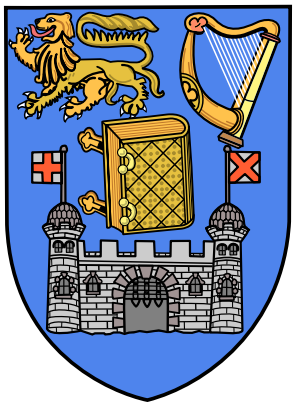
Bayesian Inference for Misaligned Irregular Time Series

with Application to Palaeoclimate Reconstruction

A thesis submitted to the University of Dublin, Trinity College
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Department of Statistics, Trinity College Dublin



August 2015

Thinh K. Doan

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement. The copyright remains solely with the author, Thinh K. Doan.

Thinh K. Doan

Dated:

Abstract

This thesis proposes new Bayesian methods to jointly analyse misaligned irregular time series. Temporal misalignment occurs when multiple irregularly spaced time series are considered together, or when the time periods defining the data points are not the same across different series. Other issues under consideration include errors in the time scales, and non-Gaussian processes for underlying latent values.

Our proposed models are hierarchical. This flexible framework is made concrete with the derivation of fast and efficient algorithms for parameter inference. The methods are general and can be used for any continuous time series processes. In addition, they open the door to other interesting possibilities for modelling and inference of multivariate time series in a spatio-temporal context.

We apply our methods to climate proxy signals to derive ancient climate histories. A major objective is to create climate data products (i.e. posterior summaries of climate on a regular time grid). One of the advantages of our data products is that they can easily be utilised to answer complex questions that are otherwise analytically intractable. We demonstrate this by using case studies of abrupt climate change events.

Acknowledgements

This thesis would not have been possible if it was not for my supervisors, John and Andrew. I'm deeply indebted to them for their close guidance and advice. I've learned a lot from their effective and exciting way of teaching and explaining. It is not an exaggeration to say that the choice to do a PhD with them has been one of the best decisions so far in my life.

I feel lucky to have been in both TCD and UCD during my time as a student. These world-class statistics departments are full of smart, kind and fun people. I have made many good friends: Tiep, Angela, Shane, Arnab, Arthur, Brett, James, Chaitanja, Shuarwei, Susie, Louis, Gernot, Cristina, Niamh, John, Berlinda and Nancy. Special mention has to go to Donnacha, Sean and Jason, with whom I shared many the whys, whats and hows in statistics, as well as my wedding preparation.

Outside of TCD and UCD, I'm grateful to Brian and Judy from Durham for providing an interesting dataset, and Eric from Cambridge for helpful comments on an incomplete report which subsequently lead to an interesting article.

No words would be sufficient enough to describe my most sincere gratitude to my parents (Long and Le), brother (Phu) and wife (Van) for their unconditional love and support. I am going to be a father for the first time in August this year, and this thesis is dedicated to this new member of our family.

This work is funded by Science Foundation Ireland grant no. 10/RFP/MTH2779.

Thinh K. Doan

Trinity College Dublin

August 2015

Publications

Published work (at the time of writing) of relevance to this thesis are listed below. More specifically, chapter 4 is an edited version of paper 1 and chapter 5 is an extension of paper 2.

1. Joint Inference of Misaligned Irregular Time Series with Application to Greenland Ice Core Data, with John Haslett and Andrew Parnell. This paper was published in *Advances in Statistical Climatology, Meteorology and Oceanography* in 2015. The methods were developed in collaboration with the co-authors. I performed all of the implementation and writing. The co-authors edited and improved the paper. The data, code and analysis are available at `http://www.scss.tcd.ie/~tdoan/research.html`.

2. Bayesian Inference for Paleoclimate with Time Uncertainty and Stochastic Volatility, with Andrew Parnell, John Haslett, James Sweeney, Michael Salter-Townshend, Judy Allen and Brian Huntley. This paper was published in *Royal Statistical Society: Series C (Applied Statistics)* in 2015. The co-authors were responsible for the model formulation, data analysis and writing up. I carried out some practical implementations, and contributed some functions to the R package Bclim which accompanies this paper.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis is concerned with three broad topics: misaligned irregular time series, Bayesian inference, and palaeoclimate reconstruction. This introductory chapter contains an overview of each topic, a succinct account of the main contribution of the work done, and an outline summary of the remaining chapters.

1.1 Statistical motivation

Many real world time series data sets are irregularly spaced in time. Furthermore, the time period over which a data point is collected can be different across series (i.e. different temporal supports) or, in some cases, uncertain. When we consider multiple series together, the series themselves are temporally misaligned. Frequently, interest lies in multiple series with different irregularities. There is a relatively small body of literature on the statistical treatment of such data. The objective of this research is to develop new methods for a joint statistical inference of underlying processes which give rise to such data.

The statistical methods in this thesis are considered in a Bayesian hierarchical context. Under our framework, there are a large number of unknown processes and parameters. This is associated with a heavy computational burden. Therefore, an important part of this thesis is concerned with efficient computational algorithms. We propose strategies to marginalise out high dimensional latent processes, where appropriate. Additionally, we derive and implement both simulation-based and simulation-free approaches for fast inference of model parameters.

A motivating application for this work is a quantitative reconstruction of climate histories from proxies recorded at multiple locations over thousands of years. Climate change is undoubtedly one of the most important environmental issues of the present day. A crucial basis for predicting future climate is its past condition. Unfortunately, instrumental weather measurements are only available for the last few hundred years. Fortunately, some palaeoclimate proxies can be used as a guide for climate in the past. Hereafter we refer to the term *reconstruction* as the exercise of using statistical methods to draw inference on ancient climate based on proxy signals. More specifically, we aim to provide methods to reconstruct and analyse aspects of climate from data preserved in ice and fossil sediment cores.

When dealing with an ice or fossil sediment core, it is important to take into account the relationship between the core's depth and calendar time. This is because the accumulation rate of deposits varies significantly with time, due to the change in climate as well as processes other than climate inside a core. Thus, time is a non-linear function of depth, such that its temporal resolution decreases from the top to bottom of a core. In other words, even if cores are regular in depth, the associated ages can be irregular. Consequently, multiple cores with different irregularities are necessarily misaligned. Formally, our objective is to derive '*climate data products*', i.e. summaries of climate information with fully quantified estimates of uncertainty from raw palaeoclimatic data.

An important component in all of our proposed models is the Gaussian Markov assumption based on the multivariate independent increment process which provides a natural vehicle for *joint* statistical inference. A joint approach allows 'borrowing strength' between different sources of information in order to study the true underlying climate processes of interest. It permits a reliable modelling of the impact of the multiple sources of uncertainty in proxies to produce a rich climate output. This is in contrast to, for instance, the reconstruction of the notorious 'hockey stick' (Mann et al. 1998) which utilises only high resolution proxy data. Similarly, the reconstruction procedure proposed by Li et al. (2010) uses the so-called 'pseudo-proxy' as opposed to raw proxy data. Both examples can be viewed as climate data products. Moreover, the time range of these reconstructed climates goes back to just about a thousand calendar years Before Present (k cal yr BP, where Present is 1950), in contrast to more than 10k cal yr BP as considered in this thesis.

1.2 Overview of contributions

The thesis focuses on Bayesian inference for three cases of temporal misalignment in the analysis of multivariate time series data. The original contributions to the literature are:

1. In chapter 3, we propose a simple hierarchical framework for joint statistical inference of misaligned irregular time series. We show that our joint approach is more efficient than independent alternatives. We also discuss and demonstrate that process histories are a useful (and in some cases, the only) way to study non-linear functions of partially observed processes of interest.

2. In chapter 4, each time series has its own measurement procedure, which leads to different support for observations across multiple series. We propose a model for this. Based on this model, we implement a fast and efficient simulation-free algorithm that completely bypasses Markov chain Monte Carlo (MCMC) methods. We apply our methods to jointly analyse multiple ice cores. To the best of our knowledge, this work represents the first attempt to perform Bayesian inference for multiple ice core time series in their raw and misaligned form.

3. In chapter 5, temporal misalignment occurs when the times of observations are subjected to some uncertainty. We develop a multivariate stochastic volatility model and derive a modularised MCMC algorithm for computation of posterior distributions. This model can deal with abrupt changes in underlying process of interest. Although formulation of a stochastic volatility model is not new in the statistical literature, we are unaware of any work that use a multivariate long-tailed distributions for stochastic interpolation of climate histories.

1.3 Outline of chapters

The rest of this thesis is divided into the following chapters:

Chapter 2: Palaeoclimate time series modelling and Bayesian statistics
In the first part of this chapter, we discuss specific features of two applications (oxygen isotope in ice cores and pollen count in sediment cores) that lead to difficulties in

using standard approaches for time series analysis. These include irregularly spaced timings and errors in the time scale corresponding to different measurement procedures. Additionally, we review and evaluate existing solutions in the statistics literature, with special emphasis on joint inference of multiple time series.

In the second part, we discuss the basis of the statistical methods that will be used in subsequent chapters. First, we review the Bayesian hierarchical modelling approach. Then, we describe the independent increment process in detail and provide justification for this choice in the context of palaeoclimate reconstruction. After this, we give a general account of techniques for inference of parameters using both simulation-based and simulation-free approaches.

Chapter 3: A toy example

This is a short but nonetheless important chapter. In this chapter we give some basic definitions that will help to establish the framework on which subsequent chapters are based. Additionally, we compare and contrast different techniques for computing and summarising data products. We also discuss the benefit of joint versus separate statistical inference alternatives.

Chapter 4: Joint inference of processes with different temporal supports

This chapter is concerned with data sets with different supports underlying their measurements. An extension to the simple model in chapter 3 is proposed. To deal with different supports, we use *change of support* theory from the geostatistics literature. We further derive a fast simulation-free algorithm for parameter inference, and avoid the use of MCMC.

The proposed methods are applied to analyse two ice cores drilled in neighbouring locations in Greenland. We show that our approach allows for a richer analysis than previously possible. Additionally, we demonstrate that our sampled process histories are useful to study non-linear functions of underlying processes of interest.

Chapter 5: Joint inference of processes with time uncertainty and stochastic volatility

This chapter presents a more general framework than that discussed in previous chapters. Under the Bayesian hierarchical structure, we incorporate a variety sources of information, including treatments for temporal uncertainty and stochastic volatility.

At the same time, we show how Bayesian modularisation can be applied to reduce the complexity of the overall model.

The main novelty of this chapter lies in the specification of a multivariate non-Gaussian stochastic volatility prior model for an underlying latent process. This subsequently leads to the derivation of a modularised MCMC framework, and development of a stochastic interpolation algorithm with heavy-tailed behaviour. The motivation for this chapter is to jointly reconstruct ancient climate histories from three sites in Finnmark, conditioning on pollen data.

Chapter 6: Conclusion and future work

The last chapter gives an overview of the results of the thesis and discusses the contributions with several remarks for future directions of research.

Chapter 2

Palaeoclimate time series and statistical methodology

This chapter presents a background for problems that will be solved in subsequent chapters. We also provide a broad literature review of relevant statistical methods.

We begin, in Section 2.1, with two example applications; oxygen isotopes from ice cores and pollen grains from lake sediments. In particular, we discuss features of palaeoclimate data which pose a number of challenging statistical issues, and announce our objective which is to create easy-to-use climate data products from these raw data sets. Next, in Section 2.2, we discuss difficulties, and review existing solutions to varying data supports, time uncertainties, irregularly-spaced time series and temporal misalignment.

Then, in Section 2.3, we provide a general description of Bayesian hierarchical models; they are the building block for the methodological contribution of this thesis. More specifically, we discuss independent increment processes as the prior model choice for the main latent process, and summarise some techniques for computing posterior distributions.

2.1 Applications

Climate proxies are indirect climate measurements. They provide a useful source of information to climatic conditions in the past. This section introduces two examples of proxies. It also outlines our objective in performing statistical inference.

2.1.1 Oxygen isotopes from ice cores

Ice cores play an important role in revealing Earth's climate history via the analysis of their chemical composition. Here, we focus on the ratio of stable and unstable oxygen isotopes (expressed as $\delta^{18}O$) 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.

A further step would be required to transform $\delta^{18}O$ into an aspect of climate which might then add considerable uncertainties. For clear presentation of new statistical methods for misaligned time series, we will work directly with $\delta^{18}O$ measurements.

Various ice core data sets are available from the National Climatic Data Center (`http://www1.ncdc.noaa.gov/pub/data/paleo/icecore/`). In this thesis, 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 kilometres deep. The locations of these two cores are about 30 kilometres apart, as shown in Fig. 2.1.

The $\delta^{18}O$ measurements are recorded as a function of depth. Whilst ice core data sets are presented in several versions, we use a raw version in which the cores are cut into sections of equal lengths; 200cm and 55cm for GISP2 and GRIP respectively. In this context, 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. 55cm) rather than volumes. A length of a core section directly maps to a period of elapsed time. Hence, we define 'support' as the period of time over which is represented by a single data point.

The dating of these cores is conducted by counting annual layers in chemicals that show a summer/winter seasonal cycle (e.g. Rasmussen et al. 2006). Intuitively, the lower the sediment is, the older the information in the core is. As pointed out, for instance, by Stuiver & Grootes (2000), measurements from these cores contain negligible dating errors, particularly for the period of relatively stable climate approximately between 0 and 11 k cal yr BP. This is the time period we consider in the application in chapter 4. We defer discussion pertaining to dating uncertainty to chapter 5,

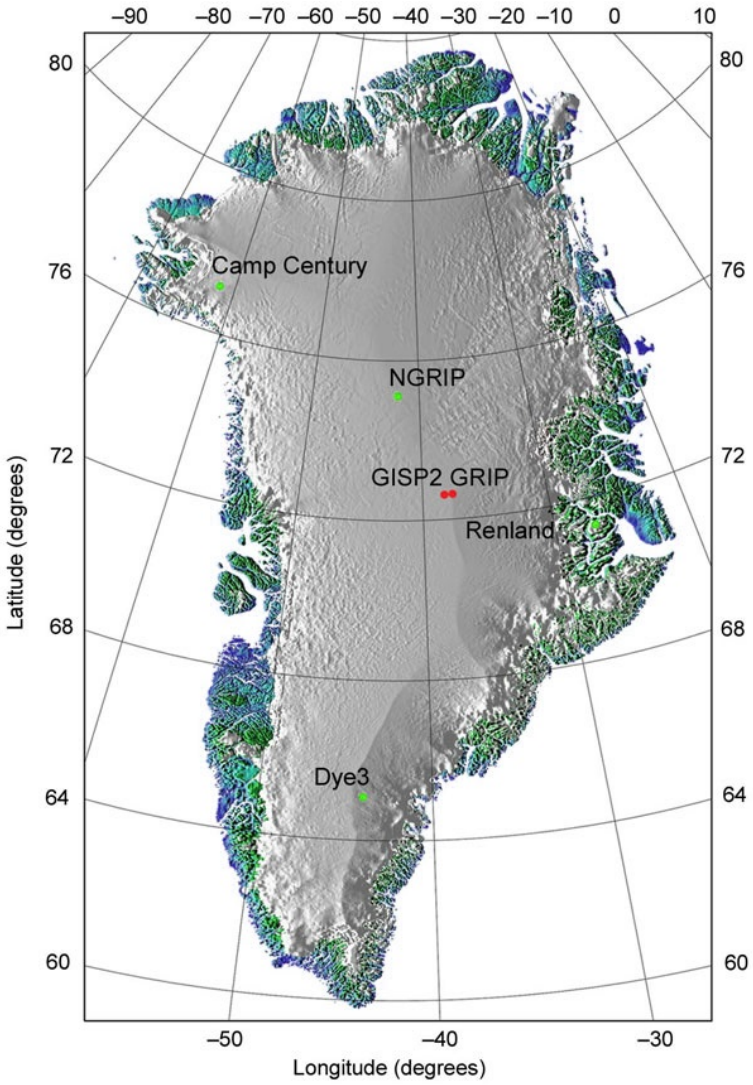


Fig. 2.1: Locations of drill sites on the Greenland Ice Sheet (modified from Andersen et al. (2004)). In particular, we consider site GISP2 (72.58^0N, 38.46^0W) and GRIP (72.59^0N, 37.64^0W), marked with red points.

when dealing with non-negligible radiocarbon dated errors from fossil pollen at lake sediments. A detailed description of pollen data sets is discussed in Section 2.1.2.

Due to ice compression over time, age and depth are not linearly related. Material at the top pushes down on material at the bottom therefore temporal compression generally increases with depth. This is most clearly seen in Fig. 2.2 where we plot age versus depth for both cores for the last 100 k cal yr BP. For this reason, a data set comprising of sections of equal length does not generally lead to a regularly spaced time series. Figure 2.3 shows the measurements of GISP2 and GRIP ice cores for the last 100 000 year. Boxplots of the age increments clearly show different irregularities

in the ages. In Chapter 4, we consider only data of approximately the last 11 k cal yr BP, a period of relatively stable climate conditions.

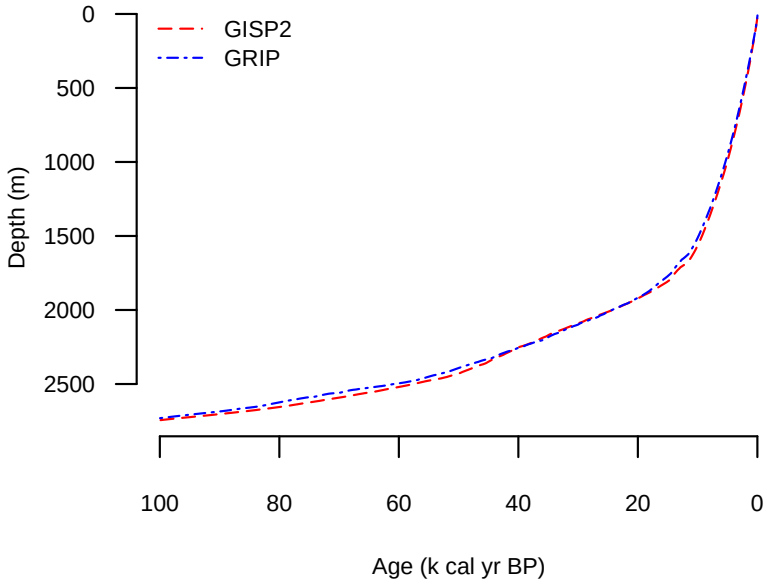


Fig. 2.2: Plot of depth versus age for GISP2 and GRIP approximately the last 100 k cal yr BP. Slope is more gradual at older ages, implying that compression generally increases with age. Here, we assume no errors associated with the estimation of time from depth. In chapter 4, we consider only data of approximately the last 11 k cal yr BP where the age-depth relationship can be approximately linear.

To sum up, a considerable challenge in a statistical analysis of multiple ice core time series is associated with the irregular timing of measurements, temporal misalignment across series, and different measurement supports.

2.1.2 Pollen data from sediment cores

Although ice provides a rich source of information for palaeoclimate reconstruction, ice cores can only be drilled at limited locations. In contrast, fossil pollen samples can be found in many more areas on Earth. An example is the European Pollen Database (Fyfe et al. 2009, see also `http://www.europeanpollendatabase.net`). A statistical analysis of this type of dataset will face issues similar to that with ice cores, namely different irregularities in time and temporal misalignment. Additional, and probably more challenging, issues are associated with time uncertainty and stochastic volatility - details of which are discussed in the rest of this section.

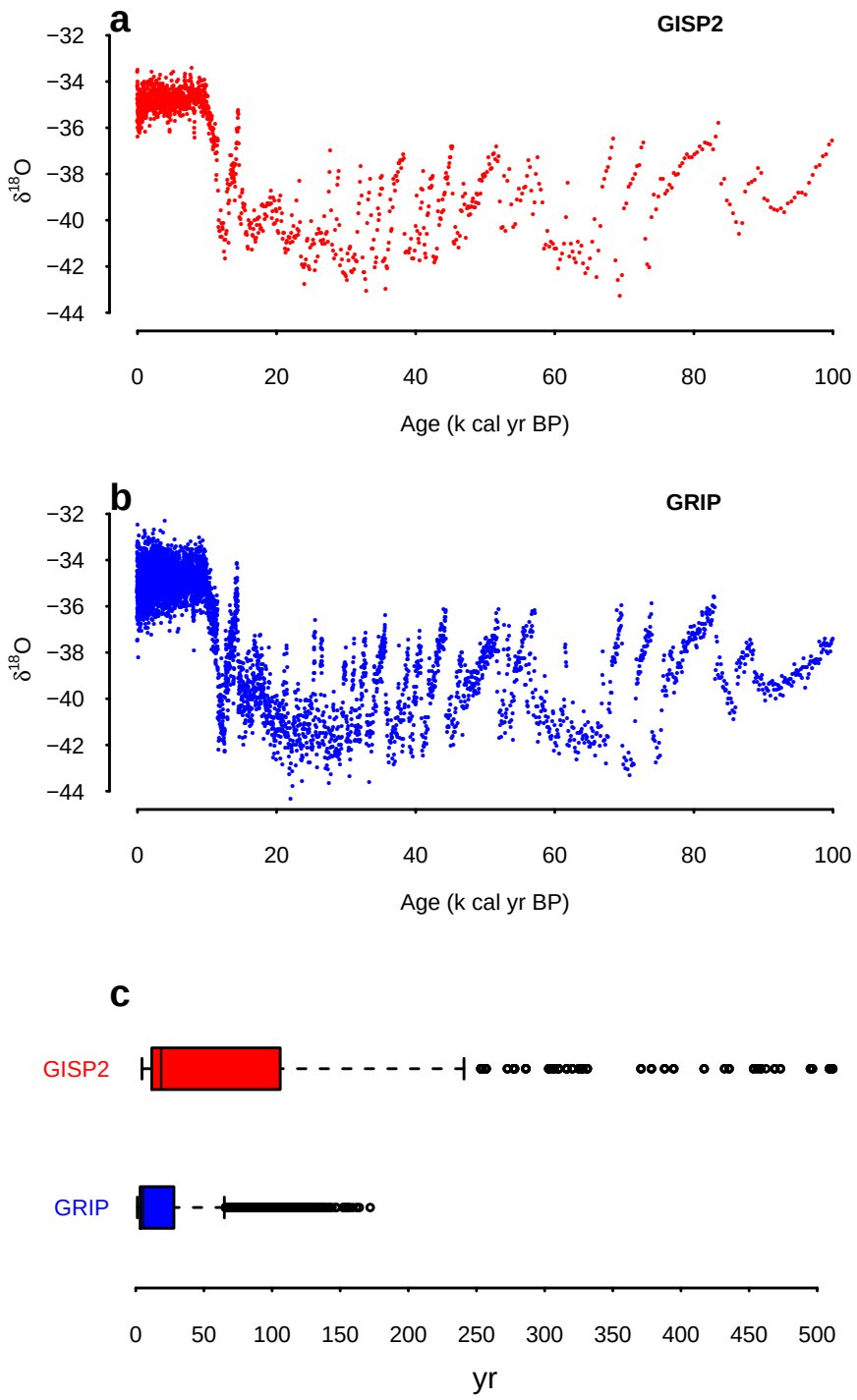


Fig. 2.3: Scatter plots of $\delta^{18}O$ measurements and ages of **(a)** GISP2 and **(b)** GRIP. **(c)** Boxplots of the age increments clearly show different irregularities in the ages.

We consider a data set comprising three cores from three lakes in Finnmark, nothern Norway. The lakes are located at Liten Cappesjavri, Over Gunnarsfjorden and Over Kobbkrokvatnet, and the corresponding sediment cores are known as M, NKH1

and PS8. The physical locations form a west-east transect in the coastal regions of Finnmark, as shown in Fig. 2.4. This region is believed to be sensitive to past climate changes that were related to the disruptions of ocean circulation variations (Huntley et al. 2013).

At a selection of depths, scientists identify, count and compute pollen concentrations that have been accumulated over thousands of years beneath lakes or bogs. Therefore, each data point reflects counts of many pollen types at a given depth. The pollen percentage diagram for a sediment core from one of the sites in Finnmark (site M) is represented in Fig. 2.5.

In addition to the type of data set described in the previous paragraph, there is a training set of modern pollen surface samples at locations around Europe. The latest European modern pollen data base is described in Davis et al. (2013). Nevertheless, in chapter 5 we use the training set detailed in Haslett et al. (2006). This is for the purpose of direct comparison to Parnell et al. (2015) who uses the latter data set. In any case, a training data set consists of a large number of known climate conditions and associated pollen counts. Statistical connection between pollen and climate is known as *forward inference*. Conversely, given fossil pollen information, it is possible to reverse the inference direction (*inverse inference*) to draw inference about ancient climate (Haslett et al. 2006, Salter-Townshend & Haslett 2012). The output of inverse inference is a layer-by-layer climate reconstruction at a selection of depths. A schematic representation of the forward and inverse inference is shown in Fig. 2.6.

Similar to the situation with ice core, depths from sediment core can be transformed to calendar ages, albeit with uncertainty. Scientists can date material at some depths to return radiocarbon ages. Radiocarbon dating is a common dating technique in palaeoecological research to determine the age of an object or event (Buck et al. 1996). Whilst calendar ages are directly comparable between cores, this is generally not true with radiocarbon ages. Thus, it is crucial to transform radiocarbon ages to calendar ages. This calibration process is often done via a calibration curve derived from a huge number of radiocarbon dates of material with known ages (Reimer et al. 2013). The implication is that measurements of time have inherent error.

Further uncertainty arises from stochastic interpolation for ages at depths where no information is present pertaining to radiocarbon ages. There has been a growing

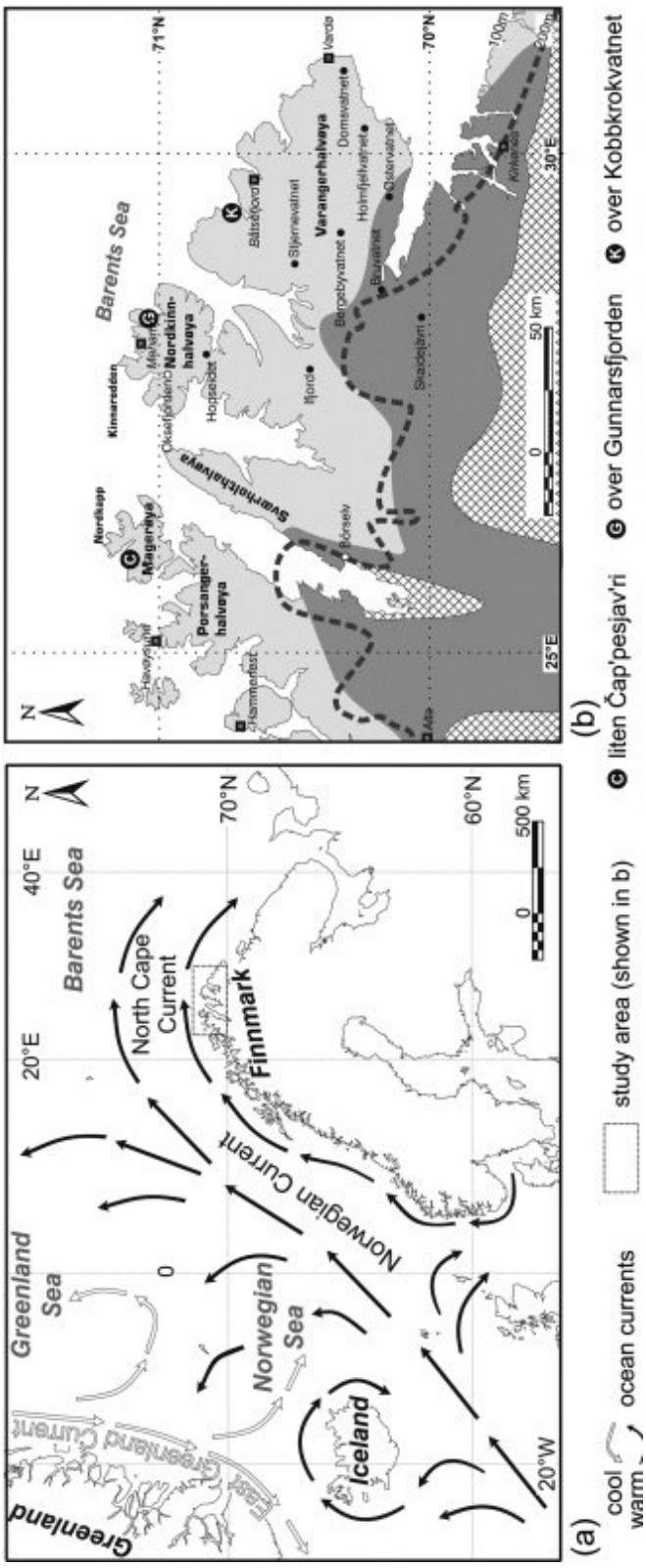


Fig. 2.4: Plots taken from Huntley et al. (2013) showing (a) the region under study in northern Norway and how it is related to ocean surface currents, and (b) the locations of the three lakes in Finnmark, which form a west-east transect.

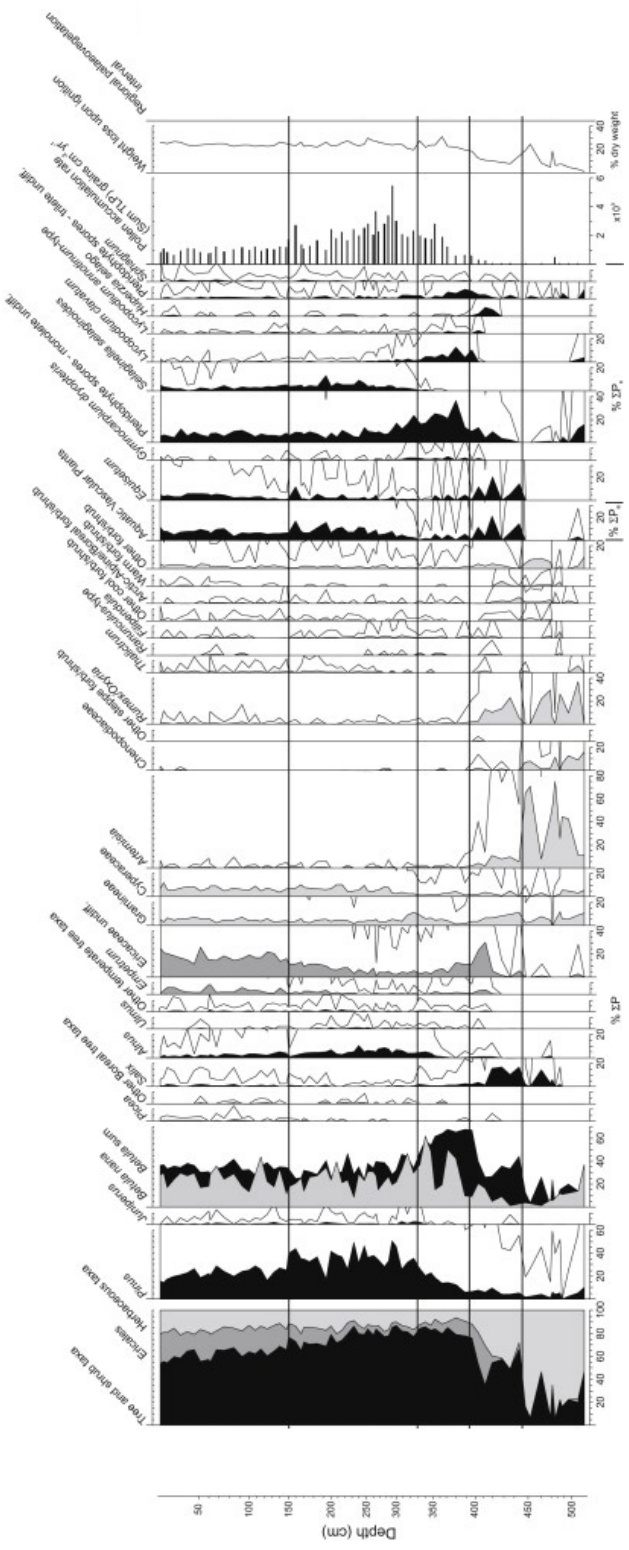


Fig. 2.5: Pollen percentage within a core at one of the sites in Finnmark (site M). For pollen details of cores from other sites, see Huntley et al. (2013). The vertical axis shows the depths of the core (sediments in the upper part of the core are the most recent ones). The horizontal axis shows the percentage of many pollen types.

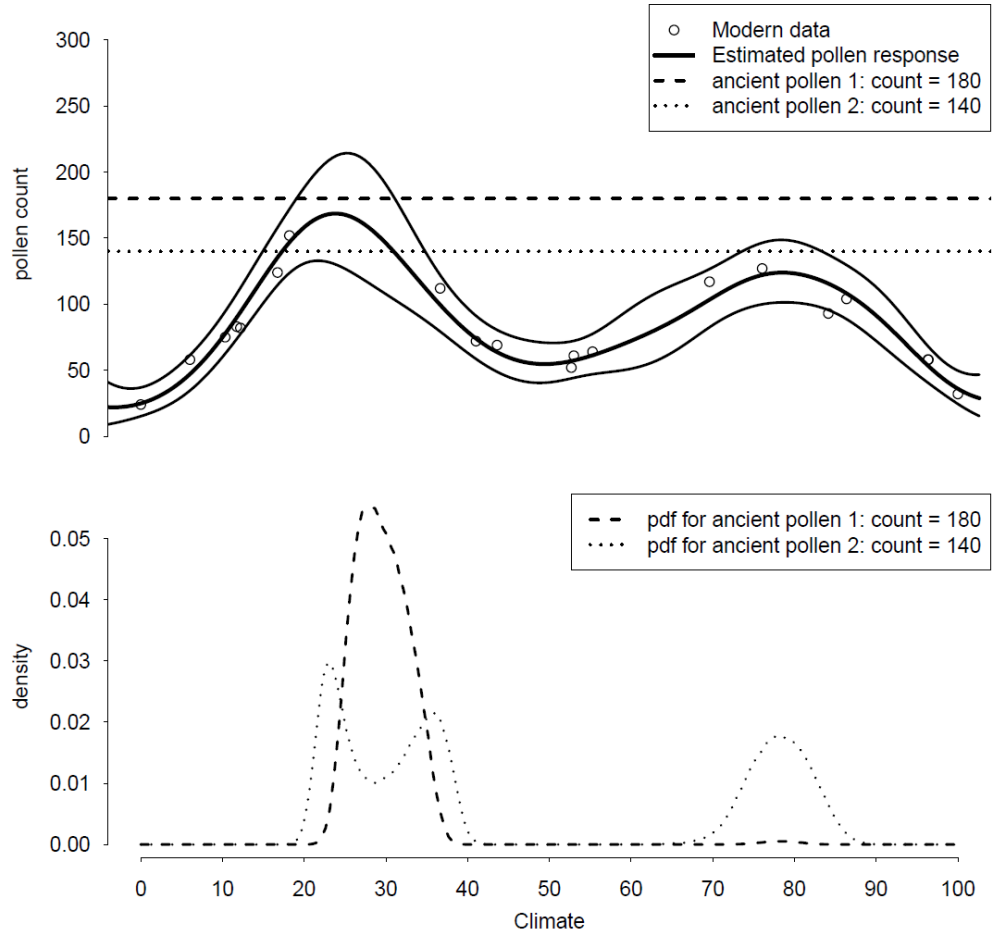


Fig. 2.6: A schematic plot of one-layer-at-a-time paleaoclimate reconstruction. The upper panel shows example of a functional relationship between climate and pollen, created from a training set of modern climate and pollen data. Given two example fossil (ancient) pollen counts, we obtain posterior estimates of unknown climate in the lower panel. This plot is modified from Parnell et al. (2015).

interest in Monte Carlo age-depth models within a Bayesian framework (Parnell et al. 2011). Figure 2.7 is a schematic representation of stochastic interpolation of ages at an arbitrary selection of depths based on the calibrated calendar ages. Each sample of interpolated ages for a selection of depths is known as a *chronology*.

