



Automated ice-core layer-counting with strong univariate signals

J. J. Wheatley¹, P. G. Blackwell¹, N. J. Abram², J. R. McConnell³, E. R. Thomas⁴, and E. W. Wolff⁴

¹School of Mathematics and Statistics, The University of Sheffield, UK

²Research School of Earth Sciences, Australian National University, Australia

³Desert Research Institute, 2215 Raggio Parkway, Reno, Nevada, 89512-1095, USA

⁴British Antarctic Survey, Cambridge, UK

Correspondence to: J. J. Wheatley (j.wheatley@shef.ac.uk)

Received: 7 June 2012 – Published in Clim. Past Discuss.: 3 July 2012

Revised: 12 October 2012 – Accepted: 17 October 2012 – Published: 22 November 2012

Abstract. We present an automated process for determining the annual layer chronology of an ice-core with a strong annual signal, utilising the hydrogen peroxide record from an Antarctic Peninsula ice-core as a test signal on which to count annual cycles and explain the methods. The signal is de-trended and normalised before being split into sections with a deterministic cycle count and those that need more attention. Possible reconstructions for the uncertain sections are determined which could be used as a visual aid for manual counting, and a simple method for assigning probability measures to each reconstruction is discussed. The robustness of this process is explored by applying it to versions of two different chemistry signals from the same stretch of the NGRIP (North Greenland Ice Core Project) ice-core, which shows more variation in annual layer thickness, with and without thinning to mimic poorer quality data. An adapted version of these methods is applied to the more challenging non-sea-salt sulphur signal from the same Antarctic Peninsula core from which the hydrogen peroxide signal was taken. These methods could readily be adapted for use on much longer datasets, thereby reducing manual effort and providing a robust automated layer-counting methodology.

1 Introduction

Ice-cores hold information about the Earth's past climate. This information is recorded by chemical and physical signals in the ice that reflect the environmental conditions when the ice was deposited. These signals are measured as a depth series, but need to be modelled against time. In some cases the chemical or isotopic signals measured from

the melt-water of ice-cores have annual cycles which can be counted to provide a chronology. Traditionally this has been undertaken by visual assignment of annual markers, with no formal systematic method for ensuring consistency or assessing uncertainty. In some cases, ice-core annual layer chronologies have been counted manually for data spanning tens of thousands of years (see Andersen et al., 2006) requiring months to years of effort. This highlights a clear need to develop an automated, statistically based methodology for application in this type of physical sciences research.

Previous attempts at developing automated cycle counting methods in the literature have generally only provided point estimates with no measure of uncertainty. Weber et al. (2010) apply automated methods for determining annual peaks and troughs to high-resolution signals induced from scanned images of tree rings, marine varves, and marine laminae – more discussion of these tools can be found in Sect. 8. Rasmussen et al. (2002) use independent component analysis to combine chemistry data in conjunction with a low-pass filter; the resulting signal is dated by counting peaks that rise above a constant threshold value. Other methods range from iteratively picking the most likely annual peaks along a core with respect to their spacing and magnitude (Taylor et al., 2004; McGwire et al., 2011) to measuring periodicities in the signals (Rupf and Radons, 2004; Svensson et al., 2005).

Uncertainty has been addressed under the Bayesian framework, applying hidden Markov models (Winstrup et al., 2012) and Markov chain Monte Carlo methods (Wheatley et al., 2012) to univariate models. The method presented here has the advantage over these approaches that it is sufficiently straightforward computationally to use on full size cores, whilst still providing a measure of uncertainty.

In this context it has to be pointed out that in the following our (as well as any other) counting method considers the records provided by some measurement technique and used for the counting to be complete. That means that we do not consider signal loss due to the true resolution of the measurement, which may be lower than the sampling frequency, or loss of information because the ice core record consists of individual snow fall events and is subject to wind reworking. Thus, annual cycles that are lost in this process cannot be counted by any method and will generally lead to a systematic error of the layer count to smaller numbers.

Hydrogen peroxide (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. Because of 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.

In this study we utilise the H_2O_2 record from an Antarctic Peninsula ice-core as a test signal on which to count annual cycles and explain the methods. We use this signal as it clearly has the strongest annual component and is therefore most reliable, and is therefore the most realistic choice within this dataset. The Gomez core extends to 134 m (Thomas et al., 2008). We use 2 cm average H_2O_2 concentrations determined from continuous measurements along the core. This core has been manually dated to 153 yr by determining the depths of the lowest points (“nadirs”) of the annual troughs. We present a simple method based on these data for counting cycles which is fully automated, requires no prior manual assessment of the chronology, and assigns probability measurements to the cycle count.

Robustness is explored by applying our method to thinned and unthinned versions of the ammonium and calcium signals from a stretch of the NGRIP ice-core (Dahl-Jensen et al., 2002), which shows more variation in annual layer thickness than the Gomez core; see Sect. 6. A more robust version of our method is applied to the non-sea-salt sulphur signal from the Gomez core as an example of how our general framework can be adapted to asymmetric data; see Sect. 7.

