Adv. Stat. Clim. Meteorol. Oceanogr., 1, 15–27, 2015
www.adv-stat-clim-meteorol-oceanogr.net/1/15/2015/
doi:10.5194/ascmo-1-15-2015

ASCMO
Open Access

Joint inference of misaligned irregular time series with application to Greenland ice core data

T. K. Doan[1], **J. Haslett**[1], **and A. C. Parnell**[2,3]

[1]School of Computer Science & Statistics, Trinity College Dublin, Dublin, Ireland
[2]School of Mathematical Sciences, University College Dublin, Dublin, Ireland
[3]UCD Earth Institute, University College of Dublin, Dublin, Ireland

Correspondence to: T. K. Doan (tdoan@tcd.ie)

Received: 22 September 2014 – Revised: 1 March 2015 – Accepted: 8 March 2015 – Published: 25 March 2015

Abstract. Ice cores provide insight into the past climate over many millennia. Due to ice compaction, the raw data for any single core are irregular in time. Multiple cores have different irregularities; and when considered together, they are misaligned in time. After processing, such data are made available to researchers as regular time series: a data product. Typically, these cores are independently processed. This paper considers a fast Bayesian method for the joint processing of multiple irregular series. This is shown to be more efficient than the independent alternative. Furthermore, our explicit framework permits a reliable modelling of the impact of the multiple sources of uncertainty. The methodology is illustrated with the analysis of a pair of ice cores. Our data products, in the form of posterior marginals or joint distributions on an arbitrary temporal grid, are finite Gaussian mixtures. We can also produce process histories to study non-linear functionals of interest. More generally, the concept of joint analysis via hierarchical Gaussian process models can be widely extended, as the models used can be viewed within the larger context of continuous space–time processes.

1 Introduction

Ice cores play an important role in revealing Earth's climate history via the analysis of their chemical composition. Data from ice cores are available in two forms: "raw" (typically irregular in time) and "data products" (typically regularly spaced in time, i.e. "gridded"). Often the latter are used as input for climate models to analyse past climate change. Alternatively, being gridded, researchers can simply combine one series with other data series, similarly gridded, representing other aspects of climate. Such data products are pre-processed from raw data using a variety of techniques: from simple running averages (Stuiver and Grootes, 2000; Thomas et al., 2007) to complex parametric smoothing (Peavoy and Franzke, 2010; Nieto-Barajas and Sinha, 2014). These imputation techniques, broadly interpolators or smoothers, are typically applied to one core at a time.

This paper proposes a joint statistical model for processing multiple raw ice core data sets. Like others in the wider field of palaeoclimate reconstruction (e.g. Haslett et al., 2006; Tingley et al., 2012), our procedure is based on hierarchical Gaussian processes approached through Bayesian inference. Our primary goal is to demonstrate the use of Bayesian inference to efficiently take into account multiple sources of information. A secondary goal is to encourage the use of Monte Carlo samples from joint Bayesian posterior distributions as a rich type of data product. Appropriately combining many sources of information should yield reduced uncertainties in cases where there is between-core correlation in the data. The combination should yield increased uncertainty in cases where information sources diverge. If such behaviour in the different sources is captured in the model structure, then we believe that combining them together will yield better estimates of uncertainty.

A pair of ice cores drilled in Greenland are used to illustrate our framework. As pointed out, for instance, by Stuiver and Grootes (2000), measurements from these cores contain negligible dating errors, particularly for the period of relatively stable climate approximately between 0 and 11 000 calendar years Before Present (k cal yr BP, where Present is 1950) as considered in this paper. We refer the interested

reader to the recent work of Parnell et al. (2015), who attempt to address temporal uncertainty in radiocarbon dated lake sediments. In this initial presentation of the methodology we do not provide treatment of this issue. Notwithstanding, a considerable challenge in an analysis of multiple ice core data sets is temporal misalignment, due to different irregularities across the cores.

Whilst our motivating example stems from climate research, multivariate data with different temporal irregularities are a common feature in many contemporary applications. For example, the ability to combine outputs at different levels of accuracy is crucial to the understanding of processes being studied through potentially expensive computer experimentation. A useful approach in such applications is to combine results from many cheap but low-resolution experiments with those from a few expensive but high-resolution experiments, by linking the data via different layers of modelling (Qian and Wu, 2008). In medical applications, one recommendation to overcome issues with misalignment is to align the times before further modelling (Cismondi et al., 2013). For more examples of misaligned time series and associated methods, see Cismondi et al. (2011) and Eckner (2012), and the references therein.

From a theoretical perspective, we view misaligned time series as a special case of spatial misalignment in spatial statistics (e.g. Gelfand et al., 2001; Wikle and Berliner, 2005). Our methodological proposal follows the "change of support problem" for data that are indexed in both space and time. This involves studying the statistical properties of a stochastic process at "supports" that are different to those associated with data. In this context, the location of an ice core is defined in space, and "support" refers to a duration of time over which a measurement is recorded. Technically, the volume of a section of an ice core provides support for a data point. However, as the thickness in all sections is identical, comparison of supports is simpler via physical lengths (e.g. 55 cm) rather than volumes. A length of a core section directly maps to a varying period of elapsed time. Hereafter, we use "support" to refer to the period of time which is represented by a single data point. In this paper, we focus on the case where the supports of observations are the same within one series but different across series, and inferences of underlying processes are at arbitrary time points, typically on a regular grid.

We introduce and discuss the raw data and their associated uncertainties in Sect. 2. In Sect. 3 we describe our hierarchical stochastic process model, and discuss an efficient procedure for inference of underlying latent processes. We report and discuss some results in Sect. 4. One of the advantages of our approach is that researchers can use simulations from a joint posterior distribution to study non-linear functionals of partially observed processes. As typically published, classical interpolants and smoothers do not facilitate such studies, for they do not typically make available the joint uncertainty associated with the product, a point made pithily by McShane and Wyner (2011) in their critical study of surface temperature reconstructions of the last millennium. Details of the so-called "8.2 ka event" (e.g. Thomas et al., 2007), an abrupt climate change period, provide specific examples. For instance, what precision do data permit in the timing of this event? We demonstrate that imputation samples from posterior distributions of our model are useful for answering this question. Finally, the paper ends with Sect. 5, where we conclude and offer our perspectives on possible directions of further work. The data, code and analysis are available online (http://www.scss.tcd.ie/~tdoan/research.html).

2 Greenland ice core data

Palaeoclimate archives such as tree rings, laminated lake sediments and ice cores are often used as a guide to past climatic conditions (e.g. Li et al., 2010; Tingley et al., 2012). This paper focuses on the stable isotope ratios of oxygen (expressed as δ^{18}O) in ice cores that are linked to past temperature. The process by which an ice core represents temperature is based on the ease with which two particular types of isotopes in water evaporate from the ocean and condense as snow. We refer interested readers to Jouzel et al. (1997) for a more detailed account of this connection.

We obtain data from the National Climatic Data Center (http://www1.ncdc.noaa.gov/pub/data/paleo/icecore/greenland/). In particular, we use the data sets from the United States' Greenland Ice Sheet Project II (Stuiver, 1999, GISP2) and Europe's Greenland Ice Core Project (Johnsen, 1999, GRIP). These are the results from drilling through the Greenland Ice Sheet to recover ice records over 3 km deep. Whilst these data sets are presented in several versions, we use a raw version in which the cores were cut into sections of equal lengths; 55 and 200 cm for GRIP and GISP2 respectively. The δ^{18}O measurements are recorded as a function of depth. For the Holocene period considered in this work, depth is transformed to age by counting annual layers in chemicals that show a seasonal cycle (Rasmussen et al., 2006). As previously mentioned, this dating method produces minor errors, and we do not consider temporal uncertainty in this paper.

