

NOTES AND CORRESPONDENCE

On the Stochastic Nature of the Rapid Climate Shifts during the Last Ice Age

PETER D. DITLEVSEN

Centre for Ice and Climate, The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

OVE D. DITLEVSEN

Department of Mechanical Engineering, Technical University of Denmark, Lyngby, Denmark

(Manuscript received 16 January 2008, in final form 1 July 2008)

ABSTRACT

The rapid climate shifts observed in the glacial climate are analyzed. The transitions into the warm interstadial states, the onsets, are easily identifiable in the record. The distribution of waiting times between consecutive onsets is well fitted, assuming that the remaining residence time in each state is independent of the past. This implies that it has a simple no-memory exponential waiting time distribution, but with the mean waiting time depending on the climate state. The mean waiting time from one onset to the next is around 2400 yr. The most likely (maximum likelihood) distribution derived solely from the onset sequence is rather insensitive to the mean waiting time in the warm interstadials in the range of 400–1200 yr. When extending the analysis to include the transitions from the warm interstadials to the cold stadials observed with a larger uncertainty, the distributions in the two states are well fitted to exponential distributions, with mean waiting times of around 800 yr in the warm state and around 1600 yr in the cold state. This observation is an important piece in the climate puzzle, because the fact that the climate is a no-memory process indicates that the transitions are noise induced and the mean residence time in one state indicates how stable that climate state is to perturbations. The possibility of a hidden periodic driver is also investigated. The existence of such a driver cannot be ruled out by the relatively sparse data series (containing only 21 onsets). However, because the record is fitted just as well by the much simpler random model, this should be preferred from the point of view of simplicity.

1. Introduction

The Dansgaard–Oeschger (DO) events of rapid climate shifts in the glacial climate observed in the Greenland ice cores are still not well understood. It is debated to which extent this is either a regional or a more global in nature phenomenon (Wunsch 2006), but it seems to be agreed that there are two quasi-stationary climate states: the glacial or stadial and the DO or interstadial states. These represent reorganizations of the climate system either related to changes in the North Atlantic oceanic circulation (Weaver et al. 1991), the extent of the ice sheets (MacAyeal 1993), sea ice (Gildor and

Tziperman 2003), or even changes in the mean state of the atmospheric flow (Wunsch 2006). The state-of-the-art global circulation models are not capable of simulating the events, which should be of high priority in order to test if the models resolve the true dynamical range of possible climate changes. The mechanism for triggering a DO event could be externally driven possibly by solar variations (Braun et al. 2005), related to either internal nonlinear free oscillations (Haarsma et al. 2001) or random noise-driven jumps between quasi-stationary climate states (Cessi 1994; Ditlevsen 1999).

To approach this enigma from an observational point of view, the waiting time statistics could be used both to distinguish between the different possibilities and to be a benchmark for modeling the events. From the spectral content of the signal it has been proposed that the triggering might be periodic (Schulz 2002). However, it was demonstrated in Monte Carlo simulations that the

Corresponding author address: Peter D. Ditlevsen, Centre for Ice and Climate, The Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, Copenhagen DK-2100, Denmark.
E-mail: pditlev@gfy.ku.dk

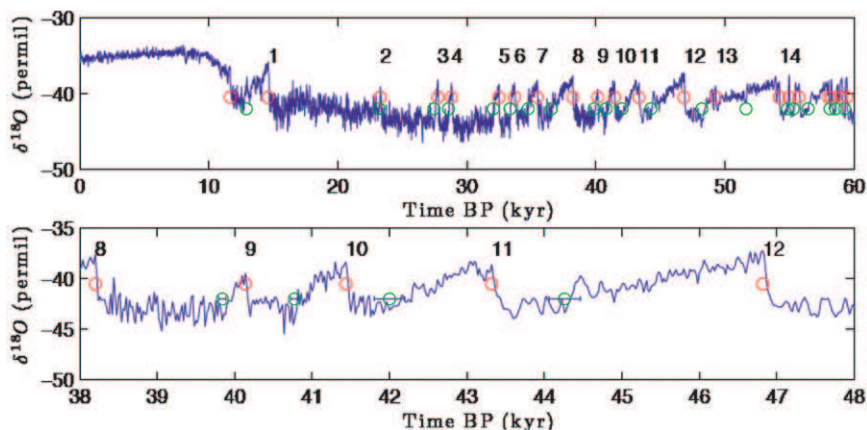


FIG. 1. (top) NGRIP isotope record with red circles indicating the transitions from the stadial to the interstadial state (off-on), while the green circles indicate the transitions from the interstadial to the stadial state (on-off). The numbers are the canonical numbering of DO events. (bottom) Close-up view of the period 38–48 kyr BP, where it is seen that the determination of the on-off transitions are much more uncertain than determining the sharp off-on transitions. This is indicated by the error bars. The transition times in years BP based on the NGRIP Greenland Ice Core Chronology 2005 (GICC05) time scale used here are as follows: off-on = [11 660, 14 600, 23 300, 27 740, 28 800, 32 480, 33 680, 35 440, 38 200, 40 140, 41 440, 43 320, 46 820, 49 260, 54 200, 54 960, 55 720, 58 020, 58 240, 59 040, 59 420]; on-off = [12 900, 23 180, 27 440, 28 560, 32 040, 33 340, 34 740, 36 540, 39 840, 40 770, 42 000, 44 260, 48 190, 51 620, 54 840, 55 280, 56 440, 58 120, 58 540, 59 260].

apparent periodicity is not significantly different from what would be found in a random signal (Ditlevsen et al. 2007). Likewise, it was demonstrated that a stochastic resonance model (Alley et al. 2001) would have too small a strength of the periodic forcing to be detected above the noise level for the DO events (Ditlevsen et al. 2005).

The waiting time statistics relies on accurate dating of the events observed in the ice core records. The dating of the North Greenland Ice Core Project (NGRIP) ice core has been completed by annual layer counts going back to 60 kyr BP with unprecedented accuracy (Svensson et al. 2008). The quoted dating uncertainty (one sigma) accumulates to 1300 yr at 60 kyr BP. Thus, an approximate uncertainty of $(1300/\sqrt{60}) \sqrt{n} \approx 168 \sqrt{n}$ y accumulates during n kyr. Herein this measuring uncertainty is not included in the analysis of the data read from the NGRIP isotope record.

The isotope record from NGRIP is shown in Fig. 1. The numbers refer to the original numbering of DO events (Dansgaard et al. 1993). The red circles indicate the initiations of DO events. A discussion of criteria for determining the initiations can be found in Ditlevsen et al. (2005). The green circles indicate the terminations of the DO events. As seen in the blow up in the lower panel, these are much less well determined than the initiations.

2. Distribution of the time between consecutive DO events

The fact that the time points of initiation of the DO events are much more precisely identified than the points of termination motivates us to first study the statistical properties of the stream of initiation points by themselves.

