Author response to reviewer comments on Regayre et al. manuscript entitled "Remaining aerosol forcing uncertainty after observational constraint and the processes that cause it."

We thank both reviewers for their careful reading of the manuscript and for their positive and constructive evaluations. We appreciate their positive assessments and constructive suggestions which have helped improve the clarity of the paper. Reviewer comments below are in blue font, followed by author responses and revised, tracked-changes text where appropriate.

Reviewer 1

Review for ACP of Regayre et al. manuscript entitled "Remaining aerosol forcing uncertainty after observational constraint and the processes that cause it."

Summary and General Comments:

This analysis uses an existing climate model perturbed parameter ensemble (PPE) and after first discussing how the progressive/sequential addition of observations yields constraints (with decreasing return) on aerosol forcing, they analyze the role of parameters on the remaining aerosol forcing uncertainty. Interestingly, they find that parameters that initially seem unimportant contribute a larger role after the optimal constraint. They then take a different approach and use a K-means clustering algorithm to find the dominant clusters of parameters that contribute to the uncertainty, and they plot these clusters onto a latitude-longitude map, enabling visualization of spatial extent of clusters relative to the latitude-longitude resolved aerosol forcing uncertainty map. The paper has numerous quantitative details of parameter/cluster contributions to the uncertainty, and one has to pay close attention so as to not get lost in details, but the analysis is interesting, and the clustering approach toward understanding uncertainty contributions is an interesting approach that warrants publication. I do not have major comments, but I do have a few questions (below), mostly about the chosen clustering approach, that the authors may wish to consider in their revision.

Specific Comments/Questions:

I see that in the PPE being analyzed, less than 10 non-aerosol parameters were perturbed, which is small. But climate models have far more than 10 physics parameters that impact surface wind, surface-atmosphere fluxes, turbulence, convective mass flux, clouds, and by extension, aerosol-convection-cloud interactions. One cannot at this point go back to re-do this particular PPE, and I applaud that there are many aerosol parameters (the other 27 parameters), but is it overall possible that the non-aerosol portion of the parameter space really was not explored much and results are muddled by a lack of effort to explore the many other parameters that impact environments and clouds? Thus, when speaking of "remaining" uncertainty, would not the full parametric uncertainty be much larger and we're really not at the level of looking at remaining parametric uncertainty?

This is an important point. We agree that free-running fully coupled Earth system models contain a huge number of parameters that will influence circulation, clouds, and surface-atmosphere interactions. A comprehensive exploration of all such parameters would indeed further expand the ΔF_{aci} uncertainty range. However, such PPEs are not yet viable due to the number of simulations required. In practice, we design PPEs according to the weights we allocate to the importance of coupling model components, having a free-running vs nudged atmosphere and the number of parameters we intend to perturb. In Regayre et al., 2023 (R23), we elected to perturb as many aerosol and physical atmosphere parameters as possible within a set of nudged simulations.

The selection of non-aerosol parameters in the R23 PPE was informed by earlier sensitivity analyses across multiple model versions (e.g. Regayre et al., 2018; Sexton et al., 2012, 2021) which demonstrated that a subset of physical atmosphere parameters make important contributions to ΔF_{aci} uncertainty in our nudged model framework. We have adapted text on line 138 to reflect that the remaining uncertainty could be larger in a free-running PPE with additional non-aerosol parameters, as stated in R23.

Revised text from line 138:

This "optimal" constraint reduced ΔF_{aci} uncertainty to the maximum limit with the~~ir~~ chosen observations and structurally imperfect model within the explored parameter space, noting that total uncertainty could be larger in free-running simulations or if additional parameters were included.

Aside from the land domain of eastern Asia/China, it is obvious that the other geographic regions of large uncertainty in Fig. 2c – namely, the stratocumulus regimes off the west coasts of the continents – are covered by a diverse range of clusters (Fig. 4, with light/dark blues, oranges, reds, greens, purples, etc.). I did not anticipate this; in general, I would have thought that there would be a mixing of clusters in geographic domains where the uncertainty was smaller in absolute W/m2. Intuitively, for example, I would think that the oceanic stratocumulus regimes would have emerged as a consistent color that then transitioned to another color to the west as those cloud types transitioned into cumulus.

