Dear Omer Roi-Cohen and co-authors,

Thank you for this interesting and timely paper. A systematic comparison of the different cloud-controlling-factor (CCF) approaches used to estimate ERFaci is valuable and much needed. I was particularly interested in the perfect-model evaluation. In Park et al. (2025), accounting for the aerosol activation rate reduced the RMSE relative to estimates that did not account for activation. I was surprised that your analysis appears to reach the opposite conclusion. I have several questions and additional analyses that may help clarify the source of this difference.

1) I understand that your ground-truth ERFaci estimates are derived from Smith et al. (2020). Did you separate the contributions from low and non-low clouds when calculating the ground-truth ERFaci? In Park et al. (2025), we separated these contributions using the cloud classification of Webb et al. (2006). This appears important because the CCF framework assumes that aerosol–cloud interactions primarily affect low clouds, with negligible contributions from non-low clouds. Therefore, consistently restricting both the CCF-based estimate and the model-derived ground truth to low-cloud contributions may be necessary for a direct comparison.

2) Have you compared the performance of the activation-rate methods separately over ocean and over the combined land+ocean domain? In the manuscript, I found this comparison only for the three best-performing methods—OLS 1x1, RIDGE 3x3, and DIST-RIDGE—but these appear to be methods in which the activation rate is not explicitly considered.

I understand that using the combined land-and-ocean domain may reduce the uncertainty in the ground-truth ERFaci estimate. However, the CCF frameworks considered here were primarily developed for marine low clouds. Cloud controlling factors and aerosol–cloud relationships over land may differ because continental clouds are influenced by meteorological controls that are not fully represented in the current CCF analysis. For this reason, all three CCF studies for ERFaci—Wall et al. (2022), Park et al. (2025), and Zelinka et al. (2026)—as well as the original CCF study of Scott et al. (2020), focused on ocean regions and excluded land. Therefore, comparing methods with and without activation over the combined land+ocean domain may not provide a proper evaluation of the intended marine low-cloud CCF framework. An ocean-only comparison would help determine whether the contrasting result relative to Park et al. (2025) is primarily related to the evaluation domain.

3) I was unsure about the meaning of labels such as OLS 1x1, RIDGE 3x3, 7x5 and 9x7 configurations. Does "OLS 1x1" mean that only the local-grid-cell CCFs are used as predictors, while the underlying data remain at the original 5° x 5° resolution? Or, does it indicate that the 5° x 5° results were regridded to a 1° x 1° spatial resolution? Similarly, for the other configurations, does 7x5 describe the number of neighboring grid cells included in the regression—for example, seven grid cells in latitude, potentially covering 35° meridionally, and five grid cells in longitude, potentially covering 25° zonally—or does it describe the spatial resolution of the analysis? A clearer explanation of this nomenclature would be helpful.

Perfect-model comparison

Below, I present an attempt to reproduce the perfect-model comparison as closely as possible following your approach. Unlike Park et al. (2025), in which the model CCF-predicted ERFaci was estimated by scaling the observationally derived CCF relationship, I performed a model-specific CCF analysis using each model's own aerosol, cloud, and meteorological variables.

I used the 1995–2014 historical simulations to derive the susceptibility, consistent with the period used in your study. The CCF relationships were estimated using ordinary least squares (OLS) regression. Changes in aerosol concentration, represented by either aerosol index (AI) or sulfate mass concentration at 925 hPa (SO_4), were calculated as the difference between the preindustrial period (1850–1859) and the present-day period (2005–2014). All variables were spatially averaged to a 5° x 5° resolution before performing the regression analysis. At present, I have analyzed only the first ensemble member (r1) from each model. Therefore, some of the models included in your study are not represented in the comparison below. The definition of ERFaci_true is similar to that of the ground-truth ERFaci used in your study. However, I separated the cloud radiative contribution using radiative kernels and isolated the low-cloud contribution using the Webb et al. (2006) cloud classification.

