https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

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

Leighton A. Regayre[1,2], Lucia Deaconu[3,4], Daniel P. Grosvenor[1,2,5], David M. H. Sexton[2], Christopher Symonds[5], Tom Langton[3], Duncan Watson-Paris[3], Jane P. Mulcahy[2], Kirsty J. Pringle[1,5,6], Mark Richardson[5], Jill S. Johnson[1,7], John W. Rostron[2], Hamish Gordon[1,8], Grenville Lister[9,10], Philip Stier[3] and Ken S. Carslaw[1]

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

Correspondence to: Leighton A. Regayre (L.A.Regayre@leeds.ac.uk)

Research Article

Keywords: Climate change, aerosol, clouds, radiative forcing, uncertainty, constraint, observations, structural error

Abstract: Aerosol radiative forcing uncertainty affects estimates of climate sensitivity and limits model skill at making climate projections. Efforts to improve the representations of physical processes in climate models, including extensive comparisons with observations, have not significantly constrained the range of possible aerosol forcing values. A far stronger constraint, in particular for the lower (most-negative) bound, can be achieved using global mean energy-balance arguments based on observed changes in historical temperature. Here, we show that structural deficiencies in a climate model, revealed as inconsistencies among observationally constrained cloud properties, limit the effectiveness of observational constraint of the uncertain physical processes. We sample uncertainty in 37 model parameters related to aerosols, clouds and radiation in a perturbed parameter ensemble of the UK Earth System Model and evaluate one million model variants (different parameter settings from Gaussian Process emulators) against satellite-derived observations over several cloudy regions. We show it is possible to

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

reduce the parametric uncertainty in global mean aerosol forcing by more than 50% to a range in close agreement with energy-balance constraints (around -1.3 to -0.1 W m^{-2}). However, incorporating observations associated with model inconsistencies weakens the constraint because the inconsistencies introduce conflicting information about relationships between model parameter values and aerosol forcing. Our estimated aerosol forcing range is the maximum feasible constraint using these observations and our structurally imperfect model. Structural model developments, targeted at the inconsistencies identified here, would enable a larger set of observations to be used for constraint, which would then narrow the uncertainty further.

.

Significance Statement: Aerosol forcing of Earth's energy balance has persisted as a major cause of uncertainty in historical climate simulations over several generations of climate model development. Here we show that potential structural deficiencies in a climate model can be exposed by comprehensively exploring its parametric uncertainty, and that these deficiencies limit how much the model uncertainty can be reduced through observational constraint. Combined consideration of parametric and structural uncertainties provides a future pathway towards building models that have greater physical realism and lower uncertainty.

Introduction

The most uncertain component of human forcing of the climate system over the industrial period is aerosol effective radiative forcing (ΔF_{aer}) (1). Uncertainty in historical ΔF_{aer} reduces our ability to confidently project near-term future changes to our climate (2–5). The best estimate of ΔF_{aer} based on current understanding of aerosols, clouds, radiation and their interactions (informed by results from global climate models and analysis of observations) ranges from -3.2 to -0.4 W m^{-2} (6). The magnitude of ΔF_{aer} has remained uncertain through all Intergovernmental Panel on Climate Change assessment reports (1), despite decades of research to improve our scientific understanding of the key processes and abundant observations with which to test models.

The lower (most-negative) bound on ΔF_{aer} is more tightly constrained by global mean energy balance arguments, which infer the magnitude indirectly based on historical emissions and changes in global mean surface temperature. Such studies suggest the lower bound may be around -1.8 to -1.7 W m^{-2} (7–9). Evidence for a weaker (less negative) lower bound on ΔF_{aer} comes from energy-balance relationships that are additionally informed by output from global climate model ensembles. For example, (10) constrain the ΔF_{aer} lower bound to around -1.5 W m^{-2} and (11) constrain the lower bound to between -1.8 and -1.2 W m^{-2}. However, tight constraint of just the

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

magnitude of historical and future global mean ΔF_{aer} does not produce a climate model that can be used to explore the full range of regional and global climatic effects. Thus, although energy-balance constraints can set the plausible bounds on historical global mean ΔF_{aer}, we also need a "process-based" approach of building reliable global climate models that can accurately simulate the observed state and behaviour of aerosols, clouds and radiation that determine the regional patterns of aerosol effects on climate.

