expressed in the space of model states through the cyclic component residual. We should stress that the exceptional property of the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ series of being almost free from observational noise makes the relationship between the innovations and the cyclic component residuals very close to a linear deterministic one and makes the method so illustrative.

As we have shown the estimated long $\hat{\mu}(t)$ and short $\hat{\psi}(t)$ timescale variations are nearly mutually uncorrelated. The auto-correlated trend component residuals contain the specific structure of the long timescale variation. The cyclic component residuals, i.e. the ones corresponding to the short timescale variation, are nearly mutually uncorrelated. The specific structure of the short timescale variations is absorbed into the dynamical propagator and is statistically summarized in the estimate of the wave frequency λ_c and the "damping factor" ρ .

More extensional study should be performed in order to understand if the total variation of non-linear time series can always be decomposed into nearly mutually uncorrelated the long and the short timescale variation components in presence of observational noise. The deficiency of our study is certainly the over-simplistic dynamical and statistical models, which are chosen to describe variability of these geological time series. The multivariate analysis has not been carried out either. But again we would like to stress that the core of the study is not to model the geological series, but to outline and investigate certain problems of data assimilation working with non-Gaussian systems, the influence of the model error specification on the estimate of the model state being one of them.

Reconstruction of the data generating mechanism is the unconditional inference. Under it, the observed time series of geological measurements is interpreted as one realisation of a stochastic process. The selection of the non-Gaussian model (12) to describe the variation of the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series is largely based on convenience. The decomposition of the total variation of a time series into the long and the short timescale variations is a conditional inference. Of cource, the statistical conditional inference is also based on subjective assumptions about the data generating mechanism.

The time-dependent Gaussian approximating model (19) is valid under the conditional environment only. The local variance function $\sigma_k^2(\hat{\mathcal{X}}_t)$ is a function of observations themselves. Furthermore, the time-dependent Gaussian model (19) is just an approximation of the original model (12). But working in the conditional environment, these two essentially different statistical models provide the same solution for decomposing the total variation into the different timescales.

5.2. Notes on the Interpretation of the time series models. The statistical comparison of the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ series is summarised in Figure 22, Figure 23 and Table 11. Figure 22 compares the same chemical measurements from the *NGRIP* and *GISP*2 ice cores. Figure 23 compares the different chemical measurements within each ice core. For the basic parameters, being the variances of the model error components $\hat{\sigma}_{\xi}$ and $\hat{\sigma}_{\psi}$, the "damping" factor $\hat{\rho}$ and the spectral frequency of the cyclic process λ_c , Table 11 shows the estimate of the original and the transformed parameters (the log-transform for σ_{ξ} , σ_{ψ} and λ_c and the logit transform for ρ) together with the standard errors of the transformed parameters in parentheses.

The standard errors of the estimates of the parameters are quite large. The too simplistic dynamical and statistical models can provoke the loss of precision in the estimation of the parameters. Still more important is the fact that the estimated variance of the (transformed) parameters estimates is just more or less crude approximation to the true uncertainty. In fact the estimated variance is a diagonal of the inverse of the Hessian (a "curvature") of the corresponding likelihood at the point of the estimates of parameters (*argmax* of the likelihood). Therefore, the estimated values of the parameters should be taken with care.

Each of four transformed parameters $\log \sigma_{\xi}, \log \sigma_{\psi}, logit\rho$ and $\log \lambda_c$ could be classified according to two factors, namely ice core (*NGRIP* or *GISP2*) and measured substance ($\delta^{18}O$, $\log Ca^{2+}$ and $\log SO_4^{2-}$). An ANOVA-type analysis (the analysis of variance) of this two-factor design is performed as a crude statistical test for significant differences between the estimates of the parameters fitted to these six different time series. Performing this ANOVA-type analysis the standard errors are used to get crude variance estimates. The influence of these two factors on all four responses (the estimates of the corresponding transformed parameters) is assumed to be additive. For each parameter the mean square of interaction between these two factors is much smaller than the standard error squared. Very small observed values of the mean squared interaction in comparison to the standard error squared can indicate that the standard error are unrealistically large. However, because of small degree of freedom with six realisations of each response only, we still prefer to stay with standard error as a crude precision estimate.

Based on the current precision of the transformed parameters estimates, the qualitative conclusions are as follows. There is no indication found in data that these parameters are different between the NGRIP and GISP ice cores. Most of transformed parameters ($\log \sigma_{\xi}$, $\log \sigma_{\psi}$ and $logit\rho$, but not $\log \lambda$) show a statistically significant (or very close to it) difference between substances. Finally, considering logratio between variances, namely $\log \sigma_{\psi} - \log \sigma_{\xi}$, there is no substantial variation in it at all, and thus there is no indication of neither ice core nor substance effect on the logratio.

