Response to the comments from Eric Wolff.

'This paper represents a huge amount of work to identify the deposition of volcanic sulfate in the polar regions and to try and draw conclusions about the occurrence of volcanism over the last 60 kyr. I really applaud the effort, and the attempt to draw large-scale conclusions from it. I have a lot of relatively minor comments on the paper (which put it somewhere between minor and major revision), and I think the authors could have made some better choices in the way they treated the data. However, I accept that they made mainly reasonable choices and so I do not propose to insist on any significant reanalysis of the data – just in places an extra sentence is needed to discuss the choices made. I found some of the messages that end up in the abstract too strong given the nature of the data and I will comment on those in the text, expecting the authors also to address them in the abstract.'

We thank the reviewer for taking time to provide constructive comments to improve this work. We will revise the manuscript accordingly.

'A final point is that the paper is extremely hard to follow – while I appreciate the need for a lot of supplementary material, the fact that the figures in the main text and the supplement seem to be called in almost random order is very unhelpful, and I suggest the authors renumber the figures to provide a more logical flow.'

We have updated the order of figures and tables in the text to follow the flow of the text.

'Before I give a detailed set of comments, I should add that (as I think several of the authors are aware) I have been involved in a similar exercise, using only the EDC ice core, but for a period of 200 kyr. This has been presented a few times and has been (in a paper involving also terrestrial and marine data) under review for a considerable time. My comments should therefore be taken with the knowledge that I have used slightly different methods and criteria but am not seeking to impose the same methodology on the authors here.'

Thanks for sharing this. We look forward to see the exciting 200 kyr profile being published.

Detailed comments:

'The paper will need a thorough copy edit as the English at times is awkward and occasionally hard to follow.'

We went through the text and hope we have succeeded in making it more readable.

'Line 165. I think the authors have misunderstood the MSA correction. This had to be done for WAIS Divide because ICPMS measures elemental S, and therefore the values obtained are (sulfate + MSA). This is not the case for any of the FIC/IC methods, where sulfate itself is measured, so no correction would be required and no correction should have been made for any other sites. Please clarify this in your text.'

Thank you for this clarification. We completely agree and we have removed the discussion of MSA, except mentioning that we cannot do the correction for WDC in the glacial period.

'Line 188. I found myself a little confused about the volcanic detection threshold. I understand the idea behind the use of RRM and RMAD. But then in the end you just use a threshold of 10 or 20 kg $km^{-2}$ so I am not sure what purpose the statistical threshold serves.'

Indeed, we do the volcanic detection in two steps. First step being the 'classical' RRM and RMAD approach that works well in the Holocene and a second step being a reasonable and bold cut-off. Ideally, the second step should not be necessary, but we did it in order to obtain a more homogenous volcanic list throughout the investigated period. With all of the variation in both climate and data resolution, we found it safest to take this approach. We did experiments with different cut-off values and the applied values are chosen conservatively to avoid including non-volcanic sulfate spikes that may not be discarded in the 'classical' approach in particular during stadial periods.

'Equation 1: please be careful to define the units when describing this equation. I assume concentration is in ppb or ng/g (as in the figures), and D is in m ice equivalent (this is important as if you had included the top 100 m where density is less than that of ice, then the equation is not correct if in real depth), and 0.917 is in g/cm$^3$. By fluke, after cancelling factors of 10 to change the units this does indeed work out as kg km$^{-2}$, but the reader needs that to be stated.'

The units for the equation have been added in the text.

'Line 209. Please also be careful to define what T is, ie layer thickness/original layer thickness. It's confusing otherwise because you talk next about a 60% reduction of the layer thickness which is equivalent to a T of 0.4. Perhaps that is the best way to deal with it, "60% reduction of the layer thickness (T=0.4)".'

This has been clarified in the text. It now says: 'T is the layer thinning (the ratio between the layer thickness in the ice core and the original layer thickness). The layer thinning has been calculated by site-specific ice-flow models (thinning function) that we applied here (Table S1 and Fig. S2 (c)). During the last glacial period, the layer thinning can be significant. For the Greenland cores, the thinning rate ranges from a 60% reduction of the original layer thickness (T=0.4) at 12 ka b2k to as much as 90% (T=0.1) at 60 ka b2k (Table S1). For Antarctica, the thinning rate is most significant for the high-accumulation WDC core, where it approaches 95% (T=0.05) at 60 ka b2k (Fudge et al., 2016; Buizert et al., 2015), whereas the EDC core exhibits a modest thinning of 30% (T=0.7) at 60 ka b2k (Fig. S2 (c)).'

