

In our response, reviewer comments are marked in bold, our responses and original text in plain text, and altered text in the paper in bold italic.

Response to reviewer 1 (J. Mülmenstädt)

We thank the reviewer for their interesting and useful comments on our manuscript. Our responses to these comments are given below.

Major comment (Section 1): Are the synthetic observations used in this analysis representative of model tuning?

The authors argue that producing one tuned configuration of a climate model underestimates the size of the parameter space (no disagreement there); that the tuning process is often done one parameter at a time, underestimating the non-additive effects of varying multiple parameters at once (only mild disagreement there); and that an analysis based on an ensemble of ERF estimates from tuned models underestimates the true ERF uncertainty. For this last conclusion to be true, the range of ERFs that a single model can plausibly produce (“plausible” meaning consistent with the observed present-day climate) must be non-negligible compared to the intermodel ERF spread, as sketched in Figure 10.

In their analysis, the authors subject the different parameter combinations of their emulated aerosol–climate model to a consistency check against nine observables of the present-day climate. This is the emulator-world equivalent of tuning a single climate model to agree with the present-day climate; the further conclusions made in the manuscript about the uncertainty on the model’s ERF estimate thus hinge crucially on whether this consistency with observations is indeed equivalent to model tuning.

I am concerned that the answer may be “no”, for the following reason. The observational dataset used here is the July mean conditions over Europe. Europe is a tiny portion of the globe, and July only samples one point in the seasonal cycle of aerosol and cloud properties. Since, as the authors point out, the model uncertainty stems from the interaction of aerosol and host model parameters, it seems like one would want to use the widest spectrum of weather and aerosol conditions available to be able to discard observationally excluded parameter combinations.

As a GCM tuning strategy, a European seasonal approach would fail because the constraint on the global-mean climate would be negligible. I expect that many of the ensemble members that are consistent with the European July observations will have outlandish global-mean TOA fluxes, cloud fields, etc., that would get them rejected in a constraint that uses global observations. This, I expect, would narrow the estimate of the single-model uncertainty compared to what the authors find using their constraint strategy.

The interesting question, in my mind, is how much using global constraints would narrow the uncertainty, i.e., whether Figure 10(b) would still look mostly as it does now or start to look more like Figure 10(a).

I realize that Europe was chosen for a reason, namely that global observations do not exist for all of the fields used as constraints (CCN; decadal trends of surface shortwave radiation and AOD). The most accurate equivalence with GCM tuning would probably be to use global fields where they are available and European fields (but for more than a single month) otherwise.

We don't tackle the problem of global mean ERF or global tuning; the paper is focused solely on Europe. The word tuning doesn't need to imply global. Aerosol models are frequently tuned or adjusted in some way (call it what you will) to agree with observations in a particular region or environment, so it's this process we are drawing comparisons with. We restrict the analysis to Europe because there is a distinct set of parameters that are important to aerosol properties and ERF uncertainty in each region (see Lee et al, 2016: 10.1073/pnas.1507050113 for a map of clusters of uncertain parameters). Because parameter sensitivities vary regionally, to constrain the global forcing you would need to constrain the different regions individually – i.e., in the aerosol world it doesn't make much sense to talk about global tuning. So our European study is like a mini version of what one would need to do globally. If it doesn't work for Europe (with a distinct set of parameters) then it won't work globally either. Using global ToA flux would have limited effect because of many regional compensating factors. We tested this in Regayre et al, 2018 (10.5194/acp-18-9975-2018; section 3.5).

Another reason to use Europe was to relate to the Cherian et al, 2014 study where European surface solar radiation trends were used as an emergent constraint on regional and global forcing. As we state in our conclusions, we think the constraint will be weaker than they calculated because of unconstrainable uncertainties in the individual models.

So to answer the question “how much using global constraints would narrow the uncertainty”, the answer is that it would narrow the *global* ERF more than we would currently achieve. But a global constraint won't have much effect on Europe (again, because of the distinct set of parameter uncertainties). We only ever refer to Europe, and we maintain that focus when we compare with the Cherian et al Europe SSR constraint on Europe-mean ERF in Fig 10.