A process-based constraint of ΔF_{aer} is a substantial undertaking, with many steps involved. It relies mainly on using complex climate models to simulate the underlying physical processes that affect changes in aerosols, clouds and radiation (and hence ΔF_{aer}), then settling on models that have been developed and refined to achieve acceptable agreement with extensive observations of these atmospheric properties and trends. It is assumed that good agreement of a model with observations ensures that the model is able to make trustworthy estimates of ΔF_{aer}, which cannot itself be observed. Yet, the process-based uncertainty range has remained far wider than estimates from energy balance approaches because models simulate a very large number of poorly understood processes that can affect the magnitude of ΔF_{aer} (12–15).

A further challenge in process-based constraint is that the range of ΔF_{aer} stems from two sources of uncertainty in climate models, structural uncertainty and parametric uncertainty. Structural deficiencies in a model are associated with coding choices related to spatial resolution, numerical methods, parametrisation schemes, and neglected processes. Model developments attempt to reduce these deficiencies and the biases they cause compared to observations, and multi-model intercomparisons (16, 17) can be used to estimate a range of ΔF_{aer} across sets of structurally different models (structural uncertainty). Within a particular model, the uncertain parameters in the process equations cause an additional uncertainty in simulations of ΔF_{aer} (parametric uncertainty). Adjustment of parameter values, or tuning, is performed during and/or following model development to further improve the goodness-of-fit to observations (18), although it is recognised that well-tuned models still have a large (and usually unquantified) parametric uncertainty (19). Perturbed parameter ensembles (PPEs) of the kind we use here (see Methods) are a substantial extension of normal model tuning that explore many combinations of parameter values across their likely uncertainty ranges and quantify their combined effects on ΔF_{aer} (12, 15, 20). The resulting unconstrained parametric uncertainty in ΔF_{aer}, from sampling all important sources of uncertainty in our model, is larger than the range based on energy balance constraints and approximately as wide as the multi-model range, suggesting that parametric uncertainties in ΔF_{aer} are as important as structural model differences.

Separation of structural and parametric sources of model uncertainty is important because they have different remedies. Structural uncertainties point to model deficiencies that require model developments, while parametric uncertainties can be reduced by matching the outputs of many model variants (parameter combinations) to historical

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

observations through a process called 'history matching' (21–25). There is currently no best practice for accounting for and separating the effects of structural and parametric uncertainties (24, 26–29). In particular, the observational constraint of parametric uncertainty cannot be cleanly separated from the effects of structural uncertainties. This is because, without accounting for potential (usually unquantified) structural errors, it may not be possible to find any parameter combinations that produce a model that is consistent with all target observations. Therefore, it is common to add a structural error term during model-observation comparison, which effectively inflates the parametric uncertainty and the overall model uncertainty to accommodate the structural errors. This approach avoids overfitting and provides an estimate of the uncertainty in ΔF_{aer} that broadly accounts for both sources of uncertainty. However, it does not provide any information about which processes cause structural model errors, nor how they weaken the constraint of parameter values and hence constraint of ΔF_{aer}

To maximise ΔF_{aer} constraint, we need to address three key challenges. First, we need to densely sample the model parametric uncertainty related to the multitude of cloud, aerosol and physical atmosphere processes that determine ΔF_{aer}. Secondly, we need to identify model variables that share causes of uncertainty with ΔF_{aer}, to prioritise associated observations for use in the constraint process (12, 25). The final challenge is to ensure any constraint on ΔF_{aer} is consistent across multiple observation types and/or quantify the limiting effect of any internal model inconsistencies on ΔF_{aer} constraint. Here, we tackle these challenges using one million variants of the UKESM1-A model (based on statistical emulators trained on output from 221 model simulations) that sample uncertainty caused by 37 aerosol, cloud and physical atmosphere model parameters (see Methods; table S1). We evaluate the causes of uncertainty in cloud properties over stratocumulus-dominated regions and observationally constrain ΔF_{aer} using tens of thousands of combinations of more than 450 satellite-derived values. This approach exposes previously hidden structural inconsistencies related to representations of cloud properties. We remove variables associated with these inconsistencies from the constraint process to produce an internally consistent constraint on ΔF_{aer}. This constraint does not make use of all available observations, so we argue that in a model free from structural inconsistencies, our approach could constrain $\Delta F_{aer,}$ and associated process uncertainties even further.

