Replies to Dr. Stephen E. Schwartz's review on: 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.

Jie ZHANG, 2025-0605

Thank Dr. Schwartz for your constructive critique, particularly for emphasizing the importance of distinguishing the "effective sulfur residence time" (ESRT) from other lifetime measures reported in the literature. We acknowledge that the term "effective sulfur residence time (ESRT)" is potentially misleading.

ESRT was envisioned primarily as a diagnostic tool (not a physical timescale) for model tuning. Because it accounts for both sulfate and SO_2 deposition, its value is typically lower than the sulfate atmospheric lifetime. As Dr. Schwartz noted, it is fundamentally a metric for model evaluation rather than a conventional definition of atmospheric residence time and should not be interpretated as such. Therefore, we propose renaming it to the "Sulfur Assessment Metric for ESMs" (SAME).

We acknowledge that sulfate lifetime remains critical for validating the model's physical realism. Therefore, we calculated sulfate lifetime as the ratio of sulfate burden to total sulfate deposition (wet plus dry) in the CMIP6 models, BCC-ESM1-1, and UKESM1-1-LL (Table A1). Sulfate lifetime ranges from 1.65 days in MIROC models to 6.57 days in EC-Earth3-AerChem, which is consistent with previous literatures, particularly the estimates in AeroCom models (Textor et al., 2006). This wide range is attributed to variations in both sulfate burden (0.33 to 0.75 Tg S) and deposition rates (0.75 to 7.58 Tg S yr^{-1} for dry deposition and 31.68 to 69.41 Tg S yr^{-1} for wet deposition).

Table 1 Sulfate burden, sulfate depositions, and sulfate lifetime in CMIP6 models, BCC-ESM1-1 and UKESM1-1-LL in PHC period.

model name	Sulfate burden (Tg S)	Sulfate Deposition (Tg S yr-1)		Sulfate lifetime (days)
		DrySO4	WetSO4	
BCC-ESM1	0.59	2.07	43.78	4.70
CESM2	0.40	6.14	35.13	3.54
CESM-FV2	0.43	5.92	32.60	4.07
EC-Earth3-AerChem	0.75	1.29	40.39	6.57
GFDL-ESM4	0.46	7.58	31.68	4.28
MIROC6	0.33	7.50	61.29	1.75
MIROC-ES2L	0.33	5.67	67.42	1.65
MPI-ESM-1-2-HAM	0.74	2.41	69.41	3.76
MRI-ESM2-0	0.53	0.75	56.96	3.35
NorESM2-LM	0.52	6.39	40.33	4.06
UKESM1-0-LL	0.63	7.00	34.79	5.50
BCC-ESM1-1	0.48	1.34	19.2	8.53
UKESM1-1-LL	0.52	5.57	27.34	5.77

Sulfate lifetimes in BCC-ESM1-1 (8.53 days) and UKESM1-1-LL (5.77 days) are generally longer than in their previous versions (4.70 and 5.50 days, respectively). The longer sulfate lifetimes in the updated models may be due to lower SO_2 in these revised models but also could be due to physical climate changes (e.g., temperatures, clouds, rainfall). Compared to prior lifetime measures reported in the literature and considering the range of lifetimes found in recent models, the sulfate lifetimes in BCC-ESM1-1 and UKESM1-1-LL appear reasonable (e.g., Charlson et al, 1992; Kristiansen et al. 2012; Textor et al., 2006).

