Response to reviewers:

We thank the reviewers for their very positive feedback and constructive suggestions. We have revised the manuscript accordingly. Please find below all reviewer comments in plain black, our responses in blue, and the revised text in dark blue.

Reviewer #1

Review response to Jia et al.: “Process-level contributions to uncertainty in aerosol effective radiative forcing: a perturbed parameter ensemble with aerosol-climate model ICON-HAM”

In their manuscript, Jia et al. describe a novel PPE developed with the ICON-HAM model framework designed to analyze both aerosol-cloud interactions as well as aerosol direct effects. Through Gaussian Process emulation, they expand their sample space allowing for robust parametric uncertainty analysis compared against both PPEs and model intercomparison projects, as well as select observational datasets. While not directly constraining the PPE with observations, their comparisons highlight parameters that may benefit from constraint, and the authors describe a natural next step in their work as constraining their ensemble to observations. In addition to the parametric uncertainty analysis, they present a novel experimental design to determine an optimal ensemble member-to-parameter ratio for the design of PPEs. This is a highly beneficial addition to the PPE community and provides much needed guidance in terms of PPE experimental design.

This manuscript reads very well and presents impactful and novel analyses to benefit the PPE community. Some minor clarifications are needed, but overall the analysis is thorough and clear. The scientific significance, quality, and presentation of the material are all excellent. Below are minor comments and suggestions. Overall, I recommend acceptance after minor revisions.

We thank the reviewer for the encouraging comments and thoughtful suggestions. We have addressed all points and revised the manuscript to improve clarity as suggested.

Major comments/questions:

Regarding the rsp analysis: (line 276) This is a very exciting finding and makes the prospect of generating PPEs easier to accomplish when limited by computational resources. Do the authors think that using a single model to sample across this rsp range introduces biases inherent in ICON-

HAM? Would they expect these results to be consistent no matter the PPE/base-model?

The reviewer raises an excellent point. If different models are perturbed within the same parameter space (i.e., using the same parameters and perturbation ranges), such that no additional non-linearity from parameter interactions is introduced, we would not expect the optimal rsp to vary substantially across models. Interestingly, despite differences in parameter selection and sampling, the CAM6 PPE with rsp~6 still shows a similar range of R2 values for LWP (0.69–0.90) to that obtained from our ICON-HAM PPE at the same rsp (0.70–0.83; Fig. 4).

Nevertheless, it is difficult to draw a definitive conclusion at the moment. We believe that this will be an interesting question to explore using the upcoming coordinated MMPPE dataset (https://github.com/aerocom-community/experiments/tree/main/phase-4/MMPPE).

We have added this discussion on Lines 778–780:

'For PPEs sampling comparable parameter spaces, and thus subject to similar levels of non-linearity arising from parameter interactions, we would not expect the near-optimal rsp to vary substantially across models. Nevertheless, this assumption will be interesting to test with the upcoming MMPPE.'

Fig. 11 and Fig. 12: With your emulator output, would it be possible to extract the fractional parametric uncertainty within the green isolines and include as additional bar plots following these figures? I think this would make the message much clearer for the reader than having to reference patterns in the supplementary figures that are quite numerous and somewhat challenging to track. This may complicate the regional estimates, but potentially your regional sectors in Fig. A1 could be applied and masked by the green sectors which may result in some empty slots due to sparse green isolines. This visually could be important information for understanding what observational products might be regionally important for constraint. I'd think these bar plots could be interpreted as an observationally informed parameter set that is most in need of observational constraint.

We thank the reviewer for this very helpful suggestion. As suggested, we have added bar plots similar to Fig. 8 and 9, but considering only highly uncertain locations within the green isolines (Fig. S14 & S18). The dominant parameters identified by these regional bar plots are overall consistent with those inferred from the spatial correspondence between model biases and parametric uncertainty.

