Response to Reviewer #1

We thank the reviewer for his/her evaluation of the manuscript and helpful comments. Below, we reply to all comments of the reviewer (red font color).

1. […] aim is to obtain mechanistic understanding of the DO cycles. This is not true in my opinion, since only statistical features are reported. These features can be used for benchmarking modelling studies testing different mechanisms, but this is not done in this paper.

We agree that this is not the best wording to describe the goal of the paper. We did not mean "mechanistic" in the sense of "physical mechanism", but rather "empirical", indicating how the properties of the cycles develop over time, as a function of previous cycles and forcings. We use the word "empricial" instead in the revised manuscript.

2. […] remove the word "causal" [...]. Only statistical similarities are tested, and no conditioning is performed to infer conditional dependencies. Also, no dynamical models are used, which could provide some hints at actual causality.

We agree that the word "causal" is problematic to use in our context and removed/replaced it throughout the manuscript.

3. I don't think that previous work on the DO cycles is sufficiently recognized by the authors. […]

We agree that a paragraph summarizing previous hypotheses is helpful and included one at the beginning of the introduction.

4. […] ice volume […] strong influence on the interstadial durations; this observation has, however, been made previously: Mitsui and Crucifix (Clim Dyn 2017) show from a statistical point of view that including this forcing is supported by the data, and Boers et al. (PNAS 2018) use it explicitly to infer the interstadial cooling rate during interstadials.

We thank the reviewer for pointing this out. We are aware of these studies, but we don't discuss them in the manuscript, because we do not actually argue for a correlation of the interstadial durations and ice volume. Even though this is not discussed in the manuscript, we find that for the whole data set, the correlation is not significant at 95% (can be seen in Fig. 5), and the influence of ice volume on interstadial durations is mostly due to the long interstadials 23.1 and 21.1 occurring at low ice volume. This fact is then further obscured by the short 23.2 and 21.2 happening during the same period.

In the paper, we are presenting results on the cooling rates as a predictor of the interstadial durations, a connection of the cooling rates and forcings (such as ice volume), referring to earlier work (Schulz 2002), and argue that CO_2 is actually a better predictor than ice volume.

5. The ultimate goal of this study is to provide the statistical basis for discriminating between different mechanisms to explain the DO events, but comparison of different mechanisms is not performed. Do the statistical features you extract give some hints at which of the prominent hypotheses listed above (point 3) are more likely? It would be nice to include at least a discussion on this at the end, as it is somewhat promised in the beginning.

We agree that this is desirable and reworked large parts of the Discussions Section in order to establish a stronger connection of our work and leading hypotheses or model experiments concerning DO events. To test our results in more detail in future studies, more models, or at least more runs of existing models, under different forcing scenarios are needed.

Response to Reviewer #2

We thank the reviewer for his/her careful evaluation of the manuscript, which helped to significantly improve the manuscript. Below, we reply to all comments of the reviewer (red font color), while splitting some longer comments into multiple paragraphs.

[…] after 31 pages of manuscript the reader has not really learned anything that was not already described in the literature. The authors confirm the 17-yr old result by Schulz (2002), briefly revisit the result by Buizert and Schmittner (2015), and confirm some of their own conclusions from Lohmann and Ditlevsen (2018) using a new method.

It is obviously our duty to convey the novelty more clearly to the reader. We hope that this comes through better in the revised manuscript. We clarify this by a more precise interpretation of our results, as outlined in the following comments.

We do not confirm all results by Schulz (2002), but on the contrary we critically examine Schulz' results by first doing an actual statistical analysis of the relationship of cooling rates and interstadial durations. Secondly, we show the limitations in time of the ice volume forcing hypothesis, as well as statistical limitations relating to simple common linear trends. Furthermore, we derive that atmospheric CO_2 is a better predictor of the cooling rates.
Similarly, we only recover the result by Buizert and Schmittner (2015) when discarding outliers. The outliers themselves require further interpretation, which relates to the comments further below.

One new element (the idea that cooling rates control interstadial duration, a variation on the Schulz argument) is based on the fallacy that correlation proves causation.

We believe this is a misunderstanding of language, which should now be clear in the revised manuscript. We do, of course, not imply that correlation proves causation. We simply observe that the interstadial durations are predictable to a reasonable degree after several hundreds of years. The physical climate process causing this cannot be identified within our data analysis and the cooling rates are solely an indicator allowing us to measure this process. (See also comments to 2))

We will clarify this in all corresponding parts of the manuscript.

Here is a brief summary of what is in our opinion the novelty:

- We develop a model for objective characterization of the DO events observed in ice core records as noisy saw-tooth "shapes".
- We classify the DO events based on their consistency with the saw-tooth shape.

- We extract a variety of features based on the model parameters obtained from the record, and look for relationships in between features and forcings. The analysis is more extensive than any previous efforts and may be used as a reference for future research. It is (partly) summarized in the heat map of figure S4. To highlight this better, we have moved it to the main text (Sec. 4).
- The interstadial durations can be predicted reasonably well a few hundred years after GI onset. This is due to the gradual cooling being sufficiently linear and consistent over time. → Thus, a noise-induced scenario is unlikely for the interstadial-stadial transitions.
- In the younger half of the glacial, atmospheric CO2 is a good predictor of the cooling rates, whereas we cannot establish a significant driver for the older half.
- Even though there is correlation with the stadial levels and insolation, the stadial durations cannot be predicted accurately. When removing forcing influences, the variability remains exponential.
- We extract warming durations from the record and analyze the distribution and temporal evolution (versus forcings). A comparison with the theory of reactive paths for stochastic multi-stable system shows that this corroborates results from the stadial durations: The stadial-interstadial transition is consistent with a noise-induced scenario.
- The DO warming amplitudes are weakly predictable from the stadial levels. However, they are not significantly correlated with external forcings, such as ice volume as has been suggested previously.

1) In deciding which events are stadials the authors rely completely on R14. [….], the authors inherit all the assumptions and historical numbering conventions that are in that study.

DO cycles come along a wide spectrum of shapes and timings, and deciding what is a stadial […] vs. a cold sub-event […] is a very arbitrary choice.

We fully agree that the careful analysis and the classification made in R14 that could be debated. It is not the focus of our paper to challenge this classification, which we assume to be a reasonable basis for our analysis.

More specifically, in R14 a threshold is needed to decide which events are sub-events and which not. However, in any conceivable scenario this would be necessary to pick events out of smaller-scale background variability. The R14 classification is based on three ice cores and two proxies and from the point of view of Greenland climate seems justified. As such, it is “Greenland-centric”, which is discussed in the comments further below.

We mention these assumptions in the revised manuscript, and acknowledge that some of our results depend on them (Changes document p3 l28 – p4 l2 and p29 l2-14).

This leads to some strange results from Lohmann and Ditlevsen, such as stadials that lasts only 20 years (P18 L24). From a climatic perspective, such short cold periods are more likely to be cold sub-events rather than true stadials.

The 20 years of GS-24.2 only refer to the time where the proxy is stable. The overall break in between adjacent interstadials (including cooling and warming) is about 170 years, and certainly more significant in amplitude than any sub-event in the Greenland records.

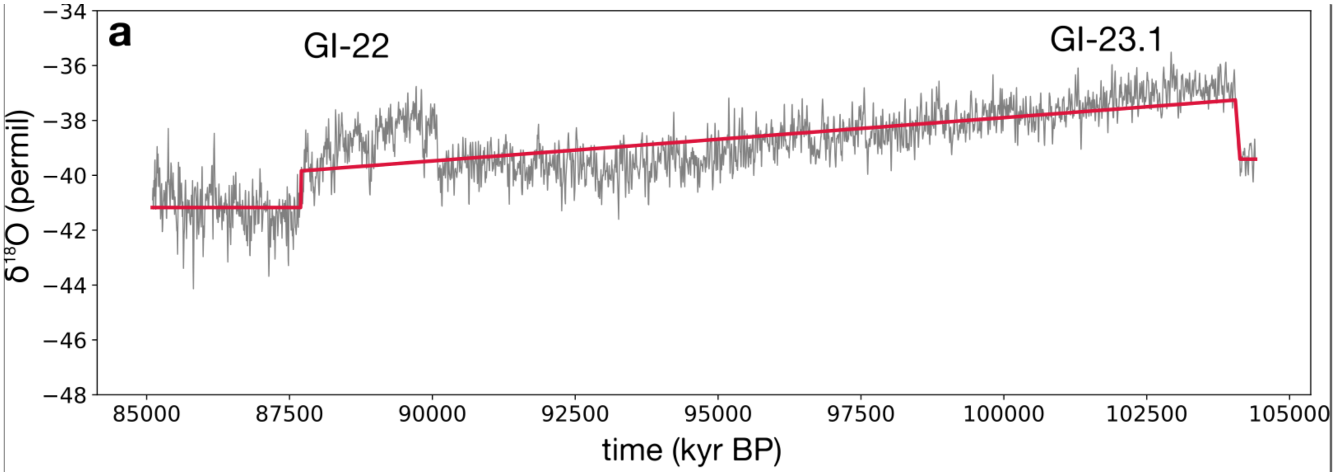
If driven by the AMOC changes, which our analysis does not assume a priori, the cycle of shutdown and resurgence would then be 170 years, which is close to the 250 years suggested by the reviewer. The numerical value of 20 years is unlikely to be important since we mostly use rank statistics in our analysis.

The cause of the cold sub-events [...] is of course not well known, but most likely they represent outburst floods (e.g. from proglacial lakes) that temporarily increase N-Atlantic sea ice cover and cool Greenland, the 8.2 ka event is the posterchild for such events. They occur most frequently during periods of ice sheet decay (MIS4-MIS3 and MIS2-MIS1 transitions) supporting this interpretation. When is a cold period a sub-event and when a true stadial? R14 applies some criterion of baseline separation; this makes GI-24.1b a “cold event” and GS24.2 a “stadial”, even though they appear nearly identical and GS24.2 may just be a larger cold event (bigger outburst flood?). In other "stadial" cases (GS 14, GS 23) baseline separation is not achieved and R14 calls them stadials only because they choose to adhere to historical DO numbering (based on older low-res cores). Many DO events have a clear warming period at their end (7a, 8a, 12a) – could DO 13 be one of those? The difference is arbitrary.

As above, this is more a discussion of the quality of the classification in R14, which we do not discuss, however, we have the following remarks:

In terms of Greenland ice cores, GI-24.1b and GS-24.2 are not nearly identical. Estimated form the high-resolution NGRIP d18O record, GS-24.2 lasts 170 years (when including warming and cooling periods) at an amplitude of ~3.8 permil, whereas GI-24.1b is 130 years at ~2.0 permil amplitude.

We agree that the case of GS-14 and GS-23 (labeled quasi-stadials by R14) is more debatable. First, if one compares GI-13 and GI-22 with the other “rebound” type events, they are simply more pronounced (longer, slightly larger amplitude). Second, we need to include them because only like this we can get a satisfactory saw-tooth fit throughout the record, and representative features for each event.
This is very pronounced for the case of GS-23 (see picture below, where we merge GI-22 and GI-23.1 in the fit), and to a lesser extent for GS-14.



We do not claim that our fit is a method to decide whether they should be considered individual events, but in order for our analysis to be meaningful, we require a reasonable fit.

While we admit that the choice of GS-14 and GS-23 might go one or the other way, since we do statistical robustness tests, our results do not rely on the presence or absence individual stadials. In fact, we performed a semi-complete analysis for a fit without using GS-23 (not discussed in the manuscript), yielding qualitatively unchanged results.

We appreciate the reviewers comment and discuss these issues in Section 2.1 of the revised manuscript (Changes document p3 l28 – p4 l2).

DO 24, 17, 16, 15 each consist of 2 DO events in the Greenland-centric R14, but in for example the Hulu speleothem record and the Iberian margin SST these events are recorded as single (and not double) events. Could these be single events from the perspective of the overturning circulation, but separated events in Greenland due to regional effects (like freshwater)?

These sub-events are always the outliers in the scatter plots (5a, 6a, 6b), making this an interesting point to consider. The authors note that “Buizert and Schmittner (2015) lump each of the interstadials 24, 23, 21, 17, 16, 15 and 2 together into one event, even though they are comprised of two events.” But are they two events? If we put 100 paleoclimate researchers in one room, my guess is that less than half would call GS 24.2 a true stadial (minimum requirement to support the claim that DO 24 is comprised of 2 events). The vote may be different for GS 15.2. I don’t want to claim here that they are 1 or 2 events, just that this depends completely on arbitrary definitions, and that reasonable people can disagree on the number of “true” DO cycles give that the cold sub-events are ubiquitous and can have a wide range of sizes.

We agree that it is still unclear whether successions of shorter events found in the Greenland ice cores are complete reorganizations of the large-scale climate system (such as AMOC shutdown and resurgence), in the same way as longer events are.

While the Iberian margin SST record is not high-resolution enough to resolve the shortest events, Asian monsoon speleothem records, including Hulu cave, now exist in sufficient resolution, and indeed resolve events such as GI-21.2, and the rapid successions of GI-17.2 to GI-15.1 (Cheng et al. 2016).

The same goes for Alpine speleothem records, which have a very detailed similarity to the Greenland ice cores. The event succession of GI-15-17 is seen (Moseley et al. 2014) as full-amplitude DO events, suggesting that the centennial variability is not regional to Greenland. Similarly, events 23.2 and 21.2 are identified (Boch et al. 2011).

Additionally, the short events are well recorded in methane records from Greenland and Antarctica (WAIS Antarctic record: Rhodes et al. 2015, Rhodes et al. 2017; Greenland NEEM record: Chappellaz et al. 2013).

Of course, we cannot argue in this way that chosen events are indeed the most prominent globally. However, this problem is very hard to address, and beyond the scope of this work. A more 'global' assessment of which centennial- to millennial-scale abrupt climate changes should be considered most important is difficult since different regions and proxies are sensitive to a different degree to the various parts of the climate system undergoing change, and thus will most likely highlight different events. For example, Asian monsoon proxies are much more sensitive to Heinrich events, so it is difficult to use them to study the statistics of DO events. An assessment including different types of proxies in different regions will require a (subjective?) weighing of some kind to extract the type of abrupt climate change one wishes to study. Furthermore, dating issues will need to be overcome.

Since the Greenland ice cores are in a region that is very sensitive to the North Atlantic climate they are ideally suited to study climate changes related to the AMOC.

Nevertheless, we agree with the reviewer that shorter events might be more specific to Greenland and could have a different trigger than longer events that are more clearly recorded in other archives. Simply due to their shortness it is not unlikely that they indeed do not represent global reorganization of the oceanic circulation or climate system.

Including the short events in the analysis is a way of identifying how they are different, in terms of their detailed features as well as in the context of forcing. Indeed, the fact that certain events show up as outliers might be because they are caused by something different than the majority of events.

We acknowledge the open problem of the cause and significance of the shorter DO events in the Discussion Section of the revised manuscript (Changes document p29 l2-14).

One could turn the question on its head, and argue that, based on the fact that the sub events of DO 24, 23, 21, 17, 16, 15 consistently show up as outliers in scatter plots, and fit the climatic trends when lumped together, they are actually single (and not double) events from the perspective of the global climate system and oceanic overturning circulation.

We agree that the evolution of the properties of DO cycles over time and their relationship to external influences becomes clearer when certain shorter Greenland events are discarded. Still, we do not believe that if discarding outliers (or combining short events) the correlation with external forcings or climatic trends is strong enough to warrant such a reasoning.

One way to make the present paper more interesting is to try out different definitions of a stadial (beyond adopting R14), and see what definition may minimize scatter in the plots. What do other climate archives like speleothems suggest these events looked like? Would a duration threshold (e.g.

cold period longer than 300 years) be a better way to define a stadial? There are probably good climatic reasons why DO timescales are linked to global climate markers (as argued by e.g. Schulz 2002 and Buizert and Schmittner 2015), so an approach of finding an objective (multi-proxy?) stadial definition that minimizes scatter is justified in my view.

We do not believe that it is good to put a constraint of stadial durations, such as 300 years, within our framework. We treat the record without preconceived notions of the underlying driver, i.e., AMOC changes or hypothesized outburst floods for short events, which has both merits and shortcomings.

In terms of multiple hypothesis testing it is risky to try out many different definitions of stadials and correlate them with many different global climate markers, in order to find a match. We would furthermore need to decide which features would we like to be explained by the global climate markers, which is unclear a priori.

Finally, doing a full analysis as presented in the paper with many different stadial definitions is unfortunately beyond our present capacity and is left for future work.

I want to emphasize this is not a critique of R14 itself, which seeks to provide a consistent nomenclature for events and has succeeded in doing so. The problem arises when R14 is mistaken for an objective and meaningful decision on which events are "true" stadials – which was never the aim of that study. The author's algorithm has the liberty to change the timing of transitions, but not the number of events, and so does not challenge the R14 definitions.

To summarize an excessively long comment, the approach by the authors has a fundamental and tenuous assumptions that is neither acknowledged nor examined. They interpret the R14 beyond its intended use as a climatically meaningful distinction of which cold events are stadials and which are not. While the shape fitting is done with much mathematical rigor, these underlying assumptions will always limit the validity of their conclusions. Trying different definitions for stadials would be an interesting research direction, as well as different proxy archives.

In our judgment, solving the problem of unique classification of DO events will require a much better globally extended set of high resolution, well dated climate records, which we do not have today and probably also a realistic theory of their cause. We consider this as a fundamental problem, but somehow disconnected from our results, which will hopefully be a steppingstone towards solving the puzzle of the DO events.

In the revised manuscripts, the assumption of a predefined set of DO events is acknowledged and discussed (Changes document p3 l28 – p4 l2 and p29 l2-14). An extension of our algorithm to also objectively detect events would be less powerful than the assessment of R14, because it is based on 3 ice cores and 2 proxies, which is why we did not design the algorithm in such a way.

We agree that the use of other proxy archives is very promising. However, within the framework of our method these are difficult to incorporate, since they do not show a DO signal that is consistent enough, and has comparable resolution and age control.

2)

One of the main conclusions of the paper is that cooling rates "control" the interstadial durations (P14 L13, P16 L21 and elsewhere). They are only correlated, which does not prove any kind of causation or control; both could be controlled by a third parameter such as AMOC strength, CO_2 or SH temperature. This correlation was discovered by Schulz (2002), who actually argues for an ice volume control.

The authors further suggest that the “interstadial duration is determined as soon as the rate is established” and that the duration is “determined” a few hundred years after the onset. However, if both cooling rate and duration are controlled by a third parameter (like AMOC strength), this interpretation is strange. In my view, Fig. 5b simply reflects the amount of data needed to determine the slope in a noisy time series, and is not some time scale on which the interstadial duration is somehow “determined” by the coupled climate system.

We have tried to clarify this misunderstanding in the manuscript. If we may be a little impertinent: If you are walking to a town 10 km away at a pace of five km/h, you’ll be there in two hours. I can predict that as soon as I know your pace. That may take me five minutes to do, but that implies neither that five minutes has any significance for the prediction (two hours), nor whether your speed and the distance to town are controlled by any third parameter. The only “control” is that the speed -kept constant- controls the arrival time.

Thus, the word “control” was meant in the context of our statistical analysis. As elaborated in the manuscript, each interstadial duration is defined by the cooling amplitude divided by the rate. If either the amplitudes or the rates have a clearly dominant variability they also effectively “control” the durations of the interstadials. We show that this is the case for the rates.
We rewrote parts of the manuscript to clarify this (Changes document p17 l24-29, p18 l11-12 and p27 l1-8).

