JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 103, NO. D5, PAGES 5649–5659, MARCH 20, 1998

New method for identification of sources for chemical time series and its application to the Greenland Ice Sheet Project ice core record

Peter D. Ditlevsen and Nigel D. Marsh

Department of Geophysics, Niels Bohr Institute, University of Copenhagen, Denmark

Abstract. Point measurements of chemical tracers transported via the atmosphere contain information regarding origins, atmospheric mixing, scavenging, and in situ chemical processes. A new technique for accessing this information is presented. The method is based on the physical constraints that are naturally imposed on the data. In order to evaluate the performance of this new technique, we have applied it to artificially generated data sets. Furthermore, the method is used to analyze eight chemical time series obtained from ice core data. Three distinct source vectors dominate the composition throughout the last 41 kyr period. By comparing ratios among the eight chemical species within each source vector, it is possible to show that they describe three source types: oceanic, continental, and biochemical land sources.

1. Introduction

It is often interesting to apportion tracers in the atmosphere, measured at one or more stations, into distinct sources. In particular, identifying contributions from sources that have a fixed mass balance among different chemical species can provide useful information related to pollution or atmospheric transport properties. However, this information is partly lost through mixing and chemical reactions en route to the receptor points where tracers are measured.

Techniques for retrieving as much information as possible on source identification are called receptor models. Two main approaches can be taken: chemical mass balance (CMB) models that assume identification of sources and in which the fixed mass balances among the different chemical species is known a priori, and multivariate or factor analysis that attempts to both identify sources and their respective loadings from the collected data alone [*Henry et al.*, 1984]. An intermediate version between these two can be generated by utilizing a priori knowledge in any realistic case.

A major problem for factor analysis models is that the obvious physical constraints, such as nonnegative source components and loadings, are not automatically satisfied. Singular value decomposition (SVD) and principle component analysis (PCA) are based on second-order statistics of the data, thus a straightforward interpretation of the data in terms of source vectors is difficult. The problem of finding source compositions is often underdetermined [*Henry*, 1987]. It is thus important when validating the results to be able to estimate how much information is obtainable from the data. In other words one should avoid unique answers to ill-posed questions.

In section 2 we present a very simple model based on the physical constraints of the data. The model permits any hybrid splitting between a priori and a posteriori additional information and constraints. In section 3 we argue that factor analysis models are unsuitable for recovering identifiable sources, and in section 4 we show how this new model performs on an artificial data set. Finally, in section 5 we apply the technique to ice core data obtained from the Greenland Ice Sheet Project (GISP) ice core [*Maycwski et al.*, 1994; *Mayewski et al.*, 1993].

2. Model

We assume that the data, $c^\alpha(t), \alpha = 1,..,\nu$, comprises ν time series of chemical tracers obtained at one station. Index t could trivially be extended to a spatial index or any combination of a spatial index and a temporal index. For simplification and to keep notation to a minimum, we assume t to be time only and we measure at one station. In order to subtract the desired information from the data, we make the following assumptions: (1) the chemical tracers originate from n different source types and (2) each of these sources has a constant relative chemical abundance over the whole time series. This means that chemical activity in the atmosphere, changing the mass balance within a given source, is assumed to result in a constant change (independent of temperature, etc.) of the mass balance. This

Paper number 97JD03623.
0148-0227/98/97JD-03623$09.00

process is represented by the source vectors, $x_i^\alpha, i = 1, n$ and $\alpha = 1, \nu$, with the relative loading from each source vector x_i^α given by coefficients $a_i(t)$. The measured concentration $c^\alpha(t)$ for a given species α is then

$$c^\alpha(t) = \sum_i a_i(t) x_i^\alpha + \eta^\alpha(t). \quad (1)$$

Here $\eta^\alpha(t)$ is a noise term accounting for contributions not originating from any of the n predominant sources. To minimize complications, we assume measuring errors to be negligible and uncorrelated between the different species.

To retain information regarding the dominant source types, we must have $n \leq \nu$, regardless of the length of the data record (or number of stations). Values of the individual species concentrations, $c^\alpha(t), \alpha = 1, \nu$, can be thought of as lying within a ray spanned by n source vectors x_i^α in ν-dimensional space. The analysis now takes two steps. First, the number of predominant sources are found, and then an error minimization is performed to determine the source vector components. The number of predominant source types n can be resolved from an embedding analysis.

The *Grassberger and Procaccia* [1983] method assumes that for points situated evenly on an n-dimensiona manifold within a ν-dimensional space, the number of points $N(r)$ within a sphere of radius r scales as $N(r) \sim r^n$. To prevent the over domination of one species due to large absolute values, all species are normalized with respect to their individual series means. Geometrically, this is just a rescaling of the axis in ν-dimensional space. This assumes the various species are equally important in determining the sources. The embedding analysis technique has been used previously in the context of climate data in an attempt to determine a climate attractor from δ^{18}O records [*Maasch*, 1989]. It was observed by *Grassberger* [1986] that spurious, low-dimensional estimates can arise from short time series and that thousands of independent points in time are required to obtain a reliable estimate of attractor dimension. This is especially true when attempting to characterize a potentially non integer dimensional attractor in a phase space of N degrees of freedom, where N is large.

The problem is twofold: (1) the number of points needed for filling the space grows as 2^N and (2) a low dimensional attractor could lie on some complicated folded manifold such that $N(r) \sim r^d$ holds only over short distances, and many more points are then required to obtain the correct dimension. However, in this analysis we want to characterize a much simpler geometric object: the data cluster within the ν-dimensional species space. In our experience this can be performed reliably with a few hundred points n of each species. In fact, in the δ^{18}O study the embedding analysis was performed on a single time series with as little as $N = 180$ points, and by assuming ergodicity, embedding attractors were constructed using *Ruelle*'s [1981] method of delays. With $\nu = 8$ and $n = 200$ we have an equivalent of $N = 1600$, points which would thus appear sufficient. By combining the embedding analysis with an a posteriori justification for model selection, penalizing for the number of degrees of freedom, we have confidence in our dimensional analysis.

When ν is found, the error minimization is performed, such that the angles between the source vectors and the number of points falling outside the ray spanned by the vectors are minimized, combined with an optimal best fit to the actual data. Points falling outside of the ray result in negative loads from one or more sources. The geometry is schematically shown for two dimensions in Figure 1, where it is obvious that the data are equally well spanned by any set of vectors spanning a larger angle, including the one found from the minimization. In this respect the problem is underdetermined, but by minimizing with respect to the angles in the ray, we assume that there are enough data in the "true" phase space, such that the "minimum angle" vectors are the most likely. With a more careful statistical analysis and uniform distribution, one would assume that the most likely angles between the source vectors are larger by an $O(1/N)$ factor. So with 100 data points the error is of the order of 1%. The minimization of the angles spanning the ray is the major difference between this method and traditional factor analysis models.