Thus, each fossil pollen core is a time series of multivariate pollen counts, the times themselves being uncertain. This is contrary to the ice core data as considered in chapter 4, in which temporal uncertainty was assumed to be inconsequential because the dating procedure is considered relatively accurate.

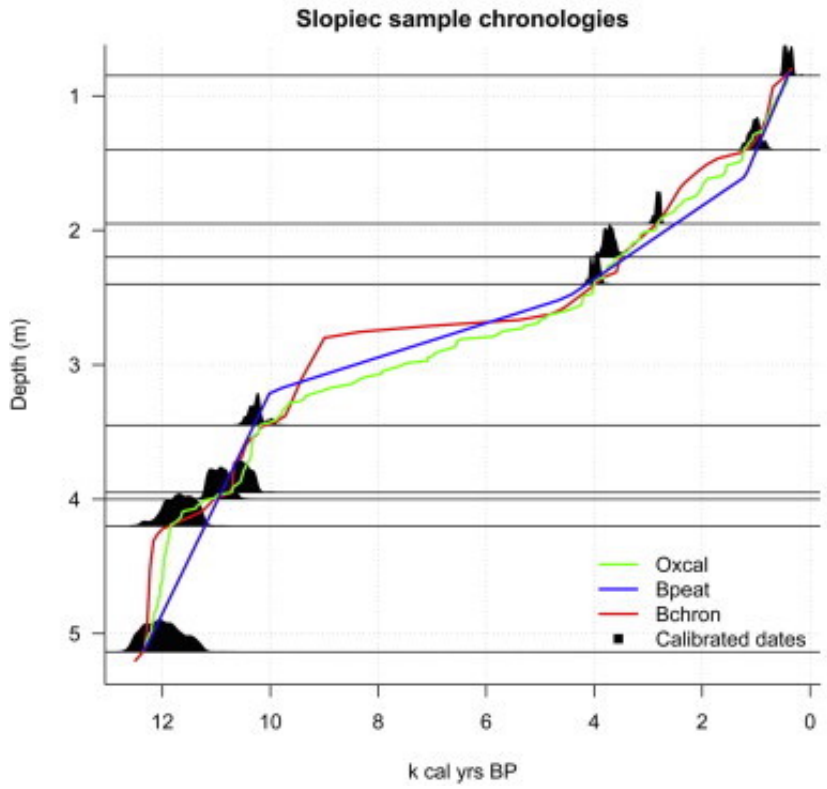


Fig. 2.7: A schematic plot of three chronology samples (generated from Oxcal Ramsey (2008), Bpeat(Blaauw & Christen 2005) and Bchron(Haslett & Parnell 2008)). First, radiocarbon ages at ten depths of a core are calibrated to ten distributions of calendar ages. Then, chronology models are fitted to stochastically interpolated calendar ages associated with a selection of depths. This plot is taken from Parnell et al. (2011).

2.1.3 Climate data products

Palaeoclimate time series data are available in two forms: 'raw' (typically irregular, or uncertain in time) and 'data products' (typically regularly-spaced in time, i.e. 'gridded'); as pointed out, for instance, by Chandler et al. (2012).

Previous methods for creating data products from ice core rely on a variety of techniques; from simple running averages (Stuiver & Grootes 2000, Thomas et al. 2007) to complex parametric smoothing (Peavoy & Franzke 2010, Nieto-Barajas & Sinha 2015). The statistical literature on creating the type of data products (defined in the previous paragraph) from raw pollen data is scarce. As far as we are aware, Parnell et al. (2015) is the only existing reference which takes into account of the challenges discussed in Section 2.1.2, to create gridded data product from raw pollen signals. However, their methods are only applicable to a single core.

To sum up, existing techniques - broadly interpolators or smoothers - are applied to only one core at a time. Therefore, an important objective of this research is to develop methods to *jointly* create easy-to-use climate data products conditional on all raw information, from multiple cores.

2.2 Statistical inference for multiple misaligned time series

From the data sets discussed in the previous sections, we see that palaeoclimate data from a core typically consists of two main sets of measurements: proxy and dating information. Table 2.1 shows a schematic representation of this.

[Depth]	Time	[Proxy data]	Climate
top	present		
⋮	⋮	⋮	⋮
bottom	past		

Table 2.1: A schematic representation of palaeocimate data from a core. Square brackets are used to distinguish observed measurements from unobserved latent variables. At some depths of the core, material has been dated to return times; the times themselves may be available with errors. Proxy data (also available at some depths) are used to learn about climate. Reconstruction is the process of turning proxy information at some depths into climate at a selection of times.

Palaeoclimate time series data have many unique features. Standard statistical time series analysis models such as autoregressive integrated moving average focus on discrete time processes (Donner 2007, Mudelsee 2010). Hence, they are not generally applicable to the types of raw palaeoclimate time series data, particularly those discussed in the previous sections. Succinctly stated, the outstanding issues which deserve immediate attention and effort from statisticians include, but are not limited, to the following:

- High level of uncertainty in the observed signal due to the sensitivity of proxy to various processes other than climate. Moreover, the relationship between proxy

signals and climate can be non-linear and require complex modelling assumptions.

- Missing or irregularly-spaced data and, in some cases, uncertainty in the data's timings. Under the multivariate setting, these series are temporally misaligned.

Relevant works include Erästö et al. (2012) who propose a Bayesian method to merge distinct paleoclimate time series to learn about the common feature of these series. However, this method is only applicable to post-processing climate output which have previously been reconstructed from raw proxy data. Another related work is Li et al. (2010) who attempt to reconstruct historical temperature using multiple sources of 'pseudo' proxies. These synthetic proxy datasets are output from other climate models. Hence, again, this method does not make use of raw data. As mentioned in the previous section, Parnell et al. (2015) seems to be the only attempt to reconstruct climate from raw pollen data whilst taking into account all of the aforementioned challenges. Alas, a limitation of this method is that it does not perform joint inference for multiple fossil 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-accuracy) experiments with those from a few expensive (high-accuracy) experiments by linking the data via different layers of modelling (Qian & Wu 2008). In medical applications, one recommendation to overcome issues with misalignment is to align the times into a regular template 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.

Misaligned time series can be viewed as a special case of spatial misalignment in spatial statistics; the so-called 'change of support problem' for data that are indexed in both space and time (e.g. Gelfand et al. 2001, Wikle & Berliner 2005). In this context, statistical inference involves studying the statistical properties of a stochastic process at 'supports' that are different to that associated with data. The methodological development in this thesis can be seen as a special case of this problem. We focus on the case where an observation is defined at an instantaneous point in time and seek

prediction of the underlying process at arbitrary new time points.

2.3 Bayesian inference for hierarchical models

The literature review in the rest of this chapter focuses on Bayesian statistics. It will be general and does not directly link to issues in palaeoclimate reconstruction.

2.3.1 Hierarchical models

The use of Bayesian inference for palaeoclimate problems has increased in recent years. The Bayesian approach considers data $\boldsymbol{y} = \{y_1, \ldots, y_n\}$, having already being observed, to be fixed. Suppose an (unobserved) parameter τ is random, to which *prior* information can be assigned to reflect our knowledge before observing any data. A whole range of different priors is available, ranging from subjective and informative expert opinion to uninformative. The conditional distribution of data given the process prior is the *likelihood* function. Letting π denote a probability distribution function, Bayes' theorem states that

$$\begin{aligned} \pi(\tau|\boldsymbol{y}) &= \frac{\pi(\boldsymbol{y}|\tau)\pi(\tau)}{\pi(\boldsymbol{y})} \\ &\propto \pi(\boldsymbol{y}|\tau)\pi(\tau) \\ \text{posterior} &\propto \text{likelihood} \times \text{prior} \end{aligned} \tag{2.1}$$

It is possible to omit the denominator on the right hand side (RHS) of the above equation and replace the equality sign with one indicating proportionality since the denominator is not a function of τ. The product of prior and likelihood is proportional to the *posterior*. Intuitively, the more data that become available, the less influence the prior has on the posterior results.

Further levels of dependency are possible. For instance, suppose data $\boldsymbol{y}$ are observed from a latent (unobserved) process $\boldsymbol{x} = \{x_1, \ldots, x_n\}$; the latter process is controlled by parameter v. We can impose a conditional independence structure on the data $\boldsymbol{y}$, given $\boldsymbol{x}$ and relevant parameters. Let $\boldsymbol{\theta} = \{\tau, v\}$. In a full Bayesian analysis, prior distributions for $\boldsymbol{\theta}$ are introduced as another stage of the hierarchy. In fact, the joint distribution of all unknown processes and parameters given the data can be written in a hierarchical structure. This is a Bayesian hierarchical model (BHM), see, for instance

Gelman et al. (2014, Chapter 5). A fairly simple hierarchical structure is

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \prod_{i=1}^{n} \pi(y(t_i)|x(t_i), \tau)\pi(\boldsymbol{x}|v)\pi(\boldsymbol{\theta}) \tag{2.2}$$

It can be seen that (2.2) is an extended version of (2.1). BHMs are simple, yet provide a powerful and unified framework for data analysis, and allow us to combine multiple sources of information. The corresponding graphical representation of the hierarchical structure in 2.2 is depicted in Fig. 2.8, where circles indicate parameters/latent random variables and boxes indicate observations.

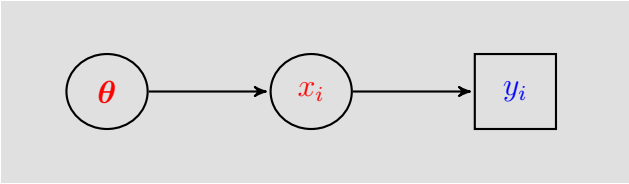


Fig. 2.8: A Directed Acyclic Graph (DAG) of the hierarchical model specified by Eq. (2.2). Circles indicate parameters/latent random variables whilst boxes indicate data. Note that y_i and x_i are defined for $i = 1, \ldots, n$ but $\boldsymbol{\theta}$ is not.

2.3.2 Independent increment process prior distributions and Markov property

A widely used prior model for temporal smoothing assumes independent increments in the first or second order of $\boldsymbol{x}$ (e.g. Haslett et al. 2006, Lindgren & Rue 2008). The first order model assumes smoothness of first differences whilst the second order model is concerned with the smoothness of the rate of change of $\boldsymbol{x}$. In this thesis, we focus on the first order independent increment model. This continuous-time stochastic process does not make any assumption on the process itself, but only on its increments. In the context of palaeoclimate reconstruction, it seems plausible that changes in climate at different series/cores/sites are similar if the sites are close in geographical location. Hence, a multivariate version of the process discussed in (2.3) below will play an important role in the models we propose in subsequent chapters.

We illustrate this process with a univariate case. Suppose $\boldsymbol{x}$ is a continuous *time* stochastic process. Hereafter we write x_i and y_i as $x(t_i)$ and $y(t_i)$ respectively. We

emphasise that $\boldsymbol{x}$ is defined at all times. However, to simplify the example we assume that time is observed on a discrete time grid. We model an increment as a zero-mean Gaussian prior with variance v that is directly proportional to a time difference, i.e.

$$x(t_{i+1}) - x(t_i) \sim \mathcal{N}(0, \delta_i v) \tag{2.3}$$

where $\delta_i = |t_{i+1} - t_i|$, and v is the squared *volatility*, governing the smoothness of $\boldsymbol{x}$. Note that the parameter v might not be adequate when underlying process of interest is highly oscillating. A remedy for this is to allow v to be time-varying (Parnell et al. 2015).

The joint distribution for $\boldsymbol{x}$, conditioning on the first value $x(t_1)$, is

$$\begin{aligned}
\pi(\boldsymbol{x}|v) &\propto v^{-(n-1)/2} \exp\Big(-\frac{v^{-1}}{2} \sum_{i=1}^{n-1} \delta_i^{-1} \big(x(t_{i+1}) - x(t_i)\big)^2 \Big) \\
&\propto v^{-(n-1)/2} \exp\Big(-\frac{1}{2} \boldsymbol{x}^T \mathbf{Q}_x \boldsymbol{x} \Big)
\end{aligned}$$

where

$$\mathbf{Q}_x = v^{-1} \begin{pmatrix}
\delta_1^{-1} & -\delta_1^{-1} & & & \\
-\delta_1^{-1} & (\delta_1^{-1} + \delta_2^{-1}) & -\delta_2^{-1} & & \\
 & & \ddots & & \\
 & & -\delta_{n-2}^{-1} & (\delta_{n-2}^{-1} + \delta_{n-1}^{-1}) & -\delta_{n-1}^{-1} \\
 & & & -\delta_{n-1}^{-1} & \delta_{n-1}^{-1}
\end{pmatrix} \tag{2.4}$$

Here, $\mathbf{Q}_x$ is the precision (inverse covariance) matrix, with the blank entries being zeroes. More details on the construction of precision matrices for univariate independent increment processes applied to irregularly spaced time series are discussed in Rue & Held (2005, Section 3.3) and Lindgren & Rue (2008).

Formally, a stochastic process with independent increments is a *Markov* process. We see that in the construction of $\boldsymbol{x}$, conditioned on information at the *present*, the past gives no new information about the future. A further look at the *full conditional* distribution of $\boldsymbol{x}$ at each time point (given all *past and future* points) shows that the process at the current time is only dependent on the process at the previous and next time. The exception is at the edge, i.e. the first and last time point. This form of dependency can be seen in the tridiagonal structure of the precision matrix $\mathbf{Q}_x$ as shown in (2.4). It turns out that this precision matrix gives information about conditional

independence properties (Rue & Held 2005, Theorem 2.2). Generally, sparse precision matrices have cheaper storage cost and lower computational burden than dense matrices. We discuss some computational aspects associated with sparse precision matrices in Section 2.3.3.4.

All models in this thesis will be set up in ways such that the resulting posterior distributions for main latent processes can inherit the Markov property from independent increment prior distributions. The embedded Markov properties will ensure that associated precision matrices are sparse. More specifically, we make use of multivariate independent increment process. In this case, the process $\boldsymbol{x}$ is multi-dimensional. The increments have multivariate Gaussian distributions; each has, for simplicity, a mean zero and cross-covariance matrix $\boldsymbol{\Sigma}$. Thus, the precision matrix for $\boldsymbol{x}$ is block-triadiagonal. We can write this matrix as a Kronecker product, denoting $\mathbf{Q}_x \otimes \boldsymbol{\Sigma}^{-1}$. A useful reference for Kronecker products is Harville (1997). Our discussion in the rest of this chapter focuses on the univariate case, for the purpose of simplicity.

2.3.3 Computation of posterior distributions

In the model specified by (2.2), the posterior distributions of interest include the joint posterior distributions $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, $\pi(\boldsymbol{x}|\boldsymbol{y})$, and *marginalised* joint posterior distributions $\pi(v|\boldsymbol{y}), \pi(\tau|\boldsymbol{y}), \pi\big(x(t_i)|\boldsymbol{y}\big)$ or some combination of i, for all i. These unknown quantities of interest correspond to the integrals:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \int \pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{x} \tag{2.5a}$$

$$\pi(\boldsymbol{x}|\boldsymbol{y}) = \int \pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta} \tag{2.5b}$$

$$\pi(v|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{y})\, d\tau \tag{2.5c}$$

$$\pi(\tau|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{y})\, dv \tag{2.5d}$$

$$\pi\big(x(t_i)|\boldsymbol{y}\big) = \int \pi\big(x(t_i)|\boldsymbol{y}, \boldsymbol{\theta}\big)\pi(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta} \tag{2.5e}$$

As a general difficulty in Bayesian inference, the above integrals are typically intractable. Furthermore, there is often high correlation within $\boldsymbol{x}$, and between $\boldsymbol{x}$ and $\boldsymbol{\theta}$. Many solutions are discussed in the literature, we will only mention the techniques used later in this thesis.

In all modelling setting in this thesis, $\dim(\boldsymbol{\theta})$ is typically small whilst $\dim(\boldsymbol{x})$ is large. Hence, the overall computational cost can be alleviated if $\boldsymbol{x}$ can be marginalised out of the overall model so that inference can initially be focused on the lower dimensional parameters.

2.3.3.1 Marginalisation of latent processes

We begin by rewriting the definition of conditional probability:

$$\begin{aligned}\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y}) &= \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\ &= \pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\end{aligned}$$

Moreover,

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi(\boldsymbol{y})}$$

Hence:

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \propto \pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{2.6}$$

In what follow we describe two case scenarios that are relevant to this thesis, and suitable solutions in a full Bayesian analysis.

Case 1: Gaussian conjugacy We make some additional assumptions to further simplify the problem. Firstly, we suppose each data point is realisation from a Gaussian distribution with common variance, i.e. $y(t_i)|x(t_i), \tau \sim \mathcal{N}(x(t_i), \tau)$. Secondly, we use the independent increment process as the prior distribution for $\boldsymbol{x}$, as discussed in Section 2.3.2. For completeness, we rewrite the full model in details as follows:

$$\boldsymbol{y}|\boldsymbol{x}, \tau \sim \prod_{i=1}^{n} \pi(y_i|x_i, \tau) = \mathcal{N}(\boldsymbol{x}, \mathbf{Q}_y^{-1}) \tag{2.7a}$$

$$\boldsymbol{x}|v \sim \prod_{i=2}^{n} \pi(x_i|x_{i-1}, v) = \mathcal{N}(\mathbf{0}, \mathbf{Q}_x^{-1}) \tag{2.7b}$$

$$\{\tau, v\} = \boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}) \tag{2.7c}$$

where $\mathbf{Q}_y$ is the precision matrix with τ^{-1}'s in the main diagonal and zeroes everywhere else. The matrix $\mathbf{Q}_x$ has been discussed in (2.4). In this modelling setting, both (2.7a) and (2.7b) are Gaussian distributions. Thus, the full conditional for $\boldsymbol{x}$ is exactly Gaussian, i.e. $\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{x|y}, \mathbf{Q}_{x|y}^{-1})$. Here, $\mathbf{Q}_{x|y} = \mathbf{Q}_x + \mathbf{Q}_y$ and $\boldsymbol{\mu}_{x|y}$ is the solution

to the system $\mathbf{Q}_{x|y}\boldsymbol{\mu}_{x|y} = \mathbf{Q}_y\boldsymbol{y}$. Thus, the full conditional for $\boldsymbol{x}$ can be evaluated up to a normalising constant and $\boldsymbol{x}$ can be analytically marginalised out of the joint distribution $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$.

Case 2: non-Gaussian components If the distribution of $\boldsymbol{y}|\boldsymbol{x}, \tau$ is non-Gaussian and $\boldsymbol{x}|v$ is Gaussian, approximation methods can be used. More specifically, we can write the full conditional distribution for $\boldsymbol{x}$ as $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$, where

$$\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}) \propto \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta}) \propto |\mathbf{Q}|_x^{1/2} \exp\Big(\sum_{i=1}^{n} \log \pi\big(y(t_i)|x(t_i), \tau\big) - \frac{1}{2}\boldsymbol{x}^T\mathbf{Q}_x\boldsymbol{x}\Big) \tag{2.8}$$

The first term inside the exponential can be written as as a Taylor expansion to the second order (Tierney & Kadane 1986). Thereafter (2.8) can be approximated as a Gaussian distribution by iteratively matching its mode and curvature at the mode (Rue & Held 2005, Section 4.4.1). Once the full conditional for $\boldsymbol{x}$ has been approximated as Gaussian, analytical integration is equivalent to case 1 described above.

Another approximation approach that we use throughout this thesis - most explicitly in Chapter 5 - follows the recent work of Parnell et al. (2015). They propose to replace $\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y}) = \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ in (2.7a) with the so-called '*marginal data posteriors*' (MDPs) of the form $\pi_{MDP}(\boldsymbol{x}|\boldsymbol{y}) = \prod_i^n \pi\big(x(t_i)|y(t_i)\big)$. In the current model setting, the computation of each MDP is:

$$\pi\big(x(t_i)|y(t_i)\big) \propto \pi\big(y(t_i)|x(t_i)\big)\pi\big(x(t_i)\big) \approx \pi\big(y(t_i)|x(t_i)\big) \tag{2.9}$$

The last equality in the above equation is possible since $\boldsymbol{x}$ is an independent increment process and $\pi\big(x(t_i)\big)$ is flat for all i. Although this is what we use here, any model choice for $\boldsymbol{x}$ would be as suitable. The computation of (2.9) also requires evaluation of the integral $\pi(y_i|x_i) = \int_\tau \pi(y_i|x_i, \tau)\pi(\tau|\boldsymbol{y})\, d\tau$ hence MDPs are typically non-Gaussian. To remedy this, Parnell et al. (2015) propose to approximate the non-Gaussian MDPs as Gaussian mixtures. For the sake of simplicity in this introduction of the method, we assume τ is a known constant. Consequently, $\pi_{MDP}(\boldsymbol{x}|\boldsymbol{y})$ is exactly Gaussian and marginalisation of $\boldsymbol{x}$ from the joint posterior distribution is the same as in the conjugate Gaussian case described above, i.e.

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi_{MDP}(\boldsymbol{x}|\boldsymbol{y})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \propto \pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{y})$$

where $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ is available, up to a normalising constant, as a Gaussian distribution.

The MDP approach can be extended to the case where both $\boldsymbol{y}|\boldsymbol{x},\tau$ and $\boldsymbol{x}|\upsilon$ are non-Gaussian. We refer to a more detailed discussion of this in Parnell et al. (2015) and again in Chapter 5.

2.3.3.2 Simulation-free inference

From (2.6), we can write down the identity

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \frac{\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})} \tag{2.10}$$

When both $\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}$ and $\boldsymbol{x}|\boldsymbol{\theta}$ are Gaussian distributions, conjugacy implies that the full conditional distribution for $\boldsymbol{x}$ is exactly Gaussian as discussed in Case 1 in Section 2.3.3.1. Moreover, $\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta}) = \pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{\theta})$ in the numerator on the RHS of (2.10). Thus, we can analytically evaluate $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ followed by $\pi(\boldsymbol{\theta}|\boldsymbol{y})$.

Provided the dimension of $\boldsymbol{\theta}$ is not too high, Eq. (2.10) can be efficiently evaluated on a parameter grid. Several grid search strategies are proposed in the recently developed '*integrated nested Laplace approximation*' method (INLA Rue et al. 2009). In what follow we review a strategy that we use in Chapter 4. It begins by optimising $\log \pi(\boldsymbol{\theta}|\boldsymbol{y})$ to locate the mode $\hat{\boldsymbol{\theta}}$ and the Hessian matrix $\mathbf{H}$ evaluated at $\hat{\boldsymbol{\theta}}$. The matrix $\mathbf{H}$ is asymptotically the precision matrix for $\hat{\boldsymbol{\theta}}$. Thus $\boldsymbol{\Sigma} = \mathbf{H}^{-1}$ is the correspondent covariance matrix. The mode and covariance are used as a guide to search for the parameter space of interest. Moreover, INLA recommends exploring $\boldsymbol{\theta}$ via the standardised variable $\boldsymbol{z}$:

$$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + \mathbf{V}\boldsymbol{\Delta}^{1/2}\mathbf{z} \tag{2.11}$$

Here, $\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^{-1}$ is the eigen-decomposition. The search begins from the mode (i.e. $\boldsymbol{z} = \mathbf{0}$) and proceed to all combination of directions according to a stopping rule; see Martino (2007, Algorithm 3) for further technical details.

At the end of a grid search, we obtain the discrete parameter space $\boldsymbol{\Theta}_J; j = 1, \ldots, J$ where J is the total number of grid points. This parameter space 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. $\tilde{\pi}(\boldsymbol{y}) \approx \sum_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_J} \pi(\boldsymbol{y}|\boldsymbol{\theta}_j)\,\pi(\boldsymbol{\theta}_j)\,\triangle\boldsymbol{\theta}_j$ given appropriate weights $\triangle\theta_j$. In our work, we choose $\boldsymbol{z}$ to have regular steps therefore we always have equal weights.

To obtain the marginal posterior distributions such as those discussed in (2.5e), we numerically integrate over the gridded parameters as follows

$$\tilde{\pi}(x_k|\boldsymbol{y}) \approx \sum_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_J} \pi(x_k|\boldsymbol{y}, \boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j|\boldsymbol{y}) \triangle\boldsymbol{\theta}_j \tag{2.12}$$

Evaluation of $\pi(\boldsymbol{x}_k|\boldsymbol{y}, \boldsymbol{\theta})$ is least challenging when $\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}$ is Gaussian, to which each marginal is itself an exact Gaussian distribution. When $\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}$ is not Gaussian, Gaussian approximation techniques can be applied to the denominator of (2.10). This gives rise to the approximated distribution $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ where:

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \approx \left.\frac{\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\theta)} \tag{2.13}$$

Here, $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional for $\boldsymbol{x}$ evaluated at $\boldsymbol{\theta}$, and the approximated conditional distribution $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ is the same as that discussed in (2.8). The expression (2.13) is the main ingredient of INLA in which the Gaussian distribution is used to marginalise out latent processes of interest. In some settings, it is also possible to use the MDP approach as discussed in Case 2 of Section 2.3.3.1.

It is sometimes of interest to evaluate the marginal posterior distribution for the parameters, as described in (2.5c) and (2.5d). These approximated marginal posterior distributions are:

$$\tilde{\pi}(\tau|\boldsymbol{y}) \approx \int \boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{y})\, dv \tag{2.14a}$$

$$\tilde{\pi}(v|\boldsymbol{y}) \approx \int \boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{y})\, d\tau \tag{2.14b}$$

where $\boldsymbol{I}$ is the interpolated function of $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ already computed during the grid exploration stage; see Martins et al. (2013, Section 3) for the technical details. Alternatively, when the dimension of $\boldsymbol{\theta}$ is not too high, we can evaluate its joint posterior distribution on a regular grid and then use the resulting values in place of $\boldsymbol{I}$ in Eq. (2.14) above. This approach requires the grid to be finer and wider. Hence, the disadvantage is that numerical integration is slower than that based on the the interpolated function. The advantage is a more accurate representation of the marginal posterior distributions.

Simulation-free Bayesian inference has been shown to work well with a broad class of models (Rue et al. 2009, Martins et al. 2013). However, it has many limitations. Most notably, the dimension of $\boldsymbol{\theta}$ must be small in order for quadrature integration in (2.14) to be feasible. In chapter 4, the maximum number of parameters are 4 hence we make

extensive use of simulation-free inference techniques. However, there are hundreds of unknown parameters in the models in in chapter 5. In this case, simulation-based methods can be a better alternative solution. We review this approach in the following subsection.

2.3.3.3 Simulation-based inference

The aim of simulation-based inference, or Monte Carlo methods, is to draw sample values from $(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$ and approximate the characteristics of the target distribution using the characteristics of the samples.

When the posterior distributions are high dimensional, direct simulation techniques (e.g. rejection sampling) are not a feasible solution. Markov chain Monte Carlo (MCMC) is an approximated sampling technique. The objective is to build a Markov chain that converges to the desired target distribution. It works by producing a chain of samples where each draw is dependent on the previous draw. The two basic algorithms used in MCMC are the Gibbs and Metropolis Hastings sampler.

Gibbs sampling is by far the most popular approach, mostly because of its versatility. The widespread use of the Gibbs sampler for general Bayesian problems is arguably due to Gelfand & Smith (1990)'s influential paper, see, for example, Robert & Casella (2011) and Tanner & Wong (2010) for a historical perspective on this. The algorithm makes use of full conditional distributions. The basic concept is simple. First, we divide all unknown processes and parameters into separate components. Next, we derive the full conditionals for each component given all other components and data. Then, we iteratively sample from each full conditional distribution. A generic Gibbs sampler in the context of this chapter proceeds as follows:

1. Choose starting values for $\boldsymbol{\theta}$ and $\boldsymbol{x}$.

2. Sample from $\pi(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y}) \coloneqq \pi(\boldsymbol{\theta}|\boldsymbol{y})$ or its marginals, $\pi(\tau|\boldsymbol{y})$ and $\pi(v|\boldsymbol{y})$.

3. Sample from $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ or its marginals, $\pi(x_k|\boldsymbol{\theta}, \boldsymbol{y})$ for all k.

4. Repeat step 2 and 3 until convergence in all relevant distributions is achieved.

Gibbs is a special case of Metropolis-Hasting (MH) rejection sampling algorithm. The MH algorithm is arguably the most general form of Markov chain sampling technique

(Chib & Greenberg 1995). It is useful when the full conditionals do not belong to any standard distribution, and samples from them are difficult to obtain. Therefore, rather than sampling directly from the full conditional, a candidate value from an arbitrary proposal distribution is drawn and accepted with a certain probability. Like the Gibbs sampler, the MH sampler updates the parameters component-by-component. For our simple example, to sample from $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, we draw new samples $\boldsymbol{\theta}^*$ from a proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$, then decide to accept/reject them with probability $\alpha(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ where

$$\alpha(\boldsymbol{\theta}^*, \boldsymbol{\theta}) = \min\left(1, \frac{\pi(\boldsymbol{\theta}^*|\boldsymbol{y})q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}|\boldsymbol{y})q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}\right)$$

A combined usage of both the Gibbs and MH sampler is the Metropolis-within-Gibbs sampler. Like in the Gibbs sampler, parameters are updated component-by-component. However, components with a non-standard full conditional are updated using a MH step. This combination is the main ingredient of our MCMC algorithm in chapter 5.

In theory, MCMC can provide nearly exact inference for the target posterior distributions, given perfect convergence. Convergence and mixing of a sampler are heavily influenced by model parameterisation, choice of the proposal densities and updating schemes. For an in depth review of advanced techniques for improving MCMC, see, for example, Brooks et al. (2011). Of course, convergence can also be achieved by running Markov chains for long periods of time. Unfortunately, we have finite running time in practice, and this must be balanced with model complexity.

A special hierarchical framework is the class of state-space models. In this setting, (2.7a) and (2.7b) can be written as the observation equation and state equation respectively. The problems are divided into state inference and parameter inference. In our example, after marginalisation of the parameters $\boldsymbol{\theta}$ from $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$, the state inference problem for $\pi(\boldsymbol{x}|\boldsymbol{y})$ has a closed form solution known as the Kalman filters and/or smoothers (Särkkä 2013, chapter 4.3). Filtering only estimates the current state of $\boldsymbol{x}$ given the history of observations $\boldsymbol{y}$, smoothing is the reconstruction of $\boldsymbol{x}$ up to and include the current time. If the observation and state equations are non-linear and non Gaussian, Kalman filters/smoothers are not appropriate. Sequential Monte Carlo (SMC) methods or particle filters can often be a better alternative, particularly if we want to perform sequential inference. Recently, Bhattacharya & Wilson (2014) shows that it is possible to perform simulation-free inference for the parameter process $\boldsymbol{\theta}$ sequentially. We do not consider this problem in this thesis.

2.3.3.4 Computational aspects related to inference with Gaussian Markov processes

As previously stated, sparse precision matrices play a key role for the models in this thesis. Many of the fast numerical algorithms for this type of matrix are associated with efficient Cholesky factorisation (Rue & Held 2005, Chapter 2.3 and 2.4).

Cholesky factorisation refers to the procedure of decomposing a n-by-n symmetric positive definite matrix $\mathbf{Q}$ as the product of a unique lower triangular matrix $\mathbf{L}$ and upper triangular matrix $\mathbf{L}^T$, i.e. $\mathbf{Q} = \mathbf{L}\mathbf{L}^T$. For instance, one way to sample $\boldsymbol{x}$ from the distribution $\mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$ is to compute $\mathbf{L}$ from $\mathbf{Q}$, generate a vector $\boldsymbol{z}$ of n standard Normal random variates and solve $\mathbf{L}^T\boldsymbol{x} = \boldsymbol{z}$. The computational complexity of a full Cholesky for a *dense* matrix is generally of the order $O(n^3/3)$. To solve a system of linear equations, the cost is $O(n^2)$. Thus, the overall cost for sampling from a Gaussian process with a mean zero and dense precision matrix is of cubic order.

If $\mathbf{Q}$ is *sparse* then the computation burden is typically lower. The main idea is to determine the zero entries in $\mathbf{L}$ and do not carry out computation for them, for instance, using Theorem 2.8 in Rue & Held (2005). A special case of sparse matrices is *band matrix*. If $\mathbf{Q}$ is a band matrix, the cost to sample from $\mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1}_{x|y})$ includes $O(nb^2 + 3nb)$ for the band-Cholesky decomposition and $O(2nb)$ for solving the system of linear equations. Here, b denotes the bandwidth. Hence, the cost is linear in n. This represents a huge speed-up in comparison to sampling from the distribution involving a dense precision matrix.