In appendix B, we investigate whether it is compatible with the observed sequence of consecutive time distances between the DO events to assume that these chronologically ordered distances are mutually independent. This is done by transforming the observed time distances into approximately normally distributed numbers, followed by calculating the “autocorrelation” coefficients of the transformed sequence. The deviations from zero of these “number lag” correlation coefficients are found to be insignificant. The property of zero correlation is not a sufficient condition for independence in the case of normal marginal distributions. However, the simplest possible consistent assumption is that the bivariate distribution is normal, implying that the two variables become independent. The dataset is far too small for rejection of the bivariate normal hypothesis with any reasonable confidence.

The simplest suggestion of a model of a random point stream with statistically independent consecutive time

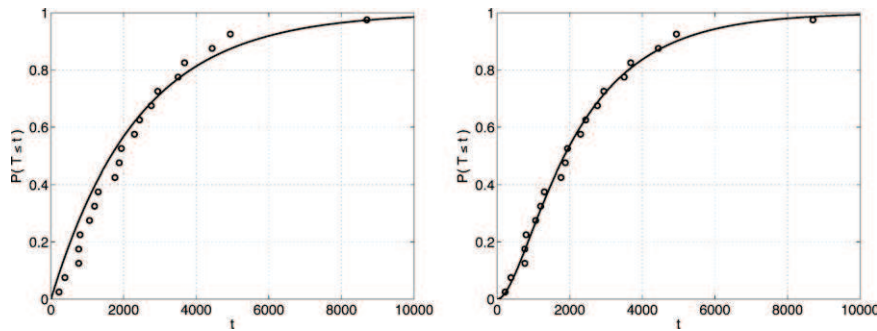


FIG. 2. (left) The exponential distribution function $F_T(t) = 1 - \exp[-t/E(T)]$, $t > 0$, corresponding to the mean $E(T) = 2388$ yr, together with the observed 20-sample distribution function. (right) A better fit by the distribution function $F_T(t) = 1 - [\tau_1 \exp(-t/\tau_1) - \tau_2 \exp(-t/\tau_2)] / (\tau_1 - \tau_2)$, $t > 0$, corresponding to the maximum likelihood estimates $\tau_1 = 493$ yr, $\tau_2 = 1895$ yr obtained from the sample so that $\tau_1 + \tau_2 = E(T)$.

distances is that these have a common probability distribution characterized by a lack of memory in the following sense: The probability distribution of the remaining time distance to the first point after any chosen instant is the same independent of how far back in time from the instant the previous point occurred. This assumption implies that the exponential distribution is the only possible distribution for the time distance between two consecutive points in the stream (Johnson and Kotz 1970). The point stream is then a so-called homogeneous Poisson process.

The observed sample consists of 20 distances and has the empirical distribution function plotted as circles in Fig. 2 [plotting position after ordering the sample as $[t_i, 1/40 + (i - 1)/20]$, where $t_1 \leq t_2 \leq \dots \leq t_i \leq \dots \leq t_{20}$]. Moreover, the exponential distribution function with the mean equal to the average 2388 yr of the observed sample is shown. The systematic misfit that appears between the empirical distribution and the exponential distribution indicates that the assumption of lack of memory may not be fully convincing.

A next step of choice of a simple and also physically reasonable model that includes memory is to assume that the system can be in two different states, and that

the only memory element is the current state of the system.

This is tantamount to the assumption that the glacial climate varied in time as a realization of a stationary on-off jump process [where the interstadials (D-Os) are the “on” state and the stadials are the “off” state] in which the jumps between the two states occur completely at random and at any time statistically independent of the past. It is emphasized that such an on-off model does not imply that there must be statistically stationary behavior (stability) within each state. At this level of modeling the within-state variations are of no concern except for the lack of memory. Physically the triggering of the jumps could come from the fast time-scale chaotic variations (noise), which are temporally uncorrelated on the observed climatic time scales. The assumed within-state lack of memory implies that the durations of the interstadials and the durations of the stadials become exponentially distributed, but not necessarily with the same mean values.

Adopting this model we have $T = T_1 + T_2$, where T_1 and T_2 are independent exponentially distributed random variables of means τ_1 and τ_2 , respectively. The sum of the two has the probability density

$$f_{T_1 + T_2}(t) = \int_0^t \frac{1}{\tau_1} e^{-s/\tau_1} \frac{1}{\tau_2} e^{-(t-s)/\tau_2} ds = \begin{cases} (\tau_1 - \tau_2)^{-1} (e^{-t/\tau_1} - e^{-t/\tau_2}) & \text{for } \tau_1 \neq \tau_2 \\ \tau^{-2} t e^{-t/\tau} & \text{for } \tau_1 = \tau_2 = \tau. \end{cases} \quad (1)$$

For $\tau_1 \neq \tau_2$, the probability density at the point (t_1, \dots, t_n) in the n -dimensional space of the n sample of T is

$$L(\tau_1, \tau_2; t_1, \dots, t_n) = (\tau_1 - \tau_2)^{-n} \prod_{i=1}^n (e^{-t_i/\tau_1} - e^{-t_i/\tau_2}), \quad (2)$$

according to the multiplication rule for statistically independent random variables. This density calculated for the actual sample t_1, \dots, t_n is called the likelihood of the parameters (τ_1, τ_2) and as function of these parameters, $L(\tau_1, \tau_2; t_1, \dots, t_n)$ is called the likelihood function. It is obviously a reasonable principle to estimate

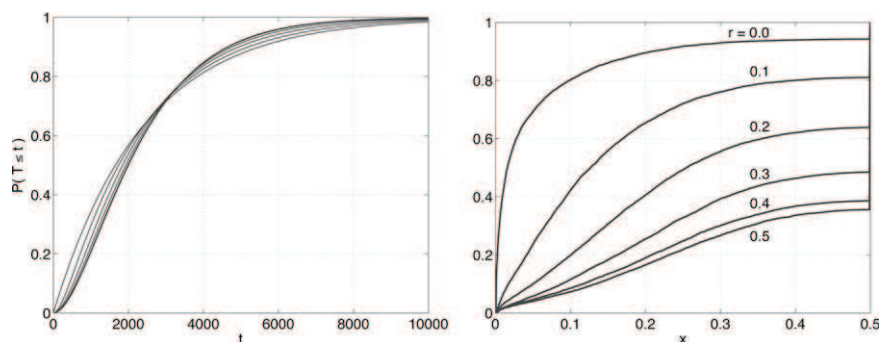


FIG. 3. Plots showing distribution parameter insensitivity. (left) The family of distribution functions $F_T(t) = 1 - [\tau_1 \exp(-t/\tau_1) - \tau_2 \exp(-t/\tau_2)]/(\tau_1 - \tau_2)$, $t > 0$, corresponding to the mean $E(T) = \tau_1 + \tau_2 = 2388$ yr and the ratios $\tau_1/(\tau_1 + \tau_2) = 0.1, 0.2, 0.3, 0.4, 0.5$ (from left to right in the lower part of the diagram). (right) The conditional distribution functions of the ratio of the maximum likelihood estimator of τ_1 and the average estimator of $\tau_1 + \tau_2$ given that $r = \tau_1/(\tau_1 + \tau_2) = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5$, respectively. At $x = 0.5$ all the distribution functions jump to the value 1, since by definition $\tau_1 \leq \tau_2$.

the true values of the parameters by the point at which their likelihood is maximal.