Figure 1 illustrates how the choice of evaluation domain can affect the conclusions, particularly for AI. The upper panels show ERFaci estimates averaged over land and ocean between 55°S and 55°N, whereas the lower panels show the corresponding results over ocean only. Because of the limited availability of ISCCP cloud fraction output, the analysis includes only six models for both AI and SO_4.

For AI over the combined land+ocean domain, including the activation rate results in a slightly larger RMSE than excluding it (0.72 vs 0.68). Although the difference is smaller than that reported in your study, the qualitative conclusion is similar. However, when the analysis is restricted to ocean regions, the conclusion changes: accounting for activation produces a slightly lower RMSE than neglecting it (0.67 vs 0.68). For SO_4, accounting for the activation rate consistently produces a lower RMSE over both the combined land+ocean domain and the ocean-only domain. Moreover, restricting the analysis to ocean regions leads to a relatively large reduction in RMSE for the activation-rate case, from 0.58 to 0.50. In contrast, the RMSE for estimates without activation remains nearly unchanged when land is excluded, increasing slightly from 0.89 to 0.91. These results suggest that the inferred effect of accounting for aerosol activation may depend on both the aerosol proxy and the evaluation domain. The differences between our results may also arise from the definition of the ground-truth ERFaci and the limited number of available models. With such a small sample, modest differences in the set of models included in the analysis could affect the resulting RMSE and the overall conclusion.