3. Comparison With Factor Analysis Models

Conventional factor analysis models are based on the correlation structure (a mean-centered statistic) present within the data. The PCA base vectors determined in this way are the orthogonal eigenvectors of the covariance matrix of the data (in meteorology often referred to as empirical orthogonal functions (EOFs)). These vectors (principle components (PCs)) will, in general, have negative components and in this form cannot be interpreted as source vectors for the following reason.

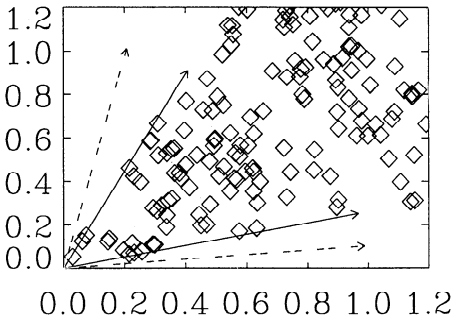

Figure 1. The data points lie within a ray spanned by the source vectors. We determine the vectors spanning the ray (solid lines), but the problem is obviously underdetermined since the data could be equally well spanned by the dashed vectors. The major assumption is that the true source source vectors collapse around the data such that the angle between them is minimized (solid vectors).

If the PCs are to be interpreted as source vectors, the mean itself, which is just the average contribution from the combined sources, must also be expressed in terms of the PCs. However, this cannot be physically sound if the PCs have negative components, as we end up with the unphysical requirement that source areas are "sucking out" chemical species from the system. Only after the mean vector is added back to the dominant PCs is it possible, though, by no means, guaranteed, for them to regain the requirement of positivity. Furthermore, projecting the data onto each of the principle components generates a respective set of coefficient time series. Since the PCs are eigenvectors of the covariance matrix, the time series are, by construction, uncorrelated (temporal orthogonality), so that negative coefficients must exist at some times. Each coefficient time series effectively describes the contribution from its related PC with respect to the mean vector. Even if the PCs are all positive as a result of adding the mean vector, the coefficient time series will still contain negative values owing to the orthogonality requirement, and we still have the unphysical problem of sucking out chemical species.

Rotation techniques, e.g., varimax [*Richman*, 1986], have been constructed in order to avoid such problems. They require that the dominant PCs are known prior to the rotation, which itself is a nontrivial problem. However, in general, the actual rotation process is not based on physical constraints of the system in question but rather, is a somewhat arbitrary numerical method designed to replace a "rotation by eye" in order to emphasize data clusters, if they exist. Relaxing the orthogonality requirement and rotating the few dominant PCs within the full space is essentially a noise reduction technique. In the current context of source attribution this could be improved by including further optimization demands based on physical constraints, similar to our model discussed in section 2. However, as discussed in section 4, the method for selecting the number of dominant PCs is marginal, and we argue that the technique of embedding analysis is more robust.

4. Artificial Data

In order to understand the properties of the model, artificial species data are generated with known source vectors x_i^α and coefficient time series c^α. Each source vector is selected to have eight components that are chosen randomly and then normalized. The coefficient time series are also generated randomly, each containing 200 points. Combining the artificial source vectors and coefficient time series as in (1) produces eight artificial species, each containing data for 200 consecutive time intervals. As long as the number of sources are less than the number of species, any number can be generated for this purpose. It is with this species data that we assume no prior knowledge and apply the model in an attempt to reproduce the source vectors and their respective coefficient time series. The choice of 200 data points allows for a direct comparison of the statistical significance of the method with the real data discussed in section 5.

An embedding analysis of the species data reveals the dimensions of the subspace containing the data cluster spanned by the source vectors. Figure 2a shows the scaling obtained from eight artificial species data sets generated from $d = 2, 3, 4$, and 5 source vectors, respectively. As seen from plotting the function $N(r) \sim r^d$ on the log-log plot, the dimensions of the data cluster are reproduced over a good range of scales. Adding 5% noise (equation (1)) has the effect of "smearing out" the data cluster, which is seen from the breakdown in scaling at small scales from a dimension equal to the number of generating source vectors to the eight dimensions of the full species space (Figure 2b). In other words, at scales $r \leq 0.2$, measured from any data point, the noise will dominate and the points surrounding the central data point will be evenly distributed in eight-dimensional space. Thus the scaling $N(r) \sim r^8$ for $r < 0.2$ is obtained. However, for $r \gg 0.2$ the noise is negligible and a limited region with the scaling $N(r) \sim r^d$ exists. From the scaling in this region the number of sources can be determined. Deviations from the linear slope as r increases are caused by boundary effects; eventually, all points will be inside a sphere when the radius is larger than the maximum distance between two points in the cluster. For large data samples, intermediate scaling regimes $< d$ can arise from points being situated within a narrow ray, thus at larger scales the data appear to lie in a lower-dimensional subspace. This feature is well captured in the analysis of ice core data in section 5 (see Figure 6a). From the examples in Figure 2b we estimate that 200 points is marginal for an embedding analysis with a 5% noise level.

The dimension of the subspace spanned by the source vectors can be determined as well by calculating how many dominating eigenvalues the correlation matrix has, i.e., the PCs. This is usually achieved by calculating the amount of relative signal variance captured in each PC and arbitrarily choosing a suitable cutoff so that the main features of the signal are retained. However, in the current context the cutoff for significant eigenvalues corresponding to source vectors is not always as obvious as the significant vectors obtained from the log-log slope of an embedding analysis. This could be a result of the fact that elements of the correlation matrix are squared differences (second-moment statistics), while the distances between points in the embedding analysis are first order quantities (first-order statistic). The two methods of analysis thus put a different weight on outlying points.

