Supplementary Material for: Systematic Observation-Based Estimate of Effective Radiative Forcing from Aerosol–Cloud Interactions

Omer Roi-Cohen[1], Gaea Hadary[1], Casey J. Wall[2,3], Paulo Ceppi[4], and Guy Dagan[1]

[1]Fredy and Nadine Herrmann Institute of Earth Sciences, Hebrew University, Jerusalem, Israel
[2]Department of Meteorology, Stockholm University, Stockholm, Sweden
[3]Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden
[4]Department of Physics, Imperial College London, London, United Kingdom

S1 Central Estimates and Uncertainty for Figure 1

The central estimate for each configuration shown in Fig. 1a is computed by multiplying the regression coefficient $[\partial R_{\mathrm{net}}/\partial \log_{10}\mathrm{Proxy}]$ by the model estimate of $[\Delta \log_{10}\mathrm{Proxy}]$ from CMIP6 historical simulations at each grid cell, averaging the product over the spatial domain, and scaling to the global mean using the central estimate of the domain-to-global scaling factor $[\gamma]$ derived from CMIP6 aerosol-only simulations.

The uncertainty calculation at each grid box i follows Wall et al. (2022). Let ϵ_i represent the half-width of the 95% confidence interval of the grid box mean $\mathrm{ERF}_{\mathrm{aci}}$. We estimate ϵ_i as

$$\epsilon_i = t_i \sigma_i \sqrt{\frac{N_{\mathrm{nom},i}}{N_{\mathrm{eff},i}}} [\Delta \log_{10}\mathrm{Proxy}]_i, \tag{S1}$$

where σ_i is the standard error (SE) of the regression coefficient, $N_{\mathrm{nom},i}$ is the nominal number of temporal degrees of freedom, $N_{\mathrm{eff},i}$ is the effective number of temporal degrees of freedom, and t_i is the critical value of a Student's t distribution at the 95% significance level. The ratio $N_{\mathrm{nom},i}/N_{\mathrm{eff},i}$ is estimated as $(1+r)/(1-r)$, where r is the temporal lag-1 autocorrelation of the R'_{net} anomaly time series.

These grid-box half-widths are combined into a domain-average observational uncertainty, δ_{obs}, accounting for spatial autocorrelation using an EOF-based estimate of the effective number of spatial degrees of freedom (Bretherton et al., 1999):

$$\delta_{\mathrm{obs}} = \frac{\sqrt{\sum_{i=1}^{N_{\mathrm{obs}}} W_i^2 \epsilon_i^2}}{\sum_{i=1}^{N_{\mathrm{obs}}} W_i} \sqrt{\frac{N^*_{\mathrm{nom}}}{N^*_{\mathrm{eff}}}}, \tag{S2}$$

where N^*_{nom} and N^*_{eff} are the nominal and effective number of spatial degrees of freedom computed from the leading EOFs of R'_{net} over the analysis domain, and W_i is the area in grid box i.

Two additional uncertainty components are then combined with δ_{obs} in quadrature. The first arises from inter-model spread in the CMIP6 estimates of $\Delta \log_{10}\mathrm{Proxy}$: $\mathrm{ERF}_{\mathrm{aci,d}}$ is computed separately for each available CMIP6 model and the 95%

half-width δ_{Proxy} is taken as half the difference between the 97.5th and 2.5th percentiles of the resulting distribution. This percentile-based approach is used in place of the range-based method of Wall et al. (2022) to accommodate the varying number of models available for different aerosol proxies. The second additional component δ_γ is the half-width of the 95% confidence interval of $\mathrm{ERF}_{\mathrm{aci,g}}$ that arises from inter-model spread in the domain-to-global scaling factor γ, computed analogously from the RFMIP model distribution. The total 95% confidence interval on the global $\mathrm{ERF}_{\mathrm{aci,g}}$ is then defined by

$$\delta_{\mathrm{total}} = \sqrt{([\gamma]\delta_{\mathrm{obs}})^2 + ([\gamma]\delta_{\mathrm{Proxy}})^2 + \delta_\gamma^2}. \tag{S3}$$

For the activation pathway configurations, the aerosol sensitivity entering the above calculations is replaced by the product $([\partial R_{\mathrm{net}}/\partial \ln N_d]) \cdot ([\partial \ln N_d/\partial \log_{10} \mathrm{Proxy}])$ following Park et al. (2025), with uncertainty propagated analytically through the same residual-based approach applied independently to each regression step.