'Supp table S1: Please explain what the column "Name" is: I assume it's the age model used. But I'm surprised because the standard for a 60ka period is GICC05. I appreciate you may have to use a model to derive a smooth acc rate, but then you need to explain the relationship between ss09seabm1 and GICC05 as few of your readers will have heard of the former.'

The column 'Name' in the supplement Table S1 is the name of the ice-flow model of each core. An explaination for 'ss09seabm1' is added to the caption of Table S1 as follows: 'ss09seabm1 is the model timescale which include the layer thinning rate.'

'Fig S1 and caption. Please clarify that in part c the y-axis "thinning function" is equivalent to "T" in the text as I don't think you ever explain that in the text. Also in the caption, what does "thinning file" mean? The second sentence of the caption doesn't make sense, please re-word.'

The 'thinning function' has been clarified as follows: 'T is the layer thinning (the ratio between the original layer thickness and the layer thickness in the ice core). The layer thinning has been calculated by site-specific ice-flow models (thinning function) that we applied here (Table S1 and Fig. S2).'.

In the caption, 'thinning file' is now changed to 'thinning profile'. The second sentence of the caption is re-worded to 'A simple linear fit is applied to the WDC thinning profile below 2800m depth, labelled as WDC (combined), as the gas derived thinning profile has potentially unrealistic wiggles (Buizert et al., 2015).'.

'Line 220. This is wrong. The volcanic sulfate flux calculation does not assume anything about wet or dry deposition: it is simply the product of concentration and snow acc rate. It's true that the flux is affected by whether there is significant wet deposition in addition to the dry deposition and this is a point worth making, but the flux as calculated is definitely correct.'

This has been reworded. We edited the passage as follows: 'For high snow accumulation sites, such as those in Greenland and coastal Antarctica, it is quite likely that wet deposition is dominating the sulfate deposition at present day (Kreutz et al., 2000; Schupbach et al., 2018). During the last glacial period, dry deposition may have played a more important role, giving rise to a potential bias of our glacial sulfate deposition estimates. We do not attempt to make any corrections for this effect.'

'Table S4, column O, kg not Kg please.'

Corrected.

'Table S5. Why is N/A written for WD2014 ages below 30 ka. WD2014 extends from the surface to 68ka even if it changes from layer counted to methane-tied at 31 ka.'

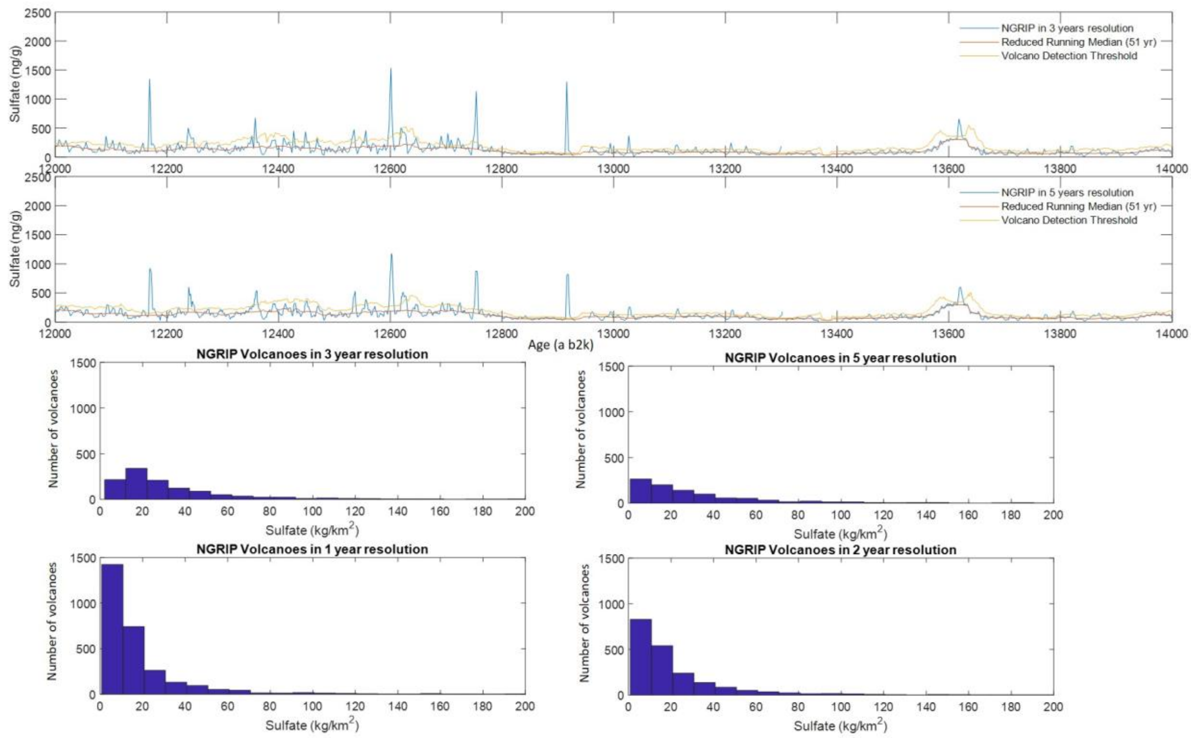
Thanks for the suggestion. The WD2014 ages have been included throughout.