Some subregional differences, however, are obscured by regional averaging. For example, over Africa, uncertainty is dominated by dust emissions in the north and OC emissions from biomass burning over central Africa, while the regional-mean bar plot shows the predominant

role of dust due to the larger AOD over Sahara. The spatial-pattern discussion is therefore still valuable, which we keep, while adding the new bar plots to the SI to provide a more direct overview of which parameter constraints may be most effective in reducing model biases in each region. We have added the following discussion on Lines 725-729:

'To provide a more direct overview of the parameters driving uncertainty in regions with both large biases and high parametric uncertainty, we repeat the analyses of Fig. 8 considering only locations within the green isolines (Fig. S14). It is shown that the dominant parameters are overall consistent with those inferred from the spatial patterns discussed above, although regional averaging can obscure subregional signals, such as the role of emi_bb_oc over central Africa. These results can inform which parameters may be most useful to constrain observationally for reducing model biases in each region.'

And on Lines 749-751:

'For the locations within the green isolines, the parameter contributions to LWP and CDNC to uncertainties in each region (Fig. S18) are broadly consistent with the spatial correspondence discussed above.'

Specific in-line comments:

Line 110-111: Consider changing to "...from those obtained using **a** satellite simulator...". I was a little confused as 'the simulator' makes it sound like the simulator had already been mentioned, but this is the first mention of it in the text. Reading on made it clear you were referencing the COSP simulator for comparisons.

Corrected. Thanks for catching this!

Line 199-200: I very much like this idea for testing the rsp. I see that you show the number of members per subset in Fig. 1. Can you please also reference the number of members corresponding to the different rsp in the text for consistency? Something like, "...corresponding to rsp ranging from 2 (n=84 members) to 8 (n=363 members)."

This is really a good idea. As suggested, we have revised the text to include the corresponding number of members for each rsp:

'For each partition, the training dataset is further resampled to generate subsets corresponding to rsp ranging from 2 (n = 84 members) to 8 (n = 336 members).'

Line 286-287: It is unclear to me how you fix the other parameter group from what I have read so far. I assume this is derived from your emulator output whereby you have control over the parameter sets and their ranges over which to emulate, but I'm not sure. Can you add a sentence elaborating on how you isolated the different parameter groups? A reference to this in the methods or clarification in the methods would also be helpful.

Yes, the reviewer is correct! We isolate the two parameter groups using the emulator, varying one group over its defined uncertainty ranges while fixing the other group at its prior medians. This was mentioned in the caption of Fig. 5, but we have now explicitly clarified this in the text on Lines 289-290:

'To isolate the contributions from the aerosol and cloud parameter groups, we use the emulator to vary the parameters in one group over their uncertainty ranges while fixing all parameters in the other group at their prior medians (Fig. 5).'

Line 350-351: Can you support this claim with a figure from the PPE/emulator or with a previous literature reference?

Thanks for the comment. Our figures 10 and S19 show that anthropogenic SO2 and BC emission parameters affect ERFari to a similar extent but in opposite directions (with the sensitivities of −2.6 and +2.4 over land, respectively; Fig. S19), reflecting the opposing radiative effects of scattering sulfate and absorbing BC aerosols. As China is a major source region of both SO2 and BC emissions (Wu et al., 2024), their radiative effects can partly compensate each other. We have clarified this interpretation and added the reference on Lines 352-355:

'The weak ERFari over China likely results from the partial cancellation between warming from absorbing BC aerosols and cooling from scattering sulfate aerosols, both of which have substantial anthropogenic emissions in this region (Wu et al., 2024). This is further supported by the sensitivity analysis presented later (Fig. S19), where ERFari shows comparable but opposite sensitivities to their emission parameters over land (−2.6 to emi_antns_so2 and +2.4 to emi_ant_bc).'

Wu, N., Geng, G., Xu, R., Liu, S., Liu, X., Shi, Q., Zhou, Y., Zhao, Y., Liu, H., Song, Y., Zheng, J., Zhang, Q., and He, K.: Development of a high-resolution integrated emission inventory of air pollutants for China, Earth Syst. Sci. Data, 16, 2893–2915, https://doi.org/10.5194/essd-16-2893-2024, 2024.