It is a simple exercise to see that the maximum likelihood estimate of τ_1 from a sample t_{11}, \dots, t_{1n} of the exponentially distributed variable T_1 is the average of the sample, and similarly for τ_2 . From this fact it does not follow that the maximum likelihood estimate of $\tau_1 + \tau_2$ is the sum of the two averages, that is, the average of the sample $t_{11} + t_{21}, \dots, t_{1n} + t_{2n}$, except if $\tau_1 = \tau_2$. However, the sample average estimator of $\tau_1 + \tau_2$ is consistent with the fact that $E(T) = \tau_1 + \tau_2$. This means that the maximum likelihood estimate of $\tau_1 + \tau_2$ is biased while the average is not. Thus, we remove this bias by maximizing L under the restriction

$$\tau_1 + \tau_2 = \frac{1}{n} \sum_{i=1}^n t_i. \quad (3)$$

The sample size is $n = 20$ and the average is $\bar{t} = 2388$ yr. The value of τ_1 for which L is maximal is $\tau_1 = 493$ yr. Figure 2 (right panel) shows the corresponding distribution function, together with the observed 20-point sample distribution function. The improvement of the fit from the left panel to the right panel is striking. However, it should be noted that there is a very small sensitivity of the distribution function in the right panel to variations of the parameter values τ_1 and τ_2 given the sum $\tau_1 + \tau_2$. This is clear from inspecting the bunch of distribution function graphs in the left panel of Fig. 3. These correspond to the distribution functions of T for different values of the ratio $\tau_1/(\tau_1 + \tau_2)$ ranging from 0 (exponential distribution) over 0.1, 0.2, 0.3, 0.4, to 0.5 (gamma distribution). For the obtained maximum likelihood estimate the ratio is $493/2388 \approx 0.21$. The right panel of Fig. 3 shows the conditional distribution

functions (each obtained from 10 000 simulated 20-point samples) of the ratio of the maximum likelihood estimator of τ_1 and the average estimator of $\tau_1 + \tau_2$, given the ratio $r = \tau_1/(\tau_1 + \tau_2)$ for $r = 0.0, 0.1, \dots, 0.5$. It is remarkable that there is a probability of about 0.36 (the jump at $x = 0.5$) that the maximum likelihood estimate of τ_1 and τ_2 becomes equal even if the true ratio $\tau_1/(\tau_1 + \tau_2)$ is as small as 0.2, close to the obtained estimate 0.21. Thus, the gamma distribution is wrongly selected with a probability of 0.36, if a sample of size of 20 is drawn and the maximum likelihood estimate of τ_1 is calculated from that sample.

It is also seen that if the exponential distribution is assumed to be the true distribution of T , then there is only about a 10% probability to get an estimate larger than 0.2 of $\tau_1/(\tau_1 + \tau_2)$. Thus, it is not unreasonable to look for a better model than the exponential. This is further supported by the physically reasonable hypothesis of the existence of two states.

It is an obvious task to investigate whether it is possible to construct an equally well fitting distribution function under some deterministic DO event triggering mechanism. For example, this mechanism might be some hidden periodic forcing. Such a model is considered in appendix A. One might defend its more complicated nature if the detected periodicity could be given some geophysical or astronomical cause. Otherwise, the simple pure randomness model is just as good. It seems difficult to go further in the direction of determinism when considering the possibility of the existence of two different states of random duration.

The indication of the existence of two states of independent exponentially distributed durations by the quite good fit obtained in the right panel of Fig. 2 in-

vites further inspection of the record to see if the states are visible and measurable with acceptable accuracy.

3. Estimation of the exponential distributions

The estimation of the parameters τ_1 and τ_2 will obviously be improved if the 20 realizations of the (T_1, T_2) pair are included in the estimation. However, in the $\delta^{18}O$ isotope record the interstadials have the characteristic saw-tooth shape, which is not present in the records of dust concentrations. By inspecting the $\delta^{18}O$ record displayed in Fig. 1 it is recognized that because of the saw-tooth shape it is considerably more difficult to accurately read the time points of transition from interstadial to stadial states than from the stadial to the interstadial states analyzed in the previous section. In addition to the duration variables T_1 and T_2 we need, therefore, to introduce an error-of-identification variable R assigned with suitable distributional properties.

Specifically write $T = X_1 + X_2$, where X_1 and X_2 are random variables that represent the identified interstadial and stadial durations, respectively. Next write $X_1 = T_1 + R, X_2 = T_2 - R$, where, for given $(T_1, T_2) = (t_1, t_2), R$ is a random variable with values that by necessity is limited to the interval $-t_1 < R < t_2$ because X_1 and X_2 both are positive. This means that the usual normal distribution model for an error distribution is not applicable here. Instead, an applicable sufficiently extensive and mathematically convenient distribution family is represented by the beta-distribution family, for which the standard form of the probability density is $f_\beta(x; p, q) = x^{p-1} (1-x)^{q-1} / B(p, q), 0 < x < 1$, where p and q are positive parameters, and $B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx$ (the beta function). Thus, for given $(T_1, T_2) = (t_1, t_2)$, the random variable $(T_1 + R)/(T_1 + T_2)$ is assumed to have the conditional probability density

$$f_{R|T_1, T_2}(r|t_1, t_2) = \frac{1}{(t_1 + t_2)B(p, q)} \left(\frac{r + t_1}{t_1 + t_2}\right)^{p-1} \left(1 - \frac{r + t_1}{t_1 + t_2}\right)^{q-1}, \quad r \in (-t_1, t_2). \tag{4}$$

From the relations $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$ and $\Gamma(p+1) = p\Gamma(p)$ it is directly seen that $\int_0^1 x^p(1-x)^{q-1} dx = B(p, q)p/(p+q)$, so that we get the conditional expectation

$$E(R|T_1 = t_1, T_2 = t_2) = (t_1 + t_2) \frac{p}{p+q} - t_1. \tag{5}$$

We will model the error correction R so that it has the conditional mean value $E(R|T_1 = t_1, T_2 = t_2) = \mu t_1 t_2$,

where μ is some constant. This model reflects that the error must vanish in the two limits $t_1 \rightarrow 0$ and $t_2 \rightarrow 0$. It follows from (5) that this property is obtained if q is defined as

$$q = \frac{t_2(1 - t_1 \mu)}{t_1(1 + t_2 \mu)} p. \tag{6}$$

To obtain the joint distribution of X_1, X_2 we note that $x_1 + x_2 = t_1 + t_2$. Thus,

$$\begin{aligned} f_{X_1, X_2|T_1, T_2}(x_1, x_2|t_1, t_2) &= \delta[x_2 - (t_1 + t_2 - x_1)] f_{X_1|T_1, T_2}(x_1|t_1, t_2) \\ &= \delta[t_2 - (x_1 + x_2 - t_1)] f_{R|T_1, T_2}(x_1 - t_1|t_1, t_2) \\ &= \delta[t_2 - (x_1 + x_2 - t_1)] \frac{1}{(t_1 + t_2)} f_\beta \left[\frac{x_1}{t_1 + t_2}; p, \frac{t_2(1 - t_1 \mu)}{t_1(1 + t_2 \mu)} p \right]. \end{aligned} \tag{7}$$