‘Tables S4 and S5: Are the values given for individual sites before or after the rescaling of Antarctic concentrations. I think before – if not then it’s hard to understand the averages you give for the Antarctic. Please clarify this in the captions, ie that the individual values are as measured while the average is after rescaling.’

This has been clarified in the supplementary caption of Table S4-S5 as follows: ‘‘Deposition’ is the sulfate deposition for individual events after correcting the layer thinning but without applying a scale factor to obtain the average Antarctic deposition. When an event is detected in three cores, the ‘Average deposition’ is the simple mean of the volcanic sulfate deposition at the three sites. When an eruption is detected in one or two cores, a scale factor is applied to obtain the averaged Antarctic deposition (see ‘Methods 2.5’).’.

‘Line 241. What matters is of course not the value of the sulfate background but its variability. I think this is what you mean but from this wording it isn’t quite clear. I think you need to explain your 10 and 20 kg m-2 criteria better – how is it derived? In my own work we estimated the “negative” peaks to understand how big a deviation from the median could be generated in a given section from noise alone, and then we used that to establish a threshold (which was 20 kg $km^{-2}$ for the 200 ka record). So I think your thresholds are probably fine, but could be explained more clearly.’

We have two main concerns: One is the inconsistent resolution of sulfate records over the investigated period due to the high-level layer thinning in the deep part of ice core. The other one is the cold Greenland periods (stadials), where the sulfate background variability is high and numerous short-lived sulfate spikes (<3 yr duration) get above the detection limit in the ‘classical’ volcanic detection scheme, in particular for the high-resolution NGRIP CFA profile. This is demonstrated in the figure below where the NGRIP CFA sulfate profile is exposed to different degrees of smoothing of 1, 2, 3 and 5 years, respectively. With the 1 and 2 year smoothing a large number of small peaks are detected, some of which may be of volcanic origin, but most likely a large fraction is related to gypsm or other sources of non-volcanic origin (as discussed in section 4.1). For Greenland, the detection of sulfate depositions larger than 20 kg $km^{-2}$ is rather insensitive to the degree of smoothing and we therefore chose this value for the cutoff. For Antarctica, the background variability is much lower and we find that 10 kg $km^{-2}$ is a ‘safe’ cutoff.




Caption of the above figure: The two plots in the upper panel show examples of background signal determination and volcanic peak detection for the NGRIP sulfate record as smoothed to 3-year and 5-year resolution, respectively. The four plots in the lower panel are the number volcanoes detected in 1-year, 2-year, 3-year, and 5-year resolution for the entire NGRIP sulfate record.

'Line 255. The Antarctic plots shown in Fig S2 are really not impressive with very small r^2. They don't seem at all a good basis for the rescaling you do. While I accept there may be higher fluxes for a given volcano at WD because it receives more wet deposition (this is the only reason the accumulation rate is relevant), the values you derive here have a huge uncertainty which you have not apparently propagated into your 26% error estimate (line 265). We know there is a huge uncertainty anyway because of local variability in deposition (as shown eg by Gautier et al (2016). I suggest you do an alternative error estimate where you use the rescaled data from the 123 Antarctic eruptions where you have 3 sites and find the average 1-sigma and/or 2-sigma between the 3 sites. This would give you an alternative way to estimate the uncertainty in the values you end up with. 26% is certainly way too small an error in the cases where you have only 1 site.'

