Automated Bayesian Layer Counting of Ice Cores

Joe Wheatley

A Thesis presented for the degree of
Doctor of Philosophy

School of Mathematics and Statistics
The University of Sheffield

January 2015

Automated Bayesian Layer Counting of Ice Cores

Joe Wheatley

Submitted for the degree of Doctor of Philosophy
January 2015

Abstract

The polar ice sheets hold a continuous record of climatic and environmental information, in the composition and concentrations of various chemicals, particles and gasses, extending back over hundreds of thousands of years. In order to interpret these data we must first learn about the underlying relationship between depth and age. Ice cores are vertical samples of the ice sheets. Some signals measured from them have annual cycles which show as quasi-periodic seasonality; layer counting uses this periodicity to obtain a chronology for the core. This is currently achieved manually, which is time-consuming and open to inconsistency and human error.

We present a method to standardise an ice core signal, isolating its seasonality, and to split it into sections with well-defined cycle counts and those with uncertain cycle counts. We show how the uncertain sections can be presented for manual assessment, and describe how the possible reconstructions can be identified and assigned probabilities based on their implied cycle lengths. We also develop a multivariate fully Bayesian approach, which models the signals as phase-shifted sine waves with continuously varying mean and amplitude. We use Markov chain Monte Carlo algorithms to enable inference about the age-depth relationship, and specifically the number of years covered by a particular section of ice core, including quantitative assessment of the uncertainty involved. We provide examples, applying our methods to several chemistry signals measured from ice cores from Greenland and Antarctica.

Declaration

The work in this thesis is based on research carried out at the School of Mathematics and Statistics, The University of Sheffield. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Acknowledgements

This work was supported by the Natural Environment Research Council (grant reference: NE/I52797X/1).

A special thank you to my supervisor Professor Paul Blackwell. I am grateful for your unwavering support and for always being generous with your time and expertise. I wish to thank my supervisors and mentors at the British Antarctic Survey, Professor Eric Wolff, Dr Nerilie Abram, Dr Rob Mulvaney and Dr Liz Thomas. Your help and guidance has been invaluable.

I would like to extend my gratitude for the warm welcome and constructive feedback I received from the teams at both the Centre for Ice and Climate in Copenhagen and the Australian Antarctic Division in Hobart.

I wrote Chapters $1-3$ at the same time as a paper, published in the open access journal Climate of the Past (Wheatley et al, 2012). There is significant overlap in the wording and content, although the work of the paper has been extended in this thesis. I would like to thank H. Fischer, M. Mudelsee and S. Rasmussen for valuable comments on an earlier version of the paper

I also thank the Division for Climate and Environmental Physics, Physics Institute, University of Bern, for providing high-resolution CFA data from the NGRIP ice core.

Contents

List of Figures

Chapter 1

Introduction

Ice cores are samples of ice that are typically removed vertically from an ice sheet using a hollow drill. The world's thickest and oldest ice sheets cover Greenland, an autonomous country within the Kingdom of Denmark, and Antarctica, the coldest, driest, highest, and windiest continent, inhabited only by extreme-cold-adapted organisms and climate scientists.

The polar ice sheets are composed of layers of snow that fell in sequence. At different sites the snowfall rate can range from as low as a few cm per year, up to more than a metre. The snow densifies to solid ice by a depth of (typically) $60 - 100$m, and thereafter thins with depth due to the flow of ice. Annually-resolved ice cores tend to come from Greenland and coastal parts of Antarctica where there is sufficiently high snowfall to preserve seasonal information. They hold a continuous record of climatic and environmental information that can extend back over hundreds of thousands of years. This information is preserved in the composition and concentrations of various chemicals, particles and gases that can be measured from ice core samples. In order to interpret these data we must first learn about the underlying depth to time relationship within a particular ice core.

Accurate dating of ice cores, with quantified uncertainties, is crucial for the understanding of climate dynamics. As an example, it is necessary to allow comparison of the sequence of events recorded in ice cores, marine sediments, speleothems, coral and so on; each dated by different means. Also, for comparing ice core events with the forcing from astronomical cycles in Earth's orbit, which can be absolutely and precisely calculated.

Some ice core signals have annual cycles which show as quasi-periodic seasonality in the depth series. Layer counting uses this periodicity to count back in time, year by year, and is currently achieved by eye with some considerable effort.

1.1 Layer counting

Various environmental factors can be determined directly or indirectly from measurements on ice core samples. An accurate translation from depth to the time of deposition is required in order to model the resulting climatic data in a meaningful way. Snow has different chemical and isotopic properties at different times of year, which results in annual layering once the snow has been compressed into ice. After these layers have been identified and counted, the age of the ice can be determined at any depth. In deeper parts of a core (or at low-accumulation sites across much of central Antarctica) annual layers can be so compressed that layer counting is no longer possible. Here, dating generally relies upon mathematical models of ice flow, accumulation and compression; see Svensson et al (2008), Lemieux-Dudon et al (2010) and Klauenberg et al (2011).

The most basic method used to identify annual layers is to visually inspect a scan or photograph of the ice, and to count the layers. This information, known as visual stratigraphy (VS), can be extracted as a signal for analysis. The VS intensity of the ice can be measured by a line-scanner as a continuous signal along a polished core and is attributed mostly to the dust content, the visible layers termed 'cloudy bands'. The VS record can show multiple layers in a year and therefore interpretation is problematic. A comprehensive explanation of VS and

how it is measured can be found in Svensson et al (2005) and Rasmussen et al (2006).

One method of extracting data from ice cores is to cut the core into discrete sub-samples, that are then melted and analysed for their chemical composition. Continuous measurements of a number of ice core constituents can also be taken at high resolution using Continuous Flow Analysis (CFA). CFA involves melting a stick of ice and feeding the flow of melt water continuously through an array of of analysis instruments. The analyses can include chemical impurities, insoluble dust, electrical conductivity and water isotope ratios. Many of these have annual cycles, and the number of cycles in the resulting signals can be counted back in time to obtain a chronology. Even the best indicators can produce signals with ambiguous annual layers. This can occur for a number of reasons: there may be no snowfall for a large part of the year; a part of the year's snowfall may be redistributed by the wind; and the underlying controls may simply have an irregular cycle. For this reason univariate counting approaches are inherently limited.

The electrical conductivity of the ice is indicative of the presence of hydrogen ions. Dips in conductivity are associated with high levels of other ions which can peak at different times of year. As with VS intensity this can result in a noisy annual signal with multiple peaks in a year; see Andersen et al (2006).

Chemical ions from impurities within the ice also preserve annual layer information. Andersen et al (2006) and Rasmussen et al (2006) describe the annual layer signatures from CFA analysis of the NGRIP ice core from northern Greenland (see section 1.5.3) in detail, the following is a summary. In warm periods the annual cycles in these signals are usually out of phase. A pronounced peak in insoluble dust and calcium ions (Ca^2+) is observed during spring. Summer is characterised by an initial increase in sulphate ions (SO_4^{-2}) followed by peaks in nitrate and ammonium ions (NO_3^- & NH_4^+). A peak in sodium ion (Na^+) concentration is seen in late winter. In colder periods these ions peak almost simultaneously, leading to much stronger annual cycles in the VS and electrical

conductivity measurements.

The isotopic composition of snow is also known to vary predictably throughout the year. There are relatively more of the heavier oxygen (O^{18}) and hydrogen (H^2; also known as deuterium) isotopes in summer as they require more energy to evaporate and move through the water cycle, providing annually cyclic signals. Water isotopes in snow can defuse after deposition and as such are often not very useful for layer counting. Their longer term trends can be used to help synchronize chronologies between different palaeoclimate records via wiggle matching. A detailed analysis of $\delta^{18}O$ from the NGRIP core can be found in Andersen et al (2004), Svensson et al (2006) and Svensson et al (2008).

Fixed time constraints can be used to aid in the dating of ice cores as they provide a known year of deposition for a certain depth. For example, large volcanic eruptions can leave a signal of sulphate or a tephra fingerprint in the ice. The layer may then be ascribed to a known event through the pattern of sulphate peaks, or by the relative concentrations of oxides in specific tephras. Fixed time constraints can be used to check or calibrate layer counting methods, and to help match up chronologies from multiple cores. Volcanic age-depth control points are discussed in Taylor et al (2004), Svensson et al (2006), and Klauenberg et al (2011).

Layer counting in ice cores is more challenging than for other palaeoclimatic archives; for example, due to the thinning of annual layers down the core. However, there is the benefit of the availability of multivariate information from a single core, which also presents statistical and analytical challenges.

1.2 Manual methods

Current methods of layer counting rely on a manual interpretation of the annual variations seen in ice core measurements. Teams of climate scientists use their comprehensive knowledge of the relative seasonality in multiple signals at different time periods to count annual cycles by eye, making judgements about which are

sufficiently distinct and appropriately spaced. For longer cores this process can take several years.

Ice cores from Greenland lend themselves to annual layer counting due to the high snow accumulation rate in that area. The Greenland Ice Core Chronology 2005 (GICC05) published by the Niels Bohr Institute at the University of Copenhagen reaches back to 60, 202 years b2k (before A.D. 2000). This timescale is based on overlapping isotope and chemical signals from three different Greenland cores, NGRIP, GRIP, and DYE-3, pieced together using fixed age-depth control points and chemical markers. The dating process took several years and is described in Vinther et al (2006) (< 7.9 ka b2k); Rasmussen et al (2006) ($7.9 - 14.7$ ka b2k) Andersen et al (2006) and Svensson et al (2006) ($14.7 - 42$ ka b2k); Svensson et al (2008) ($42 - 60$ ka b2k). The GICC05 chronology is the most comprehensive annual-layer ice core chronology achieved to date.

A summary of factors that lead to uncertainty in the chronologies obtained from layer counting can be found in Rasmussen et al (2006). The potential for missing years due to melting, wind scouring and other depositional issues is disregarded for the GICC05 chronology because of the high accumulation rates in Greenland. Gaps in the data are generally small relative to an annual layer, and tend not to affect all of the chemical signals at once. VS and electrical conductivity data are generally still available as their measurement does not rely on melting the ice. In deeper parts of the core insufficient measuring resolution can introduce uncertainty. One example of the resulting data loss is when two annual layers show as a double peak, and this type of signal alteration down a core is taken into account during manual interpretation.

When the annual signal is not clear cut, the accuracy of the chronology relies on the judgement of the interpreter. This can arise when, for example, evidence for an annual layer is not seen in all signals; a double peak is suspected to represent more than one annual layer; or when assessing long sections of missing data. Sections are dated independently by multiple investigators to assess reproducibility and layers can be marked ‘uncertain’.

Uncertain layer markers are discussed in Andersen et al (2006) and Rasmussen et al (2006). A layer marking is labelled uncertain if its probability p is judged to be between 1/4 and 3/4. These can be introduced in the situations above, or when the spacing of seasonal peaks is inconsistent with the investigators' interpretation. This raises an issue with 'certain' marks, which are treated as $p = 1$, but in some cases represent $p \geq 3/4$. However, this potential for overestimation may be offset by a number of $p \leq 1/4$ cycles not being marked.

The 'maximum counting error', or MCE, represents the accumulated error in the count. If there are N uncertain layers marked in the chronology, the MCE is calculated as $N/2$ years. Effectively, uncertain layers add $1/2 \pm 1/2$ a year to the total year count. The MCE implicitly assumes that all counting errors are correlated, and is considered to be a conservative estimate. At the other end of the scale, assuming independence gives an estimate of $\sqrt{N}/2$. Other measures, which assume varying levels of dependence, are discussed in Andersen et al (2006). In the GICC05 chronology the counting error is assumed to be Gaussian and the MCE is taken to equal twice the standard deviation, resulting in a final 1σ error estimate of $N/4$ years.

A critical review of two manual counting methods used to date the Holocene portion of the Siple Dome ice core from Antarctica is given in Taylor et al (2004). A univariate interpretation of the VS intensity was completed over four years by two groups, independently of each other. The degree of agreement between these groups was used to measure reproducibility: tests indicated around 95% agreement. A multivariate interpretation of electrical and chemical data was assessed over a three year period by three individuals. Dating was a subjective trade-off between the size and regular spacing of peaks, however this was never quantified. The resulting chronologies differed by up to 15%. This paper also presents an automated layer counting method, which is discussed in Section 1.4.

1.3 Automated methods

Manually counting the annual layers of an ice core and assessing the uncertainty of each one requires meticulous care. It can be argued that manual assessment is open to inconsistency and human error, and that the uncertainty measures on the resulting chronologies are subjective. This necessitates multiple re-counts and comparisons. Chronologies stretching as far back as the GICC05 can therefore take several years to complete. Correspondence with the Niels Bohr Institute suggests that approximately one person-month was required to date a recent 3000 year core. It is therefore clearly worthwhile to pursue a quick, well defined, and consistent statistical method of multivariate layer counting, that automatically provides a chronology with a robust uncertainty measure.

There have been several attempts at developing univariate automated cycle counting methods. These predominantly aim to replicate the manual approach, and generally have little consideration of uncertainty. This section summarises these methods; more detail of each one is given in Section 1.4.

Bandpass or lowpass filters are frequently used to remove any cycles with periods much larger or smaller than expected. Examples can be found in Rasmussen et al (2002), Rupf and Radons (2004), Taylor et al (2004), and McGwire et al (2011). In Shimohara et al (2003) the data are smoothed to aid manual assessment by taking a moving average over a range of half an annual layer thickness, estimated from ice-flow modelling.

Two of these methods iteratively pick the most likely annual peaks along a core, trading off the local height of a prospective peak with its separation in depth from the previous pick. Taylor et al (2004) present a method based on an ad-hoc algorithm, which is later developed in Smith et al (2009), to identify annual cycles in the trace element profiles of stalagmites. McGwire et al (2011) select new peaks on the basis of their vertical distance from a spline 'selection curve', fit to a small section of manually dated calibration data.

Further methods are based on measuring periodicities in the signals. In Svensson et al (2005) cloudy bands that are indicative of annual layers are identified from digital VS images. These are then thinned to satisfy conditions on their spacing and relative spectral power. Rupf and Radons (2004) model the periodicity of laminated sediment data in order to automatically detect variations in lamina thickness.

Rasmussen et al (2002) assume that short sections of unequally spaced ice core data can be treated as a time series, and use Independent Component Analysis (ICA) to combine chemistry data. When used in conjunction with a low-pass filter, this is shown to result in a signal that can be dated by counting all peaks that rise above a threshold value.

Winstrup et al (2012) present a univariate method of automated layer counting that uses a hidden Markov model (HMM) algorithm. Other than our own work (Wheatley et al, 2012) this is the first paper to address uncertainty in a robust way. The HMM method is limited by the requirement of an extensive manual study of the data prior to analysis and by the discrete and discontinuous model. For the MSc dissertation that preceded this doctoral research (Wheatley, 2007) a univariate method was developed that uses a generalised version of the model used in Winstrup et al (2012), with 'knots' at certain depths, describing the start of each annual cycle. We fitted the model using Metropolis-Hastings MCMC with a set of dimension-changing block updates that add and remove cycles, providing a measure of uncertainty on the count. This allows the hyperparameters of the model to be fit simultaneously using prior distributions to represent expert knowledge – rather than fixed from manual assessment of the signal. This was further developed to the Reversible Jump MCMC framework in a four month period of research following the MSc, funded by BAS. We have a paper in progress that provides a detailed account of this work, available on request.

During the course of my PhD research we published a preliminary univariate approach to automating the layer counting process (Wheatley et al, 2012). The methods developed in this paper are expanded upon in Chapters 2 and 3, some of

the theory is also used in later chapters. Some passages from this paper have been quoted verbatim in this thesis. We have also published a short summary of some of the ideas from Chapters 4 and 5 as a conference (BAYSM2013) proceedings paper (Wheatley et al, 2014).

1.4 Literature review

Here we provide a short summary of each paper in the literature that introduces an automated approach to the layer counting of ice cores.

1.4.1 Rasmussen et al (2002)

This paper presents a method of data enhancement termed Dynamical Decorrelation. Independent Component Analysis (ICA) is used to combine multiple annually cyclic signals into one master signal, with more regular cycles in that: multiple peaks are replaced by single peaks and all peaks have a more uniform height. This is used in combination with a low-pass filter to aid in the dating of ice cores via cycle counting. This method is demonstrated on two example sections of around 20 years in length: one from Site D in Antarctica, and one from NGRIP in Greenland. A simple method of automatically counting annual peaks in the resulting signal is also introduced.

Data Pre-processing

Missing data cannot be handled in this framework and as such are replaced using a mean-of-neighbours scheme. Abnormally large peaks are cut off to twice the average peak height, and the series is normalised to zero mean and unit variance. The signals are adjusted to peak simultaneously. One signal is fixed and the others are shifted according to the depth lag which maximises their correlation with the fixed signal.

ICA

In performing ICA the depth series is treated as being equally spaced in time - local variations in layer thickness and cycle asymmetry are disregarded. The assumption of orthogonal source series from Principal Component Analysis is replaced with the requirement that their time lagged correlations are zero. A depth lag τ is used to calculate the mixing matrix. This is treated as a tuning variable and found by trial and error.

PCA assumes that the source signals (S) are orthogonal:

$$X = AS$$

$$XX^T = AS(AS)^T = ASS^TA^T$$

where X is normalised and SS^T is diagonal due to orthogonality. An Eigen decomposition of XX^T provides A and thus S. In ICA the assumption that S is orthogonal is replaced with independence and another method is used to determine A.

In this paper $C_X(\tau)$ is defined as XX_τ^T where $X_\tau^T = \{x_{i+\tau}\}$, $Q = C_X(\tau)C_X(0)^{-1}$. Assuming an infinite series, it is stated that

$$C_X(\tau) = XX_\tau^T = ASS_\tau^TA^T = AC_S(\tau)A^T \Rightarrow Q = AC_S(\tau)C_S(0)^{-1}A^{-1}$$

where $C_S(\tau)$ is diagonal as S is independent. Q can be calculated from the data and diagonalised to find A and thus S. A suitable value for τ, and signal from the resulting S, is chosen by trial and error.

Values of τ between 1 and 30 were tried and the one that produced the source series with the most distinct annual component was selected. Fixing different signals yielded different optimal values of τ. The three signals with the strongest seasonality were chosen from the chemistry data to be candidates for the fixed signal. Approximately 100 permutations were tried, each yielding a number of

master signals. The three which performed best are presented.

Dating

Peaks in the resulting signal that rise above a threshold value l are counted. Using this method it was expected that random fluctuations in the signal could be included as peaks. To overcome this problem the signal was passed through a fifth order Butterworth digital filter.

Results

For Site D, when the counting method is applied to the filtered annual component chosen from the Dynamical Decorrelation, the same number of years as obtained by manual layer counting is found for $l \in [-.42, .71]$. For the unfiltered component the number of peaks is less well defined.

For NGRIP, as with the Site D, only the filtered series provides good countability. With $l \in [-.50, -.07]$ the result agrees with the 24 years determined by manual counting, while 23 years is found when $l \in [-.07, .55]$ due to one peak being less pronounced than the others.

Discussion

Combining multiple annually cyclic signals could reduce the manual or computational effort required for cycle counting whilst still utilising all available information. However, strong assumptions are made for ICA, and Dynamical Decorrelation is limited to short sections of signal with a reasonably consistent annual layer thickness. A robust method for use on longer or less stable sections would need to account for the continuous timescale and the dynamic phase difference between signals.

1.4.2 Taylor et al (2004)

In this paper the Holocene portion of the Siple Dome ice core from Antarctica is dated, both manually and using an automated computer algorithm. The consistency of the algorithm is presented as an improvement over the 'tedious and subjective' manual interpretation. Drilling and sampling methods are discussed along with a number of age-depth control points including known volcanic eruption events.

Computer Algorithm Interpretation

A bandpass filter is used on electrical conductivity data, preserving only features with a wavelength of $1 - 50$cm. This results in a zero mean signal which is subsequently normalised by dividing by the standard deviation of a 50m long moving window. An ad-hoc algorithm is presented, designed to mimic the thought process used by the interpreters when they manually count annual layers.

Given a starting peak, the algorithm searches for the next peak along a 15 step search sequence, stopping once one has been found. Each step:

- looks in one of five different proximities from the known peak relative to the expected year length, calculated from the preceding 20 cycles
- looks for peaks that are one of three of different sizes (small, medium, large) with respect to a chosen threshold.

The 15 steps are in a logical order chosen to avoid mis-classification.

The algorithm is forced to agree with volcanic age-control points and to a point of the GISP2 chronology tied in via the methane stratigraphy. It is stated that in the absence of age control points, and where the annual cycles are irregular, the design of the algorithm becomes difficult and subjective.

Discussion

The search sequence describes the subjective balance between the regular spacing and magnitude of annual peaks considered during manual interpretation. However, the algorithm is well defined and consistent. Results are subject to chosen thresholds and rely on age-depth control points for accuracy. There is no rigorous measure of uncertainty. The robustness of this method could be tested by applying the same algorithm to a different dataset.

This method is developed further; for use on speleothems in Smith et al (2009), and ice cores in McGwire et al (2011) (discussed below).

1.4.3 Svensson et al (2005)

This paper presents a method for estimating the average annual layer thickness for a given section of ice core, achieved by a frequency analysis of the strongest visible layers. The NGRIP core VS intensity, split into 55cm sections, is used as an example.

The VS profile was obtained by measuring light intensities along the core and storing them as digital images. Dark sections in these images, termed 'cloudy bands', are attributed mostly to the dust and Ca^{2+} content, and are thus indicative of annual layers. The contrast of an image is positively correlated with the number of bands identified, and some annual layers appear as multiple cloudy bands. This method of data sampling allows more control over the depth scale, and provides greater resolution further down the core where annual layers are no longer visible in chemical records.

Dating Methodology

The images are calibrated with respect to their intensity. From this, the onset depths of the strongest visual layers (or cloudy bands), with respect to a tuning parameter, are recorded in an array. The relative intensity of the bands is then

disregarded, each onset depth effectively being assigned the same weighting. Not all of the annual layers of the core will be represented by exactly one onset depth in the array - some may have more than one and some may have none.

The array is then iteratively thinned by the removal of closest neighbours - one of the closest pair of onset depths is removed at random. For each iteration a Gaussian filter is used to create a signal with peaks centred at each onset depth in the thinned array. A power spectrum for this induced signal is calculated, normalised to an integral value of 1, and weighted by the current size of the array to allow comparison between iteration steps.

The random thinning of the array can result in divergent frequency estimates. To avoid these outliers, the thinning process is carried out 20 times for each iteration and the resulting powers are summed. The estimated annual layer thickness for each 55cm section is calculated from the frequency with maximum power through all iterative steps.

Results

Annual layer thicknesses were estimated from a depth of 1.8km down the NGRIP core. Above this depth issues with ice storage had dampened the signal, causing the method to malfunction. Below 2.7km the estimates become unrealistic due to irregularities in the layering of the core.

This approach has a number of adjustable parameters which were tuned to the layer thicknesses of the manual ss09sea age model (Johnsen et al, 2001; Andersen et al, 2004) which is based on the NGRIP δ^{18}O signal. A strong agreement to the ss09sea chronology is shown.

Annual layers with multiple cloudy bands are addressed by the array thinning process. This frequency based approach should be robust to annual layers with bands that are too weak to be included in the array as long as the majority of layers in the section are identified. However, a significant proportion of the annual layers could not be identified in some of the sections, and in these cases

the intensity of the image was manually recalibrated.

Discussion

This paper introduces a new approach to obtaining ice core chronologies which could be applied when the limits of using chemical records have been reached. It relies on the, sometimes subjective, calibration of the digital image intensity and a number of tuning parameters. Because of this calibration, an absolute dating of the NGRIP ice core from 1.8km to 2.7km is not claimed. However, this method does correctly recognise the general pattern of the annual layer thicknesses with depth.

1.4.4 Weber et al (2010)

This work was brought to our attention by a referee of our Wheatley et al (2012) paper. It is not directly intended for use on ice core data, and was developed for glacial varves from marine sediment which show less variation in annual layer thickness. It is simple and effective; we show how the ideas can be expanded for use on ice core signals, with a measure of uncertainty, in Chapters 2 and 3 of this thesis.

This paper presents two tools: the BPMix Tool, which extracts a VS signal from the varves; and the PEAK Tool, which uses two methods to calculate a chronology from the signal. The 'zero-crossing method' algorithm iteratively finds the points at which a signal crosses a wide interval Gaussian moving average. The 'frequency truncation method' algorithm similarly finds the zero-crossing points in the signal after high-frequency noise and low-frequency shifts have been removed via Fourier transformation. Both of these methods have user-defined parameters that represent minimum layer thickness and a minimum amplitude tolerance which are adjusted to tune the count visually, along with other smoothing or frequency parameters. These methods provide a point estimate layer count, along with valuable information about the positioning of layer markings and the cycle

lengths. This paper does not address missing values or provide a measure of uncertainty for the count or markings.

1.4.5 McGwire et al (2011)

As in Taylor et al (2004), the first annual peak in the signal is fixed, then some selection criteria is used to pick subsequent annual peaks. Rather than an ad-hoc algorithm, a spline selection curve calibrated to a short section of manually dated signal is used to select the next peak.

Fitting the Spline Selection Curve

The spline selection curve is positioned relative to an estimate of the average annual layer thickness $\bar{T}_Z$ at each depth. Three different methods of estimation are compared here, with an increasing level of granularity:

1. fitting a second order polynomial to interpolate between three manually assessed layer thicknesses taken at equally spaced calibration depths
2. a running mean of the lengths of the last $16 < N < 60$ picked cycles
3. Fourier analysis, calculating the FFT power spectrum in a 10m running window and linearly interpolating between these discrete steps.

The selection curve is of length $3 \times \bar{T}_Z$ and is intended to locate the most probable annual peak within a maximum three average annual layer thicknesses from the last pick. It is positioned at some arbitrary height above the corresponding section of signal, and the peak closest vertically to the curve is chosen.

The lowest point of the curve is set at a horizontal distance of $\bar{T}_Z$ from the last peak selected, as this is the most likely place for the next annual peak. The spline curve has four other parameters that are fit to a short section of manually dated calibration data. These parameters describe how low the curve drops at each depth, and thus what amplitude a peak requires in order to be selected.

The curve describes the balance between the spacings of annual peaks and their expected magnitudes.

In each $3 \times \bar{T}_Z$ long section of data assessed for peak selection, a Blackman window low pass filter is used to filter out frequencies of less then $p \times \bar{T}_Z$. p was calibrated to 0.23 for the example.

Results

The polynomial and Fourier methods of estimating $\bar{T}_Z$ result in a close agreement to the manual counts. The direction of the algorithm up or down the core did have an effect on the final result - running it up the core proved to be more accurate. Surprisingly, increasing the amount of calibration data is shown to have little effect on the outcome. Using only 25m of manually interpreted calibration data, the Fourier method agreed with the manual dating within a tolerance of 2% over 1200m of core.

Discussion

The iterative peak picking method introduced in Taylor et al (2004) is achieved without the need to manually design and calibrate a complicated ad-hoc algorithm. Both of these methods assign only 'certain' layer markings and do not address the issue of uncertainty.

It is concluded that for short ice cores this approach may require more effort than manual assessment. However, in longer cores this approach could provide a chronology with greater consistency and less effort. This consistency is particularly valuable for conductivity measurements in the brittle zone of an ice core where multi-parameter chemical analysis is not possible.

1.4.6 Winstrup et al (2012)

This paper presents a univariate method of automated layer counting that uses a hidden Markov model (HMM) algorithm. Their approach is applied to a section

of the δ^{18}O record from the DYE-3 ice core of southern Greenland, and to three consecutive sections of the visual stratigraphy (VS) signal from the NGRIP ice core of central Greenland.

Outline of the approach

A log-normal prior distribution is assigned to the thickness, $d \in \mathbb{Z}$, of each layer:

$$d \sim \log \mathrm{N}(\mu_d, \sigma_d^2),$$

assumed independent. The likelihood is derived from how well a given section of signal fits to a discretized cycle template, determined by principal component analysis (PCA) on a number of manually picked cycles, via Bayesian regression. For a stretch of signal, $\boldsymbol{o}$, between depths t_1 and t_2 this is defined as:

$$b(\boldsymbol{o}_{t_1:t_2}) = p(\boldsymbol{o}_{t_1:t_2} \text{ represents an annual cycle as defined by the template}).$$

This likelihood is calculated for each possible subsection of the signal, up to a maximum layer thickness. The forward-backwards algorithm is used to calculate $\eta_t(j, d)$: the posterior probability that layer j is of length d and ends at depth t, for all combinations of t, j and d – up to a maximum number of layers. Summing over d gives the posterior probability that layer j ends at depth t, which can then be used to calculate a chronology.

The parameters for this model come from the Bayesian regression model used to calculate the likelihood $(\Phi, \varphi, \sigma_\epsilon)$, and from the prior distribution for the layer thicknesses (μ_d, σ_d). μ_d and φ are fit using the iterative expectation-maximisation (EM) algorithm, whereas σ_d, Φ, and σ_ϵ are estimated from a manual layer count of a portion of the signal, and are fixed in order to constrain the algorithm.

The signals are broken down into smaller batches of around 50 years, determined manually, to run the algorithm. These are pulled back together by convolving the resulting layer probability distributions. Details of this non-trivial step are

not discussed in the paper, but do feature in the thesis from which this work originates (Winstrup, 2011). This approach allows μ_d to adapt to changes in layer thickness along the core.

Results

The DYE-3 δ^{18}O data ($0 - 800$m) is first normalised with respect to the variance of a three year subsection. The algorithm is applied separately to both the raw data and its derivative after being passed through a first-order Savitzky-Golay filter; it is not clear which results are presented. Five iterations were used to tune the σ_d and φ parameters via the EM algorithm. This analysis gives a count of 1926, 95% CI $(1919, 1932)$, compared to a manual count of 1900 layers.

The more challenging NGRIP VS signal (2200m–2240m) is split into three sections to perform separate analyses: section 1 (2200m–2220m) a warm period; section 3 (2225m–2240m) a cold period; and the transition between (section 2). Each of these sections are split further into 50 year subsections as described above. The data is first log transformed and then normalised over a moving window of 10cm with respect to its maximum and minimum values. The σ_d, Φ, and σ_ϵ parameters are estimated separately from manual layer counts consisting of 1/4 of each section, which I assume was also used to calculate the layer shape template via PCA. The algorithm fit: 894 layers $(882, 905)$ to section 1 vs. a manual count of 839 $(792, 996)$; 291 years $(284, 298)$ to section 2 vs. a manual count of 278 $(266, 290)$; and 1249 years $(1234, 1265)$ to section 3 vs. a manual count of 1204 $(1142, 1266)$. This gives differences of 6.6%, 4.7% and 3.7% from the most likely manual count respectively.

Discussion

This method is not fully Bayesian as it requires an extensive manual study of the data prior to analysis, in order to assign a layer template and to fix some of the model parameters. A more general layer template, based on a sine wave, was explored in the thesis from which this work originates (Winstrup, 2011), however

this approach was abandoned by the authors. The likelihood used considers each potential layer in the signal independently, and does not take into account the data either side. The resulting chronologies give higher cycle counts than the manual analysis in each example given, suggesting that the manual counts from which the method is calibrated are an underestimate of around 5%. However, the uncertainty bounds do overlap for the NGRIP examples.

1.4.7 Gay et al (2014)

This paper introduces a univariate ice core dating method based on spectral analysis: Fourier time-frequency and wavelet transforms. The nitrate (NO_3) signal from the first 70m of a 137m ice core drilled from the summit of Nevado Illimani in the Bolivian Andes in 1999 is used to illustrate the approach.

Pre-processing

The largest peaks in the signal are first reduced in size by thresholding, with the cut-off parameter corresponding to the 98th percentile of the data. The raw signal is unequally spaced, and is oversampled onto equally spaced depths – a form of spline interpolation was chosen over linear, nearest neighbour, hermite and polynomial flavours as it yielded the most favourable signal for spectral analysis. The mean value of the resulting signal is then subtracted, and the resulting signal run through a high pass filter with cut-off frequency f_b.

Summary

Point estimates for the spectral frequency at each depth of the processed signal are calculated using a window of size N around that depth. N is chosen as the value that produced the best spectrogram plots, the criteria are not discussed. N is constant, causing the number of years over which the frequency estimates are calculated to increase down the core.

The frequency estimates are then smoothed using a moving median approach with

window length M. The smoothed frequency estimates are assumed constant over the depth increments and used to estimate the amount of time that has past over each depth step, $\hat{f}_s(d)$.

$\hat{f}_s(d)$ increases down the core to a depth of around 50m then reduces. The decreasing part of $\hat{f}_s(d)$ is reflected around its maximum to make it increasing to the final considered depth of 70m.

The $\hat{f}_s(d)$ values are summed to obtain a chronology. The chronology is tuned on two volcanic marker horizons using parameters f_b, from the pre-processing, and M.

Discussion

Spectral analysis has previously been used to estimate the average annual layer thickness at a given depth as a precursor to a dating method in McGwire et al (2011). Dating an ice core using these estimates directly, as suggested here, seems likely to introduce systematic bias into the chronology in the absence of regular fixed time constraints on which to tune the parameters. This approach hinges on getting very accurate estimates for the spectral frequency and smoothing these estimates seems counter-intuitive. It would be interesting to see this method applied to a much shorter signal with very clear seasonality where the resulting chronology can be assessed.

1.5 Example ice core signals

In this thesis we use chemistry and isotope signals from three ice cores on which to experiment and test our dating methods: the Gomez, Fletcher and NGRIP ice cores. Each is discussed below.

1.5.1 The Gomez ice core

The Gomez ice core (Thomas et al, 2008, 2009) was drilled from the Gomez Plateau, a high accumulation site on the Antarctic Peninsula, by a team from BAS in 2007 and extends to 136m. We use the Hydrogen peroxide (H_2O_2) signal to date this core (Sections 2.2.4, 3.2.6 and 7.6.2) as it clearly has the strongest annual component of the Gomez chemistry signals, and is therefore the most reliable and realistic choice within this data-set. We use 2cm average H_2O_2 concentrations determined from continuous measurements along the core.

H_2O_2 is particularly suitable as a chronological marker in ice cores. It is created in the atmosphere by a chemical reaction that requires ultraviolet light. Due to the very strong seasonality in Antarctica (from complete darkness in midwinter to 24 h daylight in midsummer), there is a strong and regular seasonality in ultraviolet light, and a theoretical basis for expecting a single quasi-sinusoidal variability in H_2O_2 concentration. Measurements of H_2O_2 dissolved in the ice therefore provide a good annual cycle. We also use the more challenging non-sea-salt sulphur (nss-s) signal to test the robustness of our classification dating method to its asymmetrical cycles in Section 3.2.8.

This core has been manually dated to 153 years by determining the depths of the lowest points ("nadirs") of the annual troughs.

1.5.2 The Fletcher ice core

The Fletcher ice core was drilled from the southern West Antarctic Ice Sheet (WAIS) on the Antarctica Peninsula in January 2012. This 650m core, covering an estimated $130,000$ years, was drilled in collaboration between BAS and Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), and the scientific program is being led by Dr Robert Mulvaney. A preliminary $\delta^{18}O$ signal (2.48m–80.60m) has been provided to us by BAS to see whether it is a strong enough signal to date using our methods. See Sections 2.3, 3.2.9 and 7.6.1.

1.5.3 The NGRIP ice core

The North Greenland Ice Core Project (NGRIP or NorthGRIP) ice core (Dahl-Jensen et al, 2002; Andersen et al, 2004) was drilled from the ice sheet of central Greenland which is over 3km thick. Drilling took place between 1996 and 2003, extracting 3.5m cores of 11cm diameter at a time. The chemistry signals are measured at 1mm intervals. Note that this represents the sampling frequency of the CFA measurements and not the nominal resolution. The latter is controlled by dispersion in the CFA system and is more in the order of 1 cm (Lambert et al, 2012).

Section 2.2.5 shows the standardisation process for all of the annually cyclic chemistry signals from the NGRIP ice core between 1440m–1465m – the depth range that we have available. These data cover part of the Holocene climatic period (Rasmussen et al, 2006) and include ammonium (NH_4), calcium (Ca), nitrate (NO_3), sodium (Na) and insoluble dust. The NGRIP ammonium and calcium signals are chosen as test datasets throughout this thesis as they are out of phase with each other and show strong annual seasonality.

1.6 Summary

It can be argued that manual layer counting is open to inconsistency and human error, leading to the necessity for multiple re-counts and comparisons, and that uncertainty measures on the resulting chronologies are subjective. It is also quite laborious and time consuming. A consistent and automated method for layer counting, on multivariate ice core signals, that provides a chronology with a robust uncertainty measure, would clearly be beneficial to the climate science community. The literature to date has only addressed the limited univariate case and, until recently, with no rigorous measure of uncertainty on layer markings. It should also be noted that whilst we focus here on ice core signals, the development of an automated method of layer counting will also have applications to other annually layered palaeoclimate records.

The main goal of the thesis is to develop a fully Bayesian approach to layer counting which provides a marginal posterior distribution for the time of year, as well as the date, at each depth of an ice core. With a simple and flexible model for the signals, we develop a Markov chain Monte Carlo approach to reconstruct the underlying periodic process – allowing for the frequent clusters of missing values. The latent chronology is sampled directly in a way that allows the number of cycles in the reconstruction to be changed without the need for dimension-changing algorithms such as Reversible Jump. We allow for the dependence in observation error and the lack of stationarity by modelling means, amplitudes and errors as continuous functions of depth.

In this thesis we present a method to standardise an ice core signal, isolating its seasonality; and to split it into sections with well defined, and uncertain, cycle counts. We show how the uncertain sections can be presented for manual assessment, and describe how the possible reconstructions can be identified and assigned probabilities based on their length. We then develop the MCMC methodology with illustrative examples; firstly on standardised signals, then raw or log signals, and finally describe how the approach can be extended to the multivariate framework and scaled for use on full size cores.

Chapter 2

Standardisation

2.1 Introduction

In this chapter we introduce a method of processing an ice core signal so that its annual cycles are expressed as a noisy sine wave on a non-linear timescale. This process has three stages:

1. if required and possible, transform the signal to improve the visual symmetry of the cycles,
2. calculate a rough estimate for the average cycle length at each depth along the signal,
3. smooth the signal with respect to this estimate:
    - de-trend the signal to leave only the seasonality,
    - normalise the de-trended signal – set constant annual cycle amplitudes.

The underlying assumption to this process is that the seasonality in the signal is annual in period and consistently follows a ... peak, descending, trough, ascending ... pattern. Once standardised, we have two methods of counting the annual

cycles and measuring our uncertainty in the count. In Chapter 3 the basic properties of a sine wave are exploited to classify the signal into seasons, or quarter years, which can then be counted. A measure of uncertainty based on linear regression modelling of the season lengths against depth can then be calculated. In Chapter 4 a fully Bayesian MCMC method of modelling the standardised signal as a noisy sine wave is explored. Further chapters look at performing the standardisation process dynamically as part of the MCMC algorithm.

2.2 Equally spaced data

2.2.1 Transformation

All ice core signals with strong annual seasonality that I have examined, through papers, talks and collaboration, can be categorised into two groups: those with (roughly) symmetrical cycles; and those with short thin peaks and long flat troughs. The former do not require any transformation whereas the latter's symmetry is generally improved with a log transformation. The ammonium (NH_4) signal from the NGRIP ice core is an example of the latter group, Figure 2.1 below shows the process of taking the natural logarithm of a short stretch of the signal (1441m–1442m). The y-axis symmetry of the cycles is clearly improved, and this is the case for all of the annually cyclic NGRIP signals (see Figures 2.10–2.14).

Taking logarithms typically does not adversely affect the cycle symmetry of signals that already have fairly symmetrical cycles. This suggests that a logarithmic transformation could be used as a first step in a fully automated dating method, however different signals may require other transformations or pre-processing.

2.2.2 Estimating cycle length – ACF

Our standardisation process requires a rough initial estimate for the average annual cycle length (in depth units) at each depth. Methods used to obtain this estimate as a precursor to layer counting in the literature include Fourier analysis

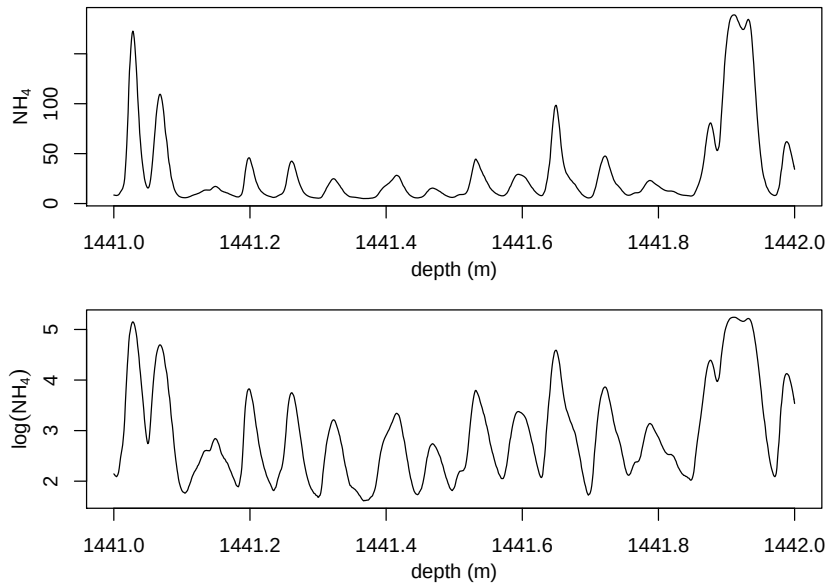

Figure 2.1: Log transformation of the NGRIP NH_4 signal (1441m–1442m) to improve symmetry in the annual cycles.

(McGwire et al, 2011) and ice flow modelling (Shimohara et al, 2003). We use the autocorrelation function (ACF) to find an estimate of cycle length in any sufficiently large subsection of an equally spaced signal. We use this estimate solely for the smoothing process, and not directly for the actual layer counting. At lag k, the ACF of a signal $\boldsymbol{x}$ of length n is defined as:

$$R_k = \frac{1}{(n-k)\sigma^2} \sum_{i=1}^{n-k} (x_i - \mu)(x_{i+k} - \mu).$$

where μ is the signal's mean and σ is the signal's variance. It measures the cross-correlation of a signal with itself as a function of difference in depth (the lag), and is a useful tool for finding repeating patterns. Note that $R_0 = 1$, always. Figure 2.2 (left) shows the ACF of a sine wave over two cycles. At a lag of 1/2 a cycle the signal is almost perfectly negatively correlated ($R \approx -1$), at a lag of 1 cycle the signal is almost perfectly positivity correlated ($R \approx 1$). This pattern goes on forever. Figure 2.2 (right) shows the ACF for the NGRIP log NH_4 signal between $1440 - 1465$m up to a lag of 124mm – approximately 2 cycles. A clear peak can be seen at a lag of 62mm ($R \approx 0.3$) which corresponds to one average

cycle length in this section of signal, 62mm in this case. Comparing the two ACF plots, there are clearly other sources of variation present in the log NH_4 signal besides the annual seasonality.

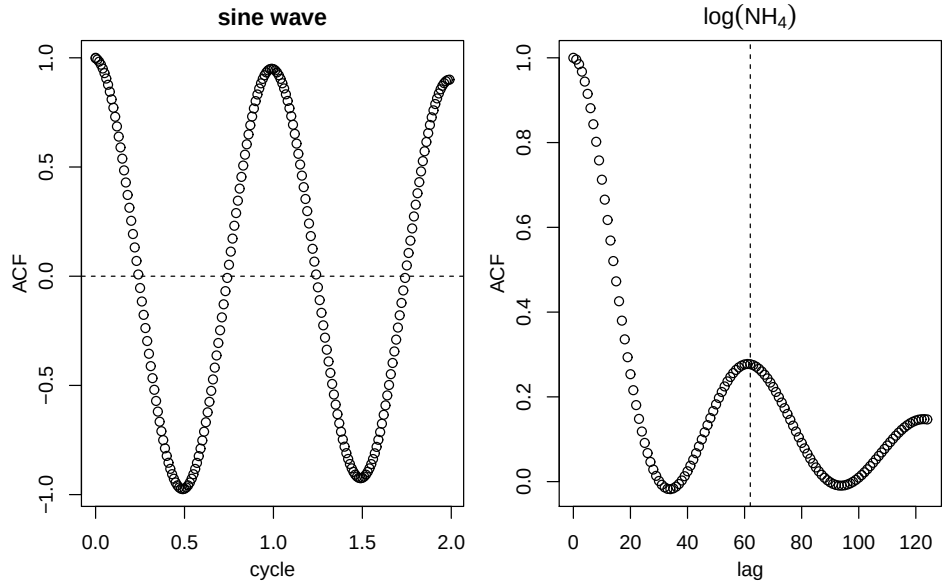

Figure 2.2: (left) the ACF of a sine wave. (right) the ACF of the NGRIP log NH_4 signal between 1440m and 1465m.

Automation

Automating the above process can answer two questions: is there a regular cyclic element to the signal and, if so, what is the average length of the cycles? The algorithm goes like this:

1. calculate the ACF of the signal up to some sensible maximum lag m: R_k, $k = 0, 1, \ldots, m$
2. estimate the average cycle length l as the first $k < m$ such that $R_{k-1} > R_{k-2}$ and $R_k < R_{k-1}$.

Figure 2.3 (top) shows the NGRIP log NH_4 signal between 1445m and 1446m, its ACF (middle) and differenced ACF ($R_k - R_{k-1}$) (bottom). This section has an estimated average cycle length of $l = 62$mm, found using a maximum lag of $m = 120$mm.

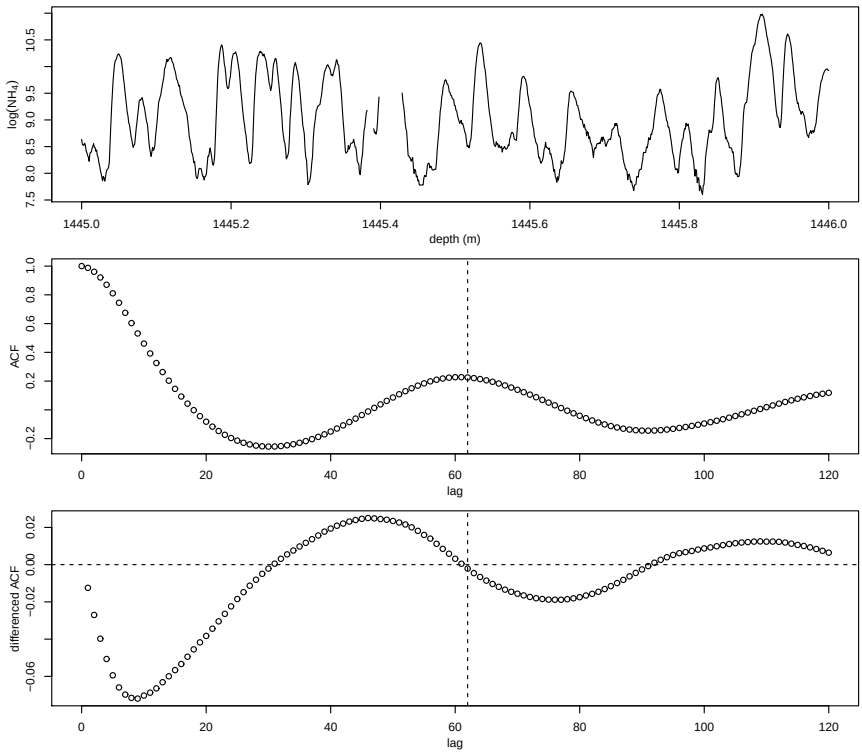

Figure 2.3: (top) the NGRIP log NH_4 signal between 1445m and 1446m. (middle) its ACF. (bottom) its differenced ACF.

Using this method we calculate the estimated average cycle length in non-overlapping 20cm windows over a 1440m–1465m section of the NGRIP log NH_4 signal. Figure 2.4 (left) shows the estimates plotted against depth with no obvious visible trend. A discussion of how to deal with trends in the estimated cycle lengths can be found below. Figure 2.4 (right) is a histogram of the estimates. They have a mean of 61.7mm and a standard deviation of 10.1 which gives a 95% confidence interval of $(59.9, 63.5)$ for the mean.

We detected regular annual cycles in 4 other signals from the NGRIP core. Figure 2.5 shows their ACFs and estimated average cycle lengths: calcium ($l = 62$), nitrate ($l = 60$), sodium ($l = 63$) and dust ($l = 62$).

Dealing with trends

In the example section of the NGRIP ice core above (1440m–1465m) the estimated average cycle length can be assumed constant at a mean level of approximately

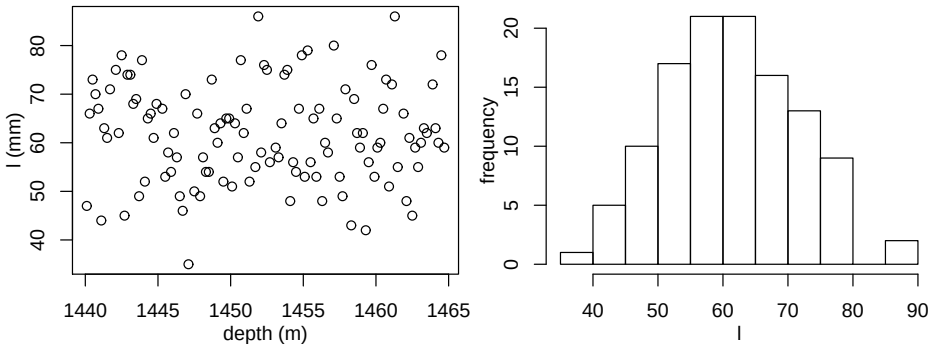

Figure 2.4: (left) estimated average cycle lengths, l, of 20cm non-overlapping sections of the NGRIP log NH_4 signal between 1440m and 1465m. (right) a histogram of these lengths.

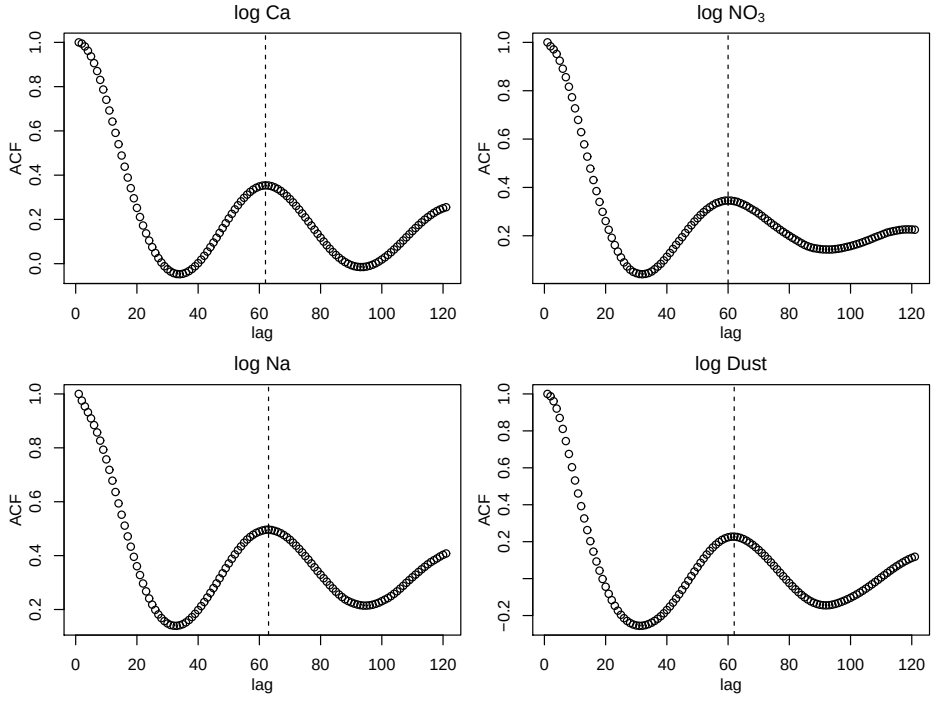

Figure 2.5: ACFs for the log calcium, nitrate, sodium and dust signals from the NGRIP ice core (1440m–1465m). The vertical dashed lines indicate the local maxima of the ACFs, used to obtain an initial estimate of cycle length.

62mm. When a trend is seen, generally a thinning of the annual cycles down the core, a little more care is required.

Figure 2.6 shows the estimated average cycle length over 4m non-overlapping subsections of the Gomez H_2O_2 signal which is sampled at 2cm depth intervals. This shows a clear decreasing trend with depth from around 2m at the top of the

core to around 0.5m at the bottom. This is caused by the compaction of snow into ice and the subsequent thinning of the ice layers under pressure from the weight above.

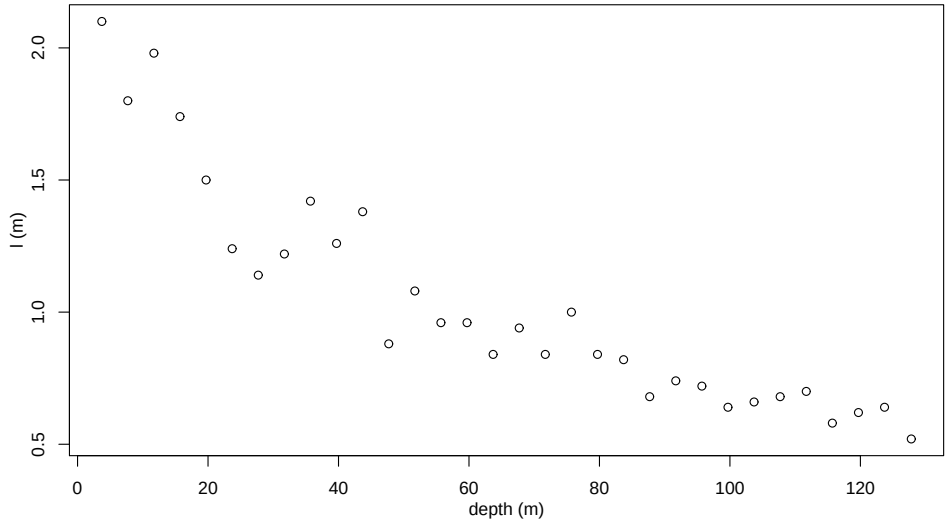

Figure 2.6: The estimated average cycle length, l, of 4m sections of the Gomez H_2O_2 signal, found using the autocorrelation Function.

For such types of signals it is necessary to estimate the average cycle length in sections. The sections ought to be long enough to obtain a good estimate, but not too long as to be averaging over too many of the trending cycles. To achieve this we split the signal $\boldsymbol{x}$ into β non-overlapping sections, each containing approximately the same number of cycles. This is achieved via an algorithm that perturbs the section boundaries with respect to the expected number of cycles in each section. Firstly the signal is split into β sections using $(\beta - 1)$ boundaries that are equally spaced in depth, then we iterate as follows:

1. calculate the expected average cycle length, l_j, for each section using its ACF, $j \in (1, 2, \ldots, \beta)$;
2. redistribute the boundaries so that section j contains $[nl_j / \sum_j l_j]$ points;
3. recalculate l_j for each section;
4. if the expected numbers of cycles in all sections are equal then stop, otherwise go to 2.

Each data point is then assigned the average cycle length for its section, to be used in the smoothing process described in Section 2.2.3.

Figure 2.7 (top) shows Gomez H_2O_2 signal split into $\beta = 6$ sections with equal expected numbers of cycles. Figure 2.7 (bottom) shows the ACFs of the second (circles) and fifth (stars) of the sections, plotted against lag, measured as a number of observations. The vertical dashed lines indicate the local maxima of the ACFs, used to obtain an initial estimate of cycle length. The second section has an estimated 52 points in an average annual cycle, and is of length 1237 points. The fifth section has an estimated 32 points in an average annual cycle, and is of length 761 points. Both sections therefore contain an estimated 23.78 annual cycles, as do all six sections. This gives an initial estimate of approximately 143 cycles contained in $\boldsymbol{x}$, which is an underestimate due to the stretches of missing values.

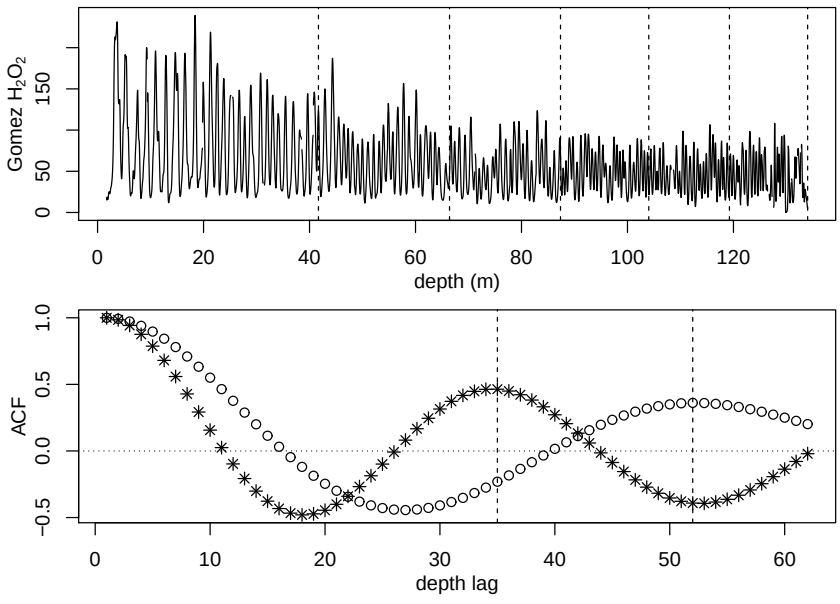

Figure 2.7: (top) the Gomez H_2O_2 signal split into $\beta = 6$ sections with an equal expected number of cycles. (bottom) the ACFs of the second (circles) and fifth (stars) of the sections, plotted against lag, measured as a number of observations. The vertical dashed lines indicate the local maxima of the ACFs, used to obtain an initial estimate of cycle length.

Treatment with a filter

The ACF method of estimating the average annual cycle length in a section of ice core signal breaks down when there is a lot of other lower (or higher) frequency noise. This is the case for the ammonium signal deeper down the NGRIP core. One option would be to use a manual (prior) estimate for such sections. However, if automation is a priority, then a high (or low) pass filter could be used with a priori settings to leave only cycles with a frequency in a reasonable bound.

Figure 2.8 (first) shows the NGRIP log NH_4 signal between $1477 - 1478$m, the second plot shows its ACF – where an estimate for the cycle length cannot be determined using our method. The third plot shows the signal after treatment with a second order high-pass Butterworth filter with a cut-off frequency of $1/50$ (Butterworth, 1930; signal developers, 2013). This should allow only cycles with a length of < 50mm to pass through. An average cycle length estimate of 46mm can then be found from its ACF (the fourth plot).

2.2.3 The standardisation process

Once we have initial estimates of cycle length, we can use them to define an appropriate local standardisation. We use a point-wise standardisation to smooth a signal $\boldsymbol{x}$: at each depth i

$$s_i = \frac{x_i - \mu_i}{\sqrt{2}\sigma_i}.$$

$\boldsymbol{\mu} = \{\mu_i\}$ is a moving average of $\boldsymbol{x}$ and $\boldsymbol{\sigma} = \{\sigma_i\}$ is a moving standard deviation of $(\boldsymbol{x} - \boldsymbol{\mu})$, both over the range of one annual layer thickness based on a preliminary estimate at each depth and centred on that depth. If l_i is the estimate for the number of points in a cycle at depth i then

$$\mu_i = \sum_{j=i-l_i/2}^{i+l_i/2} \frac{x_i}{l_i + 1}$$

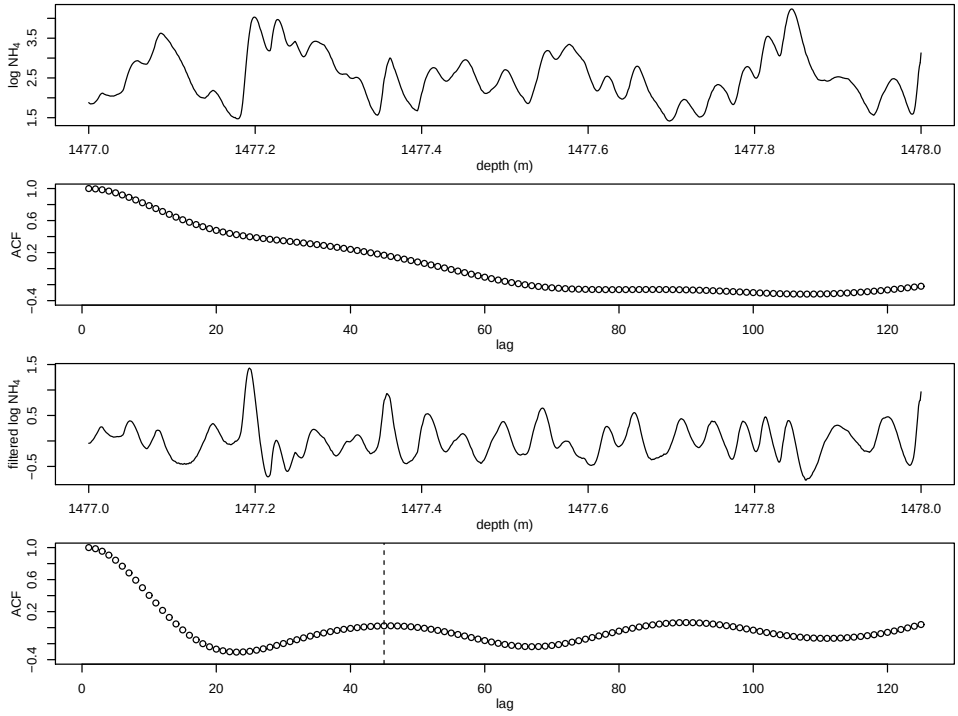

Figure 2.8: (first) NGRIP log NH_4 signal between 1477m and 1478m. (second) Its ACF. (third) The signal after treatment with a second order high-pass Butterworth filter, cut-off frequency: 1/50. (fourth) The resulting ACF.

for even l_i, odd values are rounded up. $\boldsymbol{\mu}$ measures the trend in the data, the annual seasonality being mostly averaged out due to the interval length being approximately a whole year. Subtracting $\boldsymbol{\mu}$ de-trends $\boldsymbol{x}$ and centres $\boldsymbol{s}$ on 0.

$$\sigma_i^2 = \sum_{j=i-l_i/2}^{i+l_i/2} \frac{(x_i - \mu_i)^2}{l_i + 1}$$

for even l_i. $\sqrt{2}\boldsymbol{\sigma}$ is an estimate for the annual cycle amplitude at each depth. Dividing by $\sqrt{2}\boldsymbol{\sigma}$ sets the apices and nadirs of the annual cycles in $\boldsymbol{s}$ to a magnitude of approximately 1. In the case of missing values, the corresponding $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ values are linearly interpolated from the closest surrounding points where there are sufficient data.

2.2.4 Gomez H_2O_2

Figure 2.9 illustrates the smoothing process for $\beta = 6$; each plot has vertical lines that represent the section boundaries. The top plot shows the signal $\boldsymbol{x}$, with its annual moving average $\boldsymbol{\mu}$ as a dotted line. The middle plot shows the de-trended signal, with its estimated amplitude $\sqrt{2}\boldsymbol{\sigma}$ as a dotted line. The bottom plot is of $\boldsymbol{s}$, the standardised signal. There is still some variation in mean and amplitude visible due to the crude estimate of cycle length at each depth, however this is sufficient to serve as a starting point for our automated dating scheme.

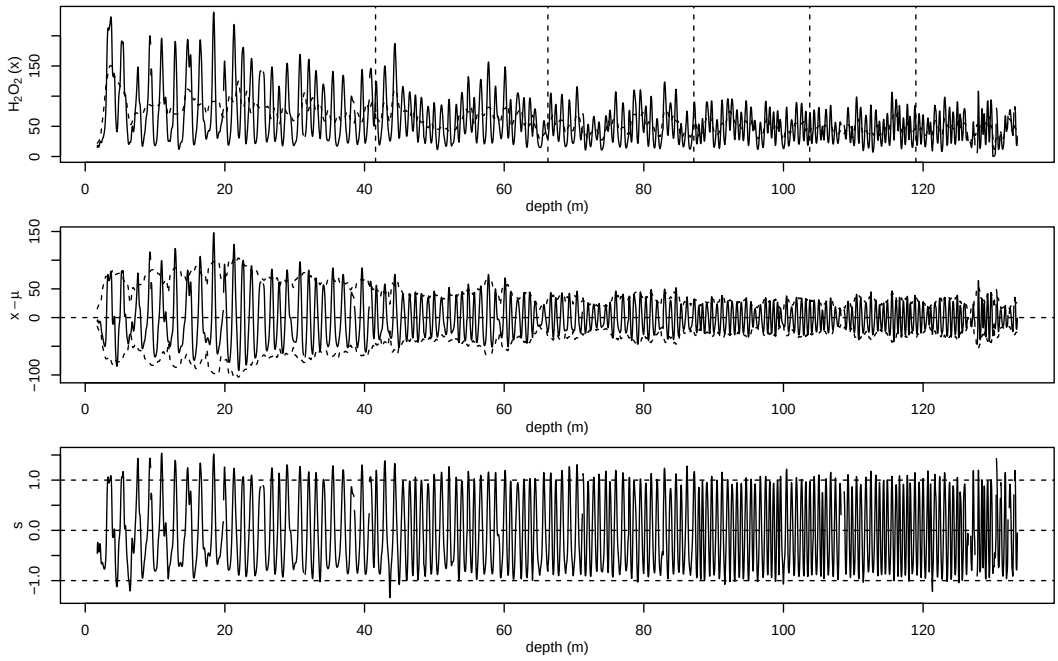

Figure 2.9: The process of standardising the Gomez H_2O_2 signal, using initial estimates for annual cycle length based on $\beta = 6$ sections, boundaries are shown as vertical lines. (top) The signal $\boldsymbol{x}$, with its annual moving average $\boldsymbol{\mu}$ as a dotted line. (middle) The de-trended signal, with its estimated amplitude $\sqrt{2}\boldsymbol{\sigma}$ as a dotted line. (bottom) The standardised signal $\boldsymbol{s}$.

2.2.5 NGRIP examples

Figures 2.10-2.14 illustrate the standardisation process for each of the annually cyclic chemistry signals from the NGRIP ice core between 1440m and 1465m. These include ammonium (NH_4), calcium (Ca), nitrate (NO_3), sodium (Na) and

insoluble dust. In each case the cycle symmetry is improved by taking logs. The first plot is the raw signal $\boldsymbol{x}$; the second is the log transformed signal; the third is the de-trended log signal $\log(\boldsymbol{x}) - \boldsymbol{\mu}$; the fourth is the standardised signal $\boldsymbol{s} = (\log(\boldsymbol{x}) - \boldsymbol{\mu})/(\sqrt{2}\boldsymbol{\sigma})$.

