Author’s Response to Referee V. Livina

We thank Valerie Livina for her thorough evaluation of our manuscript and helpful comments. In the following, we list all referee comments and our according responses.

1. *“the model parameters and statistics (RMSD, p-values) would be good to summarise in a table”*

We agree that this would be helpful as a summary. In the revised manuscript, we show a table of (1) model parameters, (2) hypothesis test results (p-values) and (3) Goodness-of-Fit of the model averages (RMSD) for both stationary and non-stationary one- and two-process models. In order for the table not to become unmanageably big, we prefer to omit parameters and results for the reduced models, which are still given in the main text.

2. *“figures 2, 4, 6, notations P, E, I, S could be used in the same style, preferably in the label of the Y-axis rather than near the curves”*

We thank the referee for pointing out the different styles in the figures. We will adjust the figures in the manuscript.

3. *“Blue lines are better to replace by black ones, for better visibility”*

We agree and adjust the figures in cases where color is not needed for other purposes.

4. *“text definitions could be accompanied by a suitable formula, see page 5, lines 13-14”*

We included formulas for both E_S and P_S and hope that this increases readability.

5. *“What is the authors’ definition of “regularity” in page 4, line 12? Stationarity?”*

We agree in the lack of clarity with the expressions “regularity”/”irregularity” and “non-stationarity”. We used “regularity” as a colloquial term rather than a mathematical definition, denoting how variable the average event frequency (measured in a sliding window) is over time. We thus use the word to describe what our test statistic E_S measures. An equally spaced sequence of events is then considered fully regular. A strongly irregular event sequence in terms of E_S can be both due to non-stationarity or a very fat-tailed distribution of event waiting times. Thus it is not synonymous with “non-stationarity”. Since we give a clear definition of E_S, we keep the colloquial use of “regularity”/”irregularity” in the manuscript and add a clear explanation of what we mean with “irregularity” in the introduction:

“In order to test the hypotheses, we regard the evolution of the number of warming events in a moving window of 20 kyr. This quantity measures how variable the average event frequency is over time, a property which we denote as irregularity, and in the DO sequence it deviates strongly from a constant occurrence frequency of events over time.”

6. *“page 2, line 19, it is better to say “consider” than “regard”. The same in page 3, line 3”*

ok

7. *“page 6, line 9, it is better to say “data source” than “archive” “*

ok

8. *“page 6, line 23: “as extreme as the observed data” (not “than”)”*

ok

9. *“page 9, line 17: “it needs not” “*

We would prefer the use of a modal auxiliary (“it need not be changing…”) in this case. To make the sentence more clear we will change it to “it need not change … “ in the manuscript.

10. *“page 9, line 19: “we still think it is” “*

ok

Author's Response to T. Mitsui

We thank Takahito Mitsui for carefully reading and evaluating of our manuscript. His comments have been very helpful and have improved the quality of the revised manuscript. In the following, we list all referee comments and our corresponding responses.

1. *"I confused about the terms, "stadial rate" and "interstadial rate". I wondered if the stadail rate is the transition rate from stadial to interstadial or vice versa. I'm temped to call them "warming (cooling) rate" or " warming (cooling) event rate".*

We agree that this terminology is confusing and will adopt a version of the reviewers proposition. Our original terminology was meant the following way: Stadial rate is the rate of transitioning from stadial to interstadial, thus it corresponds to the warming rate. In the same way, the interstadial rate can be referred to as the cooling rate. We would like to propose the following change: The transition rate from stadial to interstadial (and vice versa) is referred to as warming (cooling) transition rate.

2. *"Eq. (3) sounds counterintuitive because the insolation reduces the warming rate $\lambda 1$ and the ice volume increases the warming rate. Similarly, the insolation increases the cooling rate $\lambda 2$ and the ice volume decreases the cooling rate. Is there any possible explanation for this?"*

Thanks for this observation regarding Eq. (3). It is in fact a mistake, since we mixed up the right-hand sides for $\lambda 1$ and $\lambda 2$. When changing this, the interpretation is not counter-intuitive anymore: $\lambda 1$ is the warming transition rate (increase by insolation and decrease by ice volume) and $\lambda 2$ is the cooling transition rate (decrease by insolation and increase by ice volume). It has now been corrected.

3. *"I suggest to explicitly show the relation between the stadial rate and S(t), and that between the interstadial rate and I(t) for the reduced two-process model, like Eq. (3). Otherwise, it's not entirely clear whether the insolation (the global ice volume) indeed promotes or inhibits the warming (cooling) events."*

We agree that this will improve clarity and will add an equation in the manuscript.

4. *"The integrated insolation above 350 W/m 2 (Huybers, 2006) is chosen as a forcing. Why don't you choose the summer solstice daily-mean insolation, which is also common? Is it a consequence of some optimization? If so, it is worthy to be mentioned.*

We thank the referee for this important comment and agree that we need to address this in the manuscript. We did not use the integrated insolation forcing as a result of an optimization, but rather used it as a first choice since we believed it is the relevant quantity for the phenomenon at hand, capturing the notion of positive degree days in high latitudes. However, we also conducted the fitting routine with daily-mean summer solstice insolation at 65 deg North, and obtained results that are equivalent to the case with integrated insolation. Specifically, we again find insolation control of stadial durations and ice volume control of interstadial durations. The fitting error with full forcing is then RMSD_sum = 0.62, and thus slightly worse than when using integrated insolation. We also find a reduced model with a very good fit of RMSD_sum = 0.69. The good agreement is not surprising because the two forcings look very much alike.

We thus added the following paragraph at the end of the discussion section:

“The results do not depend critically on the specific insolation forcing we used. To illustrate this, we also tried the daily-mean summer solstice insolation at 65 deg North and obtained results that are very much in line with what has been presented here. Specifically, we again find insolation control of stadial durations and ice volume control of interstadial durations. The fitting error with full forcing is RMSD_sum = 0.62, and thus only slightly worse than when using the integrated insolation presented in this paper. We also find a reduced model with a very good fit of RMSD_sum = 0.69. With this work we do not attempt to study which kind of insolation forcing might lead to the best fit, since the results would not be statistically significant given the small sample size of DO events and the fact that we already obtain a very close fit with both integrated and summer soltice insolation.”