S2 Bootstrap Uncertainty for Opportunistic Experiments

Uncertainty in the opportunistic experiment predictions is estimated using a stationary bootstrap applied to the regression coefficients (Politis and Romano, 1994). For each bootstrap replicate, a new set of time indices is drawn from the observed monthly time series using a mean block length of 12 months, preserving the temporal autocorrelation structure of the data. The bootstrap is applied simultaneously across all grid cells using a shared set of calendar indices, so that the same months are resampled at every location within a given replicate. For each replicate, the regression is refitted on the resampled data to obtain a new set of coefficients, which are then multiplied by the observed CCF changes to produce a replicate prediction of ΔR_{SW}. The 66% and 90% confidence intervals shown in Fig. 3 are derived from the distribution of 1,000 such replicates. We use this bootstrap-based uncertainty rather than analytical confidence intervals because analytical intervals tend to widen as the number of predictors increases, which would artificially penalize the non-local regression methods and make cross-method comparison unfair. The bootstrap approach ensures that the displayed uncertainty reflects only the sampling variability of the predictions, independent of the number of predictors in each configuration.

S3 Additional Figures

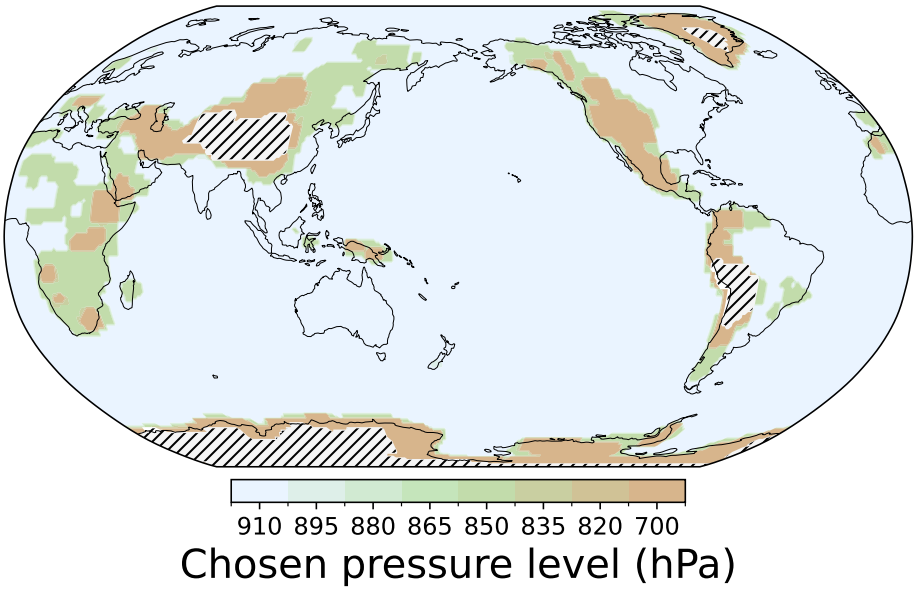

Figure S1. Global map of the selected atmospheric pressure levels (hPa) over land used for extracting the sulfate mass concentration (SO_4) proxy. To ensure spatial and temporal consistency across the MERRA-2 and CAMS reanalyses, as well as the CMIP6 historical simulations, the highest valid pressure level between 910 hPa and 700 hPa was selected for each grid cell. Over the ocean, the 910 hPa level is used uniformly. Mountainous land regions with a surface pressure below 700 hPa are excluded from the analysis.

