HAL
open science

Statistical distributions of ice core sulfate from climatically relevant volcanic eruptions

P. Naveau, C.M. Ammann

To cite this version:

P. Naveau, C.M. Ammann. Statistical distributions of ice core sulfate from climatically relevant volcanic eruptions. Geophysical Research Letters, 2005, 32 (5), 10.1029/2004GL021732 . hal-03134057

HAL Id: hal-03134057

https://hal.science/hal-03134057v1

Submitted on 8 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistical distributions of ice core sulfate from climatically relevant volcanic eruptions

P. Naveau
Laboratoire des Sciences du Climat et de l'Environnement, IPSL-CNRS, Gif-sur-Yvette, France
Department of Applied Mathematics, University of Colorado at Boulder, Boulder, Colorado, USA

C. M. Ammann
Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colorado, USA

Received 12 October 2004; revised 5 January 2005; accepted 15 February 2005; published 12 March 2005.

[1] Quantitative knowledge of external climate forcing is required for accurately attributing past climatic changes. Information on volcanic activity over the past millennium has primarily been drawn from high-latitude ice cores. A few large events with distinct signatures in the ice are well known and they are commonly used as marker events to synchronize time scales in individual ice cores. Over the past decade different efforts have been undertaken to systematically identify lesser known eruptions and to develop time series of past volcanic forcing. Here we mathematically quantify the distribution of the magnitude of volcanic events that have a climatic relevance during the past millennium. Volcanic sulfate magnitudes of such events clearly exhibit a "heavy tailed" extreme value distribution. Indeed, the climatically relevant eruptions are only the extremes of global volcanic activity. This characterization of volcanic amplitude is a fundamental step in detection and attribution studies of past natural forcing and of its effects on climate. **Citation:** Naveau, P., and C. M. Ammann (2005), Statistical distributions of ice core sulfate from climatically relevant volcanic eruptions, *Geophys. Res. Lett.*, *32*, L05711, doi:10.1029/2004GL021732.

1. Introduction

1.1. Motivation

[2] Large explosive volcanic eruptions can play an important role in climate [*Crowley*, 2000]. Their typical spike-like signature has been found in numerous long climatic time series, instrumental and proxy alike [*Lamb*, 1970; *Kelly and Sear*, 1984; *Bradley*, 1988; *Briffa et al.*, 1998]. Statistically, they can be viewed as pulse-like events, i.e. short and intense deviations from the background climate, with a somewhat delayed recovery. Except for the largest events, the signal of interest is often contaminated by strong noise. In the past, prior knowledge was often needed to allow in many cases an identification of these smaller events. Recently, a new method has been proposed for an automatic extraction of pulse-like signals [*Naveau et al.*, 2003] which offered the advantage of being an objective procedure to identify the timing of events. Additionally, a probability associated with each extracted event provided a measure of confidence in the extracted pulse-like event. An underlying assumption held in this study was that the magnitude of the extracted signal would follow a Gaussian distribution. But, since only the largest of explosive eruptions inject significant amounts of gases and aerosol into the stratosphere, spike amplitudes recovered in ice cores are probably not well characterized by a Gaussian distribution. They rather follow a skewed distribution and can take values outside of the range of a Gaussian tail (e.g. Tambora in 1815/16). *Hyde and Crowley* [2000] applied an simple exponential function to describe large events. Section 2 demonstrates that an analysis within the framework of extreme value theory is required for modeling the large, climatically relevant volcanic events. In Section 3, the largest amplitudes distribution is identified as being heavy tailed. Therefore, the exponential fit proposed by *Hyde and Crowley* [2000] does not entirely capture the amplitude of the largest peaks.

[3] Developing a parameterization of the distribution of strong volcanic events is important for at least three reasons: First, it provides a better tool to compare various reconstructions that might differ somewhat in their resolution of individual events [*Robock and Free*, 1995; *Zielinski*, 1995; *Crowley*, 2000; *Robertson et al.*, 2001]. Second, a better description of this forcing is important for detection and attribution of natural climate variability over the past centuries and millennia. Finally, a parameterization can be used to generate realistic forcing series for climate model simulations with statistically appropriate forcing scenarios [*Hyde and Crowley*, 2000].

