Clim. Past, 8, 1881–1895, 2012
www.clim-past.net/8/1881/2012/
doi:10.5194/cp-8-1881-2012
© Author(s) 2012. CC Attribution 3.0 License.

Climate of the Past

An automated approach for annual layer counting in ice cores

M. Winstrup[1,*], A. M. Svensson[1], S. O. Rasmussen[1], O. Winther[2], E. J. Steig[3], and A. E. Axelrod[4]

[1]Centre for Ice and Climate, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
[2]DTU Informatics, Technical University of Denmark, Lyngby, Denmark
[3]Quaternary Research Center & Department of Earth and Space Sciences, University of Washington, Seattle, Washington, USA
[4]Department of Electrical Engineering, University of Washington, Seattle, Washington, USA
[*]now at: Department of Earth and Space Sciences, University of Washington, Seattle, Washington, USA

Correspondence to: M. Winstrup (mai@gfy.ku.dk)

Received: 7 June 2012 – Published in Clim. Past Discuss.: 9 July 2012
Revised: 26 October 2012 – Accepted: 30 October 2012 – Published: 23 November 2012

Abstract. A novel method for automated annual layer counting in seasonally-resolved paleoclimate records has been developed. It relies on algorithms from the statistical framework of hidden Markov models (HMMs), which originally was developed for use in machine speech recognition. The strength of the layer detection algorithm lies in the way it is able to imitate the manual procedures for annual layer counting, while being based on statistical criteria for annual layer identification. The most likely positions of multiple layer boundaries in a section of ice core data are determined simultaneously, and a probabilistic uncertainty estimate of the resulting layer count is provided, ensuring an objective treatment of ambiguous layers in the data. Furthermore, multiple data series can be incorporated and used simultaneously. In this study, the automated layer counting algorithm has been applied to two ice core records from Greenland: one displaying a distinct annual signal and one which is more challenging. The algorithm shows high skill in reproducing the results from manual layer counts, and the resulting timescale compares well to absolute-dated volcanic marker horizons where these exist.

1 Introduction

An accurate chronology is of fundamental importance for the interpretation of a paleoclimatic record. For some ice cores, the temporal resolution of the records is sufficiently high to preserve seasonal information. Indeed, with the increased resolution presently being achieved in many ice core measurements (e.g. Bigler et al., 2011), an annual signal is preserved in an increasing number of different data records and spanning much longer time intervals. Absence of snowfall over an extended time, layers lost by wind reworking, diffusion processes in firn and ice, limited measurement resolution, etc. may act to decrease the number of layers that can be recognised in the ice core data, whereas small or larger-scale folds in the ice sheet may cause some layers to occur twice. Nevertheless, when the conditions are fulfilled for annual layers in the ice core to have survived the archiving and measurement processes in stratigraphic order, annual layer counting represents the most accurate method to produce a chronology for the core.

Yet, manual layer counting is a tedious and sometimes subjective method. The Greenland Ice Core Chronology 2005 (GICC05) was developed over several years and involved the efforts of many researchers: counting, comparing, and re-counting the annual layers recorded in multiple data records from several Greenland ice cores (Andersen et al., 2006b; Rasmussen et al., 2006; Svensson et al., 2006). The chronology extends back to 60 kyr b2k (kyr before 2000 CE) (Svensson et al., 2008). Decreasing annual layer thicknesses causes the annual signal in most ice core parameters to weaken and eventually disappear with depth. As reliable layer recognition becomes more challenging, a manual approach increasingly relies on subjective interpretation of the available data. Thus, time-wise and potentially also quality-wise, much may be gained by an automated and therefore reproducible approach for annual layer identification in ice core records.

Many attempts have been made to develop automated methods of annual layer counting in paleoclimatic archives (McGwire et al., 2008; Rasmussen et al., 2002; Smith et al., 2009). These have generally had limited success where the annual signal in the data is inherently ambiguous, leaving manual layer counting to still be considered the most accurate. In comparison to these previous attempts, the method developed here is quite similar to the manual approach of layer counting; multiple layer boundaries in an entire data section are determined simultaneously, while allowing for a very flexible definition of an annual layer signature in the record. Furthermore, the method allows for multi-parameter counting, using the annual information in multiple data records at once (although this is yet to be implemented). As the approach is based on statistics, it also allows for objectively inferring an uncertainty estimate of the resulting timescale. Such uncertainties are generally very hard to assess manually. It has hitherto proven difficult to incorporate these properties in automated layer-recognition methods.

The annual layer detection algorithm has here been applied to two ice core records: a stable-isotope record from DYE-3, Southern Greenland, and visual stratigraphy data from NGRIP, Central Greenland, as obtained from line-scan images (Svensson et al., 2005). The seasonality is very distinct in the selected section of the δ^{18}O record from DYE-3, and annual layers caused by the seasonal variation in temperature can be manually identified with high certainty. Moreover, the ice core contains several absolute-dated age markers in form of volcanic horizons, which provide an independent data set against which the resulting timescale can be evaluated. The NGRIP ice core, on the other hand, has particularly high temporal resolution with depth, and therefore has the potential to be dated by annual layer counting far back in time. The oldest part of the GICC05 chronology was based exclusively on data from this core. For the depth interval in consideration, the line-scan data displays a clear banding of dark and bright layers, which essentially provides a proxy for dust content in the ice core. The data is at such high resolution that it may resolve individual precipitation events. From modern-day ice core data it is known that the dust input to Greenland has a significant seasonal variation (Alley et al., 1997; Hamilton and Langway, 1968). Assuming this relationship to hold back in time, the grey-tone intensity of the line-scan images should also display a seasonal pattern. More often than not, however, the annual signal in the data is ambiguous, and when considered in isolation these are not easy to use for layer counting. Nonetheless, due to the ability of the line-scan data to resolve annual layering in sections of thin annual layers, it has potential for being utilised to extend the GICC05 chronology further back in time.

2 A hidden Markov modelling approach to annual layer recognition

The layer detection routine is based on the statistical framework of hidden Markov models. A hidden Markov model (HMM) is a stochastic signal model that can be used for modelling the output of a system displaying Markovian behaviour, i.e. a stochastic system which transitions between states, and where the next state of the system depends only on the current state. By comparison to observed data, knowledge of the nature of the underlying signal can then be obtained (Rabiner, 1989). The concept of hidden Markov models was originally introduced in the late 1960s (Baum and Petrie, 1966), and has successfully been applied to pattern recognition in the field of machine speech recognition since the mid-70s (Jelinek et al., 1975). The rich mathematical structure of hidden Markov models can form the theoretical basis for a wide range of signal modelling applications, spanning from magnetic resonance imaging (MRI) brain mapping (Faisan et al., 2005) to the analysis of financial time series (Bulla and Bulla, 2006). Here it will be applied to annual layer recognition in ice core records.

An annual layer recognition algorithm based on hidden Markov modelling is inherently of Bayesian nature. The annual layers are obtained based on prior knowledge on for example the general layer thickness probability distribution, which is updated based on observed data, and the resulting layer boundaries are given as probability distributions, both in depth and layer number. When calculating these probabilities, an entire observation sequence is taken into account at once. The likelihood that a given observation segment represents an annual layer is therefore judged not only by its own resemblance to a typical annual layer, but is judged in conjunction with the likelihood of the resulting layering on either side. Determination of layer boundaries in data sections with poorly-resolved annual layering is therefore primarily based on the positioning of surrounding clearly discernible annual layer boundaries. Sections containing missing data are treated much the same way: an appropriate number of layers is fitted into these regions, with the "appropriate number" based on knowledge of the layer thickness probability distribution in combination with the most likely positions of layer boundaries in the surrounding data. Simultaneously, the spread of the probability distribution of counted annual layers is generally enlarged, corresponding to increased uncertainty in the total number of counted layers. This is very similar to what is implicitly done when counting layers manually.

The method can relatively easily be extended to allow the incorporation of multiple data series containing an annual layer signal and infer the most likely annual layering based on all of these data series simultaneously. It therefore provides the necessary statistical framework to allow for an automated multiple-parameter counting algorithm. This property of the algorithm is important because misinterpretation of the annual layering in any single ice core record is likely

to occur due to variability of the annual layer signal and occurrence of non-annual features. For the current purpose, information has been incorporated based on the derivative of the data series as well as the observed data itself.

2.1 Hidden Markov models

In a hidden Markov model (HMM), the system under consideration is allowed to be in a finite number of states. The state of the system corresponding to any index t (e.g. time) is considered a stochastic variable S_t with the allowed state outcomes ℓ_j, $j \in \{1, 2, ..., J\}$. For the purpose of annual layer detection, ℓ_j will represent a label corresponding to the annual layer number. The use of t for indexing (t spanning from 1 to T) is owed to hidden Markov modelling usually being applied on time series; for the current purpose it will be used as an index for observation number, i.e. depth. The state of the system is assumed to change stochastically in such a way that the state sequence, $S_{1:T}$, is a first-order Markov chain, i.e. the next state of the system is only dependent on the current state.

If a direct outcome of the state sequence can be observed, it is easy to characterise the statistical nature of this signal. Very often, however, this is not possible. Instead, it may be possible to observe the influence of the state sequence on another stochastic process, the outcome of which is seen as a sequence of observations, $\boldsymbol{o}_{1:T}$. The model hereby has a two-level structure, with the state sequence providing the unknown "truth", and each observation by itself only providing incomplete information on the current state. Algorithms developed for hidden Markov models allow for the statistical inference of the underlying hidden Markovian state sequence (Rabiner, 1989).