Minor suggestion (Section 2): Some of the conclusions of the paper could be phrased as recommendations on the future direction of the aerosol forcing community

This is an interesting suggestion and we are pleased that the reviewer has identified the potential impacts that our work and results could have in the aerosol forcing community. We did try to draw out the implications. We prefer not to completely change the style of the conclusions because we think it's important that we include a synthesis of findings and integration with current knowledge, rather than recommending future studies. We think recommendations are more appropriate for perspective articles and reviews. In response to some specific suggestions we have tried to highlight the recommendations a bit.

Apart from the main finding of large single-model uncertainty, there are other interesting findings in the paper that I think are underemphasized:

- **The emulator–PPE method is a necessary step toward estimating the model uncertainty correctly. But to narrow the uncertainty range, it will also be necessary to assemble the most powerful (i.e., discriminating) combination of observables. Therefore, I encourage the authors to advocate for both of these necessary steps at once – the more they go hand-in-hand, the more efficient we are likely to be at making progress. (For example, the emulator–PPE method could point to which missing observable would provide the greatest additional constraint if it were added to the existing set of observables; see my comment on Figure 9 in the next section.) This sentiment is kind of there in the last sentence of the abstract and in the text (p. 17, l. 20; p. 23, l. 16), but it would be a missed opportunity not to phrase it more forcefully and more positively in the abstract.**

We have added a sentence on this to the end of abstract to highlight this point. The end of the abstract now reads:

“...However, the uncertainty in the aerosol ERF after observational constraint is large compared to the typical spread of a multi-model ensemble. Our results therefore raise questions about whether the underlying multi-model uncertainty would be larger if similar approaches as adopted here were applied more widely. ***The approach presented in this study could be used to identify the most effective observations for model constraint.*** It is hoped that aerosol ERF uncertainty can be further reduced by introducing process-related constraints, however, any such results will be robust only if the enormous number of potential model variants is explored.”

- **The aerosol forcing change over the coming decades is a much easier target than the forcing relative to preindustrial conditions, assuming the emissions changes are known. Perhaps it is time to move away from ERF as the holy grail of the field and instead focus on future aerosol emissions scenarios.**

We agree, and have made this point in at least a couple of previous papers. There is a paragraph (5th of the conclusions in the original manuscript; 6th paragraph in the revised manuscript) dealing with this point, but we agree it could be focused. We now write:

“Observational constraint using nine observations has the potential to reduce the uncertainty in aerosol ERF ***slightly more over a multi-decadal period than over the full industrial period:***”

The rest of the paragraph then explains why this is important. Then we add at the end of the paragraph:

“A shift of emphasis of the research community towards trying to constrain decadal forcing uncertainty, instead of industrial era forcing, is likely to accelerate progress.”

- **AOD is a terrible variable if your aim is to understand aerosol–cloud interactions. We should figure out something better. Ed Gryspeerdt’s work seems to show that aerosol index is a much better proxy for CCN. If the authors’ model diagnoses AI, it would be easy for them to refute or corroborate this result.**

Unfortunately, we cannot compute AI from the model runs of this PPE and so we cannot evaluate this here. However, we disagree that AOD is unambiguously a “terrible variable” for constraining forcing. This point of view is maybe based on the fact that AOD and CCN are not closely related but AI and CCN are more related. However, our study shows that neither CCN nor AOD alone provides a strong constraint on ERF (Fig 7). In that sense you might argue that CCN is also a terrible variable. The key for model constraint and uncertainty reduction is to find *combinations of observables* that are sensitive to the same set of process or emission uncertainties as ERF. So *all* observed variables are helpful in their own way. It is possible that AI will be useful for direct extrapolation back to the pre-industrial period, but it will not necessarily be useful for model constraint. We intend to include AI in the output diagnostics of a future PPE that is in the planning stage, with which we will be able to investigate the potential of AI for model constraint.

We have added a new paragraph in the conclusions (after the 4th paragraph in the original manuscript):

“It is often argued that AOD is a poor variable to use for understanding aerosol-cloud interactions. However, our results show that even the most strongly related measurement (CCN) also does not

provide a strong individual constraint on ERF_{Aci} (Figure 7). It is doubtful that other derived variables like aerosol index will be any better. The key to model constraint is to find combinations of observations that help to constrain ERF: Individual constraints are unlikely to be effective, although they may appear to be effective if the model uncertainty is not fully sampled.