1.2. Data Sets

[4] The volcanic data considered here include three very different data sources of volcanic activity of the past Millennium. For summaries of commonly used data, see *Zielinski* [2000] and *Robock* [2000]. The first series from the Crete ice core of South-Central Greenland [*Hammer et al.*, 1980] is the result of an electric conductivity measurement (ECM) that records the concentration of highly conducting substances such as acids (see the top panel in Figure 1). These include sulfuric acid, the major aerosol produced by large volcanic eruptions. In addition, the signal might be dependent on other acids or dust deposition as well as other factors [*Wolff et al.*, 1997]. Consequently, this Crete signal is strongly tainted by noise. In comparison, our second dataset, the sulfate record from the Greenland Ice Sheet Project - 2 (GISP2) [*Zielinski et al.*, 1994] (see the third panel in Figure 1), is more appropriate for the

Copyright 2005 by the American Geophysical Union.
0094-8276/05/2004GL021732

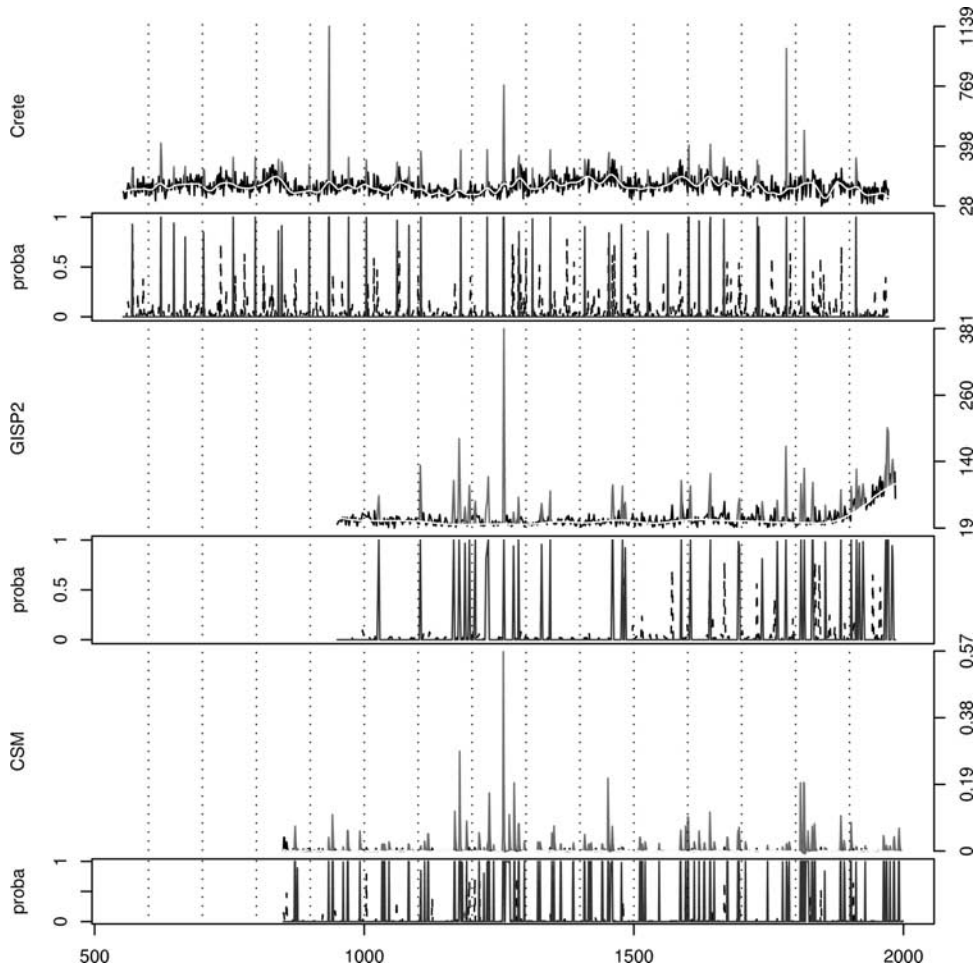

Figure 1. Panels 1, 3, and 5: Crete-ECM, GISP2-Sulfate and a composite series used in climate model simulations (CSM), see text for details and units. Panels, 2, 4, and 6: Probabilities of having a peak. In all panels, the x-axis shows the time in year, the solid, respectively dashed, lines correspond to events with a probability greater, respectively lower, than 0.8″.

