

Inverse stochastic–dynamic models for high-resolution Greenland ice core records

Niklas Boers, Mickael D Chekroun, Honghu Liu, Dmitri Kondrashov, Denis-Didier Rousseau, Anders Svensson, Matthias Bigler, Michael Ghil

▶ **To cite this version:**

Niklas Boers, Mickael D Chekroun, Honghu Liu, Dmitri Kondrashov, Denis-Didier Rousseau, et al.. Inverse stochastic–dynamic models for high-resolution Greenland ice core records. Earth System Dynamics, 2017, 8 (4), pp.1171 - 1190. 10.5194/esd-8-1171-2017 . insu-03321824

HAL Id: insu-03321824
https://insu.hal.science/insu-03321824

Submitted on 18 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Earth Syst. Dynam., 8, 1171–1190, 2017
https://doi.org/10.5194/esd-8-1171-2017
© Author(s) 2017. This work is distributed under
the Creative Commons Attribution 3.0 License.

Earth System Dynamics
Open Access
EGU

Inverse stochastic–dynamic models for high-resolution Greenland ice core records

Niklas Boers[1], Mickael D. Chekroun[2], Honghu Liu[3], Dmitri Kondrashov[2,4], Denis-Didier Rousseau[1,5], Anders Svensson[6], Matthias Bigler[7], and Michael Ghil[1,2]

[1]Geosciences Department and Laboratoire de Météorologie Dynamique (CNRS and IPSL), École Normale Supérieure and PSL Research University, Paris, France
[2]Department of Atmospheric and Oceanic Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, USA
[3]Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, USA
[4]Institute of Applied Physics of the Russian Academy of Sciences, Nizhny Novgorod, Russia
[5]Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA
[6]Centre for Ice and Climate, University of Copenhagen, Copenhagen, Denmark
[7]Physics Institute and Oeschger Center for Climate Change Research, University of Bern, Bern, Switzerland

Correspondence: Niklas Boers (nboers@lmd.ens.fr)

Received: 23 January 2017 – Discussion started: 24 January 2017
Revised: 19 June 2017 – Accepted: 12 November 2017 – Published: 15 December 2017

Abstract. Proxy records from Greenland ice cores have been studied for several decades, yet many open questions remain regarding the climate variability encoded therein. Here, we use a Bayesian framework for inferring inverse, stochastic–dynamic models from δ^{18}O and dust records of unprecedented, subdecadal temporal resolution. The records stem from the North Greenland Ice Core Project (NGRIP), and we focus on the time interval 59–22 ka b2k. Our model reproduces the dynamical characteristics of both the δ^{18}O and dust proxy records, including the millennial-scale Dansgaard–Oeschger variability, as well as statistical properties such as probability density functions, waiting times and power spectra, with no need for any external forcing. The crucial ingredients for capturing these properties are (i) high-resolution training data, (ii) cubic drift terms, (iii) nonlinear coupling terms between the δ^{18}O and dust time series, and (iv) non-Markovian contributions that represent short-term memory effects.

1 Introduction

Data-driven stochastic difference equation models have recently been successfully applied to a wide range of climatic phenomena (Kondrashov et al., 2005, 2006; Kravtsov et al., 2005, 2009). The striking success of this empirical-model reduction (EMR) approach in reproducing dynamical and statistical properties of the underlying dynamical systems has been explained by embedding EMR models in the larger class of multilayer stochastic models (MSMs); the latter models, in turn, were shown to be solidly grounded in the Mori–Zwanzig (MZ; Zwanzig, 1964; Mori, 1965) formalism of statistical physics (Kondrashov et al., 2015). In addition to enhancing our understanding of the geophysical systems under consideration, EMR–MSM models have also been shown to be well suited for predictive purposes; e.g., they have considerable skill in predicting certain key variables associated with the El Niño–Southern Oscillation (Barnston et al., 2012; Chekroun et al., 2011) and the Madden–Julian Oscillation (Kondrashov et al., 2013).

In general terms, stochastic–dynamic models are derived by approximating the discrete-time divided differences of observed time series by a deterministic function **F** plus residual noise (e.g., Hasselmann, 1976):

$$\frac{\Delta \boldsymbol{x}_i}{\Delta t_i} \approx \mathbf{F}(\boldsymbol{x}_i) + \eta_i \ ; \tag{1}$$

here $\boldsymbol{x}_i$ is the empirical observations, $\Delta\boldsymbol{x}_i = \boldsymbol{x}_{i+1} - \boldsymbol{x}_i$, and $\Delta t_i = t_{i+1} - t_i$ denotes the time spans from one observation to the next. The specific functional form of $\mathbf{F}$ may use some a priori knowledge of the system under study, while the parameters of the proposed model are always inferred by training it on a given set of time series produced by the system. In this sense, the approach is semiempirical, rather than being entirely hypothesis-free. Most existing methodologies for empirical-model derivation are based on least-squares fitting to determine optimal parameters for $\mathbf{F}$.

In the present study, we are specifically interested in fitting low-dimensional stochastic–dynamic models to high-resolution time series of two paleoclimatic proxy records, namely the δ^{18}O ratios and dust concentrations obtained by the North Greenland Ice Core Project (NGRIP; Ruth et al., 2003; Andersen et al., 2004; Gkinis et al., 2014); see Fig. 1. Low-dimensional conceptual models have a long history in paleoclimate (e.g., Källén et al., 1979; Le Treut and Ghil, 1983; Saltzman and Maasch, 1990; Ghil, 1994; Tziperman et al., 2006; De Saedeleer et al., 2013). Such models typically incorporate a few global or regional climate variables, such as global temperature and ice-sheet volume, nonlinear interactions among these, and astronomical forcing subject to possible stochastic fluctuations (Saltzman and Maasch, 1991; Crucifix and Rougier, 2009).

In contrast, we choose here a data-driven, stochastic–dynamic approach: we intend to find a system of stochastic differential equations (SDEs) to simulate time series that reproduce the statistical and dynamical properties of the observed δ^{18}O and dust time series and do so without taking into account exogenous astronomical forcing. The main issue with the naive approach of Eq. (1) is that the noise term η, which represents the unobserved variables, will typically be correlated with the state vector $\boldsymbol{x}$. To overcome this problem, we adapt the recently developed non-Markovian data-driven closure models introduced by Kondrashov et al. (2015) as follows.

Given a multivariate, low-dimensional time series $\boldsymbol{x}(t)$ of partial observations of a much higher-dimensional system, the MZ formalism yields the abstract generalized Langevin equation (GLE):

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \mathbf{F}(\boldsymbol{x}) + \int_0^t \mathbf{G}(t, s; \boldsymbol{x}(s))\mathrm{d}s + \eta(t)\,. \tag{2}$$

In our application, $\boldsymbol{x}$ is the two-vector of δ^{18}O and log(dust), while the much larger system is the climate system. The first term $\mathbf{F}(\boldsymbol{x})$ is Markovian and it accounts for the nonlinear self-interactions among the observed variables, while the non-Markovian integral term accounts for the cross-interactions between the observed and unobserved variables; in this formulation, the latter are not present in $\mathbf{F}(\boldsymbol{x})$. This non-Markovian integral term involves the past of the observed variables and thus introduces memory effects into the closed system. The term $\eta(t)$ accounts for the stochastic forcing that is now uncorrelated with $\boldsymbol{x}$, as guaranteed by the MZ formalism. The noise term, however, is not necessarily white in time or space; see Kondrashov et al. (2015) for a detailed derivation of the GLE approach and Appendix A for a sketch of the main ideas and more technical details.

The temporal evolution of both the δ^{18}O and dust time series indicates that there exist two alternative, relatively steady states for the underlying dynamical system, namely the colder stadials and the warmer interstadials (Dansgaard et al., 1993; Rasmussen et al., 2014). Transitions from the stadials to the interstadials occur very abruptly, within several decades, during the so-called Dansgaard–Oeschger (DO) events (Johnsen et al., 1992; Dansgaard et al., 1993; Ditlevsen et al., 2005), while transitions in the opposite direction are characterized by a comparably slow relaxation process that may last centuries to millennia, depending on the specific event.

This bistability suggests using a system of two coupled SDEs with a double-well potential as a model of the processes generating the two time series of δ^{18}O ratios and dust (Ditlevsen, 1999; Ditlevsen and Ditlevsen, 2009). In addition, we will take into account here possible memory effects in the climate system (Bhattacharya et al., 1982; Ghil et al., 2015) by including explicit non-Markovian terms in the model. In particular, these memory terms are included in order to reproduce the temporal asymmetry of the observed time series: the sharp transitions from the stadials to the interstadials, followed by rather smooth transitions from the interstadials back to the stadials. For these data, we thus propose a two-dimensional stochastic delay differential equation as an approximation of the GLE (2):

$$\mathrm{d}\boldsymbol{x} = \left\{ A + \sum_{s=0}^{d} B_s \boldsymbol{x}(t - s\tau) + C(\boldsymbol{x}, \boldsymbol{x}) + D(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{x}) \right\} \mathrm{d}t + \mathbf{Q}\,\mathrm{d}W(t)\,. \tag{3}$$

Here $\boldsymbol{x}$ is the two-dimensional time series of δ^{18}O and dust, which is sampled at time steps t_i in the NGRIP ice core. The model has a cubic drift term and retarded, non-Markovian arguments in the linear terms. The matrix $\mathbf{Q}$ denotes the Cholesky decomposition of the covariance matrix of the noise, and $W(t)$ denotes a multidimensional Wiener process. Model parameters will be inferred using maximum likelihood estimation (MLE) and, for comparison, ordinary least-squares fitting. It will turn out that both approaches are in fact equivalent for the specific optimization problem proposed here; see Sect. 2.2 for further details and the explicit version of this SDE that we use in practice.

Time series simulated by our empirical model will be compared to the original time series in terms of statistical properties, such as the probability density functions (PDFs) of the time series, their power spectra and the average waiting time between sharp transitions from stadials to interstadials. Furthermore, we will test the relevance of the different model

ingredients, such as the nonlinear terms, the memory and the coupling terms, using Bayesian model selection criteria.

The general potential of Bayesian parameter inference such as MLE has recently been discussed for stochastic–dynamic climate models from incomplete data (Peavoy et al., 2015). A Bayesian framework to compare different types of models has also recently been used for the specific case of the NGRIP δ^{18}O record, including a double-well potential model, a relaxation oscillator, and two versions of a mixture of locally linear stochastic models. Based on the Bayesian information criterion, it was concluded there that the two locally linear mixture models are best supported by the observations if three local models are used in each mixture (Kwasniok, 2013). Furthermore, a comprehensive Bayesian approach was employed to infer parameters for a double-well potential model from the NGRIP δ^{18}O record (Krumscheid et al., 2015). Most recently, a double-well potential model was compared to a relaxation oscillator model with different external forcings using comparative Bayesian statistics (Mitsui and Crucifix, 2017). There, the conclusion was that the oscillator model is the more likely model candidate given the data and in particular that external forcing in terms of the ice volume significantly improves the statistical model. The latter three studies, however, used a lower-resolution version of the record, and did not consider either memory terms or coupling between the δ^{18}O record and the dust record, as will be done here.

2 Data and methods

2.1 High-resolution NGRIP data

We employ proxy records of δ^{18}O ratios (Andersen et al., 2004; Gkinis et al., 2014) and dust concentrations (Ruth et al., 2003) from the same core at the NGRIP drilling site. The δ^{18}O ratios were regularly sampled every 5 cm, while the dust concentrations were sampled at a resolution of 1 mm (Ruth et al., 2002, 2003). The dust record was resampled here to the lower 5 cm resolution for consistency with the δ^{18}O record. The proxy time series for the two variables have a common chronology, referred to as GICC05 (Svensson et al., 2008), for the time interval starting at approximately 59 ka b2k. Here, 1 ka equals 1000 a and "b2k" refers to "before AD 2000." The GICC05 chronology is based on counting annual layers, which are distinguishable due to seasonal variations in the ice (Andersen et al., 2006), and the sampling intervals along the core range from 1 to 7 a.