The short timescale variations are remarkably similar for all 6 time series. The measurements of $\delta^{18}O$ are lower for the NGRIP ice core than for the GISP2 ice core. This is consistent with the fact that the NGRIP ice core was drilled north of the GISP2 location. The essential differences in the behaviour of the time series are absorbed into the estimated long timescale variation $\hat{\mu}(t)$. The long timescale variation of the $\delta^{18}O$ from NGRIP seems to lag (statistically insignificantly) the one from GISP2. The difference in lag becomes smaller and the difference in absolute value increases going back in time. For making any conclusion about systematic differences of the climate at these two locations where these two ice cores were drilled out, many factors must be taken into account, such as, for example, the terrain differences at these two projects and the procedure of synchronisation.

The measurements of $\log(SO_4^{2-})$ are in a strong inverse correspondence with the measurements of $\delta^{18}O$ for both ice cores. Smaller amount of precipitation during periods with lower temperature in high latitudes and stronger winds due to stronger spatial temperature gradients can contribute to increase, with decreased temperature, the concentration of the impurities deposited in the snow and ice, and thus making the inverse relationship between $\log(SO_4^{2-})$ and $\delta^{18}0$ clear. The short timescale variations of the $\log(Ca^{2+})$ and the $\log(SO_4^{2-})$ series are remarkably similar to each other in structure and strongly cross-correlated between substances for each ice cores and between ice cores for each substance. The time-averaged zero-lag cross-correlation between the short timescale variations $\rho_0(\log(Ca^{2+}), \log(SO_4^{2-})) = 0.96$ on the *NGRIP* ice core and $\rho_0(\log(Ca^{2+}), \log(SO_4^{2-})) = 0.96$ 0.94 on the GISP2 ice core. At the same time, the difference in the long timescale variation between the $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ decreases with increasing depth on the NGRIP ice core and it increases with increasing depth the GISP2 ice core. Because the measurements of the Ca^{2+} are considered as a terrestrial dust proxy, the dissimilarity in the estimated trend behaviour may indicate that the contribution from the marine biological sources (through the oxidation of dimethylsulfide) to the deposition of the SO_4^{2-} were different at these two locations before 50-60 ky B.P.

We have not carried out any further statistical analysis of the MS^- time series from the NGRIP and the GISP2 ice cores, for example, neither a comparison between the MS^- time series from these two ice cores nor between the MS^- and the $\delta^{18}O$, $\log(Ca^{2+})$ and $\log(SO_4^{2-})$ time series within each ice core. The fact that the linear time-invariant Gaussian model can capture the main variability of the MS^- time series from the GISP2 ice core looks reasonable and in agreement with the data processing. The averaging of the high-resolution data over 200 years periods, provided that strong deviations from the background signal are absent, can make the time series closer to a Gaussian one. The MS^- measurements represent solely the result of marine biological activity. It would be unreasonable to expect a clear linear dependence between the MS^- and $\delta^{18}O$ measurements and thus no strong cross-correlations between the MS^- and the $\delta^{18}O$ time series should be expected. Certainly, a non-linear dependence, which does not result in a high cross-correlation, exists between the series.

The behaviour of the MS^- time series from the NGRIP ice core is completely different from the behaviour of the other time series analysed in the paper. The MS^- time series from NGRIP contains a large observational noise. Deeper investigations of, in the first hand, the data processing mechanism, the spatial variability of the impurities deposited in the snow and ice and the post-depositional processes going on in the ice/snow, should be performed in order to understand the reason for the difference in behaviour between the MS^- time series from the NGRIP and the GISP2 ice cores.

5.3. **Conclusions.** The research carried out in this paper is a case study of geological time series. At the same time conclusions of this study are quite general and are extendable to other areas of geophysics.

• Non-linear time series can be approximated by linear Gaussian models with (partially) locally estimated parameters.

• The Kalman filter and Kalman smoother recursions can be used as en efficient computational algorithm to estimate parameters fitted to non-linear and non-Gaussian time series.

• A relative measure of uncertainty is a core of data assimilation by linear Gaussian models. The data assimilation by linear Gaussian models always contributes to reduction of uncertainty with the strongest impact on largest uncertainty component.

• The density of heavy-tailed distributions can be used for an on-line estimation of innovation variance. Simple diagnostic check helps to select the heavy-tailed distribution that fits time series.