Having approximately found the dimensionality of the data clusters' subspace from the embedding analysis, it is possible to search for the source vectors themselves and their coefficient time series by performing the minimization with a prescribed number of sources. A minimum requirement for this procedure is that it must

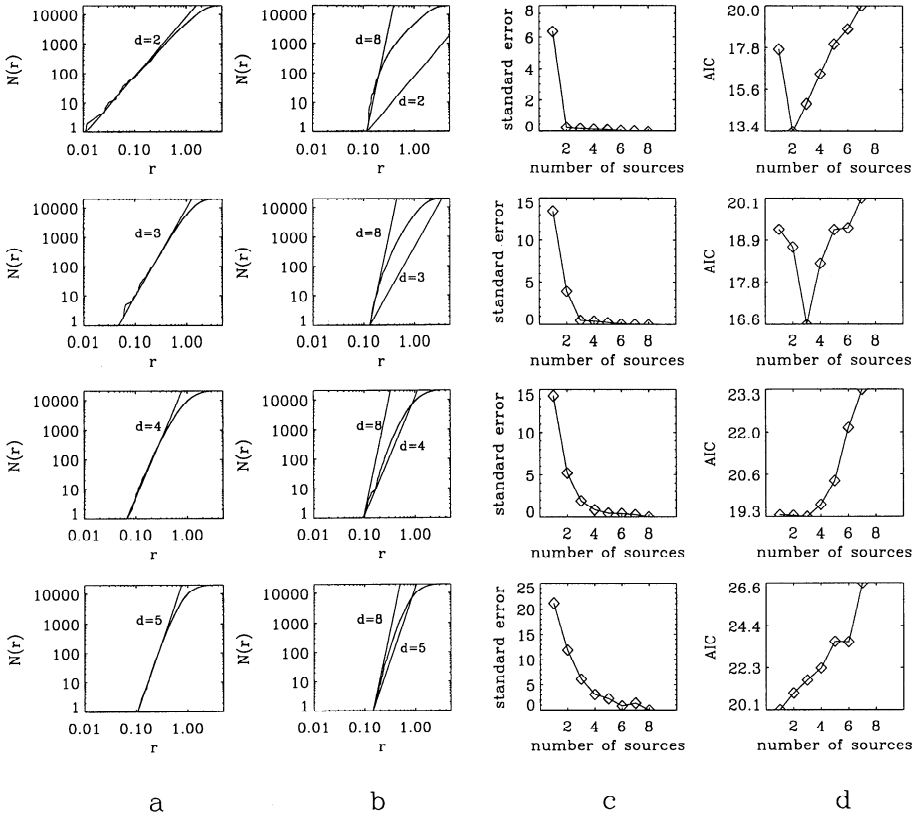

Figure 2. (a) Embedding analysis of the eight artificial species data, where the values are normalized to the same mean. The average number of points within a sphere of radius r, $N(r)$ is plotted against r. On a log-log scale the relation $N(r) \sim r^d$ is a straight line with slope d. This would correspond to the points being situated on a d-dimensional manifold embedded in the eight-dimensional space. From top to bottom the slopes of the overplotted straight lines are $d = 2$, 3, 4, and 5. (b) Embedding analysis performed on the same data as in Figure 2a but with 5% noise added. The noise blurs the data clusters' sub-space into the full eight-dimensional species space at small scales. The additional straight line has a slope of $d = 8$ in all cases. (c) The standard error and (d) Akaike information criteria (AIC) estimates for fitting the artificial species data of Figure 2b with a range of source vectors. This suggests that even with noise, it is possible to resolve the correct number of predominant sources.

produce a reconstructed set of species data that is a good fit to the original data. With no noise present, the reconstructed species data will be perfect, as long as the reconstructed species vectors span the same subspace as the original one. For illustrative purposes we perform a minimization on the species data that produce the scaling dimension, $d = 3$, in Figure 2b. Figure 3 shows that even with 5% noise added, the reconstructed data give a good fit.

Owing to the uncertainty in the number of vectors required to span the data subspace, an independent check can be performed by repeating the optimization for a range of source vectors $n = 1, 2, .., 8$. A best fit model can be found from minimizing the mean square error or standard error. For each value n the standard error is defined as the sum of the integrals of differences squared between the generated data and the artificial data (normalized). Figure 2c shows the standard error analysis performed on the artificial data for source vectors that generate the scaling observed in Figure 2b. It is clear that the error decreases as the number of source vectors are increased, reaching a plateau where the optimal number of vectors are used. Increasing the number of source vectors thereafter has a minimal effect on the error, with the additional vectors capturing the uncorrelated residual noise. To check the significance of this observation, one should include a penalty on the number of independently adjusted parameters. This can be achieved using the Akaike information criteria (AIC) [*Tong*, 1990], where the minimum of the function

$$\mathrm{AIC}(k) = -2\log(\text{maximum likelihood}) + 2k \qquad (2)$$

will occur when the most likely number of source vectors are used. Here the maximum likelihood is defined as the inverse of the standard error, and the number of independent parameters, $k = (\nu - 1) + (n - 1)$, corresponds to a model of n source vectors describing ν

