Supplementary Information for
Identifying climate model structural inconsistencies allows for tight constraint of aerosol radiative forcing

Leighton A. Regayre1,2, Lucia Deaconu3,4, Daniel P. Grosvenor1,2,5, David M. H. Sexton2, Christopher Symonds5, Tom Langton3, Duncan Watson-Paris3, Jane P. Mulcahy2, Kirsty J. Pringle1,5,6, Mark Richardson5, Jill S. Johnson1,7, John W. Rostron2, Hamish Gordon1,8, Grenville Lister9,10, Philip Stier3 and Ken S. Carslaw1

1Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK
2Met Office Hadley Centre, Exeter, Fitzroy Road, Exeter, Devon, EX1 3PB, UK
3Atmospheric, Oceanic and Planetary Physics Department, University of Oxford, Oxford, OX1 3PU
4Faculty of Environmental Science and Engineering, Babes-Bolyai University, Cluj, Romania, 400294
5Centre for Environmental Modelling and Computation, School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK
6Edinburgh Parallel Computing Centre, Bayes Centre, University of Edinburgh, EH8 9BT
7School of Mathematics and Statistics, University of Sheffield, Sheffield, S3 7RH, UK
8Department of Chemical Engineering and Center for Atmospheric Particle Studies, Carnegie Mellon University, Pittsburgh, PA 15213, USA
9Department of Meteorology, University of Reading, RG6 6AH, UK
10National Centre for Atmospheric Science, Reading, RG6 6AH, UK

Leighton Regayre
Email: L.A.Regayre@leeds.ac.uk

This PDF file includes:

- Supplementary text
- Figures S1 to S30
- Tables S1 to S4
- SI References
Model version

We used the atmosphere-only configuration of version 1 of the UK Earth System Model (UKESM1) (1) to create our perturbed parameter ensembles (PPEs). UKESM1 was the model version submitted to the 6th Coupled Model Intercomparison Project (CMIP6) (2). UKESM1 is based on the HADGEM3-GC3.1 physical climate model (3) with additional coupling to key Earth System processes (1), including the United Kingdom Chemistry and Aerosol (UKCA) model (4). The atmosphere-only configuration as used here consists of the GA7.1 atmosphere (5, 6), with additional aerosol, cloud and physical atmosphere structural updates as implemented in UKESM1 (6). GA7.1 includes several structural advancements to the aerosol component of the model which significantly affect anthropogenic aerosol radiative forcing (7). We refer to this model version as UKESM1-A.

Horizontal wind fields above around 2km in our simulations (model vertical level 17) were nudged towards ERA-Interim values for the period December 2016 to November 2017. Nudging largely removes the effects of differences in large-scale meteorology from our PPE members, meaning we can attribute differences between model variants to perturbed parameter values. We do not nudge winds within the boundary layer, as many of our parameters are intended to affect meteorological conditions in this part of the atmosphere.

The model was forced using anthropogenic SO$_2$ emissions, for the years 2014 and 1850, as prescribed in CMIP6 simulations. We calculated aerosol effective radiative forcing ($\Delta F_{\text{aer}}$) as the difference in top-of-the-atmosphere radiative fluxes between these two periods. We accounted for above-cloud aerosol in our calculation of the components of $\Delta F_{\text{aer}}$ (8) and aerosol-cloud interactions (9).

Carbonaceous aerosol from fossil fuel and residential sources match those used in CMIP6 in our early-industrial simulations. However, in our present-day simulations (2014 anthropogenic SO$_2$ emissions) we prescribed carbonaceous aerosol from biomass burning sources using emissions generated using Copernicus Atmospheric Monitoring Service Information (December 2016 to November 2017) (10) and spread these emissions between the surface and around 3km. We used emissions for the same period as prescribed wind fields, for the closest possible comparison to observed values. In our early-industrial simulations (1850 anthropogenic SO$_2$ emissions) we similarly scaled CMIP6 carbonaceous aerosol from biomass burning over model levels between the surface and around 3km.

We also prescribed, rather than simulated, sea surface temperatures and sea ice fraction for the December 2016 to November 2017 period. We prescribed land surface quantities, ocean surface concentrations of dimethylsulfide (DMS) and chlorophyll, and atmospheric concentrations of gas species (including oxidants OH and O$_3$, which we then perturb), using monthly mean output values from a fully-coupled version of the UKESM model, averaged over the 1979 to 2014 period. Additionally, we prescribe volcanic SO$_2$ emissions for continuously emitting and sporadically erupting volcanoes (11) and for explosive volcanic eruptions (12).