detection of volcanic input. This ice core obtained near the summit in Greenland was biannually sampled and actual sulfate, SO_4^{2-}, was measured. We use the original noisy sulfate data, not the volcanic sulfate series by *Zielinski et al.* [1994]. Finally, the analysis is applied to a composite time series used in climate model simulations (C. Ammann et al., Constraining the range of solar irradiance changes during the last millennium, submitted to *Journal of Climate Letters*, 2005, hereinafter referred to as Ammann et al., submitted manuscript, 2005), referred here as the CSM record and shown in the fifth panel of Figure 1. This data series was developed in an attempt to isolate volcanic sulfate from the noise using many ice cores. An ideal series of this type should reflect the statistical properties found in the individual ice cores. Since it does not contain noise, any extraction procedure should be able to essentially capture the full series.

2. Statistical Methodology

2.1. Extracting the Volcanic Signal

[5] Before quantifying the amplitude of pulse-like events, we first need to identify the timing of such events and to extract their intensities from the noise and the possible low frequency part of the signal (trends, cycles, etc).

[6] Using prior knowledge of the volcanic activity during the last 400 years [*Lamb*, 1970; *Newhall and Self*, 1982; *Simkin and Siebert*, 1994], one can visually associate a number of cooling events in any given climatic time series to known volcanic events [*Briffa et al.*, 1998]. To move from visual devices to automatic identification procedures, a variety of extraction methods has been studied during the 1990's [*Crowley et al.*, 1993; *Zielinski et al.*, 1994]. More recently, *Naveau et al.* [2003] proposed a global and automatic extraction procedure for detecting pulse-like events such as abrupt volcanic coolings. The same technique can also be applied to other time series containing spikes, namely volcanic records in ice cores. The main advantage of this automatic procedure over past methods is that it provides a probability associated with each detected pulse (see solid lines in Figure 1) and thus a measure of confidence for a more accurate selection of the events. The distribution of their intensities, however, was not clearly characterized. The main goal of the following section is to provide a mathematically sound model for evaluating the distribution of such intensities.

2.2. Extreme Value Theory

[7] Extreme value theory has long been applied to a variety of problems in finance [*Embrechts et al.*, 1997]

Table 1. Estimated Generalized Pareto Distribution Parameters Obtained From a Maximum Likelihood Estimation[a]

Data	First Year	Last Year	Threshold t	Number Above t	σ	ξ
Crete	553	1972	150	20	87.85 (37.79)	0.56 (0.37)
GISP2	950	1985	35	36	29.60 (7.42)	0.24 (0.19)
CSM	850	1999	0.1	19	0.04 (0.02)	0.59 (0.49)

[a]See equation (1) and Section 2.2. Numbers inside the parentheses represent the standard errors (the standard error is an asymptotic estimate). The threshold was chosen by implementing classical threshold selection methods: mean excess function, qq-plots, thresholds versus parameters. See Section 2.2 and *Coles* [2001] for more details on these techniques.

and hydrology [*Katz et al.*, 2002]. In recent years, the statistical methodology for extremes has also been used for a wide range of problems in climate studies [see *Smith*, 1992; *Kharin and Zwiers*, 2000; *Coles*, 2001; *Naveau and Moncrieff*, 2003]. To learn more about extreme value theory and their applications, the books by *Coles* [2001] are recommend.

[8] Probability theory for extremes indicates that the distribution of the exceedances can be asymptotically approximated by the Generalized Pareto distribution (GPD)