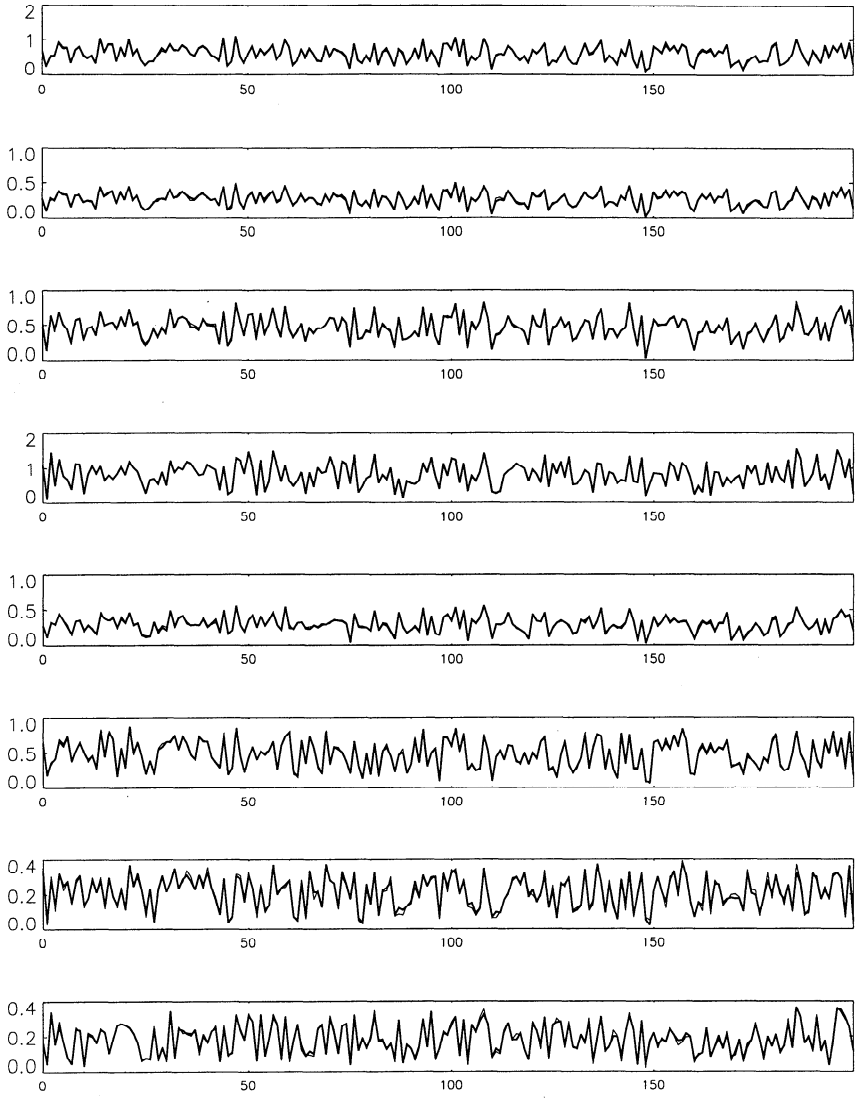

Figure 3. The eight artificially generated species data from three source vectors with 5% noise added (thick lines). Overplotted are the reconstructed data obtained after the minimization has been performed (thin lines).

species. Figure 2d shows AIC estimates that reveal a clear minimum for the correct number of vectors assuming $n = 2, 3$. However, it is marginal for $n \geq 4$ since this analysis was performed on only 200 points; with a greater number of points we would achieve more significant results for $n \geq 4$.

With knowledge of the number of sources, the source vectors can be determined by a best fit to the data, combined with collapsing the vectors such that they just touch the ray containing the data (see Figure 1). This amounts to minimizing the angles between the vectors under conditions of positive coefficients. The source vectors and corresponding coefficient time series obtained from the data that generated the scaling properties of Figure 2b produce a remarkably good fit. For two source vectors (not shown) the fit is near perfect, for three source vectors (Figure 4) and four source vectors (Figure 5) the fit is very good, and even with five source vectors the fit is reasonable. Any disagreement between original and reconstructed vectors arises from the fact that the minimization selects the source vectors that optimize a best fit to the data cluster. Only if a data point falls exactly on each of the axes of the original source vectors will they be recovered completely. The problem will, in general, be underdetermined, as seen from Figure 1. The data plotted could as well be spanned by the two dashed vectors as by the two solid vectors in Figure 1. There may be physical explanations for the true subspace not being filled completely, resulting from inherent correlations between the pure source vectors because of geographic location and transportation to the deposition site. This would result in the reconstructed source vectors possessing a projection from two or more of the pure source vectors. In effect, the vectors have been rotated into their true subspace such that only a part of it is spanned, a result of demanding the a priori

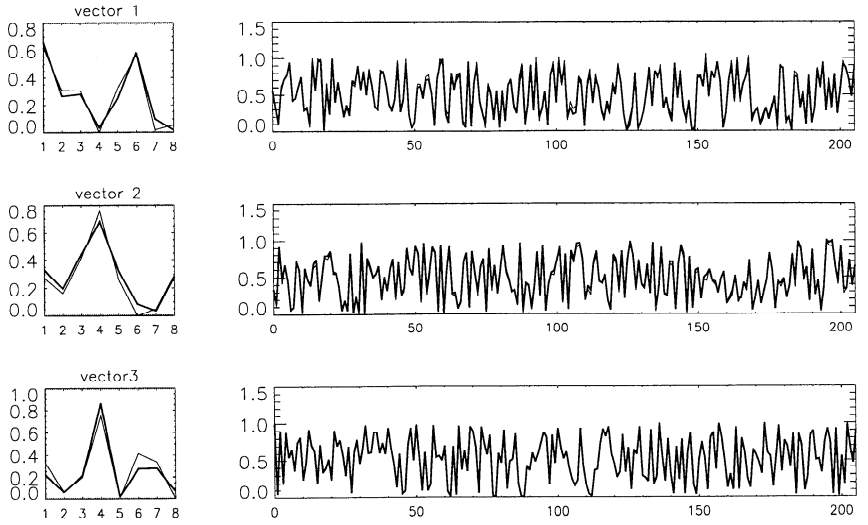

Figure 4. Reconstructions of source vectors and their respective coefficient time series performed on the artificial species data of Figure 3, showing (left) the original source vectors (thick lines) and the reconstructed source vectors (thin lines) and (right) their respective original coefficient time series (thick lines) and reconstructions (thin lines).

constraint of nonnegative components. Such a posteriori information can only be accessed when using real data and comparing the chemical mass balances within each reconstructed source vector with those from known pure sources, i.e., oceanic or continent [*Holland*, 1978]. The following example is of this type of case where a posteriori validation is necessary.

5. GISP Ice Core Data

Concentrations of eight different ions have been measured from the GISP ice core at Greenland summit [*Mayewski et al.*, 1994; *Mayewski et al.*, 1993]. Substantial changes in the concentrations of the ions are observed throughout the last glaciation. Changes in the generic source areas of the chemical species depend on; sea level changes [*MacAyeal*, 1993; *Bond and Lotti*, 1995], changes in among others vegetation (desertification), and glacial extent [*Hammer et al.*, 1985], all of which have typical timescales of several hundred years. Variations in atmospheric transport related to individual storms, on the other hand, take place at much faster timescales. It is not possible from point measurements