We use an N96 horizontal resolution, which is 1.875 $\times$ 1.25$^\circ$ (208 $\times$139 km) at the equator, with 85 vertical levels between the surface and 85km in altitude. Model vertical levels use a stretched grid such that the vertical resolution is around 13 m near the surface and around 150 to 200 m at the top of the boundary layer. We chose this resolution since it is the same as that used for long climate runs in CMIP6.

Aerosol number concentrations are treated prognostically with the GLOMAP multi-modal scheme (13, 14), which uses five log-normal aerosol size modes and includes sulfate, sea-salt, black carbon and organic carbon chemical components that are internally mixed within each size mode. Mineral dust is simulated separately using the CLASSIC dust scheme (15). GLOMAP simulates...
new particle formation, coagulation, gas-to-particle transfer, cloud processing and deposition of gases and aerosols. The activation of aerosols into cloud droplets is calculated using distributions of sub-grid vertical velocities based on available turbulent kinetic energy (16) and the removal of cloud droplets by autoconversion to rain is calculated by the host model. Aerosols are also removed by impaction scavenging of falling raindrops according to the collocation of clouds and precipitation (17, 18).

We modified some aspects of UKESM1-A in our PPE. Firstly, we define an ice mass fraction threshold (cloud_ice_thresh; table S1) above which no nucleation scavenging occurs, to allow sufficient aerosol to be transported to the Arctic (19). We assume that the wet scavenging of all aerosol particles (soluble and insoluble) is set to zero in large-scale raining clouds if the simulated ice to total water mass fraction is higher than this fixed value. This first structural change replicates the model change we implemented in (20) which is not yet in the release version of the model. We evaluated the climatic importance of this parameter as a cause of uncertainty in (21–24). Secondly, we implemented a version of look-up tables for aerosol optical properties (25) that includes optical properties for mineral dust (26) and higher-resolution increments of the imaginary part of the refractive indices, to better resolve the absorption coefficient of aerosols, especially at the low-absorption end of the spectrum. Finally, we included an organically-mediated boundary layer nucleation parametrisation (27) to enhance remote marine and early-industrial aerosol concentrations in the model.

**Perturbed Parameter Ensembles (PPEs)**

We created a new PPE of 221 UKESM1-A model simulations for this study. Each member of the PPE has a distinct combination of 37 aerosol and physical atmosphere parameter values, spanning expert elicited ranges (table S1). Parameters perturbed in previous PPEs using older versions of our model (20, 28) and identified as important causes of uncertainty in cloud active aerosol concentrations and/or aerosol forcing (21, 29, 30) are perturbed here, alongside parameters associated with structural model developments (5–7). Many parameters are described in table S1 as ‘scale factors’, which indicates we scaled the corresponding process parameter up or down over the indicated range. Other parameters are specific components of process parametrizations.

**Multi-stage experimental design**

We created our PPE in two stages, following ‘history matching’ conventions (31, 32). In the first stage, the 221 member ensemble was made by combining a simulation using median values for each parameter with 220 additional parameter combinations were drawn from a Latin hypercube optimized to ensure design points were distributed as evenly as possible across the 37-dimensional parameter space, using the ‘optimumLHS’ R function (33). We output monthly mean data for 4 months for each ensemble member and analysed output from the final month, which corresponded to anthropogenic emissions for May 2014 and horizontal wind fields for 2015.

We created statistical Gaussian process emulators (34) of multiple monthly mean output variables. For each variable, we sampled one million model variants (parameter combinations), from the corresponding emulator, that uniformly spanned the uncertain parameter space, in keeping with efforts to constrain aerosol radiative forcing uncertainty using large ensembles (22, 23). We then ruled out implausible parameter combinations that compared poorly to observations within known emulator uncertainty and assumed observational uncertainty bounds. Observations included global mean shortwave and longwave top-of-the-atmosphere radiative fluxes from the Clouds and the Earth’s Radiant Energy System experiment (35) and global mean precipitation amount from version 2 of the Global Precipitation Climatology Project (36). Additionally, we used North Pacific and North Atlantic marine only data between 10° and 60° N for low- and total-cloud fraction from the Moderate Resolution Imaging Spectroradiometer (37) and LWP from the Multi-Sensor
Advanced Climatology of Liquid Water Path data set (38). We assumed errors of 8%, 2%, 30%, 20%, 20% and 40% respectively for these observations. Of the retained model variants, we started with the parameter combination central to the retained space, then iteratively identified an additional 220 parameter combinations with the greatest Euclidean distance from existing points, until we had a new and diverse set of 221 members that span the uncertain parameter space retained from the first observational filter. Thus, the simulation to perturbed parameter ratio in our PPE is around six.