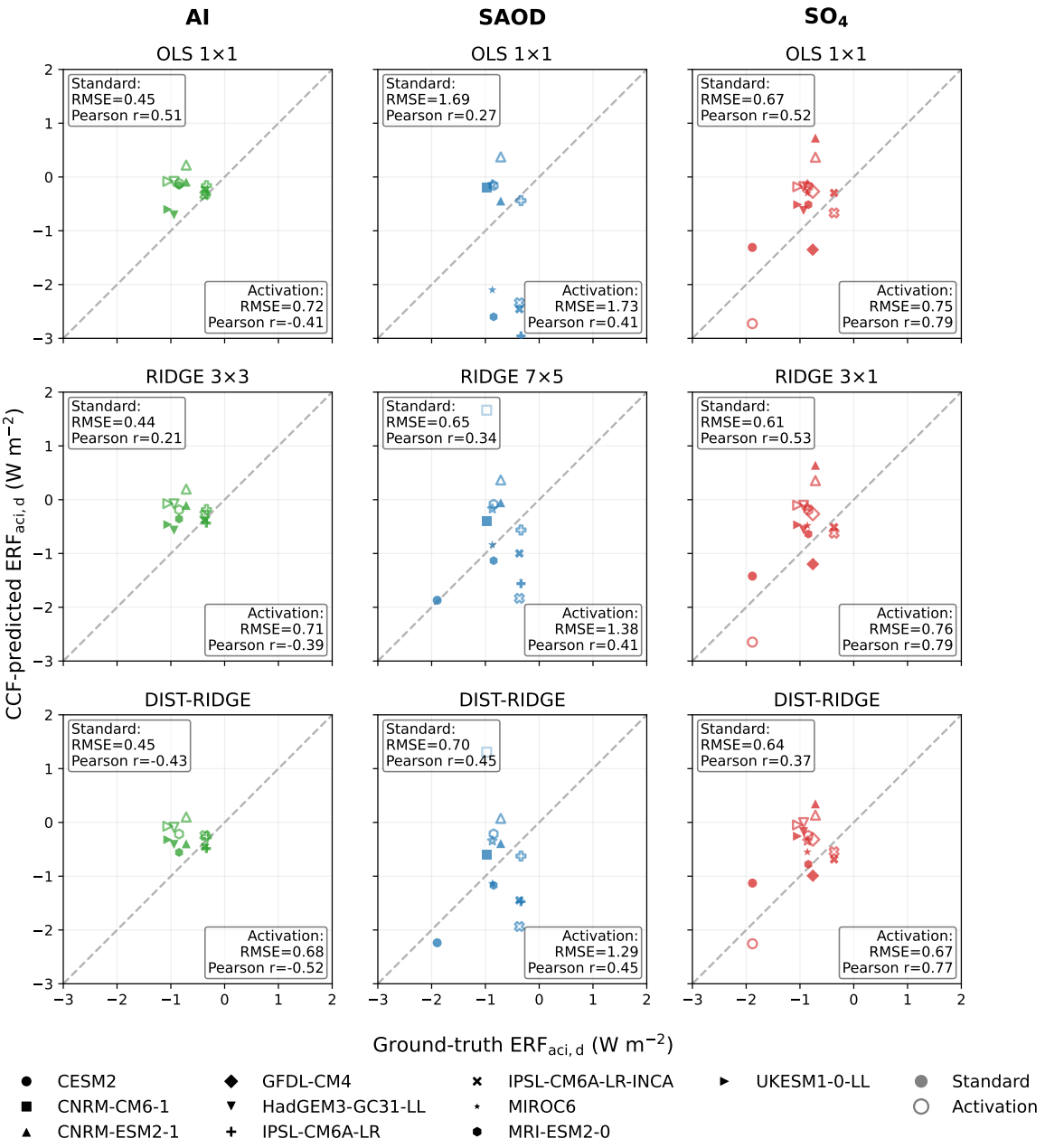

Figure S2. Per-model scatter plots of CCF-predicted versus "ground-truth" domain-averaged $\mathrm{ERF}_{\mathrm{aci,d}}$ (W m^{-2}) for all nine combinations of regression method (rows: OLS 1×1, RIDGE, and DIST-RIDGE) and aerosol proxy (columns: AI, SAOD, SO_4), evaluated over the combined ocean-and-land domain. Each marker represents one CMIP6 model, with shapes distinguishing individual models as indicated in the legend. Solid filled markers denote the standard direct proxy route; open hollow markers denote the intermediate droplet activation route following Park et al. (2025). The dashed diagonal line indicates perfect prediction (1:1 line). RMSE and Pearson correlation coefficient r are reported in text boxes separately for the standard and activation routes within each subplot. For the Ridge regression row, the domain shown corresponds to the best-performing domain size for each respective proxy (3×3 for AI, 7×5 for SAOD, 3×1 for SO_4).