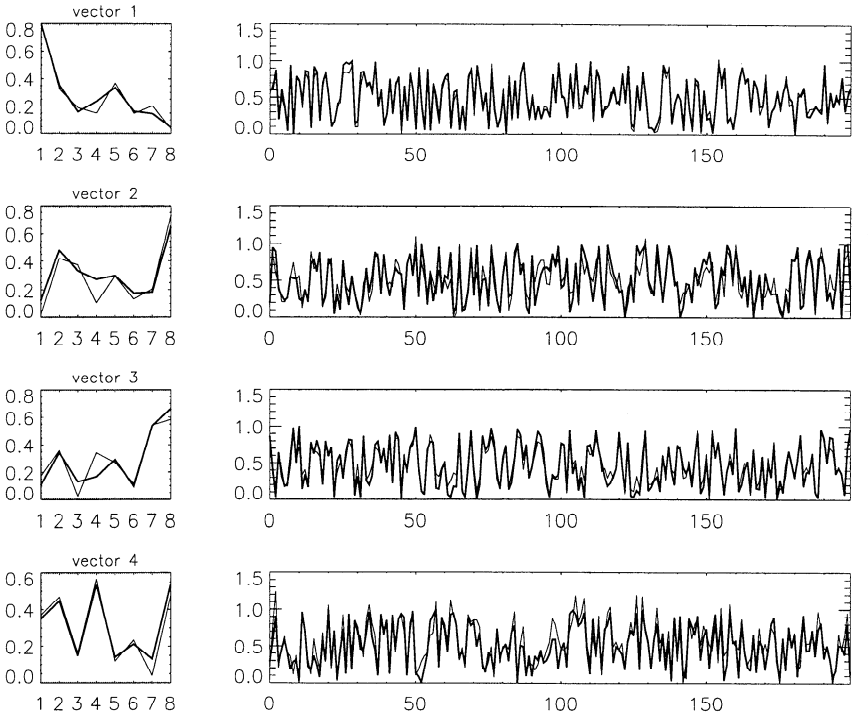

Figure 5. Same as Figure 4 but with four source vectors and coefficient time series.

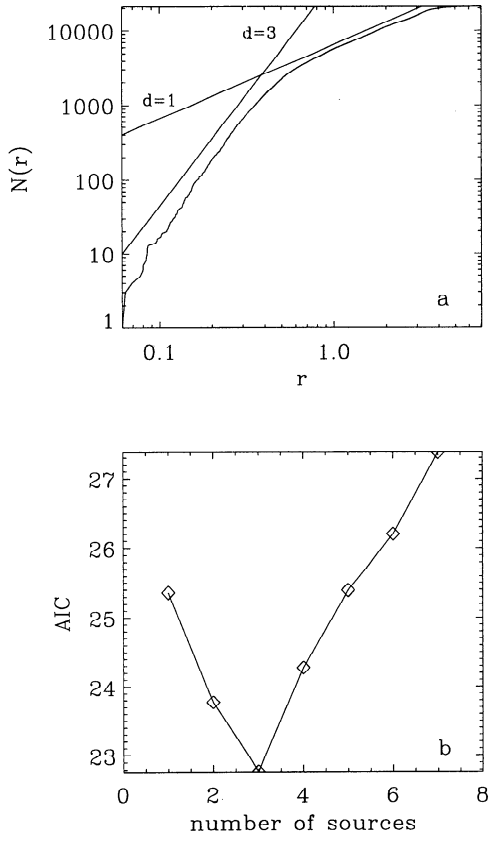

Figure 6. (a) Embedding analysis of the eight chemical time series obtained from the Greenland Ice Sheet Project ice core, (Figure 2 caption for more details). At large values of r the scaling breaks from 3 to 1 because the data are situated in a narrow three-dimensional ray. (b) AIC estimates in fitting the eight chemical species, depending on the number of source vectors assumed, strongly suggesting there are three dominant source types describing the chemical species.

to distinguish between effects resulting from changes in source area strength and transport processes [*Bales and Wolff*, 1995]. However, it is believed that variations at time scales greater than 200 years are dominated by source area changes and large-scale changes in the mean atmospheric flow pattern [*Marsh and Ditlevsen*, 1997a]. Thus we analyze the 200 years mean of these signals, providing 206 points for each of the eight measured ions; Cl^- Na^+, K^+, Ca^{2+}, Mg^{2+}, SO_4^{2-}, NH_4^+, and NO_3^- , covering the last 41 kyr. In order to allow for effects resulting from changes in precipitation levels, the ion concentrations are converted to fluxes by using the empirical relationship between snow accumulation and $\delta^{18}O$; accumulation $= 0.23 \exp[0.14(\delta - \delta_0)]$ meters ice per year [*Johnsen et al.*, 1989]. The flux is then given as flux = accumulation $\times$ concentration since wet deposition dominates [*Wolff*, 1993]. Contributions resulting from volcanic activity [*Clausen et al.*, 1993], aging snow [*Silvente and Legrand*, 1993], extraterrestrial material, etc., occur at timescales much shorter than 200 years and with magnitudes that are smoothed out into the noise level at this resolution.

Figure 6a shows the result of the embedding analysis performed on the species flux data, where the line with a slope of $d = 3$ indicates that there are three predominant sources. This analysis obtains the same number of predominant sources as does the EOF analysis of *Mayewski et al* [1993] who claimed that the first three EOFs account for 82% of the variance in the signal. The error minimization was then performed assuming three sources; the resulting vectors are displayed in Table 1, and the normalized vectors are shown graphically in Figure 7.

The number of data points (206) in the time series is arguably marginal for performing an embedding analysis. This is seen from the limited scaling regime in Figure 6a. In order to substantiate the assumption that there are three base vectors, we identify the best fit model from the standard error assuming $n = 1, 2.., 8$ different sources and penalizing against the number of degrees of freedom as described in section 4. The AIC as a function of number of source types n is displayed in Figure 6b. This confirms that three source vectors pro-

Table 1. Source Vectors Obtained From Greenland Ice Sheet Project Data

Species	Vector 1 - Oceanic	Vector 2 - Continental	Vector 3 - Biochemical
Cl^-	4.03 (0.53)	1.86 (0.25)	0.87 (0.11)
Na^+	2.25 (0.55)	1.09 (0.26)	0.21 (0.05)
K^+	0.18 (0.33)	0.19 (0.36)	0.07 (0.13)
Ca^{2+}	0.49 (0.03)	10.52 (0.66)	0.14 (0.01)
Mg^{2+}	0.52 (0.30)	0.79 (0.45)	0.07 (0.04)
SO_4^{2-}	6.76 (0.37)	5.43 (0.30)	3.14 (0.17)
NH_4^+	0.06 (0.06)	0.11 (0.10)	0.91 (0.79)
NO_3^-	3.37 (0.28)	0.99 (0.08)	6.87 (0.56)

The three source vectors are the relative abundance of each species in absolute values with units 10^{-3} gm/m^{3}yr, where the abundance of each species is first normalized with respect to its mean before performing the analysis. The values in parentheses are the normalized vector components that are non-dimensional and represent the relative importance of the different species within a specific vector.

