Extracting the annual signal from Greenland ice core chemistry and isotopic records

Sune Olander Rasmussen,[1]
Katrine Krogh Andersen,
Marie-Louise Siggaard-Andersen,
and Henrik B. Clausen

Department of Geophysics, NBIfAFG, University of Copenhagen,
Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark

Stratigraphic dating of ice cores by identification and counting of annual cycles in e.g. chemical measurements requires skill and experience. The work presented here investigates a method of data enhancement which is a first step towards an automated and more objective method of annual layer counting. The method of dynamical decorrelation is briefly introduced and is applied to data from Site D and NorthGRIP in central Greenland. With this method the measured data series are decomposed into a number of independent source series, one of which exhibits a more pronounced annual variation than the input data themselves. The annual variation is more regular in that (1) some double and triple peaks in the measured series are replaced by single peaks in the extracted signal and (2) the resulting annual peaks have a much more uniform height. A simple method of determining the number of annual peaks in a series is set up. Using this method, it is shown that it is easier to determine the number of annual peaks in the series produced by dynamical decorrelation, than in the original data series. Dynamical decorrelation may thus be used to improve data series prior to dating.

Introduction

In this paper we investigate the possibility of extracting a common annual signal from combined chemistry and isotopic records from central Greenland ice cores. Records from Site D (Steffensen, 1988) and NorthGRIP (Dahl-Jensen and others, 2001) are analysed. The objective is to extract a signal that exhibits a more pronounced annual variation than what can be found directly in the measured records. The extracted signal can be used as the basis for e.g. automated counting of annual layers in ice core records.

The method introduced here is based upon the assumption that the measured records can be considered to be linear combinations of a set of underlying source signals. The method allows us to determine the source signals from the measured records if their autocorrelation functions are different. The extracted source signals may reflect variations in physical parameters such as temperature, but they may also be purely mathematical constructions. From an ice core dating point of view, even the latter could be useful if the obtained signal shows pronounced annual variation.

---

[1]Corresponding author, e-mail olander@gfy.ku.dk.

Treating ice core records as time series

Most ice core records are sampled at constant depth intervals. If precipitation is unevenly distributed in time, ice core records are equivalent to time series sampled at varying (and unknown) time intervals. Firnification processes, thinning of annual layers due to ice flow, and changes in the amount of precipitation during different climatic periods will lead to further changes in the sampling rate with depth.

In this work, we use data series each spanning only around 20 years, corresponding to less than 10 meters. Within each period, we may assume the overall climatic situation to be unchanged, and thus that the annual precipitation does not undergo large changes. Firnification processes and thinning of annual layers may also be disregarded when using short series at these depths. It will be assumed that the precipitation is relatively evenly distributed throughout the year, which is supported by investigations by e.g. Shuman and others (1995). Adding up these assumptions, we will consider the measured records to be sampled at approximately constant time intervals, thus treating the "depth series"as ordinary time series.

When analysing longer data series, one should account for firnification in the upper layers and the thinning of annual layers due to ice flow. Also, when analysing data from periods with changing precipitation patterns and rates special care should be taken.

Dynamical Decorrelation

We assume that the measured data series (the rows of a $N \times T$ matrix $\mathbf{X}$) are formed by a linear combination of an unknown set of underlying source signals $\mathbf{S}$ (also $N \times T$) as $\mathbf{X} = \mathbf{AS}$, where the $N \times N$ matrix $\mathbf{A}$ is called the mixing matrix. A common approach is to assume that the source signals form an orthogonal set and to arrange the series in the set with decreasing variance. This approach is known as Principal Component Analysis (PCA) and is widely used in time series analysis.

However, the assumption of orthogonality of the source signals is not necessarily viable, and some other assumption about the source signals must be chosen. Methods that assume statistically independent source signals are known as Independent Component Analysis (ICA), but the assumption of independence of the source signals does not alone contain enough information to determine $\mathbf{S}$ uniquely. Many choices of additional assumptions can be made (for a short summary, see e.g. Hyvärinen (1999)), but here the method of Molgedey and Schuster (1994), further developed by Hansen and others (2000), will be applied. Only a brief overview will be given here, and only the formalistic simple version of infinitely long series is considered. When performing the analysis on real finite data series, a number of mathematical technicalities must be considered, but the basic idea remains the same.