We created full year simulations forced using 1850 and 2014 anthropogenic aerosol and precursor gas emissions for each of these 221 PPE members. We held greenhouse gas concentrations constant at 2014 levels. These 221 model simulations are the second stage of our PPE creation process. As in the first stage, we created and validated statistical emulators of global mean and regional mean variables using these 221 members, then created a sample of one million model variants from these emulators.

Measurements

Measurements: Regional mean cloud and radiative properties

Satellites carry instruments that measure atmospheric properties, then geophysical quantities are calculated using retrieval algorithms and inverse modelling methods. We compare values derived from MODIS instruments (39) to model output calculated using the Cloud Feedback Model Intercomparison MODIS satellite simulator (40, 41) where available. This simulator minimizes errors in model comparisons to MODIS retrieval data, by recreating as near as possible what the satellite would retrieve given the model-simulated atmospheric conditions.

We use MODIS retrievals of of liquid water path (LWP), liquid cloud fraction ($f_c$), cloud optical depth ($\tau_c$) and cloud droplet effective radius ($r_e$) at 1° by 1° resolution and use $\tau_c$ and $r_e$ values to calculate cloud droplet number concentration ($N_d$). We assume constant $N_d$ throughout cloud layers, which is a good approximation for stratocumulus clouds (9, 42). We compare all cloud properties to satellite-simulator output and compare $N_d$ to values calculated at model–simulated cloud tops. We use outgoing top-of-the-atmosphere shortwave radiative flux ($F_{SW}$) measurements from the Clouds and the Earth’s Radiant Energy Systems instrument (35).

We degrade all satellite-derived measurements to match our model resolution, then identify regions with high cloud fraction across the year (table S2). We evaluate constraint variables at the regional level, since there are no clear relationships between aerosol forcing and observations of global mean values (SI Fig. S26). These regions are dominated by stratocumulus cloud, have relatively high multi-model diversity in cloud amount in CMIP6 models (43) and are the most important regions for understanding the role of aerosol-cloud interactions (44). We only used values corresponding to model grid boxes with at least 50% ocean coverage in our area-weighted regional mean calculations.

Measurements: Hemispheric difference in $N_d$

The contrast between marine $N_d$ in the polluted Northern Hemisphere and relatively pristine Southern Hemisphere ($H_d$) can act as a proxy for the difference in $N_d$ between the early-industrial
and present-day atmospheres (45). We calculate $H_d$ as the difference in hemispheric mean marine $N_d$ values, using MODIS $\tau_c$ and $r_e$ values.

**Measurements: Transects from stratocumulus- to cumulus-dominated regions**

We calculate the changes in multiple measurement values along transects from regions dominated by stratocumulus cloud to those dominated by cumulus. Cloud physical and radiative properties are sensitive to changes in aerosol concentrations in these transition regions (46). We chose transects on the Eastern side of major ocean basins (Fig. S12, table S3) where air is advected from the subtropics towards the equator. We used data from July 2017 for Northern Hemisphere transects and for November 2017 for Southern Hemisphere transects, to evaluate relatively strong transitions in warmer months.

The gradients of linear relationships between observed values and distances (in meters) along these transects are used as constraint variables. We evaluate gradients of individual measurement types including $N_d$, $r_e$, $f_c$ and LWP, calculated using values that informed our regional mean calculations. Additionally, we calculate gradients of aerosol index (AI; the total aerosol optical depth at 550 nm multiplied by the Ångström exponent) using MODIS aerosol optical depth retrieval data. We additionally include gradients of ratios of observation types along each transect as constraint variables. We calculate gradients of ratios using natural logarithms following (47). We include the ratios of $N_d$ to AI, $r_e$ to $N_d$, LWP to $N_d$ and $f_c$ to $N_d$. We compare satellite-derived values to probability distributions of corresponding output from our PPE members in SI Fig. S27-30.