Removing the conditioning by use of the total probability theorem, (7) gives the joint probability density

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \frac{1}{\tau_1 \tau_2} \int_0^\infty \int_0^\infty e^{-t_1/\tau_1} e^{-t_2/\tau_2} f_{X_1, X_2|T_1, T_2}(x_1, x_2|t_1, t_2) dt_1 dt_2 \\ &= \frac{e^{-(x_1 + x_2)/\tau_2}}{\tau_1 \tau_2 (x_1 + x_2)} \int_0^{x_1 + x_2} e^{-t(1/\tau_1 - 1/\tau_2)} f_\beta \left\{ \frac{x_1}{x_1 + x_2}; p, \frac{(x_1 + x_2 - t)(1 - \mu t)}{t[1 + \mu(x_1 + x_2 - t)]} p \right\} dt, \end{aligned} \tag{8}$$

which, with the restriction $\tau_2 = a - \tau_1$ [see (3)], $a = 2388$ yr, gives the likelihood function

$$L[\tau_1, p, \mu; (x_{11}, x_{21}), \dots, (x_{1n}, x_{2n})] = \prod_{i=1}^n f_{X_1, X_2}(x_{1i}, x_{2i}). \tag{9}$$

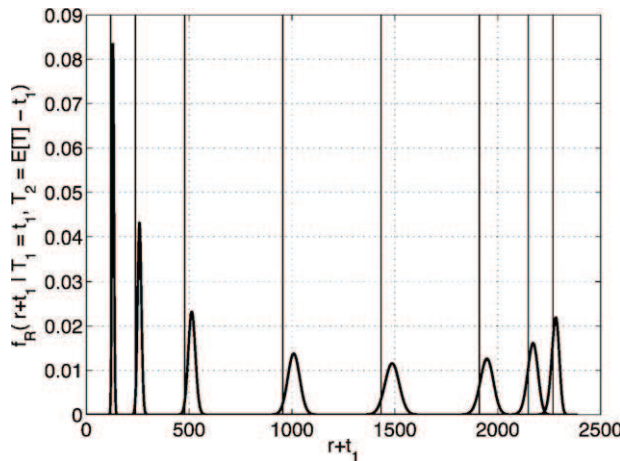


FIG. 4. Examples of estimated conditional distributions of the error term R given $(T_1, T_2) = (t_1, 2388 \text{ yr} - t_1)$. The vertical lines mark the point $r = 0$. It is noted that R is positive in the mean with a bell-shaped probability density of appearance as a normal distribution.

The maximum is obtained for $\tau_1 = 802 \text{ yr}$, $\tau_2 = 1586 \text{ yr}$, $p = 698$, and $\mu = 39.1 \cdot 10^{-6} \text{ yr}^{-1}$. The average of the X_1 sample is 846 yr , that is, 44 yr larger than τ_1 . We note that $\mu\tau_1\tau_2 \approx 50 \text{ yr}$.

Figure 4 shows the estimated conditional distributions of the error term R for a selected set of values of T_1 . Even though the conditional standard deviations of R are not insignificant for values of T_1 in the range somewhat larger than from about τ_1 to about τ_2 , the error practically averages out in its influence on the estimated distribution functions shown in Fig. 5. Only a small influence comes from the bias away from zero of the mean of R .

The obtained ratio $\tau_1/(\tau_1 + \tau_2) = 802/2388 \approx 0.34$ is not in conflict with the ratio 0.21 estimated from the on-on data in the previous section. In fact, if we assume

that 0.34 is the true value of the ratio, then it follows from the corresponding conditional distribution function in the right panel of Fig. 3 that the probability of obtaining an estimate of the ratio less than 0.21 is about 25% when drawing a random 20 samples from the assumed distribution.

A simple evaluation of the statistical uncertainty of the maximum likelihood estimators of τ_1 and τ_2 , interpreted as the mean values of two independent exponentially distributed random variables, are the standard deviations $\tau_1/\sqrt{n} = 802/\sqrt{20} = 179 \text{ yr}$, and $\tau_2/\sqrt{n} = 1586/\sqrt{20} = 355 \text{ yr}$. Thus, the standard deviation of the estimator of $a = \tau_1 + \tau_2$ is $\sqrt{179^2 + 355^2} = 398 \text{ yr}$.

The estimated distribution functions fit quite reasonably to the observed 20 -sample distribution function values shown as circles in Fig. 5. The statistically inexperienced reader may express doubt about this. However, it is important for the judgment of the quality of the fit to be aware that for a sample size of no more than 20 , deviations of a size as seen in Fig. 5 must be expected with high probability. To convince the reader, a demonstration of the effect of this kind of statistical uncertainty is given in the next section.

The obtained distribution results, based on the assumption that T_1 and T_2 are statistically independent, to some degree also support the independence hypothesis. For further support, the scatterplot of the 20 pairs of observations of (X_1, X_2) (transformed to standard normal distribution) are shown (see Fig. C1). The empirical correlation coefficient is calculated to be as small as about 0.01 . It should be noted that zero correlation between T_1 and T_2 implies that there is some correlation between $X_1 = T_1 + R$ and $X_2 = T_2 - R$. For completeness the correlation coefficient between X_1 and X_2 is calculated in appendix C only by use of the model and the obtained parameter estimates. The the-

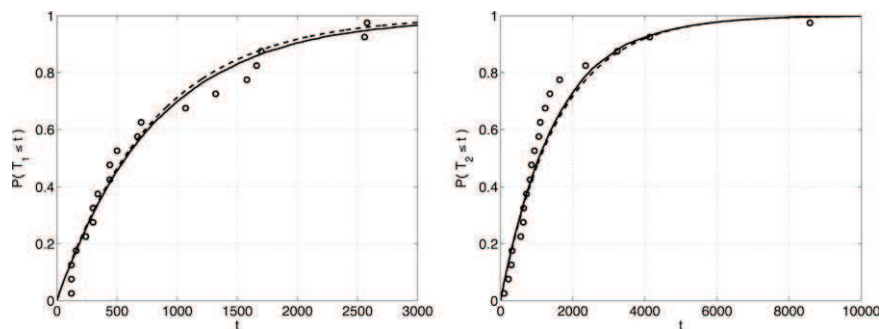


FIG. 5. Exponential distribution functions $F_{T_i}(t) = 1 - \exp(-t/\tau_i), t > 0$, corresponding to the maximum likelihood estimates (left) $\tau_1 = 802 \text{ yr}$ and (right) $\tau_2 = 1586 \text{ yr}$ obtained from the sample of (on-off, off-on) duration pairs (dashed curves) together with the estimated marginal distribution functions of X_i obtained by simulation for the maximum likelihood estimates $\tau_1 = 802 \text{ yr}$, $p = 698$, and $\mu = 39.1 \times 10^{-6}$ (full curves). The small circles show the observed 20 -sample distribution functions of X_1 (left) and X_2 (right).

