Supplement of: Recent Modelling Studies Systematically Underestimate the Warming from IMO2020 Shipping Regulations

Josh Smith[1], Matthew Henry[1], Masaru Yoshioka[2], Ben Johnson[3], and Jim Haywood[1]

[1]Department of Mathematics, University of Exeter, Exeter, UK
[2]School of Earth and Environment, University of Leeds, Leeds, UK
[3]Met Office Hadley Centre, Exeter, UK

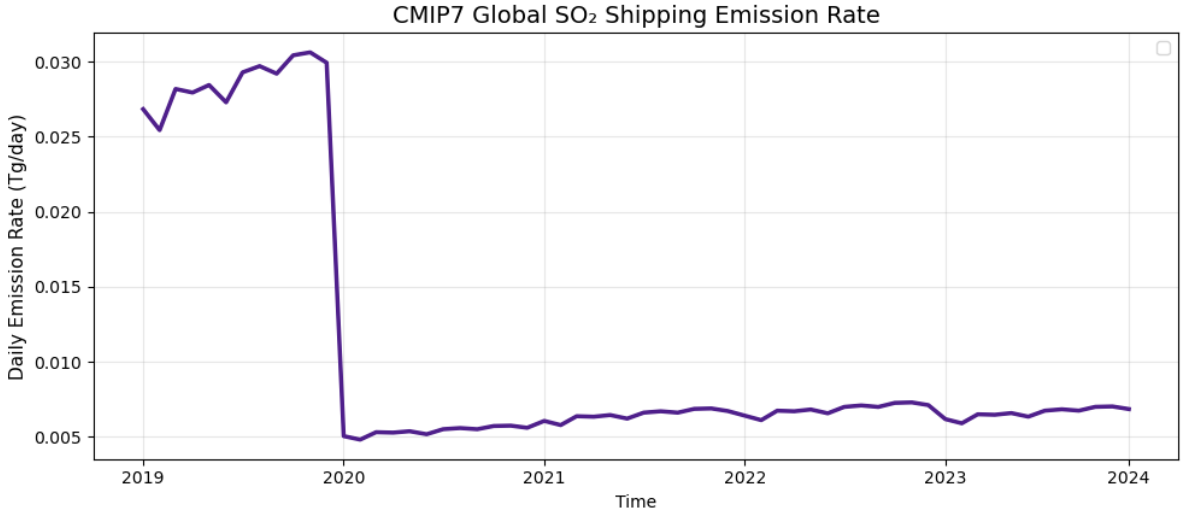

Figure S1. Global mean monthly shipping SO_2 emissions [$\mathrm{Tg\,day^{-1}}$] from the CEDS CMIP7 dataset used in our control (CTL, 2019 repeated) and IMO (2020–2023, with 2023 repeated) simulations.

Table S1. Additional sensitivity experiments and results with Base from main text for reference. All experiments use shipping SO_2 emissions injected at $\sim$20 m altitude. The κ-fix column indicates whether the updated hygroscopicity values are applied.

Experiment		Ens. size	κ-fix	Act. scheme	f_{SO_4} (%)	Marine DMS	Primary SO_4 size	ERF ($W\,m^{-2}$)
Default	CTL IMO	10 10	Off	ARG2000	2.5	1×	50 % Acc. (150 nm), 50 % Coarse (1500 nm)	0.132
Base_ARG2000	CTL IMO	10 10	On	ARG2000	2.5	1×	50 % Acc. (150 nm), 50 % Coarse (1500 nm)	0.097
Base	CTL IMO	10 10	On	ARG-G25	2.5	1×	50 % Acc. (150 nm), 50 % Coarse (1500 nm)	0.100
Ait_25_ARG2000	CTL IMO	5 5	On	ARG2000	25	1×	100 % Aitken (60 nm)	0.513
Ait_25	CTL IMO	10 10	On	ARG-G25	25	1×	100 % Aitken (60 nm)	0.552
Ait_10_2xDMS	CTL IMO	5 5	On	ARG-G25	10	2×	100 % Aitken (60 nm)	0.364