Line 384-386: Can you clarify where you are getting this these percent values? I would assume that holding emissions fixed means ERFaci (Fig. 7b), but I don't see as strong a contribution from the referenced variables as mentioned here. Please reference the table or figure used to retrieve these values and elaborate in the explanation.

Thanks for the comment. This percentage values refer to the contributions of cloud parameters to the remaining uncertainty after parameters related to anthropogenic, biomass-burning, and DMS emissions are fixed at their prior medians. These percentage contributions are shown below (Fig. R1) and are now also included in the SI as Fig. S10. We have clarified this on Lines 392-394:

'Once anthropogenic, biomass burning, and DMS emissions are fixed at their prior medians, cloud parameters dominate the remaining uncertainty (76%; Fig. S10), primarily through convection (conv_cprcon, conv_entrpen, and conv_cmfctop), cloud cover (cld_cov_crs), and auto-conversion (micro_ccsaut).'

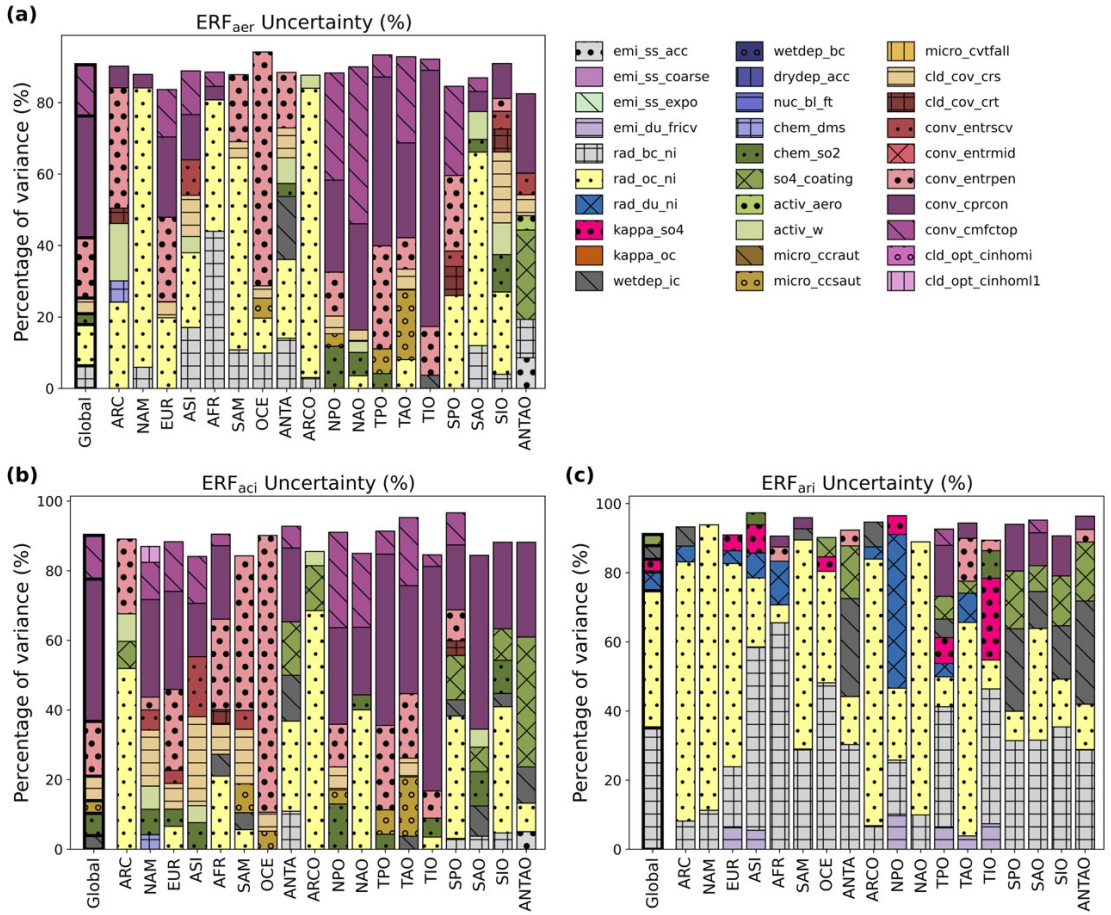


Figure R1. Same as Fig. 7 but with parameters related to anthropogenic, biomass burning, and DMS emissions fixed at their prior medians. Parameters contributing less than 3 % of the total uncertainty are not shown.

Line 434: I wonder if the contributions of biomass burning and anthropogenic emissions (along with their radiative properties) to ARC, NAM, EUR, and ASI may be related to the ERF sensitivity to higher pre-industrial emissions over these regions. Given that these uncertainties are not reflected in the present day (Figs. 8,9) it may be worth mentioning the sensitivities of these results on PI emissions and choice of PI period (e.g.,

10.1038/s41467-018-05592-9; 10.1029/2025GL121443). Namely, the role that a higher carbonaceous aerosol influenced PI may have on a PI to PD radiative forcing change. It may also be possible that if your PI emissions are incorrect, constraining to PD observations may not actually improve your ERFaci.

Thank you for this important point. We agree that uncertainties in the prescribed PI emission could affect both the PI-to-PD aerosol forcing and the parameter uncertainty contributions, and that these uncertainties cannot be constrained by PD observations alone. We have added the following discussion to clarify this limitation on Lines 442–451:

‘In addition to the parametric uncertainty, another source of the ERFaer uncertainty arises from the prescribed PI emission baseline. In existing PPEs, the spatial pattern and reference period of PI emissions are fixed, while the emission scaling parameters are perturbed in the same way as in the PD runs. Different assumptions about PI biomass-burning and the choice of PI period could substantially alter the PI aerosol background and consequently the magnitude and regional pattern of PI-to-PD aerosol forcing (Hamilton et al., 2018; Brown et al., 2026). The PI baseline may further affect the relative contributions of individual parameters presented here. For example, an emission perturbation may exert a stronger effect under a cleaner PI background than under a more polluted background because of the nonlinear response of clouds and radiation to aerosol perturbations (Jia and Quaas, 2023). Although scaling natural aerosol emissions provides some means of constraining the PI aerosol background, uncertainties associated with the prescribed PI emissions are not captured by the current PPE design and cannot be constrained by PD observations alone.’

Brown, H. Y., McCoy, D. T., Nugent, J. M., Wang, H., Wu, M., Zhang, K., and Burrows, S. M.: Sensitivity of Anthropogenic Cloud Droplet Number Change to Preindustrial Emission Inventories and Physics Parameterizations, Geophy. Res. Lett., 53,e2025GL121 443, https://doi.org/https://doi.org/10.1029/2025GL121443, 2026.

Hamilton, D. S., Hantson, S., Scott, C. E., Kaplan, J. O., Pringle, K. J., Nieradzik, L. P., Rap, A., Folberth, G. A., Spracklen, D. V., and Carslaw, K. S.: Reassessment of pre-industrial fire emissions strongly affects anthropogenic aerosol forcing, Nat. Commun., 9, 3182, https://doi.org/10.1038/s41467-018-05592-9, 2018.

Jia, H. and Quaas, J.: Nonlinearity of the cloud response postpones climate penalty of mitigating air pollution in polluted regions, Nat. Clim. Change, 13, 943–950, 2023.

Line 450: In the bar plots this is referenced as ASI but in Figure A1 it is referenced as ASIA. Please change within text or plots to be consistent throughout.

Thanks! ‘ASIA’ has now been changed to ‘ASI’ throughout the text and figures

Line 578: This seems to be the first time mentioning ‘AI’, please define the acronym.

Thanks for the reminder. ‘AI’ is now defined as aerosol index here.