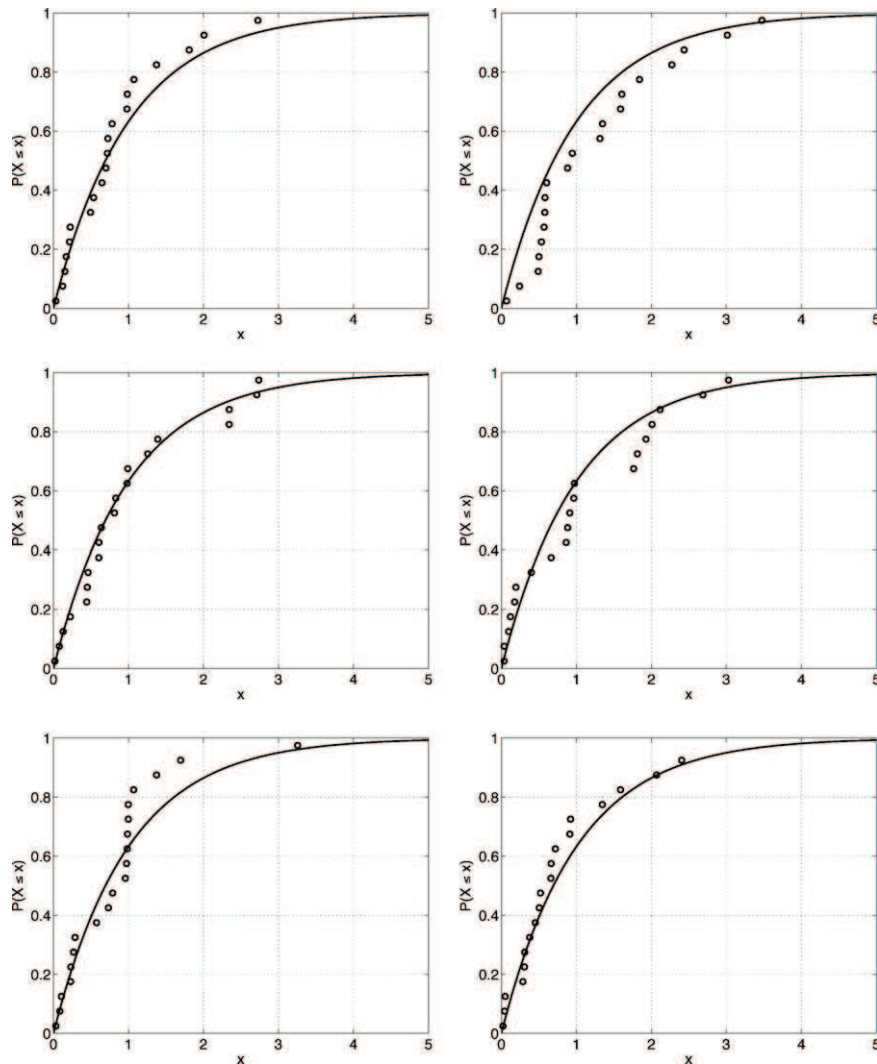


FIG. 6. Documentation of the effect of statistical uncertainty. Each panel shows the exponential distribution function $F(x) = 1 - e^{-x}$ together with the empirical distribution function of 20 samples generated from the exponential distribution according to the formula $X = -\log(U)$, where U is a uniformly distributed random variable on the interval $(0, 1)$.

oretical correlation coefficient is obtained to 0.02, that is, it is negligibly small. In the same calculation the standard deviation of R is found to be about 90 yr.

The findings in this section increase the confidence that the hypothesis of the exponential distribution of the consecutive distances between the DO events should be rejected. This is because the hypothesis that $T_1 + T_2$ has exponential distribution is inconsistent with the finding that T_1 and T_2 are uncorrelated and exponentially distributed (or just close to being exponentially distributed). The exponential probability density $a \exp(-at)$ for the sum would be generated if T_1 and T_2 are independent with gamma probability densities proportional to $t^{r-1} \exp(-at)$ and $t^{s-1} \exp(-at)$, respec-

tively, where $r + s = 1$, $r > 0$, $s > 0$. These gamma densities are infinite for $t = 0$. However, the data do not support an assumption of infinite probability densities at $t = 0$.

4. Demonstration of the actual statistical uncertainty

Figure 6 shows six independent examples of empirical distributions obtained by generating 20 samples from the standard exponential distribution with density function $f(x) = e^{-x}$, $x > 0$. The corresponding distribution function $F(x) = 1 - e^{-x}$ is shown in each panel. Thus, large deviations occur without inconsistency with

the theoretical probability distribution. Visual comparison with the deviations seen in Fig. 5 shows that there is no reason to expect a fit of better quality than that obtained there.

Returning to Fig. 2, it is seen in the perspective of Fig. 6 that the exponential distribution for the on–on time distances cannot be excluded as the valid distribution. In fact, earlier we read from the right panel in Fig. 3 that there is a probability of about 10% that the estimate of the ratio $\tau_1/(\tau_1 + \tau_2)$ becomes larger than 0.2 under the assumption that the ratio is zero. Thus, the path of observation point deviations seen in the left panel of Fig. 2 does not occur so rarely for the exponential model that one is entitled to reject the exponential model as the valid model. To reject the exponential model with confidence it is essential that supplementary information is considered, for example, such as the information that two independent states are detectable in the isotopic record.

5. Interpretations and concluding remarks

The waiting time distribution observed in the ice core record is well fitted by a random on–off process. The mean waiting times in the two states are different, indicating that the waiting time depends on the climate state. In the warm state the mean waiting time (≈ 800 yr) is markedly shorter than the mean waiting time in the cold state (≈ 1600 yr). The random nature of the waiting times indicates that the jumping is triggered by internal climatic fluctuations. A possible exception is the transition into the present warm Holocene period, which has been stable for 12 000 yr. This could indicate either that the Holocene climate is a distinctly different climate from the interstadials or that the internal climate fluctuations become so small in the Holocene climate that the triggering of jumps into the cold state is exponentially suppressed. Both possibilities could be argued to be caused by the disappearance of the Laurentide and Fennoscandian ice sheets, which makes the Holocene distinctly different from the interstadials.

From the random no-memory nature of the climatic jumps it should be expected that there are no correlations between consecutive climate periods, so that a short interstadial period is not preferentially followed by either a long or a short stadial period, and vice versa. This is independently checked and is indeed the case, as is shown in detail in appendixes B and C. In principle this observation could constraint the possibilities for how glacier buildup or Atlantic sea-saw connections influence the Dansgaard–Oeschger climatic shifts.

In the glacial period the internal climate fluctuations are observed to be larger in the cold stadial state than

in the warm interstadial state (Ditlevsen et al. 2002). Naively, one should expect that from a purely noise-induced transition the waiting time in the cold state should be shorter than in the warm state, in contrast to the observations. Thus, it can be concluded from the observations that the cold state is more stable than the warm state and the barrier for noise-induced transition is highest from the cold to the warm state. This is an important benchmark for high-resolution climate models, if these can be constructed to simulate the two climate states and the transitions between the two states. At present the two climate states have been simulated by so-called water hosing experiments in intermediate complexity models, and are thus not internally noise generated in these models (Schmittner et al. 2002).

