

Clim. Past Discuss., author comment AC3
<https://doi.org/10.5194/cp-2022-1-AC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC3

Johannes Lohmann and Anders Svensson

Author comment on "Ice core evidence for major volcanic eruptions at the onset of Dansgaard–Oeschger warming events" by Johannes Lohmann and Anders Svensson, Clim. Past Discuss., <https://doi.org/10.5194/cp-2022-1-AC3>, 2022

Below are our responses to the issues raised by the referee (*italic*).

The validity of the statistical methods relies on two key points:

-

The bipolar volcanic catalogue by Svensson et al. (2020) is complete, and

-

The choice of Null hypothesis (1 eruption per 500 years) is correct

I am not convinced that the authors have demonstrated either.

We agree with the referee that in order for our conclusions to be reliable, the true rate of eruptions corresponding to the magnitude of the eruptions observed shortly before DO onsets needs to be known.

The "two key points" identified by the referee are actually two ways of stating the same thing, since an "incomplete" volcanic catalogue would simply translate into an underestimate of the eruption rate.

As discussed in more detail further below, in the paper we give a number of arguments (based on other data sets) supporting that our null hypothesis is suitable and the statistical analysis can be applied to the data set at hand. We indeed allow for the possibility that the volcanic catalogue is incomplete and that the true eruption rate is higher.

We give quantitative constraints on how far the data might in fact be biased in this way,

and show that the results are still significant when taking this into account. In addition, in the revised version of the paper (and in this author response) we give further arguments to support this, and hope that this can convince the referee of the validity of our method and conclusions.

(1) the authors rely on the bipolar events identified by Svensson et al. (2020) using annual layer counting. The goal of that study was not to make an unbiased and complete catalogue of all bipolar events, but rather to investigate the phasing of bipolar climate change at great precision. Svensson et al (2020) write: "The bipolar layer counting is not continuous but is focused on periods of abrupt climate variability or high volcanic activity."

This implies that the Svensson et al. (2020) catalogue is biased toward periods of abrupt climate change. This is critically important. LS22 demonstrate convincingly that there are more events in the vicinity of the abrupt events, however does this reflect the DO triggering mechanism proposed, or simply a bias in the volcanic catalogue towards periods of abrupt climate change that were investigated in more detail (as suggested by Svensson 2020)?

It seems critical to me that the bipolar layer counting has to be done continuously across the full interval of interest with no regard to the presence of abrupt transitions, and not just the periods of abrupt climate change.

Anders Svensson is an author on both studies, and I am interested to hear his perspective on this issue.

Hello, this is Anders Svensson speaking.

The main reason for stating that the layer counting of Svensson et al, CP, 2020 is not continuous is to say that there is no new time scale provided with that study. Had the layer counting been covering the 12-60 ka interval continuously it would at the same time have been a revision of the GICC05 time scale.

In the right hand side of Table S2 of Svensson et al., CP, 2020, the layer counting made for that study is shown in column 'This study.' As mentioned in the paper, the section 16.5-24.5 ka b2k is not covered by the study as ice cores are difficult to synchronize in this interval. Apart from this gap most of the 12-60 ka period is, however, counted continuously in Svensson 2020. There are four gaps in the layer counting at 14.5, 27.8, 32.0 and 37.1 ka, respectively (indicated as blank fields in Table S2), spanning a total of some 6500 years or about 16.5% of the investigated 12-60 ka period (excluding the 16.5-24.5 ka section.) Thus for more than 80% of the investigated period layer counting has been performed and the major bipolar volcanic eruptions occurring in those intervals have been identified.

In Svensson 2020 Figure 1 the identified bipolar eruptions are indicated. Except for the periods where no layer counting has been performed, the eruptions are fairly evenly distributed in time. (Note that the high density of bipolar events in the GI-9-10-11 interval is due to the five Laschamp-related ^{10}Be bipolar matches that are also included in the figure.)