Similar algorithms have also been developed in case of so-called hidden semi-Markov state sequences, in which the changes in state are endowed with a Markov property, but with holding times of each state distributed according to a prescribed probability distribution $p(d)$, with the duration d given in terms of number of observations covered by each state. Such a process can be envisioned as a doubly embedded Markov chain, with generalised states consisting of both the state label and state duration. Allowing the underlying stochastic process to be semi-Markov, this variant is sometimes known in the literature as a "hidden semi-Markov model" (HSMM) (Yu, 2010), or – depending on the specific assumptions of the model, its application area and the author – a "segment model" (Ostendorf et al., 1996).

The sequence of encountered years in a seasonally resolved data record, when assuming the associated annual layers to occur in sequential order without any skipping, can be viewed as a degenerated Markov chain. Hence, the observations corresponding to a succession of annual layers in an ice core record can be regarded as the output of a hidden semi-Markov process, in that each individual data point can be labelled with a corresponding state (i.e. the layer number)

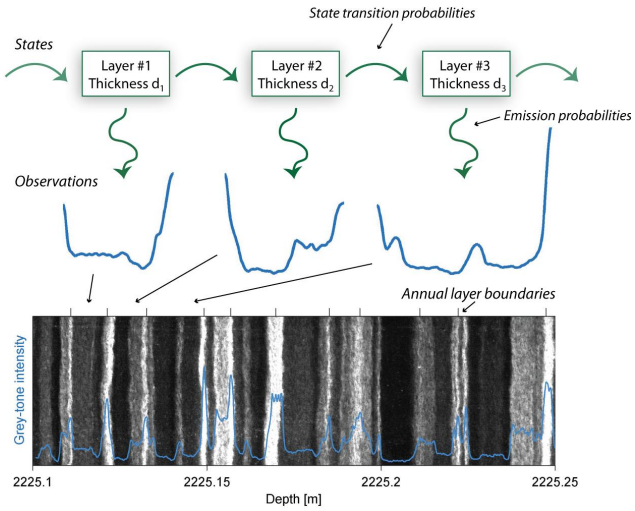

Fig. 1. Illustration of the two-level structure of the hidden semi-Markov model used for annual layer detection. The states of the system are "layer numbers", which have been endowed with a layer thickness parameter d. Changes in state are described by state transition probabilities. Each state emits a segment of observations corresponding to one layer. In this case, the observations are grey-tone intensity values from line-scan images from the NGRIP ice core (bottom). Based on the entire sequence of observations, the HMM algorithm can be used to infer the most likely state sequence (i.e. the depth-age relationship) along with its corresponding uncertainty.

as well as a duration (i.e. the layer thickness), see Fig. 1. The desired depth-age relation is then the most likely state sequence given the observations.

The application of HMMs for annual layer counting in annually laminated paleoclimate archives differs from their application in most other areas. The changes in state are simple, with one year simply following the previous. However, this simplicity is often combined with large inter-annual variability in the annual layer expressions, thereby giving rise to a rather challenging pattern-recognition problem. The general HMM algorithms must therefore be adapted to these conditions.

2.2 Characterising an annual layer

Mathematically, the layer detection algorithm works by reducing the complex issue of simultaneous pattern-matching of multiple successive layers to a given template, to the much simpler question concerning how likely a particular data segment is to represent a single annual layer. The likelihood of a given observation segment forming an annual layer is partly determined by the probability of the resulting annual layer thickness, and partly by the resemblance of the segment to an annual layer signal. These probabilities provide the criteria used by the layer detection algorithm for subsequently retrieving the most likely annual layer sequence based on the entire data record. The relative weighting of these two criteria

can be adjusted as desired. In this way the layer detection algorithm employs the same principles as used in manual layer identification.

Empirical data shows that for a given depth interval, the annual layer thicknesses in an ice core are approximately log-normal distributed (Andersen et al., 2006a). The assumed probability distribution of the layer durations is therefore taken to be a log-normal distribution described by the two parameters μ_{d} and σ_{d}:

$$p(d) = \ln N\left(\mu_{\mathrm{d}},\ \sigma_{\mathrm{d}}^2\right).$$

The parameters will be termed the location (μ_{d}) and scale parameter (σ_{d}) of the distribution. The mean annual layer thickness, λ, is a function of both of these parameters. The continuous probability density function is discretised and normalised to provide duration probabilities corresponding to a finite integer number of data points.

Resemblance of a given observation segment $\boldsymbol{o}_{t_1:t_2}$ to a typical annual layer signal can be assessed in a variety of ways. The likelihood that the segment forms an annual layer will here be determined by regression of the segment to a layer template. This so-called emission probability is formally defined as the conditional probability of observing a given sequence of observations when assuming these to form an annual layer. Following Yu (2010), we will use the shorthand notation $S_{[t_1:t_2]} = \ell_j$ to signify that layer j (i.e. state outcome ℓ_j) starts exactly at t_1 and ends exactly at t_2 (both data points included). Using this notation, the emission probability can be written as follows:

$$b\left(\boldsymbol{o}_{t_1:t_2}\right) \equiv P\left(\boldsymbol{o}_{t_1:t_2} \middle| S_{[t_1:t_2]} = \ell_j,\ \theta\right). \tag{1}$$

The inferred layer boundaries depend critically on the applied annual layer template and the model parameters herein, θ, used for judging how well the segment of observations $\boldsymbol{o}_{t_1:t_2}$ exhibits the characteristics of an annual layer.

Annual layers in especially the NGRIP line-scan data display significant variability in shape from one year to the next, and the annual layer template must be able to capture this diversity. To ensure such flexibility of the layer detection algorithm, the annual layers are modelled based on a generic layer template, which consists of an appropriate selection of analytical basis functions ($\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots$) that each describes an aspect of the shape of the layer signal. The basis functions are normalised with respect to the annual layer thickness such that their resulting shape is independent on this. Layer trajectories are then formed by linear combinations of these basis functions, and assumed to be observed with an additive Gaussian white noise component. By forming the matrix $\mathbf{X} = [\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \boldsymbol{x}_3,\ \ldots]$, the signal for layer j can be described as the noisy output of a linear system:

$$\boldsymbol{y}_j = \mathbf{X}\boldsymbol{a}_j + \boldsymbol{\varepsilon}_j.$$

For a specific annual layer, the weightings corresponding to the individual basis functions are contained in the vector $\boldsymbol{a}_j$. These weightings are assumed to be multivariate Gaussian distributed. Mean value ($\boldsymbol{\varphi}$) and covariance (Φ) of this distribution are regarded as model parameters and are included in θ, along with the variance (σ_ε^2) of the additive Gaussian white noise component, $\boldsymbol{\varepsilon}_j \sim N(\mathbf{0},\ \sigma_\varepsilon^2)$. In addition, θ contains the parameters describing the layer thickness probability distribution (μ_{d} and σ_{d}).

By using the above formulation, the description of an annual layer is divided into average layer characteristics ($\boldsymbol{\varphi}$), and random effects pertinent to each individual layer, $\boldsymbol{r}_j \sim N(\mathbf{0},\ \Phi)$, that cause it to differ from the average. An annual layer can therefore also be described as

$$\boldsymbol{y}_j = \mathbf{X}\boldsymbol{a}_j + \boldsymbol{\varepsilon}_j = \mathbf{X}\left(\boldsymbol{\varphi} + \boldsymbol{r}_j\right) + \boldsymbol{\varepsilon}_j = \mathbf{X}\boldsymbol{\varphi} + \mathbf{X}\boldsymbol{r}_j + \boldsymbol{\varepsilon}_j. \tag{2}$$

This model for the annual layers represents a two-level hierarchical model, and the emission probabilities can be evaluated by Bayesian linear regression (Bishop, 2006). See Winstrup (2011) for a more detailed description of how these computations are carried out. The employed model is an extended version of the linear trajectory models used by Gish and Ng (1996) and Russell and Holmes (1997), and bears many similarities to the one used for electrocardiographic waveform detection by Kim et al. (2004) and Kim and Smyth (2006).

The explicit modelling of the allowed range of inter-annual variability in layer shape permits the developed layer detection algorithm to address this general problem faced by any annual layer counting routine. The range of year-to-year variability in layer shapes becomes an integrated part of the layer detection algorithm, and enables it to better handle observation sequences in which the annual layer signal is noisy and irregular.

A principal components analysis (PCA) was initially conducted on the data series to constrain the allowed variability of individual layer shapes. This analysis was based on results from manual layer picking. For the NGRIP line-scan data, the first two principal components describing the layering in the cold period (Fig. 2c) and the warm period (not shown) explain respectively 73 % and 57 % of the variance in layer shapes from the mean. With three principal components the numbers increase to 84 % (cold period) and 68 % (warm period). The difference in the amount of variance explained indicates a less distinct annual signal in the line-scan data during the warm period, which can also be seen by visual comparison of the two data series. In the following, we have employed analytical functions that mimic the variability described by the principal components as basis functions, and regression to these is performed on the deviation of an observation segment from the calculated mean annual layer signal.