**Relative Importance of Parameters**

For each constraint variable, we calculated the relative importance of parameters as causes of model uncertainty using Pearson partial correlations (48). Partial correlations control for the effects of all other perturbed parameters on the variable of interest in the correlation calculations. A partial correlation between some variable and a chosen parameter is the correlation between the residuals from a) linear regression of the variable on the remaining 36 parameters and b) linear regression of the chosen parameter on the remaining 36. For each of the 37 model parameters, we define the relative importance metric, for any chosen variable, as the proportion of its partial correlation with the chosen variable to the total of the 37 partial correlations, multiplied by the sign of the gradient of the linear regression of the variable on the parameter in question. We include the sign of the gradient in the relative importance metric to convey the effect of changing parameter values on the variable, which helps develop a process-based understanding of model behavior within the uncertainty framework. Relative importance metrics are used here as a guide to our choice of variables for model constraint and inform our understanding of how they relate to $\Delta F_{aer}$. Variance-based sensitivity analyses (49) can be used to robustly quantify the percentage of variance caused by each parameter. However, the multi-stage design of our PPE leaves gaps in the parameter space that limits the interpretability of variance-based methods. Therefore, we approximate the relative importance of parameters as causes of uncertainty using a method that is suited to our data structure and purpose. We calculate relative importance metrics using 1 million model variants for Fig. 2 and 221 PPE members for SI Fig. S1-11.

**Constraint process**

We identified over 450 constraint variables for consideration as potential constraints on the $\Delta F_{aci}$ component of $\Delta F_{aer}$. This total includes monthly mean values, annual means and seasonal
amplitudes of $H_d$ and regional mean constraint variables. Gradients along transects from stratocumulus to cumulus regions were also included as constraint variables.

We previously used ‘implausibility metrics’ that quantify the implausibility of each model variant with reference to an observed value, accounting for emulator uncertainty, observational uncertainty, inter-annual variability and representation errors (22, 23). Implausibility metrics were calculated for one million model variants across more than 9000 distinct measurements and we used these implausibility values to rule out model variants as observationally implausible if they did not compare well to the full set of observations. In practice, observations associated with relatively large uncertainties had little-to-no impact on ruling out model variants. Using this approach, we constrained $\Delta F_{\text{act}}$ and our parameter space, but could not readily isolate the role of individual constraint variables on the resulting $\Delta F_{\text{act}}$ constraint and could not quantify the efficacy of total constraint in terms of improved model skill, only in terms of reduced $\Delta F_{\text{act}}$ uncertainty range.

Here, we calculated root mean squared error (RMSE) values for every model variant in our one million member sample for each of the 450 plus constraint variables. For each constraint variable, we then normalized the one million RMSE values and ranked model variants according to their normalized RMSE (NRMSE) values, to identify which model variants we could rule out as observationally implausible. To avoid over-constraining our model, we set NRMSE values to zero where the uncertainty in our emulators was large relative to the difference between observed and emulated values. For this step, we defined the emulator uncertainty as the square root of the emulator variance at that specific combination of model parameters. In this way, individual constraints were stronger for constraint variables where parameter perturbations clearly defined the response surface of the associated statistical emulators.

We did not account for inter-annual variability because we ensured large-scale meteorological features of our model variants were very similar to observed conditions. We did not include (largely unquantified) observational errors in our constraint because we compared satellite data to model output from satellite simulators, which significantly reduced the importance of this source of uncertainty in observation to model comparisons. We also neglected the effects of representation errors (50) because they are unquantified for the satellite-derived observations used here. Instead, we compared mean values with stratocumulus-dominated regions to reduce the magnitude of these errors. Thus, observational and representation errors did not influence our method to identify which model variants to reject as implausible. Instead, we retained a proportion of model variants of the same order of magnitude as earlier constraint efforts that used constraint variables with more readily quantifiable sources of model-observation comparison uncertainty (22, 23). In this way, our method avoided over-constrain the model, yet allowed us to identify model structural inconsistencies without the masking effects of additional uncertainties.

For each of the 450 individual constraint variables we retained the 5000 model variants (0.5% of our original sample) with the lowest NRMSE values. However, the number of variants retained was larger than 5000 in many cases where the standard deviation from the associated emulator is larger than the difference in observed and emulated values (NRMSEs set to zero) for multiple model variants. For combinations of constraint variables, we calculated the average NRMSE value across all variables, for each model variant, prior to ranking and rejecting model variants with the highest average NRMSE values across variables. The number of constraint variables needed to optimally constrain $\Delta F_{\text{act}}$ in our structurally imperfect model was affected by the number of model variants retained (SI Fig. S25 and table S4) because reducing the efficacy of individual constraint variables affects the potential for additional observations to further reduce the $\Delta F_{\text{act}}$ uncertainty. However, the strength of constraint (quantified as a reduction in the 90% credible interval) was largely unaffected by the number of model variants retained at each step. The constraint was improved by only around 4 percentage points when we significantly increased the constraint criteria to retain 1k variants at each step (rather than 5k, as in the main article), and decreased by only around 3
percentage points when we significantly relaxed our constraint criteria to retain 20k variants. The constrained $\Delta F_{ac}$ bounds were largely unaffected by the number of variants retained, shifting by only around 0.1 W m$^{-2}$ (table S4). Thus, the choice of retaining 5k model variants at each step was arbitrary and did not affect our interpretation of results.