When performing PCA and most versions of ICA, no temporal information from the data is used; the data could be sampled at random times or be without temporal structure at all. In our case with ice core records, the source signals are expected to be e.g. the annual signal, other (quasi-)periodic signals and noise. These signals have very different periods and thereby autocorrelation functions, and the idea of Molgedey and Schuster is to make use of this fact. Given a linear mixture of the source signals, the method is able to separate the source signals by only assuming that one can choose a time lag τ, for which the values of the autocorrelation functions of the different source signals are clearly distinct.

Methodology

The correlation function between two infinitely long normalized discrete signals $x_{i,t}$ and $x_{j,t}$, $t \in \mathbb{Z}$, can be defined as $c_{ij}(\tau) = \lim_{T \to \infty} \left[\frac{1}{2T+1} \sum_{t=-T}^{T} x_{i,t} x_{j,t+\tau} \right]$. The correlation functions are all collected in the correlation matrix $\mathbf{C}_X(\tau)_{ij} = c_{ij}(\tau)$, which for $\tau = 0$ is the crosscorrelation matrix. We now define the quotient matrix $\mathbf{Q} = \mathbf{C}_X(\tau) \mathbf{C}_X(0)^{-1}$. Using the basic assumption that $\mathbf{X} = \mathbf{AS}$, we get that $\mathbf{C}_X(\tau) = \mathbf{X}\mathbf{X}^\top = \mathbf{A}\mathbf{C}_S(\tau)\mathbf{A}^\top$, where $\mathbf{C}_S$ is the correlation matrix for the source signals, and conclude that $\mathbf{Q} = \mathbf{A}\mathbf{C}_S(\tau)\mathbf{C}_S(0)^{-1}\mathbf{A}^{-1}$. The fact that the source signals were assumed to be independent results in $\mathbf{C}_S(\tau)$ being diagonal, such that the quotient matrix is given as

$$\mathbf{Q} = \mathbf{A} \begin{bmatrix} \frac{c_{11}(\tau)}{c_{11}(0)} & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \frac{c_{NN}(\tau)}{c_{NN}(0)} \end{bmatrix} \mathbf{A}^{-1} \tag{1}$$

This closely resembles an eigenvalue decomposition of $\mathbf{Q}$, and thus $\mathbf{A}$ (and thereby $\mathbf{S}$) can be computed by diagonalizing $\mathbf{Q}$, which itself can be readily calculated from the measured data. $\mathbf{A}$ is uniquely determined as long as the diagonal in the above equation consists of significantly different values. This means that if a choice of τ exists where the autocorrelation functions of the source signals are significantly different, the source signals can be uniquely determined. There is, however, no way of choosing τ a priori, and one has to find a suitable value by trial-and-error. A number of technical and numerical details must be taken into consideration. For a thorough discussion of these, and a proof of the fact that $\mathbf{Q}$ always has a real-valued eigenvalue decomposition, see Hansen and others (2000).

Performing Dynamical Decorrelation

A suitable set of data series from the same depth interval must be available for the analysis. Only series that are anticipated to contain the annual signal should be included. For evaluation purposes, we have chosen two different data sets which have been dated by annual layer counting. This dating has been verified with the help of known historical markers (Hammer and others, 1978).

For Site D the concentrations of Cl^-, NO_3^- and SO_4^{2-} and dust mass have been measured at 5 cm resolution, whereas $\delta^{18}O$ samples were cut in 8 samples/year according to a cutting scale based on an ice flow model (Steffensen, 1988). For this application the latter thus had to be resampled in order to be used together with the chemistry data. The resampling was done using statistical cubic interpolation. The sequence covers the years 1890-1910 in the depth interval 42.2 - 51.15 m. The relatively high annual accumulation rate of 0.37 m yr^{-1} (Clausen and others, 1988) at site D combined with the fact that the selected series are taken from a shallow depth give a sampling frequency of approx. 9 samples per year.

Also new data from NorthGRIP have been analysed. The NorthGRIP(1) ice core deep drilling started during the field season of 1996 9.85 m below surface and terminated at a depth of 1371 m during the 1997 field season (Dahl-Jensen and others, 2001). From the top down to a depth of 349.8 m a continuous series of 5 cm resolution was sampled for ion chromatographic analysis. This series corresponds to a time interval of approximately 2000 years. The chemistry samples were manually decontaminated and cut in the field using a microtome knife. They were stored frozen in Coulter accuvettes until measurement. Samples representing core depths from 116.05 to 221.65 m were measured by M. Hansson at the University of Stockholm, Sweden. The rest of the series was measured at the Dept. of Geophysics, University of Copenhagen.