Even though sections within a core are regular in length, their corresponding ages are irregular. Thus, jointly, multiple ice core time series with different irregularities are temporally misaligned. The δ^{18}O and date records for all consecutive sections are presented in Fig. 1, along with boxplots of the age increments[1].

Since the δ^{18}O measurements have continuous support and are irregular in time, we use the empirical semivariograms to investigate their temporal variability. This involves calculation of half of the squared differences of all com-

[1] An age difference value of 80.6 yr between roughly 1320 and 1400 cal yr BP has been omitted in this figure to focus on other significant features of this plot.

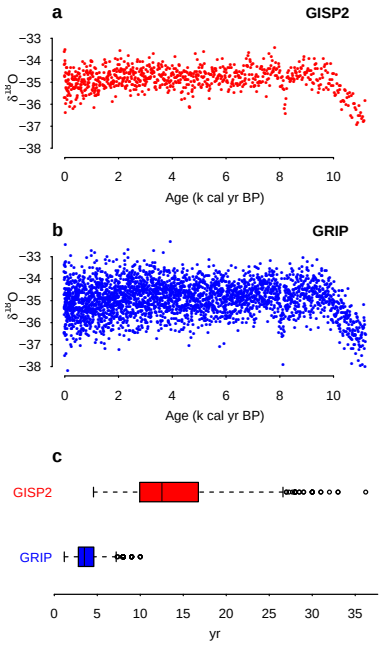

Figure 1. Scatter plots of δ^{18}O measurements and ages of **(a)** GISP2 and **(b)** GRIP. **(c)** Boxplots of the time increments clearly show different irregularities in the ages. Note that we have omitted an age difference value of 80.6 yr between roughly 1.32 and 1.4 k cal yr BP in the boxplot for GISP in order to focus on other significant features. The interquartile ranges (calculation including the omitted value) are (10.0, 12.5, 16.8) and (2.8, 3.5, 4.6) yr for GISP2 and GRIP respectively.

binations of δ^{18}O values from each core. As with all processes defined on continuous support, there is necessarily a minimum separation in the data, beyond which we have no information on which to infer the semivariogram. The usual approach is to assume one of the standard models and use this to infer the information near the origin corresponding to high-frequency behaviour (Chiles and Delfiner, 2012, Chap. 2). Figure 2 suggests that the variability in each series may be adequately modelled via a theoretical linear semivariogram, i.e. $\frac{1}{2}$ Var (first difference of the δ^{18}O values) = intercept + slope $\times$ (time lag). The intercept is known as the nugget effect[2], and the slope is twice the variance of the unit increments of a nugget-free, continuous-time stochastic process. Three key features can be seen from the empirical semivariograms. First, they are dominated by their respective nugget term. Second, the ratio of the nuggets is approximately $\frac{55}{200}$, which is the same as the ratio of the lengths of their respective ice core sections. Third, and most importantly, the slopes are approximately equal. We discuss further aspects of this in Appendix A.

[2]The term "nugget" refers to the apparent discontinuity at the beginning of a semivariogram. It is attributed to two sources of variation: the noise of data at high temporal frequency, and that which is due to uncertainty from data collection (Chiles and Delfiner, 2012, Chap. 2).

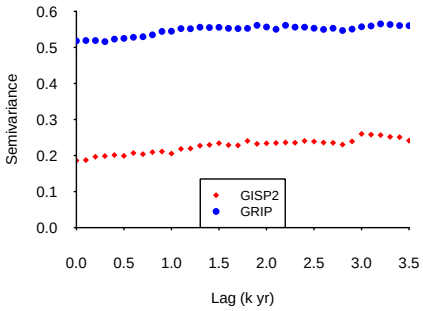

Figure 2. Empirical semivariograms of GISP2 and GRIP. They suggest that the linear semivariogram is a suitable model for both of the ice core data sets, i.e. $\frac{1}{2}$ Var (first difference in the δ^{18}O values) = intercept + slope $\times$ (first difference in ages). See the text for further details.

Our final exploratory analysis focuses on the standardised distribution of the first differences of the δ^{18}O values. To begin with, we use R package geoR (Ribeiro Jr. and Diggle, 2001) to estimate the parameters of the theoretical linear semivariogram models. Based on these estimations, we compute the standardised distribution of the δ^{18}O increments. The resulting QQ plots, as shown in Fig. 3, indicate the suitability of the Gaussian assumption. This, as well as the finding obtained via the empirical semivariograms, provides the basis for the continuous-time stochastic model proposed in the following section.

3 Joint modelling and Bayesian inference

In this section we outline our notation and describe our model for multiple misaligned irregular time series data. We show how to perform fast Bayesian inference on the parameters of this model without resorting to Markov chain Monte Carlo methods. Subsequently, we describe two efficient algorithms for imputation of a latent process of interest onto a time grid. One algorithm is simulation-free computation of posterior marginals and the other involves simple Monte Carlo simulation from posterior joint distributions.

3.1 A three-layer hierarchical statistical model

We consider a hierarchical model comprised of the data, process and parameter layer. At the data layer, $z(t_{c,i})$ are unobserved δ^{18}O values in core $c = 1, \ldots m$ at time $t_i; i = 1 \ldots n_c$. When they are observed, the instrumentation is such that

$$y(t_{c,i}) = z(t_{c,i}) + \nu(t_{c,i}). \quad (1)$$

Here, the terms $y(t_{c,i})$ denote the observations. There are $N = \sum_c^m n_c$ observations in total; we use $\boldsymbol{y}$ to denote these, and $\boldsymbol{y}_c$ for observations in core c only. The terms $\nu(t_{c,i})$ are

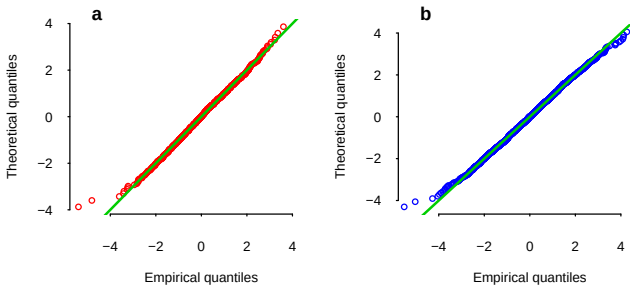

Figure 3. QQ plots of the standardised increments, i.e. the ratio of the first differences in the δ^{18}O measurements and estimated standard errors of increments for **(a)** GISP2 and **(b)** GRIP. The unusual values correspond to very large differences in consecutive pairs of δ^{18}O values.

iid, zero mean, Gaussian random variables with fixed known variance $\sigma^2_{v_c}$ corresponding to the instrumentation. For simplification, hereafter we drop the i subscript when discussing latent processes since they are defined for all times.

At a process layer, we express $z(t_c)$ as a function of latent value $x(t_c)$ via an additive Gaussian model:

$$z(t_c) = x(t_c) + w(t_c), \tag{2}$$

where $w(t_c)$ reflects micro-scale, annual-level variations and is modelled by a white noise process, independently across cores, having common $\mathrm{Var}[w(t_c)] = \sigma^2_{w_c}$. The process $x(t_c)$ may be thought of as driven by climate and thus slowly changing. We propose to model it as a continuous-time, independent increments process with increment variance such that, for each core, $\mathrm{Var}[x(t_c) - x(t_c - h)] = v^2|h|$, i.e. having a linear semivariogram with no nugget effect. Consequently we have

$$y(t_c) = x(t_c) + w(t_c) + v(t_c) := x(t_c) + \epsilon(t_c). \tag{3}$$

Here $\epsilon(t_c) \overset{\text{ind}}{\sim} \mathcal{N}(0, \sigma^2_{\epsilon_c})$, with $\sigma^2_{\epsilon_c} = \sigma^2_{w_c} + \sigma^2_{v_c}$, and the core-specific variance component $\sigma^2_{\epsilon_c}$ is the annual-level nugget effect at core c.