In several places of this thesis, we need to compute the marginal variances from a precision matrix. Although we can perform matrix inversion, i.e. $\mathbf{\Sigma} = \mathbf{Q}^{-1}$, this is often not necessary. In what follows we provide a summary of the more efficient approach discussed in (Rue & Martino 2007, Section 2). Firstly, we need to perform Cholesky decomposition to obtain the lower triangular matrix $\mathbf{L}$ from $\mathbf{Q}$. Secondly, we use the *sequential* representation when solving $\mathbf{L}^T\boldsymbol{x} = \boldsymbol{z}$. Within the second step, the starting point is $x_n = z_n/L_{nn}$ or equivalently, $x_n \sim \mathcal{N}(0, 1/L^2_{nn})$ where x_n and z_n are the n^{th} element of vectors $\boldsymbol{x}$ and $\boldsymbol{z}$ respectively, and L_{nn} is the n^{th} row and n^{th} column of matrix $\mathbf{L}$. Then,

$$x_i = \frac{1}{L_{nn}}\left(z_i - \sum_{k=i+1}^{n} L_{ki}x_k\right), \qquad i = n-1, \ldots, 1$$

Next, multiply both sides of the above equation by x_j and take expectation yields

$$\Sigma_{ij} = \frac{\delta_{ij}}{L_{ii}^2} - \frac{1}{L_{ii}} \sum_{k \in \mathcal{I}(i)}^{n} L_{ki}\Sigma_{kj}, \qquad i = n-1, \ldots, 1$$

Here, Σ_{ij} is the i^{th} row and j^{th} column of covariance matrix $\mathbf{\Sigma}$, $\delta_{ij} = 1$ if $i = j$ and zero otherwise, and $\mathcal{I}$ are those j where L_{ji} is non-zero.

2.4 Summary remarks

As a result of this chapter, it became evident that new model formulations and computational methods were required for statistical inference of multiple palaeoclimate data sets. We have reviewed some ideas associating with the Bayesian approach to statistical inference. More specifically, we focused on hierarchical modelling, with an emphasis on Gaussian Markov processes. Some methods for parameter inference have also been explored, namely MDP, INLA and MCMC. In subsequent chapters, we will engage in a detailed study of some specific models and computational methods for misaligned irregular time series.

Chapter 3

Toy example

This chapter marks the first contribution of the thesis by laying essential groundwork for the research by introducing terminology upon which later chapters are based. To elucidate the narrative in the chapter, our discussion will be accompanied by a simple model applied to a simulated data set.

The rest of the chapter is organised as follows. Formulation of a simple model for two irregular spaced time series is presented in Section 3.1. Issues with regard to stochastic interpolation are discussed in Section 3.2. In particular, we focus on the usefulness and limitations of both pointwise and pathwise posterior summaries. Section 3.3 highlights the benefit of our joint approach in comparison to the one-series-at-a-time approach, using predictive variance as the metric. The chapter concludes with a discussion in Section 3.4, where we summarise the proposed framework and outline the possible extensions to be developed in subsequent chapters.

3.1 A simple model formulation

In this section, a simple hierarchical modelling structure is set up. Technical notation will be introduced to deal with temporal misalignment in the times of observations across different series. We also introduce posterior distributions of interest.

3.1.1 A two-stage hierarchical statistical model with known parameters

A simple hierarchical model comprised of two main layers is adopted to illustrate a joint modelling approach for multiple irregular time series. We suppose $y(t_{s,i}); i = 1, \ldots, n_s$ denote an observation of series s at time t_i; as a vector, it is written as $\boldsymbol{y}_s$. There are $N = \sum_s n_s$ observations in total, and we write them as $\boldsymbol{y}$. We label the series as $s = 1, \ldots, m$. At the *data layer*, $y(t_{s,i})$ are an conditionally independent Gaussian function of the latent variable $x(t_{s,i})$ with constant variance τ, i.e.

$$y(t_{s,i}) \sim \mathcal{N}(x(t_{s,i}), \tau) \tag{3.1}$$

We consider a *process layer* with the form:

$$\begin{pmatrix} x(t_{1,i}) - x(t_{1,i-1}) \\ \vdots \\ x(t_{m,i}) - x(t_{m,i-1}) \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, v|t_i - t_{i-1}|\boldsymbol{\Sigma} \right] \tag{3.2}$$

where v is the common variance of an increment per unit time, and $\boldsymbol{\Sigma}$ is the m-by-m matrix which controls the strength of the relationship of data across the series. The latter is a unit correlation matrix with ones in the main diagonal and ρ's in the off-diagonals. Hence $\boldsymbol{x}$ is a time-continuous, multivariate, independent increments process. For simplification, hereafter we drop the i subscript when discussing latent processes since they are defined for all times. In vector form, we write the multivariate latent process for all series at all times as $\boldsymbol{x}$, and $\boldsymbol{x}(t_s)$ represents the process for series s at time t only. The Markov property implied by this continuous-time stochastic process is crucial for computational reasons. Most importantly, $\boldsymbol{x}$ is *a priori* a joint model for all series.

To keep the presentation of a toy model in this chapter simple, we assume all parameters $\boldsymbol{\theta} = (v, \tau, \rho)$ to be known, and defer issues with parameter inference to later chapters. Figure 3.1 shows a graphical representation of the full model.

There are many choices of models for (3.1) and (3.2). The hierarchical modelling approach is simple yet powerful as it explicitly separates the former from the latter. In Section 3.4 we discuss possible extensions to this simple structure to deal with more complicated situations.

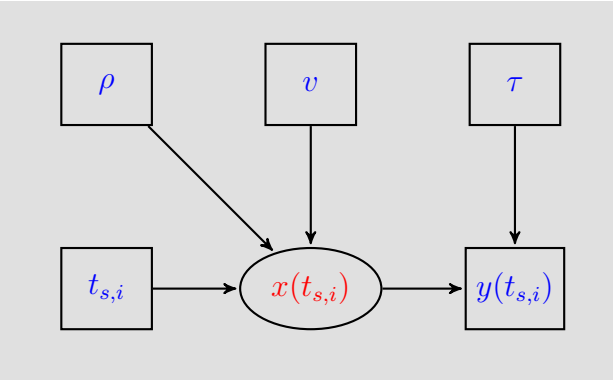


Fig. 3.1: A Directed Acyclic Graph (DAG) for the model described in Section 3.1. Circles indicate parameters/latent random variables whilst boxes indicate data.

3.1.2 Notation tricks associated with temporal misalignment

Subsequent mathematical derivations are substantially simplified by defining

1. $\boldsymbol{t_o}$ as the sorted set of all observed times at all series, i.e. $\boldsymbol{t_o} = \text{sort}\{t_{s,1}, t_{s,2}, \ldots\} = \{t_1, t_2, \ldots, t_n\}$ where n is the total number of unique times across all series. In our model, the multivariate stochastic process $\boldsymbol{x}$ is considered at these times.

2. $y_o(t_{s,i})$ as the values which coincide with the $y(t_{s,i})$ values when series s has an observation at t_i and whose values are missing otherwise. There is typically one observation at each t_i, and more than one observation if t_i is non-unique. Let $\boldsymbol{y_o}$ denote the set of all such vectors. Similarly, we let $\boldsymbol{x_o}$ be the latent multivariate process $\boldsymbol{x}$ defined on $\boldsymbol{t_o}$. Hereafter, we write $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{t_o}, \boldsymbol{\theta}) = \pi(\boldsymbol{y_o}|\boldsymbol{x_o}, \boldsymbol{\theta})$, $\pi(\boldsymbol{x}|\boldsymbol{t_o}, \boldsymbol{\theta}) = \pi(\boldsymbol{x_o}|\boldsymbol{\theta})$, and $\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{t_o}, \boldsymbol{\theta}) = \pi(\boldsymbol{x_o}|\boldsymbol{y_o}, \boldsymbol{\theta})$.

The re-writing of these terms is simply a technical 'notational trick'; the vectors $\boldsymbol{y}$ (length N) and $\boldsymbol{y_o}$ (length mn) contain the same information. The objective is to use this information, together with the hierarchical model to infer the underlying continuous-time multivariate stochastic process $\boldsymbol{x}$ with appropriate prediction intervals.

Using the above definitions, some of the equations discussed in Section 3.1.1 can be rewritten in their compact matrix form. For instance we can write (3.1) as $\boldsymbol{y_o}|\boldsymbol{x_o}, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{x_o}, \mathbf{Q}_{y_o}^{-1})$ where $\mathbf{Q}_{y_o}$ is a mn-by-mn diagonal precision matrix with entries τ^{-1} corresponding to series at times with data, and zeros where there are no data. This is

the notation trick previously introduced to write all vectors and matrices in clear order of time and series identity. We emphasise that no computation is ever necessary for components that do not represent data.

Similarly, (3.2) has the form $\boldsymbol{x_o}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{x_o}^{-1})$. Here, $\mathbf{Q}_{x_o} = \mathbf{Q}_{t_o} \otimes v^{-1}\boldsymbol{\Sigma}^{-1}$, is an inverse of the mn-by-mn block diagonal covariance matrix. This covariance matrix is a Kronecker product of the inverse of the n-by-n precision matrix $\mathbf{Q}_{t_o}$ and the m-by-m cross-covariance matrix $v\boldsymbol{\Sigma}$. In particular, $\mathbf{Q}_{t_o}$ is the precision matrix of an independent increment process for an irregularly spaced time series as discussed in Section 2.3.2. That is,

$$\mathbf{Q}_{t_o} = \begin{pmatrix} |t_2 - t_1|^{-1} & -|t_2 - t_1|^{-1} & & \\ -|t_2 - t_1|^{-1} & |t_2 - t_1|^{-1} + |t_3 - t_2|^{-1} & -|t_3 - t_2|^{-1} & \\ & & \ddots & \\ & & -|t_n - t_{n-1}|^{-1} & |t_n - t_{n-1}|^{-1} \end{pmatrix} \tag{3.3}$$

with the blank entries being zeroes.

3.1.3 Posterior distributions

The full posterior distribution for $\boldsymbol{x_o}$ is

$$\begin{aligned} \pi(\boldsymbol{x_o}|\boldsymbol{y_o}, \boldsymbol{\theta}) &\propto \pi(\boldsymbol{y_o}|\boldsymbol{x_o}, \boldsymbol{\theta})\pi(\boldsymbol{x_o}|\boldsymbol{\theta}) \\ &\propto |\mathbf{Q}_{y_o}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{y_o} - \boldsymbol{x_o})^T \mathbf{Q}_{y_o}(\boldsymbol{y_o} - \boldsymbol{x_o})\right) |\mathbf{Q}_{x_o}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{x_o}^T \mathbf{Q}_{x_o} \boldsymbol{x_o}\right) \\ &\propto \mathcal{N}\left(\boldsymbol{x_o}; \boldsymbol{\mu}_{x_o|y_o}, \mathbf{Q}_{x_o|y_o}^{-1}\right) \end{aligned} \tag{3.4}$$

where $|\mathbf{Q}_{y_o}| = \prod_{\tau \neq 0}^{mn} \tau = \tau^N$, i.e. it is a generalized determinant as a result of the notational trick introduced in the previous section. Furthermore, $|\mathbf{Q}_{x_o}| = |v\boldsymbol{\Sigma}|^{-(n-1)}$, $\mathbf{Q}_{x_o|y_o} = \mathbf{Q}_{y_o} + \mathbf{Q}_{x_o}$, and $\boldsymbol{\mu}_{x_o|y_o}$ is the solution to the system of equation $\mathbf{Q}_{x_o|y_o}\boldsymbol{\mu}_{x_o|y_o} = \mathbf{Q}_{y_o}\boldsymbol{y_o}$. Note that all of these quantities depends on $\boldsymbol{\theta}$. But, for clarity, this dependence has been suppressed in the notation.

Ultimately, the main goal is to compute posterior distributions of latent processes of interest at new times. In this thesis, we consider new times on grids that are regular and denoted by $\boldsymbol{t_g} = \{i\Delta; i = 1, \ldots, n_g\}$. We refer to the latent process $\boldsymbol{x}$ and data process $\boldsymbol{y}$ defined on the time grid $\boldsymbol{t_g}$ as $\boldsymbol{x_g}$ and $\boldsymbol{y_g}$ respectively. The posterior distribution of

$\boldsymbol{x_g}$ conditional on all data and model parameters is:

$$\begin{aligned}\pi(\boldsymbol{x_g}|\boldsymbol{y_o}, \boldsymbol{\theta}) &\propto \int \pi(\boldsymbol{x_o}, \boldsymbol{x_g}|\boldsymbol{y_o}, \boldsymbol{\theta})\, d\boldsymbol{x_o} \\ &\propto \int \pi(\boldsymbol{x_g}|\boldsymbol{x_o}, \boldsymbol{\theta})\pi(\boldsymbol{x_o}|\boldsymbol{y_o}, \boldsymbol{\theta})\, d\boldsymbol{x_o} \end{aligned} \tag{3.5}$$

The second quantity inside the integrand in (3.5) is a Gaussian distribution, as derived in (3.4). Furthermore, the process $\boldsymbol{x}$ is *a priori* a multivariate independent increment process. Thus, the joint prior distribution $\{\boldsymbol{x_o}^T, \boldsymbol{x_g}^T\}^T$ is Gaussian. Consequently, the conditional distribution that is the first quantity in the integrand in (3.5) is also a Gaussian distribution. In this setting, this term is generally known as the multivariate *Brownian bridge*, a tied-down Gaussian process (Glasserman 2003, Chapter 3). The idea of 'bridging' will be useful for the evaluation of the high-dimensional integral (5.22) in Chapter 5. Relevant technical detail of Brownian bridge and the extension considered in Chapter 5 are given in Appendix C.

For the problem in this chapter (and for a similar problem in Section 4.2.3 in Chapter 4), computation of (3.5) can in fact be simplified via the notational trick discussed in Section 3.1.2. We let the star notation denote the processes defined at both the unique (and sorted) observed times and grid, i.e. $\boldsymbol{y_*} = \left(\boldsymbol{y_o}^T, \boldsymbol{y_g}^T\right)^T$, $\boldsymbol{x_*} = \left(\boldsymbol{x_o}^T, \boldsymbol{x_g}^T\right)^T$ and $\boldsymbol{t_*} = \left(\boldsymbol{t_o}^T, \boldsymbol{t_g}^T\right)^T$. In our notation, these are vectors of length $m(n + n_g)$. The problem becomes that of calculating

$$\pi(\boldsymbol{x_*}|\boldsymbol{y_*}, \boldsymbol{\theta}) \propto \pi(\boldsymbol{y_*}|\boldsymbol{x_*}, \boldsymbol{\theta})\pi(\boldsymbol{x_*}|\boldsymbol{\theta}) \tag{3.6}$$

After which the joint posterior distribution of $\boldsymbol{x_g}$, conditional on all data and known parameters, is the subset of the joint conditional distribution given by Eq. (3.6). Computation of (3.6) is straightforward when both components on the RHS are Gaussian distributions. It begins with the completion of the quadratic form for the first two quantities to obtain a posterior mean vector and precision matrix. Specifically, the diagonal precision matrix $\mathbf{Q}_{y_*}$ in the conditional distribution of $\boldsymbol{y_*}|\boldsymbol{x_*}, \boldsymbol{\theta}$ can only have non-zero values at series and time indices where data are available. The precision matrix for *a priori* $\boldsymbol{x_*}$ is $\mathbf{Q}_{x_*} = \mathbf{Q}_{t_*} \otimes v^{-2}\boldsymbol{\Sigma}^{-1}$ with $\mathbf{Q}_{t_*}$ having the same form as (3.3) adapted to the new selection of points. By completing the quadratic form we again have, up to a normalising constant, $\boldsymbol{x_*}|\boldsymbol{y_*}, \boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{x_*}; \boldsymbol{\mu}_{x_*|y_*}, \mathbf{Q}^{-1}_{x_*|y_*}\right)$, where $\mathbf{Q}_{x_*|y_*} = \mathbf{Q}_{y_*} + \mathbf{Q}_{x_*}$ and $\boldsymbol{\mu}_{x_*|y_*}$ is the solution to $\mathbf{Q}_{x_*|y_*}\boldsymbol{\mu}_{x_*|y_*} = \mathbf{Q}_{y_*}\boldsymbol{y_*}$.

3.2 Simulation study

In Section 3.2.1 we generate two irregularly-spaced time series based on the model described in Section 3.1.1. The *joint* conditional distribution of $\boldsymbol{x_g}$, conditional on all relevant data and parameters, is the prediction target in Section 3.2.2. In Section 3.2.3, the focus is on *marginalised* joint conditional distributions. Section 3.3 compares and contrasts the properties of the two forms of prediction in the context of data products previously discussed in Section 2.1.3.

3.2.1 A simulated data set

First, we generate a bivariate time series, each with 100 data points based on a discrete bivariate independent increment models with Gaussian noise. We emphasise that the underlying process is a time-continuous stochastic process. The simulation is controlled by three parameters: a variance of a unit increment $v = 1$, a two-by-two cross-correlation matrix with one in the main diagonal and a correlation coefficient $\rho = 0.8$ elsewhere, and variance parameter $\tau = 0.01$ for all Gaussian noise term.

Then, we randomly choose 30 and 20 data points from each series, so that they can be irregularly spaced in time, and when considered together the times are necessarily misaligned. For the toy example in this chapter, $m = 2$, $n_1 = 30$, $n_2 = 20$, and $N = 50$. The simulated data set (hereafter denoted as $\boldsymbol{a}$ and $\boldsymbol{b}$) is shown in Fig. 3.2.

3.2.2 Joint posterior distributions and pathwise summaries

As previously indicated, we defer to later chapters issues of uncertainty that arise when the model parameters $\boldsymbol{\theta}$ are themselves only available through statistical inference. In what follows we focus on inference of $\boldsymbol{x_g}$, which is the process $\boldsymbol{x}$ defined on the grid $\boldsymbol{t_g} = \{0, 2, 4, 6, 8, 10\}$.

In Section 3.1.3, we showed that $(\boldsymbol{x_*}|\boldsymbol{y_*}, \boldsymbol{\theta})$ is a Gaussian distribution. Hence it is easy to generate many samples of $\boldsymbol{x_*}$ conditioning on all data and parameters. The posterior samples for $\boldsymbol{x_g}$ can then be extracted from those for $\boldsymbol{x_*}$. We refer to each sample as a *history*. Two process histories on the time grid $\boldsymbol{t_g}$ are depicted in Fig. 3.3. These sample values are also shown in Table 3.1 and discussed in Section 3.2.4. A history is a *pathwise* summary of the process $\boldsymbol{x_g}$; very many histories allow us to

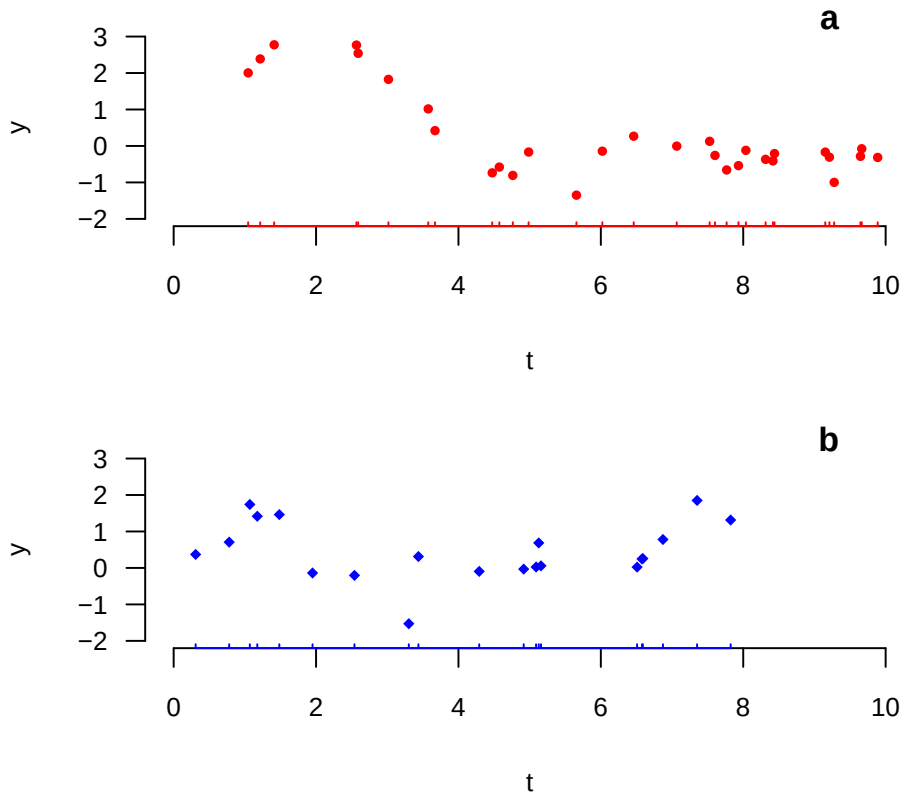


Fig. 3.2: An artificial data set used as an example throughout this chapter: two irregularly time series (**a** and ***b***) sampled from a bivariate independent increment process with a constant Gaussian noise.

compute *pointwise* summaries. We refer to Section 3.2.4 for further discussion of this.

		Sample 1	Sample 2	Mean
Time	0	-0.10	-0.68	-0.39
	2	-0.20	-0.40	-0.30
	4	0.41	-0.11	0.15
Minimum		-0.20	-0.68	
Time of minimum		2	0	

Table 3.1: Illustration of the histories representing the truth process which gives rise to series ***b*** on a time grid $\{0, 2, 4\}$, as shown in Fig. 3.3. An example of a linear functional (mean) is shown on the right most column, and some non-linear functionals (minimum and timing of minimum) are displayed in the last two rows.

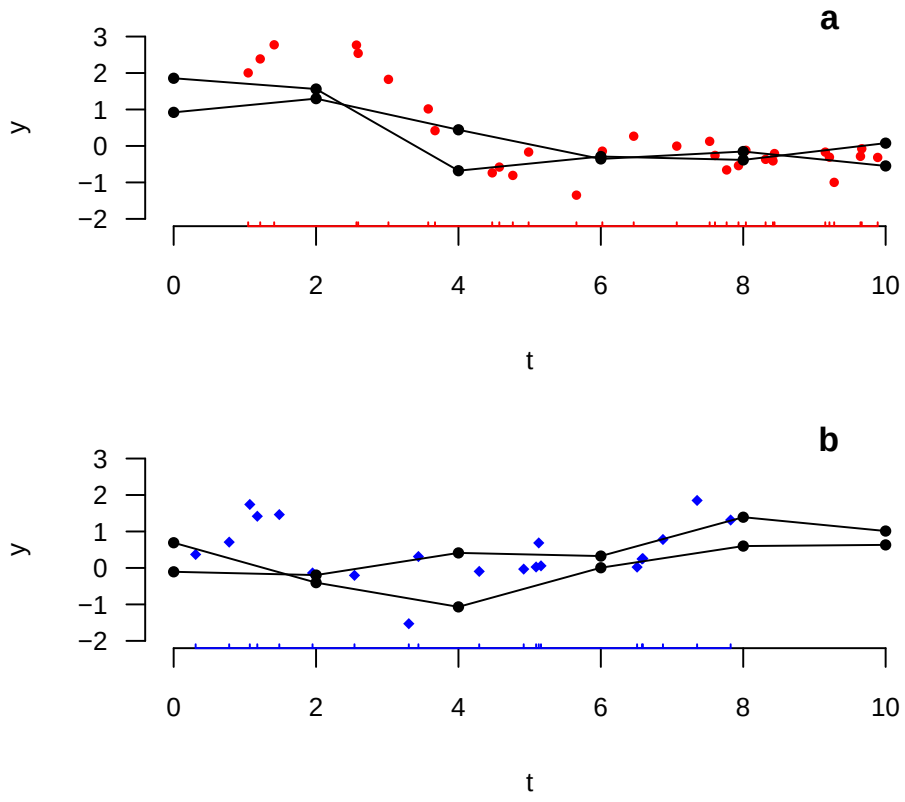


Fig. 3.3: Plots of two histories of $\boldsymbol{x}$ on a regular time grid {0,2,4,6,8,10}. They are generated from $\pi(\boldsymbol{x_g}|\boldsymbol{y}, \boldsymbol{\theta})$, the joint posterior distribution conditional on all data and parameters.

3.2.3 Marginal posterior distributions and pointwise summaries

We recall that the *joint* posterior distribution of $\boldsymbol{x}_*$ conditioning on all relevant data and parameters is a Gaussian distribution. It follows that the *marginal* posterior distribution of the l^{th} element at a specific time series corresponding to a time point is also a Gaussian distribution, i.e.

$$\pi(x_*^{(l)}|\boldsymbol{y_*}, \boldsymbol{\theta}) = \mathcal{N}\left(x_*^{(l)}; \mu_{x_*|y_*}^{(l)}, \tau_{x_*|y_*}^{(l)}\right) \tag{3.7}$$

where $\mu_{x_*|y_*}^{(l)}$ is the l^{th} element of $\boldsymbol{\mu}_{x_*|y_*}$ at the corresponding series. Similarly, each conditional posterior variance $\tau_{x_*|y_*}^{(l)}$ is the l^{th} element of a time series represented in the diagonal of the covariance matrix. As discussed in Section 2.3.3.4, marginal variance terms can be computed efficiently from the precision matrix $\mathbf{Q}_{x_*|y_*}$ without having to perform matrix inversion.

The predictive distribution in equation (3.7) is a source of the data product dis-

cussed in Section 2.1.3. Various posterior summaries (means, variances, modes, quantiles, etc.) can easily be computed based on (3.7). More crucially, all of the matrices herein are sparse and can be efficiently stored and quickly computed using fast algorithms for band matrices (Rue & Held 2005).

Figure 3.4 shows the posterior means and 95% credible intervals for $\boldsymbol{x_g}$ on a regular time grid $\boldsymbol{t_g} = \{0, 2, 4, 6, 8, 10\}$. The prediction errors, represented by the widths of the posterior marginals, are related to data availability at the corresponding series as well as the other series. We return to this in Section 3.3.

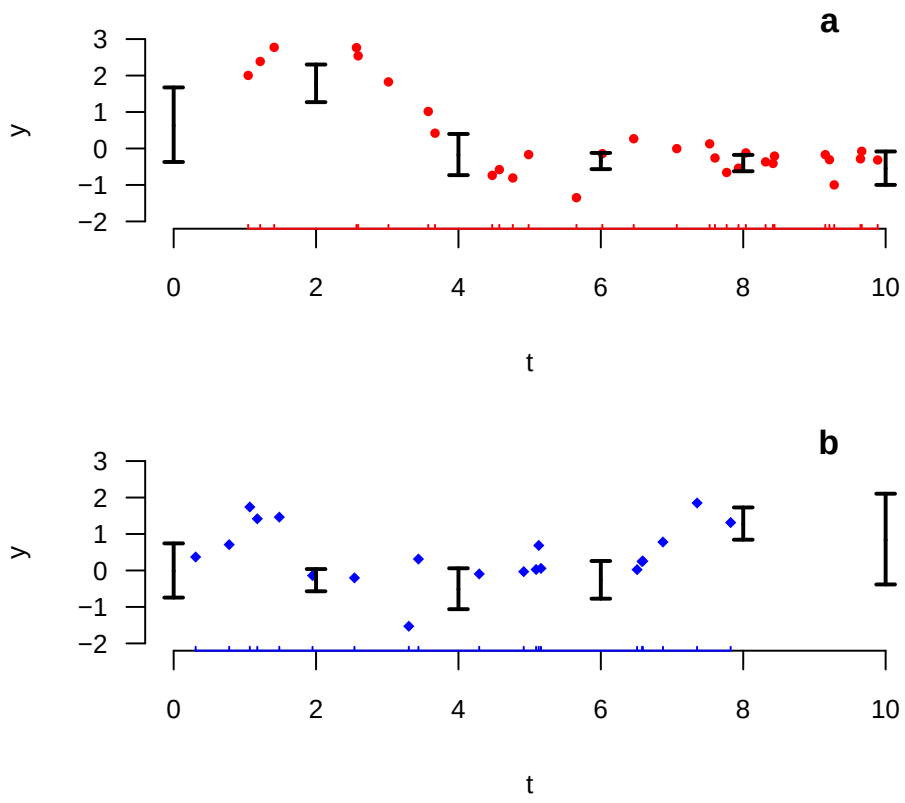


Fig. 3.4: Plots of 95% credible intervals of the marginal posterior distributions of process $\boldsymbol{x}$ on a regular time grid {0,2,4,6,8,10} conditioning on all data and parameters.

3.2.4 Comparison of interpolants

One of the advantages of pathwise over pointwise summaries is that researchers can use simulations from a joint posterior distribution to study any functionals of partially observed processes. Pointwise summarised statistics such as medians, modes, variances, quantiles, etc. 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 & 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. (3.7) are generally non-Gaussian when we take into account parameter uncertainties as considered in later chapters.

A greater challenge is posed by the investigations of non-linear functions of $\boldsymbol{x_g}$ such as minima and their timings. As pointed out, for instance, by Li et al. (2007) and Tingley & Huybers (2013), probability distributions of such random variables are generally not available from posterior marginals at each time point separately. Instead, simulations based on the joint posterior distribution (3.7) at all arbitrary time points provide a statistically consistent and easily interpreted probabilistic statement of any functionals of $\boldsymbol{x_g}$. Each such simulation represents an independent history of $\boldsymbol{x_g}$ that is statistically consistent with the model and data. As an example, we focus on the distributions of random variables $x_{\min} = \min_t x(t)$ and $t_{\min} = \arg_t \min x(t)$, being respectively the minimum value and the time at which this minimum was achieved. In Table 3.1, we illustrate how to compute a linear function (mean) and non-linear functions (minimum and timing of minimum) of $\boldsymbol{x_g}$ from two process histories at series $\boldsymbol{b}$ on a time grid $\{0, 2, 4\}$. These sampled histories were previously shown in Fig. 3.3). A more complete version of this approach uses 1000 histories, conceivably on a dense grid. For generality we compute conditional quantiles at time grid $\boldsymbol{t_g}$ and use pointwise credible intervals to quantify the uncertainty of our imputations.

Very many process histories thus provide a flexible posterior summary in their own right. The use above for mean, minima and time of minima is illustrative; any function of the process may be studied, conditional on the data. In this simple case, linear functions such as conditional means and variances are of course already available from the analytical expression in Eq. (3.7).

3.3 Joint versus separate inference

In this section we demonstrate the benefit of joint modelling in comparison to separate alternatives. Our comparison is evaluated as a function of the correlation coefficient ρ, keeping all other parameters fixed. First, we perform prediction for series $\boldsymbol{b}$ independently of $\boldsymbol{a}$. Then we carry out joint prediction using information from both $\mathbf{a}$ and $\boldsymbol{b}$ for a range of values of ρ to vary the strength of the association between the two series. Figure 3.3 illustrates that joint prediction reduces the predictive variability. The benefit of joint modelling is most pronounced in the extrapolation step from $t_g > 8$ where there is no observation from series $\boldsymbol{b}$ available.

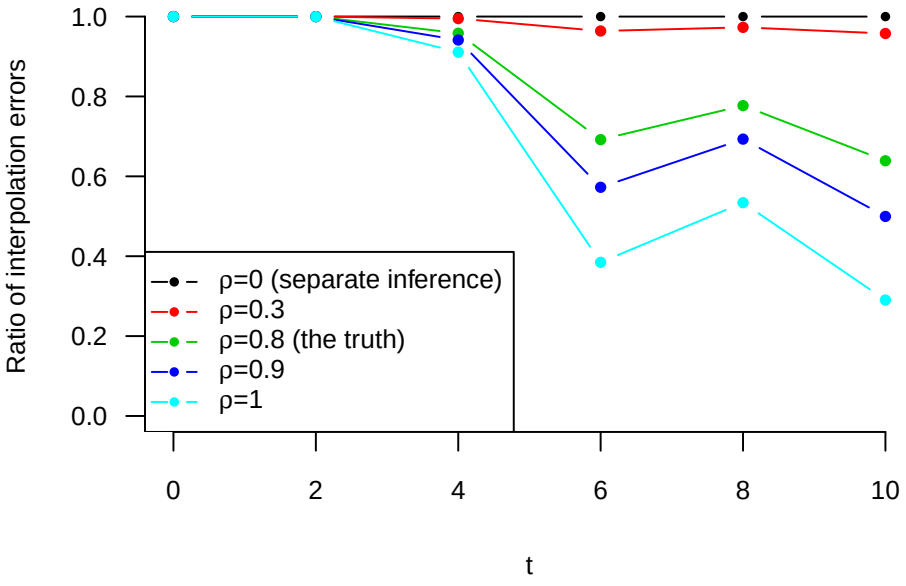


Fig. 3.5: Plot of the ratio of the interpolation errors in term of the correlation coefficient ρ. The ratio is computed by dividing the interquartile range of the marginal posterior distributions of $\boldsymbol{x}$ on a regular time grid $\{0, 2, 4, 6, 8, 10\}$ for series $\boldsymbol{b}$ conditioning on data from both of series $\boldsymbol{a}$ and $\boldsymbol{b}$, by that conditioning on only data from series $\boldsymbol{b}$ (i.e. when $\rho = 0$).

This experiment can be repeated for other parameters to learn about their effect on the predictive performance of the models under consideration. Even though our experiment is carried out on a restrictive class of model, the result is general. More specifically, it demonstrates that separate inference ignores valuable information by treating each series separately.

3.4 Summary and discussion of outstanding challenges for later chapters

This chapter has presented a hierarchical framework to perform joint inference of multiple irregular time series. An important component of our model is the Gaussian Markov assumption based on multivariate independent increments that provides a natural vehicle for joint modelling. We have shown that process histories, generated from joint posterior distributions, are extremely useful for prediction of both linear and non-linear functionals of partially observed processes of interest. We also demonstrated that joint analysis of multiple irregular time series is better than its independent alternative in the sense of lower predictive uncertainty.

The adopted framework has been shown to work well for a simple toy example. It forms the basis for extensions which are to be developed in later chapters of the thesis. In what follows, we summarise the model used in this chapter and announce the extensions to be considered in the following chapters. To keep the presentation simple, we leave out the notation trick associated with misalignment (e.g. $\boldsymbol{x_o}$, $\boldsymbol{y_o}$, etc.). Additionally, we defer the technical discussion of specific tools and techniques in the context of their usage to later chapters.

The model used in this chapter can be succinctly formulated as:

$$\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{t},\boldsymbol{\theta}) \propto \pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{t},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{t},\boldsymbol{\theta}) \tag{3.8}$$

Here, the underlying latent process $\boldsymbol{x}$ is the only unknown component.