Nevertheless, the statement from Svensson 2020 only says that the layer counting was

not done continuously. While the layer counting was indeed done in segments that comprised clusters of eruptions that could be matched as bipolar, it does not imply that the actual procedure of finding bipolar matches was not done on the whole time period as continuously as possible. Since the statement goes on to say that it focused on "periods of high volcanic activity", it should be clear that periods with large eruptions were considered, regardless of whether they happened close to DO events.

That being said, while an effort was made in Svensson 2020 to provide a bipolar volcanic data set that is as continuous as possible, it is indeed quite likely that not all of the large bipolar eruptions have been identified, especially in the intervals of no layer counting, but potentially also in some of the longer layer counted sections, where the synchronization may have failed.

This is the reason why we added a comprehensive analysis of the continuous data set of Lin et al since our first submission to *Clim Past*. This gave upper bound estimates on the number of potentially missing bipolar eruptions. Based on these estimates we allowed for up to two times more large bipolar volcanic eruptions in our robustness tests of LS22 as compared to those identified in Svensson 2020.

Here we want to stress that we have several lines of evidence to support that the dataset is in fact relatively complete, and certainly suitable for our analysis:

1. The magnitude of (most of) the eruptions, including the ones before DO events, is consistent with eruptions larger than Tambora, and thus with the observed 500 year return period. Unless the frequency of eruptions of that size was significantly larger in the glacial, there is thus no reason to suspect we are missing a large number of eruptions.
2. Even without reference to known eruptions with relatively well-known return periods like Tambora, we can constrain the number of eruptions that may be missing from our dataset, using a different, fully continuous and objective list of Greenland and Antarctic volcanic eruptions provided by Lin et al 2022. This yields our upper bound that the true occurrence rate may be up to 1.5-2 times higher. It is an upper bound, since it also includes a large fraction of regional eruptions with large deposition on only one of the poles. The results are still significant when using this rate.
3. The magnitude of the eruptions before DO events is a representative sample of the magnitudes of all of the bipolar eruptions. Thus, there is no evidence that relatively smaller eruptions (with corresponding higher recurrence rate) have been picked specifically in the vicinity of DO onsets in order to get the sought-after bipolar phasing of DO events.
4. This point is new and will be included in the revised manuscript (Sec. 3.1 and new panel in Fig. 2): In general, there is no clustering of eruptions around DO events, since there are very few cases where an eruption happens shortly after a DO onset. With "very few cases" we mean here "not more than would be expected by chance", i.e., in our data we find 0 eruptions within 20 years and 2 eruptions within 50 years after the onset. This is opposed to 5 eruptions within 20 years and 7 eruptions within 50 years before the onset.

From the point of trying to find a bipolar match to constrain the bipolar phasing of DO events (as a potential source of bias in our dataset), a match shortly after a DO onset would be just as useful as a match shortly before. Further, a match shortly after the interstadial onset should even be easier to establish due to the higher accumulation rate and lower impurity noise.

5. Apart from the occurrence of 5-7 eruptions closer to DO onsets than would happen by chance, there is no evidence of general clustering around the times of abrupt change,

since the waiting times in between events are consistent with a stationary Poisson process. This would not be the case otherwise, as it would create gaps that lead to a tail in the distribution that is not consistent with the exponential distribution.

6. Note that there are more than 70 eruptions not close to DO events. First, this should illustrate that it was not the sole purpose of Svensson 2020 to find eruptions close to DO events. Second, these eruptions were identified in patterns of typically ~ 4 eruptions, which are spaced by 500 years on average. Thus, even if Svensson 2020 would have only looked at the DO onsets, the resulting volcanic data set would already cover $20 \times 3 \times 500 = 30,000$ years (or more depending on how one interprets the edges of the eruption patterns) out of the 40,000 total years. Thus, by definition of the matching via patterns of eruptions and the given spacing of eruptions and DO events, this necessarily yields a volcanic record with fairly complete coverage even if only the vicinity of DO onsets were allowed as starting points (which was not the case).