$$P(X \leq x) = 1 - \left(1 + \xi \frac{x-t}{\sigma}\right)^{-1/\xi} \tag{1}$$

where X represents the random variable of interest (exceedances). Both the x and the fixed threshold t are real numbers for which x has to satisfy $1 + \xi\frac{x-t}{\sigma} > 0$ with $\sigma > 0$ and $x - t$ has to be positive because it is the exceedance above such a threshold. While σ is a simple scale parameter, ξ guides the tail behavior of the distribution, and thus is responsible for representing the extremes. For example, the classical exponential distribution corresponds to $\xi = 0$. If $\xi > 0$ (Fréchet type) then the distribution is "heavy tailed". This means that large values are more likely to occur than with classical "light tailed" distributions such as the Gaussian and Gamma distributions. For example, the variance is infinite if $\xi > 0.5$. If $\xi < 0$ (Weibull type), it has a bounded upper tail. Hence, the sign and value of the shape parameter are very important when modeling the behavior of exceedances. At this juncture, we stress that the distribution of exceedances is not chosen arbitrarily; rather the GPD is the unique distribution *derived theoretically* for modeling exceedances [e.g., see *Embrechts et al.*, 1997]. It is important to note that the GPD fits exceedances not only from a Gaussian sample, but also from *any* continuous distribution (e.g., exponential, uniform, Cauchy, etc.). Hence, the methodology is general and independent of specific numerical values. Analogous to the Central Limit Theorem that deals with sample averages, this result concerning exceedances distribution is asymptotic, i.e. the approximation by the GPD gets better as the sample size increases.

[9] After applying the extraction procedure described by *Naveau et al.* [2003] to our three time series (see Figure 1), we select pulse-like events for which the posterior probability is greater than 0.8 (see panels 2, 4 and 6 in Figure 1), i.e. with a high confidence that these peaks do not belong to the background noise. The next step is to apply the classical extreme value technique called *Peaks Over Threshold* [e.g., see *Katz et al.*, 2002]. As described above, if a sufficiently high threshold t is applied (see Table 1), exceedances above this threshold *have* to follow a GDP. Classically, finding an optimal threshold is a difficult problem. A value of t too high results in too few exceedances and consequently high variance estimators. For t too small estimators become biased (the theory works only asymptotically). For our cases, the pre-selection step of only keeping pulse-like events with a posterior probability of greater than 0.8 has been equivalent to remove lower values. Consequently, we are already working in the tail of the distribution and the threshold selection is less important, the approximation of the exceedances by a GPD being reasonable at this stage. Still, to optimize the threshold choice, we implement classical threshold selection methods: mean excess function, qq-plots, thresholds versus parameters. See *Embrechts et al.* [1997, section 6.5] for more details on these techniques. For each time series, Table 1 shows the selected threshold t after implementing such procedures and the number of observations above t.

[10] Finally, we estimate the GPD parameters (σ and ξ) by using the GPD density, i.e. the first derivative of equation (1):

$$\frac{1}{\sigma}\left(1 + \xi \frac{x-t}{\sigma}\right)^{-1/\xi - 1} \tag{2}$$

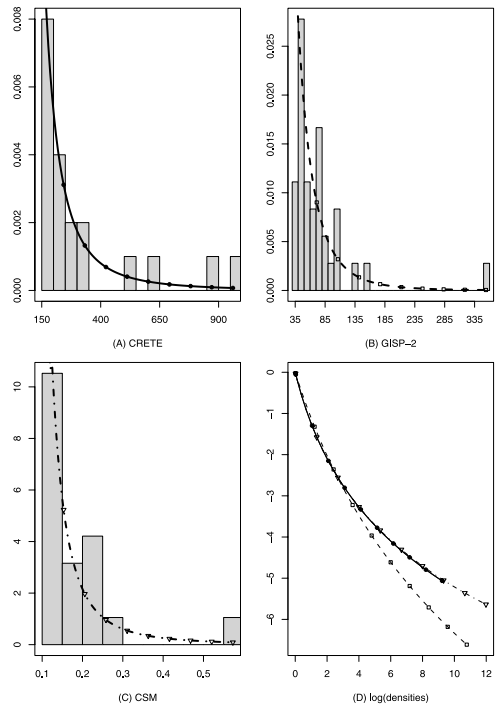

Figure 2. Panels A–C: Comparison between the histogram and the fitted GPD (see equation (2)). Panel D: Log display of the densities for the renormalized observations, see equation (1) and Section 2.2.

To obtain these parameters, we implement a maximum likelihood procedure, i.e. the probability of observing our data is maximized relative to the GPD parameters. We refer to *Coles* [2001] for a detailed presentation of maximum likelihood techniques for extremes.