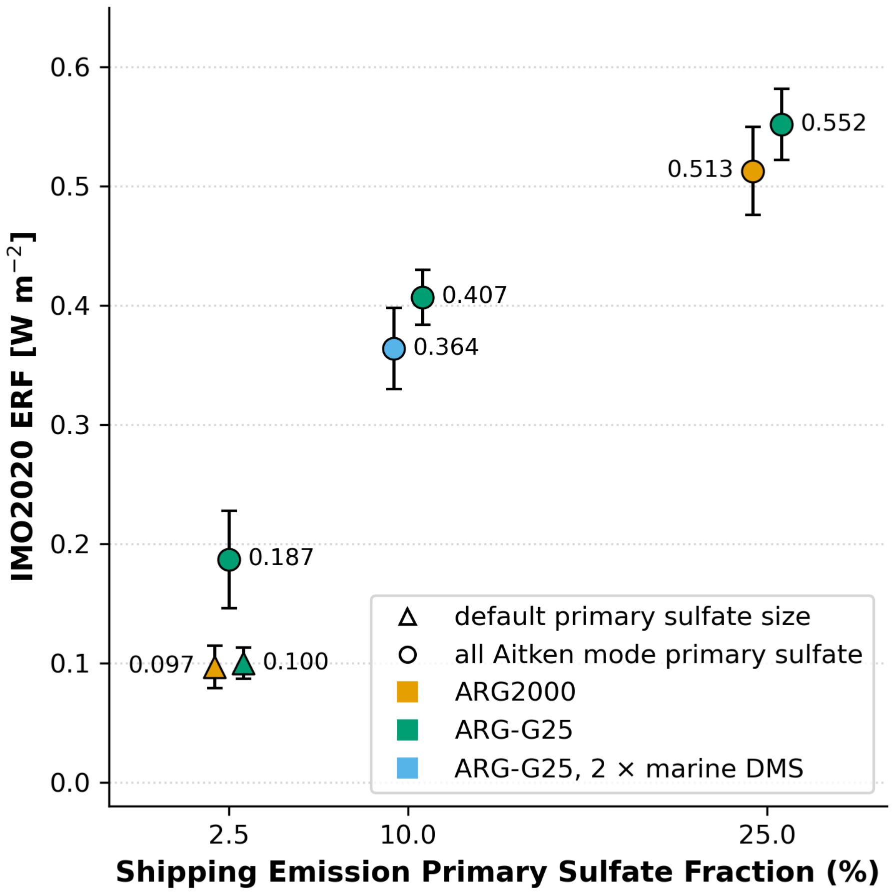

Figure S2. Results showing the relationship between shipping primary sulphate fraction and global mean ERF from IMO2020 along with sensitivities to primary sulphate size, activation scheme and background marine DMS emissions in isolation. Error bars are one ensemble standard error from the mean.