5. *“The authors mention "While the distribution of waiting times in between warming events is well modeled by an exponential distribution (not shown here)," (P9. Line 14-15). This is the fact from the observation since Ditlevsen’s early works. The exponential distribution is true for the stationary one-process model but not true for the stationary two-process model as shown by by Eq. (1). The latter inconsistency is OK because the authors rejects the model in the end. However, is the exponential distribution consistent with the non-stationary two process model? If so, why?”*

We thank the referee for this question and would like to offer the following explanation. Instead of discussing whether (samples of) the distributions of the different models are consistent with one another (e.g. exponential distribution and non-stationary two-process model), we think it is more instructive to focus directly on whether the empirical data distribution is consistent with the different model distributions.
The empirical distribution of inter-warming times in the data lies in fact very close to an exponential distribution of the same mean and is thus clearly consistent with it (using one-sided KS test: p=0.96). Interestingly, the data is also consistent with the distribution of the two-process model using parameters estimated from data, albeit with lower significance (p=0.30). When using the non-stationary two-process model, we find again strong correspondence of data and model distribution (p=0.92). The reason for this is following: The data distribution has a large dispersion (coefficient of variation CV = 1.129) , which is close to the one of the exponential distribution (CV=1.0). The stationary two-process model is, however, less dispersed (CV=0.708), as discussed in the manuscript. If we vary the parameters of the two-process model in time, the mean is varying and thus we expect the stationary distribution to be “smeared out” (i.e. more dispersed), which we indeed find for our best-fit non-stationary two-process model (CV=1.156). Thus it is close to the data distribution and presumably consistent with an exponential distribution.
We would like to omit a discussion of this in the manuscript since the results of the paper are consistent with these considerations and the aim of this work was to go beyond the stationary statistics of waiting times in between warming events.

6. *“How is the observation of the exponential distribution consistent with the following statement?: "In the limiting case of a DO cycle comprised of a very large number of independent processes, one finds a Gaussian distribution of waiting times" (P10. Line 9-10).”*

Thanks for this interesting comment. We would like to point out that these two things are not supposed to be consistent, because we do not claim that a DO cycle is comprised of such a large sequence of independent processes. If we consider two independent processes, there is already a departure from exponential statistics. Here, we merely mention that it would tend to Gaussian statistics if there were more independent processes. The variance of this Gaussian would also decrease as the number of processes is increased.

For clarity, we propose to address this by expanding the corresponding paragraph of the revised manuscript in the following way:

"This model gives rise to a more regular sequence of warming events, compared to the one-process model. This is because one DO cycle is the sum of two independent processes and thus its duration does not follow an exponential distribution (coefficient of variation CV = 1.0), but Eq. (1), which is less dispersed (CV = 0.708). In the limiting case of a DO cycle comprised of a very large sequence of N independent and stationary processes, one finds a Gaussian distribution of waiting times with decreasing variance as N grows. This would then correspond to an almost evenly-spaced sequence of events, which is not supported by the observations."

7. *"In Eq. (1), both exponents are $-\lambda$ 1 T . Is this right?"*

Thanks. The first exponent is $-\lambda 2$ T, which has been corrected in the manuscript.

8. *"P2. Line 12: Is "single events" fine?"*

Changed it to "individual events".

9. *"P3. Line 4 and in Fig. 6: "ky" -> "kyr" (if you want to correct)"*

ok

10. *"P3. Line 24-25: Svensson et al. (2006) -> (Svensson et al., 2006)"*

ok

11. *"P4. Line 1: "withing" -> "within" "*

ok

12. *" P8. Line 3: What do you mean by "range 1". Is this the value of the standard deviation? "*

With "range 1" we mean that the amplitude (maximum – minimum value) of the signal is 1. Has been clarified in the manuscript.

Random and externally controlled occurrence of Dansgaard-Oeschger events

Johannes Lohmann[1] and Peter D. Ditlevsen[1]

[1]Centre for Ice and Climate, Niels Bohr Institute, University of Copenhagen, Denmark

Correspondence: Johannes Lohmann (johannes.lohmann@nbi.ku.dk)

Abstract. Dansgaard-Oeschger (DO) events constitute the most pronounced mode of centennial to millennial climate variability of the last glacial period. Since their discovery, many decades of research have been devoted to understand the origin and nature of these rapid climate shifts. In recent years, a number of studies have appeared that report emergence of DO-type variability in fully coupled general circulation models via different mechanisms. These mechanisms result in the occurrence of DO events at varying degrees of regularity, ranging from periodic to random. When examining the full sequence of DO events as captured in the NGRIP ice core record, one can observe high irregularity in the timing of individual events at any stage within the last glacial period. In addition to the prevailing irregularity, certain properties of the DO event sequence, such as the average event frequency or the relative distribution of cold versus warm periods, appear to be changing throughout the glacial. By using statistical hypothesis tests on simple event models, we investigate whether the observed event sequence may have been generated by stationary random processes or rather has been strongly modulated by external factors. We find that the sequence of DO warming events is consistent with a stationary random process, whereas dividing the event sequence into warming and cooling events leads to inconsistency with two independent event processes. As we include external forcing, we find a particularly good fit to the observed DO sequence in a model where the average residence time in warm periods are controlled by global ice volume and cold periods by boreal summer insolation.

1 Introduction

During the last glacial period, lasting from approximately 120 kyr BP to 12 kyr BP (kiloyears before present), a large number of abrupt large-scale climate changes have been recorded in Greenland ice cores and other Northern Hemisphere climate proxies. These so-called Dansgaard-Oeschger (DO) events (Dansgaard et al., 1993) are characterized by an abrupt warming of 10-15 K from cold conditions (stadials) to warmer conditions (interstadials) within a few decades. This is typically followed by gradual cooling, lasting centuries to thousands of years, until a more abrupt jump back to cold conditions is observed. The warming events are not regularly spaced over the glacial, but rather distributed in a complex temporal pattern, as can be seen in the NGRIP ice core record in Fig. 1. This raises questions about the causes of these recurring climate changes. Could an internal oscillation of large components of the climate system under strongly varying conditions give rise to this pattern? Are the climate changes in contrast manifestations of highly sensitive, multistable climate system components, where jumps in between different states are triggered in an unpredictable way by one or possibly many different other chaotic components?