Indeed, we agree that the linear regression among three Greenland and three Antarctic ice cores is not a perfect approach for obtaining the rescaling. We used the error of propagation method to estimate the uncertainty of the sulfate deposition for individual volcanic events, considering the sulfate diffusion, the layer thinning, the uncertainty of the analytical method, and the data resolution. For EDC and EDML, we now include an additional uncertainty due to the demonstrated high local variability in sulfate deposition. The uncertainty of the local variability is 29% according to Gautier et al. (2016). Thus, the estimated error of the sulfate deposition for EDC and EDML is now up to 40%. We applied the method 'the average 1-sigma and/or 2-sigma between the 3 sites' to estimate the average uncertainty for the rescaled volcanic sulfate deposition from 124 events, where we have sulfate deposition at three sites. When we have less than three sulfate deposition values, we use the maximum uncertainty of the sulfate deposition of the individual cores.

We revised the wording of the caption for Table S4 and the Lines 268-273 'For EDC and EDML, the low accumulation leads to a large local variability in sulfate deposition of up to 29% (Gautier et al., 2016), so for those cores we arrive at an combined error estimate of 40% for individual deposition events. When there are sulfate signals from three ice cores in one hemisphere, we used the standard deviation of the rescaled volcanic sulfate depositions for all ice cores to estimate the uncertainty for the average area volcanic sulfate deposition. When there is a common signal in fewer than three cores, we use the maximum uncertainty of the volcanic sulfate deposition from the individual cores.'.

'Line 268. I am familiar with the paper by Marshall (not Martshall) but I don't know what you mean by "The volcanic sulfate deposition in Greenland and Antarctica shows a distribution pattern (Table S5 and Table S6), similar to that derived from the aerosol-climate modeling of volcanoes over past 2500 years". I can believe that you are assuming they show such a distribution and that this is how you decide on the latitude, but you have no evidence Section 2.6 and Fig S3. I apologise for my lack of technical knowledge but this section is not comprehensible to someone coming at it new. I think you need to explain what you are trying to achieve here (which I think is to use the Greenland/Antarctic ratio of known eruptions to classify the latitude of unknown bipolar eruptions). Then please try to explain SVM better. To anyone not in the know Fig S3 b and c cannot be understood.'

This paragraph has been revised as follows: 'The volcanic sulfate deposition in Greenland and Antarctica shows a distribution pattern related to the latitudinal band of the eruption site (Fig. 1) (Marshall et al., 2019). To estimate the latitudinal band of bipolar volcanic eruptions of unknown origin, we applied the Support Vector Machine (SVM) classification model of Hastie et al. (2009) and Vapnik (1998). The classification model is based on a kernel function generation and logistic regression. The model was trained using 21 eruptions for which the eruption site is known from tephra deposits in the ice (Table S6). The input values for each eruption to the model are the average Greenland sulfate deposition, the average Antarctic sulfate deposition and the latitudinal band (above 40°N, 40°N-40°S, or below 40°S) of the eruption site. The cross-validation used for tuning the algorithm is 10-fold partition for each evolution. For an optimal classification, a maximum-margin hyperplane was used to separate two classes. The kernel scale and box constraints were chosen for the model and a Bayesian optimization was used to optimize the above two parameters to yield the best classification model (Fig. S4) (more detailed descriptions are in the Hastie et al. (2009), page 17). The bipolar eruptions of unknown origin were predicted into two latitudinal bands – above 40°N (NHHL) and below 40°N (LL or SH) (Table S5) based on the trained model. Due to the low number of known volcanoes erupted in the high latitudes of the Southern Hemisphere, the method does not allow unambiguous identification of eruptions potentially located in this region.'.

The support vector machine learning model provides a statistical approach to predict the erupted latitudinal pattern for the past unknown volcanic eruptions.

'Line 282, I think you mean Fig S10, not S9.'

It has been corrected to Fig. S10.

General comment: this reminded me that the calling of figures, especially supplementary ones, in apparently random order is really confusing. Please sort this out.