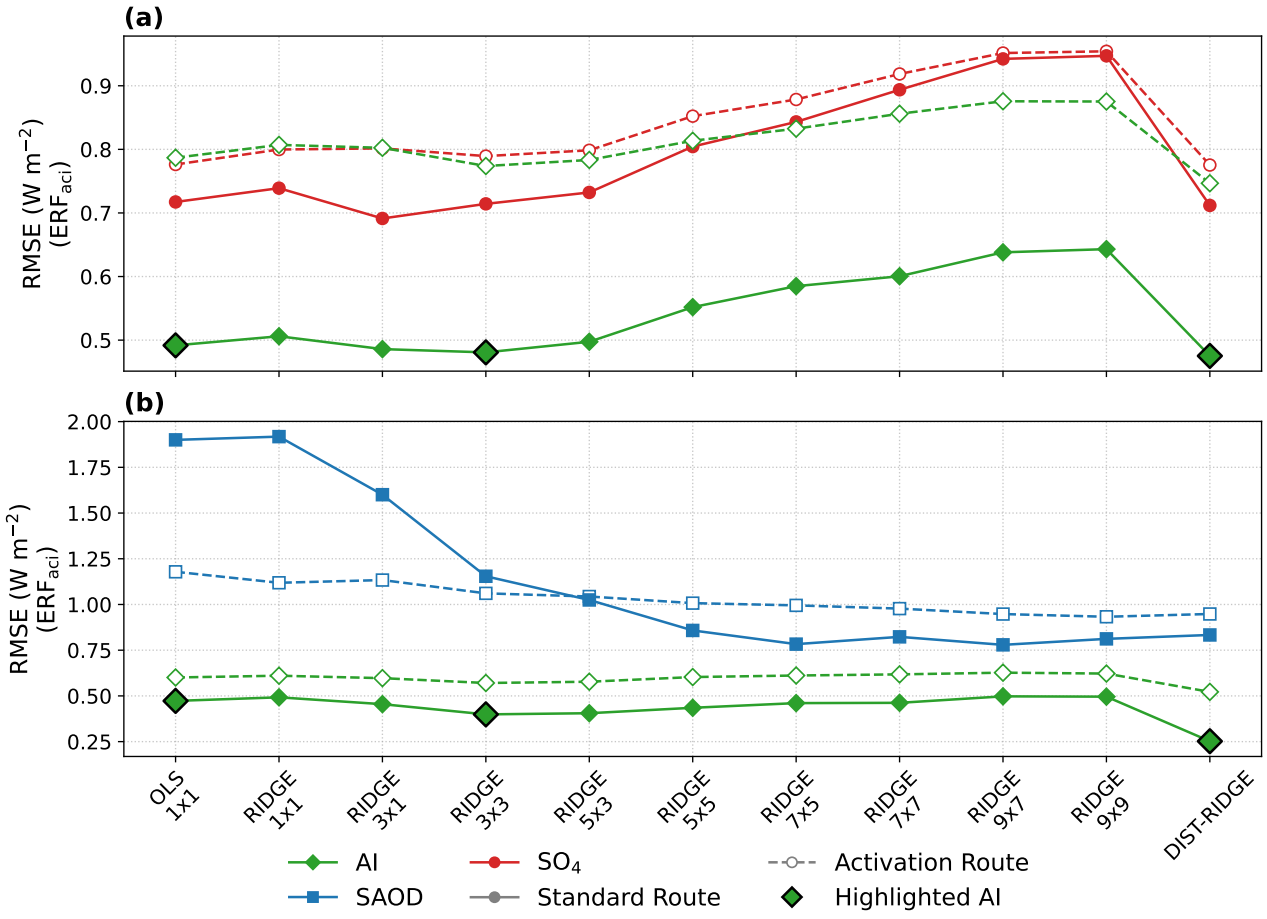

Figure S3. As in Fig. 2a, but comparing matched model subsets to isolate the effect of aerosol proxy choice from differences in model availability. Panel (a) compares AI (green diamonds) against SO_4 (red circles), using the 5 models that provide output for both proxies; panel (b) compares AI (green diamonds) against SAOD (blue squares), using the 4 models that provide output for both proxies. In each panel, only models providing output for both proxies under comparison are included, ensuring that RMSE differences reflect proxy choice alone rather than the composition of the model ensemble. Solid lines with filled markers denote the standard direct proxy route; dashed lines with open markers denote the intermediate droplet activation route. Highlighted black-outlined green diamonds mark the three best-performing AI standard-route configurations identified from Fig. 2a of the main text (OLS 1×1, RIDGE 3×3, and DIST-RIDGE). In both panels and across both physical pathways, AI achieves lower RMSE than the alternative proxy across nearly all regression configurations, confirming that the better predictive skill of the AI proxy is not an artifact of differences in the set of models available for each proxy.