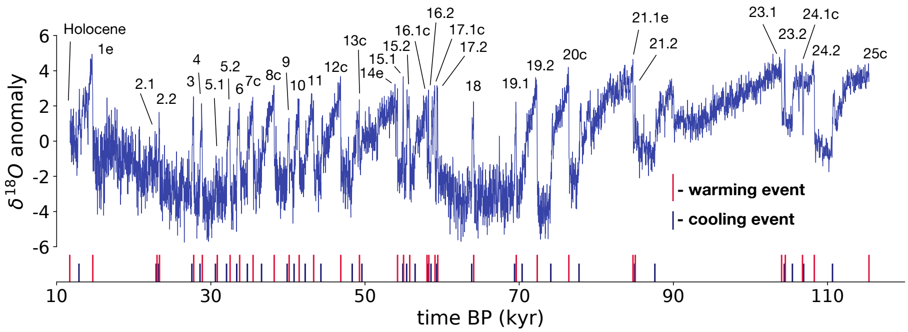

Figure 1. NGRIP oxygen isotope ice core record in 20 year binned resolution and associated Dansgaard-Oeschger warming and cooling events. The numbers above the time series indicate the warming transitions into the respective Greenland interstadials. The nomenclature is adopted from (Rasmussen et al., 2014) and only events considered in this study are marked. On the time axis we marked the timing of warming (red) and cooling (blue) events.

Since the discovery of these unexpected climate events with no known cause, questions of this kind have been addressed. Whereas high-resolution coupled climate models under glacial conditions typically lack DO-type variability, models of intermediate complexity and simpler conceptual models have been proposed to explain qualitative features of the sequence of last glacial climate changes. Starting from the discovery of an approximate 1500 year spectral signature in the GISP2 ice core record (Grootes and Stuiver, 1997) and an apparent in-phase pacing of individual events by multiples of this time period (Alley et al., 2001; Schulz, 2002; Rahmstorf, 2003), a number of competing hypotheses have been compared to the data. Among these were studies aiming to establish a mechanism for this periodicity, including direct triggering by periodic forcing (Braun et al., 2005), stochastic resonance (Alley et al., 2001), ghost resonance (Braun et al., 2007) and coherence resonance (Timmermann et al., 2003). On the other hand, it has been shown that there is limited significance to the periodic spectral signature (Braun et al., 2010) and pacing of single individual events (Ditlevsen et al., 2007). When including data reaching further back in time than 50 ky kyr BP it is found that only very weak periodic contributions to modeled switching sequences are compatible with the data and that instead it is more likely that the observed sequence of events is a realization of a purely noise-driven process (Ditlevsen et al., 2005).

In this work, we want to expand on this idea by testing whether the observed sequence of events is indeed consistent with one or more random, stationary processes, or whether the changes over time of the properties of the observed event sequence require a modulation of parameters of the governing process over time. To this end, we regard consider the whole glacial period, as opposed to previous efforts focusing on a rather regular period in the middle of the glacial. We investigate two different levels in detail of description by first only regarding the sequence of warming events and second the combined sequence of alternating transitions in between cold and warm conditions. We proceed by testing two null hypotheses: 1. The sequence of DO warming events is a realization of a Poisson process with fixed rate parameter. 2. The sequence of stadials and interstadials is a realization of two independent Poisson processes with fixed rate parameters giving rise to transitions in between stadials and interstadials. In order to test the hypotheses, we ~~regard~~ consider the evolution of the number of warming events in a moving window of 20 ~~ky~~kyr. This quantity ~~strongly deviates~~ measures how variable the average event frequency is over time, a property which we denote as irregularity, and in the DO sequence it deviates strongly from a constant occurrence frequency of events over time~~in the DO sequence~~. We test whether samples from the above mentioned stationary processes show a similar irregularity~~over time~~.

In addition to the evolution of the frequency of warming events we look at the evolution of the abundance of the stadial over the interstadial condition, which changes significantly over time in the DO sequence. This additional non-stationary structure in the data is the basis for another hypothesis test we perform. Finally, we test how the models' support with respect to the data is improved as we force the rate parameters with a combination of a global climate proxy and orbital variations of insolation, to incorporate changing background climate conditions. The main findings of this study are: 1. A Poisson process with fixed rate parameter, modeling warming transitions only, is consistent with the time variations in the NGRIP DO warming event sequence. 2. A model composed of two independent stationary Poisson processes governing transitions in between stadials and interstadials is not consistent with the time variations in the observed DO event sequence. 3. Forcing the aforementioned models with a combination of a global ice volume proxy and a summer insolation curve leads to good statistical agreement with the observed sequence. Specifically, we find good agreement for a model with two individual processes, where the average transition rate from interstadial to stadial is controlled by global ice volume forcing, obtained from independent ocean core isotope records, and the average transition rate from stadial to interstadial is controlled by boreal summer insolation.

The paper is structured in the following way. In Section 2 we introduce in more detail the data used in this study, the summary statistics used to investigate irregularity in the event series, the models used to explain the data and the hypothesis test procedure. In Section 3 we present the results of the hypothesis tests on the different models. We discuss and interpret the results in Section 4.

2 Methods and Models

Our study of the sequence of DO events is based on the refined dating represented by the GICC05 time scale ~~Svensson et al. (2006)~~(Svensso the classification of Greenland stadials (GS) and Greenland interstadials (GI) given in ~~(Rasmussen et al., 2014)~~ Rasmussen et al. (2014) and the timings reported therein. We consider all stadials and interstadials and corresponding transitions, starting with GI-25c at 115370 kyr BP and ending with the transition from GS-1 to the Holocene at 11703 kyr BP. We do not include events classified as subevents, i.e., drops in the middle interstadials to colder, but not fully stadial conditions, with the exception of GS-14. This yields a total number of 34 warming events and 33 cooling events. This increase in number from the 25 originally reported warming events is due to refined subdivision (Rasmussen et al., 2014).

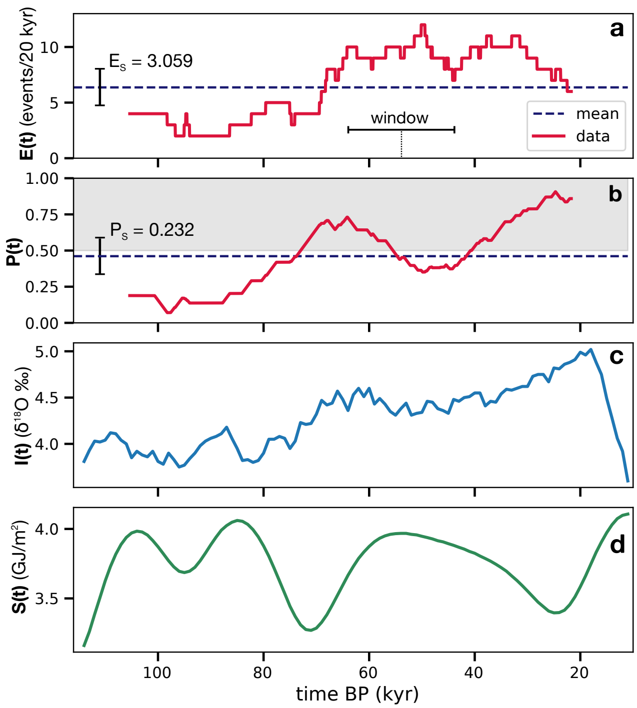

Figure 2. Time-varying ~~non-stationarity~~ irregularity indicators calculated from the NGRIP DO sequence, and climate forcings. (a): The number of warming events in a running 20 kyr window $E(t)$ (red) and the mean value ~~, corresponding to a regular sequence of events~~ (dashed blue). (b): The abundance of stadials in a 20 kyr window $P(t)$. For values greater than 0.5 (indicated by gray shading) the portion of stadials is larger than the portion of interstadials within the window. (c) Ocean sediment proxy record for global ice volume $I(t)$. (d) Integrated summer insolation at 65 ~~deg~~ degree North $S(t)$.