The samples measured in Copenhagen were melted in the Coulter accuvettes and under a clean bench they were decanted into Dionex 5 ml polyvials with filter caps. The polyvials and filter caps were precleaned in > 18 MΩ water from an Elgastat UHP from Elga Ltd.

The measurements were done using a Dionex 500 micro bore ion chromatograph with two channels. A single standard solution was used for quantification of the contents of Li^+, Na^+, NH_4^+, K^+, Mg^{2+}, Ca^{2+}, F^-, MSA^-, Cl^-, NO_3^-, and SO_4^{2-} in the samples. In the standard solution the concentrations of Li^+ and MSA^- were 0.2 and 0.3 μequiv.kg^{-1}. The concentrations of the other ions in the standard solution were ranging from 1 to 7 μequiv.kg^{-1}. The chromatograms were integrated using Peaknet 4.30 software and the peak heights were used in the evaluation except when the peak area seemed to be more reproducible.

A sequence from the depth interval 28.1-34.65 m (corresponding to 24 years) was selected because of excellent data quality and because the firnification and thinning of layers have not proceeded far at this depth. The measured concentrations of 11 ions were used. $\delta^{18}O$ was left out of the analysis (1) because the signal is lost at greater depths due to diffusion in the relatively thin annual layers and (2) in order to make the data set significantly different from that of Site D.

With a sample size of 5 cm and with an annual accumulation rate of 0.195 m yr^{-1} (Dahl-Jensen and others, 2001), the sampling frequency is approx. 5.5 samples per year.

Data preprocessing

Missing data cannot be handled in this analysis, and must therefore be replaced by realistic values. The series used here only had isolated holes and a simple mean-of-neighbours scheme was applied to replace missing data and measurements that seemed improbable. The top of abnormal high peaks ascribed to volcanic eruptions was cut off such that the height of these peaks was reduced to approx. twice the average peak height. Prior to further analysis, the series were normalized to zero mean and unit variance in order to weigh the series equally. As a consequence of the normalization, *all* series on the plots are presented without dimensions.

The concentration of different chemical components peak at different times of the year, but as the objective of this work is to extract the annual signal from all the data series and to collect them into only one component, we decided to adjust the series such that they all peak simultaneously.

First one series is chosen relative to which the others are shifted. To synchronize another series with the fixed series, the time lag which maximizes the *absolute* value of the time lagged crosscorrelation function is computed and the other series is shifted accordingly. The crosscorrelation and the time lag computed here only relate to the synchronization of the series, and they have no relation to the dynamical decorrelation described in the previous sections. The maximum allowed time lag should be less than the number of samples per year to prevent the series from being shifted a full annual cycle. It should however still be large enough to allow synchronization of winter and summer peaks. After trying different values, we have settled on a maximum time lag of 3/4 of an average year. If the obtained crosscorrelation is negative, the sign of the adjusted series is changed such that the peaks are synchronized and have the same sign. The latter is done only to facilitate comparison between series.

For Site D, we have chosen to use the $\delta^{18}O$ profile as the fixed signal and synchronized and/or inverted the others relative to this.

For the NorthGRIP data, we have systematically tried each of the series as the fixed series, and we found that using Cl^-, K^+ or Na^+ worked best. This is probably due to the fact that they all exhibit

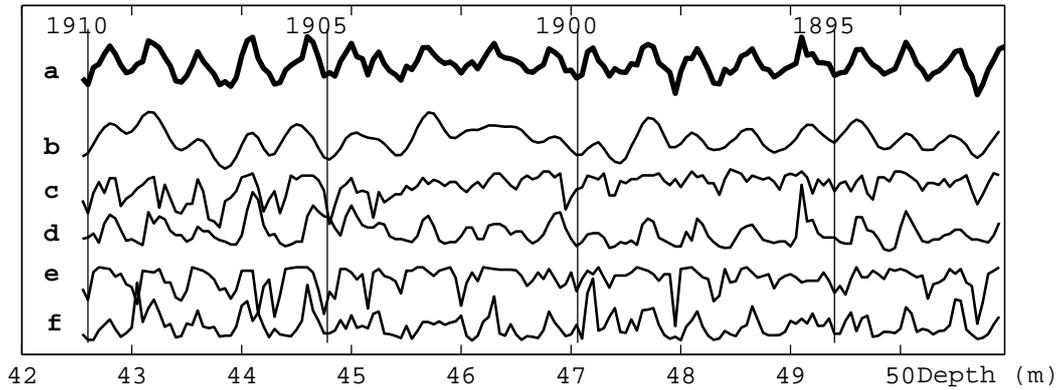Figur 1: Results of dynamical decorrelation of Site D data (a) with $\tau = 5$ compared to the raw but synchronized data series of (b) δ^{18}O, (c) SO_4^{2-}, (d) NO_3^-, (e) Cl^-, and (f) dust mass.