6. Acknowledgements

The author would like to thank Rolf Sundberg for good discussion on statistical methods and for creative comments on the manuscript. The author is grateful to Ulf Jonsell for the provided access to the data. Finally, many thanks are to Nils Gustafsson for passing through the jungle of my writing.

7. References

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Trans.* Automat. Control., **19**, 716-723.

Anderson, B.D.O. and J.B. Moore (1979). *Optimal Filtering*. Englewood Cliffs: Prentice-Hall.

Box, G.E.P. and G.C. Tiao (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

Castellano, E., S. Becagli, J. Jouzel, A. Migliori, M. Severi, and co-authors (2004), Volcanic eruption frequency over the last 45 ky as recorded in Epica-Dome C ice core (East Antarctica) and its relationship with climatic changes. *Global and Planetary Change* **42**, 195-205.

Charlson, R.J., S.E. Schwartz, J.M. Hales, R.D. Cess, J.A. Coakley, and co-authors (1992). Climate forcing by anthropogenic aerosols. *Science* **255**, 423-430.

Chiles, J.-P. and P. Delfiner (1999). *Geostatistics: Modelling Spatial Uncertainty*. John Wiley & Sons.

de Jong (1989). Smoothing and interpolation with the state space models, J. American Statistical Association, 84, 1085-1088.

Dee, D.P. (1995). On-line estimation of the error covariance parameters for the atmospheric Data Assimilation. *Monthly Weather Review*, 4, **123**, 1128-1145.

Dee, D.P. (2004). Detection and correction of the model bias during data assimilation. *Seminar on Recent developments in data assimilation for atmosphere and ocean.* ECMWF Seminar Proceedings.

Delmas, R. and Boutron, C. (1976), Sulfate content in East Antarctica firn samples. *EOS, Transactions, Americal Geophysical Union*, **57**(9), 677.

Doornik, J.A., (1998). *Object-Oriented Matrix Programming using Ox 2.0*. London:Timberlake Consultants Press.

Durbin, J., and S.J. Koopman, (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model, *Biometrika*, 84, 669-684.

Durbin, J., and S.J. Koopman, (2001). *Time series Analysis by State Space Methods*. Oxford: Oxford University Press.

Fahrmeir, L., and L. Knorr-Held, (2000). Dynamical and Semiparametrical Models. In *Smoothing and Regression: approaches, Computation and Application*, edited by M.G. Schimek, John Wiley& Sons.

Fahrmeir, L. and H. Kaufmann, (1991). On Kalman filtering, posterior mode estimation and Fischer scoring in dynamical exponential family regressions. *Metrica*, **38**, 37-60.

Fahrmeir, L. and G.Tutz, (1994). Multivariate Statistical Modelling Based on Generalized Linear Models. New York : Springer-Verlag.

Fahrmeir, L. and S. Wagenpfeil, (1997). Penalised liklihood estimation and iterative Kalman smoothing for non-Gaussian regression models, *Comput. Statistical Data Analysis*, **24**, 295-320.

Feller, W. (1966). An Introduction to Probability Theory and its Application, Vol. II. New York : John Wiley & Sons.

Fischer, H., D. Wagenbach, and J. Kipfstuhl. (1998) Sulfate and nitrate firn concentrations on the Greenland ice sheet 2. Temporal anthropogenic deposition changes. J. Geophysical Res., 103(D17), 21935-21942.

Fletcher, R. (1987). *Practical Methods of Optimisation*. 2nd edition. New York: John Wiley.

Fuhrer, K. E.W. Wolff and S.J. Johnsen (1999), Timescales for dust variability in The Greenland Ice Core Project (GRIP) ice core in the last 100,000 years, *J. Geophys. Res.*, **104** (D24), 31043-31052.

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

Hammer, C.U., (1977), Past volcanism revealed by Greenland Ice Sheet impurities. *Nature*, **288**,230-235.

Harvey, A.C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge: Cambridge University Press.

Harvey, A.C., and S.J. Koopman (1992). Diagnostic checking of unobserved components time series models, *J. Bussiness and Economic statistics*, **10**, 377-389.

Harvey, A.C., and M. Streibel (1998). Testing for nonstationary unobserved components. *Time Series Analysis*, **20**, 1-19.

Jonsell, U., M.E. Hansson, M.-L. Siggaard-Andersen, J.P. Steffensen. Comparison of Northern and Central Greenland Ice core records of methanesulfonate covering the last glacial period. Submitted to *J. Geophys. Res.* (in review).

Johnsen, S.J., D. Dahl-Jensen, N. Gundestrup, J.P. Steffensen, H.B. Clausen, H. Miller, V. Masson-Delmotte, A.E. Sveinbjrnsdottir, and. J.White (2001). Oxygen isotope and paleotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NGRIP. J. Quat. Sci., 16, 299-30, 299-307..