Given sequence and timing of transitions in between stadials and interstadials, we construct time-varying indicators of ~~non-stationarity~~ irregularity in the sequence of ~~events~~event timings, which are shown in Fig. 2a,b. To this end, we calculate the number of warming transitions ~~withing~~ within a moving window of 20 kyr at midpoint in time t, which we denote as $E(t)$. The window size of 20 kyr is chosen as trade-off ~~in~~ between resolution and statistical robustness of longer-term features in the event sequence given the characteristic time scale of event occurrence of 3.1 kyr. The window is furthermore of comparable size to dominant variations in global background climate and insolation forcing, which will be investigated below. We ~~yield a timeseries~~ obtain a time series indicating the deviation in the occurrence frequency of warming events from ~~a fully regular~~an evenly spaced (regular) event occurrence. We summarize this in a scalar test statistic E_S defined as the root mean squared deviation of the ~~timeseries $E(t)$ from the constant~~ time series $E(t_n)$ from the expectation values $\bar{E}$

$$E_S = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(E(t_n) - \bar{E})^2}, \tag{1}$$

where $\bar{E} = 6.367$ ~~, which~~ the average number of events per 20 kyr of the whole DO sequence. For ~~periodically occurring events~~ ~~,~~ events occurring periodically with a period significantly smaller that the window size, ~~we would expect~~ the test statistic E_S ~~to be~~ is close to zero. For completely randomly occurring events the test statistic will show a finite value, which depends on the moving window size relative to the ~~expected~~ average waiting time in between events. The same time-varying indicator has previously been used to complement a model comparison study aiming to quantify the influence of external forcing to conceptual models of the NGRIP ice core record (Mitsui and Crucifix, 2017). With this statistic we test whether the observed DO sequence departs significantly further from regularity as compared to what is expected by a random, uncorrelated event sequence. If this is ~~true~~the case, it would ~~point to~~ hint either at non-stationarity of the underlying process or a super-exponential event waiting time distribution. We consider the latter scenario to be less likely since no clear motivation for such a process exists.

While no significant correlation between duration of individual stadials and preceding or subsequent interstadial is observed (Pearson's $r = 0.04$ and $r = -0.15$, respectively), the data suggests long-term variations in stadial and interstadial duration distributions. If these variations are systematic for stadials and interstadials (i.e. correlated or anti-correlated) they should be detectable in the correlation of individual neighboring stadial/interstadial durations given a large enough sample size. However, due to the small sample size of events in this study and the broad distribution of event waiting times a correlation due to long-term trends is practically not observed. It is thus necessary to devise another time-varying indicator in order to capture additional detail in the structure of the DO sequence. When observing a given number of events in a time window, this may be either comprised of a combination of long stadials and short interstadials, or short stadials and long interstadials. This is not resolved in the statistic E_S. To capture this structure, we investigate the total portion of stadials within a moving window. Given the sum of the duration of all stadials $T_{st}(t)$ in a time window ~~T_{st}~~around a given midpoint in time t, the indicator is defined as ~~$P(t) = T_{st} \cdot (20\,\text{kyr})^{-1}$~~$P(t) = T_{st}(t) \cdot (20\,\text{kyr})^{-1}$. We summarize this indicator by a scalar test statistic P_S. It is defined as the root mean squared deviation from the ~~expected value~~ average value $\bar{P}$

$$P_S = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(P(t_n) - \bar{P})^2)}, \qquad (2)$$

where $\bar{P} = 0.461$ ~~, which~~ is the sum of all stadial durations divided by the total ~~length of the record~~duration of the last glacial period.

We now describe the models which are used to evaluate our hypotheses on the data using the test statistics described above. The first model used in our study models the process generating the sequence of warming events as a Poisson process with fixed rate parameter λ, i.e., we disregard the cooling transitions in between warming events. It is denoted as 'one-process model' hereafter. The inverse of the rate parameter corresponds to the average waiting time in between warming events. The Poisson process corresponds to a situation where there is no memory of the past and thus the probability for a transition is determined by λ and is independent of time. All information on climate stability is represented in the parameter λ. We set its value equal to the inverse of the empirically observed average waiting time in the full over the entire glacial record. This yields $\lambda = (3.141\ \mathrm{kyr})^{-1}$.

As a second model, labeled 'two-process model' hereafter, we propose two individual processes for generating warming transitions from stadials to interstadials and cooling transitions from interstadials back to stadials. Each is represented by a Poisson process with fixed rates a fixed rate λ_1 and λ_2, for warming and cooling, respectively. Again, the parameters are derived from the data by considering the empirical average residence times in stadials and interstadials, yielding $\lambda_1 = (1.477\ \mathrm{kyr})^{-1}$ and $\lambda_2 = (1.663\ \mathrm{kyr})^{-1}$. The model is different from the one previously introduced in that the sequence of warming transitions will not be a Poisson process, but a more regular one that is obtained from the sum of two independent processes. The probability distribution of waiting times T in between warming events is not exponential, but can be evaluated, yielding

$$P(t > T) = (\lambda_1 - \lambda_2)^{-1} \cdot (\lambda_1 e^{-\lambda_1 T - \lambda_2 T} - \lambda_2 e^{-\lambda_1 T}). \quad (3)$$

The average interstadial and stadial durations of the data seem to behave differently over the course of the glacial, as captured by our second test statistic. This motivates us to study whether this behavior is likely to be encountered by chance assuming randomness and independence of both up- and down-switcheswarming and cooling transitions.