Line 599: This was briefly mentioned above in the context of past work conducting observational constraints. Could you please define in more detail here? Based on Regayre et al. 2023, something like: "(AI; the total MODIS aerosol optical depth at 550 nm multiplied by the Ångström exponent)". Also, why is this diagnostic important in addition to the AE?

Thanks for the suggestion. We have now explicitly defined AI as AOD multiplied by AE as suggested. AE only reflects the information of particle size, whereas AI additionally incorporates aerosol loading through AOD, making it a good proxy for fine-mode aerosol abundance (Nakajima et al., 2001) and thus more closely related to SO2 emissions, which are a major source of fine-mode particles. We have clarified this on Lines 615-619:

‘Three PD diagnostics related to fine-mode aerosol concentrations also show strong sensitivities to emi_antns_so2, including AI (AOD multiplied by AE) and CCN number concentrations at the surface (CCN_S0.3) and at the top of boundary layer (CCN_BL0.3). Compared with AOD or AE alone, AI better represents fine-mode aerosol abundance by combining aerosol loading and size information (Nakajima et al., 2001), and is thus more closely related to SO2 emissions.’

Nakajima, T., Higurashi, A., Kawamoto, K., and Penner, J. E.: A possible correlation between satellite-derived cloud and aerosol microphysical parameters, Geophys. Res. Lett., 28, 1171–1174, https://doi.org/10.1029/2000GL012186, 2001.

Line 678-679: Please elaborate on this. Is this in reference to supplementary figures? If so, please reference those figures.

If we understand the reviewer’s question correctly, here we refer to the spatial overlap between the model biases and overall parametric uncertainty (green isolines), both shown in Fig. 11f. We did not intend to refer to the contributions from individual parameters shown in the supplementary figures.

Line 689-690: It is unclear to me what you are trying to say here. Are you saying that rad_oc_ni reflects wavelength dependence of OC, or are you saying that AE reflects wavelength dependence of OC? I'm also not sure how this relates to the previous sentence. Please reword or expand into two sentences to walk the reader through your reasoning.

Thanks for pointing this out. The previous sentence discusses parameters that contribute to AE uncertainty through changes in aerosol size, whereas here, rad_oc_ni provides a different pathway that is not related to particle size. We intended to say that AE's response to rad_oc_ni is caused by the wavelength-dependent response of aerosol extinction to changes in OC absorption. We have reworded the text to make this clearer on Line 708-710:

'Although rad_oc_ni also contributes to AE uncertainty over land, this does not reflect changes in particle size. Instead, changes in the OC absorption can alter aerosol extinction differently across wavelengths and thereby affect AE (Schuster et al., 2006).'

Line 772-775: Please check the bullets in this sentence as they are inconsistent.

Corrected.

Reviewer #2

In this manuscript, the authors run and emulate a large PPE for ICON-HAM, finding that aerosol-related parameters, rather than cloud parameters, dominate uncertainty in both ERFaci and ERFari. For global ERFaer, about a quarter of parametric uncertainty comes from DMS emissions, another quarter from anthropogenic sulfur dioxide emissions (excluding shipping), a third quarter from the size of emitted fossil fuel and biomass burning primary aerosol particles, and the remainder split amongst smaller contributions from the other 38 parameters. Useful results are obtained for future PPE development as well, with a recommended simulations-to-parameters ratio of around 5 nicely balancing accuracy with computational efficiency. The analysis appears overall sound and the paper is very well written — the manuscript is almost ready for publication as-is, although the authors may wish to consider the comments below in minor revisions.

We thank the reviewer for the positive assessment of our work and the constructive comments. We have considered all suggestions and revised the manuscript accordingly.

General comment: Are there any parameters accounting for uncertain VOC emissions and processing, etc., over land? This seems potentially relevant given the importance of DMS via the PI background.