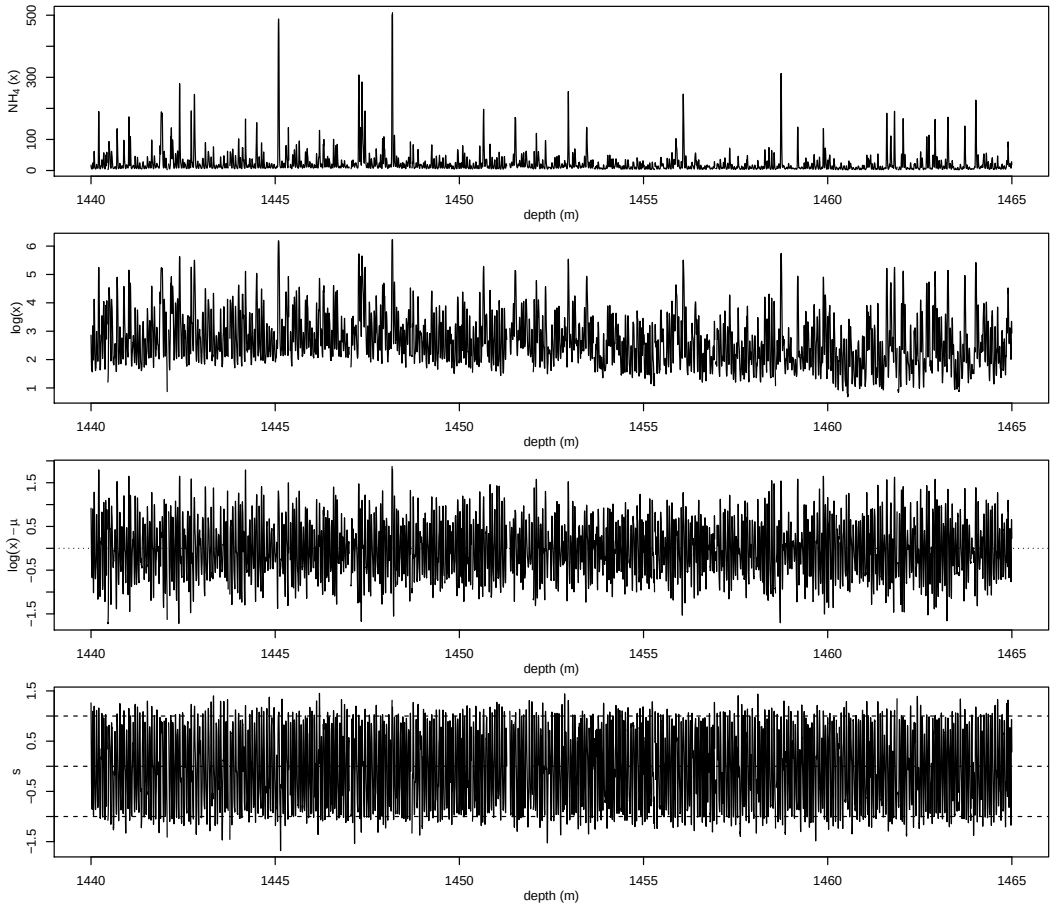

Figure 2.10: The process of standardising the NGRIP ammonium signal (1440m–1465m). (first) The raw signal $\boldsymbol{x}$. (second) The log transformed signal. (third) The de-trended log signal $\log(\boldsymbol{x}) - \boldsymbol{\mu}$. (fourth) The standardised signal $\boldsymbol{s} = (\log(\boldsymbol{x}) - \boldsymbol{\mu})/(\sqrt{2}\boldsymbol{\sigma})$.

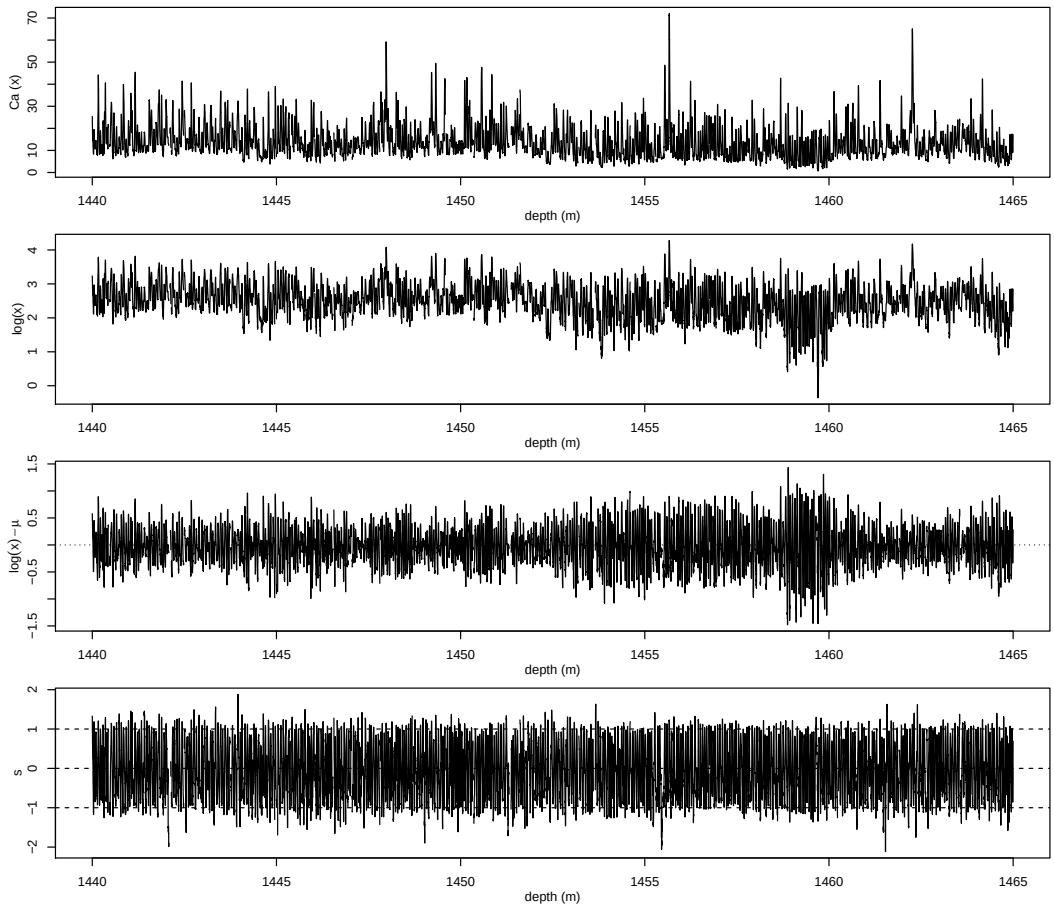

Figure 2.11: The process of standardising the NGRIP calcium signal (1440m-1465m). (first) The raw signal $\boldsymbol{x}$. (second) The log transformed signal. (third) The de-trended log signal $\log(\boldsymbol{x}) - \boldsymbol{\mu}$. (fourth) The standardised signal $\boldsymbol{s} = (\log(\boldsymbol{x}) - \boldsymbol{\mu})/(\sqrt{2}\boldsymbol{\sigma})$.

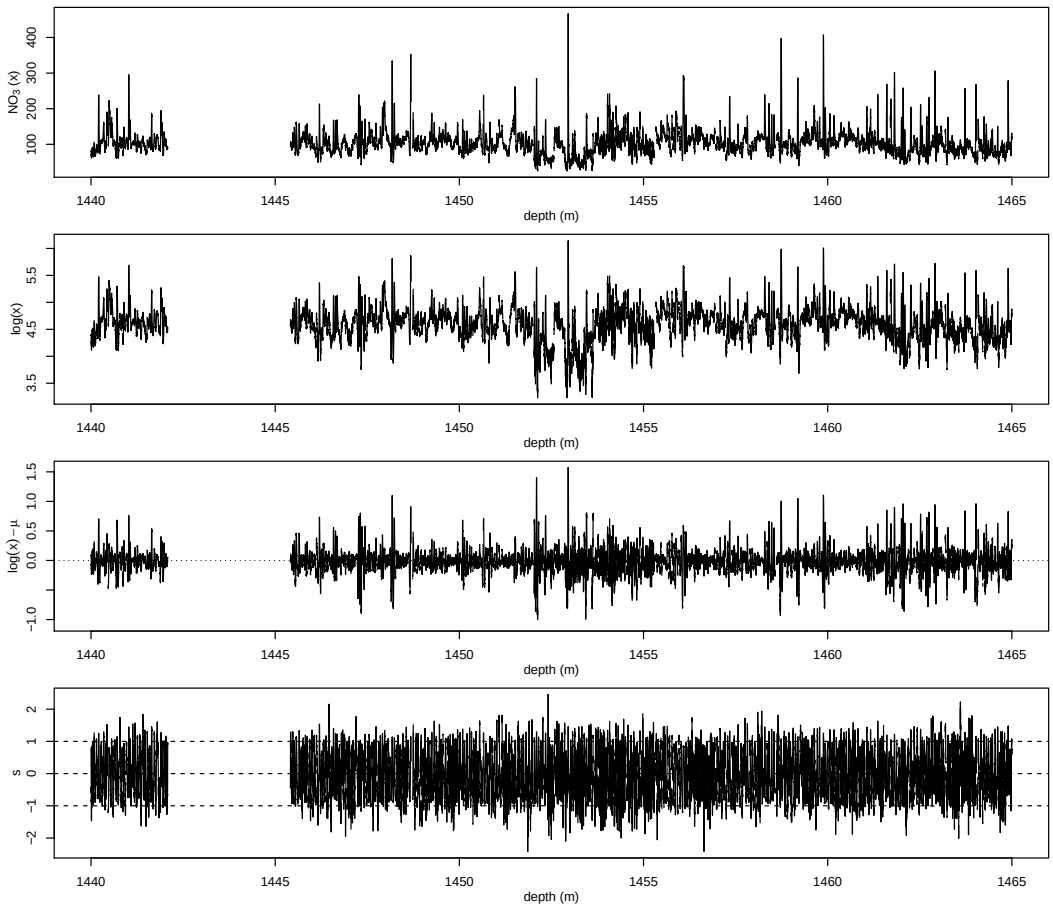

Figure 2.12: The process of standardising the NGRIP nitrate signal (1440m–1465m). (first) The raw signal $\boldsymbol{x}$. (second) The log transformed signal. (third) The de-trended log signal $\log(\boldsymbol{x}) - \boldsymbol{\mu}$. (fourth) The standardised signal $\boldsymbol{s} = (\log(\boldsymbol{x}) - \boldsymbol{\mu})/(\sqrt{2}\boldsymbol{\sigma})$.

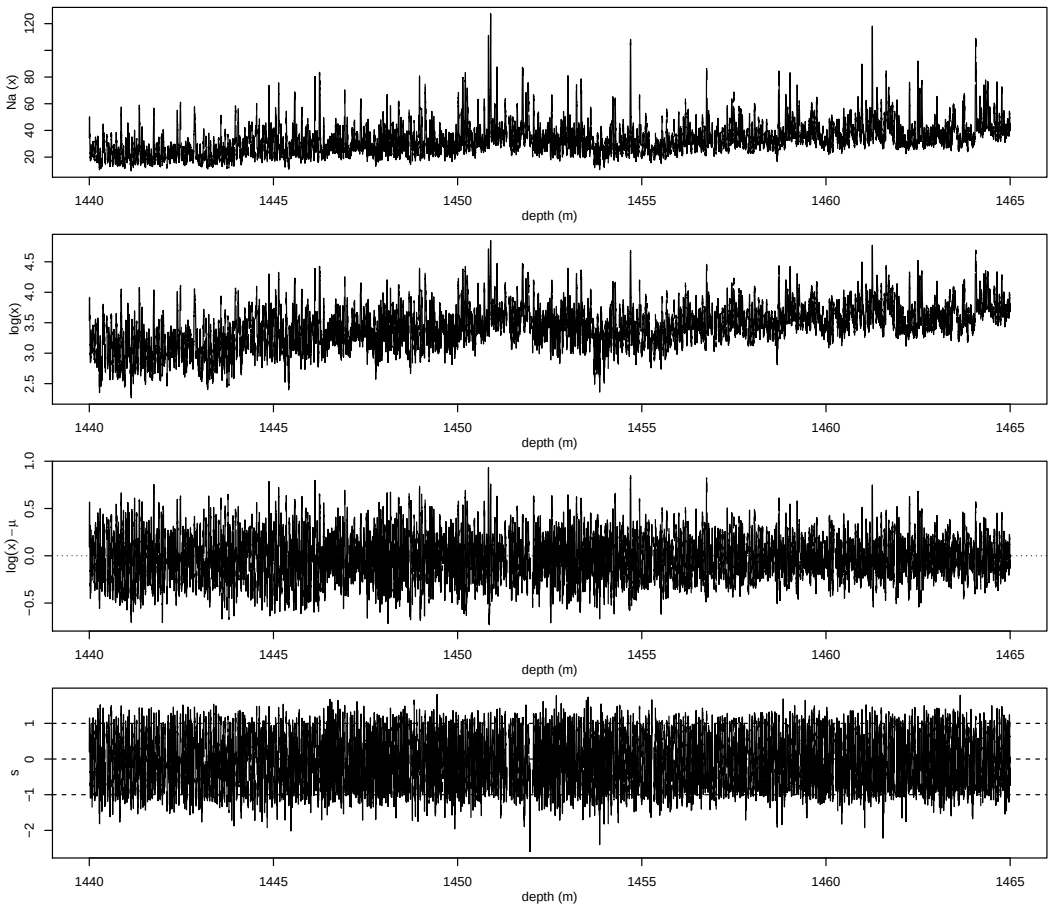

Figure 2.13: The process of standardising the NGRIP sodium signal (1440m–1465m). (first) The raw signal $\boldsymbol{x}$. (second) The log transformed signal. (third) The de-trended log signal $\log(\boldsymbol{x}) - \boldsymbol{\mu}$. (fourth) The standardised signal $\boldsymbol{s} = (\log(\boldsymbol{x}) - \boldsymbol{\mu})/(\sqrt{2}\boldsymbol{\sigma})$.

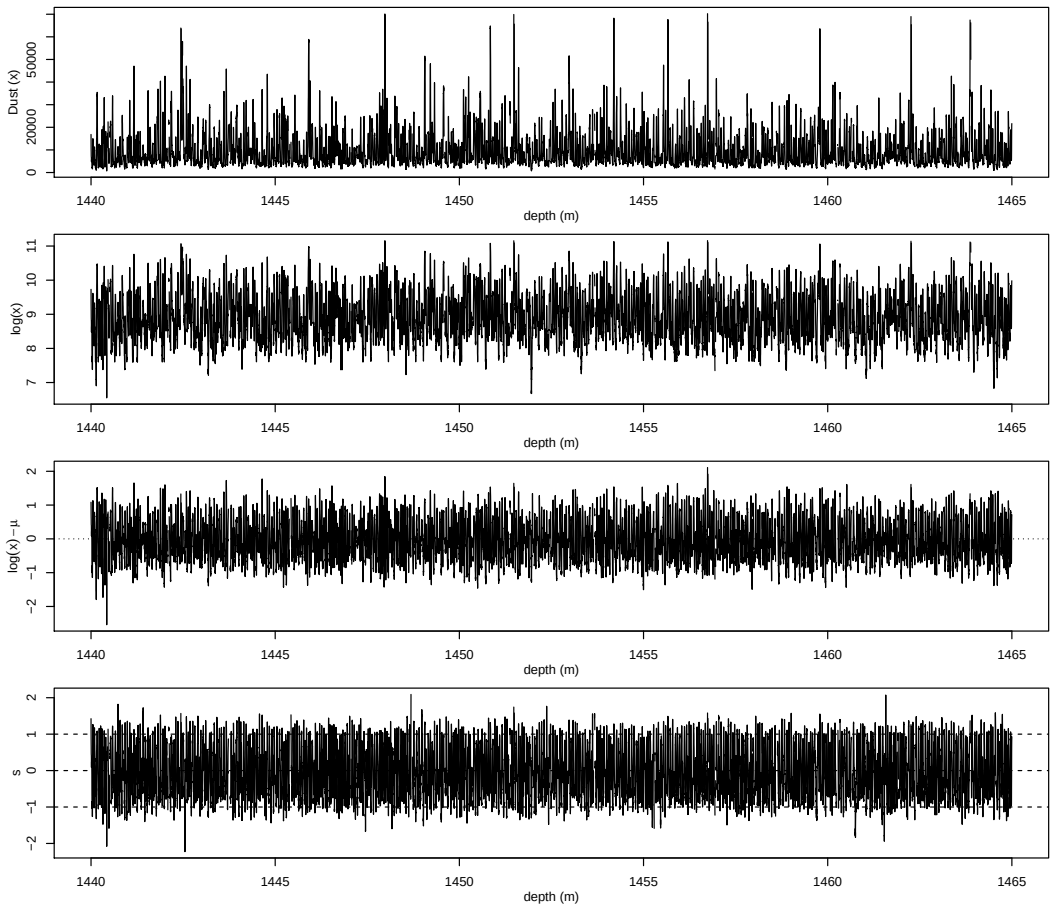

Figure 2.14: The process of standardising the NGRIP dust signal (1440m–1465m). (first) The raw signal $\boldsymbol{x}$. (second) The log transformed signal. (third) The detrended log signal $\log(\boldsymbol{x}) - \boldsymbol{\mu}$. (fourth) The standardised signal $\boldsymbol{s} = (\log(\boldsymbol{x}) - \boldsymbol{\mu})/(\sqrt{2}\boldsymbol{\sigma})$.

2.2.6 Under-smoothing and over-smoothing

Figure 2.15 illustrates what happens if the estimated average cycle length l_i, used for the interval lengths to calculate $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, is incorrect. This is the same stretch of NGRIP NH_4 signal as in Figure 2.1 (1441m–1442m). In the top plot an interval length of half the estimated average cycle length ($l_i/2$) is used – this picks up noise fluctuations that are too small to be annual cycles. In the middle plot an interval length of l_i is used – this appears to do an adequate job at creating a noisy sinusoid from the annual cycles. In the bottom plot an interval length of $2l_i$ is used – some of the variation in mean and amplitude are still present in the standardised signal, but short years may be smoothed into adjacent cycles.

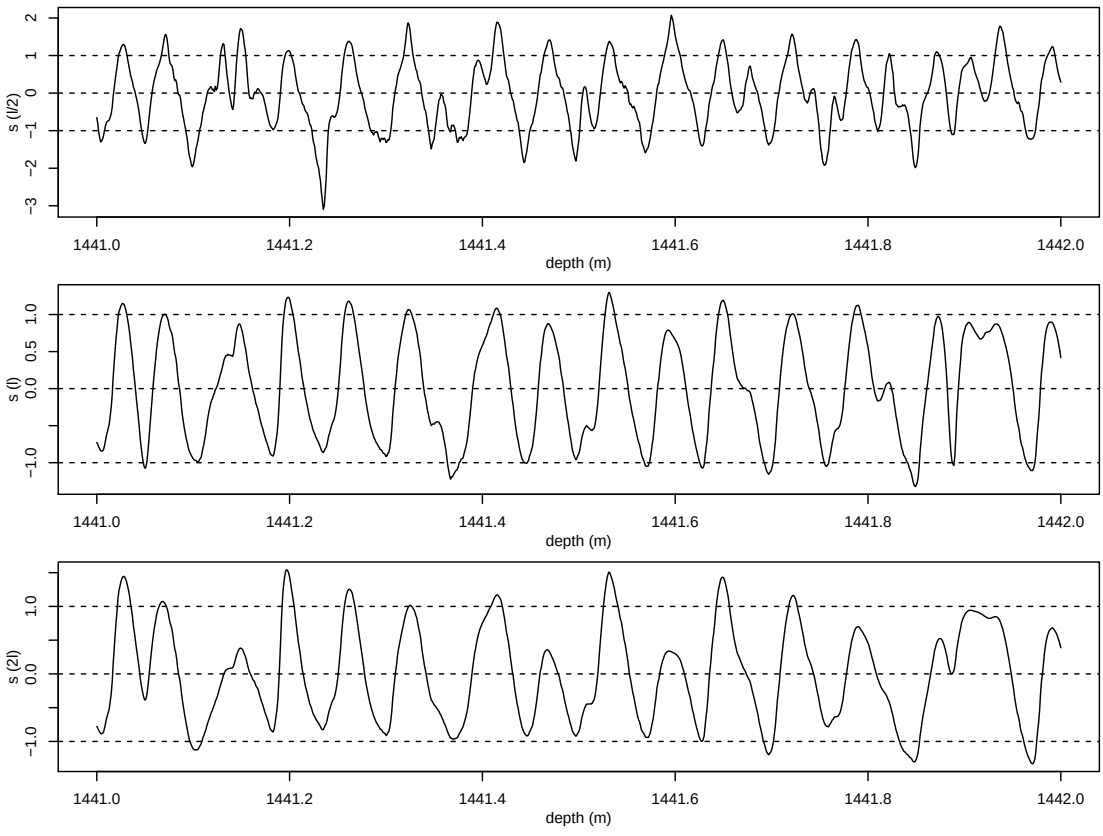

Figure 2.15: Three standardisations of the NGRIP NH_4 signal (1441m–1442m) using different interval lengths to calculate $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. (top) $l_i/2$. (middle) l_i. (bottom) $2l_i$.

The dating methods described in the next two chapters use standardised signals.

They are robust to 'small' errors when estimating l_i, generally giving the same cycle count with increasing uncertainty with error magnitude. However, errors as large as those shown above may cause them to fail. Overestimation of l_i has a less adverse effect on the resulting count than underestimation. In all of the examples in this research, the estimate calculated using the ACF as described above has been adequate.

2.3 Unequally spaced data - Fletcher δ^{18}O

The Fletcher δ^{18}O signal is sampled at unequally spaced depths. In general, each data point of an unequally spaced signal x_i is sampled at depth d_i, for depth indexes $i = 1, \ldots, n$. Figure 2.16 is a scatter plot of the depth spacing of the Fletcher δ^{18}O signal against depth.

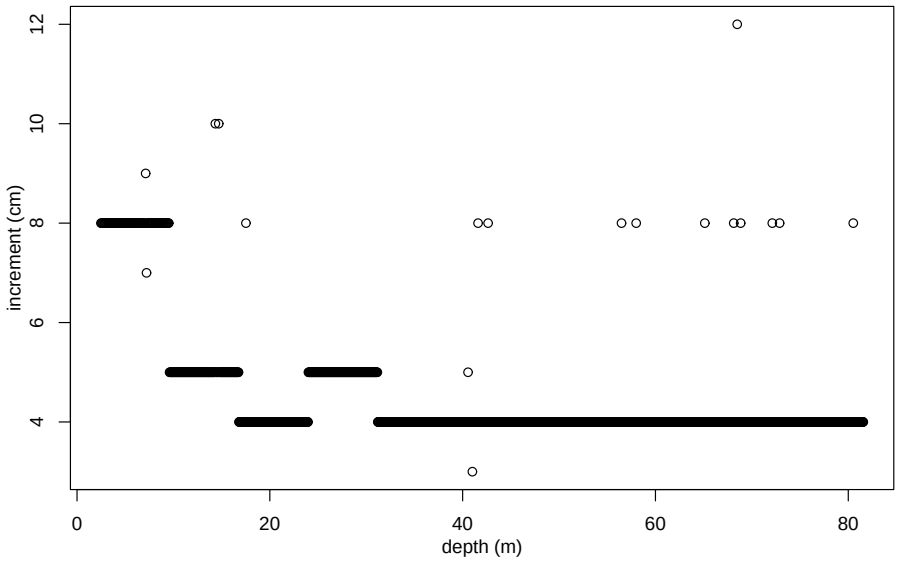

Figure 2.16: Scatter plot of the depth spacing of the Fletcher δ^{18}O signal.

One option that is commonly used in the literature is to interpolate the signal onto equally spaced depths. Assuming $d_{i-1} < d_i < d_{i+1}$ we can interpolate to $\hat{d}_j, j = 1, \ldots, \hat{n}$ of spacing Δ, take

$$d_{\min} = \Delta\lceil d_1/\Delta \rceil \quad \text{and} \quad d_{\max} = \Delta\lfloor d_n/\Delta \rfloor$$

$$\hat{n} = \frac{d_{\max} - d_{\min}}{\Delta} + 1$$

$$\hat{d}_j = d_{\min} - (j-1)\Delta$$

$$\hat{x}_j = x_i + \frac{\hat{d}_j - d_i}{d_{i+1} - d_i}(x_{i+1} - x_i)$$

where $d_i \leq \hat{d}_j \leq d_{i+1}$ are the nearest depths that bound $\hat{d}_j$. The interpolated dataset can then be treated as above.

A second option is to use a continuous analogue of the equal spacing standardisation techniques. The continuous analogue of the ACF gathers the spacing of the signal into bins and calculates the lagged correlation based on each bin. We use the `correlogram` function of the Spatial package in R (Venables and Ripley, 2002), which has the number of bins used, n_{int}, as a parameter. This is used to calculate the average cycle length, a multiple of the bin size, in a section of signal in the same way as the ACF. The signal can then be split into β sections as before. A slight adaptation to the standardisation process is required: calculating $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ from data points within an average cycle length around each interpolated depth index. Figure 2.17 illustrates the continuous standardisation of the Fletcher δ^{18}O signal with $\beta = 5$ and $n_{\text{int}} = 150$ throughout; section boundaries are shown as vertical lines. The n_{int} parameter was chosen by trial and error – although a good value could be calculated from prior knowledge of cycle lengths in the signal.

Figure 2.18 shows the continuous standardisation of the final (fifth) section to illustrate the effectiveness of the method. The continuous standardisation process is not robust to the choice of n_{int} and so this method is not desirable for a fully automated dating method. In further chapters, when working with the Fletcher δ^{18}O signal, we interpolate using $\Delta = 4$cm so that most data points remain the same. Looking at Figure 2.16 this is a natural choice for the depth spacing.

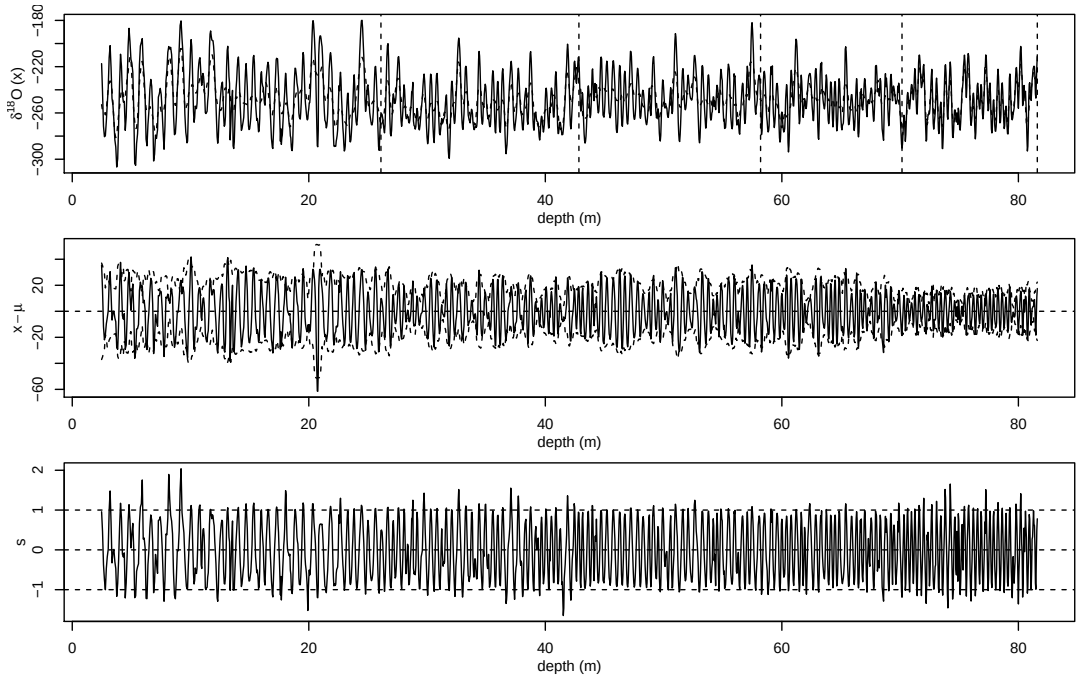

Figure 2.17: Continuous standardisation of the δ^{18}O signal from the Fletcher core, using initial estimates of annual cycle length based on $\beta = 5$ sections, section boundaries are shown as vertical lines. (top) The signal $\boldsymbol{x}$, with its annual moving average $\boldsymbol{\mu}$ as a dotted line. (middle) The de-trended signal, with its estimated amplitude $\sqrt{2}\boldsymbol{\sigma}$ as a dotted line. (bottom) The standardised signal $\boldsymbol{s}$.

2.4 Conclusions

In this chapter we have introduced a method for use on ice core signals that:

- detects whether seasonality is present,
- provides a rough estimate for the average cycle length in a given section,
- breaks the signal down into a chosen number of subsections that contain roughly the same number of cycles,
- standardises it to approximately a noisy sinusoid on a non-linear timescale.

For equally spaced data this process is fully automatic and requires no user input other than the number of subsections. This process can be used as a first step to automating layer counting and is used in some way in all further chapters.

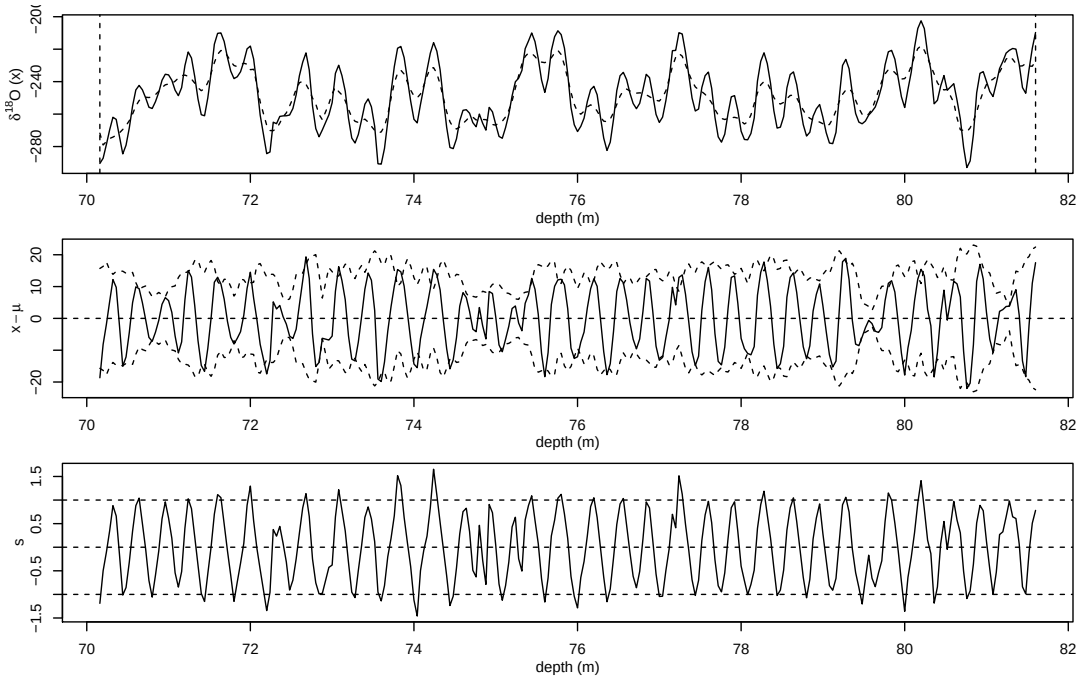

Figure 2.18: Continuous standardisation of the final (fifth) section of the δ^{18}O signal from the Fletcher core, using initial estimates of annual cycle length based on $\beta = 5$ sections – section boundaries are shown as vertical lines. (top) The signal $\boldsymbol{x}$, with its annual moving average $\boldsymbol{\mu}$ as a dotted line. (middle) The detrended signal, with its estimated amplitude $\sqrt{2}\boldsymbol{\sigma}$ as a dotted line. (bottom) The standardised signal $\boldsymbol{s}$.

Chapter 3

Classification algorithms

3.1 Introduction

Once a signal has been standardised, as is described in the previous chapter, it can be thought of as a 'noisy' sinusoid on a non-linear timescale. The visible smoothness of most ice core signals suggest that this noise is strongly correlated. We utilise the properties of a sine wave to split the signal into sections: those with a deterministic cycle count and those that need more attention. The possible reconstructions for uncertain sections have a known form and can be presented in a useful way as a visual aid for manual counting. With a view to automation, a simple method for assigning probability measures to each possible reconstruction is presented. In this chapter the univariate case is thoroughly explored and a discussion of possible multivariate extensions is provided in Section 3.3.

3.2 Univariate classification

We use the standardised Gomez H_2O_2 record as a test signal on which to count annual cycles and explain the univariate classification methods. The standardisation process for this signal is illustrated in Figure 2.9.

Robustness is explored by applying our method to thinned versions of the NGRIP

ammonium and calcium signals which show more variation in annual layer thickness than the Gomez core; see Section 3.2.7. A non-parametric version of our classification method is applied to the Gomez non-sea-salt sulphur signal, as an example of how our general framework can be adapted to accommodate asymmetrical data; see Section 3.2.8. The method is also applied to the Fletcher δ^{18}O signal after linear interpolation onto equal depth spacing in Section 3.2.9.

We assume throughout that data points are equally spaced in depth, as is the case for the Gomez and NGRIP examples, and that depth can therefore be represented as an integer index; as explained in Section 3.2.5 this is purely for notational convenience, and is not inherent in the method.

3.2.1 Certain annual cycles

In large-scale manual layer counting exercises, annual cycles have been termed "certain" if they are judged through consensus of multiple expert counters to have a probability of at least 3/4; see Andersen et al (2006). Here we present a method of automatically classifying quarter cycles, analogous to seasons in the H_2O_2 signal, as being "certain" if they are well-defined in the standardised signal with respect to our repeatable algorithm; we do not however attempt to match the '3/4 consensus ratio'. This classification process is very simple in terms of computation and is therefore very quick, taking less than a second on a modern lap-top for the Gomez core. It is consistent, repeatable, and does not require any prior manual assessment of chronology.

3.2.2 Classification into runs

We aim to segment the points of the standardised signal $\boldsymbol{s}$ into non-overlapping subsections termed *runs*, each representing either a "certain" quarter cycle, or an "issue" where manual intervention is required. Each run is a collection of consecutive points and has one of five labels:

P: peak/summer;

D: descending/autumn;

T: trough/winter;

A: ascending/spring;

χ: issue.

This classification is a three-stage process. Firstly we find potential quarter cycles; these cannot contain missing values as we are unsure of their classification. For some sensible choice of cut-off parameter ν: all runs of consecutive data points $s_i \geq \nu$ are labelled as P^*, the star meaning potential; all runs of data points $s_i \leq -\nu$ are labelled as T^*. Runs of data points for which $-\nu < s_i < \nu$ are labelled as potentially ascending and descending: A^* if they fall between a T^* and P^*; D^* if they fall between a P^* and a T^*; and χ^* otherwise. So the stretch of data in Figure 3.1a with $\nu = 1/\sqrt{2}$, has potential run label pattern:

$$\ldots, \mathrm{T}^*, \mathrm{A}^*, \mathrm{P}^*, \mathrm{D}^*, \mathrm{T}^*, \mathrm{A}^*, \mathrm{P}^*, \mathrm{D}^*, \mathrm{T}^*, \chi^*, \mathrm{T}^*, \mathrm{A}^*,$$
$$\mathrm{P}^*, \chi^*, \mathrm{P}^*, \mathrm{D}^*, \mathrm{T}^*, \mathrm{A}^*, \mathrm{P}^*, \mathrm{D}^*, \mathrm{T}^*, \mathrm{A}^*, \mathrm{P}^*, \ldots$$

In the second stage, potential runs are labelled as certain runs only if they are central to a consecutive set of size 5 whose labels obey the pattern expected from a sinusoidal signal. In the example, this gives the following:

$$\ldots, \mathrm{T}, \mathrm{A}, \mathrm{P}, \mathrm{D}, \mathrm{T}, \mathrm{A}, \mathrm{P}, \mathrm{D}^*, \mathrm{T}^*, \chi^*, \mathrm{T}^*, \mathrm{A}^*,$$
$$\mathrm{P}^*, \chi^*, \mathrm{P}^*, \mathrm{D}^*, \mathrm{T}, \mathrm{A}, \mathrm{P}, \mathrm{D}, \mathrm{T}, \mathrm{A}, \mathrm{P}, \ldots$$

Finally, consecutive data points which make up the unlabelled and potential runs are collected together into runs labelled as χ, and termed *issues*. This gives:

$$\ldots, \mathrm{T}, \mathrm{A}, \mathrm{P}, \mathrm{D}, \mathrm{T}, \mathrm{A}, \mathrm{P}, \chi, \mathrm{T}, \mathrm{A}, \mathrm{P}, \mathrm{D}, \mathrm{T}, \mathrm{A}, \mathrm{P}, \ldots$$

Each point of $\boldsymbol{s}$ is now an element of exactly one run, and each run that is not labelled as χ makes up a certain quarter annual cycle – these are termed certain runs and add 1/4 year to the chronology. In Figure 3.1 runs labelled P are coloured red, D runs are orange, T runs are blue, A runs are green, and issues are black.

The value of the threshold used here, $\nu = 1/\sqrt{2} \approx 0.707$, is natural as in a perfectly symmetric signal it would lead to the same proportions of points in each of the four types of quarter cycles. Other choices have their merits; the choice of ν is further discussed in Sections 3.2.3 and 3.2.5, and robustness to this choice is considered in Section 3.2.6.

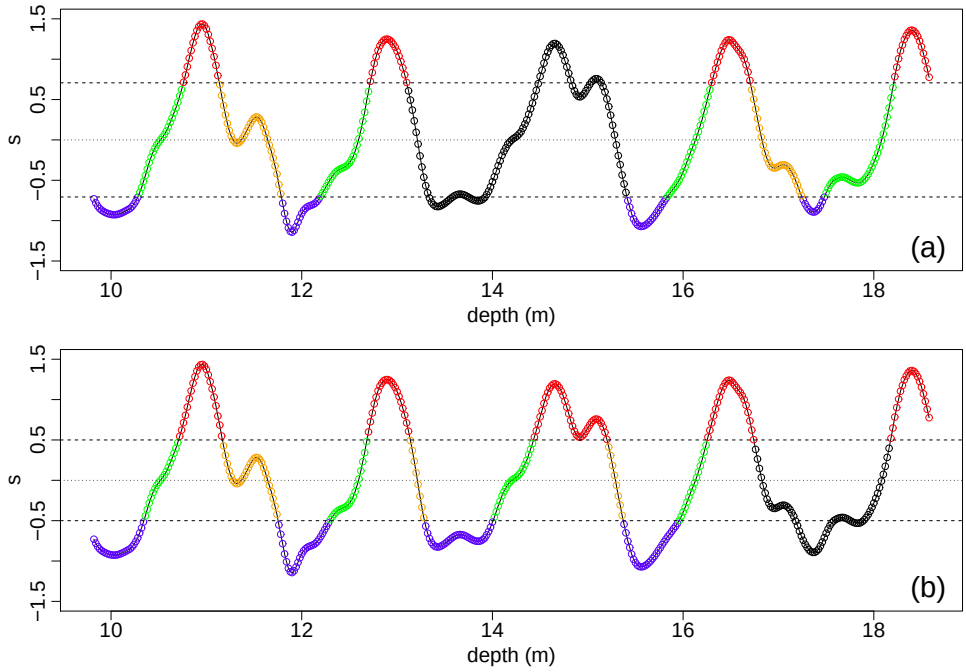

Figure 3.1: Part of the standardised Gomez H_2O_2 signal, for depths between 9.8m and 18.6m, showing the "issues" (regions of dating uncertainty) caused by fluctuations in the signal and identified using two different values of the threshold ν for classifying individual points. Points within "issues" are black; points within peaks (labelled P in main text) are coloured red, descending points (D) are orange, troughs (T) are blue, and ascending points (A) are green. **(a)** Threshold $\nu = 1/\sqrt{2}$. **(b)** Threshold $\nu = 0.5$.

3.2.3 Issues

Runs labelled χ, or *issues*, consist of consecutive data points that are missing values, that surround missing values, or that are in some other way dubious in their periodicity. Sections of signal where there are no issues have a deterministic number of cycles: 1/4 of a cycle per certain run.

Figure 3.1 shows a stretch of $\boldsymbol{s}$ between 9.8m and 18.6m, classified using $\nu = 1/\sqrt{2}$ (a) and $\nu = 1/2$ (b). The issues here are caused by the run pattern alone and contain no missing values. At $\nu = 1/\sqrt{2}$ there is an issue between 13.1m and 15.4m: a peak dips below $1/\sqrt{2}$, and a trough rises above $-1/\sqrt{2}$. At $\nu = 1/2$ the points corresponding to this issue now make up five certain runs; however there is a new issue between 16.8m and 18.2m.

Issues such as those seen in Figure 3.1 occur only in the first 20m of the Gomez core where, due to the very high resolution of the sampling with respect to the annual cycle length, small fluctuations are found in the annual cycles. Choosing different values of ν affects where data points are classified into certain runs and therefore the distribution of issues in this first part of the signal.

Figure 3.2 shows a stretch of $\boldsymbol{s}$ between 106.7m and 110.2m with $\nu = 1/\sqrt{2}$. This has an issue caused by a stretch of missing values, the non-missing data points that also form part of the issue are shown as a black line. In this case the choice of ν only affects the length of the issue, which decreases with ν.

For the Gomez H_2O_2 signal standardised using $\beta = 6$ and classified using $\nu = 1/\sqrt{2}$, $\boldsymbol{s}$ has 528 deterministic runs and 12 issues; with $\nu = 1/2$, $\boldsymbol{s}$ has 533 deterministic runs and 15 issues.

We can now split the signal into sections with a deterministic count and sections that need more attention. At this stage, the issues could be presented to experts as in Figure 3.1 so they can place certain and uncertain layer markers or their own subjective probability measures. However, the run pattern provides more information that could be exploited either to assist the expert analysis or to

allow further automation.

3.2.4 Reconstructions

The labels of the certain runs that bound an issue provide insight into the label pattern of the runs that could replace it. It is convenient to refer to peak and trough runs collectively as *extreme* runs; similarly ascending and descending runs are termed *central*. The classification algorithm described above is such that issues are always bounded by extreme runs, leading to four possible cases:

$$\ldots, \mathrm{P}, \chi, \mathrm{P}, \ldots \quad (1) \qquad \ldots, \mathrm{P}, \chi, \mathrm{T}, \ldots \quad (2)$$
$$\ldots, \mathrm{T}, \chi, \mathrm{P}, \ldots \quad (3) \qquad \ldots, \mathrm{T}, \chi, \mathrm{T}, \ldots \quad (4)$$

The points of $\boldsymbol{s}$ that correspond to an issue can be replaced with a compatible section of sine wave, spread evenly over the points, referred to as a *reconstruction*. This could be used to fill in missing values, or replace sections of data affected by a sampling problem or where the annual signal isn't sufficiently clear, but is primarily intended as a visual aid for manual counting.

The minimal reconstructions for cases (1) and (4) consist of 3 runs, as the minimal compatible sections of sine wave have run label patterns D, T, A and A, P, D, respectively. The minimal reconstructions for cases (2) and (3) consist of 1 run labelled as D and A, respectively. Further reconstructions can be found by adding in whole cycles.

Each possible reconstruction is made up of an odd number of runs. If m is the number of runs in its minimal reconstruction and k cycles are added in, the issue would contain $d = m + 4k$ runs. Issues are always bounded by extreme runs and will therefore always contain $\frac{d-1}{2}$ extreme runs and $\frac{d+1}{2}$ central runs.

Figure 3.2 shows possible reconstructions of an issue caused by missing values. The top plot shows the minimal reconstruction with 1 run, and the bottom shows the reconstruction with three additional cycles added in which consists of 13 runs. Plots like this could be used as visual aids to the manual assessment of issues.

However, the available information can be used to assign probabilities to each reconstruction based solely on their length, which would provide an automated method of layer counting, or give further guidance in semi-automated counting.

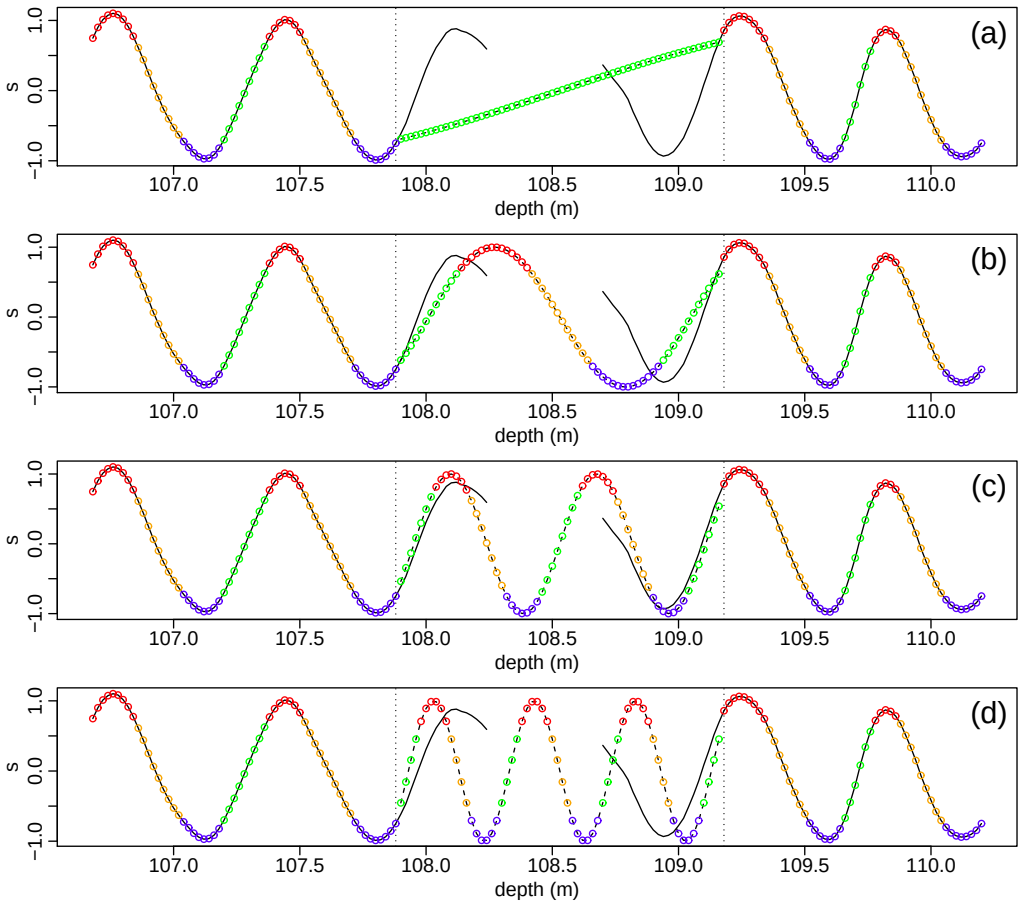

Figure 3.2: 4 possible reconstructions of an issue (region of dating uncertainty, bounded by the dotted lines) caused by missing values in the standardised Gomez H_2O_2 signal (black curve) at a depth of around 108.5m. Coloured points within the issue represent imputed values for missing observations, or conjectured "corrected" values for nearby observations; coloured points outside the issue indicate the classification of actual observations. In each case, points classified as being within peaks (labelled P in main text) are coloured red, descending points (D) are orange, troughs (T) are blue, and ascending points (A) are green. The four different reconstructions are indexed by d, the number of runs (sets of consecutive points classified in the same way) used to reconstruct the issue, including the adjacent runs; adding an extra year to the reconstructed chronology increases d by four. **(a)** The minimal reconstruction, with no additional annual cycles, involving $d = 1$ runs. **(b)** The reconstruction with one additional annual cycle and $d = 5$. **(c)** The reconstruction with two additional annual cycles and $d = 9$. **(d)** The reconstruction with three additional annual cycles and $d = 13$.

It should be noted here that the classification algorithm has a number of possible variations in the way it is set up, which each lead to different issues and thus forms for the reconstructions. The method presented here was chosen as it minimises the number of possible reconstructions.

3.2.5 Assigning probabilities

The *length* of a run, ℓ, is used here to denote the number of data points it contains. Provided that the time-depth relationship does not change too rapidly, we would expect a run of a given type to have a similar length to other such runs near to it within the core. This concept is key to existing manual and semi-automated layer counting approaches. Here we present a method of assigning probabilities to the possible reconstructions of each issue by comparing ℓ to its implied distribution. For simplicity we are assuming throughout that points are equally spaced in depth, as is the case in our examples. Relaxing that assumption is straightforward in principle; it simply requires a slight extension of the notation, to work with lengths expressed in terms of differences of depth rather than just numbers of points. Figure 3.3 is a plot of certain run lengths against depth for a range of ν for the standardised Gomez H_2O_2 signal.

We take $p(d|\ell) \propto p(\ell|d)$; that is, we take the probability of a reconstruction with d runs, given that the issue contains ℓ data points, to be proportional to the probability of those d runs having total length ℓ. This is essentially a Bayesian statistical approach, with a flat prior distribution on d. Again, using an alternative prior distribution would be straightforward, but in practice, prior information is always likely to be dominated by the other information in the core, as represented by $p(\ell|d)$.

Groups of d consecutive certain runs, of which $\frac{d-1}{2}$ are extreme, are analogous to issue reconstructions. Ideally, we would model the lengths of these directly to find the distribution of ℓ given d. However, issues are concentrated in the first 20m of the Gomez H_2O_2 signal due to the fluctuations discussed above, and the

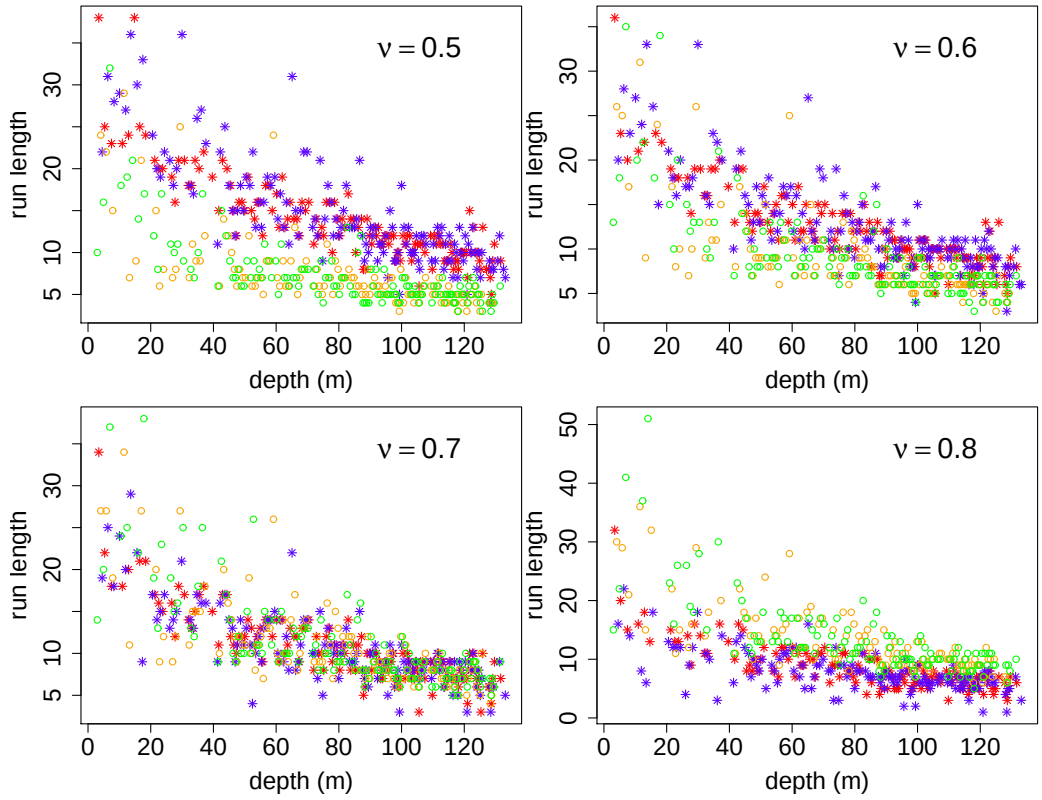

Figure 3.3: Certain run lengths from the classification of the standardised Gomez H_2O_2 signal plotted against depth for a range of the cut-off parameter, ν: runs labelled P are coloured red; D orange; T blue; and A green.

last 20m where there are regular stretches of missing values. In either case, this results in the analogous sections being concentrated at the centre of $\boldsymbol{s}$, resulting in extrapolation and a poor fit to the sections with issues.

Instead, we need to make use of information on individual certain run lengths. In the Gomez H_2O_2 signal, peak run lengths are equivalent in distribution to trough run lengths because of the symmetry in the seasonality of $\boldsymbol{s}$ and in the classification process. When ν is equal to the 75th percentile of a sine wave $(1/\sqrt{2})$ extreme and central run lengths are equivalent in distribution; for $\nu < 1/\sqrt{2}$ extreme run lengths are generally larger than central; and for $\nu > 1/\sqrt{2}$ central run lengths are generally larger than extreme. The lengths of the certain runs are non-linear, and also change in spread, as a function of depth in the Gomez core (see Figure 3.3). This non-linearity in ice core layer thickness is caused by vertical compaction of snow into ice and thinning of the ice layers caused by horizontal

flow.

Other proxy records where annual layer counting is used for establishing chronologies, such as tree rings, varves and corals, will not in general display this systematic reduction in layer thickness through the depth profile. After taking natural logs the individual certain run lengths show a linear trend for the Gomez H_2O_2 series. This is well described by a regression model which has the extreme or central run label as a factor; see Figure 3.4. The model is a linear regression, with independent Gaussian errors and constant variance, fitted using ordinary least-squares estimation within the `lm` function in R (R Development Core Team, 2011). The standard regression diagnostics and residual plots in `plot.lm` in R were all satisfactory. If we know the central depth of a run and its label, we can find its expected length from the model. This model is analogous to the results of Rasmussen et al (2006) where annual layer thicknesses from the NGRIP core are shown to be log-normally distributed after a linear strain correction.

All of the d runs that make up the reconstruction of an issue have implied central depths and labels. The expected value for the lengths of these runs, and therefore the total length of the issue implied by the reconstruction, $\hat{\ell}_d$, can be *interpolated* from the regression model. We assume that $\log(\ell) \sim \mathrm{N}(\log(\hat{\ell}_d), \sigma_d^2)$, where σ_d depends on d and is estimated using groups of certain runs, analogous to that particular reconstruction, as discussed above. Note that σ_d *cannot* be obtained directly from the model for individual runs, because of the dependence in lengths between consecutive runs.

The issue from Figure 3.2 contains $\ell = 64$ data points. The minimal reconstruction for this issue (a) is made up of one quarter cycle ($m = 1$); at this depth $\hat{\ell}_1 = 8$ and $p(\ell|d = 1) = 0$ to three decimal places after normalisation. The second reconstruction contains 5/4 cycles, $\hat{\ell}_5 = 38$, and $p(\ell|d = 5) = 0.005$. At $d = 9$: $\hat{\ell}_9 = 68$ and $p(\ell|d = 9) = 0.984$; and at $d = 13$: $\hat{\ell}_{13} = 95$ and $p(\ell|d = 13) = 0.011$. Continuing to add cycles in this way results in reconstructions with negligible probability.

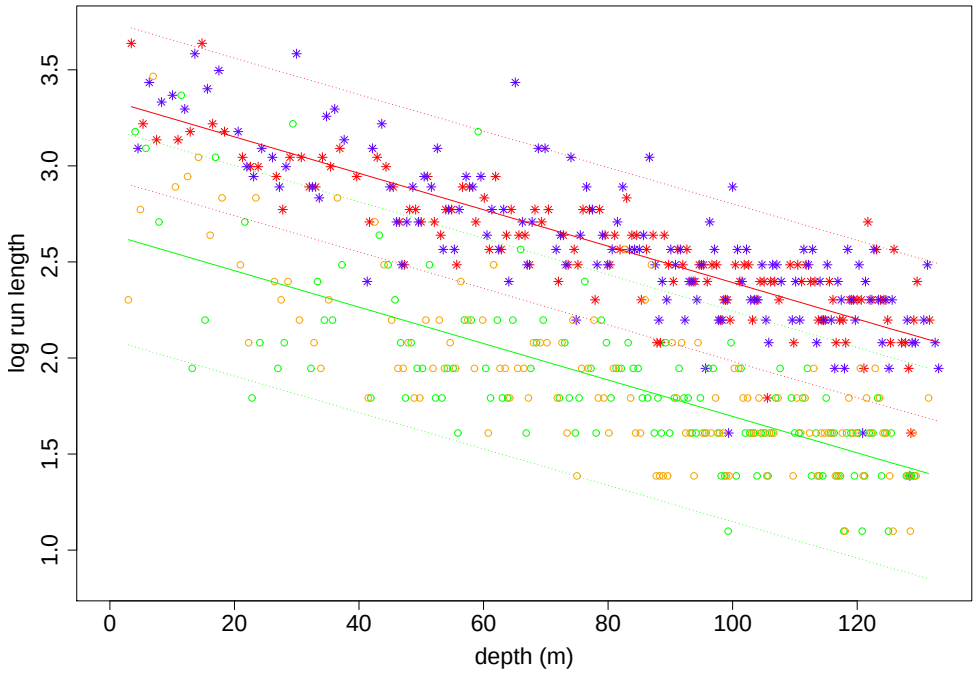

Figure 3.4: A linear regression model for the logarithm of the length of a certain run – a set of consecutive observations that can be definitely classified as belonging to the same part of an annual cycle – against depth within the Gomez core, calculated from its H_2O_2 signal with $\nu = 0.5$ and $\beta = 6$. Runs representing peaks (labelled P in main text) are represented by red stars, descending runs (D) by orange circles, troughs (T) by blue stars, and ascending runs (A) by green circles. The solid lines indicate the fitted values for peak/trough runs (red) and ascending/descending runs (green), and the red and green dotted lines represent the respective 95% predictive intervals for individual runs.

Mudelsee et al (2012) state that "Age-depth modelling must also provide simulated curves, which can then be fed into modern resampling methods of climate time series analysis, resulting in realistic measures of uncertainty in our knowledge about the climate". One benefit of our method over other layer detection schemes is that it simultaneously models the change in layer thickness with depth and, although it is not our main purpose here, it is possible to simulate timescales from the fitted regression model via an iterative scheme. This would require a minor adjustment: taking the run starting depth as a covariate instead of the run central depth. Starting at depth 1, and using the fact that run labels must follow the P, D, T, A pattern, fitted log run lengths can be generated one at a time down the core and perturbed by adding Gaussian noise (using the residual

error). One could continue to generate cycles until a required number or depth is reached. To incorporate the serial dependence between depth points into this scheme it should be possible to measure the correlation between log run lengths of all possible pairs of types, and take this into account when adding noise.

3.2.6 Results for Gomez H_2O_2

The probabilities obtained in Section 3.2.5 can be combined across the whole core, assuming separate issues to be independent given the certain parts of the classification.

Figure 3.5 shows the probability distribution for the number of annual cycle troughs in the Gomez ice core for $\beta = 6$, $\nu = 1/2$ (a) and $\nu = 1/\sqrt{2}$ (b). This is found by combining the probabilities for each possible reconstruction of each issue. The manual count gave 153yr exactly. The differences in these distributions are due to the fluctuations found in the first 20m of signal as discussed in Section 3.2.3. These relate to the date at the bottom of the core, since this is the single point most likely to be of interest, but such distributions could equally be calculated at any required depth.

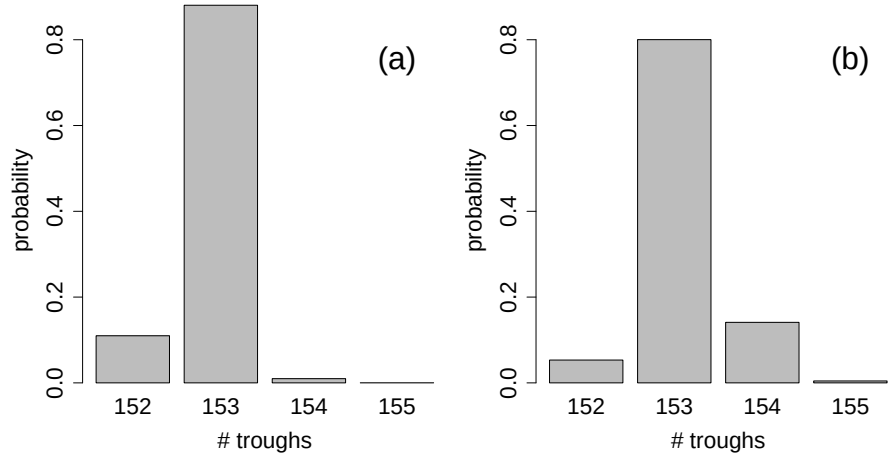

Figure 3.5: Bar plots showing the probability distributions of the number of annual troughs – a summary of the chronology – from the dating of the Gomez core using its H_2O_2, based on two possible values of the threshold ν for classifying individual points, and dividing the core into $\beta = 6$ sub-sections for the initial smoothing process. **(a)** Threshold $\nu = 0.5$. **(b)** Threshold $\nu = 1/\sqrt{2}$.

The most probable reconstruction ($p > 0.8$ in both cases) has 153 troughs and agrees closely with the manually counted solution. Figure 3.6 shows annual cycle lengths (nadir to nadir) as found by manual counting and from the model's most probable reconstruction. Each nadir placed by the model corresponds uniquely to a manually determined one, and their placements agree closely except for two cases: at the top of the core where troughs are wide and determining their nadir is subjective; at the bottom of the core where a number of troughs have been marked in stretches of missing data.

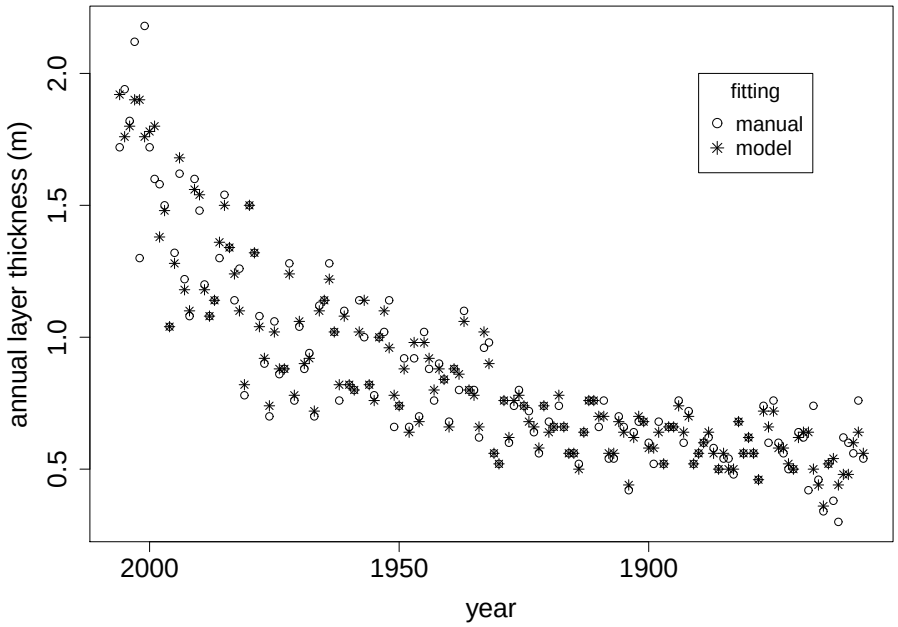

Figure 3.6: Annual cycle lengths for the Gomez H_2O_2 signal: placement from the manual assessment vs. fitted values from the most probable reconstruction of the classification model with $\beta = 6$ and $\nu = 1/\sqrt{2}$.

The signal is standardised in sections with respect to a typical cycle length, which may not be robust to an abrupt change in frequency. Depending on the cut-off ν, it is possible that a relatively short (and therefore uncertain) cycle could be either counted as certain, or missed out altogether, without flagging an issue. Labelling a short cycle as certain would result in several consecutive low valued run lengths, whereas missing a short cycle out would result in one very high valued central run length. Outliers from the regression model have been assessed to test for this possibility and no examples were found. This sense check could be automated.

The Gomez core covers the firnification process, and an exponential decay in layer thickness is observed. There are several cycles in the first 20m of the core that contain over 100 data points (2m), in contrast the last 5m of the core has several cycles with fewer than 20 data points (40cm) - a five fold decrease. This is well modelled by a linear trend on the log transformed thicknesses, fit via simple linear regression, The mean layer thickness (and 99% C.I.s) under this model for the start (@ 3m), middle (@ 65m), and end of the core (@ 132m) are respectively: 171cm (109cm, 268cm), 94cm (59cm, 147cm), and 49cm (31cm, 77cm). Note the asymmetry in the confidence intervals. To allow comparison with other datasets: under this model the probability of a random annual layer being either double or half the mean thickness at any given depth is $\approx 2.5\%$ after allowing for the exponential decay in layer thickness.

Sensitivity to ν

To test the sensitivity of this process to the choice of ν it was run for a wide range of values. Figure 3.7a shows the resulting distributions of cycle counts for $0.3 \leq \nu \leq 0.8$; the areas of the circles are proportional to the probabilities. While not identical, the results are quite stable for ν in this range. For $\nu < 0.3$ the resolution of the data is such that some of the A and D runs at the bottom of the core are of length 1 or missing, and similarly for $\nu > 0.8$ some of the P and T runs at the bottom of the core are of length 1 or missing. This adversely affects the regression model and causes the model assumptions to fail outside of this range.

Sensitivity to β

To test the sensitivity of this process to the choice of β it was run for $2 \leq \beta \leq 20$. Figure 3.7b shows the resulting distributions of cycle counts for $\nu = 1/\sqrt{2}$. For $\beta > 20$ there is not enough data in the first section with which to estimate the average cycle length. For $\beta < 5$ the sections are too long; the thinning of annual cycle length with depth causes the estimated average length to be unrepresentative of cycles at either end of the sections. This results in some of

the mean trend and variation in amplitude still being present in $\boldsymbol{s}$. For $\nu = 1/\sqrt{2}$ a number of peaks and troughs are missed, causing an underestimate in the cycle count. However, $\nu = 1/2$ has many additional issues but still works well in this range, and the equivalent plot shows a stable distribution for all β.

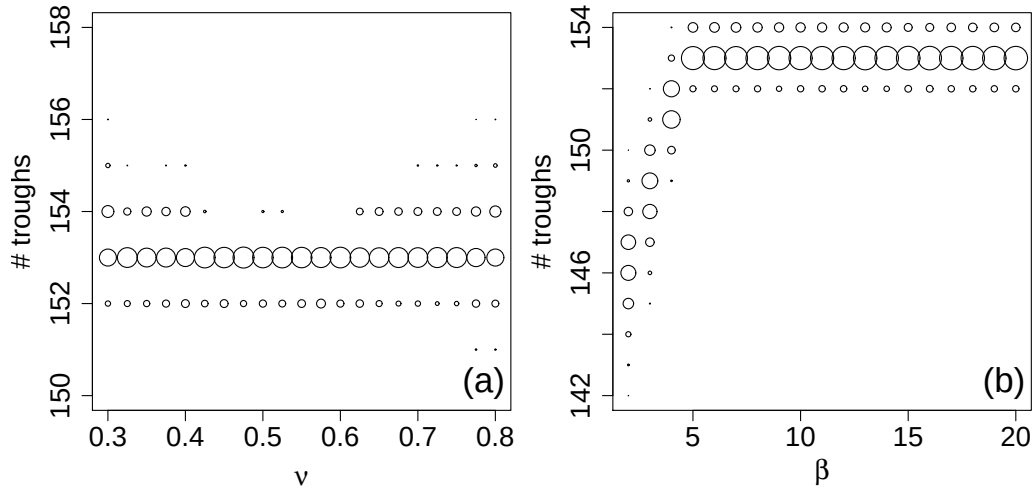

Figure 3.7: The effect of the number of sections β used in the initial smoothing of the H_2O_2 signal, and the threshold ν for classifying individual points, on the probability distribution obtained for the number of annual troughs – a summary of the chronology – in the dating of the Gomez core. Each circle has area proportional to the probability of a particular number of troughs, when the reconstruction uses particular values of β and ν. **(a)** $\nu = 0.3, \ldots, 0.8$ with $\beta = 6$. **(b)** $\beta = 2, \ldots, 20$ with $\nu = 1/\sqrt{2}$.

3.2.7 NGRIP: ammonium and calcium

In this section we analyse the standardised ammonium (NH_4) and calcium (Ca) chemistry signals from the NGRIP ice core between 1440m and 1465m. The standardisation process for these two signals is illustrated in Figures 2.13 and 2.11. We use the same methodology as on the H_2O_2 signal from the Gomez core. These signals have a slightly greater noise to annual cycle ratio than the Gomez H_2O_2, with regular fluctuations and stretches of missing values, and much higher variability in annual layer thicknesses.

To test the effect of sampling rate, we also run the analysis on three thinned down versions of both signals: taking every second point (2mm); every third point

(3mm); and every fourth point (4mm). We refer to these as the second, third and fourth thinnings. An alternative way to generate signals of lower sampling rate would be to take averages of *non-overlapping* intervals – rather than a moving average. We would expect our method to work better in that case as more information is retained.