In Chapter 4, the temporal supports for observations are different across different series. More generally, we treat parameters $\boldsymbol{\theta}$ as unknown; these parameters will be learnt from information in the model and data. That is,

$$\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{t}) \propto \pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{t})\pi(\boldsymbol{x}|\boldsymbol{t},\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{3.9}$$

The model in Chapter 5 will be the most complex. Firstly, $\boldsymbol{\theta}$ is unknown. Secondly, a new set of parameters $\boldsymbol{\psi}$ is introduced to capture the statistical relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$. These parameters are to be learnt from a separate set of training data $\boldsymbol{D_\psi}$. Thirdly, times $\boldsymbol{t}$ are uncertain; to be learnt from another calibrated data set $\boldsymbol{D_t}$. The model for this problem setting is:

$$\pi(\boldsymbol{x},\boldsymbol{t},\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y},\boldsymbol{D_t},\boldsymbol{D_\psi}) \propto \pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{t},\boldsymbol{\psi})\pi(\boldsymbol{D_\psi}|\boldsymbol{\psi})\pi(\boldsymbol{x}|\boldsymbol{t},\boldsymbol{\theta})\pi(\boldsymbol{D_t}|\boldsymbol{t})\pi(\boldsymbol{t})\pi(\boldsymbol{\theta})\pi(\boldsymbol{\psi}) \tag{3.10}$$

Chapter 4

Joint inference of multiple processes with different time supports[1]

Chapter 3 demonstrated that Bayesian hierarchical models are useful to model multiple irregular time series. Therein, all model parameters were assumed to be known.

In many real-world applications, model parameters are unknown. More crucially, data are often measured as the average over a certain support rather than a precise point in time or space. For example, the Earth's climate conditions as a physical system are defined over a continuous spatio-temporal domain but climate data (both its direct instrumental and indirect proxy measurements) are recorded and reported over a period of time for a region. This chapter proposes a simple extension to the model discussed in chapter 3 to deal with issues arising from different temporal supports for data in different time series.

A pair of ice cores drilled in Greenland will be used to illustrate our proposed framework. This dataset was previously introduced in Section 2.1.1. In particular, we consider the period of approximately the last 11 k cal yr BP. Our primary goal is to demonstrate the use of Bayesian inference to efficiently take into account multiple sources of information. Our secondary goal is to encourage the use of Monte Carlo samples from joint Bayesian posterior distributions as a rich type of data product.

In Section 4.1 we describe our hierarchical stochastic process model, and discuss an efficient procedure for inference of underlying latent processes. Section 4.2 gives an

[1] **A manuscript based on the work in this chapter is a publication in the journal of *Advances in Statistical Climatology, Meteorology and Oceanography* (Doan et al. 2015)**

efficient algorithm which comprises of two separate stages: a simulation-free procedure for the model parameters, and two computational procedures for latent process of interest. Application to Greenland ice cores are discussed in Section 4.3. Various aspects of our proposed model are placed under scrutiny in Section 4.4. A brief discussion in Section 4.5 concludes the chapter.

4.1 A three-layer hierarchical statistical model

4.1.1 Data layer

We consider a hierarchical model comprised of the data, process and parameter layer. At the data layer, $z(t_{s,i})$ are unobserved data in series $s = 1, \ldots, m$ at time $t_i; i = 1, \ldots, n_s$. When they are observed, the instrumentation is such that

$$y(t_{s,i}) = z(t_{s,i}) + \nu(t_{s,i}) \tag{4.1}$$

Here, the terms $y(t_{s,i})$ denote the observations. There are $N = \sum_s^m n_s$ observations in total; we use $\boldsymbol{y}$ to denote these, and $\boldsymbol{y}_s$ for observations in series s only. The terms $\nu(t_{s,i})$ are *i.i.d.*, zero mean, Gaussian random variables with fixed known variance τ_{ν_s} corresponding to the instrumentation. For simplification, hereafter we drop the i subscript when discussing latent processes since they are defined for all times.

4.1.2 Process layer and temporal change of support

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

$$z(t_s) = x(t_s) + w(t_s) \tag{4.2}$$

where $w(t_c)$ reflects micro-scale, annual-level variations and is modelled by a white noise process, independently across series, having common $\text{Var}\left[w(t_s)\right] = \tau_{w_s}$.

We propose to model process $x(t_s)$ as a continuous-time, independent increments process with increment variance such that for each series $\frac{1}{2}\text{Var}\left[x(t_s) - x(t_s - h)\right] = \frac{1}{2}v|h|$, i.e. having a linear semivariogram with no *nugget effect*. The nugget parameter refers to the apparent discontinuity at the beginning of a semivariogram. This effect 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 & Delfiner 2012, chapter 2). Of course, an independent increment prior model for $x(t_s)$ is one of many modelling choices. More specific to our context to be discussed in Section 4.3.1, the use of linear semivariogram (with nugget) is consistent with the generating process $y(t_s)$ hence the choice of independent increment model. More formally, our semivariogram is the sum of: (i) an underlying independent increments continuous-time process with semivariance proportional to lag; and (ii) a white noise process, manifest in the intercept or nugget effect which is dominated by τ_s. From (4.1) and (4.2), we obtain

$$y(t_s) = x(t_s) + w(t_s) + \nu(t_s) := x(t_s) + \epsilon(t_s) \tag{4.3}$$

Here $\epsilon(t_s) \overset{ind}{\sim} \mathcal{N}(0, \tau_s)$, with $\tau_s = \tau_{w_s} + \tau_{\nu_s}$, and the series-specific variance component τ_s is the annual-level nugget effect at series s. An explicit assumption here is that there exists an underlying latent process at a common level in both series that was driving the observational data. In particular, we consider the situation where the measurement processes differ for each series that lead to different supports in the observations. Mathematically,

$$z(t_s) = |t_s|^{-1} \int_{t_s} x(t)\, dt \tag{4.4}$$

where $|t_s|$ is an interval of time, i.e. the support, for which an observation is defined over. We have discussed the concept of support in the context of ice core data in Section 2.1.1. Therein, we suppose an ice core s is cut into equal, and non-overlapping sections of length l_s, i.e. we do not consider the case of overlapping sections. Moreover, we assume that the thickness of all sections are identical, and that a section of length l_s maps to a time interval t_s. In the present context, as a measurement is obtained within this time interval, we view the latter as the support of the former.

The chief implication of the difference in the supports of observations from different series is the relationship in the nugget effects. We show, in Appendix A, that the nuggets differ in proportion to the length of times that define the supports. We can reparameterise so that there is only one nugget term for one of the series (denoted as τ here), with the other having a different nugget which is a multiple of this value, i.e., $\tau_s = k_s\tau$ for a positive known series-specific value k_s based on the ratio of the measurement periods. We denote a vector of constants associated with the change in the temporal support of the measurements across different series as $\boldsymbol{k} = \{k_1, \ldots, k_m\}$.

The support transformation takes place in the observations, leaving the latent process to be modelled at another stage. In vector form, $\boldsymbol{x}$ is the multivariate latent process for all series at all times, and $\boldsymbol{x}(t)$ represents all the series at time t. The x increments at different series are allowed to be correlated if the series themselves reflect a physical process (e.g. climate) at nearby locations; we write

$$\boldsymbol{x}(t+h) - \boldsymbol{x}(t) \;\sim\; \mathcal{N}(\mathbf{0}, v|h|\boldsymbol{\Sigma}) \tag{4.5}$$

where v 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 series. The multivariate process in Eq. (4.5) forms the basic model underlying our joint approach.

4.1.3 Parameter layer

To complete the hierarchical modelling structure, prior distributions are assigned to the model parameters. We use reference priors on v and τ so that $\pi(v) \propto v^{-1}$ and $\pi(\tau) \propto \tau^{-1}$. We defer the discussion of the model choice for $\boldsymbol{\Sigma}$ to Section 4.3.2 and 4.4.1. Hereafter the parameters in our model are written as $\boldsymbol{\theta} = \{v, \boldsymbol{\Sigma}, \tau, \boldsymbol{k}\}$. The graph for our complete model can be seen in Fig. 4.1.

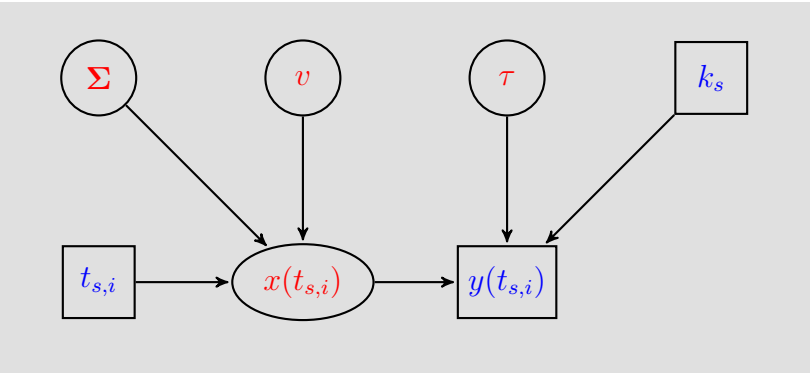


Fig. 4.1: A Directed Acyclic Graph (DAG) for the model described in Section 4.1. Circles indicate parameters/latent random variables whilst boxes indicate data. The solid lines indicate the direction of information flow.

4.2 Bayesian inference

This section discusses a strategy to perform fast inference on our model parameters without resorting to MCMC methods. Predictive posterior distributions of interest include the parameters $\boldsymbol{\theta}$, and either of the latent process $\boldsymbol{x}$ or $\boldsymbol{z}$, conditioning on data $\boldsymbol{y}$. Both $\boldsymbol{x}$ and $\boldsymbol{z}$ are legitimate targets of interest; and their computations are equally straightforward. Here we choose the former to demonstrate the computation.

4.2.1 Posterior distributions

Mathematical notation in this chapter closely follows that in chapter 3, including the notation trick discussed in Section 3.1.2. Notwithstanding, to allow this chapter to be self-contained we reintroduce our notation. We suppose $\boldsymbol{t_o}$ is the sorted set of all observed times at all series, i.e. $t_o = \text{sort}\{t_{s,1}, t_{s,2}, \ldots\} = \{t_1, t_2, \ldots, t_n\}$ where n is the total number of unique time points of observations across all series. We shall consider a stochastic system at these times. Similarly, we let $y_o(t_{s,i}); s = 1, \ldots, m$ whose values coincide with the $y(t_{s,i})$ values when series s has an observation at t_i and whose values are missing otherwise; there is typically one such series for each t_i. Let $\boldsymbol{y_o}$ denote the set of all such vectors. the vectors $\boldsymbol{y}$ (length $N = \sum_s^m n_s$) and $\boldsymbol{y_o}$ (length mn) contain the same information. We shall refer to $\boldsymbol{y}$ and $\boldsymbol{x}$ defined on $\boldsymbol{t_o}$ as $\boldsymbol{y_o}$ and $\boldsymbol{x_o}$ respectively. The data model can now be written as

$$\boldsymbol{y_o}|\boldsymbol{x_o}, \tau, \boldsymbol{k} \sim \mathcal{N}(\boldsymbol{x_o}, \mathbf{Q}_{y_o}^{-1})$$

Here, $\mathbf{Q}_{y_o}$ is a diagonal precision matrix with entries τ^{-1} and $k_s^{-1}\tau^{-1}$ corresponding to cores at times with data, and zeros where there are no data. Similarly, the process model has the form:

$$\boldsymbol{x_o}|v, \boldsymbol{\Sigma} \ \sim \ \mathcal{N}(\mathbf{0}, \mathbf{Q}_{x_o}^{-1})$$

where $\mathbf{Q}_{x_o} = \mathbf{Q}_{t_o} \otimes v^{-1}\boldsymbol{\Sigma}^{-1}$. Here, $\mathbf{Q}_{x_o}$ is a Kronecker product of the precision matrix $\mathbf{Q}_{t_o}$ and correlation matrix $\boldsymbol{\Sigma}$. At this stage, it is easy to write down the full conditional distribution of all unknown processes and parameters in our model. We could conceivably fit our model using MCMC methods, though the latent process $\boldsymbol{x_o}$ induces a large number of parameters and requires a long run. Instead, since the data model and prior can both be written as Gaussian distributions (via the notational tricks) it is

feasible to analytically integrate out $\boldsymbol{x_o}$, irrespective of the form of prior distributions for $\boldsymbol{\theta}$. More specifically,

$$\begin{aligned}\pi(\boldsymbol{x_o}, \boldsymbol{\theta}|\boldsymbol{y_o}) &\propto \pi(\boldsymbol{y_o}|\boldsymbol{x_o}, \boldsymbol{\theta})\pi(\boldsymbol{x_o}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\ &\propto |\mathbf{Q}_{y_o}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{y_o} - \boldsymbol{x_o})^T \mathbf{Q}_{y_o}(\boldsymbol{y_o} - \boldsymbol{x_o})\right) |\mathbf{Q}_{x_o}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{x_o}^T \mathbf{Q}_{x_o} \boldsymbol{x_o}\right) \pi(\boldsymbol{\theta}) \\ &\propto \ \pi(\boldsymbol{x_o}|\boldsymbol{y_o}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{y_o})\end{aligned} \tag{4.6}$$

Here, the full conditional of $\boldsymbol{x_o}$ is a Gaussian distribution defined by the precision matrix $\mathbf{Q}_{x_o|y_o} = \mathbf{Q}_{y_o} + \mathbf{Q}_{x_o}$ and mean vector $\boldsymbol{\mu}_{x_o|y_o}$ that is the solution to $\mathbf{Q}_{x_o|y_o}\boldsymbol{\mu}_{x_o|y_o} = \mathbf{Q}_{y_o}\boldsymbol{y_o}$. The full conditional of $\boldsymbol{\theta}$ has the form:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y_o}) = \left|\mathbf{Q}_{x_o|y_o}\right|^{-\frac{1}{2}} |\mathbf{Q}_{y_o}|^{\frac{1}{2}} |\mathbf{Q}_{x_o}|^{\frac{1}{2}} \exp\left(\frac{1}{2}\boldsymbol{y_o}^T \mathbf{Q}_{y_o}(\boldsymbol{\mu}_{x_o|y_o} - \boldsymbol{y_o})\right) \pi(\boldsymbol{\theta}) \tag{4.7}$$

This result allows analytical marginalisation of the latent process $\boldsymbol{x_o}$ from $\pi(\boldsymbol{x_o}, \boldsymbol{\theta}|\boldsymbol{y_o})$. Therefore the inference procedure can be divided into two separate stages as follows.

4.2.2 Stage 1: simulation-free computation of model parameters

Initial inference is focused on $\pi(\boldsymbol{\theta}|\boldsymbol{y_o})$, which is the same as $\pi(\boldsymbol{\theta}|\boldsymbol{y})$. We use R function *optim* to minimise the negative marginal likelihood (4.7) to locate its mode. We then evaluate it on a suitable parameter grid as discussed in Section 2.3.3.2. Recall that this approach allows us to approximate the continuous distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y}); \boldsymbol{\theta} \in \boldsymbol{\Theta}$ by the discrete distribution $\pi(\boldsymbol{\theta}_j|\boldsymbol{y}); \boldsymbol{\theta}_j \in \boldsymbol{\Theta}_J; j = 1, \ldots, J$. Here, $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}_J$ are, respectively, continuous and discrete parameter spaces; the latter will be used as a Riemann sum approximation to analytical integrations as discussed in the next inference stage.

4.2.3 Stage 2: summaries of latent process at an arbitrary time grid

As previously mentioned, our objective, given $\boldsymbol{y}$, is to provide posterior summaries of the latent process $\boldsymbol{x}$ on a regular time grid denoted by $\boldsymbol{t_g} = \{i\Delta; i = 1, \ldots, n_g\}$. We shall use $\boldsymbol{x_g}$ and $\boldsymbol{y_g}$ to refer to $\boldsymbol{x}$ and $\boldsymbol{y}$ defined on the grid $\boldsymbol{t_g}$.

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. $\boldsymbol{y_*} = \left(\boldsymbol{y_o}^T, \boldsymbol{y_g}^T\right)^T$ and $\boldsymbol{x_*} = \left(\boldsymbol{x_o}^T, \boldsymbol{x_g}^T\right)^T$. The main task in this stage involves marginalisation over $\boldsymbol{\theta}$:

$$\pi(\boldsymbol{x_*}|\boldsymbol{y_*}) = \int_{\boldsymbol{\Theta}} \pi(\boldsymbol{x_*}|\boldsymbol{y_*}, \boldsymbol{\theta})\ \pi(\boldsymbol{\theta}|\boldsymbol{y_*})\ d\boldsymbol{\theta} \tag{4.8}$$

Derivation of the first quantity in the above integrand is, again, by completing the quadratic form as in Eq. (4.6). The second quantity in the integrand is $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, the joint posterior distribution of the model parameters previously derived in Section 4.2.2. Importantly, the discrete approximation of the latter equation renders as summations the integrals that arise in Eq. 4.8, i.e.

$$\pi(\boldsymbol{x_*}|\boldsymbol{y_*}) \approx \sum_{\boldsymbol{\Theta}_J} \mathcal{N}\left(\boldsymbol{x_*}; \boldsymbol{\mu}_{x_*|y_*}(\boldsymbol{\theta}_j), \mathbf{Q}^{-1}_{x_*|y_*}(\boldsymbol{\theta}_j)\right)\ \pi(\boldsymbol{\theta}_j|\boldsymbol{y_*}) \triangle\boldsymbol{\theta}_j \tag{4.9}$$

where $\boldsymbol{\mu}_{x_*|y_*}$ and $\mathbf{Q}_{x_*|y_*}$ are the posterior mean and precision of $(\boldsymbol{x_*}|\boldsymbol{y_*}, \boldsymbol{\theta})$. We write $\boldsymbol{\mu}_{x_*|y_*}(\boldsymbol{\theta}_j)$ and $\mathbf{Q}_{x_*|y_*}(\boldsymbol{\theta}_j)$ 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. In Section 4.3.3 we obtain *pathwise* summaries of the full conditional for $\boldsymbol{x_g}$ by sampling for $\boldsymbol{x_*}$ from (4.9) and extract relevant values for $\boldsymbol{x_g}$.

It may be of interest to derive *pointwise* posterior summaries for $\boldsymbol{x_g}$. Since the *joint* posterior distribution is a finite mixture of Gaussian distributions, the *marginal* posterior distribution of the l^{th} element at a specific series corresponding to a temporal grid point of interest may be approximated as finite Gaussian mixture:

$$\pi(x^{(l)}_*|\boldsymbol{y_*}) \approx \sum_{\boldsymbol{\Theta}_J} \mathcal{N}\left(x^{(l)}_*; \mu^{(l)}_{x_*|y_*}(\boldsymbol{\theta}_j), \tau^{(l)}_{x_*|y_*}(\boldsymbol{\theta}_j)\right) \alpha_j \tag{4.10}$$

where $\mu^{(l)}_{x_*|y_*}(\boldsymbol{\theta}_j)$, conditional on the sample value $\boldsymbol{\theta}_j$, is the l^{th} element of the corresponding series identity from $\boldsymbol{\mu}_{x_*|y_*}$. Similarly, each conditional posterior variance $\tau^{(l)}_{x_*|y_*}(\boldsymbol{\theta}_j)$ is the l^{th} element of the series identity represented in the diagonal of the covariance matrix. Note that it is not necessary to invert the precision matrix $\mathbf{Q}_{x_*|y_*}$ to extract the marginal variances, as discussed in Section 2.3.3.4.

Pointwise posterior summaries such as the mean and variance corresponding to the

distribution in Eq. (4.10) can be computed as

$$
\begin{aligned}
E(x_*^{(l)}|\boldsymbol{y}_*) &= \sum_{\boldsymbol{\Theta}_J} \mu_{x_*|y_*}^{(l)}(\boldsymbol{\theta}_j)\alpha_j \\
Var(x_*^{(l)}|\boldsymbol{y}_*) &= \sum_{\boldsymbol{\Theta}_J} \left[\left(\mu_{x_*|y_*}^{(l)}(\boldsymbol{\theta}_j)\right)^2 + \tau_{x_*|y_*}^{(l)}(\boldsymbol{\theta}_j)\right]\alpha_j - \left(\sum_{\boldsymbol{\Theta}_J} \mu_{x_*|y_*}^{(l)}(\boldsymbol{\theta}_j)\alpha_j\right)^2
\end{aligned}
\tag{4.11}
$$

Calculations of other summarised statistics such as posterior modes, quantiles, etc. are similar. For generality, interquartile ranges (IQR) shall be used to quantify the prediction. We re-emphasise that although there are many computations with large matrices, such as solving equations, Cholesky decomposition, etc. they are only required J number of times. The fact that all of the matrices are sparse and can be efficiently stored and computed as band matrices presents a further significant computation saving.

4.3 Application to two Greenland ice core data sets

In this section we apply the model framework and inference procedures presented in previous sections to analyse the Greenland ice core data sets described in Section 2.1.1. As indicated, we work with the $\delta^{18}O$ process that is related to some aspects of climate, rather than climate itself. Moreover, we consider only data of approximately the last 11 k cal yr BP as we would like to avoid the non-linearities in the age/depth relationship, as well as non-negligible dating uncertainty.

4.3.1 Exploratory data analysis

For the Holocene period (approximately the last 11 k cal yr BP), depth is transformed to age by counting annual layers in chemicals that show a seasonal cycle (Rasmussen et al. 2006). This dating method produces minor errors and we do not consider temporal uncertainty in this chapter. The $\delta^{18}O$ and date records for all consecutive sections are presented in Fig. 4.2, along with boxplots of the age increments. 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.

Since the $\delta^{18}O$ observations are irregularly spaced in time and it may be non-stationary, it is not feasible to look at autocorrelation plots. Instead, the *empirical* semivariograms will be used to investigate the temporal variability of the measurements

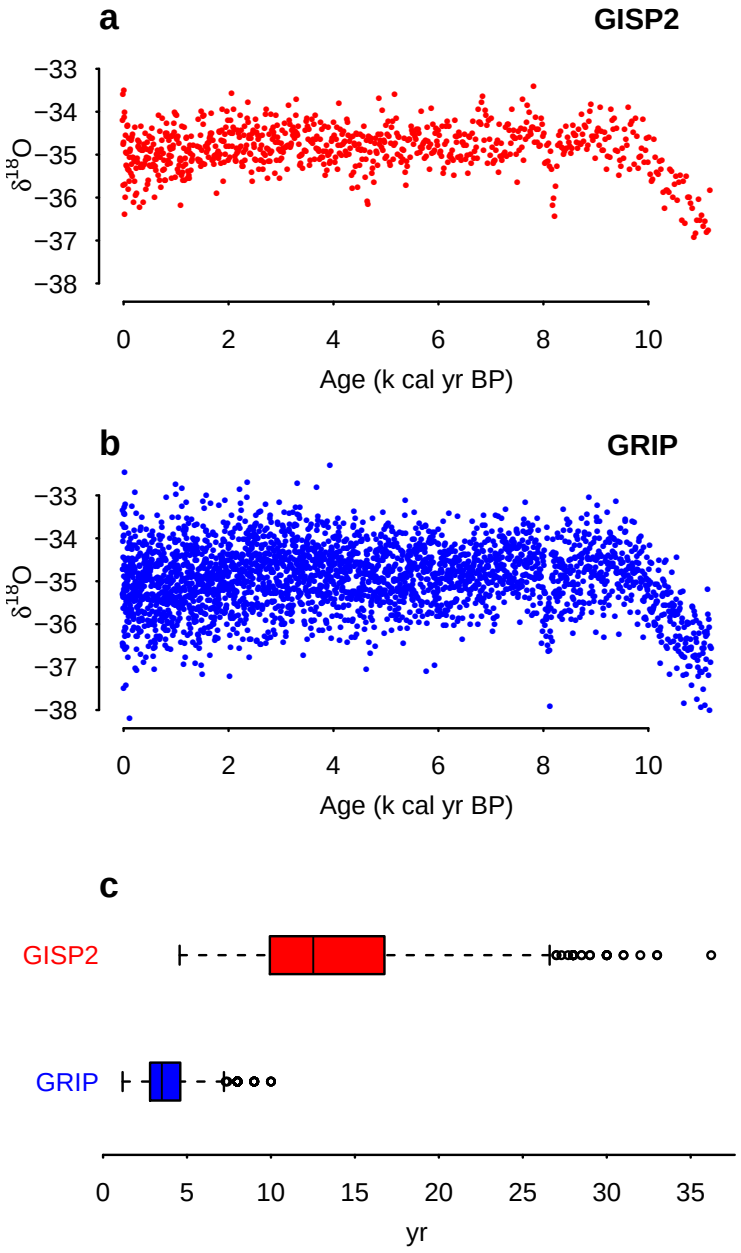


Fig. 4.2: Scatter plots of $\delta^{18}O$ measurements and ages of **(a)** GISP2 and **(b)** GRIP. **(c)** Boxplots of the time increments clearly show different irregularities in the ages. Note that we have omitted an age difference value of 80.6 yr between roughly 1.32 and 1.4 k cal yr BP in the boxplot 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) year for GISP2 and GRIP respectively.

under a continuous-time setting (Haslett 1997). This involves calculation of half of the squared differences of all combinations of $\delta^{18}O$ 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 empirical semivariogram. The usual approach is to assume one of the *theoretical* models and use this to infer the information near the origin corresponding to high-frequency behaviour (Chiles & Delfiner 2012, chapter 2).

Figure 4.3 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 $\delta^{18}O$ values) = intercept + slope $\times$ (time lag). To recap, the intercept is known as the nugget effect, and the slope is twice the variance of the unit increments of a nugget-free, continuous-time stochastic process. Specifically, the nugget term refers to the apparent discontinuity at the beginning of a semivariogram. It attributes to two sources of variation: the noise of data at high temporal frequency, and that which is due to uncertainty from data collection (Chiles & Delfiner 2012, chapter 2). 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. The least squared estimated intercepts and slopes are respectively 0.19 and 0.03 for GISP2, and 0.52 and 0.02 for GRIP. These point estimates should be compared to the Bayesian estimation depicted in Fig. 4.5.

Our final exploratory analysis focuses on the standardised distribution of the first differences of the $\delta^{18}O$ values. To begin with, we use R package geoR (Ribeiro Jr & Diggle 2001) to estimate the parameters of the theoretical linear semivariogram models. Based on these estimates, we compute the standardised distribution of the $\delta^{18}O$ increments. The resulting QQ plots, as shown in Fig. 4.4, indicate the suitability of the Gaussian assumption. This, as well as the finding obtained via the empirical semivariograms, provides the basis for the independent increment modelling assumption for the continuous-time stochastic model proposed in Section 4.1.

4.3.2 Model-fitting results

In this section the model framework and inference procedures presented in Section 4.1 are applied to analyse the pair of GISP2 and GRIP. When there are two cores, $m = 2$,

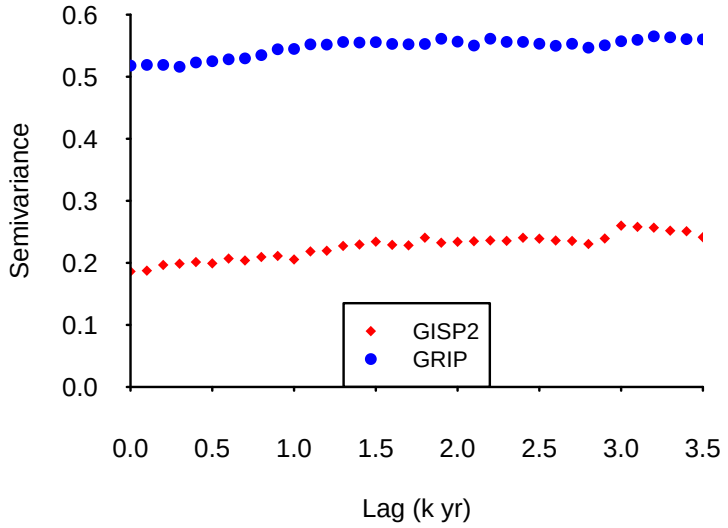


Fig. 4.3: The empirical semivariograms of GISP2 and GRIP. They suggest that the linear semivariogram is a suitable model for both of the ice core data sets, i.e. $\frac{1}{2}$ Var(first difference in the $\delta^{18}O$ values) = intercept + slope $\times$ (first difference in ages). See the text for further details.

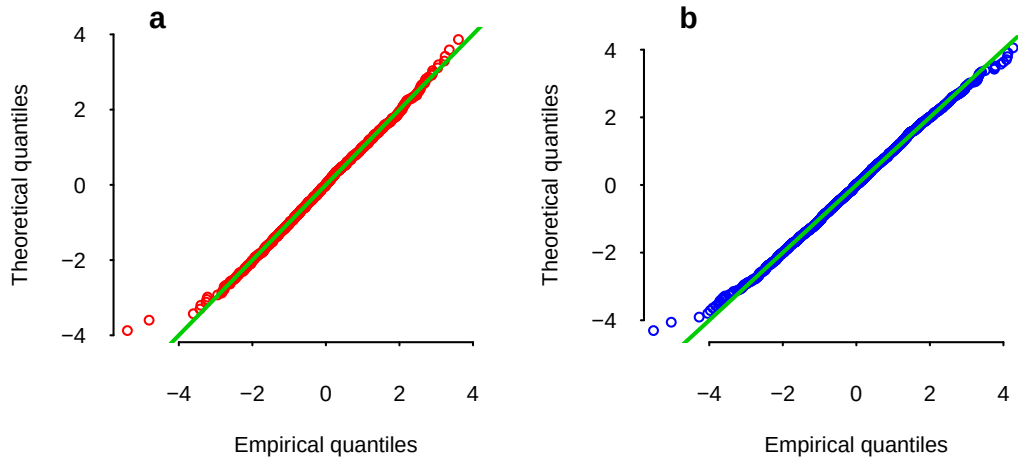


Fig. 4.4: QQ plots of the standardised increments, i.e. the ratio of the first differences in the $\delta^{18}O$ measurements and estimated standard errors of increments for **(a)** GISP2 and **(b)** GRIP. The unusual values correspond to very large differences in consecutive pairs of $\delta^{18}O$ values.

we model the cross-correlation matrix (4.5) as

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{4.12}$$

where ρ is the correlation coefficient; see also Section 4.4 for more details on this choice. As temperature processes at nearby locations always have (strongly) positively correlated increments, the same is expected for the $\delta^{18}O$ processes. Hence, we set 0.5 and 1 as the acceptable range for ρ. Furthermore, we assign a flat prior on its logit transformation [2] so that it has values on the real axis.

Previously we denoted $\boldsymbol{\theta} = \{v, \boldsymbol{\Sigma}, \tau, \boldsymbol{k}\}$. We chose to model $\boldsymbol{\Sigma}$ as a function of ρ. We also proved that all values of $\boldsymbol{k}$ can be written as a function of τ. Hereafter all unknown parameters in our model are $\boldsymbol{\theta} = \{v, \rho, \tau\}$, representing the variance per unit increment of the latent process, cross correlation coefficient and nugget effect (for the GRIP core) respectively.

Using the inference procedure outlined in Section 4.2, we obtain the discrete approximation to the marginal posterior distributions for the components of $\boldsymbol{\theta}$. For computational stability, we evaluate the posterior distribution for v and τ on a log scale, and ρ on the logit scale. For brevity of presentation, we transform the results to their original scale, apply a smoother to the gridded values and present them as in Fig. 4.5. It can be seen that the results for v and τ are consistent with Fig. 4.3, i.e. the posterior distributions contain the estimated slopes and intercepts. 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 regional temperature of Greenland.

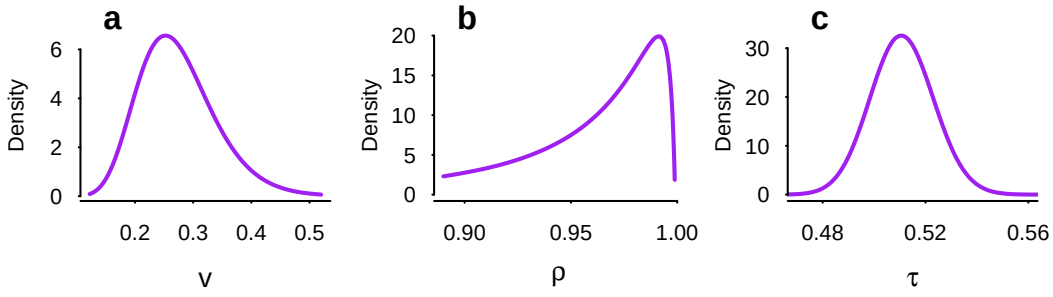


Fig. 4.5: Plots of the smoothed posterior distributions of **(a)** the variance of the unit increment of latent process $\boldsymbol{x}$, **(b)** cross correlation coefficient and **(c)** nugget effect for GRIP; the nugget effect for GISP is directly proportional (55/200) to this.

[2] $\text{logit}(\rho) = \frac{\rho - l}{h - \rho}$. Here, we set $l = 0.5$ and $h = 1$, respectively as the lower and upper bound value.

Under our model-based approach we have the choice of what summaries of the $\delta^{18}O$ histories to use. Our main interest lies in the true $\delta^{18}O$ values on a bidecadal time grid over the period of 0 to 11 k cal yr BP. As previously mentioned, we target $\boldsymbol{x_g}$ and call this true $\delta^{18}O$. The 95% and 50% credible intervals for the gridded process $\boldsymbol{x_g}$ are represented in Fig. 4.6. 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 & 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 Section 4.1, 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.