This is a very good point. We agree that uncertainty in biogenic emissions may contribute to ERF uncertainty through their influence on the PI aerosol background, similar to DMS. In our current PPE setup, however, no VOC-related parameters were perturbed because full chemistry is not

activated in our model configuration. Indeed, exploring VOC-related uncertainties could be a very interesting next step for future PPEs. We have now discussed this in Lines 383–385:

‘In particular, the role of DMS highlights the importance of the PI aerosol background. Similarly, uncertainties in biogenic volatile organic compound (VOC) emissions and processing may affect the PI aerosol background, too, but are not explored here because full chemistry is not activated in the current model configuration. ’

Specific comments:

1. Line 2: It would be helpful to spell out ARI and ACI when they’re first introduced here.

   Done.

2. Line 6: The parametric uncertainties lead to the credible range but not the mean, right?

   Thanks for the suggestion. The sentence has now been reformulated as:

   ‘Parametric uncertainties in aerosol and cloud processes yield a 90% credible range of -1.42 to -0.65 Wm-2 for ERFaer (mean: -1.04 Wm-2) for the period 2024-2025. ’

3. Line 56: Repeated phrase.

   Removed.

4. Line 125: Horizontal wind nudging eliminates the possibility of ERFaer due to large-scale circulation changes… how much this may affect your results, particularly for biomass burning emissions/black carbon, should be discussed.

   Thank you for raising this interesting point. We agree that nudging may suppress the large-scale circulation response to aerosol effects. To assess this, we performed twin simulations with nudging applied either throughout the atmospheric column or only above 2 km, with the latter allowing the lower atmosphere to respond more freely. The comparison (Fig. R2; also included in the SI as Fig. S1) shows that the nudging extent has little effect on the global-mean ERFaer but, as the reviewer anticipated, causes slight regional differences in ERFari over biomass-burning regions due to strongly absorbing BC. We have now added the following discussion on Lines 128–131:

   ‘We also compared the aerosol forcing in twin simulations with full-column and above-2km nudging and found that the global-mean ERFaer is insensitive to nudging extent (Fig. S1), consistent with previous studies (Kooperman et al., 2012), while over regions with strong biomass-burning emissions, allowing the lower atmosphere

to respond to aerosol absorption leads to a slightly more positive ERFari. ’

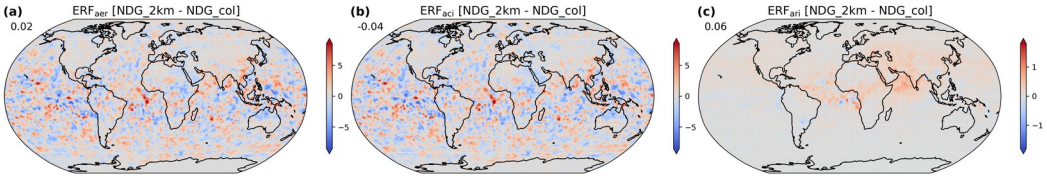


Figure R2. Differences in annual-mean (a) ERFaer, (b) ERFaci, and (c) ERF ari between the above-2km nudging (NDG_2km) and full-column nudging (NDG_col) experiments, based on single-model runs.

5. Line 200: I’m struggling to understand here how the n = 84, 126, … 336 is determined for the r_sp = 2, 3, … 8 cases, as illustrated in Fig. 1.

   Here, *n* denotes the number of simulations in each resampled PPE subset and is determined as *n*=42×rsp, where 42 is the number of perturbed parameters and rsp is the simulations-to-parameter ratio. Thus, rsp=2, 3, …, 8 corresponds to *n*=84, 126, …, 336, respectively (e.g., *n*=42x2=84 for rsp=2). We have revised Fig. 1 and clarified this in the figure caption.

6. Line 285: It seems like the emulator has a strong positive bias, which is an important caveat for the “uncertain magnitude and sign” for ERFari. It’d be useful to show ERFari and ERFaci separately in Figure 2, or perhaps in the SI if you don’t want to add more panels there.