There is very little trend in cycle length through this depth range in the NGRIP core. In each case we estimate the overall average cycle length from the ACF of the entire signal to use as the interval length when calculating $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, effectively setting $\beta = 1$. Figure 3.8 shows ACFs for both signals: (a) no thinning (estimated average cycle length 62 points); (b) second thinning (32 points); (c) third (21 points); and (d) fourth (16 points). In each case the ACF for NH_4 is shown as circles and for Ca as stars.

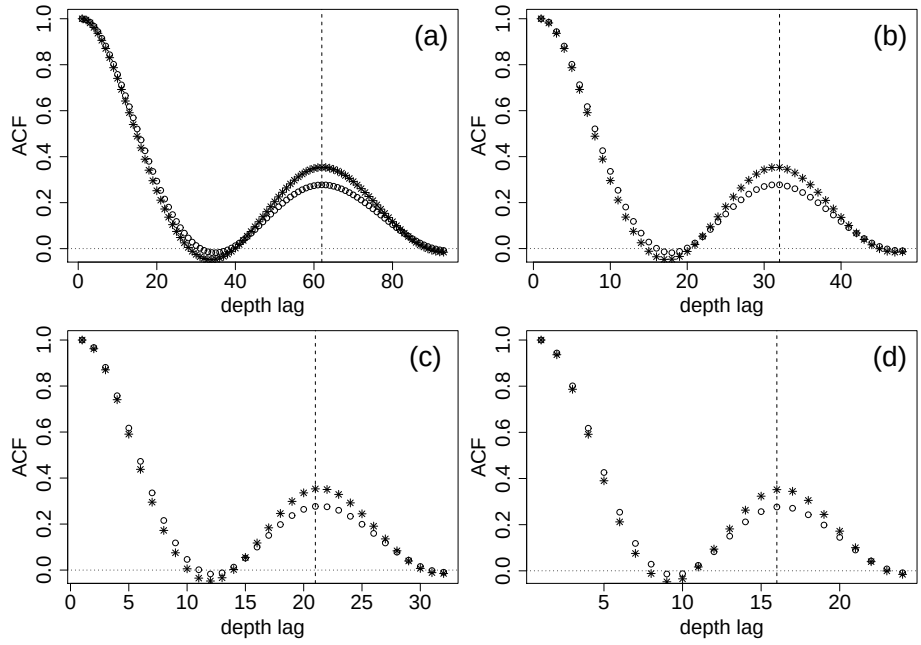

Figure 3.8: ACFs of the log NH_4 (circles) and log Ca (stars) signals from the NGRIP ice core (1440m–1465m), plotted against lag. The vertical dashed lines indicate the local maxima, used to obtain an initial estimate of cycle length. **(a)** No thinning. **(b)** Second thinning. **(c)** Third thinning. **(d)** Fourth thinning.

Figure 3.9 shows a stretch of the classified log NH_4 signal (fourth thinning) for $\nu = 0.5$ with 9 cycles: (a) is the log signal with $\boldsymbol{\mu}$ and $\boldsymbol{\mu} \pm \boldsymbol{\sigma}$ shown as dotted lines; and (b) is the standardised signal. Here there are two issues; the first is

caused by a single missing value and the second by a fluctuation in the data (a 'double peak'). Note that there are three ascending runs of length one – the reason why we could not do a fifth thinning – an average of around 16 points per cycle appears to be the limit for the classification method with this cycle shape and length distribution.

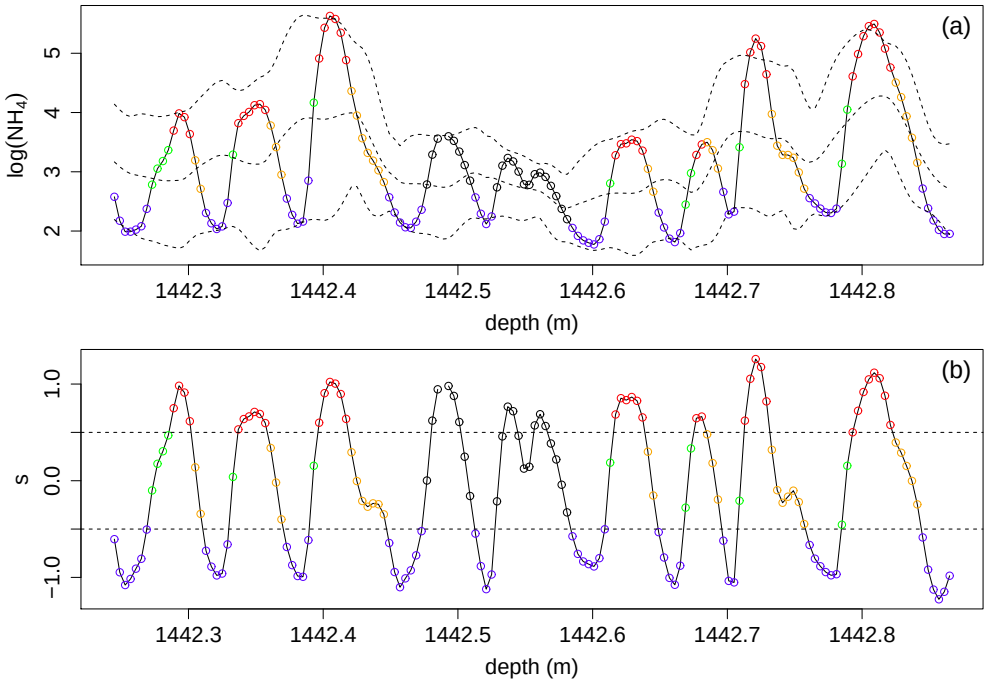

Figure 3.9: A stretch of the classified log NH_4 signal (fourth thinning) from the NGRIP ice core (1442.2m–1442.8m), for $\nu = 0.5$. Points within an issue are black, points within peaks are coloured red, descending points are orange, troughs are blue, and ascending points are green. **(a)** The log signal with $\boldsymbol{\mu}$ and $\boldsymbol{\mu} \pm \boldsymbol{\sigma}$ shown as dotted lines. **(b)** The standardised signal $\boldsymbol{s}$, with $\pm\nu$ shown as dotted lines.

Figure 3.10 shows a stretch of the classified Ca signal (second thinning) for $\nu = 0.5$ with seven cycles. There is one issue – a probability of 10% is assigned to there being two troughs in this section.

The probability distributions for the resulting chronologies for the second and fourth thinning of both signals are summarised in Figure 3.11 over a range of 8 values for ν. The no-thinning and third thinning cases are similar. The range of ν was chosen in each case as the interval over which the count is 'most stable' – in that the probability distributions are most similar. In each case a cursory

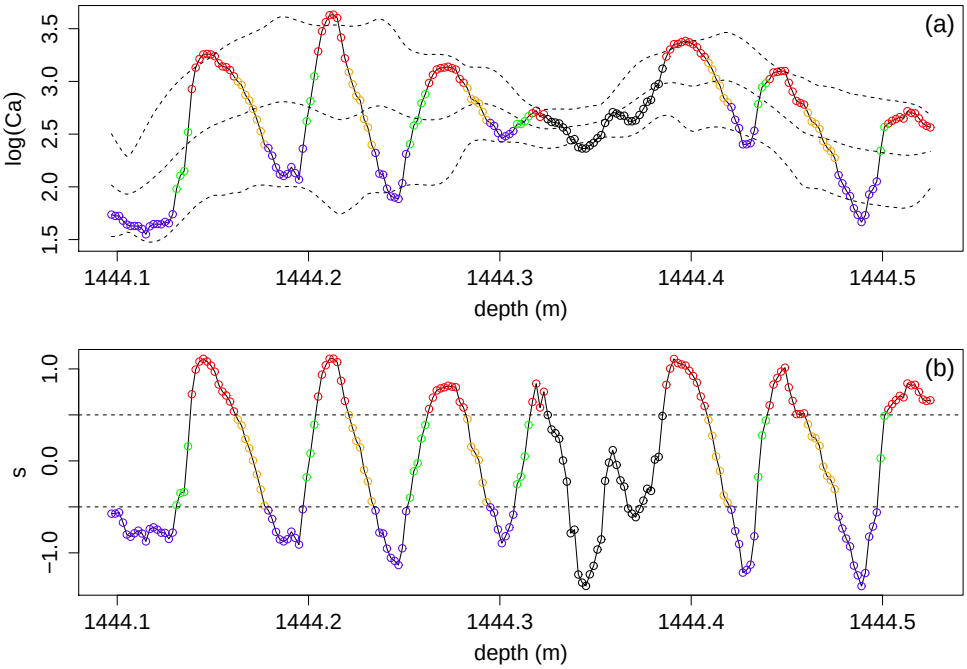

Figure 3.10: A stretch of the classified log Ca signal (second thinning) from the NGRIP ice core (1444.1m–1444.53m), for $\nu = 0.5$. Points within an issue are black, points within peaks are coloured red, descending points are orange, troughs are blue, and ascending points are green. **(a)** The log signal with $\boldsymbol{\mu}$ and $\boldsymbol{\mu} \pm \boldsymbol{\sigma}$ shown as dotted lines. **(b)** The standardised signal $\boldsymbol{s}$, with $\pm\nu$ shown as dotted lines.

check on the model's choice of 'certain' runs, and the probabilities assigned to resulting issues, was made to confirm that they are sensible. These ranges were chosen by eye and vary for each thinning; for the most part this is due to the effect of the sampling rate on the visibility of the fluctuations and identification of annual cycles. Choices of ν below these intervals generally overestimate the count – classifying fluctuations as 'certain' cycles. Choices of ν above these intervals generally underestimate the count – missing out whole cycles in the classification of 'certain' runs. One way to stabilise the 'certain' cycle count would be to do one run of the classification process, model the 'certain' run lengths, check for outliers in the distribution (abnormally short or long runs), and assign these as issues. Note that the Ca count is generally higher than the NH_4 count and has a greater uncertainty; this is due to a number of extra potential annual cycles present in the Ca signal when compared to the NH_4 signal, suggesting that a bivariate

implementation, as outlined in section 3.4, would be beneficial. However, in multivariate datasets the different nominal resolution of individual components must also be considered.

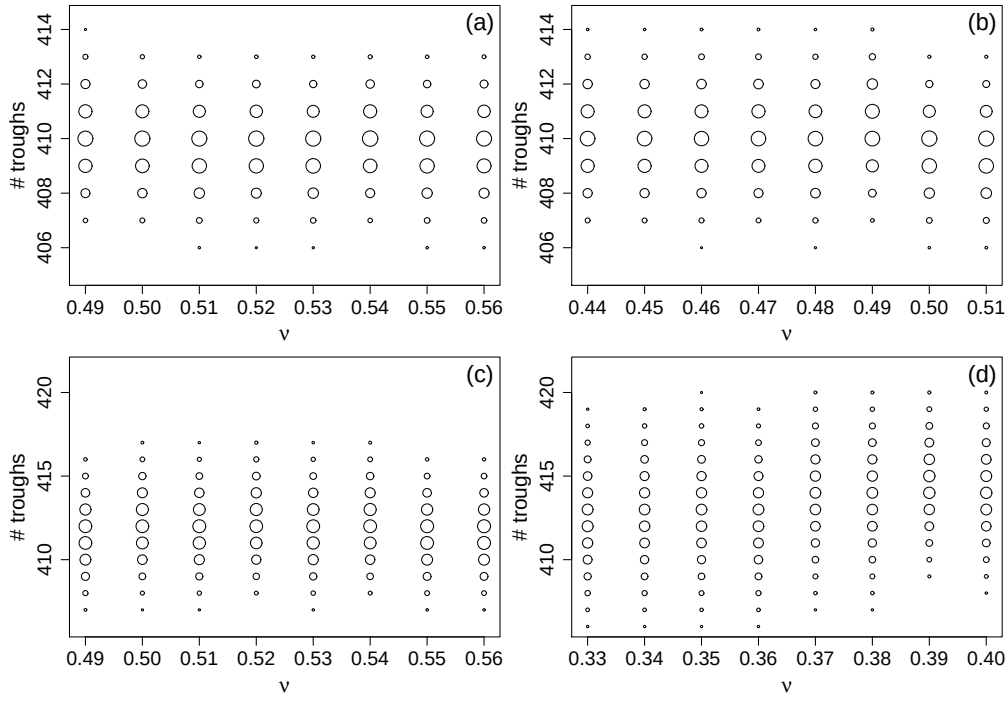

Figure 3.11: The effect of the classification parameter ν on the probability distribution obtained for the number of annual troughs in the dating of the NGRIP ice core (1440m–1465m) using its log NH_4 and Ca signals. Each circle has area proportional to the probability of a particular number of troughs. **(a)** NH_4, second thinning. **(b)** NH_4, fourth thinning. **(c)** Ca, second thinning. **(d)** Ca, fourth thinning.

The GICC05 chronology (Rasmussen et al, 2006) for the NGRIP core is published as the depths that mark the end points of 20 year intervals, making it unsuitable for direct comparison of individual layer markings. It assigns 420 certain years, with a maximum counting error (MCE) of 3, between 1439.92m and 1465.52m; and 400 certain years, with a MCE of 3, in the interval 1441.16m to 1464.37m,. This linearly interpolates to 409.1 ± 3 years between 1440m and 1465m, but this feasibly could be 408 ± 3 or 410 ± 3. This is consistent with our results.

For comparison, after correcting for the slight decreasing trend in the 'trough to trough' log cycle lengths from the most probable NH_4 chronology and modelling

them as Gaussian, the probability of a random annual layer being either double or half the mean thickness at any given depth is approximately 2.3%.

3.2.8 Gomez: non-sea-salt sulphur

To show how our classification process can be adapted for signals with asymmetrical annual cycles we now date the Gomez ice core using its non-sea-salt sulphur (nss-S) signal. This signal is challenging because the shape of its annual cycles changes with depth – from wide noisy troughs at the top of the core to cycles similar to those of the example H_2O_2, NH_4 and Ca signals at the bottom. A logarithmic transformation improves symmetry at the bottom of the core but not at the top, so the nss-S signal is not transformed for this analysis. Instead, we make a minor adaptation to the method.

The standardisation method presented above is parametric - a local mean and standard deviation is estimated for each depth; points which exceed a given number ($\nu\sqrt{2}$) of standard deviations from the mean are then classified as *potential* peaks and troughs. A more robust method of classifications is required here as symmetry cannot be induced throughout the nss-S signal – we use local quantiles (or percentiles). Along interval lengths estimated from the ACF of the nss-S, as above, we calculate the local ν_1th and ν_2th quantiles at each depth. Data points above the ν_1th quantile are then classified as potential peaks, and data points below the ν_2th quantile are classified as potential troughs. From this point on the method continues as in previous sections.

In the regression model for the Gomez nss-S, log 'certain' run lengths of the A and D ('central') classifications are found to be equivalent, whereas P and T classifications show a statistically significant difference ($p < 1\%$). There is also a statistically significant interaction between the depth index and the classification factor ($p < 1\%$). This fits a steeper gradient to the trough log lengths relative to the other classifications – modelling the change in cycle shape down the core.

Figure 3.12 shows two sections of classified nss-S signal, both with seven cycles:

(a) at the start of the core, and (b) towards the end of the core.

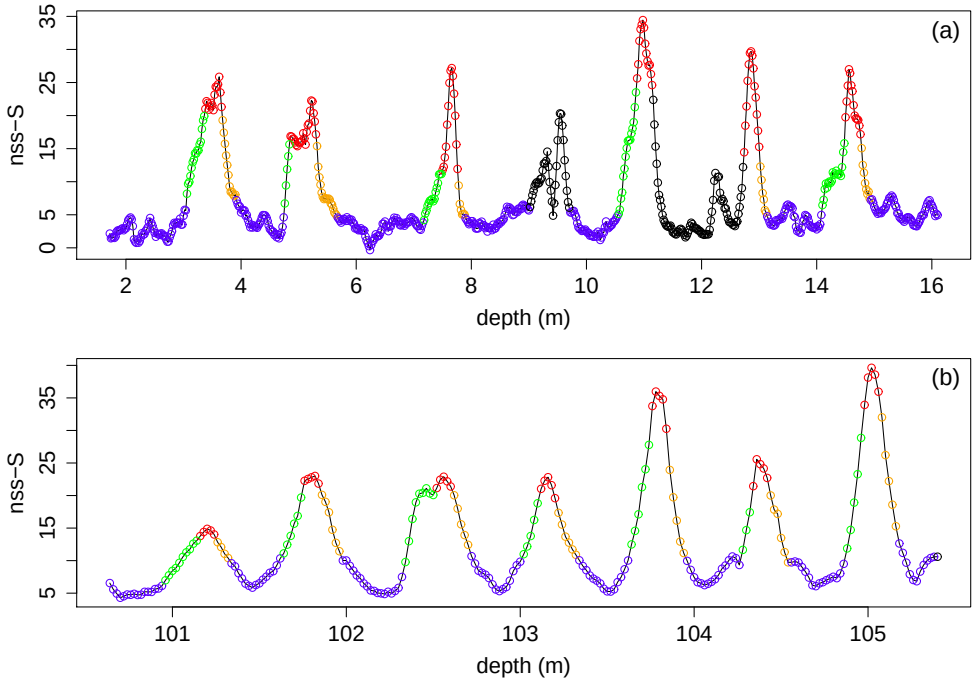

Figure 3.12: Two sections of the log nss-S signal from the Gomez ice core, classified using a non-parametric approach with $\nu_1 = 0.85$, $\nu_2 = 0.5$, and $\beta = 10$. Points within an issue are black, points within peaks are coloured red, descending points are orange, troughs are blue, and ascending points are green. **(a)** At the start of the core. **(b)** Towards the end of the core.

Figure 3.13 shows the resulting probability distribution for the chronology at $\nu_1 = 0.85$ and $\nu_2 = 0.5$. This has more uncertainty than that found from the H_2O_2 signal. The most likely chronology ($p = 0.4$) has a one-to-one trough correspondence with the most likely reconstruction found from the H_2O_2 signal and the manual count. This is true over the range $0.8 \leq v_1 \leq 0.9$ and $0.4 \leq v_2 \leq (v_1 - 0.3)$.

The log H_2O_2 signal from the Gomez ice core has symmetrical cycles and the choice of $\nu = 1/\sqrt{2}$ splits each cycle into four equal parts with equivalent length distributions. These parts are analogous to seasons as the annual cycles are directly correlated to sunlight, and allow other variables to be discussed in seasonal terms.

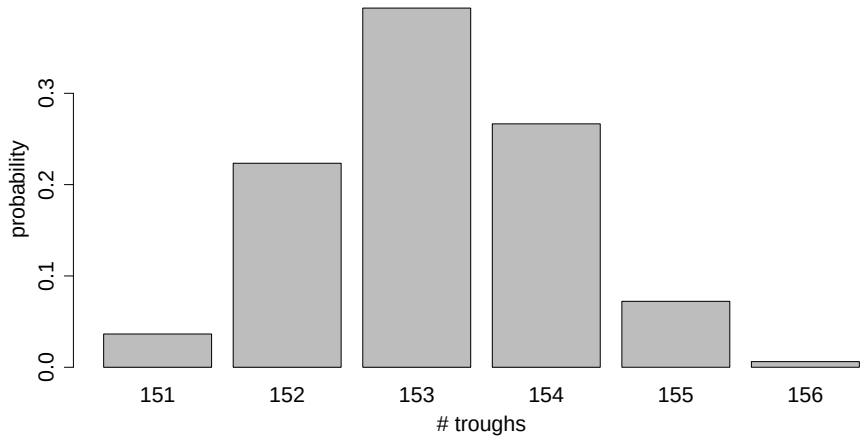

Figure 3.13: Bar plot showing the probability distribution of the number of annual troughs, a summary of the chronology, in the dating of the Gomez core using its nss-S signal for $\nu_1 = 0.85$, $\nu_2 = 0.5$, and $\beta = 10$.

3.2.9 Unequally spaced data: Fletcher δ^{18}O

Here we date the unequally spaced Fletcher δ^{18}O signal. The standardisation process for this signal is described in Section 2.3.

The univariate classification algorithm developed for equally spaced data, described above, finds 18 issues and 135 certain trough runs when applied directly to the unequally spaced δ^{18}O signal. Figure 3.14 (a) shows certain run length (cm) against depth (m) for this signal with $\nu = 1/2$. At this resolution some of the certain runs contain only one data point, causing the regression modelling to fail. To overcome this, we oversample the signal onto an equally spaced finer resolution depth scale. Figure 3.14 (b) shows the result when a depth increment of 1cm is chosen.

The method of calculating uncertainty measures for reconstructions could be adapted for use on unequally spaced data by using differences in depths instead of counts of points for the issue lengths. However, as this signal is now on an equally spaced depth scale, we can use the same approach as before. Figure 3.15 shows the effect of ν and β on the resulting chronology distribution. This process is quite sensitive to the value of β used – for a final dating of this core I would recommend

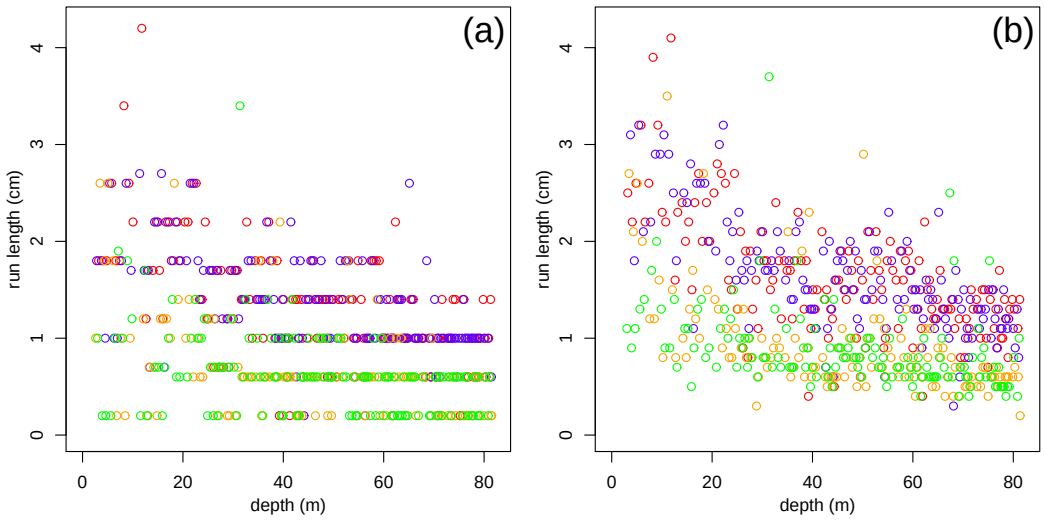

Figure 3.14: Certain run length (cm) against depth (m) from the classification of the Fletcher δ^{18}O signal (2.48–80.60m) using $\nu = 1/2$. (a) The raw unequally spaced signal. (b) The interpolated equally spaced (1cm) signal.

that an expert look over and assess all certain years and reconstructions.

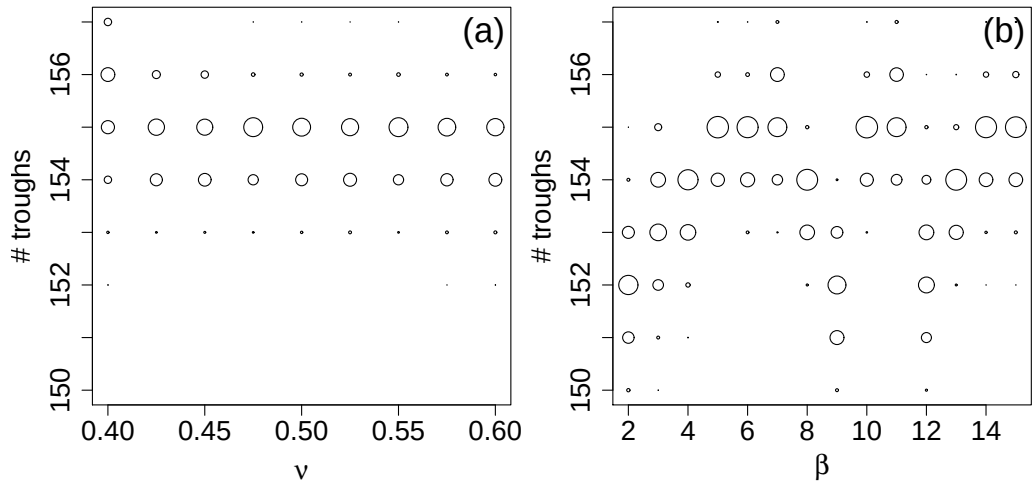

Figure 3.15: The effect of the number of sections β – used in the initial smoothing – and the threshold ν – for classifying individual points – on the probability distribution obtained for the number of annual troughs – a summary of the chronology – in the dating of the Fletcher δ^{18}O signal (2.48m–80.60m). Each circle has area proportional to the probability of a particular number of troughs, when the reconstruction uses particular values of β and ν. **(a)** $\nu = 0.4, \ldots, 0.6$ with $\beta = 6$. **(b)** $\beta = 2, \ldots, 15$ with $\nu = 1/2$.

3.3 Multivariate classification

Here we discuss some possibilities for extending the classification method to a multivariate framework – using multiple signals to group depths into certain runs and issues described by length and label. Once this has been achieved, the same regression based method of assigning probabilities to issue reconstructions as described above can be used. The work presented in this section is not a fully developed methodology, it is intended only to demonstrate the potential of the approach to be used on multiple signals. A fully Bayesian multivariate MCMC based approach is presented in Chapter 6.

Annually cyclic chemistry data recorded from Greenland ice cores tend to have seasonality which is in-phase during cold periods and out-of-phase during warm periods (Andersen et al, 2006; Rasmussen et al, 2006) and different algorithms are required in each case. For a fully automated method we would want to detect and estimate the phase difference between signals to determine which method to use. One possible way to accomplish this is discussed in Section 3.3.1.

3.3.1 Detecting and estimating phase differences

Once a signal has been standardised it is (approximately) a noisy sine wave on a non-linear timescale. Standardised signals from the same ice core, and on the same depth range, can be thought of as sharing some common underlying timescale – shifted by some phase difference. If $\boldsymbol{s}_k$ and $\boldsymbol{s}_l$ are two such signals, with constant phases $\psi_k, \psi_l \in (-\pi, \pi)$, this can be written as:

$$\boldsymbol{s}_k \approx \sin(2\pi\boldsymbol{\tau} + \psi_k) \text{ and } \boldsymbol{s}_l \approx \sin(2\pi\boldsymbol{\tau} + \psi_l).$$

Applying the angle sum trigonometric identity to $\boldsymbol{s}_k$ gives

$$\boldsymbol{s}_k \approx \sin(2\pi\boldsymbol{\tau} + \psi_k) = \cos(\psi_k)\sin(2\pi\boldsymbol{\tau}) + \sin(\psi_k)\cos(2\pi\boldsymbol{\tau})$$

for some unequally spaced timescale $\boldsymbol{\tau}$. Over a long stretch of signal, with *exactly* m cycles, we would expect that $\sum_i \sin(2\pi\tau_{k,i})/n \approx 0$, the peaks cancelling with the troughs and any remainder small when compared to n. Similarly, if we have m full cycles and two part cycles either side then we would expect the sum of the points in the part cycles to be small compared to n, giving

$$E(\boldsymbol{s}_k) \approx 0$$

for large enough n. The variance of a sine wave with *exactly* m cycles and equally spaced observations is 1/2. Under the assumption that the distribution of $\boldsymbol{\tau}$ is the same in peaks and troughs, for a large enough number of cycles and number of points per cycle we would expect the variance of $\sin(2\pi\boldsymbol{\tau})$ to be approximately 1/2 also. Thus we have

$$var(\boldsymbol{s}_k) \approx 1/2.$$

Then we can write

$$E(\boldsymbol{s}_k^2) = var(\boldsymbol{s}_k) + E(\boldsymbol{s}_k)^2 \approx 1/2.$$

$$E(\boldsymbol{s}_k \boldsymbol{s}_l) \approx E\left[(\cos(\psi_k)\sin(2\pi\boldsymbol{\tau}) + \sin(\psi_k)\cos(2\pi\boldsymbol{\tau})) \times \right.$$

$$\left.(\cos(\psi_l)\sin(2\pi\boldsymbol{\tau}) + \sin(\psi_l)\cos(2\pi\boldsymbol{\tau}))\right]$$

$$= \cos(\psi_k)\cos(\psi_l)E(\sin^2(2\pi\boldsymbol{\tau})) + \cos(\psi_k)\sin(\psi_l)E(\sin(\tau)\cos(\tau)) +$$

$$\sin(\psi_k)\cos(\psi_l)E(\cos(\tau)\sin(\tau)) + \sin(\psi_k)\sin(\psi_l)E(\cos^2(\tau))$$

$$= \frac{\cos(\psi_k)\cos(\psi_l) + \sin(\psi_k)\sin(\psi_l)}{2} = \frac{\cos(\psi_k - \psi_l)}{2}$$

as $2\sin(\boldsymbol{\tau})\cos(\boldsymbol{\tau}) = \sin(2\boldsymbol{\tau}) + \sin(0)$. The absolute phase difference $|\psi_k - \psi_l|$ can

be approximated by:

$$\cos^{-1}(2cov(\boldsymbol{s}_k, \boldsymbol{s}_l)).$$

It is essentially the same, and computationally safer, to use $\cos^{-1}(cor(\boldsymbol{s}_k, \boldsymbol{s}_l))$ which is robust to the case: $|cov(\boldsymbol{s}_k, \boldsymbol{s}_l)| > 1/2$. The estimated absolute phase differences (divided by 2π) between the standardised 1440m–1465m NGRIP annually cyclic signals are shown in table 3.1 below. This suggest that they are mostly out-of-phase, as we would expect during the (warm) Holocene period (Rasmussen et al, 2006).

Table 3.1: Estimated absolute phase differences, NGRIP (1440m–1465m)

signal	NH_4	Ca	NO_3	Na	Dust
NH_4	0.000	0.279	0.113	0.295	0.250
Ca	0.279	0.000	0.296	0.243	0.091
NO_3	0.113	0.296	0.000	0.287	0.268
Na	0.295	0.243	0.287	0.000	0.248
Dust	0.250	0.091	0.268	0.248	0.000

3.3.2 In-phase signals

In-phase signals (those with an estimated phase difference of approximately zero) could be assessed independently using the univariate classification method (Section 3.2). The resulting run patterns could then be compared, with 'multivariate issues' defined as: the union of the individual 'univariate issues', and where the certain run classifications disagree between the signals.

3.3.3 Out-of-phase signals

In the univariate example we initially split the signal of interest into three types of runs along the real line: high, medium and low. In a similar way, pairs of data points of the same depth index from two out-of-phase standardised signals, $\boldsymbol{x}'$ and $\boldsymbol{y}'$, can be split into labelled runs on the bivariate plane. Plotting $\boldsymbol{x}'$ against $\boldsymbol{y}'$ helps to visualise this idea. These runs can than be classified as 'certain' or 'issues' in ways analogous to those discussed in Section 3.2.2.

A natural way to achieve this is to collect the points into quadrants:

Q_1: $x' > 0$ and $y' > 0$

Q_2: $x' < 0$ and $y' > 0$

Q_3: $x' < 0$ and $y' < 0$

Q_4: $x' > 0$ and $y' < 0$.

These could be treated as potential runs as above, and classed as certain if they are central to an odd number, g, of consecutive runs that follow a

$$\ldots Q_1 \to Q_2 \to Q_3 \to Q_4 \to Q_1 \ldots$$

pattern. For the univariate method in Section 3.2 we used $g = 5$ throughout. The method of defining possible reconstructions and assigning probabilities using regression modelling would be the same as the univariate version from here.

Figure 3.16 (top row) illustrates this quadrant method applied to the 1440m–1465m NGRIP ammonium and calcium signals with $g = 5$. The left hand plot shows the standardised NH_4 ($\boldsymbol{x}'$) vs. the standardised Ca ($\boldsymbol{y}'$). Points in Q_1 are shown in red, points in Q_2 are shown in blue, points in Q_3 are shown in green and those in Q_4 are shown in orange. The middle plot shows the corresponding certain run lengths against depth and the right hand plot is a histogram of the certain run lengths. These show a good fit to the $G(4.8, 3.1)$ distribution without transformation – probabilities for the issue reconstruction lengths could be calculated via a scheme based on this distribution.

Similar bivariate classification methods can be defined by splitting the bivariate plane into any number of sections. Another possibility with four sections – which aims to split the signals into seasons as in the univariate method – is shown on the middle row of Figure 3.16 with $g = 5$. This 'bivariate season' classification is described as:

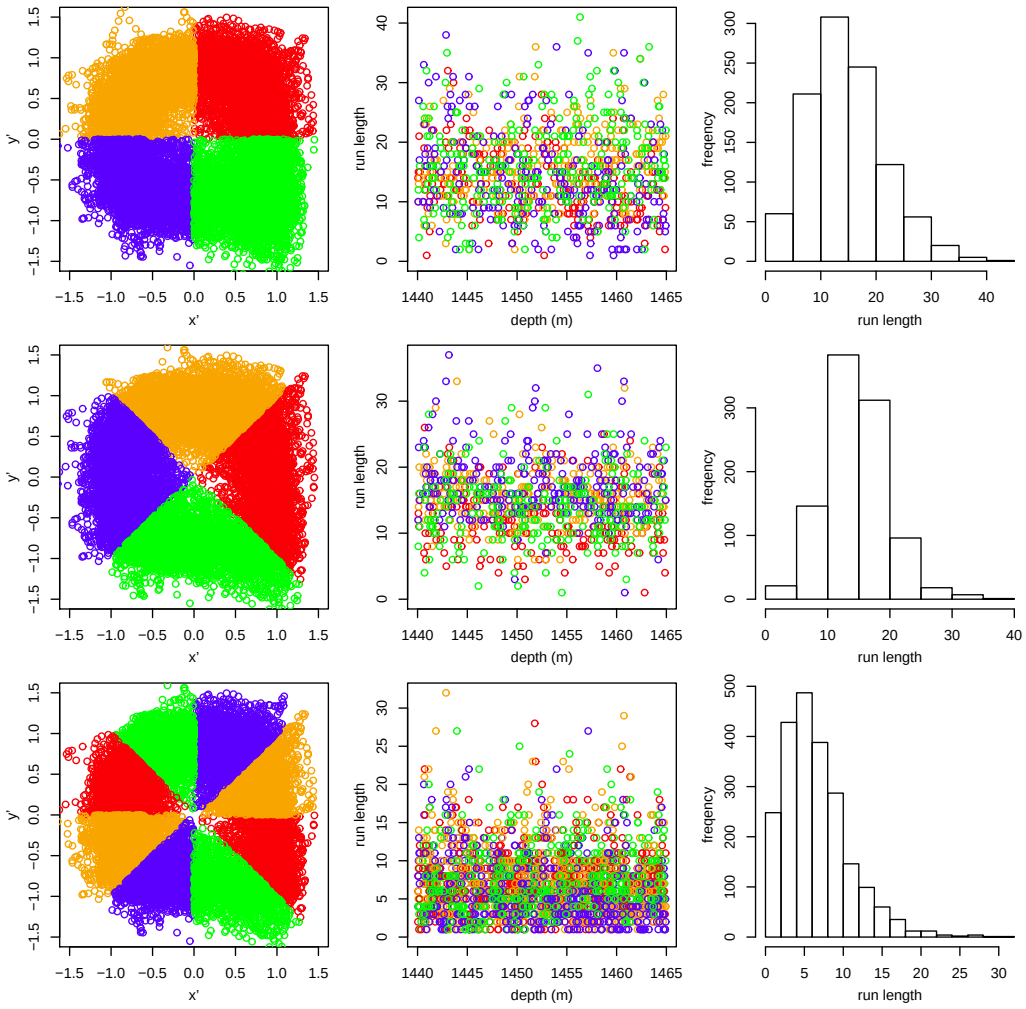

Figure 3.16: Three methods of bivariate classification – splitting pairs of data points of the same depth index from two out-of-phase standardised signals: NGRIP NH_4 ($\boldsymbol{x}'$) vs Ca ($\boldsymbol{y}'$) into runs (1440m–1465m). The left hand plots show $\boldsymbol{x}'$ vs. $\boldsymbol{y}'$. The middle plots show run lengths against depth. The right hand plots are histograms of the run lengths. (top and middle) four run types. (bottom) eight run types.

S_1: $x' > 0$ and $|x'| > |y'|$

S_2: $y' > 0$ and $|x'| < |y'|$

S_3: $x' < 0$ and $|x'| > |y'|$

S_4: $y' < 0$ and $|x'| < |y'|$

In this case the certain run lengths follow a $\mathrm{G}(8.9, 1.7)$ distribution. The bottom

row of Figure 3.16 illustrates a further method which splits the bivariate plane into 8 sections using a mixture of the quadrant and seasonal methods. Here the certain run lengths follow a $G(2.8, 2.6)$ distribution.

Figure B.1, in the appendix, shows an example section of the standardised NH_4 and Ca signals from the NGRIP ice core (1452.54m–1453.10m) classified into runs using the bivariate quadrant based method. An issue can be seen in the centre caused by a cycle which is present in the ammonium signal but not the calcium.

Figure B.2 shows the same example section of the standardised NH_4 and Ca signals classified into runs using the bivariate seasons based method. Here the issue seen in Figure B.1 has not been picked up, however a new issue can be seen around depth 1452.7m caused by the cycle shapes.

Figure B.3 shows the same example section classified into runs using a mixture of the bivariate quadrant and seasonal based methods. Here both issues have been identified – suggesting that this approach may be preferable.

3.4 Conclusions

With regards to developing an automated method for determining the annual layer chronology in an ice core with a strong annually cyclic signal, in this chapter we have presented univariate methods to

- split a standardised signal into sections with a deterministic cycle count and those that need more attention,
- display possible reconstructions for the uncertain sections that could be used as visual aids for manual counting, and
- assign probability measures to each reconstruction based on its length and the classification of the bordering deterministic sections,

which together provide a stable count with an uncertainty measure on various signals from the Gomez, Fletcher and NGRIP ice cores. Some of the work carried

out for the examples would need to be repeated for each ice core: for example, to determine the best value of β for a particular case, or to assess the data for trends in layer thickness, which can arise for climatic as well as glaciological reasons. Our run classification process requires a signal with a strong seasonal component and would need to be adapted for use on noisier signals such as ECM or Visual Stratigraphy.

These methods could be adapted for use on much longer datasets, thereby reducing manual effort and providing a robust methodology. Future work to develop this method for broad application in physical science research, including but not limited to ice core palaeoclimate research, may involve extending the methodology to take in information from multivariate datasets with more uncertain annual cyclicity and being able to provide solutions for optimally fitting annual chronologies between fixed points of known age. We have provided a discussion of how the ideas introduced in this chapter could be extended for use on multiple signals simultaneously.

We attempt to show that this process is robust to the tuning parameters. Plots are provided to show the effect of changing the parameter ν on the resulting probability distributions for the cycle count. In practice, since this method does not aim to give definitive probabilities, it could be argued that the robustness of the classification method, and of the issues found, is of more importance. In all analyses, different issues arise when varying ν in the ranges presented; see Figure 3.11 for an example. However, in all cases where one value of ν has an issue over a depth range where another value of ν gives a 'certain' count, see Figure 3.1, the reconstruction which corresponds to the 'certain' runs always has very high probability and in most cases is assigned a probability of 1 after normalisation. Thus the key message about which parts can be confidently classified, and which are genuinely uncertain, is highly robust.

One new and useful aspect of this method is that it splits the signal into seasons as well as years. This has already been used by scientists at BAS to compare objectively determined summer and winter values (P and T runs) of the Gomez

H_2O_2 signal with time.

Chapter 4

Univariate MCMC - standardised signals

4.1 Introduction

In this chapter we present a model for annually cyclic ice core signals that have been standardised via the method introduced in Chapter 2. This model is a precursor to those used in Chapters 5 and 6 for raw or non-standardised signals. Here we introduce the continuous latent timescale $\boldsymbol{\tau}$ and its prior distribution, and discuss how it can be sampled using a fully Bayesian MCMC approach. Examples from the standardised NGRIP ammonium signal are used to help illustrate the model and the methods used to fit it to the data. It should be noted that whilst we do not recommend this method to be used directly for dating ice cores, some of the theory developed below is used and referenced in the algorithms of the further chapters.

4.2 The timescale: $\boldsymbol{\tau}$

$\boldsymbol{\tau}$ is an ordered vector of elements $\tau_i \in \mathbb{R}$ and is a model parameter. It describes the date in years at each depth index $i \in (1, \ldots, n)$ of an ice core. It can be

parametrised directly as the date that the ice formed (or snow fell) or as the elapsed time from the top of a core, or over a subset of depth. Its integer part, $\lfloor \tau_i \rfloor$, describes the year or the cycle count; its non-integer or decimal part, $\tau_i - \lfloor \tau_i \rfloor$, describes the time of year or distance through a cycle. The absolute difference between two elements of $\boldsymbol{\tau}$, $|\tau_i - \tau_j|$, gives the elapsed time between depth indexes j and i.

4.2.1 The model

Once standardised as described in Chapter 2, an annually cyclic ice core signal, $\boldsymbol{x}'$, can be thought of as a noisy sine wave on an unequally spaced or non-linear timescale: $\boldsymbol{\tau}$. With this in mind we model $\boldsymbol{x}'$ as

$$x'_i = \sin(2\pi\tau_i) + \epsilon_i, \ i = 1, \ldots n.$$

The hierarchical prior structure for $\boldsymbol{\tau}$ is discussed in Section 4.2.2 and the error structure for $\boldsymbol{\epsilon}$ in Section 4.2.3.

It should be noted that the use of a sine wave in the model equation is not essential. Any function $f()$ which is cyclic or periodic in the sense that $f(\tau) = f(\tau + 1) \ \forall \ \tau$ could be used:

$$x'_i = f(\tau_i) + \epsilon_i, \ i = 1, \ldots n.$$

The function $f()$ could have a number of other local or global parameters which could be fitted simultaneously under the MCMC framework discussed in Section 4.3. However, a sine wave, $f(\tau) = \sin(2\pi\tau)$, works well for all signals considered in this research.

4.2.2 Prior distribution for $\boldsymbol{\tau}$

The elapsed times over each depth increment are labelled $\delta_i = \tau_i - \tau_{i-1}$, $i = 2, \ldots, n$. For the NGRIP core these are equivalent to how long it snowed to

create 1mm of ice at a given depth. We assume that $\delta_i > 0 \; \forall \; i$, that is, that the ordering of the ice layers has not changed since it was deposited. We also assume that their distribution is positively skewed; we see this feature in the 'certain' run length distributions for all ice cores in Chapter 3 and in the log-normally distributed manually assigned cycle lengths in Andersen et al (2006). Under these assumptions it is natural to assign the continuous elements of $\boldsymbol{\delta}$ a Gamma prior distribution.

One convenient choice is to model the elements of $\boldsymbol{\delta}$ as independent Gamma distributed variables with global shape ψ and rate λ. The assumption of independence is only ever appropriate over short sections of signal where no systematic change in the accumulation rate, such as compression thinning or firnification, is observed. In the example NGRIP CFA chemistry data, with an average of 60 data points in an annual cycle, we observe within-cycle correlation in the accumulation rate. The distribution of $\boldsymbol{\delta}$ changes slowly and continuously with depth. This is caused by the mixing of melt water in the CFA apparatus and is therefore assumed to be correlated in depth. Such local correlation is not observed in data collected in discrete cuts (or bags), such as those signals from the Gomez and Fletcher ice cores.

In later chapters, with higher dimensional models, we thin the CFA sampled data so it is closer to its nominal resolution (1cm for the NGRIP core) to reduce autocorrelation. We then model the elements of $\boldsymbol{\delta}$ as being independent, with constant shape ψ. In this chapter we investigate a method of modelling local correlation in the timescale. We consider a more general case of the gamma prior where the elapsed times are independent given a local shape ψ_i that forms some correlated process $\boldsymbol{\psi}$. To achieve this in a way that is convenient for the MCMC updates we set $\boldsymbol{\psi}$ to be a Gaussian random walk in depth with constant variance,

$$\psi_i \sim \mathrm{N}(\psi_{i-1}, \sigma^2_\psi), \;\; i = 3, \ldots, n.$$

From here on these are referred to as the 'constant shape, ψ' case and 'correlated

shape, $\boldsymbol{\psi}$' case. The theory of the former is a clear simplification of the latter. Both cases are discussed in Section 4.3 which explains how we fit the model using MCMC. In the correlated $\boldsymbol{\psi}$ case we have

$$\delta_i \sim \mathrm{G}(\psi_i, \lambda), \quad f(\delta_i) = \frac{\lambda^{\psi_i}\delta_i^{\psi_i-1}e^{-\lambda\delta_i}}{\Gamma(\psi_i)} \propto \delta_i^{\psi_i-1}e^{-\lambda\delta_i}.$$

The δs are independent given $\boldsymbol{\psi}$ and so have joint distribution function:

$$f(\boldsymbol{\delta}) = \prod_i f(\delta_i) \propto \prod_i \delta_i^{\psi_i-1}e^{-\lambda\delta_i} = e^{-\lambda\sum_i \delta_i}\prod_i \delta_i^{\psi_i-1}.$$

Conditioned on their sum, any subset of $\boldsymbol{\delta}$ follow a Dirichlet distribution:

$$f(\boldsymbol{\delta}|\Sigma_i\delta_i) \propto \prod_i \delta_i^{\psi_i-1}.$$

Although $\boldsymbol{\psi}$ is usually constrained by the signal, it has to behave appropriately in sections of missing data; we discuss a method of eliciting the rate (or scale) parameter λ for this prior in the second example of this chapter (Section 4.6.2).

4.2.3 Error structure

It is clear from looking at the standardised signals in Chapters 2 and 3 that their smooth deviations from a sine wave, the residuals $\epsilon_i = x_i' - \sin(2\pi\tau_i)$, are correlated. Furthermore it is assumed that the error is a result of the standardisation process which suggests that $\boldsymbol{\epsilon}$ is correlated in time not depth. We therefore want to model this correlation with respect to the unequally spaced in timescale $\boldsymbol{\tau}$ and not the depth index. To achieve this we use a zero-mean Ornstein-Uhlenbeck or OU process, the continuous time analogue of the AR(1) autoregressive process. The OU process was originally used to describe the velocity of a Brownian particle under the influence of friction (Uhlenbeck and Ornstein, 1930). It is Markovian, stationary, Gaussian, mean reverting and defined by the stochastic differential equation:

$$d\epsilon_i = \eta(\mu - \epsilon_i)d\tau + \sigma dW_i$$

where $\eta > 0$ is a correlation parameter which defines how strongly the system corrects perturbations. $\mu \in \mathbb{R}$ is the mean of the system; we set $\mu = 0$. $\sigma > 0$ measures the variation or noise level and W_τ is a Wiener process on the continuous timescale parameter $\boldsymbol{\tau}$:

$$W_0 = 0, \quad W_i - W_j \sim N(0,\ \tau_i - \tau_j).$$

Asymptotically $\epsilon_i \sim N(0, \sigma^2)$. This process has solution

$$\epsilon_i \mid \epsilon_{i-1} \sim N\left(\mu + (\epsilon_{i-1} - \mu)e^{-\eta\delta_i}, \sigma^2(1 - e^{-2\eta\delta_i})\right),$$

and since our process is zero-mean we have

$$\epsilon_i \mid \epsilon_{i-1}, \delta_i \sim N\left(\epsilon_{i-1}e^{-\eta\delta_i}, \sigma^2\left(1 - e^{-2\eta\delta_i}\right)\right)$$

$$\epsilon_i - \epsilon_{i-1}e^{-\eta\delta_i} \sim N\left(0, \sigma^2\left(1 - e^{-2\eta\delta_i}\right)\right)$$

$$\frac{\epsilon_i - \epsilon_{i-1}e^{-\eta\delta_i}}{\sigma\sqrt{1 - e^{-2\eta\delta_i}}} \sim N(0, 1),$$

$i = 2, \ldots, n$. The likelihood for our model comes directly from this last equation:

$$f(\boldsymbol{x}'|\boldsymbol{\tau}) = \prod_{i=2}^{n} \varphi\left(\frac{\epsilon_i - \epsilon_{i-1}e^{-\eta\delta_i}}{\sigma\sqrt{1 - e^{-2\eta\delta_i}}}\right)$$

where $\varphi()$ is the density of the standard normal distribution.

4.3 Fitting $\boldsymbol{\tau}$ via MCMC

We use an MCMC scheme to sample from the posterior distribution of $\boldsymbol{\tau}$, the distribution of the chronology given the standardised signal and prior assumptions. The algorithms for single component updates and block updates are discussed in Sections 4.3.1 and 4.3.2 respectively. The methods for proposing $\boldsymbol{\tau}$, and in the correlated shape case $\boldsymbol{\psi}$, presented here are used and referred to in later chapters. The updates discussed in this section only affect the time of year part

of the timescale; updates that affect the cycle (or year) count are discussed in Section 4.3.3.

4.3.1 Updating τ_i – one at a time

We first discuss how $\boldsymbol{\tau}$ can be updated one element at a time via a Metropolis-Hastings scheme, generally referred to as single component updates or SCU. Each new τ_i is proposed from its prior distribution conditional on all the other elements of $\boldsymbol{\tau}$, denoted $\boldsymbol{\tau}_{-i}$. There are two cases to consider: constant ψ and correlated $\boldsymbol{\psi}$. In each case prior and proposal ratios cancel, leaving the likelihood ratio as the acceptance probability (or more precisely as the Hastings ratio: see Section A.2 in the appendix).

Proposing τ_i with constant shape, ψ

τ_i' is proposed from the independent Gamma prior on its increments, conditional on $\boldsymbol{\tau}_{-i}$. $\boldsymbol{\tau}$ is Markovian, and so this is equivalent to conditioning on τ_{i-1} and τ_{i+1}. Recall that $\delta_i = \tau_i - \tau_{i-1}$ and $\delta_i \sim \mathrm{G}(\psi, \lambda),\ i = 2, \ldots, n$. Defining $\Delta = \delta_i + \delta_{i+1} = \tau_{i+1} - \tau_{i-1}$, we have $\Delta \sim \mathrm{G}(2\psi, \lambda)$, and so

$$\frac{\delta_i}{\Delta} = \frac{\tau_i - \tau_{i-1}}{\tau_{i+1} - \tau_{i-1}} \sim \mathrm{Beta}(\psi, \psi).$$

We can therefore propose $\tau_i',\ i = 2, \ldots, n-1$ from their conditional prior by drawing a perturbation $u \sim \mathrm{Beta}(\psi, \psi)$ and setting

$$\tau_i' = \tau_{i-1} + u(\tau_{i+1} - \tau_{i-1}).$$

Note that for the cases $i = 1$ and $i = n$ we draw $u \sim \mathrm{G}(\psi, \lambda)$ and set either: $\tau_1' = \tau_2 - u$ or $\tau_n' = \tau_{n-1} + u$ respectively.

Proposing τ_i and ψ_i jointly with correlated shape, $\boldsymbol{\psi}$

In this case we have $\delta_i \sim \mathrm{G}(\psi_i, \lambda)$ where $\psi_i \sim \mathrm{N}(\psi_{i-1}, \sigma_\psi^2)$, $i = 3, \ldots, n$. Let $\Delta = \delta_i + \delta_{i+1} = \tau_{i+1} - \tau_{i-1}$ as before. Now $\Delta \sim \mathrm{G}(\psi_i + \psi_{i+1}, \lambda)$ giving:

$$\frac{\delta_i}{\Delta} = \frac{\tau_i - \tau_{i-1}}{\tau_{i+1} - \tau_{i-1}} \sim \mathrm{Beta}(\psi_i, \psi_{i+1}).$$

For $2 < i < n-1$, updating τ_i conditional on τ_{i-1} and τ_{i+1} is equivalent to updating δ_i and δ_{i+1} conditional on their sum, Δ. For efficiency we update $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ together, in one step. We first propose ψ_i' and ψ_{i+1}' from their prior distribution conditioned on ψ_{i-1} and ψ_{i+2}:

$$\begin{pmatrix} \psi_i' \\ \psi_{i+1}' \end{pmatrix} \sim \mathrm{N}\left(\boldsymbol{\mu}_\psi, \Sigma_\psi\right)$$

$$\text{where } \boldsymbol{\mu}_\psi = \begin{pmatrix} \psi_{i-1} + (\psi_{i+2} - \psi_{i-1})/3 \\ \psi_{i-1} + 2(\psi_{i+2} - \psi_{i-1})/3 \end{pmatrix}$$

$$\text{and } \Sigma_\psi = \sigma_\psi^2 \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}.$$

We then propose τ_i from its conditional prior by drawing a perturbation $u \sim \mathrm{Beta}(\psi_i', \psi_{i+1}')$ and setting $\tau_i' = \tau_{i-1} + u(\tau_{i+1} - \tau_{i-1})$ as before. There are four special cases at the ends:

$i = 1$ We draw $\psi_2' \sim \mathrm{N}(\psi_3, \sigma_\psi^2)$ and $\tau_1' = \tau_2 - u$ where $u \sim \mathrm{G}(\psi_2', \lambda)$.

$i = 2$ We draw $\psi_3' \sim \mathrm{N}(\psi_4, \sigma_\psi^2)$ and $\psi_2' \sim \mathrm{N}(\psi_3', \sigma_\psi^2)$; which is equivalent to

$$\begin{pmatrix} \psi_2' \\ \psi_3' \end{pmatrix} \sim \mathrm{N}\left(\boldsymbol{\mu}_\psi, \Sigma_\psi\right)$$

$$\text{where } \boldsymbol{\mu}_\psi = \begin{pmatrix} \psi_4 \\ \psi_4 \end{pmatrix} \text{ and } \Sigma_\psi = \sigma_\psi^2 \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

We then set $\tau'_2 = \tau_1 + u(\tau_3 - \tau_1)$ where $u \sim \text{Beta}(\psi'_2, \psi'_3)$.

$i = n-1$ We draw $\psi'_{n-1} \sim \text{N}(\psi_{n-2}, \sigma^2_\psi)$ and $\psi'_n \sim \text{N}(\psi'_{n-1}, \sigma^2_\psi)$; equivalent to

$$\begin{pmatrix} \psi'_{n-1} \\ \psi'_n \end{pmatrix} \sim \text{N}\left(\boldsymbol{\mu}_\psi, \Sigma_\psi\right)$$

$$\text{where } \boldsymbol{\mu}_\psi = \begin{pmatrix} \psi_{n-2} \\ \psi_{n-2} \end{pmatrix} \text{ and } \Sigma_\psi = \sigma^2_\psi \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

We then set $\tau'_{n-1} = \tau_{n-2} + u(\tau_n - \tau_{n-2})$ where $u \sim \text{Beta}(\psi'_{n-1}, \psi'_n)$.

$i = n$ We draw $\psi'_n \sim \text{N}(\psi_{n-1}, \sigma^2_\psi)$ and $\tau'_n = \tau_{n-1} + u$ where $u \sim \text{G}(\psi'_n, \lambda)$.

4.3.2 Updating $\boldsymbol{\tau}$ using block updates

It is more efficient to update $\boldsymbol{\tau}$ in blocks than to use single component updates. Furthermore, it is necessary if we want to change the cycle count (see Section 4.3.3). Here we explain how to update $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ jointly for the correlated shape case. The simplification to the constant shape, ψ, case is discussed where relevant. As in the single component update case, the proposal and prior ratios cancel leaving the likelihood ratio as the acceptance probability.

Choosing the blocks

We first need to define the block, or interval, of $\boldsymbol{\tau}$ that we are going to update. This is denoted by a subscript I and is defined by depth indexes $s, f \in \mathbb{Z}$. The subsection of $\boldsymbol{\tau}$ that we are going to update, $\boldsymbol{\tau}_I = \{\tau_{s+1}, \ldots, \tau_{s+m}\}$, is bounded by τ_s and τ_f and contains $m = f - s - 1$ data points; let $\Delta = \tau_f - \tau_s$. We will use the notation $\tau_{I,j}$, $j = 1, \ldots, m$ to refer to the elements of $\boldsymbol{\tau}_I$; note that $\tau_{I,j} = \tau_{s+j}$.

We propose $\boldsymbol{\tau}'_I = \{\tau'_{s+1}, \ldots, \tau'_{s+m}\}$ from their Gamma prior conditioned $\boldsymbol{\tau}_{-I}$, which is equivalent to conditioning on τ_s and τ_f because of the Markov property of the timescale. Furthermore, this is equivalent to proposing new $\{\delta'_{s+1}, \ldots, \delta'_f\}$, which we will denote $\boldsymbol{\delta}_I$ given their sum: Δ. Note that there are $m+1$ elements

in $\boldsymbol{\delta}_I$ as it also includes δ_f. We update the corresponding shape parameters $\boldsymbol{\psi}_I = \{\psi'_{s+1}, \ldots, \psi'_f\}$ jointly with $\boldsymbol{\tau}_I$, also from their prior conditioned on $\boldsymbol{\psi}_{-I}$. This is equivalent to conditioning on ψ_s and ψ_{f+1}.

To select an interval we must firstly choose its length, ℓ. This is defined in years, which is necessary to balance switching equations when changing the cycle count (see Section 4.3.3), and therefore with respect to the latent timescale $\boldsymbol{\tau}$. ℓ is chosen uniformly with respect to a maximum cut-off l, which is discussed in the first example of this chapter (Section 4.6.1).

$$\ell \sim U(0, l) \text{ where } l > 0.$$

We choose the update interval I to be the smallest possible such that $\Delta \geq \ell$ and $m > 0$. In the correlated shape case we condition on ψ_s and ψ_{f+1} which implies that $s > 1$ and $f < n$, so $s < n - 2$. s can take any value such that

$$2 \leq s \leq \max(i < n - 2 \ : \ \tau_{n-1} - \tau_i \geq \ell).$$

This gives the maximum possible number of intervals of length ℓ in a given timescale $\boldsymbol{\tau}$ as

$$n_\ell = \max(i < n - 2 \ : \ \tau_{n-1} - \tau_i \geq \ell) - 1.$$

In the constant shape case this is $1 \leq s \leq \max(i < n - 1 \ : \ \tau_n - \tau_i \geq \ell)$. For a given s the corresponding right hand boundary index is

$$f = \min(i > s + 1 \ : \ \tau_i - \tau_s \geq \ell).$$

It is possible that $m = 1$ for small ℓ, in this case we use the theory presented in Section 4.3.1, otherwise we continue as below.

Proposing $\boldsymbol{\psi}_I$ and $\boldsymbol{\tau}_I$

In the correlated shape case, once we have chosen the interval to update we then propose $\boldsymbol{\psi}'_I$ from its prior conditioned on $\boldsymbol{\psi}_{-I}$. $\boldsymbol{\psi}'_I$ is drawn from the $\mathrm{N}\left(\boldsymbol{\mu}_\psi, \Sigma_\psi\right)$

distribution where

$$\mu_{\psi,j} = \psi_s + \frac{j(\psi_{f+1} - \psi_s)}{m+2} \text{ and } \Sigma_{\psi,jk} = \sigma_\psi^2 \left(\min(j,k) - \frac{jk}{m+2} \right),$$

for $j, k = 1, 2, \ldots, m+1$ (see Section A.3 in the appendix). We now propose $\boldsymbol{\tau}_I$, conditional on $\boldsymbol{\tau}_{-I}$, and $\boldsymbol{\psi}'_I$ by drawing m perturbations, $\boldsymbol{d}'$, from the Dir($\boldsymbol{\psi}'_I$) distribution and setting

$$\boldsymbol{\tau}'_{I,j} = \tau_s + \Delta \sum_{i=1}^{j} d'_i \ : \ j = 1, \ldots, m.$$

Note that in the constant shape case we have $\boldsymbol{\psi}'_{I,j} = \psi \ \forall \ j$. In the equations above the interval I is bounded at both sides, we also require a method of updating the ends of the signal. Below is a scheme for achieving this in the correlated shape case - the constant shape case is a simplified version of this.

LHS Set $f = \min(i > 1 \ : \ \tau_i - \tau_1 \geq \ell)$ and $m = f - 1$. If $m = 1$ continue as in Section 4.3.1; if $m > 1$ set $\mu_{\psi,j} = \psi_{f+1}$ and $\Sigma_{\psi,jk} = \sigma_\psi^2 \min(j,k), \ j, k = 1, 2, \ldots, m+1$. Then propose $\boldsymbol{\tau}'_I = \{\tau'_1, \ldots, \tau'_m\}$ conditional on τ_f: draw perturbations $d'_j \overset{iid}{\sim} \mathrm{G}(\psi'_j, \ \lambda) \ : \ j = 1, \ldots, m$ and set $\tau'_{f-j} = \tau_f - \sum_{k=1}^{j} d'_k$.

RHS Set $s = \max(i < n \ : \ \tau_n - \tau_i \geq \ell)$ and $m = n - s$. If $m = 1$ continue as in Section 4.3.1; if $m > 1$ set $\mu_{\psi,j} = \psi_s$ and $\Sigma_{\psi,jk} = \sigma_\psi^2 \min(j,k), \ j, k = 1, 2, \ldots, m+1$. Then propose $\boldsymbol{\tau}'_I = \{\tau'_{s+1}, \ldots, \tau'_n\}$ conditional on τ_s: draw perturbations $d'_j \overset{iid}{\sim} \mathrm{G}(\psi'_j, \ \lambda) \ : \ j = 1, \ldots, m$ and set $\tau'_{s+j} = \tau_s + \sum_{k=1}^{j} d'_k$.

4.3.3 Switching – updating the cycle count

In this section we discuss how the block updates for $\boldsymbol{\tau}$ described in Section 4.3.2 can be modified to increase or decrease the cycle count in the timescale by 1. This is only done in bounded intervals, not at the edges of the signal. These updates *do not* affect the model dimension which is determined by the number of observations, not the number of cycles. The cycle addition and removal updates use the same l_1 and l_2 parameters, discussed below.

Cycle addition

To propose the addition of a cycle into a subsection of $\boldsymbol{\tau}$ we first choose $\ell^+ \sim U(l_1, l_2)$, for some $0 < l_1 < l_2$, then follow the method described in Section 4.3.2 – apart from when proposing new values for the timescale in I (denoted $\boldsymbol{\tau}_I^+$), in this case we set:

$$\boldsymbol{\tau}_{I,j}^+ = \tau_s + (\Delta + 1) \sum_{i=1}^{j} d_i' \ : \ j = 1, \ldots, m.$$

In the correlated shape parameter case the proposal and prior ratios reduce to:

$$\frac{p(\boldsymbol{\delta}^+)q(\boldsymbol{\delta})}{p(\boldsymbol{\delta})q(\boldsymbol{\delta}^+)} = \frac{(\Delta+1)^{(\Sigma_j \psi_{I,j}' - 1)} e^{-\lambda(\Delta+1)}}{\Delta^{(\Sigma_j \psi_{I,j} - 1)} e^{-\lambda\Delta}} = e^{-\lambda} \frac{(\Delta+1)^{(\Sigma_j \psi_{I,j}' - 1)}}{\Delta^{(\Sigma_j \psi_{I,j} - 1)}}.$$

In the constant shape parameter case this reduces further to:

$$e^{-\lambda} \left(\frac{\Delta + 1}{\Delta} \right)^{(m+1)(\psi - 1)}.$$

(see Section A.2 in the appendix). To avoid machine precision errors in R, we use logs to calculate this and then take exponentials:

$$\log \left(\frac{p(\boldsymbol{\delta}^+)q(\boldsymbol{\delta})}{p(\boldsymbol{\delta})q(\boldsymbol{\delta}^+)} \right) = \lambda(\Sigma_j \psi_{I,j} - 1) \log(\Delta) - \lambda(\Sigma_j \psi_{I,j}' - 1) \log(\Delta + 1).$$

This compares the incremented and original interval lengths (in years) to the $\mathrm{G}((\Sigma_j \psi_{I,j} - 1), \lambda)$ distribution. If the update is accepted we then increment τ_i,

$i \geq f$, by 1.

Figure 4.1 illustrates how $\boldsymbol{\tau}_I$ is proposed without affecting the cycle count (top) and when increasing the cycle count via a switching update (bottom). In both cases $\boldsymbol{\tau}$ and the proposed $\boldsymbol{\tau}^+$ are on the left and the corresponding reconstructions are shown on the right, the interval I is between two circular points.

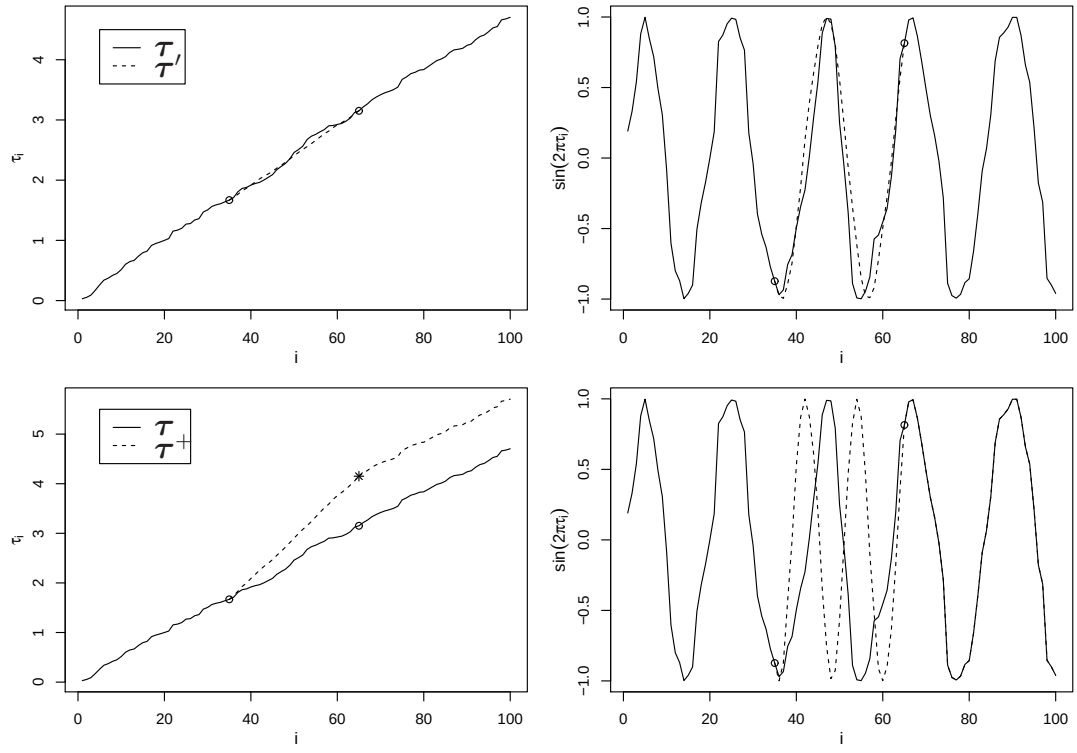

Figure 4.1: Updating the timescale $\boldsymbol{\tau}$ in a time interval using block updates. (top) Without affecting the cycle count (bottom) Increasing the cycle count via a switching update. In both cases $\boldsymbol{\tau}$ and the proposed $\boldsymbol{\tau}^+$ are on the left and the corresponding reconstructions are shown on the right, the interval I is between two circular points.

Cycle removal

To propose the removal of a cycle from a subsection of $\boldsymbol{\tau}$ we choose $\ell^- \sim U(l_1 + 1, l_2+1)$, for the same l_1 and l_2 values as above, to ensure that corresponding birth and death steps are equally likely. This way the acceptance probabilities for the switching updates cancel to the prior proposal ratio multiplied by the likelihood ratio. We then follow the method described in Section 4.3.2 apart from when

proposing new values for the timescale in I (denoted $\boldsymbol{\tau}_I^-$); in this case we set:

$$\boldsymbol{\tau}_{I,j}^- = \tau_s + (\Delta - 1)\sum_{i=1}^{j} d_i' \ : \ j = 1, \ldots, m.$$

The proposal and prior ratios reduce to:

$$\frac{p(\boldsymbol{\delta}^-)q(\boldsymbol{\delta})}{p(\boldsymbol{\delta})q(\boldsymbol{\delta}^-)} = e^{-\lambda}\frac{(\Delta-1)^{(\Sigma_j \psi'_{I,j}-1)}}{\Delta^{(\Sigma_j \psi_{I,j}-1)}}$$

which compares the decremented and original interval lengths (in years) to the $\mathrm{G}((\Sigma_j \psi_{I,j} - 1), \lambda)$ distribution. If the update is accepted we then decrease τ_i, $i \geq f$, by 1.