These results have a direct bearing on where the aerosol forcing community would best invest its efforts in order to reduce the forcing uncertainty. I understand that the authors want to keep the focus on the main result (the large single-model uncertainty), but they might think about ways to give these other findings a prominent place as well. For example, in our recent review article on radiative forcing by aerosol–cloud interactions (<https://doi.org/10.1007/s40641-018-0089-y>), we eschewed the traditional re-listing of conclusions at the end of the paper in favor of making recommendations to the forcing community. Mentioning these recommendations in the abstract may be appropriate as well.

We understand how useful a set of recommendations to the aerosol forcing community would be, but we have decided not to fully re-work our conclusions in this paper to a set of recommendations. The presented study, based only on synthetic observations, corresponds to a specific stage in our overall work to understand and constrain the aerosol forcing uncertainty in this aerosol-climate model. Our current and future work is now expanding this work to use real observations in the constraint methodology, and we are learning even more as we move forward with this approach. We feel that recommendations to the community must take into account findings from the complete constraint process in which real observations are used. We therefore elect to defer producing such a set of recommendations at this earlier stage to a future time when we also have a comprehensive evaluation of model constraint with real observations.

Minor Comment 1: Table 1: A bit more detail on parameters 8 and 9 would be nice. (Threshold in what variable? Rate of change with respect to what?)

Full descriptions of the parameters perturbed in this PPE are given in Yoshioka et al, 2018 and Regayre et al, 2018. We therefore do not repeat this full detail in this publication. We have made it clearer to the reader where more detailed descriptions of the parameters can be found by moving the reference for this from the end of paragraph 3 (page 8 line 25 in the original manuscript) to paragraph 2 in Section 2.3. Paragraph 2 of Section 2.3 now reads:

“The 27 perturbed parameters are listed in Table 1. They are categorized as either aerosol (aer) or atmospheric (atm) according to their role in the model. To define the set of parameters we used expert elicitation and carried out one-at-a-time parameter perturbation screening experiments to quantify the effect of individual parameter perturbations away from the default setting. **The selected parameters are described in more detail in related papers (Regayre et al., 2018; Yoshioka et al., 2018)**”

For the specific parameters mentioned here, we have adapted the text in Table 1 to provide further clarity.

For parameter 8: Dec_Thres_Cld, we have changed the description to: ***“Threshold for the ratio of buoyancy consumption to production before decoupling occurs”***

For parameter 9: Fac_Qsat, we have updated the description to: “Rate of change in convective parcel maximum condensate *with altitude*”

Minor Comment 2: p. 10, l. 8: Do we know that the validation carries over to this new emulator? What is the rationale for not simply using the emulator that is definitely validated?

We are confident that the validation does carry over to the new emulator built using all simulations. The rationale for doing this is to incorporate all of the information that we have from our PPE into the emulator model we use.

We do investigate the validity of the final emulator using a process called ‘leave-one-out’ validation on the merged set of simulations. In this procedure we remove a simulation from the set, build a new emulator and then use that emulator to predict the removed simulation, repeating this for all simulations in the set in turn. This procedure has returned good validation prediction results in all cases, and Figure A below shows an example of this validation for our present-day AOD emulator. The plot shows that the real model output and corresponding Leave-one-out emulator predictions (with 95% credible intervals) strongly correlate along a line of equality. Hence, this alternate validation indicates that the addition of the extra simulations in the construction of the emulator does not deteriorate the emulator quality and we have confidence in this approach.

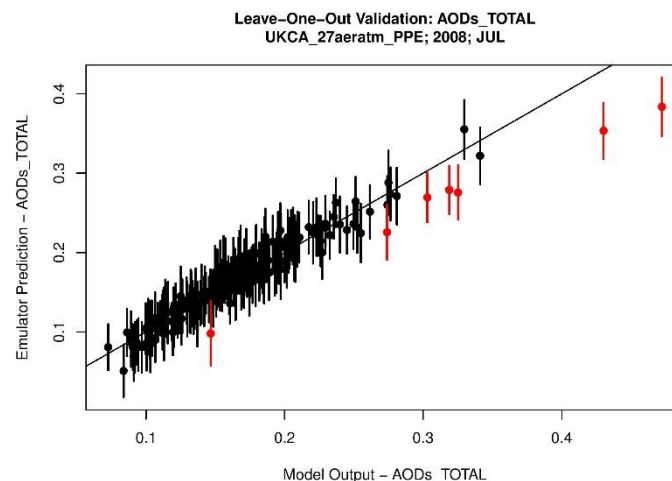


Figure A: Example output of a ‘Leave-one-out’ validation test of the final emulator.