Regardless of all this, we do want to reiterate that the dataset does not even need to be fully complete/continuous, as long as we can constrain the true rate of eruptions, and show the eruptions before DO events in question are of the correct magnitude for the given rate, both of which we do in the manuscript. We do admit that the manuscript in the present form did not name several of the potential biases and concerns explicitly, which will be added in the revised manuscript (Sec.'s 2.3 and 3.2, as well as Discussion) along with the more detailed reasoning above.

(2) the choice of the null hypothesis is critical here. The authors base their choice on the assumption that the Svensson 2020 bipolar catalogue is complete, which yields 1 large bipolar event per 500 years.

(2a) However, the quote from Svensson (2020) implies that the catalogue is not complete, but instead focused on periods of abrupt climate change or large volcanic activity. The recurrence time of such events should thus be shorter.

See previous comment.

(2b) It is clear that the Svensson catalogue has more events during the deglaciation (12-16 ka) than during the glacial (24-60 ka). Performing a simple Student t-test on the recurrence time distribution of these two intervals will show this I believe. This makes intuitive sense given the larger annual layer thickness in the former interval. What would happen if the deglacial recurrence time was used as the basis of the Null hypothesis? There are also many bipolar links identified in the Holocene (e.g., AICC2012). What is the recurrence time of those?

We already perform statistical tests on whether the deglacial eruption frequency is higher and report this in the manuscript. Taken individually, indeed two 1,500 year periods have a significantly higher occurrence rate than the whole population (at 90% confidence). However, in the larger context of the glacial, this is not statistically significant due to multiple testing.

Our analysis thus shows that in the bigger picture of the glacial eruptions this increase of eruptions does not appear as statistically significant, even though it may well be a genuine feature. Given that it is not significant there is no sound basis for us to include this in our null model.

This is even more so since the increased frequency may well be an artifact of the higher measurement resolution of the proxies during the deglaciation (as you say), thus allowing one to observe eruptions of smaller magnitude (and thus higher frequency), compared to the rest of the glacial. On top of this, using the observed frequency during the deglaciation would not be a good null hypothesis since a) it is a very small sample to reliably estimate a recurrence time, and b) there are enough potential physical explanations for why the frequency of eruptions during the deglaciation may not be representative of the whole glacial.

Further, since a) the deglaciation is only a short segment of the entire time period, and b) we already test our results against severe undercounting of events, this potential issue does not influence our results. We will still include a better discussion of it in Sec. 2.3 of the revised manuscript.

Regarding the recurrence time of eruptions in the Holocene, this is already mentioned in the paper on page 6. There we write that 80 eruptions with bipolar signature have been found in the last 2,500 years in Sigl et al 2015. Thus there could be in principle one bipolar eruption every 30 years, and we are fully transparent in the manuscript that this higher number of eruptions should exist in reality. But the point is that nowhere nearly as many eruptions could be identified in the glacial period due to decreased measurement resolution and increased noise in the proxy archives, allowing one to only detect and match eruptions of the largest magnitudes, which are at the same time the eruptions that can be expected to have significant climate impact.

(2c) The choice of large, 1-in-500 year bipolar events as the sole trigger for events is somewhat arbitrary. The model simulations suggest the system is most sensitive to NH forcing. So would it not make sense to use the largest NH eruptions instead?

We agree that the eruption latitude is important for the climate response, but we are not so confident as to the exact impact of eruptions at different latitudes on the dynamics of DO cycles.

The data set of Lin et al gives a binary classification (below 40degN or Southern Hemisphere, LLSH; versus Northern Hemisphere High Latitude, NHHL) of the eruption latitude based on the relative Greenland-Antarctic deposition. We do not believe, however, that this data is informative and reliable enough to incorporate it into our statistical analysis, or even use it as the main data of the analysis.