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; as explained in Sect. 4.2 this is purely for notational convenience and is not inherent in the method.

2 Data pre-processing

The Gomez H_2O_2 signal, indexed by its depth i , is transformed by taking logarithms to improve the symmetry in its annual periodicity, giving the log signal $\mathbf{x} = \{x_i\}$.

In order to simplify the process of cycle counting, \mathbf{x} is detrended and the annual cycle amplitudes are set to approximately 1. This is achieved by smoothing the signal with respect to an initial rough estimate of the average cycle length at each depth; see Sect. 2.1 below for details. While hydrogen peroxide seems to (and is expected to) yield a sine wave, the precise shape of the normalised signal has little effect on the methods presented below. The underlying assumption is that the seasonality follows a ...peak, descending, trough, ascending... pattern.

2.1 Smoothing

Band-pass or low-pass filters are frequently used to remove cycles with periods deemed to be larger or smaller than expected (see Rasmussen et al., 2002; Taylor et al., 2004; Rupf and Radons, 2004; McGwire et al., 2011). In Shimohara et al. (2003) the data are smoothed by taking a moving average over a range of half an estimated annual layer thickness. Here we use a point-wise standardisation: at each depth i

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

where $\boldsymbol{\mu} = \{\mu_i\}$ is a moving average of \mathbf{x} and $\boldsymbol{\sigma} = \{\sigma_i\}$ is a moving standard deviation of $(\mathbf{x} - \boldsymbol{\mu})$, both over the range of one annual layer thickness based on a preliminary estimate – see Sect. 2.2 below.

$\boldsymbol{\mu}$ measures trend in the data, the annual seasonality being averaged out due to the interval length being a whole year. Subtracting $\boldsymbol{\mu}$ de-trends \mathbf{x} and centres s on 0. $\sqrt{2}\boldsymbol{\sigma}$ 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 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 Initial estimate of cycle length

An initial estimate for the average cycle length in any sufficiently large subsection of the signal can be found using the autocorrelation function (ACF); see Fig. 1. This is used for the smoothing and not directly for the actual layer-counting. \mathbf{x} is split 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 from its ACF, $j \in (1, 2, \dots, \beta)$;
2. redistribute the boundaries so that section j contains $[nl_j / \sum_j l_j]$ points;

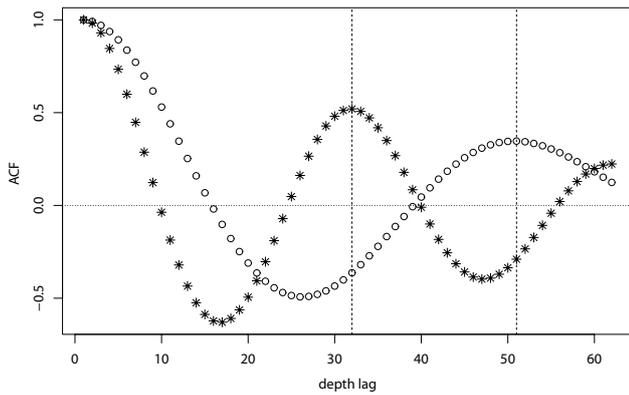


Fig. 1. The autocorrelation functions (ACFs) of the log H₂O₂ from the second (circles) and fifth (stars) of the $\beta = 6$ sections of the Gomez ice-core, plotted against lag measured as a number of observations. The vertical dashed lines indicate the local maxima of the ACFs, used to obtain an initial estimate of cycle length.

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 as the interval length for calculating μ_i and σ_i . Other methods used to obtain this estimate as a precursor to layer counting include Fourier analysis (McGwire et al., 2011) and ice flow modelling (Shimohara et al., 2003).

Figure 1 shows the ACF for the second (circles) and fifth (stars) sections of the Gomez ice-core when $\beta = 6$. The second section has an estimated 51 points in an average annual cycle and is of length 1233 points. The fifth section has an estimated 32 points in an average annual cycle, and is of length 774 points. Both sections therefore contain an estimated 24.2 annual cycles, as do all 6 sections. This gives an initial estimate of approximately 145 cycles contained in s , which is an underestimate due to the stretches of missing values.

Figure 2 illustrates the smoothing process for $\beta = 6$; each plot has vertical lines that represent the section boundaries. Figure 2a shows the log signal x , with its annual moving average μ as a dotted line. Figure 2b shows the de-trended signal, with its estimated amplitude ($\sqrt{2}\sigma$) as a dotted line. Figure 2c is of 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.

The choice of β affects the number of cycles available to estimate the average cycle length in each section; there is a trade-off between having sufficient data in each section and averaging over many cycles. In Sect. 5.2 we further examine the sensitivity of the method to the choice of β .