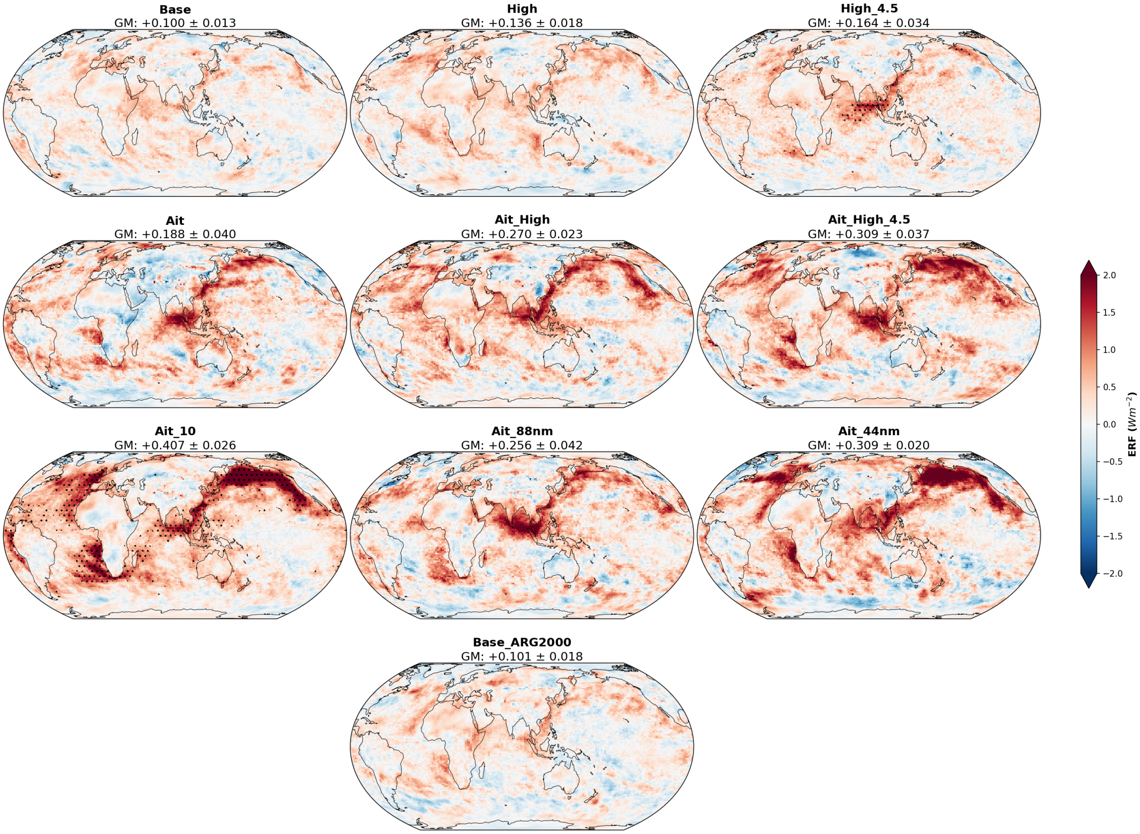

Figure S3. Ensemble mean IMO2020 ERF maps and global mean values from all experiments listed in Table 1 of the main text. Error intervals shown are one ensemble standard error from the mean. Stippling indicates grid elements where the local null hypothesis, assessed using the Wilcoxon signed-rank test, is rejected after controlling the expected false discovery rate at 5 % using the False Discovery Rate method (main text Figure 5).

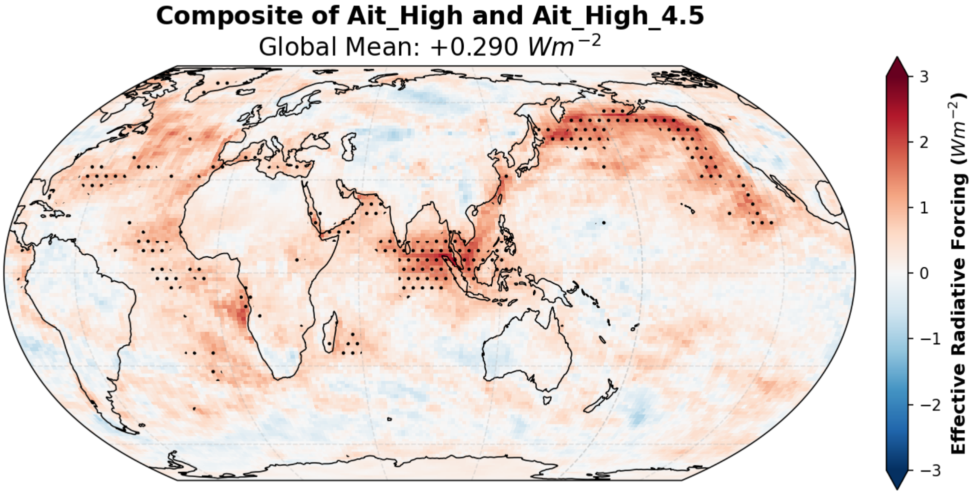

Figure S4. Composite ERF map and mean of Ait_High and Ait_High_4.5 experiments. Stippling indicates grid elements where the local null hypothesis, assessed using the Wilcoxon signed-rank test, is rejected after controlling the expected false discovery rate at 5 % using the False Discovery Rate method (main text Figure 5).