Observe that $\boldsymbol{o}_{1:T}$ is allowed to be a sequence of vector observations, and hence each data point may, for example, contain a collection of measured chemistry data. As the assessment of the probability $b(\boldsymbol{o}_{t_1:t_2})$ can be made based on

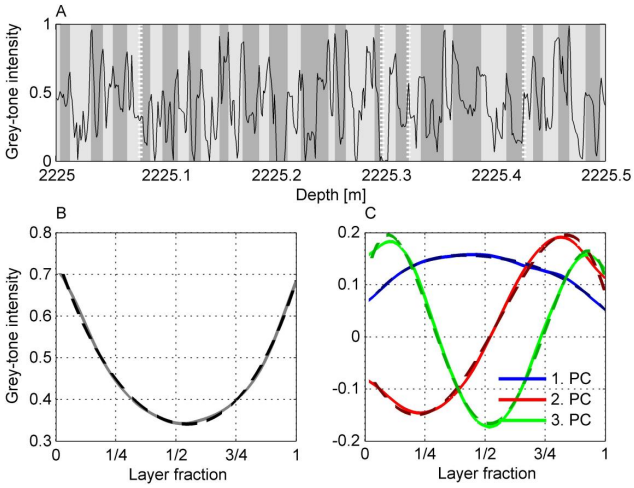

Fig. 2. (A) Section of line-scan data with manually counted annual layers indicated by alternating bright and dark grey bands. White dashed layer boundaries are regarded as uncertain. Prior to analysis, data were treated by taking the logarithm of the grey-tone intensity values and normalising according to minimum/maximum values. **(B)** Mean annual layer trajectory and **(C)** the first three principal components of the deviation from the mean trajectory during the cold period GS-13 (depth: 2225–2240 m). 4th order polynomial fits to the mean and principal components are shown as dashed lines.

multiple data sets, it allows for the annual layer signal in all of these to be taken into account simultaneously.

We have here considered both the observed data series as well as its derivative as two separate, yet related, observation sequences. Although the two data sequences essentially contain the same information, the noise component on the derivative data series is generally "whiter", but has the caveat that the signal-to-noise ratio is smaller. Consequently, by combining information from both data series, the underlying assumption is relaxed that all variance of an annual signal not explained by the layer template can be considered white noise. Indeed, the noise corresponding to successive data points is correlated, and when applied to the line-scan data in particular, much better estimates of the annual layering were obtained when utilising also the derivative of the observed sequence of grey-tone intensities.

2.3 Inference of a most likely layer sequence and its uncertainty

Having calculated the emission probabilities (Eq. 1) for all possible layer start and ending positions, the most likely layer at any given location can be inferred. Consider the option of using brute force to examine all possible state sequences, one at a time, calculating the respective probabilities, and adding up the contributions from those that give rise to the same state ℓ_j at each t. In this way, the most likely layer at t can be found, and the result will be based on the entire observation sequence in consideration. In reality such an approach is rarely feasible, but the same probabilities can be efficiently calculated by recursion by the forward-backward algorithm, which routinely is applied to HMMs, and here is used in the extended version applicable for HSMMs (Yu, 2010).

The name "forward-backward algorithm" is derived from the way the algorithm makes the judgment of a most likely state sequence in a rigorous yet efficient manner by executing a double pass of the observation sequence: a forward pass, which passes on the information included in all previous data, and a backward pass, containing information from all subsequent data. The best estimate of the hidden state sequence is then found by combining the two sources of information. In this way, the entire data sequence is used for inferring the most likely layering, and also the involved uncertainties can be evaluated. These provide an estimate of the counting error relatively similar to that of the manually layer counted GICC05 chronology, in which uncertain layers were counted as $1/2 \pm 1/2$ yr (Andersen et al., 2006b; Rasmussen et al., 2006).

In the forward message pass, the forward variable $\alpha_t(j,\ d)$ is calculated. It is defined as the joint probability of ending state ℓ_j with duration d at t and observing the partial observation sequence $\boldsymbol{o}_{1:t}$. It can be recursively calculated by

$$\alpha_t(j,\ d) \equiv P\left(S_{[t-d+1:t]} = \ell_j,\ \boldsymbol{o}_{1:t}\right) = p(d)\, b\left(\boldsymbol{o}_{t-d+1:t}\right) \sum_{d'} \alpha_{t-d}\left(j-1,\ d'\right). \quad (3)$$

Similarly for the backward message pass, a backward variable β_t is defined as the probability of observing the partial observation sequence $\boldsymbol{o}_{t+1:T}$ when conditioned on state ℓ_j ending at t. Using the shorthand notation $S_{t]} = \ell_j$ to indicate that layer j ends at t, without providing any information on where the layer begins, the backward variable can be calculated as

$$\beta_t \equiv \sum_j P\left(\boldsymbol{o}_{t+1:T} \,|\, S_{t]} = \ell_j\right) = \sum_d p(d)\, b\left(\boldsymbol{o}_{t+1:t+d}\right) \beta_{t+d}. \quad (4)$$

As initialisation condition for the two recursive passes of the data series, the first observation is assumed known to be in state ℓ_1, whereas the state corresponding to the last observation is considered unknown, and thus all possible states have equal probability. Due to the flexibility of the forward-backward algorithm to allow for inter-annual variations in layer thickness, the algorithm is not very strongly dependent on the specifics of the applied initialisation conditions.

These equations can be derived from the general equations of the forward-backward algorithm for HSMMs, as given for example in Yu (2010), when using the following simplifying assumptions: encountered layers are assumed to occur in sequential order without any skipping, the thickness of a layer is assumed independent on thickness and label of the previous layer, and all layers are assumed to be an outcome of the same process, implying that the layer signal (shape as well as thickness) is independent on the layer number. Previous studies (Andersen et al., 2006a; Fisher et al., 1985) point to the

existence of a slightly negative correlation between the thickness of successive layers, and such correlation is not taken into account in the above simplified formulation. Also the algorithm does not impose any restrictions regarding continuity of the fitted layer trajectories across layer boundaries: for each layer, the best layer trajectory is determined independently from those of its neighbours. In practise, however, none of these shortcomings represent a major limitation of the method. A formal derivation of the appropriate formulation of the forward and backward variables, and a more explicit description of the initialisation conditions used, can be found in Winstrup (2011).

By multiplying the forward and backward variables (Eqs. 3 and 4), one obtains the posterior probability of ending layer j with duration d at t, based on the entire observation sequence (Yu, 2010):

$$\eta_t(j,\ d) \equiv P\left(S_{[t-d+1:t]} = \ell_j,\ \boldsymbol{o}_{1:T}\right) = \alpha_t(j,\ d)\,\beta_t. \quad (5)$$

Summing over all values of d gives the probability of ending layer j at t regardless of its duration:

$$\tilde{\eta}_t(j) \equiv P\left(S_{t]} = \ell_j,\ \boldsymbol{o}_{1:T}\right) = \sum_d \eta_t(j,\ d).$$

The probability of being in layer j at t can then be calculated recursively as the probability of being in layer j at $t-1$, subtracting the probability of ending layer j at $t-1$, and adding the probability of starting the layer at t:

$$\gamma_t(j) \equiv P\left(S_t = \ell_j,\ \boldsymbol{o}_{1:T}\right) = \gamma_{t-1}(j) - \tilde{\eta}_{t-1}(j) + \tilde{\eta}_{t-1}(j-1). \quad (6)$$

This recursive formulation quantifies the probability of being in layer j at point t in the observation sequence. In the forward-backward algorithm, the most likely state sequence is understood as the state sequence in which each state individually has maximum posterior probability when conditioned on the entire observation sequence. The maximum a posteriori (MAP) estimate of the state corresponding to the observation at t is therefore:

$$\ell_{\mathrm{MAP}}(t) \equiv \underset{\ell_j}{\arg\max}\left\{P\left(S_t = \ell_j \middle| \boldsymbol{o}_{1:T}\right)\right\} = \underset{\ell_j}{\arg\max}\left\{\gamma_t(j)\right\}. \quad (7)$$

This is the most likely layer at t. The corresponding layer boundaries can be found as locations where $\ell_{\mathrm{MAP}}(t)$ shifts. Furthermore, as the full probability distributions are computed, the uncertainty associated with inference of the most likely layer at a given position can be estimated by considering the width of the derived distributions. A chronology consisting of both a best estimate ($\ell_{\mathrm{MAP}}(t)$, the mode of the distribution) and an uncertainty estimate (here taken as distribution percentiles) can therefore be derived.

However, it must be stressed that the validity of these uncertainty estimates is contingent on the annual layer template and corresponding model parameters to be valid. The derived uncertainty estimate does not take into account possible inaccuracies in these, and should therefore always be regarded as a lower bound estimate.

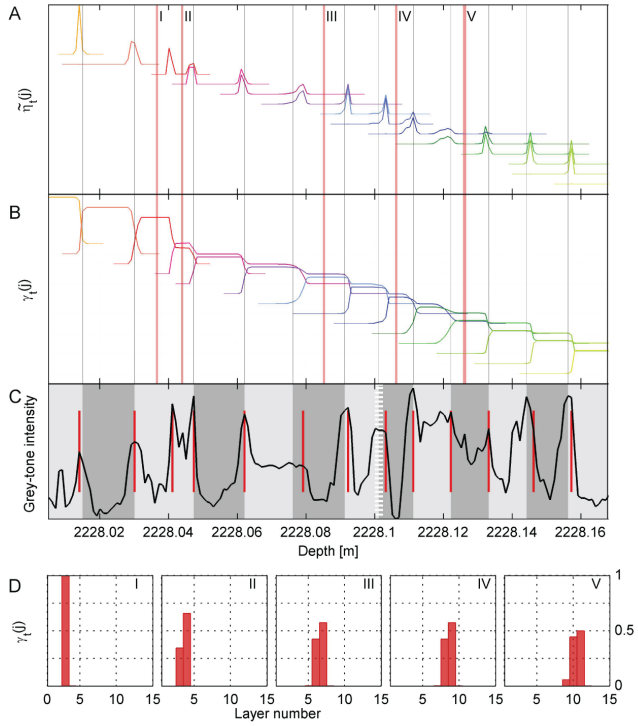

Fig. 3. Layer detection algorithm applied to line-scan data from NGRIP. **(A)** and **(B)**: the posterior probabilities $\tilde{\eta}_t(j) \equiv P(S_{t]} = \ell_j,\ \boldsymbol{o}_{1:T})$ and $\gamma_t(j) \equiv P(S_t = \ell_j,\ \boldsymbol{o}_{1:T})$. Probabilities corresponding to individual layers are drawn in different colours. All probabilities are plotted to the same scale, but shifted slightly downwards with increasing layer number to enhance visibility. **(C)** Line-scan data from NGRIP (black line). The alternating bright and dark grey banding shows the annual layers in the GICC05 chronology, an uncertain layer boundary is marked with white dashes. Superimposed (red lines) are the most likely layer boundaries as inferred from the forward-backward algorithm. These are found as those places where the most likely layer number – i.e. mode of the $\gamma_t(j)$ distributions with t fixed, see examples in **(D)** – changes with depth. **(D)** Inferred probability distributions of the annual layer number ($\gamma_t(j)$) at the positions I–V, indicated with red bars in **(A)** and **(B)**.