We removed constraint variables from our constraint process where the associated emulator average standard deviation across our sample of points was larger than the standard deviation of emulated values. That is, we discounted constraint variables where the emulator uncertainty was larger than the changes in the emulated response surface. This was the case for a small number of transect constraint variables and for the seasonal amplitude of $f_c$ in the Southern Ocean. Additionally, we removed transect measurements from our set of constraint variables where the observed values were outside of the 90% credible interval of corresponding values in our sample, since such discrepancies are indicative of structural model inadequacies and/or unaccounted for observational errors (SI Fig. S27-30).

In each region we identified a subset of constraint variables as being pairwise consistent with $N_d$. Individual monthly mean $N_d$ values in each region were used to identify which other constraint variables could be considered pairwise consistent. The months used were September, October, December, March and the annual mean for the North Atlantic, North Pacific, South Atlantic, South Pacific and Southern Ocean respectively. In these months, $N_d$ was determined to be most consistent on average with $N_d$ in other months, as represented by the effect on average NRMSE in the associated constraint (Fig. 4 and SI Fig. S18-21). At this stage, we assumed constraint variables that are consistent with $N_d$ in these specific months in these regions were also consistent with $N_d$ (and other selected constraint variables) in other regions. Our strategy here was to rule out constraint variables that are clearly inconsistent, rather than to assure internal consistency between all remaining constraint variables. Across all regions, we retained 225 constraint variables (out of more than 450) which we considered consistent with $N_d$.

We identified an optimal set of constraint variables by first identifying the individual constraint variable with the greatest impact on $\Delta F_{ac}$ uncertainty (our target model variable), then progressively added constraint variables that most improved the overall constraint. We continued to add constraint variables to the optimal set, that weakened the $\Delta F_{ac}$ constraint the least, in case our constraint was a local maximum. The effects on $\Delta F_{ac}$ uncertainty are shown in Fig. 5 of the main article. At each of the more than twenty thousand steps in this process, we evaluated the average NRMSE values for each of the one million model variants, for every possible additional constraint. The blue and purple lines in Fig. 5 are synthetic examples of how our constraint may be improved with fewer, or no, remaining structural model inadequacies. These values used to create these lines are chosen to exemplify our point and do not correspond to actual constraints of our model.