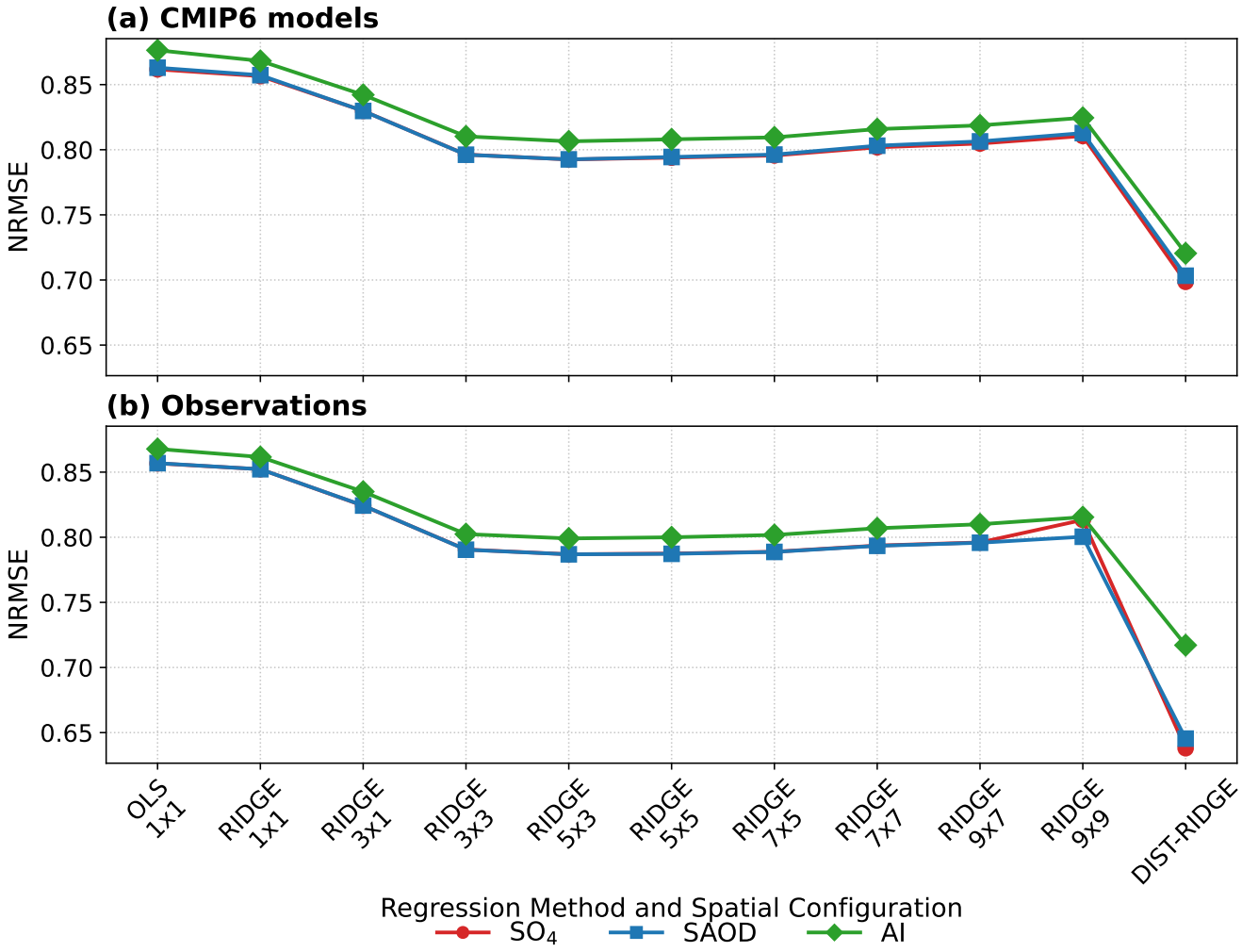

Figure S4. Out-of-sample prediction skill across regression methodologies for models and observations. Normalized Root Mean Square Error (NRMSE) for predicting low-cloud radiative anomalies using leave-one-year-out cross-validation. Panel (a) evaluates the CMIP6 historical model ensemble, while panel (b) evaluates CERES observations. Colored lines denote the tested aerosol proxies: SO_4 (red circles), SAOD (blue squares), and AI (green diamonds). The x-axis organizes the methods by increasing spatial complexity, from local OLS to various Ridge regression domain sizes, up to the global Distance-Weighted Ridge (DIST-RIDGE). Both models and observations exhibit nearly identical behavioral shapes across the regression techniques: prediction error drops from OLS to Ridge 3×3, plateaus or slightly increases, and then falls sharply to reach its lowest value under the DIST-RIDGE configuration.