Figure 3 shows the output of the layer detection algorithm when applied to a small section of line-scan data from NGRIP. The computed values of $\tilde{\eta}_t(j)$ (Fig. 3a) estimate the probability of ending layer j at t, and thus peaks are locations of highly probable layer boundaries. A general decrease in peak height with distance from the beginning of the data series is caused by an increasing uncertainty as to which layer the boundary belongs. The total probability of ending any layer at a given location t can be calculated by summing the individual contributions from all layers. By subsequent conversion to $\gamma_t(j)$, the probability of being in each layer at a given position can be assessed (Fig. 3b). The further away from the beginning of the data series, the less certain the layer number can be determined. As a result, the maximum

probabilities corresponding to a given layer j slowly decrease as the spread of the distribution of possible annual layers at a specific location (Fig. 3d) increases. The most likely annual layering (Fig. 3c, red lines) can be determined from the changing mode of these distributions (Eq. 7) down the core.

For the line-scan data in Fig. 3, several interesting features of the output of the layer detection algorithm can be noted. First observe how well the inferred most likely layer boundaries correspond to those in the manually obtained GICC05 chronology (Fig. 3c), despite this chronology being based on the annual layer signal in multiple data series. Only one extra layer boundary (depth: 2228.04 m) is reconstructed by the algorithm. Considering the data in this interval, it is seen that indeed the annual layer signal in the line-scan data here is ambiguous: the line-scan data displays a distinct peak, but the distance to the surrounding layers is relatively small.

When considering in detail the inferred values for $\tilde{\eta}_t(j)$, peaks in this probability variable is seen to correspond to peaks in the line-scan data which may or may not indicate an annual layer. In general, only some of these peaks give rise to an inferred layer boundary, while others act to increase the uncertainty estimate of the counted number of layers. Uncertain positioning of a specific layer boundary may also occur, in which case a broad peak is observed in $\tilde{\eta}_t(j)$ (see e.g. around 2228.12 m). Also observe how an ambiguous layer boundary at one location (e.g. at 2228.04 m) results in $\tilde{\eta}_t(j)$ having multiple peaks down core for a given layer j.

The annual layering in the first part of the data series is relatively distinct, and the algorithm recovers the manually deduced annual layering with high confidence, which can be seen from the inferred very spiky $\gamma_t(j)$ probability distributions (Fig. 3d, I). The equivocal interval around 2228.04 m is picked out by the algorithm as containing a possible layer boundary, which is judged to be more likely than not: comparing Fig. 3d, I and Fig. 3d, II, it is seen that after 2228.04 m, there is about 1/3 chance of still being in layer 3, and 2/3 chance of now being in layer 4. The probability distribution is subsequently slowly broadened in a non-trivial way as the algorithm encounters other uncertain layer boundaries (Fig. 3d, III–V).

2.4 Layer detection in successive batches of data

It is not feasible, nor desirable, to run the forward-backward algorithm on several hundred metres of ice core data at once. While also increasing the computational complexity, doing so would require a homogeneous data series, in which the layer thickness distribution as well as the annual layer signal is essentially stationary. Neither of these conditions are satisfied: the mean annual layer thickness changes down the ice core as the combined result of climate-induced variations in past accumulation rates and a general thinning of layers with depth due to ice flow. Furthermore, in different climate regimes the seasonal influx of dust and other impurities to the ice sheet may differ, hereby altering the general annual layer signal in the ice core data (Andersen et al., 2006b).

A better strategy is to divide the total data series into smaller batches, apply the layer detection algorithm to one of these at a time, and subsequently stitch them together by convolving the resulting layer probability distributions. Each of these batches must be sufficiently long to fully exploit the HMM's optimal estimation of layer boundaries, while being short enough that the assumption of a fixed layer thickness distribution and layer signal is reasonable. Here, we have run the layer detection algorithm on batches of data covering approximately 50 yr each. During fast climatic shifts, however, the mean annual layer thickness may display significant changes even more abruptly than this (Alley et al., 1993; Steffensen et al., 2008).

By dividing the full observation sequence into batches and running the layer detection algorithm on each of these individually, some of the information contained in the full data series is lost. Close to the edges of each batch, the lack of surrounding data will in general cause the annual layer boundaries here to be placed less accurately. To a large extent, however, such knowledge can be recovered by choosing data batches in consecutive order and in an overlapping fashion. Furthermore, some of the information inferred from the layering in one batch of data may be incorporated into the next. This will be discussed in Sect. 2.5.

2.5 Optimisation of applied model parameters

The outcome of the forward-backward algorithm is an evaluation of the layering in a data series when given a set of model parameters (θ) to describe the annual layer signal. In addition, however, the forward-backward algorithm can be utilised for evaluating the likelihood of the employed set of model parameters based on the appearance of inferred layers in the observation sequence (Rabiner, 1989; Yu, 2010). This is possible because the computed posterior probabilities, Eqs. (5) and (6), are joined with the probability of the entire observation sequence and – though not always explicitly stated in the preceding equations – conditioned on the model parameters. By normalising the posterior probabilities, the likelihood of the applied model parameters can be determined (see e.g. Winstrup, 2011).

This feature of the annual layer detection algorithm provides the opportunity to implement a learning process, which can select the model parameters best suited for modelling the observations. In other words, the model is able to improve on an initial guess of the appropriate model parameters based on how the data actually appear. By doing so, the model is able to continuously adjust itself to temporal changes in annual layer appearance as well as layer thickness probability distribution, and it gives rise to high performance of the algorithm even when only imperfect knowledge on the appropriate parameter values is available. The abruptness of some climatic events (Steffensen et al., 2008), influencing both the mean

annual layer thickness and the annual layer signal recorded in the ice core records, makes this a very important issue. The opportunity of such training of the annual layer detection algorithm is a major advantage of the hidden Markov modelling approach.

The optimisation has here been performed using the expectation-maximisation (EM) algorithm (Baum et al., 1970; Dempster et al., 1977; Gupta and Chen, 2011). The EM algorithm presents an efficient iterative method for obtaining a point estimate of the (possibly local) maximum likelihood set of parameters. See Appendix A for update equations of the respective model parameters.

Due to this optimisation procedure, it is in principle possible to determine the annual layering in an observation sequence without any prior information on the appearance of layers in the data. However, given the relatively high noise level in the line-scan data, it was found that these data by themselves did not contain sufficient information to produce robust statistics and adequately constrain the layer detection algorithm. Thus, when employed on the line-scan data, some model parameters (σ_d, Φ, and σ_ε) were estimated based on manual layer counts within a smaller data section and subsequently held fixed. Other parameters (μ_d and $\boldsymbol{\varphi}$) were allowed to change with depth and were determined separately for each batch of data.

Another option for constraining the algorithm is to incorporate prior knowledge on the appropriate parameter values. In this case, a modified version of the EM algorithm can be used to retrieve the set of parameter values of (possibly local) maximum posterior probability (see e.g. Gupta and Chen, 2011). Including prior knowledge on parameter values will generally act to stabilise the performance of the layer detection algorithm by lessening its demand on data quality and volume. Prior information may consist of knowledge derived from layering in previous data batches and/or be driven by information based on, for example, the δ^{18}O record, which can provide the algorithm with information on the current climate regime. The high correlation between climate and accumulation rate (Dahl-Jensen et al., 1993) implies that δ^{18}O data may be used for quantifying prior knowledge on relative changes in mean annual layer thickness over shorter depth intervals, within which the ice-flow induced layer thinning can be assumed reasonably constant. By incorporating priors, the layer counting algorithm will be allowed to continually adjust itself to temporal changes in annual layer signature in a more controlled, yet still very flexible, manner.

The potential advantages of incorporating prior knowledge will be explored in future studies, but has here been avoided in order not to guide the layer detection algorithm unnecessarily. The maximum likelihood solution, as employed here, represents the most transparent implementation of the algorithm, which is an important property when assessing its performance.

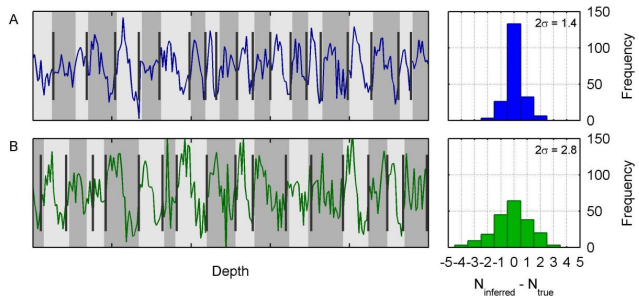

Fig. 4. Test of layer detection algorithm on ensembles of two types of synthetic data. Left panel: example data series. Grey banding indicates the original layer boundaries, based on which the data has been produced. Dark grey bars signify the location of inferred layer boundaries. Right panel: differences between inferred (N_inferred) and true (N_true) number of layers in the data based on an ensemble of 200 data series with approximately 50 layers each. **(A)** Data is produced as a sequence of sinusodials generated by the parameters $\boldsymbol{\varphi} = 1$, $\Phi = 0.5^2$, and $\sigma_\varepsilon = 0.5$ (Eq. 2). **(B)** As **(A)**, but with higher amounts of noise and variability in the layer signal ($\boldsymbol{\varphi} = 1$, $\Phi = 1$, $\sigma_\varepsilon = 1$).