We agree with the reviewer that a mutual control by a third parameter is likely, which is also what we are investigating extensively in the paper. In fact, we think it is not only likely but necessarily the case. The gradual cooling rate is merely an indicator of the global climate reorganizing on a specific timescale. The termination of the interstadial, and thus its observed duration, is the final consequence of it. Both durations and cooling rate are driven by the same process, even though the actual interstadial termination might be governed by additional processes that influence the climatic threshold.
Our empirical finding is that this timescale seems to be consistently expressed via the linear cooling already early on in the glacial. As a result, the end of the interstadial is already anticipated and can be predicted.
We thus obtain the novel result that the interstadial-stadial transition is not purely noise-induced.

We agree with the reviewer that our result of the time elapsed until the durations can be predicted (150y-350y) is influenced by the amount of data needed to determine the slope. However, this time elapsed is only relevant because the cooling trend is actually “linear enough”, which is not obvious a priori and which is what we implicitly test for.

On a side note, I am uncomfortable with language of transitions being “determined” thousands of years in advance. Interstadial terminations occur because at that time interactions between components of the climate system are favorable for such a transition (including “noise” components). The climate system is not a decision-making entity that plans things centuries to millennia ahead. The word “predicted” seems more appropriate. So please revise.

We are unsure what is the basis for this argumentation.
Concerning the climate system planning ahead: Some models would suggest that the DO cycles are self-sustained oscillations, with a period potentially slowly modulated by external forcing. In this case,

interstadial terminations would indeed be determined centuries/millennia in advance due to the periodicity.
We do not argue that this scenario is the case, but only want to point out that it is conceivable that climate transitions are determined ahead due to slow, deterministic processes that may be measured a long time before the actual transition.

Our use of the word “determined” is simply based on the observation that 1) the gradual cooling is sufficiently linear and 2) the variability of the rates is larger than of the amplitudes. As a result, the interstadials are reasonably well determined already a few hundred years after interstadial onset.

If the authors want to argue for their mechanism (cooling rate controls duration) they will have to provide a meaningful climatic pathway for such control, which is currently lacking. At the very least they have to clarify all the language suggesting causation.

As above, we hope we can clarify this with better terminology and language in the revised manuscript. We do not aim to find a mechanism by which the cooling rates control the durations. It simply follows from our analysis that there is a time scale in the climate system, which manifests itself in the rate of the roughly linear interstadial cooling, which is established soon after interstadial onset, and which approximately predicts the duration of the interstadial.

We do look for an actual control among the external forcings and find that CO_2 is the best predictor for the younger half of the glacial. However, whether CO_2 really acts as a forcing remains to be seen as it also shows millennial-scale variability in line with the (Antarctic) DO cycle, and is likely a response to it. The mechanism by which CO_2 influences interstadials depends on what drives the DO cycles in the first place, which is not known a priori and is hard to establish from our analysis.

3) The discussion section is basically a lengthy summary of the preceding chapters, which is not the function of a discussions section. This will need to be rewritten. There are many caveats and assumptions that need to be addressed, and the work can be placed in a broader paleoclimatic context.

We agree and rewrote large parts of this section.

4) In earlier work, Ditlevsen suggested that the DO transitions are purely noise-driven – others have probably made that suggestion also (Ditlevsen et al., 2007; Ditlevsen et al., 2005). Given that event durations are clearly correlated to global climate parameters, is that still your view?

In this study, we regard the warming and cooling transitions separately. Our analysis shows that neither the stadial nor the interstadial durations can be clearly matched unambiguously with a climate forcing over the entire glacial. Only the general trends may be explained by forcing, an exception being evidence for good correlation of CO_2 and the interstadial cooling rates in the younger half of the glacial.

Conceptually, evidence for a modulation of the occurrence frequency of events over time does not exclude that the individual transitions are purely noise-driven. The modulation only affects the expected value of the occurrence frequency.
Nevertheless, this study shows that the interstadial-stadial transition appears to be less consistent with a noise-induced process, due to their predictability.

On the other hand, we provide new independent evidence for a noise-induced scenario for the stadial-interstadial transition by comparing the DO warming durations to reactive trajectories.
Of course, this is not the end of the story since new data/analysis could find evidence to the contrary.

5) The referencing of published material is very minimal for a paper of this length on a topic that is so extensively written on. Much is in fact known about the DO cycle (despite the author's claims to the contrary). Marine sediment data clearly show a link to the Atlantic ocean circulation (see e.g. review by Lynch-Stieglitz, 2017), and climate modeling studies clearly implicate sea ice cover in the North-Atlantic. Many remote teleconnections have been clearly described, and several drivers have been proposed. Also, many more papers have used similar statistical techniques on the DO cycle that should be referenced.

We agree that a review of published hypotheses and model experiments is useful and added a paragraph at the beginning of the introduction. If the reviewer is aware of more papers with similar techniques that are relevant to our results, we are happy to include them.

6) The authors ignore Heinrich events, while it is commonly believed that H-events lengthen the stadials in which they occur by putting freshwater into the North Atlantic. Please discuss this in the stadial duration section.

We thank the reviewer for pointing this out. Indeed, we omitted a discussion of Heinrich events, because it is difficult to find manifestations of them in the Greenland records, and due to difficulties of timing them relative to Greenland ice cores. Nevertheless we did some statistical tests and included the following as a new sub-section (4.2.3) in the manuscript:

“Besides DO events, Heinrich events are the other major mode of millennial-scale climate variability during the last glacial period. These events correspond to massive discharges of ice rafted debris found in ocean sediment cores (Heinrich, 1988), with large climatic impacts that are well-documented in numerous proxy records at various locations. While their duration and timing needs to be better constrained, we follow Seierstad et al. (2014) for the temporal link of Heinrich events and the GICC05 chronology. This yields the set of Heinrich events H2, H3, H4, H5a, H5, H6, H7a, H7b and H8, which overlap with stadials 3, 5.2, 9, 13, 15.1, 18, 20, 21.1 and 22, respectively. Since some of these Heinrich events might be less established in the community, we also look at the reduced set of H2, H3, H4, H5 and H6.
We test whether these ’Heinrich stadials’ have significantly different properties than the remaining stadials, such as longer durations, by randomly sampling 9 stadials (5 for the reduced set) from the entire set without replacement and calculating the mean duration of this subset. We repeat many times until we can estimate the probability of trials yielding a higher mean duration than the actual set of ’Heinrich stadials’. If this is less than 5% (corresponding to $p=0.05$) we reject the hypothesis that ’Heinrich stadials’ have the same mean duration as the remaining stadials at 95% confidence. This test gives essentially the same results as a one-sided t-test.

For the full (reduced) set of Heinrich events we find $p = 0.028$ ($p = 0.022$). It is not obvious whether this should be considered significant in the sense of a hypothesis that Heinrich events prolong stadials. A better statistical test is needed, since if the events were to occurr randomly during the course of stadials they would naturally be found preferentially in longer stadials. We leave a resolution of this for upcoming work. Based on the idea that ’Heinrich stadials’ are colder than normal, we perform a test on the stadial levels, yielding p = 0.052 (p = 0.047). Again, this is probably not significant since Heinrich events mostly occur in the younger glacial half with generally lower levels. We can reject the notion that DO events following Heinrich events are 'stronger'. A statistical test on the DO warming amplitudes yields p = 0.102 (p = 0.472), whereas a test on the interstadial durations yields p = 0.403 (p = 0.583). This might depend on the precise timing of H3, which in our analysis precedes the especially weak GI-5.1."

7) The paper is very long and could be shortened substantially.

We tried our best, but believe that most of the content needs to remain in order to support our results and demonstrate statistical significance.

8) There is no data availability statement.

Will do in the final version.

9) Are the d18O data corrected for mean ocean d18O? This would of course influence the cooling rates of long interstadials.

The data are not corrected for mean ocean d18O. We agree that doing so will in principle influence the cooling rates. However, the changes in mean ocean d18O during the longest interstadial (GI-23.1) are only about 0.2 permil (Shackleton 2000; Waelbroeck et al. 2002), which is an order of magnitude smaller than the Greenland isotopic changes, and thus negligible for our analysis.

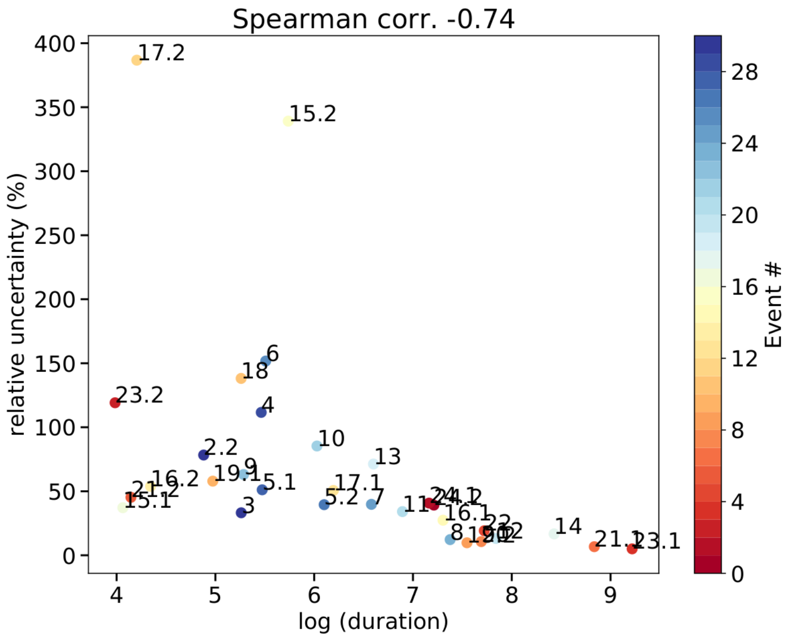
10) The work mostly just confirms earlier work. I would encourage the authors to think about ways to broaden or improve the scope, to reward the reader with something new.

With our response comment (specifically the beginning and 1) and 2)) and corresponding revisions to the manuscript, we hope to convince the reviewer that our paper does not merely confirm earlier work, but instead critically assesses it, and furthermore provides a number of new results.

11) The uncertainties in the fitting parameters are carefully estimated. Could they be listed in table 2? I imagine that for the very short interstadials (<400 yrs) things like the gradual cooling slope are not well constrained.

It is hard to list all uncertainties in this table, since we preferentially need two numbers to specify the uncertainties in terms of 1-sigma ranges. This would result in a very squeezed table.

Regarding the cooling slopes of shorter interstadials: Uncertainty in the rates needs to be assesses with care. We propose to look at the ratio of the 1-sigma range to the best fit value.
We find that indeed the cooling rates of the shorter interstadials are more uncertain than those of the longer interstadials. There is a significant negative correlation (Spearman r = -0.74) of the relative uncertainty in the rates and the duration, as expected due to the smaller amount of data available to estimate the slope for shorter events. See the following plot of relative uncertainty in the slope vs. duration.



There are two events that we indeed could say are not well constrained: GI-17.2 and 15.2. We also exclude these from our analysis investigating the predictability of interstadials. (Sec 4.1.3)

In the manuscript, error bars for the logarithmic cooling rates can be seen in Fig. 6 c/d. Taking into account the many orders of magnitude of the cooling rates, we conclude that apart from the two outliers (can be seen as largest error bars in the plot), the other cooling rates are well constrained, as can be seen by the error bars on the lgarithmic scale. Simply put, fast rates are fast even when taking into account the error, which in some cases can be more than 100% relative.

12) Why do you use a constant stadial level? Do the stadial levels resemble Antarctica, as suggested by (Barker and Knorr, 2007).

With few exceptions, the stadials do not show a clear trend, but rather fluctuate around a mean. For the purpose of estimating the stadial or interstadial onset times, using a constant stadial part works sufficiently good.
Looking into the finer details of trends within the stadials is certainly of interest, but was beyond the scope of our paper.

We have no reason to believe that the Greenland d18O stadials resemble the Antarctic record, as observed by Barker and Knorr for a detrended/dejumped Greenland dust record. The constant stadial levels do correlate with the Antarctic EDML record at stadial onset, however this is mostly due to common linear trend, as mentioned in Section 4.2.2. A more convincing correlation is found with insolation.

13) Have you tried including other records? The Ca record has much better signal to noise, allowing for better timing determination. I think marine and speleothem records have much to add to the problem also.

We have not tried to include records other than Greenland d18O records. The Greenland Ca record has a more complicated shape that is not as well captured by a piecewise-linear fit throughout the whole glacial.
We agree that marine and speleothem records are important to answer further questions relating to DO events. However, as mentioned in our comments to 2), they do not have such a consistent DO signal (which there are probably good reasons for). Furthermore, they are less ideal in terms of resolution and timescale. The layer-counted Greenland ice core records provide excellent relative time constraints allowing for the determination of stadial, interstadial, and abrupt warming durations.

P1L7: remove “mechanistic”. This is not attempted, in my view. Climate dynamics are not discussed.

Ok, we use the word “empricial” instead.

P1L14: “largely unknown”: A lot is known, the gist of which could easily be summarized in a few sentences.

We agree and added a new opening paragraph with some references to published hypotheses and modeling work.

P2L16: “we do not have to rely on any subjective choice of stadial and interstadial onset or levels”. This is a misrepresentation, in my view. The subjective choices were all made by R14. The algorithm does not have the ability to decide independently how many events there are, and whether individual cold periods are stadials or just cold sub-events.

We agree that the sentence as it is can be misleading.
It was meant to refer to choices of what constitutes an event, but to choices of a segmentation of the record if we would ramp-fit each event individually.
We reformulated and hope it is clearer now: “Thus, our results are not sensitive to subjective choices of cutting the record at predefined times before and after a transition.”

P2L24: “In contrast. . .state” But the transitions are not purely noise-driven, since the period durations are linked to global climate, no?

We indeed find some correlation of the stadial durations with insolation. However, since the correlation is not very strong, we interpret this as a modulation in time of the expected residence time in the stadial state, which does not mean that the transitions themselves are not noise driven. In the paper we present that even if one removes the trend due to forcing on the stadial durations, we still find an exponential distribution, as expected for noise-induced phenomena.

If one considers a slow modulation of the average residence time as part of the driver, then the term ‘purely’ noise-driven might be problematic.

P3L13: Do the short cold events influence the fitting?

Although they might influence the fitting in some minor way, you can visually verify that they are mostly 'ignored' by the linear fit, and the fit just follows the linear slope of the remaining parts of the interstadial. (see e.g. GI-24.1, GI-17.1, GI-16.1, GI-14, GI-13)

P4L29: Instead, we use . . . basin-hopping. This means very little to much of your readership. Please elaborate or leave out.

We agree and omit the term here. Basin-hopping is explained in Appendix A, and a citation is given.

P5L8: "climate features" should be "d18O features"

Ok, we leave out 'climate'.

P5L12: "mechanistic" I don't think this is the right word. Climate dynamics are barely discussed.

We agree and use the word 'empirical' instead.

P7L1: but you use 90% confidence, correct? How many false positives for 0.9?

Indeed we mostly consider correlations etc. at 90% for further testing, and at 90% there are 45 false positives, which we added in the revised manuscript. The p<0.1 mentioned in the manuscript is just a guideline of how we proceed with this large number of statistical tests. Note that we often also investigate correlations that are not significant 90% for the entire data set, as mentioned earlier in the manuscript.

P7 section 3: why is DO 1 omitted?

As mentioned in the manuscript, GS-2.1 is non-stationary and would influence both GI-2.1 and GI-1 in a way that many features, such as warming amplitude and durations, would need to be discarded as outliers for these events. The d18O values during this stadial rise higher for a long period of time than the maximum of the preceding GI-2.1.

Fig 2: can you comment on DO 23.1? You routine picks its termination 3 ka before R14 does.

From the NGRIP d18O and the notion of stadials being defined as roughly constant low values after a gradual cooling this seems justified. In NGRIP, GI-23.1 consists of roughly 10kyr of gradual cooling, followed by 3.5kyr of approximately constant values leading up to the warming of GI-22.
Since our analysis is only based on one ice core and proxy we however do not suggest that this is sufficient to argue for a definition of GI/GS-23.1, which is why we do not discuss this in the manuscript. Still, our analysis is unlikely to be sensitive to the precise parameters of a single event.

Section 4.1.1: Is there any conceivable mechanism by which cooling rates determine interstadial duration? Correlation and causation are falsely equated here.

We refer the reviewer to our previous comments on 2).

Our analysis suggests that the cooling rates and their associated timescale are a manifestation of a deterministic climate process that pre-determines the interstadial durations. From our analysis we cannot establish the nature of this climate process.
We adjusted the wording in this paragraph to not give the impression we equate correlation and causation.

P13 L13: What is lambda? Wasn't the cooling rate called s_2? Also specify whether you're talking about the gradual or fast cooling rate

We agree that this might cause confusion. It is a new definition that we added to make the notation more intuitive, along with D_I for the interstadial durations and A for the warming amplitudes. For consistency, we added these definitions to Table 1 in the revised manuscript.

P14L3: fixed cooling rate. . . You mean all interstadials would have the same cooling rate? We know this to be untrue. I do not understand this hypothetical scenario.

We are explaining the two limiting scenarios of either the amplitudes of the rates to be perfectly constant. Of course neither is true, but we test in the following which scenario is closer to the truth.

P14L8: "strong control". Correlation is not causation

We rephrased to the following: "Thus, from the perspective of linear interstadial cooling, the interstadial durations over the entire glacial are indeed largely governed by the cooling rates, in agreement with..."

P14L22: log-normal: is this meaningful since you only fit the tail of the log-normal distribution?

We are not only fitting the tail of the log-normal distribution. Maybe the reviewer refers to the AD test being tail-sensitive? In this case, we also use the Cramer-von Mises and Kolmogorov-Smirnov tests, which yield qualitatively unchanged results.
We mentioned this too far below in Section 4.3.1, but in the revised manuscript moved this to Section 2.4, where it is more approriate.

P15L10: "determined" change wording. The climate system does not plan ahead millennia. If you want to argue for such a mechanism you should at least provide a dynamical pathway, even if speculative.

Changed sentence to: "If the rates govern the durations much more so than the cooling amplitudes, then the durations can already be approximately predicted as soon as the rate is established, which might happen early in the interstadial."

P16 L19: Not necessarily. Both duration and cooling rate are controlled by heat transport and interactions within the climate system, and appear to correlate to a third parameter like CO_2, ice volume, or similar. Within a few hundred years you can detect the cooling trend within the noisy time series, and because of the correlation you can predict the interstadial duration with reasonable accuracy at that time. Nothing is "determined" a few hundred years into the event.

We refer the reviewer to our response to Comment 2) above.

P16L21: rewrite.

“Given the previous result, we investigate...”

P18L16: see comments above. Whether these are one or two events depends on ones definition of a stadial. There is no widely accepted definition, and R14 provides nomenclature only and is not the final authority on this matter. Other archives should also weigh in on this question, since Greenland may reflect regional effects.

We refer the reviewer to our previous comments above.
We now worded differently: “This disagreement comes from the fact that Buizert and Schmittner (2015) view each of the interstadials 24, 23, 21, 17, 16, 15 and 2 as one event, whereas we consider these as two events each, as suggested by the analysis of Rasmussen et al. (2014).”

P18L24: a 20yr stadial is probably not a stadial in most people’s definition. A time threshold (250 yrs?) may be of use in defining stadials. 250 yrs provides a rough timescale of Atlantic overturning (volume divided by rate), and makes some intuitive sense for that reason.

See our response to Comment 1)

P18L27: the data ARE consistent with. . . (data are plural).

Ok.

Fig 7a: all outliers are in MIS5. Is that relevant?

We agree that this might be relevant and it holds not only for Fig 7a, but for all models we use for the stadial durations, as mentioned in the manuscript.

Since the correlation of the remaining events is not very strong but “the models fit only the overall trend and leave unexplained variability on top of it”, we prefer not to interpret the fact that the outliers are in MIS5.