We have added the following sentence to page 10, line 9 of the original manuscript (also page 10, line 9 of the revised manuscript) to detail that we use a leave-one-out validation procedure to verify the quality of the final emulator:

“A ‘leave-one-out’ validation procedure (where each simulation in turn is removed from the merged set, and a new emulator is constructed and used to predict that removed simulation) is applied to additionally verify the quality of our final emulator.”

Further explanation on our rationale for this is given below for the reviewer’s interest.

Given a good validation of an initial emulator based only on the original training runs, we could certainly use this emulator for our analysis and have confidence in the results. However, the validation simulations are purposely placed into gaps between the locations of the training simulations and so including this extra information enables us to better guide/tweak the predicted

emulator response in these gaps to be even more in-line with the underlying 'true' model output. We believe that if a good emulator is achieved based on only the training runs, then a further emulator based on both the training and validation runs must be of similar quality, if not better, given the inclusion of the extra information that these extra runs provide.

Minor Comment 3: p. 10, l. 14: Do the two approaches (elicited vs uniform PDF) give very different results?

In general, we have found that the overall constraint achieved on regional average model outputs (aerosol properties and forcings) is of a very similar nature under both approaches (using the elicited v's using the uniform pdfs). However, there are differences in how the generated large sample of model variants (size 4 million) covers the multi-dimensional parameter uncertainty and hence the proportion of that sample that is retained on application of the constraint methodology.

Using elicited pdfs rather than uniform pdfs when sampling over the multi-dimensional parameter uncertainty space essentially applies a further constraint to that parameter space, sampling much less in areas that the experts believe to be highly unlikely. This means that the sample of model variants tested against the observations in the constraint procedure is more concentrated towards the most likely (as evaluated by our experts) parameter values in the parameter ranges. The areas sampled less tend to be around the edges of the multi-dimensional space.

On constraint we have found that using the elicited pdfs means we retain more of the 4 million model variants and leads us to a slightly stronger constraint on the parameter space itself (e.g. on the marginal parameter space in Figure 9). We reject more variants using uniform pdfs as the corresponding sample under this assumption covers more densely the parts of the parameter space that were deemed unlikely to be plausible by the experts in the elicited pdfs. This result shows some consistency between the model behaviour and the expert judgements, which gives us confidence in using the elicited pdfs. Hence, in the results presented we choose to use the elicited pdfs and so include all available information in the sampling in order to determine the full potential of possible constraint on ERF that is achievable with our approach.

Minor Comment 4: p. 11, l. 10: This does not make intuitive sense to me; if the "truth" is chosen to lie at one the edge of parameter space, shouldn't ensemble members from the opposite direction of that edge be penalized much more strongly than they would be if the "truth" were chosen to lie in the center?

Yes, it is correct that choosing the "truth" to lie in a different part of the parameter space can affect the corresponding 'observed' output values and so change the model variants (and hence parts of parameter space) that are retained on constraint. However, the effect of this change on model output quantities like aerosol forcing is dependent on the way in which the response surfaces of these outputs vary over the multi-dimensional parameter space, and so it is difficult to predict. The compensating effects of the different parameters in the multi-dimensional space can mean that different areas of parameter space lead to similar output values.

For the model run we investigated as a marginal set of synthetic observations (our validation run 27, which had several parameter values towards the edges of their uncertainty range), we found that the achieved overall constraint on each of the forcing variables (using all observations together) was slightly greater and we also retained fewer model variants in the constrained sample. For ERF we

saw a further reduction in the standard deviation of around 12%, however this reduction was much lower for ERF_{ARI} at only 1%. Even so, the relative individual constraint effect of each observational type (Figure 7) was of the same nature/order for all four forcing variables (ToA flux is the most effective for ERF and ERF_{ACI} , the sulphate concentration is most effective for ERF_{ARI} and ERF_{ARIClr} , and in general the achieved constraint from the individual observable aerosol properties is weak). Hence, the overall conclusions of the study did not change on using a different observation set. The resulting constrained set of model variants still corresponds to a large multi-dimensional area of parameter space with a lot of equifinality.