The order these constraint variables were chosen may affect the outcome. That is, a stronger constraint may have been achieved using a different set of ‘optimal’ constraint variables. However, we could not calculate NRMSE values for one million model variants across all possible combinations of 225 consistent constraint variables. Instead, we tested the effect of starting with all 225 consistent constraint variables and progressively removing one variable at a time. This is the most distinct test of reordering the constraint variables, from the method we used in the main article. This approach yielded a similar constraint on $\Delta F_{ac}$ as achieved by progressively adding constraint variables (90% CI between -1.4 and -0.2 W m$^{-2}$, or -1.2 to -0.0 W m$^{-2}$ depending on which local maxima is used) and very similar constraints on marginal parameter distributions (equivalent to SI Fig. S23, 24). These tests revealed there are multiple ways to combine sets of consistent constraint variables to achieve a similar constraint on $\Delta F_{ac}$, highlighting the degree of redundancy in using multiple observations of the same variable for constraint.
Fig. S1. Relative importance of model parameters as causes of uncertainty in $H_d$. Relative importance metrics are calculated for each month (December 2016 to November 2017), for the annual mean (Ann) and the seasonal amplitude (Amp). Relative importance metrics lower than 4% are not shown.
Fig. S2. Relative importance of model parameters as causes of uncertainty in global mean $N_d$. Figure features are identical to Fig. S1.
**Fig. S3.** Relative importance of model parameters as causes of uncertainty in global mean $F_{SW}$. Figure features are identical to Fig. S1.
Fig. S4. Relative importance of model parameters as causes of uncertainty in global mean $f_c$. Figure features are identical to Fig. S1.
**Fig. S5.** Relative importance of model parameters as causes of uncertainty in global mean LWP. Figure features are identical to Fig. S1.
**Fig. S6.** Relative importance of model parameters as causes of uncertainty in global mean $\tau_c$. Figure features are identical to Fig. S1.
**Fig. S7.** Relative importance of model parameters as causes of uncertainty in global mean $r_e$. Figure features are identical to Fig. S1.
Fig. S8. Relative importance of model parameters as causes of uncertainty in North Atlantic transect constraint variables. Figure features are identical to Fig. S1.
Fig. S9. Relative importance of model parameters as causes of uncertainty in North Pacific transect constraint variables. Figure features are identical to Fig. S1.
Fig. S10. Relative importance of model parameters as causes of uncertainty in South Atlantic transect constraint variables. Figure features are identical to Fig. S1.
Fig. S11. Relative importance of model parameters as causes of uncertainty in South Pacific transect constraint variables. Figure features are identical to Fig. S1.
Fig. S12. Transects from stratocumulus to cumulus cloud dominated regions in a) July and b) November, superimposed on MODIS liquid cloud fraction values for the corresponding month.
Fig. S13. Median and standard deviations of annual mean $\Delta F_{aer}$, $\Delta F_{aci}$, and $\Delta F_{ari}$, across the 221 PPE members. Values are calculated in each model grid box independently.
Fig. S14. Probability distributions of North Pacific regional mean output from our sample of model variants, satellite-derived measurements and the default UKESM1-A model, for individual months spanning December 2016 to November 2017 and the annual mean.
Fig. S15. Probability distributions of South Atlantic regional mean output from our sample of model variants, satellite-derived measurements and the default UKESM1-A model, for individual months spanning December 2016 to November 2017 and the annual mean.
Fig. S16. Probability distributions of South Pacific regional mean output from our sample of model variants, satellite-derived measurements and the default UKESM1-A model, for individual months spanning December 2016 to November 2017 and the annual mean.
Fig. S17. Probability distributions of Southern Ocean regional mean output from our sample of model variants, satellite-derived measurements and the default UKESM1-A model, for individual months spanning December 2016 to November 2017 and the annual mean.
Fig. S18. Pairwise comparisons of North Pacific and $H_d$ constraint variables. Figure features are identical with Fig. 3.
**Fig. S19.** Pairwise comparisons of South Atlantic and $H_d$ constraint variables. Figure features are identical with Fig. 3.
**Fig. S20.** Pairwise comparisons of South Pacific and $H_d$ constraint variables. Figure features are identical with Fig. 3.
Fig. S21. Pairwise comparisons of Southern Ocean and $H_d$ constraint variables. Figure features are identical with Fig. 3.
**Fig. S22.** Probability density functions for global, annual mean a) $\Delta F_{aer}$, b) $\Delta F_{aci}$ and c) $\Delta F_{ari}$ in the original one million member sample and after optimal constraint.
Fig. S23. Probability density functions of model parameters after constraint using our optimal set of constraint variables. In the original sample of 1 million model variants, these pdfs would be uniformly distributed on this scale. Non-shaded sections indicate a proportion of model variants with corresponding parameter values have been ruled out as implausible.
Fig. S24. Probability density functions of model parameters after constraint using our optimal set of constraint variables. Features are identical to Fig. S23.
Fig. S25. Constraint of $\Delta F_{\text{act}}$ and the effect of varying the number of constraint variables used and the number of model variants retained (percentage of original 1 million) at each stage of the constraint (legend). The constraints achieved by retaining 5000 model variants at each stage is identical to the constraints shown in Fig. 5.