P19L6: common forcing makes most sense, right? Weak AMOC means low temps and long stadials?

We agree that it makes intuitive sense for long stadials to be also the coldest. However, we don’t know whether this would be due to background forcing or as a result of the mechanism driving the climate transitions, such as AMOC changes, or both.
From the perspective of our study the drivers and (most relevant) forcings are not known a priori, which is why we are testing evidence for it.

Discussion: This is just a lengthy summary. Please discuss the strengths and short-comings of your method, and place it in a broader context.

We agree and rewrote the Section.

P24L5: “cooling rates clearly determine” rewrite

We rewrote: “Because these durations are much more strongly governed by the cooling rates than by the cooling amplitudes they can be predicted to a good approximation as soon as the cooling rates have stabilized.”

Literature

Boch, R. et al.: Clim. Past, 7, 1247–1259 (2011)

Buizert, C. and Schmittner, A.: Paleoceanography, 30, 1595–1612 (2015)

Chappellaz, J. et al.:Clim. Past, 9, 2579–2593 (2013)

Cheng, H. et al.: Nature, 534, 640-646 (2016)

Moseley, G. E. et al.: Geology, 42 (12), 1043–1046 (2014)

Rashid, H. et al. Paleoceanography, 18 (4), 1077 (2003)

Rasmussen, S. O. et al.: Quat. Sc. Rev., 106, 14–28 (2014)

Rasmussen, T. L. et al.: Paleoceanography, 18 (1), 1018 (2002)

Rhodes, R. H. et al.: Science, 348, 1016-1019 (2015)
Rhodes, R. H. et al.: Global Biogeochem. Cycles, 31, 575–590 (2017)

Sanchez Goni, M. F. and Harrison, S. P.: Quat. Sc. Rev., 29, 2823–2827 (2010)

Schulz, M.: Geophys. Res. Lett., 29, 1002 (2002)

Seierstad, I. K. et al.: Quat. Sc. Rev., 106, 29-46 (2014)

Shackleton, N. J.: Science, 289 (5486), 1897-1902 (2000)

Waelbroeck, C. et al.: Quat. Sc. Rev., 21, 295-305 (2002)

Objective extraction and analysis of statistical features 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. The strongest mode of centennial to millennial climate variability in the paleoclimatic record are the Dansgaard-Oeschger (DO) cycles. Despite decades of research their dynamics and physical mechanism remain poorly understood. Valuable insights can be obtained by studying high-resolution Greenland ice core proxies, such as the NGRIP δ^{18}O record. However, conventional statistical analysis is complicated by the high noise level, the cause of which is partly due to glaciological effects unrelated to climate, and which is furthermore changing over time. We remove the high-frequency noise and extract the most robust features of the DO cycles, such as rapid warming and interstadial cooling rates, by fitting a consistent piecewise-linear model to three Greenland ice core records. With statistical hypothesis tests we aim to obtain an empirical ~~, mechanistic~~ understanding of what controls the of amplitudes and durations of the DO cycles. To this end, we investigate distributions of and ~~causalities~~ correlations in between different features, as well as modulations of them in time by external climate factors, such as CO_2 and insolation. Our analysis suggests different mechanisms underlying warming and cooling transitions due to contrasting distributions and external influences of the stadial and interstadial durations, as well as the fact that the interstadial durations can be predicted to some degree by the linear cooling rates already shortly after interstadial onset.

1 Introduction

~~The~~ Different physical mechanism(s) ~~and cause of the~~ underlying Dansgaard-Oeschger (DO) events ~~are largely unknown and debated. Modeling and simulations of the events are guided by the~~ have been proposed in the literature. Most of these are characterized by changes in between different modes of operation of the Atlantic Meridional Overturning Circulation (AMOC) that accompany the warm and cold phases of a DO cycle. This is supported by marine sediment data evidence linking DO cycles and changes in the AMOC (Henry et al., 2016; Lynch-Stieglitz, 2017). Different drivers for these AMOC changes have been proposed including North Atlantic freshwater forcing (Ganopolski and Rahmstorf, 2001; Timmermann et al., 2003; Kageyama et al., 2013 variations in ice sheet topography (Zhang et al., 2014) and atmospheric CO_2 (Zhang et al., 2017). On the other hand, unforced millennial-scale oscillations involving the AMOC have been reported in comprehensive climate models (Vettoretti and Peltier, 2016; Klock In these oscillations, sea ice variability in ocean convection areas plays an important role, which has been proposed previously (Li et al., 2010; Dokken et al., 2013; Petersen et al., 2013) and is supported by recent proxy records (Sadatzki et al., 2019). Another scenario underlying DO cycles might be spontaneous climate transitions due to extremes in the chaotic atmospheric dynamics (Drijfhout et al., 2013; Kleppin et al., 2015).

Modeling of DO events is guided by proxy records, among which the stable water isotope records from Greenland ice cores are very prominent. DO-type transitions in models range in their dynamics from stochastic to excitable and oscillatory, and are sensitive to different forcings. A statistical analysis of the DO cycles extracted from Greenland ice core records can thus be useful to evaluate the proposed models. The records are noisy, and since there are no established theories about how they should evolve, there is no obvious filter to extract the large-scale climate signal. A common characteristic of the DO cycles seems to be an abrupt temperature increase from the cold stadial conditions to a maximum temperature in the warm interstadial state followed by a gradual cooling until there is another abrupt jump back into the stadial state. This is referred to as the saw-tooth shape of the events.

Due to the high noise level in the record it is however difficult to discern this specific structure in all of the events. Some events do not seem to follow the generic shape. Furthermore, there are very short events where it is difficult to speak of a gradual cooling episode. Even other events are interrupted by shorter cooling episodes, referred to as sub-events (Capron et al., 2010). As interpretation of noisy time series are often biased, subjective and one is prone to recognize patterns that can arise by chance, we seek a quantitative evaluation of the record. Assuming the saw-tooth shape of the events, we develop an algorithm for fitting the saw-tooth shape to the entire NGRIP δ^{18}O record of the last glacial, similar to ramp-fitting a jump in a noisy record.

~~Firstly~~First, our method gives an objective basis of the validity of the generic saw-tooth description of the DO cycles and ~~identify~~ identifies which individual cycles fall outside this description. Secondly, with a piecewise-linear fit, we obtain estimates for the stadial and interstadial levels, the abruptness of the transitions and the gradual cooling rate in the interstadial periods. By bootstrapping, we estimate the uncertainty in extracting these parameters from the noisy background. Finally, we perform a comprehensive statistical analysis of the fit parameters across the DO events ~~of the last glacial period~~and their relation to external forcings, in order to obtain ~~a mechanistic~~ an empirical understanding of what controls the evolution of amplitudes and durations of the DO cycles. This can potentially be used for identifying or excluding proposed mechanisms and for benchmarking model results.

Previous efforts to extract robust DO event features from the record were conducted on only part of the record and were focused on single or very few features. In Schulz (2002), linear fits to the interstadials were used to infer the cooling rates starting with Greenland interstadial 14 (GI-14). Estimates for abruptness of warming transitions and durations of interstadials have been derived by Rousseau et al. (2017), starting at GI-17.1. A comprehensive survey of onset times of all interstadial and stadial periods is given in Rasmussen et al. (2014). Our work is different in that we derive all features that can be extracted from a saw-tooth shaped fit to all events at once, by using a fit that is consistent and continuous throughout the record. ~~We thus do not have to rely on any subjective choice of stadial and interstadial onsets or levels~~Thus, our results are not sensitive to subjective choices of cutting the record at predefined times before and after a transition. We do, however, not ~~attempt to~~ define the DO events themselves from the NGRIP δ^{18}O record, but instead use the ~~fixed~~ set of all previously classified events ~~(Rasmussen et al., 2014)~~in Rasmussen et al. (2014), which have been derived from three ice cores and two proxy records each.

In this study, we show that a characteristic saw-tooth waveform can be fit to all DO cycles. However, almost half of the cycles do not actually display a significant rapid cooling episode after the more gradual interstadial cooling. A subsequent statistical analysis of ~~the~~ DO event features hints at different mechanisms underlying warming and cooling transitions. ~~Specifically, we find different~~ First, this follows from the distributions of the ~~stadials and interstadial durations ,~~ durations of the stadials as well as ~~different external factors that could influence them. Furthermore~~the rapid DO warmings on the one hand, and of the interstadials on the other hand. Secondly, the influence of external forcing is contrasting, with stronger evidence for insolation influence on stadials and CO_2 or ice volume influence on interstadials Thirdly, the interstadial durations can be predicted to some degree by the linear cooling rates within a few hundred years after interstadial onset. In contrast, the stadial and rapid warming durations are consistent with the stadial-interstadial transitions as spontaneous, noise-induced escapes from a metastable state.

The paper is structured in the following way: In Sec. 2 we introduce the data used in the study and its pre-processing, the iterative fitting algorithm, the features we extract from the saw-tooth shape fit to the events and the statistical tools to analyze these features. In Sec. 3 we report the results of the fit. Section 3.1 discusses the appropriateness of the saw-tooth fit to the events, and Sections 3.2 and 3.3 treat the uncertainty in estimating the fit parameters and the derived features. In Sec. 4 we analyze in detail the features characterizing the stadial, interstadial and abrupt warming periods. The results of the fit and the implications of the subsequent data analysis are discussed in Sec. 5. We conclude in Sec. 6.

2 Methods and Materials

2.1 Data

The basis of our study is the $\delta^{18}O$ Greenland ice core record of the last glacial period (120 - 12 kyr BP, kyr BP = one thousand years before present). In the NGRIP ice core, $\delta^{18}O$ has been measured in 5 cm samples (NGRIP Members, 2004; Gkinis et al., 2014; Rasmussen et al., 2014). These raw depth measurements are transferred to the GICC05 time scale (Svensson et al., 2006). This results in an unevenly spaced time series with a resolution of 3 years at the end to 10 or more years at the beginning of the last glacial period. Because it greatly simplifies our analysis, we transfer this to an evenly-spaced time series by oversampling to 1 year resolution using nearest-neighbor interpolation. Thus, we do not alter the actually measured values, adding or removing any variability. For subsequent comparison, the high-resolution $\delta^{18}O$ record from the GRIP ice core on the GICC05 time scale has been used (Johnsen et al., 1997; Rasmussen et al., 2014), and processed in the same way.

Our fit is performed on a previously classified set of events from Greenland, which has been reported by Rasmussen et al. (2014) together with the time stamps. These time stamps are used to initialize our iterative fitting procedure, and are subsequently refined during the process. We do not treat sub-events, which are small dips to colder conditions during a warm period, as separate events, but instead fit them as part of the interstadial periods. On the other hand, we do include the interstadials 22 and 13, which follow the stadials 23.1 and 14 (denoted 'quasi-stadials' by Rasmussen et al. (2014)). These interstadials have some resemblance to the so-called rebound events of, e.g., interstadials 12, 8 and 7. However, they are longer and larger in amplitude. Even though the stadials 23.1 and 14 do not fully reach the low values of the stadials before and after, a saw-tooth fit where these are included in the interstadial is not satisfactory because the resulting gradual linear cooling is not representative of the actual trends. Our choice to consider them as separate events is difficult to justify on the basis of the NGRIP $\delta^{18}O$ record alone. Nevertheless, it is unlikely to change the conclusions of our analysis, which are based on statistical robustness tests.

In our data analysis, we use several other data sets that are not derived from Greenland ice cores. These are loosely referred to as external forcings, although not all are truly external to the climate system, but rather obtained from independent data sources. As proxy for global ice volume, we use the LR04 ocean sediment stack (Raymo and Lisiecki, 2005). To represent Antarctic temperatures, we choose the $\delta^{18}O$ record of the EDML ice core on the AICC12 time scale (EPICA Community Members, 2010). This data was processed by interpolation to an equidistant 20 year grid and subsequently smoothing by convolution with a 600 year Hamming window. Greenhouse gas forcing is represented by a composite CO_2 record from different Antarctic ice cores on the AICC12 gas time scale (Bereiter et al., 2015). Furthermore, we consider changes in insolation due to orbital variations. Firstly, we use incoming solar radiation at 65 degree North integrated over the summer (referred to as 65Nint hereafter), which we define as the annual sum of the radiation on days exceeding an average of 350 W/m^2 (Huybers, 2006). Secondly, we use incoming solar radiation at 65 degree North at summer solstice (referred to as 65Nss hereafter) (Laskar et al., 2004). In addition, we also consider the raw orbital parameters of obliquity, eccentricity, and precession index (Laskar et al., 2004).

2.2 Fitting routine

A naive approach to obtain a piecewise-linear fit of the whole record could proceed in the following way: Considering the stadials as constant, first cut the time series at a predefined beginning and end of two consecutive stadials. Then fit a saw-tooth shape to the event within the two stadials. The end of the fit to this event then determines the start of the next stadial, used to fit the following event. However, the point at which one initially cut the stadials influences the constant levels of the two stadials that have been used to determine the fit to the event. For a consistent fit, the start and end points of a stadial must rather be determined by the fits to two neighboring events. In this way, the fit to each event depends on both its neighboring events before and after, and we cannot fit the events sequentially. One solution is to fit the whole time series at once to a piecewise linear model with 186 parameters, corresponding to 6 parameters for each of the 31 DO events. However, due to high noise and abundance of sub-event features, such a fit will be difficult to achieve without invoking very complicated constraints.

Instead, we propose an iterative fitting routine that converges to a consistent fit, as detailed in the following. We start with a guess for the stadial onset and end times, which determine the constant stadial levels. Then we fit a saw-tooth shape individually to each event. Thereafter, we update the stadial onset and end times according to the fit and repeat the procedure. When after some iterations the onset and end times do not change significantly anymore the fit has converged and is consistent.

We start with an initial guess of the onsets and ends of the stadial periods, based on the timings reported by Rasmussen et al. (2014), which are kept fixed throughout the iterations of our algorithm. The time series is divided in segments at these times. For each event i, we take a segment consisting of a stadial and interstadial period plus the following stadial period. These segments are fitted individually to a piecewise-linear model, as shown in Fig. 1. The model starts with a constant line at the beginning of the stadial. The constant is fixed to the mean level of the stadial l_i^s, where the stadial beginnings and ends are determined by the initial guess, or the previous iteration. A first break point (parameter b_1) of this constant is determined, followed by a linear up-slope (parameter s_1). The slope ends at the second break point (parameter b_2). After this break point there is a linear down-slope (parameter s_2), which ends at a break point (parameter b_3). After this break point there is a steeper down-slope until a last break point (parameter b_4), which is at the level of the next stadial l^s_{i+1} that is determined from the previous iteration. After all events have been fit, the parameters b_4 and b_1 of each event update the beginnings and ends of the stadials. The updated stadials yield a new segmentation of the time series and new stadial levels, which are then used as constant segments in the next iteration. The idea of this approach is that if the problem is well behaved, the beginnings and ends of the stadials do not change significantly anymore after a certain number of iterations, meaning that a consistent fit of the entire time series is obtained. An algorithm for this routine, along with details of the optimization procedure to fit each event, is given in Appendix A.

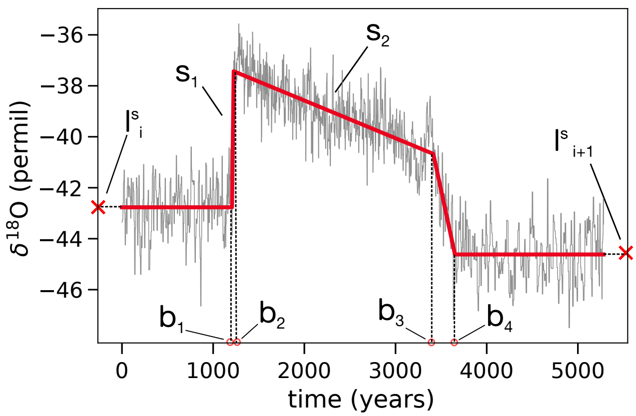


Figure 1. Piecewise-linear model fit to DO event 20, where the time series consists of GS-21.1, GI-20 and GS-20. The parameters of the piecewise-linear model are the four break-points $b_{1,2,3,4}$, the up-slope s_1 and the down-slope s_2. The constant levels l^s_i and l^s_{i+1} of GS-21.1 and GS-20 are constant during an iteration of the fitting routine, and are updated when after each iteration all breakpoints have been determined.

The fitting procedure outlined above yields a single best fit that we hope to be close to the absolute global minimum of the optimization problem and furthermore as consistent as possible, meaning that the stadial sections that were used for the fit in the last iteration are identical to the stadial sections defined by the resulting fit. Additionally to this best fit we assess the uncertainty in each of the parameters that arise due to noise in the record. We cannot estimate this from the output of our fitting procedure in a straightforward way. Instead, we use bootstrapping to repeatedly generate synthetic data for each transition and optimize the parameters~~with basin-hopping~~. Like this, we yield a distribution on each parameter. Due to computational demands, we do not combine this with our iterative procedure, but rather resample and fit every transition independently. Thus, we neglect the co-variance structure of the errors in the parameters of neighboring transitions. However, we still consider it to be a very good estimate of the uncertainty due to the noise in the record. The detailed procedure is given in Appendix C.

2.3 DO event features

From the best fit parameters of each DO cycle a variety of features follow, which form the core of our analysis. For each rapid warming period, gradual interstadial cooling period, as well as rapid cooling period at the end of an interstadial, we consider the duration, rate of change and the amplitude. Furthermore, several absolute levels are of interest, including the constant stadial levels, the interstadial levels after the abrupt warming and the interstadial level before the rapid cooling. As a level relative to each event, we consider the level before the rapid cooling above the previous stadial level, which is given by the rapid warming amplitude minus the gradual cooling amplitude. Finally, the gradual cooling amplitude divided by the rapid warming amplitude measures the position of the point of rapid cooling within the event amplitude. In total, we consider 15 interdependent ~~climate~~ features, which are listed in Tab. 1.

Table 1. List of DO event features obtained from the fit that are analyzed in this study.

Feature	Definition
Warming duration	$b_2 - b_1$
Warming rate	s_1
Warming amplitude	$\cancel{s_1(b_2-b_1)}\,A \equiv s_1(b_2-b_1)$
Gradual cooling dur.	$\cancel{b_3-b_2}\,D_I \equiv b_3-b_2$
Gradual cooling rate	$\cancel{-s_2}\,\lambda_c \equiv -s_2$
Gradual cooling ampl.	$s_2(b_2-b_3)$
Fast cooling dur.	$b_4 - b_3$
Fast cooling rate	$\frac{s_1(b_2-b_1)+s_2(b_3-b_2)-(l^s_{i+1}-l^s_i)}{(b_4-b_3)}$
Fast cooling ampl.	$s_1(b_2-b_1)+s_2(b_3-b_2)-(l^s_{i+1}-l^s_i)$
Stadial duration	b_1
Stadial level	l^s
Interstadial level	$s_2(b_2-b_3)+l^s$
Interstadial end level	$s_1(b_2-b_1)+s_2(b_3-b_2)+l^s$
Relative Int. end level	$s_1(b_2-b_1)+s_2(b_3-b_2)$
Cooling/warming ampl.	$s_2(b_2-b_3)\cdot[s_1(b_2-b_1)]^{-1}$

2.4 Data analysis

Our aim is to develop ~~a mechanistic~~ an empirical understanding of the evolution of the DO cycles. To this end, we employ several tools to search for relations between different features, as well as between features and external climate factors. Additionally, the distributions of the individual features themselves hold important information, especially when there is no strong external modulation in time. We test the distributions using Anderson-Darling (AD), Cramer-von Mises and Kolmogorov-Smirnov tests. Since the AD test is typically most powerful, and the other tests yield qualitatively unchanged results in all of our analyses, we only report results for the AD test.