Dating uncertainties were reported for each measurement of the raw data, and they accumulate toward the more remote past, as a consequence of the layer-counting procedure, which starts from the top of the core (Svensson et al., 2008; Rasmussen et al., 2014). The dating uncertainties are reported to be roughly 5 % and reach a maximum counting error of 2573 a at an age of 59 420 a b2k. Note that there are no uncertainties in the relative timing of the δ^{18}O and dust, since they are obtained from the same ice core.

The δ^{18}O ratios are interpreted as proxies for surface air temperature variability, with algebraically higher ratios corresponding to warmer temperatures (Johnsen et al., 1992, 2001a; Dansgaard et al., 1993; Andersen et al., 2004). The dust concentrations have been proposed as proxies of large-scale atmospheric circulation changes, with higher particle counts being associated with stronger winds and thus with larger equator-to-pole temperature differences (Fischer et al., 2007; Steffensen et al., 2008). Although the two proxies thus represent very distinct climatic features, they exhibit a high degree of co-variability for the time interval 59–22 ka b2k, as previously reported, e.g., by Johnsen et al. (1997), Ruth et al. (2003) and Rasmussen et al. (2014). This co-variability is best seen when considering the negative natural logarithm $-\log(\text{dust})$ of the data along with the δ^{18}O data (Fig. 1c).

The raw time series (blue curves in Fig. 1a and b) are preprocessed as follows: first, both time series are interpolated to an equidistant time axis with 5 a intervals. Second, gaps of varying length present in the dust time series and totaling about 6 % of the data points are filled by next-neighbor interpolation. Third, the multimillennial trend is removed from both time series using a 40 ka running mean. Fourth, the noise level of the δ^{18}O record is reduced by applying a Butterworth low-pass filter of order 4, with a cutoff frequency of 0.02, corresponding to a period of 50 a.

2.2 Approximating the GLE in practice

It has recently been shown in general terms how to approximate the GLE (2) by a set of SDEs that is relatively easy to derive from the observables $\boldsymbol{x}$ (Kondrashov et al., 2015). Their MSMs both generalize EMR models and provide the correct time-continuous limit of such models. Based on their work, we directly derive here an approximation of Eq. (2) in terms of Eq. (3).

Compared to the GLE (2), the non-Markovian term with general kernel $\mathcal{G}$ is discretized and replaced by a sequence of Dirac kernels to obtain the second term on the right-hand side of Eq. (3), while the cubic term $D(\boldsymbol{x},\boldsymbol{x},\boldsymbol{x})$ guarantees the stability of the solutions. Essentially, the proposed model is thus a dynamical system with a two-dimensional double-well potential that accounts for two alternative stable states. The additional memory term modulates the transitions between the two wells, which are stochastically forced by Gaussian white noise.

In practical applications, the SDE (3) is approximated by the following system of k coupled, discrete difference equations with delays:

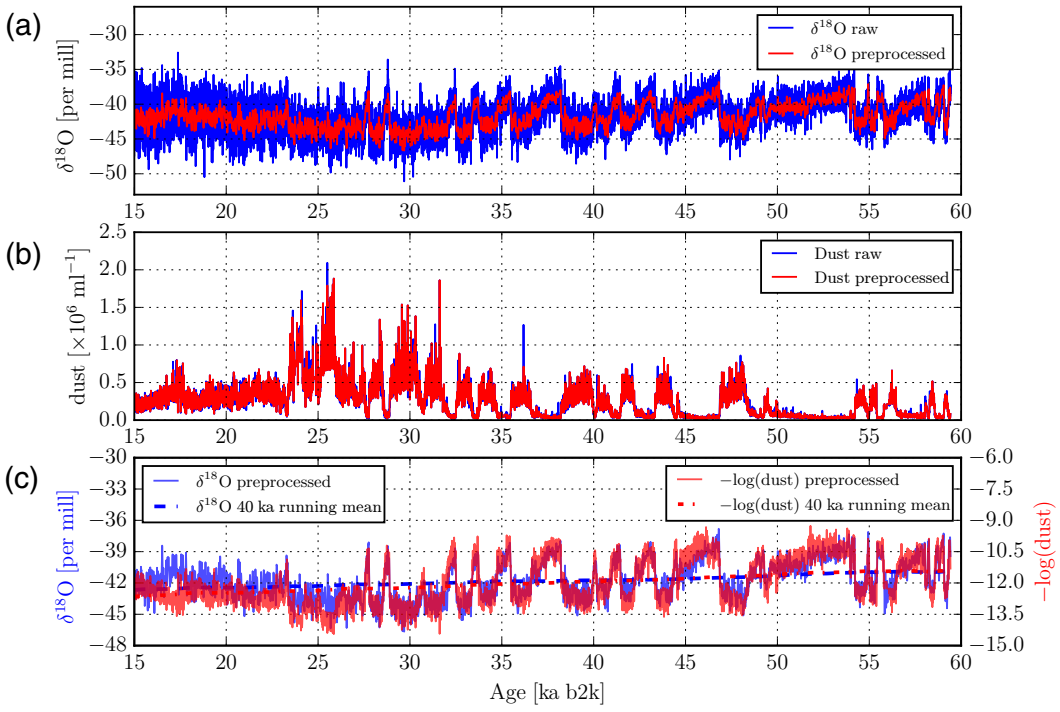


Figure 1. Time series of δ^{18}O and dust from the NGRIP record for the time interval 15–59 ka b2k; following paleo-record and geochemical use, the time axis points from the present (at left) towards the past (at right). **(a)** Raw (blue) and preprocessed (red) δ^{18}O time series. **(b)** Raw (blue) and preprocessed (red) dust time series. The two records are visually almost indistinguishable; the blue is visible only where it exceeds the red. **(c)** Low-pass filtered δ^{18}O (blue) and $-\log(\text{dust})$ (red), together with their 40 ka running means (dashed lines). Note the strong co-variability between the two preprocessed time series during the interval 59–22 ka b2k; the Pearson correlation coefficient on this interval equals $r^{\mathrm{P}} = 0.84$. For details on the preprocessing of the records see Sect. 2.1.

$$\boldsymbol{x}_{n+1} - \boldsymbol{x}_n = \left\{ A + \sum_{s=0}^{d} B_s \boldsymbol{x}_{n-s\tau} + C(\boldsymbol{x}_n, \boldsymbol{x}_n) + D(\boldsymbol{x}_n, \boldsymbol{x}_n, \boldsymbol{x}_n) \right\} \delta t_n + \mathbf{Q} (\delta t_n)^{1/2} \chi_n \,. \tag{4}$$

Here $\boldsymbol{x}_n \in \mathbb{R}^k$ denotes the k-component observed variable $\boldsymbol{x}(t)$ at time $t = t_n$, with $1 \le n \le N$, and the $\chi_n \in \mathbb{R}^k$ denote k-dimensional, independent white-noise increments. Note that for modeling the NGRIP records, we have $k = 2$ variables given by the δ^{18}O and dust measurements.

In contrast to all other model parameters, the values for d and τ used in the linear memory part in Eq. (4) are not varied in the parameter optimization: in accordance with the MZ formalism, choosing $d = 2$ suffices to obtain residuals that are sufficiently uncorrelated with the observations $\boldsymbol{x}$. Optimal values of the memory step width τ are chosen in an outer loop so as to have the averaged PDFs of the model-simulated time series, along with the average waiting time between subsequent transitions, as close as possible to those of the observed ones. A choice of $\tau = 12$ time steps, corresponding to 60 a, will provide optimal approximations of these statistical characteristics; see Fig. 5 below.

The explicit coupled SDE system governing our stochastic–dynamic model is hence given by

$$\begin{aligned} x_{n+1} - x_n = \Bigg(& A_1 + \sum_{s=0}^{2} \left(B_{11}^s x_{n-s\tau} + B_{12}^s y_{n-s\tau} \right) \\ & + C_{11} x_n^2 + C_{12} x_n y_n + C_{13} y_n^2 \\ & + D_{11} x_n^3 + D_{12} x_n^2 y_n + D_{13} x_n y_n^2 + D_{14} y_n^3 \Bigg) \delta t_n \\ & + \left(Q_{11} \xi_n^1 + Q_{12} \xi_n^2 \right) \delta t_n{}^{1/2} \,, \end{aligned} \tag{5a}$$

$$\begin{aligned} y_{n+1} - y_n = \Bigg(& A_2 + \sum_{s=0}^{2} \left(B_{21}^s x_{n-k\tau} + B_{22}^s y_{n-k\tau} \right) \\ & + C_{21} x_n^2 + C_{22} x_n y_n + C_{23} y_n^2 \\ & + D_{21} x_n^3 + D_{22} x_n^2 y_n + D_{23} x_n y_n^2 + D_{24} y_n^3 \Bigg) \delta t_n \\ & + \left(Q_{21} \xi_n^1 + Q_{22} \xi_n^2 \right) \delta t_n{}^{1/2} \,, \end{aligned} \tag{5b}$$

where $\boldsymbol{x}_n = (x_n, y_n)$ and $\tau = 60$ a. In total, there are thus 31 parameters to be estimated. Note that the total set of model parameters includes the standard deviations of the noise residuals, as well as their correlation.

2.3 Maximum likelihood estimation

Recently, Chorin and Lu (2015) introduced a general methodological framework for discrete stochastic parameterizations. In principle, an optimal parameter set Π^* for the forms A, B, C and D of order 0–3 can be determined by regressing the right-hand side of Eq. (4) onto the observed increments $\delta \boldsymbol{x}_n = \boldsymbol{x}_{n+1} - \boldsymbol{x}_n$, e.g., by ordinary least-squares (LSQ) optimization:

$$\Pi^* = \arg\min_{\Pi} \left\| \frac{\delta \boldsymbol{x}_n}{\delta t_n} - \mathbf{F}^{\Pi}(\boldsymbol{x}_n, \ldots, \boldsymbol{x}_{n-s\tau}) \right\|_2, \tag{6}$$

where $\mathbf{F}^{\Pi}$ denotes the operator corresponding to the right-hand side of Eq. (4), dropping the noise term and using a parameter combination Π.

Following, e.g., Kwasniok (2013), Chorin and Lu (2015), Krumscheid et al. (2015), Mitsui and Crucifix (2017), we rely here on Bayesian parameter inference for reduced stochastic models. For the present modeling task, we propose the Gaussian likelihood function

$$\begin{aligned} \mathcal{L}_{\mathcal{D}}(\Pi) = & \prod_n \frac{1}{2\pi |\Sigma|^{1/2}} \\ & \times \exp\left\{ -\frac{1}{2} \left(\frac{\delta \boldsymbol{x}_n}{\delta t_n} - \mathbf{F}^{\Pi}(\boldsymbol{x}_n, \ldots, \boldsymbol{x}_{n-s\tau}) \right)^T \Sigma^{-1} \right. \\ & \left. \times \left(\frac{\delta \boldsymbol{x}_n}{\delta t_n} - \mathbf{F}^{\Pi}(\boldsymbol{x}_n, \ldots, \boldsymbol{x}_{n-s\tau}) \right) \right\}; \end{aligned} \tag{7}$$

here Σ is the covariance matrix of the noise, estimated from the residuals of the least-squares optimization.

Note that the functional form of the likelihood function in Eq. (7) assumes that the residuals are normally and independently distributed. The MZ formalism only ensures that there exists a GLE with noise forcing that is uncorrelated with $\boldsymbol{x}$. The additional assumption that the noise is white in time is not theoretically guaranteed. Therefore, one has to check empirically how well Gaussian white noise approximates the residual. Given that the stochastic difference equation (Eq. 4) only approximates the theoretical GLE provided by the MZ formalism, one also needs to validate empirically that the residuals are uncorrelated with the observations $\boldsymbol{x}$ (Kondrashov et al., 2015).