In Sect. 2, we argue that the linear semivariogram is a suitable model for both ice core data sets. More formally, each is consistent with a generating process which is the sum of (i) an underlying independent increments continuous-time process with semi-variance proportional to lag; and (ii) a white noise process, manifest in the intercept or nugget effect which is dominated by σ^2_ϵ. For our application, the measurement processes differ for each core. In particular, both involve averaging a section of core. The chief implication of this is that the nugget effects differ in proportion to the time durations associated with the sections. We can reparameterise so that there is only one nugget term σ^2_ϵ for one of the cores, with the other having a different nugget which is a multiple of this value, i.e., $\sigma^2_{\epsilon_c} = k_c\sigma^2_\epsilon$ for a positive known core-specific value k_c based on the ratio of the measurement periods. For GRIP and GISP2, we set this constant to be the ratio of average supports of 0.275. We outline this choice in more detail in Appendix A.

The x increments at different cores will be highly correlated since the cores themselves reflect climate at nearby locations. In vector form, $\boldsymbol{x}$ is the multivariate latent process for all cores at all times, and $\boldsymbol{x}(t)$ represents all the cores at time t; we write

$$\boldsymbol{x}(t+h) - \boldsymbol{x}(t) \;\sim\; \mathcal{N}(\boldsymbol{0}, v^2\,|h|\,\boldsymbol{\Sigma}), \tag{4}$$

where v^2 is the common variance of an increment per unit time, and $\boldsymbol{\Sigma}$ is an m-by-m matrix which controls the strength of the relationship of data across the cores. The multivariate process in Eq. (4) forms the basic model underlying our joint approach.

To complete the hierarchical modelling structure, prior distributions are assigned to the model parameters. We use reference priors on v^2, σ^2_ϵ so that $\pi(v^2) \propto v^{-2}$ and $\pi(\sigma^2_\epsilon) \propto \sigma_\epsilon^{-2}$. When there are two cores, we model $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \tag{5}$$

where ρ is the correlation coefficient; see also Appendix B for more details on this choice. Since temperature processes at nearby locations always have (strongly) positively correlated increments, we set 0.5 and 1 as the acceptable range for ρ. Furthermore, we assign a flat prior on its logit transformation so that it has values on the real axis. Hereafter the unknown parameters in our model are written as $\boldsymbol{\theta} = \{v^2, \rho, \sigma^2_\epsilon\}$.

3.2 Imputation at a regular time grid

Our objective, given $\boldsymbol{y}$, is to provide imputed values of δ^{18}O for all cores on a regular time grid denoted by $\boldsymbol{t}_\mathrm{g} = \{i\Delta; i = 1, \ldots, n_\mathrm{g}\}$. These values will be proposed below as expected values of suitable random variables, conditional jointly on $\boldsymbol{y}$. Our contribution in this paper is to focus on joint conditioning, in contrast to conditioning on data $\boldsymbol{y}_c$ in a single core, and on the issues of uncertainty that arise when the various parameters above are themselves only available through statistical inference.

There are two choices of random variables for imputation. We could focus on $E[z(t_c)|\boldsymbol{y}]$ or $E[x(t_c)|\boldsymbol{y}]$ for each core c and time $t \in \boldsymbol{t}_\mathrm{g}$. Both are legitimate targets of interest, and their computations are equally straightforward. In this paper we choose the latter; see Eq. (10). Given the parameters appropriate to our data, and in particular that $\sigma^2_{w_c} \gg \sigma^2_{v_c}$, there will be little difference between them, except for cases where a time grid effectively coincides with a time of an observation, and for the corresponding core only. The associated variances, $\mathrm{Var}[z(t_c)|\boldsymbol{y}]$ and $\mathrm{Var}[x(t_c)|\boldsymbol{y}]$, will however differ; conditional on the parameters, this difference will be approximately $\sigma^2_{w_c}$.

3.3 Bayesian inference and computation

Our later mathematical derivations are substantially simplified by defining

1. $\boldsymbol{t}_{\mathrm{o}}$ as the sorted set of all observed times at all cores, i.e. $t_{\mathrm{o}} = \mathrm{sort}\{t_{c,1}, t_{c,2}, \ldots\} = \{t_1, t_2, \ldots, t_n\}$, where n is the total number of unique time points of observations across all cores.

2. $y_{\mathrm{o}}(t_{c,i}); c = 1, \ldots, m$ whose values coincide with the $y(t_{c,i})$ values when core c has an observation at t_i and whose values are missing otherwise; there is typically one such core for each t_i. Let $\boldsymbol{y}_{\mathrm{o}}$ denote the set of all such vectors.

This is simply a technical "notational trick"; the vectors $\boldsymbol{y}$ (length N) and $\boldsymbol{y}_{\mathrm{o}}$ (length mn) contain the same information. The objective is to use this information, together with a hierarchical model of the continuous-time multivariate stochastic process $\boldsymbol{x}$, to impute its values on $\boldsymbol{t}_{\mathrm{g}}$. We shall refer to $\boldsymbol{y}$ and $\boldsymbol{x}$, defined on $\boldsymbol{t}_{\mathrm{g}}$ as $\boldsymbol{y}_{\mathrm{g}}$ and $\boldsymbol{x}_{\mathrm{g}}$. Moreover, we let $\boldsymbol{x}_{\mathrm{o}}$ be the latent multivariate process $\boldsymbol{x}$ defined on $\boldsymbol{t}_{\mathrm{o}}$.

3.3.1 Posterior distributions

In our new notation we can rewrite some of the equations discussed in Sect. 3.1 in their compact matrix form. For instance, Eq. (3) can be written as $\boldsymbol{y}_{\mathrm{o}}|\boldsymbol{x}_{\mathrm{o}}, \sigma_\epsilon^2, \boldsymbol{k} \sim \mathcal{N}(\boldsymbol{x}_{\mathrm{o}}, \boldsymbol{Q}_{\epsilon_{\mathrm{o}}}^{-1})$. Here, $\boldsymbol{k}$ is a vector of constants associated with the change in the support of the measurements across different cores, i.e. $\boldsymbol{k} = \{k_1, \ldots, k_m\}$. Furthermore, $\boldsymbol{Q}_{\epsilon_{\mathrm{o}}}$ is a diagonal precision matrix with entries zeros where there are no data, and $(k_c\sigma_\epsilon^2)^{-1}$ corresponding to a core c at times with data. This is the notation trick we previously introduced to allow us to write all vectors and matrices in clear order of time and ice core identity. Note that no computation is necessary for components that do not represent data.

Similarly, Eq. (4) has the form $\boldsymbol{x}_{\mathrm{o}}|v^2, \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_{x_{\mathrm{o}}}^{-1})$. Here, $\boldsymbol{Q}_{x_{\mathrm{o}}} = \boldsymbol{Q}_{t_{\mathrm{o}}} \otimes v^{-2}\boldsymbol{\Sigma}^{-1}$ is an inverse of the separable space-time covariance matrix (Cressie and Wikle, 2011, Chap. 6.1). More specifically, it is a Kronecker product of $\boldsymbol{Q}_{t_{\mathrm{o}}}$, the precision matrix of an independent increment process for irregular time series as discussed in Rue and Held (2005, Sect. 3.3), and the cross-covariance matrix $v^2\boldsymbol{\Sigma}$.