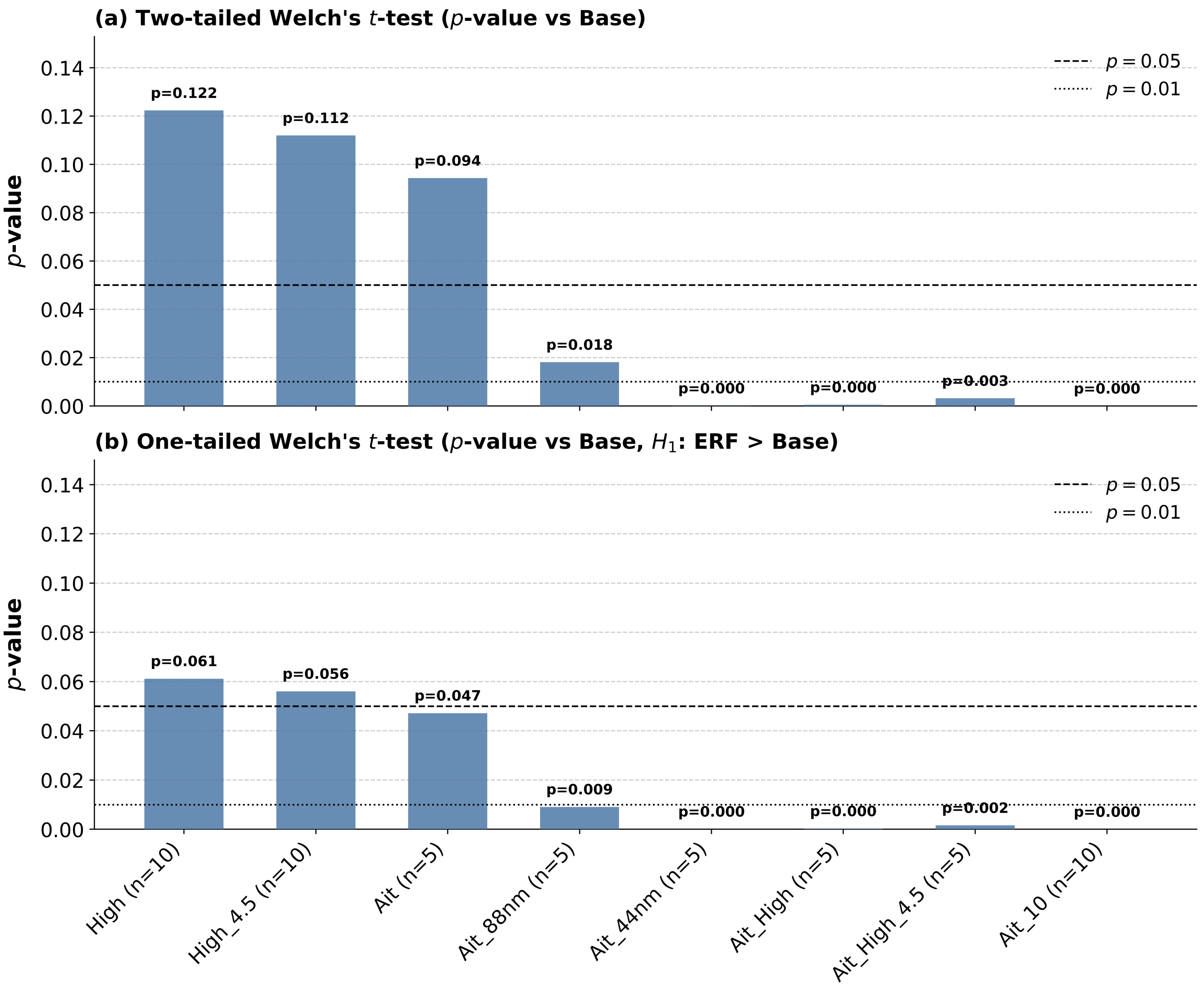

Figure S5. Statistical significance of ERF differences between each sensitivity experiment and *Base*, assessed using Welch's t-test on annual-mean ERF values across ensemble members. (a) Two-tailed test (H_1: ERF $\neq$ Base). (b) One-tailed test (H_1: ERF $>$ Base). Ensemble sizes (n) are indicated for each experiment. Dashed and dotted horizontal lines mark the $p = 0.05$ and $p = 0.01$ significance thresholds respectively. The accumulation- and coarse-mode sensitivity experiments (*High*, *High_4.5*, *Ait*) do not produce ERF differences from *Base* that are significant at the $p = 0.05$ level under the two-tailed test, whereas all Aitken-mode experiments incorporating updated primary sulphate size ($D_p \leq 88\,\mathrm{nm}$) yield highly significant ERF increases.

