Review of: Unveiling Sulfate Aerosol Persistence as the Dominant Control of the Systematic Cooling Bias in CMIP6 Models: Quantification and Corrective Strategies by Jie Zhang et al, for ACP. Version 2

Stephen E. Schwartz, Reviewer; 2025-0715; modified 2025-0716

I remain unpersuaded by the changes in this manuscript relative to Version 1.

Rather than change the definition of the quantity previously denoted Effective Sulfate Retention Time ESRT to something like a turnover time for a species, or yield for sulfate, which would have physical meaning, the authors simply changed the name of this quantity to Sulfur Assessment Metric for Earth System model evaluation SAME. Changing the name of a physically non-meaningful quantity does not make it physically meaningful.

I think that the key extensive measure of sulfate influence would simply be the sulfate burden, as that is the quantity that does the forcing, or the sulfate forcing itself. Burden anomaly vs SAT anomaly is shown in Fig 2a (Better the other way around, Burden anomaly being the independent variable and SAT anomaly being the dependent quantity). This burden is a consequence of model parameters that govern formation and removal, such as reaction coefficients and removal coefficients, or the SO2 emission rate.

Given the anticorrelation between change in SAT and sulfate burden, the authors are evidently trying to find the reason for the differences in sulfate burden across the models. This is laudable. Such difference might be due to different emissions, different sulfate yield per emitted SO2 or different sulfate lifetime (turnover time) or some combination. The latter two are physically meaningful intensive properties of the system. I argued previously that the ESRT, now renamed SAME, although an intensive property, is not a physically meaningful quantity. Figure 4a seems to make a strong case that the differences in sulfate burden among models is due to different total SO2 + sulfate sink anomaly, the slope being ESRT, now renamed SAME, which differs for the different models. But showing this correlation does not point to a path forward in terms of assessing the accuracy of representation of processes in models. Modelers cannot directly modify representation of SAME in their models because SAME is not a physical quantity.

It would also seem essential to compare the relative values of SO2 sink rate and sulfate sink rate; does it make sense to add these quantities if one or the other is dominant? Are the differences across models due to differences in sulfate sink rate or SO2 sink rate? These questions speak to differences in the consequences of different representations of processes. ESRT, now renamed SAME, does not.

One wonders what a plot of sulfate burden anomaly vs sulfate sink rate anomaly like Fig 4a would look like; this slope would be sulfate turnover time, a physically meaningful quantity that can be compared to measurements such as the studies by Cambray and by Kristiansen that I called attention to in my previous review.

One wonders whether there is appreciable difference between such plots during PHC and non PHC periods. The quantities are intensive, depending on the same representation of physical processes, so in principle they depend on the governing physics, not on whether the data are from the PHC or non-PHC time periods. If the slopes are different in non PHC periods, that would be interesting. That would direct attention to representation of processes in models.

The authors should be able to evaluate sulfate yield (rate of production of sulfate/rate of SO2 emission) across the models. Again, this is a physically meaningful quantity. Is there appreciable difference across the models.

If there is little correlation across the models between sulfate burden vs yield, and between sulfate burden vs sulfate dep rate, that would suggest some anticorrelation across the models between yield and sulfate lifetime, which would be an important finding.

The authors state (abstract) that that their metric SAME is overestimated by almost all the CMIP6 models. One would ask what is the basis for this statement. Overestimated relative to what standard?

Minor comment: I wonder whether the terms “anomalous sulfate deposition rate” etc. might be replaced by “sulfate deposition rate anomaly” etc. throughout the text, as used in figure axis labels, and consistent with commonly used “temperature anomaly”

Minor comment: I do not care for the terminology "pot hole cooling period" (PHC). A simple reference to the time period 1960-1990, during which GMST decreased (or perhaps better, did not exhibit same rate of increase as in earlier and later times). Language matters.

Minor comment. Equations, such as the one given for SAME should be written with algebraic symbols, not words.

Important comment for authors and editor, should this paper reach the acceptance stage: Tables should be presented of all graphical quantities such as time series, as in Fig 1, and x-y plots as in Figs 2-4, in an appendix or supplementary material. These are highly processed data. It is wholly unacceptable simply to state that the data can be freely downloaded from the EGSF nodes.

Some further thoughts on the manuscript 2025-0716

In my review I did not pay much attention to the assumption inherent in Fig 2 that the change in global surface air temperature would be proportional to sulfate burden. The sulfate forcing F_S, the rate of change in Earth heat content due to sulfate, $\mathrm{d}H_S/\mathrm{d}t$, is proportional to the sulfate burden S,

$$F_S = \mathrm{d}H_S/\mathrm{d}t = -\mathrm{k}S$$

where k is the constant of proportionality; the minus sign denotes cooling. One would not necessarily expect temperature change $\mathrm{d}T/\mathrm{d}t$ to be proportional to $\mathrm{d}H_S/\mathrm{d}t$, because of time lag and damping. The time lag is only a few years (e.g., Held et al., 2010) so there would seem to be a possibility of discerning change in T resulting from a change in sulfate burden, especially if the forcing is sustained. I think the authors should discuss this.

I also did not pay much attention to Table 2 of the Revision, an addition to the manuscript. It should be explained how the quantities in the table were determined, and the time period to which they pertain. In principle all the quantities – sulfate burden, sulfate wet dep rate, sulfate dry dep rate, sulfate lifetime – are time dependent. What would be much more valuable would be time series of these quantities, also rate coeffs for wet and dry deposition (ratio of dep rate to stock) rather than a simple table. Also sulfate yield. Are those intensive quantities more or less constant with time? Do they change during the time period of interest (1960-1990)? Are these quantities coherent with time series of SAT? Such analyses could turn this paper into an important study.

Held, I.M., Winton, M., Takahashi, K., Delworth, T., Zeng, F. and Vallis, G.K., 2010. Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, *23*(9), pp.2418-2427.