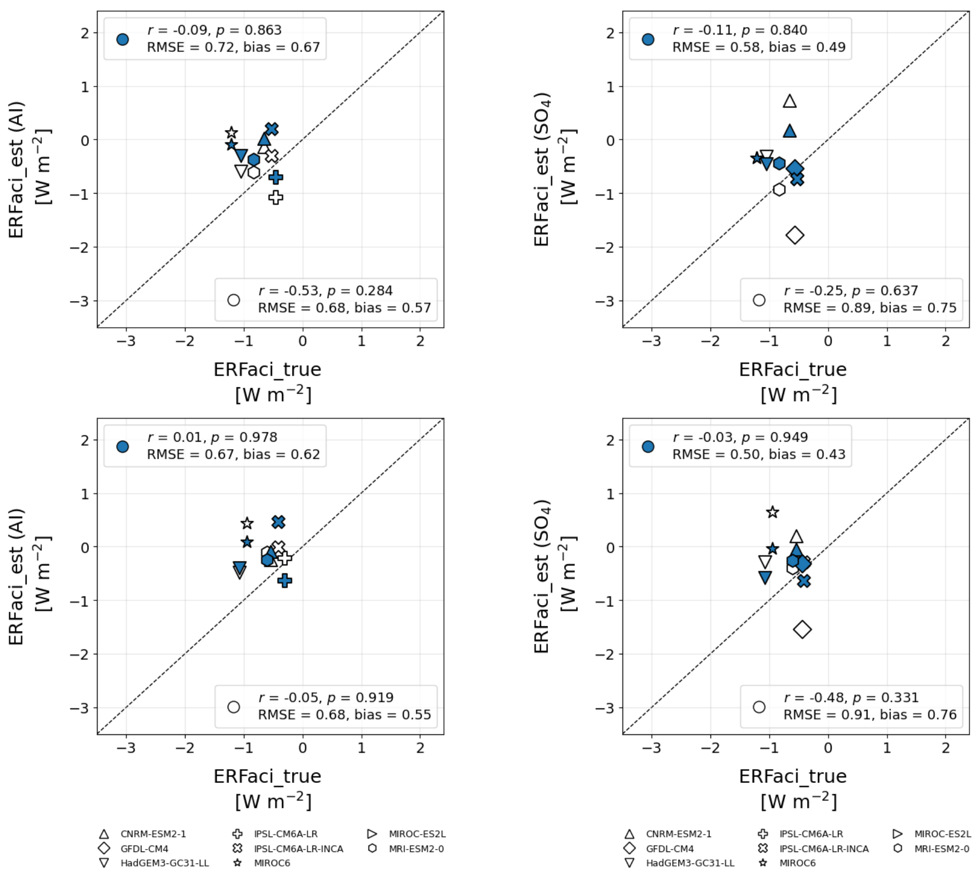

Figure 1. Perfect-model cross-validation of ERFaci estimates averaged over 55°S–55°N for land and ocean combined (upper panels) and ocean-only (lower panels). The left panels show the ground-truth ERFaci (ERFaci_true) versus ERFaci estimated using a cloud controlling factor analysis with aerosol index (AI) as the aerosol proxy. The right panels show the corresponding results using sulfate mass concentration at 925 hPa (SO_4). Filled blue markers represent estimates that account for the aerosol activation rate, whereas open hollow markers represent estimates that

do not account for activation. The correlation coefficient (r), associated p-value, root mean square error (RMSE), and bias are shown in either the upper-left or lower-right corner of each panel. Bias is defined as the mean absolute deviation from the 1:1 line, shown by the dashed line.

To increase the number of models available for the perfect-model comparison, I examined whether ERFaci_est derived from cloud radiative kernel (CRK) estimates of non-obscured low-cloud radiative anomalies is consistent with ERFaci_est derived using the radiative kernel combined with the Webb et al. (2006) cloud classification. As shown in Fig. 2, the two approaches produce generally consistent ERFaci_est values for most models. The main exceptions are IPSL-CM6A-LR and IPSL-CM6A-LR-INCA, which show substantial deviations from the relationship found for the other models. Given the overall agreement, I used the radiative-kernel-derived low-cloud radiative anomalies to expand the sample to 10 models for AI and 12 models for SO_4, while excluding the two IPSL models. This provides a larger model sample while maintaining consistency with the CRK-derived ERFaci estimates.