Because of the large number of possible combinations of features, we first pre-select significant and potentially relevant relationships and thereafter investigate in detail, whether the results are robust to outliers, among other things. In some cases we also consider relationships of features and forcing that are not significant for the whole data set, but for a large subset. This might highlight that there were qualitatively different periods within the last glacial, or that some DO events are of different nature than most.

We first consider Pearson and Spearman correlation coefficients of pairs of features and external climate factors. We pre-select combinations with p-values $p < 0.1$, which are estimated by permutation tests that assume independent samples. For a given number of data points in a sequence, the true p-values should often be higher due to autocorrelation. Along with other potential artifacts, this is investigated individually for the pre-selected combinations.

Next, in order to find relations between more than two variables, we search for multiple linear regression models to explain selected features of the data. For this, we often use logarithmic quantities because with many features it is otherwise unlikely to find linear relationships that are not dominated by outliers. Given a feature as response variable, we fit linear regression models of combinations of two other features or forcings and pre-select models with the largest coefficients of determination, in order to further analyze the fit.

Furthermore, in order to find subsets of events that have distinct properties or ~~causal relations~~ relationships that are only valid in part of the data, we perform a clustering analysis on the data, using two different algorithms (K-means and Agglomerative Hierarchical Clustering). Given our sample size of 31 events, we search for clusterings with 2 or 3 clusters. Potentially relevant clusterings are pre-selected by considering the mean Silhouette coefficient, which is a distance-based measure for the validity of clusters. With the abovementioned tools, we perform an analysis on the entire set of features and forcings. From the results obtained, we report selected findings, which are most robust and relevant, in Sec. 4 of this paper.

It is important to be aware of the problem of multiple comparisons when interpreting the significance of such an analysis. Tests for significant correlations of many pairs of features using, e.g., the Spearman correlation coefficient, yield a non-negligible number of false positives when using confidence levels that are reasonable for our purposes. We consider both features of the same and neighboring events, yielding $\frac{15\cdot 14}{2} = 105$ and $15 \cdot 15 = 225$ tests, respectively. Furthermore, we test the correlation of all features and forcings, yielding another $15 \cdot 8 = 120$ tests. Assuming these are all independent tests, the total expected number of false positives is 45 at 90%, 22.5 at 95%~~confidence, while~~ , and 4.5 at 99% ~~it is 4.5~~confidence. Since we derive 15 features from only 7 independent parameters (including stadial levels) for each DO cycle, many pairs of features within the same DO cycle are related and thus we expect true positives for correlation tests. For instance, this is true for warming amplitude and interstadial level, or relative interstadial level and gradual cooling amplitude. Similarly, due to the constraints on the fit parameters, the rates and durations of fast and gradual cooling are correlated. These types of correlations are not relevant and thus this reduces the number of pairwise correlations to consider. For combinations of amplitude, duration and rates of a given linear segment we also expect correlation, because they are trivially related: duration = amplitude $\cdot$ (rate)$^{-1}$. However, it is interesting to test whether the rates or the amplitudes more strongly determine the durations, which essentially depends on which of the two has a larger variability. We investigate this for the different periods of the DO cycles below.

There are sophisticated methods to control the multiple comparisons problem. These could be helpful to better detect false positives from our analysis, but depend on being able to properly estimate the significance of individual correlations in between features with autocorrelation and assess the statistical dependence of the hypothesis tests due to the dependence of some of the features. For simplicity, we do not consider such an analysis, but consider individually significant correlations as suggestions to be investigated further.

3 Piecewise-linear fit of NGRIP record

The iterative fitting routine is performed for 40 iterations, so that the initial fluctuations in the parameters have died out and converged to a consistent fit, as detailed in Appendix B. In Fig. 2, the resulting fit is superimposed onto the high-resolution NGRIP time series. We fit 31 DO events in total, starting with DO 24.2 and ending at DO 2.2, excluding the two outermost events of the last glacial, because they are very non-stationary in their stadial parts. Table 2 shows all parameters obtained from the fit. Instead of $b_{1,2,3,4}$ for each transition, we show the corresponding times of stadial end, interstadial onset, interstadial end and stadial onset.

3.1 Saw-tooth shape of DO events

In our fit, all transitions follow the characteristic saw-tooth shape. For a few events, this is because of the constraints we use in the fitting algorithm. Typically, the constraints do not strictly bound the best fit parameters, but they force the fit into another local minimum that is consistent with the saw-tooth shape, which often yields parameters that are still clearly within the constraints. There are, however, four events where the best fit parameters actually lie very close to the bounds set by the constraints. This happens for GI-5.1 and GI-3, which both have ratios of rapid to gradual cooling rates very close to the constraint value of 2.0. Similarly, for GI-15.2 and GI-6 the ratio of gradual to rapid cooling duration is close to 2.0. Detailed pictures of each transition and the corresponding fit are shown in Fig. S2 in the supplementary material.

The fact that some constraints are needed in order to ensure that the fit of each event follows a saw-tooth shape can be used to classify which events fall outside of this description. To this end, we perform another run of the iterative fitting routine without using constraints 3, 4, 6 and 7 listed in Appendix A. From the resulting fit we then analyze, which of the events are not consistent with the saw-tooth shape. For this, we use 4 criteria: 1. The abrupt cooling rate is at least twice as large as the gradual cooling rate. 2. The gradual cooling lasts at least twice as long as the abrupt cooling. 3. There is gradual cooling after the rapid warming, i.e., the gradual cooling rate is negative. 4. The abrupt cooling amplitude is larger than 0.5 permil. Criterion 1 is not met by events 23.1, 19.2, 15.1, 11, 5.1, 3 and 2.2, criterion 2 by events 21.2, 19.2, 17.2, 15.2, 15.1, 11, 10, 9, 8, 6, 5.1, 3 and 2.2, criterion 3 by event 11, and criterion 4 by events 23.1 and 15.1. By demanding that all of these criteria are met, we thus conclude that the following 14 out of 31 events fall outside of the saw-tooth description: 23.1, 21.2, 19.2, 17.2, 15.2, 15.1, 11, 10, 9, 8, 6, 5.1, 3 and 2.2.

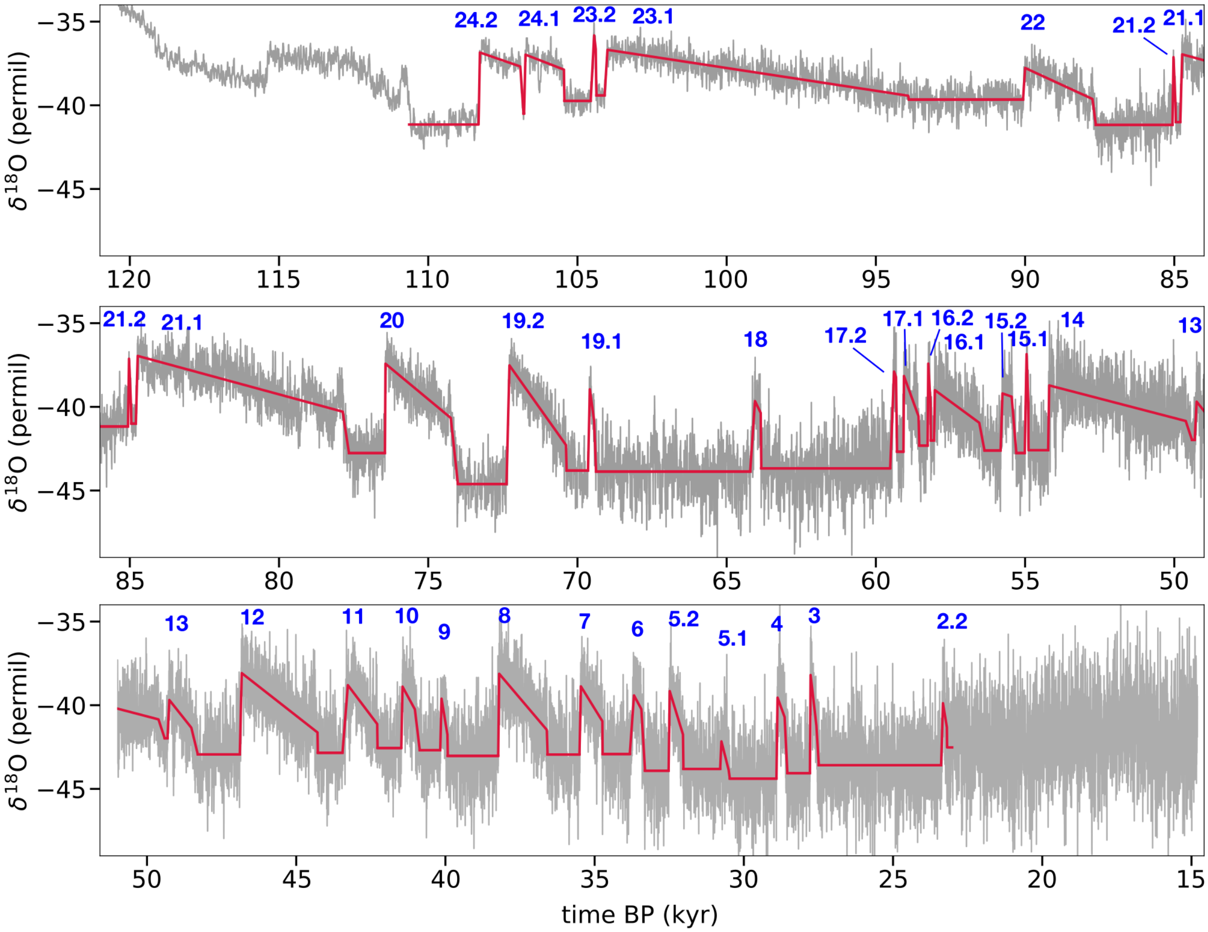


Figure 2. High-resolution NGRIP δ^{18}O time series and the piecewise-linear fit obtained by our method. The numbers above the interstadials indicate the names of the DO cycles considered in this study.

3.2 Uncertainty of fit parameters and features

From the best fit, we estimate the uncertainty of each parameter via bootstrapping, as explained in Appendix C. As an example, we show distributions of the parameters for DO event 20 in Fig. 3. In Tab. 3 we show the durations and amplitudes of the rapid warmings for each event along with a bootstrap confidence interval consisting of the 16- and 84-percentiles, which would correspond to the $\pm\sigma$ range if the distributions were Gaussian. The actual distributions are often skewed, so that the best fit values lie close to the edges of the confidence intervals, or even outside of the intervals. In these cases, the $\pm\sigma$ confidence intervals are not the best indicator for the uncertainty, because they barely include the mode of the very skewed distributions.

The magnitude of the uncertainty varies from event to event. In the case of the warming durations, the average bootstrap standard deviation is 20.0 years, with a minimum of 3.4 years for GI-16.2 and a maximum of 57.4 years for GI-18. We observe that shorter warmings typically also have smaller uncertainties. As comparison, the durations of the rapid coolings at the end of

Table 2. Parameters resulting from the fitting routine on the NGRIP data.

Event	Stadial End (yr BP)	Interstadial Onset (yr BP)	Interstadial End (yr BP)	Stadial Onset (yr BP)	Warming Rate (permil/yr)	Cooling Rate (permil/yr)
24.2	108313	108270	106914	106810	0.0992	0.00062
24.1	106790	106743	105452	105439	0.0744	0.00069
23.2	104556	104441	104387	104366	0.0340	0.01555
23.1	104090	103996	93916	93898	0.0290	0.00027
22	90069	89999	87743	87631	0.0270	0.00082
21.2	85060	85027	84964	84952	0.1230	0.03992
21.1	84799	84737	77866	77659	0.0655	0.00049
20	76452	76434	74245	74009	0.2935	0.00148
19.2	72377	72280	70385	70365	0.0730	0.00251
19.1	69646	69587	69443	69381	0.0831	0.01262
18	64212	64051	63858	63846	0.0262	0.00367
17.2	59520	59390	59323	59294	0.0446	0.00503
17.1	59076	59061	58571	58549	0.2951	0.00491
16.2	58266	58245	58168	58162	0.2340	0.03107
16.1	58051	58023	56536	56364	0.1059	0.00131
15.2	55821	55759	55449	55296	0.0554	0.00062
15.1	55011	54950	54892	54887	0.0981	0.05104
14	54228	54193	49617	49410	0.1092	0.00046
13	49315	49253	48517	48301	0.0367	0.00223
12	46890	46826	44286	44277	0.0761	0.00140
11	43450	43271	42285	42278	0.0225	0.00236
10	41479	41439	41024	40864	0.0910	0.00326
9	40175	40131	39933	39928	0.0699	0.01096
8	38231	38199	36602	36583	0.1549	0.00210
7	35508	35461	34741	34735	0.0859	0.00289
6	33822	33681	33434	33314	0.0250	0.00334
5.2	32521	32485	32039	32028	0.1324	0.00583
5.1	30794	30752	30514	30473	0.0394	0.00695
4	28908	28871	28635	28544	0.1302	0.00485
3	27786	27765	27572	27492	0.2762	0.01529
2.2	23389	23328	23196	23191	0.0607	0.01098

an interstadial have a larger uncertainty with an average bootstrap standard deviation of 53.6 years. This is expected, because the rapid cooling is typically less well pronounced in the record compared to the rapid warming. The coolings also have a larger spread in the bootstrap standard deviations with a minimum of 4.6 years for GI-16.2 and a maximum of 209.9 years for GI-23.1, because some events have a very clearly defined rapid cooling, while others do not. Similarly, the onset times of the rapid warmings have an average bootstrap standard deviation of 11.4 years, whereas the onset of the stadial periods have a corresponding average uncertainty of 31.7 years.

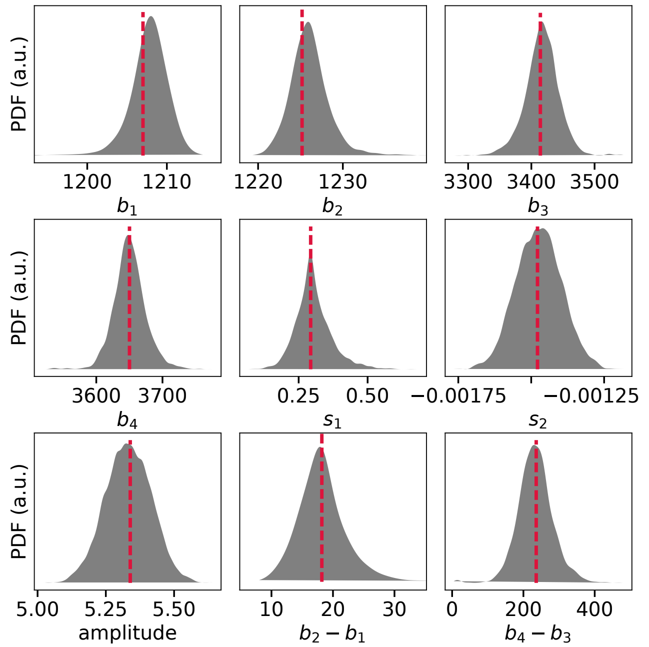


Figure 3. Gaussian kernel density of the density of model parameters and some derived quantities for the DO event 20 after 5000 iterations of the bootstrap resampling procedure. The parameter values for the best fit, as reported in Sec. 3, are indicated with *red dashed* lines. The amplitude feature is given by $s_1(b_2 - b_1)$.

3.3 Comparison of NGRIP and GRIP records

As complementary approach to assess the uncertainties of the features, we compare them to those derived in the same way from another Greenland ice core. We chose the δ^{18}O record of the GRIP ice core, which is measured at a similar resolution to the NGRIP record and has been transferred to the GICC05 time scale starting at the onset of GI-23-1. We thus started fitting the record from GS-22 with 40 iterations of our algorithm, using the same constraints and hyperparameters. Again, the algorithm converges to a consistent fit, where each of the events is well approximated by a saw-tooth shape. We now describe how well the features of NGRIP and GRIP correspond for the 26 mutual events.

For the gradual cooling rates, the Pearson (Spearman) correlation is $r_p = 0.64$ ($r_s = 0.65$). Here, the discrepancy in between the two records is only due to a handful of short events that show a clear linear cooling slope in one record, but are more plateau-like in the other. This happens for the interstadials 18, 16.2 and 5.1, which don't show a slope in GRIP, and 17.2, which doesn't show a strong slope in NGRIP. If we remove these events, the correlation is $r_p = 0.97$ and $r_s = 0.98$. The warming durations show a correlation of $r_p = 0.55$ and $r_s = 0.63$. There are no outliers, but a rather large spread, indicating that the warming duration is a less robust feature compared to the cooling rate. With 69 years on average, the GRIP warmings are 8 years shorter than the NGRIP average. The average absolute deviation of warming durations in the two cores is 31 years, with a maximum discrepancy of 103 years for GI-10, where we find a warming of 40 years for NGRIP and 143 years for GRIP.

Table 3. Durations and amplitudes of the rapid warmings inferred from the fit, together with a confidence interval obtained by bootstrapping.

Event	**Warming duration** (yr)			**Amplitude** (permil)		
	Best fit	16-p	84-p	Best fit	16-p	84-p
24.2	43.4	36.3	47.8	4.30	4.23	4.40
24.1	47.4	34.9	45.0	3.53	3.42	3.61
23.2	115.2	96.1	126.1	3.92	3.72	4.11
23.1	94.1	78.9	127.3	2.73	2.69	2.75
22	70.0	70.3	91.8	1.89	1.78	1.95
21.2	33.0	25.7	39.9	4.06	3.61	4.10
21.1	61.7	53.5	79.6	4.05	3.98	4.09
20	18.2	14.7	21.6	5.34	5.25	5.42
19.2	97.2	74.3	98.1	7.09	6.93	7.19
19.1	58.5	37.7	60.0	4.86	4.45	4.97
18	161.0	74.6	194.0	4.21	3.99	4.51
17.2	129.7	83.7	158.0	5.79	5.47	6.20
17.1	15.3	13.9	27.0	4.53	4.14	4.72
16.2	21.0	18.6	24.0	4.92	4.59	5.19
16.1	28.4	28.9	84.0	3.01	2.88	3.16
15.2	61.6	39.0	100.0	3.41	3.38	3.67
15.1	60.6	56.4	69.2	5.94	5.68	6.12
14	35.5	38.0	79.0	3.87	3.78	3.95
13	62.4	63.4	101.2	2.29	2.07	2.60
12	63.9	45.7	73.8	4.86	4.71	4.94
11	179.5	143.0	201.0	4.05	3.86	4.17
10	40.3	41.3	80.2	3.67	3.40	3.97
9	44.2	37.5	86.4	3.09	2.66	3.22
8	31.7	29.8	53.0	4.91	4.78	4.98
7	47.4	45.3	90.2	4.07	3.86	4.27
6	140.4	110.6	172.1	3.51	3.41	3.92
5.2	36.0	31.1	54.6	4.76	4.45	4.93
5.1	41.8	41.4	82.0	1.65	1.51	1.89
4	37.2	27.1	41.8	4.84	4.47	5.11
3	21.3	18.0	25.0	5.88	5.40	5.92
2.2	61.2	42.0	91.6	3.72	3.21	3.75

Such deviations can arise if there is a slight step in the record before the most rapid warming and the algorithm includes this in the fit.

The warming amplitudes are very well correlated with $r_p = 0.87$ and $r_s = 0.83$. The average amplitude of 3.87 permil in GRIP is 0.45 permil lower than the NGRIP average. The stadial levels are also well correlated with $r_p = 0.78$ and $r_s = 0.66$. There is a quite consistent offset in between the cores of 1.84 permil due to differences in altitude and latitude of the GRIP and NGRIP sites. Exceptions include GS-21.1, which does not follow the offset but is at a very similar level in both GRIP and NGRIP, and GS-14, which is difficult to define and thus vulnerable to give different results due to different noise in the cores.