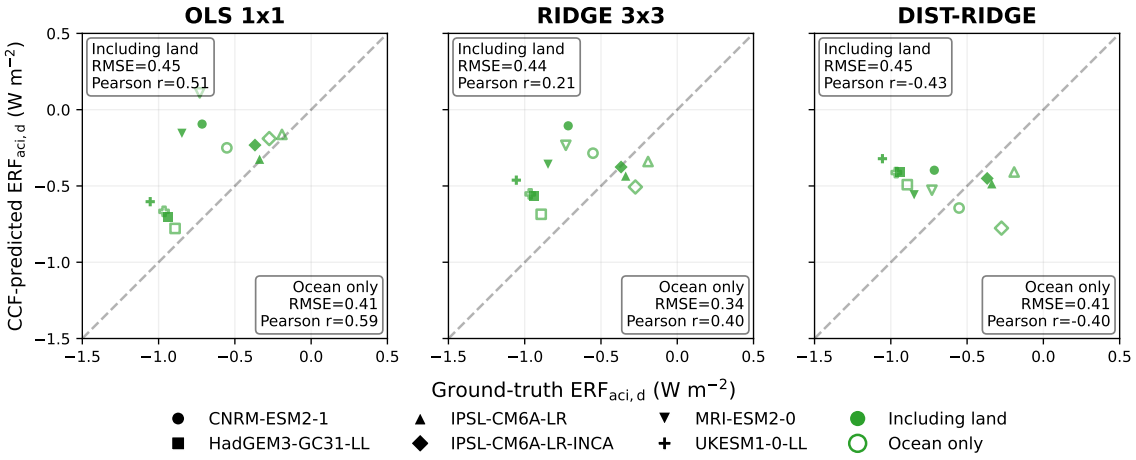

Figure S5. Scatter plots of CCF-predicted versus "ground-truth" $\mathrm{ERF}_{\mathrm{aci,d}}$ (W m^{-2}) for the three best-performing AI configurations (OLS 1×1, RIDGE 3×3, and DIST-RIDGE), comparing the combined ocean-and-land spatial domain (solid filled green markers) against the ocean-only domain (open hollow green markers). Each marker represents one CMIP6 model, with shapes distinguishing individual models as indicated in the legend. The dashed diagonal line indicates perfect prediction (1:1 line). RMSE and Pearson correlation coefficient r are reported separately for each spatial domain within each subplot. Raw RMSE values are similar between the two domains across all three configurations, indicating that expanding the analysis domain to include land does not degrade $\mathrm{ERF}_{\mathrm{aci,d}}$ prediction accuracy.