4.3.4 Updating the hyperparameters

In the correlated case $\boldsymbol{\psi}$ is updated jointly with $\boldsymbol{\tau}$. λ is updated in a separate Gibbs step via its conjugate Gamma prior. We have

$$p(\boldsymbol{\delta}|\lambda) \propto \prod_i \lambda_i^{\psi} e^{\lambda\delta_i} = \lambda^{\Sigma_i \psi_i} e^{\lambda \Sigma_i \delta_i}$$

so if $\lambda \sim G(a, b)$ we have

$$p(\lambda|\boldsymbol{\delta}) \propto p(\boldsymbol{\delta}|\lambda)p(\lambda) \propto \lambda^{\Sigma_i \psi_i + a} e^{\lambda(\Sigma_i \delta_i + b)}$$

which means that λ can be drawn from the $\mathrm{G}(a + \Sigma_i \psi_i,\ \Sigma_i \delta_i + b)$ distribution. We use an uninformative prior for λ, setting $a = b = 1/1000$.

In the constant case ψ is updated via a Metropolis-Hastings step with a improper uniform prior on $[1, \infty)$. λ is updated via a Gibbs step, given a non-informative conjugate Gamma prior. It is drawn from the $\mathrm{G}\left(1/1000 + (n-1)\psi,\ 1/1000 + \Sigma_i \delta_i\right)$ distribution.

η is set at 4 giving some dependency in the residuals after half a cycle ($e^{-4/2} = 0.14$), and slight dependency after a whole cycle ($e^{-4} = 0.02$). σ is updated via a

Gibbs step with a non-informative conjugate inverse-gamma prior.

4.4 Starting values

The MCMC algorithm is robust to the starting values for the model parameters, which do not need to be close to their posterior distribution. Even with very unlikely starting values for $\boldsymbol{\tau}$, as long its elements are ordered, convergence is achieved via the switching updates. Here we describe a scheme for obtaining feasible starting values for $\boldsymbol{\tau}$ that is used in all further examples in this thesis. Although not necessary to fit this model, it does speed convergence, which is conducive to an automated methodology.

Firstly we date the signal using the methods developed in Chapters 2 and 3. The univariate classification method for standardised signals provides a set of change points (or knots) at quarter cycles or seasons. Working along the most likely run sequence for the signal we can assign values for $\boldsymbol{\tau}$ at these depths. For example the depth index at the start of a peak run corresponds to a value of 0.125 for the decimal part of $\boldsymbol{\tau}$. Similarly for the start of descending, troughs and ascending runs the corresponding decimal values are 0.375, 0.625 and 0.875. Keeping track of the count, once we have assigned values for $\boldsymbol{\tau}$ at the depth indexes that correspond to the start of all runs we can linearly interpolate the remaining values of $\boldsymbol{\tau}$. Staring values for $\boldsymbol{\psi}$, λ and σ are calculated from the starting values for $\boldsymbol{\tau}$, $\boldsymbol{\tau}_{(0)}$ say, unless otherwise stated in the example. If $\boldsymbol{\delta}_{(0)}$ is the differenced $\boldsymbol{\tau}_{(0)}$ we have:

$$\psi_{(0)} = E(\boldsymbol{\delta}_{(0)})^2 / var(\boldsymbol{\delta}_{(0)}).$$

In the correlated shape case we set each element of $\boldsymbol{\psi}_{(0)}$ equal to $\psi_{(0)}$, the starting value for σ_ψ has to be chosen manually. Starting values for the other hyperparameters can be obtained by performing Gibbs steps, as described in Section 4.3.4 using the staring values for $\boldsymbol{\psi}_{(0)}$ and $\boldsymbol{\tau}_{(0)}$.

4.5 The MCMC algorithm

All of the examples below use the correlated $\boldsymbol{\psi}$ case to show how it could be used on a relatively simple model. In further chapters the constant shape case is used. For each iteration of the MCMC algorithm we follow:

1. choose ℓ
2. for a sample of A intervals, drawn from all possible intervals of length ℓ without replacement, update:

$$\boldsymbol{\tau}_I, \boldsymbol{\psi}_I \mid \boldsymbol{x}'_I, \lambda, \sigma$$

3. choose $p \sim U(0,1)$
   - if $p < 1/2$:
     - choose ℓ^+
     - for a sample of B intervals, drawn from all possible intervals of length ℓ^+ without replacement, propose a cycle addition update
   - if $p \geq 1/2$:
     - choose ℓ^-
     - for a sample of B intervals drawn from all possible intervals of length ℓ^- without replacement, propose a cycle removal update
4. update the hyperparameters in random order
   - update $\sigma \mid \boldsymbol{\tau}, \boldsymbol{x}'$
   - update $\lambda \mid \boldsymbol{\tau}, \boldsymbol{\psi}$
   - update $\sigma_\psi \mid \boldsymbol{\psi}$

The hyperparameters λ and σ are updated once per iteration. A block updates of $\boldsymbol{\tau}$ are proposed per iteration, with $A \leq n_\ell$. B cycle addition *or* cycle removal

updates are proposed per iteration, with $B < A$. A and B are tuning parameters. The value of A affects the convergence of the within cycle timescale; the ratio of $A : B$ affects the mixing of the switching updates.

4.6 Examples

In this section we present examples of fitting this model to real data. It is fitted to three subsections of standardised ammonium signal from the NGRIP ice core. In each case the Gamma prior shape parameter $\boldsymbol{\psi}$ is modelled as a Gaussian random walk process, the correlated shape case. The first example is on a short section of 'well behaved' signal where the entire stretch is classified into 'certain' runs in the classification process described in Section 3.2.2. The aim of this example is to see how the hyperparameters, particularly λ behave when the model is fitted to a section of signal where the seasonality is well defined. Although switching updates are proposed they are only ever accepted during the convergence of the chain. The other two examples have uncertainty in the cycle count – that is, switching occurs after convergence. The second example has an 'issue' caused by a long stretch of missing values, the third has an 'issue' caused by ambiguity in the cycles. In each case we choose $A = n_\ell$ and $B = 30$ which gives good mixing of the switching updates over the $10,000$ iterations.

4.6.1 Example 1 – a 'well behaved' section

Here we fit the model to a section of 'well behaved' standardised signal $\boldsymbol{x}'$, NGRIP NH_4 between 1454.1m and 1454.4m, which has no missing values and very clear seasonality. $\boldsymbol{x}'$ is the solid line of Figure 4.2, the posterior mean reconstruction, $\sin(2\pi\hat{\boldsymbol{\tau}})$, is shown as a dotted line. The mean reconstruction is a close fit to the $\boldsymbol{x}'$ and comparing the two illustrates the correlation in the residual error process. In this case there are 300 data points, $n = 300$. Over the $10,000$ iterations none of the proposed switching updates were accepted after 1000 iterations of burn-in, giving no uncertainty on the cycle count.

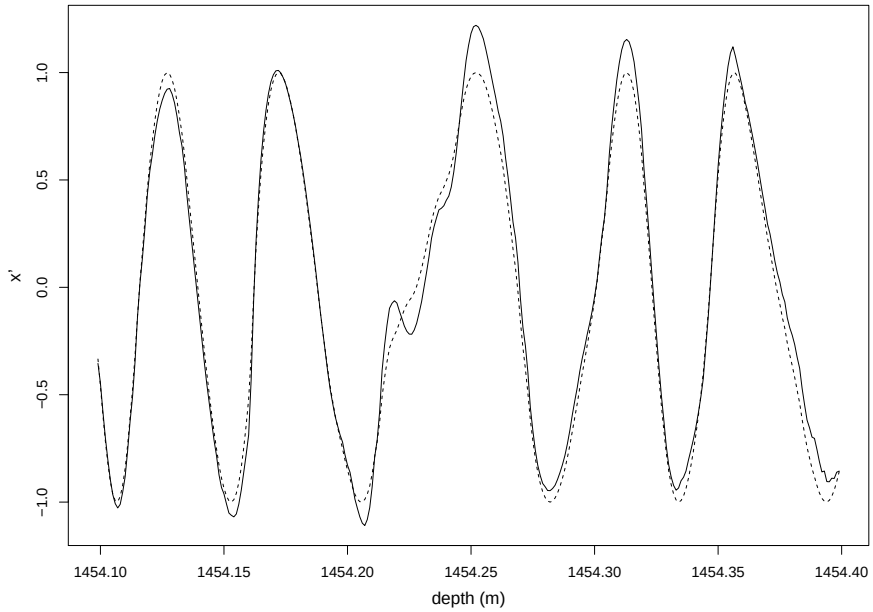

Figure 4.2: The standardised NGRIP NH_4 ($\boldsymbol{x}'$) signal, between 1454.1m and 1454.4m. Its posterior mean reconstruction, calculated from an MCMC run of $10,000$ iterations with a burn-in of 1000 is shown as a dotted line.

Trace plots for the hyperparameters σ, λ and σ_ψ are provided in the appendix as Figure B.4; these suggest convergence of the chain within 1000 iterations. Also included are histograms of their posterior distributions, after a burn-in of 1000 iterations. These give a posterior mode for σ of around 0.076 with $\eta = 4$ (see Section 4.3.4); for λ around 925; and for σ_ψ around 1.16. λ is given an uninformative $G(1/1000, 1/1000)$ prior. The starting value for σ_ψ was chosen as 0.15 to be similar to the converged values from previous experimental runs on similar stretches of signal.

For this example we set $l = 1$: the block update interval lengths, ℓ, are chosen uniformly between 0 and 1 years (see Section 4.3.2). Figure 4.3 is a scatter plot of ℓ against the acceptance rate for the block updates. For very small ℓ, for which we have $m = 1$, the acceptance rate is around 0.7 reducing to around 0.15 for $\ell = 1/2$. In all further examples in this thesis we set $l = 1/2$.

Trace plots for the timescale at the very start and end of the signal, τ_1 and τ_n, are provided in the appendix as Figure B.5. These do not show any evidence

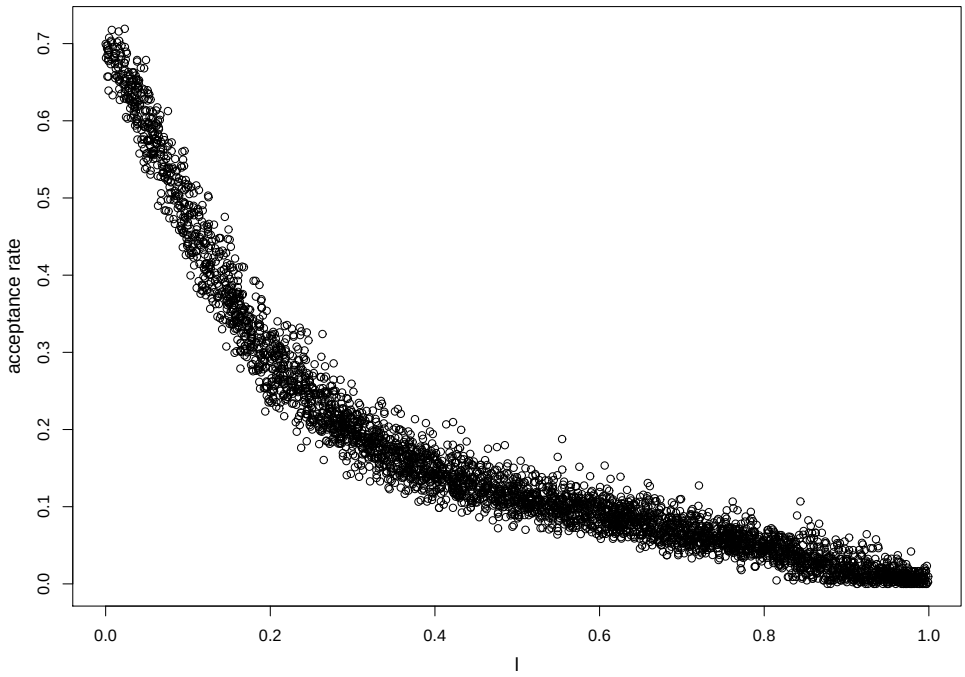

Figure 4.3: Scatter plot of the block update interval length ℓ (in years) against acceptance rate at each iteration of the $10,000$ iteration MCMC run on the standardised NGRIP NH_4 signal (1454.1m–1454.4m).

against the chosen burn-in period of 1000 iterations. Histograms of their posterior distributions, after the burn-in period, are also provided. The posterior mode for τ_1 is approximately 0.555 with a 95% posterior interval (PI) of $(0.546, 0.562)$ – found by calculating the 5th and 95th percentiles of the posterior sample. This implies that the signal starts a little bit after half way through a cycle, in terms of the chosen sine wave cycle shape, which corresponds to what we see in Figure 4.2. The posterior mode for τ_n is approximately 5.838 with a 95% PI of $(5.820, 5.849)$. None of the switching updates were accepted in this MCMC run – which provides a constant posterior cycle count of five, and suggests that the signal ends just after 3/4 of the way through a cycle. Trace plots for the 298 other elements of $\boldsymbol{\tau}$ look very similar to these.

Figure 4.4 is a posterior density heat plot of $\boldsymbol{\psi}$ against depth. Note the dipping feature between 1454.20m and 1454.25m where the process takes relatively low values with greater certainty. This suggests that the δs – the elapsed times over each 1mm depth interval – are relatively large in this section, i.e. the cycle is

relatively 'stretched out'. This can be clearly seen in Figure 4.2. A similar looking plot can be produced for the posterior sample of the $\boldsymbol{\delta}$, the posterior mean for $\boldsymbol{\delta}$ is 0.017 suggesting an average of around 59 points per cycle for this small section of the NGRIP ice core.

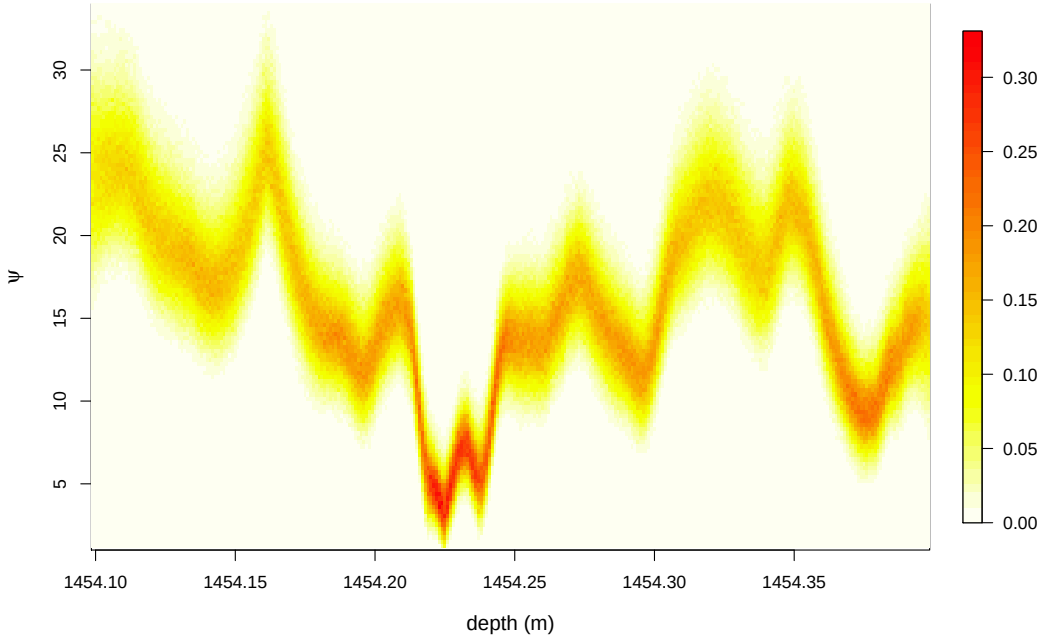

Figure 4.4: Density heat plot of the posterior distribution for the shape process parameter $\boldsymbol{\psi}$ against depth, from the $10,000$ iteration MCMC run on the standardised NGRIP NH_4 signal between 1454.1m–1454.4m.

4.6.2 Example 2 – a stretch of missing values

Here we fit the model to a section of the standardised signal between 1451.1m and 1451.6m, $\boldsymbol{x}'$. This section has a large stretch of missing values, 83mm, which is the longest stretch in the example NGRIP ammonium signal (1440m–1465m).

Through experimentation we have found that problems can occur when fitting this model, with the random walk prior for $\boldsymbol{\psi}$, to data with long stretches of missing values, where $\boldsymbol{\psi}$ is not constrained by the signal. The elements of $\boldsymbol{\psi}$ tend to converge to relatively high values over the stretch of missing values. This is caused by negative proposed values of $\boldsymbol{\psi}$, which are not allowed, being

proposed (and rejected) during the block and switching updates. This causes an undesirable positive skew to the resulting posterior distribution. We choose λ to be large enough such that this 'bias' for larger values of $\boldsymbol{\psi}$ does not occur. It should be noted here that this is not an issue in the constant ψ case which we use in later chapters and that we do not propose this method of MCMC is used to date ice cores. A starting values of $\sigma_\psi = 0.15$ is used, as for the previous example.

To get a feel for what values of λ are appropriate, we investigate the behaviour of $\boldsymbol{\psi}$ over 83 points, when not constrained by data, at different values of λ. For a given value of λ we simulate $20,000$ Gaussian random walks of length 83 points constrained to start and end at $\lambda/60$ – the expected mean of $\boldsymbol{\psi}$ assuming that the mean for $\boldsymbol{\delta}$ is $1/60$ – with standard deviation of of 1.16 (the posterior mode for σ_ψ in example 1). Figure 4.5 is a density plot for the distribution of $\boldsymbol{\psi}/\lambda$ against λ from these simulations for a range of λs between 0 and 5000.

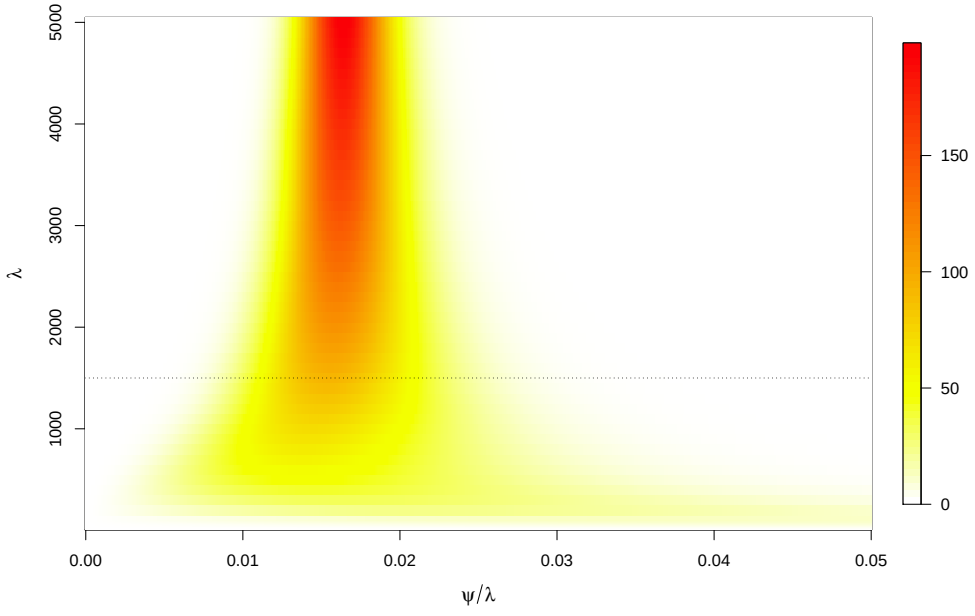

Figure 4.5: Density heat plot of $\boldsymbol{\psi}/\lambda$ against λ from a simulation of $20,000$ random walk $\boldsymbol{\psi}$ processes, constrained to start and end at $\lambda/60$, for a range of λ between 0 and 5000.

The variation in $\boldsymbol{\psi}$ decreases as λ increases. $\lambda = 1500$ is the smallest value such that the average of the simulated $\boldsymbol{\psi}$s is exactly $\lambda/60$. In this and the following

example we set $\lambda = 1500$. In the example with no missing values, above, the posterior mode for λ was around 925 – so we are inflating λ to accommodate the stretch of missing values.

$\boldsymbol{x}'$ is the solid line of Figure 4.6; in this case there are 500 data points, $n = 500$. The algorithm was run for $10,000$ iterations. Switching did occur following convergence, between two possible reconstructions for the signal. The posterior mean reconstruction calculated from iterations where the cycle count is eight is shown as a red dotted line. The posterior mean reconstruction from iterations where the cycle count is nine is shown as a blue dotted line. Where these agree, the reconstruction is shown as a purple dotted line. Visually both of these are feasible, and we discuss how posterior probabilities are attached to them below.

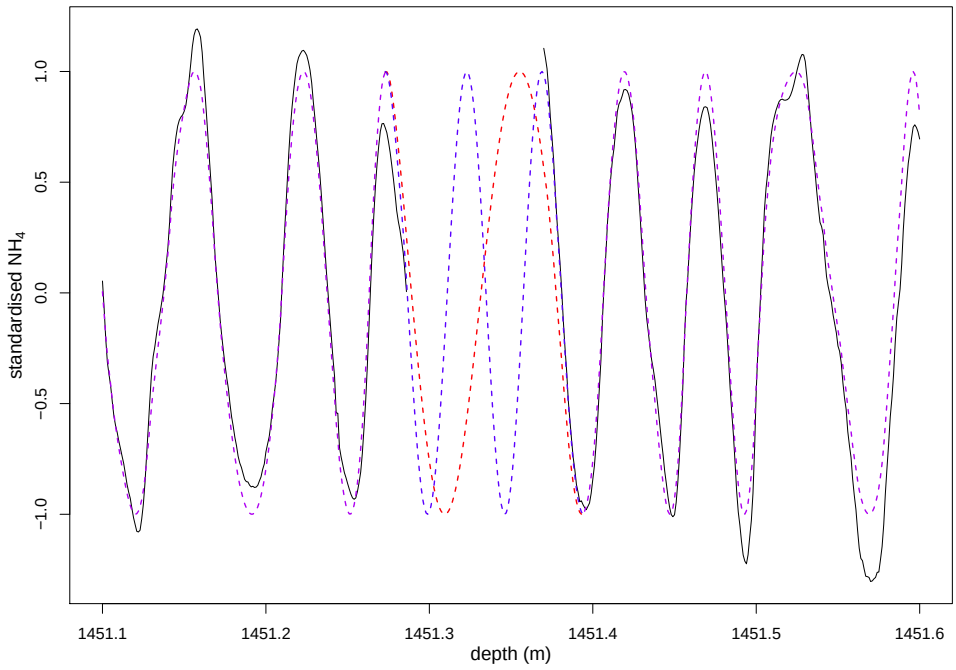

Figure 4.6: The standardised NGRIP NH_4 ($\boldsymbol{x}'$) signal, between 1451.1m and 1451.6m. The posterior mean reconstruction when the cycle count is eight is shown as a red dotted line, when the cycle count is nine is shown as a blue dotted line, where they agree is shown as a purple dotted line. Calculated from an MCMC run of $10,000$ iterations with a burn-in of 1000.

Trace plots for hyperparameters σ and σ_ψ are provided in the appendix as Figure B.6; these suggest that the chain converged within 1000 iterations. Figure B.6 also shows histograms of their posterior distributions, after a burn-in period of

1000 iterations. These give a posterior mode for σ of around 0.065 with $\eta = 4$; and around 1.09 for σ_ψ with $\lambda = 1500$

The trace plot for the end of the timescale, τ_n, is provided as Figure 4.7 along with a histogram of its posterior distribution after a burn-in of 1000 iterations. The chain spent 2041 iterations with $\lfloor \tau_n \rfloor = 8$, and 6959 iterations with $\lfloor \tau_n \rfloor = 9$. This gives the posterior probability for the red and blue reconstructions (Figure 4.6) of 0.23 and 0.77 respectively. The posterior mode of the non-integer part of τ_n, $\tau_n - \lfloor \tau_n \rfloor$, is around 0.353 with a 95% PI of $(0.343, 0.369)$.

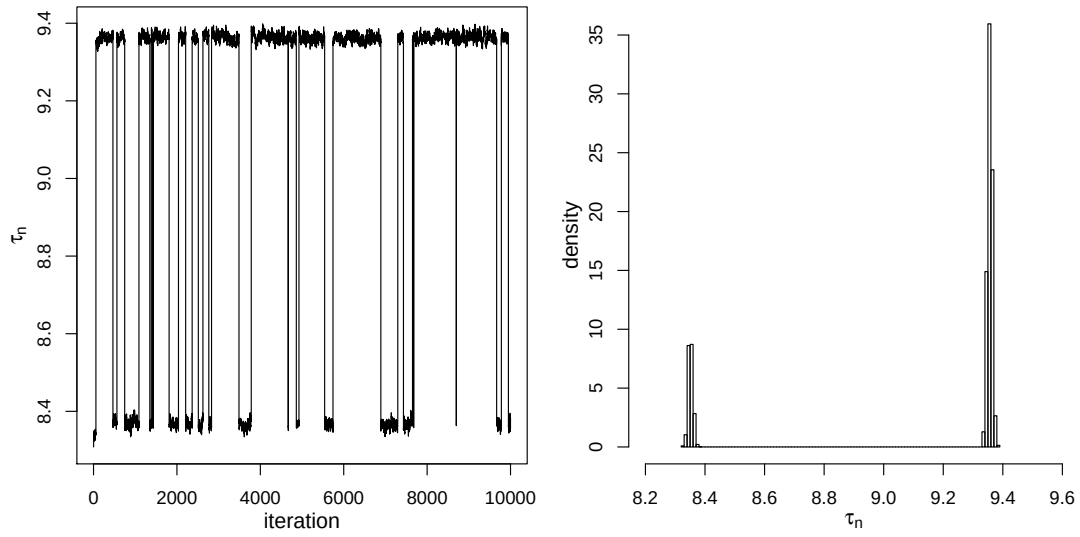

Figure 4.7: A trace plot for the end of the timescale, τ_n, from an MCMC run on the standardised NGRIP NH_4 signal (1451.1m–1451.6m), with a histogram of its posterior distribution after a burn-in of 1000 iterations.

Figure 4.8 is a posterior density plot of the process $\boldsymbol{\psi}$ against depth. Note the feature between 1451.3m and 1451.4m which corresponds to the stretch of missing values in the standardised signal $\boldsymbol{x}'$. There is less certainty about the values of $\boldsymbol{\psi}$ in this range as they are not constrained by the data. We can see that there is higher density for larger values of $\boldsymbol{\psi}$ in this range – which correspond to the reconstruction with a cycle count of nine. $\boldsymbol{\psi}$ are distant from zero over the entire section, suggesting the choice of $\lambda = 1500$ was sufficient in this case, if a little over-kill.

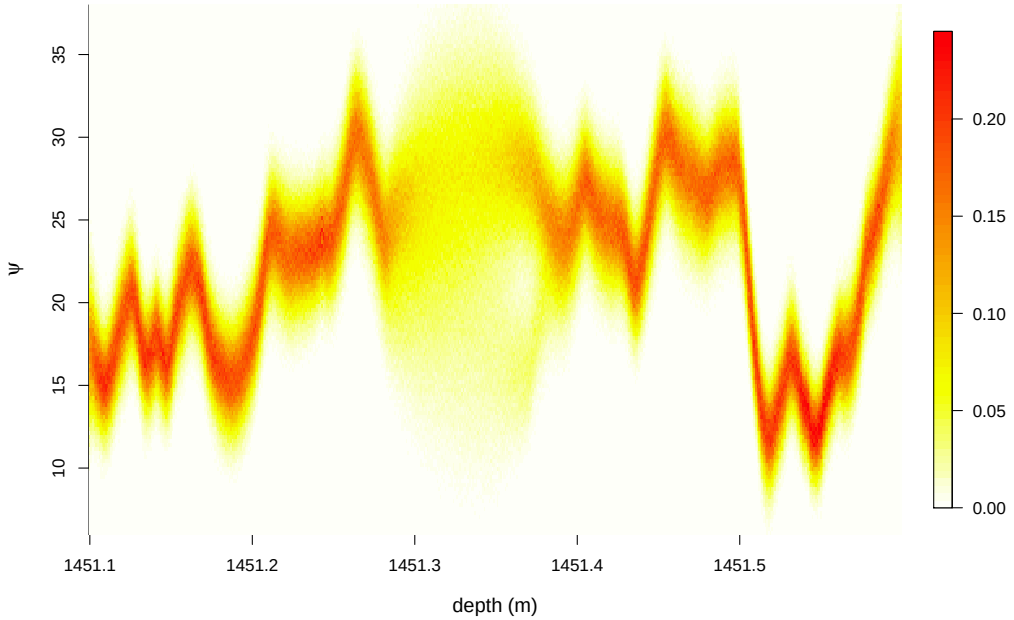

Figure 4.8: Density heat plot of the posterior distribution for the shape process parameter $\boldsymbol{\psi}$ against depth, from the $10,000$ iteration MCMC run on the standardised NGRIP NH_4 signal between 1451.1m and 1451.6m.

4.6.3 Example 3 – switching

Here we fit the model to a section of standardised NGRIP NH_4 signal, $\boldsymbol{x}'$, between 1449.4m and 1450.0m. As with the previous example, we set $\lambda = 1500$. In this case we also set $\sigma = 0.076$, its posterior mode from example 1. This gives better mixing for the switching updates than when updating σ via a Gibbs step. It should be noted that *much less* informative priors are required in the methods presented in later chapters.

$\boldsymbol{x}'$ is the solid line of Figure 4.9, in this case there are 600 data points, $n = 600$, and the algorithm was again run for $10,000$ iterations. Switching updates were accepted following convergence, between two possible reconstructions for the signal. The posterior mean reconstruction calculated from iterations where the cycle count was nine is shown as a red dotted line. The posterior mean reconstruction for iterations when the cycle count was ten is shown as a blue dotted line. Where

the mean reconstructions agree they are shown as a purple dotted line. There is a short stretch of missing values, 13 points, at around 1449.6m – no switching occurred in this section. In the depth range where switching did occur, around 1449.75m, both reconstructions are a poor fit to the signal.

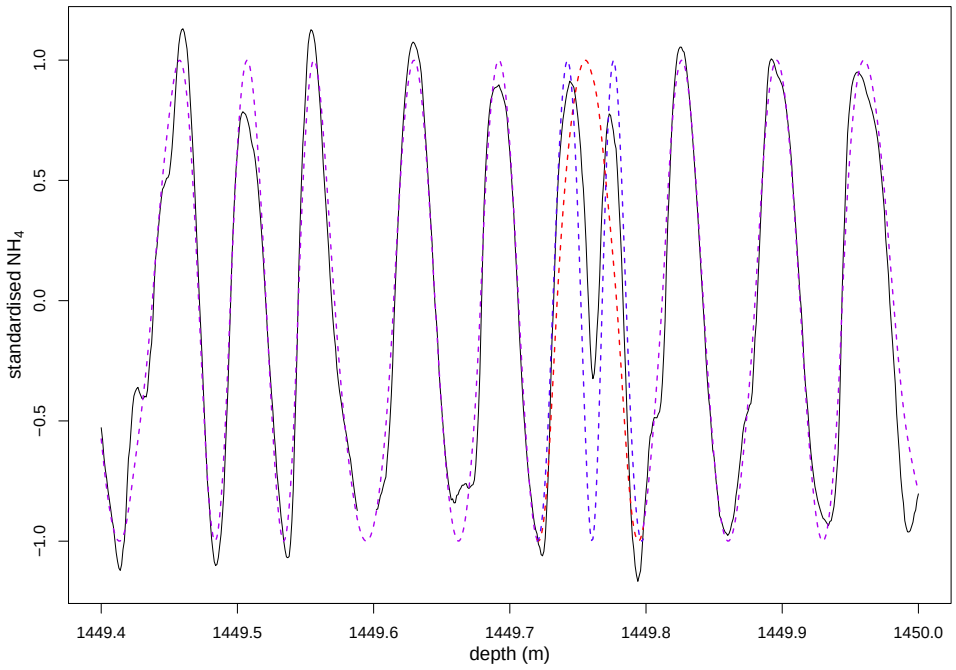

Figure 4.9: The standardised NGRIP NH_4 ($\boldsymbol{x}'$) signal, between 1449.4m and 1450.0m. The posterior mean reconstruction when the cycle count is nine is shown as a red dotted line, when the cycle count is ten is shown as a blue dotted line, where they agree is shown as a purple dotted line. Calculated from an MCMC run of $10,000$ iterations with a burn-in of 2000 iterations.

The trace plot for last element of the timescale, τ_n, is provided as Figure 4.10 along with a histogram of its posterior distribution after a burn-in of 2000 iterations – convergence was slower than in the last two examples. The chain spent 2502 iterations with $\lfloor \tau_n \rfloor = 9$ and 5498 iterations with $\lfloor \tau_n \rfloor = 10$. This gives the posterior probability for the red and blue reconstructions of Figure 4.9 as 0.31 and 0.69 respectively. The posterior mode of the non-integer part of τ_n, $\tau_n - \lfloor \tau_n \rfloor$, is approximately 0.647 with a 95% PI of $(0.633, 0.661)$.

A trace plot for σ_ψ is provided in the appendix as Figure B.7 with a histogram of its posterior distribution after a burn-in of 2000 iterations. This gives a posterior mode for σ_ψ around 1.01 with $\lambda = 1500$.

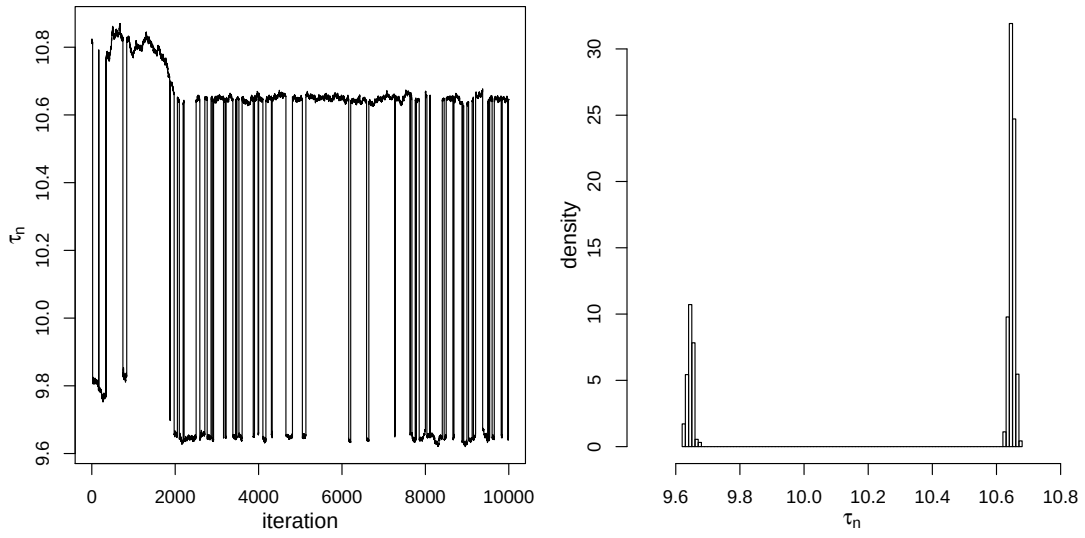

Figure 4.10: Trace plot for the end of the timescale, τ_n, from an MCMC run on the standardised NGRIP NH_4 signal (1449.4m–1450.0m), with a histogram of its posterior distribution after a burn-in of 2000 iterations.

Figure 4.11 is a posterior density heat plot of the shape process $\boldsymbol{\psi}$ against depth. Note the feature between 1449.7m and 1449.8m where switching occurred. There is more certainty about the values of $\boldsymbol{\psi}$ in this range than in the feature we see in Figure 4.8 as $\boldsymbol{\psi}$ is constrained by the data. $\boldsymbol{\psi}$ spent more iterations exploring the top end of this feature, which corresponds to a cycle count of ten. As in the previous example, all values of $\boldsymbol{\psi}$ were distant from zero for all iterations over the entire depth range considered, suggesting that the choice of $\lambda = 1500$ was sufficient.

4.7 Conclusions

In this chapter we introduced the timescale $\boldsymbol{\tau}$, which is the basis for the models that are developed in the following chapters. We showed how it can be proposed from its conditional prior, using either single component or block updates. Two forms for the prior of $\boldsymbol{\tau}$, both based on assuming a Gamma distribution for the amount of time elapsed over a single depth increment, were considered. Example runs for the 'correlated shape parameter' case - which aims to model correlation

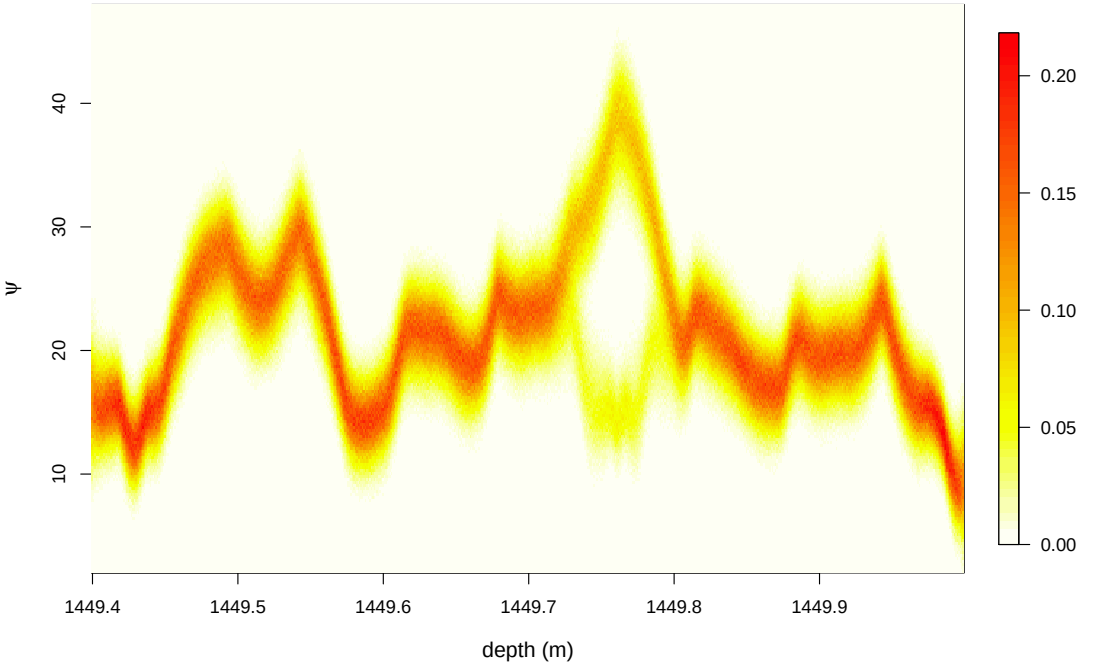

Figure 4.11: Density heat plot of the posterior distribution for the shape process parameter $\boldsymbol{\psi}$ against depth, from the $10,000$ iteration MCMC run on the standardised NGRIP NH_4 signal between 1449.4m and 1450.0m.

in the timescale – required strong priors for the hyperparameters λ and σ, even for this simple model. We therefore only consider the 'constant shape parameter' case in the models going forward, to avoid over-influencing the fit with prior beliefs that are difficult to elicit.

We introduced switching updates for intervals of $\boldsymbol{\tau}$, defined in terms of time, which propose a unit change in the cycle count. These updates do not affect the dimension of the model, avoiding methods like Reversible-Jump MCMC.

The theory introduced here is influenced of the choice of standardisation process and its parameters, β and ν – although the residual error structure of the model is chosen to compensate for this. In further chapters we show how the model can be extended for use on raw or log transformed signals, and how this framework can be extended to the multivariate framework.

Chapter 5

Univariate MCMC - raw signals

5.1 Introduction

In this chapter we extend the univariate model that was introduced in Chapter 4 for use on raw or log transformed annually cyclic ice core signals, using a universal and minimal choice of prior assumptions. This model effectively performs the standardisation process of Chapter 2 as part of the MCMC algorithm, removing the influence of this ad-hoc and somewhat subjective method.

We fit this model to three short example sections of the NGRIP NH_4 signal to illustrate the modelling process. In this and later chapters we only consider the constant shape prior distribution shape case for the timescale $\boldsymbol{\tau}$. Autocorrelation in the timescale is dealt with by thinning the signal closer to its nominal resolution, which leads to improved mixing of the switching updates. In Chapter 6 we show how this model can be extended to a multivariate framework and in Chapter 7 we show how it can be used to fit a chronology to the whole Gomez H_2O_2 and Fletcher $\delta^{18}O$ signals.

5.2 The model

The proposed model for a raw or log transformed annually cyclic signal, $\boldsymbol{x}$, is:

$$x_i = \alpha_i f(\tau_i) + \beta_i + \epsilon_i$$

$i = 1, \ldots, n$, where the timescale $\boldsymbol{\tau}$ is as described in Section 4.2. We again use the cycle shape function $f(\tau_i) = \sin(2\pi\tau_i)$ throughout this chapter, and to simplify notation we define the vector $\boldsymbol{s}$ where $s_i = \sin(2\pi\tau_i)$.

$\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ model the amplitude and mean level of the signal, and are analogous to $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ from Chapter 2. They are intended to be slow-moving correlated processes. For simplicity and mathematical convenience their prior takes the form of two Gaussian random walks,

$$\alpha_i \sim N(\alpha_{i-1}, \sigma_\alpha^2) \text{ and } \beta_i \sim N(\beta_{i-1}, \sigma_\beta^2),$$

for $i = 2, \ldots, n$. We assume that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are independent when fitting the models, though this is not essential, and below we discuss how dependence can be built in where appropriate.

We incorporate independent residuals, $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$, as any correlated noise in the signal is modelled by the autocorrelated multiplicative $\boldsymbol{\alpha}$ and additive $\boldsymbol{\beta}$ processes. We expect the residuals to generally be very low valued relative to the data. They are included to allow anomalous outlier data-points, particularly at the edges of stretches of missing values, and to allow the chain to converge from unlikely starting values. An alternative approach might be to remove outliers manually, and to use a zero-error model where $\epsilon_i = 0 \; \forall \; i$, that is the model:

$$x_i = \alpha_i f(\tau_i) + \beta_i.$$

We discuss any simplifications that can be made in the theory for this model throughout this chapter. These simplifications produce more efficient code. For

this reason it may be beneficial to switch to the zero-error model once σ has dropped below some cut-off value.

5.3 Fitting $\boldsymbol{\tau}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ via MCMC

We use an MCMC scheme which is a mixture of Metropolis-Hastings and Gibbs sampling to sample from the joint posterior distribution of $\boldsymbol{\tau}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The parameters are updated in intervals, denoted I, as introduced in Section 4.3.2. I is made up of m points and is bounded by depth indexes s and f; the details for cases $m = 1$ and $m > 1$ are discussed separately below in Sections 5.3.1 and 5.3.2.

A single update can be broken down into two steps: we firstly propose $\boldsymbol{\tau}'_I$ from its prior conditioned on $\boldsymbol{\tau}_{-I}$ and then secondly we draw $\boldsymbol{\alpha}'_I$ and $\boldsymbol{\beta}'_I$ from their joint distribution conditioned on the signal, $\boldsymbol{\tau}'_I$, $\boldsymbol{\alpha}_{-I}$ and $\boldsymbol{\beta}_{-I}$. We show below that the acceptance probability can be calculated after the first step, allowing us to draw $\boldsymbol{\alpha}'_I$ and $\boldsymbol{\beta}'_I$ only if the update has been accepted.

The Hastings ratio for each update can be broken down into three parts, the likelihood ratio; the prior ratio; and the proposal ratio:

$$\frac{p(\boldsymbol{x}|\boldsymbol{\tau}', \boldsymbol{\alpha}', \boldsymbol{\beta}')}{p(\boldsymbol{x}|\boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta})} \frac{p(\boldsymbol{\tau}', \boldsymbol{\alpha}', \boldsymbol{\beta}')}{p(\boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta})} \frac{q(\boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta})}{q(\boldsymbol{\tau}', \boldsymbol{\alpha}', \boldsymbol{\beta}')}$$

respectively. The likelihood ratio cancels to

$$\frac{p(\boldsymbol{x}|\boldsymbol{\tau}', \boldsymbol{\alpha}', \boldsymbol{\beta}')}{p(\boldsymbol{x}|\boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta})} = \frac{p(\boldsymbol{x}_I|\boldsymbol{\tau}'_I, \boldsymbol{\alpha}'_I, \boldsymbol{\beta}'_I)}{p(\boldsymbol{x}_I|\boldsymbol{\tau}_I, \boldsymbol{\alpha}_I, \boldsymbol{\beta}_I)}.$$

Allowing for dependence in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the prior ratio cancels to

$$\frac{p(\boldsymbol{\tau}', \boldsymbol{\alpha}', \boldsymbol{\beta}')}{p(\boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta})} = \frac{p(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I}) p(\boldsymbol{\alpha}'_I, \boldsymbol{\beta}'_I|\boldsymbol{\alpha}_{-I}, \boldsymbol{\beta}_{-I})}{p(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I}) p(\boldsymbol{\alpha}_I, \boldsymbol{\beta}_I|\boldsymbol{\alpha}_{-I}, \boldsymbol{\beta}_{-I})}$$

as all of the prior structures are Markovian. This cancels further to

$$\frac{p(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I}) p(\boldsymbol{\alpha}'_I|\boldsymbol{\alpha}_{-I}) p(\boldsymbol{\beta}'_I|\boldsymbol{\beta}_{-I})}{p(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I}) p(\boldsymbol{\alpha}_I|\boldsymbol{\alpha}_{-I}) p(\boldsymbol{\beta}_I|\boldsymbol{\beta}_{-I})}$$

if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are assumed to be independent a priori.

$\boldsymbol{\tau}_I$ is proposed from $q(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})$ as explained in Section 4.3. We update $\boldsymbol{\alpha}_I$ and $\boldsymbol{\beta}_I$ via a Gibbs step: from their distribution conditioned on the signal, $\boldsymbol{\tau}'_I$, $\boldsymbol{\alpha}_{-I}$ and $\boldsymbol{\beta}_{-I}$. This gives proposal ratio

$$\frac{q(\boldsymbol{\tau},\boldsymbol{\alpha},\boldsymbol{\beta})}{q(\boldsymbol{\tau}',\boldsymbol{\alpha}',\boldsymbol{\beta}')} = \frac{q(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})p(\boldsymbol{\alpha}_I,\boldsymbol{\beta}_I|\boldsymbol{x}_I,\boldsymbol{\tau}_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}{q(\boldsymbol{\tau}'_I|\boldsymbol{\tau}_{-I})p(\boldsymbol{\alpha}'_I,\boldsymbol{\beta}'_I|\boldsymbol{x}_I,\boldsymbol{\tau}'_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}.$$

From Bayes' Theorem:

$$p(a|b,c) = \frac{p(b|a,c)p(a|c)}{p(b|c)}.$$

Setting $a = \{\boldsymbol{\alpha}_I,\boldsymbol{\beta}_I\}$, $b = \boldsymbol{x}_I$ and $c = \{\boldsymbol{\tau}_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I}\}$, this gives

$$p(\boldsymbol{\alpha}_I,\boldsymbol{\beta}_I|\boldsymbol{x}_I,\boldsymbol{\tau}_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})$$

$$= \frac{p(\boldsymbol{x}_I|\boldsymbol{\alpha}_I,\boldsymbol{\beta}_I,\boldsymbol{\tau}_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})p(\boldsymbol{\alpha}_I,\boldsymbol{\beta}_I|\boldsymbol{\tau}_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}{p(\boldsymbol{x}_I|\boldsymbol{\tau}_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}$$

$$= \frac{p(\boldsymbol{x}_I|\boldsymbol{\tau}_I,\boldsymbol{\alpha}_I,\boldsymbol{\beta}_I)p(\boldsymbol{\alpha}_I,\boldsymbol{\beta}_I|\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}{p(\boldsymbol{x}_I|\boldsymbol{\tau}_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})},$$

giving proposal ratio

$$\frac{q(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})}{q(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I})}\frac{p(\boldsymbol{x}_I|\boldsymbol{\tau}_I,\boldsymbol{\alpha}_I,\boldsymbol{\beta}_I)p(\boldsymbol{\alpha}_I,\boldsymbol{\beta}_I|\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}{p(\boldsymbol{x}_I|\boldsymbol{\tau}_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}\frac{p(\boldsymbol{x}_I|\boldsymbol{\tau}'_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}{p(\boldsymbol{x}_I|\boldsymbol{\tau}'_I,\boldsymbol{\alpha}'_I,\boldsymbol{\beta}'_I)p(\boldsymbol{\alpha}'_I,\boldsymbol{\beta}'_I|\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}$$

$$= \frac{q(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})}{q(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I})}\frac{p(\boldsymbol{x}_I|\boldsymbol{\tau}_I,\boldsymbol{\alpha}_I,\boldsymbol{\beta}_I)}{p(\boldsymbol{x}_I|\boldsymbol{\tau}'_I,\boldsymbol{\alpha}'_I,\boldsymbol{\beta}'_I)}\frac{p(\boldsymbol{\alpha}_I,\boldsymbol{\beta}_I|\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}{p(\boldsymbol{\alpha}'_I,\boldsymbol{\beta}'_I|\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}\frac{p(\boldsymbol{x}_I|\boldsymbol{\tau}'_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}{p(\boldsymbol{x}_I|\boldsymbol{\tau}_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}.$$

The Hastings ratio therefore reduces to:

$$\frac{p(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I})q(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})}{p(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})q(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I})}\frac{p(\boldsymbol{x}_I|\boldsymbol{\tau}'_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}{p(\boldsymbol{x}_I|\boldsymbol{\tau}_I,\boldsymbol{\alpha}_{-I},\boldsymbol{\beta}_{-I})}$$

which is independent of $\boldsymbol{\alpha}'_I$ and $\boldsymbol{\beta}'_I$. For non-switching updates the ratio

$$\frac{p(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I})q(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})}{p(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})q(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I})}$$

cancels, for switching updates see Section 4.3.3 (and Section A.2 in the appendix).

5.3.1 Details for the case: $m = 1$

We first propose τ_i' via the scheme explained in Section 4.3.1. From the Gaussian random walk prior for $\boldsymbol{\alpha}$ we know that a single α_i has conditional prior

$$\alpha_i \mid \boldsymbol{\alpha}_{-i} \sim \mathrm{N}(\mu_a, \sigma_a^2)$$

$$\text{where } \mu_a = \frac{\alpha_{i+1} + \alpha_{i-1}}{2} \text{ and } \sigma_a = \frac{\sigma_\alpha}{2},$$

similarly for $\boldsymbol{\beta}$. If x_i is missing, and we accept τ_i', we draw α_i and β_i separately from their conditional prior distributions.

If x_i is not missing, the conditional distribution for x_i required for the acceptance probability is

$$x_i \mid \tau_i', \boldsymbol{\alpha}_{-i}, \boldsymbol{\beta}_{-i} \sim \mathrm{N}(s_i'\mu_a + \mu_b,\ s_i'^2\sigma_a^2 + \sigma_b^2 + \sigma^2).$$

In the zero error case this reduces to

$$x_i \mid \tau_i', \boldsymbol{\alpha}_{-i}, \boldsymbol{\beta}_{-i} \sim \mathrm{N}(s_i'\mu_a + \mu_b,\ s_i'^2\sigma_a^2 + \sigma_b^2).$$

Once we have proposed and accepted τ_i', we draw new α_i' and β_i' via a Gibbs step. After conditioning on τ_i', $\boldsymbol{\alpha}_{-i}$ and $\boldsymbol{\beta}_{-i}$ the joint distribution of $s_i'\alpha_i$, β_i and x_i is multivariate normal

$$\begin{pmatrix} s_i'\alpha_i \\ \beta_i \\ x_i \end{pmatrix} \sim N\left(\begin{pmatrix} s_i'\mu_\alpha \\ \mu_\beta \\ s_i'\mu_\alpha + \mu_\beta \end{pmatrix}, \begin{pmatrix} s_i'^2\sigma_\alpha^2 & 0 & s_i'^2\sigma_\alpha^2 \\ 0 & \sigma_\beta^2 & \sigma_\beta^2 \\ s_i'^2\sigma_\alpha^2 & \sigma_\beta^2 & s_i'^2\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_x^2 \end{pmatrix}\right).$$

We can partition and condition to find that the joint distribution of $s_i\alpha_i$ and β_i conditional on the signal, τ_i', $\boldsymbol{\alpha}_{-i}$ and $\boldsymbol{\beta}_{-i}$ is multivariate normal with expectation

$$\begin{pmatrix} s_i'\mu_\alpha \\ \mu_\beta \end{pmatrix} + \begin{pmatrix} s_i'^2\sigma_\alpha^2 \\ \sigma_\beta^2 \end{pmatrix} \frac{x_i - s_i'\mu_\alpha - \mu_\beta}{s_i'^2\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_x^2}$$

and covariance matrix

$$\begin{pmatrix} s'^2_i\sigma^2_\alpha & 0 \\ 0 & \sigma^2_\beta \end{pmatrix} - (s'^2_i\sigma^2_\alpha + \sigma^2_\beta + \sigma^2_x)^{-1} \begin{pmatrix} s'^2_i\sigma^2_\alpha \\ \sigma^2_\beta \end{pmatrix} (s'^2_i\sigma^2_\alpha, \sigma^2_\beta)$$

$$= \begin{pmatrix} s'^2_i\sigma^2_\alpha & 0 \\ 0 & \sigma^2_\beta \end{pmatrix} - (s'^2_i\sigma^2_\alpha + \sigma^2_\beta + \sigma^2_x)^{-1} \begin{pmatrix} s'^4_i\sigma^4_\alpha & s'^2_i\sigma^2_\alpha\sigma^2_\beta \\ s'^2_i\sigma^2_\alpha\sigma^2_\beta & \sigma^4_\beta \end{pmatrix}$$

(see Section A.1 in the appendix). We draw $(s'_i\alpha_i)'$ and β'_i from this distribution and set $\alpha'_i = (s'_i\alpha_i)'/s'_i$. In the zero-error case we draw α_i from its conditional distribution

$$\alpha_i \mid x_i, \boldsymbol{\tau}_{-i}, \boldsymbol{\alpha}_{-i}, \boldsymbol{\beta}_{-i} \sim N\left(\mu_\alpha + \frac{s'_i\sigma^2_\alpha}{s'^2_i\sigma^2_\alpha + \sigma^2_\beta}(x_i - s'_i\mu_\alpha - \mu_\beta),\ \frac{\sigma^2_\alpha\sigma^2_\beta}{s'^2_i\sigma^2_\alpha + \sigma^2_\beta}\right)$$

and set $\beta'_i = x_i - s'_i\alpha'_i$.

5.3.2 Details for the case: $m > 1$

We first propose $\boldsymbol{\tau}'_I$ as explained in Section 4.3.2. To simplify notation, we define the vector $\boldsymbol{s}_I$ where $s_{I,j} = \sin(2\pi\tau_{I,j}),\ j = 1, \ldots, m$. We also define S as an $m \times m$ matrix with s_I along the diagonal and zeros elsewhere, such that

$$\boldsymbol{x}_I = S\boldsymbol{\alpha}_I + \boldsymbol{\beta}_I + \boldsymbol{\epsilon}_I.$$

Switching updates follow this same scheme, with the proposal distribution for $\boldsymbol{\tau}$ as explained in Section 4.3.3. There are three cases to consider: all of $\boldsymbol{x}_I$ is missing; none of $\boldsymbol{x}_I$ is missing; some of $\boldsymbol{x}_I$ is missing. We discuss each case below.

All of $\boldsymbol{x}_I$ missing

In this case, if we accept the proposed values for $\boldsymbol{\tau}_I$, we draw $\boldsymbol{\alpha}_I$ and $\boldsymbol{\beta}_I$ from their conditional priors

$$\boldsymbol{\alpha}_I \mid \boldsymbol{\alpha}_{-I} \sim \mathrm{N}(\boldsymbol{\mu}_\alpha,\ \Sigma_\alpha), \quad \boldsymbol{\beta}_I \mid \boldsymbol{\beta}_{-I} \sim \mathrm{N}(\boldsymbol{\mu}_\beta,\ \Sigma_\beta),$$

where

$$\mu_{\alpha,j} = \alpha_s + \frac{j(\alpha_f - \alpha_s)}{m+1} \text{ and } \Sigma_{\alpha,jk} = \sigma_\alpha^2 \left(\min(j,k) - \frac{jk}{m+1} \right),$$

for $j, k = 1, \ldots, m$, recalling that the interval I is bounded by depth indexes s and f (see Sections 4.3, A.3). Similarly for $\boldsymbol{\beta}_I$.

None of $\boldsymbol{x}_I$ missing

The relevant conditional distribution to calculate the acceptance probability is

$$\boldsymbol{x}_I \mid \boldsymbol{\tau}'_I, \boldsymbol{\alpha}_{-I}, \boldsymbol{\beta}_{-I} \sim \mathrm{N}(S\boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\beta,\ S\Sigma_\alpha S^T + \Sigma_\beta + \sigma^2 1_m)$$

where 1_m is the identity matrix of order m. Note that for the model with zero error this simplifies to

$$\boldsymbol{x}_I \mid \boldsymbol{\tau}'_I, \boldsymbol{\alpha}_{-I}, \boldsymbol{\beta}_{-I} \sim \mathrm{N}(S\boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\beta,\ S\Sigma_\alpha S^T + \Sigma_\beta)$$

Once we have proposed and accepted $\boldsymbol{\tau}'_I$, we draw new $\boldsymbol{\alpha}'_I$ and $\boldsymbol{\beta}'_I$ via a Gibbs step. After conditioning on $\boldsymbol{\tau}'_I$, $\boldsymbol{\alpha}_{-I}$ and $\boldsymbol{\beta}_{-I}$ the joint distribution of $S\boldsymbol{\alpha}_I$, $\boldsymbol{\beta}_I$ and $\boldsymbol{x}_I$ is multivariate normal.

$$\begin{pmatrix} S\boldsymbol{\alpha}_I \\ \boldsymbol{\beta}_I \\ \boldsymbol{x}_I \end{pmatrix} \sim N\left(\begin{pmatrix} S\boldsymbol{\mu}_\alpha \\ \boldsymbol{\mu}_\beta \\ S\boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\beta \end{pmatrix}, \begin{pmatrix} S\Sigma_\alpha S^T & \mathbf{0} & S\Sigma_\alpha S^T \\ \mathbf{0} & \Sigma_\beta & \Sigma_\beta \\ S\Sigma_\alpha S^T & \Sigma_\beta & S\Sigma_\alpha S^T + \Sigma_\beta + \sigma_x^2 I_m \end{pmatrix} \right)$$

We can partition and condition to find that the joint distribution of $S\boldsymbol{\alpha}_I$ and $\boldsymbol{\beta}_I$ conditional on the signal, $\boldsymbol{\tau}'_I$, $\boldsymbol{\alpha}_{-I}$ and $\boldsymbol{\beta}_{-I}$ is multivariate normal with expectation

$$\begin{pmatrix} S\boldsymbol{\mu}_\alpha \\ \boldsymbol{\mu}_\beta \end{pmatrix} + \begin{pmatrix} S\Sigma_\alpha S^T \\ \Sigma_\beta \end{pmatrix} (S\Sigma_\alpha S^T + \Sigma_\beta + \sigma_x^2 I_m)^{-1} (\boldsymbol{x}_I - S\boldsymbol{\mu}_\alpha - \boldsymbol{\mu}_\beta)$$

and covariance matrix

$$\begin{pmatrix} S\Sigma_\alpha S^T & \mathbf{0} \\ \mathbf{0} & \Sigma_\beta \end{pmatrix} - \begin{pmatrix} S\Sigma_\alpha S^T \\ \Sigma_\beta \end{pmatrix} (S\Sigma_\alpha S^T + \Sigma_\beta + \sigma_x^2 I_m)^{-1} (S\Sigma_\alpha S^T,\ \Sigma_\beta)$$

(see Section A.1 in the appendix). We draw $(S\boldsymbol{\alpha}_I)'$ and $\boldsymbol{\beta}_i'$ from this distribution, and set $\boldsymbol{\alpha}_I' = S^{-1}(S\boldsymbol{\alpha}_I)'$. In the zero error case we can save on computational expenditure by drawing $\boldsymbol{\alpha}_I'$ from the lower dimensional multivariate normal distribution with mean

$$\boldsymbol{\mu}_\alpha + \Sigma_\alpha S^T (S\Sigma_\alpha S^T + \Sigma_\beta)^{-1} (\boldsymbol{x}_I - S\boldsymbol{\mu}_\alpha - \boldsymbol{\mu}_\beta)$$

and covariance matrix

$$\Sigma_\alpha S^T - \Sigma_\alpha S^T (S\Sigma_\alpha S^T + \Sigma_\beta)^{-1} S\Sigma_\alpha S^T,$$

and setting $\boldsymbol{\beta}_I' = \boldsymbol{x}_I - S\boldsymbol{\alpha}_I'$.

Some of $\boldsymbol{x}_I$ missing

There are m data points in the interval I, of which p are present and $m - p$ are missing. Define $\boldsymbol{x}_p$ as the sub-vector of $\boldsymbol{x}_I$ containing the non-missing values, still ordered by depth. Let P be the $(p \times m)$ matrix that satisfies:

$$P\boldsymbol{x}_I = \boldsymbol{x}_p.$$

The relevant conditional distribution to calculate the acceptance probability is

$$\boldsymbol{x}_p \mid \boldsymbol{\tau}_I', \boldsymbol{\alpha}_{-I}, \boldsymbol{\beta}_{-I} \sim \mathrm{N}(P(S\boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\beta),\ P(S\Sigma_\alpha S^T + \Sigma_\beta + \sigma_x^2 1_m)P^T).$$

For the zero error model this is

$$\boldsymbol{x}_p \mid \boldsymbol{\tau}_I', \boldsymbol{\alpha}_{-I}, \boldsymbol{\beta}_{-I} \sim \mathrm{N}(P(S\boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\beta),\ P(S\Sigma_\alpha S^T + \Sigma_\beta)P^T).$$

Once we have proposed and accepted $\boldsymbol{\tau}'_I$ we draw new $\boldsymbol{\alpha}'_I$ and $\boldsymbol{\beta}'_I$ via a Gibbs step. After conditioning on $\boldsymbol{\tau}'_I$, $\boldsymbol{\alpha}_{-I}$ and $\boldsymbol{\beta}_{-I}$ the joint distribution of $S\boldsymbol{\alpha}_I$, $\boldsymbol{\beta}_I$ and $\boldsymbol{x}_p$ is multivariate normal with mean

$$\begin{pmatrix} S\boldsymbol{\mu}_\alpha \\ \boldsymbol{\mu}_\beta \\ P(S\boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\beta) \end{pmatrix}$$

and covariance matrix

$$\begin{pmatrix} S\Sigma_\alpha S^T & \mathbf{0} & S\Sigma_\alpha S^T P^T \\ \mathbf{0} & \Sigma_\beta & \Sigma_\beta P^T \\ PS\Sigma_\alpha S^T & P\Sigma_\beta & P(S\Sigma_\alpha S^T + \Sigma_\beta + \sigma_x^2 I_m)P^T \end{pmatrix}.$$

We can partition and condition to find that the joint distribution of $S\boldsymbol{\alpha}_I$ and $\boldsymbol{\beta}_I$ conditional on the signal, $\boldsymbol{\tau}'_I$, $\boldsymbol{\alpha}_{-I}$ and $\boldsymbol{\beta}_{-I}$ is multivariate normal with expectation

$$\begin{pmatrix} S\boldsymbol{\mu}_\alpha \\ \boldsymbol{\mu}_\beta \end{pmatrix} + \begin{pmatrix} S\Sigma_\alpha S^T P^T \\ \Sigma_\beta P^T \end{pmatrix} (P(S\Sigma_\alpha S^T + \Sigma_\beta + \sigma_x^2 I_m)P^T)^{-1}(\boldsymbol{x}_p - P(S\boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\beta))$$

and covariance matrix

$$\begin{pmatrix} S\Sigma_\alpha S^T & \mathbf{0} \\ \mathbf{0} & \Sigma_\beta \end{pmatrix} - \begin{pmatrix} S\Sigma_\alpha S^T P^T \\ \Sigma_\beta P^T \end{pmatrix} (P(S\Sigma_\alpha S^T + \Sigma_\beta + \sigma_x^2 I_m)P^T)^{-1}(PS\Sigma_\alpha S^T, P\Sigma_\beta)$$

(see Section A.1 in the appendix). We draw $(S\boldsymbol{\alpha}_I)'$ and $\boldsymbol{\beta}'_i$ from this distribution and set $\boldsymbol{\alpha}'_I = S^{-1}(S\boldsymbol{\alpha}_I)'$. $\boldsymbol{\alpha}_I$ and $\boldsymbol{\beta}_I$ must also be drawn jointly in the zero error case.

5.4 Prior assumptions and updating the hyper-parameters

For this model we use a minimal choice of prior assumptions that works on all annually cyclic signals considered in this research. We only consider the constant shape prior distribution shape case for the timescale $\boldsymbol{\tau}$ (see Section 4.2.2). Autocorrelation in the timescale is reduced by thinning the signal closer to its nominal resolution if necessary, which leads to much improved mixing in the switching updates for the NGRIP ammonium examples below. We update ψ using a Metropolis-Hastings step with a Gaussian prior

$$\psi \sim \mathrm{N}(3, 1/9).$$

We have run the algorithm on a variety of 'well behaved' stretches of ice core signal with an improper prior on ψ ($p(\psi) \propto 1$), and its posterior mean was always in the range $(2, \ldots, 4)$. This prior should ensure that $\psi > 1$ as we want to keep all of the elapsed times in a depth interval $\delta_i > 0$. One issue when fitting this model to subsections of signal with more ambiguous seasonality is that ψ can explore some large values ($\psi > 10$) before the chain converges. This causes the algorithm to fit an equally spaced timescale, and for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ to compensate. Having an upper limit for ψ prevents this happening. Keeping ψ roughly in the range $(2, \ldots, 4)$ puts the probability of an individual δ_i being half or twice the mean level in the range $(0.18, 0.35)$. This sort of logic could be used to elicit the prior for ψ from an expert. λ is given an uninformative conjugate Gamma prior, $\lambda \sim \mathrm{G}(1/1000, 1/1000)$ and is updated via a Gibbs step.

Another issue that can arise when fitting this model to subsections of signal with more ambiguous seasonality is that, before convergence of the chain, $\alpha_i \to 0\ \forall i$. This reduces the model to

$$x_i = \beta_i + \epsilon_i$$

and $\boldsymbol{\beta}$ fits to the signal. To avoid this we set $\alpha_i > c_\alpha$ for some tuning parameter

c_α. For a conservative value for c_α, we first perform the standardisation process introduced in Chapter 2, and set $c_\alpha = \min(\boldsymbol{\sigma})/2$. This keeps the layer counting process automatic with universal prior definitions between signals and subsections of signals.

σ, σ_α and σ_β are given uninformative conjugate Inverse-Gamma priors, with shape and scale equal to 1/1000, and are sampled via Gibbs steps.