Given these results, and so as not to over-complicate the presented study, we made the decision to focus on just one observation case and the centralised set (with parameters set to their elicited median values) seemed most appropriate for a general example of our approach.

Minor Comment 5: p. 14, l. 2: Much as I hate linear correlation coefficients, perhaps it would be useful to tabulate them to make it easier to follow this discussion. I am having trouble reading them off Figure 2.

We have added the following table of linear correlation coefficients into Section 3.1 of the manuscript:

ToA Flux	0.20	0.00	0.23	-0.30	0.16	0.10	0.04	-0.03	-0.59	-0.65	0.25	-0.05
	ASSR	-0.04	0.22	-0.72	0.12	0.49	-0.04	0.11	-0.32	-0.22	-0.52	-0.61
		CCN	0.46	-0.20	0.21	0.03	0.37	0.09	-0.14	-0.14	-0.01	-0.09
			AOD	-0.50	0.88	0.59	0.66	0.59	-0.33	-0.27	-0.33	-0.55
				ΔAOD	-0.32	-0.71	-0.17	-0.20	0.48	0.38	0.55	0.76
					PM_{2.5}	0.54	0.69	0.54	-0.21	-0.15	-0.33	-0.44
						Sulphate	0.47	0.64	-0.41	-0.28	-0.67	-0.77
							OC	0.62	-0.34	-0.31	-0.20	-0.30
								BC	-0.27	-0.22	-0.28	-0.46
									ERF	0.98	0.16	0.48
										ERF_{ACI}	-0.04	0.33
											ERF_{ARI}	0.83
												ERF_{ARIClr}

Table 3: Pearson linear correlations (r) between the PPE member regional mean model outputs for Europe in July, for the aerosol properties used as constraints, corresponding to the pairwise scatter plots in Figure 2.

We have edited the first paragraph of Section 3.1 to say:

“Figure 2 shows pairwise scatter plots of the PPE member output (Europe July-mean), which provides an overview of the spread of the model outputs as well as the relationships between the variables. *We further quantify any linear relationships between the variables using the Pearson correlation coefficient (r) in Table 3.*”

We have adjusted the numbering of all following results tables in the manuscript to incorporate this table into Section 3.1.

Minor Comment 6: Speaking of Figure 2:

- **Nice to run into fellow R users.**
- **All the labels are tiny! `par(cex)` may help.**

It is difficult with so many small plots in the same figure to make the labels any larger. We have used '`par(cex)`' to adjust the size of the axis labels to gain the best possible labelling size whilst avoiding axis values overlapping and becoming unreadable.

- **Point clouds are hard to interpret. Perhaps the relationships between variables would be easier to interpret as color maps of the 2D densities of the emulator results?**

The purpose of this figure is to look at the direct model output from the 191 individual model runs before any interpolation over the parameter uncertainty space (using the emulators) is applied. With only 191 points in each case, we do not have enough data to generate robust density contours and therefore use scatter plots.

- **Raster graphics are an abomination. R supports PDF output; this is the best option in general, and in particular for an information-rich figure like this one, where the reader will want to zoom in on interesting features.**

Figure 2 has been re-made using postscript and PDF graphics, and the new version is much improved for zooming in on interesting features. The new Figure 2 is below (at the end of this comment response)

These comments apply to Figure 3 as well, where additionally I see funny boxes around some of the panels at certain magnifications.

We have looked in detail at Figure 3 and cannot find any issues with the figure on magnification. We therefore have not changed this figure. We will check the proofs.