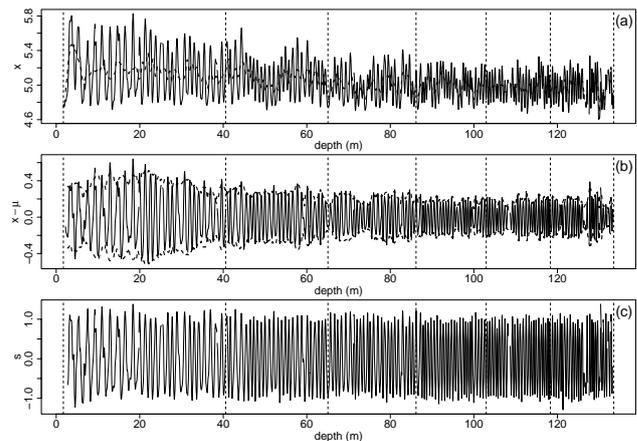


Fig. 2. The process of smoothing the H₂O₂ signal from the Gomez core, using initial estimates of annual cycle length based on $\beta = 6$ sections – boundaries shown as vertical lines. (a) The log signal x , with its annual moving average μ as a dotted line. (b) The de-trended signal, with its estimated amplitude ($\sqrt{2}\sigma$) as a dotted line. (c) The standardised signal s .

3 “Certain” annual cycles

In large-scale manual layer counting exercises, annual cycles have been termed “certain” if they are judged through consensus of multiple counters to have probability at least 3/4; see Andersen et al. (2006). Here we present a method of automatically classifying quarter cycles, analogous to seasons in the H₂O₂ signal, as being “certain” if they are well-defined in the smoothed 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 laptop for the Gomez core. It is consistent, repeatable, and does not require any prior manual assessment of chronology.

3.1 Classification into runs

We aim to segment the points of s into non-overlapping subsections called *runs*, each representing either a “certain” quarter cycle or season, 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/spring;
- T: trough/winter;
- A: ascending/autumn;
- χ : issue.

This classification is a two-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 Fig. 3a with $\nu = 1/\sqrt{2}$ has potential run label pattern:

..., $T^*, A^*, P^*, D^*, T^*, A^*, P^*, D^*, T^*, \chi, T^*, A^*,$
 $P^*, \chi, P^*, D^*, T^*, A^*, P^*, D^*, T^*, A^*, P^*, \dots$

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

..., $T, A, P, D, T, A, P, D^*, T^*, \chi^*, T^*, A^*,$
 $P^*, \chi^*, P^*, D^*, T, A, P, D, T, A, P, \dots$

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

..., $T, A, P, D, T, A, P, \chi, T, A, P, D, T, A, P, \dots$

Each point of s is now in exactly one run, and each run that is not labelled as χ makes up a certain quarter annual cycle – these are termed certain runs. In Fig. 3 runs labelled P are coloured red, D runs orange, T runs blue, A runs green, and issues black.

The value of the threshold used here, $\nu = 1/\sqrt{2} \approx 0.707$, is natural in that it would lead to the same proportions of points in each of the four types of quarter cycles, in an ideal signal. Other choices have their merits; the value used is further discussed in Sects. 3.2 and 4.2, and robustness to the choice is considered in Sect. 5.1.

3.2 Issues

Runs labelled χ , or *issues*, consist of consecutive data points that are missing values, surround missing values or are in some 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 shows a stretch of s between 9.8 m and 18.6 m, 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 13.1 m and 15.4 m: 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 5 certain runs; however there is a new issue between 16.8 m and 18.2 m.

Issues such as those seen in Fig. 3 occur only in the first 20 m of the Gomez core where, due to the very high resolution of the sampling with respect to the annual cycle length,

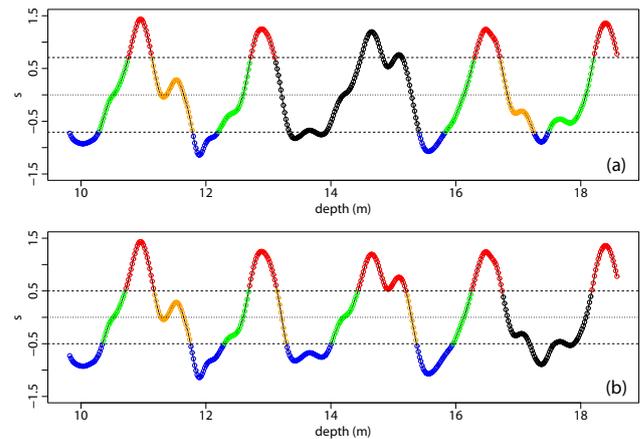


Fig. 3. Part of the standardised H_2O_2 signal, for depths between 9.8 m and 18.6 m, showing the “issues” (regions of dating uncertainty) caused by fluctuations in the signal and identified using two different values of the threshold ν for classifying individual points. Points within “issues” are black; 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 $\nu = 0.5$.

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 4 shows a stretch of s between 106.7 m and 110.2 m with $\nu = 1/\sqrt{2}$. This issue is 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. The choice of ν affects only the issue’s length, which decreases with ν .

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

We can now split the signal into sections with a deterministic count and those that need more attention. At this stage, the “issues” could be presented to experts as in Fig. 3 so they can place certain and uncertain layer markers. However, the run pattern provides more information that could be exploited either to assist the expert analysis or to allow further automation.