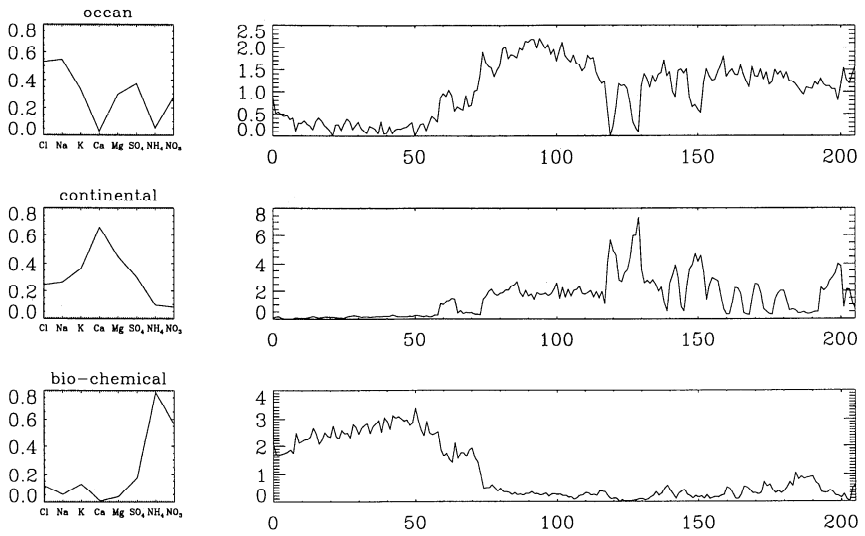

Figure 7. Reconstructed source vectors and coefficient time series from minimizing the fit to the eight normalized chemical species. (left) Dominant species in each source vector that generates the (right) corresponding coefficient time series. Table 1 displays the non normalized data values (see text).

vide a best fit minimum for the model. With only one source a vector similar to the dominating EOF found by *Mayewski et al* [1993] is obtained, while assuming two sources gives vectors similar to those corresponding to the ocean and continental sources in Table 1. With the number of sources $n > 3$, the three vectors of Table 1 are retained, with the remaining vectors describing noise. At a resolution of 200 years the noise term has buried within it any unresolved sources, including effects resulting from aging snow processes at timescales of days [*Silvente and Legrand*, 1993] and volcanic deposits at timescales of a few years [*Clausen et al.*, 1993] which will be smoothed out.

The eight generated species data, resulting from the reconstructed vectors and coefficient time series (Figure 7), are seen to match almost perfectly with the measured data in Figure 8. With confidence established for these reconstructed source vectors, their chemical mass balance must be analyzed before further interpretations can be made.

6. Interpretation and a Posteriori Validation

By considering possible source areas for the eight species and highlighting the dominant chemical components, we present a physical interpretation of the three vectors as follows. Vector 1 is strongly dominated by chlorine and sodium, with an absolute value ratio of 1.8 as observed in bulk sea water (BSW) [*Legrand and Delmas*, 1988; *Holland*, 1978], suggesting this vector could represent an oceanic source. Sulfate, also with significant dominance, has a much higher ratio than expected from BSW, which can be explained when considering that only a fraction of sulfate is released directly from the ocean into the atmosphere through ocean spray [*Legrand*, 1993], while a much larger oceanic contribution comes from the oxidation of dimethyl sulfide (DMS), a gas released by seawater that can account for up to two thirds of natural atmospheric sulfate [*Burgermeister et al.*, 1990] . However, potassium and magnesium, which are strong contributors to vector 1, possess an enrichment factor [*Legrand and Delmas*, 1988] twice that of BSW ratios. By repeating the analysis with one vector forced to contain the BSW ratios of chlorine, sodium, potassium, and magnesium [*Legrand and Delmas*, 1988], no observable change occurred in the remaining "free" components of all three vectors. Probably because of resolution, the analysis may have rotated the vector slightly. Calcium, and ammonium are of very little significance to this vector.

The strongest contributors to vector 2 are calcium, magnesium, potassium and sulfate. Calcium is a good indicator of continental dust [*Grip members*, 1993], resulting from weathered rock, which can also include magnesium, if silicate rich rocks are present ($MgSiO_3$), sulfate in the form of gypsum ($CaSO_4$), and potassium [*Hammer et al.*, 1985; *Wayne*, 1991]. The alkaline calcium aerosols are capable of absorbing gaseous sulfur compounds in the atmosphere [*Legrand and Delmas*, 1988], which is likely to have been an influential source of sulfate in Greenland ice during the last glacial with its massive increase of calcium content [*Grip members*, 1993] over present-day values. This vector is then dominated mainly by continental source types, although there are also substantial components of chlorine and sodium, with a ratio of 1.8 as expected from BSW. If our initial assumption is correct, i.e., that the relative chemical abundance in source areas is constant throughout the 41 kyr period, then the continental coefficient