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

Results

Sampling uncertainty in ΔF_{aer}

Industrial period ΔF_{aer} ranges from around -3.5 to 3.0 W m^{-2} in our set of one million UKESM1-A model variants (see Methods), with a 90% credible interval of -1.8 to 0.9 W m^{-2} (Fig. 1). This unconstrained 90% credible range (2.7 W m^{-2}) is as wide as the credible range (2.8 W m^{-2}) based on an in-depth review of evidence from models and observations related to aerosol-cloud and aerosol-radiation interactions (6), and therefore spans a wide spectrum of model behaviour. Our range includes positive ΔF_{aer} values that stem from positive forcing contributions from aerosol-cloud interactions (ΔF_{aci}) and aerosol-radiation interactions (ΔF_{ari}), which the Bellouin review discounts (6). These positive ΔF_{aci} and ΔF_{ari} values arise from individually plausible parameter values that produce seemingly implausible model output when combined. We expect the associated model variants to be amongst the first we will rule out as observationally implausible.

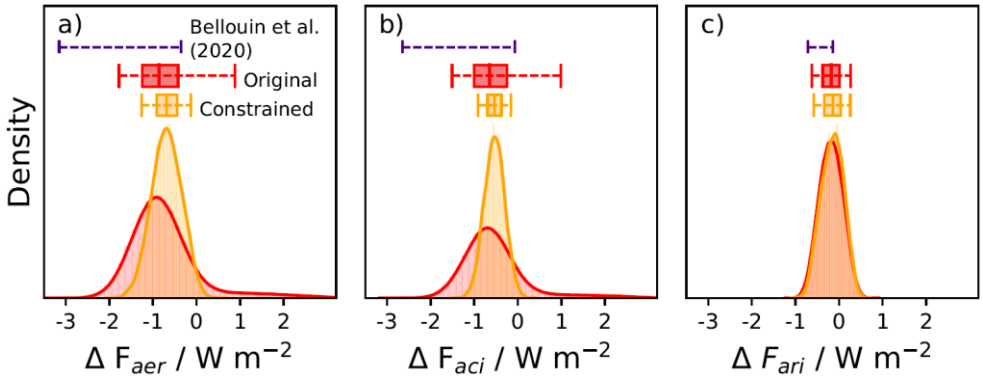

Figure 1. Probability density functions for global, annual mean effective radiative forcings from 1850 to 2014. a) ΔF_{aer}, b) ΔF_{aci} and c) ΔF_{ari} in the original one million member sample and after optimal constraint (see Optimal constraint of aerosol forcing). Box plots show the 5th, 25th, 50th, 75th and 95th percentiles. The 5th and 95th percentiles from (6) are also shown.

Shared causes of uncertainty imply potential for constraint

Our aim is to constrain ΔF_{aer} as tightly as possible using a set of observations that constrain all processes and associated model parameters that cause ΔF_{aer} uncertainty. One way to prioritise which observations to use for constraint is to quantify the overlap in causes of uncertainty between ΔF_{aer} and model variables associated with the observations. Figure 2 shows global mean ΔF_{aer} is sensitive to around 10 model parameters (see Methods and table S1). Here, we prioritise the constraint of global mean ΔF_{aer} because it is the quantity most commonly used to inform

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

policy decisions (1). We are particularly motivated to constrain processes that cause uncertainty in ΔF_{aci}, since it is the larger and more uncertain component of ΔF_{aer} (Fig. 1) and the ΔF_{ari} component can be more readily constrained using available aerosol observations (24, 30). Thus, we seek model variables that share causes of uncertainty with global mean ΔF_{aer} and ΔF_{aci}. Sharing causes of uncertainty (or parameter sensitivity) with ΔF_{aer} is a necessary, but not sufficient, condition for constraint (19). Model variables and ΔF_{aer} must also share parameter dependencies (responses to high-dimensional parameter combinations). It is highly unlikely that any one model variable will share exactly the same set of dependencies on uncertain model parameters with ΔF_{aer} and ΔF_{aci} (19, 25). Thus, to constrain model uncertainty we anticipate needing multiple observations that share some causes of uncertainty and parameter dependencies with ΔF_{aer} and ΔF_{aci}.