We will however discuss it in the revised manuscript (Sec. 3.7 and Discussion), in relation to our proposed mechanism of the volcanic trigger, giving the following details:

There are 5 (LLSH) vs 2 (NHHL) eruptions within 50 years before a DO onset, and 4 (LLSH) vs 1 (NHHL) eruptions within 20 years. This is out of 34 (LLSH) vs 48 (NHHL) total bipolar eruptions.

While the sample size of eruptions before DO onsets is very small, it still indicates that eruptions classified as LLSH are more likely to cause a DO warming. We understand this

could be taken as evidence that our hypothesis of a North Atlantic (NA) cooling to trigger the DO onsets is not so likely.

However, the real climatic impact and its latitudinal footprint is still very much unconstrained, since the classification of Lin et al 2022 only tells us whether there has been relatively more deposition of sulfur in Greenland versus Antarctica.

The first obvious impediment is the large uncertainty in the deposition estimates of individual eruptions, due to inter-core variability and other factors. But even if this uncertainty was zero, there is still a massive uncertainty in the climatic impact, because we do not know where most of the sulfur aerosols actually were transported (troposphere or stratosphere) and how long they were present in the atmosphere. So there is the possibility that an eruption with a large sulfur deposition in Greenland compared to Antarctica still produced less NA cooling than a tropical eruption classified as LLSH.

Further, the latitudinal estimate "LLSH" includes NH eruptions up to 40N. So based on this classification we do not have a clear separation of eruptions with and without a NA climate impact.

Note that even tropical eruptions tend to have a hemispherically asymmetric cooling towards the NH, as discussed in the manuscript.

All in all, the latitude classification by Lin et al 2022 does not warrant us to focus the analysis on either latitudinal subset. Even more so because there is yet no other evidence for our NA cooling hypothesis that would warrant us to have the hypothesis inform the data analysis.

We will, however, include a detailed discussion on this in the manuscript (Sec. 3.7 and Discussion), so the reader can decide to what degree this is in conflict with our physical hypothesis.

(3) The discussion of literature is not very balanced, and heavily favors self-citation over important prior work by others.

We are happy to receive suggestions on important literature that was overlooked, and will add some references in the Introduction to other prior work regarding the influence of external forcing on DO cycles, as detailed further below. We still believe that all existing self-citations are appropriate as they are directly related to our study and cannot be replaced by others, as explained in the following.

Works with primary contributions by the authors:

Svensson et al 2020 and Lin et al 2022 are crucial because these are the two datasets that this study is based on.

Lohmann and Ditlevsen 2019: We use the method present in this paper to estimate the stadial onset times.

Lohmann and Ditlevsen 2021: This is crucial because it is the first that uses the Veros model (besides a benchmarking study that is also cited), and it contains methodology and further information on the tipping point of the AMOC in this model.

Lohmann 2019: This is to our knowledge the only paper that shows actual predictability of DO onsets. There exist other papers on e.g. early-warning signals of DO events, but these

have no predictive power, and the observed signals may also arise due to other reasons. Since the fact that DO events are predictable to some degree at first glance contradicts the idea of a volcanic trigger, the paper needs to be discussed.

Lohmann and Ditlevsen 2018: This paper is used to highlight that there is evidence for an external modulation of DO cycles. Here we agree that other studies may be cited, which evolve around the same issue. We will add a few important studies on data analysis and data-driven modeling (Buizert/Schmittner *Paleoceanography* 2015; Kawamura et al. *Sci. Adv.* 2017; Mitsui/Crucifix *Clim Dyn* 2017), as well as Earth system modeling (Brown/Galbraith *Clim Past* 2016; Zhang et al *Nature Geosc.* 2021; Kuniyoshi et al *GRL*2022; Vettoretti et al *Nature Geosc* 2022).