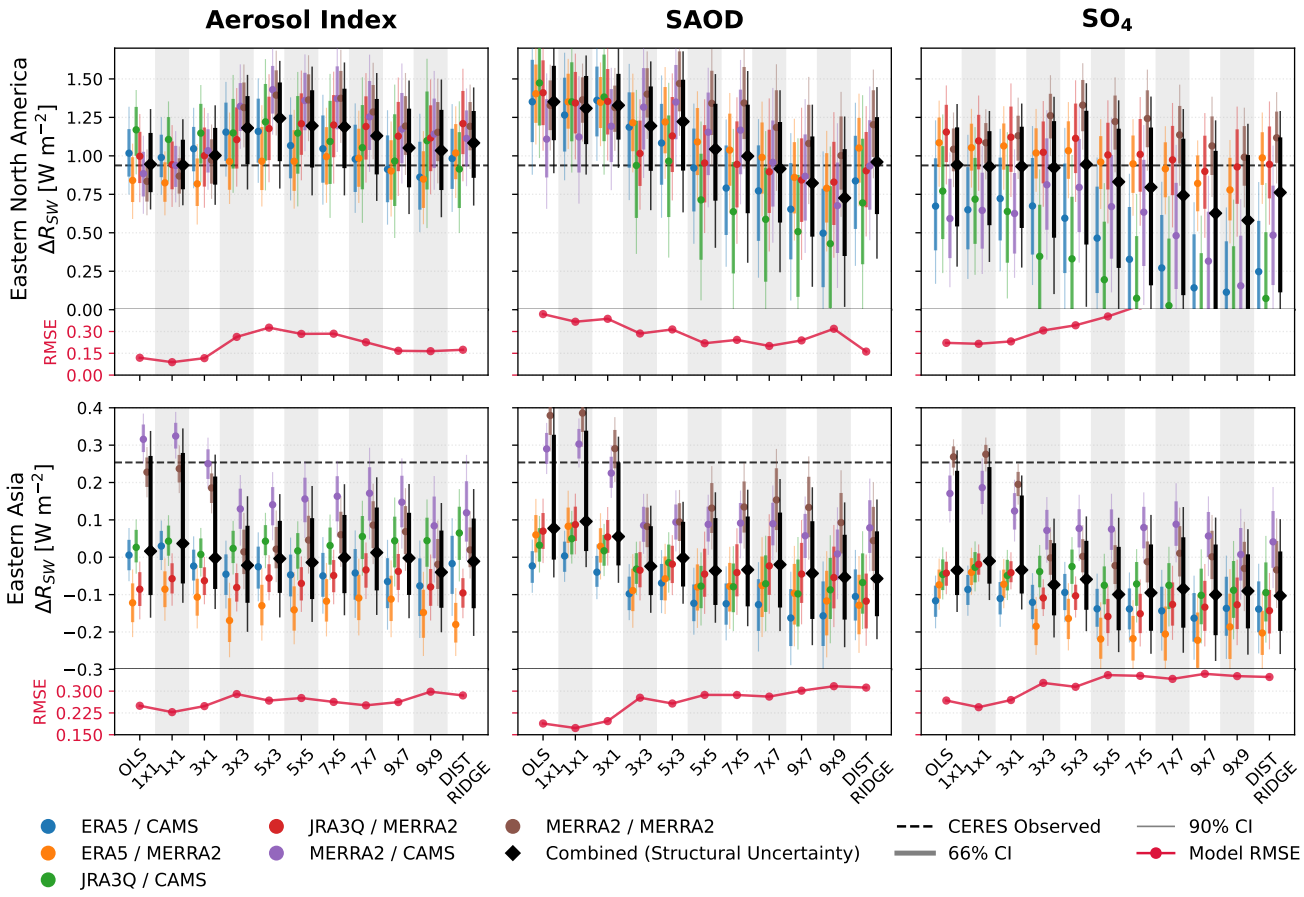

Figure S6. Full opportunistic experiment validation results for all tested regression configurations, aerosol proxies, and reanalysis combinations, extending Fig. 3 of the main text. Rows correspond to the two validation regions (Eastern North America, top; Eastern Asia, bottom) and columns to the three aerosol proxies (AI, SAOD, SO_4). Within each panel, the x-axis organizes configurations by regression method and domain size, progressing from local OLS through increasing Ridge domain sizes up to the global DIST-RIDGE, and the y-axis shows the predicted decadal change in shortwave low-cloud radiative effect (ΔR_{SW}, W m^{-2}), defined as the 2012–2019 average minus the 2003–2010 average. Colored vertical lines show the 66% (thick) and 90% (thin) confidence intervals derived from a stationary bootstrap (Sect. S2 in the Supplement) for each reanalysis combination (colors as indicated in the legend). The black diamond at each configuration position represents the combined estimate across all reanalysis combinations, reflecting the full structural uncertainty. The dashed horizontal black line indicates the CERES-observed ΔR_{SW} ground truth. The red line in the lower sub-panel of each subplot shows the RMSE of the individual reanalysis-combination predictions relative to the CERES observation as a function of regression configuration.