We remark that the approximation of the divided differences $\delta \boldsymbol{x}/\delta t$ by the function $\mathbf{F}^{\Pi}$ is in fact a multivariate multiple linear regression, with regressors chosen to be the polynomials in Eq. (5). It can easily be seen that MLE and LSQ are equivalent for the case of univariate data and Gaussian errors, since the same term $\delta \boldsymbol{x}_n/\delta t_n - \mathbf{F}^{\Pi}$ is minimized in Eqs. (6) and (7). In fact, subject to the assumption of uncorrelated – but not necessarily Gaussian – errors with zero mean and identical variance, the Gauss–Markov theorem ensures that parameters obtained from LSQ, i.e., from Eq. (6), are the best linear unbiased estimators for such a regression. These estimators are *best* in the sense that they have the lowest variance.

On the other hand, the MLE is asymptotically optimal. In particular, it is efficient in the sense that the variance of the parameter estimates achieves the so-called Cramér–Rao lower bound, which is optimal, as the number of samples tends to infinity (Andersen, 1970). Srivastava (1965) extended the Gauss–Markov theorem to the multivariate case, where correlations among the variables lead to correlations between the error terms of the distinct variables and thus result in cross-terms in the exponent of Eq. (7). Although both optimization approaches are thus equivalent in the case at hand, the MLE approach has the advantage that the parameter estimates can be interpreted as the most likely ones, given the observed data.

3 Results

As a training set for the parameter optimization of our stochastic–dynamic model in Eq. (5), we choose the time interval 59–22 ka b2k; this interval roughly coincides with Marine Isotope Stage 3 (approximately 60–28 ka b2k). Our choice results in $N = 7529$ data points for each time series. The reason for this choice is that the layer-counted chronology has only been carried out until 59 ka b2k and that the co-variability between δ^{18}O and log(dust) is substantially reduced for the more recent part of the record, as already noted by Ruth et al. (2002, 2003) and apparent in Fig. 1c here.

We use the natural logarithm of the dust time series (i) because of the large range of dust concentration values and (ii) because it has high co-variability with δ^{18}O (cf. Fig. 1c); it also guarantees that the simulated dust values are positive. This logarithmic scale requires, however, high accuracy in the modeling of the dust concentrations to resolve the multiplicity of abrupt variations that span several orders of magnitude; cf. Fig. 1b.

Our results indicate that for $d = 2$ memory steps, the residuals of the least-squares optimization Eq. (6) are indeed uncorrelated with the observations $\boldsymbol{x}$ – the correlations are less than 10^{-8} – and they are approximately Gaussian distributed (Fig. A1), while their autocorrelations decay very fast (not shown). These tests empirically support our choice of a Gaussian likelihood function of the form Eq. (7) for the MLE. Furthermore, these results allow us to integrate the stochastic–dynamic model given in Eq. (5) with an Euler–Maruyama scheme.

The specific values of the coefficients of Eq. (5), as derived via MLE, are given in Table A1. Note, in particular, the crucial nonlinear and lagged cross-interaction terms. For the reasons explained above, the parameters that minimize the least-squares problem Eq. (6) are identical to the most likely parameter values as determined via MLE up to numerical precision, i.e., with a relative error less than 10^{-5}.

In order to simulate optimal sample time series for the δ^{18}O and log(dust) variables, the parameter combination Π^* that maximizes the likelihood function (Eq. 7) is used. The

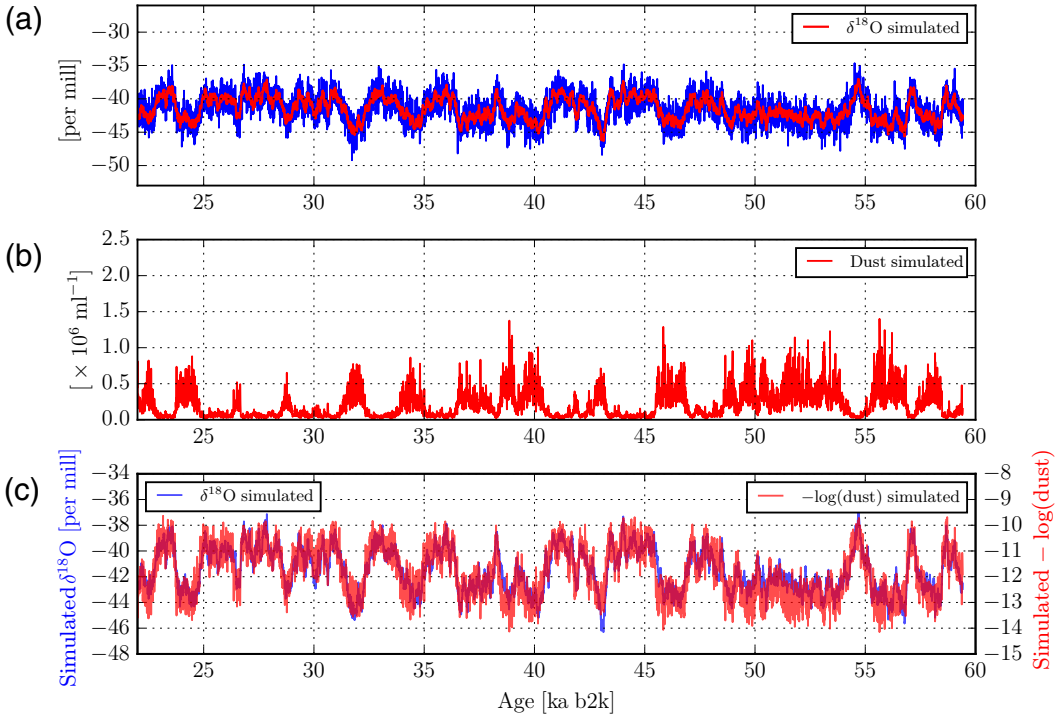


Figure 2. Simulated δ^{18}O and dust time series. **(a)** Simulated δ^{18}O time series (red). Gaussian white-noise approximation of the residual removed during the low-pass filtering of the original time series is added back in to obtain the full time series (blue). **(b)** Simulated dust time series, obtained as the exponential of the simulated log(dust). **(c)** Simulated δ^{18}O time series (blue, same as in A) together with $-\log(\text{dust})$ (red). Note that the strong co-variability between the two variables is captured very well by the model: Pearson's correlation coefficient, averaged over 1000 simulated time series, is $r^{\mathrm{P}} = 0.85$.

model is then integrated by the Euler–Maruyama scheme with a uniform step size of $\delta t = 10^{-5}$. Compared to the observational data, this step size corresponds to the sampling size of 5 a. Essentially, we have thus rescaled the time unit in order to guarantee a stable numerical integration. The stochastic forcing is given by two-dimensional Gaussian white-noise increments multiplied by the Cholesky matrix $\mathbf{Q}$. The residuals of the δ^{18}O and log(dust) are correlated with a Pearson's correlation coefficient $r^{\mathrm{P}} = -0.13$, resulting in a non-diagonal covariance matrix $\Sigma = \mathbf{Q}\mathbf{Q}^{\mathrm{T}}$.

Illustrative time series of δ^{18}O and dust, simulated using the most likely model parameter combinations, are shown in Fig. 2. In Fig. 3a and b, the dashed lines show averages and uncertainties in PDFs of 1000 time series simulated using the most likely parameter combinations. Figure 3c and d show the average spectral densities of the latter time series. In this case, error bars would not be visible and are therefore omitted.

The means and standard deviations of the observed time series are reproduced well by the simulations: for the preprocessed δ^{18}O, the mean and standard deviation equal -41.79 ± 1.72 compared to -41.82 ± 1.73 for the simulations; the corresponding preprocessed and simulated values for log(dust) are 11.98 ± 0.97 and 12.00 ± 0.92. For the simulations, these values are computed as averages over 1000 simulated sample time series, obtained using the parameter combinations that maximize the likelihood function Eq. (7).

The simulated time series (Fig. 2) exhibit abrupt changes that resemble the so-called Dansgaard–Oeschger events, which mark the sharp transitions from colder stadials to warmer interstadials (Dansgaard et al., 1993; Johnsen et al., 2001b; Rasmussen et al., 2014) in the original time series (Fig. 1). Given the more gradual temperature changes from interstadials to stadials, the red curve in Fig. 2a has a dominant sawtooth-shape pattern. Recall that time here, as in Fig. 1, runs from right to left, as it does in Fig. 3a and c of (Krumscheid et al., 2015), which also display a high qualitative resemblance between their model-simulated and observed δ^{18}O time series.

The observed time series are, due to the sawtooth-shaped transitions between stadials and interstadials, not symmetric under time reversal. A quantitative measure of the time-reversal asymmetry is provided by the third-order moment (Kwasniok, 2013):

$$M(\theta) = \langle x(t)x^2(t+\theta) - x^2(t)x(t+\theta)\rangle_t \,.$$

Figure 4b shows that, for values of θ up to roughly 1000 a, $M(\theta)$ computed from dust simulations does exhibit a some-

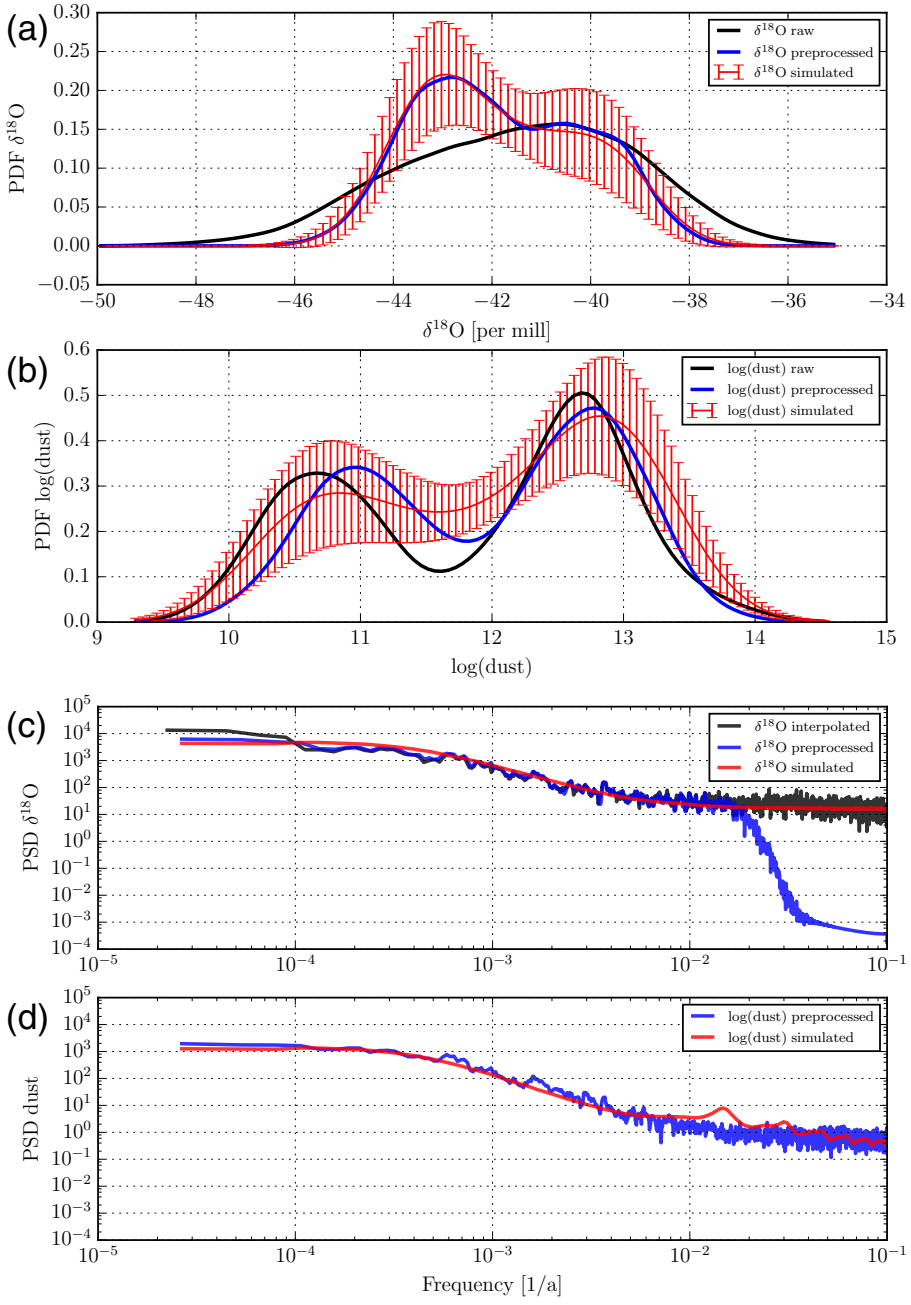


Figure 3. Statistical properties of the observed and simulated δ^{18}O and dust time series. **(a)** PDFs for the raw, preprocessed, simulated δ^{18}O. Note how the bimodality, corresponding to the transitions between stadials and interstadials, only arises after applying a low-pass filter to the time series, leading to the preprocessed time series. **(b)** PDFs for the raw, preprocessed, and simulated log(dust) time series. PDFs are obtained as averages over 1000 simulated time series, each obtained from the most likely model parameters (solid red). Error bars indicate the 2σ range of uncertainties derived from theses 1000 sample time series. **(c)** Power spectral densities for the interpolated, preprocessed (i.e., interpolated and smoothed) and simulated δ^{18}O. **(d)** Power spectral densities (PSDs) for the preprocessed and simulated log(dust). For the simulated time series, Gaussian white noise with the same standard deviation as the residual of the LSQ (Eq. 6) was added before computing the PSD. The PSDs are estimated using the multi-taper method (Vautard et al., 1992), with a total of seven tapers. The PSD shown for the simulations is an average over 1000 simulated time series and is therefore strongly smoothed. For the spectra, the uncertainties would be hardly visible and are therefore omitted.