Other self-citations with A. Svensson are mainly standard references to the underlying datasets:

Svensson et al 2006, 2008, Seierstad et al 2014 (standard references for GICC05); Rasmussen et al 2014 (definition of Greenland stadials and interstadials, including estimates for their duration); Rasmussen et al 2013 (NEEM matching to GICC05); NGRIP members 2004 (NGRIP isotope data; Gkinis et al 2021 (NEEM high-resolution isotope data).

Further, there are these studies with A. Svensson as co-author:

Abbott et al 2021: Most recent study on the Younger Dryas volcanic hypothesis.

Capron et al 2021: One of the recent studies that investigate the timing of the onsets at decadal precision. We also cite all others that we are aware of.

Schüpbach et al 2018: Important reference regarding the impurity signals in ice cores during the glacial.

(4) Work by Sigl et al. (2015) suggests that the largest volcanic eruptions influence climate by up to 10 years. Why do you use a 50 year threshold instead? Is there a basis for this number?

We do not use a 50 year threshold, but investigate all lags from 10 to 70 years. The specific numbers of eruptions using a 20-year and 50-year tolerance are simply used in order to present concrete numbers to the readers, allowing them to judge the results depending on whether they a priori believe in only a short-term (20-year) or more longer-term (50-year) impact of eruptions.

Sigl et al 2015 look at the response of tree ring data, which mostly capture atmospheric temperature in regions with temperature-limited tree growth. A mechanism that triggers abrupt climate change from volcanism in the last glacial period is likely to involve other feedbacks than the immediate tropospheric cooling, which are not necessarily captured in the tree ring signal. There are numerous modeling studies (e.g. Pausata et al., *PNAS* 2015; Stenchikov et al., *JGR* 2009; Swingedouw et al., *Nature Comm* 2015) that show that the impact of a large eruption, especially on the ocean circulation, can last for longer than 10 years.

P1 L16: remove "past".

Ok.

P2 L5-10: this seems contradictory. The models now show unforced oscillations, so wouldn't this obviate the need for triggers / drivers?

Indeed, at first glance this may seem contradictory. But we discuss why the notions of unforced, deterministic DO cycles and a short-term volcanic trigger are not mutually exclusive at various places in the manuscript (P2 L17-22, P15 L16-22, P16 L1-3, P17 L6-9, P17 L11-17, P19L35ff).

Based on previous evidence regarding studies on the predictability of DO events (Lohmann GRL 2019) and self-sustained oscillations obtained in models (Brown/Galbraith Clim Past 2016; Vettoretti/Peltier GRL 2016, and others), we interpret our finding (the statistical relationship of eruptions and DO events) such that a trigger is not necessary for DO events to occur, but may still operate to lead to a transition that occurs earlier than otherwise (or potentially also a delayed transition in case of the interstadial-stadial transition).

However, at this stage this is only a hypothesis, and the point of our statistical analysis is to precisely assess whether there is evidence for a trigger of the events, regardless of whether a priori other evidence points towards the fact that DO events are an unforced, deterministic phenomenon.

Note that it is not certain whether the current models are in the correct dynamical regime. At different boundary conditions, the dynamics may be excitable for instance, where a driver would be needed while the same physical mechanisms are at play. This is why in this text passage we also state that there is no consensus whether there are any concrete drivers of DO events, or whether there is a need for any drivers. Studies that investigate the effect of volcanic eruptions on self-sustained DO-like cycles in comprehensive models are underway.

P2 L12: Though this depends really on the subjective choice of what one decides to call a stadial or interstadial. The events that follow the traditional numbering are all over 1000 years.

Indeed, but we would argue that the traditional numbering is based on fewer and lower-resolution records. We follow the widely applied stratigraphic framework of Rasmussen et al, QSR, 2014.

The shorter interstadials therein are now confirmed also in speleothem records (see e.g. NALPS record), so it is not subjective that these short events exist. If they should be called GI/GS or something else would depend on whether they are actually created by a

different mechanism, which is not known at this point. In fact, our study contributes to make progress regarding this aspect.