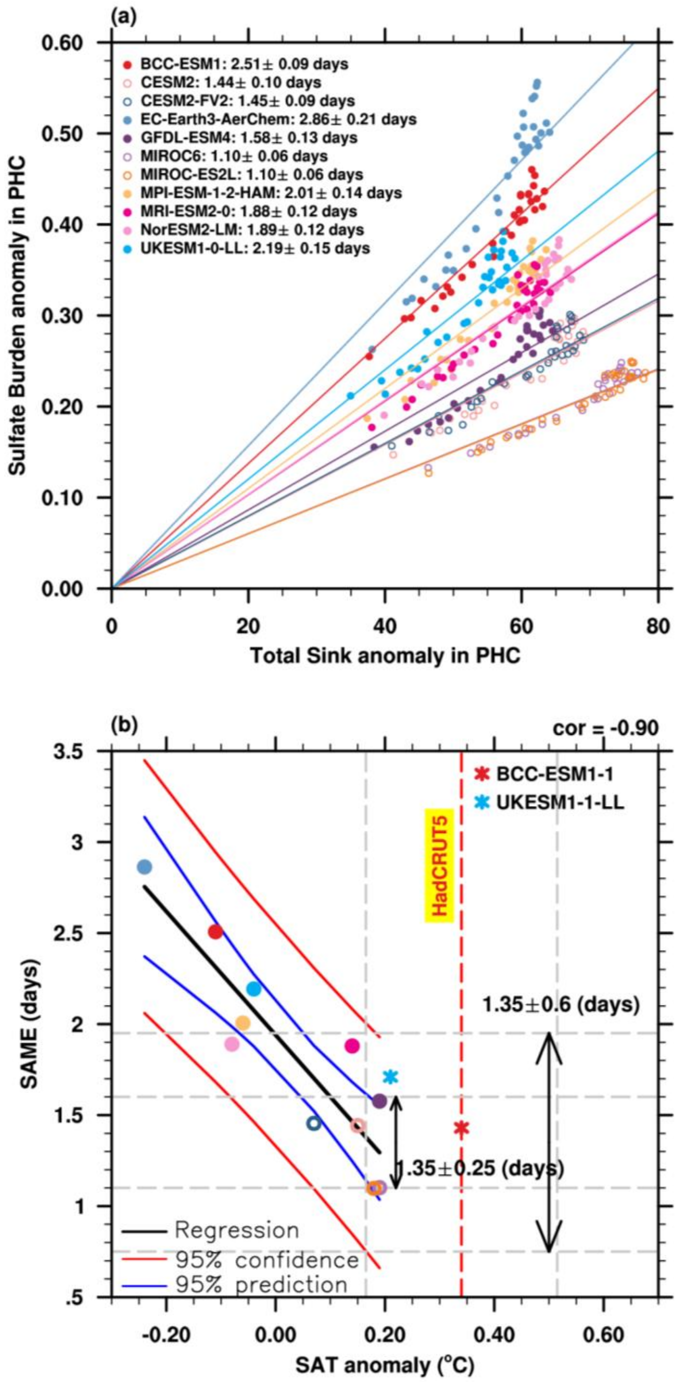

Figure A1. (a) Scatter plots of yearly total sulfur sink anomaly (x-axis, Tg S yr-1) versus sulfate burden anomaly (y-axis, Tg S) in PHC period in relative to 1850~1900 mean. Number in legend shows the mean and standard deviation of ratio between sulfate burden anomaly and total sulfur sink anomaly in PHC period, defined as SAME, units: days. (b) The SATa (^{o}C, x-axis) versus SAME (days, y-axis) in PHC period for each model. The black solid line is the linear fitting. The blue and red solid lines are the 95% confidence interval (CI) and 95% prediction interval (PI), respectively. SAT anomaly in HadCRUT5 and its 0.175^{o}C boundaries are shown by the red dashed line and parallel gray dashed lines, respectively. The red and blue asterisks are the results in the two post-CMIP6 models BCC-ESM1-1 and UKESM1-1-LL, respectively.

To eliminate the effect of differing climatological states across models, the SAME metric is defined as the ratio of sulfate anomaly during the PHC period to the sum of sulfate and SO_2 deposition anomalies:

SAME = loadSO4a / (Daso4 + Daso2)

where:

- loadSO4a is the total sulfate loading anomaly in the atmosphere,
- Daso4 denotes the total (wet plus dry) sulfate deposition anomaly, and
- Daso2 denotes the total (wet plus dry) SO_2 deposition anomaly during the PHC period.

The SAME ranges from 1.1 days in MIROC models to 2.86 days in EC-Earth3-AerChem (Fig. A1a). The correlation coefficient between SATa and SAME is -0.90 (Fig. A1b). In addition to the linear regression between SATa and SAME (black line in Fig. A1b), we also show the 95% confidence interval (CI, blue curves) and the 95% prediction interval (PI, red curves).

There are 4 models with SAT values around 0.165^{o}C (the lower limit of observation), giving a range of SAME between 1.1 to 1.58 days. Since most models underestimate SATa, it is difficult to predict SAME values when SATa is higher than the observation (0.34^{o}C). Results from BCC-ESM1-1 suggest that SAME may not need to decrease at the regression line rate beyond the lower limit (0.165 °C). Therefore, we recommend a SAME value of 1.35 ± 0.25 days based on the 95% CI (encompassing the four CMIP6 models and BCC-ESM1-1), or a wider range of 1.35 ± 0.6 days by the 95% PI that includes the UKESM1-1-LL.

Generally, the SAME metric is used to facilitate model tuning of the aerosol load, ensuring that models do not overestimating the aerosol cooling effect over the historical period, as was the case in CMIP6 and is a current concern for model performance in the upcoming CMIP7 experiments. To ensure model credibility, the sulfate atmospheric lifetime must align with observations (literature) values, which our calculations above confirm. The manuscript will be amended to include both the sulfate lifetime in CMIP6 models and the two updated models, along with the SAME metric and its recommendations for model performance.