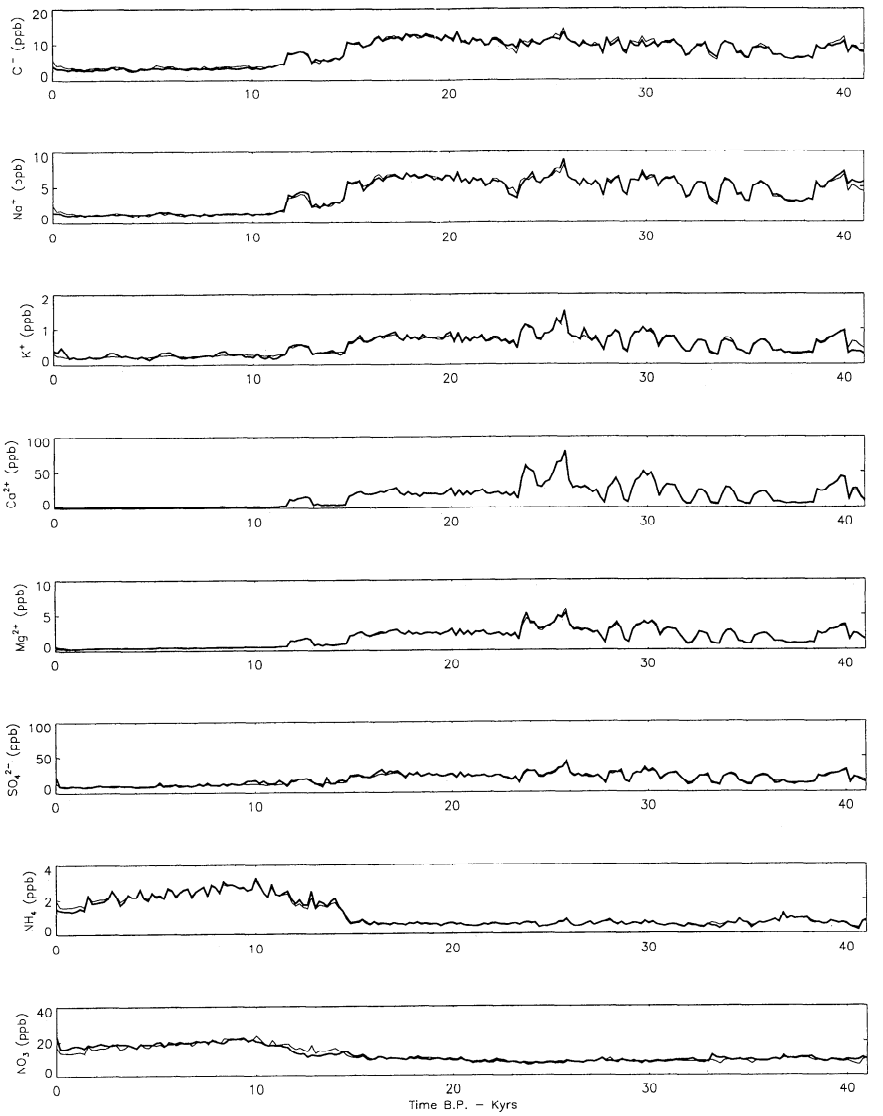

Figure 8. The 200 year averages of the chemical species concentrations through the ice core. Thick lines are measured data, while thin lines are the concentrations obtained from optimizing the three predominant source types.

time series contains signals of pure continental source areas and a correlated, fixed ratio contribution from ocean source components.

Ammonium is by far the strongest contributor to vector 3, and it is well known that its sources are the result of the decomposition of organic waste and aerobic biology in soils [*Wayne*, 1991]. Other than nitrate, all species in this vector are of insignificant importance, which is interpreted as an indicator of biochemical source types.

The origins of nitrate in the Greenland ice are still uncertain, at present. However, the major sources are probably tropospheric lightning, stratospheric sources, and soil exhalation [*Wolff*, 1993]. Soil exhalation naturally occurs over vegetated regions', thus an element of nitrate fluxes would be expected to correlate with ammonium. Comparison between the species time series of nitrate and ammonium over the last 41 kyr reveals similarities in agreement with the strong presence of nitrate in vector 3 [*Wolff*, 1993]. The more dominant sources are from lightning, which occurs predominantly over land and in the tropics [*Orville and Hendersen*, 1986] and the photochemical oxidation of N_2O in the stratosphere. Both these processes are likely to contribute to the production of a more global background source, though the low levels of nitrate in the continental vector might then seem contradictory. However, only three vectors are considered, none of which captures trends in the global background, which partially explains the relatively poor reconstruction of nitrate in Figure 7. Contributions from the global sources of nitrate may then artificially spill into any of the three vectors without affecting the ratios between the other chemical components. To confirm this, the analysis was repeated on seven species, excluding nitrate, which made only a minimal change to the distributions of the

other chemical species in the source vectors. It may well be that an analysis with higher-resolution data would reveal an additional, fourth vector describing a background signal.

Following the previous discussion, the three source vectors are described as oceanic, continental, and biochemical, respectively. However, as mentioned earlier, it is not possible to directly separate source area and transport effects, thus the resulting vectors are unlikely to represent pure sources but, instead, will be mixed with additional contributions due to transporting air parcels with an "unclean" background. The corresponding coefficient time series $a_i(t)$ (Figure 7), will then depend on contributions from the loading of the dominant pure source and the strength of contamination from global sources and/or other much weaker sources that are not revealed by the analysis at this resolution. Correlations exist between these series with correlation coefficients of

$$\begin{aligned} \text{oceanic-continental} &= 0.38 \\ \text{oceanic-biochemical} &= -0.66 \\ \text{continental-biochemical} &= -0.73 \end{aligned}$$

These time series are then of a different nature than those obtained from an EOF analysis, which, by construction, are uncorrelated [*Preisendorfer*, 1988].

The continental vector, in particular, is the most contaminated. Though dominated by terrestrial dust, it contains substantial chemical components associated with seawater and possessing ratios similar to those seen in the ocean vector. This could be the result of two processes: (1) air transporting terrestrial dust takes a path or mixes with air that passes over an ocean source area or (2) the dust originates from newly exposed continental shelves that are rich in sea salt [*Marsh and Ditlevsen*, 1997b]. As a result, the ocean source vector must then contain (1) excesses or (2) a completely separate ocean signal originating from air/sea interactions. Since calcium, a good tracer for terrestrial dust, is only present in this continental vector, the trends of its species time series (Figure 8) dominate the continental source vector time series. Thus, for periods of very strong calcium content, the continental vector will be strong and, subsequently, so will, the related ocean components. Depending on the extent of ocean loading, this may be enough to capture all contributions from the ocean source, if it is correlated with continental sources. Any excess to the ratio between these dust/ocean components would then appear in the ocean time series correlated to the continental time series. However, if the ocean and continental time series possess negative correlations, the ocean source could be of a different origin than any ocean components captured in the continental source. As such, the calcium content of the continental vector is then acting as an identification tag for the correlated ocean source. By understanding the geographic location of calcium sources, i.e., glacial desertified areas, and physically reasonable transport routes to Greenland, it should be possible to separate the total ocean signal into components originating from two different regions, e.g., the North Atlantic and the Pacific.

A more detailed discussion of the coefficient time series and their implications for glacial-deglacial dynamics is given by *Marsh and Ditlevsen* [1997b].

7. Summary

A new method has been developed for analyzing multiple time series obtained from a single spatial receptor point. The method can be trivially generalized to the case of more receptor points. By maintaining the physical restriction that any reduced system must be positive and based on the assumption that source areas maintain a constant relative abundance in time, the source vectors can be isolated. We have demonstrated that the method captures the structure of an artificially generated noisy signal.

The chemical data obtained from the GISP ice core have been analyzed using this new technique. Three predominant sources are identified. Through interpretation of the chemical ratios within each vector (a posteriori), it has been possible to show that they capture the basic chemistry of oceanic, continental, and biochemical source types. The respective coefficient time series possess correlations, that indicate, that although the source types are geographically independent, their transport routes within the atmosphere are not. This is quite different from previous EOF studies where the source types and corresponding time series are all independent. Ready to use codes for performing this analysis can be obtained freely by request to pditlev@pditlev.gfy.ku.dk.