[11] To assess the quality of fit, the GPD distributions with respect to our data, the GPD density (see equation (2)) is superimposed to the histogram of the exceedances in the three panels of Figure 2. Comparing densities with histograms indicates the fit is good overall. In addition, it shows that the largest values of the distribution are captured by the fitted densities. This will not be the case with a simple exponential function, and even less with a Gaussian fit. Other diagnostic tools to assess the GPD fit quality (mean excess function, qq-plots, thresholds versus parameters, profile-likelihood, etc.) have also been used. These graphs are available upon request from the authors. Note that the number and width of the histograms in Figure 2 were chosen in order to emphasize the three following points: (1) exceedances do not have a symmetric distribution (in contrast with the Gaussian case), (2) the GDP fit is reasonable for our data and (3) the largest values are well captured by the fitted GDP.

3. Volcanic Distribution Comparisons

[12] The GPD model allows us to compare the strength of extreme events among the three records. Table 1 summarizes the GPD characteristics of each record. For all three time series, the shape parameter is clearly positive (see the standard errors associated with ξ in Table 1). This indicates that, overall, the distributions of exceedances are clearly heavy tailed, although the GISP2 tail (ξ around 0.24 compared to while for Crete and CSM a value of around 0.6 is found, Table 1). This difference is also illustrated in the lower-right panel of Figure 2. After rescaling each time series by subtracting the threshold t and dividing by the scale parameter σ (because the data do not have the same unit), the difference among the tails can be observed on a log-scale. Crete and CSM event magnitudes are essentially the same while the GISP2 data exhibits a somewhat less heavy tail (though still with significantly positive ξ). This result might arise because the GISP2 record is a biannual record with a mean of the sulfate loading spread over time. This way the peak of volcanic spikes will be somewhat underestimated. Additionally, the location of GISP2 is further north compared to the Crete site. Particularly the Icelandic events (see AD 1783) are much less pronounced in the GISP2 data.

4. Conclusion and Future Work

[13] It has been shown that the magnitudes of volcanic eruptions as represented by different parameters in polar ice cores are best described by a Generalized Pareto Distribution (GPD). As expected from extreme value theory, this GPD class is universal (mathematically derived), as long as large extremes are the object of interest. This generality of the GDP class is illustrated by studying three time series with different signal to noise ratios. The GPD shape parameter becomes one of the most important indicators when comparing the intensity of volcanic spikes and it emphasizes the heavy tailed aspect of the explosive volcanic signal in these three time series.

[14] The natural next step would be to test if and how the extreme value distribution of the volcanic sulfate is preserved in the forcing and ultimately in the climate response signals. There is some indication that the resulting radiative forcing might not linearly scale with the sulfate mass [*Pinto et al.*, 1989]. This analysis will be performed in a forthcoming study.

[15] Finally, we would like to conclude with a note of caution. Modeling extreme event distributions in time series is by nature associated with a large margin of error. By definition, extreme events are rare and therefore very few large values are available to estimate the appropriate parameters. Extreme value theory helps by providing a theoretically sound model such as the GDP. Still, the estimation procedure is generally performed on a very small sample and it is important that the largest values retain the most important weight in the analysis. Without this attention, the important structure of the heavy tail can be lost.

[16] **Acknowledgments.** This work was supported by the NSF-GMC grant (ATM-0327936) and by The Weather and Climate Impact Assessment Science Initiative (NCAR). The authors would like to thank an anonymous reviewer and the editor. Their comments had greatly helped to improve the clarity of the manuscript.

References

Bradley, R. S. (1988), The explosive volcanic eruption signal in Northern Hemisphere continental temperature records, *Clim. Change*, *12*, 221–243.

Briffa, K. R., P. D. Jones, F. H. Schweingruber, and T. J. Osborn (1998), Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years, *Nature*, *393*, 450–455.

Coles, S. G. (2001), *An Introduction to Statistical Modeling of Extreme Values*, *Springer Ser. Stat.*, vol. 14, 208 pp., Springer, New York.

Crowley, T. J. (2000), Causes of climate change over the past 1000 years, *Science*, *289*, 270–277.

Crowley, T. J., T. A. Criste, and N. R. Smith (1993), Reassessment of Crete (Greenland) ice core acidity/volcanism link to climate change, *Geophys. Res. Lett.*, *20*, 209–212.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997), *Modelling Extremal Events for Insurance and Finance*, *Appl. Math.*, vol. 33, 645 pp., Springer, New York.

Hammer, C. U., H. B. Clausen, and W. Dansgaard (1980), Greenland ice sheet evidence of post-glacial volcanism and its climatic impact, *Nature*, *288*, 230–235.

