General comments

This study investigates the scaling factors (or loading to deposition—LTD—factors) that are used to translate ice core sulfur measurements to estimates of the sulfate aerosol loading (or related quantities) from past large volcanic eruptions. This scaling procedure is a key step in the reconstruction of volcanic aerosol radiative forcing used in studies of past climate variability. The study revisits the estimates of Gao et al. (2007), which have been used in many subsequent reconstructions, and supplements that prior analysis with updated ice core data and some new analysis.

The topic is certainly important and relevant to the scope of Climate of the Past. The manuscript generally fairly presents the background of the research topic and illustrates the difficulty in obtaining an accurate and precise LTD factor from observations. The results of the study and its conclusions I find rather vague, and it remains unclear if the analysis actually contributes anything new to the topic.

The main conclusion of the study is that “the conversion factor may vary significantly among different eruptions.” There are two ways the LTD factor could vary between eruptions. First, there will be random perturbations in the sulfate deposition pattern over the area of the globe (and thus the ice sheets) and in the ice sheet average flux from eruption to eruption due to meteorological variability and other factors. This would lead to a different LTD factor when calculated for individual eruptions, but for many “equivalent” eruptions one would expect those values to cluster around a mean value. This mean value is really what we are interested in, since we can use it to estimate the most likely amount of aerosol loading given a certain ice cap average flux. An idea of the random variability in deposition for a given loading would be useful in order to associate an uncertainty to the estimated loading due to random uncertainty in the LTD factor—this requires consideration of many eruptions in order to have sufficient statistics. Toohey et al. (2013) show results
from model ensembles that provides some information on the variability in ice cap
deposition, and those results have been used to provide uncertainties on the sulfur
emission values given by Toohey and Sigl (2017).

The 2\textsuperscript{nd} way that the LTD could vary between eruptions—and the way the authors are
suggesting based on text in the Conclusions and elsewhere, is that the LTD could
vary \textit{systematically} for eruptions with different characteristics, e.g., latitude, season, etc.
This is of course possible, but to support this conclusion, one needs to show statistically
significant different LTD factors between eruptions of different characteristics. In other
words, one would need to show that differences in ice sheet flux between eruptions of
different eruption source parameters are larger than can be explained by the random
variations discussed above, and the uncertainties in the estimated LTD factor discussed
below. The present study does not do this, and so I find does not provide any concrete
evidence to support the notion that “the conversion factor may vary significantly for
individual eruption, depending on the volcano location, eruptive characteristic and
magnitude, etc.” as stated in the conclusions.

An estimate of LTD for any individual eruption from observations depends on information
on both the loading (hemispheric or global) and the ice sheet average deposition.
Uncertainty in the LTD factor is based on the uncertainties in both those quantities. This
paper focuses mainly on the uncertainty in the ice sheet deposition, but the uncertainty in
the loading is perhaps even more important. Even for Pinatubo, the best observed large
eruption we have, the best estimate of the SO\textsubscript{2} injection is 14-23 Tg (Guo et al., 2004). It
is not clear from the paper whether this uncertainty is included in the analysis. For
Tambora and Agung, it stands to reason since there are no satellite estimates available,
there is even more uncertainty in the injected sulfur amounts. For Tambora, the authors
address this to some degree by using apparent upper and lower limits to the Tambora
SO\textsubscript{2} injection, of 60 and 80 Tg. It is unclear though 1. where this upper limit comes from
(Self et al. 2004, estimate 53-58 Tg SO\textsubscript{2}, and the Gertisser et al.2012 study do not seem
to provide an alternate estimate) and 2. How well the 60-80 Tg SO\textsubscript{2} range really reflects
the true uncertainty in Tambora’s SO\textsubscript{2} injection. Oppenheimer (2003) lists estimates from
petrology that span 10-100 Tg S (so 20-200 Tg SO\textsubscript{2}), and petrological methods have been
found to have quite high uncertainties when compared to other estimates such as satellite
observations or ice cores. Part of the reason the Self et al. (2004) value is widely quoted I
believe is that it agrees well with ice core-based estimates for Tambora, which leads to the
potential for a circular argument. In any case, even if the Self et al. (2004) estimate of
around 60 Tg can be assumed to be the best estimate independent of the ice cores, the
real question is what is the uncertainty of that estimate. The small range quoted by Self et
al. (2004) seems to me unrealistically small, given the already mentioned uncertainties
noted in comparing petrologic versus satellite measurements. It is possible that the
uncertainty in the SO\textsubscript{2} injection from Tambora is so large that the range of LTD values
one would obtain from it would so wide as to provide no constraint compared to the other
data (from bomb tests or Pinatubo). The challenge for the authors is to make a convincing
argument that Tambora does provide some information to the discussion. For Agung, I did
not even find an explanation of where the SO\textsubscript{2} injection (or aerosol loading) estimate is
obtained from or what its uncertainty is. A related major difficulty for Agung is the
hemispheric partitioning of the aerosol: the authors mention they assume a 2:1 SH:NH
spread of aerosol (without reference), but other studies point to an even stronger SH
component (e.g., Stothers et al. 2001). This uncertainty needs to be included in the
analysis.
Specific comments:

L10: “CMIP5 and CMIP6 volcanic forcing” is ambiguous, since there were different forcings used in those 2 projects and within each project, for example for simulations of the historical era and on paleo timescales. More specificity is required.

L10: Language needs to be more generally understandable—a reader not deeply familiar with the subject likely wouldn’t understand what is meant by “Tambora deposition”.

L11: I do not believe the Monte Carlo procedure used herein constitutes a “model”. A model is a mathematical framework that describes a system.

L16: What model is referred to here?

L22: achieves->archives

L23: You can remove the word “may” since of course there are uncertainties, and change “introduce” to something like “involve”—the method has uncertainty but it is still decreasing our overall uncertainty compared to other methods (or else we wouldn’t use it).

L24: the second part of this sentence (“limitations in the conversion factor”) is a tautology, needs more explanation.

L27: combing->combining

L36: Reference needed for Pinatubo aerosol loading estimate

L41: Again, need to be specific about the CMIP experiments which use this information.

L42: Not all the models here have comprehensive chemistry modules, so they should not be referred to as chemistry-climate models.
L52: “overlooked” is too strong, this is largely the motivation for the Toohey et al. (2013) and Marshall et al. (2018) studies.

L52: What is meant by “high depth resolution”?

L61: “Multimodel”

L68: L cannot be either the mass of SO2 or sulfate aerosols, of course you will get very different answers if you use one or the other (without some sort of conversion).

L75: PARCA needs to be explained if it is used.

L78: This sentence says that the authors have performed an extraction of the volcanic sulfate flux from these two ice cores independent of the work of Sigl et al. (2015). It would be helpful then to briefly present how the flux values the authors compute compare to those of Sigl et al. (2015).

L86: Not quite clear if here again the authors have performed an independent estimation of the volcanic sulfate flux for these Antarctic cores? If so, a comparison with Sigl et al., (2015) would be quite useful.

L91: How comparable the magnitudes of sulfur emission from Pinatubo and Agung are is arguable. In any case, does it really matter if they are, and does it matter that Agung is at a similar latitude as Tambora? If so, why?

L112: Where does the 60-80 Tg SO2 range come from? Self et al. (2004) quotes 53-58 Tg SO2, and Gertisser et al. (2012) don’t seem to provide any independent estimate.

L125: There’s a logical problem here. If the flux to Antarctica and Greenland is similar, to assume this means the aerosol partitioning is symmetric assumes a similar LTD factor for the two ice sheets. But then you use the assumption of even partitioning to calculate the LTD factor for the two ice sheets. This is circular.

L145: The results stated here seem rather obvious results of the resampling procedure, but miss the point that the width of the distributions increases for smaller sample sizes, meaning that the uncertainty of a single sample (of some few ice cores) increases as n
decreases. That the SD of the LTD decreases as 1/\sqrt{n} simply confirms that the typical standard error of the mean (SD/\sqrt{n}) is a suitable assumption for the data, but it is not clear if this has any physical meaning or utility for the present purposes.

L150: To say that “the precision of LTD values is related to the limit in the number of cores” is trivial, this is basic statistics. If you want to say “the precision of LTD values is ONLY related to the limit in the number of cores” then I would argue this is incorrect, because it depends on the random error of each ice core (from measurement noise or other factors), which is likely quite variable between different ice cores. If an ice core is particularly noisy, then adding it to the composite may increase the overall uncertainty. This analysis tells us nothing about the potentially very different errors of the individual ice cores.

L151: The convention of quoting precision is unclear to me, I am used to percent precision for a value which is \( x \pm y \) as \( 100\cdot \frac{y}{x} \), so a smaller percent precision means more precise. That doesn’t seem to be the case here.

L182: I think your argument here has to do with aerosol particle size distribution, but this needs to be explained more clearly.

L193: “A series of...”?

L216: If BTD is the same thing as LTD, please use the same name for it.

L226: flux->aerosol

L267: This last statement makes no sense. If you use the LTD derived from Tambora on the ice core values for Tambora, you will of course get a loading estimate that is equal to the loading you used to calculate the LTD! You might as well just use the original loading estimate.

References