As seen from appendix A, other statistical models, including hidden periodic triggers, can explain the data record as well. This is an unfortunate consequence of the limited size of the record. From a philosophical point of view one should, in our opinion, refer to the principle of Occam's razor, favoring the simplest model with fewest assumptions to explain the data, unless independent evidence selects a different model. In that sense the purely random model seems at present to be the first choice.

As a final remark it should be mentioned that the observation of a simple statistical structure of the climate shifts does not imply a simple underlying climate dynamics. The jump process and threshold crossing dynamics is a highly nonlinear phenomenon.

APPENDIX A

Model with Hidden Periodic Trigger

In the last paragraph of section 2 the question is posed whether the distribution of the ice core on–on time data can be constructed from a model by which the DO events are triggered by some hidden periodic mechanism of period τ . To investigate this problem we adopt the probabilistic model

$$T_i = t_0 + N_i\tau + S_i \quad (\text{A1})$$

for the times T_1, T_2, \dots of occurrences of the consecutive DO events. Here t_0 is some unknown time origin, N_1, N_2, \dots is an increasing random sequence of integers, and S_1, S_2, \dots are independent random slack times (i.e., delay times that can be positive as well as negative) of identical probability density, which is zero outside a bounded interval of length at most equal to the period τ . The model is formulated so that not all the mechanism periods necessarily trigger a DO event. As suggested by P. Hedegård (2007, personal communica-

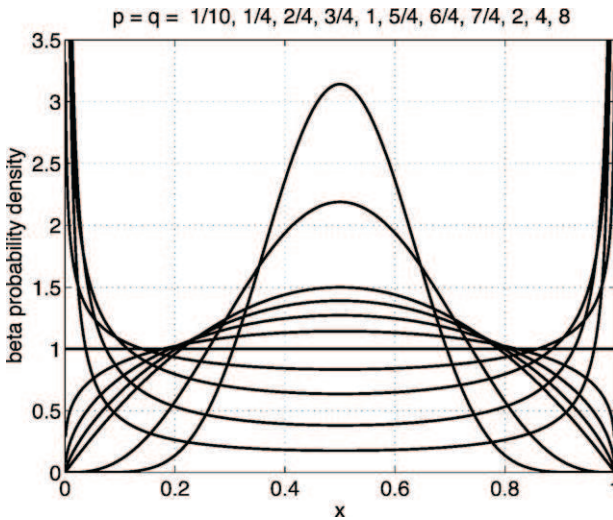


FIG. A1. Graphs of the beta density $f_{\beta}(x; p, p) = [x(1-x)]^{p-1}/B(p, p)$ for different values of the parameter p . The probability density (A4) is obtained by the transformation $s = \tau(x - 1/2)$.

tion), the successful periods with respect to DO events are reasonably chosen at random with some probability α .

Instead of proceeding with (A1), we change from absolute time to relative time considering only the time differences

$$T_{i+1} - T_i = (N_{i+1} - N_i)\tau + S_{i+1} - S_i, \quad i = 1, \dots, n, \quad (\text{A2})$$

whereby t_0 is eliminated. The variables $X_i = N_{i+1} - N_i$ are defined as mutually independent random variables with the common frequency function

$$P(X_i = x) = \alpha(1 - \alpha)^{x-1}, \quad x = 1, 2, \dots \quad (\text{A3})$$

(the geometric distribution). For convenience we write $T_{i+1} - T_i$ as T_i , with observations t_1, \dots, t_n . As probability density type for the mutually independent random variables S_1, \dots, S_{n+1} we will adopt the beta distribution density (see Fig. A1),

$$f_S(s) = \frac{1}{\tau B(p, p)} \left[\frac{1}{4} - \left(\frac{s}{\tau} \right)^2 \right]^{p-1}, \quad s \in (-\tau/2, \tau/2) \quad (\text{A4})$$

with mean $E(S) = 0$ and variance

$$\text{Var}(S) = \frac{1}{4(2p+1)} \tau^2, \quad (\text{A5})$$

which we will denote as σ^2 . The likelihood function of the unknown parameters α , τ , and p becomes too complicated even for numerical calculations to be completed within reasonable time. Instead, the much sim-

pler method of moments is applied. Noting that the identically distributed random variables given by (A2) are of the form as $T = X\tau + S_2 - S_1$, where $E(X) = 1/\alpha$, $E(X^2) = (2 - \alpha)/\alpha^2$ {easily obtained by repeated differentiation of the probability generating function $\psi(u) = E(X^u) = \alpha u/[1 - (1 - \alpha)u]$ and setting $u = 1$ }, and $E(S_2 - S_1) = 0$, $E[(S_2 - S_1)^2] = 2\sigma^2$, it follows that

$$E(T) = \tau/\alpha, \quad (\text{A6})$$

$$E(T^2) = (\tau/\alpha)^2(2 - \alpha) + 2\sigma^2. \quad (\text{A7})$$

Elimination of τ and substitution of σ^2 gives the equation

$$\alpha^2 - 2(2p+1)\alpha + 2(2p+1) \left[2 - \frac{E(T^2)}{E(T)^2} \right] = 0 \quad (\text{A8})$$

in α for each given value of p . With the given data of $n = 20$ sample values and the sample means $E(T^j) \approx n^{-1} \sum_{i=1}^n t_i^j$, $j = 1, 2$, we get the results displayed in Fig. A2. By visual judgment it seems most reasonable to adopt the uniform distribution for S , which corresponds to $p = 1$. The corresponding parameter estimates are $\tau = 854$ yr and $\alpha = 0.357$. It is difficult to distinguish the goodness of fit from that of the right panel in Fig. 2.

It is remarkable that with this simple model the most probable period is considerably shorter than the period of 1470 yr proposed in the literature (Schulz 2002). Furthermore, the obtained point process model is quite peculiar. The time distance from a point to the next point is generated almost correctly by tossing a dice repeatedly until either one dot or two dots are obtained (probability 1/3 per toss approximating α). Next, the period τ is multiplied by the used number of tosses. Finally the obtained time is corrected by adding two numbers chosen independently and completely at random from the interval $(-\tau/2, \tau/2)$. The authors consider this model as an artificial mathematical approximation construct rather than an indication of a hidden period. Moreover, the model contains no room for the existence of two states. With a sample size as small as 20 it is possible to construct many different mathematical models that cannot be rejected by a statistical test based on the sample. We have shown in this paper that whatever sophisticated model is constructed for generating the point process, it might as well be generated by a simple two-state process with complete randomness from small time interval to next small time interval (i.e., a so-called two-state Markov process).