This has been sorted.

'Line 360. I don't really understand your statement that "the layer thinning becomes stronger which makes it increasingly difficult to detect smaller eruptions signals". If that is really so then your threshold of 10 or 20 kg $km^{-2}$ is too low. The whole point of it should be to ensure that detection is the same throughout.'

Indeed, the threshold or cut-off at 10 or 20 kg $km^{-2}$ does the main part in homogenizing of the dataset over the investigated period. However, due to the very high resolution of the NGRIP CFA sulfate (1 mm resolution) and the WDC CFA sulfur (1 cm resolution), in order to discuss the frequency of events over time we also do a homogenization of those datasets by smoothing them to their 'effective' resolution that we deduce from their power spectra (Figure S6). We do not make this smoothing at an earlier stage because that would make us unable to detect smaller volcanic events for example during the deglacial period. It is only when we want to study the frequency of eruptions across the entire 51 ka period that we smoothen the high-resolution records to their lowest resolution during the cold periods early in the investigated period.

''Section 3.4. The previous comment raises a more general comment on this section. It's really obvious that you shouldn't use cores where the resolution is worse than the expected peak width. However, it's difficult to know what that means: in the case of EDC, if thinning were the only thing happening then it would be hopeless to detect eruptions reliably in the early parts of the record with resolution 5 years. However, it turns out that the diffusion at Dome C keeps the peak width quite constant with age, and therefore allows resolution of peaks where the age resolution is 5 years (because the peaks at this depth have diffused to 10 or 20 years wide). However, I'd be very surprised if you can expect to resolve most volcanoes, even large ones, in GISP2 with the stated resolution in Table S1. For NEEM, given the variable depth resolution, I suspect the age resolution is normally much better than 10 years (I suggest checking this again), but if it were 10 years again I think detection would be hopeless. I think you need a somewhat deeper discussion here, and also to consider whether it is worthwhile to contaminate a great dataset by including data from sites where detection must be severely degraded.''

We agree that this is a delicate balance. On the one hand side, we want to include as many datasets as possible to improve statistics, on the other hand, we are risking to 'contaminate' the dataset with records of too low resolution, in particular in the deeper/older part where sample resolution is decreasing for most records. Again, we have made a compromise, where we do include the lower resolution datasets all the way, but only for large eruptions, where the volcanic spike is clearly above threshold. In order to compensate for merging of several smaller spikes into a low-resolution data point, we manually check the high-resolution ECM and DEP records across the large eruptions, to see if there are several peaks involved. If there are several events occurring in one low-resolution data point we either discard that data point or we distribute the sulfate deposition into fractions based on the relative magnitude of the spikes in the high-resolution records. This procedure is now documented in the last columns of Table S3, where it is indicated if a large spike has been split and what the fractions are. We are aware that this is pushing the data a bit to the limit, but because of the large variability in sulfate deposition, we want to take advantage of all the available records in order to constrain the sulfate deposition as much as possible.

Fig S5 and line 375. Please be careful here. I believe the Rasmussen analysis makes some sense for the kind of CFA set up used by Bigler where there is a large reaction cell that is being continually mixed leading to quite a large mixing volume. In such a setup there can be a nominal data resolution of mm which is not real. The FIC setups are a bit different as they inject a slug of liquid every 2 cm (or whatever each system uses) so the resolution could never be better than that in any sense. I'm not sure your analysis really makes that distinction.

Indeed, we agree, and that is why we do the smoothing of the NGRIP CFA sulfate record to its ‘effective’ resolution as determined from the power spectra (as described two comments above).

Line 389. You call Fig 4 before Figs 2 and 3.

This has been changed.

Line 419. ‘I found this very confusing. I am looking at Fig 2 where you show the Holocene, deglacial, stadial and interstadial. You then state that the number of eruptions is higher in cold periods when the figure shows the most in the deglacial and really no difference between the glacial and Holocene. Eventually I realised that you are only talking about stadials versus interstadials within the glacial. So firstly please make this clear throughout this section. But really is that statement true at a significant level: perhaps so in the 1year data in Fig S6, but the reader isn’t looking at that, they are seeing Fig 2. You need to be much clearer that you are using Fig S6 to discuss whether the difference is an artefact and that the conclusion, as illustrated in Fig 2 is that it probably is. I realise you do say this in the end, but by pointing at Fig 2 initially you leave the reader who doesn’t take the trouble to look at the supplement lost.’