We thank the reviewer for this insight which highlights one of the key strengths of our approach. By going beyond apriori assumptions about homogeneity of behaviours in known important regions for aerosol-cloud interactions, our approach highlights more subtle sub-regional differences in the causes of unconstrained and constrained ΔF_{aci} uncertainty.

These clusters reflect meaningful differences in emission sources, removal pathways, microphysical processes, and non-aerosol parameters that control cloud and radiative properties. Transitions between clusters are physically interpretable, and their spatial patterns provide insight into regional-scale processes, offering greater power than analyses based solely on intuitive large-scale cloud transitions. The differences in the drivers of ΔF_{aci} uncertainty between Northern and Southern Hemisphere stratocumulus regimes are somewhat unexpected, as discussed in the paragraph starting on line 659. Nevertheless, Northern Hemisphere clusters share several causes of uncertainty with clusters found in Southern Hemisphere regions of persistent stratocumulus, although the relative importance of these parameters differs among clusters. Additionally, within any given ocean basin distinct clusters have overlapping causes of remaining uncertainty, again with differences in their ranking of importance. So, regional ΔF_{aci} uncertainty in oceanic stratocumulus regimes is governed by a common set of key parameters across clusters, though deeper understanding is gained by examining how their relative contributions vary and identifying additional parameters that are prominent in an individual cluster. This perspective challenges assumptions about the spatial uniformity of the drivers of ΔF_{aci} uncertainty in regions of persistent stratocumulus cloud and along stratocumulus-to-cumulus transition pathways.

With the above paragraph in mind, how does one interpret this spatial mixing of clusters in Fig 4 coincident with the bullseyes of larger absolute W/m2 uncertainties in Fig. 2c? Might this imply that these clusters as defined are not as useful and could this have arisen due to arbitrary partitioning of a continuous distribution into 10-ish clusters? K-means clustering is most appropriate for data that does exhibit unique modes or clusters, but of course, one can impose k-means clustering on any continuous dataset, even a continuous one not suitable for clustering. Which brings up my second question – how were the number of clusters determined? Finally, instead of clusters connected to geographic domains, would it make more sense to define clusters according to cloud controlling factors (i.e., diagnostics related to stability, moisture, dynamic indices) and then check to see if similar clusters emerge no matter where you are on Earth, so long as you have a similar thermodynamic and/or dynamic environment?

The reviewer raised some thoughtful considerations of our clustering approach, and they are correct that the choice of 10 clusters is somewhat arbitrary. We selected 10 clusters to balance interpretability and resolution. With 10 clusters, the causes of uncertainty are distinct both in order of importance and proportion of uncertainty caused, allowing for meaningful comparisons. Adding more clusters obscures broader spatial patterns, whilst including fewer clusters merges regions with different dominant drivers of uncertainty. For example, with only 6 clusters in total, clusters 3, 4 and 10 are merged as they share multiple key causes of uncertainty, masking important regional differences in the shifting importance of these parameters as causes of remaining uncertainty across northern hemisphere marine regions.

We acknowledge that deeper insight could be gained by analysing clusters on a monthly or seasonal scale. This type of analysis might reveal alignment of clusters with cloud-controlling factors. Such an analysis is likely to highlight regimes defined by stability, moisture and dynamics, as the reviewer suggests, but is beyond the scope of this paper.

Finally, the k-means clustering is an effective choice that helps us achieve the goals of this research: identifying regions where ΔF_{aci} uncertainty persists despite the "optimal" R23 constraint and identifying both novel and existing observations that could further reduce remaining model uncertainty. The fact that the cluster boundaries in Fig. 4 do not perfectly align with the "bullseyes" of largest remaining ΔF_{aci} uncertainty in Fig. 2c underscores the importance of evaluating ΔF_{aci} uncertainty systematically and robustly to make progress in constraining model uncertainty and improving climate projection skill.