As comparison to our hypothesis of stationary random processes, we consider the same models with time-varying rate parameters, which are given by a linear combination of two external climate factors: $\lambda = \lambda_0 + a_1 S(t) + a_2 I(t)$$\lambda = \tilde{\lambda} + aS(t) + bI(t)$. Firstly, we use a measure of incoming solar radiation at 65 deg North integrated over the summer $S(t)$ (Huybers, 2006). It is defined as the annual sum of the insolation on days exceeding an average of 350 W/m^2. Secondly, we use the LR04 ocean sediment record stack as proxy for global ice volume $I(t)$ (Lisiecki and Raymo, 2005). We note that, in contrast to insolation, global ice volume is not an external factor in the strict sense. However, it's its dominant variability is on longer timescales than DO events and most importantly it is obtained from an independent archivedata source. Time series of these forcings are shown in Fig. 2c,d. The models' parameters are chosen such that the time-varying statistics indicators are on average closest to those of the data. Specifically, by Monte Carlo simulation we generate many realizations for a fixed model parameter, compute time-varying statistics indicators for each realization and then construct an average curve. Finally, the root-mean-square deviation (RMSD) from this curve with respect to the time-varying data statistic is computed. For best fit, we search for the least RMSD on a grid in parameter space. This corresponds to a numerical calculation of the maximum likelihood fit to the observed data. The two-process model is fitted to both statistics $E(t)$ and $P(t)$ simultaneously, by minimizing the normalized sum of the errors $\mathrm{RMSD}_{E,P}$ to each of the statistics, defined as $\mathrm{RMSD}_{sum} = \mathrm{RMSD}_P / E_S + \mathrm{RMSD}_P / P_S$.

The hypothesis tests are performed in the following way. For a given model we simulate a large number of realizations, which are collections of subsequent events with the same total duration as the record (104 kyr). For each realization we calculate the time-varying quantity indicator of interest and the corresponding scalar test statistic. We then use the distribution of test statistics for a one-sided hypothesis test. The test simply counts how many test statistics in the ensemble are as large as or larger than the test statistic obtained from the data. Divided by the sample size, this yields a p-value, which estimates the probability of generating a random realization under the null hypothesis model that is at least as extreme ~~than~~ as the observed data. We can reject the null hypothesis at a confidence level α if the p-value is smaller than $1-\alpha$.

3 Results

The results of the hypothesis test on the stationary one- and two-process models are shown in Fig. 3a-c. The plots show distributions test statistics distributions of the respective null models and the corresponding test statistic value of the data. For the one-process model, the data test statistic lies well within the distribution, yielding a p-value of ~~0.16~~ $p_E = 0.16$, as seen in Fig. 3a. Thus, we cannot rejected the null hypothesis at a level >85%. This indicates that the variations in the timing of warming transitions are consistent with a stationary Poisson process, i.e., without invoking variations in the rate parameter. ~~Figures~~ Fig.'s 3b,c show the hypothesis tests of the two-process model, yielding low p-values p_E and p_P for both test statistics. The stationary two-process model is thus rejected by the hypothesis tests with both test statistics at high confidence >98%.

To better visualize the outcomes of the hypothesis tests, we show confidence bands for the time-varying indicators from our Monte Carlo simulations in Fig. 4. The indicator $E(t)$ of the data lies within the 95% point-wise confidence band of the one-process model. Moreover, this band can be calculated analytically from the fact that the probability distribution of observing k events in a time period T is given by the Poisson distribution $P(k,T) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$. The cumulative distribution thereof allows us to calculate the probabilities of observing the minimal and maximal number of events per 20 kyr found in the data indicator $E(t)$. We find the probability to observe 2 or less events is $P = 0.047$ and to observe 12 or more events $P = 0.030$. This confirms that we cannot exclude the possibility of observing only 2 events and as much as 12 events during 20 kyr of the record at 95% confidence. The 95% confidence band of $E(t)$ for the two-process model in Fig. 4b is narrower and does not include the most extreme parts of the data curve. The same holds for the indicator $P(t)$, thus confirming that the two-process model can be ruled out with high confidence as null model for the observed sequence of events.

In the following we present the hypothesis tests performed on the one- and two-process models forced with insolation $S(t)$ and ice volume $I(t)$, which are both scaled to zero mean and range 1. ~~Figure~~ Fig. 6 shows the time dependent transition rates as obtained from the parameter fit. For the one-process model we obtain

$$\lambda(t) = 0.32 + 0.43 \cdot S(t) + 0.82 \cdot I(t). \tag{4}$$

The best-fit two-process model has ~~stadial~~ warming transition rate $\lambda_1(t)$ and ~~interstadial~~ cooling transition rate $\lambda_2(t)$

$$\begin{aligned} \lambda_1(t) &= 0.97 + 1.60 \cdot S(t) - 0.57 \cdot I(t) \\ \lambda_2(t) &= 0.97 - 1.96 \cdot S(t) + 2.56 \cdot I(t). \end{aligned} \tag{5}$$

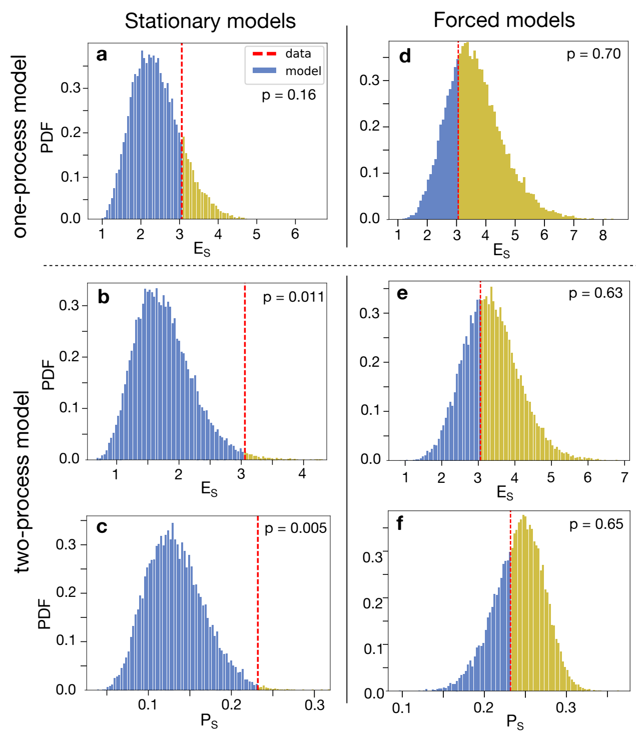

Figure 3. Empirical distributions from Monte Carlo simulation of the test statistics for the stationary (a)-(c) and the forced null model (d)-(f). The test statistics E_S of the one-process models are shown in panels (a) and (d). The test statistics E_S and P_S of the two-process models are shown in panels (b) and (e), and (c) and (f), respectively. The position of the data test statistic within the distribution is marked in red and determines the p-value of the hypothesis tests.