what similar behavior to the $M(\theta)$ computed from the observed dust. Quantitatively, the similarity between $M(\theta)$ computed for the observed and simulated dust is also supported by Kendall's τ value (red curve in Fig. 5d). The temporal asymmetry is, however, not reproduced by the δ^{18}O simulations (Fig. 4a), as also confirmed by Kendall's τ (blue curve in Fig. 5d).

Given the very strong correlations r^{P} between the two variables, with $r^{\mathrm{P}} \approx 0.9$ for both observations and simulations, the discrepancies between the resulting $M(\theta)$ given δ^{18}O or dust data are rather surprising. These discrepancies suggest that $M(\theta)$ might not be the best measure for quantifying whether sawtooth-shaped oscillations are present or not.

Following Krumscheid et al. (2015), we define the average waiting time τ_{DO} between Dansgaard–Oeschger events as the sum of the average residence times in stadials and interstadials. For this purpose, stadials and interstadials are determined as intervals for which the time series are, respectively, below or above the mean of the series. Due to their comparably high noise level, the high-resolution time series employed here are further smoothed by singular spectrum analysis (SSA) (Vautard et al., 1992; Ghil et al., 2002) using a window size of 500 a and keeping only the five leading reconstructed components. For the observed δ^{18}O and log(dust), we obtain $\tau_{\mathrm{DO}} = 1506$ and 1744 a, respectively, compared to $\tau_{\mathrm{DO}} = 1370 \pm 224$ a and 1559 ± 346 a for the simulations, computed as averages over 1000 simulated time series. Note that this definition of waiting times may not be optimal in view of the high noise level of the time series, leading to different values for δ^{18}O and log(dust). This definition is nevertheless employed here for the purpose of facilitating comparison with the results of Krumscheid et al. (2015).

For the dust concentrations (Fig. 2b), there is a striking similarity between the observed and simulated time series in terms of episodes with low dust concentrations of variable durations, as well as the presence of burst episodes of variable magnitudes. Over the training interval, the preprocessed time series are correlated at $r^{\mathrm{P}} = -0.84$, which equals the average correlation between the simulated time series.

The PDFs of the simulated δ^{18}O and log(dust) (red solid lines), obtained as averages over 1000 time series, are quite similar in shape to those of the preprocessed time series (blue solid lines) of both observed variables; see Fig. 3a and b. In particular, the bimodality of the PDFs, which reflects the relative persistence of stadials and interstadials, is reproduced by our inverse model. Uncertainties in the PDFs are derived on the basis of ensembles of 1000 simulated time series, produced by sampling from the most likely parameter combination. Error bars on the PDFs in Fig. 3a and b reflect the 2σ range of these ensembles.

The spectra of both the δ^{18}O and the log(dust) are well reproduced by our model for the entire frequency range, including sub-centennial periodicities; see the black and red curves in Fig. 3c.

4 Discussion

We studied the δ^{18}O and dust time series obtained from the high-resolution NGRIP record for the 37 000 a interval 59–22 ka b2k. The results described above show that the statistical properties of these time series – such as their PDFs, spectra and average waiting times – can be approximated quite well by the proposed stochastic–dynamic model of Eq. (3).

The main features of our inverse model are (i) the nonlinear terms in the Markovian part, (ii) the inclusion of non-Markovian memory terms and (iii) the coupling terms between the two time series. Cubic terms have previously been used to model the δ^{18}O time series of the NGRIP record (Ditlevsen et al., 2007; Kwasniok, 2013; Krumscheid et al., 2015). Non-Markovian contributions in an SDE model have, so far, only been considered in very few paleoclimatic studies. Pelletier (2003) studied the influence of delayed ice volume feedbacks on glacial–interglacial variability. In a deterministic setting, Rial (2004) used a forced logistic delay differential equation model to study DO cycles as well as glacial–interglacial transitions, and Berger (1999) discussed memory contributions in the context of the 100 ka cycle and its association with Milankovitch forcing. We are not aware of modeling efforts that take advantage of the co-variability between the δ^{18}O and log(dust) variables. It is shown in the following that the coupling terms between the two variables are crucial in order to capture the statistical characteristics of the measured NGRIP time series presented above, while the non-Markovian contribution is also significant.

The nonlinear terms can be physically motivated by the fact that the observed time series oscillate between two quasi-equilibria, namely the stadials and interstadials. If only linear terms were used, the bimodality of the observed time series, and hence the existence of two quasi-stable states, could not be reproduced (see Figs. A2 and A3).

The purely Markovian form of the model approximates the PDFs of the observed time series less well (Figs. 5c and A5). In particular, the bimodality of the PDFs for both δ^{18}O and log(dust) is weaker in this case, indicating that the memory terms do contribute to an appropriate modeling of the persistence in the stadials and interstadials, as well as of the transitions between them. In addition, the average waiting times between DO-like transitions are $\tau_{\mathrm{DO}} = 1404$ a for the δ^{18}O and 1254 a for log(dust), and they are thus too short for the log(dust) time series, in particular.

When removing all coupling terms between the two variables from the inverse-model Eq. (4), sawtooth-shaped transitions are completely absent from the simulated time series (Fig. A6). Furthermore, the bimodality of the PDFs is missed when excluding the couplings (Fig. A7), and the average waiting times are much too short, namely $\tau_{\mathrm{DO}} = 1208$ a for the δ^{18}O record and 867 a for the log(dust) record.

Following previous authors (Kwasniok, 2013; Krumscheid et al., 2015; Mitsui and Crucifix, 2017), we also compare the different model versions in terms of the Bayesian (BIC)

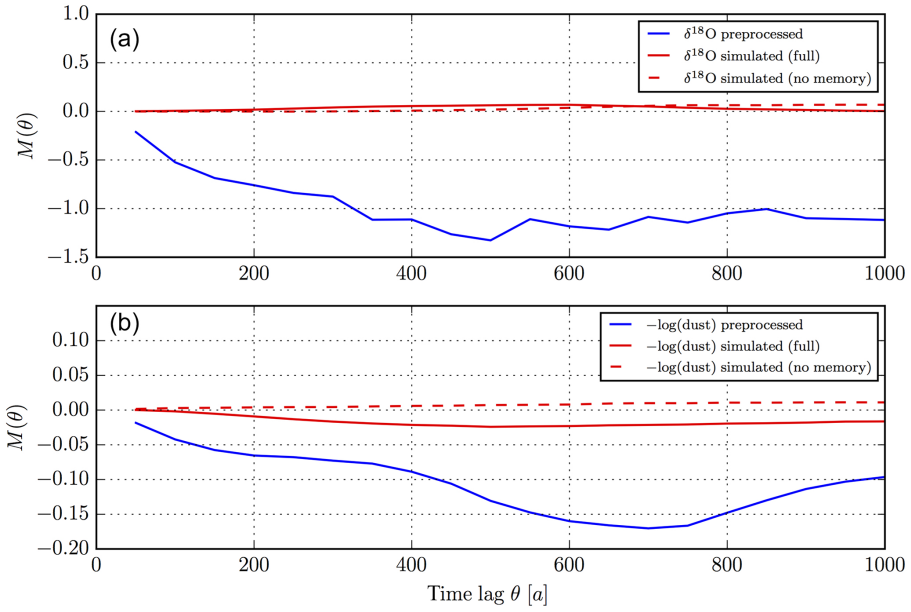


Figure 4. Third-order statistical moment $M(\theta) = \langle x(t)x^2(t+\theta) - x^2(t)x(t+\theta)\rangle_t$ for the observed NGRIP time series (solid blue), the full model including memory terms with step size $\tau = 60$ a (solid red) and the model without memory terms (dashed red). **(a)** $M(\theta)$ for the δ^{18}O time series and **(b)** $M(\theta)$ for the dust time series. For the simulations, the average of $M(\theta)$ over 1000 samples is shown. Note that, for increasing delays θ, the values of $M(\theta)$ are affected more and more by the nonstationarity of the data and should therefore be interpreted with care.

Table 1. Sample-size-corrected Akaike information criteria (AICc) and Bayesian information criteria (BIC) for the different model versions. Note that the model parameters include the standard deviations and correlation that appear in the respective model's noise term. For the models with memory, we chose $d = 2$ memory steps with step size $\tau = 12$, corresponding to 60 a.

Nonlinear	Memory	Coupling	No. of parameters	AICc	BIC
×	×	×	31	4501.93	4287.57
	×	×	17	4701.05	4583.44
×		×	23	5089.18	4929.87
×	×		14	5228.15	5131.28
		×	9	6058.83	5996.52
×	×	×	31	4287.57	4501.93
	×	×	17	4583.44	4701.05
×		×	23	4929.87	5089.18
×	×		14	5131.28	5228.15
		×	9	5996.52	6058.83

and sample-size-corrected Akaike (AICc) information criteria; these are defined as $\mathrm{AICc} = 2pn/(n-p-1) - 2\log(\mathcal{L}^*)$ and $\mathrm{BIC} = p\log(n) - 2\log(\mathcal{L}^*)$. Here, p denotes the number of model parameters, $n = N - 1 = 7528$ the total number of data points (a total number of data points N leads to $N-1$ increments to be approximated) and $\mathcal{L}^*$ the maximum value of the likelihood function Eq. (7). The lower the value of AICc or BIC for a given model, the higher the relative confidence in that model.

For the case at hand, both AICc and BIC consistently favor the full model, which includes nonlinear, memory and coupling terms. This is followed by the linear coupled model with memory terms, the nonlinear coupled model without memory terms, the nonlinear model with memory but without coupling terms, and finally the linear model without memory terms (Table 1).

Note that the AICc penalizes higher numbers p of model parameters less strongly than the BIC. Although the AICc was found, under certain conditions, to be optimal in selecting model candidates (e.g., Burnham and Anderson, 2002; Yang, 2005), notable counterexamples are known (e.g., Penland et al., 1991).