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 defer to Section 4.4.1 for a more formal discussion of these separate models. Figure 4.7 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, 8.2 k cal BP in GISP2) in Fig. 4.7 are a direct implication of either the associated age gaps, or abrupt changes in the $\delta^{18}O$ measure-

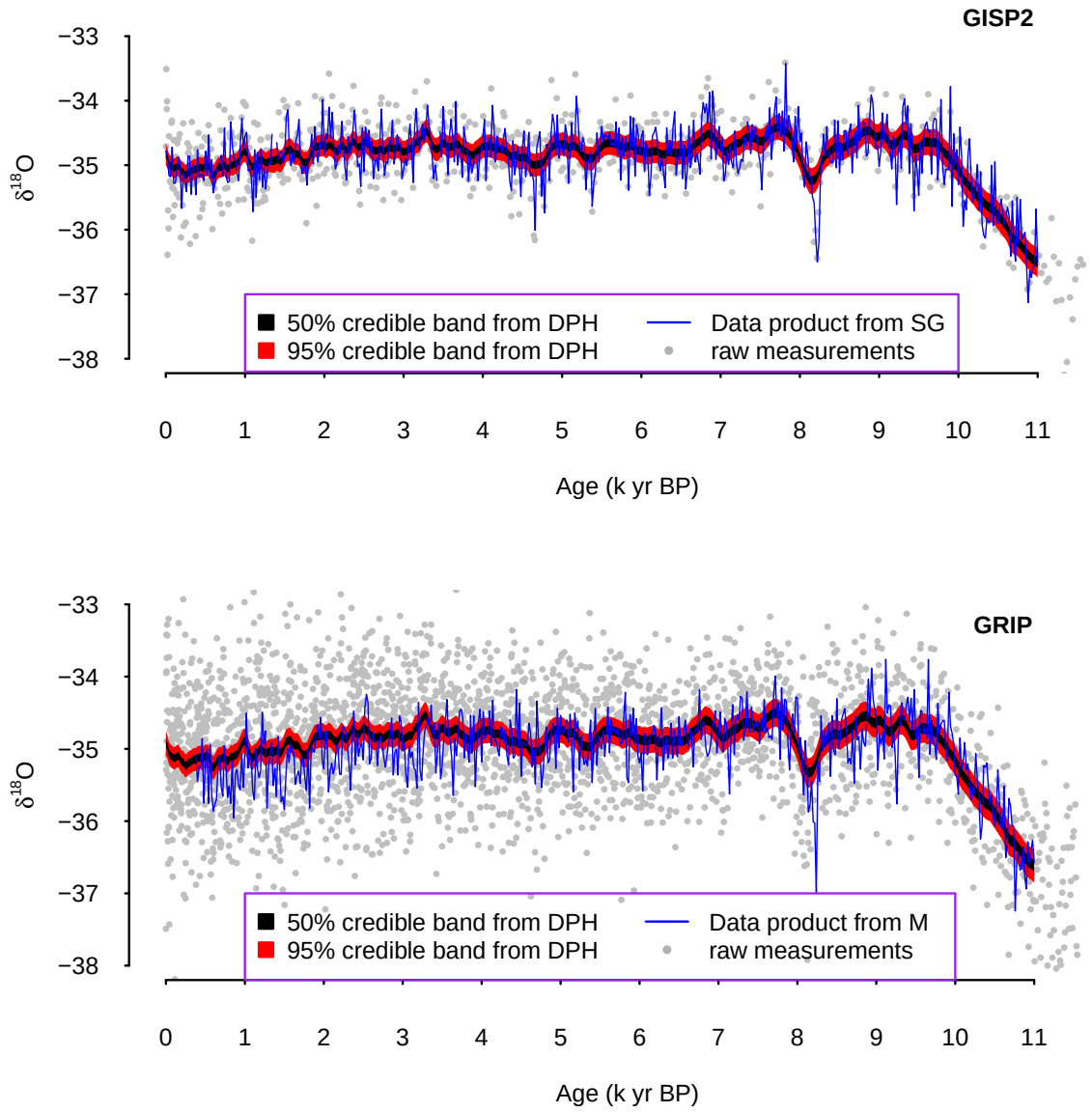


Fig. 4.6: Plots of quantile-based 50 and 95% credible intervals of the marginal posterior distributions of process $\boldsymbol{x_g}$ on a bidecadal time grid over the period of 0 to 11k cal BP conditional on both GISP2 and GRIP. We also show the bidecadal data product reported in Stuiver & Grootes (2000, SG) and Mogensen (2001, M).

ments, or both [3]. Specifically at 8.2 k cal yr BP, the spikes are more influenced by abrupt changes in the $\delta^{18}O$ measurements. In general, all of these spikes are higher and sharper in the separate models, in comparison to the joint model. The reverse

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

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) at 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. Further, it is an indication a possible limitation of the assumption of constant support within a core.

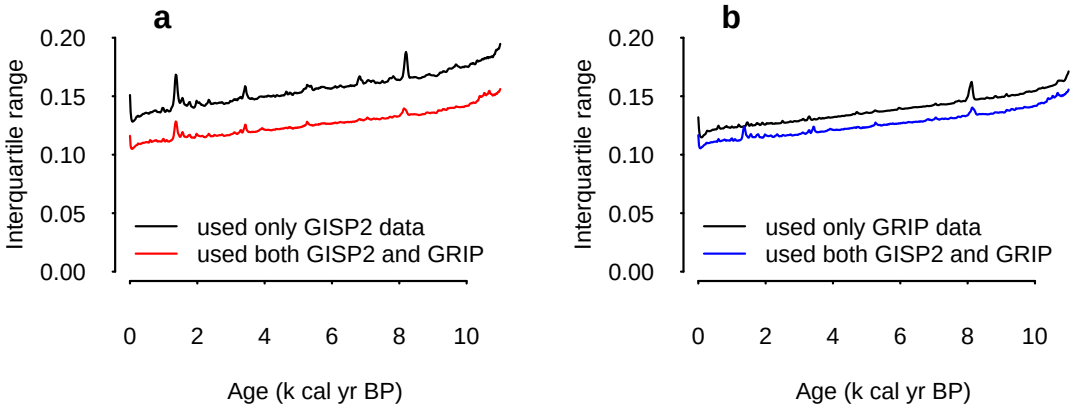


Fig. 4.7: Plots of the interquartile ranges (IQR; corresponding to the width of the 50% credible band in Fig. 4.6) of the marginal posterior distributions of $\boldsymbol{x_g}$ on a bidecadal time grid over the period of 0 to 11k cal BP at cores **(a)** GISP2 and **(b)** GRIP. The main features from 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.

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.

4.3.3 Case study of the 8.2ka event

A great challenge is posed by 'events' such as the 8.2ka event, the sudden reduction in North Atlantic temperature during a period around 8.2 k cal yr BP (Thomas et al. 2007). It is believed to be related to 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.

Like others in the literature, we define the 8.2ka event by an attainment of a minimum in the temperature value during a specific time period (Thomas et al. 2007). 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 $\delta^{18}O$ sampled histories from the joint posterior distribution of $\boldsymbol{x_g}$ as shown in Eq. (4.9), to illustrate a satisfactory approach. Formally we focus on the distributions of random variables $x_{\min} = \min_t x(t)$ and $t_{\min} = \arg_t \min x(t)$, being respectively the minimum value and the time at which this minimum was achieved. Furthermore, we focus on the period between 7.90 and 8.50 k cal yr BP that is believed to bracket the 8.2ka event, to distinguish from the possible long-term climate trend (Morrill & Jacobsen 2005).

Using the procedure previously introduced in Section 3.2.4, we obtain summary for 1000 minima of the δ^{18} histories as represented in Fig. 4.8. 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), Kobashi et al. (2007).

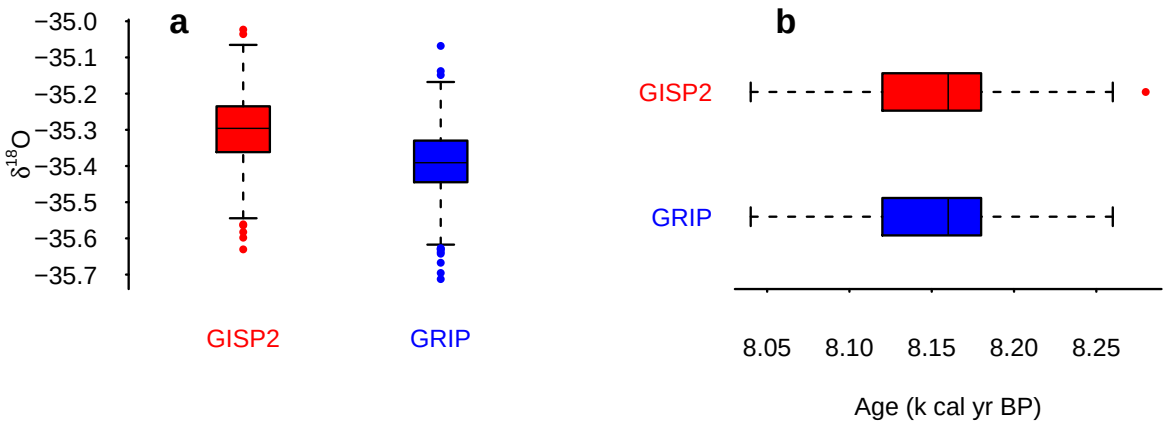


Fig. 4.8: **(a)** Boxplots of the minima and **(b)** times of the minima from GISP2 and GRIP over the range of 7.9 to 8.5 k cal yr BP. The interquartile range of the timing of the event from both cores is (8.12, 8.16, 8.18) k cal yr BP. All estimates are based on 1000 process histories.

The $\delta^{18}O$ 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.

4.4 Model validation

In this section, we perform two checks. First, within the context of our application we discuss the model choice for $\boldsymbol{\Sigma}$ in Eq. (4.5) by fitting and comparing several choices and choosing the best among them. Then, we investigate the identifiability of our model using simulated data under various circumstances.

4.4.1 Model choice for the cross-correlation function

Let M_1 denote the model with covariance function (4.12) 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 Section 4.3.2. It comprises of two separate models; one for each core. We have demonstrated, via Fig. 4.7, that the joint model utilises information from both cores more effectively than the separate models.

The previous paragraph re-emphasises the superiority of model M_1 over M_2. Next, we further propose model M_3, to be compared with M_1. The covariance of this model assumes varying variances for different cores, i.e.

$$\boldsymbol{\Sigma}_a = \begin{pmatrix} 1 & \rho\sqrt{a} \\ \rho\sqrt{a} & a \end{pmatrix} \tag{4.13}$$

We assign a reference prior for a, so that $\pi(a) = a^{-1}$. Its marginal posterior distribution is shown in Fig. 4.9. 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. 4.5. This result suggests that model M_3 is just a more conservative version of M_1.

We compare model M_1 and M_3 based on their deviance. In our notation, deviance of a model is defined as $-2\log(\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}))$. Indeed, this quantity is a random variable in the Bayesian framework so that the mean deviance is often chosen in lieu. Furthermore, there is an apparent pitfall of over-fitting since models with more parameters are

guaranteed to fit a data set better. Hence, in additional to the deviance, we would like a penalty term to deal with unnecessary terms in a model. In the literature, e.g. Gelman et al. (2014, Chapter 7), common measures of model fit and complexity are known as AIC, BIC, and DIC. For this work, we choose the DIC proposed by Spiegelhalter et al. (2002). It has the form:

$$-2\log\left[\pi(\boldsymbol{y}|\bar{\boldsymbol{x}},\hat{\boldsymbol{\theta}})\right]+2p_D$$

Here, $p_D=\log\left[\pi(\boldsymbol{y}|\bar{\boldsymbol{x}},\hat{\boldsymbol{\theta}})\right]-\mathrm{E}_{\boldsymbol{\Theta}_J}\{\log[\pi(\boldsymbol{y}|\bar{\boldsymbol{x}},\boldsymbol{\theta})]\}$, the process $\bar{\boldsymbol{x}}$ is short for the posterior mean of $\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}$, the parameter $\hat{\boldsymbol{\theta}}$ is the posterior mode of $\boldsymbol{\theta}|\boldsymbol{y}$, and the expectation term is evaluated with respect to the parameter grid $\boldsymbol{\Theta}_J$. All aforementioned quantities are readily available from the inference stages discussed in Section 4.2.2 and 4.2.3 hence calculation of DIC is straightforward. We obtained DIC values of 82.5 and 81.9 for model M_1 and M_3 respectively. The lower is this value, the better is the model. Although in this case, it could be misleading to choose M_3 over M_1 based on the small difference of 0.6 in the DIC values.

We have obtained some evidences that the extra parameter a in model M_3 can be made redundant, and that the common variance assumption in model M_1 is more suitable for GISP2 and GRIP. Including parameter a slightly decreases the DIC value; but such an extension may not be necessary when we take into account the additional computational cost associated with a more complex configuration of parameter grid $\boldsymbol{\Theta}_J$ and more expensive quadrature resulting from using more grid points.

4.4.2 Checking for model identifiability

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,ρ,τ) 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 of support that occurs in our ice core example. We then fit the model as described in Section 4.2, 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

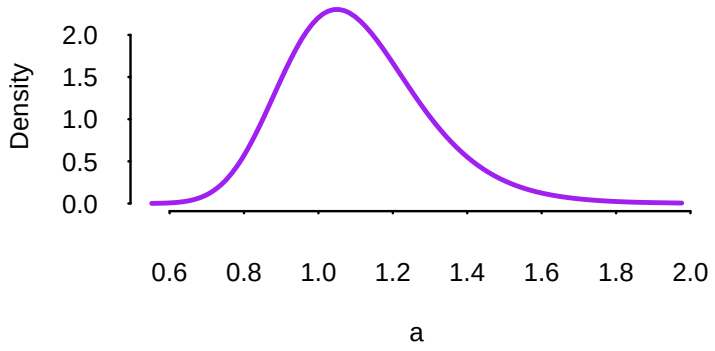


Fig. 4.9: Plot of the smoothed marginal posterior distribution of the parameter a in the covariance matrix of model M_3 described in Section 4.4.1. The mode is roughly 1 which suggests that a is redundant.

with the 90% interval.

The results of our simulations are shown in Table 4.1. 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.

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

Table 4.1: Performance of the model fitting algorithm. All results were based on 1000 simulation runs.

4.5 Discussion

This chapter has presented a hierarchical model to jointly analyse misaligned irregular time series. Our proposed framework is simple but useful to combine multiple time series, allowing each series to have a different temporal support. We applied the change of support theory from the geostatistical literature to deal with issues with different

supports. An important component of our model is the Gaussian Markov assumption based on multivariate independent increments. We derived and implemented a fast algorithm for parameter inference and imputation based on this model. We further demonstrated that the joint approach utilises information from multiple time series more efficiently than one-series-at-a-time alternatives.

We applied the method to create climate data product from a pair of Greenland ice cores. Some parameters in our model were formulated according to the respective lengths of ice core sections. Firstly, this approach is likely to be problematic in the study of sections longer than the Holocene. Secondly, we have implicitly assumed perfect knowledge of the timing of all observations, which is generally not true in many other palaeoclimate records. Hence, a generic extension to our hierarchical structure is to incorporate uncertainty in the timescale. Thirdly, the Gaussian assumption for the increments is conceivably too restrictive for capturing of abrupt climate change events. We offer solutions to these issues in the next chapter.

Chapter 5

Joint inference of multiple volatile processes with time uncertainty [1]

Chapter 4 developed a method based on a three-layer hierarchical model to deal with multiple irregular time series data, with each series having a different temporal support. A generic version of this model has the form:

$$\pi(\boldsymbol{x}, \boldsymbol{\psi}, \boldsymbol{\theta} | \boldsymbol{y}, \boldsymbol{t}) \propto \pi(\boldsymbol{y} | \boldsymbol{x}, \boldsymbol{t}, \boldsymbol{\psi}) \pi(\boldsymbol{x} | \boldsymbol{t}, \boldsymbol{\theta}) \pi(\boldsymbol{\psi}, \boldsymbol{\theta}) \tag{5.1}$$

At the highest layer, a vector of data $\boldsymbol{y}$ observed at times $\boldsymbol{t}$ is linked to a latent process $\boldsymbol{x}$ by a simple model with parameters $\boldsymbol{\psi}$. In the next layer, a vector of parameters $\boldsymbol{\theta}$ specifies a distribution for the process $\boldsymbol{x}$. The lowest level comprises of prior distributions for parameters $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$. Linear Gaussian processes are the central focus of the above model. Moreover, uncertainties in the times of observations were neglected. Whilst these settings provided a fast and easy method for a joint statistical analysis, they are only applicable to a limited set of situations.

This chapter further generalises the problem setting considered in Eq. (5.1). The aim is to develop efficient statistical treatments for multiple non-Gaussian time series processes, with the times of the observations themselves being uncertain. More specifically, we consider the following extensions:

1. Weakening of Gaussian assumptions to allow data $\boldsymbol{y}$ and latent process $\boldsymbol{x}$ to have more realistic / flexible likelihood functions,

[1] **The work in this chapter is an extension to a publication in the *Journal of the Royal Statistical Society: Series C* (Parnell et al. 2015).**

2. Allowing uncertainty in times $\boldsymbol{t}$ at which data points are measured,

3. Incorporating stochastic volatility effect for latent process $\boldsymbol{x}$.

In Section 5.1, we describe the relevant palaeoclimate data sets and formalise the problem setting for the chapter. Section 5.2 presents our four-layer hierarchical model. Section 5.3 introduces a modularised algorithm which divides the inference procedure to three different stages and give rises to a fast MCMC algorithm for for posterior inference. The chapter continues in Section 5.4 with a real data example; a joint analysis of pollen data extracted from three sediment cores taken beneath three lakes in Finnmark. We apply our methods to study a rapid climate change period known as the Younger Dryas. Section 5.5 reports a simulation study for the purpose of model checking. The chapter concludes in Section 5.6 with a summarised discussion of our model and its variants.

5.1 Problem setting

The statistical challenges for this chapter are motivated by the reconstruction of climate histories from pollen data previously introduced in Section 2.1.2. Therefore, our discussion throughout the chapter will be associated with the terminology and statistical issues in palaeoclimatology.

To recap, each fossil sediment core comprises two sets of observations: pollen counts recorded at some instantaneous depths, and dating of the depth layers themselves. Table 5.1 illustrates a schematic representation of the information available in a single sediment fossil core (we have many cores). At some layers of the core, pollen data are available. At some other layers, materials are extracted and radiocarbon dated.

In addition to the fossil pollen information, there is a large set of training data collected from thousands of locations around the world. It can be seen from Table 5.2 that each pair of data points consists of known climate and associated pollen counts. This data set is used for calibration, i.e., to build a statistical relationship between pollen and climate. The science is simple; change in climate causes change in vegetation which is quantified, in this case, by counts of pollen grains.

Both modern and fossil pollen data sets are utilised to make statements about underlying unknown climate processes on a calendar time scale. The times are unknown

Fossil data			Point reconstruction	
Depth	Fossil pollen	Radiocarbon age	Calendar age	Climate
$[d_{s,1}]$	$[y_{s,1}]$	$[r_{s,1}]$	$t_{s,1}$	$x(t_{s,1})$
$[d_{s,2}]$	$[y_{s,2}]$		$t_{s,2}$	$x(t_{s,2})$
$[d_{s,3}]$	$[y_{s,3}]$	$[r_{s,2}]$	$t_{s,3}$	$x(t_{s,3})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[d_{s,n_s-1}]$	$[y_{s,n_s-1}]$	$[r_{s,l_s}]$	t_{s,n_s-1}	$x(t_{s,n_s-1})$
$[d_{s,n_s}]$	$[y_{s,n}]$		t_{s,n_s}	$x(t_{s,n_s})$

Table 5.1: A schematic representation of information from *one* fossil sediment core (core 's'). Square brackets are used to distinguish observed measurements from unobserved latent variables. At layers corresponding to depths $\{d_{s,1}, \ldots, d_{s,n_s}\}$, pollen counts $\{y_{s,1}, \ldots, y_{s,n_s}\}$ are available. This information is used (in addition to a training data set as represented in Table 5.2 below) to learn about climate $\{x(t_{s,1}), \ldots, x(t_{s,n_s})\}$ at times $\boldsymbol{t} = \{t_{s,1}, \ldots, t_{s,n_s}\}$; the times themselves are stochastically interpolated from radiocarbon ages $\{r_{s,1}, \ldots, r_{s,l_s}\}$ at l_s slices of depth in the core.

Training data	
Modern pollen	Modern climate
x_1^m	y_1^m
$\vdots$	$\vdots$
x_K^m	x_K^m

Table 5.2: Modern (training) data set of modern climate conditions $\boldsymbol{x^m} = \{x_1^m, \ldots, x_K^m\}$ and pollen counts $\boldsymbol{y^m} = \{y_1^m, \ldots, y_K^m\}$.

but can be derived from radiocarbon dated material at some depths of the core. We emphasise that our focus is on *joint* inference at all times, for multiple fossil cores.

Parnell et al. (2015) recently propose a hierarchical model to pool multiple sources of uncertainty into a unified framework, and stochastically interpolate climate onto an arbitrary time grid. Two novel ingredients of their methods are the use of the Normal Inverse Gaussian ($\mathcal{NIG}$) process as a prior to model the smoothness of climate over time, and a novel modularised MCMC algorithm for posterior inference. This method

utilises the modern data set, but it works with each fossil sediment core separately.

The $\mathcal{NIG}$ process is arguably the heart of the model developed by Parnell et al. (2015), and thus the extension in this chapter. We propose to use the multivariate version of the $\mathcal{NIG}$ to model climate from multiple fossil sediment cores jointly. Our aim is to allow for appropriate 'strength borrowing' for more efficient use of data, as opposed to the one-core-at-a-time approach. Increased model complexity is inevitably associated with increased computational burden. Simulation-free inference for model parameters, as discussed in Chapter 4, is no longer implementable. We will extend the aforementioned modularised MCMC algorithm to the multiple cores context.

5.2 A four-layer hierarchical statistical model

In this section, we introduce main notation, highlight posterior distributions of interest, announce main assumptions, and describe a four-layer hierarchical model. The layers are named the data, calibration, process and parameter layer. This chapter as a whole is built on the recent work of Parnell et al. (2015). Therefore, whilst discussion of the data, calibration and parameter layer is brief, we provide a more detailed discussion of the process layer; wherein lies our main contribution.

5.2.1 Main notation and posterior distributions

Since this chapter is an extension of the work in preceding chapters, the basic notation used therein is employed again. Necessarily, several new pieces of notation are also introduced. The joint distribution of all latent processes and unknown parameters is

$$\begin{aligned}
&\pi(\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{v}, \boldsymbol{\Sigma}, \boldsymbol{\psi}) | \boldsymbol{y}, \boldsymbol{x^m}, \boldsymbol{y^m}, \boldsymbol{d}, \boldsymbol{r}, \boldsymbol{\theta}) \propto \\
&\underbrace{\prod_{s=1}^{m} \prod_{i=1}^{n_s} \pi\Big(y(t_{s,i}) | x(t_{s,i}), \boldsymbol{\psi}\Big) \prod_{s=1}^{m} \prod_{j=1}^{l_s} \pi\Big(r_{s,j} | t_{s,j}, x(t_{s,j})\Big)}_{\text{data layer}} \times \underbrace{\prod_{k=1}^{K} \pi(y_k^m | x_k^m, \boldsymbol{x}, \boldsymbol{\psi})}_{\text{calibration layer}} \times \\
&\underbrace{\prod_{u=2}^{n} \pi\Big(\boldsymbol{x}(t_u) | \boldsymbol{x}(t_{u-1}), \boldsymbol{v}_{u-1}, \boldsymbol{\Sigma})\Big) \prod_{s=1}^{m} \prod_{g=1}^{n} \pi(t_{s,g} | d_{s,g}) \pi(\boldsymbol{v} | \boldsymbol{t}, \boldsymbol{\theta})}_{\text{process layer}} \times \underbrace{\pi(\boldsymbol{\Sigma}, \boldsymbol{\psi})}_{\text{parameter layer}} \qquad (5.2)
\end{aligned}$$

The notation, in order of appearance on the RHS of the above equation, are

- $y(t_{s,i})$ is an observed (multivariate) count of pollen grains from core $s; \{s = 1, \ldots, m\}$ at an instantaneous time point $t_i \{i = 1, \ldots, n_s\}$. There are $N = \sum_{s}^{m} n_s$ total number of times across all series, n of which are unique, $n \leq N$. In vector form, $\boldsymbol{y}$ is the N-by-1 multivariate pollen counts from all cores.

- $x(t_{s,i})$ is the (multivariate) latent climate process at time t_i on core s, with $\boldsymbol{x}$ being the corresponding vector notation. Note that in scalar form the latent variable is expressed as $x(t_s)$ because it is defined for all times.

- Parameter vector $\boldsymbol{\psi}$ describes the climate/pollen relationship.

- At the j^{th} slice of core s, $r_{s,j} \{j = 1, \ldots, l\}$ is a radiocarbon determination of material at depth $d_{s,j}$. This, in turn, carries information about the unobserved calendar age $t_{s,j}$. As a result, there is uncertainty associated with the times of the observation $y(t_{s,i})$. In vector notation, we denote the radiocarbon ages, selection of depths, and calendar ages respectively as $\boldsymbol{r}_s$, $\boldsymbol{d}_s$ and $\boldsymbol{t}_s$ for core s, and $\boldsymbol{r}$, $\boldsymbol{d}$ and $\boldsymbol{t}$ for all cores.

- We write the training data sets in vector form as $\boldsymbol{x^m} = \{x_1^m, \ldots, x_K^m\}$ and $\boldsymbol{y^m} = \{y_1^m, \ldots, y_K^m\}$. Here, K is the total number of locations around the world where known climate conditions and pollen counts are available.

- The variance-covariance matrix of a multivariate increment of climate between time t_i and t_{i+1} has variance v_i and cross-correlation matrix $\boldsymbol{\Sigma}$. The variance process for all increments are denoted $\boldsymbol{v}$.

- The variance process $\boldsymbol{v}$ is the function of parameter vector $\boldsymbol{\theta}$. As can be seen from Eq. (5.2), we fix the values for $\boldsymbol{\theta}$ as they carry little information forward. That is, in the context of the application considered in this chapter, we do not expect to learn about this set of higher level parameter to any significant degree. We return to this in Section 5.2.6.

5.2.2 Main assumptions

1. The set of parameter $\boldsymbol{\psi}$ can be fully learnt from the training set $\{\boldsymbol{x}^m, \boldsymbol{y}^m\}$.

2. The dependency between climate $\boldsymbol{x}$ and the processes which generate either radiocarbon ages $\boldsymbol{r}$ or $\boldsymbol{t}$ is assumed negligible.

3. Conditional independence: (a) pollen counts $\boldsymbol{y}^m$ and $\boldsymbol{y}$ at different layers and cores are independent given climate $\boldsymbol{x}^m$ and $\boldsymbol{x}$ and parameters $\boldsymbol{v}$ and $\boldsymbol{\psi}$; (b) radiocarbon dates $\boldsymbol{r}$ are independent given calendar ages $\boldsymbol{t}$ at different layers and across all cores.

In comparison to Parnell et al. (2015), assumption 1 is exactly the same, whilst assumptions 2 and 3 are herein in the context of multiple cores.

5.2.3 Data layer

The model for each fossil pollen observation conditioning a true but unknown climate, $y(t_{s,i})|x(t_{s,i}), \boldsymbol{\psi}$, is an independent zero-inflated negative binomial distribution (ZINB). A technical review of this model is given in Section 5.3.1.1.

It follows from assumption 2 in Section 5.2.2 that $\pi(\boldsymbol{r}|\boldsymbol{t}, \boldsymbol{x}) \approx \pi(\boldsymbol{r}|\boldsymbol{t})$. Furthermore, each radiocarbon age conditioning on the corresponding unknown calendar age, $r_{s,j}|t_{s,j}$, can be modelled as a Gaussian distribution. Section 5.3.1.2 gives more details on this model, as well as prior choice for $\boldsymbol{t}$.

5.2.4 Calibration layer

The sheer amount of the training data makes assumption 1 in Section 5.2.2 feasible. That is, the set of parameters $\boldsymbol{\psi}$ represents all the information in the calibration layer, and $\pi(\boldsymbol{y}^m|\boldsymbol{x}^m, \boldsymbol{x}, \boldsymbol{\psi}) \approx \pi(\boldsymbol{y}^m|\boldsymbol{x}^m, \boldsymbol{\psi})$. As in the case of fossil pollen data, we apply the conditional independence assumption outlined in Section 5.2.2 to the training pollen data, given associated climate and parameters. Again, each of these terms is a ZINB model.

5.2.5 Process layer

One of the novel contributions of this chapter is the specification of a joint prior distribution for $\boldsymbol{x}$. Therefore, we provide a more detailed discussion of this model layer.

Before discussing our proposal, we briefly recap the multivariate Gaussian independent increments model used in both Chapter 3 and 4 via Eq. (3.2) and (4.5) respectively. Let $\boldsymbol{x}(t)$ be the latent process $\boldsymbol{x}$ defined at time t. A Gaussian multivariate independent increment model has the form:

$$\boldsymbol{x}(t+h) - \boldsymbol{x}(t) \;\sim\; \mathcal{N}(\mathbf{0}, v\ |h|\ \boldsymbol{\Sigma}) \tag{5.3}$$

where v represents the variance of a unit increment, h is an arbitrary time increment, and $\boldsymbol{\Sigma}$ is the cross-correlation matrix capturing the relationship between cores. Hence the variance of an increment of any unit is directly proportional to its corresponding time difference. It is the sum of the variances of many smaller incremental units. That is, the variance between time t_j and t_k is $|t_k - t_j|v$ and $\int_{t_j}^{t_k} v dt$ for the discrete and continuous case respectively. Their calculation is straightforward since v is a constant. However, climate variability is a *stochastic* system. It is a slowly varying background process that changes smoothly with time, occasionally with large changes. Unfortunately, model (5.3), by design, can not capture the stochastic variability.

We propose to extend (5.3) to accommodate stochastic variability in the underlying increment as follows:

$$\begin{aligned} \boldsymbol{x}(t+h) - \boldsymbol{x}(t) &\sim \mathcal{N}(0, v_t \boldsymbol{\Sigma}) \\ v_t &\sim \mathcal{IG}(\theta_1|h|, \theta_2|h|) \end{aligned} \tag{5.4}$$

where v_t is the variance random variable of the i^{th} increment of $\boldsymbol{x}$, depending on time lag $|h|$ and parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$ controlling scale and shape of the variances of the increments. Let $h = 1$ for clear presentation, the $\mathcal{IG}$ process has the following formulation, as given in Betro & Rotondi (1991):

$$\left(\frac{\theta_1\theta_2}{2\pi}\right)^{1/2} v_t^{-3/2} \exp(\theta_2) \exp\left[\frac{-\theta_2}{2}\left(\frac{v_t}{\theta_1} + \frac{\theta_1}{v_t}\right)\right] \tag{5.5}$$

This is is flexible model as its distribution function can have a wide range variety of shapes. We believe it is a suitable candidate for modelling variances for several reasons. Firstly, it always returns positive values. Secondly, it is *infinitely divisible* so it can be represented as the sum of an arbitrary number of *i.i.d.* random variates. Thirdly, it is *closed under convolution* in the sense that the distribution of the sum of arbitrary components is also a $\mathcal{IG}$ distribution. Finally, and most importantly for this work, the process $\boldsymbol{x}$ is multivariate Gaussian conditional on the stochastic variances

$\boldsymbol{v} = \{v_1, \ldots, v_{n-1}\}$. In other words, the process $\boldsymbol{x}$ is marginally a multivariate Normal Inverse Gaussian ($\mathcal{NIG}$) process.

The Normal, inverse Gaussian, and indeed the Normal Inversion Gaussian distributions are not only infinitely divisible but also closed under addition. This means that the shape of the distribution does not change if we change the size of the time increment. This is a crucial feature for the bridging algorithm that we propose in Section 5.3.1.3.

On the other hand, a distribution such as the t distribution is infinitely divisible but not closed under addition. To see this, we revisit the case study of the t-distribution with 8 df as proposed by Haslett et al. (2006) to model changes in climate histories. As reviewed in Ahsanullah et al. (2014, Chapter 5), whilst it is possible to derive the sum of n *i.i.d.* t-distributed random variables, the resulting sum does not guarantee to have a t distribution. Therefore, the t distribution is infinitely divisible but not generally closed under addition; the latter is true only in the limit when the degree of freedom is infinity. One of the novel contribution of Parnell et al. (2015), in extending the work of Haslett et al. (2006), is the use of the NIG distribution to model climate changes in the past, in lieu of the t-distribution. In this chapter, we propose the multivariate NIG, thus making further contributions based on the aforementioned works.

From a different view point, model (5.3) can be seen as associating with a *deterministic* linear variogram, as discussed in Section 4.3.1 in the context of ice core modelling. In contrast, model (5.4) is richer as it is associated with *stochastic* linear variograms, as can be seen in Fig. 5.1.

The process layer also includes unknown latent calendar ages $\boldsymbol{t}$. We use the Compound Poisson Gamma process (Haslett & Parnell 2008) for $\pi(\boldsymbol{t}|\boldsymbol{d})$, applied to each core separately. Section 5.3.1.2 gives further technical details of this model.

5.2.6 Parameter layer

- As already indicated, we fix values for $\boldsymbol{\theta}$. More specifically, we let $\theta_1 = 3$ and $\theta_2 = 55$, approximately the mode of the prior distributions as used in Parnell et al. (2015). Note that as θ_2 approaches ∞, model (5.4) becomes the multivariate independent increment models discussed in Eq. (5.3). Therefore $\mathcal{NIG}$ is a more general and flexible model than the Gaussian process.

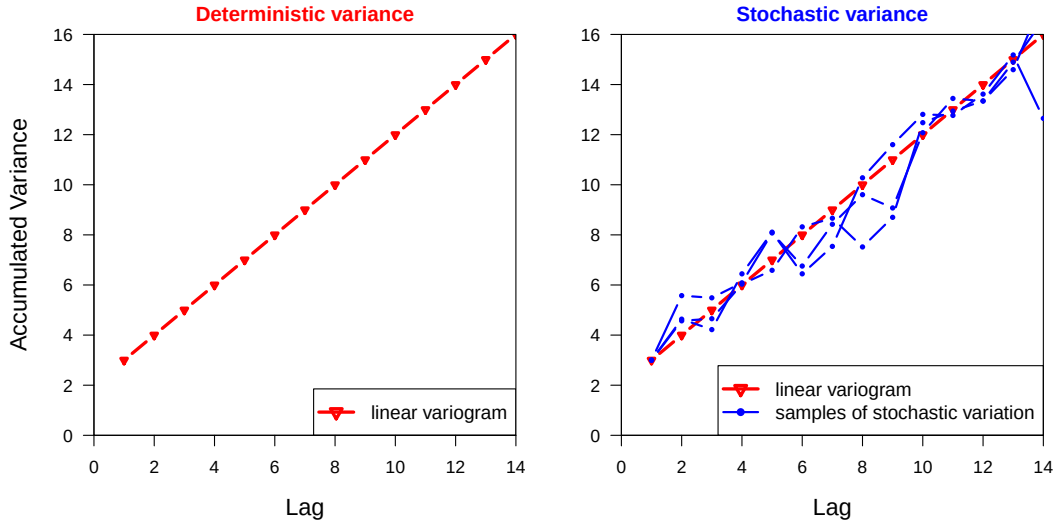


Fig. 5.1: A schematic representation of the variogram associated with model (5.3) (left figure) and (5.4) (right figure).

- Initially, we use the exchangeable correlation structure for matrix $\boldsymbol{\Sigma}$. In other words, this matrix has values of ones in the main diagonal an ρ's everywhere else. We assign a strong prior knowledge for ρ. Specifically, we allow it to be between 0.7 and 1 to take into account the potentially high correlation between climate from nearby locations. For computational convenience, we first use the general logit function to transform ρ to have values on the real axis (as done in footnote 2 of page 54). Then we assign a flat prior on this logit function. We discuss other model choices for $\boldsymbol{\Sigma}$ in Section 5.6.

- Prior specification for $\boldsymbol{\psi}$ is discussed in Section 5.3.1.1.