Yes indeed, this point was not clear from the context. Section 4.1 is meant to be a discussion of stadial versus interstadial periods. We now state this clearly in the beginning of the section: ‘This observed difference leaves us two options: either there is a true millennium-scale volcanic eruption variability related to northern hemispheric climate variability, or the dependency is an artefact caused by the very different properties of the Greenland sulfate record in stadial and interstadial periods.’ and make a reference to Fig S7.

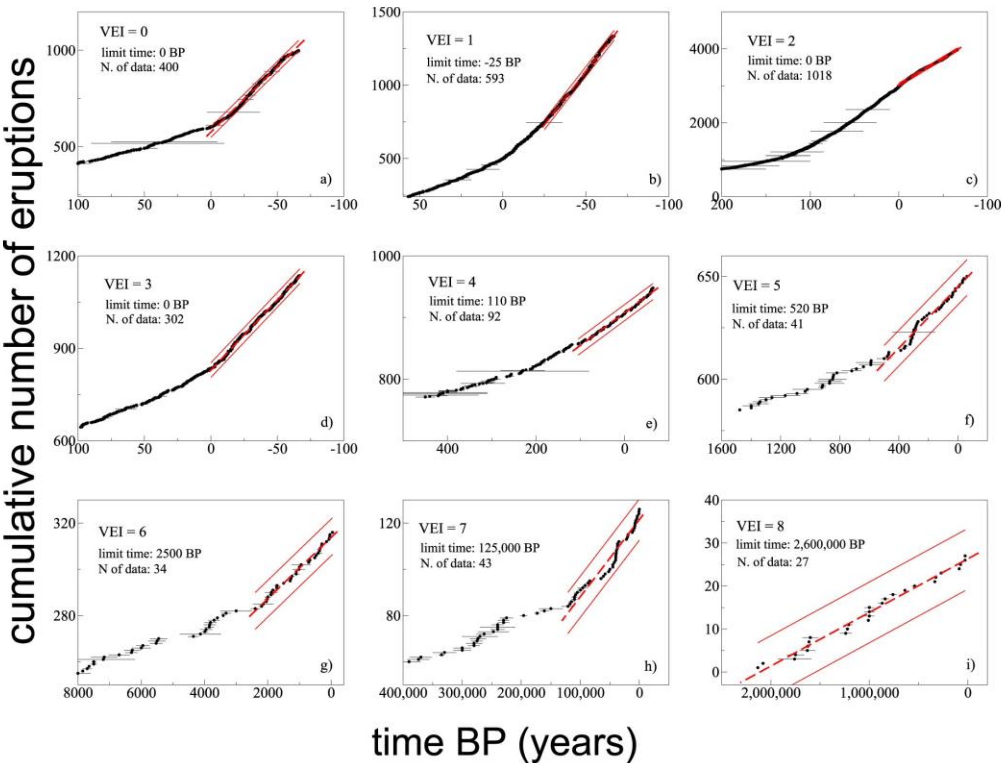
‘Line 488 (2-0) Do you mean 2-0 ka? When I look at Fig 2, I don’t really see a 34% increase. At least for the small (blue) category, there is no change. Perhaps describe this more precisely as to what you mean.’

Yes, we mean ‘2-0 ka’, and yes, the increase is for the large eruptions only. This is now stated more clearly in the text as follows: ‘In our 3-core Greenland sulfate deposition compilation, we also found an increased volcanic eruption frequency over the deglacial period, which is 34% higher than the late Holocene period (2-0 ka), especially for the fraction of large sulfate deposition events (Fig. 3 and Fig. 4).’.

‘Line 535-544. I don’t find this comparison of the ice core eruption rates against Lameve data helpful. Firstly you should also look at the analysis by Rougier et al (Rougier, J., R. S. J. Sparks, K. V. Cashman, and S. K. Brown (2018), The global magnitude–frequency relationship for large explosive volcanic eruptions, Earth planet. Sci. Lett., 482, 621-629, doi:https://doi.org/10.1016/j.epsl.2017.11.015.). It is very clear in that paper that the community is well aware that Lameve under-reports so I don’t really see the value of the comparison you are making.’