The hypothesis tests for the fitted models are shown in Fig. 3d-f and yield high p-values, where the data statistic lies near the mode of the distributions. Note that we only measure the deviation of the time-varying statistics from a constant average value. Thus, the statistical test is not targeted at evaluating the fit to the data, but merely at probing whether the fluctuations ~~in~~ over time of the indicators are of the right magnitude. Goodness-of-fit can be seen by means of the confidence bands and mean of the time-varying indicators, as shown in Fig. 7. For both models, the mean indicators lie close to the data curves, which consequently lie within 95% confidence bands. We summarize all model parameters, hypothesis test results and goodness-of-fit values in the table in Fig. 5.

We additionally report how the goodness-of-fit of the forced models changes when using only partial forcing and thus a reduced number of parameters. When forcing the one-process model with both ice volume and insolation, we yield a RMSD of the model mean $E(t)$ from the data curve of 1.42. Forcing with ice volume (insolation) only yields a best-fit RMSD of 1.64 (3.00). As baseline comparison, the RMSD from the unforced model to the data curve is equal to E_S, i.e. has a RMSD of 3.05. The model forced with ice volume fits the data only marginally worse than the model with both forcings and for comparison

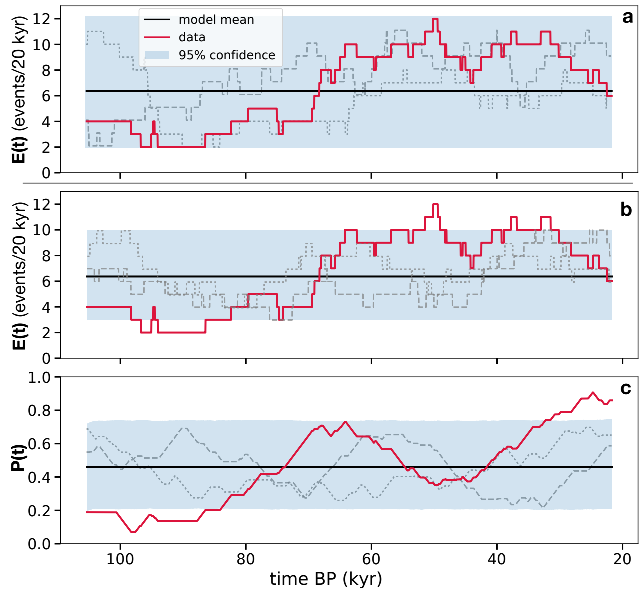

Figure 4. Point-wise 95% confidence bands and model mean (~~blue~~ black line) for the time-varying indicators $E(t)$ and $P(t)$ from Monte Carlo simulations. (a) $E(t)$ for the stationary one-process model. (b) and (c), $E(t)$ and $P(t)$ for the stationary two-process model. The indicators for the data are shown in red and two typical model realizations are shown in gray.

Model	Parameters	p-Value	Goodness-of-Fit
Stationary one-process	$\lambda = 0.32$	$p_E = 0.16$	$\text{RMSD} = 3.05$
Stationary two-process	$\lambda_1 = 0.68$ $\lambda_2 = 0.60$	$p_E = 0.011$ $p_P = 0.005$	$\text{RMSD}_{sum} = 2.0$
Non-stat. one-process	$\tilde{\lambda} = 0.32$ $a = 0.43$, $b = 0.82$	$p_E = 0.70$	$\text{RMSD} = 1.42$
Non-stat. two-process	$\tilde{\lambda}_1 = 0.97$, $\tilde{\lambda}_2 = 0.97$ $a_1 = 1.60$, $b_1 = -0.57$ $a_2 = -1.96$, $b_2 = 2.56$	$p_E = 0.63$ $p_P = 0.65$	$\text{RMSD}_{sum} = 0.59$

Figure 5. Summary of model parameters, hypothesis test results and Goodness-of-Fit of the mean model time-varying indicators with respect to the data.

we show the mean time-varying indicator $E(t)$ for this model in Fig. 7a with a green dashed curve. For the two-process model, we considered all combinations where both ~~stadial and interstadial~~ warming and cooling processes are only forced by either ice volume or insolation. Goodness-of-fit in the two-process model is given by the sum of errors of both indicators RMSD_{sum}. For the best-fit model with full forcing we find $\text{RMSD}_{sum} = 0.59$, whereas the baseline of an unforced model gives $\text{RMSD}_{sum} = 2.0$. Forcing both ~~interstadial and stadial~~ warming and cooling processes with ice volume (insolation) only yields a best-fit of $\text{RMSD}_{sum} = 0.90$ (1.60). Using insolation for the stadials warming transitions and ice volume forcing for the interstadials cooling transitions yields $\text{RMSD}_{sum} = 0.68$, while the converse choice yields $\text{RMSD}_{sum} = 1.68$. Thus, the only reduced two-process model yielding a comparable goodness-of-fit compared to the model with full forcing is the model with stadial insolation and interstadial insolation forcing on warming transitions and ice volume forcing . on cooling transitions. It is defined by

$$\lambda_1(t) = 0.81 + 1.54 \cdot S(t) \; \lambda_2(t) = 0.80 + 2.39 \cdot I(t). \tag{6}$$

We show the mean time-varying indicators for this model in Fig. 7b,c with a green dashed curve.

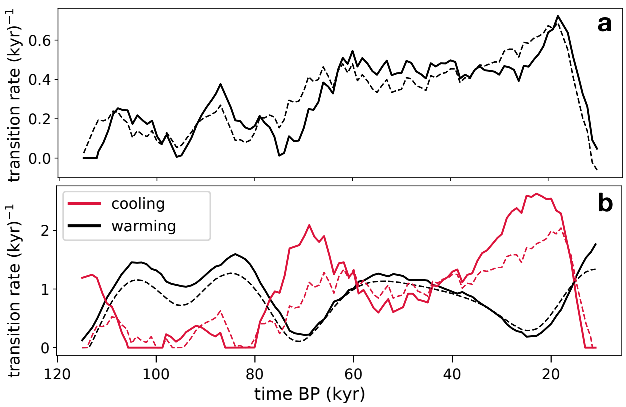

Figure 6. Time-varying transition rate parameters $\lambda(t)$ of the best-fit one-process $\lambda(t)$ (a) and two-process (b) models $\lambda_1(t)$ and $\lambda_2(t)$, as well as of the reduced models (dashed lines).