Reviewer 2

The manuscript "Remaining aerosol forcing uncertainty after observational constraint and the constraints of aerosol and cloud properties to analyze and decrease the uncertainty in aerosol forcing determined using UKESM climate model. The manuscript makes a substantial contribution to scientific progress within the scope of Atmospheric Chemistry and Physics, mainly through its innovative approach to diagnosing remaining uncertainties and its ability to provide guidance for future model development and what kind of observations are needed to decrease model uncertainty. The manuscript addresses highly relevant questions within Atmospheric Chemistry and Physics by diagnosing the causes of aerosol cloud forcing uncertainty. It introduces novel concepts and methods. The study reaches substantial and actionable conclusions, recommending targeted existing and new observations to constrain the parameters causing uncertainty in the aerosol forcing. The scientific approach is sound and clearly described, and the results adequately support the conclusions regarding shifting uncertainty drivers. Result traceability is high, with both the analysis code and underlying ensemble data made available, and appropriate credit is given to the foundational work (R23) on which the study builds. The title and abstract accurately reflect the content, the manuscript is well structured, and the number and quality of references and appendices are appropriate. I can recommend publishing the manuscript after the following minor comments have been addressed.

- The manuscript relies very much on the Regayre et al., (2023) paper. It would be good to repeat what observational constraints for certain months, such as August Hd means although it has already been explained in Regayre et al.

We agree the use of monthly mean data in R23 could be clearer and have changed the text in the first paragraph of section 2.3 starting on line 234. We define H_d on line 241, its role as a proxy for PI to PD N_d on line 368, and the effect of H_d constraint on model parameter values on lines 370-382, so do not elaborate on these aspects any further.

Revised text from line 234:

Regayre et al. (2023) constrained ΔF_{aci} using multiple satellite-derived cloud and radiation properties. Observations used for constraint included liquid water path (LWP), liquid cloud fraction (f_c), cloud optical depth (τ_c), and cloud droplet effective radius (r_e) from the MODIS instruments (King et al., 2003). τ_c and r_e values were used to calculate cloud droplet number concentration (N_d) values. Observational constraints also included outgoing top-of-the-atmosphere shortwave radiative flux (F_{SW}) measurements from the Clouds and the Earth's Radiant Energy System experiment (CERES; Loeb et al., 2018). Regional mean observations were derived for regions of persistent stratocumulus cloud in the North and South Atlantic, North, and South Pacific and Southern Ocean. R23 also used hemispheric differences in marine *N*d for constraint (H_d). For each observation type, monthly means, annual means and seasonal amplitudes were treated as distinct observations. R23 additionally made use of~~, as well as~~ multiple observed relationships between aerosol, cloud, and radiation properties along transects from stratocumulus- to cumulus-dominated regions during hemispheric summer months.

- Line 197: why only OH and O3 are mentioned?

These species were specifically mentioned because they were perturbed in R23.

Revised text from line 194:

Carbonaceous aerosol emissions were prescribed using CMIP6 (1850) and Copernicus Atmospheric Monitoring Service (CAMS; 2016-17) data, whilst ocean surface concentrations of dimethylsulfide (DMS) and chlorophyll, as well as atmospheric concentrations of gas species (including oxidants OH and O3, which R23 perturbed between 70% to 130% of baseline values) were prescribed using monthly mean output from a fully coupled version of the UKESM model averaged over the 1979 to 2014 period.

- Line 254: in R23 it says 225 observations

Corrected.

- Line 285: Why does high number concentrations of 10 nm particles suppress cloud formation? They are unlikely to activate being so small. I would expect that their size rather than number suppresses activation.

This is a somewhat counter-intuitive result. Classically, we assume only aerosol larger than around 50 nm diameter affect cloud albedo. Aerosol smaller than 10 nm diameter have extremely high Kelvin curvature so will not readily activate to form cloud droplets. The explanation relates to how a large number of small emitted particles affects the aerosol size distribution, and hence activation efficiency. Total aerosol surface area will be far larger when an abundance of implausibly small sulfate particles is emitted. Higher total surface area will spread the limited amount of condensable vapours (e.g. H_2SO_4 and low-volatility organics) across more particles which will limit particle growth to cloud-active diameters.