only one major annual peak which is rather broad. In this work, the importance of the synchronization has been investigated by performing the analysis both with and without the synchronizing of peaks. Although the difference is moderate, synchronizing the peaks gives somewhat better results.

Results

The dynamical decorrelation method was applied with values of the time shift τ ranging from 1 to 30, corresponding to a few annual cycles. This interval should be sufficiently large to reveal differences in the autocorrelation functions of the source series. The value of τ that produced the source series with the most distinct annual component was individually chosen.

For Site D, the method produced the most pronounced annual signal for $\tau = 5$. The source signal displaying the annual variation is in figure 1 displayed together with the 5 input data series. The vertical lines indicate the dating used by Steffensen (1988). It is seen that the produced annual signal somewhat resembles the NO_3^--series, but that it is significantly improved especially at year 1901, where the triple peak of the NO_3^--series is replaced by a broad single maximum and at year 1908, where a double peak is replaced by a single peak.

As stated in the previous section, the synchronization of the NorthGRIP data series was performed prior to the analysis with different choices of the fixed series. Depending on the choice of fixed series, different values of τ performed best. A total number of approx. 100 tests were performed with different combinations of fixed series and time lags τ. The 3 combinations that performed best are displayed in figure 2 together with the series that was held fixed. In general our results show that the annual signal produced by dynamical decorrelation shows the closest resemblance to the one of the input series that was held fixed.

It is seen that the produced signals exhibit clear annual variation with peaks of comparable size and that the 3 source series agree well on where the peaks should be situated, although the sizes of the peaks vary considerably across the series.

To test the robustness of the results, the analysis was performed on two overlapping series containing either the first 60% or the last 60% of the series from Site D respectively, using the same value of τ. The results agreed almost perfectly in the interval where the two parts overlapped. Also 10 samples were removed from each end of the series and the result was compared with the result from the analysis
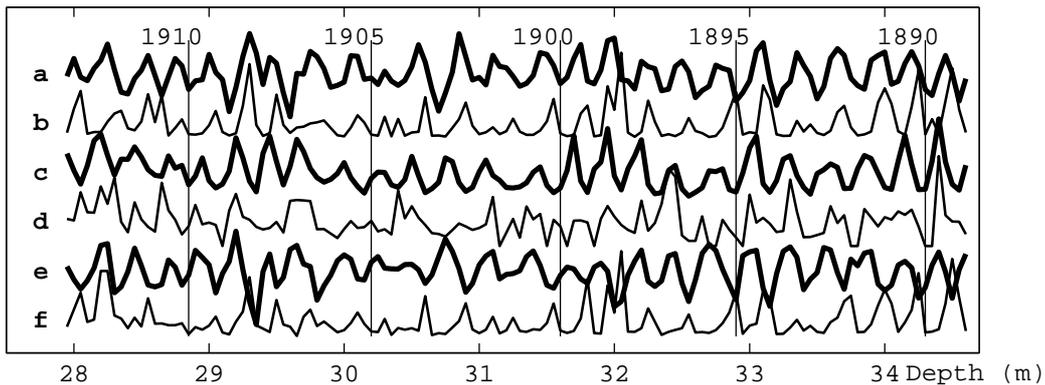
Figur 2: Examples of results from dynamical decorrelation (DD) of NorthGRIP data compared to the raw but synchronized data series. (a) DD with Na^+ fixed and $\tau = 3$. (b) Raw Na^+. (c) DD with K^+ fixed and $\tau = 5$. (d) Raw K^+. (e) DD with Cl^- fixed and $\tau = 2$. (f) Raw Cl^-.

with the full data set. Again almost perfect agreement was achieved, and from these two tests we thus conclude that dynamical decorrelation produces results that are largely indifferent to even considerable changes in input data selection.