P2 L15-L22: Please refrain from self-citation in favor of a balanced review of the literature. For example, a long literature exists on the dependence of the DO timing characteristics on background climate conditions that predates the work by the author himself.

Ok, see reply above.

P3 L10: Is there a published bipolar ice core chronology? Or just the volcanic ties?

It is just the bipolar volcanic match points in the cores that are available. These are used to transfer the Greenland GICC05 chronology to the Antarctic ice cores. Will rephrase the sentence.

P4 L3: Are the

Not sure what is meant here.

P6 L10-15: "Thus, most eruptions in the dataset of Svensson et al. (2020) fall into the category of 1-in-500 year events in terms of their magnitude". This is unclear to me. Do the 82 bipolar events you identify overlap with the 69 events by Lin 2021?

Further down it seems this is addressed in Section 3.2. Please consolidate this information into one place to avoid confusion. Although section 3.2 refers back to 2.3, so it is not clear how the analysis was done.

Yes the 69 events by Lin 2022 are a subset of the N=82 bipolar set in Svensson 2020. We will add a Section 2.2. where we more clearly describe the two volcanic data sets and how they relate.

P6 L16: I strongly disagree with this conclusion. The Svensson (2020) paper itself clearly states that it is incomplete: "The bipolar layer counting is not continuous but is focused on periods of abrupt climate variability or high volcanic activity."

See response above.

P6L 22: what do these p values signify? What threshold do you use?

I am not sure what is meant here with thresholds, but for instance a p-value below 0.05 would signify that the empirical distribution of the waiting times is inconsistent with an exponential distribution at 95% confidence. The statistical tests used are standard. Since we clearly find $p > 0.1$ (or whatever confidence level one prefers), there is no evidence that the data is inconsistent with a Poisson process.

P6 L32: It seems clear that the deglaciation (12-17ka) and glacial (24-60 ka) have different volcanic statistics. It is true one expects 2 false positives at 90% confidence, but that ignores the fact that BOTH of these are consecutive, and BOTH are in the youngest segment with thicker annual layers and thus higher detection probability for volcanic layers.

A t-test of waiting time distributions on either side of the data gap would probably suggest these are different.

We agree that it is likely that this is a systematic effect and also state this on page 6. But from the point of statistics, we do not have enough evidence that this is really the case. It is thus not reasonable to construct our statistical framework in a way that would somehow take this into account.

For more, see our more detailed response to your comment further above.

The fact that the intervals in question are (almost!) consecutive will not really change the statistical (non-)significance, since this indicates that there is a (plausible) correlation in the occurrence frequency of 2 kyr segments, which would reduce the number of effective data points in the multiple testing context.

Figure 3: Do you use the estimates from Lin 2021 for the bipolar volcanoes? In Fig 3c, what is the aerosol loading based on?

Yes, Lin 2021 is the only available source for the deposition estimates. This is stated in the figure caption. See Lin 2021 for more details on the calculation of the aerosol loading.

P11 L8: I don't understand the logic. Antarctica also has several large local volcanoes in close proximity (on the continent). Also, these NH eruptions probably had more impact on the AMOC that you care about.

The logic is the following:

Since the Greenland data set contains more eruptions than that of Antarctica, and a bipolar match requires an eruption in both Greenland and Antarctica, there are by definition more "local/regional/unipolar" eruptions in the Greenland data set. One could spin this the other way around and say that simply more eruptions are missing in the Antarctic records.

But neither interpretation would influence our analysis since then also the dataset in Svensson 2020 (based on the same ice cores) would be missing these eruptions, as they require deposition in both Greenland and Antarctica, and this would be the case regardless of whether they occurred close to DO onsets or not. Thus, as already stated in the manuscript, it is logical to use the smaller Antarctic dataset to make our inference on the potential number of missing bipolar candidates.