Our main message is to provide an ice-core estimate of the frequency of large global volcanic eruptions during the last glacial period (Table 1). It is the first time such an estimate is possible on a long ice-core based record beyond the last 2500 yrs and we think it is quite relevant. We mention the geological record because so far that record has been the only source of information of global volcanism on those time scales. The geological record applies the VEI index that we cannot deduce from the ice core record. Instead, we compare the magnitude of the glacial eruptions to those of well-studied historical eruptions for which the VEI index is known, such as the Tambora and Salamas eruptions.

Rougier et al., 2018 analyzed volcanic frequency based on LaMEVE database and found that this volcanic record shows a deterioration of information with age, not the real variation. The GVP database (http://volcano.si.edu) has a similar problem. For the recent time, the small size volcanic eruptions are well recorded, and for the past time the large size eruptions are much easier to discover in the geological record. Furthermore, the large size volcanic eruptions show a linear pattern reflecting a stationary process, as seen for the VEI 8 index volcanoes over the past 2 Ma BP (Fig.1 in Papale et al. (2018)).




Caption: Cumulative plots of the number of eruptions with time for each individual VEI class. The plots represent a zoom in a (relatively) recent time region where approximately linear trends emerge. The horizontal axis refers to years Before Present (BP), conventionally set to zero at year AD 1950. Best-fit linear trends are represented by the thick red dashed lines, while the thin solid red lines bound the 95% confidence band for expected statistical variability of data23. Wider confidence limits reflect lower number of data. Each panel reports the time limit of the linear region, which generally increases with increasing VEI, and the number of eruptions in such a region, which generally decreases with increasing VEI. (Papale et al., 2018, Fig.1).

'I am unconvinced by section 4.4 Given the huge uncertainties, especially for eruptions where there is only 1 core represented in one of the ice sheets, the values are so uncertain that trying to call out individual positions in the medal table seems a bit pointless. At the least please add a column where you state how many cores are represented in each ice sheet. But I feel this section of the paper is given too much weight and should be shortened. Remember also that you certainly show no eruptions that are bipolar between 16.4 and 24.5 ka, but this is not because they were absent but only because the tie points are not good enough to know whether they are bipolar. Given this kind of issue I think the league table, while it can be included should be downplayed and put into context.'

We think there is a general interest in knowing which volcanic eruptions are the largest during the last glacial period, even if the list does not cover the complete period 0-60 ka. We now include columns in Table 2 showing in how many ice cores the individual volcanic events are identified. The majority of those large eruptions are identified in 4-6 ice cores and only a few of the lower ranking eruptions are only identified in one ice core in Greenland. Of course, the list may be improved with time as we obtain more and hopefully also more accurate estimates of the polar sulfate depositions, and of course the list will become more complete whenever we are able to include the entire Holocene and the LGM and possibly part of the earlier last glacial. But are those good arguments for not presenting a list of the largest eruptions over some 40 ka of the last glacial period investigated here for the first time in a global context? In our view, there is so much interest in this list that if we do not include it in the manuscript it will be produced by the readers, and we prefer to be the first ones to comment on it. Section 4.4 takes up less than 10 % of the manuscript text and we see it as one of the major outcomes of our work. The discussion does put our results into context in that it discusses the potential origins of several of the major eruptions that can be investigated in future studies concerned with tephra, modelling and other paleo-archives.

References:

Marshall, L., Johnson, J. S., Mann, G. W., Lee, L., Dhomse, S. S., Regayre, L., . . . Schmidt, A. (2019). Exploring How Eruption Source Parameters Affect Volcanic Radiative Forcing Using Statistical Emulation. *Journal of Geophysical Research-Atmospheres, 124*(2), 964-985. doi:10.1029/2018jd028675

Yan, J. P., Zhang, M. M., Jung, J. Y., Lin, Q., Zhao, S. H., Xu, S. Q., & Chen, L. Q. (2020). Influence on the conversion of DMS to MSA and SO42- in the Southern Ocean, Antarctica. *Atmospheric Environment, 233*. doi:10.1016/j.atmosenv.2020.117611

Zeitz, M., Levermann, A., & Winkelmann, R. (2020). Sensitivity of ice loss to uncertainty in flow law parameters in an idealized one-dimensional geometry. *Cryosphere, 14*(10), 3537-3550. doi:10.5194/tc-14-3537-2020