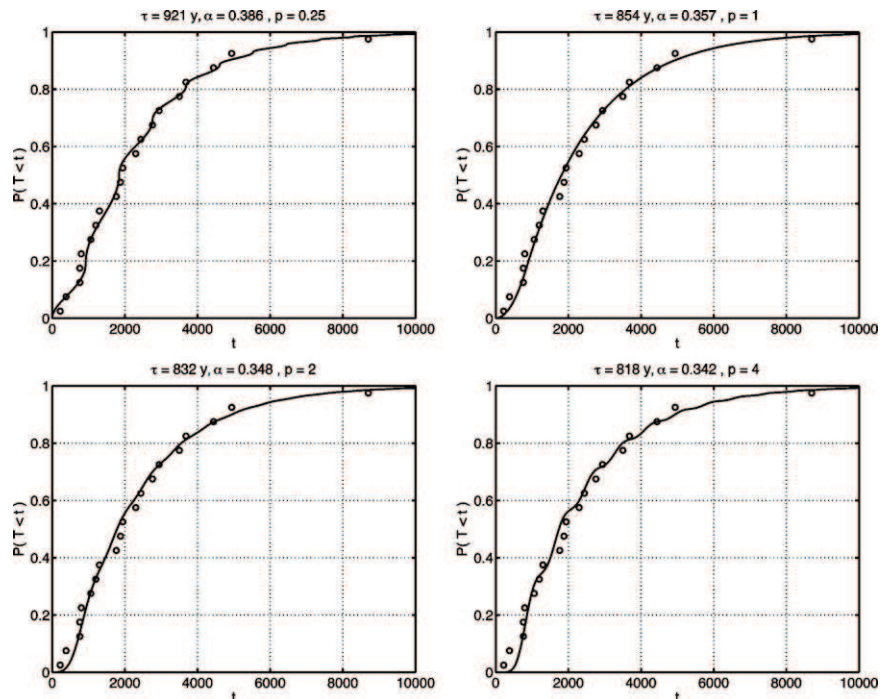


FIG. A2. Estimated distribution function of on–on time differences in periodic model with beta distributed S variables for given $p = 0.25, 1, 2, 4$, and the corresponding empirical distribution function defined by the given sample of 20 values. The standard deviations of the S variables are 441, 247, 186, 136 yr, respectively. The judgmentally most reasonable fit is seen in the upper-right panel and is obtained for the period $\tau = 854$ yr, S uniformly distributed in the interval $(-427, 427)$ yr, and the probability of 0.36 for DO event occurrence.

APPENDIX B

Test of Lack of Correlation between Consecutive DO Point Distances

The estimation of the distribution function for the time distances between the DO events described in section 2 is not dependent on whether or not there is some dependence structure in the sequence of consecutive time distances. Any reordering of the sequence leads to the same distribution function, of course. To see whether the sequence taken in chronological order as measured shows signs of a dependent structure that is essentially different from that of the same sequence of time distances obtained by randomizing the order, the 15 first “number lag” correlation coefficients are shown in Fig. B1 for the measured sequence and three other randomized versions of the measured sequence. To be more specific, the measured sequence of time distances are first transformed one to one into a sequence y_1, y_2, \dots, y_{20} with standard normal distribution of the elements. Next, the sequence of $20 - i$ pairs $(y_1, y_{i+1}), (y_2, y_{i+2}), \dots, (y_{20-i}, y_{20})$ is constructed for each $i = 1, 2, \dots, 15$, where i is called the number lag, and the

empirical correlation coefficient is calculated for each i . The results are shown as circle marks in the upper-left panel of Fig. B1. The other three panels in Fig. B1 shows the empirical correlation coefficients as small square marks for the three randomized versions. The dashed curves are distribution function quantiles of the correlation coefficient estimator under the assumption that the correlation coefficient is zero, an assumption that is true for the randomized versions of the sequence, of course. Comparison of the upper-left panel with the three other panels gives no reason to reject the hypothesis that the measured distances comply with a model where the number lag correlation coefficients are all zero.

APPENDIX C

Test of Lack of Correlation between Interstadial Duration and the Next Following Stadial Duration

The scatterplot in Fig. C1 shows that the assumption of independence between T_1 and T_2 in no way is contradicted by the observed data. However, to be careful, the size of the correlation induced between X_1 and X_2

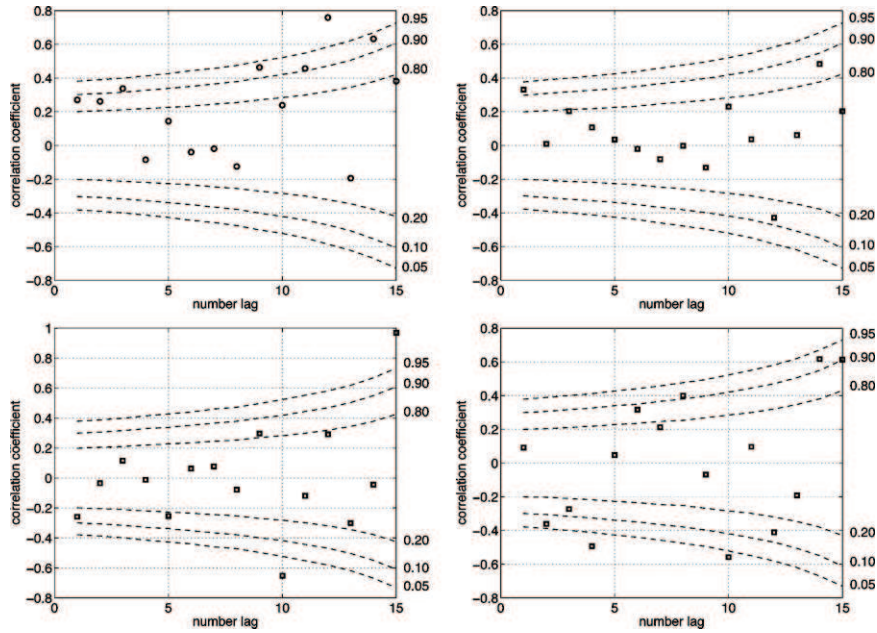


FIG. B1. Estimated number lag correlation coefficients for the sequence of consecutive time distances between DO events. (top left) The correlation coefficients for the measured sequence with the distances in chronological order. The three other panels correspond to order randomized versions of the measured sequence. The dashed curves are distribution function quantiles of the estimator given that the correlation coefficient number lag correlation coefficient is zero.

by the correcting term R must be evaluated. The following derivations address this.

Assuming that T_1 and T_2 are independent, we have the expectation $E(T_1 T_2) = E(T_1)E(T_2)$ and the covariance $\text{Cov}(T_1, T_2) = E(T_1 T_2) - E(T_1)E(T_2) = 0$. More-

over, by assumption we have $E(R|T_1, T_2) = \mu T_1 T_2$. Finally, $E(T_1) = \tau_1, E(T_2) = \tau_2, \text{Var}(T_1) = \tau_1^2, \text{Var}(T_2) = \tau_2^2$ because of the exponential distribution. By the rules of calculations with expectations and covariances it then follows that

$$\text{Cov}(T_1 + R, T_2 - R) = \text{Cov}(R, T_2) - \text{Cov}(R, T_1) - \text{Var}(R), \tag{C1}$$

$$\begin{aligned} \text{Cov}(R, T_1) &= \text{Cov}[E(R|T_1, T_2), E(T_1|T_1, T_2)] + E[\text{Cov}(R, T_1|T_1, T_2)] \\ &= \mu \text{Cov}(T_1 T_2, T_1) = \mu E(T_2) \text{Var}(T_1) = \mu \tau_2 \tau_1^2, \end{aligned} \tag{C2}$$