We agree that there are also local eruptions in Antarctica. There is no reliable data yet as to whether Greenland or Antarctica features more "local" eruptions, whatever the definition thereof. We will rephrase to not give the impression that there are only local eruptions close to Greenland.

Regarding your last point, we don't believe we have reliable data yet to choose a subset of eruptions for our analysis that we think has a priori more impact on the AMOC. See also our responses to your previous points regarding the latitudinal issue.

P11 L10: Out of these 84 largest eruptions, how many correspond to bipolar events?

None. The point of this analysis is to find all eruptions with a large sulfur deposition that have not been identified as bipolar by Svensson 2020. This will include some that actually had a bipolar imprint, but also plenty of eruptions that only have a large sulfur deposition because the eruption site was close to the ice coring site. That is why we say this is an upper bound estimate.

Since this has not been brought out clearly enough in the manuscript, we will add a Section 2.2., where the relationship of the Lin et al 2022 and Svensson 2020, as well as the unipolar subset of Lin et al 2022, is addressed explicitly.

P13 L20: Personally I think at 50 years it would already be unlikely. Sigl et al. 2015 shows a ~ 10 yr impact on climate of the largest eruptions.

As pointed out above, Sigl et al 2015 only show the impact on atmospheric temperature in tree ring records. The following response of the ocean circulation could certainly take decades before an actual regime shift of the circulation starts, and it is not so clear how this would be manifested in tree ring data. Leads and lags in between the eruptions and the DO onsets below 15 years are unfortunately hard to assess as this is within the uncertainty of the onset detection.

Section 3.5: Figure 5b is hard to interpret. Fig. 5a is much more intuitive. Can you also plot the result when doubling the events there?

We agree that panel a is easier to understand, but it also contains less information, since from panel a one cannot infer the significance of the results in a quantitative way. Because of the discrete nature of the distributions, the confidence bands are only defined to the nearest integer (being conservative, here we choose the next larger integer).

We will improve our explanation of panel b so that it is immediately clear what is shown. It is simply the p-value of our null hypothesis as a function of the free parameter.

Thus, we will include the expectation value of the doubled occurrence rate in panel a, but not the confidence bands. The latter would be misleading as they do not allow one to judge whether the data is really significant at 90% or 95% confidence in case the data is close to these bounds, as is the case here for the doubled occurrence rate (see Fig. 5b).

Section 3.7: It is not surprising that your model is more sensitive to NH cooling. Does this not imply that you should be evaluating the proximity of DO event onsets to large NH volcanoes?

It may not be surprising in relation to our physical hypothesis, but it still not clear a priori that this is indeed the case in a model. As stated previously above, we believe our statistical analysis is more powerful when not constrained by a preconceived hypothesis on a particular subset of eruptions. Otherwise we might be accused of pursuing a preformed hypothesis.

As also stated above, this is especially true since we do not actually have reliable information on the latitude of the eruptions, and even less on their true climatic impact on the global or NH climate.

P19 L5: My main concern is the incompleteness, and the fact that the focus of the identification of bipolar events focused on periods of abrupt climate change.

This is exactly what we mean here. In the revised manuscript we will address the issue of the potential systematic bias around DO events explicitly, including providing more evidence to the contrary, as discussed above.

P20 L11: These kind of statements are somewhat tentative, as this depends a lot on the weather conditions that distributed the volcanic deposits to both polar regions, and depositional processes and redistribution on the ice sheet surface.

Unfortunately it is not clear which statement this refers to, and what the referee want us to change.

P20 L12: "arguably more accurate"? I would say certainly more accurate!

Yes in a way they are clearly more accurate, but in another way they are not: Observations of 2,500 years do not give a precise estimate on the magnitude of a 1-in-500 year eruption.

P20 L19: I believe this is in violation of the data policy; all data should be made available. I will let the editor weigh in.

We will leave this up to the editor.