The rapid cooling durations, i.e. $b_4 - b_3$, show an average absolute deviation in between the two cores of 59 years, with $r_p = 0.46$ and $r_s = 0.53$. This corroborates the fact that this feature is harder to define than the rapid warmings. The break points b_3 and b_4 are very susceptible to noise and can yield qualitatively different results for different cores. As a result, the abrupt cooling of GI-19.2 lasts 208 years in GRIP and only 20 years in NGRIP, and for GI-12 294 years in GRIP and only 9 years in NGRIP. Conversely, the abrupt coolings of GI-19.1, GI-10 and GI-6 last much longer in NGRIP, with 62, 160 and 120 years in NGRIP versus 2, 5 and 2 years in GRIP, respectively. Consequently, we do not report any results concerning the rapid cooling period in this paper.

The stadial and interstadial durations are very well correlated with $r_s = 0.99$ and $r_s = 0.97$, respectively. The average absolute deviation is 59 years for interstadials and 73 years for stadials, which is small compared to the average durations. The biggest discrepancies in between the two cores come from the indeterminacy in the rapid coolings of certain events, as described above.

In summary, the uncertainties obtained by bootstrapping and by comparison with the GRIP ice core are compatible. The average bootstrap standard deviation of rapid warming and cooling durations is 20 and 54 years, respectively. This compares well to the average absolute deviation in between GRIP and NGRIP of warming and cooling durations of 31 and 59 years, respectively. The discrepancy of 31 years for warming durations also includes a systematic bias of on average 8 year longer warmings in GRIP. Thus the unbiased uncertainty is likely even closer to the one obtained by bootstrapping. Shorter time scale features like rapid warming durations are not fully representative for every single event in one core. However, the overall trends are consistent, as seen by significant correlation. Features on longer time scale, such as most of the cooling slopes and stadial levels, are clearly representative. The same holds for stadial and interstadial durations.

4 Statistical analysis of DO event features

In Fig. 4 we show histograms of all the DO event features derived from the fit parameters that we consider in this study, as defined in Sec. 2.3. From the histograms we can see that the features have different types of distributions. Whether an event feature should be considered an independent sample from a distribution depends on whether it shows a significant trend over time. If we consider the event-wise sequence of one feature as evenly spaced time series we can calculate the autocorrelation and determine by a permutation test, whether the value at a given lag is significantly larger than what could be expected in an uncorrelated sample for a given confidence. By considering autocorrelations up to lag 5, we find that the three different levels (stadial, interstadial and level before rapid cooling) show significant autocorrelation at 95% confidence until a lag of 3. We also find significant autocorrelation for four other features at only one specific lag value each, which we believe are false positives. In fact, when independently testing the hypothesis of significant autocorrelation at 95% confidence for 15 different time series (features) at 5 lags, there is an expected value of 3.75 false positives. The corresponding data is shown in Fig. S3 in the Supplementary material. As a result, in most cases we can consider the features of each event as independent samples. This helps to assess the significance of correlations between features with permutation tests. ~~While a~~ A general overview of the correlations between different features and forcings ~~is given~~ can be seen in Fig. 5, and is explained further in Appendix D, ~~the~~ The most important results of our statistical analysis are presented in the following Sections.

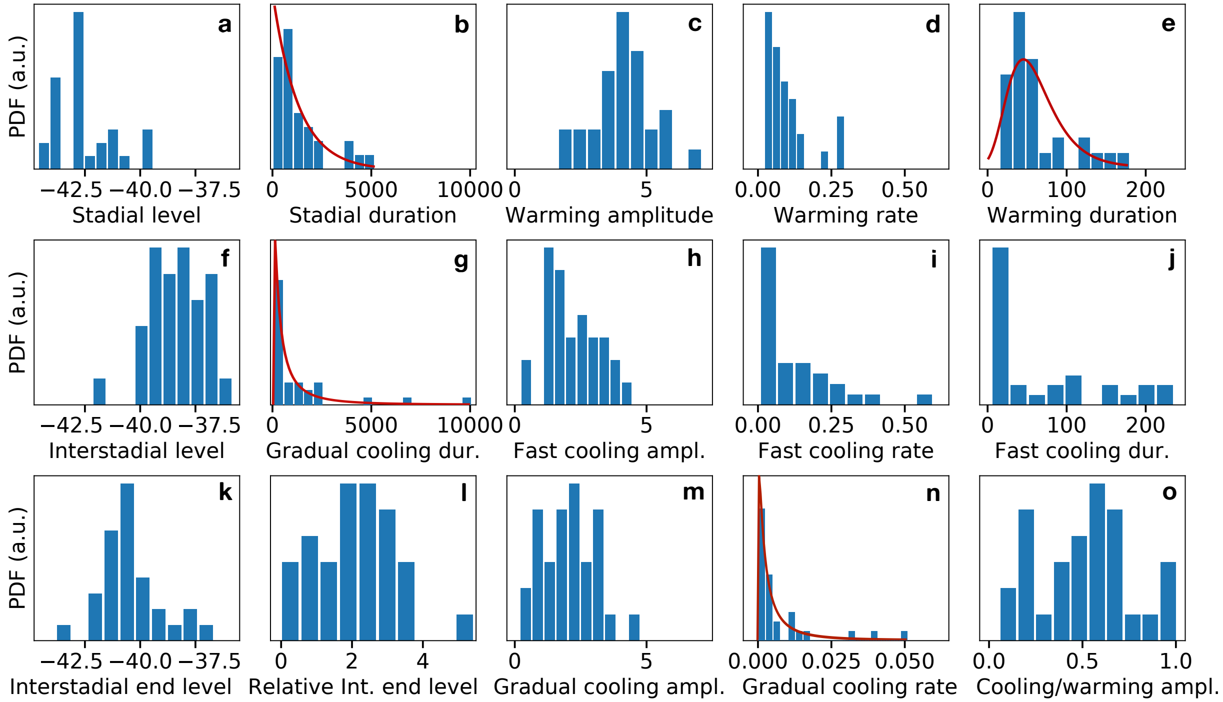


Figure 4. Histograms of our sample of 31 events for all features considered in this study, as defined in Tab. 1. The *red* curves in panels b) and e) are fits with the exponential and Gumbel distributions, respectively, whereas those in panels g) and n) are fits with the log-normal distribution.

4.1 Interstadial periods

4.1.1 Cooling rates ~~determine~~ govern interstadial durations

We focus on the factors influencing the durations of the interstadial periods ~~, defined as $b_3 - b_2$. As noted previously for the younger half of the last glacial in the GISP2 ice core, there is a correlation of interstadial durations D_I and their respective cooling rates $\lambda_c = AD_I^{-1}$, where A~~ $D_I = b_3 - b_2$. In our fit, these durations are furthermore defined by $D_I = A\lambda_c^{-1}$, where A is the amplitude ~~of the cooling (Schulz, 2002)~~and λ_c the rate of the gradual cooling. If for every interstadial the gradual cooling would be perfectly linear and the jump back to stadial conditions would always occur after the same magnitude of cooling $\bar{A}$, the interstadial duration would be inversely proportional to the cooling rate $D_I = \bar{A}\lambda_c^{-1}$. Conversely, if the interstadials would have a fixed cooling rate $\bar{\lambda}_c$ and the jump back to stadial conditions happened after variable cooling magnitudes, the interstadial durations would be proportional to the cooling amplitudes $D_I = A\bar{\lambda}_c^{-1}$.

We test which of the two scenarios is better supported by the data. This depends on whether either the cooling amplitudes or the cooling rates have a larger spread than the other. The coefficient of variation for the amplitudes is $CV = 0.51$, whereas for the rates we find $CV = 1.49$. The Spearman correlation of durations and cooling rates is $r_s = -0.89$, which is clearly

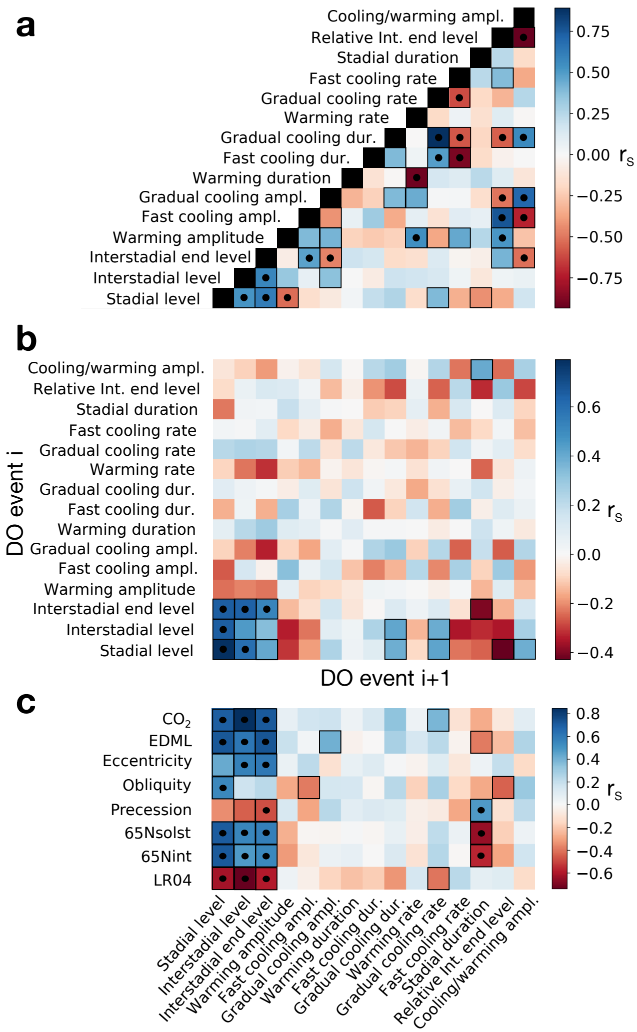


Figure 5. ~~Histograms~~ Spearman correlation heat map of ~~our sample~~ **a)** pairs of ~~31 events for all~~ features ~~considered in this study~~within the same DO cycle, ~~as defined in Tab. 1. The redcurves in panels~~ **b)**~~**and c) are fits with the exponential and Gumbel distributions, respectively, whereas those**~~ pairs of features in ~~panels g~~adjacent DO cycles, and **c**) pairs of one feature and ~~n~~one external forcing at the relevant time point of the feature. Correlations that are significant at 95% (99%) according to a permutation test are ~~fits~~ highlighted with ~~the log-normal distribution~~a black frame (dot).

significant given the samples size of 31 events and weak autocorrelation of the sequence of interstadial durations and rates. On the contrary, for durations and cooling amplitudes we find $r_s = 0.40$, which is mainly due to two outliers, the two longest interstadials GI-23.1 and GI-21.1. Removing these reduces the correlation to $r_s = 0.30$, which is not significant at 95% confidence. As a result, there is no relationship between durations and amplitudes that goes beyond outlier events, as opposed to durations and cooling rates. Furthermore, the correlation of cooling amplitudes and rates is not significant. Thus~~there is indeed a strong~~

~~control by the cooling rates on~~ , from the perspective of linear interstadial cooling, the interstadial durations over the entire glacial are indeed largely governed by the cooling rates, in agreement with the findings for the younger half of the glacial ~~by Schulz (2002)~~obtained by Schulz (2002) from the GISP2 ice core.

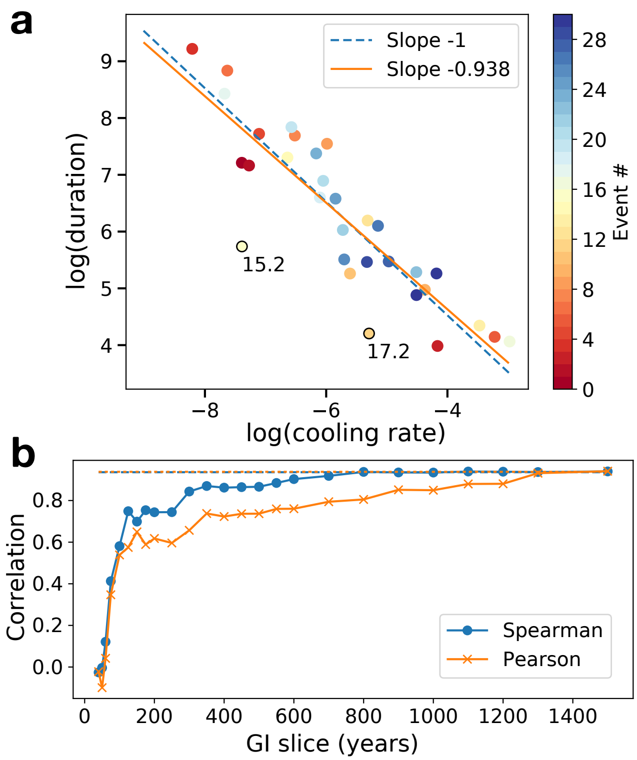


Figure 6. a) Scatterplot of the logarithms of interstadial durations and cooling rates. The color coding indicates the temporal sequence of the events starting with GI-24.2 as event #0. Two linear fits obtained by ordinary least squares are shown. For one of them we fixed the slope to -1 and varied only the intercept. **b)** Correlation coefficients of the logarithms of interstadial duration and the linear slope fitted to a slice of the beginning of the interstadial as a function of the length of that slice. The values of the Spearman (Pearson) correlation coefficients using slopes obtained from the full interstadials is marked with a *dashed* (*dotted*) line.

In Fig. 6a we show a scatterplot of $\log \lambda_c$ and $\log D_I$ along with a linear regression yielding a slope of -0.94. The 95% confidence interval of this slope obtained via bootstrapping is [-1.12, -0.75]. Because we do not account for errors in the rates estimated from the data the regressed slope is biased towards 0 due to attenuation and the true slope will be closer to -1. The model $D_I \propto \lambda_c^{-1}$ is consistent with the data, where the spread is caused by the fact that the jump back to stadial conditions happens after varying cooling amplitudes, which have a mean of 2.04 and standard deviation of 1.04.

4.1.2 Distribution of interstadial cooling rates and durations

The relationship between interstadial durations and cooling rates also manifests itself in the respective distributions. As seen in Fig.'s 4g and 4n, both features have strongly skewed histograms. Both are consistent with log-normal distributions, as shown by ~~Anderson-Darling~~ AD tests with $p = 0.47$ and $p = 0.89$ for durations and rates, respectively. A fit with this distribution is illustrated in the figure. Because the two features are inversely related with $D_I = \bar{A} \cdot \lambda_c^{-1}$, the fact that one is a log-normal random variable implies that the other is, too. If D_I is distributed log-normally with parameters μ and σ, then λ_c^{-1} follows a log-normal distribution with parameters $-\mu + \ln(\bar{A})$ and σ. In our data this relation holds: We estimate μ and σ from the data D_I and use the observed average amplitude $\bar{A} = 2.04$. An ~~Anderson-Darling~~ AD test with $p = 0.33$ shows that the data λ_c^{-1} is consistent with a log-normal distribution with $-\mu + \ln(2)$ and σ.

As opposed to other skewed distributions like the exponential, Gumbel and power law, both durations and cooling rates are also consistent with an inverse Gaussian distribution. The observation that the durations and rates and are both well fitted by the inverse Gaussian despite their inverse relation is explained by the similar shape of the reciprocal inverse Gaussian distribution. If a variable is distributed as inverse Gaussian $X \sim IG(x)$, then the distribution of $Y = \frac{\bar{A}}{X}$ is reciprocal inverse Gaussian $Y \sim \frac{\bar{A}}{x^2} IG(\bar{A}/x)$. A moderately sized sample of Y is still likely to be consistent with an inverse Gaussian distribution, due to the similarity of the two. The inverse Gaussian could make an appealing model for the interstadial durations, since it arises as distribution of first hitting times of a constant level for Brownian motion with a constant drift. However, the proxy time series in interstadials look qualitatively different than what is expected from this model, because they are quite linear and yet have strongly varying slopes. In order for the model to produce roughly linear time series, the drift has to be high, which results in very similar slopes of the time series with the resulting distribution of first hitting times converging to a Gaussian. We leave a further discussion on which mechanism could yield log-normal or inverse Gaussian distributions of durations or cooling rates for upcoming studies. Instead, in the following we focus on implications of the approximate linearity of the interstadial time series.

4.1.3 Predictability of interstadial durations

The ~~simple~~ strong relationship of interstadial durations and cooling rates might have some implications on the understanding of DO event dynamics. If the rates ~~imply the durations~~ govern the durations much more so than the cooling amplitudes, then the durations ~~are already determined~~ can already be approximately predicted as soon as the rate is established, which might happen early in the interstadial. This is different from the idea that the ~~transition from interstadial to stadial might be~~ interstadial-stadial transition is a noise-induced escape from one metastable state to another. To test this, we take small slices of the beginnings of each interstadial, fit a linear slope s to them and then calculate how well these slopes ~~determine~~ correlate with the durations of the full interstadials as we increase the length of the slices. Due to noise in the beginning of the interstadials, for some interstadials a small positive slope is being detected. We reset these slopes instead to $s = -0.0001$, because in our analysis we use the logarithms of slopes and durations. For the relatively short events 15.2 and 17.2, no negative slopes are obtained when fitting the whole interstadial part independently, as opposed to the slopes obtained in the fit of the entire time series. We thus have to exclude these two outliers in the following. In Fig. 6b we show how the correlation between the logarithm of the slopes $-\log(-s)$ of these slices and the durations $\log D_I$ evolves as we increase the length of the slices. The correlation of the slopes estimated from the full interstadials and the durations, when excluding events 15.2 and 17.2, is $r_s = 0.94$ ($r_p = 0.94$), and is indicated by a dashed (dotted) line. We can see that the correlation rapidly increases up to a length of 150 years. Thereafter the correlation stabilizes until another more rapid increase at a length of 350 years. The rapid increase in correlation is partly due to a non-negligible number of events already being at full length (6 events at 150 years and 12 events at 350 years). Still, also the slopes of the remaining events already correlate well with the durations. At 350 years, the durations are almost as well determined by the slopes estimated from the slices as they are from the full interstadials. The remaining indeterminacy comes from a handful of longer interstadials (23.2, 22, 14 and 11) that do not settle to a clear negative slope after 350 years. For the latter three events, this is due to sub-events that occur shortly after the interstadial onset.

Our interpretation is that the cooling rate is an indicator of a timescale of large-scale climate reorganization, which can already be measured relatively early in the interstadial and which remains approximately constant. Although we can see that there are exceptions, we conclude that for most events the interstadial duration is determined ~~at a relatively early stage within~~ a few hundred years after the rapid warming.

4.1.4 Influence of external forcing

~~Having established control of the interstadial durations by the cooling rates~~Given the previous result, we investigate whether the variability in the ~~rates~~ timescale associated with the cooling rate can be explained by other features of the DO cycles, or by external forcing. Among correlations of the cooling rates with other features deemed significant by a permutation test, none of them are relevant, either because they are caused by few outliers or else directly due to their definition and parameter constraints.

Considering external climate factors, we find a correlation of $r_s = 0.40$ and $r_p = 0.35$ of the logarithm of the cooling rates with LR04 at the time of the interstadial maximum. This correlation is, however, influenced by a common trend of the two quantities, and is not significant anymore at 90% confidence when removing a linear trend. On the other hand, there is a large sub-set of events which appears to be linearly related.