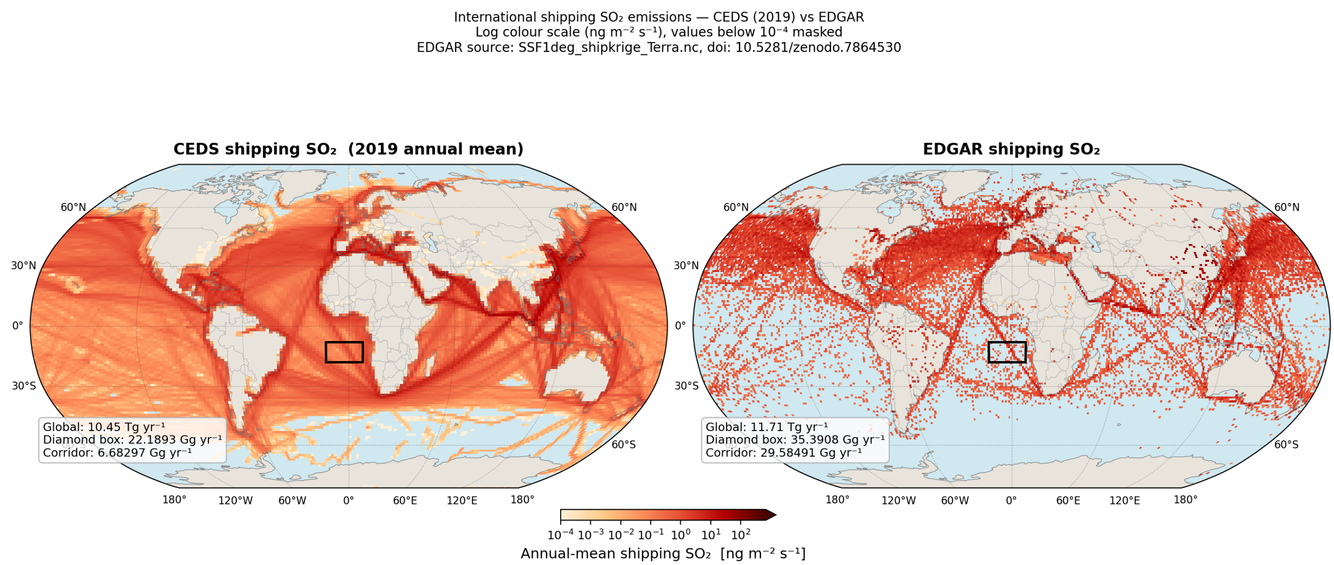

Figure S6. Comparison of annual-mean international shipping SO_2 emissions between the CEDS (2019; left) and EDGAR (right) inventories, plotted on a logarithmic colour scale ($\mathrm{ng\,m^{-2}\,s^{-1}}$); values below 10^{-4} are masked. The black box denotes the Southeast Atlantic analysis region used by Diamond (2023). Inset values report the global total emissions, the total within Diamond's analysis box, and the total within Diamond's corridor mask. The two inventories report comparable global totals (CEDS: $10.45\,\mathrm{Tg\,yr^{-1}}$; EDGAR: $11.71\,\mathrm{Tg\,yr^{-1}}$) but differ substantially in spatial distribution within the SEA analysis region. This contrast complicates direct application of Diamond's kriging-based clean-reference assumption to model output driven by CEDS emissions.