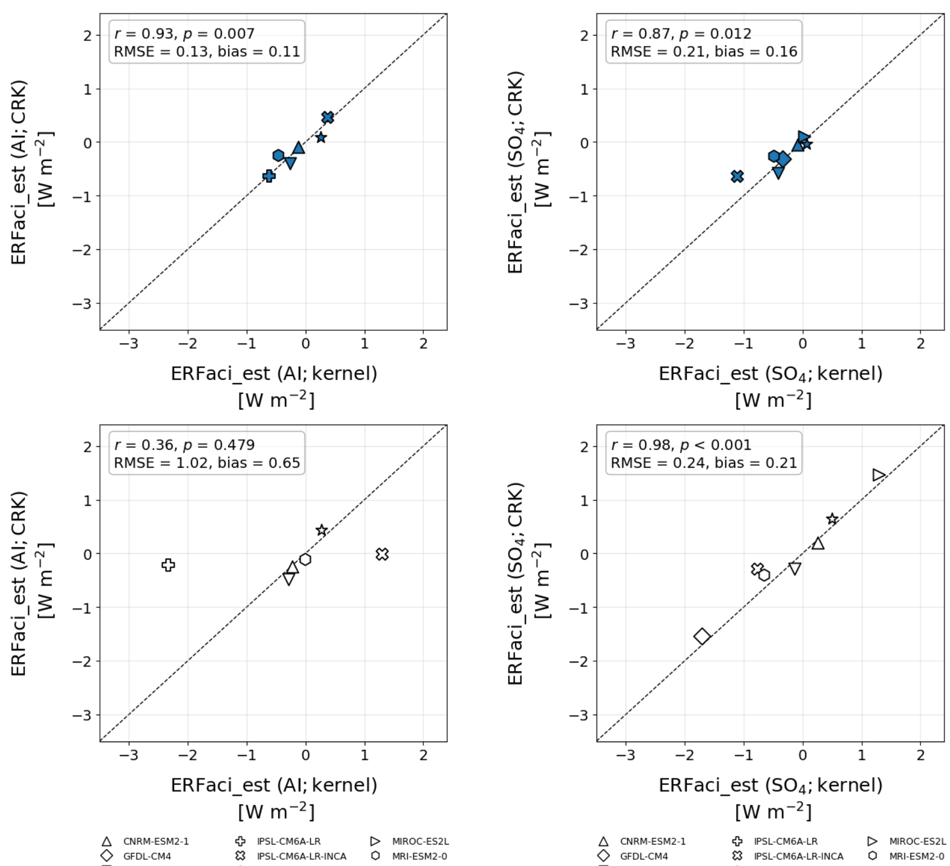

Figure 2. Comparison of ERFaci_est calculated using two different definitions of low-cloud radiative anomalies. The y-axis shows estimates based on a cloud radiative kernel (CRK) applied

to non-obscured low clouds, whereas the x-axis shows estimates based on the radiative kernel combined with the Webb et al. (2006) cloud classification to isolate low-cloud radiative anomalies. All ERFaci_est values are averaged over the ocean-only domain from 55°S to 55°N. The upper row shows estimates that account for the aerosol activation rate, while the lower row shows estimates without activation. The left column uses aerosol index (AI) as the aerosol proxy, whereas the right column uses sulfate mass concentration at 925 hPa (SO_4).

With the expanded model sample, we again find that accounting for the aerosol activation rate improves the perfect-model comparison. In Fig. 1, the AI-based analysis over the ocean-only domain showed little difference in RMSE between estimates that included activation and those that omitted it. However, after increasing the number of models in Fig. 3, the activation-included estimates clearly yield a lower RMSE. For SO_4, the results are more robust across both the original and expanded model samples: estimates that account for activation consistently agree better with the ground-truth ERFaci than estimates that neglect activation. Overall, the expanded analysis further supports the conclusion of Park et al. (2025) that explicitly accounting for aerosol activation improves CCF-based ERFaci estimates in perfect-model comparisons.

Although Park et al. (2025) used a simplified approach to estimate ERFaci_est, the present analysis further confirms that accounting for aerosol activation improves its relationship with ERFaci_true in aerosol-only forcing experiments. Because your study reaches the opposite conclusion based on a land-inclusive evaluation domain, a relatively small subset of models, and a ground-truth ERFaci estimate that does not isolate the low-cloud contribution, I believe the manuscript would benefit from a discussion of the possible reasons for this discrepancy.

Thank you again for this valuable study. I would appreciate clarification of these points during the revision.