We thus suggest interpreting the values in Table 1 with caution. For example, in the case at hand both AICc and BIC

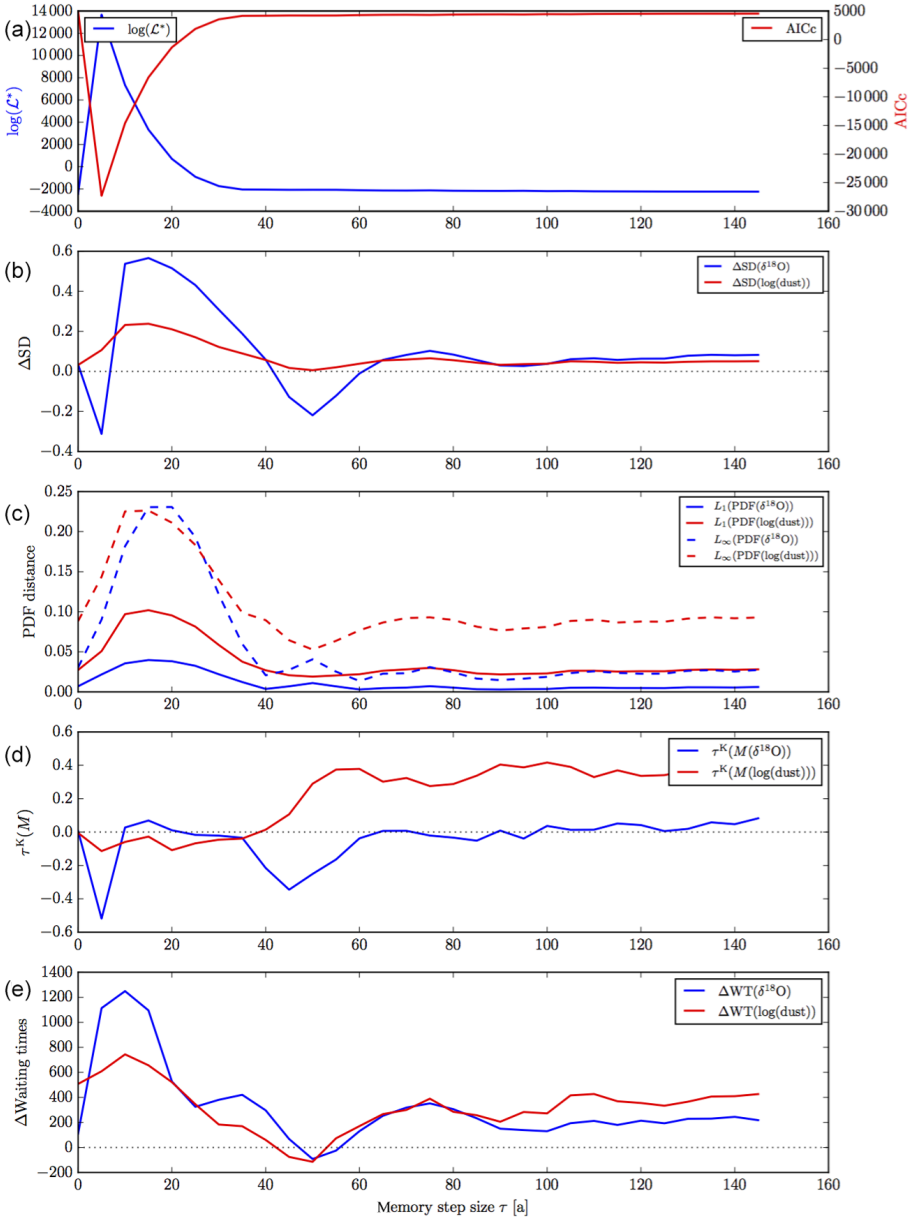


Figure 5. Summary statistics for different memory step sizes τ. **(a)** Log-likelihood and AICc values for different memory step sizes τ. The AICc is computed as $\mathrm{AICc} = 2pn/(n-p-1) - 2\log\mathcal{L}^*$. **(b)** Difference between observed and simulated standard deviations. **(c)** L^2 and L^∞ distances between observed and simulated PDFs. **(d)** Kendall's τ statistic between the third-order moments $M(\theta)$, computed for observed and simulated time series. **(e)** Difference between observations and simulations in terms of average waiting times between subsequent transitions from stadials to interstadials. In **(b–e)**, statistics for simulated time series are obtained as averages over 1000 simulations using the full model.

propose higher confidence in the linear model than in the model without memory terms. This would suggest that the memory terms are more important than the higher-order parameters that correspond to the double-well shape of the potential. However, the PDFs of the cubic model without memory terms are considerably closer to the observed PDFs than the PDFs obtained from the linear model including memory terms. In particular, by construction, the latter model cannot reproduce the bimodality of the observed PDFs. We would thus still argue that the nonlinear contributions are more important than the memory terms. Nevertheless, the AICc and BIC values presented herein provide information-theoretic evidence that the inclusion of memory terms does substantially improve the model.

We emphasize that the full model proposed herein has the highest number of parameters out of the different candidates but is still the one with the lowest BIC and AIC. Therefore, it can be argued that this number of parameters is not too high,

and it is not likely that the full model over-fits the observed data.

Furthermore, it should be noted that the values of AICc and BIC can only be compared on the basis of the same underlying data. Since we use a higher-resolution version of the NGRIP data as compared to the previous authors (Kwasniok, 2013; Krumscheid et al., 2015; Mitsui and Crucifix, 2017), the values for AICc and BIC presented here cannot be compared to the values reported in those studies.

As noted above, we chose for the memory step size $\tau = 60$ a, corresponding to 12 time steps. In fact, the smallest values of both AICc and BIC are obtained at a considerably smaller step size of $\tau = 1$ time step (5 a) (Fig. 5a). However, the approximation of the correct standard deviations, PDFs, third-order moments and average waiting times by the simulations is strongly improved for larger values of τ (Fig. 5b–e). Therefore, we suggest that a choice of $\tau = 60$ a provides a good tradeoff between the AICc and BIC, on the one hand, and the statistical characteristics of the simulated time series, on the other. In particular, an accurate reproduction of the correct average waiting time between subsequent transitions from stadials to interstadials is only achieved for $55\,\text{a} \lesssim \tau \lesssim 60\,\text{a}$. It should be emphasized, moreover, that both the AICc and BIC criteria consistently favor the models with memory terms over the models without memory terms, regardless of the specific value of τ.

The model results presented here appear only in the high-resolution version of the NGRIP ice core record, which was originally sampled every 5 cm, a depth sampling that yielded temporal step sizes between 1 and 7 a. For example, interpolating the raw data to a uniform grid with $\Delta t = 10$ a, instead of 5 a, leads to substantially less accurate approximations of the observed statistical properties of both the δ^{18}O and dust time series (not shown). On the other hand, interpolating the raw data with a uniform sampling step of 3 a is problematic because of the original temporal step sizes, and does not further improve the results (not shown). It would thus appear that the 5 a uniform grid size is nearly optimal, given the irregularities in the sampling and the uncertainties in both the dating and the values of the records.

5 Conclusions

We have shown that a coupled, two-dimensional stochastic–dynamic model with cubic drift term and linear delay terms is capable of reproducing the statistical properties of δ^{18}O and dust time series derived from the high-resolution NGRIP record for the interval 59–22 ka b2k, which roughly corresponds to Marine Isotope Stage 3. These statistical properties are expressed in terms of the PDFs of the time series, their power spectra and the waiting times between sharp transitions from stadials to interstadials.

Key ingredients for an accurate simulation of the observed time series are as follows.

1. High-resolution time series have to be used as training data, indicating that the high-frequency variability present in the records plays a vital role for the overall evolution of the climate processes that generated the NGRIP ice core. Interpolation of the raw data, which is sampled at depth intervals of 5 cm in the core, to 5 a intervals in the preprocessed time series was found to be optimal in our inverse-model setup. This finding is qualitatively consistent with the assertion of Rypdal (2016) that an increase in decadal-scale variability may be a statistical precursor for the abrupt transitions from stadials to interstadials during Dansgaard–Oeschger events.

2. Cubic terms need to be included in the Markovian part of the model. This can be physically motivated by the presence of two quasi-equilibria in the observed time series – the stadials and interstadials – that could not be modeled without two such quasi-stable states in the underlying dynamical system. Cubic terms have already been considered in previous attempts to model the δ^{18}O time series of the NGRIP record (Ditlevsen et al., 2007; Kwasniok, 2013; Krumscheid et al., 2015); these attempts, however, did not include the dust series, used lower-resolution data and did not consider memory effects.

3. Coupling terms between the δ^{18}O and dust variables substantially improve the statistical characteristics of the simulated time series: the reproduction of the bimodality of the PDFs, as well as of the correct average waiting time between subsequent transitions from stadials to interstadials, are substantially improved when coupling terms are included.

4. Non-Markovian terms that account for memory effects are helpful. Their inclusion allows, to some extent, reproducing the time-reversal asymmetry of the dust time series. The main contribution of including memory terms into the model is, however, to improve the average simulated waiting times between subsequent transitions from stadials to interstadials (cf. Fig. 2e) for $55\,\text{a} \lesssim \tau \lesssim 60\,\text{a}$. The memory terms also help improve the PDFs of simulated time series, in particular for the δ^{18}O time series. Furthermore, in general, the AICc and BIC criteria support the model versions that include memory terms.

Our results demonstrate that the statistical characteristics of the roughly 40 ka long, high-resolution NGRIP time series of δ^{18}O and dust considered here can be reproduced by a nonlinear inverse model without taking into account exogenous forcing, whether astronomical, solar or volcanic. Based on our results alone, there would thus be no reason to assume that the temporal evolution of the δ^{18}O ratios and dust concentrations – and hence that of the climatic variabilities they

represent, in particular the DO transitions between stadials and interstadials – are externally forced.

We note, though, that Mitsui and Crucifix (2017) have recently found evidence that including orbital forcing in modeling the 20 a averaged NGRIP Ca^{2+} time series improves their model's BIC score. This forcing – however important on longer, astronomical timescales – does not appear to be directly relevant for the millennial-scale DO transitions, according to the present results.

The predictive power of the proposed stochastic–dynamic model for the abrupt transitions from stadials to interstadials should be addressed in future work.

Data availability. The high-resolution NGRIP data used in this study are available online at http://www.icecores.dk.

Appendix A: Key ideas in deriving the GLE

The approach to data-driven stochastic–dynamic modeling taken here is rooted in the MZ formalism of statistical mechanics (Zwanzig, 1964; Mori, 1965; Chorin et al., 2002; Chorin and Hald, 2013), which proposes an integro-differential closed form for model inference from partial data. While the derivation of such an MZ model does not allow one to easily simulate its solutions, it is possible, under suitable hypotheses, to obtain a good approximation thereof by a finite number of coupled SDEs (Chekroun et al., 2011; Kondrashov et al., 2015).

Assume that z is a high-dimensional state vector, whose temporal evolution is governed by the following system of ordinary differential equations, which is not explicitly known:

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \mathbf{F}(z), \qquad z \in \mathbf{R}^n, \tag{A1}$$

where $\mathbf{R}$ denotes the set of real numbers. Assume furthermore that z can be decomposed into a sum of an observed vector $\boldsymbol{x}$ and an unobserved vector $\boldsymbol{y}$, i.e., $z = \boldsymbol{x} + \boldsymbol{y}$.

By orthogonally projecting Eq. (A1) onto the subspace spanned by the observed variables $\boldsymbol{x}$, via $Pz = \boldsymbol{x}$, we obtain

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = P\mathbf{F}(\boldsymbol{x}+\boldsymbol{y})\,, \tag{A2}$$

which depends on the unobserved variable $\boldsymbol{y}$. By introducing the averaging

$$\overline{P\mathbf{F}(\boldsymbol{x})} := \int_Y P\mathbf{F}(\boldsymbol{x}+\boldsymbol{y})\,\mathrm{d}\mu_{\boldsymbol{x}}(\boldsymbol{y}), \tag{A3}$$

where $\mu_{\boldsymbol{x}}$ denotes the probability distribution of $\boldsymbol{y}$ conditioned on $\boldsymbol{x}$, we obtain

$$P\mathbf{F}(\boldsymbol{x}+\boldsymbol{y}) = \underbrace{\overline{P\mathbf{F}(\boldsymbol{x})}}_{\text{averaged part}} + \underbrace{\left(P\mathbf{F}(\boldsymbol{x}+\boldsymbol{y}) - \overline{P\mathbf{F}(\boldsymbol{x})}\right)}_{\text{fluctuating part}}. \tag{A4}$$

The parameterization of the fluctuating part in Eq. (A4) is at the core of any stochastic parameterization method – whether linear (Penland and Sardeshmukh, 1995) or nonlinear (Majda et al., 2001) – as well as of the MZ formalism.