5.5 The MCMC algorithm

Each iteration of the MCMC algorithm:

1. choose ℓ
2. for a sample of A intervals I, drawn from all possible intervals of length ℓ without repetition, propose: $\boldsymbol{\tau}'_I \mid \boldsymbol{\tau}_{-I}, \lambda, \psi$
    - if accepted draw $\boldsymbol{\alpha}'_I, \boldsymbol{\beta}'_I \mid \boldsymbol{\tau}'_I, \boldsymbol{\alpha}_{-I}, \boldsymbol{\beta}_{-I}, \boldsymbol{x}_I, \sigma_\alpha, \sigma_\beta, \sigma$
3. choose $p \sim U(0,1)$
    - if $p < 1/2$:
        - choose ℓ^+
        - for a sample of B intervals, drawn from all possible intervals of length ℓ^+ without repetition, propose a cycle addition update
    - if $p \geq 1/2$:
        - choose ℓ^-
        - for a sample of B intervals, drawn from all possible intervals of length ℓ^- without repetition, propose a cycle removal update
4. update the hyperparameters in random order
    - update $\sigma \mid \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{x}$ (Gibbs)
    - update $\lambda \mid \boldsymbol{\tau}, \psi$ (Gibbs)

- update $\psi \mid \boldsymbol{\tau}, \lambda$ (Gibbs)
- update $\sigma_\alpha \mid \boldsymbol{\alpha}$ (M-H)
- update $\sigma_\beta \mid \boldsymbol{\beta}$ (Gibbs)

The hyperparameters λ, σ, σ_α, σ_β and ψ are updated once per iteration. A block updates of $\boldsymbol{\tau}$ are proposed per iteration, with $A \leq n_\ell$. B cycle addition or removal updates are proposed per iteration, with $B < A$. A and B are tuning parameters. The value of A affects the convergence of the within cycle timescale; the ratio of $A : B$ affects the mixing of the switching updates.

5.6 Starting values

The following is a scheme to select all starting values automatically without any need for user input. These values also ensure quick convergence of the chain.

The starting value for $\boldsymbol{\tau}$ is calculated as described in Section 4.4. The starting values for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are taken to be $\boldsymbol{\alpha}_0 = \boldsymbol{\sigma}$ and $\boldsymbol{\beta}_0 = \boldsymbol{\mu}$, from the standardisation process described in Chapter 2. Given its prior, the starting value for ψ is always taken as $\psi_0 = 3$. All of the other hyperparameters of the model are updated via Gibbs steps – their starting values can be obtained by updating them with respect to the starting values for $\boldsymbol{\tau}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

5.7 Examples

Here we present examples of fitting this model to real data using three subsections of the log transformed ammonium signal from the NGRIP ice core. This signal has been thinned, taking every sixth data point, to minimise correlation in the signal and accommodate the constant shape ψ Gamma prior for the timescale. Each depth index now corresponds to a 6mm elapsed depth, the nominal resolution for this signal is 1cm, however thinning the signal this much leads to little information remaining on the seasonality. The first two examples are the same stretches of

signal as for the examples of Section 4.6. The third example is a different stretch of data with no missing values where switching occurs for this model.

$C_\alpha = \min(\boldsymbol{\sigma})/2$ always. The standard deviation for the Metropolis-Hastings updates for ψ is set at 0.45, which has been tuned in experimental runs of the algorithm on similar data to achieve an acceptance rate of approximately 25%. In each example we choose $A = n_\ell$ and $B = 8$ which gives good mixing of the switching updates over the $10,000$ iterations in the examples where they occur.

5.7.1 Example 1 – the 'well behaved' section

Here we fit the model to a section of 'well behaved' log transformed signal $\boldsymbol{x}$, the thinned NGRIP NH_4 between 1454.1m and 1454.4m, which has no missing values and very clear seasonality. In this case $C_\alpha = 0.11$ and there are 51 data points after thinning, $n = 51$. Over the $10,000$ iterations none of the proposed switching updates were accepted, giving no uncertainty on the cycle count. Figure 5.1 shows the model fit to the signal, and each sub-plot is described in the paragraphs below.

Figure 5.1 (top): this shows the signal $\boldsymbol{x}$ as circular points. The posterior mean reconstruction, $\hat{\boldsymbol{x}} = E(\boldsymbol{\alpha}\sin(2\pi\boldsymbol{\tau}) + \boldsymbol{\beta})$, is shown as a dashed line. This reconstruction is an extremely close fit to $\boldsymbol{x}$; the posterior mean for σ, after a burn in of 1000 iterations, is 0.037 – which is small on the scale of this data and suggests that we could have swapped over to the zero-mean model after convergence. $\hat{\boldsymbol{\beta}}$, the posterior mean reconstruction for $\boldsymbol{\beta}$ is shown as a blue dashed line. $\hat{\boldsymbol{\beta}} \pm \hat{\boldsymbol{\alpha}}$ is shown as two red dashed lines. We can see how these have compensated for noise in the signal just after 1454.2m. Although these mean values look smooth, at a given iteration the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ processes can be relatively noisy.

Figure 5.1 (middle): this shows the de-trended signal $\boldsymbol{x} - \hat{\boldsymbol{\beta}}$ as circular points, its posterior mean reconstruction $E(\boldsymbol{\alpha}\sin(2\pi\boldsymbol{\tau}))$ is shown as a dashed line. $\hat{\boldsymbol{\alpha}}$, the posterior mean reconstruction for $\boldsymbol{\alpha}$ is shown as a dotted red line.

Figure 5.1 (bottom): this shows the de-trended and normalised signal $(\boldsymbol{x} - \hat{\boldsymbol{\beta}})/\hat{\boldsymbol{\alpha}}$ as circular points, $\sin(2\pi\hat{\boldsymbol{\tau}})$ is shown as a dotted line. We can see how the fitted timescale has also compensated for the feature just after 1454.2m.

The equivalent plot showing the fit of the starting values for the chain is included in the appendix as Figure B.8. The starting values calculated for the hyperparameters were: $\sigma = 0.060$, $\sigma_\alpha = 0.051$, $\sigma_\beta = 0.070$ and $\lambda = 29.1$.

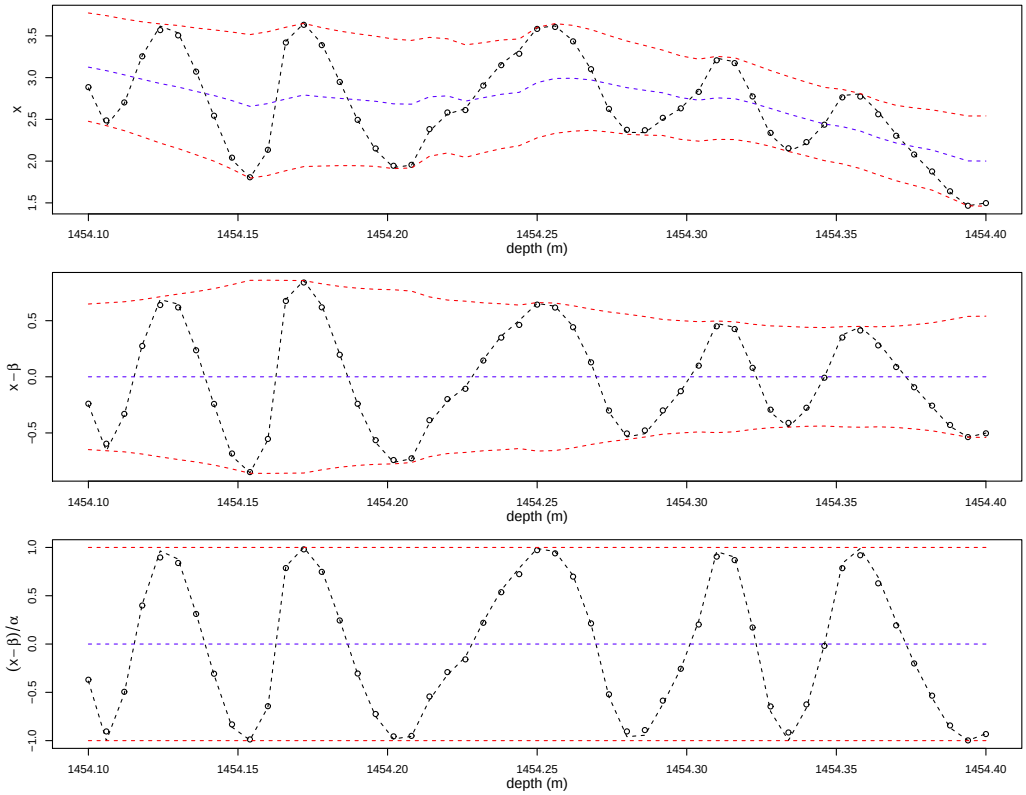

Figure 5.1: The posterior mean model fit to the log NGRIP NH_4 ($\boldsymbol{x}$) signal (1454.1m–1454.4m), thinned to 6mm resolution, calculated from an MCMC run of $10,000$ iterations with a burn-in of 1000. **(top)** $\boldsymbol{x}$ as circular points, the reconstruction of $\boldsymbol{x}$ as a black dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

Trace plots for hyperparameters σ, σ_α, σ_β, λ and ψ are provided in the appendix as Figure B.9, these suggest almost immediate convergence of the chain from their signal-specific starting values. The acceptance rate for ψ was 37.2%, which

is a little higher than intended.

For this example we set $l = 1/2$, that is the block update interval lengths ℓ are chosen uniformly between 0 and 1/2 years. Figure 5.2 is a scatter plot of ℓ against the acceptance ratio at each iteration. For very small ℓ, where we have $m = 1$, the acceptance rate is around 0.65 reducing to around 0.2 for $\ell = 1/2$.

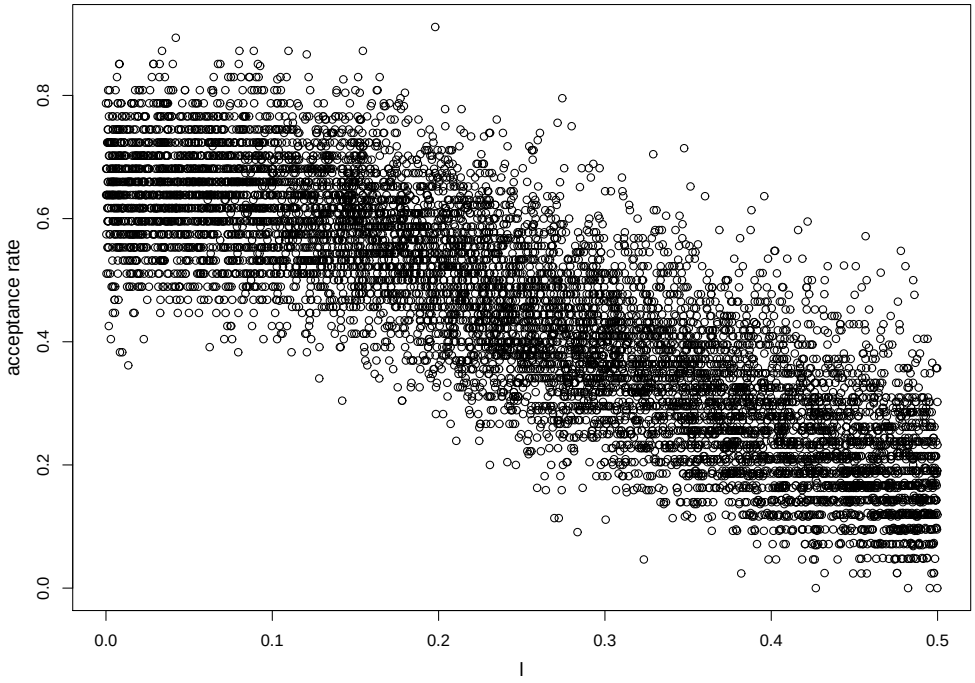

Figure 5.2: Scatter plot of the block update interval length ℓ (in years) against acceptance rate at each iteration of the $10,000$ iteration MCMC run on the log NGRIP NH_4 signal (1454.1m–1454.4m).

A trace plot for the timescale at the end of the signal, τ_n, is provided as Figure 5.3 (left). A histogram of its posterior distribution is also provided. The posterior mean for τ_n is 5.811 with a 95% PI of $(5.739, 5.881)$. There was no switching, so this gives a constant posterior cycle count of five and suggests that the signal ends just after 3/4 way through a cycle. Trace plots for the 299 other elements of $\boldsymbol{\tau}$ are similar to this.

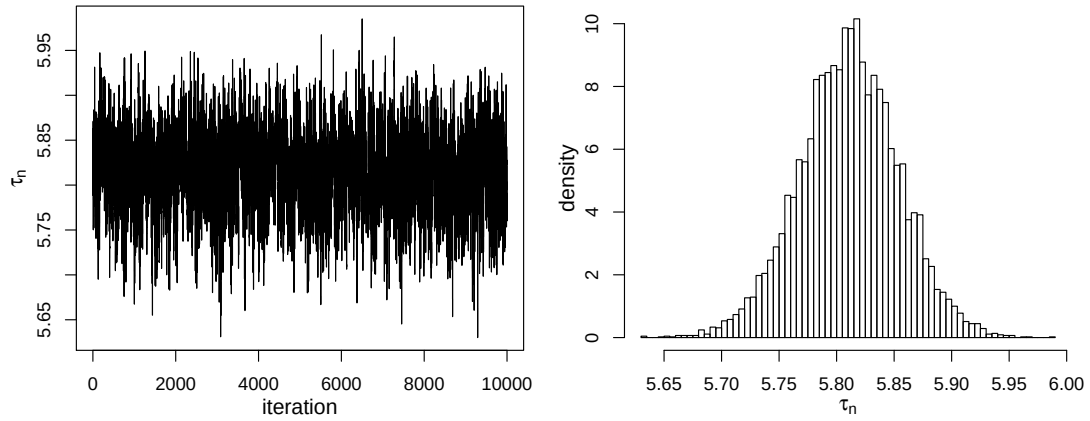

Figure 5.3: Trace plot for the end of the timescale, τ_n, from an MCMC run on the log NGRIP NH_4 signal (1454.1m–1454.4m), with a histogram of its posterior distribution after a burn-in of 1000 iterations.

The posterior distributions for $\boldsymbol{\alpha}$ (top) and $\boldsymbol{\beta}$ (bottom) are provided as Figure 5.4. These are in the form of boxplots at each thinned 6mm depth index. Note that $\boldsymbol{\alpha}$ has a lower bound of $C_\alpha = 0.11$, which appears to have affected the process towards the end, whereas $\boldsymbol{\beta}$ is only constrained by the signal.

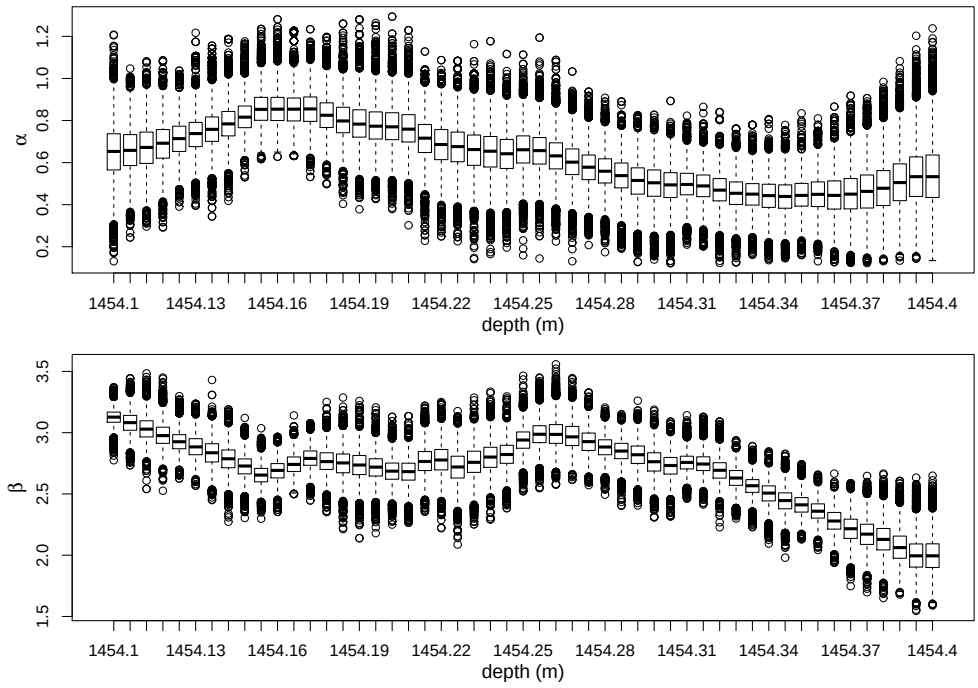

Figure 5.4: Posterior densities of the elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ – shown as boxplots evaluated at each thinned 6mm depth index – from the $10,000$ iteration MCMC run on the log NGRIP NH_4 signal between 1454.1m and 1454.4m.

5.7.2 Example 2 – a stretch of missing values

Here we fit the model to the log transformed NGRIP NH_4 signal between 1451.1m and 1451.6m, $\boldsymbol{x}$, after thinning to 6mm resolution. This section has a large stretch of missing values, 83mm, which is the longest in our example stretch of NGRIP ammonium signal. Unlike the previous chapter, no adaptation to the model's hyperparameters is required here. In this case $C_\alpha = 0.21$.

A trace plot for the timescale at the end of the signal, τ_n, is provided as Figure 4.7, a histogram of its posterior distribution is also provided. The chain explored four possible cycle counts: 7, 8, 9 and 10 years; with posterior probabilities of 2.3%, 85.3%, 12.0% and 0.3% respectively.

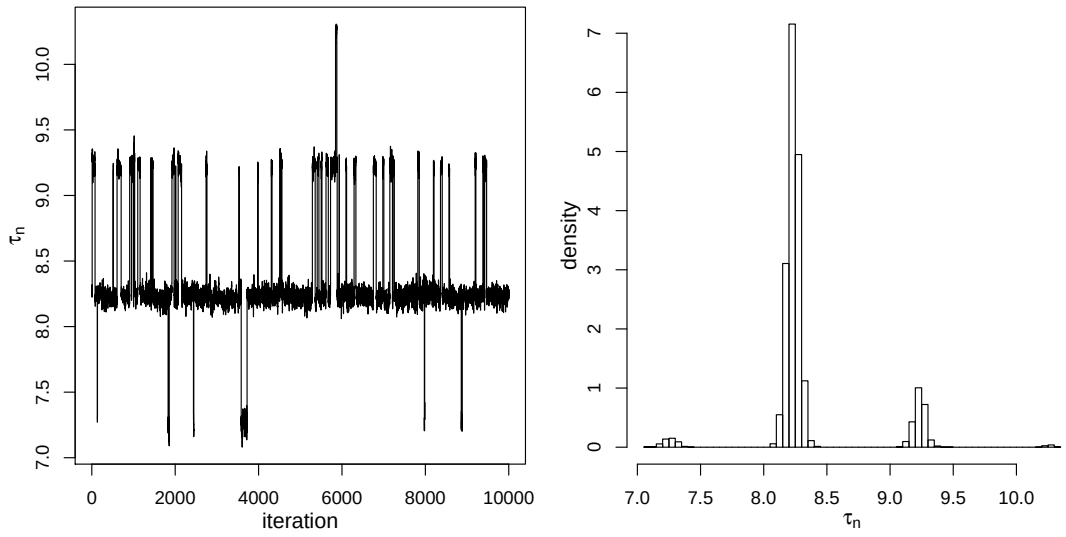

Figure 5.5: Trace plot for the end of the timescale, τ_n, from an MCMC run on the log NGRIP NH_4 signal (1451.1m–1451.6m), with a histogram of its posterior distribution after a burn-in of 1000 iterations.

Figure 5.6 shows the mean posterior model fit to the signal, each sub-plot is described in the paragraphs below.

Figure 5.6 (top): this shows $\boldsymbol{x}$ as circular points. The posterior mean reconstruction, $\hat{\boldsymbol{x}} = E(\boldsymbol{\alpha}\sin(2\pi\boldsymbol{\tau}) + \boldsymbol{\beta})$ where the chain was exploring a cycle count of eight years is shown as a dotted line; for nine years as a dashed line; and where

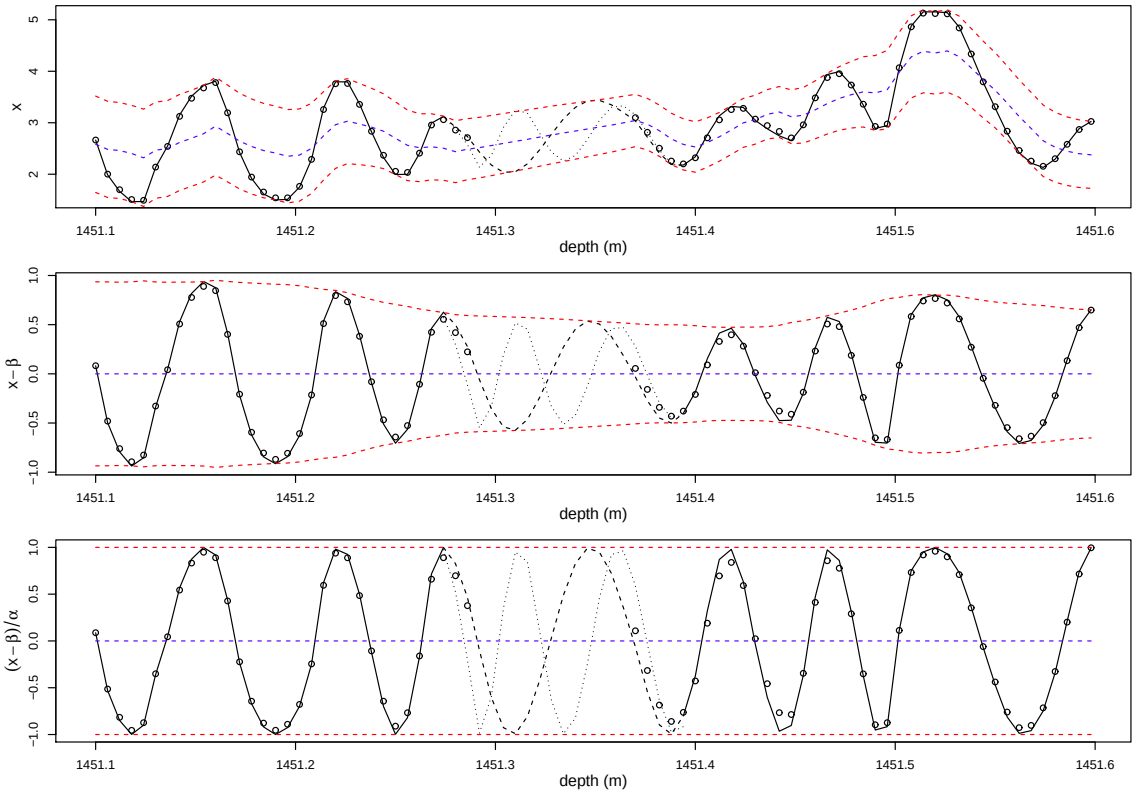

Figure 5.6: The posterior mean model fit to the log NGRIP NH_4 ($\boldsymbol{x}$) signal (1451.1m–1451.6m), thinned to 6mm resolution, calculated from an MCMC run of 10,000 iterations with a burn-in of 1000. **(top)** $\boldsymbol{x}$ as circular points, the reconstruction of $\boldsymbol{x}$ with eight years as a dotted line and with nine years as a dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction with eight years as a dotted line and with nine years as a dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction with eight years as a dotted line and with nine years as a dashed line.

these agree as a solid line. These mean reconstructions are close fits to $\boldsymbol{x}$; the posterior mean for σ is 0.013. $\hat{\boldsymbol{\beta}}$, the overall posterior mean reconstruction for $\boldsymbol{\beta}$ is shown as a dashed blue line. $\hat{\boldsymbol{\beta}} \pm \hat{\boldsymbol{\alpha}}$ is shown as two red dashed lines.

Figure 5.6 (middle): this shows the de-trended signal $\boldsymbol{x} - \hat{\boldsymbol{\beta}}$ as circular points, $E(\boldsymbol{\alpha}\sin(2\pi\boldsymbol{\tau}))$ is shown as a dotted line. Again this is dotted for a cycle count of eight, dashed for nine and solid where they agree. $\hat{\boldsymbol{\alpha}}$, the overall posterior mean reconstruction for $\boldsymbol{\alpha}$ is shown as a dashed red line.

Figure 5.6 (bottom): this shows the de-trended and normalised signal $(\boldsymbol{x} - \hat{\boldsymbol{\beta}})/\hat{\boldsymbol{\alpha}}$ as circular points, $\sin(2\pi\hat{\boldsymbol{\tau}})$ is shown as before.

The equivalent plot showing the fit of the starting values for the chain is included in the appendix as Figure B.10. The starting values calculated for the hyperparameters were: $\sigma = 0.121$, $\sigma_\alpha = 0.053$, $\sigma_\beta = 0.093$ and $\lambda = 34.2$.

Trace plots for hyperparameters σ, σ_α, σ_β, λ and ψ are provided in the appendix as Figure B.11. The acceptance rate for ψ was 31.4%. The posterior distributions for $\boldsymbol{\alpha}$ (top) and $\boldsymbol{\beta}$ (bottom) are provided as Figure 5.7, these are in the form of density heat plots. Note that $\boldsymbol{\alpha}$ has a lower bound of $C_\alpha = 0.21$ whereas $\boldsymbol{\beta}$ is only constrained by the signal. There is clearly more uncertainty in these process during the stretch of missing values.

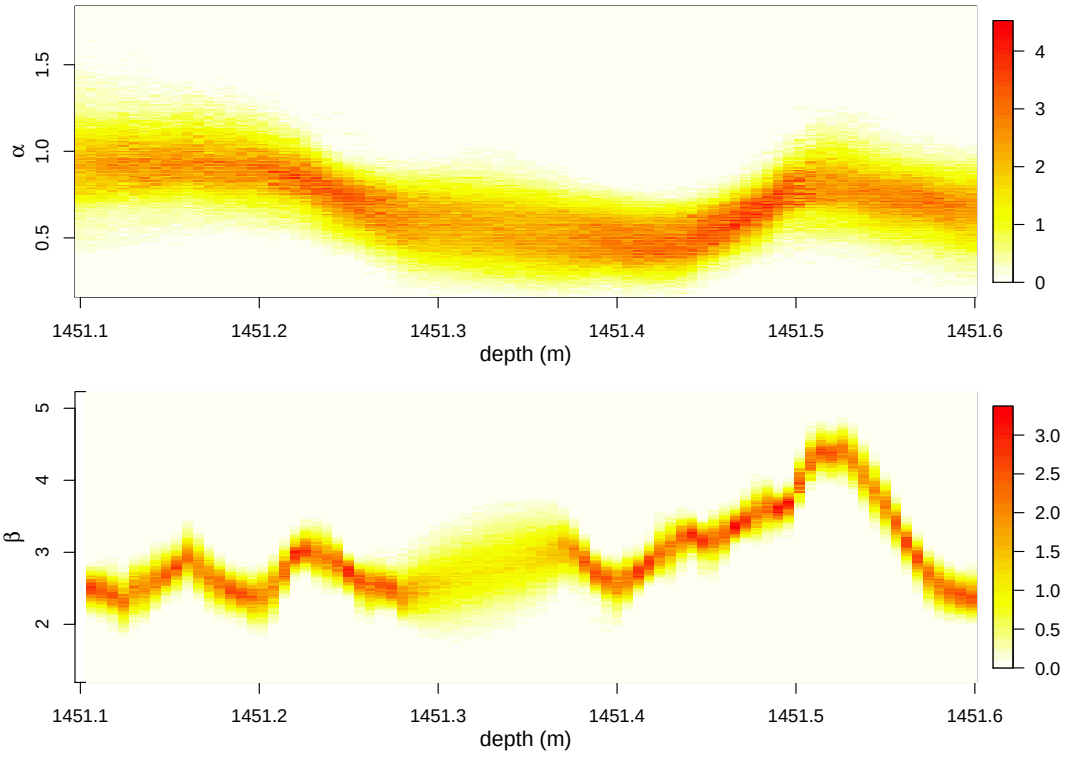

Figure 5.7: Density heat plots of the posterior distributions for the mean and amplitude processes, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, against depth, from the $10,000$ iteration MCMC run on the log NGRIP NH_4 signal between 1451.1m and 1451.6m.

5.7.3 Example 3 – switching

Here we fit the model to a section of log transformed NGRIP NH_4 signal between 1458.7m and 1459.5m, $\boldsymbol{x}$. This section has a large peak at around 1459.2m with smaller peaks either side. This sudden change in amplitude causes the model to switch and, therefore, uncertainty in the cycle count. Due to these low amplitude cycles, in this case we have $C_\alpha = 0.095$. We run the algorithm for $20,000$ iterations.

A trace plot for the timescale at the end of the signal, τ_n, is provided as Figure 5.8, a histogram of its posterior distribution is also included. The chain explored three possible cycle counts: 12, 13 and 14 years; with posterior probabilities of 17.2%, 47.4% and 35.4% respectively. The posterior mean for the decimal part of $\boldsymbol{\tau}$, $\boldsymbol{\tau} - \lfloor \boldsymbol{\tau} \rfloor$, is 0.54, with a 95% PI of $(0.46, 0.62)$.

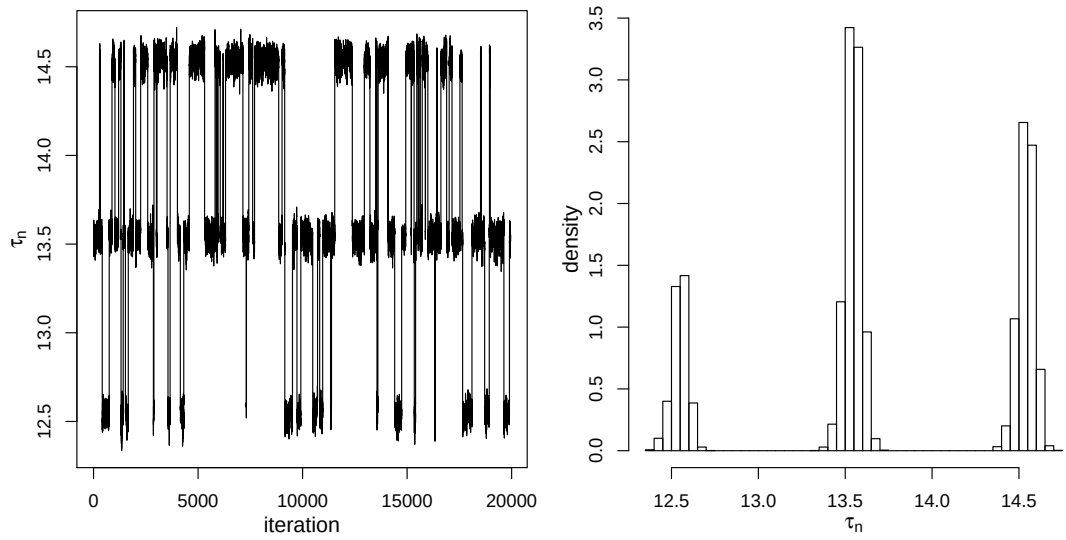

Figure 5.8: A trace plot for the end of the timescale, τ_n, from an MCMC run on the log NGRIP NH_4 signal (1458.7m–1459.5m), with a histogram of its posterior distribution after a burn-in of 1000 iterations.

Figures 5.9 and 5.10 show the posterior mean model fit to the signal over the iterations where the cycle count was 12 and 14 respectively. The equivalent plots showing the fit: for the starting values of the chain (Figure B.12) and where the cycle count was 13 (Figure B.13) are included in the appendix.

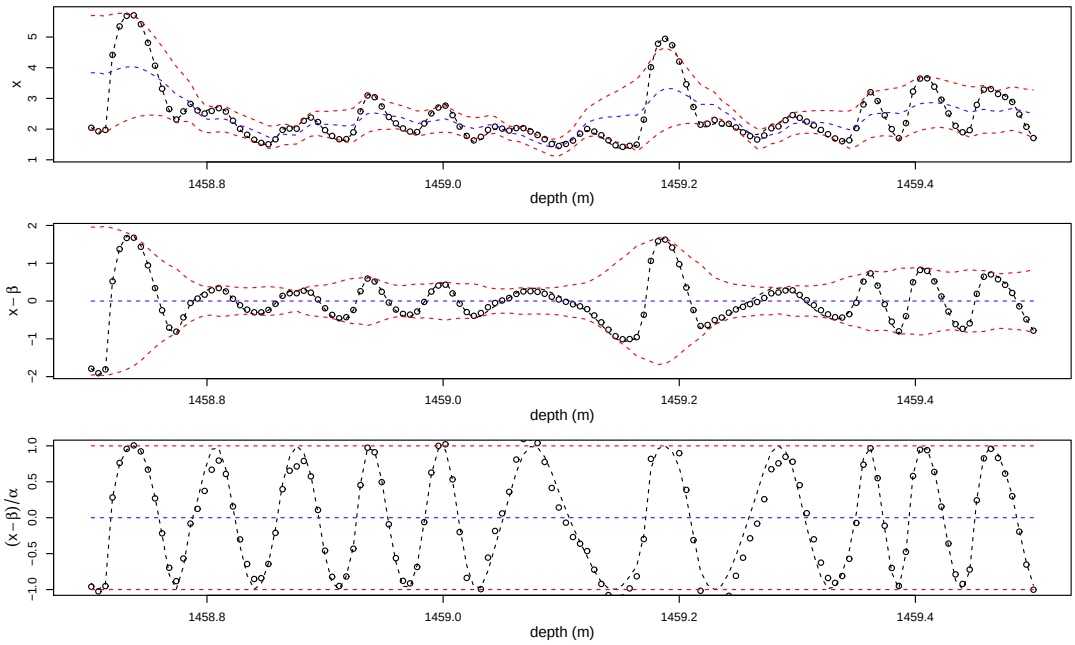

Figure 5.9: The posterior mean model fit to the log NGRIP NH_4 ($\boldsymbol{x}$) signal (1458.7m–1459.5m), thinned to 6mm resolution, over iterations where the cycle count was 12. Calculated from an MCMC run of 10,000 iterations with a burn-in of 1000. **(top)** $\boldsymbol{x}$ as circular points, the reconstruction of $\boldsymbol{x}$ as a black dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

The fit with 12 cycles looks a little unlikely at first sight, however cycles of a length similar to that at depth 1459.1m are not uncommon in the signal. The amplitude is so low in this part of the signal that the mean is relatively stable. If such reconstructions are deemed very unlikely, a stronger prior could be put on the scale parameter for the prior distribution of the timescale.

We include as Figure 5.11 a zoomed in, to between 1459.1m and 1459.4m, stretch of the signal to illustrate how cycles are added and removed by the mean process, $\boldsymbol{\beta}$, 'crossing' the signal. The left plot is the 14 cycle case and the right plot is the mean posterior fit with 13 cycles.

Trace plots for hyperparameters σ, σ_α, σ_β, λ and ψ are provided in the appendix as Figure B.14. The acceptance rate for ψ was 28.9%. The starting values

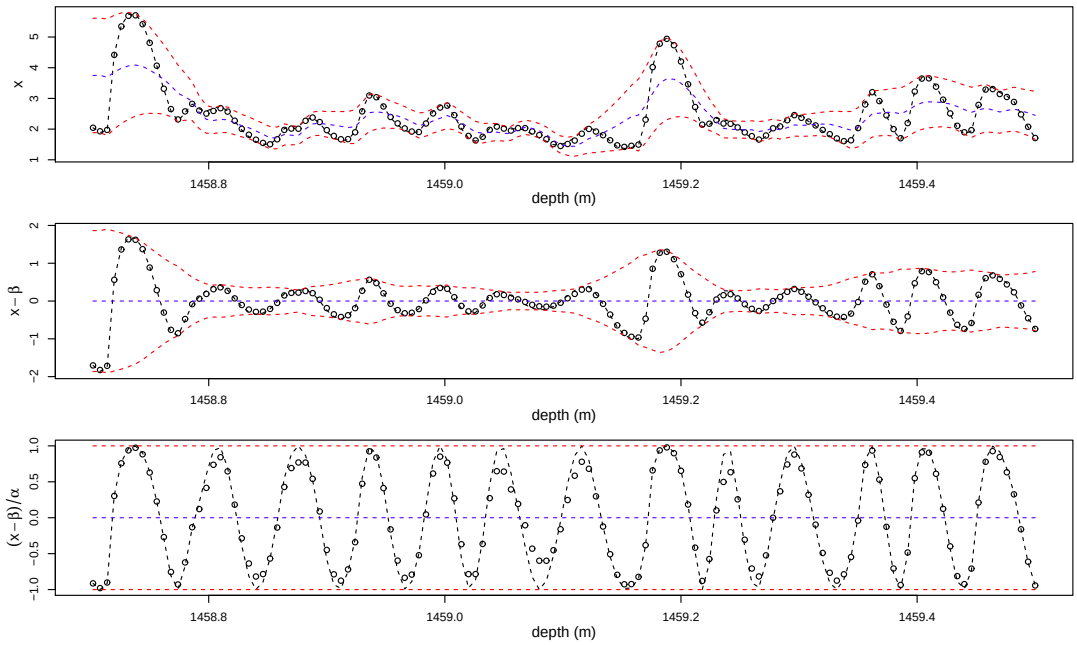

Figure 5.10: The posterior mean model fit to the log NGRIP NH_4 ($\boldsymbol{x}$) signal (1458.7m–1459.5m), thinned to 6mm resolution, over iterations where the cycle count was 14. Calculated from an MCMC run of 10,000 iterations with a burn-in of 1000. **(top)** $\boldsymbol{x}$ as circular points, the reconstruction of $\boldsymbol{x}$ as a black dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

calculated for the hyperparameters were: $\sigma = 0.112$, $\sigma_\alpha = 0.075$, $\sigma_\beta = 0.111$ and $\lambda = 30.6$.

5.8 Conclusions

We have introduced a model for raw or log transformed annually cyclic ice core signals, with their latent timescale as a parameter. The MCMC algorithm used to fit this model samples from combinations of correlated mean and amplitude processes, jointly with the timescale, to recreate the signal under the model and prior assumptions. An automated method for choosing starting values for all model parameters, using the theory developed in Chapters 2 and 3, along with a universal choice of prior assumptions allow this method to be fully automated –

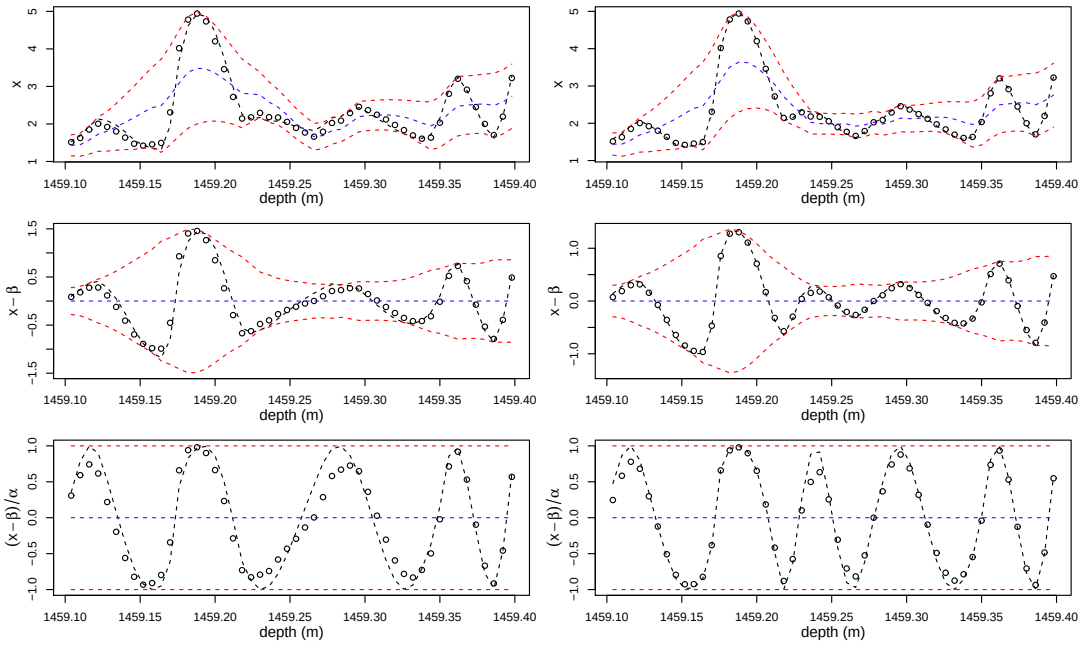

Figure 5.11: The posterior mean model fit to the log NGRIP NH_4 ($\boldsymbol{x}$) signal (1458.7m–1459.5m), thinned to 6mm resolution, zoomed in to (1459.1m–1459.4m). Calculated from an MCMC run of 10,000 iterations with a burn-in of 1000. In the left plots the reconstructions are calculated over iterations where the cycle count was 14; on the right, where there were 13 cycles. **(top)** $\boldsymbol{x}$ as circular points, the reconstruction of $\boldsymbol{x}$ as a black dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

ready to use on any new signal without tuning.

The updating scheme we have developed achieves much better acceptance rates than if vanilla Metropolis-Hastings scheme was used to fit the model – improving the performance of the algorithm.

The examples provided illustrate how the model can be fit using natural prior distributions, providing a distribution function for the number of cycles contained in a subsection of signal. The prior distribution on the shape of the timescale increments could be elicited and used to represent the prior beliefs of an expert in the field of ice core dating. All other hyperparameters are assigned improper, uninformative, prior distributions.

In the following chapters we discuss how the model can be extended into a multivariate framework, and how it can be used on longer sections of signal.

Chapter 6

Multivariate MCMC - raw signals

6.1 Introduction

In this chapter we extend the univariate theory that was introduced in Chapter 5 for use on multiple raw or log transformed annually cyclic ice-core signals. We first discuss how to fit the model to any number of signals simultaneously, each sharing the same timescale. We then explore the bivariate case using examples on the ammonium and calcium signals from the NGRIP ice-core. In Chapter 7 we show how this model can be applied to longer sections of signal.

6.2 The model

To fit a single timescale $\boldsymbol{\tau}$ simultaneously to J annually cyclic ice-core signals, $\boldsymbol{x}_j$ $j \in (1, \ldots, J)$, we propose the model

$$x_{i,j} = \alpha_{i,j} f_j(2\pi(\tau_i + \rho_j)) + \beta_{i,j} + \epsilon_{i,j}$$

for $i \in (1, \ldots, n)$ and where $\rho_1 = 0$ and $\rho_j,\ j \in (2, \ldots, J)$ are the phase differences between the signal j and signal 1. This model can accommodate different

cycle shapes, f_j, for each signal. Each signal also has its own amplitude and mean processes $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ and errors $\boldsymbol{\epsilon}_j$, which are assumed independent between signals. We again incorporate independent normally distributed errors where: $\epsilon_{i,j} \sim \mathrm{N}(0, \sigma_j^2)$.

Although the timescale, $\boldsymbol{\tau}$, is fitted using information from all of the signals, for this parametrisation it is tied directly to the cycles in signal $\boldsymbol{x}_1$. It can be transformed to correspond to any of the other signals by adding the relevant phase difference, ρ_j. It is convenient to tie the timescale to a single strong signal for the theory developed in Chapter 7, where long sections of signal are broken down into subsections to run the MCMC, and then rejoined.

We assume that all $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$, $j \in (1, \ldots, J)$, are independent of each other and of the timescale $\boldsymbol{\tau}$, each process having prior structures as discussed in Chapters 4 and 5, with their own hyperparameters:

$$\alpha_{i+1,j} - \alpha_{i,j} \sim \mathrm{N}(0, \sigma_{\alpha,j}^2), \quad \beta_{i+1,j} - \beta_{i,j} \sim \mathrm{N}(0, \sigma_{\beta,j}^2).$$

6.3 Fitting $\boldsymbol{\tau}$, $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ via MCMC

We use a modified version of the algorithm introduced in Section 5.3 to sample from the joint posterior distribution of $\boldsymbol{\tau}$, $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$, $j \in (1, \ldots, J)$. The parameters are again updated in intervals of m data points, denoted I, as introduced in Section 4.3.2.

A single update can be broken down into three steps. We firstly propose $\boldsymbol{\tau}_I$ from $q(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})$ as explained in Section 4.3. Secondly we accept or reject the proposal; under the assumption that $\boldsymbol{\tau}$, $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ are independent of each other the acceptance probability reduces to

$$\frac{p(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I})q(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})}{p(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})q(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I})} \prod_j \frac{p(\boldsymbol{x}_{I,j}|\boldsymbol{\tau}'_I, \rho_j, \boldsymbol{\alpha}_{-I,j}, \boldsymbol{\beta}_{-I,j})}{p(\boldsymbol{x}_{I,j}|\boldsymbol{\tau}_I, \rho_j, \boldsymbol{\alpha}_{-I,j}, \boldsymbol{\beta}_{-I,j})}.$$

The $p(\boldsymbol{x}_{I,j}|\boldsymbol{\tau}_I, \rho_j, \boldsymbol{\alpha}_{-I,j}, \boldsymbol{\beta}_{-I,j})$ terms correspond to those from the univariate case, as discussed in Section 5.3, with the timescale for signals $j \in (2, \ldots, J)$ transformed by adding their phase difference, ρ_j, to the timescale.

For non-switching updates the ratio

$$\frac{p(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I})q(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})}{p(\boldsymbol{\tau}_I|\boldsymbol{\tau}_{-I})q(\boldsymbol{\tau}'_I|\boldsymbol{\tau}'_{-I})}$$

cancels; for switching updates, this terms simplifies exactly analogously to the univariate case – see Section 4.3.3.

If the update is accepted we then independently for $j \in (1, \ldots J)$ draw $\boldsymbol{\alpha}'_{I,j}$ and $\boldsymbol{\beta}'_{I,j}$ from their joint distributions conditioned on $\boldsymbol{x}_{I,j}$, ρ_j, $\boldsymbol{\tau}'_I$, $\boldsymbol{\alpha}_{-I,j}$ and $\boldsymbol{\beta}_{-I,j}$ as discussed in Section 5.3 with $\boldsymbol{s}$ replaced by $\boldsymbol{s}_j$ in the equations, where $s_{i,j} = \sin(2\pi(\tau_i + \rho_j))$.

6.4 Updating the hyperparameters

ψ and λ are updated as explained in Section 5.4, with the same prior distribution. The signal-specific hyperparameters σ_j, $\sigma_{\alpha,j}$ and $\sigma_{\beta,j}$ are given uninformative conjugate Inverse-Gamma priors, with shape and scale equal to 1/1000, and are sampled via Gibbs steps independently with respect to their signal $\boldsymbol{x}_j$.

6.5 Bivariate case

In the remainder of this chapter and in the example runs (Section 6.6) we focus on the bivariate case of the model and algorithm. We assume that the information held in the strongest two signals, in terms of seasonality, measured from a given ice-core will usually be sufficient to obtain an accurate chronology distribution. This is the case for the NGRIP data that we have available; where two signals, ammonium and calcium, have clear cycles that correspond much more strongly to the GICC05 chronology (see Section 1.2) than the others.

If the use of more than two signals is necessary, the methodology described above could be coded up to accommodate further signals. Another possibility is to reduce the set of cyclic signals to the bivariate case via principal component analysis (PCA), avoiding a large increase in computational cost, see Section 6.5.5.

For notational convenience we refer to the two signals as $\boldsymbol{x}$ and $\boldsymbol{y}$ here, rather than $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ as above. The model simplifies to:

$$x_i = \alpha_{i,x} \sin(2\pi\tau_i) + \beta_{i,x} + \epsilon_{i,x}$$

$$y_i = \alpha_{i,y} \sin(2\pi\tau_i + \rho) + \beta_{i,y} + \epsilon_{i,y}$$

where ρ is the phase difference between signals $\boldsymbol{x}$ and $\boldsymbol{y}$.

6.5.1 Updating the phase difference ρ

The starting value for ρ is approximated as discussed in Section 3.3.1 using

$$\rho_0 = \cos^{-1}(cor(\boldsymbol{x}', \boldsymbol{y}'))$$

where $\boldsymbol{x}'$ and $\boldsymbol{y}'$ are the standardised $\boldsymbol{x}$ and $\boldsymbol{y}$ signals, using the theory of Chapter 2 and an appropriate choice of β.

ρ is updated once per iteration via a Metropolis-Hastings step with Gaussian proposals and assigned an informative prior,

$$\rho \sim \mathrm{N}(\mu_\rho, \sigma_\rho^2),$$

for some prior choice of μ_ρ and σ_ρ. One option would be to use the starting value estimate for the phase difference as the mean of this distribution, $\mu_\rho = \rho_0$. Although this is a useful choice in terms of automation, as no user input is required, it is not Bayesian. We suggest a value of around 0.03 for σ_ρ, as the phase difference of the two signal signals is likely to be known within a quarter cycle (in-phase or out-of-phase).

6.5.2 Starting values

Starting values for $\boldsymbol{\alpha}_x$ and $\boldsymbol{\beta}_x$ are set as $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$ from the standardisation process of signal $\boldsymbol{x}$ described in Chapter 2; similarly for $\boldsymbol{y}$. The starting value for $\boldsymbol{\tau}$ is calculated from the 'quadrant' bivariate classification algorithm of Section 3.3.3, using a scheme that is analogous to that explained in Section 4.4. Given its prior, the starting value for ψ is always taken as $\psi_0 = 3$. All of the other hyperparameters of the model are updated via Gibbs steps – their starting values can be obtained by updating them with respect to the starting values for $\boldsymbol{\tau}$, $\boldsymbol{\alpha}_x$, $\boldsymbol{\beta}_x$, $\boldsymbol{\alpha}_y$ and $\boldsymbol{\beta}_y$.

6.5.3 The MCMC algorithm

For each iteration of the MCMC algorithm:

1. choose ℓ
2. for a sample of A intervals, drawn from all possible intervals of length ℓ without repetition, propose: $\boldsymbol{\tau}'_I \mid \boldsymbol{\tau}_{-I}, \lambda, \psi$
   - if accepted draw

$$\boldsymbol{\alpha}'_{I,x}, \boldsymbol{\beta}'_{I,x} \mid \boldsymbol{\tau}'_I, \boldsymbol{\alpha}_{-I,x}, \boldsymbol{\beta}_{-I,x}, \boldsymbol{x}_I, \sigma_{\alpha,x}, \sigma_{\beta,x}, \sigma$$

$$\boldsymbol{\alpha}'_{I,y}, \boldsymbol{\beta}'_{I,y} \mid \boldsymbol{\tau}'_I, \rho, \boldsymbol{\alpha}_{-I,y}, \boldsymbol{\beta}_{-I,y}, \boldsymbol{x}_I, \sigma_{\alpha,y}, \sigma_{\beta,y}, \sigma$$

3. choose $p \sim U(0,1)$
   - if $p < 1/2$:
     - choose ℓ^+
     - for a sample of B intervals, drawn from all possible intervals of length ℓ^+ without repetition, propose a cycle addition update

- if $p \geq 1/2$:
  - choose ℓ^-
  - for a sample of B intervals, drawn from all possible intervals of length ℓ^- without repetition, propose a cycle removal update

4. update the hyperparameters and ρ in random order

- update $\rho \mid \boldsymbol{\tau}, \boldsymbol{\alpha}_y, \boldsymbol{\beta}_y, \boldsymbol{y}, \sigma$ (M-H)
- update $\sigma_x \mid \boldsymbol{\tau}, \boldsymbol{\alpha}_x, \boldsymbol{\beta}_x, \boldsymbol{x}$ (Gibbs)
- update $\sigma_y \mid \boldsymbol{\tau}, \boldsymbol{\alpha}_y, \boldsymbol{\beta}_y, \boldsymbol{y}$ (Gibbs)
- update $\lambda \mid \boldsymbol{\tau}, \psi$ (Gibbs)
- update $\psi \mid \boldsymbol{\tau}, \lambda$ (Gibbs)
- update $\sigma_{x,\alpha} \mid \boldsymbol{\alpha}_x$ (M-H)
- update $\sigma_{x,\beta} \mid \boldsymbol{\beta}_x$ (Gibbs)
- update $\sigma_{y,\alpha} \mid \boldsymbol{\alpha}_y$ (M-H)
- update $\sigma_{y,\beta} \mid \boldsymbol{\beta}_y$ (Gibbs)

where the tuning parameters A and B are as described in Section 5.5.

6.5.4 Other possible parametrisations

We considered two alternate parametrisations for the bivariate case. These were both programmed in R and experimented with, yielding timescale distributions that do not differ greatly from the original model on any of the examples below. The parametrisation presented above was chosen as it is convenient to tie the timescale to a single strong signal for the theory developed in Chapter 7.

Anchored $\boldsymbol{\tau}$

Here we fix the start of the timescale by setting $\tau_1 = 0$, and not updating it. This re-defines τ as the elapsed time over the section of signal that is being

analysed. The timescale is no longer fixed to a given signal, and so we assign phase differences to both signals, giving the the model:

$$x_i = \alpha_{i,x} \sin(2\pi\tau_i + \rho_x) + \beta_{i,x} + \epsilon_{i,x}$$

$$y_i = \alpha_{i,y} \sin(2\pi\tau_i + \rho_y) + \beta_{i,y} + \epsilon_{i,y}.$$

In this case ρ_x and ρ_y are updated independently, conditional on their respective signals. This parametrisation, as with the original, easily translates to three or more signals.

Symmetrical phase difference

Another possibility that uses only one phase difference parameter is the model:

$$x_i = \alpha_{i,x} \sin(2\pi\tau_i + \rho) + \beta_{i,x} + \epsilon_{i,x}$$

$$y_i = \alpha_{i,y} \sin(2\pi\tau_i - \rho) + \beta_{i,y} + \epsilon_{i,y}$$

where the joint timescale $\boldsymbol{\tau}$ is now central to the timescales of $\boldsymbol{x}$ and $\boldsymbol{y}$. In this case ρ is updated with respect to both signals. This parametrisation introduces a nice symmetry to the model, but does not easily translate to three or more signals.

6.5.5 Principal component analysis – dimension reduction

Principal component analysis (PCA) is a statistical method in which an orthogonal transformation is used to convert a set of correlated variables into a set of linearly uncorrelated scores (or principal components), see Moore (1981); Jolliffe (2005). We propose that a set of strong annually cyclic ice-core signals can be reduced to two orthogonal, or exactly out-of-phase, 'score signals' that contain most of the information required for dating. The bivariate MCMC model and methodology described above could then be applied. Here we assume that the first two score signals, the ones that account for the most variation in the set of

signals, will represent the seasonality or annual cycle information. The timescale underlying the first and strongest score signal, $\boldsymbol{x}$, could be modelled as a sine wave; the timescale underlying the orthogonal second score signal, $\boldsymbol{y}$, a cosine wave – or vise versa depending on the orientation. In this case the bivariate model is:

$$x_i = \alpha_{i,x} \sin(2\pi\tau_i) + \beta_{i,x} + \epsilon_{i,x}$$

$$y_i = \alpha_{i,y} \cos(2\pi\tau_i) + \beta_{i,y} + \epsilon_{i,y}$$

Performing PCA on the five cyclic signals from the example stretch of the NGRIP ice-core, identified in Section 2.2 using the autocorrelation function (ACF), gives the following loadings for the score signals.

Table 6.1: PCA score loadings for cyclic NGRIP chemistry signals (1440m–1465m)

signal	PC 1	PC 2	PC 3	PC 4	PC 5
NH_4	0.97	0	0	-0.12	-0.18
Ca	0	0.56	-0.38	-0.72	0.15
NO_3	0.21	0	0	0.22	0.95
Na	0	0	0.89	-0.41	0.19
Dust	0	0.83	0.23	0.50	0

In this case the first score signal is calculated using $0.97 \times NH_4 + 0.21 \times NO_3$, and the second using $0.56 \times \text{Ca} + 0.83 \times \text{Dust}$. Sodium is not included in either of the first two scores. The loadings for, or variance described by, each score are respectively: 0.90, 0.84, 0.33, 0.29, 0.25. So most (two thirds) of the variation in the five variables is described by the first two scores.

The two resulting score signals, and their ACFs, look just like the chemistry signals of which they consist, with the same estimated average annual cycle lengths of 61mm. Running the bivariate MCMC algorithm on short sections of the PCA score signals gives very similar results to those when using pairs of individual signals for the NGRIP data.

The PCA score signals can only be calculated when there is complete information for all of the signals. Regular long stretches of missing values in the NO_3 signal

cause it to be inappropriate for dating in this section of the NGRIP ice-core. The dust signal has much more correlated noise than the other chemistry signals, with regular sub-annual cycles. For these reasons we use the ammonium and calcium signals, which are approximately orthogonal, in the following examples.

6.6 Bivariate examples

Here we present three examples of fitting the bivariate version of the model to real data. We use three subsections of the log ammonium ($\boldsymbol{x}$) and log calcium ($\boldsymbol{y}$) signals from the NGRIP ice-core. Both signals are thinned as before, taking every sixth data point. The first two examples are the same stretches of signal that form the examples of Sections 4.6 and 5.7. The third example is a different stretch of data where switching occurs, but is not due to missing values.

The standard deviation for the Metropolis-Hastings updates for ψ and ρ are set at 0.8 and 0.012 respectively. These have been tuned in experimental runs of the algorithm on similar data to achieve an acceptance rate of approximately 20%. In each example we choose $A = n_\ell$ and $B = 12$ which gives good mixing of the switching updates over the $25,000$ iterations, where they occur.

6.6.1 Example 1 – the 'well behaved' section

Here we fit the model jointly to a small 'well behaved' subsection (1454.1m–1454.4m) of the two chemistry signals from the NGRIP core. They both contain no missing values and have very clear seasonality over this depth range. In this case the lower bound for $\boldsymbol{\alpha}_x$ is calculated as $C_{\alpha,x} = 0.11$, as in the univariate example, and $C_{\alpha,y} = 0.14$. Over the $25,000$ iterations none of the proposed switching updates were accepted, giving no uncertainty on the cycle count. Figure 6.1 shows the model fit to the signal, each sub-plot is described below.

Figure 6.1 (top left): this shows $\boldsymbol{x}$ as circular points. The posterior mean reconstruction, $\hat{\boldsymbol{x}} = E(\boldsymbol{\alpha}_x \sin(2\pi\boldsymbol{\tau}) + \boldsymbol{\beta}_x)$, is shown as a dotted line. The pos-

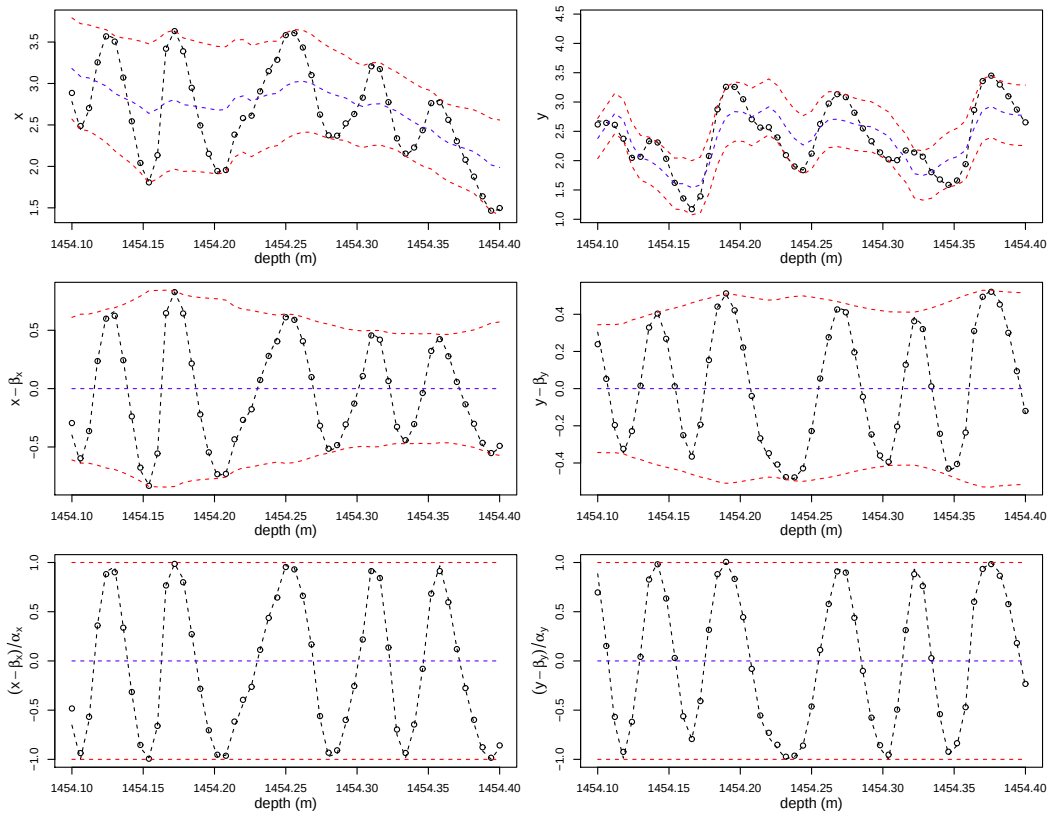

Figure 6.1: The bivariate model fit jointly to the log NGRIP NH_4 ($\boldsymbol{x}$) and Ca ($\boldsymbol{y}$) signals (1454.1m–1454.4m), both thinned to 6mm resolution, calculated from an MCMC run of 20,000 iterations with a burn-in of 2000. The plots for $\boldsymbol{x}$ are on the left; those for $\boldsymbol{y}$ are on the right. **(top)** The signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

terior mean reconstruction for $\boldsymbol{\beta}_x$ is shown as a dotted blue line. $\hat{\boldsymbol{\beta}}_x \pm \hat{\boldsymbol{\alpha}}_x$ is shown as two red dotted lines. The equivalent plot for $\boldsymbol{y}$ is top right, where $\hat{\boldsymbol{y}} = E(\boldsymbol{\alpha}_y \sin(2\pi(\boldsymbol{\tau} + \rho)) + \boldsymbol{\beta}_y)$. The reconstruction is almost exact, the posterior means for σ_x and σ_y were both 0.041.

Figure 6.1 (middle left): this shows the de-trended signal $\boldsymbol{x} - \hat{\boldsymbol{\beta}}_x$ as circular points, $E(\boldsymbol{\alpha}_x \sin(2\pi\boldsymbol{\tau}))$ as a dotted line. $\hat{\boldsymbol{\alpha}}_x$, the posterior mean reconstruction for $\boldsymbol{\alpha}_x$ is shown as a dotted red line. The equivalent plot for $\boldsymbol{y}$ is middle right.

Figure 6.1 (bottom left): this shows the de-trended and normalised signal $(\boldsymbol{x} - \hat{\boldsymbol{\beta}}_x)/\hat{\boldsymbol{\alpha}}_x$ as circular points, $E(\sin(2\pi\boldsymbol{\tau}))$ is shown as a dotted line. The equivalent plot for $\boldsymbol{y}$ is bottom right.

The equivalent plot to Figure 6.1, showing the fit of the starting values for the chain, is included in the appendix as Figure B.15. The starting values calculated for the hyperparameters were: $\sigma_x = 0.156$, $\sigma_y = 0.243$, $\lambda = 25.2$, $\sigma_{x,\alpha} = 0.051$ (as before), $\sigma_{x,\beta} = 0.070$ (as before), $\sigma_{y,\alpha} = 0.067$, $\sigma_{y,\beta} = 0.077$.

For this example we set $l = 1/2$, that is the block update interval lengths, ℓ, are chosen uniformly between 0 and 1/2 years. Figure 6.2 is a scatter plot of ℓ vs. acceptance ratio at each iteration. For very small ℓ, for which we have $m = 1$, the acceptance rate is around 0.6, reducing to around 0.2 for $\ell = 0.5$ - this plot is similar to that for the univariate version of the model (Figure 5.2).

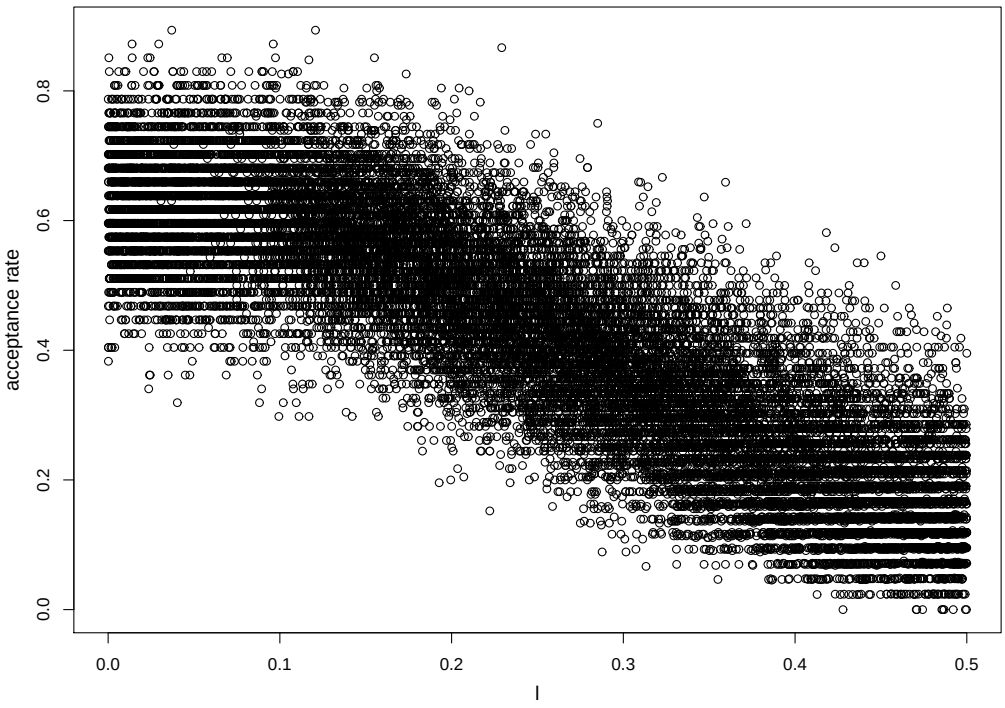

Figure 6.2: Scatter plot of the block update interval length ℓ (in years) against acceptance ratio at each iteration of the 20,000 iteration bivariate MCMC run on the log NGRIP NH_4 and Ca signals (1454.1m–1454.4m).

Trace plots for hyperparameters σ_x, σ_y, λ and ψ are provided in the appendix as Figure B.16 along with histograms of their posterior distribution after a con-

servative burn-in of 2000 iterations. The acceptance rate for ψ was 26.2%. The equivalent plot for $\sigma_{x,\alpha}$, $\sigma_{x,\beta}$, $\sigma_{y,\alpha}$ and $\sigma_{y,\beta}$ is provided as Figure B.17.

A trace plot for the timescale at the end of the signal, τ_n, is provided as Figure 6.3 (top left). A histogram of its posterior distribution is also provided. The posterior mean for τ_n is 5.826 with a 95% PI of $(5.744, 5.910)$; the posterior mean for $\tau_n + \rho$ is 5.540 with a 95% PI of $(5.451, 5.625)$. There was no switching, so this gives a constant posterior cycle count of five.