5.3 Bayesian inference

In this section, we perform parameter inference on the model discussed in the previous section. First, we divide the overall process into three separate stages; a Bayesian modularisation approach. Then, we utilise a notational trick for misaligned irregular time series to marginalise out a high dimensional climate process. Next, we derive fast MCMC algorithms to sample for the left over unknown parameters. Finally, conditioning on the MCMC samples, we propose a bridge sampling algorithm based on the multivariate $\mathcal{NIG}$ process to sample for an interpolated climate process.

5.3.1 A modularised Bayes approach

Based on Eq. (5.2), we can write down the full conditional distributions of latent processes and parameters of interest. Once these distributions are determined, generic Gibbs and / or Metropolis-Hastings can be used to generate posterior samples. In our case, as the number of parameters involved is high and the associated full conditional distributions are not available in closed form, a plain 'vanilla' MCMC is not feasible. We seek an algorithm that allows a 'once-off' calculation of expensive terms even under a sampling-based inference scheme. This is a key underlying principle of the modularised MCMC algorithm developed by Parnell et al. (2015), which we will extend to our framework.

Simply put, modularisation is an act of reducing the precision in the estimation of parameters, or, as described by Parnell et al. (2015), '*a conservative assumption*'. In practice, this is done by removing the underlined terms in the complete conditionals as follows:

$$\pi(\boldsymbol{\psi}|\ldots) \propto \underline{\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\psi})}\pi(\boldsymbol{y}^m|\boldsymbol{x}^m,\boldsymbol{\psi})\pi(\boldsymbol{\psi}) \approx \pi(\boldsymbol{y}^m|\boldsymbol{x}^m,\boldsymbol{\psi})\pi(\boldsymbol{\psi}) \tag{5.6}$$

$$\pi(\boldsymbol{t}|\ldots) \propto \pi(\boldsymbol{r}|\boldsymbol{t})\underline{\pi(\boldsymbol{x}|\boldsymbol{v},\boldsymbol{t},\boldsymbol{\Sigma})\pi(\boldsymbol{v}|\boldsymbol{t},\boldsymbol{\theta})}\pi(\boldsymbol{t}|\boldsymbol{d}) \approx \pi(\boldsymbol{r}|\boldsymbol{t})\pi(\boldsymbol{t}|\boldsymbol{d}) \tag{5.7}$$

where '...' denotes all other data and parameters. This is essentially assumptions 1 and 2 outlined in Section 5.2.2, where we propose to *cut feedback* between $\boldsymbol{\psi}$ and $\boldsymbol{y}$, and once more between $\boldsymbol{t}$ and $\{\boldsymbol{x}, \boldsymbol{v}\}$.

After modularisation, we see that (5.2) can be written in a simpler form:

$$\pi(\boldsymbol{x},\boldsymbol{t},\boldsymbol{v},\boldsymbol{\Sigma},\boldsymbol{\psi}|\boldsymbol{y},\boldsymbol{x^m},\boldsymbol{y^m},\boldsymbol{d},\boldsymbol{r},\boldsymbol{\theta}) \propto \pi(\boldsymbol{\psi}|\boldsymbol{x^m},\boldsymbol{y^m})\,\pi(\boldsymbol{t}|\boldsymbol{r},\boldsymbol{d})\,\pi(\boldsymbol{x},\boldsymbol{v},\boldsymbol{\Sigma}|\boldsymbol{y},\boldsymbol{\psi},\boldsymbol{t},\boldsymbol{\theta}) \tag{5.8}$$

Here, the first and second term on the RHS of the above equation are, respectively, Eq. (5.6) and (5.7). The third term is the full conditional distribution

$$\pi(\boldsymbol{x},\boldsymbol{v},\boldsymbol{\Sigma}|\ldots) \propto \pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\psi})\pi(\boldsymbol{x}|\boldsymbol{v},\boldsymbol{t},\boldsymbol{\Sigma})\pi(\boldsymbol{v}|\boldsymbol{t},\boldsymbol{\theta})\pi(\boldsymbol{\Sigma}) \tag{5.9}$$

From a computational perspective, the key achievement of modularisation is the division of a large complex model in (5.2) into smaller modules. Equations (5.6), (5.7) and (5.8) are dubbed respectively as the *modern analogue*, *chronology*, and *reconstruction* module. The full hierarchical model and the effect of modularisation can be seen in Fig. 5.2. The rest of this section discusses the modules in detail. We utilise the

work of Sweeney (2012) and Haslett & Parnell (2008) for the modern analogue and chronology module respectively. Therefore, the discussion on these modules will be brief. The focus of out work will be on the reconstruction module.

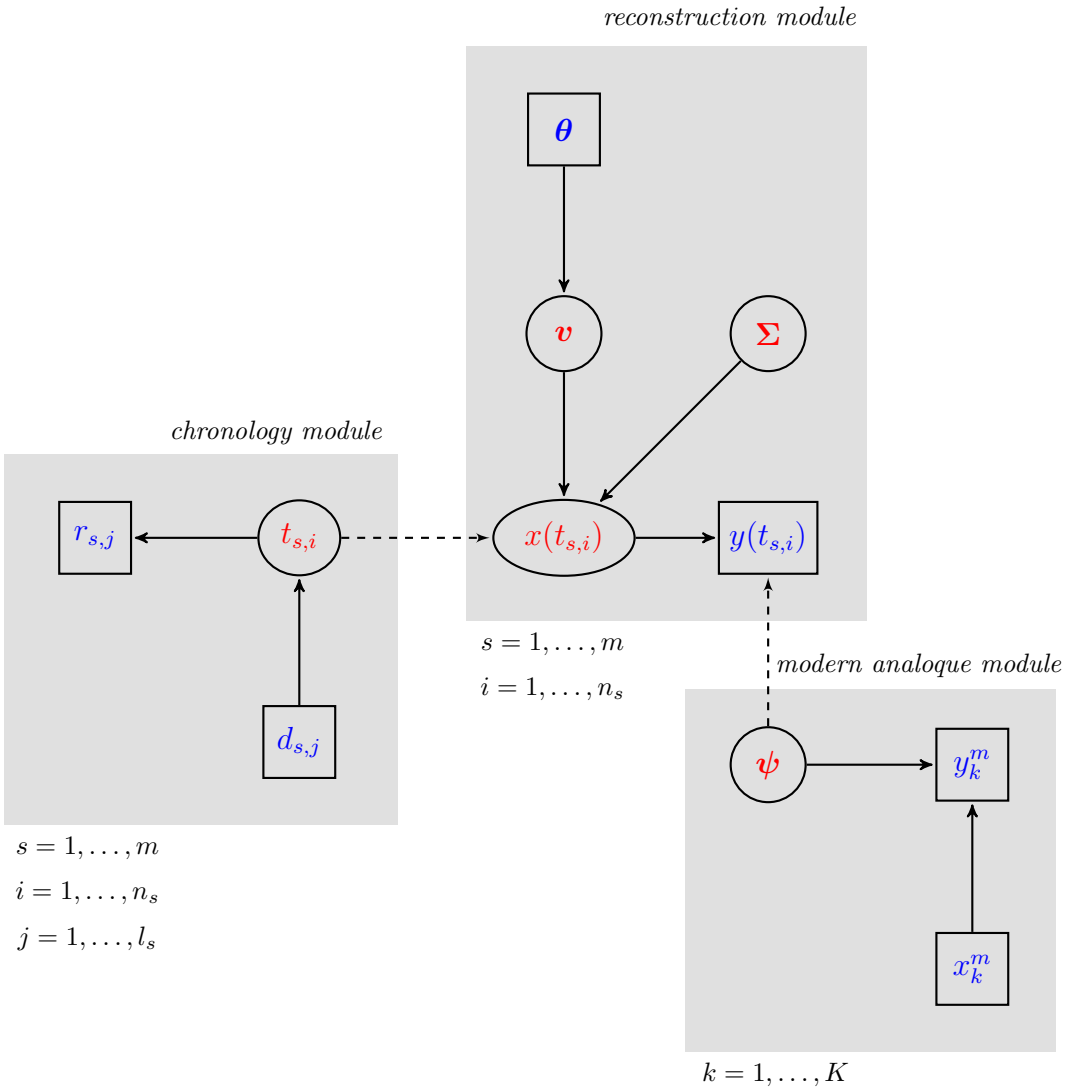


Fig. 5.2: A Directed Acyclic Graph (DAG) of the main model with different modules indicated in grey boxes. Circles indicate parameters/latent random variables whilst boxes indicate data. The solid lines indicate the direction of information flow, whilst the dashed lines indicate relationships where modularisation occurs.

5.3.1.1 Modern analogue module

This subsection deals with the training data set comprising of known climate and pollen information, as derived in Eq. (5.6):

$$\pi(\boldsymbol{\psi}|\boldsymbol{x^m}, \boldsymbol{y^m}) \propto \pi(\boldsymbol{y^m}|\boldsymbol{x^m}, \boldsymbol{\psi})\pi(\boldsymbol{\psi})$$

The model used in this chapter is a proposal of Sweeney (2012), in which $\pi(\boldsymbol{y^m}|\boldsymbol{x^m}, \boldsymbol{\psi})$ is the nested, zero-inflated Negative Binomial (ZINB), and $\pi(\boldsymbol{\psi})$ is *a priori* an intrinsic Gaussian Markov process in multivariate dimensional climate space. Sweeney (2012) deploy the Gaussian Markov random field approximation (Rue & Held 2005, Chapter 4.4) to approximate the full conditional distribution of $\boldsymbol{\psi}$ as Gaussian. Higher level parameters for $\boldsymbol{\psi}$ (not shown here) are analytically evaluated via the Laplace approximation (Rue et al. 2009).

5.3.1.2 Chronology module

For this module, we choose to use the model developed by Haslett & Parnell (2008), although any of the alternative models reviewed in Parnell et al. (2011) are also suitable. We begin by considering Eq. (5.7):

$$[\boldsymbol{t}|\boldsymbol{r}, \boldsymbol{d}] \propto \pi(\boldsymbol{r}|\boldsymbol{t})\pi(\boldsymbol{t}|\boldsymbol{d})$$

in which each posterior sample of $\boldsymbol{t}$ is termed a *chronology*. The first component on the RHS of the above equation is the likelihood of the radiocarbon ages, reported in the form of Gaussian distributions. In the second component, the Compound Poisson-Gamma process is assigned as the prior distribution for the increments of calendar ages. More specifically,

$$t_i - t_{i-1} = \sum_{j=1}^{N(d_i - d_{i-1})+1} \gamma_j \tag{5.10}$$

where each random number $N(d_i - d_{i-1})$ is modelled as a Poisson distribution, and an increment γ_i is modelled as a Gamma distribution which depends on another layer of unknown parameters (not shown here). We refer the reader to Haslett & Parnell (2008) for further details concerning this module.

5.3.1.3 Marginal data posteriors and reconstruction module

The reconstruction module requires the evaluation of (5.8). We rewrite this equation as

$$\pi(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\Sigma}, \boldsymbol{t}, \boldsymbol{\psi}| \ldots) \propto \pi(\boldsymbol{\psi}|\boldsymbol{x^m}, \boldsymbol{y^m})\, \pi(\boldsymbol{t}|\boldsymbol{r}, \boldsymbol{d}) \times \prod_{s=1}^{m} \prod_{i=1}^{n_s} \pi\Big(y(t_{s,i})|x(t_{s,i}), \boldsymbol{\psi}\Big)\, \pi(\boldsymbol{x}|\boldsymbol{v}, \boldsymbol{t}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{v}|\boldsymbol{t}, \boldsymbol{\theta})\, \pi(\boldsymbol{\Sigma}) \tag{5.11}$$

To briefly recap, the first term on the RHS of the above equation is the modern analogue module, the second term is the chronology module, each component of the third term is the ZINB distribution, the fourth term is the Gaussian distribution, the fifth term the $\mathcal{IG}$ distribution and the last term is a flat distribution of the cross-correlation coefficient ρ on a logit transformation scale.

Importantly, the term $\boldsymbol{x}$ appears twice on the RHS of (5.11), in the third and fourth term. As mentioned, the fourth term has a Gaussian distribution. Hence, it is possible to analytically marginalise $\boldsymbol{x}$ out from the reconstruction module if the third term can also be written in the form of a Gaussian distribution. This is the main purpose for the use of of '*marginal data posterior*' (MDP Parnell et al. 2015), a concept that was generically discussed in Section 2.3.3.1.

Marginal data posteriors Recall that the low-dimensional parameter $\boldsymbol{\psi}$ is learnt exclusively from a training data set. Thus, we can use Monte Carlo methods to numerically integrate out $\boldsymbol{\psi}$ from each conditionally independent term $\pi(y(t_{s,i}|x(t_{s,i}), \boldsymbol{\psi})$ as follows:

$$\pi\Big(y(t_{s,i})|x(t_{s,i})\Big) \propto \int \pi\Big(y(t_{s,i})|x(t_{s,i}), \boldsymbol{\psi}\Big)\, \pi(\boldsymbol{\psi}|\boldsymbol{x^m}, \boldsymbol{y^m})\, d\boldsymbol{\psi} \tag{5.12}$$

Subsequently,

$$\pi\Big(x(t_{s,i})|y(t_{s,i})\Big) \propto \pi\Big(y(t_{s,i})|x(t_{s,i})\Big)\pi\Big(x(t_{s,i})\Big) \approx \pi\Big(y(t_{s,i})|x(t_{s,i})\Big) \tag{5.13}$$

for a suitable prior distribution for $x(t_{s,i})$. Note that in our framework, prior distributions are assigned on the increments of $\boldsymbol{x}$ such that the prior for its marginals are flat. The LHS of (5.13) is a MDP. It is in fact a reconstruction of climate at a layer (or an uncertain point in time) of a core, conditional on pollen data at that layer.

In full generality the MDP samples considered in this work are non-Gaussian distributions. Parnell et al. (2015) propose further treatment for non-Gaussian MDPs, by considering the following Gaussian mixture approximation:

$$\pi\Big(x(t_{s,i})|y(t_{s,i})\Big) \approx \sum_{g=1}^{G} \kappa_{s,ig}\, \mathcal{N}\Big(x(t_{s,i}); m(t_{s,ig}), \lambda(t_{s,ig})\Big) \tag{5.14}$$

where, for time slice t_i from core s, m is a mixture mean, λ is a mixture variance, $g = 1, \ldots, G$ with G being the total number of mixture components, and κ is a weight for the associated mixture component. In practice, this step is performed using R package Mclust; estimation of the mixture components is via an Expectation-Maximisation algorithm and does not require MCMC simulations (Fraley et al. 2012). There is no problem with regarding to 'label switching that is often encountered in mixture estimation as we are content with a point solution. We follow the recommendation of Parnell et al. (2015) and fix G with value 10. Other parameters are functions of observation $y(t_{s,i})$. However for brevity of presentation this dependency is suppressed in the notation.

Reconstruction module After marginalising out $\boldsymbol{\psi}$, deriving and approximating the MDPs as Gaussian mixtures, (5.11) has the new form:

$$\pi(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\Sigma}, \boldsymbol{t}| \ldots) \propto \pi_{MDP}(\boldsymbol{x}|\boldsymbol{y})\pi(\boldsymbol{x}|\boldsymbol{v}, \boldsymbol{t}, \boldsymbol{\Sigma})\pi(\boldsymbol{v}|\boldsymbol{t}, \boldsymbol{\theta})\pi(\boldsymbol{t}|\boldsymbol{r}, \boldsymbol{d})\pi(\boldsymbol{\Sigma}) \tag{5.15}$$

where $\pi_{MDP}(\boldsymbol{x}|\boldsymbol{y}) = \prod_{s=1}^{m}\prod_{i=1}^{n_s} \pi\Big(x(t_{s,i})|y(t_{s,i})\Big)$. Following Parnell et al. (2015, Appendix A), an auxiliary vector $\boldsymbol{\alpha} = \{\alpha_{s,ig}; s = 1, \ldots, m; i = 1, \ldots, n; g = 1, \ldots, G\}$ is introduced at this stage. The g^{th} component of $\boldsymbol{\alpha}_{s,i}$ has a value of 1 if the MDP sample is in mixture component g, and 0 otherwise. The goal here is to retrieve $\pi(x(t_{s,i})|y(t_{s,i}), \boldsymbol{\alpha}_{s,i})$ as a Gaussian distribution to permit simple analytical marginalisation of $\boldsymbol{x}$ from (5.15). This full conditional distribution can now be rewritten as

$$\pi(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\Sigma}, \boldsymbol{t}, \boldsymbol{\alpha}| \ldots) \propto \pi_{MDP}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha})\pi(\boldsymbol{x}|\boldsymbol{v}, \boldsymbol{t}, \boldsymbol{\Sigma})\pi(\boldsymbol{v}|\boldsymbol{t}, \boldsymbol{\theta})\pi(\boldsymbol{t}|\boldsymbol{r}, \boldsymbol{d})\pi(\boldsymbol{\Sigma})\pi(\boldsymbol{\alpha}|\boldsymbol{\kappa}) \tag{5.16}$$

where each of $\boldsymbol{\alpha}_{s,i}$ is a multinomial distribution with parameters $\boldsymbol{\kappa}_{s,i}$. The vector $\boldsymbol{\kappa} = \{\kappa_{s,ig}; s = 1, \ldots, m; i = 1, \ldots, n; g = 1, \ldots, G\}$ has already been learnt from the Gaussian mixture approximation step discussed in (5.14).

It is worth re-emphasising that we model times as continuous. Therefore, it is not possible to obtain the same time point across different cores. Inevitably, temporal misalignment in the MDPs occurs if we consider multiple cores together. To remedy this, we are going to use the notation trick that was introduced in Section 3.1.2 and again in Section 4.2.1.

For each chronology sample $\boldsymbol{t} = \{t_{s,1}, t_{s,2}, \ldots\}$, we suppose $\boldsymbol{t_o} = \text{sort}\{\boldsymbol{t}\} = \{t_1, t_2, \ldots, t_n\}$ is the sorted set of all unique times at all cores. Correspondingly, $y_o(t_{s,i})$ is an observation that is precisely $y(t_{s,i})$ when core s has an observation at t_i, otherwise missing. In vector form, these are $\boldsymbol{y_o}$ (length mn, as opposed to $\boldsymbol{y}$ which has length N). All vectors in this thesis are stacked in time. For instance, $\boldsymbol{y_o}(t)$ is a $m \times 1$ vector of observations at time slice t which has typically only one observation and $m-1$ missing observations. Analogous to the definition of $\boldsymbol{y_o}$ are $\boldsymbol{x_o}$ and $\boldsymbol{v_o}$. Consequently, Eq. (5.16) emerges as

$$\pi(\boldsymbol{x_o}, \boldsymbol{v_o}, \boldsymbol{\Sigma}, \boldsymbol{t_o}, \boldsymbol{\alpha}| \ldots) \propto \pi_{MDP}(\boldsymbol{x_o}|\boldsymbol{y_o}, \boldsymbol{\alpha})\pi(\boldsymbol{x_o}|\boldsymbol{v_o}, \boldsymbol{t_o}, \boldsymbol{\Sigma})\pi(\boldsymbol{v_o}|\boldsymbol{t_o}, \boldsymbol{\theta})\pi(\boldsymbol{t_o}|\boldsymbol{r}, \boldsymbol{d})\pi(\boldsymbol{\Sigma})\pi(\boldsymbol{\alpha}|\boldsymbol{\kappa}) \tag{5.17}$$

The number of unknown terms in Eq. (5.17) is $n \times m$ for $\boldsymbol{x_o}$, n for $\boldsymbol{t_o}$, $n-1$ for $\boldsymbol{v_o}$, n for $\boldsymbol{\alpha}$ and the parameter space of $\boldsymbol{\Sigma}$ which is 1. The new notation allows us to write the joint distribution of the MDPs in vector form as

$$\pi_{MDP}(\boldsymbol{x_o}|\boldsymbol{y_o}, \boldsymbol{\alpha}) \;\sim\; \mathcal{N}(\boldsymbol{m}_\alpha, \boldsymbol{\Lambda}_\alpha^{-1}) \tag{5.18}$$

Here, $\boldsymbol{m}_\alpha$ is the vector of known mixture means, and $\boldsymbol{\Lambda}_\alpha$ is a diagonal matrix of known mixture precisions, whose values are conditional on the allocation given by $\boldsymbol{\alpha}$, and zeros where there are no data. This notation trick also applies to the multivariate $\mathcal{NIG}$ model and we can write:

$$\pi(\boldsymbol{x_o}|\boldsymbol{v_o}, \boldsymbol{t_o}, \boldsymbol{\Sigma}) \;\sim\; \mathcal{N}(\boldsymbol{0}, \mathbf{Q}_{v_o}^{-1} \otimes \boldsymbol{\Sigma}) \tag{5.19}$$

Here, $\mathbf{Q}_{v_o}$ denotes the precision matrix of the random walk of first order, $\otimes$ is the Kronecker product, and $\boldsymbol{\Sigma}$ is the cross-correlation between cores.

$$\mathbf{Q}_{v_o} = \begin{pmatrix} v_1^{-1} & -v_1^{-1} & & \\ -v_1^{-1} & v_1^{-1} + v_2^{-1} & -v_2^{-1} & \\ & & \ddots & \\ & & -v_{n-1}^{-1} & v_{n-1}^{-1} \end{pmatrix}$$

The fact that both (5.18) and (5.19) can be written as Gaussian distributions results in a trivial calculation, up to a normalising constant, of the full conditional for $\boldsymbol{x_o}$:

$$\pi(\boldsymbol{x_o}|\ldots) \;\sim\; \mathcal{N}(\boldsymbol{\mu}_{x_o|y_o}, \mathbf{Q}^{-1}_{x_o|y_o}), \tag{5.20}$$

where $\mathbf{Q}_{x_o|y_o} = \boldsymbol{\Lambda}_\alpha + \mathbf{Q}_{v_o} \otimes \boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\mu}_{x_o|y_o}$ is the solution to $\mathbf{Q}_{x_o|y_o}\boldsymbol{\mu}_{x_o|y_o} = \boldsymbol{\Lambda}_\alpha \boldsymbol{m}_\alpha$. Thus, after deriving $\boldsymbol{t_o}$ and $\boldsymbol{\psi}$ from the chronology and modern analoque module respectively, we analytically marginalise $\boldsymbol{x_o}$ out from (5.17) and are left with:

$$\begin{aligned} \pi(\boldsymbol{v_o}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}|\ldots) \;\propto\; |\boldsymbol{\Lambda}_\alpha|^{1/2} \prod_{i=1}^{n-1} |v_i \boldsymbol{\Sigma}|^{-1/2} \;\; \left|\mathbf{Q}_{x_o|y_o}\right|^{-1/2} \\ \times \;\; \exp\left(\frac{1}{2}\boldsymbol{m}_\alpha^T \boldsymbol{\Lambda}_\alpha (\boldsymbol{\mu}_{x_o|y_o} - \boldsymbol{m}_\alpha)\right) \pi(\boldsymbol{v_o}|\boldsymbol{\theta}) \; \pi(\boldsymbol{\alpha}|\boldsymbol{\kappa}) \; \pi(\boldsymbol{\Sigma}) \end{aligned} \tag{5.21}$$

We present, in Appendix B, the full MCMC algorithm to sample for the unknown processes and parameters from (5.21). The key advantage of our algorithm over that developed by Parnell et al. (2015) is the modelling extension to the multiple core setting, and computational ability to deal with terms involving Kronecker products. Thereafter, realisations of $\boldsymbol{x_o}$ from (5.20) are easily obtained using the plug-in samples of the parameters from (5.21).

The final step involves stochastic interpolation of climate and climate volatilities at an arbitrary time grid, based on the MCMC samples discussed above. We suppose $\boldsymbol{t_g}$, $\boldsymbol{x_g}$ and $\boldsymbol{v_g}$ denote the known time grid, unknown climate and unknown squared volatility process of interest. We also let $\boldsymbol{y}_* = \left(\boldsymbol{y_o}^T, \boldsymbol{y_g}^T\right)^T$, $\boldsymbol{x}_* = \left(\boldsymbol{x_o}^T, \boldsymbol{x_g}^T\right)^T$, $\boldsymbol{t}_* = \left(\boldsymbol{t_o}^T, \boldsymbol{t_g}^T\right)^T$ and $\boldsymbol{v}_* = \left(\boldsymbol{v_o}^T, \boldsymbol{v_g}^T\right)^T$. Stochastic interpolation is the computation of

$$\begin{aligned} \pi(\boldsymbol{x_g}|\ldots) &= \int \pi(\boldsymbol{x}_*, \boldsymbol{t}_*, \boldsymbol{v}_*, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{\psi}|\ldots)\, d\boldsymbol{x_o}\, d\boldsymbol{t}_*\, d\boldsymbol{v}_*\, d\boldsymbol{\Sigma}\, d\boldsymbol{\alpha}\, d\boldsymbol{\psi} \\ &\propto \int \pi(\boldsymbol{x_g}|\boldsymbol{x_o}, \boldsymbol{v_g}, \ldots)\pi(\boldsymbol{v_g}|\boldsymbol{v_o}, \ldots)\, \pi(\boldsymbol{x_o}|\boldsymbol{y_o}, \boldsymbol{t_o}, \boldsymbol{v_o}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\psi}) \\ &\quad \pi(\boldsymbol{v_o}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{t}_*, \boldsymbol{\theta}, \boldsymbol{\psi})\, \pi(\boldsymbol{t}_*|\boldsymbol{r}, \boldsymbol{d})\, \pi(\boldsymbol{\psi}|\boldsymbol{x}^{\boldsymbol{m}}, \boldsymbol{y}^{\boldsymbol{m}})\, d\boldsymbol{x_o}\, d\boldsymbol{t}_*\, d\boldsymbol{v_o}\, d\boldsymbol{\Sigma}\, d\boldsymbol{\alpha}\, d\boldsymbol{\psi} \end{aligned} \tag{5.22}$$

This is a high dimensional integration problem and stochastic simulation presents the best solution. Integration with respect to $\boldsymbol{t}_*, \boldsymbol{v_o}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ have been performed via the MCMC algorithm described in Appendix B. On the other hand, evaluation of the first quantity inside the integrand (5.22) involves, essentially, sampling from a multivariate Brownian bridge. Whilst evaluation of the second quantity inside the integrand is sampling from an $\mathcal{IG}$ bridge. We present our bridging algorithm in Appendix

C. The objective is to draw equally probable climate histories that are consistent with both data and model assumptions. These represent climate data products as discussed in Section 2.1.3.

5.3.2 Summary of the modularised MCMC algorithm

Our modularised approach permits the inference procedure to be carried out in three different stages separately. In summary, the basic framework to transform raw proxy data to climate data products involves three main steps.

1. ***Modern analogue module*** (Section 5.3.1.1). This module focus on building a climate-pollen relationship using the training data set comprising of known pollen information and associated climate at sites around the world. The output of this layer is a 'response surface' and a selection of parameters controlling the relationship between climate and pollen. This is a once-off job; the procedure is discussed in Sweeney (2012) and the results can be loaded from R package Bclim (Parnell et al. 2015, see also `http://cran.r-project.org/web/packages/Bclim/index.html`).

2. ***Chronology module*** (Section 5.3.1.2). The purpose of this module is to draw inferences about calendar ages based on radiocarbon dating information at some depths of the three sediment cores. Radiocarbon determinations and depths are used as the input for R package Bchron (Haslett & Parnell 2008, see also `http://cran.r-project.org/web/packages/Bchron`). The output are chronologies samples associated with the depths of the MDPs. In this thesis, we use an independent chronology model for each core.

3. ***Reconstruction module*** (Section 5.3.1.3). Initially, we use suitable functions from R package Bclim to create layer-independent climate, given the fossil pollen counts from the three sites in Finnmark, the response surface and appropriate parameters. Subsequently, these layer-by-layer climate are standardised, and approximated as mixtures of Gaussian using R package Mclust (Fraley et al. 2012, see also `http://cran.r-project.org/web/packages/mclust`). At the end of this step we have analytical expression for the MDPs (in the form of Gaussian mixture distributions) at a selection of depths of the three cores; the

depths will have been transformed to calendar ages via the chronology module. Note that the MDPs are independent across all cores and depths.

The steps hereafter represent the main novelties of this chapter; we propose to use all of the results above as the inputs for our algorithms in order to jointly reconstruct climate at multiple sites. Firstly, we use the MCMC algorithm described in Appendix B to sample for all unknown terms in the model, conditional on all data. Secondly and as a final step of the overall process, we apply the bridging algorithm described in Appendix C to sample for climate and climate volatility at a regular time grid across all three cores. For the purpose of presentation, we transform the posterior climate samples back to their original scale. It is unclear what the scale of volatility is hence we report it in its standardised climate scale.

Figure 5.3 is a schematic representation of all of the above steps. This plot closely follows the DAG shown in Fig. 5.2.

5.4 Application to three Finnmark pollen data sets

This section demonstrates an application of the methods discussed in the previous sections using the Finnmark data set.

5.4.1 Background

The data set comprises of three fossil sediment cores found at separate lakes in Finnmark, Norway. Each core contains pollen grains at several depths, and radiocarbon dating determinations often at fewer number of depths. The former provides information about historical climate over many thousands of years at that location, and the latter gives information about times. Recently, Huntley et al. (2013) used this data set to investigate whether there is a variability in ancient climate between these sites. The authors use the so-called 'direct analogues method' to obtain point estimate of climate. By design this method does not quantify the uncertainty associated with the climate estimate, nor uncertainty in the parameters of their model. Moreover, the analysis was carried out on data set from each site separately.

Bayesian statistical inference is thus preferred. Minimally, it allows easy mixing of uncertainty from many different data sources. Most importantly, spatio-temporal

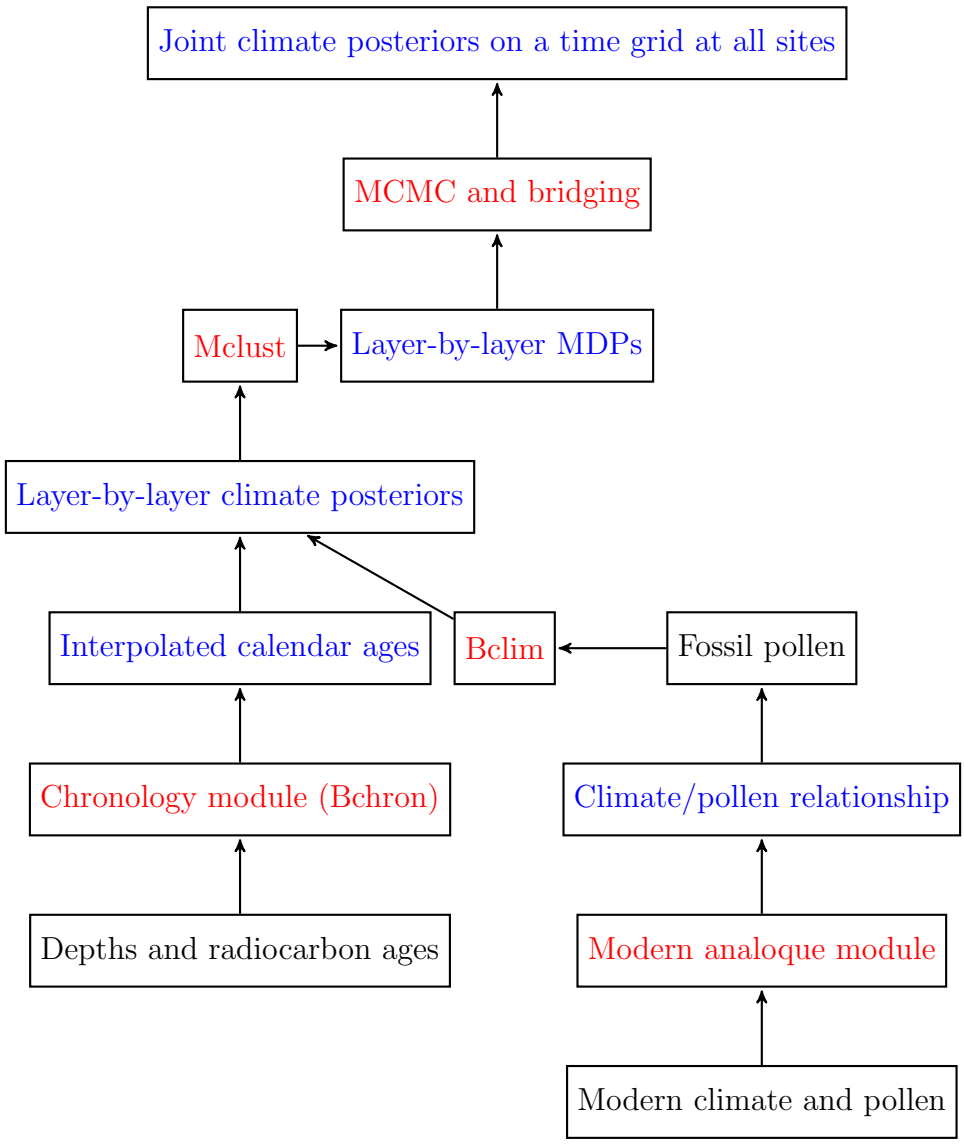


Fig. 5.3: A schematic representation of the full algorithm to reconstruct ancient climate based on pollen data and radiocarbon dates in multiple fossil pollen cores, and a modern data set comprising of known climate and pollen responses. Data are indicated in black, latent processes are indicated in blue, and models/algorithms are indicated in red.

dependency will be taken into account since data from three locations are likely to be correlated in both space and time. Our objective is to make meaningful probabilistic statements about the unknown ancient climate conditions by taking into account all (or as many as possible) uncertainties inherent in the data.

Our approach, being an extension to that developed by Parnell et al. (2015), produces results for multi-dimensional climate. For clear presentation of the methodology,

we only discuss the results for the measure of mean temperature of the coldest month (MTCO), a quantity indicative of the harshness of the winter.

5.4.2 Model-fitting results

For each core, the chronology module generates ten thousands posterior samples of chronologies representing ten thousands possible histories of times. For each core, the modern analoque module produces a vector of MDPs; each of which is a list of mixture means, precisions and weights. Figure 5.4 shows a plot of a chronology and the corresponding MDPs; we plot the latter using the mixture means associated with the highest mixture weights at a time point.

The output from the modern analogue and chronology module are used as input for the reconstruction module. Figure 5.5 shows several types of results. The pointwise 95% credible intervals of DMPs at all corresponding chronology samples are represented in blue 'blobs'. Our approach's pointwise 50% and 95% credible intervals of interpolated centennial MTCO histories over the period 0 to 14 k cal BP are represented in red. A climate history is represented by the black lines. The results obtained using the methods of Huntley et al. (2013) (green dots) and Parnell et al. (2015) (blue credible bands) are also shown since these are obvious candidates for direct comparison with our method.