4 Discussion and Conclusions

Our first result considers only the warming events. While the distribution of waiting times in between warming events is well modeled by an exponential distribution (not shown here), we show here that the number of events in a moving window of 20 kyr (and thus the mean waiting time) is clearly changing over time, but no more than would be expected from a realization of a stationary Poisson process. Thus, if there is a unique process giving rise to the warming transitions, it need not be changing over time due to external factors. Although the description of DO events solely by the timing of the abrupt warming is very simplistic, we still think it is a useful result since the abrupt warmings are the most robust feature in ice core and other proxy records and are commonly used to assess synchronicity and pacing of abrupt climate change in the last glacial.

The second result indicates, however, limits to the stationarity in the sequence of events as we increase the detail of description. Assuming two independent processes giving rise to transitions from stadials to interstadials and vice versa, the null hypothesis of stationarity can be rejected with both our statistics. Specifically, both the variations in over time of the number

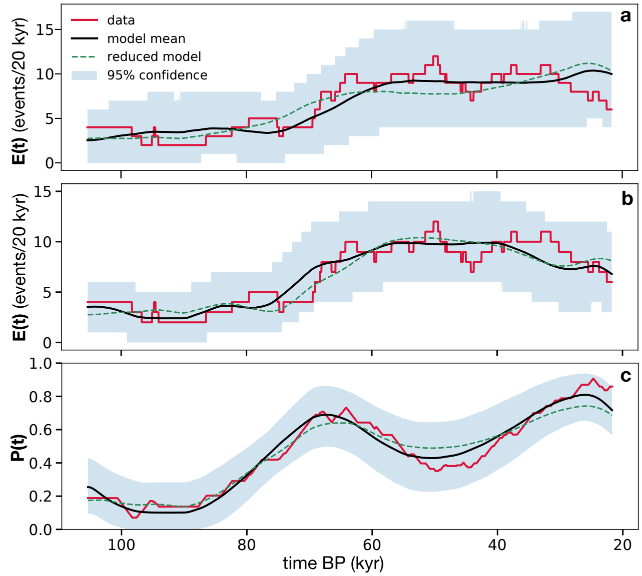

Figure 7. Point-wise 95% confidence bands and model mean (~~blue~~ black curve) for the time-varying indicators $E(t)$ and $P(t)$ from Monte Carlo simulations. (a) $E(t)$ for the best-fit non-stationary one-process model with full forcing. (b) and (c), $E(t)$ and $P(t)$ for the best-fit non-stationary two-process model with full forcing. The indicators for the data are shown in red. The green dashed line indicates the best-fit model mean curves for the respective models with reduced forcing, as described in the main text.

of warming events and the relative durations of stadials and interstadials are too large to be consistent with our two-process model using constant parameters. This model gives rise to a more regular sequence of warming events, compared to the one-process model. This is because one DO cycle is the sum of two independent processes and thus its duration ~~is not distributed exponentially~~does not follow an exponential distribution (coefficient of variation $\mathrm{CV} = 1.0$), but Eq. 3, which is less dispersed ($\mathrm{CV} = 0.708$). In the limiting case of a DO cycle comprised of a very large ~~number of independent~~ sequence of N independent and stationary processes, one finds a Gaussian distribution of waiting times with decreasing variance as N grows. This would then correspond to an almost evenly-spaced sequence of events, which is not supported by the observations.

Next, we investigated improvements of the consistency of the models with the data by allowing their parameters to vary over time as ~~a~~ linear combination of two climate forcings. Choosing the best fit linear combination of forcings, we found the average time varying indicators of both models to match very well to the data curve. Thus, whereas the data was seen as a rather out-lying realization that is consistent with a one-process model but not with a two-process model, when introducing forcings it becomes the expected behavior of the models. The goodness-of-fit follows from the correlation of the time-varying indicators and the forcings, which can be seen in Fig. 2. For the ice volume proxy we find a Pearson correlation of $r^2 = 0.78$ with $P(t)$ and $r^2 = 0.58$ with $E(t)$. We assess significance of this correlation by fitting an AR(1) process to the linearly detrended ice volume and conduct a hypothesis test yielding a correlation with $P(t)$ of 0.33 with a p-value of $p = 0.035$ and thus significance at 95% confidence. In contrast, the correlation of the indicator $E(t)$ and ice volume does not go beyond the linear trend. We do not assess the significance of correlations of insolation with the time-varying indicators, since it is difficult to find a good null model in this case.

Finally, we discuss the importance of ice volume and insolation in the best-fit one- and two-process models. In the one-process model, Eq. ~~(4)~~ 4 shows that both increased insolation and ice volume lead to higher occurrence rates of events, with the contribution of ice volume approximately twice as large as that of insolation. We note that the decrease in DO activity towards the Last Glacial Maximum is not captured by the model because the ice volume forcing is dominating. The individual contributions in the best-fit model do not completely capture the importance of the two forcings. There are directions in the likelihood landscape of parameters which are very flat. As a result, we found that the best-fit model with only ice volume forcing yields a fit only marginally worse than the best-fit model with both forcings. We thus conclude that ice volume is clearly the more important control on the sequence of warming events. This is consistent with the findings ~~of~~ in Mitsui and Crucifix (2017), where Bayesian model selection criteria show that global ice volume is a more important forcing than insolation in stochastic dynamical systems as models for Greenland ice core records.

In the two-process model the ~~transition rates in stadials and interstadials~~ warming and cooling transition rates are influenced by the forcing in opposite ways, as can be seen from Eq. ~~(6): Transitions~~ 6: Warming transitions from stadials to interstadials become more likely for higher insolation and lower ice volume, and vice versa for cooling transitions from interstadials to stadials. For ~~transitions from stadials~~warming transitions, the contribution of ~~ice volume~~ insolation is slightly larger than that of ice volume. The ~~interstadial~~ cooling transition rate is dominated by insolation, which contributes three times more than ~~insolation~~ice volume. With this model, the overall trend of mean waiting times in between warming events and of the stadial abundance is well captured, including the decrease in activity towards the LGM. Similar to the one-process model, we found a more parsimonious model which fits the data almost as well as the best-fit model with full forcing. This model uses only insolation forcing for stadials and ice volume forcing for interstadials, which complements the analysis of the individual contributions in the fully forced model. We thus hypothesize that based on our study there is evidence for insolation control on average stadial duration and ice volume control on average interstadial duration. This finding could hint at two distinct mechanisms responsible for transitions in between regimes.