Reivsed text from line 284:

For *prim_so4_diam* values lower than 10 nm, extremely high aerosol number concentrations ~~are so high that they implausibly~~ lead to unrealistically large total aerosol surface area and smaller cloud condensation nuclei that results in an implausible suppression of cloud formation in the simulated present-day atmosphere, hence these values were ruled out by the R23 constraint.

- Line 356: Can you be more specific, what kind of in-depth analysis would reveal good observational constraints?

We agree that "in-depth" was vague and have altered the text to be more descriptive.

Revised text from line 355:

Parameters that initially seem unimportant contribute more than a few percent to the remaining uncertainty after constraint, which suggests in-depth analysis of spatial variation in the process-level drivers of remaining uncertainty, before and after optimal constraint, may reveal observations with potential to further constrain ΔF_{aci}.

- Figure 2: Why is the uncertainty so high over the southern oceans? I would expect that anthropogenic aerosol does not have a significant effect on clouds there, still the uncertainty looks to be similar to what it is over Europe. In Figure 3. the size of primary sulfate seems to be a significant source of uncertainty over that area. How does primary sulfate affect those regions?

Southern Ocean ΔF_{aci} uncertainty (defined in the text on line 138 as the 90% credible interval range, following R23) is less than 5 W m^{-2} in the unconstrained case and less than 3 W m^{-2} after constraint – much lower than other marine regions. However, the Southern Ocean is far from pristine for much of the year (see Fig. 3 in Hamilton et al., 2014), so some ΔF_{aci} uncertainty is expected.

The most likely source of anthropogenic aerosol is from the shipping sector. This is the only emission source of primary sulfate in the Southern Ocean in our simulations, though some sulfate aerosol may be advected from industrial sources along western continental coastlines. Our Fig. 3 shows the prim_so4_diam parameter does contribute to uncertainty in this region, but to a far lesser degree than sea salt emissions (sea_salt) and physical model parameters (bparam and two_d_fsd_factor).

- Line 607: Alternatively: cloud droplet activation → updraft velocity in cloud droplet activation

We think this comment refers to the direction of control implied by order of phrases, so have adapted text to make causality clear.

Revised text from line 608:

Cloud droplet activation ~~The importance of updraft velocity~~ in these regions is strongly controlled by updraft velocity, consistent with an updraft-limited regime ~~of cloud droplet activation when~~at high aerosol concentrations ~~are high~~ (Reutter et al., 2009).

- I expect that all simulations are done using the model tuning setup of the default setup. Can different parameter combinations cause the model to go out-of-tune (for example with respect to radiative balance) and thus causing problems in PPE analysis?

The first wave of history-matching used in creating the R23 PPE used global mean top-of-the atmosphere shortwave radiative flux as a constraint on parameter space, alongside other observational metrics, to ensure first-order agreement before creating the larger PPE used in the R23 analysis within this pre-constrained parameter space. We make this clear with the text change starting on line 211. It is therefore extremely unlikely that the PPE analysis will be negatively affected by being “out-of-tune” – global mean radiative fluxes will only vary within the range allowed by the pre-constrained parameter space, though will vary noticeably at the regional scale according to parameter perturbation effects. Models that are tuned to match observed global mean metrics are very likely to be “right for the wrong reasons” due to compensating errors at the regional and seasonal scales. R23 showed the tuned UKESM model version likely has specific structural deficiencies related to compensating errors in liquid water path and cloud droplet concentrations - all tuned climate models very likely suffer from similar compensating errors.

Revised text from line 211:

The Regayre et al. (2023) PPE was created in two stages using a history-matching style approach (Craig et al., 1997; Williamson et al., 2013) to ensure that the 221 ensemble members (parameter combinations) spanned the 37-dimensional parameter space whilst achieving acceptable agreement with large-scale climate metrics including the global mean outgoing shortwave radiative flux.