Fig. S26. Density plots of global, annual mean output from 221 PPE members for $\Delta F_{\text{aer}}$, $\Delta F_{\text{aci}}$, and global mean $F_{SW}$, LWP, $\Delta F_{\text{ari}}$ and $N_d$, $f_c$, $\tau_c$, and $r_e$. Diagonal panels show probability density functions for individual variables.
Fig. S27. Probability density functions of North Atlantic transect constraint variables.
Fig. S28. Probability density functions of North Pacific transect constraint variables.
**Fig. S29.** Probability density functions of South Atlantic transect constraint variables.
Fig. S30. Probability density functions of South Pacific transect constraint variables.
Table S1. Parameters perturbed in our PPE, the ranges they were perturbed over and default values as prescribed in the release version of the model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Default</th>
<th>Parameter Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bl_nuc</td>
<td>0.1</td>
<td>10</td>
<td>1</td>
<td>Boundary layer nucleation rate scale factor</td>
</tr>
<tr>
<td>ait_width</td>
<td>1.2</td>
<td>1.8</td>
<td>1.59</td>
<td>Modal width of Aitken modes (nm)</td>
</tr>
<tr>
<td>cloud_ph</td>
<td>1e-7</td>
<td>2.51e-5</td>
<td>1e-5</td>
<td>Cloud droplet pH</td>
</tr>
<tr>
<td>carb_ff_diam</td>
<td>30</td>
<td>90</td>
<td>60</td>
<td>Emission diameter of carbonaceous aerosol from fossil fuel sources (nm)</td>
</tr>
<tr>
<td>carb_bb_diam</td>
<td>90</td>
<td>300</td>
<td>110</td>
<td>Emission diameter of carbonaceous aerosol from biomass burning sources (nm)</td>
</tr>
<tr>
<td>carb_res_diam</td>
<td>90</td>
<td>500</td>
<td>150</td>
<td>Emission diameter of carbonaceous aerosol from residential sources (nm)</td>
</tr>
<tr>
<td>prim_so4_diam</td>
<td>3</td>
<td>100</td>
<td>150</td>
<td>Emission diameter of 50% of new sub-grid sulfate particles (nm). Remaining 50% emitted into the larger coarse mode (nm)</td>
</tr>
<tr>
<td>sea_spray</td>
<td>0.25</td>
<td>4</td>
<td>1</td>
<td>Sea spray emission flux scale factor</td>
</tr>
<tr>
<td>anth_so2</td>
<td>0.6</td>
<td>1.5</td>
<td>1</td>
<td>Anthropogenic SO2 emission flux scale factors. Applied independently to European, North American, Chinese, Asian regions and the rest of the world</td>
</tr>
<tr>
<td>volc_so2</td>
<td>0.71</td>
<td>2.38</td>
<td>1</td>
<td>Volcanic SO2 emission flux scale factor</td>
</tr>
<tr>
<td>bvoc_soa</td>
<td>0.32</td>
<td>3.68</td>
<td>1</td>
<td>Biogenic monoterpine production rate of secondary organic aerosol scale factor</td>
</tr>
<tr>
<td>dms</td>
<td>0.33</td>
<td>3</td>
<td>1</td>
<td>Dimethyl-sulfide emission flux scale factor</td>
</tr>
<tr>
<td>prim_moc</td>
<td>0.4</td>
<td>6</td>
<td>1</td>
<td>Primary marine organic carbon emission flux scale factor</td>
</tr>
<tr>
<td>dry_dep_ait</td>
<td>0.5</td>
<td>2</td>
<td>1</td>
<td>Dry deposition velocity of Aitken mode aerosol</td>
</tr>
<tr>
<td>dry_dep_acc</td>
<td>0.1</td>
<td>10</td>
<td>1</td>
<td>Dry deposition velocity of accumulation mode aerosol</td>
</tr>
<tr>
<td>dry_dep_so2</td>
<td>0.2</td>
<td>5</td>
<td>1</td>
<td>Dry deposition velocity of SO2</td>
</tr>
<tr>
<td>kappa_oc</td>
<td>0.2</td>
<td>0.65</td>
<td>0.65</td>
<td>Hygroscopicity parameter κ for organic aerosol – affects wet diameter and clear-sky radiative flux</td>
</tr>
<tr>
<td>sig_w</td>
<td>0.25</td>
<td>1.75</td>
<td>1</td>
<td>Standard deviation of shallow-cloud updraft velocity scale factor</td>
</tr>
<tr>
<td>rain_frac</td>
<td>0.3</td>
<td>0.7</td>
<td>0.3</td>
<td>Fraction of cloud covered area where rain forms</td>
</tr>
<tr>
<td>cloud_ice_thresh</td>
<td>0.1</td>
<td>0.5</td>
<td>N/A</td>
<td>Threshold of cloud ice water fraction for scavenging</td>
</tr>
<tr>
<td>conv_plume_scav</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>Scavenging efficiency (as a fraction of total aerosol removed) of Aitken mode aerosol in convective clouds</td>
</tr>
<tr>
<td>bc_ri</td>
<td>0.2</td>
<td>0.8</td>
<td>0.565</td>
<td>Imaginary part of the black carbon refractive index</td>
</tr>
<tr>
<td>oxidant_oh</td>
<td>0.7</td>
<td>1.3</td>
<td>1</td>
<td>Offline oxidant OH concentration scale factor</td>
</tr>
<tr>
<td>Description</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>oxidants_o3</td>
<td>0.7</td>
<td>1.3</td>
<td>1</td>
<td>Offline oxidant O₃ concentration scale factor</td>
</tr>
<tr>
<td>bparam</td>
<td>-0.15</td>
<td>-0.13</td>
<td>-0.14</td>
<td>Coefficient of the spectral shape parameter β for effective radius</td>
</tr>
<tr>
<td>two_d_fsd_factor</td>
<td>1</td>
<td>2</td>
<td>1.4</td>
<td>Scale factor for the 2D relationship between cloud condensate variance, cloud cover and convection. Controls sub-grid cloud heterogeneity</td>
</tr>
<tr>
<td>c_r_correl</td>
<td>0</td>
<td>1</td>
<td>0.9</td>
<td>Cloud and rain sub-grid horizontal spatial colocation</td>
</tr>
<tr>
<td>autoconv_exp_lwp</td>
<td>2.15</td>
<td>3.31</td>
<td>2.47</td>
<td>Exponent of liquid water path in the power law for initiating autoconversion</td>
</tr>
<tr>
<td>autoconv_exp_nd</td>
<td>-3</td>
<td>-1</td>
<td>-1.79</td>
<td>Exponent of cloud droplet concentration (N_d) in the power law for initiating autoconversion</td>
</tr>
<tr>
<td>dbsdtbd_turb_0</td>
<td>0</td>
<td>1e-3</td>
<td>1.5e-4</td>
<td>Cloud erosion rate (s⁻¹)</td>
</tr>
<tr>
<td>ai</td>
<td>0</td>
<td>5e-2</td>
<td>2.57e-2</td>
<td>Scaling coefficient for the dependence of ice mass on diameter</td>
</tr>
<tr>
<td>m_ci</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>Ice fallspeed scale factor</td>
</tr>
<tr>
<td>a_ent_1_rp</td>
<td>0</td>
<td>0.5</td>
<td>0.23</td>
<td>Cloud top entrainment rate scale factor</td>
</tr>
</tbody>
</table>
Table S2. Regions of persistent stratocumulus cloud used to calculate regional mean constraint variables.