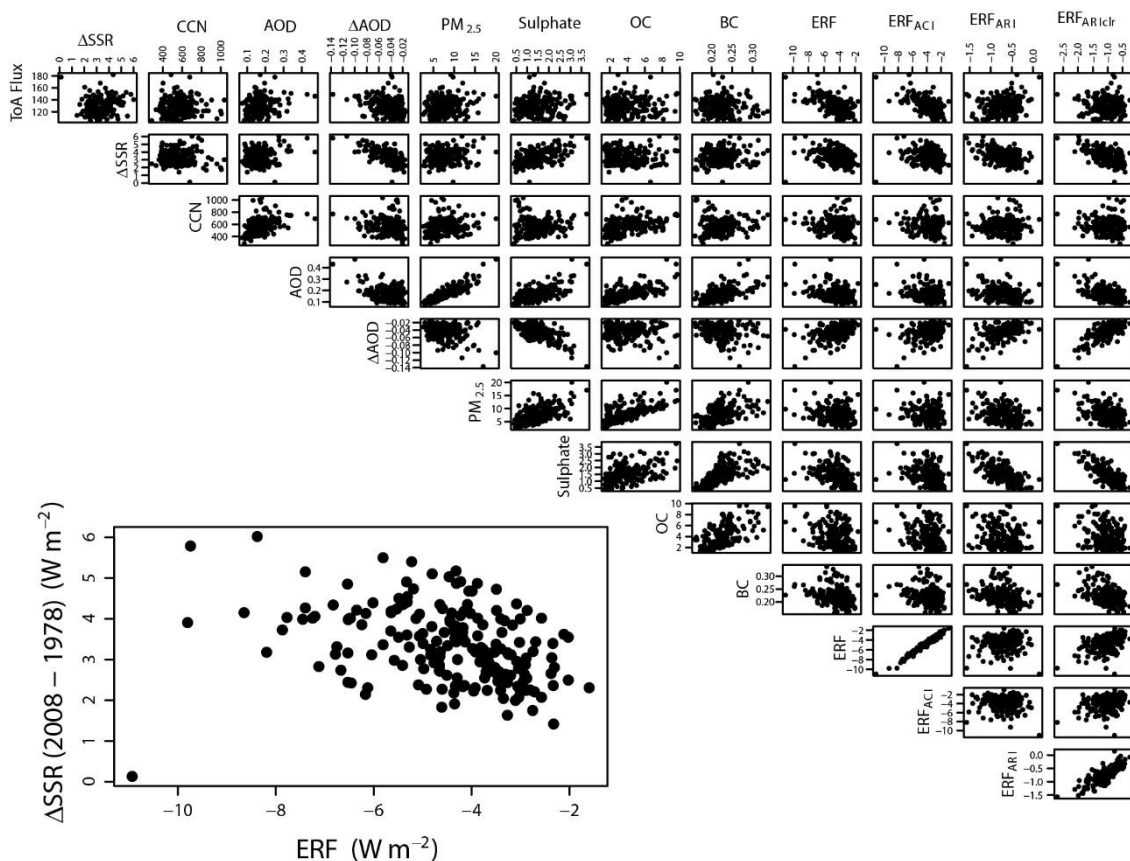


Figure 1. Pairwise scatter plots of the PPE member regional mean model output for Europe in July, for the aerosol properties used as constraints: ToA flux (W m^{-2}), change in SSR (ΔSSR , W m^{-2}) between 1978 and 2008, CCN conc. (cm^{-3}), AOD, surface mass concentrations of $\text{PM}_{2.5}$, Sulphate, OC, and BC ($\mu\text{g m}^{-3}$), the changes in AOD (ΔAOD , W m^{-2}) between 1978 and 2008, and the 1850-2008 forcing variables: aerosol ERF, ERF_{ACI} , ERF_{ARI} and $\text{ERF}_{\text{ARIClr}}$ (W m^{-2}).

Minor Comment 7: Figure 4: In this model, ERF variability is dominated by ERF_{ACI} variability (see Figure 2). Yet the model does not care one bit about aerosol–precipitation interactions, at least not wet scavenging. I believe this is quite different from other GCMs. Any idea why? (Not necessarily something that needs to go into the paper.)

The cloud state is sensitive to precipitation scavenging, but the ERF_{ACI} is not. The ERF is obviously a change from pre-industrial to present-day, so it is not obvious that if you change the aerosol/cloud state in both periods that the ERF will remain sensitive to scavenging.

Minor Comment 8: p. 25, l. 15: In GCM tuning, this would routinely be done; see my main point of criticism.

OK.

Minor Comment 9: Figure 9: Does the converse also hold (that observations of decoupling would be a useful constraint on the model)?

We do not know what the reviewer means by “observations of decoupling”, and so we are not sure on what is being referred to here. We are therefore unable to comment.

Minor Comment 10: p. 27, l. 5: But Cherian et al. do this using models tuned to reproduce the global climate; see my main point of criticism.