3 Sensitivity tests

The layer detection algorithm was first tested on synthetic data containing a recurring "annual" signal simulating that in the observed data. These data series were constructed as a succession of sinusoidal layer shapes displaying varying amounts of inter-annual variability in amplitude and added Gaussian white noise. Two such data series are shown in Fig. 4. Based on ensembles of synthetic data series, statistics on the performance of the algorithm could be assessed.

The algorithm was observed to work very robustly, and no bias towards either too thick or too thin annual layers was noted. It should be no surprise that the annual layers are easiest to identify when individual layer thicknesses are fairly similar, the layer shapes relatively identical, and the signal-to-noise ratio is high. However, even with individual layers displaying a wide range of amplitudes and containing much noise, the outcome of the layer detection algorithm was very accurate. Consider, for example, an ensemble of synthetic data series, each consisting of approximately 50 layers, in which the annual layer signal displays a moderate amount of noise and inter-annual variability (an example is shown in Fig. 4a). Based on the inferred distribution of error in the resulting most likely layer estimates, the maximum counting error (taken as 2σ of the distribution) was found to be 1.4 yr, i.e. $< 3\,\%$. When increasing the noise and inter-annual layer variability further (e.g. Fig. 4b), the maximum counting error grows, but in none of the considered scenarios did it exceed 6 %. Some of the synthetic data series hereby produced were so noisy that it would be problematic to reliably identify the annual layers by eye. The automated counting algorithm, on the contrary, is able to count these in an unbiased manner due to knowledge of the annual layer thickness probability

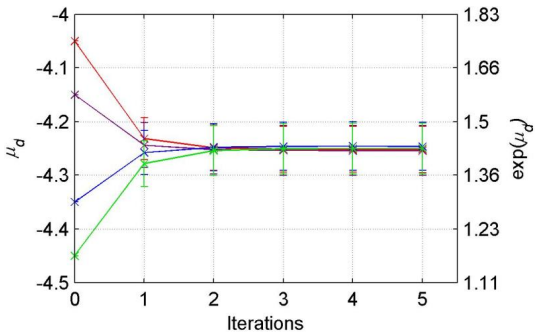

Fig. 5. Evolution of the derived layer thickness parameter μ_d as function of the number of EM iterations, when given incorrect input values of this parameter. Median of the resulting layer thickness distribution is calculated as $\exp(\mu_\mathrm{d})$. Data points are based on ensembles of 200 synthetic data series (sinusoidals produced with parameters $\boldsymbol{\varphi} = 1$, $\Phi = 0.5^2$, and $\sigma_\varepsilon^2 = 0.5^2$) that each contains approximately 50 layers. Error bars signify the resulting 1σ spread within ensemble members. For the considered range of initial values, the derived value of μ_d converged to the original input value of $\mu_\mathrm{d} = -4.25$ after just a few iterations.

distribution in combination with a robust statistical measure for judging what is most likely to be an annual layer.

Inherent to the forward-backward algorithm is the simultaneous derivation of a confidence interval for the number of annual layers within a data section. The accuracy of this uncertainty estimate was evaluated by counting the number of ensemble members for which the true number of annual layers in the data series was outside the estimated uncertainty bounds. For this simple case of synthetic data with known annual layer template and model parameters, the estimated uncertainties were found to be obtained very reliably.

We also tested the capability of the layer detection algorithm to iteratively find a best set of model parameters based on the appearance of layers in the data. For an ensemble of synthetic data series with a moderate amount of inter-annual variability and white noise, and a sensible choice of initial parameter values, the number of iterations required before convergence was reached was generally 5 or less (Fig. 5). More details are available in Winstrup (2011). However, the ability of the algorithm to swiftly converge towards an appropriate set of parameter values is highly dependent on the overall difficulty of recognising the annual layers in the data as well as on the appropriateness of the initial parameter input to the algorithm. In unfavourable conditions, correct estimates of the parameters may indeed be impossible to infer. Yet, in all investigated cases, the algorithm quickly converged to the set of parameters used for constructing the data, with a probability distribution as expected when using relatively small samples containing just 50 layers.

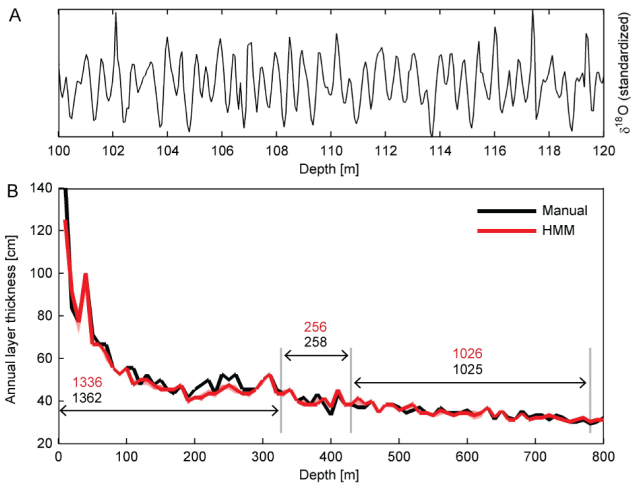

Fig. 6. (A) Example of the DYE-3 δ^{18}O-record. The data was measured with approximately 8 samples per year. **(B)** Evolution in mean layer thicknesses for the upper 800 m of the record, spanning approximately the last 2000 yr (1979–79 CE). Annual layer thicknesses are calculated over 10 m sections. The 95 % confidence interval derived for the automated counting is shown as a (very narrow) red shaded band. Uncertainties for the manual counting are negligible. The depth of three historically dated volcanic reference horizons (Öraefajökull, Hekla, Vesuvius) are indicated with grey bars. Red numbers denote automated layer counts within each section, which are to be compared to the historically known time durations (given in black).

4 Results

The layer detection algorithm was implemented for two data sets: a stable-isotope record from DYE-3, displaying a distinct seasonal pattern, and visual stratigraphy data from NGRIP, in which the annual layer signal may sometimes be ambiguous, but is maintained to great depths.

For DYE-3, the algorithm was run over the depth interval 0–779.99 m, which covers the time period 1979–79 CE (Vesuvius eruption). The accumulation rate at DYE-3 is sufficiently high that firn diffusion has not significantly altered the annual oscillations in the stable-isotope record, and although melt layers occasionally obscure the annual signal, the core can be manually layer counted with high confidence. In addition, the record contains several historically dated volcanic horizons (Clausen et al., 1997). Using such reference horizons for synchronising individual ice cores, the manual dating of this core was not only based on the stable-isotope data considered here, but also included data from GRIP and NGRIP (Vinther et al., 2006). Together, these data constituted the basis for the upper part of the GICC05 chronology.

Before analysis, the δ^{18}O record was normalised with respect to the data variance over an interval covering approximately three years (Fig. 6a). As previously discussed, the observed data and its derivative were employed as two separate data series. The derivative was calculated based on

Table 1. Results from automated and manual annual layer counts. For the automated counting, the 95 % confidence intervals (numbers in parentheses) are given based on the derived annual layer probability distributions. Note that these distributions are not necessarily symmetric around the most likely annual layer count. For the manual counting, the 95 % uncertainty estimate is based on the maximum counting error (MCE) from the GICC05 chronology. For DYE-3, the ice core section is delimited by two absolute dated horizons (top of core, Vesuvius eruption), and hence no uncertainty is associated with the total number of layers in the interval.

Data	Automated counting	Manual counting	Difference
DYE-3	(1919) 1926 (1932)	1900	1.4 %
NGRIP, GI-12	(882) 894 (905)	(792) 839 (886)	6.6 %
NGRIP, Onset of GI-12	(284) 291 (298)	(266) 278 (290)	4.7 %
NGRIP, GS-13	(1234) 1249 (1265)	(1142) 1204 (1266)	3.7 %

Table 2. Comparison of time intervals between three historically dated volcanic eruptions that are visible as marker horizons in the DYE-3 ice core. Depths are from Vinther et al. (2006). ΔT_{hist} are the historically known time intervals, ΔT_{calc} indicates the calculated time intervals and the inferred 95 % confidence intervals.

Eruption	Depth [m]	Historical age (CE)	ΔT_{hist} [yr]	Calculated age (CE)	ΔT_{calc} [yr]
Öraefajökull	326.70	1362		1336	
			258		(254) 256 (257)
Hekla	429.24	1104		1080	
			1025		(1022) 1026 (1031)
Vesuvius	779.99	79		54	

a smoothed version of the observed record as obtained by applying a first-order Savitzky-Golay filter (Savitzky and Golay, 1964).

Given the generally very well-defined annual signal in the δ^{18}O record, the layer detection algorithm was run in the most autonomous way possible, in which all parameters included in the layer model (i.e. those defining the layer thickness probability distribution as well as those defining layer shape and its inter-annual variability) were allowed to change with depth. These were optimally chosen by the algorithm according to the layer signal in the data (Appendix A). The algorithm was allowed to perform five iterations for each batch of data, at which point it was assumed that convergence had been reached. The resulting set of model parameter values was subsequently used as input to the next batch.