4 Reconstructions and probabilities

4.1 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*. Issues are always bounded by extreme runs, leading to four possible cases:

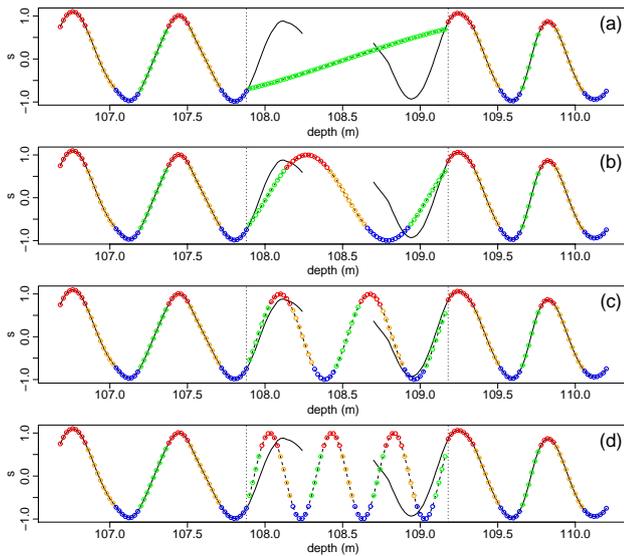


Fig. 4. Four possible reconstructions of an “issue” (region of dating uncertainty, bounded by the dotted lines) caused by missing values in the standardised H₂O₂ signal (black curve) at a depth of around 108.5 m. Coloured points within the issue represent imputed values for missing observations, or conjectured “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) orange, troughs (T) blue, and ascending points (A) 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 4. **(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 $d = 13$.

$$\begin{aligned} \dots, P, \chi, P, \dots & \quad (1) & \quad \dots, P, \chi, T, \dots & \quad (2) \\ \dots, T, \chi, P, \dots & \quad (3) & \quad \dots, T, \chi, T, \dots & \quad (4) \end{aligned}$$

The points of 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 is not sufficiently clear cut, 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 4 shows possible reconstructions of an issue caused by missing values. The top plot shows the minimal reconstruction with one run, and the bottom shows the reconstruction with 3 additional cycles added 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.

4.2 Assigning probabilities

The *length* of a run, ℓ say, is used here to denote the number of data points it contains. Provided that the time-depth relationship does not change too quickly, 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.

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 20 m of the Gomez H₂O₂ signal due to the fluctuations discussed above, and the last 20 m where there are regular stretches of missing values. In either case, this results in the analogous sections being concentrated at the centre of s , resulting in a poor fit to the sections with issues.

Instead, we need to make use of information on individual certain run lengths. In the Gomez H₂O₂ signal, peak run lengths are equivalent in distribution to trough run lengths because of the symmetry in the seasonality of 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. 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. However, 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 Fig. 5. The model is a linear regression, with independent Gaussian errors with 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 $\ln(\ell) \sim N(\ln(\hat{\ell}_d), \sigma_d^2)$, 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 can *not* be obtained directly from the model for individual runs because of the dependence in lengths between consecutive runs.)

The issue from Fig. 4 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.

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, while it is not our main purpose here, timescales can be simulated from the fitted regression model via an iterative scheme. This would require a minor adjustment: taking the run starting

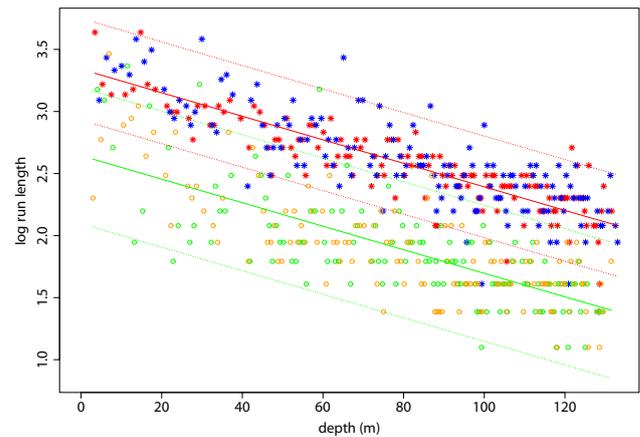


Fig. 5. A linear model for the logarithm of the length of a certain run – a set of consecutive observations that can be definitely classified as belonging to the same part of an annual cycle – against depth within the Gomez core with $\nu = 0.5$ and $\beta = 6$. Runs representing peaks (labelled P in 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.

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.

5 Results for Gomez H_2O_2

The probabilities obtained in Sect. 4.2 can be combined across the whole core, assuming separate issues to be independent given the certain parts of the classification.

Figure 6 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 153 yr exactly. The differences in these distributions are due to the fluctuations found in the first 20 m of signal as discussed in Sect. 3.2. 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.