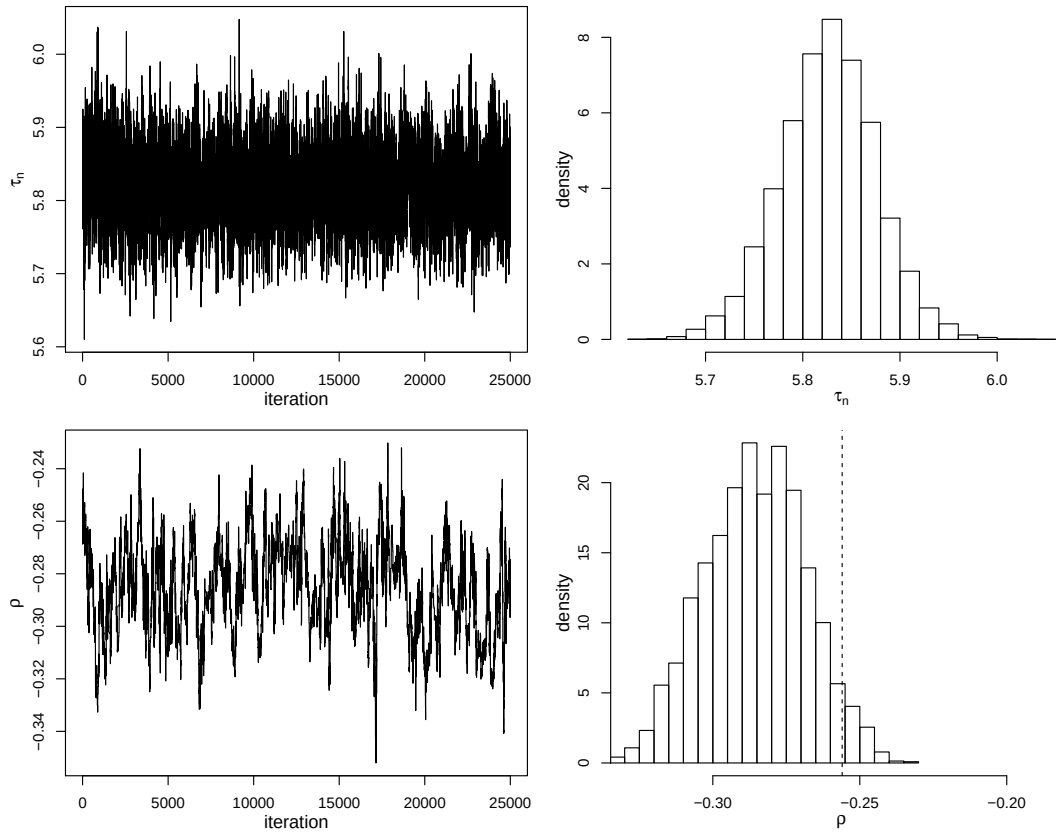

Figure 6.3: Trace plots for the end of the timescale, τ_n, and the phase difference, ρ, from a bivariate MCMC run on the log NGRIP NH_4 and Ca signals (1454.1m–1454.4m), with a histograms of their posterior distributions after a burn-in of 2000 iterations. The starting value for ρ, estimated from the signals, is also shown as a vertical dashed line.

A trace plot for ρ is provided as Figure 6.3 (bottom left) with a histogram of its posterior distribution. The posterior mean for ρ is -0.314, with a 95% PI of $(-0.314, -0.258)$. This suggests that the starting value of $\rho_0 = -0.256$, which was estimated from the standardised signals and is shown on the plot as a vertical

dashed line, was an overestimate. The acceptance rate for ρ was 27.0%.

6.6.2 Example 2 – a stretch of missing values

Here we fit the model to the log transformed and thinned NGRIP NH_4 ($\boldsymbol{x}$) and Ca ($\boldsymbol{y}$) signals between 1451.1m and 1451.6m. This subsection has a large stretch of missing values in both signals: indexes $33 - 45$ for $\boldsymbol{x}$ and $37 - 45$ for $\boldsymbol{y}$, where $n = 51$. $C_{\alpha,x} = 0.21$, as in the corresponding univariate example, and $C_{\alpha,y} = 0.05$.

A trace plot τ_n is provided as Figure 6.4 with a histogram of its posterior distribution. The chain explored three possible cycle counts: seven, eight and nine years; with posterior probabilities of 6.2%, 91.0% and 2.8% respectively.

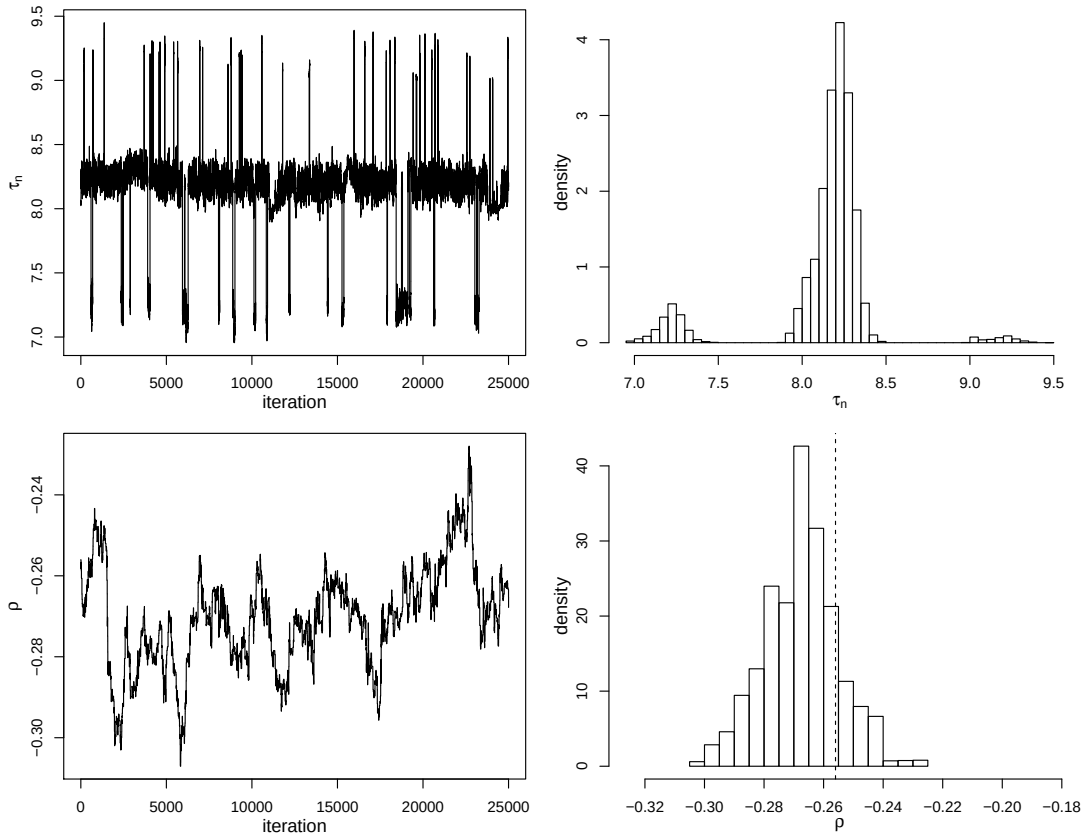

Figure 6.4: Trace plots for the end of the timescale, τ_n, and the phase difference, ρ, from a bivariate MCMC run on the log NGRIP NH_4 and Ca signals (1451.1m–1451.6m), with a histograms of their posterior distributions after a burn-in of 2000 iterations. The starting value for ρ, estimated from the signals, is also shown as a vertical dashed line.

A trace plot for the phase difference, ρ, is provided as Figure 6.4 (bottom left). It is clearly less stable than the corresponding plots for examples 1 and 3 (Figures 6.3 and 6.6) and shows little sign of convergence. The acceptance rate for ρ was only 10.7%. A histogram of its posterior distribution is also provided, the posterior mean for ρ is -0.274 with a 95% PI of $(-0.308, -0.246)$.

Figure 6.5 is analogous to Figure 6.1 which is explained in example 1, above. It shows the posterior mean model fit to the signal for the most likely (eight year) cycle count. Equivalent plots for the seven and nine year reconstructions are provided in the appendix as Figures B.18 and B.19.

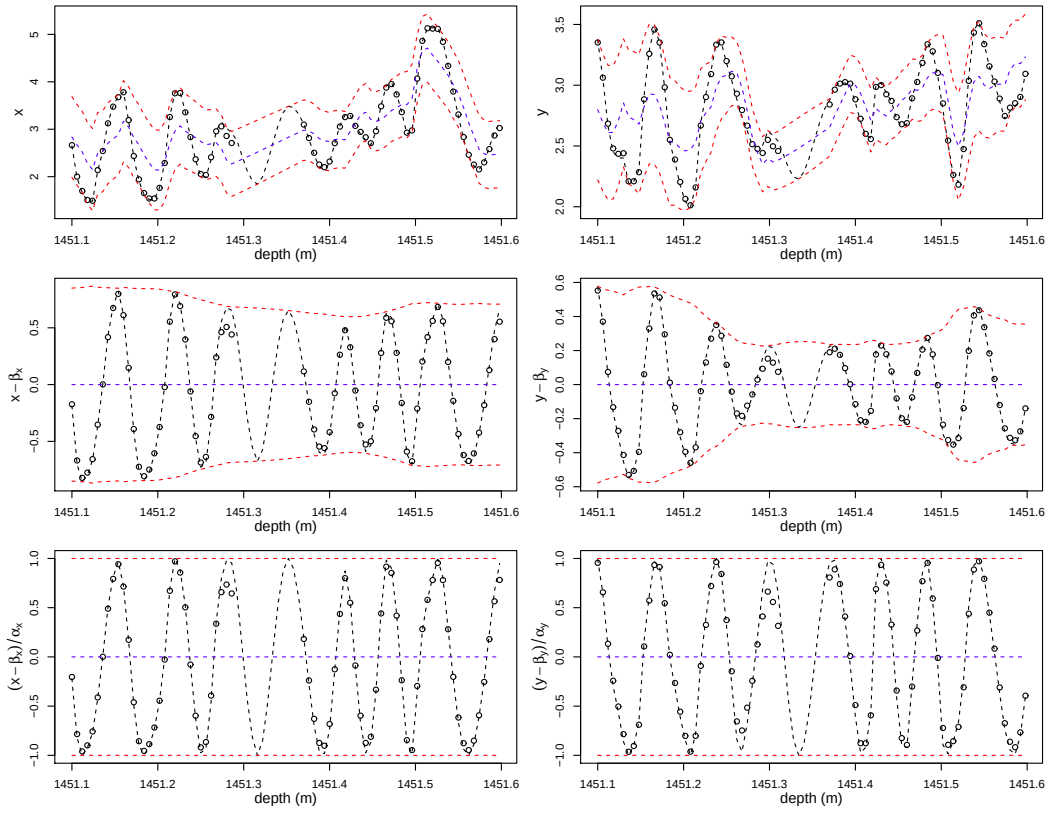

Figure 6.5: The bivariate model fit jointly to the log NGRIP NH_4 ($\boldsymbol{x}$) and Ca ($\boldsymbol{y}$) signals (1451.1m–1451.6m), both thinned to 6mm resolution, calculated from an MCMC run of 20,000 iterations with a burn-in of 2000 where the cycle count was eight. The plots for $\boldsymbol{x}$ are on the left; those for $\boldsymbol{y}$ are on the right. **(top)** The signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

Trace plots for hyperparameters σ_x, σ_y, λ and ψ are provided in the appendix as Figure B.20 along with histograms of their posterior distribution after a conservative burn-in of 2000 iterations. The acceptance rate for ψ was 23.9%. The equivalent plot for $\sigma_{x,\alpha}$, $\sigma_{x,\beta}$, $\sigma_{y,\alpha}$ and $\sigma_{y,\beta}$ is provided as Figure B.21. The starting values calculated for the hyperparameters were: $\sigma_x = 0.237$, $\sigma_y = 0.118$, $\lambda = 37.6$, $\sigma_{x,\alpha} = 0.053$ (as before), $\sigma_{x,\beta} = 0.093$ (as before), $\sigma_{y,\alpha} = 0.037$, $\sigma_{y,\beta} = 0.040$.

6.6.3 Example 3 – switching

Here we fit the bivariate model jointly to a short log transformed and thinned subsection of the NGRIP NH_4 and Ca signals ($\boldsymbol{x}$ and $\boldsymbol{y}$), between 1461.6m and 1462.1m. There is a small stretch of missing values present in both signals over this depth range: indexes $49 - 52$ for $\boldsymbol{x}$, and $50 - 52$ for $\boldsymbol{y}$, $n = 84$. In this case the amplitude lower bounds, calculated from the standardised signals, are $C_{\alpha,x} = 0.29$ and $C_{\alpha,y} = 0.08$.

A trace plot for the timescale at the end of the signal, τ_n, is provided as Figure 6.6 with a histogram of its posterior distribution. The chain explored two possible cycle counts: eight and nine years; with posterior probabilities of 85.3% and 14.7% respectively.

A trace plot for the phase difference, ρ, is provided as Figure 6.6 (bottom left). A histogram of its posterior distribution is also provided with its starting value, calculated from the standardised signals, shown as a vertical dashed line. The posterior mean for ρ is -0.315 with a 95% PI of $(-0.345, -0.290)$, suggesting that ρ_0 was an overestimate. The acceptance rate for ρ was 21.4%.

Figure 6.7 is analogous to Figure 6.1 which is explained in example 1, above. It shows the posterior mean model fit to the signal for the most likely (eight year) cycle count. Equivalent plots for the nine year reconstruction is provided in the appendix as Figure B.22. There was no switching over the short stretch of missing values, the uncertainty in the cycle count comes from the features at around 1461.78m where switching occurred.

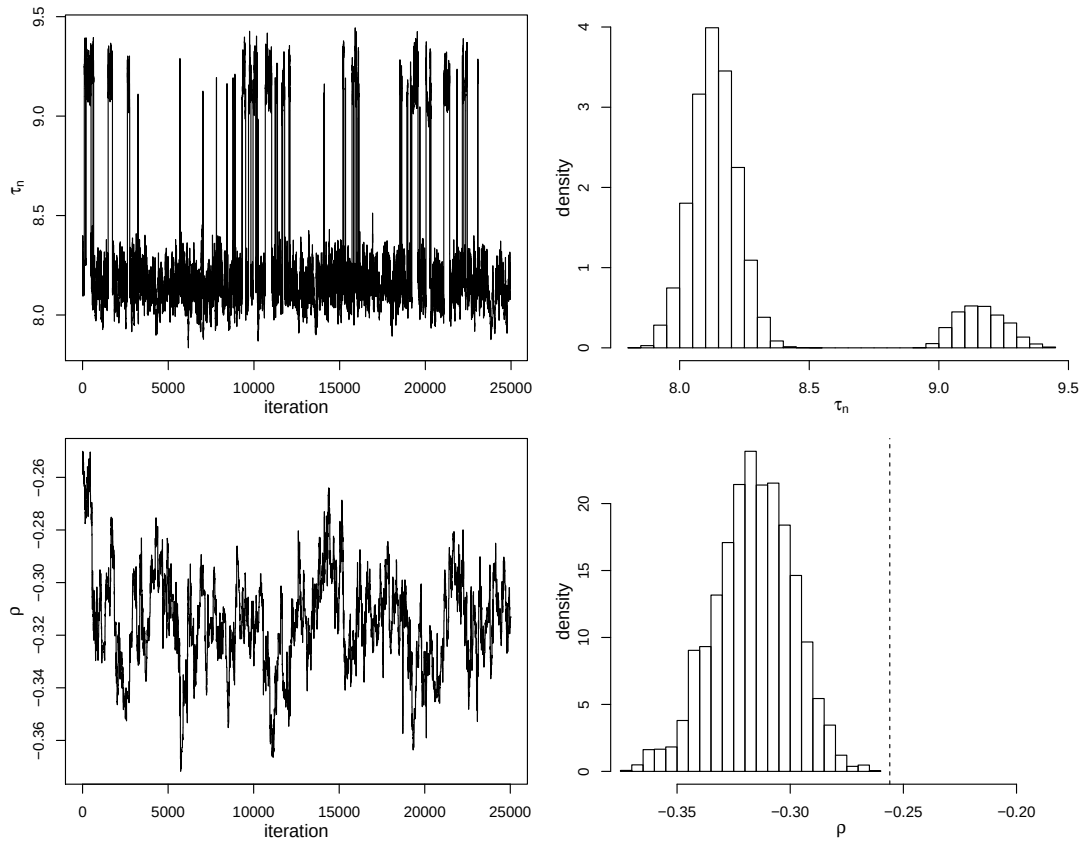

Figure 6.6: Trace plots for the end of the timescale, τ_n, and the phase difference, ρ, from a bivariate MCMC run on the log NGRIP NH_4 and Ca signals (1461.6m–1462.1m), with a histograms of their posterior distributions after a burn-in of 2000 iterations. The starting value for ρ, estimated from the signals, is also shown as a vertical dashed line.

Trace plots for hyperparameters σ_x, σ_y, λ and ψ are provided in the appendix as Figure B.23, along with histograms of their posterior distribution after a conservative burn-in of 2000 iterations. The acceptance rate for ψ was 23.9%. The equivalent plot for $\sigma_{x,\alpha}$, $\sigma_{x,\beta}$, $\sigma_{y,\alpha}$ and $\sigma_{y,\beta}$ is provided as Figure B.24. The starting values calculated for the hyperparameters were: $\sigma_x = 0.560$, $\sigma_y = 0.197$, $\lambda = 36.4$, $\sigma_{x,\alpha} = 0.086$, $\sigma_{x,\beta} = 0.171$, $\sigma_{y,\alpha} = 0.027$, $\sigma_{y,\beta} = 0.034$.

The posterior distributions for the $\boldsymbol{\alpha}_x$, $\boldsymbol{\beta}_x$, $\boldsymbol{\alpha}_y$ and $\boldsymbol{\beta}_y$ processes are provided as Figure 6.8, these are in the form of density heat plots. Note the feature present in the plots for the $\boldsymbol{\beta}$ processes during the stretch of missing values where there is more uncertainty on the mean levels, and also at around 1461.78m where switching occurred.

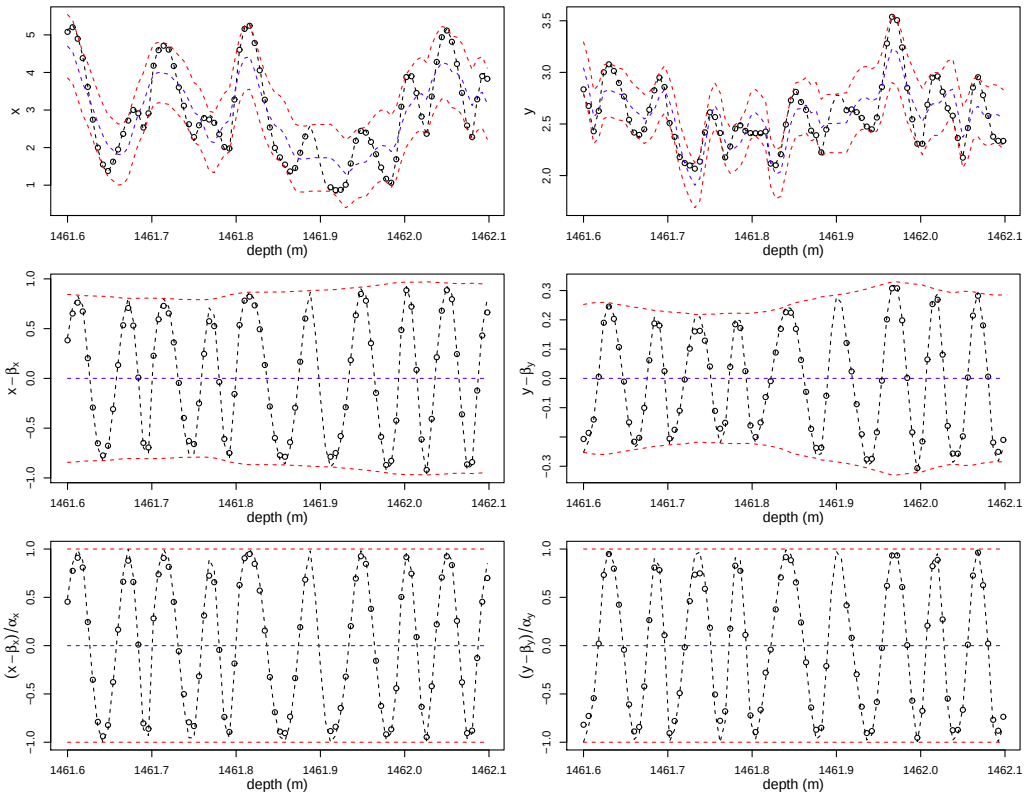

Figure 6.7: The bivariate model fit jointly to the log NGRIP NH_4 ($\boldsymbol{x}$) and Ca ($\boldsymbol{y}$) signals (1461.6m–1462.1m), both thinned to 6mm resolution, calculated from an MCMC run of 20,000 iterations with a burn-in of 2000 where the cycle count was eight. The plots for $\boldsymbol{x}$ are on the left; those for $\boldsymbol{y}$ are on the right. **(top)** The signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

6.7 Conclusions

In this chapter we have shown how the theory introduced in Chapter 5 can be extended for use on multiple raw or log transformed annually cyclic ice-core signals. This allows us to fit one timescale simultaneously using the information from two or more signals. We have given examples for the bivariate case; the settings of this algorithm should be applicable to any other pair of annually cyclic signals, from the same ice-core, that show strong annual seasonality.

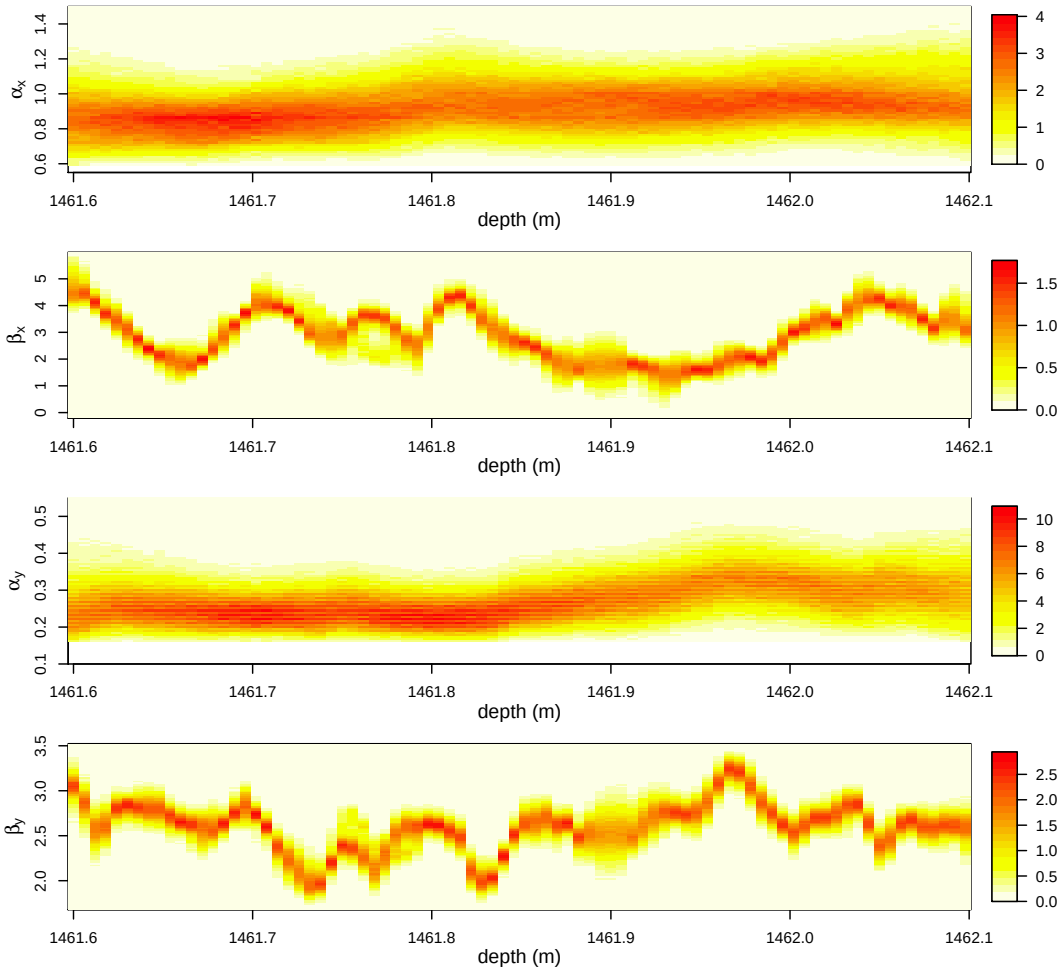

Figure 6.8: Density heat plots of the posterior distributions for the mean and amplitude processes, $\boldsymbol{\alpha}_x$, $\boldsymbol{\beta}_x$, $\boldsymbol{\alpha}_y$ and $\boldsymbol{\beta}_y$, from a bivariate MCMC run on the log NGRIP NH_4 and Ca signals (1461.6m–1462.1m).

Chapter 7

Scaling the MCMC algorithms

7.1 Introduction

In this chapter we discuss how the MCMC methods developed in the previous chapters can be used to fit timescales to longer sections of annually cyclic ice core signals. We provide chronologies for the entire Fletcher and Gomez ice cores; and for the whole example stretch of NGRIP data, using the ammonium and calcium signals both individually and jointly. Although these are still relatively short sections of signal, the theory discussed here could be used to date signals of any length.

7.2 Overview

The MCMC algorithms for use on raw or log signals presented in the previous two chapters have been tuned to mix well when working with signals containing up to 40 cycles. They do not scale very well for use on much longer sections all in one go. This is true both in terms of tuning the algorithm – setting A and B and the standard deviations for the Metropolis-Hastings steps – and for computational efficiency. Note that the 25m section of the NGRIP core we have available contains $25,000$ data points, even after thinning to every sixth point

that is $25,000/6 \times 3 + 5 = 75,005$ parameters for the (simpler) univariate case. Dating a 4km ice core in a single run therefore becomes unwieldy. It is necessary, then, to break the longer signals down into shorter sections on which to run the MCMC, and for this we suggest the following scheme:

1. split the signal into *overlapping* subsections with carefully chosen joining depths;
2. run the MCMC *independently* on each of these sections of signal;
3. join the resulting timescales together using a re-sampling approach.

This method copes with changes in the hyperparameters down the core by fitting them independently to the sections. It is also fully parallelizable, each section can be analysed at the same time. With enough processors available, dating an entire core could take as long as dating the longest subsection.

7.3 Splitting the signal into subsections

Notation We split the signal(s) into K overlapping subsections, defined by start and end depth indexes, b_k and e_k, $k = 1, \ldots, K$. Section k contains $n_k = e_k - b_k + 1$ data points, $\sum_k n_k > n$. We denote section k of signal $\boldsymbol{x}$ as

$$\boldsymbol{x}^{(k)} = x_i, \ i = b_k, \ldots, e_k.$$

In the bivariate case the subsections of $\boldsymbol{y}$ correspond to the same depth indexes:

$$\boldsymbol{y}^{(k)} = y_i, \ i = b_k, \ldots, e_k.$$

We use the same parenthesised superscript notation for the model parameters in each overlapping subsection. We choose depth indexes, $g_k, \ k \in 1 \ldots K-1$, at which to join the timescales from subsections k and $k+1$ back together: $b_{k+1} < g_k < e_k$.

Choosing the cut-offs We first standardise and classify the depth indexes of the signal(s) into 'certain runs' and 'issue runs' using the methods explained in Chapters 2 and 3. The certain runs represent quarter cycles and contain signal where the seasonality is clear, issue runs represent sections of signal where we know less about the cycles and represent an unknown amount of time. We choose b_1 to be the starting depth index of the first certain run and e_K to be the ending depth index of the last certain run. We aim to choose $g_k,\ k = 1, \ldots, K-1$ to be depth indexes that are central to a reasonably long section of well defined seasonality, so it is unlikely that switching will occur at the section boundaries.

We first note all consecutive groups of certain runs of size $2V$, for some chosen $V \in \mathbb{Z}^+$. We then choose the $K-1$ of these groups that split the signal into K sections that contain, as close as possible, the same number of runs. The first of these chosen groups will be the overlap between sections 1 and 2. b_2 is set as the depth at which the first run of this group starts, g_1 is the depth bounding runs V and $V+1$, and e_1 is the depth index at the end of run $2V$. This way, sections 1 and 2 overlap by $V/2$ strong cycles and the joining index is in the center of this overlap, in terms of cycles. Similarly for the other sections.

7.4 The MCMC algorithm

The model parameters for each overlapping section of signal are fit independently, using either the univariate algorithm from Chapter 5 or the quadrant bivariate algorithm from Chapter 4. In each overlapping subsection the algorithms are run for the same number of iterations, with the same A and B parameters that control the ratio of data points to block updates, and block updates to switching updates.

7.4.1 Hyperparameters

The prior distributions for the model parameters have been chosen so that they apply to *any* section of annually cyclic ice core signal. The prior distribution

assigned to the timescale $\boldsymbol{\tau}$ is parametrised in terms of the elapsed time over each depth index δ_i:

$$\delta_i \sim \mathrm{G}(\lambda, \psi),\ i = 2, \ldots, n.$$

The hyperpriors only have direct influence over the shape of this distribution with

$$\psi \sim \mathrm{N}(3, 1),$$

as discussed in Section 5.4. The scale hyperparameter has $p(\lambda) \propto 1$ and can therefore accommodate changes in the average cycle length, between signal sections, along the cores.

The standard deviation hyperparameters for the mean and amplitude processes, σ_β and σ_α, along with the error standard deviation, σ, are all given uninformative inverse gamma priors and are updated via Gibbs steps in each case.

7.4.2 Starting values

The entire signal has already been through the standardisation and classification process to accommodate splitting it into sections. The starting values for $\boldsymbol{\alpha}^{(k)}$ and $\boldsymbol{\beta}^{(k)}$ are set as the relevant sections of $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$, from b_k to e_k. The lower bound on the amplitude process, C_α, is calculated separately for each section.

We know that each section of signal starts and ends at a 'certain run'. It is therefore straightforward to pick starting values for $\boldsymbol{\tau}$ as discussed in Section 4.4. Starting values for the hyperparameters are calculated separately for each section using the starting values for the parameters as discussed in Section 5.6.

7.5 Joining the parameters

Re-joining $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ after fitting the model separately to the K sections is trivial: the posterior sample is twice the size for the overlapping indexes,

$$i = b_{k+1}, \ldots, e_k,\ k = 1, \ldots, K-1.$$

Re-joining $\boldsymbol{\tau}$ requires a little more care. The posterior sample of the timescale between the start of section 1, b_1, and the first joining depth index, g_1, is taken from the MCMC run on the first section: for each iteration

$$\boldsymbol{\tau}_{b_1:(g_1-1)} = \boldsymbol{\tau}^{(1)}_{b_1:(g_1-1)}.$$

The posterior probability density function for the integer part of g_1 is calculated from the posterior sample from section 1. This distribution is combined with the posterior distribution of $\boldsymbol{\tau}^{(2)}$, giving $\boldsymbol{\tau}^{(2)'}$, so that the integer part of $\tau^{(2)'}_{g_1}$ follows that distribution. In this way the information about the cycle count from the switching updates in section 1 is passed to section 2. This is achieved by, at each iteration, subtracting the integer part of $\tau^{(2)}_{g_1}$ from each element of $\boldsymbol{\tau}^{(2)}$ and then adding an integer drawn from the distribution. It is assumed, and should be checked, that the integer part of $\tau^{(2)}_{g_1}$ is constant throughout the posterior sample, after burn-in. This is due to the choice of the joining depth g_1 being central to several certain runs, making switching very unlikely in this part of the signal, and is the case for each of the examples of this chapter. At each iteration of the sample we set

$$\boldsymbol{\tau}_{g_1:(g_2-1)} = \boldsymbol{\tau}^{(2)'}_{g_1:(g_2-1)}.$$

The posterior probability density function for the integer part of g_2 is calculated from the posterior sample of $\boldsymbol{\tau}^{(2)'}$. This distribution is re-sampled onto the posterior distribution of $\boldsymbol{\tau}^{(3)}$, giving $\boldsymbol{\tau}^{(3)'}$, so that the integer part of $\tau^{(3)'}_{g_2}$ follows that distribution. At each iteration we set

$$\boldsymbol{\tau}_{g_2:(g_3-1)} = \boldsymbol{\tau}^{(3)'}_{g_2:(g_3-1)}.$$

This process is continued sequentially until finally setting:

$$\boldsymbol{\tau}_{g_{(K-1)}:e_K} = \boldsymbol{\tau}^{(K-1)'}_{g_{(K-1)}:e_K}.$$

$\boldsymbol{\tau}_{e_K}$ holds information on the cycle count from all preceding sections. An example of how this works for the Fletcher ice core is given at the end of Section 7.6.1

below using Figure 7.3.

7.6 Examples

We provide chronologies for the entire Fletcher and Gomez ice cores; and for the example stretch of the NGRIP core using its ammonium and calcium signals, both individually and jointly.

In each case the algorithm is run for $15,000$ iterations per section. The standard deviation for the Metropolis-Hasings updates of ψ is set at 0.8, which has been tuned in experimental runs of the algorithm to achieve an acceptance rate of approximately 20% in the examples of the previous chapters. For each section we set $A = n_\ell$ and $B = 12$ giving good mixing of the switching updates, where they occur.

7.6.1 The Fletcher ice core

The unequally spaced δ^{18}O signal from the Fletcher core is sampled onto 4cm resolution using linear interpolation for this analysis. This is the modal resolution of the data. The signal is shown in Section 2.3 which describes the standardisation and interpolation process.

We know from the analysis of Section 3.2.9 that there are around 155 years covered by the Fletcher ice core. In this case we choose $K = 5$, and so expect around 31 cycles in each section. We use $V = 8$ to split the signal.

The interpolated signal has $n = 6615$ data points between 2.48m and 81.60m. The univariate classification algorithm with $\nu = 1/2$ outputs 635 runs, of which 598 are classified as certain. The first certain run, run 3, starts at depth index $i = 6$ (2.68m), so we set the start of the first section as $b_1 = 6$. The last certain run, run 633, ends at depth index 1976 (81.48m) so we set the end of the last section as $e_5 = 1976$. There are 33 issue runs and 631 runs in total between these depths.

We want to split the runs as equally as possible into the $K = 5$ distinct groups, so that each group starts and ends with $2V = 16$ consecutive certain runs. The signal can be split into four groups of 126 runs and one of 127 runs: the first comprising runs $3 - 128$, then $129 - 254$, $255 - 380$, $381 - 506$ and $506 - 633$. Adjusting the boundaries to the closest with $V = 8$ certain runs either side gives groups of size 126, 133, 124, 132 and 116: the first comprised of runs $3 - 128$, then $129 - 261$, $262 - 385$, $386 - 517$ and $518 - 633$. The first bounding set of 16 consecutive certain runs consists of runs $121 - 137$. We take the first joining index, $g_1 = 574$ (25.40m), as that which bounds runs 128 and 129; the ending index of section 1, $e_1 = 600$ (26.44m), to be the end of run 137; and the start index of section 2, $b_2 = 539$ (24.00m), to be the start of run 121. Similarly we have:

$$b_1 = 6 \ \ (2.68\text{m}),\ b_2 = 539 \ \ (24.00\text{m}),\ b_3 = 979 \ \ (41.60\text{m}),$$
$$b_4 = 1351 \ \ (56.48\text{m}),\ b_5 = 1689 \ \ (70.00\text{m})$$

$$g_1 = 574 \ \ (25.40\text{m}),\ g_2 = 1007 \ \ (42.72\text{m}),$$
$$g_3 = 1373 \ \ (57.36\text{m})\ g_4 = 1707 \ \ (70.72\text{m})$$

$$e_1 = 600 \ \ (26.44\text{m}),\ e_2 = 1027 \ \ (43.52\text{m}),\ e_3 = 1399 \ \ (58.40\text{m}),$$
$$e_4 = 1723 \ \ (71.36\text{m}),\ e_5 = 1974 \ \ (81.40\text{m})$$

Figure 7.1 below shows the posterior distributions of the hyperparameters for $\boldsymbol{\tau}$ – ψ and λ – as box-plots, one for each section. The thinning of the annual layer thickness down the core can be seen in the systematic reduction in λ along the sections. The distribution of ψ stays relatively stable, except for in section 1 where the cycle lengths are less variable. Equivalent plots for the posterior distributions of the standard deviation hyperparameters σ_α, σ_β and σ are included in the appendix as Figure B.25. These suggest a marked reduction in the variability of

the amplitude process $\boldsymbol{\alpha}$ in the last two sections, with a corresponding increase in the variability of the mean process $\boldsymbol{\beta}$.

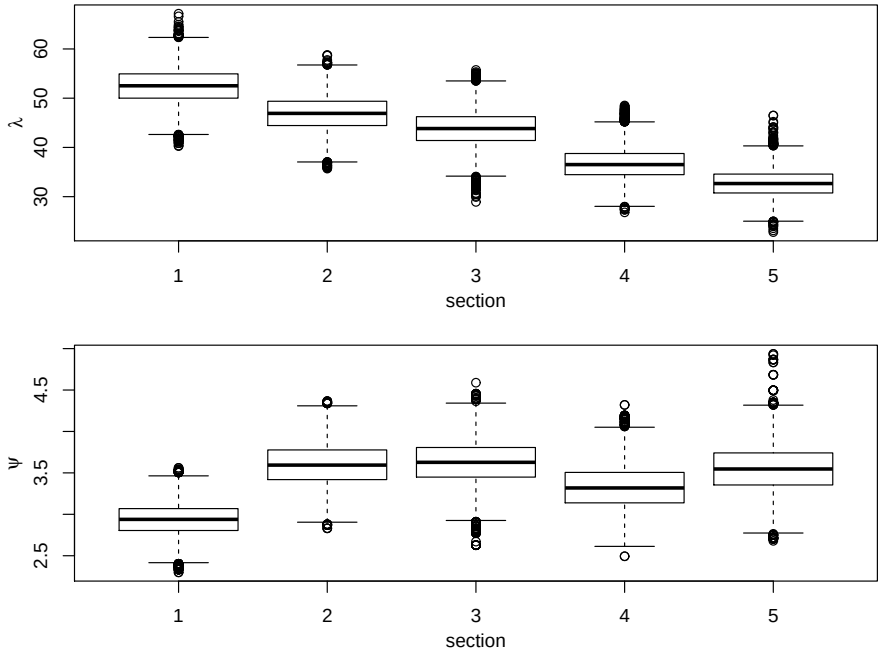

Figure 7.1: Boxplots for the posterior distributions of ψ and λ, the timescale hyperparameters, from independent MCMC runs on the Fletcher δ^{18}O signal after being split into five overlapping subsections containing roughly the same number of cycles.

The posterior distributions for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are provided below as Figure 7.2 in the form of density heat plots. The vertical dotted lines show the joining depths for the five sections. We observe a general decrease in amplitude with depth with a marked decrease in variation for the last two sections. There is no obvious trend in the mean of $\boldsymbol{\beta}$, however the last section displays more variability.

Figure 7.3 is included to illustrate the re-sampling process that is used to join together the posterior distributions of $\boldsymbol{\tau}$, evaluated for the five overlapping sections. This plot is described in detail in the four paragraphs below. Trace plots for the timescale evaluated at g_k for sections $k \in (1, \ldots, 4)$, and e_5 for section 5, are provided as Figure B.26 in the appendix for reference. From these plots we chose to use a burn-in of 4000 iterations for each section.

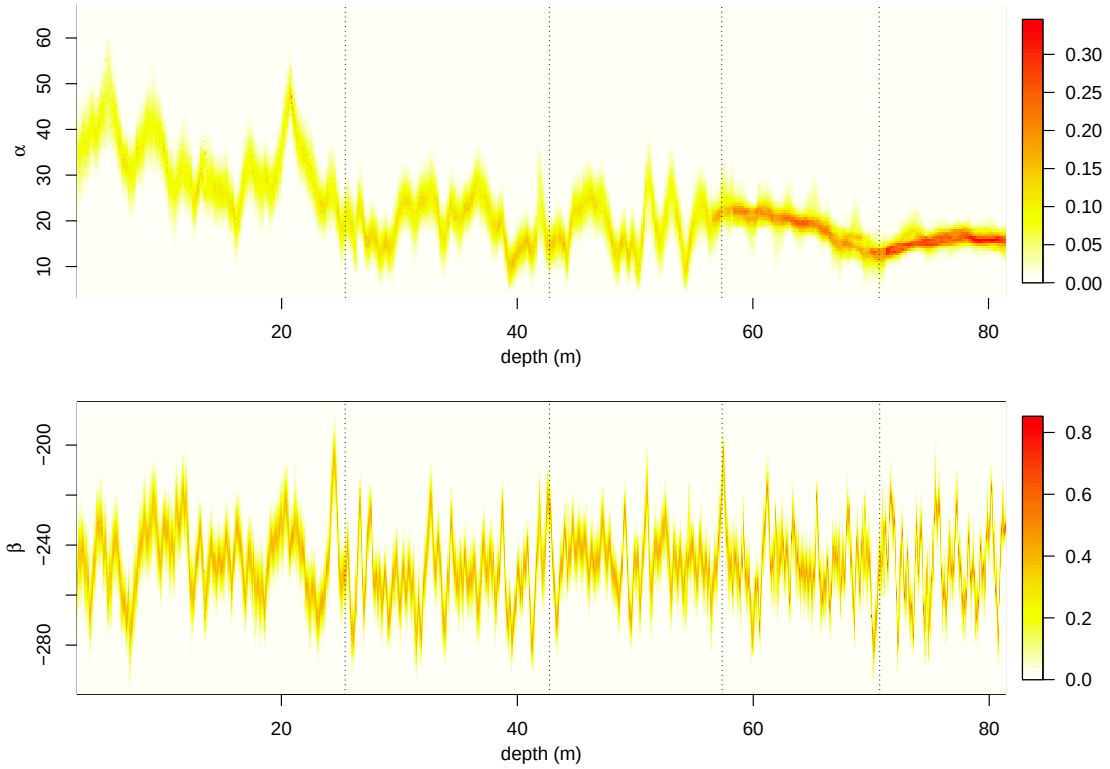

Figure 7.2: Density heat plots of the posterior distributions of the amplitude and mean processes, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, from independent MCMC runs on the Fletcher δ^{18}O signal after being split into five overlapping subsections containing roughly the same number of cycles. The vertical dotted lines show the joining depths for the five sections.

Figure 7.3: the first row of plots. The first plot shows the probability distribution for the number of cycles at depth index g_1, calculated from the MCMC run on section 1. No switching occurred in that run (after burn-in), giving a count of 31 years with a probability of 1. The second plot shows the posterior distribution of $\tau_{g_1}^{(2)}$ as a histogram. With the setting of $V = 8$ we would expect two cycles to have elapsed by this index of section 2, plus the starting time-of-year: a 95% PI for $\tau_{b_2}^{(2)}$ is $(0.64, 0.78)$. The third and fourth plots show the posterior distributions of $\tau_{g_1}^{(2)}$ and $\tau_{g_2}^{(2)}$ after re-sampling (in this case adding 29 at each iteration) so that the integer part of $\tau_{g_1}^{(2)}$ follows the distribution of plot 1: giving $\tau_{g_1}^{(2)\prime}$ and $\tau_{g_2}^{(2)\prime}$ respectively.

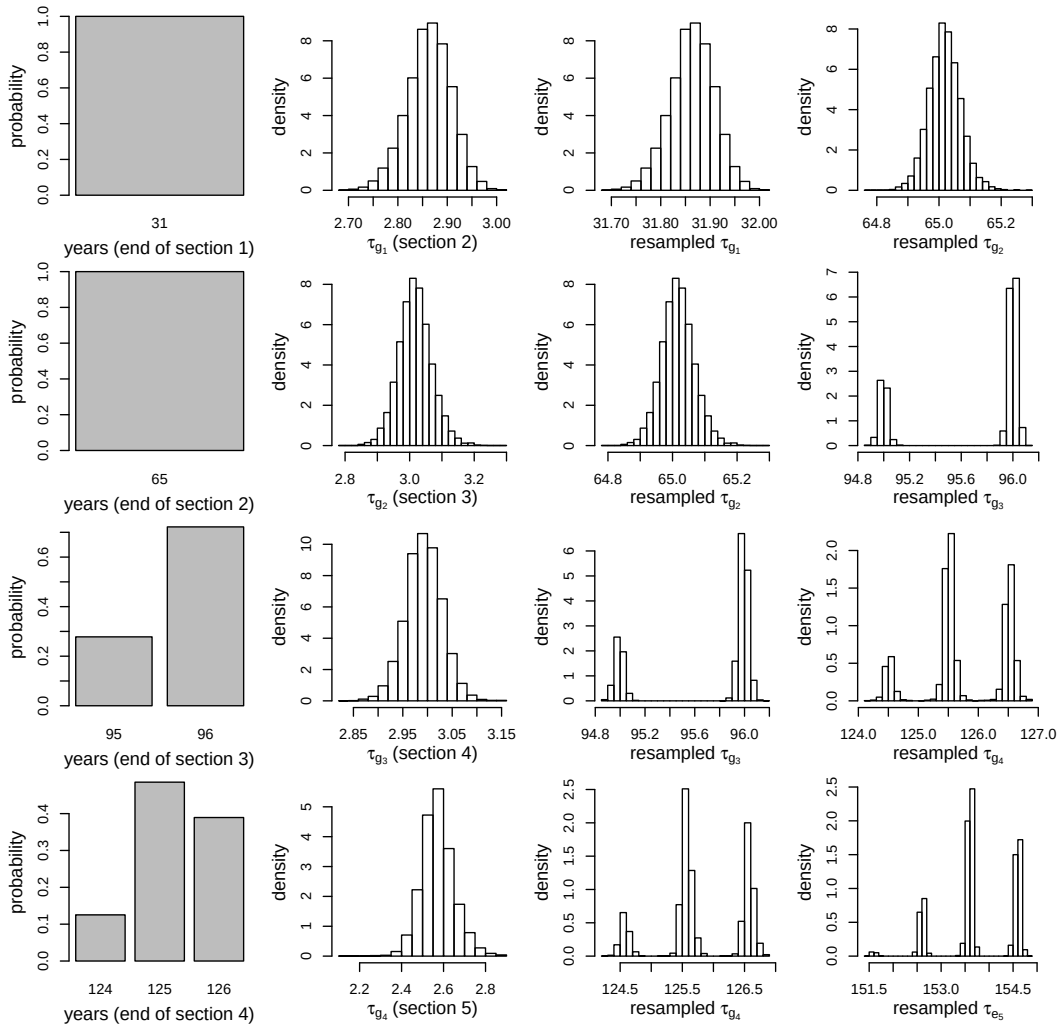

Figure 7.3: An illustration of the re-sampling process used to join together the posterior distributions of $\boldsymbol{\tau}$, from independent MCMC runs on the Fletcher δ^{18}O signal split into for the five overlapping sections. See the main text for details.

Figure 7.3: the second row of plots. The first plot is the probability distribution for the number of cycles at depth index g_2 from the MCMC run on section 2 after re-sampling to take into account the MCMC run on section 1, as shown in the fourth plot of the first row. Again, no switching occurred in section 2 (after burn-in) giving 65 years with a probability of 1. The second plot shows the posterior distribution of $\tau_{g2}^{(3)}$. The third and fourth plots show the posterior distributions of $\tau_{g2}^{(3)}$ and $\tau_{g3}^{(3)}$ after re-sampling at each iteration so that the integer part of $\tau_{g2}^{(3)}$ follows the distribution of plot 1: giving $\tau_{g1}^{(3)\prime}$ and $\tau_{g3}^{(3)\prime}$ respectively. Switching did occur in section 3.

Figure 7.3: the third row of plots. The first plot is the probability distribution for the number of cycles at depth index g_3 from the re-sampled section 3: this is 95 and 96 years with a probabilities of 0.28 and 0.72 respectively. The second plot shows the posterior distribution of $\tau_{g_3}^{(4)}$. The third and fourth plots show the posterior distributions of $\tau_{g_3}^{(4)'}$ and $\tau_{g_4}^{(4')}$ (after re-sampling at each iteration). Note that the third plot should be similar to the fourth plot of row 2. Switching also occurred in section 4.

Figure 7.3: the last row of plots. The first plot is the probability distribution for the number of cycles at the final joining depth index, g_4, from the re-sampled section 4: this is 124, 125 and 126 years with a probabilities of 0.13, 0.48 and 0.39. The second plot shows the posterior distribution of $\tau_{g_4}^{(5)}$. The third and fourth plots show the posterior distributions of $\tau_{g_4}^{(5)}$ and $\tau_{e_5}^{(5)}$ after re-sampling at each iteration. Switching also occurred in section 5. The final plot gives the distribution of cycles at the last depth evaluated, e_5: with probabilities of 0.01, 0.16, 0.48 and 0.35 assigned to the cycle counts of $151-154$ years respectively. It should be noted that this is less than the manual count of 155 as approximately one cycle was excluded from this analysis by the method used to choose b_1 and e_K.

7.6.2 The Gomez ice core

The H_2O_2 signal from the Gomez core is thinned to every second point (4cm resolution) for this analysis. We know, from the manual count and from Sections 3.2.6 and 3.2.8 of this thesis, that there are around 153 years covered by the Gomez ice core. We again choose $K = 5$, expecting around $30-31$ cycles in each section, and $V = 8$ to split the signal.

The signal has $n = 6615$ data points, from 1.72m to 134m. Running the univariate classification algorithm with $\nu = 1/2$ gives 629 runs, of which 556 are classified as certain. The first certain run (run 3) starts at depth index $i = 68$ (3.06m), we set the start of the first section as $b_1 = 68$. The last certain run (run 627) ends

at depth index 6589 (133.48m), we set the end of the last section as $e_5 = 6589$. There are 69 issue runs between these depths, 625 runs in total.

We want to split the runs as equally as possible into the five distinct groups, so that each group starts and ends with eight certain runs. The signal can be split into five groups of exactly 125 runs: the first comprising runs $3 - 127$, then $128 - 252$, $253 - 377$, $378 - 502$ and $503 - 627$. It turns out that the eight runs either side of all of these boundaries are certain runs so that no adjustment is required: the V overlap runs for sections 1 and 2 are $120 - 135$. We choose the first joining index as that which bounds runs 127 and 128: $g_1 = 2171$ (45.12m); the end of section 1 to be the last index of run 135, $e_1 = 2267$ (47.04m); and the start index of section 2 to be the first of run 120, $b_2 = 2015$ (42.00m). Similarly we have:

$$b_1 = 68 \ \ (3.06\text{m}),\ b_2 = 2015 \ \ (42.00\text{m}),\ b_3 = 3522 \ \ (72.14\text{m}),$$
$$b_4 = 4725 \ \ (96.20\text{m}),\ b_5 = 5711 \ \ (115.92\text{m}),$$

$$g_1 = 2171 \ \ (45.12\text{m}),\ g_2 = 3606 \ \ (73.82\text{m}),$$
$$g_3 = 4807 \ \ (97.84\text{m})\ g_4 = 5769 \ \ (117.08\text{m}),$$

$$e_1 = 2267 \ \ (47.04\text{m}),\ e_2 = 3685 \ \ (75.40\text{m}),\ e_3 = 4863 \ \ (98.96\text{m}),$$
$$e_4 = 5819 \ \ (118.08\text{m}),\ e_5 = 6589 \ \ (133.48\text{m}).$$

Figure 7.4 below shows the posterior distributions of hyperparameters for the timescale $\boldsymbol{\tau}$, ψ and λ, as box-plots, one for each of the sections. The thinning of the annual layer thickness down the core can be seen in the systematic reduction in λ along sections $3 - 5$. The distribution of ψ stays relatively stable in sections $2 - 4$, in sections 1 and 5 cycle lengths exhibit less variability. Equivalent plots for the posterior distributions of the standard deviation hyperparameters σ_α, σ_β and σ are included in the appendix as Figure B.27. These suggest increasing

variability in the mean, amplitude and error processes from sections $2-5$.

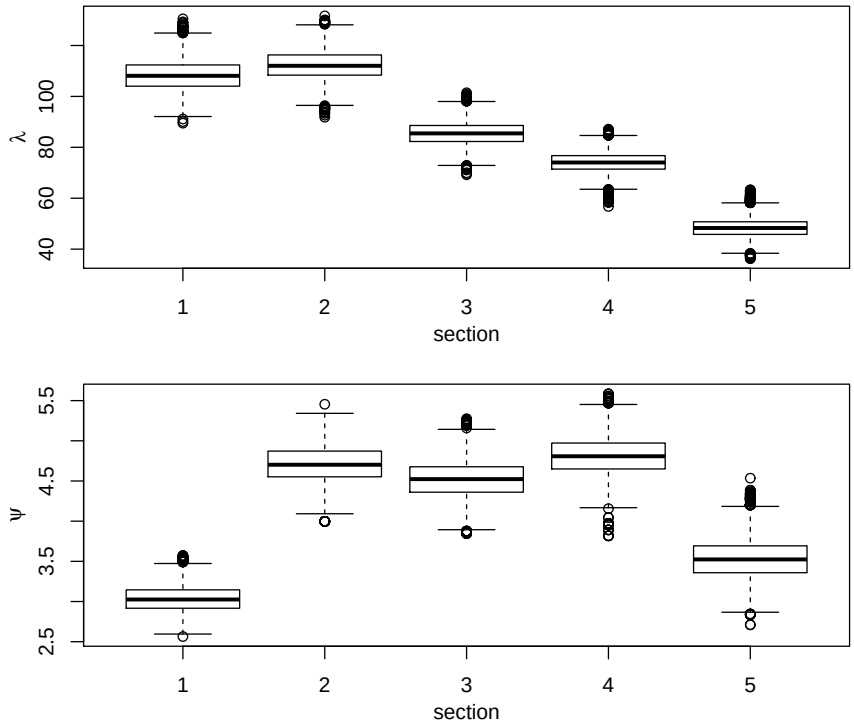

Figure 7.4: Boxplots for the posterior distributions of ψ and λ, the timescale hyperparameters, from five independent MCMC runs on the Gomez H_2O_2 signal.

The posterior distributions for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are provided in the appendix as Figure B.28 as density heat plots. The vertical dotted lines show the joining depths for the five sections. Trace plots for the timescale $\boldsymbol{\tau}$, evaluated at g_k for sections $k \in (1, \ldots, 4)$ and at e_5 for section 5, are provided as Figure B.29 in the appendix – we use a burn-in of 4000 iterations for each section. Switching only occurs (after burn-in) for the MCMC run of section 5.

Figure 7.5 below gives the distribution of cycles at the last depth index evaluated, e_5: with probabilities of 0.17, 0.76 and 0.07 assigned to cycle counts of 151, 152 and 153 years. It should be noted that this is less than the manual count of 153, which is due to approximately one cycle being excluded from this analysis by the method used to choose b_1 and e_K.

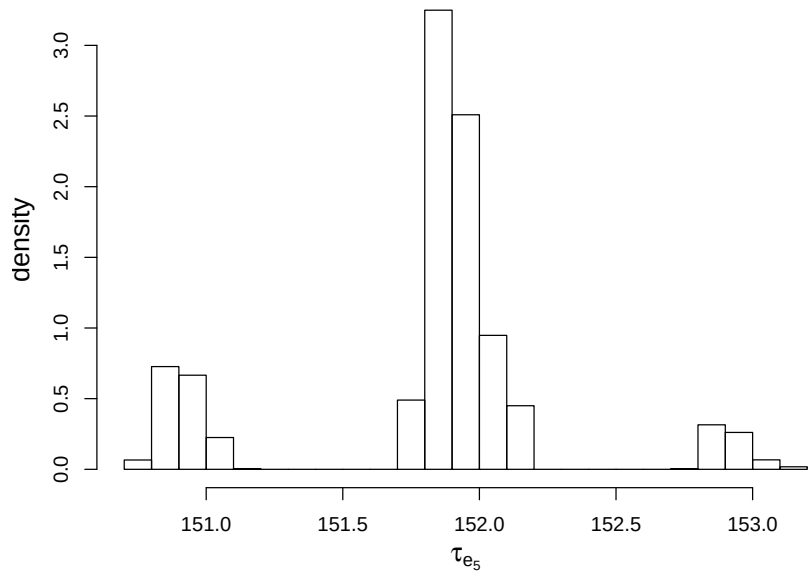

Figure 7.5: Histogram for the posterior distribution of the end of the timescale, τ_{e_5}, fit to the entire Gomez H_2O_2 signal from five independent MCMC runs.

7.6.3 A section of the NGRIP ice core

The ammonium and calcium signals from the NGRIP ice core (1440m–1465m) are thinned to every sixth point (6mm resolution) for this analysis. We know from the manual count, and from the analysis in Section 3.2.7, that there are around 410 years covered by this stretch of the NGRIP ice core. We choose $K = 20$ to split the signal, expecting around $20 - 21$ cycles in each section. We discuss the bivariate case in detail and present the results of the univariate cases. We use $V = 4$ to split the sections. The standard deviation for the Metropolis-Hastings updating of ρ is set at 0.012, which has been tuned in experimental runs of the algorithm (those of Chapter 6) to achieve an acceptance rate of approximately 20%.

We perform the standardisation and classification, along with choosing the splitting indexes of the signal, prior to thinning. The section boundaries and joining indexes are adapted for the thinned signal to be at the same depths. The starting values for the mean and amplitude process parameters are thinned along with the signal.

The (un-thinned) signal has $n = 25,000$ data points between 1440m and 1465m. After running the quadrant based bivariate classification algorithm of Section 3.3, there are 1786 runs of which 1108 are certain runs. The first certain run (run 3) starts at depth index $i = 21$ (1440.02m), so we set the start of the first section as $b_1 = 21$. The last certain run (run 1782) ends at depth index $24,961$ (1464.96m), so we set the end of the last section as $e_5 = 24,961$. There are 672 issue runs, and 1780 runs in total, between these depths. We want to split the runs as equally as possible into the $K = 20$ distinct groups, so that each group starts and ends with $V/2 = 2$ certain runs.

Figure B.30, in the appendix, shows the posterior distributions of ψ and λ, the hyperparmeters for $\boldsymbol{\tau}$, as box-plots, one for each section. Both are relatively stable along this stretch of the NGRIP core. Equivalent plots for the posterior distributions of the standard deviation hyperparameters $\sigma_{\alpha,x}$, $\sigma_{\beta,x}$, $\sigma_{\alpha,y}$, $\sigma_{\beta,y}$, σ_x and σ_y are included in the appendix as Figure B.31. The posterior distributions for $\boldsymbol{\alpha}_x$, $\boldsymbol{\beta}_x$, $\boldsymbol{\alpha}_y$ and $\boldsymbol{\beta}_y$ are provided below as Figure 7.6 in the form of density heat plots. The vertical dotted lines show the joining depths for the 20 sections.

Figure B.32 in the appendix is equivalent to Figure 7.3 – but is evaluated at only five randomly chosen sections: 6, 10, 15, 18 and 20. This is provided to give a feel for the re-sampling process used to join the timescale.

This analysis was also run using the univariate approach separately for the thinned ammonium and calcium signals, using the same section boundaries and joining indexes (b_k, g_k and e_k). The final joined up and sequentially re-sampled posterior distribution for the timescale evaluated at e_{20} for each case is provided as Figure 7.7. The univariate Na case (top) agrees closely with the bivariate case (bottom), which shows relatively less variability in the cycle count and time-of-year. The univariate Ca case (middle) provides a distribution with fewer cycles.

The most likely number of years in this section of the NGRIP ice core, from both the bivariate and univariate NH_4 MCMC runs, is 404. There is a trough in the beginning part ($\boldsymbol{\tau} < 1$) of the chronologies, giving the most likely number

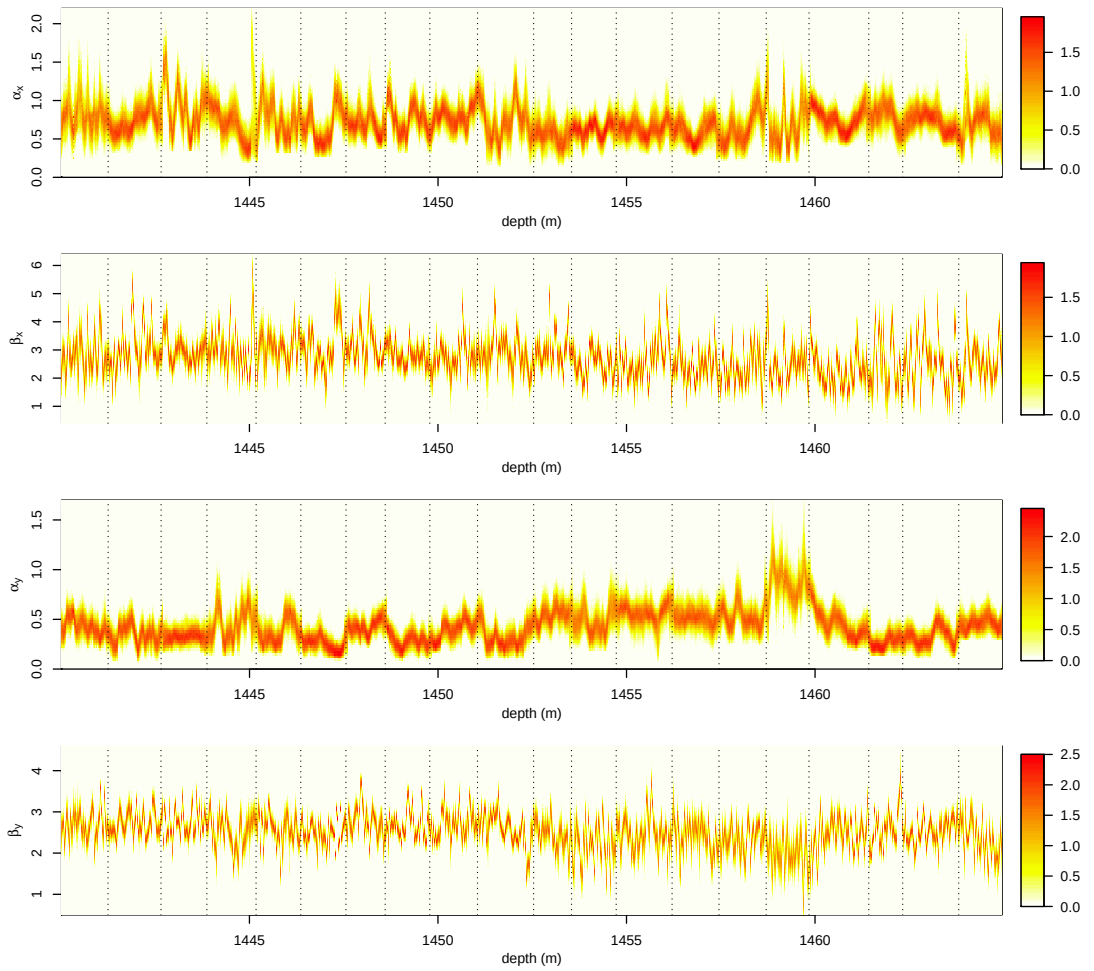

Figure 7.6: Density heat plots for the posterior distributions of the amplitude and mean processes, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, from independent bivariate MCMC runs on the NGRIP ammonium and calcium signals (1440m–1465m) after being split into 20 overlapping subsections containing roughly the same number of cycles. The vertical dotted lines show the joining depths for the 20 sections.

of troughs as 405, for comparison with 410 for the regression based chronology calculated in Chapter 3. This is a 1.2% difference. Both the univariate MCMC and the regression based methods put a reasonable probability measure on 408 troughs. The difference in these chronologies, in the language of Chapter 3, is caused by the probability measures on the possible reconstructions of each issue. The MCMC method uses all of the information in the data, rather than a summary such as run lengths, and is therefore much more likely to place cycles with certainty in an issue where the regression model would give a (high) probability.

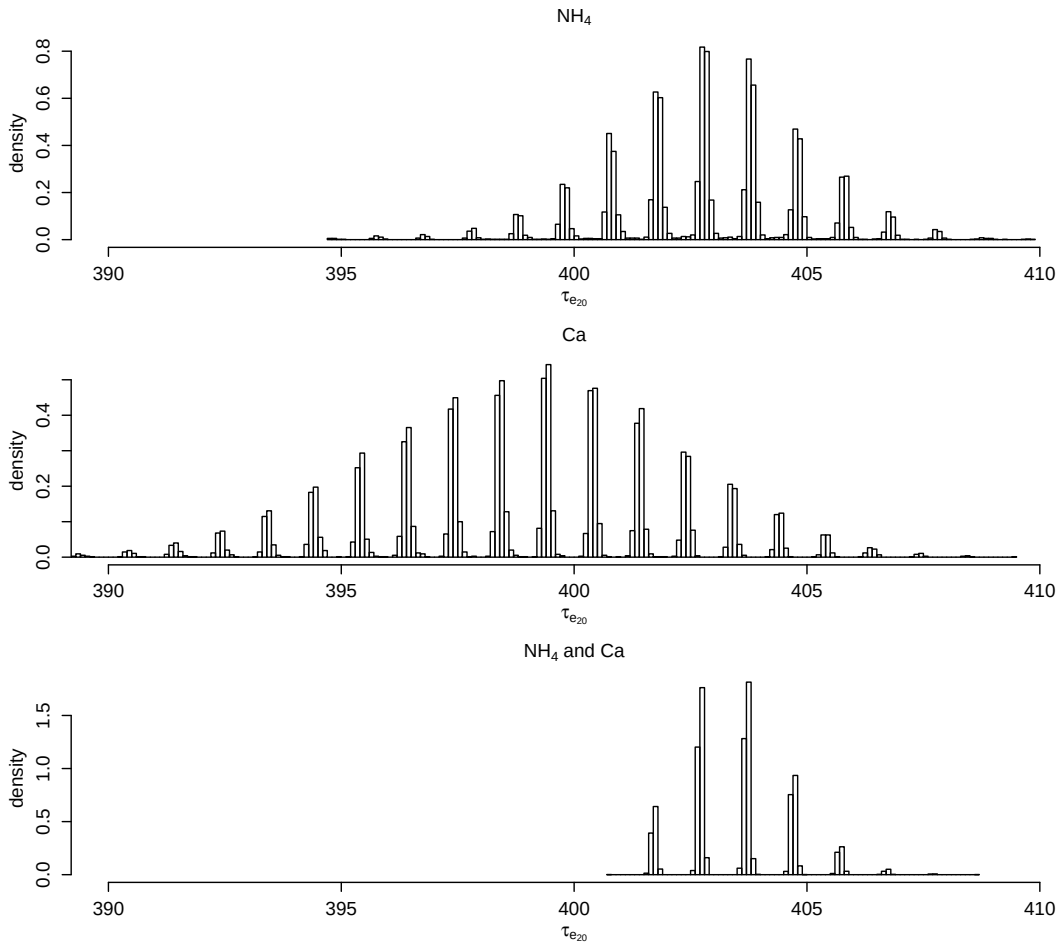

Figure 7.7: Histograms for the posterior distribution of the end of the timescale, τ_{e_5}, fit to the NGRIP ice core (1440m–1465m) ammonium and calcium signals from independent MCMC runs after being split into 20 overlapping subsections containing roughly the same number of cycles. **(top)** Dated using the univariate model on the ammonium signal. **(middle)** Dated using the univariate model on the calcium signal. **(bottom)** Dated using the bivariate model on the both signals.

Under the assumption used to calculate the regression chronologies, that issues are independent, this does not have to happen many times for the most likely number of troughs to differ between these methods as we see in this example.

To provide some insight into why the univariate Ca ($\boldsymbol{y}$) case provides a distribution with fewer cycles than the bivariate or univariate NH_4 ($\boldsymbol{x}$) cases we provide an example section of 1454.1m to 1454.4m for all three cases below as Figure 7.8. Note that this is the first example section of signal used in Chapters 4, 5 and

6. The top row shows the mean posterior reconstruction when the signals are fit separately, using the univariate approach. In this case the timescale fit to the NH_4 signal covers 5.3 years whereas that fit to the Ca signal covers only 3.5 years. The bottom row shows the mean posterior reconstruction when the signals are fit jointly using the bivariate approach. In this case the (joint) timescale covers 5.3 years, note how the reconstruction for the Ca signal has adapted. No switching occurred in any of the three runs, each of 10,000 iterations.