In Fig. 6b, the inferred evolution in mean layer thickness with depth is compared to that manually obtained. The algorithm performs very well: over the entire depth interval, the layer detection algorithm infers a total of 1926 ± 7 layers (compared to the actual number of 1900 yr) (Table 1). It is a discrepancy of merely 1.4 %, most of which arises within a short interval in the upper part of the core (200-270 m depth). This interval coincides with the upper part of an anomalous increase in layer thicknesses with depth caused by upstream depositional effects (Reeh et al., 1985; Vinther et al., 2006), which may be the cause of an exceptionally large inter-annual variation in layer thicknesses observed within this section. The detailed similarity of the two layer counts can be seen from the two records of mean layer thicknesses, which share most of the short-period variability. Over the considered time interval, three historical eruptions occurred, and these were used as tie-points in the manually counted GICC05 chronology. A comparison of the inferred time intervals between the three volcanic reference horizons shows excellent agreement with discrepancies of just 1 and 2 yr, respectively (Table 2).

Note, however, that the inferred uncertainty of the automated layer estimate is very small (0.4 %) and does not include the true age. There are several reasons for this. One reason is that the algorithm has not been allowed to fully include the uncertainty that lies within the range of parameters that might reasonably describe an annual layer signal in the data series: the algorithm is allowed to select the most likely set of model parameters, but for the derivation of the corresponding confidence intervals, the uncertainty associated with this choice is assumed negligible. Additionally, the layer counting algorithm is assumed not to be biased towards counting either too few or too many layers. An extra layer at one position is therefore likely to be counteracted by a missing layer later. With the individual uncertainties for each layer being allowed to partly compensate for each other, the resulting uncertainty estimate will grow increasingly slowly with depth. Consequently, when applied to a long data section, a relatively narrow uncertainty interval is inferred. However, the small uncertainty bounds resulting from the assumption of unbiased counting are likely to be optimistic when dealing with real data.

The maximum counting error (MCE) estimate of the GICC05 chronology is a conservative estimate of the involved uncertainties (Andersen et al., 2006b). Although MCE can be regarded as a 2σ-error bound, it takes into account that the manual counting procedure may be biased, and the uncertainty estimates of individual layer boundaries are simply summed up ($\pm$ 1/2 yr per uncertain year). This leads to an almost linear increase in uncertainty with depth. Under the assumption of unbiased counting, for comparison, the variance of the annual layer distribution will grow approximately linearly with depth, and the 1σ uncertainty bounds will accordingly grow with the square root.

The algorithm was then tested over three sections of line-scan data representative for the deeper part of the NGRIP ice core: a warm period (Greenland Interstadial 12 – GI-12, depth 2200–2220 m, duration approx. 1200 yr), the preceding cold period (Greenland Stadial 13 – GS-13, depth 2225–2240 m, duration approx. 840 yr), and the transition between the two. The associated time interval is approximately 45.9–48.3 kyr b2k. The inferred layering is compared to the manually counted layers in the GICC05 chronology. For this depth interval, the GICC05 chronology is mainly based on the high-resolution records of electrolytic melt water conductivity, ECM (an acidity index) and visual stratigraphy from the line-scan data. During the interstadial, increased layer thicknesses also permitted records of water-soluble ion concentrations, particularly Na^+, to be employed (Svensson et al., 2008). The manual counting uncertainty was estimated to be around 5 %.

Before application of the layer detection algorithm to the line-scan data, the data was treated with the aim of increasing the similarity of individual layers. This was done by first log-transforming the data (to stabilise the peak heights of individual years) and subsequently normalising data according to its minimum and maximum values (to remove the differences in peak heights between different climate regimes). The window length used for normalisation was 10 cm and hence contained several annual layers. An example of the transformed grey-tone intensity data is shown in Fig. 2a.

Climate conditions were relatively stable during both the stadial and interstadial period under investigation (Fig. 7a). Most of the parameters describing how an annual layer is expressed in the line-scan data are therefore expected to stay more or less unchanged throughout each of the two periods. For each depth interval, several of the model parameters (σ_d, Φ, σ_ε) were estimated based on manual layer counts in the upper 25 % of the data section and maintained as fixed values throughout the interval. The location parameter for the annual layer thickness distribution, μ_d, and the vector parameter $\boldsymbol{\varphi}$ describing the mean layer trajectory, were allowed to change with depth and chosen according to appearance of the data.

During the cold GS-13, the algorithm performs very well (although certainly inferior to when applied to the much more clearly defined layer signals in the DYE-3 δ^{18}O record).

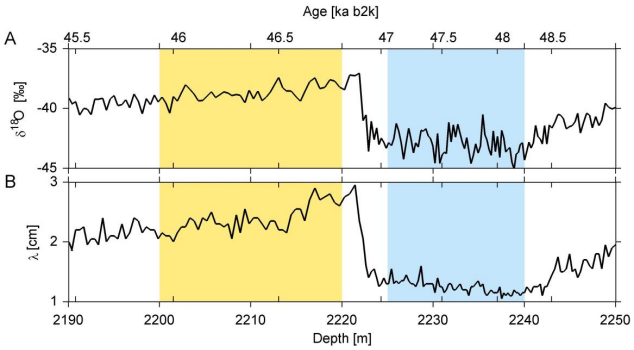

Fig. 7. Evolution of 20-yr means of **(A)** δ^{18}O and **(B)** manually obtained annual layer thicknesses (λ) over the selected NGRIP depth interval. Note the similarity of the two curves. Yellow area is the selected section from the warm period GI-12. Blue area is the considered part of the previous stadial GS-13.

The most likely duration of the selected period is by the algorithm inferred to be 1249 yr, which is to be compared to an estimated duration of 1204 yr in the GICC05 chronology (Table 1). It corresponds to a relative discrepancy of just 3.7 %, and the obtained layer count is within the uncertainty of the manual layer count. The similarity between the manual and automated layer counts can again be observed from the general evolution in layer thicknesses with depth (Fig. 8c) that displays the same decreasing trend as the manually obtained annual layer thicknesses.

From the mean layer thicknesses in Fig. 8c, it is seen that only in two small sections are the inferred layer counts outside the maximum counting error on the manual layer estimate. In particular a spike without manual counterpart occurring around a depth of 2233 m is evident. The existence of such spikes is at least partly due to the algorithm being negligent of prior information from previous batches and determining the most likely model parameters (describing e.g. the mean annual layer thickness) separately for each batch. If by chance the layering within a batch is particularly ambiguous, the algorithm may therefore be led astray here. Manual counting within a section is guided by experience from previous data on how layers appear, and hence not affected as much by ambiguities within a confined interval. More data series than the line-scan data were employed for the manual layer counting, which helps to stabilise the performance of the manual counting procedure too. These differences between the manual and automated approach can also explain the slightly higher degree of variability in mean layer thickness between successive intervals observed in the automated layer count.

If comparing in details the layer boundaries obtained from the automated counting algorithm to those of GICC05, it becomes clear that, to a large extent, disparities are caused by genuine ambiguities in the annual layering of the line-scan data. Furthermore, most of the ambiguous layers do influence

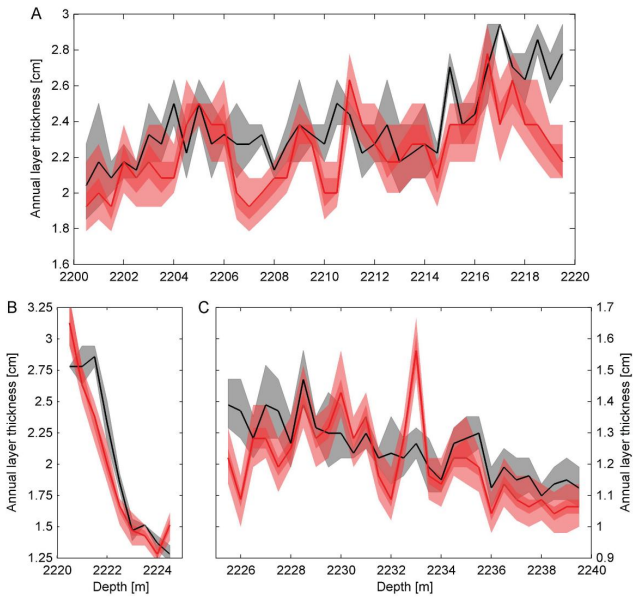

Fig. 8. The evolution in mean annual layer thicknesses during three climatic periods: **(A)** the warm interstadial GI-12, **(B)** the onset of GI-12, and **(C)** the cold stadial GS-13, as inferred by the automated layer detection algorithm (red) and manually counted annual layers (black). The corresponding approximate 2σ-uncertainty bounds are shown as grey and red shaded areas, respectively. Mean annual layer thicknesses are taken over sections of 50 cm. Note the different y-axes in the individual plots.

the inferred number of counted annual layers by broadening the annual layer probability distribution. Disagreements in inferred total number of counted layers are mainly caused by differences in how many of these ambiguous layers are being counted as years.

Given that the annual layer signal in the line-scan data is most distinct during cold periods, the annual layer detection algorithm is likely to work best here. When applied to data from the interstadial GI-12, the performance of the algorithm is indeed weaker. Nevertheless, the percentage-wise deviation from the number of manually counted layers within the considered data section is still just 6.6 %, and the two associated 95 % confidence intervals are overlapping, although only with their respective tails (Table 1). To some extent, the inferred layer thicknesses does match those of the GICC05 chronology (Fig. 8a), but not as closely as for the cold period. Generally, the algorithm seems to be slightly too optimistic in its layer detection for this section, with most of the layers deemed uncertain in the manually counted chronology here being counted as actual layers. In future work, this issue will be addressed by careful investigation of the nature of non-annual features in the data series, and it can be dealt with by simply allowing a higher degree of white noise in the individual layer signals, leading to fewer ambiguous layers being counted as annual layers.

Finally, the layer detection algorithm was run over the data section containing the onset of interstadial GI-12. As for the previous data sections, the algorithm was run downwards through the core, starting in ice deposited during the warm interstadial (with large annual layer thicknesses), and going towards the older and deeper ice deposited during the stadial (small layer thicknesses). Based on manual layer counts, the mean layer thicknesses across the transition is known to change with more than a factor two (Fig. 7b), and also the appearance of the line-scan data changes dramatically over the transition.