Let π denote a probability distribution function; the full posterior of our model is

$$
\begin{aligned}
\pi(\boldsymbol{x}_{\mathrm{o}}, \boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}}, \boldsymbol{k}) &\propto \pi(\boldsymbol{y}_{\mathrm{o}}|\boldsymbol{x}_{\mathrm{o}}, \boldsymbol{k}, \sigma_\epsilon^2)\pi(\boldsymbol{x}_{\mathrm{o}}|v^2, \boldsymbol{\Sigma})\pi(\boldsymbol{\theta}) \qquad (6)\\
&\propto |\boldsymbol{Q}_{\epsilon_{\mathrm{o}}}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{y}_{\mathrm{o}} - \boldsymbol{x}_{\mathrm{o}})^T \boldsymbol{Q}_{\epsilon_{\mathrm{o}}}(\boldsymbol{y}_{\mathrm{o}} - \boldsymbol{x}_{\mathrm{o}})\right)\\
&\quad |\boldsymbol{Q}_{x_{\mathrm{o}}}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{x}_{\mathrm{o}}{}^T \boldsymbol{Q}_{x_{\mathrm{o}}}\boldsymbol{x}_{\mathrm{o}}\right)\pi(\boldsymbol{\theta})\\
&\propto \mathcal{N}\left(\boldsymbol{x}_{\mathrm{o}}; \boldsymbol{\mu}_{x_{\mathrm{o}}|y_{\mathrm{o}}}, \boldsymbol{Q}_{x_{\mathrm{o}}|y_{\mathrm{o}}}^{-1}\right)|\boldsymbol{Q}_{x_{\mathrm{o}}|y_{\mathrm{o}}}|^{-\frac{1}{2}}|\boldsymbol{Q}_{\epsilon_{\mathrm{o}}}|^{\frac{1}{2}}\\
&\quad |\boldsymbol{Q}_{x_{\mathrm{o}}}|^{\frac{1}{2}} \exp\left(\frac{1}{2}\boldsymbol{y}_{\mathrm{o}}{}^T \boldsymbol{Q}_{\epsilon_{\mathrm{o}}}(\boldsymbol{\mu}_{x_{\mathrm{o}}|y_{\mathrm{o}}} - \boldsymbol{y}_{\mathrm{o}})\right)\pi(\boldsymbol{\theta}).
\end{aligned}
$$

Here $|\boldsymbol{Q}_{x_{\mathrm{o}}}| = |v^2\boldsymbol{\Sigma}|^{-(n-1)}$, $\boldsymbol{Q}_{x_{\mathrm{o}}|y_{\mathrm{o}}} = \boldsymbol{Q}_{\epsilon_{\mathrm{o}}} + \boldsymbol{Q}_{x_{\mathrm{o}}}$, and $\boldsymbol{\mu}_{x_{\mathrm{o}}|y_{\mathrm{o}}}$ is the solution to $\boldsymbol{Q}_{x_{\mathrm{o}}|y_{\mathrm{o}}}\boldsymbol{\mu}_{x_{\mathrm{o}}|y_{\mathrm{o}}} = \boldsymbol{Q}_{\epsilon_{\mathrm{o}}}\boldsymbol{y}_{\mathrm{o}}$. Consequently, $\boldsymbol{x}_{\mathrm{o}}$ can be analytically integrated out of $\pi(\boldsymbol{x}_{\mathrm{o}}, \boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}})$, yielding $\pi(\boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}})$, before further computation. Hence, our inference procedure comprises two separate stages as follows.

3.3.2 Inference stage 1: simulation-free computation of model parameters

Initial inference is focused on $\pi(\boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}})$. We begin with an optimisation routine on the logarithm of $\pi(\boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}})$ to locate the mode $\hat{\boldsymbol{\theta}}$ and the Hessian matrix evaluated at $\hat{\boldsymbol{\theta}}$; the latter is asymptotically the precision matrix for $\hat{\boldsymbol{\theta}}$. This information will be used as a guide to explore the parameter space of $\pi(\boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}})$ via a search strategy proposed by Rue et al. (2009, Sect. 3.1). We approximate the continuous distribution for $\boldsymbol{\theta}$, namely $\pi(\boldsymbol{\theta}); \boldsymbol{\theta} \in \boldsymbol{\Theta}$ by the discrete distribution $\pi(\boldsymbol{\theta}_j); \boldsymbol{\theta}_j \in \boldsymbol{\Theta}_J; j = 1, \ldots, J$. Here, $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}_J$ are, respectively, continuous and discrete parameter spaces. Specifically, the latter will be used as a Riemann sum approximation to analytical integrations. For instance, the normalising constant can be discretely evaluated when there are few parameters, i.e. $\pi(\boldsymbol{y}_{\mathrm{o}}) = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \pi(\boldsymbol{y}_{\mathrm{o}}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \approx \sum_{\boldsymbol{\theta}_j\in\boldsymbol{\Theta}_J} \pi(\boldsymbol{y}_{\mathrm{o}}|\boldsymbol{\theta}_j)\,\pi(\boldsymbol{\theta}_j)\,\triangle\boldsymbol{\theta}_j$.

3.3.3 Inference stage 2: imputation

Our next goal is to derive the joint posterior distribution of $\boldsymbol{x}_{\mathrm{g}}$ given all observations. We repeat the aforementioned notational trick by letting the star notation denote the processes defined at both the unique (and sorted) observed times and grid, i.e. $\boldsymbol{y}_* = \left(\boldsymbol{y}_{\mathrm{o}}{}^T, \boldsymbol{y}_{\mathrm{g}}{}^T\right)^T$ and $\boldsymbol{x}_* = \left(\boldsymbol{x}_{\mathrm{o}}{}^T, \boldsymbol{x}_{\mathrm{g}}{}^T\right)^T$. The problem becomes that of evaluating

$$
\pi(\boldsymbol{x}_*|\boldsymbol{y}_*) = \int_{\boldsymbol{\Theta}} \pi(\boldsymbol{x}_*|\boldsymbol{y}_*, \boldsymbol{\theta})\,\pi(\boldsymbol{\theta}|\boldsymbol{y}_*)\,\mathrm{d}\boldsymbol{\theta}. \qquad (7)
$$

Derivation of the first quantity in the above integrand is, again, by completing the quadratic form as in Eq. (6). The second quantity is effectively $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, the joint posterior distribution of the model parameters previously derived in Sect. 3.3.2. Importantly, the discrete approximation of the

latter renders as summations the integrals that arise in Eq. (7), i.e.

$$\pi(\boldsymbol{x}_*|\boldsymbol{y}_*) \approx \sum_{\boldsymbol{\Theta}_J} \mathcal{N}\left(\boldsymbol{x}_*; \boldsymbol{\mu}_{x_*|y_*}(\boldsymbol{\theta}_j), \boldsymbol{Q}^{-1}_{x_*|y_*}(\boldsymbol{\theta}_j)\right) \pi(\boldsymbol{\theta}_j|\boldsymbol{y}_*) \triangle\boldsymbol{\theta}_j, \quad (8)$$

where $\boldsymbol{\mu}_{x_*|y_*}$ and $\boldsymbol{Q}_{x_*|y_*}$ are the posterior mean and precision of $(\boldsymbol{x}_*|\boldsymbol{y}_*, \boldsymbol{\theta})$. We write $\boldsymbol{\mu}_{x_*|y_*}(\boldsymbol{\theta}_j)$ and $\boldsymbol{Q}_{x_*|y_*}(\boldsymbol{\theta}_j)$ to emphasise that they are the functions of $\boldsymbol{\theta}_j$. Thus, $\pi(\boldsymbol{x}_*|\boldsymbol{y}_*)$ is a Gaussian mixture over the posterior samples $\boldsymbol{\Theta}_j$ with weights $\alpha_j = \pi(\boldsymbol{\theta}_j|\boldsymbol{y})\triangle\boldsymbol{\theta}_j$ already computed in the first inference stage. Moreover, since the posterior *joint marginal* distribution is a finite mixture of Gaussian distributions, a posterior *marginal* of the lth element at a specific core corresponding to a time point is a finite mixture of Gaussian distributions. Therefore, it follows that

$$\pi(\boldsymbol{x}^{(l)}_*|\boldsymbol{y}_*) \approx \sum_{\boldsymbol{\Theta}_J} \mathcal{N}\left(\boldsymbol{x}^{(l)}_*; \mu^{(l)}_{x_*|y_*}(\boldsymbol{\theta}_j), \tau^{(l)}_{x_*|y_*}(\boldsymbol{\theta}_j)\right) \alpha_j, \quad (9)$$

where $\mu^{(l)}_{x_*|y_*}(\boldsymbol{\theta}_j)$, conditional on the sample value $\boldsymbol{\theta}_j$, is the lth element of the corresponding core identity from $\boldsymbol{\mu}_{x_*|y_*}$. Similarly, each conditional posterior variance $\tau^{(l)}_{x_*|y_*}(\boldsymbol{\theta}_j)$ is the lth element of the core identity represented in the diagonal of the covariance matrix. In fact, the latter can be computed efficiently from the precision matrix $\boldsymbol{Q}_{x_*|y_*}$ without having to perform matrix inversion (Rue and Martino, 2007, Sect. 2). We also emphasise that although there are many computations with large (sparse) matrices, such as solving equations, Cholesky decomposition, etc., we only have to do this J times. The fact that all of the matrices are sparse and can be efficiently stored and computed as band matrices presents a further huge reduction in the computation cost.