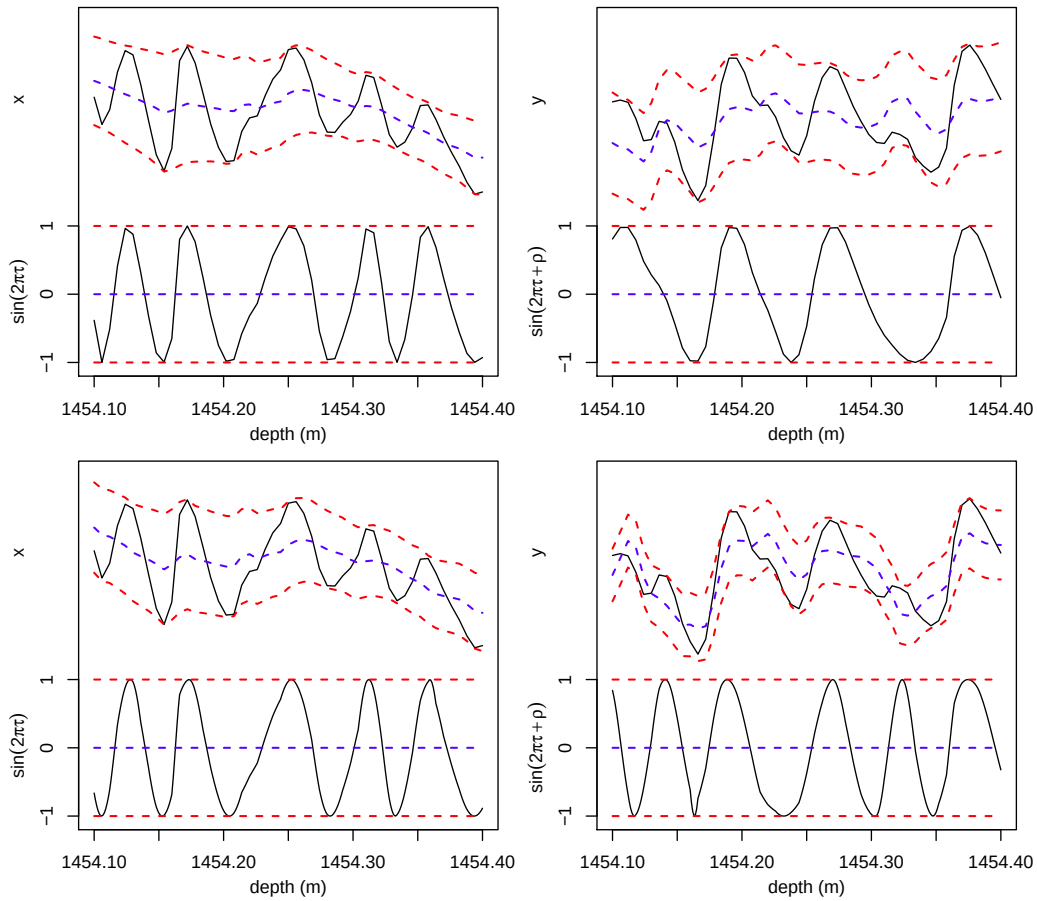

Figure 7.8: The univariate and bivariate model fit to the log NGRIP NH_4 ($\boldsymbol{x}$) and Ca ($\boldsymbol{y}$) signals (1451.1m–1451.6m), both thinned to 6mm resolution, calculated from MCMC runs of 10,000 iterations with a burn-in of 2000. The plots for $\boldsymbol{x}$ are on the left; those for $\boldsymbol{y}$ are on the right. **(top)** Fit separately using the univariate approach. **(bottom)** Fit jointly using the bivariate approach.

7.7 Conclusions

We have shown how the MCMC methods developed in the previous chapters can be used to fit timescales to longer sections of annually cyclic ice core signals. We have provided chronologies for the entire Fletcher and Gomez ice cores using the univariate approach; and for the example stretch of NGRIP data we have available using the ammonium and calcium signals, both individually and jointly.

This approach is automated in the sense that it could be applied to a new ice core with minimal user input – just the values for K, the number of sections, and V, the amount of overlap in terms of certain runs. It requires no tuning of any technical parameters such as those associated with the MCMC algorithms – the same code was used, unchanged, for each example. Each example presented was run overnight in under eight hours on an ordinary laptop.

The amount of information gathered from these signals using this method aside from the timescale is immense – see Figures 7.2 and 7.6 for example.

Chapter 8

Summary, discussion and future work

8.1 Summary

In this thesis we have presented a number of techniques that together provide an automated method for the dating of ice cores and other palaeoclimatic records using layer counting.

We have shown that, using the autocorrelation function, it is possible to detect whether seasonality is present in a signal, and have developed a simple method to estimate the average cycle length. Using an iterative algorithm, the signal can be split into a chosen number of subsections that contain approximately the same number of cycles. A bespoke technique can then be used to standardise an annually cyclic signal to what is essentially a sinusoid, isolating the seasonality.

We have presented univariate and bivariate methods to segment standardised signals into classified subsections: those with a deterministic cycle count, termed 'certain runs'; and those that require more attention, termed 'issues'. The certain runs are labelled, and represent a specific portion of an annual cycle. This allows the possible reconstructions of issues to be identified, and displayed as a visual

aid for manual counting. Further to this, probability measures can be assigned to each reconstruction, based on regression modelling of their length and the labelling of the bordering deterministic sections.

We have developed a flexible Markov chain Monte Carlo modelling approach for use on raw or log transformed annually cyclic signals. The algorithm produces a chronology consisting of a posterior distribution for the cycle count, as well as the time of year, at each depth. An automated choice of starting values and universal prior assumptions allows this method to be fully automated, and ready to use on any reasonably well behaved new signal with minimal or no tuning. The updating scheme we have developed for the model parameters achieves efficient acceptance rates, which allow the algorithm to be used in a realistic time frame. Crucially, this model can be used on multivariate datasets, with examples provided for the bivariate case. We have shown how this method can be scaled up for use on signals of any length in a fully parallelisable framework.

The chronologies we have calculated from the Gomez H_2O_2 and nss-s signals agree closely with the independent manual count and layer placings for that core. The chronologies calculated from the Fletcher $\delta^{18}O$ signal were validated, with each certain trough placement and issue reconstruction probability manually assessed and verified by scientists at BAS.

Our dating of the NGRIP ice core between 1440m and 1465m using the standardisation and classification approach agrees almost exactly with the GICC05 chronology. The MCMC approach suggests 1.5% fewer years (404 vs. 410) in this section. This difference may be explained by the process that was used to place certain manual layer markings for the GICC05 chronology. Certain layers are defined as those which fall into a $75 - 100\%$ subjective probability range. Even with complete agreement on the initial probability measures for these layers, this summary could result a lower 'most likely' count. It could also be due to information held in the other signals used for the multivariate GICC05 layer count. The MCMC approach was assigned very loose or uninformative priors when applied to the examples in this thesis, which do not represent the GICC05 manual

counters' prior beliefs. A fair test of this method against manual counting would be to elicit the true prior beliefs of a manual counting group, then for them and the algorithm to date the same section using the same signals.

8.2 Discussion

Layer counting of ice cores is generally considered to be more challenging than for other palaeoclimatic archives such as lake varves, tree rings, speleothem layers and coral. The models presented in this thesis should therefore be applicable to, or could be adapted for use on, these signals provided they are recorded at sufficiently high resolution.

Modelling three or more signals simultaneously is achievable using our MCMC methodology, and is a simple adaptation to the existing code. However, we have discussed how a multivariate dataset of strong and complete signals can be reduced to the bivariate case via Principal Component Analysis. This is recommended to avoid an increase in computational expense unless inference about the mean, amplitude or error structure of the signals is a priority.

We have discussed how the models used in the fully Bayesian MCMC approach can be adapted for signals with different or changing cycle shapes. Under the MCMC framework these could be modelled by replacing the sine wave with different periodic function with minimal adaptation to the existing code. We have shown how the classification algorithm can be adapted to a non-parametric form for use on the Gomez nss-S signal.

Fixed time constraints, such as a sulphate signals or a tephra fingerprints in the ice from a large volcanic eruption, would fit naturally into this framework as part of the prior distribution for the timescale. The difficulty arises in how to incorporate this information when the signal has been split into sections, as opposed to one run, for data that contains two or more such constraints. Fixed time constraints could be incorporated into the probability measures used in the classification method by conditioning on combinations of issue reconstructions

with a given cycle count and then normalising.

Modelling the phase difference between signals as a continuous correlated process, rather than as constant over each section or run, would be a valuable adaptation to the multivariate MCMC model. It could be assigned a Gaussian random walk prior, similar to the mean and amplitude processes. Initial experimentation with this possibility has proven promising.

Another consideration is how the resulting chronologies and their uncertainty measures can be used to model the environmental and climatic data from ice cores, such as inferred historical temperatures and concentrations of atmospheric gases, against time. One option is to use the most likely timescale, taking the reconstruction with the highest probability for each issue, or area where switching occurred for the MCMC approach. One advantage of the detailed MCMC output for the chronology is that inference can be made by integrating over the posterior distribution of the timescale.

8.3 Future work

The MCMC methodology was developed to be fully automated for reasonably well behaved signals, such as the examples presented in this thesis. Elicitation of stronger prior distributions, to reflect the beliefs of a dating expert, could be used obtain a more subjective chronology. It may be desirable, however, to use the knowledge of these experts to its fullest extent and, rather than replace them, provide them with a tool to aid manual layer counting. I have developed a preliminary manually guided program, written in R, in collaboration with Dr Liz Thomas and Dr Rob Mulvaney at the British Antarctic Survey. It is currently based on the univariate theory of Chapters 2 and 3 and allows the user to:

1. read in and view a signal from a .csv file
2. transform the signal, if required
3. iteratively standardise the signal

- try different values for the tuning parameter β
- view summaries and images for each choice

4. run the univariate (parametric) classification algorithm
    - try different values for the tuning parameter ν
    - manually assess where issues have been placed by the algorithm
    - create additional (manual) issues, if necessary
5. iterate through the issues
    - view plots of each issue, with both the raw and standardised signals
    - view each possible reconstruction, superimposed on the signal, and their regression based probabilities
    - view an illustration of how the probabilities were calculated
    - choose either to accept the calculated probabilities, or manually override them
6. view summaries and plots of the resulting chronology
7. select from various outputs

Note that during step 4 of the program, the manual assessor agrees and validates the certain runs and can add their own issues. Certain runs can be assigned to issues in step 5 by assigning one of the reconstructions a probability of 1.

The current outputs from the program include: a table of each possible cycle count and its calculated probability (the probability function); and the onset depths of the assigned runs for the most likely reconstruction (season markings). This program is currently in use by Dr Liz Thomas and Dr Rob Mulvaney at the British Antarctic Survey on ice core signals, and Nerilie Abram at the Australian National University on lake and coral records. This program could be made more user friendly by adding a GUI. Although the regression model built into the program works well for the examples we have tried so far, other signals may

have different modelling requirements. One possible future direction is to replace the regression based probabilities at step 5 with MCMC runs on the issues. This is quite different to dating the whole signal using MCMC, providing a highly detailed chronology for the issues only. Prior elicitation would be important, and could be included in the program. Another advancement, still within the theory developed in this thesis, would be to make the program multivariate.

The standardisation and classification approach to layer counting, developed in Chapters 2 and 3, have been well documented in Wheatley et al (2012). A summary of the univariate MCMC approach has been provided in a conference proceedings paper (Wheatley et al, 2014) and a full paper on the multivariate MCMC theory will follow this thesis. The most important next step for this research is the application of our modelling approach to other, longer ice core signals. The dating of a full size polar ice sheet core is within the scope of our methodology.

Bibliography

Andersen KK, Azuma N, Barnola JM, Bigler M, Biscaye P, Caillon N, Chappellaz J, Clausen HB, 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

Andersen KK, Svensson A, Johnsen SJ, Rasmussen SO, Bigler M, Röthlisberger R, Ruth U, Siggaard-Andersen ML, Peder Steffensen J, Dahl-Jensen D, Vinther BM, Clausen HB (2006) The greenland ice core chronology 2005, 15-42 ka. part 1: constructing the time scale. Quaternary Sci Rev 25(23-24):3246–3257

Butterworth S (1930) On the theory of filter amplifiers. Wireless Engineer 7:536–541

Dahl-Jensen D, Gundestrup NS, Miller H, Watanabe O, Johnsen SJ, Steffensen JP, Clausen HB, Svensson A, Larsen LB (2002) The northgrip deep drilling programme. Ann Glaciol 35(1):1–4

Eaton ML (1983) Multivariate statistics: a vector space approach. Wiley New York

Gay M, De Angelis M, Lacoume JL (2014) Dating a tropical ice core by time–frequency analysis of ion concentration depth profiles. Climate of the Past 10(5):1659–1672

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian data analysis. CRC press

Johnsen SJ, Dahl-Jensen D, Gundestrup N, Steffensen JP, Clausen HB, Miller H, Masson-Delmotte V, Sveinbjörnsdottir AE, White J (2001) Oxygen isotope and palaeotemperature records from six greenland ice-core stations: Camp century, dye-3, grip, gisp2, renland and northgrip. Journal of Quaternary Science 16(4):299–307

Jolliffe I (2005) Principal component analysis. Wiley Online Library

Klauenberg K, Blackwell PG, Buck CE, Mulvaney R, Röthlisberger R, Wolff EW (2011) Bayesian glaciological modelling to quantify uncertainties in ice core chronologies. Quaternary Science Reviews 30(21):2961–2975

Lambert F, Bigler M, Steffensen JP, Hutterli M, Fischer H (2012) Centennial mineral dust variability in high-resolution ice core data from dome c, antarctica. Climate of the Past 8(2):609–623

Lemieux-Dudon B, Blayo E, Petit JR, Waelbroeck C, Svensson A, Ritz C, Barnola JM, Narcisi BM, Parrenin F (2010) Consistent dating for antarctic and greenland ice cores. Quaternary Science Reviews 29(1-2):8–20

McGwire KC, Taylor KC, Banta JR, McConnell JR (2011) Identifying annual peaks in dielectric profiles with a selection curve. J Glaciol 57(204):763–769

Moore B (1981) Principal component analysis in linear systems: Controllability, observability, and model reduction. Automatic Control, IEEE Transactions on 26(1):17–32

Mudelsee M, Fohlmeister J, Scholz D (2012) Effects of dating errors on nonparametric trend analyses of speleothem time series. Clim Past Discuss 8:1973–2005

R Development Core Team (2011) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org/`, ISBN 3-900051-07-0

Rasmussen SO, Andersen KK, Siggaard-Andersen ML, Clausen HB (2002) Extracting the annual signal from greenland ice-core chemistry and isotopic records. Ann Glaciol 35(1):131–135

Rasmussen SO, Andersen KK, Svensson AM, Steffensen JP, Vinther BM, Clausen HB, Siggaard-Andersen ML, Johnsen SJ, Larsen LB, Dahl-Jensen D, Bigler M, Röthlisberger R, Fischer H, Goto-Azuma K, Hansson ME, Ruth U (2006) A new greenland ice core chronology for the last glacial termination. J Geophys Res 111(D6):D06,102, doi:10.1029/2005JD006,079

Rupf I, Radons G (2004) New approaches for automated data processing of annually laminated sediments. Nonlinear Proc Geoph 11(5/6):599–607

Shimohara K, Miyamoto A, Hyakutake K, Shoji H, Takata M, Kipfstuhl S (2003) Cloudy band observations for annual layer counting on the grip and ngrip, greenland, deep ice core samples (scientific note). Mem Natl Inst Polar Res, Spec Issue 57:161–167

signal developers (2013) signal: Signal processing. URL `http://r-forge.r-project.org/projects/signal/`

Smith C, Fairchild I, Spötl C, Frisia S, Borsato A, Moreton S, Wynn P (2009) Chronology building using objective identification of annual signals in trace element profiles of stalagmites. Quaternary Geochronology 4(1):11–21

Svensson A, Nielsen SW, Kipfstuhl S, Johnsen SJ, Steffensen JP, Bigler M, Ruth U, Röthlisberger R (2005) Visual stratigraphy of the north greenland ice core project (northgrip) ice core during the last glacial period. J Geophys Res 110(D2):D02,108, doi:10.1029/2004JD005,134

Svensson A, Andersen K, Bigler M, Clausen H, Dahl-Jensen D, Davies S, Johnsen S, Muscheler R, Rasmussen S, Röthlisberger R, et al (2006) The greenland ice core chronology 2005, 15-42 ka. part 2: Comparison to other records. Quaternary Science Reviews 25(23-24):3258–3267

Svensson A, Andersen K, Bigler M, Clausen H, Dahl-Jensen D, Davies S, Johnsen S, Muscheler R, Parrenin F, Rasmussen S, et al (2008) A 60 000 year greenland stratigraphic ice core chronology. Climate of the Past 4(1):47–57

Taylor KC, Alley RB, Meese DA, Spencer MK, Brook EJ, Dunbar NW, Finkel RC, Gow AJ, Kurbatov AV, Lamorey GW, Mayewski PA, Meyerson EA, Nishizumi K, Zielinski GA (2004) Dating the siple dome (antarctica) ice core by manual and computer interpretation of annual layering. J Glaciol 50(170):453–461

Thomas E, Dennis P, Bracegirdle TJ, Franzke C (2009) Ice core evidence for significant 100-year regional warming on the antarctic peninsula. Geophysical Research Letters 36(20)

Thomas ER, Marshall GJ, McConnell JR (2008) A doubling in snow accumulation in the western antarctic peninsula since 1850. Geophys Res Lett 35:L01,706, doi:10.1029/2007GL032,529

Uhlenbeck GE, Ornstein LS (1930) On the theory of the brownian motion. Physical review 36(5):823

Venables WN, Ripley BD (2002) Modern Applied Statistics with S, 4th edn. Springer, New York, URL `http://www.stats.ox.ac.uk/pub/MASS4`, iSBN 0-387-95457-0

Vinther BM, Clausen HB, Johnsen SJ, Rasmussen SO, Andersen KK, Buchardt SL, Dahl-Jensen D, Seierstad IK, Siggaard-Andersen ML, Steffensen JP, et al (2006) A synchronized dating of three greenland ice cores throughout the holocene. Journal of Geophysical Research: Atmospheres (1984–2012) 111(D13)

Weber ME, Reichelt L, Kuhn G, Pfeiffer M, Korff B, Thurow J, Ricken W (2010) Bmpix and peak tools: New methods for automated laminae recognition and counting – application to glacial varves from antarctic marine sediment. Geochem Geophys Geosyst 11(1):Q0AA05, doi:10.1029/2009GC002,611

Wheatley JJ (2007) Dating and layer counting in ice core records. Master's thesis, Department of Probability and Statistics, The University of Sheffield

Wheatley JJ, Blackwell PG, Abram NJ, McConnell JR, Thomas ER, Wolff EW (2012) Automated ice-core layer-counting with strong univariate signals. Climate of the Past 8:1869–1879

Wheatley JJ, Blackwell PG, Abram NJ, Wolff EW (2014) Bayesian layer counting in ice-cores: Reconstructing the time scale. In: The Contribution of Young Researchers to Bayesian Statistics, Springer, pp 121–125

Winstrup M (2011) An automated method for annual layer counting in ice cores: and an application to visual stratigraphy data from the ngrip ice core. PhD thesis, Københavns Universitet, Faculty of Science, The Niels Bohr Institute, Ice and Climate

Winstrup M, Svensson A, Rasmussen SO, Winther O, Steig E, Axelrod A (2012) An automated approach for annual layer counting in ice cores. Climate of the Past Discussions 8(6):1881–1895

Appendices

Appendix A

Supplementary mathematics and statistics

A.1 Conditioning a multivariate normal distribution.

This standard result, given in e.g. Gelman et al (2013), is used several times in the theory of this thesis. If $\boldsymbol{x}$ is a multivariate normally distributed vector, partitioned to $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

and we want to know the distribution of $\boldsymbol{x}_1|\boldsymbol{x}_2$, we first note that this is normally distributed (Eaton, 1983). We need to calculate the mean and variance of this distribution, define

$$\boldsymbol{y} = \boldsymbol{x}_1 + Q\boldsymbol{x}_2 \text{ where } Q = -\Sigma_{12}\Sigma_{22}^{-1}.$$

This gives:

$$cov(\boldsymbol{y}, \boldsymbol{x}_2) = cov(\boldsymbol{x}_1, \boldsymbol{x}_2) + cov(Q\boldsymbol{x}_2, \boldsymbol{x}_2)$$

$$= \Sigma_{12} + Qvar(\boldsymbol{x}_2) = \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = 0.$$

$\boldsymbol{y}$ and $\boldsymbol{x}_2$ are uncorrelated and therefore independent, since they are jointly normal. $E(\boldsymbol{y}) = \mu_1 + Q\mu_2$, and therefore

$$E(\boldsymbol{x}_1|\boldsymbol{x}_2) = E(\boldsymbol{y} - Q\boldsymbol{x}_2|\boldsymbol{x}_2) = E(\boldsymbol{y}|\boldsymbol{x}_2) - E(Q\boldsymbol{x}_2|\boldsymbol{x}_2)$$

$$= E(\boldsymbol{y}) - Q\boldsymbol{x}_2 = \boldsymbol{\mu}_1 + Q(\boldsymbol{\mu}_2 - \boldsymbol{x}_2) = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2).$$

and

$$var(\boldsymbol{x}_1|\boldsymbol{x}_2) = var(\boldsymbol{y} - Q\boldsymbol{x}_2|\boldsymbol{x}_2)$$

$$= var(\boldsymbol{y}|\boldsymbol{x}_2) + var(Q\boldsymbol{x}_2|\boldsymbol{x}_2) - Qcov(\boldsymbol{y}, -\boldsymbol{x}_2) - cov(\boldsymbol{y}, -\boldsymbol{x}_2)Q^T$$

$$= var(\boldsymbol{y}|\boldsymbol{x}_2) = var(\boldsymbol{y}) = var(\boldsymbol{x}_1 + Q\boldsymbol{x}_2)$$

$$= var(\boldsymbol{x}_1) + Qvar(\boldsymbol{x}_2)Q^T + Qcov(\boldsymbol{x}_1, \boldsymbol{x}_2) + cov(\boldsymbol{x}_2, \boldsymbol{x}_1)Q^T$$

$$= \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$= \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

A.2 Switching: the prior proposal ratio

We have independent Gamma distributed time increments,

$$\delta_{I,j} \overset{iid}{\sim} \mathrm{G}(\psi_{I,j}, \lambda),\ j \in (1, \ldots, m+1),$$

with pdf

$$p(\boldsymbol{\delta}_I) = \prod_j \frac{\lambda^{\psi_{I,j}} \delta_{I,j}^{\psi_{I,j}-1} e^{-\lambda\delta_{I,j}}}{\Gamma(\psi_{I,j})} = \lambda^{\sum_j \psi_{I,j}} e^{-\lambda \sum_j \delta_{I,j}} \prod_j \frac{\delta_{I,j}^{\psi_{I,j}-1}}{\Gamma(\psi_{I,j})}.$$

Recall that $\Delta = \sum_j \delta_{I,j}$, we know that $\Delta \sim \mathrm{G}(\Sigma_j \psi_{I,j}, \lambda)$ with pdf:

$$p(\Delta) = \frac{\lambda^{\sum_j i\psi_{I,j}} \Delta^{\sum_j \psi_{I,j}-1} e^{-\lambda\Delta}}{\Gamma\left(\sum_j \psi_{I,j}\right)}.$$

Defining $d_j = \delta_{I,j}/\Delta$ as with the perturbations $\boldsymbol{d}'$ in section 4.3.2 we can define a one-to-one transformation:

$$\{\delta_{I,1}, \delta_{I,2}, \ldots, \delta_{I,m+1}\} \to \left\{\frac{\delta_{I,1}}{\Delta}, \frac{\delta_{I,2}}{\Delta}, \ldots, \frac{\delta_{I,m}}{\Delta}, \Delta\right\} = \{\boldsymbol{d}, \Delta\}.$$

We know that $\boldsymbol{d}|\Delta$ is Dirichlet and use this when drawing $\boldsymbol{d}'$. For ease of notation defining $d_{m+1} = 1 - \Sigma_{j=1}^m d_j = \delta_{I,m+1}/\Delta$ we have:

$$p(\boldsymbol{d}|\Delta) = B(\boldsymbol{\psi})^{-1} \prod_j d_j^{\psi_{I,j}-1} = \Gamma\left(\Sigma_j \psi_{I,j}\right) \prod_j \frac{d_j^{\psi_{I,j}-1}}{\Gamma(\psi_{I,j})}$$

Our transformation has inverse:

$$\{d_1, d_2, \ldots, d_m, \Delta\} \to \left\{d_1\Delta, d_2\Delta, \ldots, d_m\Delta, \left(1 - \Sigma_{j=1}^m d_j\right)\Delta\right\} = \boldsymbol{\delta}_I$$

giving the Jacobian for the transformation as:

$$J = \begin{pmatrix} \Delta & 0 & 0 & 0 & 0 & d_1 \\ 0 & \Delta & 0 & 0 & 0 & d_2 \\ 0 & 0 & \Delta & 0 & 0 & d_3 \\ 0 & 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & 0 & 0 & \Delta & d_m \\ -\Delta & -\Delta & -\Delta & \ldots & -\Delta & 1 - \Sigma_{j=1}^m d_j \end{pmatrix}$$

which can be changed to the following using elementary row operations:

$$J \to \begin{pmatrix} \Delta & 0 & 0 & 0 & 0 & d_1 \\ 0 & \Delta & 0 & 0 & 0 & d_2 \\ 0 & 0 & \Delta & 0 & 0 & d_3 \\ 0 & 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & 0 & 0 & \Delta & d_m \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \to \begin{pmatrix} \Delta & 0 & 0 & 0 & 0 & 0 \\ 0 & \Delta & 0 & 0 & 0 & 0 \\ 0 & 0 & \Delta & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & 0 & 0 & \Delta & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} = J'$$

and we know that $|J| = |J'| = \Delta^m$. Now we can obtain the joint pdf of $\boldsymbol{d}$ and Δ,

$$p(\boldsymbol{d}, \Delta) = p_{\boldsymbol{\delta}}((\boldsymbol{d}, \Delta)^{-1})|J|$$

$$= \lambda^{\sum_j \psi_{I,j}} e^{-\lambda \sum_j d_j \Delta} \prod_j \frac{(d_j \Delta)^{\psi_{I,j}-1}}{\Gamma(\psi_{I,j})} \times \Delta^m$$

$$= \lambda^{\sum_j \psi_{I,j}} \Delta^{\sum_j \psi_{I,j}-1} e^{-\lambda\Delta} \prod_j \frac{d_j^{\psi_{I,j}-1}}{\Gamma(\psi_{I,j})}$$

$$= \left(\frac{\lambda^{\sum_j \psi_{I,j}} \Delta^{\sum_j \psi_{I,j}-1} e^{-\lambda\Delta}}{\Gamma\left(\Sigma_j \psi_{I,j}\right)} \right) \left(\Gamma\left(\Sigma_j \psi_{I,j}\right) \prod_j \frac{d_j^{\psi_{I,j}-1}}{\Gamma(\psi_{I,j})} \right)$$

$$= p(\Delta) p(\boldsymbol{d}|\Delta).$$

Similarly we have $p(\boldsymbol{\delta}) = p(\boldsymbol{d}, \Delta)/\Delta^m$. Thus the prior proposal ratio reduces to

$$\frac{p(\boldsymbol{\delta}')}{p(\boldsymbol{\delta})} \frac{q(\boldsymbol{d})}{q(\boldsymbol{d}')} = \frac{p(\boldsymbol{d}', \Delta+1)}{p(\boldsymbol{d}, \Delta)} \frac{p(\boldsymbol{d}|\Delta)}{p(\boldsymbol{d}'|\Delta+1)}$$

$$= \frac{p(\Delta+1)p(\boldsymbol{d}'|\Delta+1)}{p(\Delta)p(\boldsymbol{d}|\Delta)} \frac{p(\boldsymbol{d}|\Delta)}{p(\boldsymbol{d}'|\Delta+1)} = \frac{p(\Delta+1)}{p(\Delta)}$$

$$= \frac{f_{\mathrm{G}(\Sigma_j \psi'_{I,j}, \lambda)}(\Delta+1)}{f_{\mathrm{G}(\Sigma_j \psi_{I,j}, \lambda)}(\Delta)}.$$

A.3 Conditional prior for $\boldsymbol{\alpha}_I \mid \boldsymbol{\alpha}_{-I}$

Recall that $\boldsymbol{\alpha}$ has prior distribution

$$\alpha_i \sim \mathrm{N}(\alpha_{i-1}, \sigma_\alpha^2), \quad i \in (2, \ldots, n).$$

We want to find the distribution of $\boldsymbol{\alpha}_I = \{\alpha_{s+1}, \ldots, \alpha_{s+m}\}$ conditioned on α_s and $\alpha_f = \alpha_{s+m+1}$. For neatness we use $\sigma \equiv \sigma_\alpha$ in this section of the appendix only. Let $u_i = \alpha_{s+i} - \alpha_{s+i-1} \overset{iid}{\sim} \mathrm{N}(0,\ \sigma^2)$ for $i \in (1, \ldots, m+1)$, note that $\alpha_i = a_s + \sum_{j=1}^{i} u_j$ and $\sum_{i=1}^{m+1} u_i = \alpha_f - \alpha_s$. Define

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \\ \sum_i u_i \end{pmatrix} \sim \mathrm{N}(\boldsymbol{\mu}_u,\ \Sigma_u) \ \text{ where:}$$

$$\boldsymbol{\mu}_u = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \mu_2 \end{pmatrix} \text{ where } \boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ and } \mu_2 = 0;$$

$$\Sigma_u = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \text{ where } \Sigma_{11} = \begin{pmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & \ldots & 0 & \sigma^2 \end{pmatrix},$$

$$\Sigma_{12} = \begin{pmatrix} \sigma^2 \\ \sigma^2 \\ \vdots \\ \sigma^2 \end{pmatrix}, \quad \Sigma_{21} = (\sigma^2,\ \sigma^2, \ldots,\ \sigma^2) \text{ and } \Sigma_{22} = (m+1)\sigma^2.$$

Using standard results on conditioning of the multivariate normal, this gives

$$E\left(\boldsymbol{u} \mid \sum u_i = \alpha_f - \alpha_s\right) = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\alpha_f - \alpha_s - \mu_2) = \begin{pmatrix} \frac{\alpha_f-\alpha_s}{m+1} \\ \frac{\alpha_f-\alpha_s}{m+1} \\ \vdots \\ \frac{\alpha_f-\alpha_s}{m+1} \end{pmatrix},$$

so $E\left(u_i \mid \sum u_i\right) = \frac{\alpha_f-\alpha_s}{m+1}$, $i \in (1, \ldots, m)$, which implies that

$$E\left(\alpha_i \mid \sum u_i\right) = \alpha_s + E\left(\sum_{j=1}^{i} u_j \mid \sum u_i\right) = \alpha_s + \frac{i(\alpha_f - \alpha_s)}{m+1}$$

We also have $var\left(\boldsymbol{u} \mid \sum u_i\right) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

$$= \begin{pmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & \ldots & 0 & \sigma^2 \end{pmatrix} - \frac{1}{(m+1)\sigma^2} \begin{pmatrix} \sigma^4 & \sigma^4 & \ldots & \sigma^4 \\ \sigma^4 & \sigma^4 & \ldots & \sigma^4 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^4 & \sigma^4 & \ldots & \sigma^4 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{m\sigma^2}{m+1} & \frac{-\sigma^2}{m+1} & \cdots & \frac{-\sigma^2}{m+1} \\ \frac{-\sigma^2}{m+1} & \frac{m\sigma^2}{m+1} & \cdots & \frac{-\sigma^2}{m+1} \\ \frac{-\sigma^2}{m+1} & \frac{-\sigma^2}{m+1} & \ddots & \frac{-\sigma^2}{m+1} \\ \frac{-\sigma^2}{m+1} & \cdots & \frac{-\sigma^2}{m+1} & \frac{m\sigma^2}{m+1} \end{pmatrix}.$$

So $var\left(u_i \mid \sum u_i\right) = \frac{m\sigma^2}{m+1}$ and $cov\left(u_i,\ u_j \mid \sum u_i\right) = \frac{-\sigma^2}{m+1}$, which implies that

$$var\left(\alpha_i \mid \sum u_i\right) = var\left(\sum_{j=1}^{i} u_j \mid \sum u_i\right)$$

$$= i \times var\left(u_i \mid \sum u_i\right) + 2\binom{i}{2} cov\left(u_i,\ u_j \mid \sum u_i\right)$$

$$= \frac{im\sigma^2}{m+1} - \frac{i(i-1)\sigma^2}{m+1} = \frac{i(m+1-i)\sigma^2}{m+1}$$

$$cov(\alpha_i,\ \alpha_j) = cov\left(\sum_{k=1}^{i} u_k,\ \sum_{l=1}^{j} u_l\right) = \sum_{k=1}^{i}\sum_{l=1}^{j} cov(u_k,\ u_l)$$

which, for $i < j$,

$$= i \times var(u_i) + (ij - i) \times cov(u_i,\ u_j)$$

$$= \frac{im\sigma^2}{m+1} - \frac{(ij-i)\sigma^2}{m+1} = \frac{i(m+1-j)\sigma^2}{m+1},$$

and, in general,

$$= \frac{\min(i,j)(m+1-\max(i,j))\sigma^2}{m+1}$$

$$= \frac{\min(i,j)(m+1) - \min(i,j)\max(i,j)}{m+1} = \min(i,j) - \frac{ij}{m+1}.$$

Defining Σ_m where

$$\Sigma_{m,ij} = \min(i,j) - \frac{ij}{m+1}, \quad i,j \in (1,\ldots,m)$$

gives:

$$\boldsymbol{\alpha}_I \mid \alpha_s, \alpha_f \sim \mathrm{N}(\boldsymbol{\mu}_\alpha,\ \Sigma_\alpha = \sigma_\alpha^2 \Sigma_m),$$

$$\text{where } \ \boldsymbol{\mu}_{\alpha,i} = \alpha_s + \frac{i(\alpha_f - \alpha_s)}{m+1}, i \in (1,\ldots,m);$$

and, similarly,

$$\boldsymbol{\beta}_I \mid \beta_s, \beta_f \sim \mathrm{N}(\boldsymbol{\mu}_\beta,\ \Sigma_\beta = \sigma_\beta^2 \Sigma_m),$$

$$\text{where } \ \boldsymbol{\mu}_{\beta,i} = \beta_s + \frac{i(\beta_f - \beta_s)}{m+1}, i \in (1,\ldots,m).$$

Appendix B

Supplementary figures

B.1 Chapter 3

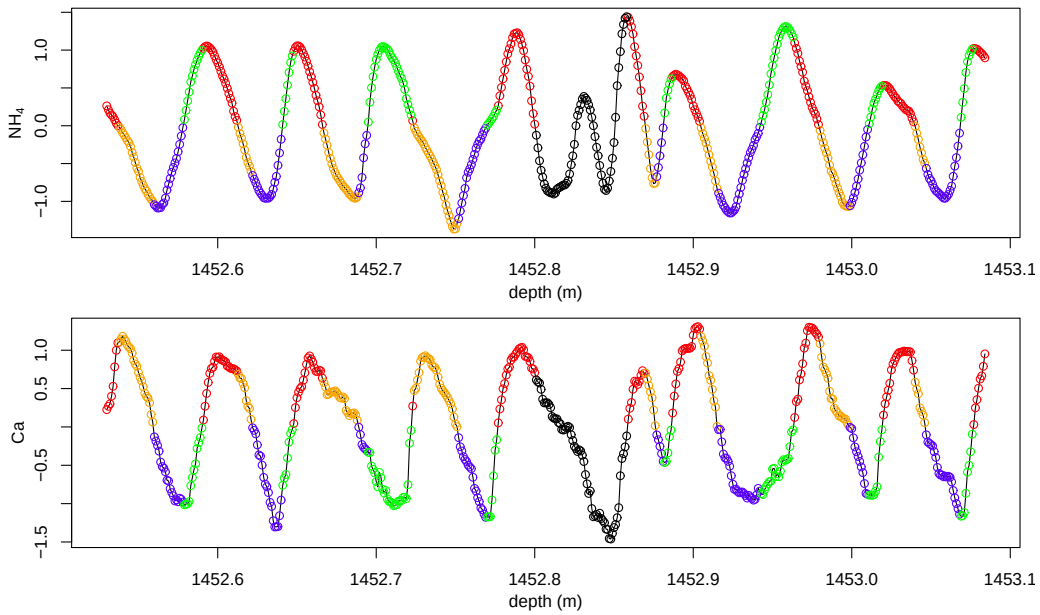

Figure B.1: A section of the standardised NH_4 and Ca signals from the NGRIP ice core (1452.54m–1453.10m) classified into runs using the bivariate quadrant based method. Points in Q_1 are shown in red, points in Q_2 are shown in blue, points in Q_3 are shown in green and those in Q_4 are shown in orange. Points that form an issue are black.

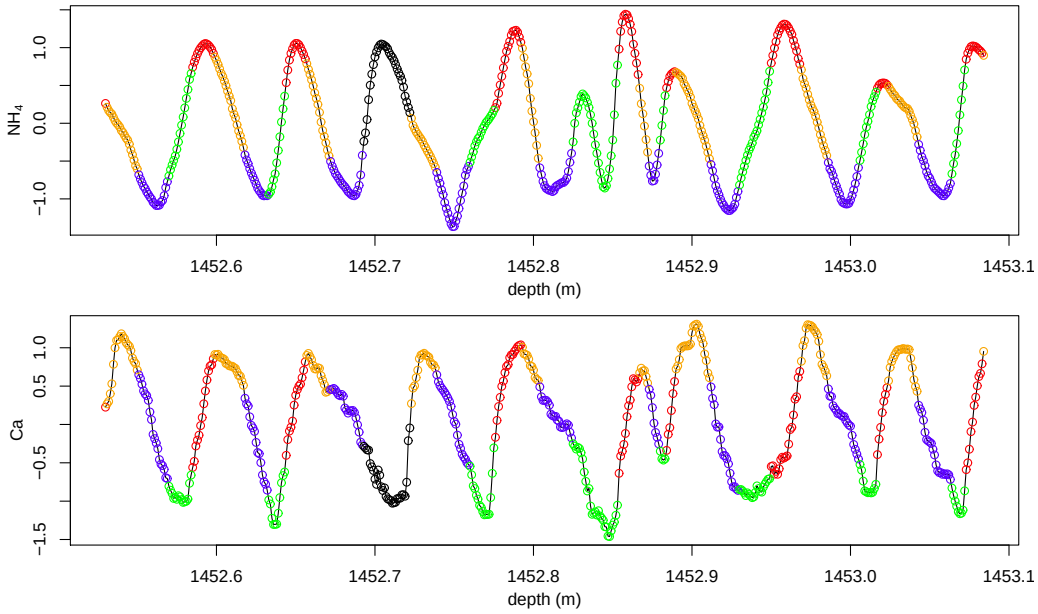

Figure B.2: A section of the standardised NH_4 and Ca signals from the NGRIP ice core (1452.54m–1453.10m) classified into runs using the bivariate seasonal based method. Points in S_1 are shown in red, points in S_2 are shown in blue, points in S_3 are shown in green and those in S_4 are shown in orange. Points that form an issue are black.

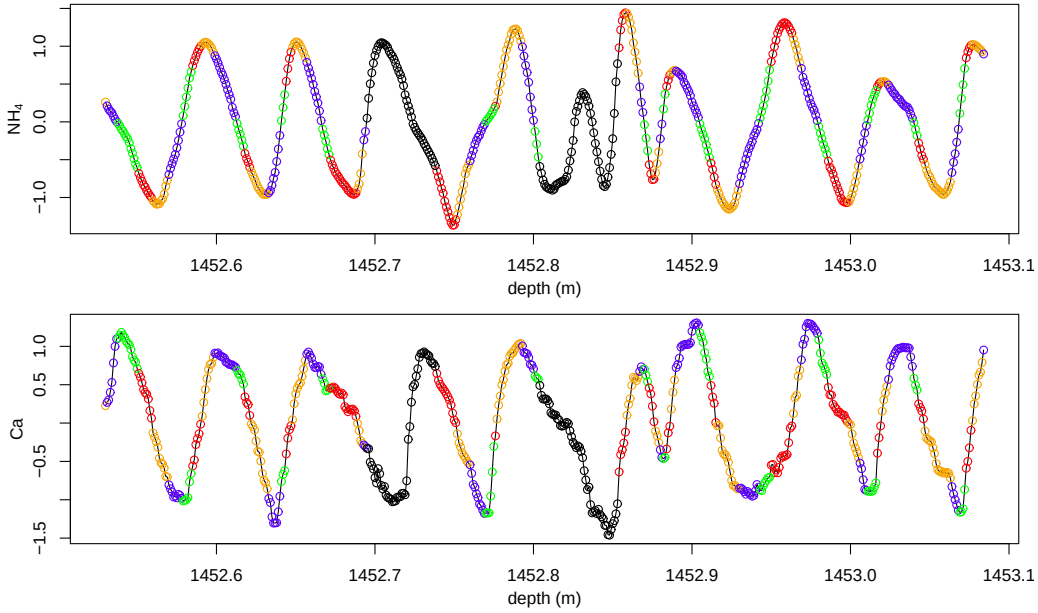

Figure B.3: A section of the standardised log NH_4 and Ca signals from the NGRIP ice core (1452.54m–1453.10m), classified into runs using a mixture of the bivariate quadrant and seasonal based methods.

B.2 Chapter 4

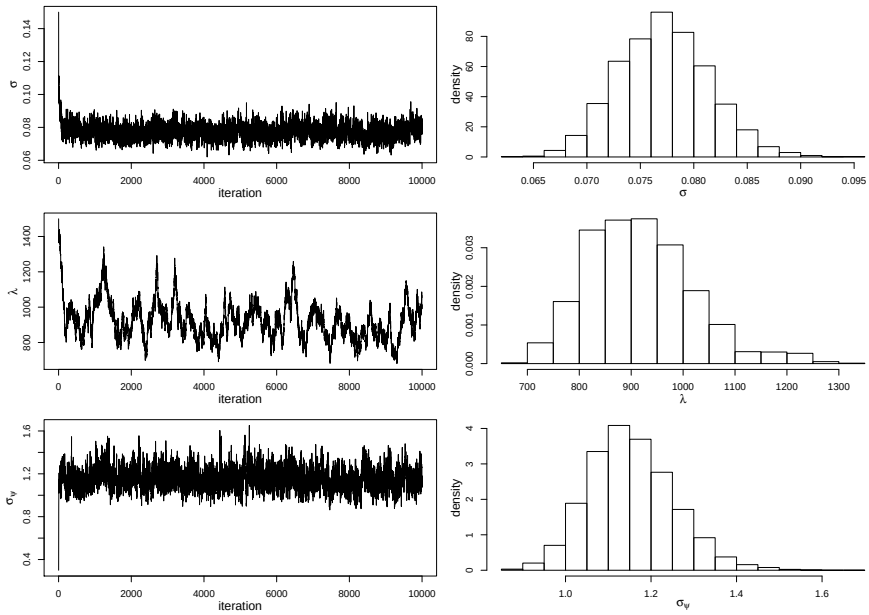

Figure B.4: Trace plots for model hyperparameters σ, λ and σ_ψ from Example 1, an MCMC run on the standardised NGRIP NH_4 signal (1454.1m–1454.4m), with histograms showing the posterior distributions after a burn-in of 1000 iterations.

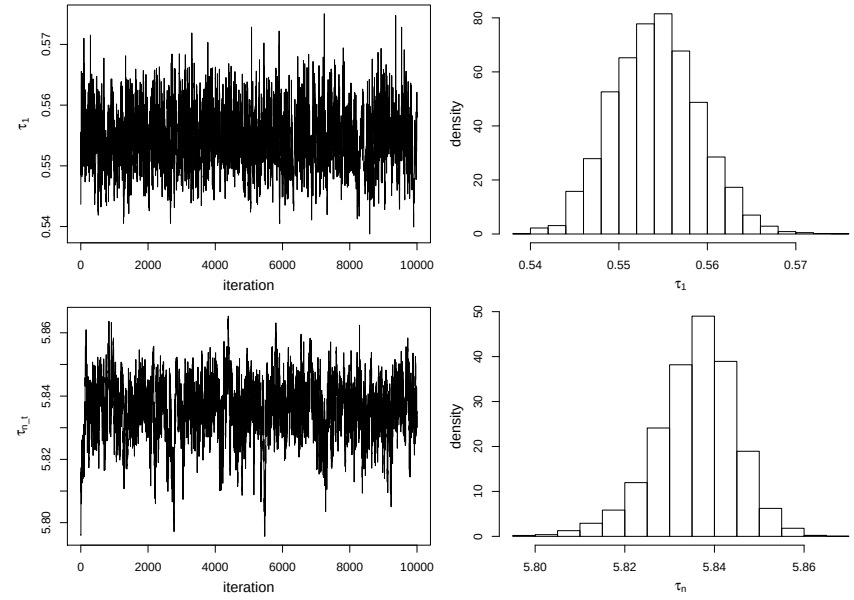

Figure B.5: Trace plots for the start and end of the timescale, τ_1 and τ_n, from Example 1, an MCMC run on the standardised NGRIP NH_4 signal (1454.1m–1454.4m), with histograms showing the posterior distributions after a burn-in of 1000 iterations.

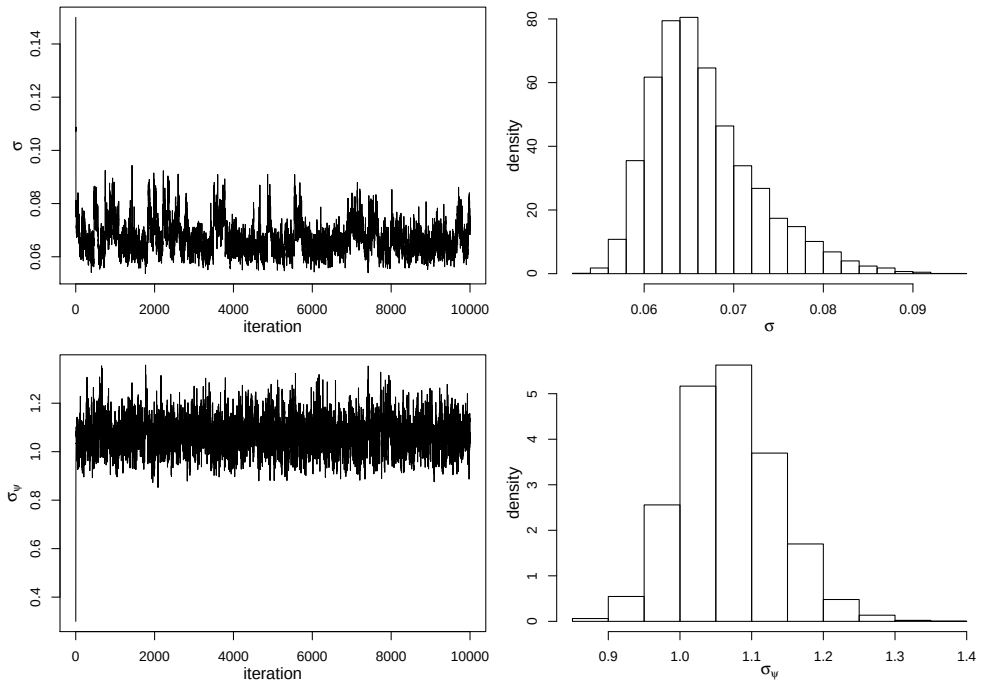

Figure B.6: Trace plots for model hyperparameters σ and σ_ψ from Example 2, an MCMC run on the standardised log NGRIP NH_4 signal (1451.1m–1451.6m), with histograms showing the posterior distributions after a burn-in of 1000 iterations.

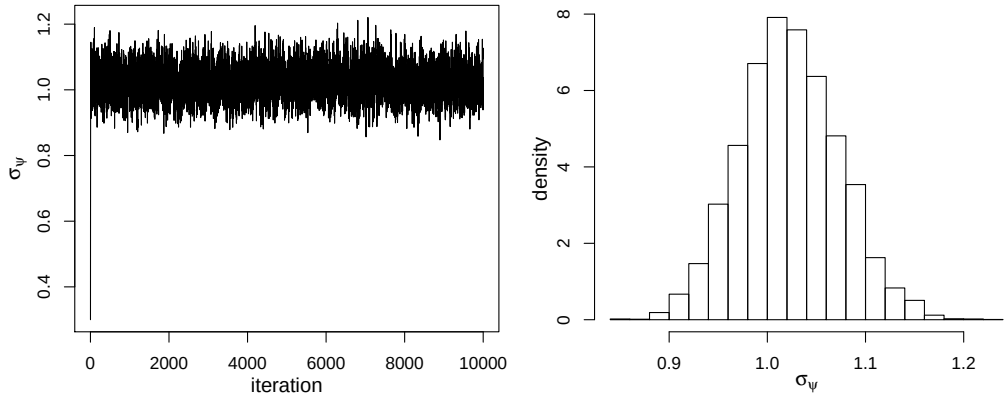

Figure B.7: Trace plot for model hyperparameter σ_ψ from Example 3, an MCMC run on the standardised NGRIP NH_4 signal (1449.4m–1450.0m), with a histogram showing its posterior distributions after a burn-in of 2000 iterations.

B.3 Chapter 5

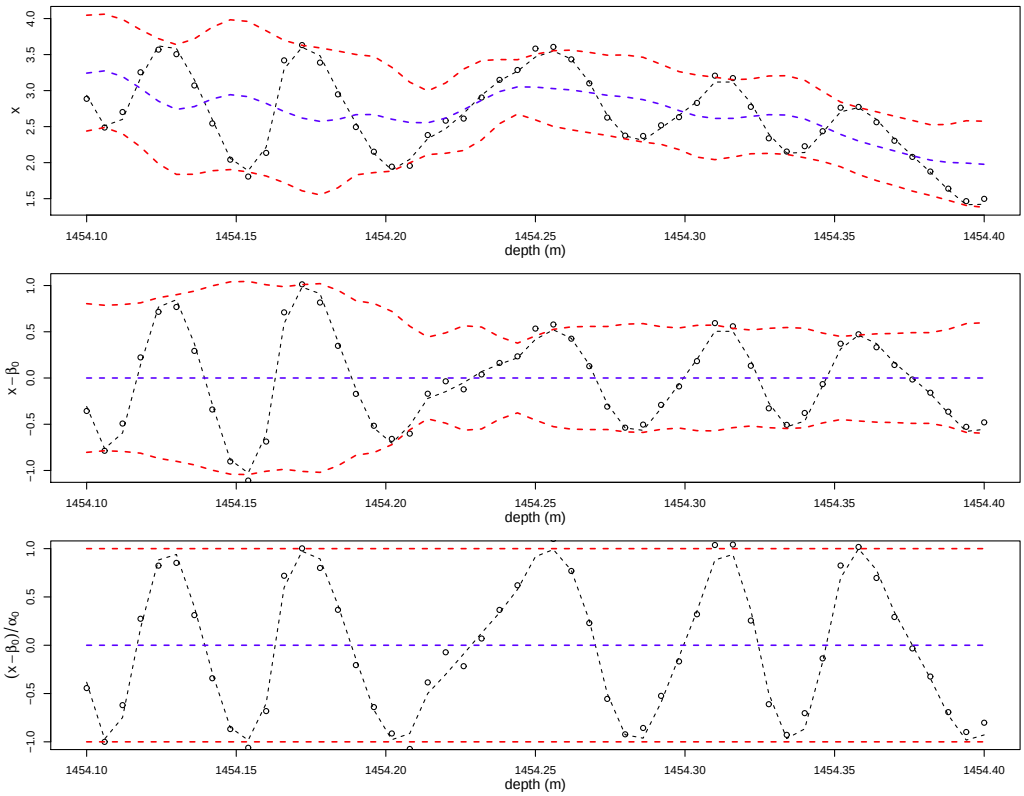

Figure B.8: Reconstructing the log NGRIP NH_4 ($\boldsymbol{x}$) signal (1454.1m–1454.4m), thinned to 6mm resolution, using its staring values. **(top)** $\boldsymbol{x}$ as circular points, the reconstruction of $\boldsymbol{x}$ as a black dashed line, the staring values for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the starting values for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

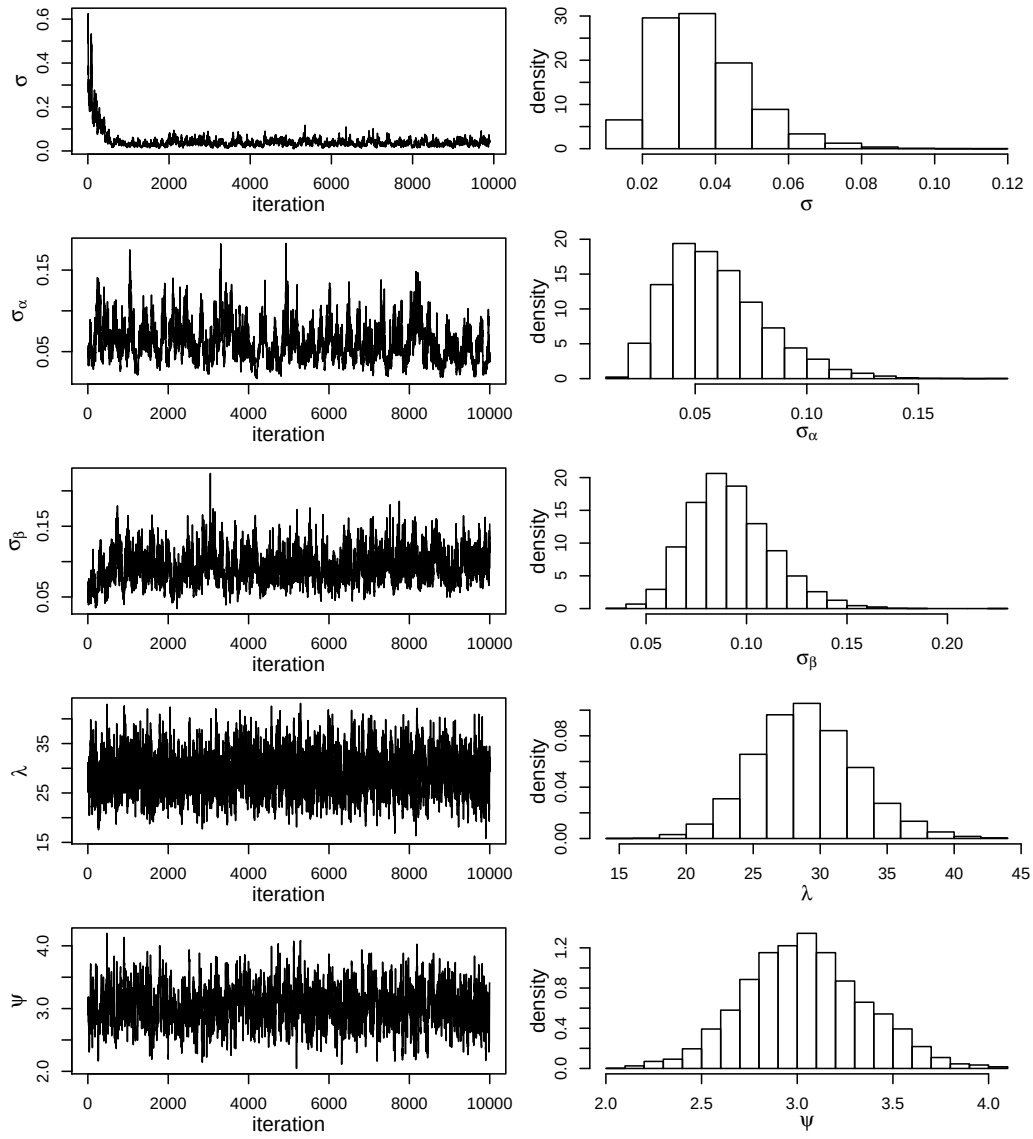

Figure B.9: Trace plots for model hyperparameters σ, σ_α, σ_β, λ and ψ from example 1, a $10,000$ iteration MCMC run on the log NGRIP NH_4 signal (1454.1m–1454.4m), with histograms showing their posterior distributions after a burn-in period of 1000 iterations.

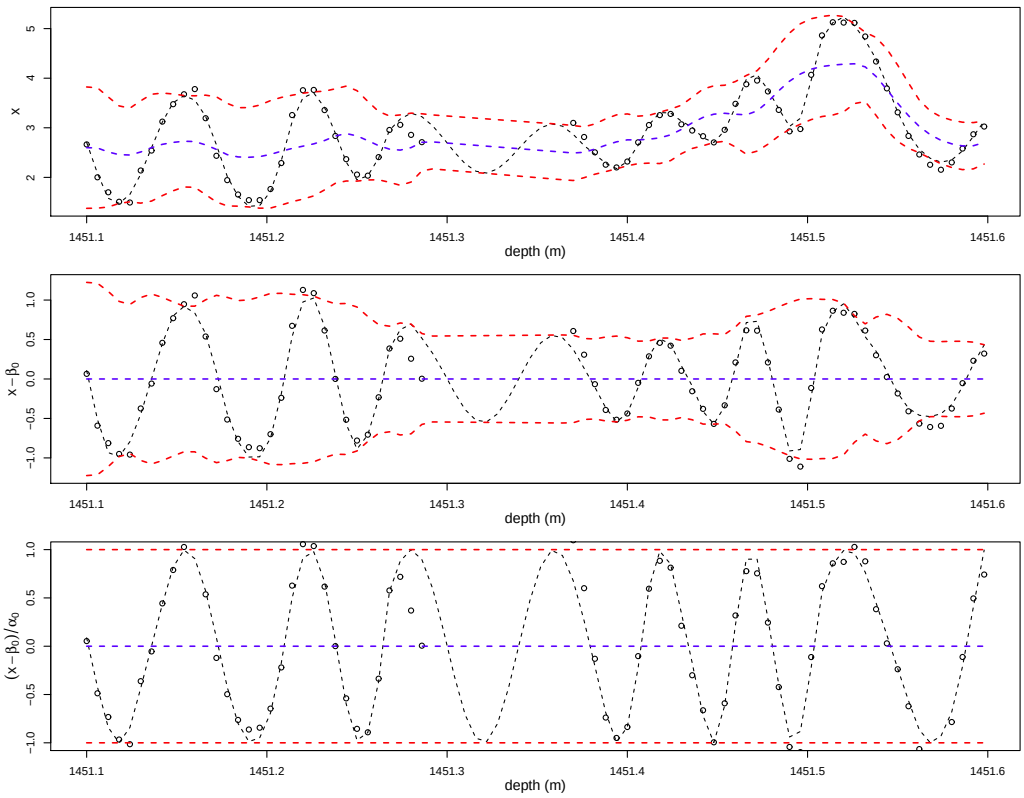

Figure B.10: Reconstructing the log NGRIP NH_4 ($\boldsymbol{x}$) signal (1454.1m–1454.4m), thinned to 6mm resolution, using its calculated staring values. **(top)** $\boldsymbol{x}$ as circular points, the reconstruction of $\boldsymbol{x}$ as a black dashed line, the staring values for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the starting values for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

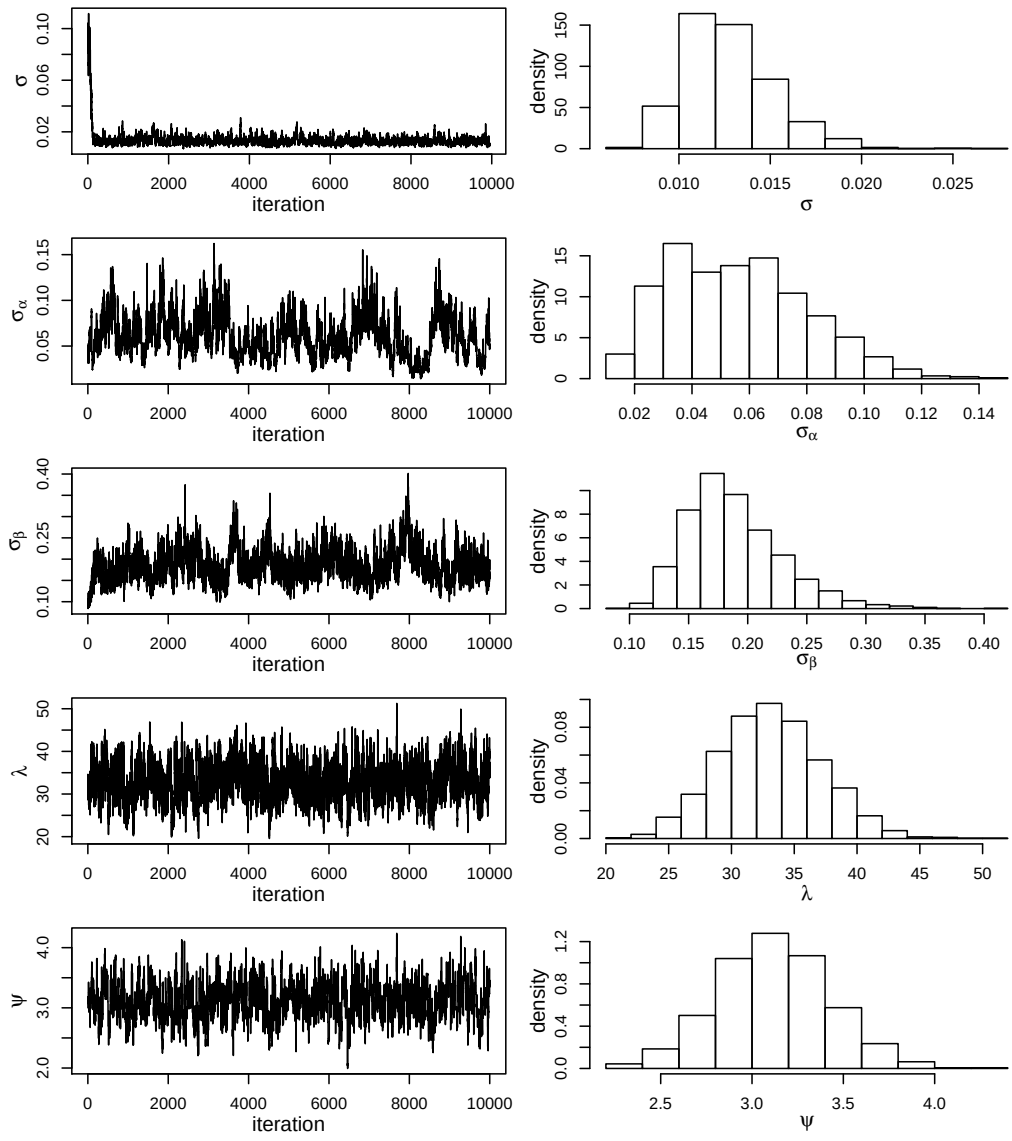

Figure B.11: Trace plots for model hyperparameters σ, σ_α, σ_β, λ and ψ from example 2, a $10,000$ iteration MCMC run on the log NGRIP NH_4 signal (1454.1m–1454.4m), with histograms showing their posterior distributions after a burn-in period of 1000 iterations.

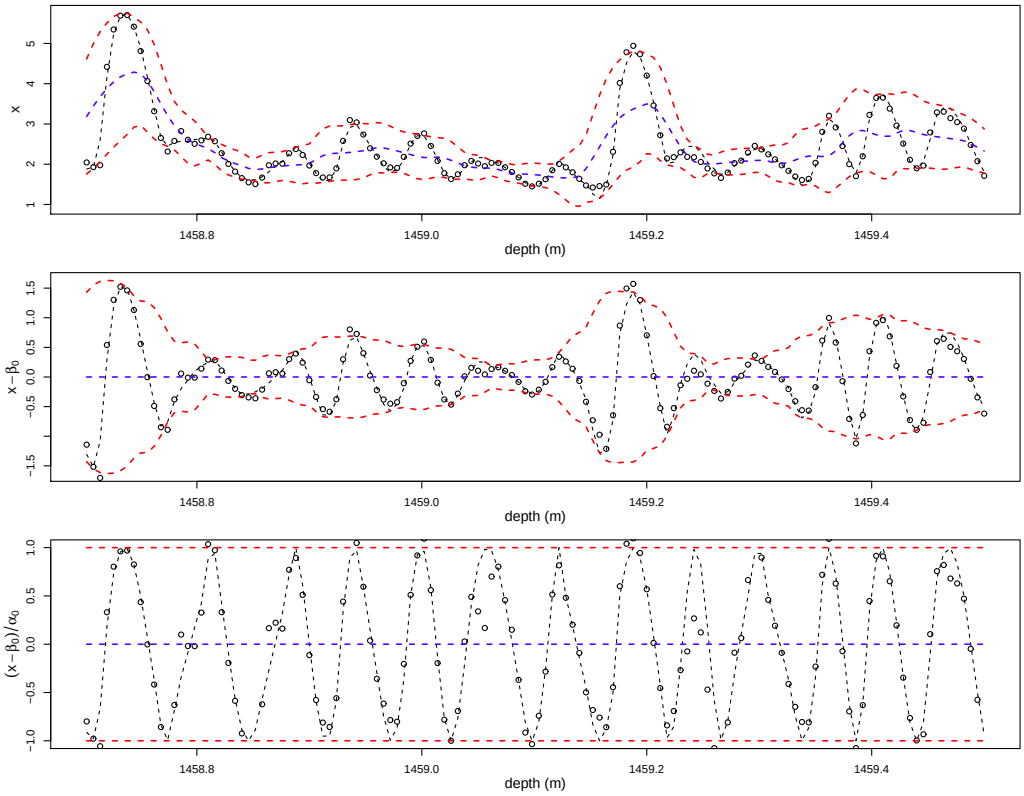

Figure B.12: Reconstructing the log NGRIP NH_4 ($\boldsymbol{x}$) signal (1458.7m–1459.5m), thinned to 6mm resolution, using its calculated staring values. **(top)** $\boldsymbol{x}$ as circular points, the reconstruction of $\boldsymbol{x}$ as a black dashed line, the staring values for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the starting values for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

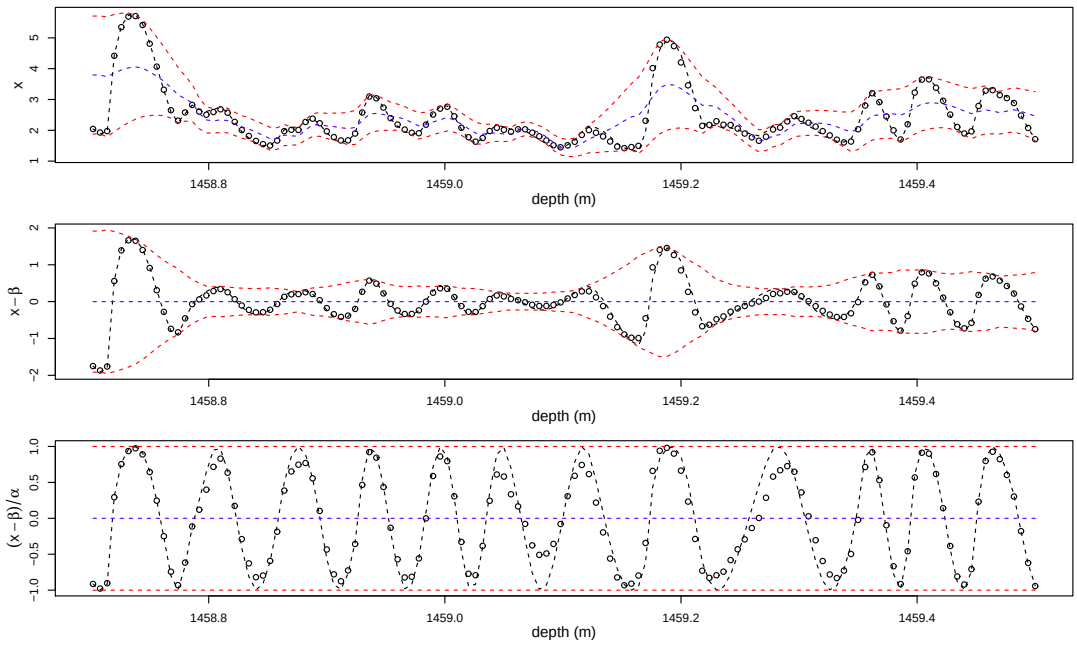

Figure B.13: The posterior mean model fit to the log NGRIP NH_4 ($\boldsymbol{x}$) signal (1458.7m–1459.5m), thinned to 6mm resolution, over iterations where the cycle count was 13. Calculated from an MCMC run of $10,000$ iterations with a burn-in of 1000. **(top)** $\boldsymbol{x}$ as circular points, the reconstruction of $\boldsymbol{x}$ as a black dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

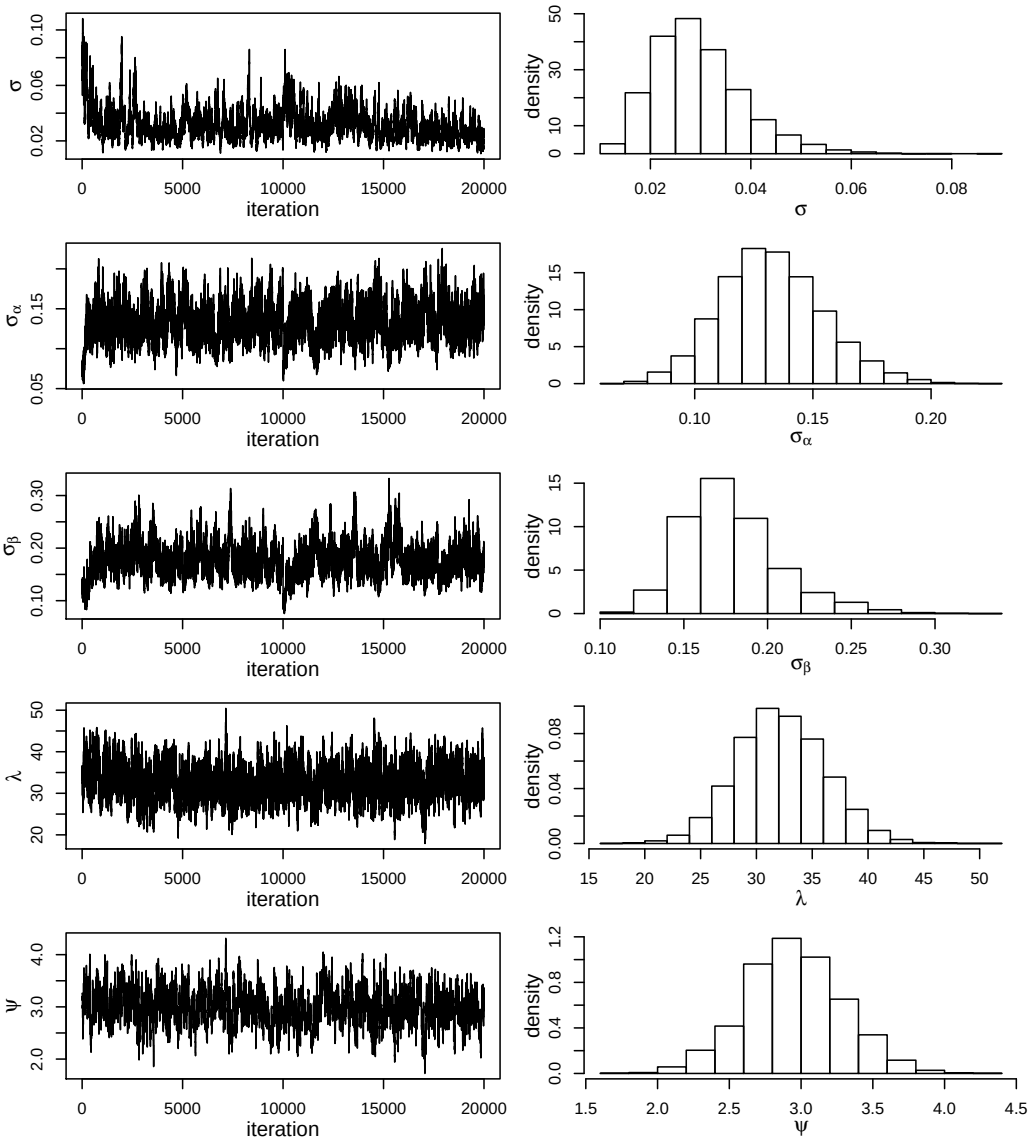

Figure B.14: Trace plots for model hyperparameters σ, σ_α, σ_β, λ and ψ from example 3, a $10,000$ iteration MCMC run on the log NGRIP NH_4 signal (1458.7m–1459.5m), with histograms showing their posterior distributions after a burn-in period of 1000 iterations.