The most probable reconstruction ($p > 0.8$ in both cases) has 153 troughs and agrees closely with the manually

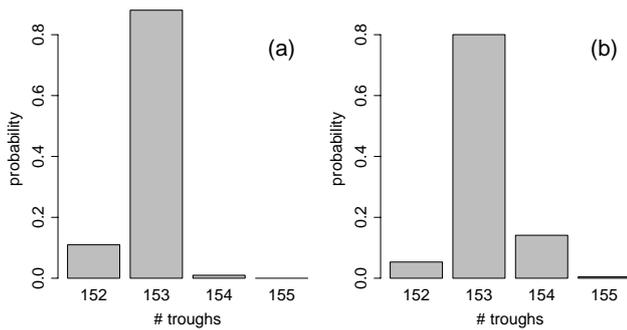


Fig. 6. Bar plots showing the probability distributions of the number of annual troughs – a summary of the chronology – in the dating of the Gomez core, based on two possible values of the threshold ν for classifying individual points, and dividing the core into $\beta = 6$ subsections for the initial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

counted solution. Each nadir placed by the model is placed in one-to-one accordance with, and in very close proximity to, the corresponding one from the manual assessment.

The signal is standardised in sections with respect to a typical cycle length, which may not be sensitive 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.

The Gomez core covers the firnification process, and an exponential decay in layer thickness is observed. There are several cycles in the first 20 m of the core that contain over 100 data points (2 m); in contrast the last 5 m of the core has several cycles with fewer than 20 data points (40 cm) – 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 thicknesses (and 99 % confidence intervals) under this model for the start (at 3 m), middle (at 65 m), and end of the core (at 132 m) are respectively 171 cm (109 cm, 268 cm), 94 cm (59 cm, 147 cm), and 49 cm (31 cm, 77 cm). 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 $\approx 0.025\%$ after allowing for the exponential decay in layer thickness.

5.1 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 7a shows the resulting distributions of cycle counts for $0.3 \leq \nu \leq 0.8$; the areas

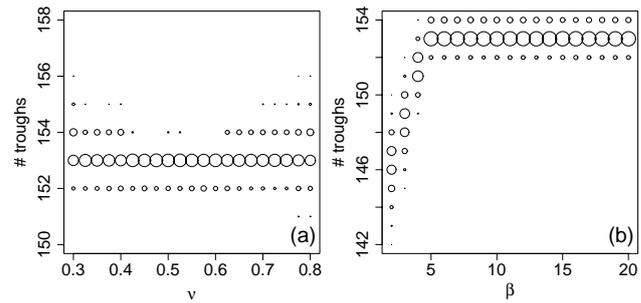


Fig. 7. The effect of the number of sections β used in the initial smoothing of the H_2O_2 signal, and the threshold ν for classifying individual points, on the probability distribution obtained for the number of annual troughs – a summary of the chronology – in the dating of the Gomez core. Each circle has area proportional to the probability of a particular number of troughs, when the reconstruction uses particular values of β and ν . **(a)** $\nu = 0.3, \dots, 0.8$ with $\beta = 6$. **(b)** $\beta = 2, \dots, 20$ with $\nu = 1/\sqrt{2}$.

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; 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.

5.2 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 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 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 .

6 NGRIP: ammonium and calcium

In this section we analyze the ammonium (NH_4) and calcium (Ca) chemistry signals from the NGRIP ice-core between 1440.49–1464.81 m, measured by the University of Bern ice-core group using continuous flow analysis at 1 mm 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 on the order of 1 cm. These data cover part of the

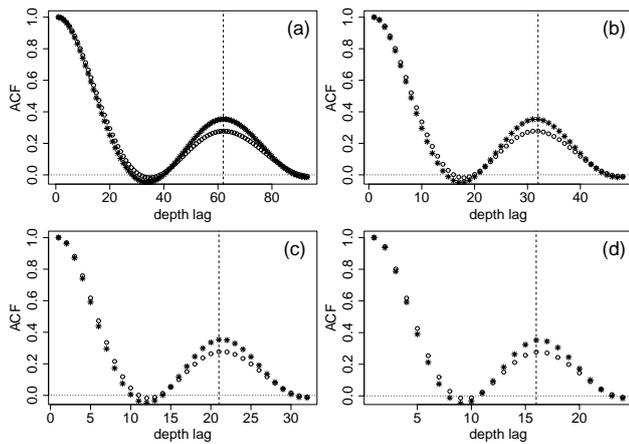


Fig. 8. The ACFs of the log NH₄ (circles) and log Ca (stars) signals from the NGRIP ice-core (1440.49–1464.81 m), plotted against lag. The vertical dashed lines indicate the local maxima, used to obtain an initial estimate of cycle length. (a) No thinning. (b) Second thinning. (c) Third thinning. (d) Fourth thinning.

Holocene climatic period (Rasmussen et al., 2006). We use the same methodology as used on the H₂O₂ signal from the Gomez core. These signals have a slightly greater noise to annual cycle ratio than the Gomez H₂O₂, 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 (2 mm); every third point (3 mm); and every fourth point (4 mm). 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 – not a moving average – and we would expect our method to work better in that case.

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 μ and σ , effectively setting $\beta = 1$. Figure 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 NH₄ is shown as circles and for Ca as stars.