Ergodic-type arguments show that the averaged part can in principle be learned from a time series, assuming the existence of a "nice" invariant measure (Chekroun et al., 2011; Kondrashov et al., 2015):

$$\operatorname{argmin}_{f\in\mathcal{E}}\left(\lim_{T\to\infty}\frac{1}{T}\int_0^T\left\|\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} - f(\boldsymbol{x}(t;\boldsymbol{y}_0))\right\|^2\mathrm{d}t\right) = \overline{P\mathbf{F}(\boldsymbol{x})}, \tag{A5}$$

which holds for almost all $\boldsymbol{y}_0$ with respect to the Lebesgue measure; see Lemma 4.1 in Kondrashov et al. (2015) for further details.

Table A1. The coefficients of the explicit SDE system, obtained from MLE.

	$\delta^{18}\mathrm{O}(x)$	$\mathrm{Dust}(y)$
	$A_1 = -7.54\times10^6$	$A_2 = 2.63\times10^7$
x_n	$B^0_{11} = -3.16\times10^5$	$B^0_{21} = 9.03\times10^5$
y_n	$B^0_{12} = 8.01\times10^5$	$B^0_{22} = -3.69\times10^6$
$x_{n-\tau}$	$B^1_{11} = 1.26\times10^3$	$B^1_{21} = 9.03\times10^2$
$y_{n-\tau}$	$B^1_{12} = 5.49\times10^3$	$B^1_{22} = 1.38\times10^4$
$x_{n-2\tau}$	$B^2_{11} = -3.89\times10^1$	$B^2_{21} = 2.31\times10^3$
$y_{n-2\tau}$	$B^2_{12} = 9.19\times10^2$	$B^2_{22} = 6.30\times10^3$
x_n^2	$C_{11} = -5.09\times10^3$	$C_{21} = 1.38\times10^4$
$x_n y_n$	$C_{12} = 1.68\times10^4$	$C_{22} = -6.51\times10^4$
y_n^2	$C_{13} = -3.97\times10^4$	$C_{23} = 1.92\times10^5$
x_n^3	$D_{11} = -3.48\times10^1$	$D_{21} = 1.26\times10^2$
$x_n^2 y_n$	$D_{12} = 4.09\times10^1$	$D_{22} = 9.07\times10^1$
$x_n y_n^2$	$D_{13} = -6.30\times10^2$	$D_{23} = 3.09\times10^3$
y_n^3	$D_{14} = 3.33\times10^2$	$D_{24} = -1.72\times10^3$
ξ_n^1	$Q_{11} = 77.61$	$Q_{21} = -13.06$
ξ_n^2	$Q_{21} = 0$	$Q_{22} = 99.97$

Note that Eq. (3) can be approximated by a Markovian SDE by using the Galerkin approximation techniques of Chekroun et al. (2016) (their Appendix C and Remark 5.1). In that respect, Eq. (3) can be put within an SDE format consistent with, although different from, the MSMs discussed by Kondrashov et al. (2015).

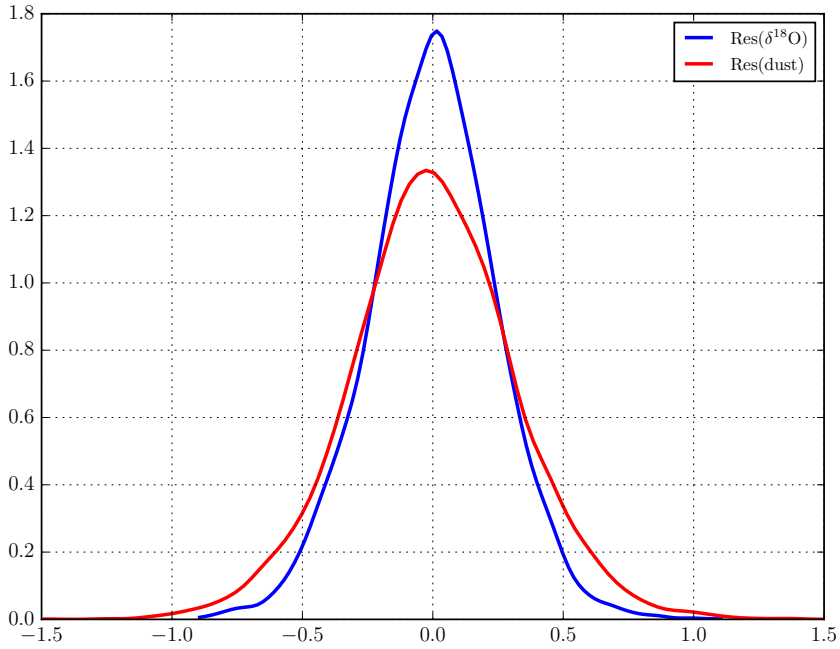


Figure A1. Probability density of the least-squares residuals for δ^{18}O (blue) and log(dust) (red). The residuals for the log(dust) can be approximated very well by a Gaussian, but for the δ^{18}O the approximation is also sufficiently good to justify the choice of a Gaussian likelihood function for the maximum likelihood estimation.

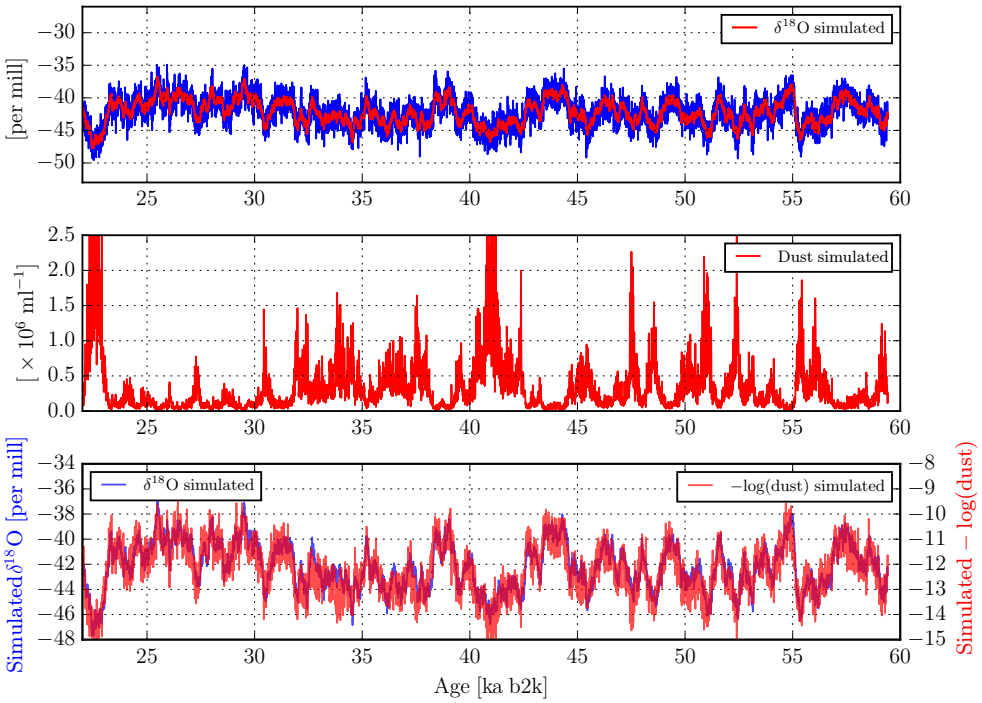


Figure A2. Same as Fig. 2 in the main text but for the model without nonlinear terms.

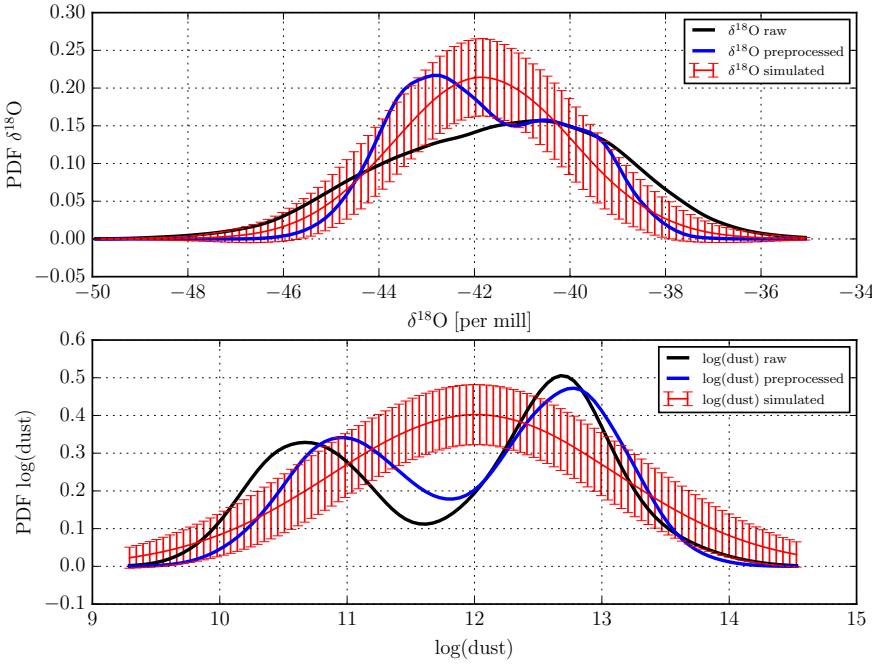


Figure A3. Same as Fig. 3a and b in the main text but for the model without nonlinear terms.

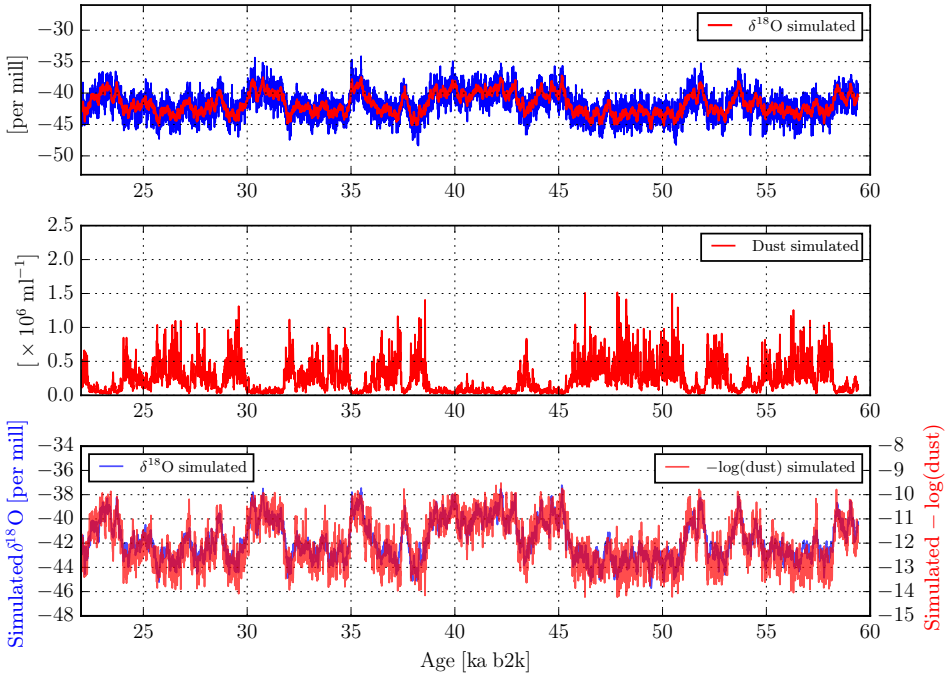


Figure A4. Same as Fig. 2 in the main text but for the model without memory terms.