3.4 Comparison of data products obtained from posterior marginals and joint distribution

The joint and marginal predictive distributions in Eqs. (8) and (9) are the source of any imputations proposed in Sect. 3.2. For example, an imputed value corresponding to the distribution in Eq. (9) is

$$E(\boldsymbol{x}^{(l)}_*|\boldsymbol{y}_*) = \sum_{\boldsymbol{\Theta}_J} \mu^{(l)}_{x_*|y_*}(\boldsymbol{\theta}_j)\,\alpha_j. \quad (10)$$

This estimate is thus a linear combination of all the data available. It is optimal given the covariance structure and robust to uncertainties in that covariance structure by being Bayesian and including posterior uncertainty in the parameters. More generally, posterior means are a summary of posterior marginal distributions. Calculations of other pointwise summarised statistics such as median, modes, variance, quantiles, etc., are also straightforward. However, unaccompanied by appropriate measures of uncertainty, such summaries are of limited value as a basis for serious statistical research, for these are "*pointwise* confidence intervals and are not confidence curves for the entire *sample path*" (McShane and Wyner, 2011). That is to say, although the gridded values are jointly based on all the available observations, the joint conditional uncertainty of the unobserved process histories is not available from pointwise uncertainties. One simple consequence is that uncertainties for changes between different times are not available, for these minimally require covariances, even in a simple case in which posterior marginals are themselves Gaussian distributions. It is worth noting that our posterior marginals in Eq. (9) are generally non-Gaussian.

A greater challenge is posed from investigations of extreme events such as minima and their timing. These are examples of non-linear functionals of $\boldsymbol{x}_\mathrm{g}$. Probability distributions of such events are generally not available from posterior marginals at each time point separately (Li et al., 2007; Tingley and Huybers, 2013). Instead, simulations based on the joint posterior distribution (Eq. 8) at all arbitrary time points provide a statistically consistent and easily interpreted probabilistic statement of any functionals of $\boldsymbol{x}_\mathrm{g}$. Each such simulation represents an equally probable history of $\boldsymbol{x}_\mathrm{g}$ that is statistically consistent with the model and data. Moreover, pointwise summaries such as posterior means are simply an average of all the possible histories. We emphasise that, although a posterior mean on its own is not useful for answering the interesting questions discussed herein, it is of course more efficiently computed from its analytical expression as discussed in Eq. (10).

For brevity, we use Table 1 to illustrate an example of two δ^{18}O histories from one core on a grid of 5 k yr between times 0 and 10 k cal yr BP. The histories, each of length 3, represent independent realisations from a Gaussian mixture. Specifically, they are simulated from $\pi(\boldsymbol{x}_*|\boldsymbol{y}_*, \boldsymbol{\theta})$, with parameters $\boldsymbol{\theta}$ being a sample from $\pi(\boldsymbol{\theta}|\boldsymbol{y}_*)$; both distributions are components of the integral (Eq. 7) or its approximation (Eq. 8). The functionals of interest can then be directly computed for each process history and summarised. In this paper we focus on the distributions of random variables $x_\mathrm{min} = \min_t x(t)$ and $t_\mathrm{min} = \arg_t \min x(t)$, being respectively the minimum value and the time at which this minimum was achieved. We return to this in Sect. 4.2.

A more complete version of this approach uses 1000 histories. For generality we compute conditional quantiles at time grid $\boldsymbol{t}_\mathrm{g}$ and use pointwise credible intervals to quantify the uncertainty of our imputations.

4 Results

In this section we apply the model framework and inference procedures presented in Sect. 3 to analyse the Greenland ice core data sets described in Sect. 2. Section 4.1 presents the parameter estimates and associated data products. It compares the uncertainty levels between different data products to illustrate the advantage of joint modelling under the Bayesian framework. Section 4.2 uses a case study to address

Table 1. Illustration of process histories of length 3 on a time grid of 5 k yr intervals. For each sample, we simulate the parameters from a joint posterior distribution, followed by a realisation of the latent process $\boldsymbol{x}_g$. Two examples of non-linear functionals of $\boldsymbol{x}_g$ are displayed in the last two rows.

		Sample 1	Sample 2	...
Parameters		(0.2, 0.8, 0.4)	(0.3, 0.9, 0.5)	...
Time	0	−34.3	−34.7	...
	5	−35.0	−34.2	...
	10	−34.4	−34.4	...
Minimum		−35.0	−34.7	...
Time of minimum		5	0	...

some difficult questions that can only be answered using an ensemble of δ^{18}O histories.

4.1 Model-fitting results

Using the data from both cores jointly, we obtain the discrete approximation to the marginal posterior distributions of $\boldsymbol{\theta}$. For brevity of presentation, we apply a smoother to these and present their smoothed versions in Fig. 4. It can be seen that the results for v^2 and σ_ϵ^2 are consistent with the slope and intercept shown in Fig. 2. The posterior distribution of ρ peaks at a high value close to 1, which indicates a strong spatial relationship between the increments in two cores. This is not surprising, since the cores were drilled from nearby locations and both reflect the historical changes in the regional temperature of Greenland.

Under our model-based approach we have the choice of what summaries of the δ^{18}O histories to use. Our main interest, as discussed in Sect. 3.2, lies in the true δ^{18}O values on a bidecadal time grid over the period of 0 to 11 k cal yr BP. The 95 and 50 % credible intervals for the gridded process $\boldsymbol{x}_g$ are represented in Fig. 5. The latter interval is more commonly known as the interquartile range (IQR). We compare our results (labelled as DPH) with those from the algorithm-based approach of Stuiver and Grootes (2000, see also http://depts.washington.edu/qil/datasets/) and Mogensen (2001, see also http://www.gfy.ku.dk/~www-glac/data/gripdelta.dat), labelled as SG and M respectively. The construction of SG is "based on averaging the measurements of samples of shorter duration", and M has been similarly sampled with respect to SG's timescale. The set of posterior imputed values from DPH can be seen to be much smoother than both SG and M because the nugget is so much larger. For both SG and M, the nugget is assumed to be the measurement error, which is set very small. As we previously discussed in Sect. 3.1, our nugget comprises both the measurement error and micro-scale variation. Our choice removes the variation due to both sources of variation, resulting in a smoother latent process.

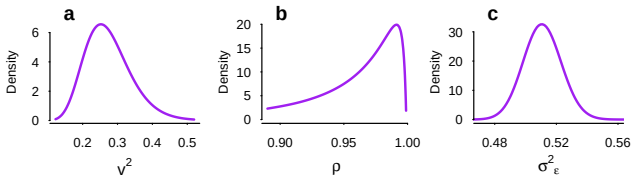

Figure 4. Plots of the smoothed posterior distributions of **(a)** the variance of the unit increment of latent process $\boldsymbol{x}$, **(b)** the cross-correlation coefficient and **(c)** the nugget effect for GRIP (the nugget effect for GISP is directly proportional to this). These are results based on the joint model discussed in Sect. 3.1 applied to GISP2 and GRIP.