Figure 9 shows a stretch of the classified log NH₄ signal (fourth thinning) for $\nu = 0.5$ with 9 cycles: (a) is the log signal with μ and $\mu \pm \sigma$ 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 3 ascending runs of length 1 – 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.

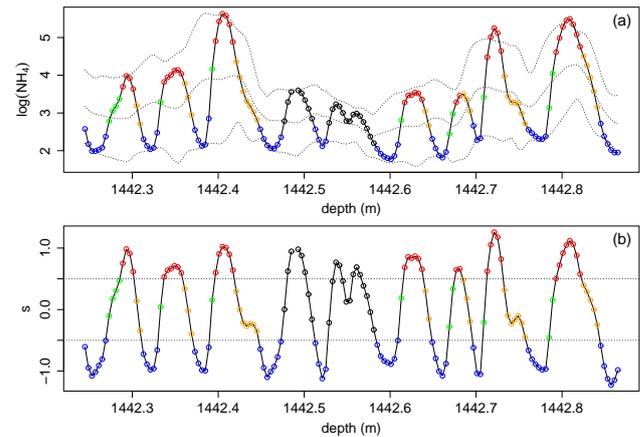


Fig. 9. A stretch of the classified log NH₄ signal (fourth thinning) from the NGRIP ice-core (1442.25–1442.87 m), for $\nu = 0.5$. Points within an issue are black, points within peaks coloured red, descending points orange, troughs blue, and ascending points green. (a) The log signal with μ and $\mu \pm \sigma$ shown as dotted lines. (b) The standardised signal s .

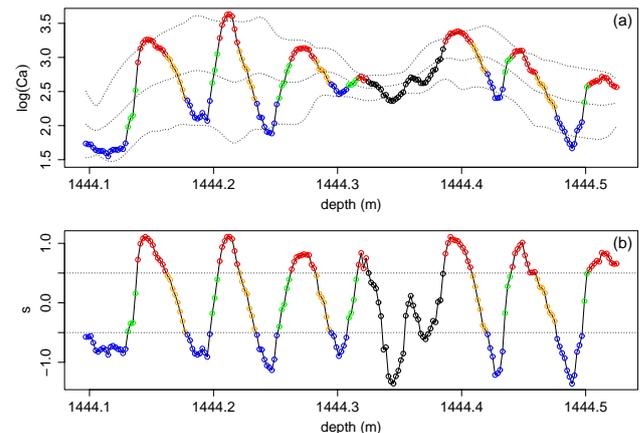


Fig. 10. A stretch of the classified log Ca signal (second thinning) from the NGRIP ice-core (1444.10–1444.53 m), for $\nu = 0.5$. Points within an issue are black, points within peaks coloured red, descending points orange, troughs blue, and ascending points green. (a) The log signal with μ and $\mu \pm \sigma$ shown as dotted lines. (b) The standardised signal s .

Figure 10 shows a stretch of the classified Ca signal (second thinning) for $\nu = 0.5$ with 7 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 Fig. 11 over a range of eight ν values. 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 check on the model’s

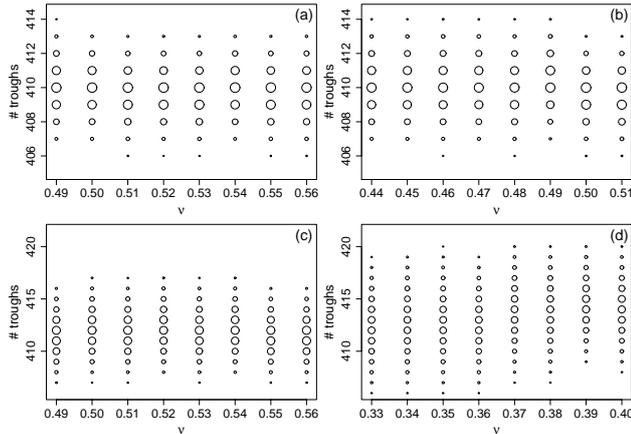


Fig. 11. The effect of ν on the probability distribution obtained for the number of annual troughs in the dating of the NGRIP ice-core (1440.49–1464.81 m) using its log NH_4 and Ca signals. Each circle has area proportional to the probability of a particular number of troughs. **(a)** NH_4 , second thinning. **(b)** NH_4 , fourth thinning. **(c)** Ca, second thinning. **(d)** Ca, fourth thinning.

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 Sect. 8, would be beneficial. However, in multivariate datasets the different nominal resolution of individual components must also be considered.

The GICC05 chronology (Rasmussen et al., 2006) assigns 420 certain years and 3 uncertain years between 1439.92 m and 1465.52 m for the NGRIP core, which 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 %.