The same method has also been applied to a different sequence of NorthGRIP data from a depth of approx 110 m. The results were not convincing, indicating that the method demands data with a resolution of at least 4-5 samples per year.

A measure of performance

In order to evaluate the performance of the method, a simple "dating"scheme has been set up. For a given series, the number of peaks rising above a certain threshold is counted. The number of peaks should in the optimal case be equal to the number of years for a wide variety of threshold levels, such that the exact value of the threshold level is not critical to the result. Applying the scheme to count the number of peaks of e.g. $\sin(2\pi t), t \in [0, 10]$, the number of peaks will be 10 for threshold levels $l \in]-1, 1[$, and zero for all other l. If the number of peaks p is plotted against l, it is from this graph alone easily verified that the number of peaks must be 10, and that the result does not depend on the exact choice of l.

In general the series' "countability"will be evaluated using this counting method. A series will be considered easy to count if the (l, p)-graph exhibits a plateau of (almost) constant p over a wide range of l, while the absence of such a plateau indicates that the series cannot be dated using this simple counting method.

This method of counting annual peaks is crude, and should therefore not necessarily yield perfect results. On the other hand it is our assumption that if the method of dynamical decorrelation improves the countability of the series when using such a simple counting method, it is most likely that the produced source series exhibit a stronger annual signal than the original data series and will be easier to date using almost *any* counting method.

The method counts peaks of any size as long a they cross the threshold level. A peak consisting of only one measurement that just barely exceeds the threshold level is thus counted as a peak. It is thus anticipated that noisy data series will have too many peaks when counted with this counting method.
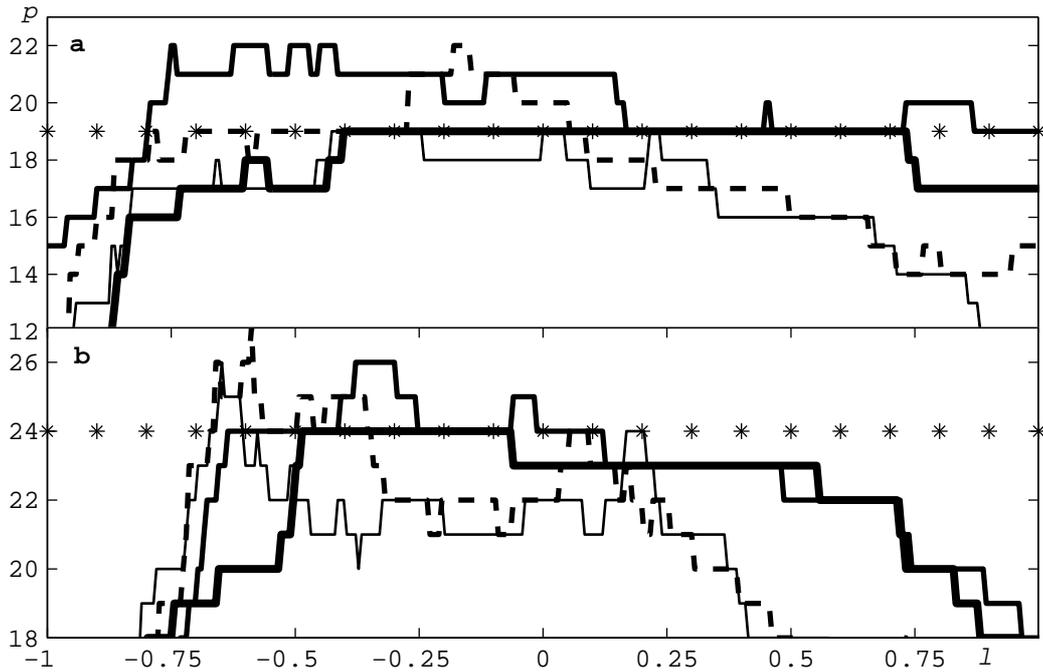
Figur 3: The number of peaks p versus threshold level l for (a) Site D and (b) NorthGRIP. Results from the dynamical decorrelation filtered with a low-pass filter (bold line), DD unfiltered (medium line), raw filtered (thin line) and raw unfiltered (dashed line). The number of years obtained by manual counting (19 for Site D and 24 for NorthGRIP) is marked with asterisks.

To overcome this problem, we have tried to run the data through a low-pass filter before counting the number of peaks. A 5th order Butterworth digital low-pass filter with a cutoff frequency of 4.5 yr^{-1} provided good results.