An exhaustive investigation of whether ~~aforementioned finding and~~ our model description ~~is~~ and subsequent findings are consistent with governing mechanisms for DO-type variability inferred from detailed data and realistic model studies is ~~out of~~ beyond the scope of this paper. Nevertheless we conclude the discussion with some interpretations which are more speculative in nature. We begin with insolation control on stadial duration. Boreal summer insolation might influence the occurrence frequency of warming transitions by modulating the ice-ocean albedo feedback, which amplifies break-up or export of larger areas of sea ice. Sea ice decrease could subsequently cause rapid warming through ~~release of~~ subsurface ocean heat release (Dokken et al., 2013). Initial openings of the sea ice cover might be created by wind stress. Evidence for stochastic wind stress forcing and subsequent sea ice changes have been reported in unforced model studies of rapid climate transitions (Drijfhout et al., 2013; Kleppin et al., 2015). To explain global ice volume control on interstadial duration we invoke different influences on the strength and stability of the interstadial (~~warm~~strong) mode of the Atlantic Meridional Overturning Circulation (AMOC). If we consider global ice volume as an indicator of mean global climate, we find consistency with coupled climate simulations that show correlation of the stability of the ~~warm~~ strong AMOC branch to freshwater hosing and mean climate state (Kawamura et al., 2017). We furthermore note the study ~~of~~ in Buizert and Schmittner (2015), where a correlation of individual interstadial duration and Antarctic temperatures from ice cores is established and explained by influences of Southern Ocean processes on the strength and stability of the AMOC. Given the strong similarity of the global ice volume record and Antarctic ice core records on longer time scales, this is closely related to our findings. We finally note that in our model description, the trigger for warmings and coolings is stochastic and thus different from near-periodic DO cycles (Peltier and Vettoretti, 2014).

In conclusion, we show that the long-term variations in DO warming event frequency, often described as millennial climate activity, is consistent with a memory-less stationary random process. From the data at hand we cannot exclude the possibility that the long-term variations have occurred by chance. If we however divide a DO cycle into two independent processes governing warming and cooling, this is not true anymore and significant time-varying structure is detected. We thus propose a model that incorporates long-term variations through forcing of the parameters with external climate factors. We find good agreement with the data in a model where the mean duration of interstadial phases of the DO cycle are controlled by global ice volume and the stadial phases by boreal summer insolation. This finding can help to distinguish in between different mechanisms that have been proposed to cause DO events.

Competing interests. The authors declare no competing interests.

Acknowledgements. This project has received funding from the European Union's Horizon 2020 research and innovation Programme under the Marie Sklodowska-Curie grant agreement No 643073.

References

Alley, R. B., Anandakrishnan, S., and Jung, P.: Stochastic Resonance in the North Atlantic, Paleoceanography, 16, 190–198, 2001.

Braun, H., Christl, M., Rahmstorf, S., Ganopolski, A., Mangini, A., Kubatzki, C., Roth, K., and Kromer, B.: Possible solar origin of the 1,470-year glacial climate cycle demonstrated in a coupled model, Nature, 438, 208–211, 2005.

Braun, H., Ganopolski, A., Christl, M., and Chialvo, D. R.: A simple conceptual model of abrupt glacial climate events, Nonlin. Processes Geophys., 14, 709–721, 2007.

Braun, H., Ditlevsen, P. D., Kurths, J., and Mudelsee, M.: Limitations of red noise in analysing Dansgaard-Oeschger events, Clim. Past, 6, 85–92, 2010.

Buizert, C. and Schmittner, A.: Southern Ocean control of glacial AMOC stability and Dansgaard-Oeschger interstadial duration, Paleoceanography, 30, 1595–1612, 2015.

Dansgaard, W. et al.: Evidence for general instability of past climate from a 250-kyr ice-core record, Nature, 364, 218, 1993.

Ditlevsen, P. D., Kristensen, M. S., and Andersen, K. K.: The Recurrence Time of Dansgaard–Oeschger Events and Limits on the Possible Periodic Component, J. Climate, 18, 2594–2603, 2005.

Ditlevsen, P. D., Andersen, K. K., and Svensson, A.: The DO-climate events are probably noise induced: statistical investigation of the claimed 1470 years cycle, Clim. Past, 3, 129–134, 2007.

Dokken, T. M., Nisancioglu, K. H., Li, C., Battisti, D. S., and Kissel, C.: Dansgaard-Oeschger cycles: Interactions between ocean and sea ice intrinsic to the Nordic seas, Paleoceanography, 28, 491–502, 2013.

Drijfhout, S., Gleeson, E., Dijkstra, H. A., and Livina, V.: Spontaneous abrupt climate change due to an atmospheric blocking–sea-ice–ocean feedback in an unforced climate model simulation, PNAS, 110, 19 713–19 718, 2013.

Grootes, P. M. and Stuiver, M.: Oxygen 18/16 variability in Greenland snow and ice with 10^{-3} to 10^{5}-year time resolution, J. Geoph. Research, 102, 26,455, 1997.

Huybers, P.: Early Pleistocene Glacial Cycles and the Integrated Summer Insolation Forcing, Science, 313, 508–511, 2006.

Kawamura, K. et al.: State dependence of climatic instability over the past 720,000 years from Antarctic ice cores and climate modeling, Sci. Adv., 3, e1600 446, 2017.

Kleppin, H., Jochum, M., Otto-Bliesner, B., Shields, C. A., and Yeager, S.: Stochastic Atmospheric Forcing as a Cause of Greenland Climate Transitions, J. Climate, 28, 7741–7763, 2015.

Lisiecki, L. E. and Raymo, M. E.: A Pliocene-Pleistocene stack of 57 globally distributed benthic δ^{18}O records, Paleoceanography, 20, PA1003, 2005.

Mitsui, T. and Crucifix, M.: Influence of external forcings on abrupt millennialscale climate changes: a statistical modelling study, Clim. Dyn., 48, 2729, 2017.

Peltier, W. R. and Vettoretti, G.: Dansgaard-Oeschger oscillations predicted in a comprehensive model of glacial climate: A “kicked” salt oscillator in the Atlantic, Geophys. Res. Lett., 41, 7306–7313, 2014.

Rahmstorf, S.: Timing of abrupt climate change: A precise clock, Geophys. Res. Lett., 30, 1510, 2003.

Rasmussen, S. O. et al.: A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy, Quat. Sc. Rev., 106, 14–28, 2014.

Schulz, M.: On the 1470-year pacing of Dansgaard-Oeschger warm events, Paleoceanography, 17, 4–1–4–9, 2002.

Svensson, A. et al.: The Greenland Ice Core Chronology 2005, 15–42 ka. Part 2: comparison to other records, Quat. Sc. Rev., 25, 3258–3267, 2006.

Timmermann, A., Gildor, H., Schulz, M., and Tziperman, E.: Coherent Resonant Millennial-Scale Climate Oscillations Triggered by Massive Meltwater Pulses, J. Climate, 16, 2569–2585, 2003.