To some extent, the changes in appearance of the line-scan data over the onset of the interstadial was dealt with by the previously described normalisation procedure applied to the grey-tone intensity profile. However, the annual signal in the visual stratigraphy changes more with climate, and hence with depth, than can be rectified by merely adjusting the peak height of the signal. We have here allowed the mean annual layer signal to change down the ice core, but all other changes in annual layer signal have not been taken into account. The generally enhanced amount of noise in the annual layer signal during warmer periods is, for example, neglected. Estimates for the non-varying model parameters were again determined based on line-scan data in the upper 25 % of the section.

Despite the abrupt change in mean layer thickness and annual layer signal, the layer detection algorithm is able to adapt to the changing environment and find an appropriate value of the mean annual layer thickness throughout the transitional zone (Fig. 8b). The resulting number of counted layers within the transitional period differs just 4.7 % from the manual layer counts, which is within the uncertainty of the manual counting (Table 1).

We also tested the performance of the layer detection algorithm on the line-scan data when describing the annual layer signal with three instead of two principal components. In this case, the discrepancy in annual layer count during GS-13 was reduced to just 0.3 %, and resulted in near perfect layer identification throughout the interval. However, using three principal components led to significantly larger discrepancies in layer counts for the interstadial and the onset of the interstadial period (~ 8 %), where the increased flexibility in annual layer template allowed a larger number of ambiguous layers to be identified as annual layers.

5 Conclusions

For paleoclimate archives with sufficient resolution to resolve sub-annual variability, annual layer counting provides a means of obtaining a very accurate chronology. Establishing such chronologies has so far predominantly been accomplished manually. The algorithm developed here represents a first step towards a high-quality automated method of annual layer counting. Based on the statistical framework of hidden

Markov modelling, originally developed for machine speech recognition, it presents a mathematically rigorous yet efficient method to determine the most likely layering in a data series and its associated uncertainty. The algorithm imitates the manual procedures, while being based on objective yet flexible criteria for annual layer recognition.

The layer detection algorithm has been applied to a stable-isotope record from DYE-3 and to visual stratigraphy data from the NGRIP ice core, Greenland, in which the annual signal appears to be maintained to great depths. This data series may therefore potentially be used for extending the manually counted GICC05 chronology further back in time. However, for the current purpose of demonstrating the integrity of the algorithm, it has been run over data sections where manual annual layer counting has already been carried out.

The algorithm shows high skill in reproducing the manually counted timescales. For the DYE-3 stable-isotope record, which displays a very clear annual layer signal, the total number of inferred layers is just 1.4 % higher than the correct number, which is known based on the Vesuvius eruption (79 CE) reference horizon in the core. The inferred timescale also shows excellent agreement with the time interval between other historically dated volcanic horizons. In the NGRIP line-scan data, the annual layer signal is more ambiguous. Nevertheless, for a data section from a cold climate period, the resulting number of inferred annual layers deviates just 3.7 % from the manually counted timescale. During warm climate regimes, the annual layer signal in the visual stratigraphy is even less apparent. Consequently, the algorithm was less confident in identifying the annual layers here, but it still obtained an annual layer count within 6.6 % of the manual layer count. The algorithm was also run over a short line-scan data section experiencing the onset of an interstadial. It showed high skill adapting to the changing environment while passing over this transitional period, during which the annual layer thicknesses were halved over merely five metres.

Although the layer detection algorithm was developed with ice core data in mind, it has been developed in a general setting which allows it to be adapted for use with other kinds of annually laminated data. When applied to ice core records, one of the most interesting prospects may be the possible development of the algorithm into a multi-parameter method that will be undertaken in future studies.

Appendix A

Update equations for the model parameters

By means of the EM algorithm, the forward-backward algorithm may be trained by successive updates of the model parameters used in the characterisation of an annual layer. An initial set of parameters is used as input to the forward-backward algorithm, and a first segmentation of the observations into annual layers is proposed. The obtained segmentation along with the observed data is then used for computing the set of parameter values having the highest conditional likelihood. This set of parameter values incorporates knowledge from the appearance of the data, and it can be shown (see e.g. Gupta and Chen, 2011) that its likelihood will always be equal or higher than that of the previous set of parameter values. By iteratively computing the layer segmentation probabilities with the forward-backward algorithm and updating the annual layer parameters, the algorithm is able to learn the appropriate parameter values, and it will converge towards a (local) maximum likelihood of these.

In the $(k+1)$-th iteration, the maximum likelihood re-estimation equations for the forward-backward algorithm, when applied to annual layer detection based on the described annual layer template (Eq. 2), are given below.

Layer thickness distribution parameters:

$$\hat{\mu}_{\mathrm{d}} = \frac{\sum_{t,j,d} \eta_t(j,\ d) \ln d}{\sum_{t,j,d} \eta_t(j,\ d)}$$

$$\hat{\sigma}_{\mathrm{d}}^2 = \frac{\sum_{t,j,d} \eta_t(j,\ d) \left(\ln d - \hat{\mu}_{\mathrm{d}}\right)^2}{\sum_{t,j,d} \eta_t(j,\ d)}.$$

In the update equations of the remaining model parameters, the following notation is introduced: the set of maximum likelihood parameter values obtained during the k-th iteration is denoted $\theta^{(k)}$. A proposed layer segment is denoted $\boldsymbol{O}_j$, and is a vector consisting of both the observed data as well as its derivative. The design matrix $\mathbf{X}$ is correspondingly appended to contain not only the applied basis functions used for characterising the annual layer signal in the line-scan data, but also the derivative of these. The residuals of layer segment $\boldsymbol{O}_j$ relative to the layer template are collected in the vector $\boldsymbol{E}_j$. The difference in variance of the white noise component in the two data series is given by the $2d \times 2d$ diagonal matrix $\mathbf{W}$, with d being the duration of a layer. The vector $\boldsymbol{r}_j$ is the random component previously introduced in Eq. (2).

An updated value of the mean annual layer signal parameter ($\boldsymbol{\varphi}$), the random component covariance matrix (Φ), and the variance of the added white noise component (σ_ε^2) can then be calculated by

$$\hat{\boldsymbol{\varphi}} = \left(\sum_{t,j,d} \eta_t(j,\ d) \mathbf{X}^{\mathrm{T}} \mathbf{W}^{-1} \mathbf{X}\right)^{-1} \sum_{t,j,d} \eta_t(j,\ d) \mathbf{X}^{\mathrm{T}} \mathbf{W}^{-1} \left(\boldsymbol{O}_j - \mathbf{X} \mathbb{E}\left[\boldsymbol{r}_j \middle| \boldsymbol{O}_j,\ \theta^{(k)}\right]\right),$$

$$\hat{\Phi} = \frac{\sum_{t,j,d} \eta_t(j,\ d) \mathbb{E}\left[\boldsymbol{r}_j \boldsymbol{r}_j^{\mathrm{T}} \middle| \boldsymbol{O}_j,\ \theta^{(k)}\right]}{\sum_{t,j,d} \eta_t(j,\ d)},$$

and

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{\sum_{t,j,d} \eta_t(j,\ d)\, \mathbb{E}\left[\boldsymbol{E}_j^{\mathrm{T}} \mathbf{W}^{-1} \boldsymbol{E}_j \middle| \boldsymbol{O}_j,\ \theta^{(k)} \right]}{2 \sum_{t,j,d} \eta_t(j,\ d)\, d},$$

with the expectation values in the equations above given as

$$\mathbb{E}\left[\boldsymbol{r}_j \boldsymbol{r}_j^{\mathrm{T}} \middle| \boldsymbol{O}_j,\ \theta^{(k)} \right] = \mathbb{E}\left[\boldsymbol{r}_j \middle| \boldsymbol{O}_j,\ \theta^{(k)} \right] \mathbb{E}\left[\boldsymbol{r}_j \middle| \boldsymbol{O}_j,\ \theta^{(k)} \right]^{\mathrm{T}} + \mathrm{cov}\left[\boldsymbol{r}_j \middle| \boldsymbol{O}_j,\ \theta^{(k)} \right]$$

$$\mathbb{E}\left[\boldsymbol{E}_j^{\mathrm{T}} \mathbf{W}^{-1} \boldsymbol{E}_j \middle| \boldsymbol{O}_j,\ \theta^{(k)} \right] = \left(\boldsymbol{O}_j - \mathbf{X}\left(\hat{\boldsymbol{\varphi}} + \mathbb{E}\left[\boldsymbol{r}_j \middle| \boldsymbol{O}_j,\ \theta^{(k)} \right] \right) \right)^{\mathrm{T}} \mathbf{W}^{-1} \left(\boldsymbol{O}_j - \mathbf{X}\left(\hat{\boldsymbol{\varphi}} + \mathbb{E}\left[\boldsymbol{r}_j \middle| \boldsymbol{O}_j,\ \theta^{(k)} \right] \right) \right) + \mathrm{tr}\left(\mathbf{X}^{\mathrm{T}} \mathbf{W}^{-1} \mathbf{X}\, \mathrm{cov}\left[\boldsymbol{r}_j \middle| \boldsymbol{O}_j,\ \theta^{(k)} \right] \right).$$

See Winstrup (2011) for a derivation and more detailed review of the above equations.