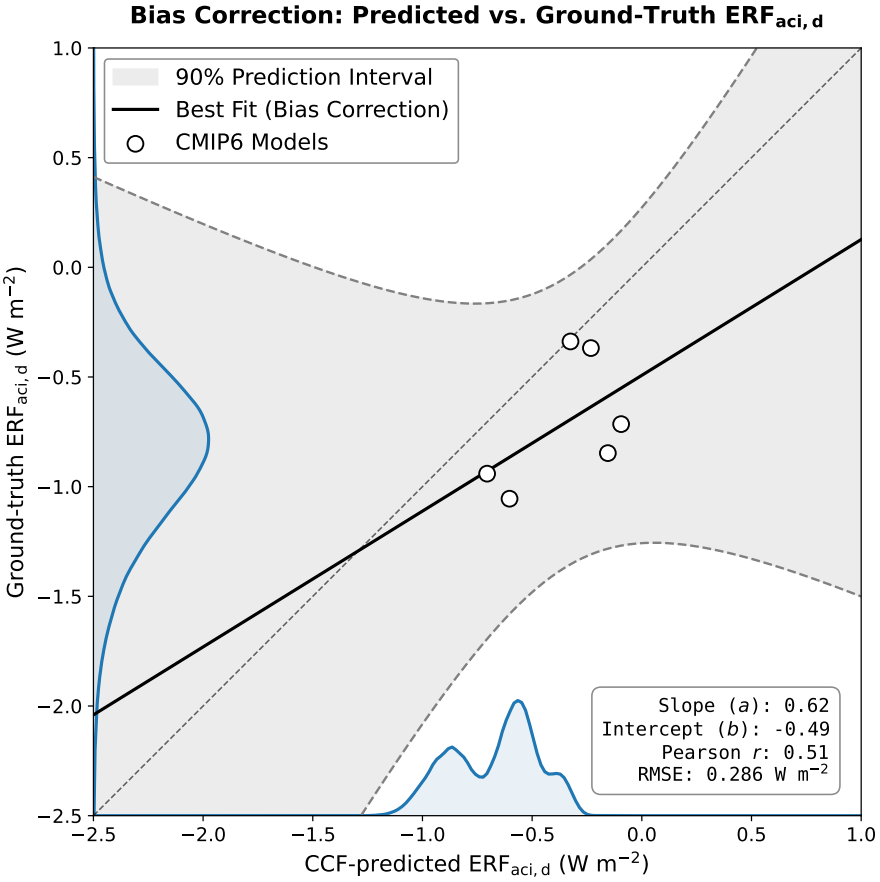

Figure S7. Emergent-constraint bias correction applied to the AI OLS 1×1 methodology over the combined ocean-and-land domain. Each open circle represents one CMIP6 model, plotting the uncalibrated domain-averaged $\mathrm{ERF}_{\mathrm{aci,d}}$ predicted by the CCF regression framework (x-axis) against the corresponding RFMIP aerosol-only ground-truth $\mathrm{ERF}_{\mathrm{aci,d}}$ (y-axis). The solid black line shows the ordinary least-squares best fit used to map raw predictions onto the true forcing; its slope (a) and intercept (b) define the bias correction applied in the Monte Carlo uncertainty propagation (Sect. 3.2). The dashed gray lines and shaded region indicate the 90% prediction interval, which propagates the calibration uncertainty into the final $\mathrm{ERF}_{\mathrm{aci,g}}$ estimate. The bottom blue distribution shows the probability density of the one-million uncalibrated global $\mathrm{ERF}_{\mathrm{aci,g}}$ samples (without the bias correction, using stationary bootstrap instead for uncertainty as written in Sect. S2 in the Supplement), while the left blue distribution shows the samples with applying the linear bias correction (the main text method). The calibration reveals that the raw CCF method systematically underestimates the magnitude of aerosol cooling, this confirms that the CCF regression capture residual meteorological noise, which masks the full aerosol signal. For the observations predicted $\mathrm{ERF}_{\mathrm{aci}}$, the calibration shifts the central estimate from an uncalibrated -0.67 W m^{-2} to a bias-corrected -0.84 W m^{-2}.

S4 Additional Tables

Table S1. Summary of CMIP6 models utilized in this study and their specific applications. Models used for the domain-to-global scaling factor are indicated in the γ column. The availability of each aerosol proxy (AI, SAOD, SO_4) is marked with an "x" for both the pre-industrial to present-day change (Δ Aerosol) and the regression framework evaluation (CMIP Test). The number of ensemble members used for the analysis is also provided. CMIP5 models were also evaluated, but none provided the complete set of variables required for this analysis.

	Scaling	**Δ Aerosol (PD$-$PI)**			**CMIP Regression Test**			
Model Name	γ	**AI**	**SAOD**	**SO_4**	**AI**	**SAOD**	**SO_4**	**Members**
ACCESS-CM2	x							1
ACCESS-ESM1-5	x							1
BCC-ESM1	x		x	x				1
CanESM5	x							1
CESM2	x		x	x		x	x	1
CESM2-FV2			x	x				1
CESM2-WACCM			x	x				1
CESM2-WACCM-FV2			x	x				1
CNRM-CM6-1	x		x			x		3
CNRM-ESM2-1	x	x	x	x	x	x	x	1
EC-Earth3	x							1
EC-Earth3-AerChem		x	x	x				1
GFDL-CM4	x			x			x	1
GFDL-ESM4	x	x	x					1
GISS-E2-1-G	x		x	x				1
GISS-E2-1-H			x	x				1
GISS-E2-2-G			x					1
HadGEM3-GC31-LL	x	x		x	x		x	5
INM-CM4-8			x					1
INM-CM5-0			x					1
IPSL-CM5A2-INCA			x	x				1
IPSL-CM6A-LR	x	x	x		x	x		11
IPSL-CM6A-LR-INCA	x	x	x	x	x	x	x	1
KIOST-ESM				x				1
MIROC-ES2H			x					1
MIROC-ES2L			x	x				1
MIROC6	x		x	x		x	x	1
MPI-ESM-1-2-HAM	x	x	x	x				1
MRI-ESM2-0	x	x	x	x	x	x	x	1
NorESM2-LM	x	x	x	x				1
NorESM2-MM	x	x	x	x				1
UKESM1-0-LL	x	x		x	x		x	14

References

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., and Bladé, I.: The Effective Number of Spatial Degrees of Freedom of a Time-Varying Field, Journal of Climate, 12, 1990 – 2009, https://doi.org/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2, 1999.

Park, C., Soden, B. J., Kramer, R. J., L'Ecuyer, T. S., and He, H.: Observational constraints suggest a smaller effective radiative forcing from aerosol–cloud interactions, Atmospheric Chemistry and Physics, 25, 7299–7313, https://doi.org/10.5194/acp-25-7299-2025, 2025.

Politis, D. N. and Romano, J. P.: The Stationary Bootstrap, Journal of the American Statistical Association, 89, 1303–1313, https://doi.org/10.1080/01621459.1994.10476870, 1994.

Wall, C. J., Norris, J. R., Possner, A., McCoy, D. T., McCoy, I. L., and Lutsko, N. J.: Assessing effective radiative forcing from aerosol–cloud interactions over the global ocean, Proceedings of the National Academy of Sciences, 119, e2210481 119, https://doi.org/10.1073/pnas.2210481119, 2022.