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. J. Basic Engineering, Transactions ASME, series D, 82, 35-45.

Kitagawa, G. (1987). Non-Gaussian state space modelling of non-stationary time series (with discussion). J. American Statist. Assoc., 82, 1032-1063.

Kohn, R. and Ansley, C.F. (1989). A fast algorithm for signals extraction, influence and cross-validation, *Biometrika*, **76**, 65-79.

Koopman, S.J. (1993). Disturbance smoother for state space models, *Biometrika*, **80**, 117-126.

Koopman, S.J., A.C. Harvey, J.A. Doornik, and N. Shephard (2000). *Stamp 6.0: Structural Time Series Analyser, Modeller and Predictor.* London: Timberlake Consultants.

Koopman, S.J., N. Shephard, and J.A. Doornik (1998). Fitting non-Gaussian state space models in econometrics: Overview, development and software, mimeo, Nuffield College, Oxford.

Koopman, S.J., N. Shephard, and J.A. Doornik (1999). Statistical algorithms in state space form using SsfPack2.2. *Econometrics J.*, **2**, 113-166. http://www.ssfpack.com/.

Lindsey, J.K. (1996). *Parametric statistical inference*. Oxford: Oxford University Press.

Ljung, G.M., and G.E.P. Box, (1978). On a measure of lack of fit in time series models. *Biometrika*, **66**, 67-72.

Masreliez, C.J. (1975). Approximate non-Gaussian filtering with linear state and observation relaions. *IEEE Trans. Automatic. Control*, AC-20, 1044-1050.

Masreliez, C.J., and R.D. Martin (1977). Robust Bayesian estimation for the linear models and robustifying the Kalman filter. *IEEE Trans. Automatic. Control*, **AC-22**, 361-371.

Mayewski, P.A., L.D. Meeker, M.S. Twickler, S. Whitlow, Y. Qinzhao, L.W. Berry, and M. Prentice (1997), Major features and forcing of high-latitude Northern Hemisphere atmospheric circulation using 110 000-year-long glaciochemical series, *J. Geophys. Res.*, **102** (C12), 26345-26366.

North Greenland Ice Core Preject members (2004), High-resolution record of Northern Hemisphere climate extending into the last interglacial period, *Nature*, **431**, 147-151.

Parzen, E.(1962). On estimation of probability density functions and mode. Ann. Math. Statistics, **33**, 1065-1076.

Ripley, B.D., (1987). Stochastic simulation. New York: Wiley.

Roe, G.H., and E.J. Steig, (2004), Characterization of millennial-scale climate variability. *J. Climate*, 1929-1944.

Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function, Ann. Math. Statist., 27, 832-837.

Röthlisberger, R., M. Bigler, M. Hutterli, S. Sommer, B. Stauffer, H.G. Junghans, and D. Wagenbach (2000a), Technique for continuous high-resolution analysis of trace substances in firm and ice cores, *Environ. Sci. Technol.*, **34**(2), 338-342.

Sarda, P., and P. Vieu (2000), Kernel Regression. In *Smoothing and Regression* edited by M.G. Schimek. John Wiley & Sons.

Saltzman, E.S. (1995), Ocean/Atmosphere Cycling of Dimethylesulfide. in *Ice Core studies of Global Biogeochemical Cycles*, edited by R.J. Delmas , pp 65-89, Springer-Verlag.

Saltzman, E.S., P.Y. Whung and P.A. Mayewski (1997), Methansulfonate in the Greenland Ice Core Sheet Project 2 ice core, *J. Geophys. Res.*, **102** (C12), 26649-26657.

Schweppe, F. (1965). Evalution of likelihood functions for Gaussian signals, *IEEE Transactions on Information Theory*, **11**, 61-70.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series, *Biometrika*, 84, 653-667.

Sugarawa, K. (1961). Chemistry of ice, snow and other water substances in Antarctica. Association Internationale d'Hydrologie Scientifique **55**, 49-52.

Tong, H., (1990) Non-linear Time Series. A Dynamical System Approach, Oxford: Oxford University Press.

White, J. (2004), GISP2 Stable Isotopes (Deuterium, Deuterium Excess, and Oxygen). National Snow and Ice Data Center.

http:// nsidc.org/data/docs/arcss/arcss137/index.html.

Young, P.C., K. Lane, C.N. Ng, D. Palmer, (1991). Recursive forecasting, smoothing and seasonal adjustment of nonstationary environmental data. *J. of Forecasting*, **10**, 57-89.