

General comments:

The authors have made a number of significant improvements to the manuscript, and I appreciate the work they have done to address the comments of the first round. I do however have a few remaining issues that need to be dealt with before publication.

While the authors acknowledge the difficulty of accurately estimating the actual global mean temperature response to the aerosol forcing associated with Pinatubo, they continue to refer to the change in temperature response produced by their new dataset and parameterization scheme as an improvement. For example, in the abstract, they now say: “the new scheme accurately reproduces the observed global mean temperature response but observed and modelled climate variability preclude statements as to the significance of this improvement.” If you have a difference, but you can’t say if the difference is significant, how do you know it is an improvement?

As mentioned in the last review, a number of studies (e.g., Thompson et al., 2009, Santer et al., 2001) estimate that the cooling associated directly with Pinatubo was on the order of 0.4°C, and that the El Nino event of 1992 perturbed global mean temperatures such that the actual observed global mean temperature anomaly was much less, i.e., around 0.2°C as seen in the raw temperature time series. If one believes these studies, then the new aerosol scheme of this study actually worsens the global mean temperature response!

The objective and important piece of information that is missing from the abstract is that the new method decreases the global mean temperature anomaly for Pinatubo. I also agree that this produces a closer agreement between the simulated ensemble mean temperature response and that observed after Pinatubo. But all statements which refer to this as a general “improvement” are not justified, and need to be removed from the manuscript.

Specific comments:

Page 1, line 18-22: This may be true of the ensemble mean, but the observed temperature time series lies within the variability of the ensemble, therefore, the correct interpretation must be that the model ensemble and the observations are not inconsistent with each other.

P1, l27: Still not clear on what justification this is called an improvement. In any case, the objective piece of information important for the reader here would be to state that the new scheme *decreases* the global mean temperature response by half.

P2, l4: the Zanchettin et al. PAGES magazine article nicely makes this point: http://www.pages-igbp.org/download/docs/magazine/2015-2/PAGESmagazine_2015%282%29_54-55_Zanchettin.pdf

p2: l17: “each of the colossal eruptions over the second half of the 20th century”:

Actually, the model response to Fuego seems to be smaller than the observed temperature change. For El Chichon and Agung, it looks like the ensemble mean response is about the same as the observations. For Pinatubo, the CMIP5 ensemble mean lies right on the observations, and the observations are definitely within the range of the model results. There is just no clear evidence here for the statement that the models are overestimating the cooling response to volcanic forcing.

P2, l20: "Most models...": The rest of the sentence does not apply to most models. I am not aware of any model other than CCSM4 that used SAD to drive the radiative anomalies in CMIP5. Better to remove this part, or write a more general sentence about most models, and a second about CCSM4.

P3, l12 (and elsewhere): please remove the word "fully" – I doubt the reference really contains all possible information... a relative descriptor like "more fully" or "in more detail" would be believable, but a "full" description is impossible.

P3, l23: Stenchikov gives a forcing for Pinatubo based on satellite observations. Ammann's database is based on a parameterized transport scheme, and the Ammann paper reference Stenchikov only in regards to estimates of effective radius. I don't see how Ammann's reconstruction is "built upon" Stenchikov's.

P5, l25: what is MIEV0? A software routine I assume?

P6, l19: There is no description of aerosol in this Eyring et al. 2010 reference. There must be a better reference that at least describes what this forcing file is based on. Satellite observations? Model results?

P7, l28: The Eyring and Lamarque article in the SPARC newsletter has actually some description of the stratospheric aerosol forcing set used for CCMI, and seems to be the (only) reference that matters here:
http://www.sparc-climate.org/fileadmin/customer/6_Publications/Newsletter_PDF/40_SPARCnewsletter_Jan2013_web.pdf

p7, line 28: More detail about "this file" needed? What is it? Where does it come from? Where can it be obtained?

P8, l31-32: If this statement belongs in the paper at all, it should be in a discussion/outlook nearer the end of the paper. Also, it should also be noted then that an update of the CCMI/SAGE-4lambda dataset spanning the years 1850-present is also in preparation, and will be the recommended forcing for CMIP6 historical simulations.

P9, l26: also important is that the aerosol information varies with latitude.

P11, l4: does "new... parameterization" include the newly applied CCMI forcing data set? Most readers would think no. I suggest adding the dataset to the sentence.

P11, l20-23: This statement is quite wrong: it's the Ammann et al, "Original CCSM4/CESM1" that misses the background aerosol.

P12, l13: this reduction does not appear to be "significant". The 1-sigma ranges overlap. With 5 ensemble members, 1-sigma is the same as a 2-sigma standard error of the mean, so I would not expect these values to be significantly different at 95% or greater confidence. "Significant" has to be removed, or a more in depth and convincing analysis applied.

P13, l15: improvements-> changes. It's not convincingly shown that this change is an improvement.

P14, l1: this statement applies only to the Pinatubo eruption which was shown in the paper... it is not necessarily a general improvement.

P14, l1: again, remove "improves". It is clear that it decreases the global mean temperature response, this is the important point.

P14, l14: "Input data availability": Please include here where the original CCMI forcing data set is made available.

Editorial comments:

Page 1, line 12 and 14: "parameterisation" spelled differently in two uses.

P1, l20: remove "by"

P1, l20: observed->reconstructed, or similar. GISTEMP is not an instrument that observes.

P1, l24: schemes

P2, l3: ".. to stratospheric aerosol" or "to the stratospheric aerosol layer"

P2, l3: ill-represented in climate model simulations, not in reality

P2, l27: these models are not simulating the evolution of aerosol plumes.

P2, l31-32: these references seem ill-placed, since the focus of the sentence should really be on the CCMI forcing data set.

P3, l4: "prescriptions"

P3, l6: "all configurations of"

P3, l15: replace or remove "full"

P4, l14: replace or remove "full"

P4, l23: properties-> factors

P4, l24: long wave band (or similar)

P5, l1: I think it's "independent" you mean

Eq. 3: Q_{exasmt} is a typo I assume?

P7, line 29: it varies in latitude as well.

P9, l11: spelling

P11, l23: again, "full"

P13, l29: "larger": than what?

Table 1: missing subscript of r_g in column 7, row 5.

Figure 2 caption: specify AOD at 550 nm

Figure 3 caption: specify wavelength of SAOD

Figure 6 caption: specify latitude range of tropical average.

References:

Santer, B. D., Wigley, T. M. L., Doutriaux, C., Boyle, J. S., Hansen, J. E., Jones, P. D., Meehl, G. A., Roeckner, E., Sengupta, S. and Taylor, K. E.: Accounting for the effects of volcanoes and ENSO in comparisons of modeled and observed temperature trends, *J. Geophys. Res.*, 106(D22), 28033–28059, doi:10.1029/2000JD000189, 2001.

Thompson, D. W. J., Wallace, J. M., Jones, P. D. and Kennedy, J. J.: Identifying Signatures of Natural Climate Variability in Time Series of Global-Mean Surface Temperature: Methodology and Insights, *J. Clim.*, 22(22), 6120–6141, doi:10.1175/2009JCLI3089.1, 2009.