Yes, we agree that Cherian et al, 2014 achieve their constraint using a set of models that have each been ‘tuned’ in some way to re-produce the global climate. We compare with our ‘tuned’ model for Europe (which includes the frequently-tuned quantity, the ToA flux). However, in the selection of each individual tuned model, the influence of parametric uncertainty has not been rigorously explored. This means that there will be many other equally plausible tuned versions of each given model that agree with the observations and could have been selected, as we have shown in this study for the single tuned model HadGEM3-UKCA. We have shown that the predicted forcing from the set of the equally plausible model variants (obtained through comparison to a diverse set of aerosol observations) is wide-ranging. This implies that tuning does not necessarily directly reduce the model spread and implies that emergent constraints such as the one obtained by Cherian et al, 2014 likely underestimate the true spread of forcing.

Minor Comment 11: p. 28, l. 10: AOD multidecadal change appears to be double-counted.

This has been corrected in the manuscript. This sentence in the first paragraph of the conclusions (page 30 lines 5-9 of the revised manuscript) now reads:

“The primary objective of our study was to determine how much uncertainty could remain in an aerosol-climate model when it is constrained to match combinations of observations that define the base state of the model: top-of-atmosphere upward shortwave flux, aerosol optical depth, PM2.5, cloud condensation nuclei, concentrations of sulphate, black carbon and organic material as well as multi-decadal change in surface shortwave radiation and aerosol optical depth.”

Minor Comment 12: p. 29, l.1: Would AI or fine-mode fraction work better? I think the authors have the opportunity to make a significant statement here about whether there is a way forward from AOD, which is known to be a poor CCN proxy, via other proxies. See my point in the recommendations section above.

Please see our response to the Minor suggestion (Section 2) above (at the end of page 3). We are not able to evaluate the effects of AI with this PPE, but we doubt that AI would provide any better constraint on the aerosol ERF than AOD. AI might be a better extrapolation variable, but this has not yet been directly demonstrated.

Minor Comment 13: p. 31, l. 6: I don’t understand the point about cancellation of correlated errors, but I would like to. Perhaps the authors could elaborate.

We mean that a model might have a large bias in AOD and in CCN (so neither can be simulated well) but if both of these model variables are biased for the same reason (e.g., incorrect aerosol deposition rates), then the ratio of CCN/AOD might be accurately simulated by the model.

Minor Comment 14: Craig 1997 has a bunch of cryptic initials instead of editor names.

This reference has been corrected in the manuscript.

Minor Comment 15: Gryspeerdts 2017 a and b are the same publication.

This has been corrected in the manuscript.

Minor Comment 16: Stier ACPD 2015 has been superseded by Stier 2016 ACP (<https://www.atmos-chem-phys.net/16/6595/2016/acp-16-6595-2016.html>).

This reference has been corrected in the manuscript.

Minor Comment 17: Penner 2011: the DOI looks strange.

This reference has been corrected in the manuscript.

Minor Comment 18: Pujol 2008: check that this is still up to date with citation("sensitivity").

This reference has been checked in R and corrected in the manuscript.

Minor Comment 19: Zhang 2016: Toshi Takemura's name is misspelled.

This reference has been corrected in the manuscript.

References:

Cherian, R., Quaas, J., Salzmann, M. and Wild, M.: Pollution trends over Europe constrain global aerosol forcing as simulated by 25 climate models, *Geophys. Res. Lett.*, 41(6), 2176–2181, doi:10.1002/2013GL058715, 2014.

Lee, L. A., Reddington, C. L. and Carslaw, K. S.: On the relationship between aerosol model uncertainty and radiative forcing uncertainty, *Proc. Natl. Acad. Sci. U. S. A.*, 113(21), 5820–7, doi:10.1073/pnas.1507050113, 2016.

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, *Atmos. Chem. Phys.*, 18, 9975–10006, doi:10.5194/acp-18-9975-2018, 2018.

Yoshioka, M., Regayre, L., Pringle, K. J., Mann, G. W., Sexton, D. M. H., Johnson, C. E. and Carslaw, K. S.: Perturbed parameter ensembles of the HadGEM-UKCA composition-climate model to explore aerosol and radiative forcing uncertainty, *J. Adv. Model Earth Syst.*, in-prep, 2018.