<table>
<thead>
<tr>
<th>Region</th>
<th>Latitude range</th>
<th>Longitude range</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Atlantic</td>
<td>34.4° to 54.4° N</td>
<td>329.1° to 347.8° E</td>
</tr>
<tr>
<td>North Pacific</td>
<td>14.4° to 48.1° N</td>
<td>197.8° to 231.6° E</td>
</tr>
<tr>
<td>South Atlantic</td>
<td>30.6° to 10.6° S</td>
<td>347.8° to 2.8° E</td>
</tr>
<tr>
<td>South Pacific</td>
<td>30.6° to 15.6° S</td>
<td>254.1° to 284.1° E</td>
</tr>
<tr>
<td>Southern Ocean</td>
<td>30.6° to 50.6° S</td>
<td>0° to 360° E</td>
</tr>
</tbody>
</table>
Table S3. Transects from stratocumulus- to cumulus-dominated regions.

<table>
<thead>
<tr>
<th>Region</th>
<th>Start position</th>
<th>End position</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Atlantic</td>
<td>54.4° N, 336.6° E</td>
<td>45.6° N, 330.9° E</td>
</tr>
<tr>
<td>North Pacific</td>
<td>30.6° N, 229.7° E</td>
<td>19.4° N, 227.8° E</td>
</tr>
<tr>
<td>South Atlantic</td>
<td>11.9° S, 357.2° E</td>
<td>11.9° S, 345.9° E</td>
</tr>
<tr>
<td>South Pacific</td>
<td>20.6° S, 282.2° E</td>
<td>15.6° S, 269.1° E</td>
</tr>
</tbody>
</table>
Table S4. Effect of varying the number of model variants retained at each stage of constraint. We show the number of measurements needed to optimally constrain $\Delta F_{aer}$ and the 90% CI in each case.

<table>
<thead>
<tr>
<th>Number of model variants retained</th>
<th>Number of measurements used</th>
<th>Lower, negative $\Delta F_{aer}$ bound</th>
<th>Upper $\Delta F_{aer}$ bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>27</td>
<td>-1.15</td>
<td>-0.07</td>
</tr>
<tr>
<td>2000</td>
<td>31</td>
<td>-1.23</td>
<td>-0.10</td>
</tr>
<tr>
<td>5000</td>
<td>13</td>
<td>-1.26</td>
<td>-0.13</td>
</tr>
<tr>
<td>10000</td>
<td>29</td>
<td>-1.30</td>
<td>-0.13</td>
</tr>
<tr>
<td>20000</td>
<td>15</td>
<td>-1.33</td>
<td>-0.13</td>
</tr>
</tbody>
</table>
**SI References**


10. GFAS, CAMS global biomass burning emissions based on fire radiative power (GFAS): data documentation, https://confluence.ecmwf.int/display/CKB/CAMS+global+biomass+burning+emissions+based+on+fire+radiative+power+%28GFAS%29%29+A+data+documentation, (2022).