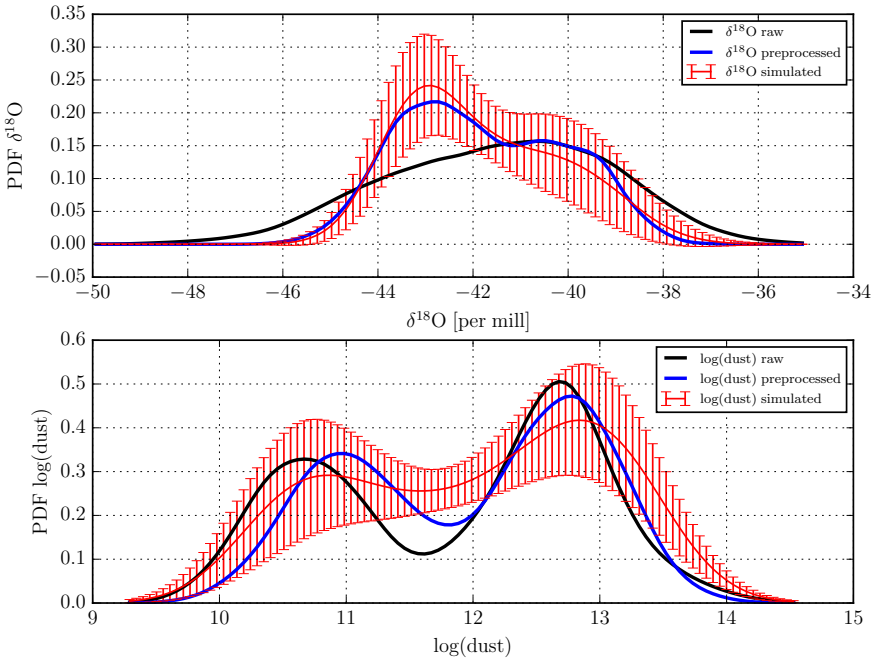


Figure A5. Same as Fig. 3a and b in the main text but for the model without memory terms.

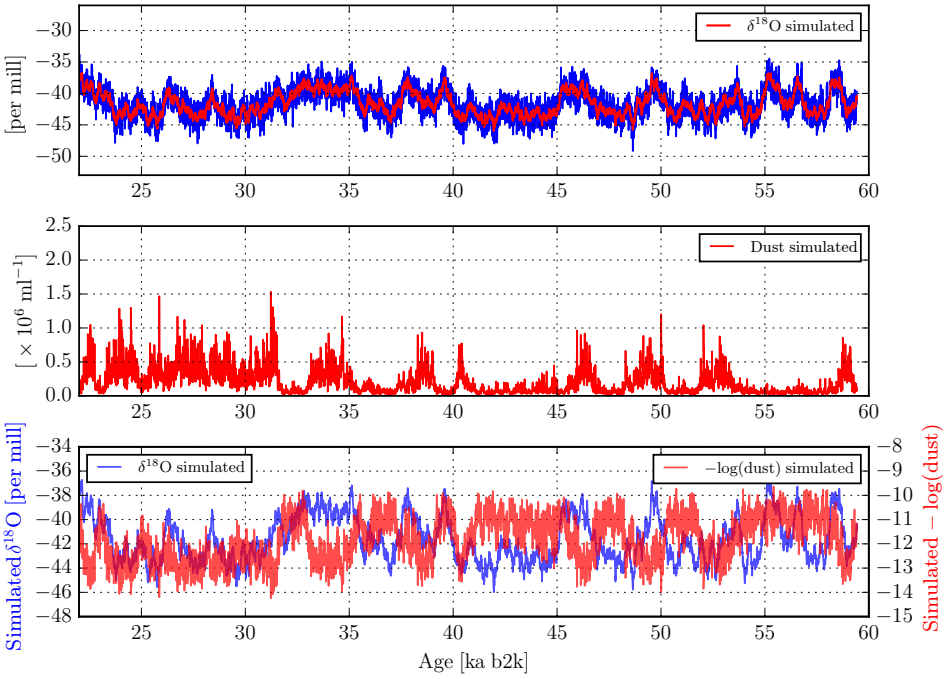


Figure A6. Same as Fig. 2 in the main text but for the model without coupling terms. Note that a coevolution of the observed δ^{18}O and log(dust) time series can in this case not be expected to be reproduced.

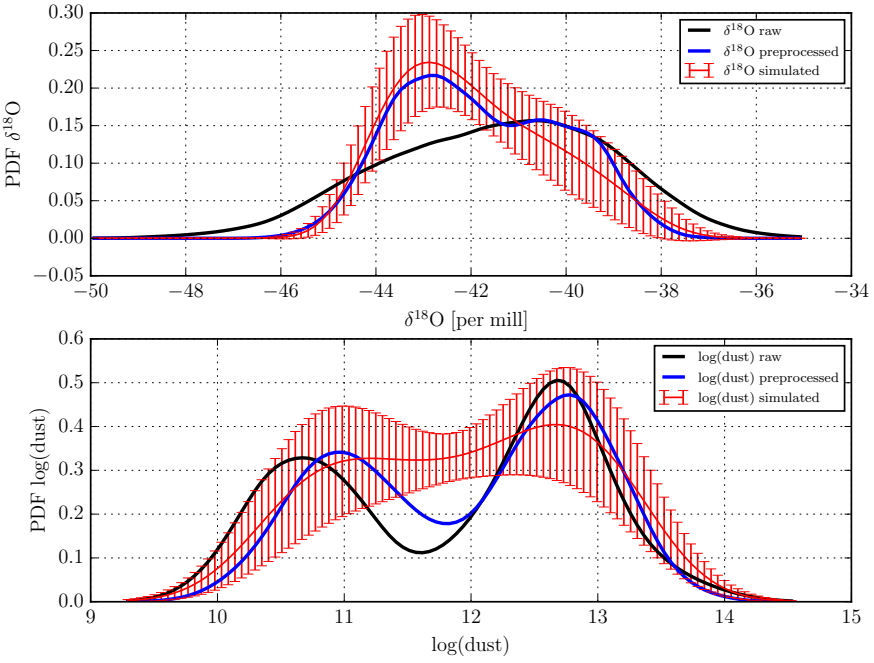


Figure A7. Same as Fig. 3a and b in the main text but for the model without coupling terms.

Author contributions. MB, D-DR and AS provided the data. NB, MDC, MG, DK and HL conceived of the research. NB conducted the numerical analysis and prepared the manuscript. All authors discussed the results, drew conclusions and edited the manuscript.

Competing interests. The authors declare that they have no competing financial interests.

Acknowledgements. This research was initiated by a collaboration between D.-D. Rousseau and the late Sigfús Johnsen, to whose memory it is dedicated. N. Boers acknowledges funding by the Alexander von Humboldt Foundation and the German Federal Ministry for Education and Research. M. D. Chekroun, M. Ghil, D. Kondrashov and H. Liu acknowledge support by grant N00014-16-1-2073 from the Multidisciplinary University Research Initiative (MURI) of the Office of Naval Research and by National Science Foundation grant OCE-1243175. M. D. Chekroun and H. Liu also acknowledge support by National Science Foundation grants DMS-1616981 and DMS-1616450, respectively. D. Kondrashov also acknowledges support by the Government of the Russian Federation (agreement no. 14.Z50.31.0033 with the Institute of Applied Physics of RAS). This is LDEO contribution no. 8167. NGRIP is directed and organized by the Ice and Climate research group, Niels Bohr Institute, University of Copenhagen. It is supported by funding agencies in Denmark (FNU), Belgium (FNRS-CFB), France (IPEV and INSU/CNRS), Germany (AWI), Iceland (RannIs), Japan (MEXT), Sweden (SPRS), Switzerland (SNF) and the USA (NSF, Office of Polar Programs).

Edited by: Anders Levermann
Reviewed by: Takahito Mitsui and one anonymous referee

References

Andersen, E. B.: Asymptotic properties of conditional maximum-likelihood estimators, J. Roy. Stat. Soc. B Met., 32, 283–301, 1970.

Andersen, K. K., Azuma, N., Barnola, J.-M., Bigler, M., Biscaye, P., Caillon, N., Chappellaz, J., Clausen, H. B., Dahl-Jensen, D., Fischer, H., Flückiger, J., Fritzsche, D., Fujii, Y., Goto-Azuma, K., Grønvold, K., Gundestrup, N. S., Hansson, M., Huber, C., Hvidberg, C. S., Johnsen, S. J., Jonsell, U., Jouzel, J., Kipfstuhl, S., Landais, A., Leuenberger, M., Lorrain, R., Masson-Delmotte, V., Miller, H., Motoyama, H., Narita, H., Popp, T., Rasmussen, S. O., Raynaud, D., Rothlisberger, R., Ruth, U., Samyn, D., Schwander, J., Shoji, H., Siggard-Andersen, M.-L., Steffensen, J. P., Stocker, T., Sveinbjörnsdóttir, a. E., Svensson, A., Takata, M., Tison, J.-L., Thorsteinsson, T., Watanabe, O., Wilhelms, F., and White, J. W. C.: High-resolution record of Northern Hemisphere climate extending into the last interglacial period, Nature, 431, 147–151, https://doi.org/10.1038/nature02805, 2004.

Andersen, K. K., Svensson, A., Johnsen, S. J., Rasmussen, S. O., Bigler, M., Rüthlisberger, R., Ruth, U., Siggaard-Andersen, M. L., Peder Steffensen, J., Dahl-Jensen, D., Vinther, B. M., and Clausen, H. B.: The Greenland Ice Core Chronology 2005, 15–42 ka. Part 1: Constructing the time scale, Quaternary Sci. Rev., 25, 3246–3257, https://doi.org/10.1016/j.quascirev.2006.08.002, 2006.

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., and Dewitt, D. G.: Skill of real-time seasonal ENSO model predictions during 2002–11: is our capability increasing?, B. Am. Meteorol. Soc., 93, 631–651, https://doi.org/10.1175/BAMS-D-11-00111.1, 2012.

Berger, W. H.: The 100-kyr ice-age cycle: Internal oscillation or inclinational forcing?, Int. J. Earth Sci., 88, 305–316, https://doi.org/10.1007/s005310050266, 1999.

Bhattacharya, K., Ghil, M., and Vulis, I. L.: Internal variability of an energy-balance model with delayed albedo effects, J. Atmos. Sci., 39, 1747–1773, https://doi.org/10.1175/1520-0469(1982)039<1747:IVOAEB>2.0.CO;2, 1982.

Burnham, K. and Anderson, D.: Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, vol. 172, Springer, New York, 2 edn., https://doi.org/10.1016/j.ecolmodel.2003.11.004, 2002.

Chekroun, M. D., Kondrashov, D., and Ghil, M.: Predicting stochastic systems by noise sampling, and application to the El Niño-Southern Oscillation, P. Natl. Acad. Sci. USA, 108, 11766–11771, https://doi.org/10.1073/pnas.1015753108, 2011.

Chekroun, M. D., Ghil, M., Liu, H., and Wang, S.: Low-dimensional Galerkin approximations of nonlinear delay differential equations, Discrete Cont. Dyn.-A, 36, 4133–4177, https://doi.org/10.3934/dcds.2016.36.4133, 2016.

Chorin, A. J. and Hald, O. H.: Stochastic Tools in Mathematics and Science, Vol. 58 of Texts in Applied Mathematics, Springer, New York, third edn., 2013.

Chorin, A. J. and Lu, F.: Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics., P. Natl. Acad. Sci. USA, 112, 9804–9809, https://doi.org/10.1073/pnas.1512080112, 2015.

Chorin, A. J., Hald, O. H., and Kupferman, R.: Optimal prediction with memory, Physica D, 166, 239–257, 2002.

Crucifix, M. and Rougier, J.: On the use of simple dynamical systems for climate predictions, Eur. Phys. J.-Spec. Top., 174, 11–31, 2009.

Dansgaard, W., Johnsen, S. J., Clausen, H. B., Dahl-Jensen, D., Gundestrup, N. S., Hammer, C. U., Hvidberg, C. S., Steffensen, J. P., Sveinbjörnsdóttir, Á. E., Jouzel, J., and Bond, G. C.: Evidence for general instability of past climate from a 250-kyr ice-core record, Nature, 364, 218–220, https://doi.org/10.1038/364218a0, 1993.

De Saedeleer, B., Crucifix, M., and Wieczorek, S.: Is the astronomical forcing a reliable and unique pacemaker for climate? A conceptual model study, Clim. Dynam., 40, 273–294, 2013.