It is easy to see that the levels of uncertainty vary across different outputs. Firstly, the MDPs are both time-independent and site-independent. Hence their marginal uncertainty bands are widest. Secondly, the separate inference approach by Parnell et al. (2015) produces joint posterior samples that are temporally correlated but spatially independent across all sites. Therefore their predictive credible intervals are *typically* smaller compared to those of the MDPs, but wider in comparison to those of our model (see the next paragraph for further discussion of this). Thirdly, the degree of uncertainty also depends upon the temporal resolution of the data sets. Here, for example, site NKH1 has the highest resolution but the oldest depth is about 10.5 k cal yr BP. Consequently, the uncertainty estimates of climate from year 0 k to 10.5 k is narrower than those after 10.5 k cal yr BP. Intuitively, this is most visible in the separate inference approach, i.e. interpolation versus extrapolation. Finally, we note that the point estimates by Huntley et al. (2013) give no measure of uncertainty.

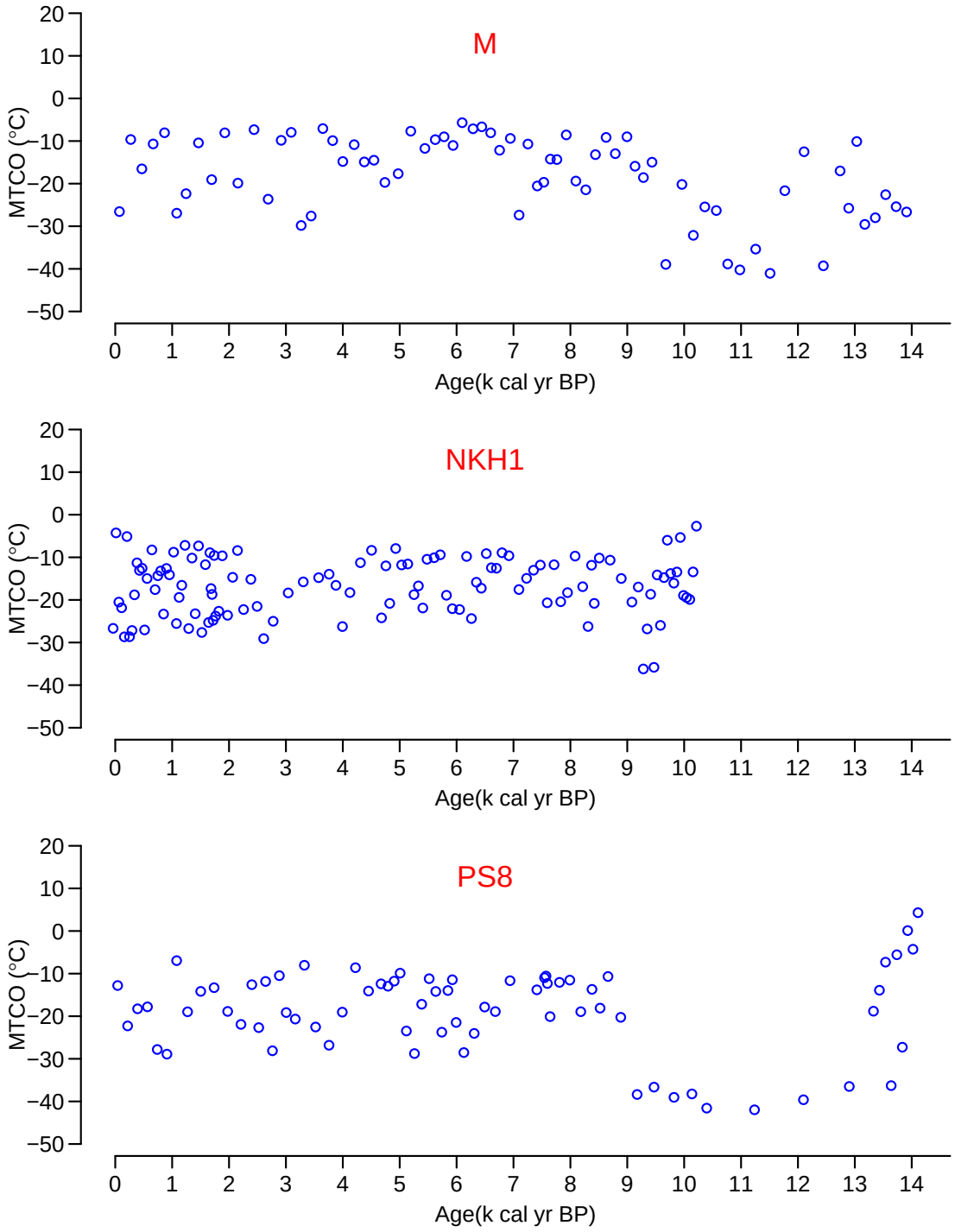


Fig. 5.4: A plot of a chronology sample and its corresponding mean temperature of the coldest month (MTCO) for three sites in Finnmark. At each time point, the MTCO value is represented by the marginal data posteriors' (MDPs) mixture mean that is associated with the highest mixture weight.

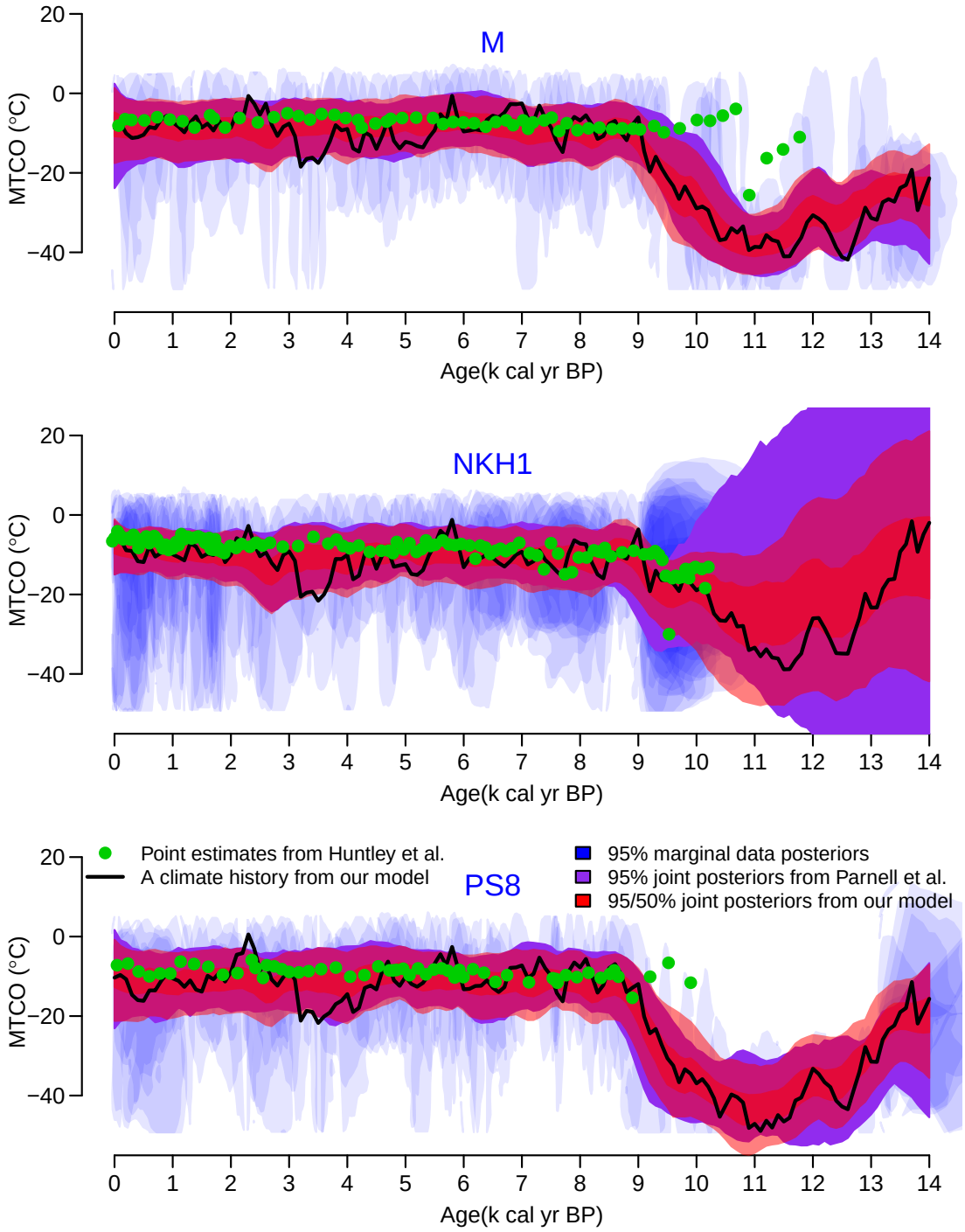


Fig. 5.5: Plots of MTCO over the period 0 to 14 k cal yr BP at three sites in Finnmark. The pointwise 95% credible intervals of DMPs at all corresponding chronology samples are represented in blue 'blobs'. The pointwise 95% credible intervals of climate histories from the model of Parnell et al. (2015) are represented in purple. Our pointwise 50% and 95% credible intervals are represented in red. A climate history from our approach is represented by the black lines. Finally, point estimates from Huntley et al. (2013) are the green dots.

Absolute differences in the interquartile range (IQR) of the posterior marginals (or pointwise summaries) between our approach and the one-core-a-time inference are represented in Fig. 5.6. There are broadly two case scenarios. In the first case, the joint model yields reduced uncertainties where there is between-core correlation in the data. Conversely, in the second case, it yields increased uncertainty where information sources diverge. The latter case also demonstrates the limitation of the joint model, arising from sharing of volatility rather than making it flexible to each site. This restriction is useful when the volatility across all three sites are approximately equal and data are empirically correlated. Otherwise new (possibly biased) information can be introduced resulting in higher uncertainty levels. Most crucially, the advantage of our approach over the one-core-at-a-time alternative is that we are able take into account both aforementioned case scenarios. This supports our claim that combining information from multiple spatial data sources yields more reliable estimates of uncertainty.

5.4.3 Case study of the Younger Dryas event

From Fig. 5.5, we can see an extreme cold interval at approximately 9 to 13.5 k cal yr BP. This period corresponds to an extreme climate change event known as the Younger Dryas (e.g. Muscheler et al. 2008). This event represents a period of cold that was rapidly followed by a warming in temperatures. We propose to use our data products to answer some questions about YD in a similar manner to that for the 8.2ka event as discussed in Section 4.3.3. We note that the 8.2ka event is not clearly visible in Fig. 5.5, most likely due to the substantial amount of uncertainty presented in pollen data.

The interesting questions here are concerned with the period of maximum change from cold to warm, and from warm to cold. We will use the period between these two maxima as our estimation for the duration of YD. Our estimation is based on an ensemble of simulations of interpolated climate between 8.5 and 14k cal yr BP, a conservative period bracketing YD. More specifically, we focus on the distributions of random variables $t_{\max \triangle} = \max_t \left(\triangle x(t)\right)$ and $t_{\min \triangle} = \arg_t \min \left(\triangle x(t)\right)$, being respectively the time of maximum positive change and maximum negative change. We compute these estimates for each history and summarise the results from 1000 simulated climate histories.

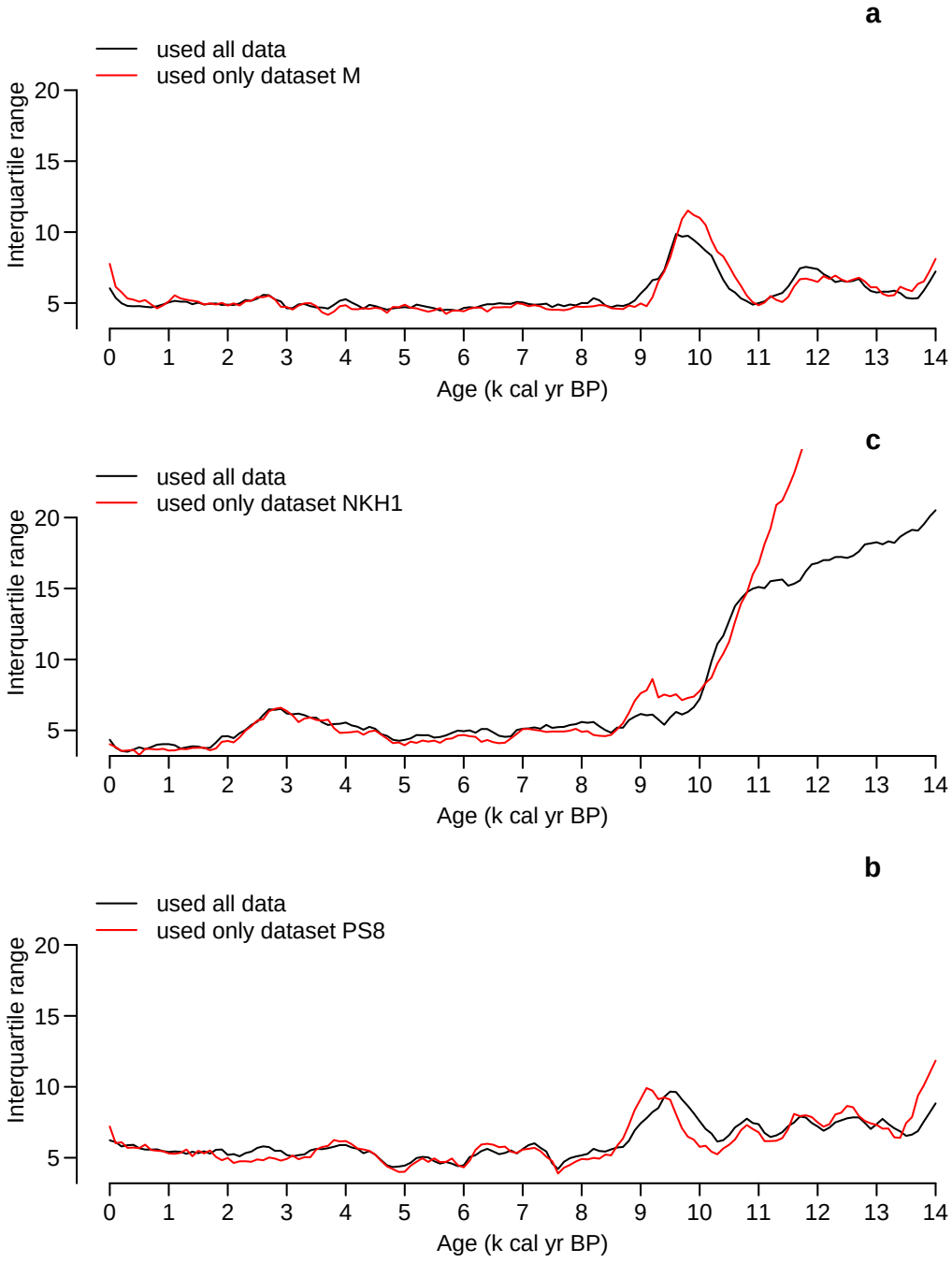


Fig. 5.6: Plots of the interquartile ranges (IQR) of centennial histories of MTCO over the period of 0 to 14 k cal yr BP at three sites in Finnmark. The IQR of the one-core-at-a-time model are generally higher than those of the joint model. However, the converse may occurs in certain circumstances. See the text for further details.

Figure 5.7 shows estimates of the century with the maximum and minimum MTCO changes. The IQRs (in units of k cal yr BP) for centuries with maximum changes are [9, 10.3], [8.9, 11.6] and [8.8, 10.1] for site M, NKH1 and PS8 respectively. Similarly, timings of centuries with minimum changes are [10.7, 12.6], [9.9, 12.6] and [11.1, 12.8]. Additionally, we compute the summary of $t_{\max \triangle} - t_{\min \triangle}$ to estimate the time interval which contain the largest negative followed by positive changes, as represented in Fig. 5.8. These are estimates of the length of YD: 2.3, 2.5 and 2.6 k cal yr for site M, NKH1 and PS8 respectively. Our estimates are wider comparing to climatologically theory (e.g. Fiedel 2011, Muscheler et al. 2008, Gulliksen et al. 1998). This is due to the fact that we attempt to incorporate several sources of uncertainty whilst, crucially, there are only data from three sites. Hence, we feel that this approach can be considered as a proof-of-concept method. More spatial information are necessary for a more comprehensive investigation into the spatial extent of abrupt climate changes in a global context.

5.5 Sensitivity analysis

This section checks if the model parameters are identifiable and if the inference procedure gives rise to satisfactory results. A model is non-identifiable if distinct values for its parameters generate the same probability distribution for the data.

We simplify the relationship between climate and proxy, and assume no error in the time scale. The model specification from Section 5.2 and its variation are applied to generate several versions of data sets. We then apply the inference procedure discussed in Section 5.3 to a obtain posterior distribution of all unknown parameters and latent processes. In each simulation, we find out if the true climate lie within a posterior summary.

At each simulation, we draw 99 trios of climate from a tri-variate Gaussian independent increments model regularly spaced, fixed time points. The shared variance of increments are generated from an Inverse Gaussian distribution with parameters θ_1 and θ_2. The cross-correlation matrix $\mathbf{\Sigma}$ is a uniform correlation matrix with ones in the main diagonal and the off-diagonal is filled with a correlation coefficient ρ. We use flat priors for climate as the starting time points. We then create pseudo-proxy $\boldsymbol{y}$

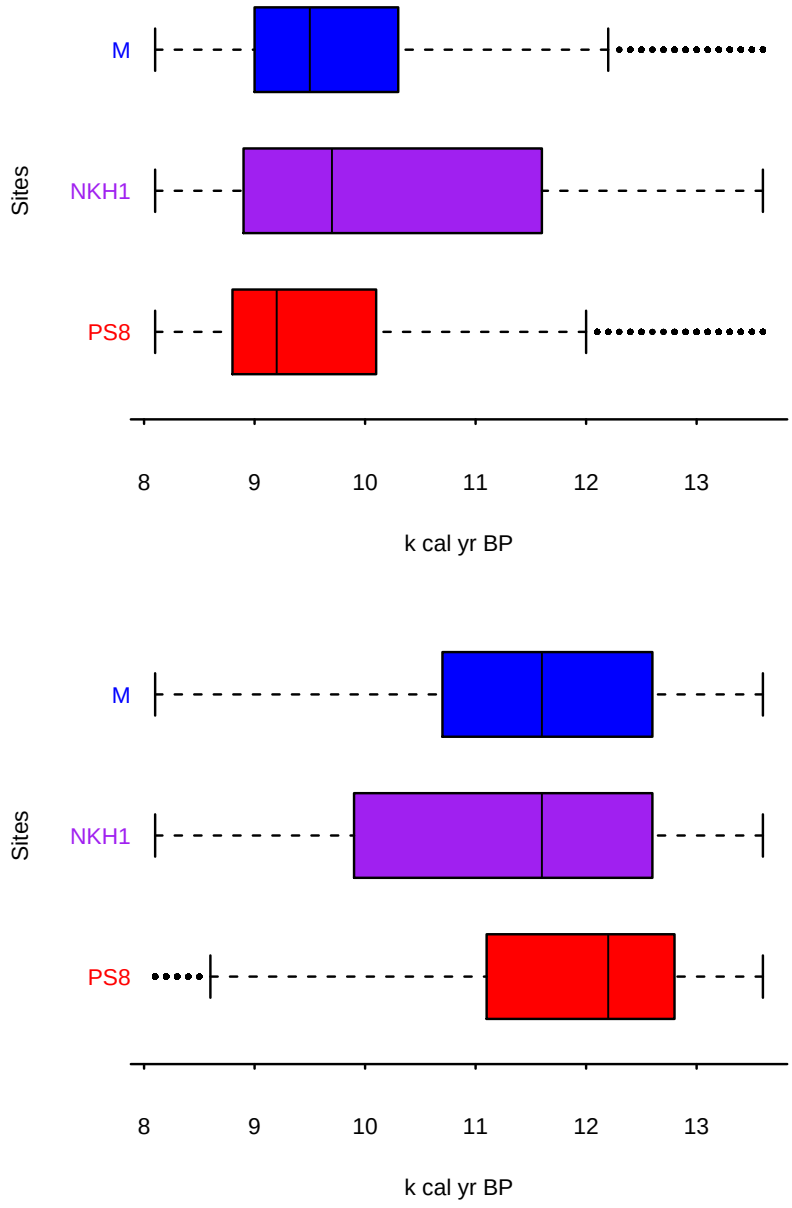


Fig. 5.7: Boxplots of estimates of the century with the maximum **(a)** positive change and (*b*) negative change. The underlying process is the centennial MTCO, computed from process histories between 8.5 - 14 k cal yr BP at three sites in Finnmark.

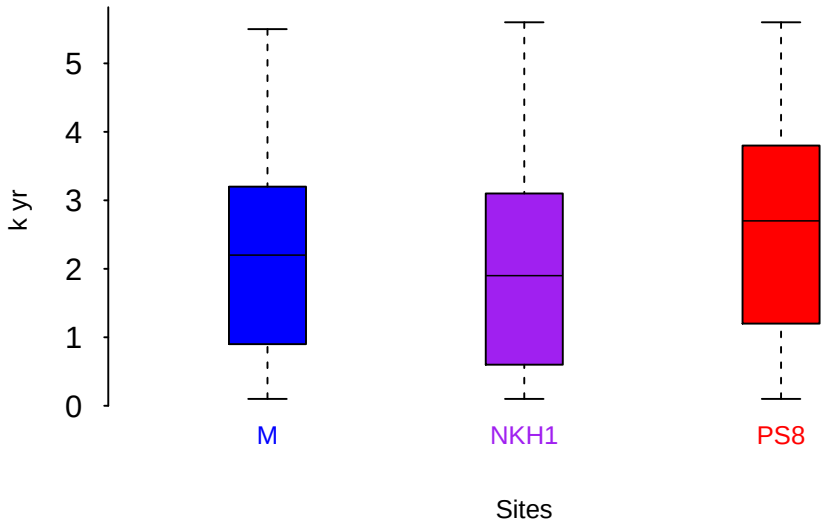


Fig. 5.8: Boxplots of time intervals which contain the largest negative followed by largest positive change in the centennial MTCO, computed from process histories between 8.5 - 14 k cal yr BP at three sites in Finnmark.

from climate using appropriate likelihood models. Temporal misalignment is created by randomly choosing 33 data points for each series. Specific details of the simulations in each scenario are as follows.

1. Case 1. From $\theta_1 = 3$, $\theta_2 = 55$, we generate $v(t_i) \sim \mathcal{IG}(\theta_1, \theta_2)$; $i = 1, \ldots, 98$ and $\rho \sim \mathcal{U}(0, 1)$. Starting from a flat prior for climate at time 1, we generate 99 trios of climate using the trivariate independent increment model and create temporal misalignment as described above. Finally, pseudo proxy $y(t_{s,i}) \sim \mathcal{N}(x(t_{s,i}), \tau_s)$; $s = 1, 2, 3$. Here, we use $\tau_s \sim \mathcal{U}(0.02, 2)$. From proxy information, creation of MDPs is trivial in this case; e.g., $\pi(x(t_{s,i})|y(t_{s,i}) \sim \mathcal{N}(y(t_{s,i}), \tau_s)$. We apply the rest of our modularised MCMC algorithm to derive information on climate $\boldsymbol{x}$.

2. Case 2. Squared volatilities and climate are generated as above. Here, we consider the zero-inflated Poison model for generation of proxy signals. We suppose $\boldsymbol{y}(t_{s,i}) \sim \mathcal{ZIP}(p_s, 3a_s\boldsymbol{x}^2(t_{s,i}))$ where $p_s \sim \mathcal{U}(0, 0.2)$ and $a_s \sim |\mathcal{N}(0, 1)|$. We use importance sampling to create MDPs, and use a 5 mixture components Gaussian approximation. As before, we derive posterior samples for climate using our modularised MCMC algorithm.

3. Case 3. The set up for this case is the same as the second case, except that the

number of mixture component is 2 (instead of 5).

4. Case 4. The set up for this case is the same as the second case. However, in the MCMC algorithm, θ_1 and θ_2 are fixed at values that are lower than the truth, i.e. $\theta_1 \in U(0.01, 2)$ and $\theta_2 \in U(0.01, 50)$.

5. Case 5. The set up for this case is the same as the second case. However, in the MCMC algorithm, θ_1 and θ_2 are fixed at values that are higher than the truth, i.e. $\theta_1 \in U(4, 10)$ and $\theta_2 \in U(70, 100)$.

We repeat each simulation and inference procedure 1000 times and summarise the results in Table 5.3. Under all considered case scenarios the model seems to perform well.

Scenario	Proportion inside 90% CI	Proportion inside 50% CI
1	89	49
2	91	54
3	91	52
4	90	50
5	91	55

Table 5.3: Performance of the model fitting algorithm. All results were based on 1000 simulation runs.

5.6 Discussion

This chapter has extended the work of Parnell et al. (2015) to reconstruct ancient climate from multiple fossil pollen cores jointly. The primary challenges under consideration are multivariate time series modelling with temporal uncertainty and stochastic volatility. We have provided a full Bayesian treatment to these problems. Our inference procedure is reasonably straightforward to implement as a result of modularisation. The modularisation approach permits easy mixing of several sources of uncertainty and modelling features. The adopted model is flexible. Some of its special cases are briefly discussed below.

Cross-correlation structure. We used, *a priori*, the uniform correlation matrix with ones in the diagonal and a common correlation coefficient ρ in the off-diagonal entries that can only take values between 0.7 and 1. The resulting posterior distribution is largely dependent on the specified prior, which implies that the likelihood contains little information. As a sensitivity analysis, we performed inference using *a priori* varying bounds. As expected, the resulting predictive error from interpolation remains largely unchanged while extrapolation intervals get wider with larger boundary ranges.

A obvious extension to the uniform correlation matrix with one common correlation parameter is that with two correlation parameters. In general, a matrix of correlations is not necessarily a valid correlation matrix because of the notorious requirement of non-negative definiteness. Mathematically, the determinant of correlation matrix must be strictly positive (Rousseeuw & Molenberghs 1994). Thus for a 3-by-3 matrix with correlation coefficient ρ_{12} and ρ_{13}, the correlation coefficients must satisfy the equation $1 + 2\rho_{13}\rho_{12}^2 - 2\rho_{12}^2 - \rho_{13}^2 > 0$, which implies that $2\rho_{12}^2 - 1 < \rho_{13} < 1$ or $0 < \rho_{12} < \sqrt{(1+\rho_{13})/2}$. As an application to the Finnmark data set, the posteriors for both parameters are flat and equal to their prior forms. Furthermore, a combination of joint analysis of two cores simultaneously offer no evidence of different correlation across the three sites. On this basis we feel that the uniform correlation function with one parameter is adequate for accurate modelling.

We also took into account of *spatial* dependency between the sites by using an isotropic exponential covariance model of the form $\exp(-||s_j - s_i||/\beta)$. Here s's are coordinates of the sites and $||.||$ is a physical distance, β is the range parameter controlling the rate of decay in correlation as a function of distance. We follow Gelfand et al. (2010, Chapter 7) and assign a discrete uniform prior for β. The bounds of the prior distribution are decided by the practical range of the data set which can be empirically estimated as 3ρ for an exponential model (Wackernagel 2003, chapter 8). The results from this model are similar to those from the case where a uniform correlation matrix is used. With only three series considered in the application, parametrisation of this form is less interpretable. Future work with an additional number of time series data sets may benefit from using this correlation function.

Marginal data posteriors (MDP) We experimented with different values for the number of mixture components G of the Gaussian approximation for the MDPs. The

case where $G = 1$ produces the narrowest joint posterior intervals of stochastically interpolated climate, and less 'bumpy' interpolated volatility. This effect wears off as we increase the value for G. Intuitively, the higher the number of mixtures the more accurate the approximation becomes. However, attention must also be paid to the trade off between accuracy and computational burden. For our final model, we follow Parnell et al. (2015) and fix $G = 10$.

Time uncertainty We fitted a multivariate $\mathcal{NIG}$ independent increments model with no time uncertainty, i.e. times of observations are known precisely. An even more simpler and special case of this model is the multivariate Gaussian independent increments model discussed in Chapter 4, again with no errors in the time scale. Under both model settings, we obtained narrower uncertainty levels for the climate posteriors, in comparison to the main model with time uncertainty. Notwithstanding, we believe that ignoring any available source of uncertainty can result in overconfidence in a statistical inference. Additionally, salient features of underlying processes can be overlooked if all available sources are not explicitly considered. Therefore we do not recommend ignoring time uncertainty when dealing with non-negligible dating information.

Chapter 6

Conclusion

The final chapter summarises the contributions of the thesis. It also discusses the outstanding challenges that beckon for further exploration.

6.1 Summary

In this thesis, new statistical methods have been proposed to directly address the joint behaviour of multiple time series in their misaligned form. Other major issues under consideration include data support, temporal uncertainty, and stochastic volatility.

All of our proposed models fall under the hierarchical framework. We began with a simple two-stage hierarchical structure in chapter 2, applied to an irregularly spaced time series with all parameters known. Then, in chapter 3, we extended this model to the case of multiple irregularly spaced series. Next, in chapter 4, we proposed a three-stage structure, with all parameters unknown, to deal with data from different series having different supports. Finally, chapter 5 further extended the model in all previous chapters to a four-stage hierarchical model, allowing for errors in the times of the observations, and stochastic volatility in an underlying latent process. This model also incorporates external knowledge from a separate set of training data for more accurate statistical inference.

Computation of posterior distributions for our models is challenging, particularly due to the complex modelling structure and high dimensional parameter space. To alleviate the computational burden, we have used a combination of the following assumptions and strategies. Firstly, we took advantage of nice properties from the Gaus-

sian Markov assumption for efficient storage and fast calculation of sparse matrices. Secondly, by combining the Gaussian Markov assumption with the concept of marginal data posteriors, we were able to marginalise out a high-dimensional latent process from the overall model, leaving it for inference in a separate stage. Thirdly, we exercised the concept of Bayesian modularisation to divide a big model into smaller modules.

The methodologies developed in this thesis should improve the quality of climate data products. They have been applied to raw and noisy oxygen isotopes from ice cores and pollen counts from sediment cores to create data products in the form of histories. The histories are realisations from non-Gaussian posterior predictive distributions, conditioning on all available data. In climate reconstruction, this is precisely as proposed by Tingley et al. (2012) and Chandler et al. (2012). We compared our results with those from the independent analysis of each core separately, both within our modelling framework and from the relevant literature. We concluded that our data products are richer than what have been previously available. We demonstrated the usefulness of our data products with case studies of extreme climate change events and found results that are consistent with the physical climate system. More broadly, this contribution has uses far beyond those presented in this thesis. We believe that our data products will be more useful in the hand of researchers in the field of climatology.

6.2 Future directions

There are a number of useful extensions to the methods presented in this thesis. We conclude this chapter by listing some ideas that could be pursued in the the future.

Varying data supports In chapter 4, we considered the case in which data are allowed to have different definitions of support across different series. But we constrained the supports for all data to be the same in each series. A generalisation of this problem is to allow the support of a data point to vary according to the underlying measurement procedure. This allows the value of k_s, as introduced in Section 4.1 and derived in Appendix A, to vary according to the empirical ratios of the supports of associated observations. More specifically, if there are m number of time series and each series has its own data support, there are $m-1$ known nugget parameters, all of which is a function of only one unknown nugget parameter. In the generic case where all $\sum_s^m n_s$

data points from m series have their own definition of supports, the number of nugget parameters will increase by the many different supports. However, as before, there will be only one unknown nugget effect parameter.

Space-time interpolation and computational extensions We believe the research in this thesis is only an initial attempt at joint palaeoclimate inference, upon which other work can be built. In this initial presentation, we illustrated our framework using three sediment cores. We did not take into account the physical locations where the time series were recorded. We have emphasised throughout that our model can be extended to a wide class of spatio-temporal processes, for instance, by modifying the cross-covariance function to include physical distances. The European pollen database beckons for attention from climate researchers. Hence, the continuation of this project is to apply the developed methods to the whole of Europe to produce a pan-European spatio-temporal map of climate history. This would allow us to gain insight into the spatial extent to which certain climate events happen across Europe.

Increasing the number of series/cores/sites will inevitably be associated with a higher computational burden, as the size of matrices increases at the rate of $\mathrm{O}(N^2)$ where N is total number of data points from all series. The number of parameters also increases quickly with the number of series. In particular, our proposed approach separates the inference of smoothed latent climate process from the gridded interpolation step, which seems rather inefficient as the number of series increases. Simply put in notation used in chapter 3, 4 and 5, the current approach requires the evaluation of $\pi(\boldsymbol{x_o}|\boldsymbol{y_o})$ before $\pi(\boldsymbol{x_g}|\boldsymbol{x_o})$. An alternative approach is to define a latent field on the grid in the first place, for instance, by re-writing $\boldsymbol{x_o}$ as a linear interpolator of $\boldsymbol{x_g}$. For example,

$$\boldsymbol{y_o} = \boldsymbol{x_o} + \boldsymbol{\epsilon} := \mathbf{B}\mathbf{x}_\mathbf{g} + \boldsymbol{\xi} + \boldsymbol{\epsilon} \tag{6.1}$$

Here, $\mathbf{B}$ is the so-called observation matrix that link $\boldsymbol{x_g}$ and $\boldsymbol{x_o}$ (Wikle & Berliner 2005). In our context, both $\boldsymbol{x_o}$ and $\boldsymbol{x_g}$ are independent increment Gaussian processes. Hence $\mathbf{B}$ is a strict linear interpolant. Times can be fixed, as considered in chapter 3 and 4, or random in 5. Furthermore, this approach necessarily introduces the interpolation error terms $\boldsymbol{\xi}$. In our modelling setting, it is easy write down the joint distribution for $\boldsymbol{x}_* = \left(\boldsymbol{x_o}^T, \boldsymbol{x_g}^T\right)^T$ and the derivation of $\boldsymbol{\xi}$ trivially follows. In other cases, this

calculation may significantly increase the overall computational complexity. A remedy is to ignore these terms by using a dense rectangular interpolation grid (Paciorek 2013), although it is not clear how fine the grid should be. Nonetheless, this is an interesting algorithmic aspect worthy of further investigation.