Hyde, W. T., and T. J. Crowley (2000), Probability of future climatically significant volcanic eruptions, *J. Clim.*, *13*, 1445–1450.

Katz, R., M. Parlange, and P. Naveau (2002), Extremes in hydrology, *Adv. Water Resour.*, *25*, 1287–1304.

Kelly, P. M., and C. B. Sear (1984), Climatic impact of explosive volcanic eruptions, *Nature*, *311*, 740–743.

Kharin, V., and F. Zwiers (2000), Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean gcm, *J. Clim.*, *13*, 3760–3788.

LaMarche, V. C., Jr., and K. K. Hirschboeck (1984), Frost rings in trees as records of major volcanic eruptions, *Nature*, *307*, 121–126.

Lamb, H. H. (1970), Volcanic dust in the atmosphere; with a chronology and assessment of its meteorological significance, *Trans. R. Philos. Soc. London, Ser. A*, *266*, 425–533.

Leadbetter, M., G. Lindgren, and H. Rootzén (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

Naveau, P., and M. Moncrieff (2003), A statistical formulation of convective mass fluxes, *Q. J. R. Meteorol. Soc.*, *129*, 2217–2233.

Naveau, P., C. Ammann, H. Oh, and W. Guo (2003), An automatic statistical methodology to extract pulse like forcing factors in climatic time series, in *Volcanism and the Earth's Atmosphere*, *Geophys. Monogr. Ser.*, vol. 139, pp. 177–186, AGU, Washington, D. C.

Newhall, C. G., and S. Self (1982), The volcanic explosivity indes (vei): An estimate of explosive magnitude for historical volcanism, *J. Geophys. Res.*, *87*, 1231–1238.

Pinto, J. P., R. P. Turco, and O. B. Toon (1989), Self-limiting physical and chemical effects in volcanic eruption clouds, *J. Geophys. Res.*, *94*, 11,165–11,174.

Reiss, T. (1997), *Statistical Analysis of Extreme Values*, Springer, New York.

Robertson, A., et al. (2001), Hypothesized climate forcing time series for the last 500 years, *J. Geophys. Res.*, *106*, 14,783–14,803.

Robock, A. (2000), Volcanic eruptions and climate, *Rev. Geophys.*, *38*, 191–219.

Robock, A., and M. P. Free (1995), Ice cores as an index of global volcanism from 1850 to the present, *J. Geophys. Res.*, *100*(D6), 11,549–11,567.

Robock, A., and J. Mao (1995), The volcanic signal in surface temperature observations, *J. Clim.*, *8*, 1086–1103.

Simkin, T., and L. Siebert (1994), *Volcanoes of the World*, 2nd ed., 349 pp., Geoscience Press, Tucson, Ariz.

Smith, R. (1992), Extreme value theory, in *Handbook of Applicable Mathematics*, vol. 7, edited by W. Ledermann, pp. 437–471, John Wiley, New York.

Wolff, E., W. Miners, J. Moore, and J. Paren (1997), Factors controlling the electrical conductivity of ice from the polar regions—A summary, *J. Phys. Chem.*, *101*, 6090–6094.

Zielinski, G. A. (1995), Stratospheric loading and optical depth estimates of explosive volcanism over the last 2100 years derived from the Greenland Ice Sheet Project 2 ice core, *J. Geophys. Res.*, *100*, 20,937–20,955.

Zielinski, G. A. (2000), Use of paleo-records in determining variability within the volcanism-climate system, *Quat. Sci. Rev.*, *19*, 417–438.

Zielinski, G. A., P. A. Mayewski, L. D. Meeker, S. Whitlow, M. S. Twickler, M. Morrison, D. A. Meese, A. J. Gow, and R. B. Alley (1994), Record of volcanism since 7000 B.C. from the GISP2 Greenland ice core and implications for the volcano-climate system, *Science*, *264*, 948–952.

C. M. Ammann, Climate and Global Dynamics Division, National Center for Atmospheric Research, 1850 Table Mesa Drive, Boulder, CO 80307-3000, USA. (ammann@ucar.edu)

P. Naveau, Department of Applied Mathematics, 526 University of Colorado, Boulder, CO 80309-0526, USA. (naveau@colorado.edu)