Acknowledgements. This work was supported by the Danish National Research Foundation under Centre for Ice and Climate at the University of Copenhagen. The authors would also like to thank Inge Lehmann Fonden that contributed with financial support. The work is a contribution to the NorthGRIP ice core project, which is directed and organised by the Department of Geophysics at the Niels Bohr Institute for Astronomy, Physics and Geophysics, University of Copenhagen. It is being supported by funding agencies in Denmark (SNF), Belgium (FNRS-CFB), France (IFRTP and INSU/CNRS), Germany (AWI), Iceland (RannIs), Japan (MEXT), Sweden (SPRS), Switzerland (SNF) and the United States of America (NSF). The work is also a contribution to the Copenhagen Ice Core Dating Initiative that is supported by a grant from the Carlsberg Foundation.

Edited by: H. Fischer

References

Alley, R. B., Meese, D. A., Shuman, C. A., Gow, A. J., Taylor, K. C., Grootes, P. M., White, J. W. C., Ram, M., Waddington, E. D., Mayewski, P. A., and Zielinski, G. A.: Abrupt increase in Greenland snow accumulation at the end of the Younger Dryas event, Nature, 362, 527–529, doi:10.1038/362527a0, 1993.

Alley, R. B., Shuman, C. A., Meese, D. A., Gow, A. J., Taylor, K. C., Cuffey, K. M., Fitzpatrick, J. J., Grootes, P. M., Zielinski, G. A., Ram, M., Spinelli, G., and Elder, B.: Visual-stratigraphic dating of the GISP2 ice core: Basis, reproducibility, and application, J. Geophys. Res.-Oceans, 102, 26367–26381, doi:10.1029/96JC03837, 1997.

Andersen, K. K., Ditlevsen, P. D., Rasmussen, S. O., Clausen, H. B., Vinther, B. M., Johnsen, S. J., and Steffensen, J. P.: Retrieving a common accumulation record from Greenland ice cores for the past 1800 years, J. Geophys. Res.-Atmos., 111, D15106, doi:10.1029/2005jd006765, 2006a.

Andersen, K. K., Svensson, A., Johnsen, S. J., Rasmussen, S. O., Bigler, M., Rothlisberger, R., Ruth, U., Siggaard-Andersen, M. L., Steffensen, J. P., 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, doi:10.1016/j.quascirev.2006.08.002, 2006b.

Baum, L. E. and Petrie, T.: Statistical Inference for Probabilistic Functions of Finite State Markov Chains, Ann. Math. Stat., 37, 1554–1563, doi:10.1214/aoms/1177699147, 1966.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N.: A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains, Ann. Math. Stat., 41, 164–171, 1970.

Bigler, M., Svensson, A., Kettner, E., Vallelonga, P., Nielsen, M. E., and Steffensen, J. P.: Optimization of High-Resolution Continuous Flow Analysis for Transient Climate Signals in Ice Cores, Environ. Sci. Technol., 45, 4483–4489, doi:10.1021/es200118j, 2011.

Bishop, C. M.: Pattern Recognition and Machine Learning, Information Science and Statistics, Springer, New York, 738 pp., 2006.

Bulla, J. and Bulla, I.: Stylized facts of financial time series and hidden semi-Markov models, Comput. Stat. Data An., 51, 2192–2209, doi:10.1016/j.csda.2006.07.021, 2006.

Clausen, H. B., Hammer, C. U., Hvidberg, C. S., Dahl-Jensen, D., Steffensen, J. P., Kipfstuhl, J., and Legrand, M.: A comparison of the volcanic records over the past 4000 years from the Greenland Ice Core Project and Dye 3 Greenland ice cores, J. Geophys. Res.-Oceans, 102, 26707–26723, doi:10.1029/97JC00587, 1997.

Dahl-Jensen, D., Johnsen, S. J., Hammer, C., Clausen, H. B., and Jouzel, J.: Past accumulation rates derived from observed annual layers in the GRIP ice core from Summit, Central Greenland, in: Ice in the Climate System, edited by: Peltier, W. R., NATO ASI Series I, 12, Springer, New York, 517–532, 1993.

Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum likelihood from incomplete data via the EM Algorithm, J. Roy. Stat. Soc. B. Met., 39, 1–38, 1977.

Faisan, S., Thoraval, L., Armspach, J. P., Metz-Lutz, M. N., and Heitz, F.: Unsupervised learning and mapping of active brain functional MRI signals based on hidden semi-Markov event sequence models, IEEE T. Med. Imag., 24, 263–276, doi:10.1109/Tmi.2004.841225, 2005.

Fisher, D. A., Reeh, N., and Clausen, H. B.: Stratigraphic noise in time series derived from ice cores, Ann. Glaciol., 7, 76–83, 1985.

Gish, H. and Ng, K.: Parametric trajectory models for speech recognition, in: Proceedings of ICSLP 96, Fourth International Conference on Spoken Language Processing, Philadelphia, PA, USA, 466–469, doi:10.1109/ICSLP.1996.607155, 1996.

Gupta, M. R. and Chen, Y.: Theory and use of the EM algorithm, Found. Trend. Signal Proc., 4, 223–296, doi:10.1561/2000000034, 2011.

Hamilton, W. L. and Langway, C. C.: A Correlation of Microparticle Concentrations with Oxygen Isotope Ratios in 700 Year Old Greenland Ice, Earth Planet. Sc. Lett., 3, 363–366, doi:10.1016/0012-821X(67)90062-3, 1968.

Jelinek, F., Bahl, L. R., and Mercer, R. L.: Design of a Linguistic Statistical Decoder for Recognition of Continuous Speech, IEEE T. Inform. Theory, 21, 250–256, doi:10.1109/TIT.1975.1055384, 1975.

Kim, S. and Smyth, P.: Segmental Hidden Markov Models with Random Effects for Waveform Modeling, J. Mach. Learn. Res., 7, 945–969, 2006.

Kim, S., Smyth, P., and Luther, S.: Modeling waveform shapes with random effects segmental hidden Markov models, in: Proceedings of UAI 04, 20th Conference on Uncertainty in Artificial Intelligence, Banff, Canada, 309–316, 2004.

McGwire, K. C., McConnell, J. R., Alley, R. B., Banta, J. R., Hargreaves, G. M., and Taylor, K. C.: Dating annual layers of a shallow Antarctic ice core with an optical scanner, J. Glaciol., 54, 831–838, doi:10.3189/002214308787780021, 2008.

Ostendorf, M., Digalakis, V. V., and Kimball, O. A.: From HMMs to segment models: A unified view of stochastic modeling for speech recognition, IEEE T. Speech Audi. P., 4, 360–378, doi:10.1109/89.536930, 1996.

Rabiner, L. R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proc. IEEE, 77, 257–286, doi:10.1109/5.18626, 1989.

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., Rothlisberger, R., Fischer, H., Goto-Azuma, K., Hansson, M. E., and Ruth, U.: A new Greenland ice core chronology for the last glacial termination, J. Geophys. Res.-Atmos., 111, D06102, doi:10.1029/2005jd006079, 2006.

Reeh, N., Johnsen, S. J., and Dahl-Jensen, D.: Dating the DYE 3 deep ice core by flow model calculations, in: Greenland Ice Core: Geophysics, Geochemistry, and the Environment, edited by: Langway, C. C., Oeschger, H., and Dansgaard, W., Geophys. Monogr. Series, 33, AGU, Washington, D.C., 57–65, doi:10.1029/GM033p0057, 1985.

Russell, M. J. and Holmes, W. J.: Linear trajectory segmental HMMs, IEEE Signal Proc. Lett., 4, 72–74, 1997.

Savitzky, A. and Golay, M. J. E.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem., 36, 1627–1639, doi:10.1021/ac60214a047, 1964.

Smith, C. L., Fairchild, I. J., Spotl, C., Frisia, S., Borsato, A., Moreton, S. G., and Wynn, P. M.: Chronology building using objective identification of annual signals in trace element profiles of stalagmites, Quat. Geochronol., 4, 11–21, doi:10.1016/j.quageo.2008.06.005, 2009.

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

Svensson, A., Nielsen, S. W., Kipfstuhl, S., Johnsen, S. J., Steffensen, J. P., Bigler, M., Ruth, U., and Rothlisberger, R.: Visual stratigraphy of the North Greenland Ice Core Project (NorthGRIP) ice core during the last glacial period, J. Geophys. Res.-Atmos., 110, D02108, doi:10.1029/2004jd005134, 2005.

Svensson, A., Andersen, K. K., Bigler, M., Clausen, H. B., Dahl-Jensen, D., Davies, S. M., Johnsen, S. J., Muscheler, R., Rasmussen, S. O., Rothlisberger, R., Steffensen, J. P., and Vinther, B. M.: The Greenland Ice Core Chronology 2005, 15–42 ka, Part 2: comparison to other records, Quaternary Sci. Rev., 25, 3258–3267, doi:10.1016/j.quascirev.2006.08.003, 2006.

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

Vinther, B. M., Clausen, H. B., Johnsen, S. J., Rasmussen, S. O., Andersen, K. K., Buchardt, S. L., Dahl-Jensen, D., Seierstad, I. K., Siggaard-Andersen, M. L., Steffensen, J. P., Svensson, A., Olsen, J., and Heinemeier, J.: A synchronized dating of three Greenland ice cores throughout the Holocene, J. Geophys. Res.-Atmos., 111, D13102, doi:10.1029/2005jd006921, 2006.

Winstrup, M.: An Automated Method for Annual Layer Counting in Ice Cores – and an application to visual stratigraphy data from the NGRIP ice core, PhD thesis, Centre for Ice and Climate, Niels Bohr Institute, http://www.iceandclimate.nbi.ku.dk/publications/theses/M.Winstrup2011.pdf, last access: July 2012, University of Copenhagen, Copenhagen, 173 pp., 2011.

Yu, S. Z.: Hidden semi-Markov models, Artif. Intell., 174, 215–243, doi:10.1016/j.artint.2009.11.011, 2010.