An obvious advantage here is that the associated matrices are always the same size irrespective of the size of the data set. At this point, the recently developed method by Lindgren et al. (2011) is useful as it permits modelling using the more flexible covariance function while performing inference using sparse precision matrices discretely defined by some Markov models. The additional benefit of working under the SPDE framework is its flexibility for extension. For instance, Bolin (2014) has extended the SPDE approach to a accommodate a version of the Normal Inverse Gaussian process that we considered in chapter 5.

Joint chronology modelling In chapter 5, we provided solutions to allow for uncertainty in radiocarbon dated sediment samples. Our framework is conceivably applicable to other proxy types that use different dating methods such as tree rings, ice cores, etc. A shortcoming of this approach is that we dealt with time uncertainty from each series/cores/sites separately. In extreme cases where the underlying sedimentation rates from crossed-correlated series are very different, there is less benefit from joint inference. However, if there is reason to believe that their sedimentation rates are similar, a separate model for each series effectively throws away information. Hence, one obvious thing to do is to extend the jointness of our existing model by modifying the prior model (5.10) to be a multivariate monotone stochastic process.

Multiple climate proxies We used point mass prior distributions for the parameter values at the highest hierarchical layer of the model in chapter 5. A more flexible prior specification would be to use the information learnt from ice cores in chapter 4 as prior information for the highest layer in chapter 5.

More generally, we have not considered utilising multiple proxies for our reconstruction. Each proxy has its own advantages and limitations, and combining different proxies may yield complementary information on different aspects of climate (Li et al. 2010). On the other hand, it is important to note that different proxies are sensitive to different climate aspects.

Other applications Finally, we note that interesting challenges arising from quantitative palaeoclimate reconstruction motivates the research of this thesis. However, many other real world time series data sets are also temporally misaligned. For example, in clinical trials the patient's health condition may be observed at irregular time intervals, and different patients are usually observed at different points in time (Cismondi et al. 2013). In finance, raw and original trade data is known as 'tick data' (Gençay et al. 2001). They are typically recorded at irregular time intervals and at ultra-high frequencies. An ideal method should be able to produce output of not only price movements, but also volatility and correlation dynamics at an arbitrary time scale, preferably regular.

Appendix A

Implication of different data supports in chapter 4

In this appendix we discuss our treatment for the process variance and nugget parameter in the model proposed in Chapter 4. Specifically, we examine the change in the theoretical semivariogram when there is a change in the underlying support of the data.

An example of change of support theory

Without loss of generality, we suppose that the data process $\boldsymbol{y} = \{y(t); t = 1, \ldots, n\}$ can be modelled by a univariate independent increments process with Gaussian noise. In other words, it possesses a theoretical linear semivariogram of the form $\Gamma_y(h)$ where

$$\Gamma_y(h) = \frac{1}{2}\,\mathrm{Var}\Big(y(t+h) - y(t)\Big) = \tau + \frac{1}{2}v|h| \tag{A.1}$$

Here, the nugget τ is the intercept, the process variance v is twice the slope value, and h is an arbitrary time lag.

We create the new process $\tilde{\boldsymbol{y}}$ 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 $\tilde{\boldsymbol{y}}$.

To simplify the problem, we break up the underlying process into two separate component: a pure nugget process and a pure independent increment process. If $\boldsymbol{y}$ is a pure nugget process, it is easy to show that the semivariogram of the averaged process

is $\Gamma_{\tilde{y}}(h)$ where

$$\Gamma_{\tilde{y}}(h) = \begin{cases} \frac{\tau}{w}, & \text{if } |h| \geq w \\ \frac{\tau|h|}{w^2}, & |h| < w \end{cases} \tag{A.2}$$

For the independent increments case on its own, the semivariogram has the form

$$\Gamma_{\tilde{y}}(h) = \begin{cases} \frac{1}{2}v|h| - \frac{wv}{6}, & \text{if } |h| \geq w \\ \frac{v|h|^2}{2w} - \frac{v|h|^3}{6w^2}, & |h| < w \end{cases} \tag{A.3}$$

We refer to Chiles & Delfiner (2012, Chapter 2.4) for the technical details of the aforementioned results. It is worth noting that, in practice, the only lags available for calculation of empirical semivariograms are $|h| \geq w$.

Hence, for a process that has both the nugget and independent increments, the implication is twofold. First, $\boldsymbol{y}$ and $\tilde{\boldsymbol{y}}$ share the same process variance. 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 function. Moreover, 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 τ and v.

Implication of different supports for data set GISP2 and GRIP

In the application discussed in Section 4.3, τ dominates v in both empirical semivariograms of GISP2 and GRIP - as seen in Fig. 4.3. Furthermore, if we assume that the nugget effect is at an annual level (denoted as τ_{annual}), then the nuggets for GISP2 and GRIP are, respectively, $\tau_{\text{GISP2}} = \frac{\tau_{\text{annual}}}{w_{\text{GISP2}}}$ and $\tau_{\text{GRIP}} = \frac{\tau_{\text{annual}}}{w_{\text{GRIP}}}$, where factors w denote the time support for each series. We can reparameterise in terms of $\tau = \tau_{\text{GRIP}}$, such that $\tau_{\text{GISP2}} = \tau \times \frac{w_{\text{GRIP}}}{w_{\text{GISP2}}} = \tau \times k_{\text{GISP2}}$. We set $\frac{55}{200}$ or 0.275 as the value for k_{GISP2}, corresponding to the respective length of support of GISP2 and GRIP. This value is also consistent with the descriptive statistics of the age increments, as represented in Fig. 4.2(**c**). Each section, being 200cm and 55cm in length respectively for GISP2 and GRIP, is negligible compared to the total length of the complete ice core which

is about 3km (approximately 1.6 km of which covers the Holocene period). Hence we feel that this is a reasonable approximation.

Appendix B

MCMC for the reconstruction module in chapter 5

This appendix describes the MCMC algorithms for fitting the reconstruction module represented by Eq. (5.17) as discussed in Section 5.3.1.3. Note that for clearer presentation, the subscript notation for time is omitted in the notation throughout this appendix. For instance, we simple rewrite v_t as v.

Where appropriate, we use functions from R package *INLA* (Martins et al. 2013) and *spam* (Furrer & Sain 2010) which contain optimised functions for sparse matrices. We also take advantage of efficient computational strategies for matrices involving Kronecker products and Sherman-Morrison-Woodbury identity as discussed, for example, in Harville (1997).

Updating v_o

It follows from (5.21) that the calculation of the Metropolis-Hastings (MH) ratio $\beta(v_o, v_o^*)$ is required to decide whether to replace v_o with v_o^*, where:

$$\beta(v_o, v_o^*) = \frac{v_o^{3/2} \left|\mathbf{Q}_{x|y}\right|^{1/2} \exp(\frac{1}{2}\boldsymbol{m}_\alpha^T \boldsymbol{\Lambda}_\alpha \boldsymbol{\mu}_{x|y}^*)\ \pi(v_o|v_o^*)\ \pi(v_o^*|\boldsymbol{\theta})}{(v_o^*)^{3/2} \left|\mathbf{Q}_{x|y}^*\right|^{1/2} \exp(\frac{1}{2}\boldsymbol{m}_\alpha^T \boldsymbol{\Lambda}_\alpha \boldsymbol{\mu}_{x|y})\ \pi(v_o^*|v_o)\ \pi(v_o|\boldsymbol{\theta})} \tag{B.1}$$

Direct computation of the expression in (B.1) is tedious as it involves expensive matrix manipulations of Kronecker products of matrices. Simplification of this term is possible

from observing that

$$\begin{aligned}\mathbf{Q}^*_{x|y} &= \mathbf{Q}_{x|y} + \{[(v_o^*)^{-1} - v_o^{-1}]\boldsymbol{b}\boldsymbol{b}^T\} \otimes \boldsymbol{\Sigma}^{-1} \\ &= \mathbf{Q}_{x|y} + [(v_o^*)^{-1} - v_o^{-1}](\boldsymbol{b}\boldsymbol{b}^T) \otimes \boldsymbol{\Sigma}^{-1} \\ &= \mathbf{Q}_{x|y} + \mathbf{R}^T\boldsymbol{\Upsilon}\mathbf{R}\end{aligned}$$

Here, $\mathbf{R} = \boldsymbol{b}^T \otimes \boldsymbol{\Sigma}^{-1/2}$, with $\boldsymbol{b}$ being a vector of length n with -1 at time t, 1 at time $t+1$ and 0's elsewhere, and $\boldsymbol{\Upsilon}$ is the m-by-3 diagonal matrix with element $[(v_o^*)^{-1} - v_o^{-1}]$ in the main diagonal. Therefore computation of $\mathbf{Q}^{*-1}_{x|y}$ follows instantly from $\mathbf{Q}^{-1}_{x|y}$. Consequently,

$$\begin{aligned}\frac{|\mathbf{Q}_{x|y}|}{|\mathbf{Q}^*_{x|y}|} &= \frac{|\mathbf{Q}_{x|y}|}{|\boldsymbol{\Upsilon}^{-1} + \mathbf{R}\mathbf{Q}^{-1}_{x|y}\mathbf{R}^T|\ |\mathbf{Q}_{x|y}|\ |\boldsymbol{\Upsilon}|} \\ &= \frac{1}{|\boldsymbol{\Upsilon}^{-1} + \mathbf{R}\boldsymbol{S}|\ |\boldsymbol{\Upsilon}|}\end{aligned}$$

where $\boldsymbol{S}$ is the solution to $\mathbf{Q}_{x|y}\boldsymbol{S} = \mathbf{R}^T$. Here, we have utilised the Woodbury formulae. Similarly:

$$\begin{aligned}\mathbf{Q}^{*-1}_{x|y} - \mathbf{Q}^{-1}_{x|y} &= \mathbf{Q}^{-1}_{x|y} - \mathbf{Q}^{-1}_{x|y}\mathbf{R}^T(\boldsymbol{\Upsilon}^{-1} + \mathbf{R}\mathbf{Q}^{-1}_{x|y}\mathbf{R}^T)^{-1}\mathbf{R}\mathbf{Q}^{-1}_{x|y} - \mathbf{Q}^{-1}_{x|y} \\ &= -\boldsymbol{S}(\boldsymbol{\Upsilon}^{-1} + \mathbf{R}\boldsymbol{S})^{-1}\boldsymbol{S}^T\end{aligned}$$

Putting all the terms together yields the simplified ratio

$$\beta(v_o, v_o^*) = \frac{v_o^{3/2}}{(v_o^*)^{3/2}\,|\boldsymbol{\Upsilon}^{-1} + \mathbf{R}\boldsymbol{S}|^{1/2}\ |\boldsymbol{\Upsilon}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{m}_\alpha^T\boldsymbol{\Lambda}_\alpha\mathbf{S}(\boldsymbol{\Upsilon}^{-1} + \mathbf{R}\boldsymbol{S})^{-1}\mathbf{S}^T\boldsymbol{\Lambda}_\alpha\boldsymbol{m}_\alpha\right) \times \frac{\pi(v_o|v_o^*)\ \pi(v_o^*|\boldsymbol{\theta})}{\pi(v_o^*|v_o)\ \pi(v_o|\boldsymbol{\theta})}$$

Computation of the above ratio is fast since all calculations of the determinant and inversion are only necessary for matrices of size m-by-m. In the application considered in Section 5.4, $m = 3$.

Updating $\boldsymbol{\Sigma}$

To explore the sensitive posterior region of $\boldsymbol{\Sigma}$ based on Eq. (5.21) using MH, we require the calculation of the ratio $\beta(\boldsymbol{\Sigma}^*, \boldsymbol{\Sigma})$, where:

$$\beta(\boldsymbol{\Sigma}^*, \boldsymbol{\Sigma}) = \frac{|\boldsymbol{\Sigma}|^{\frac{n-1}{2}}\ |\mathbf{Q}_{x|y}|^{1/2}\ \exp(\frac{1}{2}\boldsymbol{m}_\alpha^T\boldsymbol{\Lambda}_\alpha\boldsymbol{\mu}^*_{x|y})\ \pi(\boldsymbol{\Sigma}^*)}{|\boldsymbol{\Sigma}^*|^{\frac{n-1}{2}}|\mathbf{Q}^*_{x|y}|^{1/2}\ \exp(\frac{1}{2}\boldsymbol{m}_\alpha^T\boldsymbol{\Lambda}_\alpha\boldsymbol{\mu}_{x|y})\ \pi(\boldsymbol{\Sigma})}$$

Here, $|\mathbf{Q}_{x|y}| = \prod_i\ \mathbf{L}_{ii}$ where $\mathbf{L}$ is the Cholesky decomposition of $\mathbf{Q}_{x|y}$. All other computational details have been discussed elsewhere in this appendix.

Updating $\boldsymbol{\alpha}$

We begin with full conditional posterior distribution for $\boldsymbol{\alpha}$ from (5.21):

$$\pi(\boldsymbol{\alpha}|\ldots) \;\propto\; |\boldsymbol{\Lambda}_\alpha|^{1/2}\;|\mathbf{Q}_{x|y}|^{-1/2}\;\exp\left(\frac{1}{2}\boldsymbol{m}_\alpha^T\boldsymbol{\Lambda}_\alpha(\boldsymbol{\mu}_{x|y}-\boldsymbol{m}_\alpha)\right)\;\pi(\boldsymbol{\alpha}|\boldsymbol{\kappa})$$

Next, we let $\mathbf{Q}^*_\alpha = \boldsymbol{\Lambda}_\alpha + \mathbf{1}\big((\boldsymbol{\lambda}^*_\alpha)^{-1}-(\boldsymbol{\lambda}_\alpha)^{-1}\big)\mathbf{1}^T$ and $\mathbf{Q}^*_{x|y} = \mathbf{Q}_{x|y}+\mathbf{1}\big((\boldsymbol{\lambda}^*_\alpha)^{-1}-(\boldsymbol{\lambda}_\alpha)^{-1}\big)\mathbf{1}^T$. Here, $\mathbf{1}$ is a indicator vector with a value 1 corresponding to the series identity and time t whence parameter values are being update. The Woodbury identity is useful again here. It can be shown that the MH ratio required to update $\boldsymbol{\alpha}$ is $\beta(\boldsymbol{\alpha}^*,\boldsymbol{\alpha})$, where

$$\begin{aligned}\beta(\boldsymbol{\alpha}^*,\boldsymbol{\alpha}) =& \frac{|\mathbf{Q}^*_\alpha|^{1/2}\;|\mathbf{Q}_{x|y}|^{1/2}\;\exp\left(\frac{1}{2}\boldsymbol{m}^{*\,T}_\alpha\mathbf{Q}^*_\alpha(\boldsymbol{\mu}^*_{x|y}-\boldsymbol{m}^*_\alpha)\right)\;\pi(\boldsymbol{\alpha}^*)}{|\boldsymbol{\Lambda}_\alpha|^{1/2}\;|\mathbf{Q}^*_{x|y}|^{1/2}\;\exp\left(\frac{1}{2}\boldsymbol{m}^T_\alpha\boldsymbol{\Lambda}_\alpha(\boldsymbol{\mu}_{x|y}-\boldsymbol{m}_\alpha)\right)\;\pi(\boldsymbol{\alpha})}\\ =& \frac{(\boldsymbol{\lambda}_\alpha)^{1/2}}{(\boldsymbol{\lambda}^*_\alpha)^{1/2}}\left(1+\mathbf{1}^T\boldsymbol{b}(\boldsymbol{\lambda}^{-1}_{*\alpha}-\boldsymbol{\lambda}^{-1}_\alpha)\right)^{-1/2}\frac{\exp\left(\frac{1}{2}\boldsymbol{m}^{*\,T}_\alpha\mathbf{Q}^*_\alpha(\boldsymbol{\mu}^*_{x|y}-\boldsymbol{m}^*_\alpha)\right)}{\exp\left(\frac{1}{2}\boldsymbol{m}^T_\alpha\boldsymbol{\Lambda}_\alpha(\boldsymbol{\mu}_{x|y}-\boldsymbol{m}_\alpha)\right)}\;\frac{\boldsymbol{\kappa}^*_\alpha}{\boldsymbol{\kappa}_\alpha}\end{aligned}$$

with $\boldsymbol{b}$ is the solution to $\mathbf{Q}_{x|y}\boldsymbol{b}=\mathbf{1}$. It is re-emphasised that, according to our notational trick first introduced in Section 3.1.2, computation is deemed unnecessary where data is not available. Here, for instance, $|\boldsymbol{\Lambda}_\alpha| = \prod_{\boldsymbol{\lambda}_\alpha\neq 0}\;\boldsymbol{\lambda}_\alpha^{-1/2}$.

Updating $\boldsymbol{x_o}$

Recall that $\boldsymbol{x_o}$ has been marginalised out from the full posterior distribution as discussed in (5.20) and (5.21). More crucially, the full conditional for $\boldsymbol{x_o}$ is available in closed-form, i.e. $\boldsymbol{x_o}|\ldots \sim \mathcal{N}(\mathbf{Q}^{-1}_{x|y}\boldsymbol{\Lambda}_\alpha\boldsymbol{m}_\alpha, \mathbf{Q}^{-1}_{x|y})$. Therefore, posterior sampling for this component is straightforward:

1. Perform Cholesky decomposition $\mathbf{Q}_{x|y} = \mathbf{L}\mathbf{L}^T$
2. Generate nm standard normal distribution values $\boldsymbol{z}$
3. Solve $\boldsymbol{L}^T\boldsymbol{u} = \boldsymbol{\Lambda}_\alpha\boldsymbol{m}_\alpha$
4. Solve $\boldsymbol{L}\boldsymbol{x_o} = \boldsymbol{u}+\boldsymbol{z}$

Appendix C

Bridge sampling

Univariate Brownian bridge

In the univariate temporal setting, a Brownian *bridge* is a Gaussian process conditional on at least one known value (Glasserman 2003, Chapter 3). We first consider the general case by conditioning on two values at the time of the beginning and ending of the process of interest. In our notation, the aim is to interpolate $x(t_2)$ conditional on the values of $x(t_1)$ and $x(t_3)$. Equivalently, we evaluate the conditional density function $\pi(x(t_2)|x(t_1), x(t_3))$. Let v_1 and v_2 denote the variances of the increments from $x(t_1)$ to $x(t_2)$ and from $x(t_2)$ to $x(t_3)$ respectively. Using Bayes theorem,

$$\begin{aligned}
\pi(x(t_2)|x(t_1), x(t_3)) &= \frac{\pi(x(t_1), x(t_2), x(t_3))}{\pi(x(t_1), x(t_3))} \\
&= \frac{\pi(x(t_2) - x(t_1))\pi(x(t_3) - x(t_2))}{\pi(x(t_3) - x(t_1))} \\
&\propto exp\left[-\frac{1}{2}\left(\frac{(x(t_2) - x(t_1))^2}{v_1} + \frac{(x(t_3) - x(t_2))^2}{v_2}\right)\right]
\end{aligned}$$

Further simple algebra gives $x(t_2)|x(t_1), x(t_3) \sim \mathcal{N}\left(\frac{x(t_1)v_2 + x(t_3)v_1}{v_1+v_2}, \frac{v_1 v_2}{v_1+v_2}\right)$.

The multivariate bridge sampling algorithm in chapter 5

In Chapter 5, the challenge is to numerically evaluate the high dimensional integral (5.22). Conditioned on the suitable MCMC samples, the leftover tasks are sampling from the multivariate $\mathcal{NIG}$ bridge. This involves sampling from the $\mathcal{IG}$ bridge followed

by sampling from the multivariate Gaussian distribution. Therefore, at the j^{th} climate increment our algorithm proceeds as follows:

1. Multiply the crossed-correlated posterior samples of the j^{th} observed increments of $\boldsymbol{x_o}$ by the inverse of the Cholesky decomposition of the correlation matrix $\boldsymbol{\Sigma}$ to obtain independent samples of climate; these represent the bridged points that will be conditioned upon simulation of the gridded climate process.

2. Generate gridded variance values conditioned on v_j (the variance of the j^{th} increment) and its parameters using the $\mathcal{IG}$ bridge algorithm discussed in Ribeiro & Webber (2003).

3. Conditioned on the gridded $\mathcal{IG}$ available from step 2, the gridded climate process at different sites are independent Gaussian random variables. That is, the conditional process $\boldsymbol{x_g}$ are, marginally, Brownian bridges.

4. As the final step, we multiply the independent Brownian bridges with the Cholesky decomposition of $\boldsymbol{\Sigma}$ to obtain the cross-correlated Brownian bridges which represent posterior realisations from the full conditionals of $\boldsymbol{x_g}$.

Bibliography

Ahsanullah, M., Kibria, B. G. & Shakil, M. (2014), *Normal and Student's T Distributions and Their Applications*, Springer.

Andersen, K. K., Azuma, N., Barnola, J.-M., Bigler, M., Biscaye, P., Caillon, N., Chappellaz, J., Clausen, H. B., Dahl-Jensen, D., Fischer, H. et al. (2004), 'High-resolution record of northern hemisphere climate extending into the last interglacial period', *Nature* **431**(7005), 147–151.

Betro, B. & Rotondi, R. (1991), 'On bayesian inference for the inverse gaussian distribution', *Statistics & probability letters* **11**(3), 219–224.

Bhattacharya, A. & Wilson, S. P. (2014), 'Sequential Bayesian Inference for Dynamic State Space Model Parameters', *ArXiv e-prints* .

Blaauw, M. & Christen, J. A. (2005), 'Radiocarbon peat chronologies and environmental change', *J. Roy. Statist. Soc. Ser. C* **54**(4), 805–816.

Bolin, D. (2014), 'Spatial Matérn fields driven by non-Gaussian noise', *Scand. J. Stat.* **41**(3), 557–579.
URL: *http://dx.doi.org/10.1111/sjos.12046*

Brooks, S., Gelman, A., Jones, G. L. & Meng, X.-L., eds (2011), *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, CRC Press, Boca Raton, FL.

Buck, C. E., Cavanagh, W. G. & Litton, C. D. (1996), *Bayesian approach to interpreting archaeological data*, Wiley Chichester.

Chandler, R. E., Thorne, P., Lawrimore, J. & Willett, K. (2012), 'Building trust in

climate science: data products for the 21st century', *Environmetrics* **23**(5), 373–381.
URL: *http://dx.doi.org/10.1002/env.2141*

Chib, S. & Greenberg, E. (1995), 'Understanding the metropolis-hastings algorithm', *The american statistician* **49**(4), 327–335.

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

Cismondi, F., Fialho, A. S., Vieira, S. M., Reti, S. R., Sousa, J. & Finkelstein, S. N. (2013), 'Missing data in medical databases: Impute, delete or classify?', *Artificial intelligence in medicine* **58**(1), 63–72.

Cismondi, F., Fialho, A. S., Vieira, S. M., Sousa, J. M., Reti, S. R., Howell, M. D. & Finkelstein, S. N. (2011), Computational intelligence methods for processing misaligned, unevenly sampled time series containing missing data, *in* 'Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on', IEEE, pp. 224–231.

Davis, B. A., Zanon, M., Collins, P., Mauri, A., Bakker, J., Barboni, D., Barthelmes, A., Beaudouin, C., Bjune, A. E., Bozilova, E. et al. (2013), 'The european modern pollen database (empd) project', *Vegetation history and archaeobotany* **22**(6), 521–530.

Doan, T. K., Haslett, J. & Parnell, A. C. (2015), 'Joint inference of misaligned irregular time series with application to greenland ice core data', *Advances in Statistical Climatology, Meteorology and Oceanography* **1**, 15–27.
URL: *http://dx.doi.org/10.5194/ascmo-1-15-2015*

Donner, R. (2007), Advanced methods for analysing and modelling multivariate palaeoclimatic time series, PhD thesis, University of Potsdam.

Eckner, A. (2012), 'A framework for the analysis of unevenly spaced time series data', *Preprint* . (Access date: 17-03-2015).

Erästö, P., Holmström, L., Korhola, A., Weckström, J. et al. (2012), 'Finding a consensus on credible features among several paleoclimate reconstructions', *The Annals of Applied Statistics* **6**(4), 1377–1405.

Fiedel, S. (2011), 'The mysterious onset of the younger dryas', *Quaternary International* **242**(2), 262–266.

Fraley, C., Raftery, A. E., Murphy, T. B. & Scrucca, L. (2012), *mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*.

Furrer, R. & Sain, S. R. (2010), 'spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields', *Journal of Statistical Software* **36**(10), 1–25.

Fyfe, R. M., de Beaulieu, J.-L., Binney, H., Bradshaw, R. H., Brewer, S., Le Flao, A., Finsinger, W., Gaillard, M.-J., Giesecke, T., Gil-Romera, G. et al. (2009), 'The european pollen database: past efforts and current activities', *Vegetation History and Archaeobotany* **18**(5), 417–424.

Gelfand, A. E., Diggle, P., Guttorp, P. & Fuentes, M. (2010), *Handbook of spatial statistics*, CRC Press.

Gelfand, A. E. & Smith, A. F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *J. Amer. Statist. Assoc.* **85**(410), 398–409.

Gelfand, A. E., Zhu, L. & Carlin, B. P. (2001), 'On the change of support problem for spatio-temporal data', *Biostatistics* **2**(1), 31–45.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2014), *Bayesian data analysis*, Vol. 2, Taylor & Francis.

Gençay, R., Dacorogna, M., Muller, U. A., Pictet, O. & Olsen, R. (2001), *An introduction to high-frequency finance*, Academic press.

Glasserman, P. (2003), *Monte Carlo methods in financial engineering*, Vol. 53, Springer Science & Business Media.

Gulliksen, S., Birks, H. H., Possnert, G. & Mangerud, J. (1998), 'A calendar age estimate of the younger dryas-holocene boundary at kråkenes, western norway', *The Holocene* **8**(3), 249–259.

Harville, D. A. (1997), *Matrix algebra from a statistician's perspective*, Springer-Verlag, New York.

Haslett, J. (1997), 'On the sample variogram and the sample autocovariance for non-stationary time series', *Journal of the Royal Statistical Society: Series D (The Statistician)* **46**(4), 475–484.

Haslett, J. & Parnell, A. (2008), 'A simple monotone process with application to radiocarbon-dated depth chronologies', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **57**(4), 399–418.

Haslett, J., Whiley, M., Bhattacharya, S., Salter-Townshend, M., Wilson, S. P., Allen, J., Huntley, B. & Mitchell, F. (2006), 'Bayesian palaeoclimate reconstruction', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **169**(3), 395–438.

Huntley, B., Long, A. J. & Allen, J. R. (2013), 'Spatio-temporal patterns in lateglacial and holocene vegetation and climate of finnmark, northernmost europe', *Quaternary Science Reviews* **70**, 158–175.

Johnsen, S. J. (1999), 'Grip oxygen isotopes'.
URL: *http://doi.pangaea.de/10.1594/PANGAEA.55091*

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. & White, J. (1997), 'Validity of the temperature reconstruction from water isotopes in ice cores', *Journal of Geophysical Research: Oceans* **102**(C12), 26471–26487.
URL: *http://dx.doi.org/10.1029/97JC01283*

Kobashi, T., Severinghaus, J. P., Brook, E. J., Barnola, J.-M. & Grachev, A. M. (2007), 'Precise timing and characterization of abrupt climate change 8200 years ago from air trapped in polar ice', *Quaternary Science Reviews* **26**(9), 1212–1222.

Li, B., Nychka, D. W. & Ammann, C. M. (2007), 'The hockey stickand the 1990s: a statistical perspective on reconstructing hemispheric temperatures', *Tellus A* **59**(5), 591–598.

Li, B., Nychka, D. W. & Ammann, C. M. (2010), 'The value of multiproxy reconstruction of past climate', *J. Amer. Statist. Assoc.* **105**(491), 883–895. With supplementary material available online.

Lindgren, F. & Rue, H. (2008), 'On the second-order random walk model for irregular locations', *Scand. J. Statist.* **35**(4), 691–700.

Lindgren, F., Rue, H. & Lindström, J. (2011), 'An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach', *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**(4), 423–498. With discussion and a reply by the authors.

Mann, M. E., Bradley, R. S. & Hughes, M. K. (1998), 'Global-scale temperature patterns and climate forcing over the past six centuries', *Nature* **392**(6678), 779–787.

Martino, S. (2007), Approximate bayesian inference for multivariate stochastic volatility models, Technical report, Technical report, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

Martins, T. G., Simpson, D., Lindgren, F. & Rue, H. (2013), 'Bayesian computing with INLA: new features', *Comput. Statist. Data Anal.* **67**, 68–83.

McShane, B. B. & Wyner, A. J. (2011), 'A statistical analysis of multiple temperature proxies: are reconstructions of surface temperatures over the last 1000 years reliable?', *Ann. Appl. Stat.* **5**(1), 5–44.

Mogensen, I. (2001), A study of rapid climate changes, Dansgaard-Oeschger events, PhD thesis.

Morrill, C. & Jacobsen, R. M. (2005), 'How widespread were climate anomalies 8200 years ago?', *Geophysical Research Letters* **32**(19), n/a–n/a.

Mudelsee, M. (2010), *Climate time series analysis: classical statistical and bootstrap methods*, Vol. 42, Springer.

Muscheler, R., Kromer, B., Björck, S., Svensson, A., Friedrich, M., Kaiser, K. & Southon, J. (2008), 'Tree rings and ice cores reveal 14c calibration uncertainties during the younger dryas', *Nature Geoscience* **1**(4), 263–267.

Nieto-Barajas, L. & Sinha, T. (2015), 'Bayesian interpolation of unequally spaced time series', *Stochastic Environmental Research and Risk Assessment* **29**(2), 577–587. **URL:** *http://dx.doi.org/10.1007/s00477-014-0894-3*

Paciorek, C. J. (2013), 'Spatial models for point and areal data using Markov random fields on a fine grid', *Electron. J. Stat.* **7**, 946–972.

Parnell, A. C., Buck, C. E. & Doan, T. K. (2011), 'A review of statistical chronology models for high-resolution, proxy-based holocene palaeoenvironmental reconstruction', *Quaternary Science Reviews* **30**(21), 2948–2960.

Parnell, A. C., Sweeney, J., Doan, T. K., Salter-Townshend, M., Allen, J. R., Huntley, B. & Haslett, J. (2015), 'Bayesian inference for palaeoclimate with time uncertainty and stochastic volatility', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **64**(1), 115–138.

Peavoy, D. & Franzke, C. (2010), 'Bayesian analysis of rapid climate change during the last glacial using greenland δ 18 o data', *Climate of the Past* **6**(6), 787–794.

Qian, P. Z. & Wu, C. J. (2008), 'Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments', *Technometrics* **50**(2), 192–204.

Ramsey, C. B. (2008), 'Deposition models for chronological records', *Quaternary Science Reviews* **27**(1), 42–60.

Rasmussen, S. O., Andersen, K. K., Svensson, A., Steffensen, J. P., Vinther, B. M., Clausen, H. B., Siggaard-Andersen, M.-L., Johnsen, S. J., Larsen, L. B., Dahl-Jensen, D. et al. (2006), 'A new greenland ice core chronology for the last glacial termination', *Journal of Geophysical Research: Atmospheres (1984–2012)* **111**(D6).

Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Ramsey, C. B., Buck, C. E., Cheng, H., Edwards, R. L., Friedrich, M. et al. (2013), 'Intcal13 and marine13 radiocarbon age calibration curves 0–50,000 years cal bp', *Radiocarbon* **55**(4), 1869–1887.

Ribeiro, C. & Webber, N. (2003), 'A monte carlo method for the normal inverse gaussian option valuation model using an inverse gaussian bridge'.

Ribeiro Jr, P. J. & Diggle, P. J. (2001), 'geor: A package for geostatistical analysis', *R news* **1**(2), 14–18.

Robert, C. & Casella, G. (2011), 'A short history of Markov chain Monte Carlo: subjective recollections from incomplete data', *Statist. Sci.* **26**(1), 102–115.

Rousseeuw, P. J. & Molenberghs, G. (1994), 'The shape of correlation matrices', *Amer. Statist.* **48**(4), 276–279.

Rue, H. & Held, L. (2005), *Gaussian Markov random fields*, Vol. 104 of *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC, Boca Raton, FL. Theory and applications.

Rue, H. & Martino, S. (2007), 'Approximate Bayesian inference for hierarchical Gaussian Markov random field models', *J. Statist. Plann. Inference* **137**(10), 3177–3192.

Rue, H., Martino, S. & Chopin, N. (2009), 'Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(2), 319–392.

Salter-Townshend, M. & Haslett, J. (2012), 'Fast inversion of a flexible regression model for multivariate pollen counts data', *Environmetrics* **23**(7), 595–605.

Särkkä, S. (2013), *Bayesian filtering and smoothing*, Vol. 3 of *Institute of Mathematical Statistics Textbooks*, Cambridge University Press, Cambridge.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. (2002), 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4), 583–639.

Stuiver, M. (1999), 'Gisp2 bidecadal oxygen isotope data'.
URL: *http://doi.pangaea.de/10.1594/PANGAEA.55531*

Stuiver, M. & Grootes, P. M. (2000), 'Gisp2 oxygen isotope ratios', *Quaternary Research* **53**(3), 277–284.

Sweeney, J. (2012), Advances in Bayesian model development and inversion in multivariate inverse inference problems with application to palaeoclimate reconstruction, PhD thesis, PhD Thesis.

Tanner, M. A. & Wong, W. H. (2010), 'From EM to data augmentation: the emergence of MCMC Bayesian computation in the 1980s', *Statist. Sci.* **25**(4), 506–516. **URL:** *http://dx.doi.org/10.1214/10-STS341*

Thomas, E. R., Wolff, E. W., Mulvaney, R., Steffensen, J. P., Johnsen, S. J., Arrowsmith, C., White, J. W., Vaughn, B. & Popp, T. (2007), 'The 8.2 ka event from greenland ice cores', *Quaternary Science Reviews* **26**(1), 70–81.

Tierney, L. & Kadane, J. B. (1986), 'Accurate approximations for posterior moments and marginal densities', *J. Amer. Statist. Assoc.* **81**(393), 82–86.

Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E. & Rajaratnam, B. (2012), 'Piecing together the past: statistical insights into paleoclimatic reconstructions', *Quaternary Science Reviews* **35**, 1–22.

Tingley, M. P. & Huybers, P. (2013), 'Recent temperature extremes at high northern latitudes unprecedented in the past 600 years', *Nature* **496**(7444), 201–205.

Wackernagel, H. (2003), *Multivariate geostatistics*, Springer.

Wanner, H., Solomina, O., Grosjean, M., Ritz, S. P. & Jetel, M. (2011), 'Structure and origin of holocene cold events', *Quaternary Science Reviews* **30**(21), 3109–3123.

Wikle, C. K. & Berliner, L. M. (2005), 'Combining information across spatial scales', *Technometrics* **47**(1), 80–91.