To gain a better understanding of the benefit of joint modelling over separate alternatives, we fit an independent increments model with Gaussian noise to each core separately. Note that, in contrast to the joint approach, the relationship between two nugget parameters is suppressed in the separate approach. Thus, each model has two parameters (a process variance and a nugget parameter). We refer to Appendix B for a more formal discussion of these separate models. Figure 6 shows that the IQR in the separate model is always higher than that of the joint model. This reflects the difference in the temporal resolution, i.e. the number of available data points. Moreover, it indicates that the joint approach utilises information from both cores more effectively when the relationship between the cores (here measured by ρ) is strong. We anticipate that the benefit of joint inference would be more apparent as the number of correlated cores increases.

The "spikes" (e.g. 0, 1.36, 3.4, and 8.2 k cal BP in GISP2) in Fig. 6 are a direct implication of either the associated age gaps or abrupt changes in the δ^{18}O measurements, or both[3]. Specifically at 8.2 k cal yr BP, the spikes are more influenced by abrupt changes in the δ^{18}O measurements. In general, all of these spikes are higher and sharper in the separate models, in comparison to the joint model. The reverse happens around 1.4 k cal BP in GRIP. This is a direct consequence of the (lack of) data points from the other core (GISP2) in that period. Finally, we note that the IQR in all cores – in both the separate and joint settings – slightly increases with time. This occurs because temporal resolution decreases when the cores are sampled over sections of identical lengths.

Although this is not a full uncertainty comparison of our method with other methods – for neither standard deviations nor IQR are available – it suggests that these ignore valuable information by treating each core separately.

4.2 Case study: the 8.2 ka event

In this subsection, we discuss the 8.2 ka event, a sudden reduction in North Atlantic temperature during a period

[3]Note that the spikes at times 0 and 11 are the modelling artefact known as the "boundary effect"; see, for instance (Rue and Held, 2005).

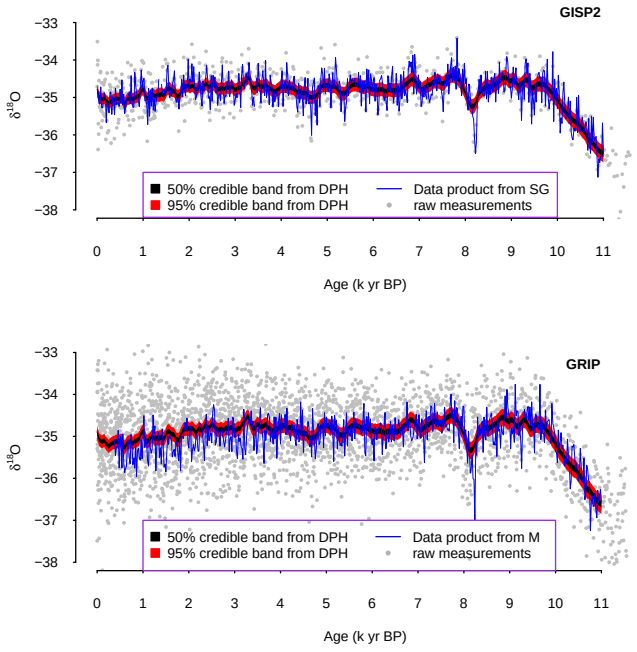

Figure 5. Plots of quantile-based 50 and 95 % credible intervals of the marginal posterior distributions of process $\boldsymbol{x}_g$ on a bidecadal time grid over the period of 0 to 11 k cal BP conditional on both GISP2 and GRIP. We also show the bidecadal data product reported in Stuiver and Grootes (2000, SG) and Mogensen (2001, M).

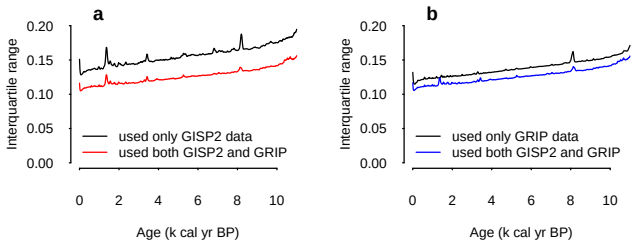

Figure 6. Plots of the interquartile ranges (IQR; corresponding to the width of the 50 % credible band in Fig. 5) of the marginal posterior distributions of $\boldsymbol{x}_g$ on a bidecadal time grid over the period of 0 to 11 k cal BP at cores **(a)** GISP2 and **(b)** GRIP. The main features from these plots are that (i) interquartile ranges of the separate models are always higher than those of the joint model; (ii) the gaps of the differences in **(a)** are consistently larger than those in **(b)**; (iii) there are several spikes; and (iv) a slight tendency for increased IQR further back in time. See the text for a detailed discussion.

around 8.2 k cal yr BP. This event is associated with a transient change in the North Atlantic overturning circulation. Consequently, the amount of evaporated water in the ocean that became ice in Greenland is amongst its best sources of evidence. The date corresponding to the local minimum of the averages is not a satisfactory estimator of the time of such event. We propose to use our data product, the δ^{18}O sampled histories from the joint posterior distribution of $\boldsymbol{x}_g$, to illustrate a satisfactory approach.

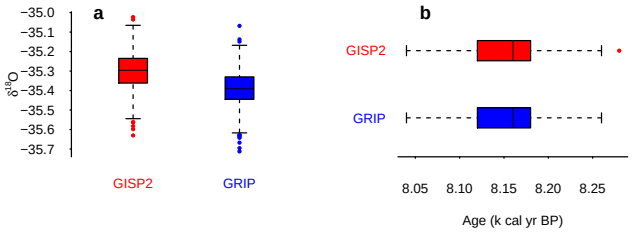

Figure 7. (a) Boxplots of the minima and **(b)** times of the minima from GISP2 and GRIP over the range of 7.9 to 8.5 k cal yr BP. The interquartile range of the timing of the event from both cores is (8.12, 8.16, 8.18) k cal yr BP. All estimates are based on 1000 process histories.

Like others in the literature, we define the 8.2 ka event by an attainment of a minimum in the temperature value during a specific time period (Thomas et al., 2007). In this study, we focus on the period between 7.90 and 8.50 k cal yr BP that is believed to bracket the 8.2 ka event, to distinguish it from the possible long-term climate trend (Morrill and Jacobsen, 2005). Using the procedure discussed in Sect. 3.4, we obtain a summary for 1000 minima of the δ^{18} histories as represented in Fig. 7. We estimate the interquartile ranges of the timing of the event from both cores to be (8.12, 8.16, 8.18) k cal yr BP. Our findings are consistent with previous studies reported elsewhere, but no quantification of the uncertainty has previously been attempted; see, for instance, Thomas et al. (2007) and Kobashi et al. (2007).

Monte Carlo samples from joint posterior distributions thus provide a flexible data product in its own right. Indeed, in climate reconstruction, this is precisely as proposed by Tingley et al. (2012). Our δ^{18}O histories can be utilised to investigate other interesting climate events (Wanner et al., 2011). The use above for minima and timings of minima herein is illustrative; any function of the process may be studied, conditional on the data.

5 Conclusions

We have presented a hierarchical model to jointly analyse multiple irregular time series. An important component of our model is the Gaussian Markov assumption based on multivariate independent increments that gives us a natural vehicle for joint modelling. We further derived and implemented a fast algorithm for parameter inference and imputation based on this model. We demonstrated that the joint approach utilises information from multiple time series more efficiently than one-series-at-a-time alternatives.