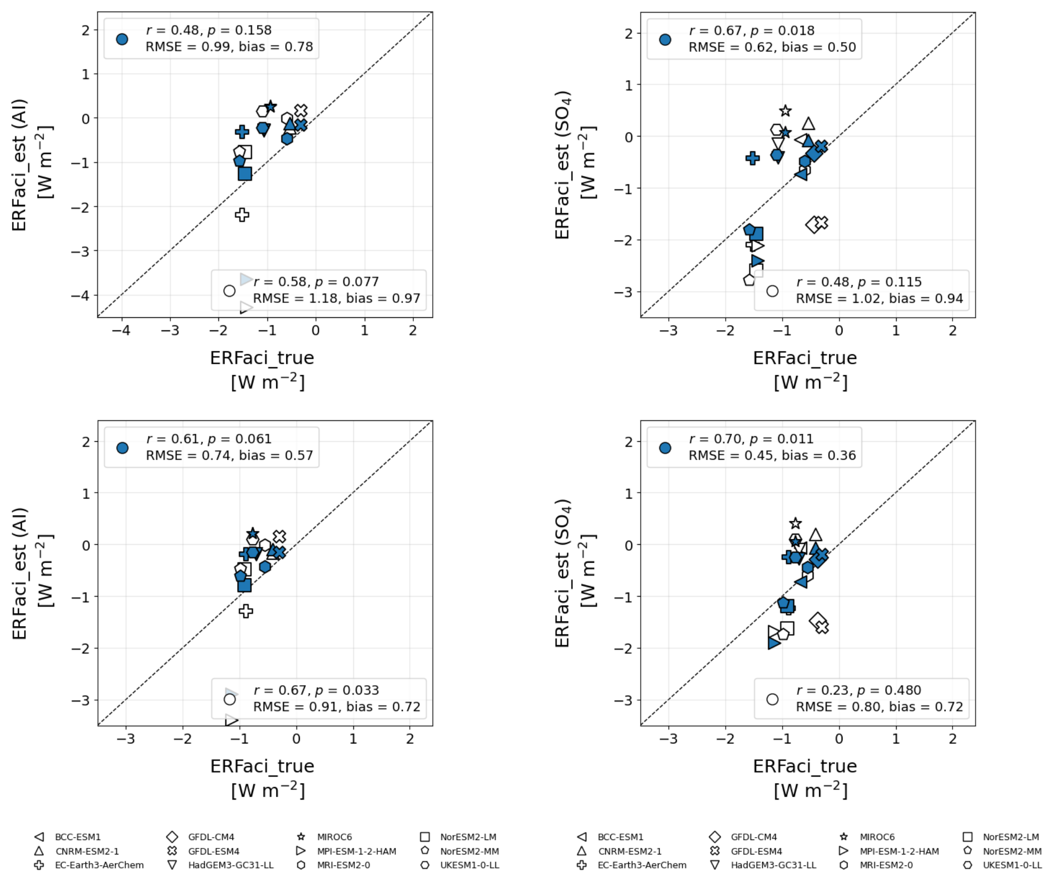

Figure 3. Same as Fig. 1, but using ERFaci_est derived from the radiative-kernel plus cloud-classification method in order to expand the number of models available for the perfect-model comparison. The upper panels show ERFaci estimates averaged over the ocean-only domain from 55°S–55°N. The lower panels show global ERFaci estimates obtained by scaling the 55°S–55°N ocean-only estimates by the ratio of global ERFaci_true to ocean-only 55°S–55°N ERFaci_true.

References

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.

Scott, R. C., Myers, T. A., Norris, J. R., Zelinka, M. D., Klein, S. A., Sun, M., and Doelling, D. R.: Observed Sensitivity of Low-Cloud Radiative Effects to Meteorological Perturbations over the Global Oceans, Journal of Climate, 33, 7717–7734, https://doi.org/10.1175/JCLI-D-19-1028.1, 2020.

Smith, C. J., Kramer, R. J., Myhre, G., Alterskjær, K., Collins, W., Sima, A., Boucher, O., Dufresne, J.-L., Nabat, P., Michou, M., Yukimoto, S., Cole, J., Paynter, D., Shiogama, H., O'Connor, F. M., Robertson, E., Wiltshire, A., Andrews, T., Hannay, C., Miller, R., Nazarenko, L., Kirkevåg, A., Olivié, D., Fiedler, S., Lewinschal, A., Mackallah, C., Dix, M., Pincus, R., and Forster, P. M.: Effective radiative forcing and adjustments in CMIP6 models, Atmospheric Chemistry and Physics, 20, 9591–9618, https://doi.org/10.5194/acp-20-9591-2020, 2020.

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.

Webb, M. J., Senior, C. A., Sexton, D. M. H., Ingram, W. J., Williams, K. D., Ringer, M. A., McAvaney, B. J., Colman, R., Soden, B. J., Gudgel, R., Knutson, T., Emori, S., Ogura, T., Tsushima, Y., Andronova, N., Li, B., Musat, I., Bony, S., and Taylor, K. E.: On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles, Clim Dyn, 27, 17–38, https://doi.org/10.1007/s00382-006-0111-2, 2006.

Zelinka, M. D., Myers, T. A., Qin, Y., Chao, L.-W., Klein, S. A., Po-Chedley, S., Ma, P.-L., Wall, C. J., Ceppi, P., and Gettelman, A.: Recent cloud trends and extremes reaffirm established bounds on cloud feedback and aerosol-cloud interactions, Communications Earth & Environment, https://doi.org/10.1038/s43247-026-03461-8, 2026.