The counting method has been used on both filtered and unfiltered results from the dynamical decorrelation and on raw data for comparison. The produced (l, p)-plots show considerable improvement in the countability of the number of peaks (figure 3).

Especially for Site D (figure 3a) the results are promising. When the counting method is applied on the filtered annual component from the dynamical decorrelation, the same number of years as obtained by manual layer counting is found for a wide interval of threshold levels $l \in [-0.42, 0.71]$. For the unfiltered signal, the number of peaks is less well-defined. The number of peaks in the raw data is counted in the same way but provides no usable results for dating, nor does the filtered but otherwise raw data.

For NorthGRIP, only the (l, p)-plot for the case where Na^+ is fixed and $\tau = 3$ is shown (figure 3b). Much like the Site D results, only the filtered dynamical decorrelation series provides good countability. With $l \in [-0.50, -0.07]$ the result agrees with the 24 years determined by manual counting, while 23 years are found when $l \in [-0.07, 0.55]$. This arises from the fact that the peak representing the year 1904 (see figure 2) is less pronounced than the other annual peaks and thus only is counted if the threshold level is lower than -0.07. Features like the "two-level-plateau" in figure 3b can thus be used to identify doubtful annual layers.

7

Discussion

When comparing the series in figure 1 and 2, it is seen that the method of dynamical decorrelation is able to extract a signal that exhibits a more pronounced annual variation than the input series themselves. It is demonstrated that the method cures the series of some problems of double or triple maxima and that the peaks in the produced signal have a relatively well-defined height. This facilitates automatic counting of the peaks by a simple method as demonstrated in figure 3.

The results produced are sensible to the choice of τ and to some extent to which series is held fixed, but is otherwise fairly robust. By comparing the 3 series of figure 2 that are the results of dynamical decorrelation, one sees that although the produced series vary somewhat for different choices of fixed series and τ, the overall appearance is similar.

Dynamical decorrelation is thus able to improve data series significantly prior to e.g. dating. Nevertheless the benefit in practical applications is at this moment limited by the need of choosing how to synchronize the data, choosing a suitable value for τ, and selecting an appropriate cut-off frequency for the low-pass filter. Future investigations will show whether it is possible to make choices that will work well for longer data series from the same ice core. It is hoped that especially values of τ can be found that will work on longer records from the same climatic period.

Acknowledgments

References

Clausen and others, 1988: Clausen H.B., N.S. Gundestrup, and S.J. Johnsen 1988. Glaciological investigations in the Crête area, central Greenland: a search for a new deep-drilling site. Ann. Glaciol. 10, 10-15.

Dahl-Jensen and others, 2001: Dahl-Jensen D, N.S. Gundestrup, and others 2001. The NorthGRIP deep drilling program. This volume : 35A151.

Hammer and others, 1978: Hammer, C.U., H.B. Clausen, W. Dansgaard, N.S. Gundestrup, S.J. Johnsen, and N. Reeh 1978. Dating of Greenland ice cores by flow models, isotopes, volcanic debris, and continental dust. J. Glaciol. 20, 3-26.

Hansen and others, 2000: Hansen, L.K., J. Larsen, and T. Kolenda 2000. On Independent Component Analysis for Multimedia Signals. In Guan, L., S.Y. Kung and J. Larsen (eds.) Multimedia Image and Video Processing. Boca Ratton, FL, CRC Press, 175-199.

Hyvärinen, 1999: Hyvärinen, A. 1999. Survey on Independent Component Analysis. Neural Computing Surveys, 2, 94-128.

Molgedey and Schuster, 1994: Molgedey, L., and H. Schuster 1994. Separation of Independent Signals using Time-Delayed Correlations. Phys. Rev. Let. 72(23), 3634-3637.

Shuman and others, 1995: Shuman, C.A., R.B. Alley, S. Anandakrishnan, J.W.C. White, P.M. Grootes, and C.R. Stearns 1995. Temperature and Accumulation at the Greenland Summit: Comparison of high-resolution isotope profiles and passive microwave brightness temperature trends. J. Geophys. Res. 100(D5), 9165-9177.

Steffensen, 1988: Steffensen, J.P. 1988. Analysis of the seasonal variation in Dust, Cl^-, NO_3^-, and SO_4^{2-} in two central Greenland firn cores. Ann. Glaciol. 10, 171-177.