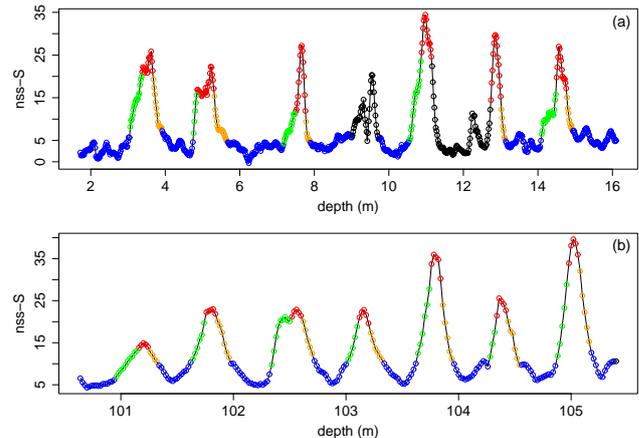


Fig. 12. Two portions of the classified log nss-S signal from the Gomez ice-core for $\nu_1 = 0.85$, $\nu_2 = 0.5$, and $\beta = 10$. Points within an issue are black, points within peaks coloured red, descending points orange, troughs blue, and ascending points green. **(a)** At the start of the core. **(b)** Towards the end of the core.

7 Gomez: non-sea-salt sulphur

To show how our classification process can be adapted for asymmetry, we re-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 $\text{H}_2\text{O}_2/\text{NH}_4/\text{Ca}$ 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 effectively estimates a local mean and standard deviation for each depth; points that exceed a given number ($\nu\sqrt{2}$) of standard deviations from the mean are then classified as *potential* peaks and troughs. As symmetry cannot be induced throughout the nss-S signal, we require a more robust method of classification – we use local quantiles or percentiles. Along interval lengths estimated from the ACF of the nss-S, as above, we calculate the local ν_1 -th and ν_2 -th quantiles at each depth. Data points above the ν_1 -th quantile are then classified as potential peaks, and data points below the ν_2 -th quantile are classified as potential troughs. From this point on the method continues as in previous sections.

In the regression model for the 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.

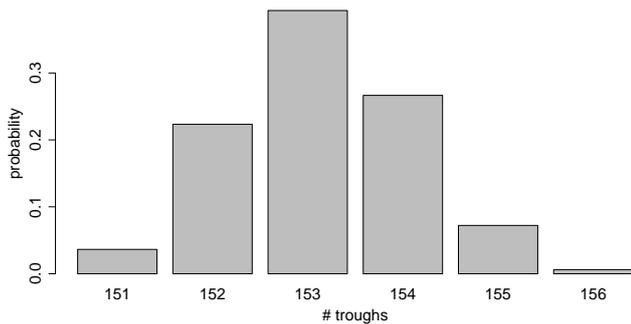


Fig. 13. Bar plot showing the probability distribution of the number of annual troughs – a summary of the chronology – in the dating of the Gomez core using its nss-S signal for $\nu_1 = 0.85$, $\nu_2 = 0.5$, and $\beta = 10$.

Figure 12 shows sections of classified nss-S with 7 cycles, (a) at the start of the core, and (b) towards the end of the core.

Figure 13 shows the resulting probability distribution for the chronology at $\nu_1 = 0.85$ and $\nu_2 = 0.5$, which has more uncertainty than that found from the H_2O_2 . 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 $0.8 \leq \nu_1 \leq 0.9$ and $0.4 \leq \nu_2 \leq (\nu_1 - 0.3)$.

8 Conclusions

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

- split the signal into sections with a deterministic cycle count and those that need more attention, using a tuning parameter, ν ,
- 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 signals from the Gomez 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.

The log H_2O_2 signal from the Gomez 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.

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.

If the classification method can be extended to a multivariate framework – using multiple signals to group depths into “certain” runs described by length and label – then the same method of assigning probabilities to issues can be utilised. In the univariate example we split the signal into three groups along the real line; it is possible to split two out-of-phase standardised signals into four groups in the bivariate plane. Plotting x' against y' the points can be collected into quadrants: Q_1 where $x' > 0$ and $y' > 0$, Q_2 where $x' < 0$ and $y' > 0$, Q_3 where $x' < 0$ and $y' < 0$, and Q_4 where $x' > 0$ and $y' < 0$. These could be treated as potential runs as above, and the method would be the same from that point.

We attempt to show that this process is robust to the tuning parameters. Plots are provided to show the effect of changing 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 four analyses, different issues arise when varying ν in the ranges presented; see Fig. 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 Fig. 3), the reconstruction that 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.

We are currently developing a method that uses a model-based statistical approach to make inferences about a chronology in much more detail. It is more computationally demanding than the method described here and uses a more flexible model formulation than that described in Wheatley et al. (2012). One possible strategy would be to use the method presented here to define and investigate issues; those issues that cannot be readily resolved (i.e. where no single reconstruction is obviously correct, based on run-lengths) could then be analysed in more detail, using the model-based approach to refine the chronology and the probabilities.

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. The PEAK tools of Weber et al. (2010) achieve a

similar goal and work with great effect on noisy data. The “zero-crossing method” algorithm iteratively finds the points at which a signal crosses a wide interval Gaussian moving average, and 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. In the nomenclature of the present paper, these methods essentially segment the signal into “certain” runs representing peaks and troughs. 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 the 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. Their most recent paper does not address missing values or provide a measure of uncertainty.