Our paper has been tailored to creating a climate data product from Greenland ice cores. Our proposed framework is simple but useful to combine and mix multiple ice core time series, allowing each series to have a different temporal support. To the best of our knowledge, this work is a first attempt at directly addressing the joint behaviour of multiple

ice cores in their raw and misaligned form. Like others in the literature, we are proposing the use of linear combinations of all the data available to perform imputation. Ours is optimal given the covariance structure and robust to uncertainties in that covariance structure by being Bayesian and including posterior uncertainty in the parameters. Our data products, in the form of non-Gaussian posterior predictive distributions, are richer than what have been previously possible. More importantly, our process histories can easily be utilised by other researchers to answer complex questions that are otherwise analytically intractable. This advantage was demonstrated using a case study of an abrupt climate change event.

We believe this to be an initial attempt at joint palaeoclimate inference, upon which other work can build. Some parameters in our model were formulated according to the respective lengths of ice core sections. This approach is likely to be problematic in the study of sections longer than the Holocene. A more realistic approach is to allow the value of k_c, as introduced in Sect. 3.1 and derived in Appendix A, to vary according to the empirical ratios of the supports of associated ice sections. We also envisage extension of our hierarchical model structure to incorporate uncertainty in the timescale, in the spirit of the approach recently discussed by Parnell et al. (2015). More generally, since this paper is a proof-of-concept building on the well-established ground of space–time modelling, our model can be immediately extended to a wide class of spatio-temporal process, for instance, by modifying Eq. (4).

Appendix A: Implication of different supports in the measurements

In this appendix we discuss our treatment for the process variance and nugget via the *change of support theory*. More specifically, we examine the change in the theoretical semivariogram when there is a change in the underlying support of the data. Without loss of generality, we suppose that the data process $\boldsymbol{y}$ can be modelled by a noisy univariate independent increments process with a theoretical linear semivariogram of the form $\hat{\sigma}^2 + \frac{1}{2}\hat{v}^2|h|$ where the nugget $\hat{\sigma}^2$ is the intercept, the process variance $\hat{v}^2$ is twice the slope value, and h is an arbitrary time lag. We create the new process $\widetilde{\boldsymbol{y}}$ on a new support by averaging $\boldsymbol{y}$ at every non-overlapping w "window" of time. We are interested in the relationship between the semivariogram of $\boldsymbol{y}$ and that of $\widetilde{\boldsymbol{y}}$.

If $\boldsymbol{y}$ is a pure nugget process, it is easy to show that the semivariogram of the averaged process is $\frac{\hat{\sigma}^2}{w}$ when $|h| \geq w$ and $\frac{\hat{\sigma}|h|}{w^2}$ when $|h| < w$. For the independent increments case, this is $\frac{1}{2}\hat{v}^2|h| - \frac{w\hat{v}^2}{6}$ when $|h| \geq w$ and $\frac{\hat{v}^2|h|^2}{2w} - \frac{\hat{v}^2|h|^3}{6w^2}$ when $|h| < w$. We refer to Chiles and Delfiner (2012, Chap. 2.4) for the technical details of the aforementioned results. In practice, the only lags available for semivariogram calculation are $|h| \geq w$. Thus, for a process that has both the nugget and independent increments, the implication is twofold. First, $\boldsymbol{y}$ and $\widetilde{\boldsymbol{y}}$ share the same process variance $\hat{v}^2$. Second, their nugget parameters are (approximately) directly proportional. This approximation affects the semivariogram near the origin, where we use the linear function to account for the true cubic. From the analytical expression it can be seen that the accuracy of the approximation depends on the value of w, and the relative difference between $\hat{\sigma}^2$ and $\hat{v}^2$. It is re-emphasised that for both GISP2 and GRIP, $\hat{\sigma}^2$ dominates $\hat{v}^2$, as seen in Fig. 2.

If we assume that the nugget effect is at an annual level (denoted as σ^2_{annual}), then the nuggets for GISP2 and GRIP are, respectively, $\sigma^2_{\text{GISP2}} = \frac{\sigma^2_{\text{annual}}}{w_{\text{GISP2}}}$ and $\sigma^2_{\text{GRIP}} = \frac{\sigma^2_{\text{annual}}}{w_{\text{GRIP}}}$, where factors w denote the time support for each series. We can reparameterise in terms of $\sigma^2_\epsilon = \sigma^2_{\text{GRIP}}$, such that $\sigma^2_{\text{GISP2}} = \sigma^2_\epsilon \times \frac{w_{\text{GRIP}}}{w_{\text{GISP2}}} = \sigma^2_\epsilon \times k_{\text{GISP2}}$. We set $\frac{55}{200}$ or 0.275 as the value for k_{GISP2}, corresponding to their respective length of support. This value is also consistent with the descriptive statistics of the age increments, as represented in Fig. 1c. Each section, being 200 and 55 cm in length respectively for GISP2 and GRIP, is negligible compared to the total length of the complete ice core, which is about 3 km (approximately 1.6 km of which covers the Holocene period). Thus we feel that this is a reasonable approximation.

Appendix B: Model choice for the cross-covariance

In this appendix, we discuss the model choice for $\boldsymbol{\Sigma}$ within the context of our application by comparing it with alternative models and choosing the best among them. Let M_1 denote the model with covariance (Eq. 5) for $\boldsymbol{\Sigma}$. This covariance structure assumes equal variance of increments across the cores.

To formally measure the benefit of joint modelling, we compare model M_1 with M_2, which has covariance that ignores cross-correlation between cores. In fact, M_2 has been informally introduced in Sect. 4. It comprises of two separate models, one for each core. We have demonstrated, via Fig. 6, that the joint model utilises information from both cores more effectively than the separate models.

We further propose M_3, which has a covariance that assumes varying variances for different cores as follows:

$$\boldsymbol{\Sigma}_a = \begin{pmatrix} 1 & \rho\sqrt{a} \\ \rho\sqrt{a} & a \end{pmatrix}. \tag{B1}$$

We assign a reference prior for a. Its marginal posterior distribution is shown in Fig. B1. Its mode centres around 1, while the marginal posteriors for other parameters in model M_3 (not shown here) are practically the same as those in model M_1 previously shown in Fig. 4. The results suggest that the common variance assumption in model M_1 is most suitable for GISP2 and GRIP. Model M_3 is essentially a more conservative version of M_1, but such an extension may not be necessary when we take into account the additional computational cost associated with an extra hyperparameter. Finally, Table B1 provides further evidence that the extra parameter can be made redundant, as the Bayesian information criterion (BIC) of Schwarz (1978) suggests that M_1 is superior to M_3.

Table B1. Values of the Bayesian information criterion (BIC) obtained from different model settings. The bold value highlights the model with the best fit.

Model	Covariance structure for two cores	$-2\log L$	Penalty	BIC
M_1	$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$	44 998	25	**45 023**
M_3	$\begin{pmatrix} 1 & \rho\sqrt{a} \\ \rho\sqrt{a} & a \end{pmatrix}$	44 998	33	45 031

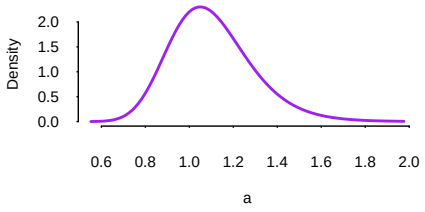

Figure B1. Plot of the smoothed marginal posterior distribution of the parameter a in the covariance matrix of model M_3 described in Appendix B. The mode is roughly 1, which suggests that a is redundant.

Appendix C: Model validation

As a final model checking step, we determine whether the parameters in our model are identifiable or not. We do this by simulating model parameters $(v^2, \rho, \sigma_\epsilon^2)$ based on the results of the data analysis of Greenland ice core, thence the latent bivariate process $\boldsymbol{x}$ and consequently artificial data $\boldsymbol{y}$. We partially average the sequences so as to match the change in support that occurs in our ice core example. We then fit the model as described in Sect. 3, and determine whether the 50 and 90 % posterior intervals contain the true values. We repeat these steps 1000 times, and count up the proportion of occurrences where the intervals contain the true values. A properly calibrated and identifiable 50 % interval should contain the true value 50 % of the time, and similarly with the 90 % interval.