- Table A1: For kappa_oc “affects wet diameter and clear-sky radiative flux” can be removed.

Removed.

Technical comments:

School of Earth and Environment is twice in School of Earth and Environment, School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK

Fixed

Please fix journal abbreviations, fix typos (for example, Regayre 2014, 2015, 2024)

Fixed

References

Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A.: Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments, in: Case Studies in Bayesian Statistics, vol. 121, edited by: Gatsonis, C., Hodges, J. S., Kass, R. E., McCulloch, R., Rossi, P., and Singpurwalla, N. D., Springer New York, New York, NY, 37–93, https://doi.org/10.1007/978-1-4612-2290-3_2, 1997.

Hamilton, D. S., Lee, L. A., Pringle, K. J., Reddington, C. L., Spracklen, D. V., and Carslaw, K. S.: Occurrence of pristine aerosol environments on a polluted planet, Proc. Natl. Acad. Sci., 111, 18466–18471, https://doi.org/10.1073/pnas.1415440111, 2014.

King, M. D., Menzel, W. P., Kaufman, Y. J., Tanre, D., Bo-Cai Gao, Platnick, S., Ackerman, S. A., Remer, L. A., Pincus, R., and Hubanks, P. A.: Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS, IEEE Trans. Geosci. Remote Sens., 41, 442–458, https://doi.org/10.1109/TGRS.2002.808226, 2003.

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., Liang, L., Mitrescu, C., Rose, F. G., and Kato, S.: Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product, J. Clim., 31, 895–918, https://doi.org/10.1175/JCLI-D-17-0208.1, 2018.

Regayre, L. A., Johnson, J. S., Yoshioka, M., Pringle, K. J., Sexton, D. M. H., Booth, B. B. B., Lee, L. A., Bellouin, N., and Carslaw, K. S.: Aerosol and physical atmosphere model parameters are both important sources of uncertainty in aerosol ERF, Atmospheric Chem. Phys., 18, 9975–10006, https://doi.org/10.5194/acp-18-9975-2018, 2018.

Regayre, L. A., Deaconu, L., Grosvenor, D. P., Sexton, D. M. H., Symonds, C., Langton, T., Watson-Paris, D., Mulcahy, J. P., Pringle, K. J., Richardson, M., Johnson, J. S., Rostron, J. W., Gordon, H., Lister, G., Stier, P., and Carslaw, K. S.: Identifying climate model structural inconsistencies allows for tight constraint of aerosol radiative forcing, Atmospheric Chem. Phys., 23, 8749–8768, https://doi.org/10.5194/acp-23-8749-2023, 2023.

Reutter, P., Su, H., Trentmann, J., Simmel, M., Rose, D., Gunthe, S. S., Wernli, H., Andreae, M. O., and Pöschl, U.: Aerosol- and updraft-limited regimes of cloud droplet formation: influence of particle number, size and hygroscopicity on the activation of cloud condensation nuclei (CCN), Atmospheric Chem. Phys., 9, 7067–7080, https://doi.org/10.5194/acp-9-7067-2009, 2009.

Sexton, D. M. H., Murphy, J. M., Collins, M., and Webb, M. J.: Multivariate probabilistic projections using imperfect climate models part I: outline of methodology, Clim. Dyn., 38, 2513–2542, https://doi.org/10.1007/s00382-011-1208-9, 2012.

Sexton, D. M. H., McSweeney, C. F., Rostron, J. W., Yamazaki, K., Booth, B. B. B., Murphy, J. M., Regayre, L., Johnson, J. S., and Karmalkar, A. V.: A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: part 1: selecting the parameter combinations, Clim. Dyn., 56, 3395–3436, https://doi.org/10.1007/s00382-021-05709-9, 2021.

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., and Yamazaki, K.: History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble, Clim. Dyn., 41, 1703–1729, https://doi.org/10.1007/s00382-013-1896-4, 2013.