$$\text{Cov}(R, T_2) = \mu E(T_1) \text{Var}(T_2) = \mu \tau_1 \tau_2^2, \tag{C3}$$

and

$$\text{Var}(R) = \text{Var}[E(R|T_1, T_2)] + E[\text{Var}(R|T_1, T_2)] = \mu^2 \text{Var}(T_1 T_2) + E\left[\frac{T_1^2(1 - \mu T_1)T_2(1 + \mu T_2)^2}{T_1(1 + \mu T_2) + p(T_1 + T_2)}\right], \tag{C4}$$

where

$$\text{Var}(T_1 T_2) = \text{Var}(T_1)\text{Var}(T_2) + E(T_1)^2 \text{Var}(T_2) + E(T_2)^2 \text{Var}(T_1) = 3\tau_1^2 \tau_2^2 \tag{C5}$$

and the last term is obtained from the variance $\text{Var}(W) = pq/[(p + q)^2(p + q + 1)]$ of the standard beta distribution combined with (6) and the linear transformation

$W = (R + t_1)/(t_1 + t_2)$ that defines the probability density (4). This complicated expectation term must be calculated by numerical integration or simple Monte Carlo

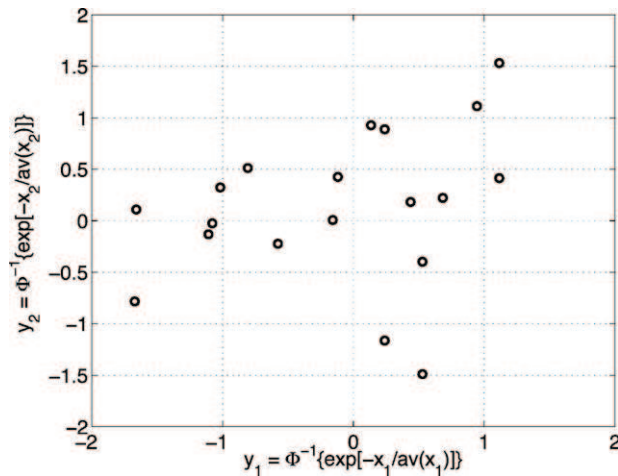


FIG. C1. Scatterplot of the 20 observation pairs (x_1, x_2) after transformation into the approximately standard normally distributed pairs (y_1, y_2) , $y_i = \Phi^{-1}\{e^{-x_i}/\text{av}(x_i)\}$, $\text{av}(x_i)$ = average of x_i sample $i = 1, 2$. The correlation coefficient is estimated to the insignificant value of 0.01.

simulation. To obtain the correlation coefficient we further need

$$\begin{aligned} \text{Var}(T_1 + R) &= \text{Var}(T_1) + \text{Var}(R) + 2\text{Cov}(T_1, R) \\ &= \tau_1^2 + \mu\tau_1^2\tau_2 + \text{Var}(R), \end{aligned} \quad (\text{C6})$$

$$\begin{aligned} \text{Var}(T_2 - R) &= \text{Var}(T_2) + \text{Var}(R) - 2\text{Cov}(T_2, R) \\ &= \tau_2^2 - \mu\tau_1\tau_2^2 + \text{Var}(R). \end{aligned} \quad (\text{C7})$$

By substitution of the estimated parameters $\tau_1 = 802$ yr, $\tau_2 = 1586$ yr, $p = 698$, and $\mu = 39.1 \text{ yr}^{-1}$, we find that the standard deviation of R becomes estimated to $\sqrt{\text{Var}(R)} \approx 90$ yr, and that the correlation coefficient between X_1 and X_2 becomes $\text{Cov}(T_1 + R, T_2 - R) / \sqrt{\text{Var}(T_1 + R)\text{Var}(T_2 - R)} \approx 0.02$, that is, a value so close to zero that it has no importance for the testing the independence of T_1 and T_2 by use of the observation pairs (x_1, x_2) .

REFERENCES

- Alley, R. B., S. Anandakrishnan, and P. Jung, 2001: Stochastic resonance in the North Atlantic. *Paleoceanography*, **16**, 190–198.
- Braun, H., M. Christi, S. Rahmstorf, A. Ganopolski, A. Mangini, C. Kubatski, K. Roth, and B. Kromer, 2005: Possible solar origin of the 1,470-year glacial climate cycle demonstrated in a coupled model. *Nature*, **438**, 208–211.
- Cessi, P., 1994: A simple box model of stochastically forced thermohaline flow. *J. Phys. Oceanogr.*, **24**, 1911–1920.
- Dansgaard, W., and Coauthors, 1993: Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature*, **364**, 218–220.
- Ditlevsen, P. D., 1999: Observation of alpha-stable noise and a bistable climate potential in an ice-core record. *Geophys. Res. Lett.*, **26**, 1441–1444.
- , S. Ditlevsen, and K. K. Andersen, 2002: The fast climate fluctuations during the stadial and interstadial climate states. *J. Glaciol.*, **35**, 457–462.
- , M. S. Kristensen, and K. K. Andersen, 2005: The recurrence time of Dansgaard–Oeschger events and limits on the possible periodic component. *J. Climate*, **18**, 2594–2603.
- , K. K. Andersen, and A. Svensson, 2007: The DO-climate events are probably noise induced: Statistical investigation of the claimed 1470 years cycle. *Climate Past*, **3**, 129–134.
- Gildor, H., and E. Tziperman, 2003: Sea-ice switches and abrupt climate change. *Philos. Trans. Roy. Soc. London*, **361A**, 1935–1944.
- Haarsma, R. J., J. D. Opsteegh, F. M. Selten, and X. Wang, 2001: Rapid transitions and ultra-low frequency behaviour in a 40 kyr integration with a coupled climate model of intermediate complexity. *Climate Dyn.*, **17**, 559–570.
- Johnson, N. L., and S. Kotz, 1970: *Distributions in Statistics: Continuous Univariate Distributions I*. John Wiley and Sons, 300 pp.
- MacAyeal, D. R., 1993: Binge/purge oscillations of the Laurentide ice sheet as a cause of the North Atlantic's Heinrich events. *Paleoceanography*, **8**, 775–783.
- Schmittner, A., M. Yoshimori, and A. J. Weaver, 2002: Instability of glacial climate in a model of the ocean-atmosphere-cryosphere system. *Science*, **295**, 1489–1493.
- Schulz, M., 2002: On the 1470-year pacing of Dansgaard-Oeschger warm events. *Paleoceanography*, **17**, 1014, doi:10.1029/2000PA000571.
- Svensson, A., and Coauthors, 2008: A 60 000 year Greenland stratigraphic ice core chronology. *Climate Past*, **4**, 1–11.
- Weaver, A. J., E. S. Sarachik, and J. Marotzke, 1991: Freshwater forcing of decadal and interdecadal oceanic variability. *Nature*, **353**, 836–838.
- Wunsch, C., 2006: Abrupt climate change: An alternative view. *Quat. Res.*, **65**, 191–203.