Acknowledgements. This work was supported by the Natural Environment Research Council (grant reference: NE/I52797X/1). We thank the Division for Climate and Environmental Physics, Physics Institute, University of Bern, for providing high-resolution CFA data from the NGRIP ice-core to test our method. We are grateful for valuable comments on an earlier version from H. Fischer, M. Mudelsee and S. Rasmussen.

Edited by: H. Fischer

References

- Andersen, K. K., Svensson, A., Johnsen, S. J., Rasmussen, S. O., Bigler, M., Röthlisberger, R., Ruth, U., Siggaard-Andersen, M. L., Peder Steffensen, J., Dahl-Jensen, D., Vinther, B. M., and Clausen, H. B.: The Greenland ice core chronology 2005, 15–42 ka. Part 1: constructing the time scale, *Quaternary Sci. Rev.*, 25, 3246–3257, 2006.
- Dahl-Jensen, D., Gundestrup, N. S., Miller, H., Watanabe, O., Johnsen, S. J., Steffensen, J. P., Clausen, H. B., Svensson, A., and Larsen, L. B.: The NorthGRIP deep drilling programme, *Ann. Glaciol.*, 35, 1–4, 2002.
- McGwire, K. C., Taylor, K. C., Banta, J. R., and McConnell, J. R.: Identifying annual peaks in dielectric profiles with a selection curve, *J. Glaciol.*, 57, 763–769, 2011.
- Mudelsee, M., Fohlmeister, J., and Scholz, D.: Effects of dating errors on nonparametric trend analyses of speleothem time series, *Clim. Past Discuss.*, 8, 1973–2005, doi:10.5194/cpd-8-1973-2012, 2012.
- R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>, last access: 12 October 2012, ISBN3-900051-07-0, 2011.
- Rasmussen, S. O., Andersen, K. K., Siggaard-Andersen, M. L., and Clausen, H. B.: Extracting the annual signal from Greenland ice-core chemistry and isotopic records, *Ann. Glaciol.*, 35, 131–135, 2002.
- Rasmussen, S. O., Andersen, K. K., Svensson, A. M., Steffensen, J. P., Vinther, B. M., Clausen, H. B., Siggaard-Andersen, M. L., Johnsen, S. J., Larsen, L. B., Dahl-Jensen, D., Bigler, M., Röthlisberger, R., Fischer, H., Goto-Azuma, K., Hansson, M. E., and Ruth, U.: A new Greenland ice core chronology for the last glacial termination, *J. Geophys. Res.*, 111, D06102, doi:10.1029/2005JD006079, 2006.
- Rupf, I. and Radons, G.: New approaches for automated data processing of annually laminated sediments, *Nonlinear Proc. Geoph.*, 11, 599–607, 2004.
- Shimohara, K., Miyamoto, A., Hyakutake, K., Shoji, H., Takata, M., and Kipfstuhl, S.: 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, 2003.
- Svensson, A., Nielsen, S. W., Kipfstuhl, S., Johnsen, S. J., Steffensen, J. P., Bigler, M., Ruth, U., and Röthlisberger, R.: Visual stratigraphy of the North Greenland Ice Core Project (North-GRIP) ice core during the last glacial period, *J. Geophys. Res.*, 110, D02108, doi:10.1029/2004JD005134, 2005.
- Taylor, K. C., Alley, R. B., Meese, D. A., Spencer, M. K., Brook, E. J., Dunbar, N. W., Finkel, R. C., Gow, A. J., Kurbatov, A. V., Lamorey, G. W., Mayewski, P. A., Meyerson, E. A., Nishiizumi, K., and Zielinski, G. A.: Dating the Siple Dome (Antarctica) ice core by manual and computer interpretation of annual layering, *J. Glaciol.*, 50, 453–461, 2004.
- Thomas, E. R., Marshall, G. J., and McConnell, J. R.: A doubling in snow accumulation in the western Antarctic Peninsula since 1850, *Geophys. Res. Lett.*, 35, L01706, doi:10.1029/2007GL032529, 2008.
- Weber, M. E., Reichelt, L., Kuhn, G., Pfeiffer, M., Korff, B., Thurow, J., and Ricken, W.: BMPix and PEAK tools: New methods for automated laminae recognition and counting – Application to glacial varves from Antarctic marine sediment, *Geochem. Geophys. Geosyst.*, 11, Q0AA05, doi:10.1029/2009GC002611, 2010.
- Wheatley, J. J., Blackwell, P. G., Buck, C. E., Lee, L., Röthlisberger, R., and Wolff, E. W.: Statistical layer-counting in ice-cores, *Environ. and Ecol. Stat.*, submitted, 2012.
- Winstrup, M., Svensson, A. M., Rasmussen, S. O., Winther, O., Steig, E. J., and Axelrod, A. E.: An automated approach for annual layer counting in ice cores, *Clim. Past Discuss.*, 8, 2519–2555, doi:10.5194/cpd-8-2519-2012, 2012.