   Thanks for the comment. Below, we show plots similar to Fig. 2 for ERFaer, ERFaci, and ERFari separately. We see that the positive emulator bias for strongly negative ERFaer arises primarily from ACI, whereas ARI shows good agreement between the prediction and truth. Therefore, the “uncertain magnitude and sign” for ERFari is unlikely to result from emulator bias. This component-wise comparison has now also been included in the SI (Fig. S4).

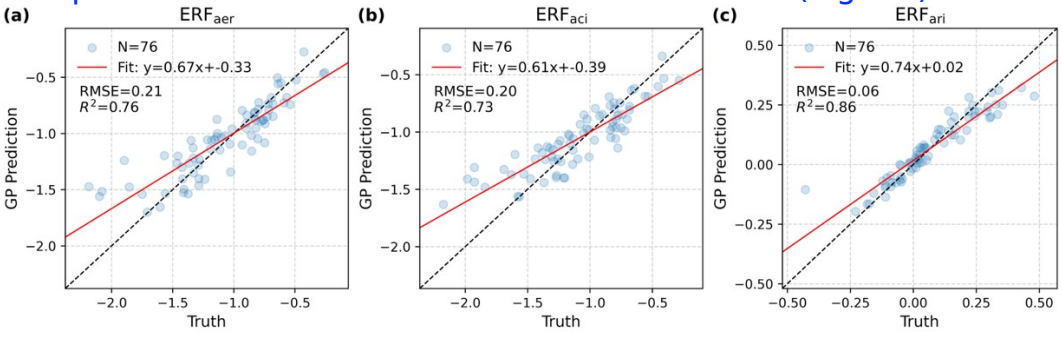


Figure R3. Comparison of globally averaged GP emulator predictions with test data (truth) for (a) ERFaer, (b) ERFaci, and (c) ERFari.

7. Lines 325-326: Do observational estimates necessarily increase the total uncertainty in the IPCC compared to MMEs?

The reviewer raises a good point. Observational estimates alone do not necessarily increase the IPCC uncertainty range. Rather, the larger IPCC uncertainty reflects the consideration of multiple lines of evidence and sources of uncertainty beyond MMEs alone. We have therefore revised the sentence on Lines 327–329:

"IPCC assessments consider a broader range of evidence than MMEs alone, including individual model studies, uncertainties from missing or poorly represented processes, independent observational, and energy-balance constraints (Myhre et al., 2013b; Forster et al., 2021), thus resulting in broader ERFaer uncertainty ranges."

8. Tables 1 and 2: Double-check units and (Abs) versus (Rel)… For example, *coating_so4* seems like it should be (Rel) instead of (Abs) and if *activ_w* were (Rel), shouldn't the default be 1?

   Thanks for the suggestion. *coating_so4* represents the number of sulfate monolayers required to transfer an insoluble particle to a soluble mode and is therefore perturbed as an absolute value. Parameters labeled "Rel" are scaling factors applied to variable values produced by the original, untuned parameterisation. Thus, control values different from 1 represent tuning factors used in the default model configuration. We have clarified this in the caption of Table 1:

   'Parameters labeled "Rel" are scaling factors applied to modelled variables produced by the original parameterisation, also known as tuning parameters, whereas those labeled "Abs" represent absolute parameter values. '

9. Figure 11: I don't understand what the criterion for the green contours means, is it possible to rephrase? I'm specifically confused about what it means to compare a regional parametric 2-sigma uncertainty to the "90th percentile of the global distribution".

   Thank you for pointing this out. We meant that the green contours enclose locations with parametric uncertainty in the highest 10% globally. We have rephrased the figure caption as:

   "Green isolines enclose regions with parametric uncertainty in the highest 10% globally."

10. Figure 12: There's a suspicious linear-looking feature in ICON-HAM at ~130 W, especially in liquid cloud fraction but also seemingly in LWP and CDNC…

    Thank you for catching this! The linear-looking feature was caused by a bug in the CIS time-collocation routine, which has now been fixed.