The results of our simulations are shown in Table C1. As can be seen, the intervals contain slightly fewer than the desired proportion of true values, so our posterior intervals are over-precise. However, this effect appears small, and the model seems generally identifiable.

Table C1. Performance of the model fitting algorithm. All results were based on 1000 simulation runs.

Parameter	Proportion inside 50 % CI	Proportion inside 90 % CI
$\boldsymbol{x}$	51 %	89 %
v^2	48 %	88 %
ρ	49 %	90 %
σ_ϵ^2	48 %	90 %

Acknowledgements. This research was funded by Science Foundation Ireland (grant no. 10/RFP/MTH2779). The authors thank Eric W. Wolff for his helpful comments on the Greenland ice core data set.

Edited by: R. Donner
Reviewed by: three anonymous referees

References

Chiles, J.-P. and Delfiner, P.: Geostatistics: modeling spatial uncertainty, Vol. 497, John Wiley & Sons, 2012.

Cismondi, F., Fialho, A., Vieira, S., Sousa, J., Reti, S., Howell, M., and Finkelstein, S.: Computational intelligence methods for processing misaligned, unevenly sampled time series containing missing data, in: 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), 224–231, doi:10.1109/CIDM.2011.5949447, 2011.

Cismondi, F., Fialho, A. S., Vieira, S. M., Reti, S. R., Sousa, J., and Finkelstein, S. N.: Missing data in medical databases: Impute, delete or classify?, Artif. Intell. Med., 58, 63–72, 2013.

Cressie, N. and Wikle, C. K.: Statistics for spatio-temporal data, John Wiley & Sons, 2011.

Eckner, A.: A framework for the analysis of unevenly spaced time series data, Preprint, available at: http://eckner.com/papers/unevenly_spaced_time_series_analysis.pdf (last access: 20 March 2015), 2012.

Gelfand, A. E., Zhu, L., and Carlin, B. P.: On the change of support problem for spatio-temporal data, Biostatistics, 2, 31–45, 2001.

Haslett, J., Whiley, M., Bhattacharya, S., Salter-Townshend, M., Wilson, S. P., Allen, J. R. M., Huntley, B., and Mitchell, F. J. G.: Bayesian palaeoclimate reconstruction, J. Roy. Statist. Soc. A Stat., 169, 395–438, doi:10.1111/j.1467-985X.2006.00429.x, 2006.

Johnsen, S. J.: GRIP Oxygen Isotopes, Pangea, doi:10.1594/PANGAEA.55091, 1999.

Jouzel, J., Alley, R. B., Cuffey, K. M., Dansgaard, W., Grootes, P., Hoffmann, G., Johnsen, S. J., Koster, R. D., Peel, D., Shuman, C. A., Stievenard, M., Stuiver, M., and White, J.: Validity of the temperature reconstruction from water isotopes in ice cores, J. Geophys. Res.-Oceans, 102, 26471–26487, 1997.

Kobashi, T., Severinghaus, J. P., Brook, E. J., Barnola, J.-M., and Grachev, A. M.: Precise timing and characterization of abrupt climate change 8200 years ago from air trapped in polar ice, Quaternary Sci. Rev., 26, 1212–1222, 2007.

Li, B., Nychka, D. W., and Ammann, C. M.: The "hockey stick" and the 1990s: a statistical perspective on reconstructing hemispheric temperatures, Tellus A, 59, 591–598, 2007.

Li, B., Nychka, D. W., and Ammann, C. M.: The value of multiproxy reconstruction of past climate, J. Am. Stat. Assoc., 105, 883–895, doi:10.1198/jasa.2010.ap09379, 2010.

McShane, B. B. and Wyner, A. J.: A statistical analysis of multiple temperature proxies: are reconstructions of surface temperatures over the last 1000 years reliable?, Ann. Appl. Stat., 5, 5–44, doi:10.1214/10-AOAS398, 2011.

Mogensen, I.: A study of rapid climate changes, Dansgaard-Oeschger events, PhD thesis, Niels Bohr Inst., University of Copenhagen, 2001.

Morrill, C. and Jacobsen, R. M.: How widespread were climate anomalies 8200 years ago?, Geophys. Res. Lett., 32, L19701, doi:10.1029/2005GL023536, 2005.

Nieto-Barajas, L. E. and Sinha, T.: Bayesian interpolation of unequally spaced time series, Stoch. Env. Res. Risk A., 29, 577–587, doi:10.1007/s00477-014-0894-3, 2014.

Parnell, A. C., Sweeney, J., Doan, T. K., Salter-Townshend, M., Allen, J. R., Huntley, B., and Haslett, J.: Bayesian inference for palaeoclimate with time uncertainty and stochastic volatility, J. R. Stat. Soc. C-App., 64, 115–138, doi:10.1111/rssc.12065, 2015.

Peavoy, D. and Franzke, C.: Bayesian analysis of rapid climate change during the last glacial using Greenland δ^{18}O data, Clim. Past, 6, 787–794, doi:10.5194/cp-6-787-2010, 2010.

Qian, P. Z. G. and Wu, C. F. J.: Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments, Technometrics, 50, 192–204, doi:10.1198/004017008000000082, 2008.

Rasmussen, S. O., Andersen, K. K., Svensson, A. M., Steffensen, J. P., Vinther, B. M., Clausen, H. B., Siggaard-Andersen, M.-L., Johnsen, S. J., Larsen, L. B., Dahl-Jensen, D., Bigler, M., Röthlisberger, R., Fischer, H., Goto-Azuma, K., Hansson, M. E., and Ruth, U.: A new Greenland ice core chronology for the last glacial termination, J. Geophys. Res.-Atmos., 111, D06102, doi:10.1029/2005JD006079, 2006.

Ribeiro Jr., P. J. and Diggle, P. J.: geoR: A package for geostatistical analysis, R news, 1, 14–18, 2001.

Rue, H. and Held, L.: Gaussian Markov random fields, Vol. 104 of Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, Boca Raton, FL, doi:10.1201/9780203492024, 2005.

Rue, H. and Martino, S.: Approximate Bayesian inference for hierarchical Gaussian Markov random field models, J. Stat. Plan. Infer., 137, 3177–3192, doi:10.1016/j.jspi.2006.07.016, 2007.

Rue, H., Martino, S., and Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, J. R. Stat. Soc. B-Met., 71, 319–392, doi:10.1111/j.1467-9868.2008.00700.x, 2009.

Schwarz, G.: Estimating the dimension of a model, Ann. Stat., 6, 461–464, 1978.

Stuiver, M.: GISP2 Bidecadal Oxygen Isotope Data, Pangea, doi:10.1594/PANGAEA.55531, 1999.

Stuiver, M. and Grootes, P. M.: GISP2 oxygen isotope ratios, Quaternary Res., 53, 277–284, 2000.

Thomas, E. R., Wolff, E. W., Mulvaney, R., Steffensen, J. P., Johnsen, S. J., Arrowsmith, C., White, J. W., Vaughn, B., and Popp, T.: The 8.2 ka event from Greenland ice cores, Quaternary Sci. Rev., 26, 70–81, 2007.

Tingley, M. P. and Huybers, P.: Recent temperature extremes at high northern latitudes unprecedented in the past 600 years, Nature, 496, 201–205, 2013.

Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B.: Piecing together the past: statistical insights into paleoclimatic reconstructions, Quaternary Sci. Rev., 35, 1–22, 2012.

Wanner, H., Solomina, O., Grosjean, M., Ritz, S. P., and Jetel, M.: Structure and origin of Holocene cold events, Quaternary Sci. Rev., 30, 3109–3123, 2011.

Wikle, C. K. and Berliner, L. M.: Combining information across spatial scales, Technometrics, 47, 80–91, doi:10.1198/004017004000000572, 2005.