As shown in Fig. 7a and c, the furthest outliers from an approximate linear relationship are the interstadials 23.2, 21.2, 16.2 and 15.1. When removing these outliers, the correlation is $r_p = 0.79$, which clearly goes beyond a common trend with $r_p = 0.63$ after linearly detrending. When only considering the younger half of the record starting with GI-14, the correlation is $r_p = 0.84$. This corresponds to the finding by Schulz (2002), who reports that the interstadial cooling rates starting from GI-14 are forced by global sea level. We note, however, that the correlation after GI-14 is largely due to common trend, as we find $r_p = 0.37$ after linear detrending, which is not significant at 90% confidence. Nevertheless, as shown above, when discarding a few outliers there is evidence for significant correlation as we include older parts of the record.

A better predictor of the interstadial cooling rates of the more recent DO cycles is given by the CO_2 composite record. Whereas for the older half of the glacial there is no significant correlation, when starting at GI-14, we find a linear relationship

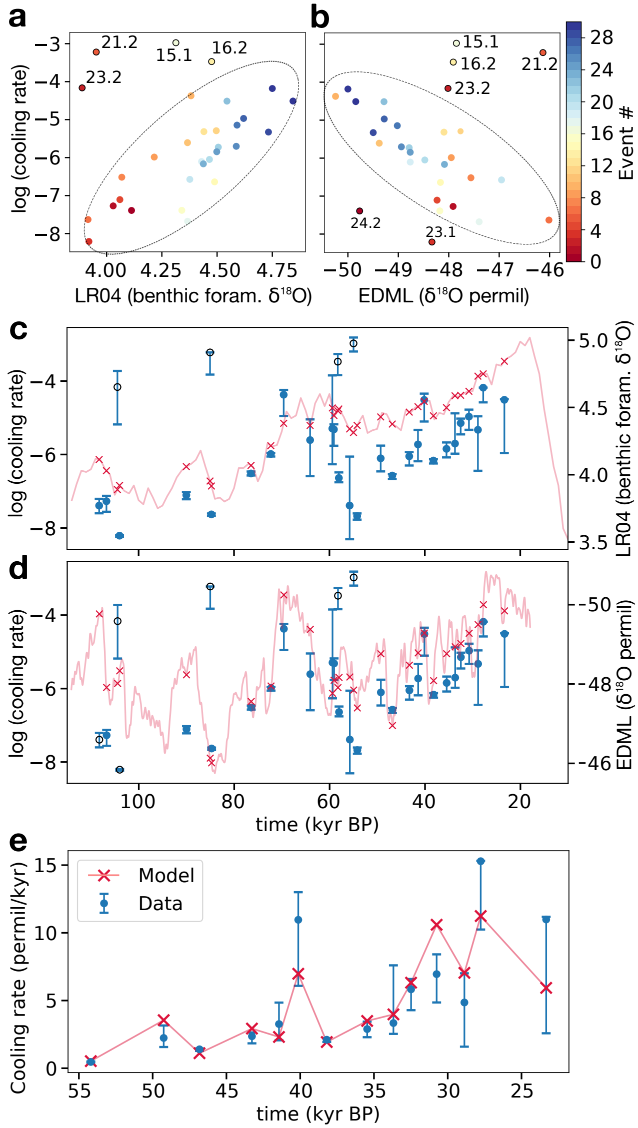


Figure 7. a-b) Scatterplot of the logarithm of the interstadial cooling rates and the **b)** LR04 and **b)** EMDL at time points corresponding to the interstadial onsets. **c)** Time series of the cooling rates (*dots*) and the LR04 stack (*crosses*). The error bars on the cooling rates are given by the 16- to 84-percentile obtained by bootstrapping. **d)** Time series of the cooling rates (*dots*) and the EDML stack (*crosses*). Note the inverted axis for EDML. **e)** Time series of the interstadial cooling rates starting at GI-14 and of a linear regression model of the CO_2 record fitted to the logarithm of the cooling rates.

with $r_p = -0.93$ and $r_p = -0.81$ after linear detrending. In Fig. 7e we illustrate how well the cooling rates of this period can be predicted from CO_2 by linearly regressing CO_2 onto the logarithm of the cooling rates and then exponentiating the result.

Additionally, in a subset of the events, there is a linear relationship between the logarithm of the cooling rates and EDML at the interstadial onsets. While the correlation of the entire data set is not significant at 90% confidence with $r_p = -0.19$ and $r_s = -0.23$, when removing events 24.2, 23.2, 23.1, 21.2, 16.2 and 15.1, the remaining events appear to have an approximate linear relationship, as indicated in Fig. 7b and d. The correlation then becomes $r_p = -0.81$ and $r_s = -0.78$, or $r_p = -0.72$ and $r_s = -0.61$ after linearly detrending, which is significant at 99% confidence. Thus, in this subset there is evidence for anti-correlation beyond a simple linear trend. Again, the linear relationship is strongest for the younger half of the record, which starts at GI-14 and does not have outliers. Here, we find $r_p = -0.89$, and $r_p = -0.70$ after linearly detrending, which is significant at 99% confidence.

A corresponding linear relationship between the logarithms of interstadial durations and Antarctic temperature in different ice cores has been noted before by Buizert and Schmittner (2015). In our data we find correlations of these quantities of $r_p = 0.29$ and $r_s = 0.27$ which are not significant at 90% confidence. This disagreement comes from the fact that Buizert and Schmittner (2015) ~~lump~~ view each of the interstadials 24, 23, 21, 17, 16, 15 and 2 ~~together into~~ as one event, ~~even though they are comprised of two events~~ whereas we consider these as two events each, as suggested by the analysis of Rasmussen et al. (2014). If we remove the events 24.2, 23.2, 23.1, 21.2, 17.2, 16.2 and 15.1, we find a strong linear relationship of $r_p = 0.91$, comparable to the findings by Buizert and Schmittner (2015). It is robust to linear detrending with $r_p = 0.87$. Most of these outliers are very short events, and discarding them removes a lot of the variability of the interstadial durations, similar to lumping them together with adjacent longer events.

4.2 Stadial periods

4.2.1 Stadial duration distribution

The stadial periods are defined to start after the rapid cooling and end at the onset of the rapid warming, and their duration is thus b_1. These durations are highly variable, ranging from 20 years for GS-24.2 to 5169 years for GS-19.1, with an average of 1328 years. Due to our definition of stadials GS-24.2 is exceptionally short, because the proxy shows rapid warming again right after the rapid cooling without stabilizing. Figure 4b shows that the stadial duration distribution is skewed. The data ~~is~~ are consistent with an exponential ($p = 0.79$ with ~~Anderson-Darling~~ AD test) and a log-normal distribution ($p = 0.18$). A relative likelihood test prefers the exponential distribution by a factor of 16 over the log-normal. An exponential fit to the data is illustrated in Fig. 4b. This distribution might be relevant in this context, as it arises in the low noise limit of noise-induced escape times from asymptotically stable equilibria in dynamical systems (Day, 1987).

4.2.2 Influence of stadial levels and forcing on durations

In the following we discuss whether the stadial duration variability is influenced by other features in the data, or external factors. Among external factors, the stadial durations are best correlated with 65Nss ($r_s = -0.64$). The only DO feature that is significantly and robustly correlated with the stadial durations, is the stadial levels with $r_s = -0.43$. In Fig. 8a we show a scatterplot of the stadial levels and the logarithms of the durations. If one discards the first 6 events of the record, there is a

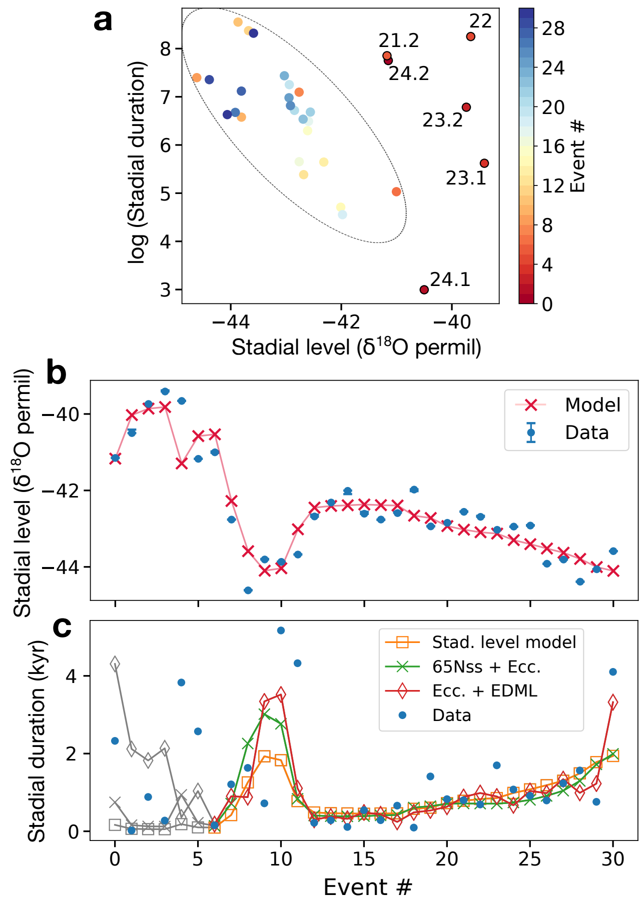


Figure 8. a) Scatterplot of stadial levels and logarithmic durations. Outliers from an approximate linear relationship are labeled. **b)** Event series of observed stadial levels and those modeled by $L_{mod} = 3.52 \cdot X_1 + 98.84 \cdot X_2 - 57.96$, where X_1 is 65Nint and X_2 the eccentricity. **c)** Models predicting the observed stadial durations (*crosses*). The first 6 events, indicated by *gray* markers, were discarded when fitting the models. The model based on predicted stadial levels from insolation (*squares*) is given by $\log(D_{mod}) = -0.90 \cdot L_{mod} - 32.18$. The second model (*circles*) is given by $\log(D_{mod}) = -0.037 \cdot X_1 - 27.11 \cdot X_2 + 25.24$, where X_1 is 65Nss and X_2 eccentricity. The third model (*diamonds*) is given by $\log(D_{mod}) = -0.90 \cdot X_1 + 75.39 \cdot X_2 + 38.71$, where X_1 is EDML and X_2 eccentricity.

linear anti-correlation of $r_p = -0.80$, or $r_p = -0.76$ after linear detrending. This could be either due to common forcing or due to the levels directly ~~controlling~~ influencing the durations.

While the stadial levels correlate well with LR04 and EDML due to a common linear trend, there is better correlation with insolation, as seen by $r_p = 0.60$ (GS-24.2 and GS-22 are outliers) for 65Nss. Removing outliers yields $r_p = 0.82$, which does not change when linearly detrending. To see whether this forcing explains most of the correlation of durations and levels, we remove a linear fit to 65Nss from each variable and find a remaining correlation of $r_p = -0.38$. Even though the significance of this correlation is unclear due to the autocorrelation of the stadial levels, this could imply that there is more information in the stadial levels about the durations than simply common insolation forcing. On the other hand, insolation components in addition to 65Nss might explain more of the observed variability.

We investigate whether multiple linear regression models with two predictors explain the stadial levels and durations significantly better. A model comprised of 65Nint and eccentricity determines the stadial levels very well ($R^2 = 0.86$), as shown in Fig. 8b. These modeled levels also correlate well with the logarithm of the stadial durations ($r_p = -0.64$ when excluding the earliest 6 events). We check whether this is a good model for the durations by linearly regressing the modeled levels onto the logarithm of the durations and exponentiating the result. In Fig. 8c we compare this to two other models that directly regress external forcings on the logarithm of the durations. None of the models fits the first 6 events adequately. Thereafter, all three models produce a similar trend. The model based on predicted stadial levels, and a model with direct forcing by 65Nss and eccentricity show similar skill with $R^2 = 0.29$ and $R^2 = 0.30$, respectively. The third model based on eccentricity and the EDML record is slightly better with $R^2 = 0.46$, mainly because it fits two of the longest stadials better. Still, all of the models fit only the overall trend and leave unexplained variability on top of it. Unless the correlation is nearly perfect, a linear correlation of the logarithm still leaves a lot of room for scatter in the original scale.

The exponential tail in the variability of the stadial durations is not a result of the modulation by the external forcings we consider. To demonstrate this, we remove the forcing influence by fitting a linear model of one or more forcings to the logarithm of the stadial durations. We obtain detrended data by adding the mean of the logarithmic data to the residuals of the fit and then exponentiating. When using 65Nss as forcing, we find $p = 0.15$ in an ~~Anderson-Darling~~ AD test on the exponential distribution. With the model of both eccentricity and 65Nss, as introduced above, we find $p = 0.29$. Thus, the distribution of the detrended data is still long-tailed and consistent with an exponential distribution.

4.2.3 Are stadials with Heinrich events special?

Besides DO events, Heinrich events are the other major mode of millennial-scale climate variability during the last glacial period. These events correspond to massive discharges of ice rafted debris found in ocean sediment cores (Heinrich, 1988), with large climatic impacts that are well-documented in numerous proxy records at various locations. While constraints on their duration and timing need to be improved, we follow Seierstad et al. (2014) for the temporal link of Heinrich events and the GICC05 chronology. This yields the set of Heinrich events H2, H3, H4, H5a, H5, H6, H7a, H7b and H8, which overlap with stadials 3, 5.2, 9, 13, 15.1, 18, 20, 21.1 and 22, respectively. Since some of these Heinrich events might be less established in the community, we also look at the reduced set of H2, H3, H4, H5 and H6.

We test whether these 'Heinrich stadials' have significantly different properties than the remaining stadials, such as longer durations, by randomly sampling 9 stadials (5 for the reduced set) from the entire set without replacement and calculating the mean duration of this subset. We repeat many times until we can estimate the probability of trials yielding a higher mean duration than the actual set of 'Heinrich stadials'. If this is less than 5% (corresponding to $p = 0.05$) we reject the hypothesis that 'Heinrich stadials' have the same mean duration as the remaining stadials at 95% confidence. This test gives essentially the same results as a one-sided t-test.

For the full (reduced) set of Heinrich events we find $p = 0.028$ ($p = 0.022$). It is not obvious whether this should be considered significant in the sense of a hypothesis that Heinrich events prolong stadials. A better statistical test is needed, since if the events were to occur randomly during the course of stadials they would naturally be found preferentially in longer stadials. We leave a resolution of this for upcoming work. Based on the idea that 'Heinrich stadials' are colder than normal, we perform a test on the stadial levels, yielding $p = 0.052$ ($p = 0.047$). Again, this is probably not significant since Heinrich events mostly occur in the younger glacial half with generally lower levels. We can reject the notion that DO events following Heinrich events are 'stronger'. A statistical test on the DO warming amplitudes yields $p = 0.102$ ($p = 0.472$), whereas a test on the interstadial durations yields $p = 0.403$ ($p = 0.583$). This might depend on the precise timing of H3, which in our analysis precedes the especially weak GI-5.1.

4.3 Abrupt warming periods

4.3.1 Warming durations

The rapid warming transitions in NGRIP as determined by our piecewise-linear fit have an average duration of 63.2 years. There is quite a large spread with a minimum duration of 15.3 years for GI-17.1 and a maximum of 179.5 years for GI-11. There is no clear trend, as we find both short and long warmings in early and later parts of the record. The distribution is skewed as seen in Fig. 4e. We find 5 transitions that last for more than a hundred years (interstadials 6, 11, 17.2, 18, 23.2). For each of them there is not only a single abrupt warming, but also a systematic departure from stadial to warmer values before, as can be seen in Fig. S1 of the supplemental material. Our algorithm includes these early warming trends into the warming transition. Clearly, other methods to define the abrupt warmings might give different results in these cases. In Rousseau et al. (2017), the transition onsets are defined by the derivative of the signal and consequently the warming transitions into interstadials 6 and 11 are reported to be much shorter. Given our definition of abrupt warmings, we can at least argue that the longest warming transitions are not a result of noise, because in our fit of the GRIP record the same transitions are also among the longest and are clearly above average.

Within the framework of our analysis, we cannot identify any DO cycle features, external forcings, or combinations thereof that can explain a significant part of the variability in the warming durations. Thus, we aim to infer something about the mechanism of the warming transitions from the distribution of their durations. ~~To assess which distributions are consistent with the data we use the Anderson-Darling test. Using the Cramer-von Mises or Kolmogorov-Smirnov test yields qualitatively unchanged results for all tests reported in this section. The Anderson-Darling test shows~~ AD tests show that the log-normal (~~p=0.63~~$p = 0.63$), Gumbel (~~p=0.053~~$p = 0.053$) and inverse Gaussian (~~p=0.95~~$p = 0.95$) distributions cannot be rejected at 95% confidence by the data. A fit with the Gumbel distribution is illustrated in Fig. 4e. By computing the relative likelihood from the Akaike information criterion, we find that the inverse Gaussian distribution is 9.7 times more likely than the Gumbel distribution, and the log-normal distribution is 7.6 times more likely than the Gumbel distribution. We cannot choose in between log-normal and inverse Gaussian with any confidence.

4.3.2 A model for the stadial-interstadial transition

In the following we compare the warming durations to what is expected in the framework of noise-induced transitions in multi-stable systems. The DO warming durations are much shorter than the time spent in the stadial state. If we consider the stadial-interstadial transition as a noise-induced transition from one metastable state to another, starting at the stadial onset, most of the time is spent in the vicinity of the stadial state. The part of the trajectory that leaves this vicinity for the last time and then moves towards the other state (interstadial) is referred to as the reactive trajectory. Because of the high noise level in the record, an unknown part of which is non-climatic or regional and changes over time, we do not estimate reactive trajectories by defining neighborhoods of two metastable states. Instead, we believe the warming periods obtained by our piecewise-linear fit are reasonable estimates. Figure 9a illustrates the reactive trajectory (in green) leading up to GI-20. In Fig. 9b the different parts of this stadial-interstadial transition are projected onto an arbitrary potential that features two metastable states. For overdamped motion driven by additive noise in such a potential, it has been proven that the reactive trajectory durations converge to a Gumbel distribution in the zero noise limit (Cérou et al., 2013). Similarly, there is numerical evidence for the Gumbel distribution applying to one-dimensional spatially extended systems for low noise (Rolland et al., 2016). Because in our data we cannot separate true climatic noise, potentially driving the observed large-scale climate transitions, from other types of noise, it is hard to say whether a low noise condition is met and a Gumbel distribution should be expected.

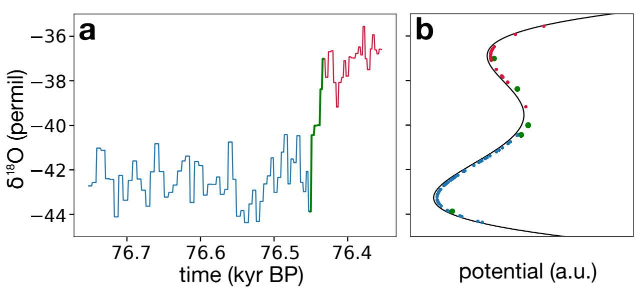


Figure 9. a) Stadial-interstadial transition leading up to GI-20 (*red*) including our estimate of the so-called reactive part of the trajectory (*green*) preceded by 350 years of the stadial GS-21.1. **b)** Data points of the same time series projected onto an arbitrary one-dimensional potential function with two minima as conceptual model for the transition.

With a small numerical experiment we address the case of finite noise levels and small sample sizes. We use stochastic motion in a double well potential as generic model for a noise-induced transition out of a metastable state to another. It is given by the stochastic differential equation $dX_t = \left(-\frac{dV(X_t)}{dx}\right) dt + \sigma dW_t$, with the potential $V(x) = x^4 - x^2$ and the Wiener process W_t. For zero noise, there are two fixed points at $x = -1$ and $x = 1$. We initialize the system at $x = -1$ and repeatedly collect reactive trajectories, which start when they last leave $x < -0.9$ and end as they enter $x > 0.9$. Small samples of 31 reactive trajectory durations are indeed typically consistent with a Gumbel distribution for a range of different noise levels, but can be consistent with other distributions, too.