11. Lines 741-750: The sizeable positive bias in ERF in the emulator seems worth mentioning here.

    Good point! We added the following to the text on Lines 771–772:

‘We note that the emulator shows a positive bias for strongly negative ERFaer, which arises primarily from ACI.’

12. Data/code availability: Can’t the emulator be shared publicly?

Great suggestion! We have now made the trained emulators publicly available on Zenodo (https://doi.org/10.5281/zenodo.22780345), together with the auxiliary data and an example for running the emulators. The corresponding link has been included in the Code availability section.

13. Figures S5,6,11: Ship emissions contributing zero uncertainty everywhere, including the northern Atlantic and Pacific, seems implausible. This should be discussed somewhere in the paper.

Thanks for the comment. The contribution from shipping emissions appears close to ‘zero’ in Figs. S5, S6, and S11 mainly because it is much smaller than that from other parameters. Even though shipping emissions are smaller than emissions from other anthropogenic sectors (e.g., industry, energy, and domestic sources), we still see distinct spatial patterns of ERF and CDNC uncertainty associated with shipping emissions when using a narrower color-bar range (Fig. R4). The ERF uncertainty due to shipping emissions is estimated to be about 0.01 W m-2, broadly consistent with, although somewhat smaller than, estimates from other models (Jordan et al., 2024; Skeie et al., 2024). The relatively weak effect of shipping emissions in our model may partly be explained by the underestimated cloud cover over oceans. This has now been discussed in Lines 793–796:

‘Although shipping emissions are an important target for reductions, e.g., through the IMO 2020 regulations (Diamond, 2023), they contribute only a small ERFaer uncertainty (0.01 W m−2) here, mainly due to their smaller magnitude compared with other anthropogenic emission sectors and the underestimated marine cloud cover limiting their effect through ACI in our model.’

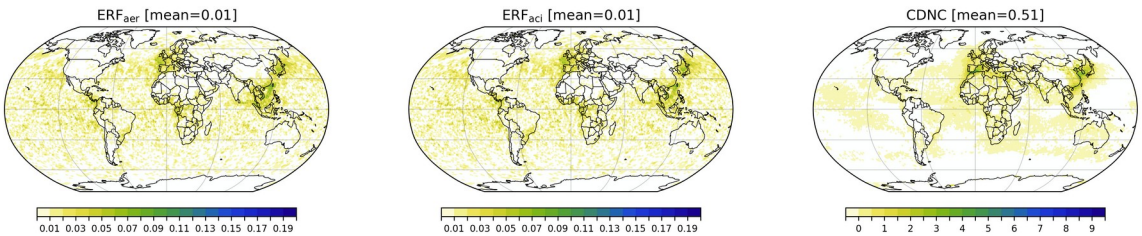


Figure R4. Parametric uncertainties (2σ across 200,000 emulated model variants) of ERFaer (W m−2), ERFaci (W m−2), and CDNC (# cm-3) over 1850–2025, caused by shipping SO2 emissions (*emi_shp_so2*). The fields are the same as those shown in Figs. S7, S8, and S15, but are displayed here using narrower colorbar ranges.

Diamond, M. S.: Detection of large-scale cloud microphysical changes within a major shipping corridor after implementation of the International Maritime Organization 2020 fuel sulfur regulations, Atmos. Chem. Phys., 23, 8259–8269, https://doi.org/10.5194/acp-23-8259-2023, 2023.

Jordan, G. and Henry, M.: IMO2020 regulations accelerate global warming by up to 3 years in UKESM1, Earth's Future, 12, e2024EF005011, https://doi.org/10.1029/2024EF005011, 2024.

Skeie, R. B., Byrom, R., Hodnebrog, Ø., Jouan, C., and Myhre, G.: Multi-model effective radiative forcing of the 2020 sulfur cap for shipping, Atmos. Chem. Phys., 24, 13361–13370, https://doi.org/10.5194/acp-24-13361-2024, 2024.