B.4 Chapter 6

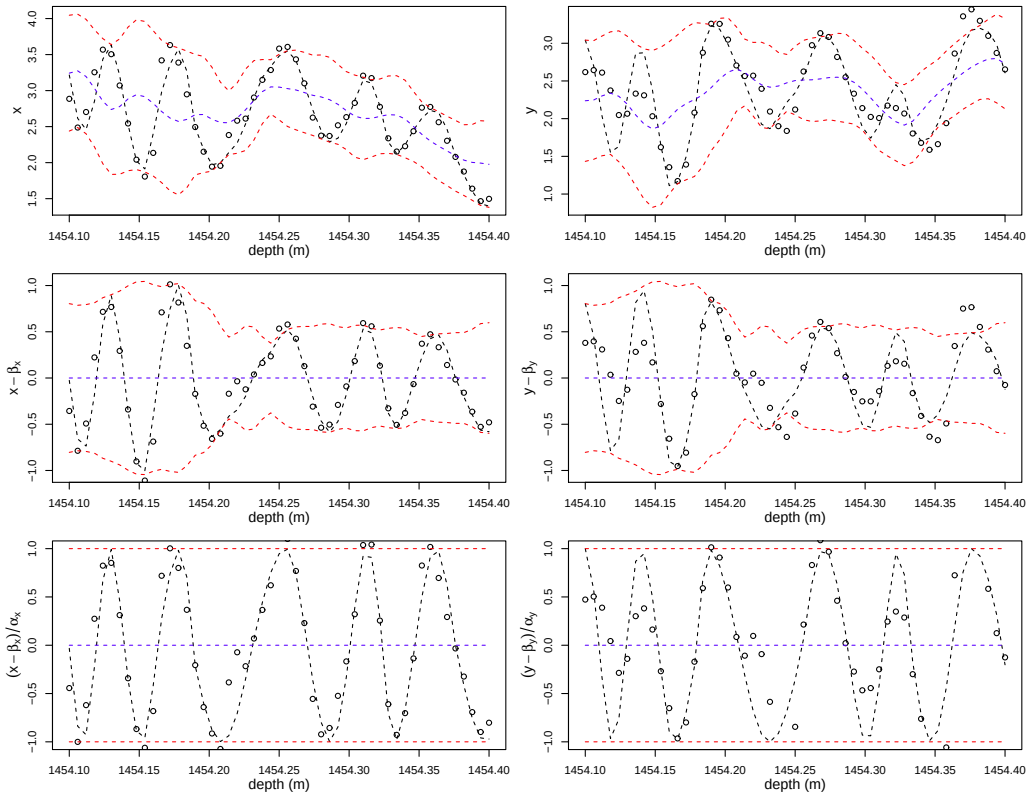

Figure B.15: The bivariate model fit jointly to the log NGRIP NH_4 ($\boldsymbol{x}$) and Ca ($\boldsymbol{y}$) signals (1454.1m–1454.4m), both thinned to 6mm resolution, using the staring values from example 1. The plots for $\boldsymbol{x}$ are on the left; those for $\boldsymbol{y}$ are on the right. **(top)** The signal as circular points, its reconstruction as a black dashed line, the starting value for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the starting value for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

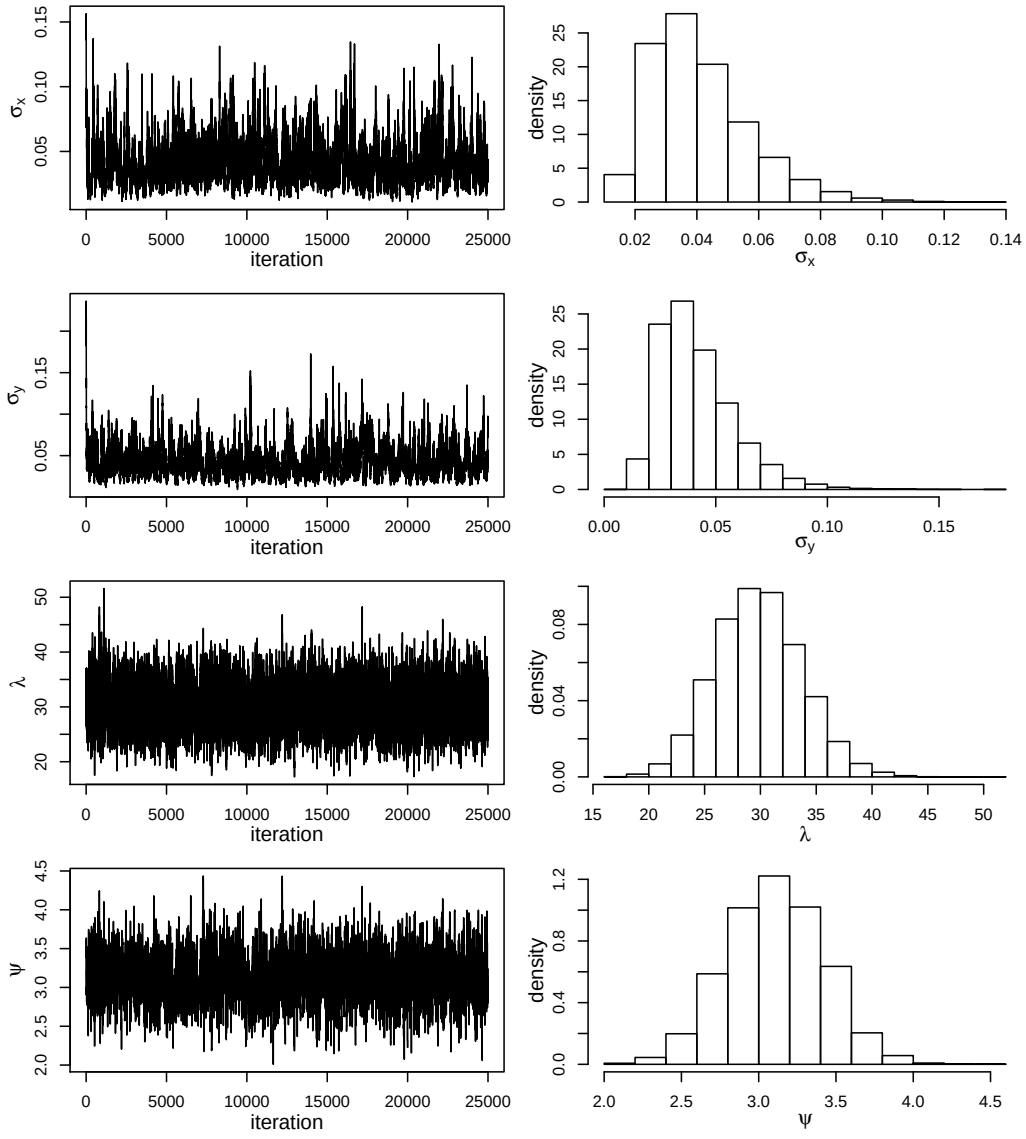

Figure B.16: Trace plots for model hyperparameters σ, σ_α, σ_β, λ and ψ from example 1, a 20,000 iteration bivariate MCMC run on the log NGRIP NH_4 and Ca signals (1454.1m–1454.4m), with histograms showing their posterior distributions after a burn-in period of 1000 iterations.

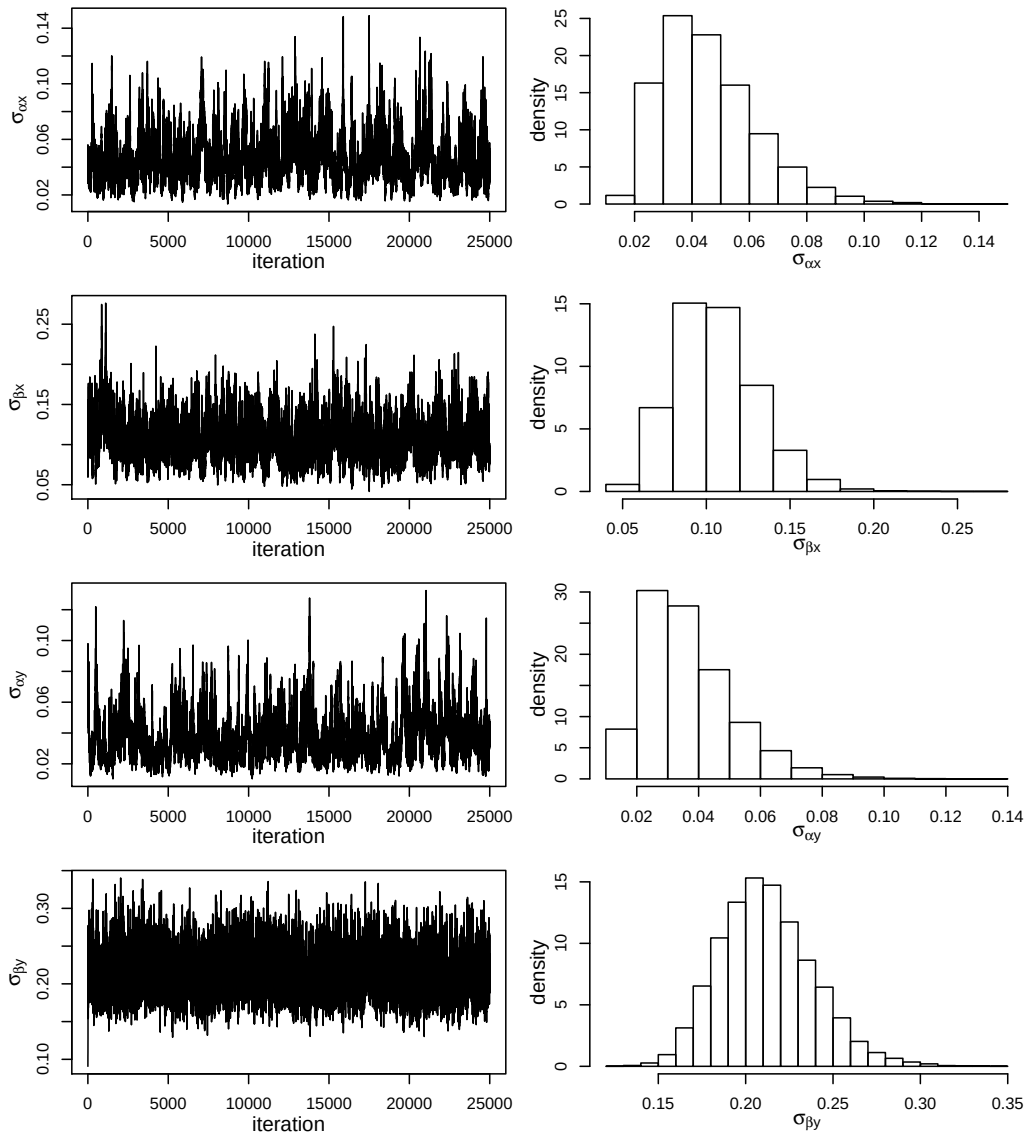

Figure B.17: Trace plots for model hyperparameters $\sigma_{\alpha,x}$, $\sigma_{\beta,x}$, $\sigma_{\alpha,y}$ and $\sigma_{\beta,y}$ from example 1, a 20,000 iteration bivariate MCMC run on the log NGRIP NH_4 and Ca signals (1454.1m–1454.4m), with histograms showing their posterior distributions after a burn-in period of 1000 iterations.

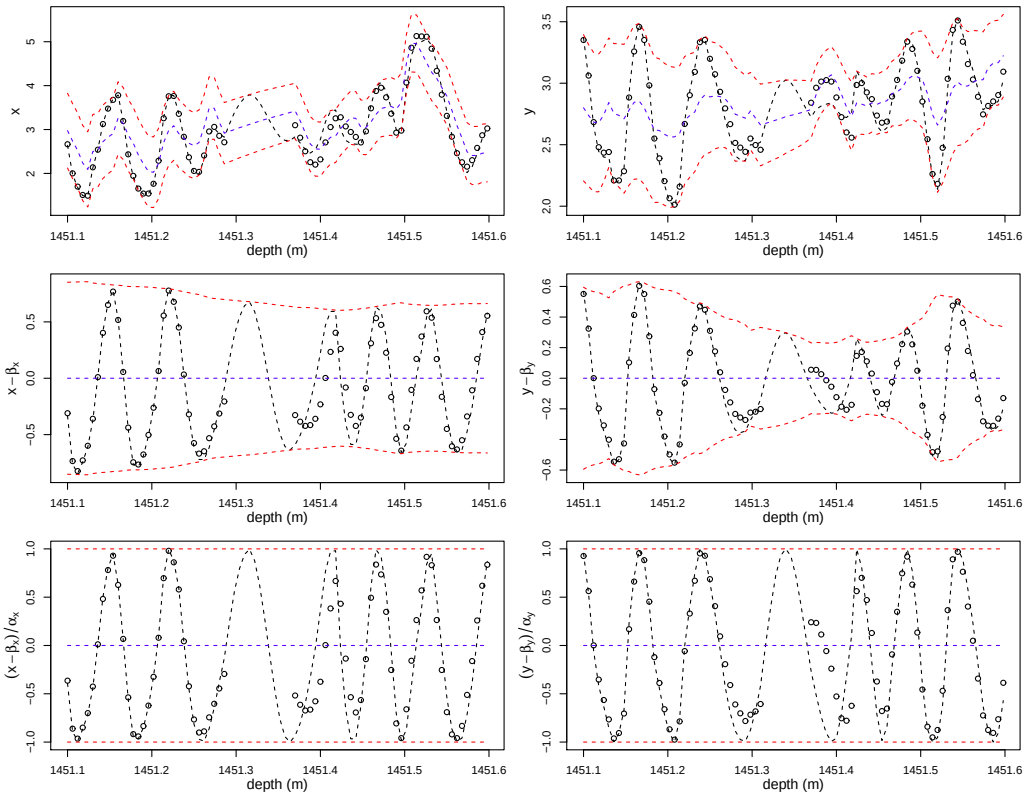

Figure B.18: The bivariate model fitted jointly to the log NGRIP NH_4 ($\boldsymbol{x}$) and Ca ($\boldsymbol{y}$) signals (1451.1m–1451.6m), both thinned to 6mm resolution, calculated from an MCMC run of $20,000$ iterations with a burn-in of 2000 where the cycle count was seven. The plots for $\boldsymbol{x}$ are on the left; those for $\boldsymbol{y}$ are on the right. **(top)** The signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

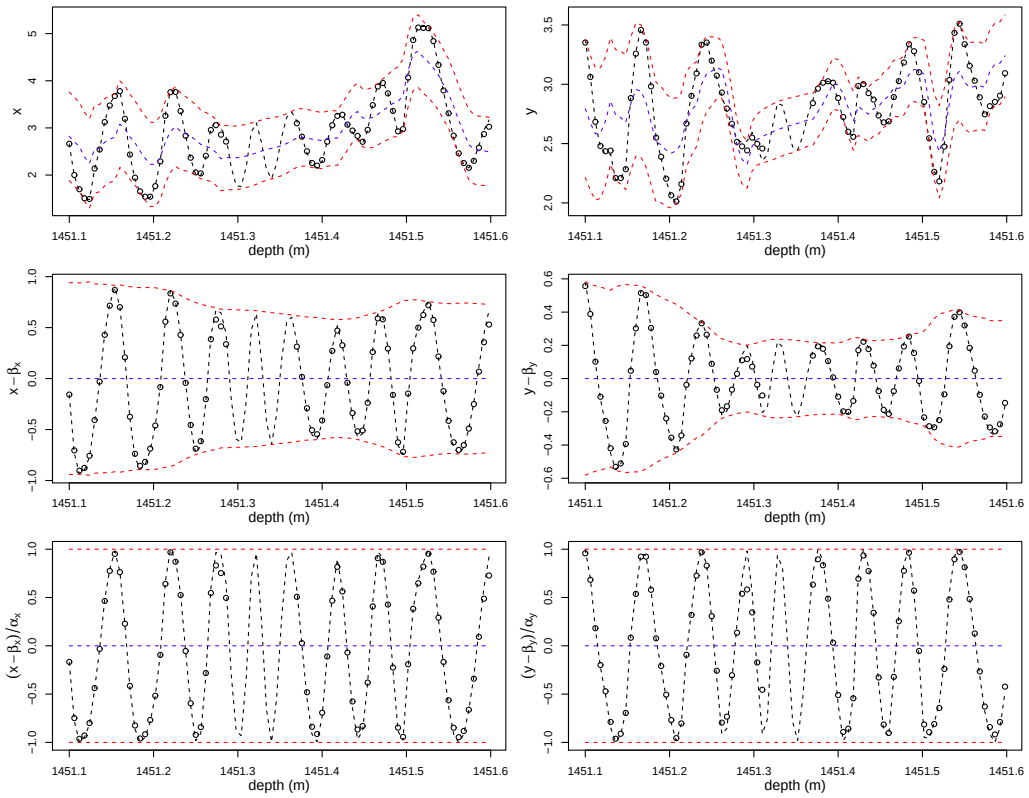

Figure B.19: The bivariate model fit jointly to the log NGRIP NH_4 ($\boldsymbol{x}$) and Ca ($\boldsymbol{y}$) signals (1451.1m–1451.6m), both thinned to 6mm resolution, calculated from an MCMC run of 20,000 iterations with a burn-in of 2000 where the cycle count was ten. The plots for $\boldsymbol{x}$ are on the left; those for $\boldsymbol{y}$ are on the right. **(top)** The signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

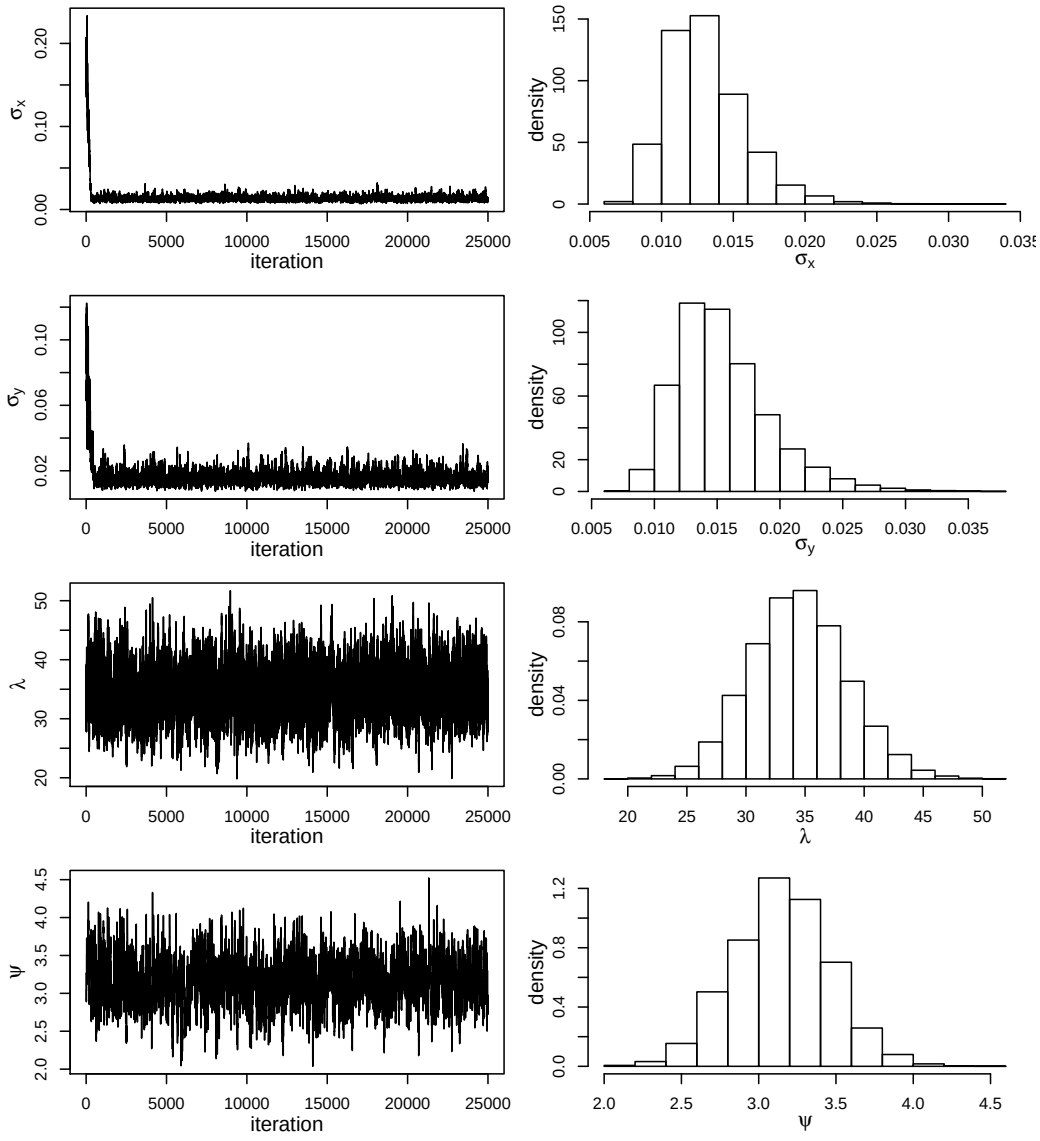

Figure B.20: Trace plots for model hyperparameters σ, σ_α, σ_β, λ and ψ from example 2, a $20,000$ iteration bivariate MCMC run on the thinned log NGRIP NH_4 and Ca signals (1451.1m–1451.6m), with histograms showing their posterior distributions after a burn-in period of 1000 iterations.

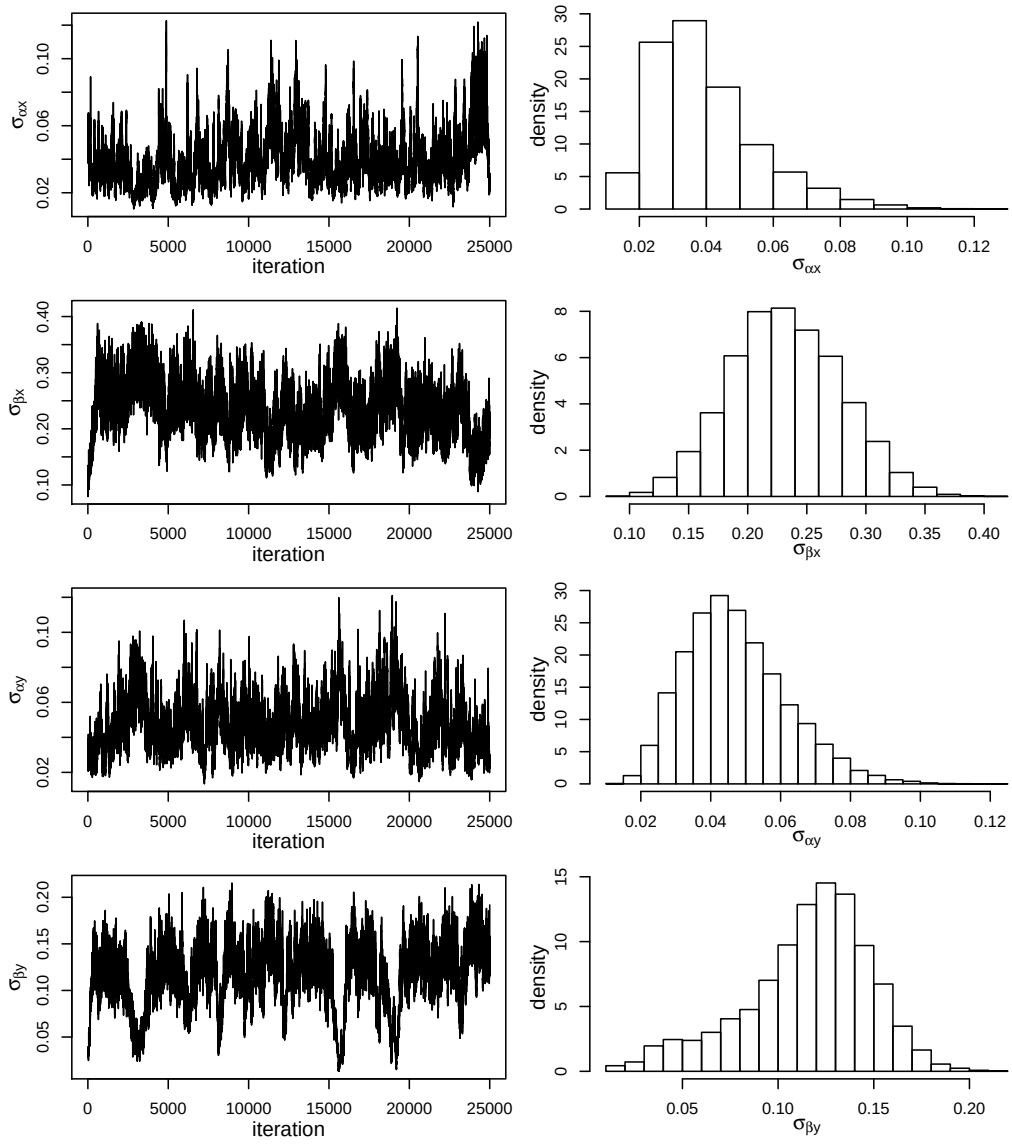

Figure B.21: Trace plots for model hyperparameters $\sigma_{\alpha,x}$, $\sigma_{\beta,x}$, $\sigma_{\alpha,y}$ and $\sigma_{\beta,y}$ from example 2, a 20,000 iteration bivariate MCMC run on the log NGRIP NH_4 and Ca signals (1451.1m–1451.6m), with histograms showing their posterior distributions after a burn-in period of 1000 iterations.

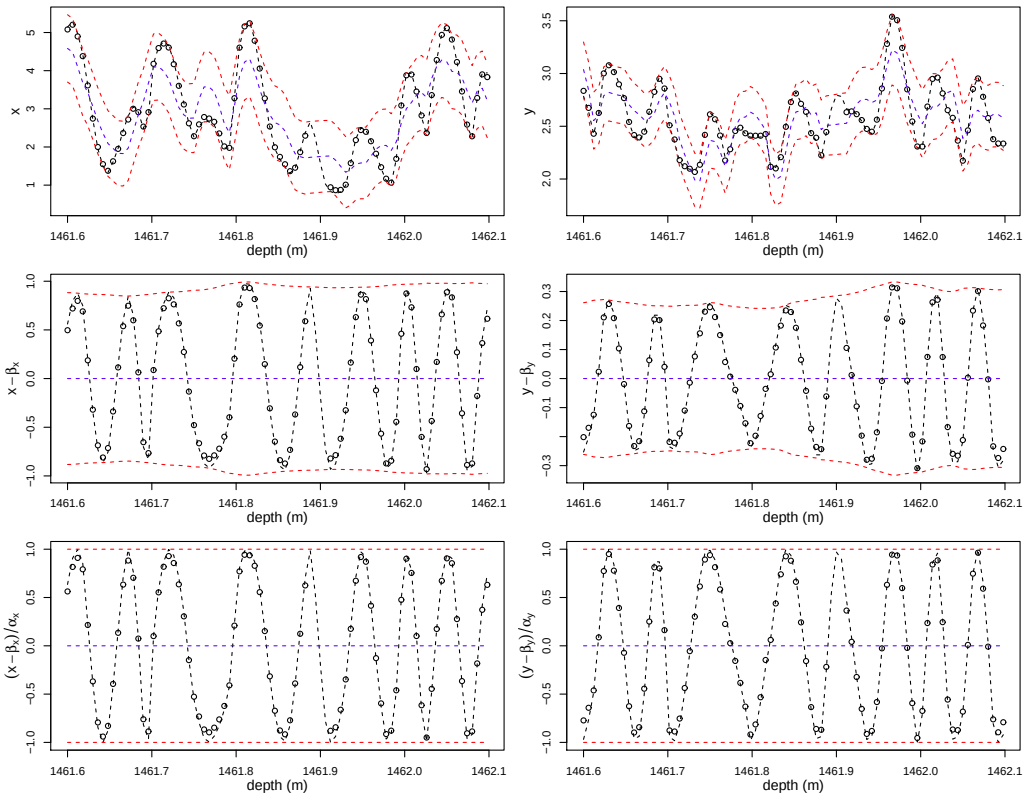

Figure B.22: The bivariate model fit jointly to the log NGRIP NH_4 ($\boldsymbol{x}$) and Ca ($\boldsymbol{y}$) signals (1461.6m–1462.1m), both thinned to 6mm resolution, calculated from an MCMC run of 20,000 iterations with a burn-in of 2000 where the cycle count was nine. The plots for $\boldsymbol{x}$ are on the left; those for $\boldsymbol{y}$ are on the right. **(top)** The signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\boldsymbol{\beta}$ as a blue dashed line, for $\boldsymbol{\beta} \pm \boldsymbol{\alpha}$ as two red dashed lines. **(middle)** The de-trended signal as circular points, its reconstruction as a black dashed line, the posterior mean for $\pm\boldsymbol{\alpha}$ as red dashed lines. **(bottom)** The de-trended and normalised signal as circular points, its reconstruction as a black dashed line.

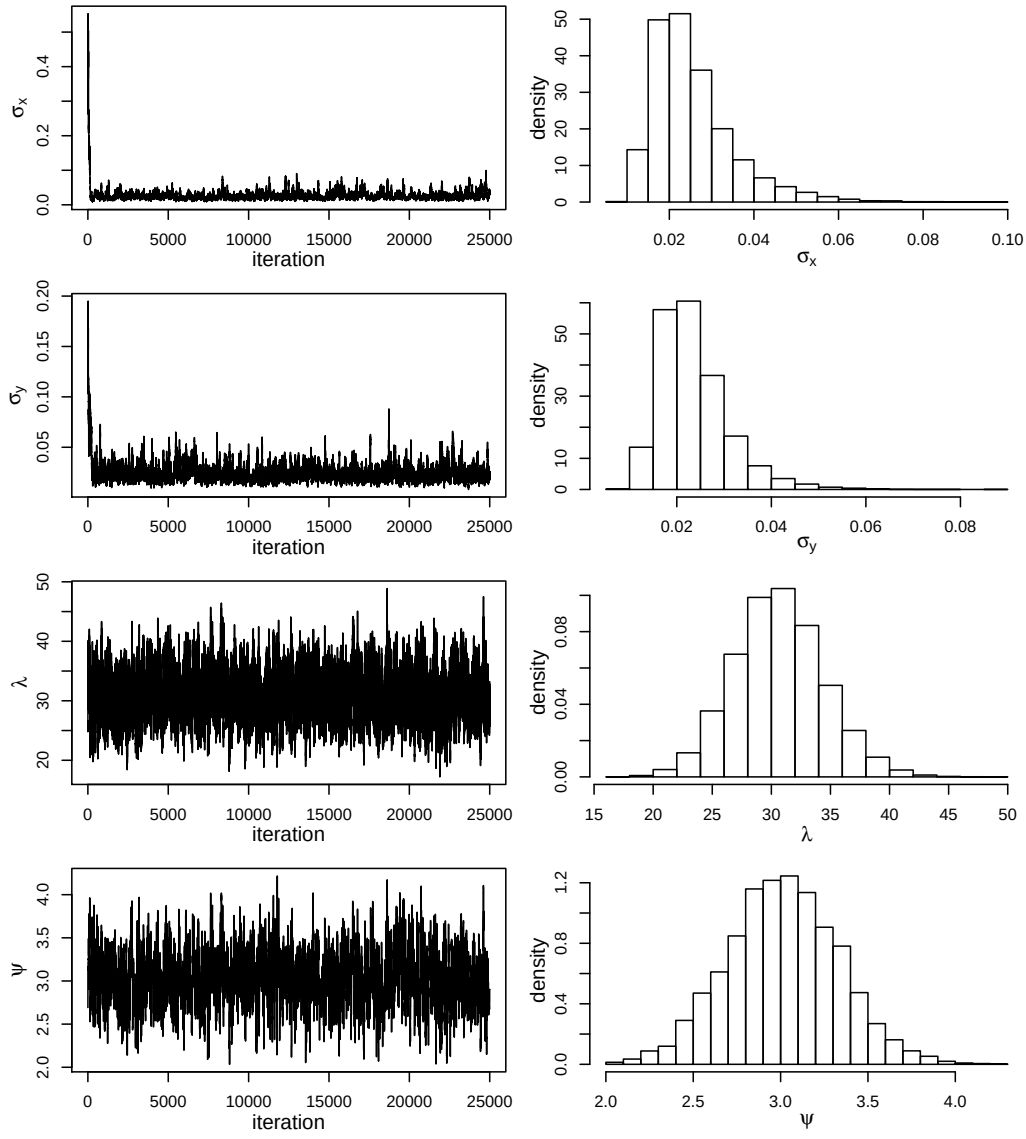

Figure B.23: Trace plots for model hyperparameters σ, σ_α, σ_β, λ and ψ from example 3, a 20,000 iteration bivariate MCMC run on the log NGRIP NH_4 and Ca signals (1461.6m–1462.1m), with histograms showing their posterior distributions after a burn-in period of 1000 iterations.

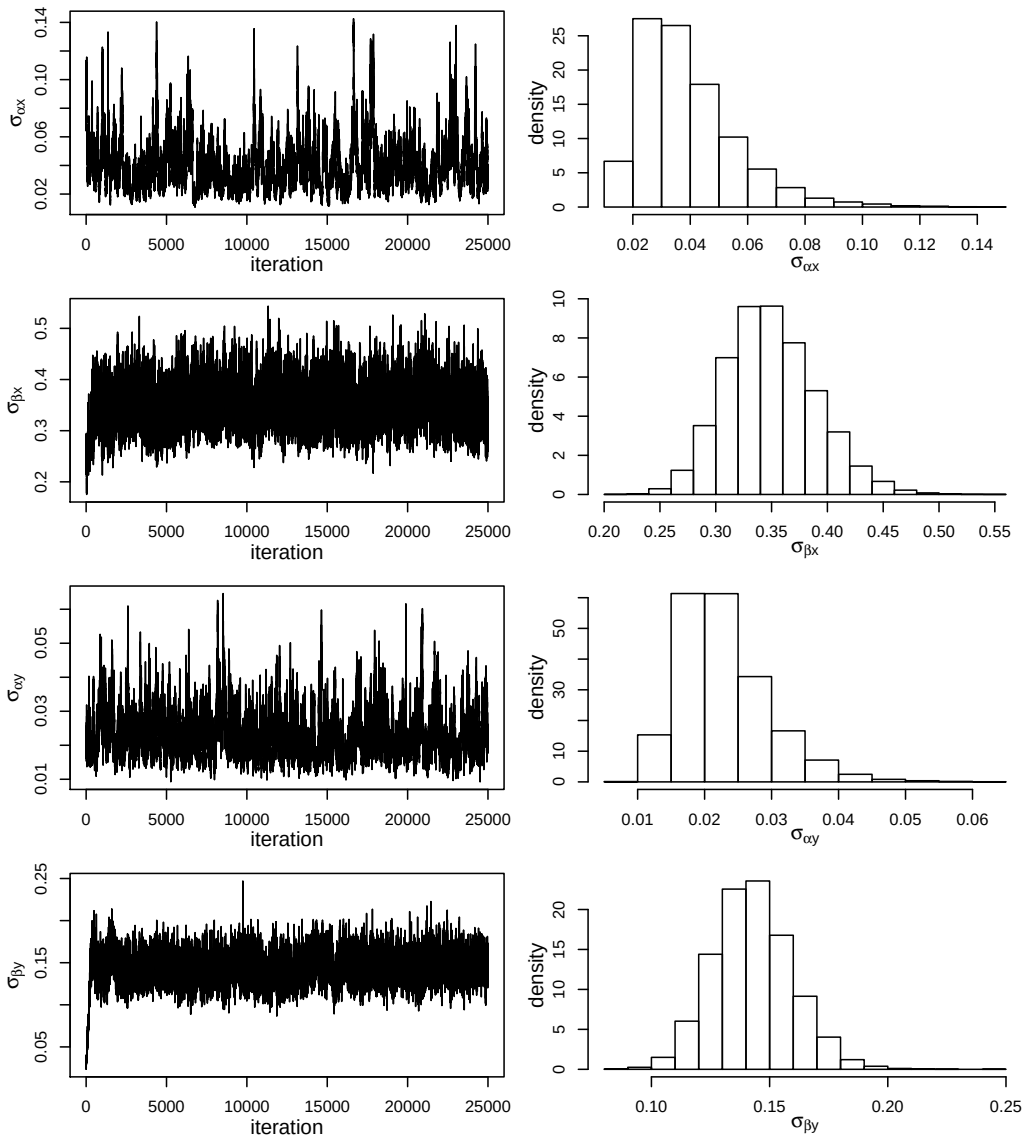

Figure B.24: Trace plots for model hyperparameters $\sigma_{\alpha,x}$, $\sigma_{\beta,x}$, $\sigma_{\alpha,y}$ and $\sigma_{\beta,y}$ from example 3, a 20,000 iteration bivariate MCMC run on the log NGRIP NH_4 and Ca signals (1461.6m–1462.1m), with histograms showing their posterior distributions after a burn-in period of 1000 iterations.

B.5 Chapter 7

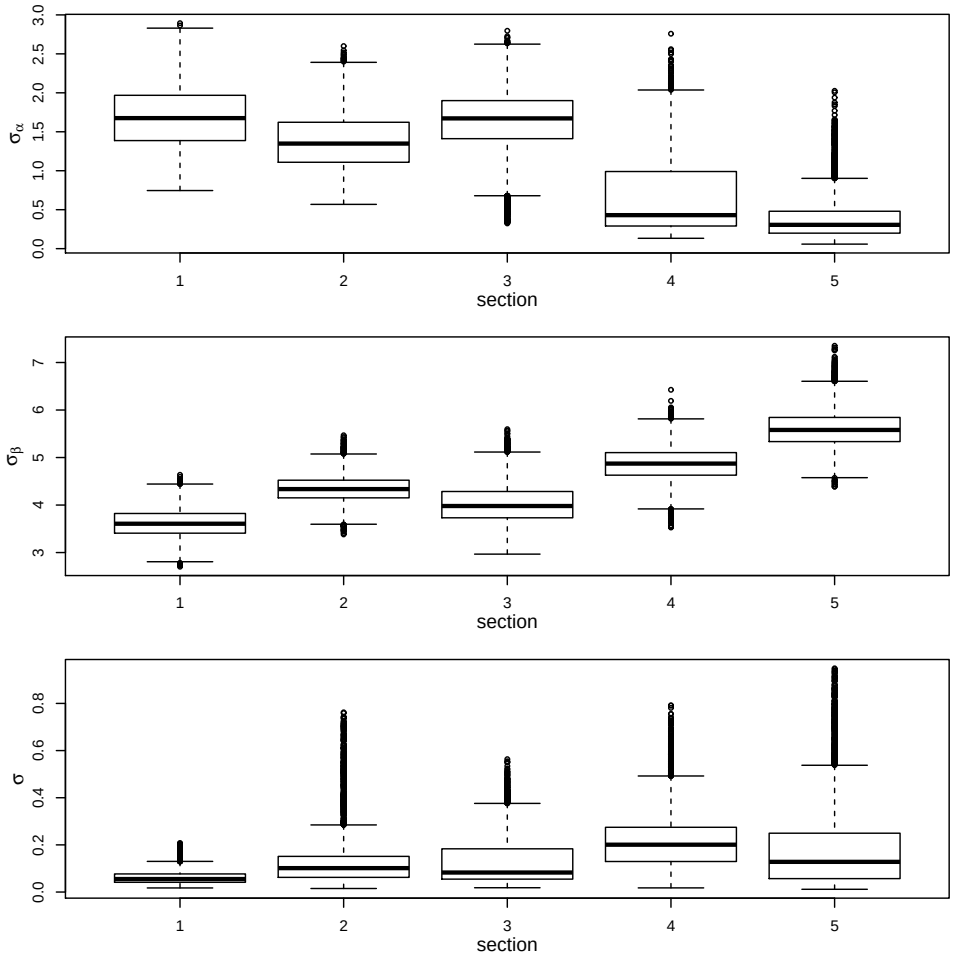

Figure B.25: Boxplots for the posterior distributions of the hyperparameters σ_α, σ_β and σ from independent MCMC runs on the Fletcher δ^{18}O signal, after being split into five overlapping subsections containing roughly the same number of cycles.

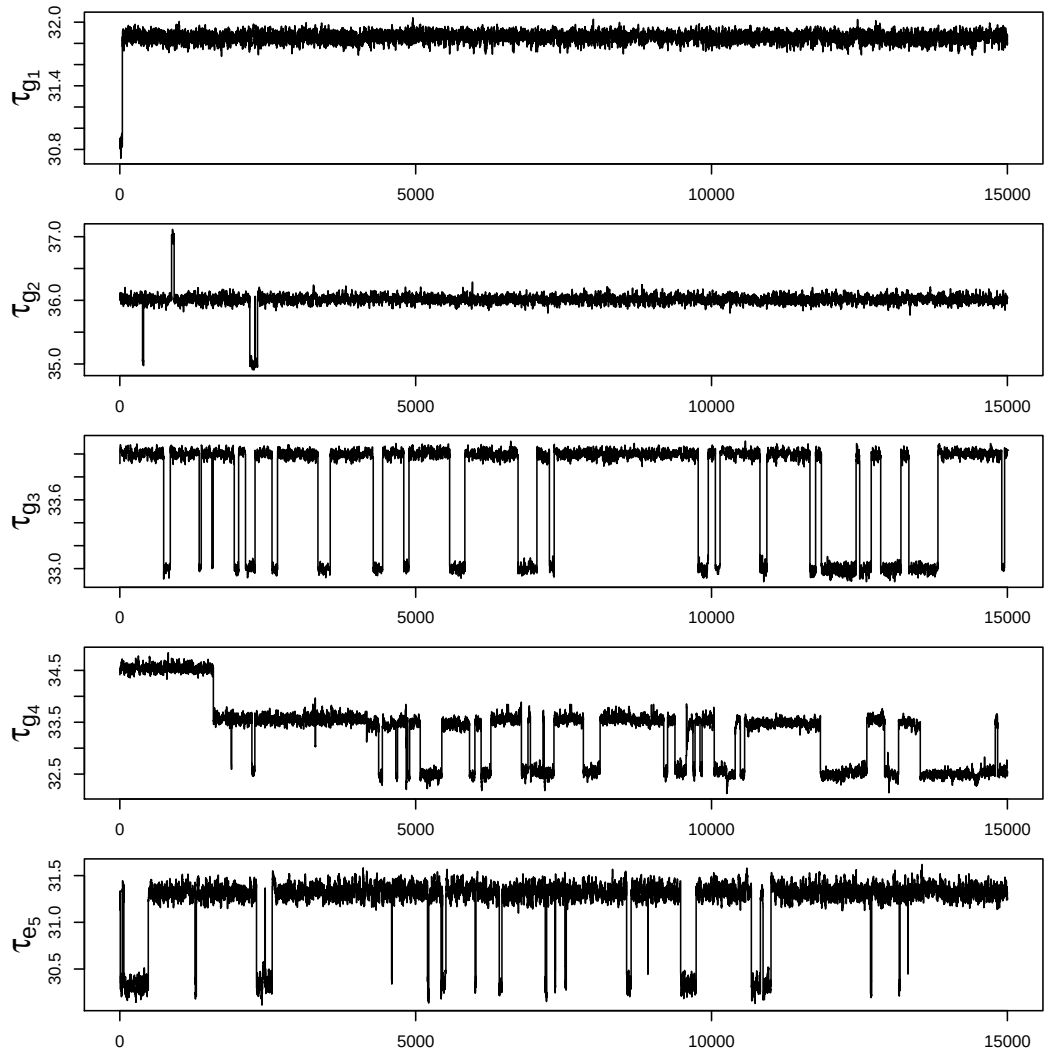

Figure B.26: Trace plots for the timescale evaluated at joining indexes g_k for sections $k \in (1, \ldots, 4)$, and at the final index e_5 for section from independent MCMC runs on the Fletcher δ^{18}O signal after being split into five overlapping subsections containing roughly the same number of cycles.

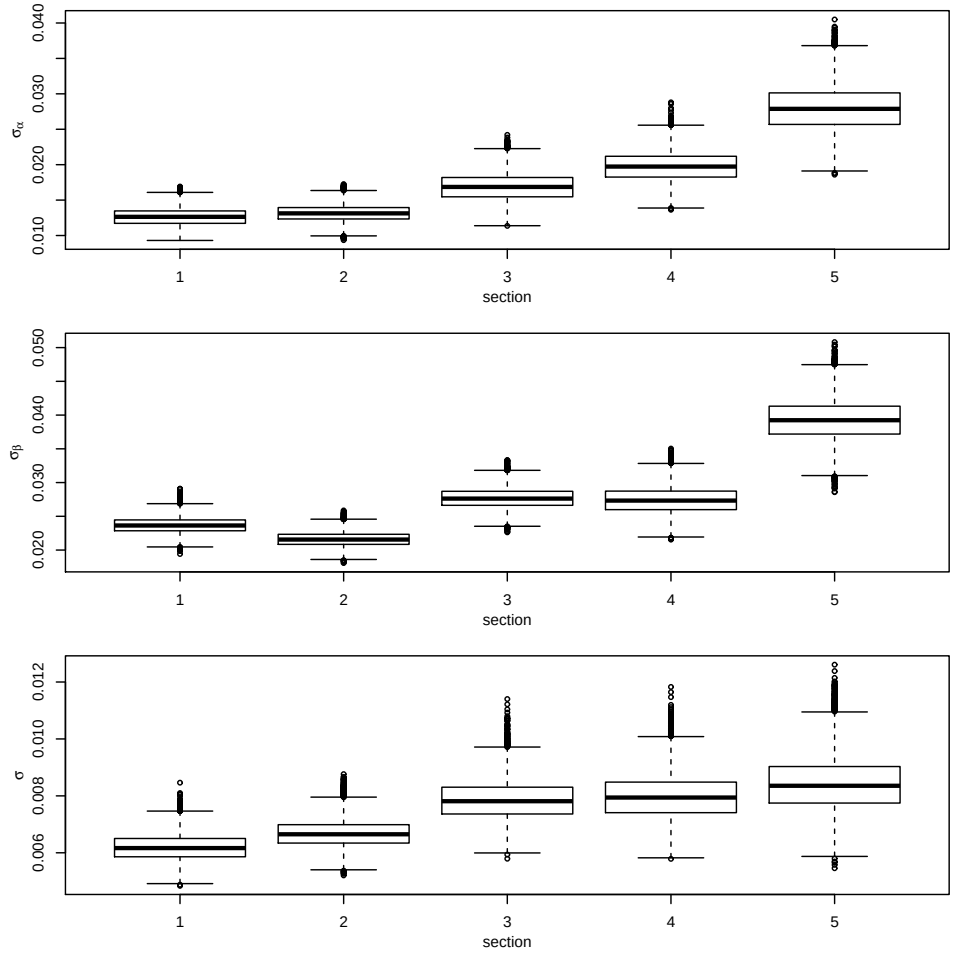

Figure B.27: Boxplots for the posterior distributions of the hyperparameters σ_α, σ_β and σ from independent MCMC runs on the Gomez H_2O_2 signal after being split into five overlapping subsections containing roughly the same number of cycles.

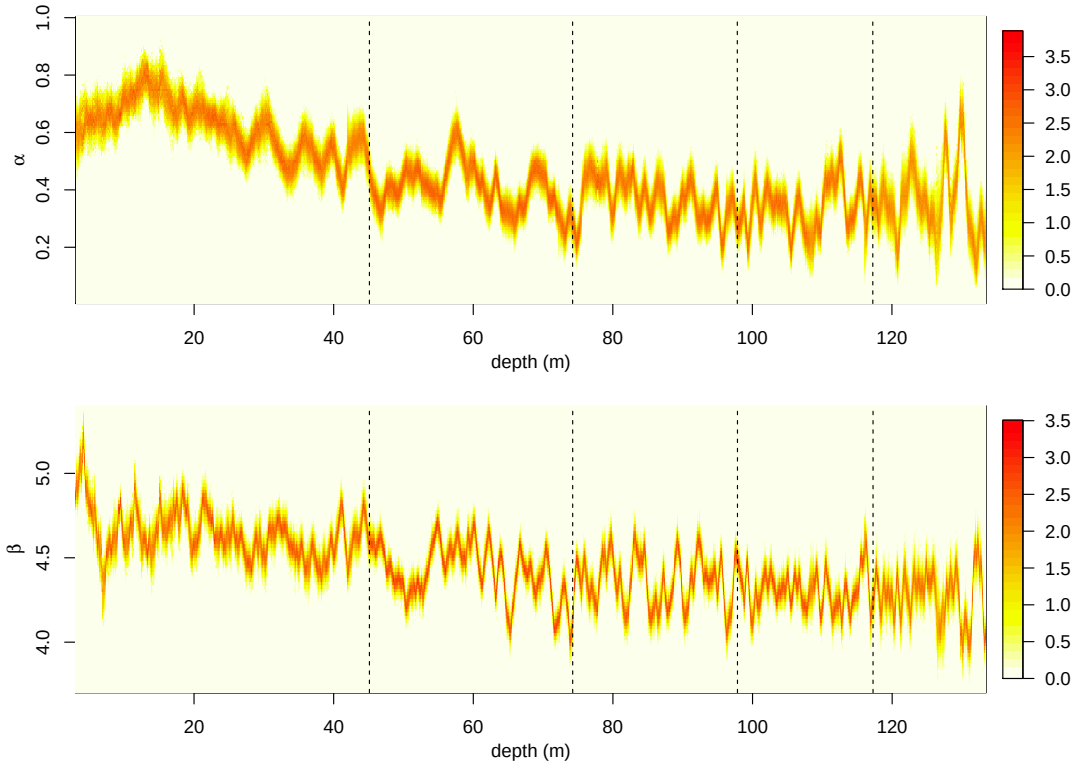

Figure B.28: Density heat plots of the posterior distributions of the amplitude and mean processes, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, from independent MCMC runs on the Gomez H_2O_2 signal after being split into five overlapping subsections containing roughly the same number of cycles. The vertical dotted lines show the joining depths for the five sections.

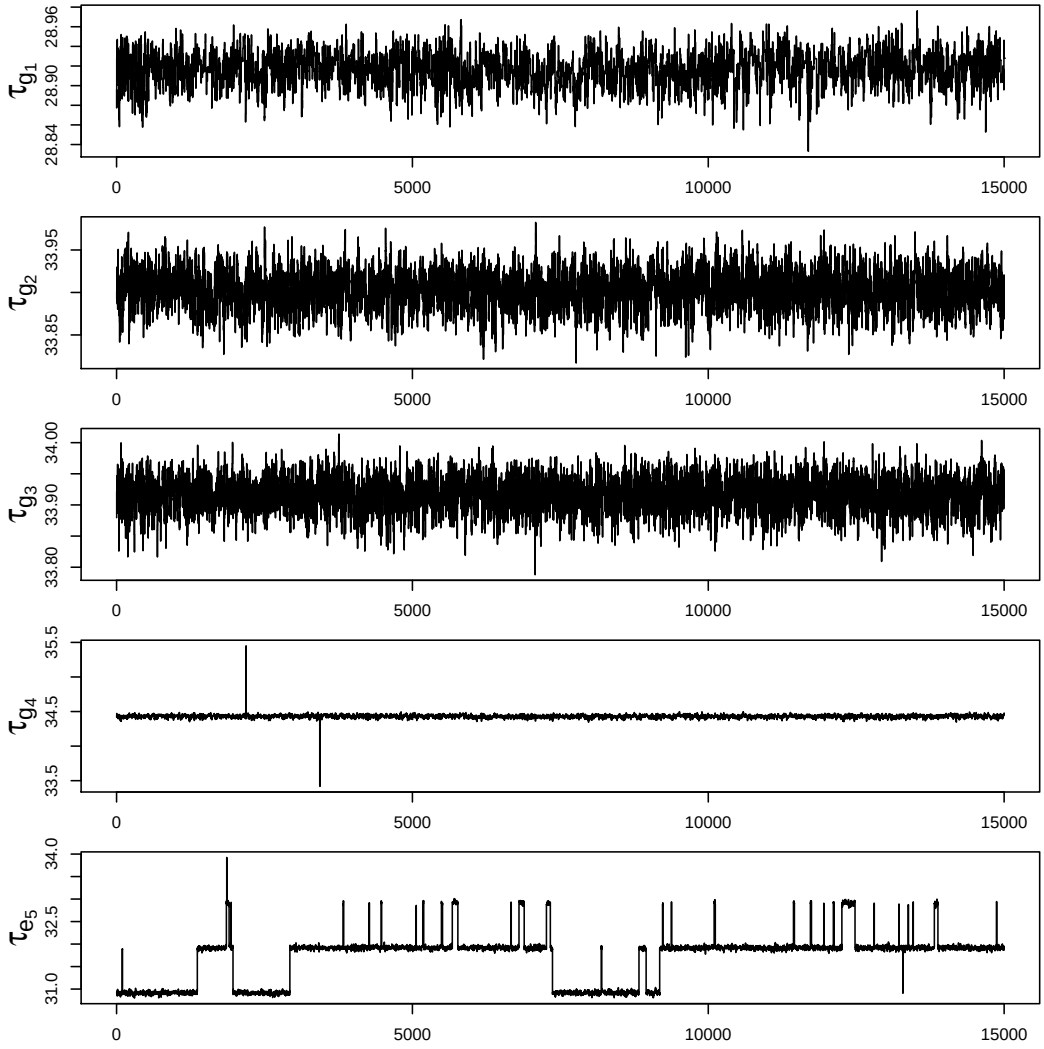

Figure B.29: Trace plots for the timescale evaluated at joining indexes g_k for sections $k \in (1, \ldots, 4)$, and at the final index e_5 for section from independent MCMC runs on the Gomez H_2O_2 signal after being split into five overlapping subsections containing roughly the same number of cycles.

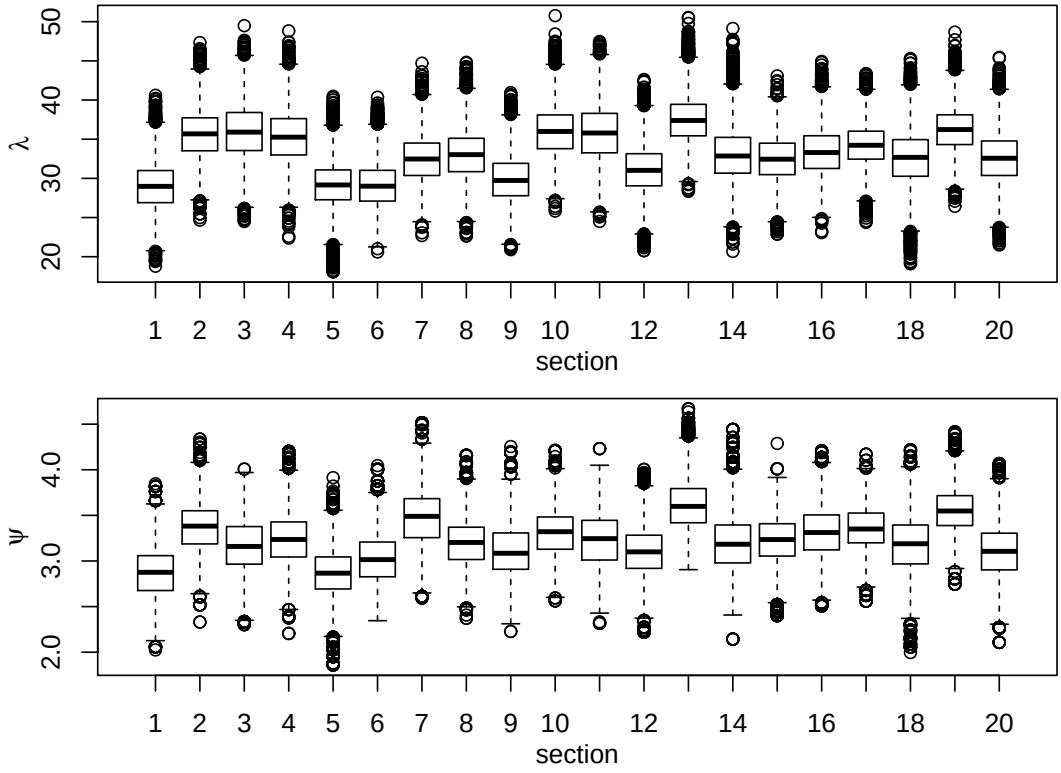

Figure B.30: Boxplots for the posterior distributions of ψ and λ, the timescale hyperparameters, from independent bivariate MCMC runs on the NGRIP ammonium and calcium signals (1440m–1465m) after being split into 20 overlapping subsections containing roughly the same number of cycles.

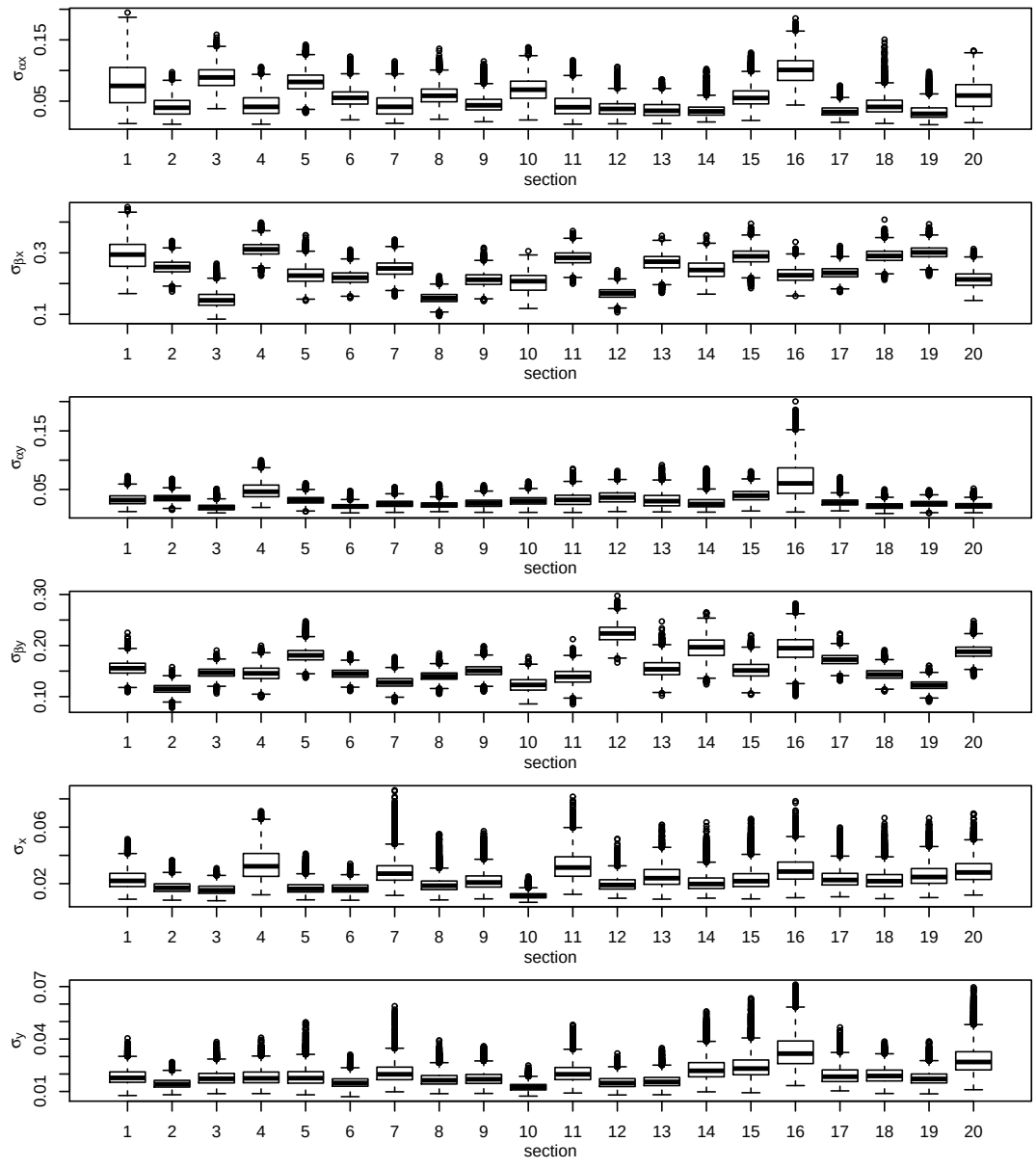

Figure B.31: Boxplots for the posterior distributions of the hyperparameters σ_α, σ_β and σ from independent bivariate MCMC runs on the NGRIP ammonium and calcium signals (1440m–1465m) after being split into 20 overlapping subsections containing roughly the same number of cycles.

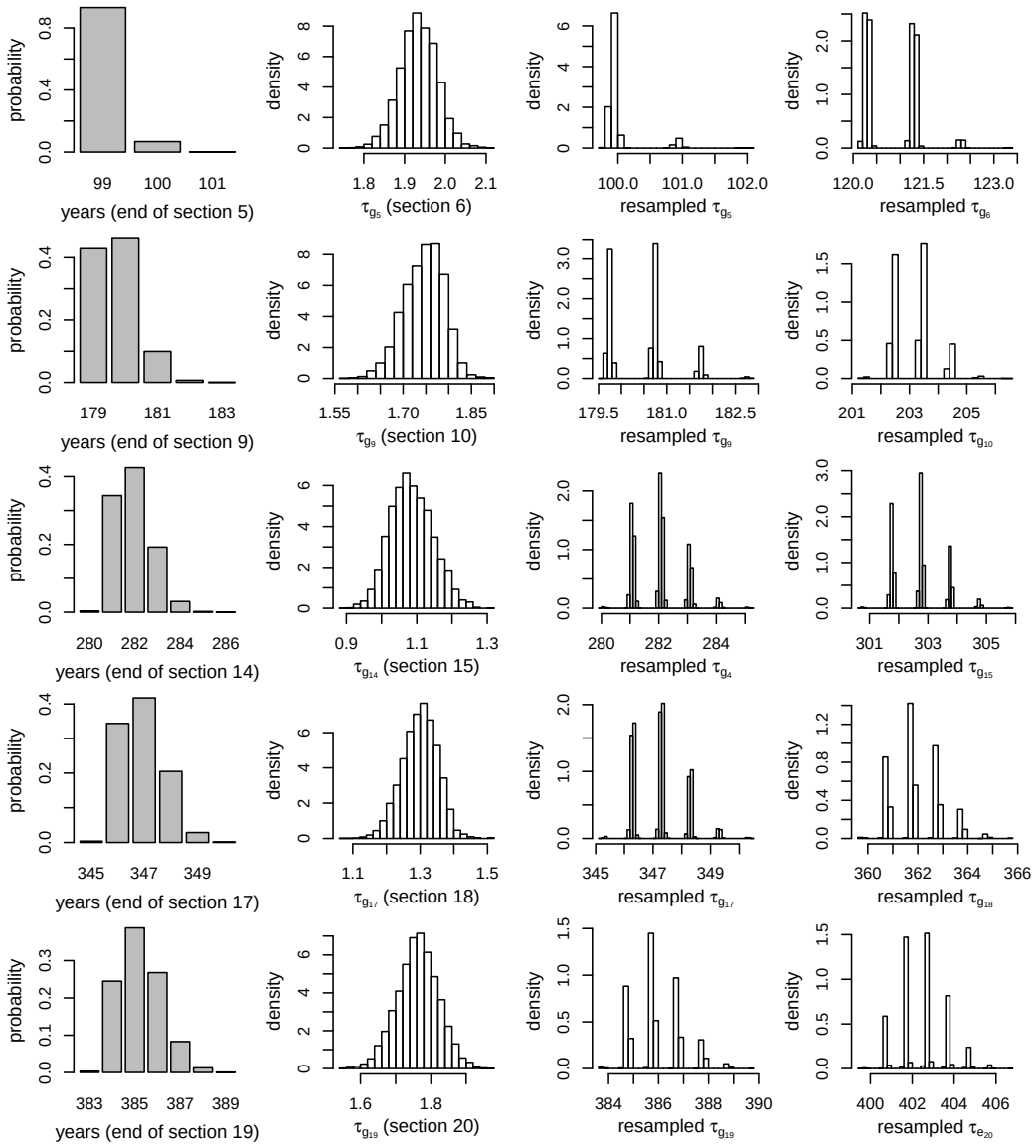

Figure B.32: An illustration of the re-sampling process used to join together the posterior distributions of $\boldsymbol{\tau}$, from independent MCMC runs on the NGRIP NH_4 and Ca signals (1440m–1465m) signal split into for the 20 overlapping sections. See the main text for details.