To show this, we collect p-values of ~~Anderson-Darling~~ AD tests on many small samples. For the Gumbel distribution at a low noise level of $\sigma = 0.00045$, 96.3% of the p-values are above 0.05. Thus, in this case, very rarely a sample of 31 reactive trajectory durations is rejected by a hypothesis test on the Gumbel distribution. For a higher noise level of $\sigma = 0.5$, still 80.1% yield $p > 0.05$. However, the log-normal distribution fits equally well with 95.4% (93.6%) yielding $p > 0.05$ for $\sigma = 0.00045$ ($\sigma = 0.5$). The distribution that most reliably fits the data is the inverse Gaussian with >99.9% (>99.9%) yielding $p > 0.05$ for $\sigma = 0.00045$ ($\sigma = 0.5$), despite the fact that in the zero noise limit the correct distribution is Gumbel. It has been noted that the inverse Gaussian also fits well for large sample sizes (Cérou et al., 2011). Even non-skewed distribution can be consistent with the samples, as seen for the Gaussian distribution, with 55.1% (22.8%) yielding $p > 0.05$ for $\sigma = 0.00045$ ($\sigma = 0.5$). Similar values are obtained for the logistic distribution.

These results imply that from a small sample of 31 reactive trajectories we cannot reliably identify the true distribution and thus a potential mechanism. Still, the data are at least consistent with the expected behavior of noise-induced escape from a metastable state. Clearly, other simple mechanisms can be consistent with the data, too. For example, as mentioned above, the inverse Gaussian is the distribution of time elapsed for a Brownian motion with drift to reach a fixed level.

4.3.3 Warming amplitudes

The average amplitude of the rapid warmings is 4.2 permil, with most events clustering around this value. The most extreme values are 7.1 permil for GI-19.2 and 1.7 permil for GI-5.1. The latter is not surprising, because GI-5.1 is almost not visually discernible as an event in the δ^{18}O series. We find an anti-correlation of the warming amplitudes and the preceding stadial levels. When discarding GI-5.1, the correlation is significant at 99% confidence with $r_s = -0.63$, which is largely due to GI-19.2 being preceded by a very deep stadial, and GS-23.1 and GS-22, which are preceded by very shallow stadials, as they happen early in the glacial. When discarding these events the remaining correlation is still significant at 99% confidence with $r_s = 0.50$. Thus, to some degree, the warming amplitudes are predictable in a statistical sense: When residing in a shallow stadial, the amplitude of the next DO warming will be small, and vice versa for a deep stadial. We also assess whether the variability can be explained by external forcing. Our analysis does not show a relationship between the DO event amplitudes and global ice volume (LR04), as has been proposed by McManus et al. (1999); Schulz et al. (1999). It should be noted, however, that these studies have a quite different notion of DO event amplitudes. Our approach, based on fitting high resolution data, seems well suited to estimate the actual amplitude of rapid transitions, as opposed to low-pass filtering that reduces the amplitude of shorter events. Instead of ice volume, we find a correlation with 65Nint of $r_p = -0.36$ and $r_s = -0.31$, which is significant at 90% confidence. However, the correlation is visually not striking. Removing GI-19.2, which occurs close to an insolation minimum, yields a correlation that is not significant at 90% confidence.

5 Discussion

This work presents a ~~consistent fitting routine allowing to extract robust features of DO events from noisy,~~ piecewise-linear fit for the high-resolution ~~ice core data, such as the~~ NGRIP δ^{18}O record ~~. The algorithm converges to a continuous piecewise-linear~~

~~fit of the whole time series~~of the last glacial period, where each DO event ~~cycle is given by a constant stadial period, an abrupt warming period, a gradually cooling interstadial period and an abrupt cooling period. The fit is satisfactory in the sense that each event receives a reasonable~~ receives a saw-tooth shaped fit~~, which is characteristic for the ensemble of DO events as a whole~~. Not for all events this is necessarily the overall best piecewise-linear fit~~. For example,~~ , since there are transitions that do not have a significant rapid cooling period at the end of the interstadials, but rather cool gradually until reaching roughly constant stadial values. In Sec. 3.1 we showed that 14 out of the 31 DO events analyzed in this study do not strictly follow the saw-tooth shape that is often reported as being generic for all DO events.

The uncertainties of the fit parameters are assessed ~~by using~~ with a bootstrap resampling technique (Sec. 3.2) and alternatively by comparison to a different ice core (GRIP, Sec. 3.3). The ~~average absolute deviations of the GRIP and NGRIP features and the average standard deviations obtained by bootstrapping are of very similarmagnitude. This gives us~~ results are very similar, giving confidence in the ~~validity of the~~ uncertainty estimates of the ~~latter method. From the uncertainties~~ former method. However, it follows that some ~~of the~~ shorter time scale features have to be taken with care, such as the rapid warming durations. Here, not all individual values might be reliable. However, the comparison with GRIP shows that the overall trends and distributions, also of the shorter time scale features, are robust. Still, different methods or models to define the features might alter the results. As an example, our piecewise-linear method yields quite different estimates of the abrupt warming durations as compared to Rousseau et al. (2017), where abrupt warmings are defined by an estimated derivative of the time series. Our results have an average absolute deviation of 25 years (26 years) compared to their algorithmically (visually) determined warming durations starting at GI-17.1.

~~We subsequently analyzed different features that describe~~ From the fit parameters, we derived different features characterizing each DO cycle and ~~can be derived from the fit parameters. These include the proxy levels in the stadials, at the interstadial maxima and ends, the durations and rates of warming and cooling periods, as well as stadial and interstadial durations. In general, all features except for the proxy levels~~ performed an extensive statistical analysis on these and various forcings. This could provide a basis for future modeling studies to compare characteristics of modeled DO cycles to the ice core data. Even though most features develop rather irregularly from event to event, ~~as shown by the absence of significant autocorrelation, and many of them are very broadly distributed. We evaluated which distributions and corresponding processes describe the individual features best. This can give some insight into the nature of the mechanisms giving rise to the abrupt climate changes. Furthermore, we investigated whether the variability in some features can be explained by other features, external forcings or combinations of them. This is done by a brute force search of significant correlations, multiple linear regression models and clusterings, which is subsequently narrowed down by assessing robustness to outliers and trends. We synthesized the most important findings in terms of the interstadial, stadial, and abrupt warming periods of the DO cycle, as summarized in the following~~and show only weak correlations with each other and external forcings, some useful information can be extracted from the analysis.

The interstadial periods have highly variable durations and are characterized by a roughly linear cooling with rates that vary strongly from one DO cycle to the next (Sec. 4.1.1). ~~The cooling rates clearly determine the interstadial durations , as opposed to the cooling amplitudes, which cannot robustly explain the variability of the durations. Interstadial durations and cooling~~

~~rates are consistent with a simple inverse relationship. As a result, the interstadial durations are determined~~ Because these durations are much more strongly governed by the cooling rates than by the cooling amplitudes they can be predicted to a good approximation as soon as the cooling rates have stabilized. We estimate from the data that for most DO cycles this happens within the first 150 to 350 years of the interstadial (Sec. 4.1.3). Our interpretation is that after interstadial onset a large-scale reorganization of the climate system takes place on a time scale, which, even though very different from event to event, can be inferred from the cooling rate and stays constant throughout the interstadial. This reorganization is the major driving force that determines when the interstadial-stadial transition takes place, which as a result cannot be purely noise-induced.

~~The influence of external factors on the large variability of the interstadialfeatures is assessed and compared to previously proposed forcing mechanisms (Sec. 4.1.4). Based on the GISP2 ice core record of the younger half of the last glacial it has been proposed ,~~ External forcing might explain the large variation of this time scale from event to event, as proposed by Schulz (2002), who argues that the interstadial cooling rates are controlled by global sea level~~(Schulz, 2002). While we can confirm this finding by observing a linear correlation of the logarithm of the cooling rates with the LR04 record for global ice volume, the relation seems~~ , based on the GISP2 ice core record of the younger half of the last glacial. We find this relation to be weak in the older half of the glacial due to a ~~handful of outliers . The interstadial duration might be controlled by~~ number of outliers (Sec. 4.1.4). A physical pathway for such a forcing might be influences of global ice volume on the strength and stability of the interstadial (strong) mode of the ~~Atlantic Meridional Overturning Circulation (AMOC)~~AMOC. However, ~~contrary to our finding, the influence of ice volume on AMOC stability reported by studies with globally coupled models is often such that increases~~ studies with coupled climate models show that enlargement of Northern Hemisphere ice sheets actually ~~enhance~~ enhances the stability of the strong AMOC state ~~, which would intuitively~~ (Zhang et al., 2014; Ullman et al., 2014; Muglia and Schmittner, 2015). Intuitively, this would result in longer interstadials~~(Zhang et al., 20~~ which is in contrast to what the ice core data suggest. This has been addressed by Buizert and Schmittner (2015), where Southern Ocean processes are invoked to influence interstadial durations. We find that the ~~control~~ correlation of Antarctic temperature ~~on~~ and interstadial durations reported in this study is only valid if certain outliers are discarded. Finally, for the younger half of the glacial ~~,~~ starting at GI-14, we find that ~~the~~ atmospheric CO_2 ~~record is the best~~ is a much better predictor of the cooling rates.

~~In our analysis, certain interstadials frequently show up as outliers. These include the short events 23.2 and 21.2, which occur early in the glacial and are surrounded by longer interstadials, as well as the events 18, 17.2 and 15.2, which are short, but do not show a clear gradual cooling. Their presence either showcases the strong irregularity and variability of the processes underlying~~ A sensitive dependence of the strong AMOC state to CO_2 has been verified in model experiments by Kawamura et al. (2017). However, more model experiments with active carbon cycle are needed to clarify whether CO_2 should be considered as forcing or response to the DO cycle~~, or could indicate that not all transitions are caused by the same trigger. We find more groups of outliers using clustering analysis. These could be studied more thoroughly, and compared with other paleoclimatic indicators~~. Yet other model experiments showed that changes in CO_2 could in fact even be the trigger for DO-type transitions (Zhang et al., 2017).

The stadials have different properties compared to the interstadials, going beyond the approximately constant temperature within stadials. The duration distribution closely resembles an exponential, and ~~is thus consistent with noise-induced escape from a metastable state to another (Sec. 4.2.1). The large dispersion of this distribution~~ its large dispersion cannot be explained by external forcing alone. Instead, the distribution is still consistent with an exponential after detrending with the best fit to insolation forcing (Sec. 4.2.2). ~~Although there are different possibilities to obtain a good fit of the trend of the stadial durations, a forcing by insolation seems most plausible. We additionally find indications for a control of the stadial durations by the levels, but it is difficult to conclude from our data whether there is a true causal link, or merely common insolation forcing on both variables.~~

~~The piecewise-linear fit furthermore gives estimates for the amplitudes, rates and durations of the rapid DO warmings. We find no evidence for influence of other DO features or external factors on the warmings. As an exception, the warming amplitudes are anti-correlated with the stadial levels~~ The stadial durations are thus consistent with noise-induced escape from a metastable state to another (Sec. ~~4.3.3). Consequently, they may be predicted in a probabilistic sense. There is considerable variability of the warming durations, and we find the distribution to be~~ 4.2.1). This is corroborated by our finding that the distribution of DO warming durations is consistent with the durations of so-called reactive trajectories in systems with noise-induced escapes in between multiple metastable equilibria (Cérou et al., 2013; Rolland et al., 2016) (Sec. 4.3.2).

Thus, ~~our analysis suggests that both the stadial period durations and subsequent warming durations are consistent with~~ the stadial-interstadial transition ~~as~~ is consistent with a noise-induced escape from a metastable state~~. This is different from the interstadial-stadial transition, which occurs in a more predictable fashion, because the linear cooling rates anticipate the interstadial durations. Additionally, the interstadial durations are not consistent with an exponential distribution. It has, however, recently been suggested~~, and thus with spontaneous, unforced climate transitions, such as the ones observed in Drijfhout et al. (2013); Kleppin et al. (2015). However, evidence for a different scenario might arise with new data or analyses, as in the studies by Rypdal (2016) and Boers (2018) who suggest that there is a bifurcation ~~before DO events~~ in a fast climate sub-system ~~of the climate , which was based on evidence for~~ before DO warmings, evidenced by critical slowing down in the high-frequencies of the ice core record prior to ~~a significant number of DO warming transitions (Rypdal, 2016; Boers, 2018)~~some events. If this is the case, it would mean that ~~there is some predictability of the warming transitions, too. Additionally, we find that~~ the transitions are not purely noise-induced, but predictable to some degree and potentially part of a self-sustained oscillation such as in the model experiments by Vettoretti and Peltier (2016) and Klockmann et al. (2018).

Our analysis, however, allows for the interpretation of the DO cycle as a manifestation of an excitable system, as proposed by Ganopolski and Rahmstorf (2002) and Timmermann et al. (2003), with a noise-induced transition out of the stadial state to the marginally unstable interstadial state, and a deterministic excursion back to the stadial state. However, the vastly different time scales for this excursion still need an explanation.

The mechanisms underlying warming and cooling transitions are furthermore likely to be different, because the influence of external forcing is different for stadial and interstadial periods, with more evidence for insolation forcing on stadials and ice volume or CO_2 on interstadials, which is related to the findings by Lohmann and Ditlevsen (2018). Except for a common forcing envelope of stadial and interstadial levels, there is no strong relationship in between features across the different periods of the DO cycle.

In this work we rely on the classification of Greenland ice core centennial to millennial variability into a set of DO events by Rasmussen et al. (2014). Notably, this classification includes short events such as 23.2 and 21.2, which occur early in the glacial and are surrounded by the longest interstadials, as well as the events 18, 17.2 and 15.2, which are short, but do not show a clear gradual cooling. In our analysis, these interstadials frequently show up as outliers. Their presence either showcases the strong irregularity and variability of the processes underlying the DO cycle, or could indicate that the events are caused by a different trigger and do not represent large-scale reorganizations of the climate system in the same way as longer events do. Some conclusions of our work might change if all short events were systematically discarded. They do not appear to be local to Greenland, since they are seen in atmospheric methane records (Rhodes et al., 2017) and speleothem records from the Alps (Boch et al., 2011; Moseley et al., 2014) and Asia (Cheng et al., 2016). However, their significance in comparison to longer DO events is an open problem that needs to be evaluated as more precisely dated, high-resolution records from outside of Greenland become available.

6 Conclusion

We developed ~~an iterative~~ a method to fit a continuous piecewise-linear waveform to the entire last glacial NGRIP δ^{18}O record ~~, which converges well. By using parameter constraints, we~~ that can fit a characteristic saw-tooth shape to every DO event. However, we find that for many of the transitions this is ad-hoc. Almost half of the events do not show a distinct and significant rapid cooling episode after the more gradual interstadial cooling. An analysis of the DO event features derived from the fit confirms the irregularity and randomness that is evident from visual inspection of the record. There is hardly any evidence for relationships linking the features that describe the stadial, interstadial and abrupt warming periods, except for a common envelope that governs the stadial and interstadial levels via external forcing. A statistical analysis hints at different mechanisms underlying warming and cooling transitions. This ~~manifests itself in different distributions and external influences of the stadials and interstadial durations~~follows from the distributions of the stadial and rapid DO warming durations on the one hand, and the interstadial durations on the other hand. It is furthermore supported by the different importance of CO_2, ice volume and insolation forcing to explain the stadial and interstadial properties, as well as the fact that the interstadial durations can be predicted to some degree by the linear cooling rates shortly after interstadial onset.

Appendix A: Iterative algorithm to fit piecewise-linear model

Algorithm 1 Pseudocode for fitting algorithm

1: **while** $j < J$ **do**
2:   **if** $j = 0$ **then**
3:     $\{t_i^b\} = \{t_i^{b0}\}$; $\{t_i^e\} = \{t_i^{e0}\}$
4:   **else**
5:     Set Stadial durations: $\{D_i^{St}\} = \{b_{1,i} + t_i^b - t_{i-1}^b - b_{4,i-1}\}$
6:     Set Interstadial durations: $\{D_i^{Is}\} = \{b_{4,i} - b_{1,i}\}$
7:     Set Stadial beginning times: $\{t_i^b\} = \left\{\sum_{n=0}^{i-1}(D_n^{St} + D_n^{Is})\right\}$
8:     Set Stadial end times: $\{t_i^e\} = \left\{\sum_{n=0}^{i} D_n^{St} + \sum_{n=0}^{i-1} D_n^{Is}\right\}$
9:   **end if**
10:   Define Stadial levels: $\{l_i^s\} = \left\{\left\langle X_{t_i^b,\ldots,t_i^e}\right\rangle\right\}$
11:   Cut into segments: $\{s_i\} = \left\{X_{t_i^b,\ldots,t_{i+1}^e}\right\}$
12:   **while** $i < N$ **do**
13:     **if** j=0 **then**
14:       Initial conditions: $\theta_i^* = \theta_i^0$
15:     **else**
16:       Initial conditions: $\theta_i^* = \theta_i$
17:     **end if**
18:     Find optimal θ_i^{new} of segment s_i with θ_i^*, l_i^s and l_{i+1}^s
19:     $i = i + 1$
20:   **end while**
21:   Update parameters $\{\theta_i\} = \{\theta_i^{new}\}$
22:   $j = j + 1$
23: **end while**

In the following, we detail the optimization procedure to find the best saw-tooth shaped fit for each event, i.e., line 18 of the algorithm above. To determine the 6 parameters at each transition, we minimize the root mean squared deviation of the fit from the time series segment. Due to the high noise level, there are many local minima in this optimization problem. Thus, either a brute-force parameter search on a grid or an advanced algorithm is needed to find a global minimum. We chose an algorithm called basin-hopping, which is described in Olson et al. (2012) and is included in the Scientific Python package scipy.optimize, where it can also be customized. The basic idea of the algorithm is the following: Given a initial coordinates in terms of the parameter vector θ_0, one searches for a local minimum of the goal function $f(\theta)$, e.g., with a Newton, quasi-Newton or other method. The argument to this local minimum θ_n is then randomly perturbed by a Kernel to yield new coordinates $\theta*$, which are the starting point of a new local minimization. Next, there is a Metropolis accept or reject step: We accept the argument of the local minimization θ_{n+1} as new coordinates if the local minimum is deeper than the previous one $f(\theta_{n+1}) < f(\theta_n)$, or else with probability $e^{-(f(\theta_{n+1})-f(\theta_n))/T}$, where T is a parameter relating to the typical difference in depth of adjacent local minima. Now we go back to the perturbation step either with old coordinates θ_n or, if accepted, with new coordinates θ_{n+1}, and repeat. The iterative procedure is repeated for a large number of iterations and the result is the argument to the lowest function value found.

Within basin-hopping, one has the freedom of choosing any local minimizer as well as perturbation Kernel. These have to be adapted to our optimization problem. We have several constraints on the parameters that need to be satisfied by the optimization. For instance, we demand that all segments of the fit are present and do not overlap ($b_1 < b_2 < b_3 < b_4$). Other constraints ensure that the characteristic shape of DO events is fit as good as possible for all events. Among other things, we thus demand the gradual slope to be significantly longer and less steep than the fast cooling transition at the end of an interstadial. An overview of all the constraints we used is given further below. To satisfy them, we chose a multivariate Gaussian perturbation Kernel, which is truncated at the respective parameter constraints. The local minimizer choice requires further consideration. Our goal function landscape is very rough and not differentiable. Thus, methods like gradient descent give very poor results in our case. A method that does not depend on derivatives and can handle constraints is called Constrained Optimization by Linear Approximation (COBYLA), and we found it to work well in our case.