Ditlevsen, P. D.: Observation of alpha stable noise induced millenial climate changes from an ice-core record, Geophys. Res. Lett., 26, 1441–1444, 1999.

Ditlevsen, P. D. and Ditlevsen, O. D.: On the stochastic nature of the rapid climate shifts during the last ice age, J. Climate, 22, 446–457, https://doi.org/10.1175/2008JCLI2430.1, 2009.

Ditlevsen, P. D., Kristensen, M. S., and Andersen, K. K.: The recurrence time of Dansgaard–Oeschger events and limits on the possible periodic component, J. Climate, 18, 2594–2603, https://doi.org/10.1175/JCLI3437.1, 2005.

Ditlevsen, P. D., Andersen, K. K., and Svensson, A.: The DO-climate events are probably noise induced: statistical investiga-

tion of the claimed 1470 years cycle, Clim. Past, 3, 129–134, https://doi.org/10.5194/cp-3-129-2007, 2007.

Fischer, H., Siggaard-Andersen, M., Ruth, U., Röthlisberger, R., and Wolff, E. W.: Glacial/interglacial changes in mineral dust and sea-salt records in polar ice cores: sources, transport, and deposition, Rev. Geophys., 45, 1–26, https://doi.org/10.1029/2005RG000192, 2007.

Ghil, M.: Cryothermodynamics: the chaotic dynamics of paleoclimate, Physica D, 77, 130–159, https://doi.org/10.1016/0167-2789(94)90131-7, 1994.

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P.: Advanced spectral methods for climatic time series, Rev. Geophys., 40, 1003, https://doi.org/10.1029/2000RG000092, 2002.

Ghil, M., Chekroun, M. D., and Stepan, G.: A collection on "Climate Dynamics: multiple scales and memory effects", Proc. R. Soc. A, 471, 20150097, https://doi.org/10.1098/rspa.2015.0097, 2015.

Gkinis, V., Simonsen, S. B., Buchardt, S. L., White, J. W. C., and Vinther, B. M.: Water isotope diffusion rates from the NorthGRIP ice core for the last 16,000 years – Glaciological and paleoclimatic implications, Earth Planet. Sc. Lett., 405, 132–141, https://doi.org/10.1016/j.epsl.2014.08.022, 2014.

Hasselmann, K.: Stochastic climate models Part I. Theory, Tellus, 28, 473–485, https://doi.org/10.1111/j.2153-3490.1976.tb00696.x, 1976.

Johnsen, S., Clausen, H., Dansgaard, W., Fuhrer, K., Gundestrup, N., Hammer, C., Iversen, P., Jouzel, J., Stauffer, B., and Steffensen, J.: Irregular glacial interstadials recorded in a new Greenland ice core, Nature, 359, 311–313, https://doi.org/10.1038/359311a0, 1992.

Johnsen, S. J., Clausen, H. B., Dansgaard, W., Gundestrup, N. S., Hammer, C. U., Andersen, U., Andersen, K. K., Hvidberg, C. S., Dahl-Jensen, D., Steffensen, J. P., Shoji, H., Sveinbjornsdottir, A. E., White, J., Jouzel, J., and Fisher, D.: The d18O record along the Greenland Ice Core Project deep ice core and the problem of possible Eemian climatic instability, J. Geophys. Res.-Oceans, 102, 26397–26410, https://doi.org/10.1029/97JC00167, 1997.

Johnsen, S. J., Dahl-Jensen, D., Gundestrup, N., Steffensen, J. P., Clausen, H. B., Miller, H., Masson-Delmotte, V., Sveinbjörnsdottir, A. E., and White, J.: Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP, J. Quaternary Sci., 16, 299–307, https://doi.org/10.1002/jqs.622, 2001a.

Johnsen, S. J., Dahl-Jensen, D., Gundestrup, N., Steffensen, J. P., Clausen, H. B., Miller, H., Masson-Delmotte, V., Sveinbjornsdottir, A. E., and White, J.: Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP, J. Quaternary Sci., 16, 299–307, https://doi.org/10.1002/jqs.622, 2001b.

Källén, E., Crafoord, C., Ghil, M., and Kaumllleacuten, E.: Free oscillations in a climate model with ice-sheet dynamics, J. Atmos. Sci., 36, 2292–2303, https://doi.org/10.1175/1520-0469(1979)036<2292:foiacm>2.0.co;2, 1979.

Kondrashov, D., Kravtsov, S., Robertson, A. W., and Ghil, M.: A hierarchy of data-based ENSO models, J. Climate, 18, 4425–4444, https://doi.org/10.1175/JCLI3567.1, 2005.

Kondrashov, D., Kravtsov, S., and Ghil, M.: Empirical mode reduction in a model of extratropical low-frequency variability, J. Atmos. Sci., 63, 1859–1877, https://doi.org/10.1175/JAS3719.1, 2006.

Kondrashov, D., Chekroun, M. D., Robertson, A. W., and Ghil, M.: Low-order stochastic model and "past-noise forecasting" of the Madden–Julian Oscillation, Geophys. Res. Lett., 40, 5305–5310, https://doi.org/10.1002/grl.50991, 2013.

Kondrashov, D., Chekroun, M. D., and Ghil, M.: Data-driven non-Markovian closure models, Physica D, 297, 33–55, https://doi.org/10.1016/j.physd.2014.12.005, 2015.

Kravtsov, S., Kondrashov, D., and Ghil, M.: Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability, J. Climate, 18, 4404–4424, https://doi.org/10.1175/JCLI3544.1, 2005.

Kravtsov, S., Kondrashov, D., and Ghil, M.: Empirical model reduction and the modelling hierarchy in climate dynamics, and the geosciences, in: Stochastic Physics and Climate modeling, edited by Palmer, T., and Williams, P., Cambridge University Press, Cambridge, available at: http://www.atmos.ucla.edu/tcd/PREPRINTS/BookEMR_Text.pdf, 35–72, 2009.

Krumscheid, S., Pradas, M., Pavliotis, G. A., and Kalliadasis, S.: Data-driven coarse graining in action: Modelling and prediction of complex systems, Phys. Rev. E, 92, 042139, https://doi.org/10.1103/PhysRevE.92.042139, 2015.

Kwasniok, F.: Analysis and modelling of glacial climate transitions using simple dynamical systems, Philos. T. R. Soc. A, 371, 1–22, 2013.

Le Treut, H., and Ghil, M.: Orbital forcing, climatic interactions, and glaciation cycles, J. Geophys. Res.-Oceans, 88, 5167–5190, https://doi.org/10.1029/JC088iC09p05167, 1983.

Majda, A. J., Timofeyev, I., and Vanden Eijnden, E.: A mathematical framework for stochastic climate models, Commun. Pur. Appl. Math., 54, 891–974, 2001.

Mitsui, T. and Crucifix, M.: Influence of external forcings on abrupt millennial-scale climate changes: a statistical modelling study, Clim. Dynam., 48, 2729–2749, https://doi.org/10.1007/s00382-016-3235-z, 2017.

Mori, H.: A continued-fraction representation of the time-correlation functions, Prog. Theor. Phys., 34, 399–416, 1965.

Peavoy, D., Franzke, C. L. E., and Roberts, G. O.: Systematic physics constrained parameter estimation of stochastic differential equations, Comput. Stat. Data An., 83, 182–199, 2015.

Pelletier, J. D.: Coherence resonance and ice ages, J. Geophys. Res.-Atmos., 108, 1–14, https://doi.org/10.1029/2002JD003120, 2003.

Penland, C. and Sardeshmukh, P. D.: The optimal growth of tropical sea surface temperature anomalies, J. Climate, 8, 1999–2024, 1995.

Penland, C., Ghil, M., and Weickmann, K.: Adaptive filtering and maximum entropy spectra with application to changes in atmospheric angular momentum, J. Geophys. Res.-Atmos., 96, 22659–22671, https://doi.org/10.1029/91JD02107, 1991.

Rasmussen, S. O., Bigler, M., Blockley, S. P., Blunier, T., Buchardt, S. L., Clausen, H. B., Cvijanovic, I., Dahl-Jensen, D., Johnsen, S. J., Fischer, H., Gkinis, V., Guillevic, M., Hoek, W. Z., Lowe, J. J., Pedro, J. B., Popp, T., Seierstad, I. K., Steffensen, J. P., Svensson, A. M., Vallelonga, P., Vinther, B. M., Walker, M. J. C., Wheatley, J. J., and Winstrup, M.: A strati-

graphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: Refining and extending the INTIMATE event stratigraphy, Quaternary Sci. Rev., 106, 14–28, https://doi.org/10.1016/j.quascirev.2014.09.007, 2014.

Rial, J. A.: Abrupt climate change: chaos and order at orbital and millennial scales, Global Planet. Change, 41, 95–109, https://doi.org/10.1016/j.gloplacha.2003.10.004, 2004.

Ruth, U., Wagenbach, D., Bigler, M., Steffensen, J. P., Röthlisberger, R., and Miller, H.: High-resolution microparticle profiles at NorthGRIP, Greenland: case studies of the calcium-dust relationship, Ann. Glaciol., 35, 237–242, https://doi.org/10.3189/172756402781817347, 2002.

Ruth, U., Wagenbach, D., Steffensen, J. P., and Bigler, M.: Continuous record of microparticle concentration and size distribution in the central Greenland NGRIP ice core during the last glacial period, J. Geophys. Res.-Atmos., 108, 1–12, https://doi.org/10.1029/2002JD002376, 2003.

Rypdal, M.: Early-Warning Signals for the onsets of Greenland Interstadials and the Younger Dryas–Preboreal transition, J. Climate, 29, 4047–4056, https://doi.org/10.1175/JCLI-D-15-0828.1, 2016.

Saltzman, B. and Maasch, K. A.: A first-order global model of late Cenozoic climate, T. Roy. Soc. Edin.-Earth, 81, 315–325, 1990.

Saltzman, B. and Maasch, K. A.: A first-order global model of late Cenozoic climatic change I I. Further analysis based on a simplification of CO_2 dynamics, Clim. Dynam., 5, 201–210, 1991.

Srivastava, J. N.: A multivariate extension of the gauss-markov theorem, Ann. I. Stat. Math., 17, 63–66, https://doi.org/10.1007/BF02868153, 1965.

Steffensen, J. P., Andersen, K. K., Bigler, M., Clausen, H. B., Dahl-Jensen, D., Fischer, H., Goto-Azuma, K., Hansson, M., Johnsen, S. J., Jouzel, J., Masson-Delmotte, V., Popp, T., Rasmussen, S. O., Röthlisberger, R., Ruth, U., Stauffer, B., Siggaard-Andersen, M.-L., Sveinbjörnsdóttir, Á. E., Svensson, A., and White, J. W. C.: High-resolution Greenland ice core data show abrupt climate change happens in few years, Science, 321, 680–684, https://doi.org/10.1126/science.1157707, 2008.

Svensson, A., Andersen, K. K., Bigler, M., Clausen, H. B., Dahl-Jensen, D., Davies, S. M., Johnsen, S. J., Muscheler, R., Parrenin, F., Rasmussen, S. O., Röthlisberger, R., Seierstad, I., Steffensen, J. P., and Vinther, B. M.: A 60 000 year Greenland stratigraphic ice core chronology, Clim. Past, 4, 47–57, https://doi.org/10.5194/cp-4-47-2008, 2008.

Tziperman, E., Raymo, M. E., Huybers, P., and Wunsch, C.: Consequences of pacing the Pleistocene 100 kyr ice ages by nonlinear phase locking to Milankovitch forcing, Paleoceanography, 21, PA4206, https://doi.org/10.1029/2005PA001241, 2006.

Vautard, R., Yiou, P., and Ghil, M.: Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, Physica D, 58, 95–126, https://doi.org/10.1016/0167-2789(92)90103-T, 1992.

Yang, Y.: Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation, Biometrika, 92, 937–950, https://doi.org/10.1093/biomet/92.4.937, 2005.

Zwanzig, R.: On the identity of three generalized master equations, Physica, 30, 1109–1123, 1964.