Acknowledgments. Thanks are extended to NCAR for hospitality and J.S. Pedersen for valuable discussions. The work was funded partly by the Carlsberg Foundation and partly by the Environment Program of the Commission for the European Communities. The GISP2 data was obtained from the archive at NSIDC.

References

Bales, R.C., and E.W., Wolff, Interpreting natural climate signals in ice cores, *Eos Trans. AGU, 76(47)*, 477, 482-483, 1995.

Bond, G.C., and R., Lotti, Iceberg discharges into the North Atlantic on millennial time scales during the last glaciation, *Science, 267*, 1005-1010, 1995.

Burgermeister, S., et al, On the biogenic origin of dimethylsulfide: Relation between chlorophyll, ATP, organismic DMSP, phytoplankton species, and DMS distribution in Atlantic surface water, *J. Geophys. Res., 95*, 20,607-20,615, 1990.

Clausen, H.B., et al., 1250 years of global volcanism as revealed by central Greenland ice cores, in *Ice Core Studies of Global Biogeochemical Cycles, NATO ASI Ser., Ser. I*, (edited by, R.J., Delmas) pp. 175-194, 1993.

Dansgaard, W., J.W.C., White, and S.J., Johnsen, The abrupt termination of the Younger Dryas climate event, *Nature, 339*, 532-533, 1989.

Ditlevsen, P., H., Svensmark, and S.J. Johnsen, Contrasting atmospheric and climate dynamics of the last glacial and Holocene periods, *Nature, 379*, 810-812, 1996.

Grassberger, P., and I., Procaccia, Characterisation of strange attractors, *Phys. Rev. Lett., 50*, 346-349, 1983.

Grassberger, P., Do climate attractors exist?, *Nature, 323*, 609-612, 1986.

GRIP Members, Climate instability during the last interglacial period recorded in the GRIP ice-core, *Nature, 364*, 203-207, 1993.

Hammer, C.U., et al, Continuous impurity analysis along the Dye 3 deep core, in *Greenland Ice Core: Geophysics, Geochemistry, and the Environment, Geophys. Monogr. Ser.*, vol. 33, edited by C.C., Langway, Jr., H., Oeschger, and W., Dansgaard), pp. 90-94, AGU, Washington, D.C., 1985.

Henry, R. C., Current factor analysis receptor models are ill-posed, *Atmos. Environ.*, 21, 1815-1820, 1987.

Henry, R. C., et al., Review of receptor model fundamentals, *Atmos. Environ., 18*, 1507-1515, 1984.

Herron, M.M., and C.C., Langway, Jr., Chloride, nitrate and sulfate in the Dye 3 and Camp Century, Greenland ice cores, in *Greenland Ice Core: Geophysics, Geochemistry, and the Environment, Geophys. Monogr. Ser.*, vol. 33, edited by C.C., Langway, Jr., H., Oeschger, and W., Dansgaard), pp.90-94, AGU, Washington, D.C., 1985.

Holland, H.D., *The Chemistry of Atmosphere and Oceans*, pp.153-249, Wiley Intersci., New York, 1978.

Johnsen, S. J., W., Dansgaard, and J., White, The origin of Arctic precipitation under present and glacial conditions, *Tellus, Ser. B, 41*(4), 452-468, 1989.

Lawson, C.L., and R.J., Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, N.J., 1974.

Legrand, M.R., Sulphur-derived species in polar ice: A review, in *Ice Core Studies of Global Biogeochemical Cycles, NATO ASI Ser., Ser. I*, (edited by, R.J., Delmas) pp. 91-120, 1993.

Legrand, M.R., and R.J., Delmas, Formation of HCl in the Antarctic atmosphere, *J. Geophys. Res., 93*, 7153-7168, 1988.

Maasch, K.A., Calculating climate attractor dimension from $\delta^{18}O$ records by the Grassberger-Proccaccia algorithm, *Clim. Dyn., 4*, 45-55, 1989.

MacAyeal, D.R., Binge/purge oscillations of the Laurentide ice sheet as a cause of the North Atlantic's Heinrich events, *Paleoceanography, 8(6)*, 775-783, 1993.

Marsh, N.D., and P.D., Ditlevsen, Observation of atmospheric and climate dynamics from a high resolution ice core record of a passive tracer over the last glaciation, *J. Geophys. Res., 102*, 11,219-11,224, 1997a.

Marsh, N.D., and P.D., Ditlevsen, Climate during glaciation and deglaciation identified through chemical tracers in ice-cores, *Geophys. Res. Lett., 24*, 1319-1322, 1997b.

Mayewski, P., et al., Changes in atmospheric circulation and ocean cover over the North Atlantic during the last 41,000 years, *Science, 263*, 1747-1751, 1994.

Mayewski, P., et al., The atmosphere during the Younger Dryas, *Science, 261*, 195-197, 1993.

Orville, R.E., and R.W., Hendersen, Global distribution of midnight lightning: September 1977 to August 1978, *Mon. Weather Rev., 114*, 2640-2653, 1986.

Pleisendorfer, R.W., *Principle Component Analysis in Meteorology and Oceanography*, Elsevier Sci., New York, 1988.

Richman, M.B., Review article: Rotation of principal components, *J. Clim., 6*, 293-335, 1986.

Ruelle, D., Chemical kinetics and differentiable dynamical systems, in: *Nonlinear Phenomena in Chemical Dynamics*, edited by A. Pacault and C. Vidal, pp.30-37, Springer-Verlag, New York, 1981.

Silvente, E. and M., Legrand, A preliminary study of the air-snow relationship for nitric acid in Greenland, in *Ice Core Studies of Global Biogeochemical Cycles, NATO ASI Ser., Ser. I*, (edited by, R.J., Delmas) pp.225-240, 1993.

Tong, H., *Non-linear Time Series*, Oxford Uni. Press, New York, 1990.

Wayne, R.P., *Chemistry of Atmospheres*, 2nd ed., Clarendon, Oxford, England, 1991.

Wolff, E.W., Nitrate in polar ice, in *Ice Core Studies of Global Biogeochemical Cycles, NATO ASI Ser., Ser. I*, (edited by, R.J., Delmas) pp. 195-224, 1993.

P. D. Ditlevsen and N. D. Marsh, Department of Geophysics, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen OE, Denmark. (e-mail: pditlev@gfy.ku.dk; ndm@gfy.dk.ku)

(Received April 15, 1997; revised November 26, 1997; accepted December 4, 1997.)