Two hyper-parameters have to be specified in the basin-hopping algorithm: The variance of the perturbation Kernel, and the parameter T used in the Metropolis criterion. These should both be comparable to typical differences in goal function (temperature) and arguments (perturbation width) of neighboring local minima in the minimization problem. We chose these parameters empirically by observing how the goal function changes as we slightly change the fit. Although this varies significantly from transition to transition, we determined single values as a compromise for all transitions. For the Kernel variance in the directions of $b_{1,2,3,4}$ we chose a value of 15, and for s_1 and s_2 we chose 0.004 and 0.0015, respectively.

The following list contains all constraints used in the optimization problem in order to ensure convergence of the algorithm to a fit within the qualitative limits of the desired characteristic waveform. Specifically, constraints 3 and 4 shall guarantee that there is a distinction in between gradual cooling and rapid cooling at the end of an interstadial. With these constraints we can prevent that our algorithm splits an interstadial in half with two very similar slopes, which can easily happen because there are interstadials which arguably have a rather gradual cooling all the way down to the next stadial with no easily discernable steep cooling at the end. The lower limit of constraint 6 shall help to only fit to the steep part of warming transitions, which might have a slight warming prior to it. The upper limit of constraint 7 is needed in order to force a small negative slope on very short transitions which otherwise could also be viewed as plateaus.

1. No overlap of segments:
   $b_2 > b_1$, $b_3 > b_2$ and $b_4 > b_3$

2. Gradual slope cannot go below following stadial level l^s_{i+1}:
   $s_1(b_2 - b_1) + s_2(b_4 - b_3) > l^s_{i+1}$

3. Gradual slope must be twice as long as steep drop:
   $b_3 - b_2 > 2 \cdot (b_4 - b_3)$

4. Drop at end of interstadial must be at least twice as steep as gradual slope:
   $2 \cdot s_2 < \frac{s_1(b_2-b_1)+s_2(b_3-b_2)-l^s_{i+1}+l^s_i}{b_4-b_3}$

5. Stadial period not shorter than 20 years:
   $b_1 > 20$, $b_2 > 20$, $b_3 < (D^{St} + D^{Is} - 20)$
   and $b_4 < (D^{St} + D^{Is} - 20)$

6. Limit steepness of up-slope (permil y^{-1}):
   $0.02 < s_1 < 1.5$

7. Limit steepness of down-slope (permil y^{-1}):
   $-0.3 < s_2 < -0.0001$

For the basin-hopping algorithm we use a multivariate Gaussian Kernel of fixed variance with $\sigma_{b_1} = 15$, $\sigma_{b_2} = 15$, $\sigma_{b_3} = 15$, $\sigma_{b_4} = 15$, $\sigma_{s_1} = 0.004$ and $\sigma_{s_2} = 0.0015$.

Appendix B: Convergence of iterative fitting routine

We repeatedly run our iterative fitting routine and monitor whether the individual parameters converge, so that a consistent fit is obtained in the end. Critical for obtaining a consistent fit is that the stadial levels do not change substantially, as explained in the Methods section. In Fig. B1a we show the evolution over 40 iterations of the incremental deviations of the stadial levels compared to the previous iteration. Most stadial levels converge rapidly so that their increments stay below 0.05 permil. Two short stadials keep fluctuating until around iteration 20 before they converge. Because of the convergence of stadial levels, we consider our fit to be consistent. Furthermore, the best fit parameters are robust, which can be seen in Fig. B1b. Here, we show the average absolute incremental deviations to the break point parameters at each iteration. After 15 iterations the procedure is stable, with average incremental deviations of roughly 0.4 years for b_1 and b_2 to 0.5 years for b_3 and b_4, which result from the stochastic fitting algorithm. Note that these values are already well below the smallest sample spacing of the original unevenly spaced time series.

Appendix C: Uncertainty estimation of fitting parameters

Because of the nature of the data, care has to be taken when generating synthetic data. The properties of the data changes throughout the record and are also quite different in between adjacent stadials and interstadials. Stadials have both a larger variance and a larger effective sample spacing in time than the interstadials. For this reason, synthetic data will be created for each stadial and interstadial period individually. The original data is unevenly spaced, which would provide difficulties on its

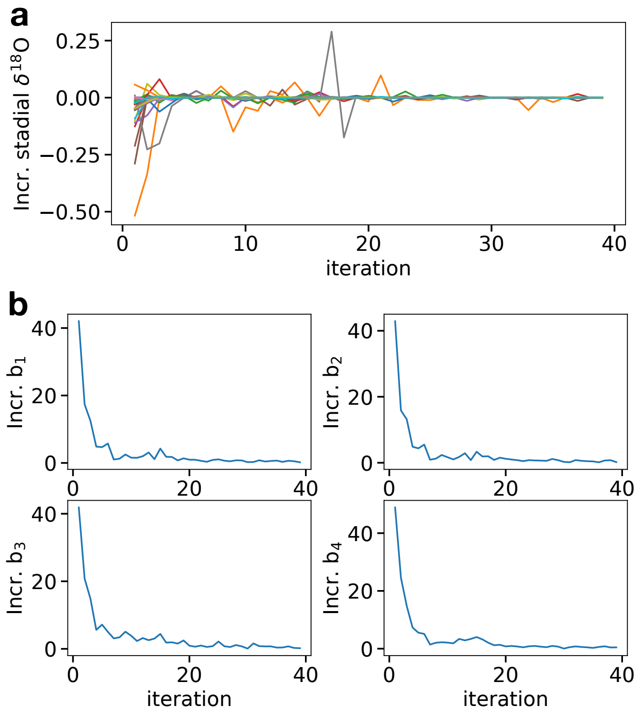


Figure B1. a) Evolution of the incremental change of all stadial levels compared to the previous iteration for all 40 iterations of the fitting routine. **b)** Average over all transitions of the incremental change (absolute value) of the break point parameters b_1, b_2, b_3 and b_4.

own, while our data is nearest-neighbor interpolated and oversampled to a 1-year resolution. This means that there typically are multiple neighboring point with the same value, making it challenging to find a valid autoregressive or ARMA model for the residuals to generate synthetic data. Instead, we use a block bootstrap resampling technique to keep all relevant structure in the data. We chose a simple block bootstrap where non-overlapping blocks of fixed length of the time series are randomly ordered, because it preserves the correct mean of the individual stadial and interstadial residuals. More involved methods, such as the stationary bootstrap, could be applied, but it likely will not change any of our conclusions.

In the following, we present the procedure for uncertainty estimation. We denote the original data time series of a given transition as $\{X_t\}$, the fit obtained by the data as $\{Y_t\}$ and the residuals to the fit as $\{R_t\} = \{X_t - Y_t\}$. We furthermore use the break points $b_{1,2,3,4}$ obtained in the fit of this transition.

1. Divide the residuals into four segment R_t^i at the breakpoints:
   $\{R_t^i\} = \{R_t\}_{t=b_{i-1}...b_i}$ for $i = 1...4$, where $b_0 = 0$.
   Denote the length of $\{R_t^i\}$ as n_i.

2. For each segment: Divide into n_i/l blocks of length l.
   Append remaining data points to last block if n_i/l non-integer.
   The block length l is determined by the length of the segment, as explained below.

3. For each segment: Randomly sample blocks without replacement and concatenate until all blocks have been used. This yields resampled segments $\{\bar{R}_t^i\}$.

4. Concatenate the four resampled segments and add the fit to get synthetic data
$\{X_t^*\} = \{Y_t\} + \{\{\bar{R}_t^1\}, \{\bar{R}_t^2\}, \{\bar{R}_t^3\}, \{\bar{R}_t^4\}\}$

5. Fit $\{X_t^*\}$ to a piecewise-linear model with the basin-hopping algorithm.

6. Repeat from step 2.

In order to also be able to resample the shortest segments, while also preserving the autocorrelative structure in all but the shortest segments, we choose the following scheme for the block length l: If the segment length n_i is larger than 40 years, choose $l = 20$. If $40 > n_i \geq 20$ choose $l = 10$. If $20 > n_i \geq 10$ choose $l = 5$. If $n_i < 10$ do not resample and simply return original segment. The scheme has been determined by looking at the residuals of each segments in all transitions and observing that the autocorrelation drops to non-significant values for all segments after 10-15 years. It thus seems reasonable to use the same block length rule for all transitions and segments.

Appendix D: Correlation analysis of features and forcings

In the following, we give an overview of the pairwise correlations in between different features and forcings. We show the Spearman correlation coefficients of all tests and their significance in Fig. ~~S4 of the Supplementary Material~~5. Considering Spearman correlations with $p < 0.05$, we find 81 positives at 95% and 50 positives at 99% confidence, which is clearly more than expected by chance. However, as detailed in the Methods Section, many of these are due to construction and will not be discussed here. We will furthermore omit correlations which are not robust due to the presence of outliers.

Among features within the same DO cycle, the three different levels yield a strong correlation with each other. However, the significance is overestimated due to their autocorrelation, and after linear detrending, the correlations are not significant anymore. Thus, the correlation comes mostly from a common trend associated with evolution of the background climate state during the glacial. Furthermore, we find significant correlations of fast cooling, gradual cooling and warming amplitudes, and a correlation of interstadial levels and gradual cooling amplitudes. This implies a certain consistency of DO cycles, where a large amplitude warming is typically also followed by a large amplitude cooling (gradual and/or fast). This is equivalent to the fact that the stadial levels are autocorrelated. In Sec. 4 we furthermore discuss the correlation of the gradual cooling durations with the gradual cooling amplitudes and rates, as well as the correlation of the stadials levels with the stadial durations and warming amplitudes.

For features in adjacent DO cycles, we do not expect any true positives a priori, because no features are related by construction. Significant correlations at 99% confidence are only found for the levels. Due to their autocorrelation, the significance determined by permutation tests are not reliable, however. Detrending shows that the correlations are dominated by a common linear trend due to the slowly changing background climate state. The remaining 8 correlations significant at 95% confidence could either be false positives, or a result of common external forcing. This is because 7 of the 8 correlations involve the levels, which are clearly influenced by forcing, as detailed below.

We furthermore correlate the features with all forcings at the onset times of the respective periods within the DO cycles. The tests indicate clearly more significant correlations than expected by chance. However, due to autocorrelation, the significance is overestimated by permutation tests. In particular, the levels yield significant correlation with most forcings, however, both are autocorrelated. By linearly detrending and discarding outliers where necessary, we find that the interstadial levels are best correlated with LR04, EDML and CO_2, the interstadial end levels with 65Nss and precession, and the stadial levels with LR04, 65Nint, 65Nss, obliquity and eccentricity. Additional significant correlations we found are discussed in Sec. 4 and include those of gradual cooling rates with the LR04 and CO_2 forcings, as well as those of stadial durations and different insolation forcings.

Competing interests. The authors declare no competing interests.

Acknowledgements. We gratefully acknowledge discussions of this work with Sune O. Rasmussen. 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

Bereiter, B., Eggleston, S., Schmitt, J., Nehrbass-Ahles, C., Stocker, T. F., Fischer, H., Kipfstuhl, S., and Chappellaz, J.: Revision of the EPICA Dome C CO_2 record from 800 to 600 kyr before present, Geophys. Res. Lett., 42, 542–549, 2015.

Boch, R., Cheng, H., Spötl, C., Edwards, R. L., Wang, X., and Häuselmann, P.: NALPS: a precisely dated European climate record 120–60 ka, Clim. Past, 7, 1247–1259, 2011.

Boers, N.: Early-warning signals for Dansgaard-Oeschger events in a high-resolution ice core record, Nature Comm., 9, 2556, 2018.

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

Capron, E., Landais, A., Chappellaz, J., Schilt, A., Buiron, D., Dahl-Jensen, D., Johnsen, S. J., Jouzel, J., Lemieux-Dudon, B., Loulergue, L., Leuenberger, M., Masson-Delmotte, V., Meyer, H., Oerter, H., and Stenni, B.: Millennial and sub-millennial scale climatic variations recorded in polar ice cores over the last glacial period, Clim. Past, 6, 345–365, 2010.

Cérou, F., Guyader, A., Lelièvre, T., and Pommier, D.: A multiple replica approach to simulate reactive trajectories, J. Chem. Phys., 134, 054 108, 2011.

Cérou, F., Guyader, A., Lelièvre, T., and Malrieu, F.: On the length of one-dimensional reactive paths, ALEA, Lar. Am. J. Probab. Math. Stat., 10, 359–389, 2013.

Cheng, H. et al.: The Asian monsoon over the past 640,000 years and ice age terminations, Nature, 534, 640–646, 2016.

Day, M. V.: Recent progress on the small parameter exit problem, Stochastics, 20, 121–150, 1987.

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.

EPICA Community Members: Stable oxygen isotopes of ice core EDML, PANGAEA, p. https://doi.org/10.1594/PANGAEA.754444, 2010.

Ganopolski, A. and Rahmstorf, S.: Rapid changes of glacial climate simulated in a coupled climate model, Nature, 409, 153–158, 2001.

Ganopolski, A. and Rahmstorf, S.: Abrupt Glacial Climate Changes due to Stochastic Resonance, Phys. Rev. Lett., 88, 038 501, 2002.

Gkinis, V., Simonsen, S. B., Buchardt, S. L., White, J. W. C., and Vinther, B. M.: Water isotope diffusion rates from the NorthGRIP ice core for the last 16,000 years – Glaciological and paleoclimatic implications, Earth and Plan. Sc. Lett., 405, 132–141, 2014.

Heinrich, H.: Origin and Consequences of Cyclic Ice Rafting in the Northeast Atlantic Ocean During the Past 130,000 Years, Quaternary Research, 29, 142–152, 1988.

Henry, L. G., McManus, J. F., Curry, W. B., Roberts, N. L., Piotrowski, A. M., and Keigwin, L. D.: North Atlantic ocean circulation and abrupt climate change during the last glaciation, Science, 353, 470–474, 2016.

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

Johnsen, S. J., Clausen, H. B., Dansgaard, W., Gundestrup, N., Hammer, C. U., et al.: The δ^{18}O record along the Greenland Ice Core Project deep ice core and the problem of possible Eemian climatic instability, J. Geoph. Research, 102, 26 397–26 410, 1997.

Kageyama, M., Merkel, U., Otto-Bliesner, B., Prange, M., Abe-Ouchi, A., Lohmann, G., Ohgaito, R., Roche, D., Singarayer, J., Swingedouw, D., and Zhang, X.: Climatic impacts of fresh water hosing under Last Glacial Maximum conditions: a multi-model study, Clim. Past, 9, 935–953, 2013.

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.

Klockmann, M., Mikolajewicz, U., and Marotzke, J.: Two AMOC States in Response to Decreasing Greenhouse Gas Concentrations in the Coupled Climate Model MPI-ESM, J. Climate, 31, 7969–7984, 2018.

Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A. C. M., and Levrard, B.: A long term numerical solution for the insolation quantities of the Earth, Astronomy and Astrophysics, pp. 261–285, 2004.

Li, C., Battisti, D. S., and Bitz, C. M.: Can North Atlantic Sea Ice Anomalies Account for Dansgaard–Oeschger Climate Signals?, J. Climate, 23, 5457–5475, 2010.

Lohmann, J. and Ditlevsen, P. D.: Random and externally controlled occurrences of Dansgaard–Oeschger events , Clim. Past, 14, 609–617, 2018.

Lynch-Stieglitz, J.: The Atlantic Meridional Overturning Circulation and Abrupt Climate Change, Annu. Rev. Mar. Sci., 9, 83–104, 2017.

McManus, J. F., Oppo, D. W., and Cullen, J. L.: A 0.5-Million-Year Record of Millennial-Scale Climate Variability in the North Atlantic, Science, 283, 971–975, 1999.

Moseley, G. E., Spötl, C., Svensson, A., Cheng, H., Brandstätter, S., and Edwards, R. L.: Multi-speleothem record reveals tightly coupled climate between central Europe and Greenland during Marine Isotope Stage 3, Geology, 42, 1043–1046, 2014.

Muglia, J. and Schmittner, A.: Glacial Atlantic overturning increased by wind stress in climate models, Geophys. Res. Lett., 42, 9862–9869, 2015.

NGRIP Members: High-resolution record of Northern Hemisphere climate extending into the last interglacial period, Nature, 431, 147–151, 2004.

Olson, B., Hashmi, I., Molloy, K., and Shehu, A.: Basin Hopping as a General and Versatile Optimization Framework for the Characterization of Biological Macromolecules, Advances in Artificial Intelligence, 2012, 674 832, 2012.

Petersen, S. V., Schrag, D. P., and Clark, P. U.: A new mechanism for Dansgaard-Oeschger cycles, Paleoceanography, 28, 24–30, 2013.

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.

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

Rhodes, R. H., Brook, E. J., McConnell, J. R., Blunier, T., Sime, L. C., Faïn, X., and Mulvaney, R.: Atmospheric methane variability: Centennial-scale signals in the Last Glacial Period, Global Biochem. Cycles, 31, 575–590, 2017.

Rolland, J., Bouchet, F., and Simonnet, E.: Computing Transition Rates for the 1-D Stochastic Ginzburg–Landau–Allen–Cahn Equation for Finite-Amplitude Noise with a Rare Event Algorithm, J. Stat. Phys., 162, 277–311, 2016.

Rousseau, D.-D., Boers, N., Sima, A., Svensson, A., Bigler, M., Lagroix, F., Taylor, S., and Antoine, P.: (MIS3 & 2) millennial oscillations in Greenland dust and Eurasian aeolian records - A paleosol perspective, Quat. Sc. Rev., 169, 99–113, 2017.

Rypdal, M.: Early-Warning Signals for the Onsets of Greenland Interstadials and the Younger Dryas–Preboreal Transition, J. Climate, 29, 4047–4056, 2016.

Sadatzki, H., Dokken, T. M., Berben, S. M. P., Muschitiello, F., Stein, R., Fahl, K., Menviel, L., Timmermann, A., and Jansen, E.: Sea ice variability in the southern Norwegian Sea during glacial Dansgaard-Oeschger climate cycles, Sci. Adv., 5, eaau6174, 2019.

Schulz, M.: The tempo of climate change during Dansgaard-Oeschger interstadials and its potential to affect the manifestation of the 1470-year climate cycle, Geophys. Res. Lett., 29, 1002, 2002.

Schulz, M., Berger, W. H., Sarntheim, M., and Grootes, P. M.: Amplitude variations of 1470-year climate oscillations during the last 100,000 years linked to fluctuations of continental ice mass, Geophys. Res. Lett., 26, 3385–3388, 1999.

Seierstad, I. K. et al.: Consistently dated records from the Greenland GRIP, GISP2 and NGRIP ice cores for the past 104 ka reveal regional millennial-scale δ^18O gradients with possible Heinrich event imprint, Quat. Sc. Rev., 106, 29–46, 2014.

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.

Ullman, D. J., LeGrande, A. N., Carlson, A. E., Anslow, F. S., and Licciardi, J. M.: Assessing the impact of Laurentide Ice Sheet topography on glacial climate, Clim. Past, 10, 487–507, 2014.

Vettoretti, G. and Peltier, W. R.: Thermohaline instability and the formation of glacial North Atlantic super polynyas at the onset of Dansgaard-Oeschger warming events, Geophys. Res. Lett., 43, 5336–5344, 2016.

Zhang, X., Lohmann, G., Knorr, G., and Purcell, C.: Abrupt glacial climate shifts controlled by ice sheet changes, Nature, 512, 290–294, 2014.

Zhang, X., Knorr, G., Lohmann, G., and Barker, S.: Abrupt North Atlantic circulation changes in response to gradual CO2 forcing in a glacial climate state, Nat. Geoscience, 10, 518–523, 2017.