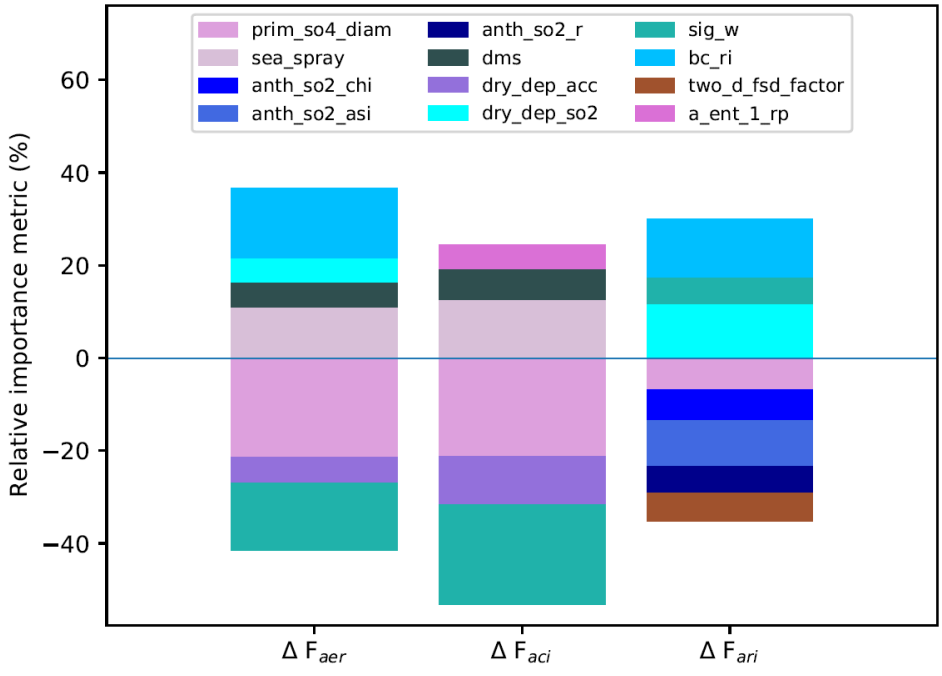

Figure 2. Relative importance of model parameters as causes of uncertainty in global mean ΔF_{aer} and its components ΔF_{aci} and ΔF_{ari}. Only parameters with a relative importance of 5% or larger are shown. Positive values correspond to parameters where increasing the parameter value causes median values of ΔF_{aer}, ΔF_{aci} or ΔF_{ari} to become weaker (less negative) across the set of one million model variants.

We evaluate the potential of several types of observations, related to clouds and aerosol-cloud interactions in multiple locations and times of the year, to serve as global mean ΔF_{aer} constraints and refer to them collectively as 'constraint variables'. We use satellite-derived observations over 5 regions of persistent stratocumulus cloud

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

(Methods) known to be key regions for interrogating processes that affect aerosol-cloud interactions (31, 32). To find a set of useful constraint variables (i.e. a set that collectively share causes of uncertainty and parameter dependencies with ΔF_{aer}), we evaluate a diverse set of constraint variables, with the aim to include as many distinctly informative variables as possible and thus account for all ΔF_{aer} parameter sensitivities and dependencies. We do not explicitly quantify 37-dimensional parameter dependencies for each constraint variable. Causes of uncertainty are likely to vary regionally and seasonally (13), so observations of the same type over multiple regions and months may all inform the constraint, but with some redundancy.

We use satellite-derived observations of outgoing top-of-the-atmosphere radiative flux (F_{SW}), cloud drop number concentrations at cloud top (N_d), and bulk cloud properties of cloud fraction (f_c), optical depth (τ_c), liquid water path (LWP) and droplet effective radius (r_e) from Moderate Resolution Imaging Spectroradiometer (33). Model output is designed to match MODIS retrievals by calculating bulk cloud properties using the Cloud Feedback Model Intercomparison MODIS satellite (COSP) simulator (34, 35). The constraint variables are defined as monthly mean, annual mean, or seasonal amplitude (difference between maximum and minimum monthly mean values) over a region of persistent stratocumulus cloud (SI table S2). So, for one observation type (e.g., N_d) we have 70 constraint variables (12 months, annual mean and seasonal amplitude, all over 5 regions). In our set of one million model variants, these constraint variables share causes of uncertainty (SI Fig. S1-S11), so in practice only a subset may be needed to constrain ΔF_{aer} and others will effectively be redundant. We also include hemispheric differences in marine N_d (H_d; annual and monthly means, and seasonal amplitude), which is a useful proxy for cloud albedo changes over the industrial period that has been shown to strongly constrain ΔF_{aci} (36). In addition to these state variables, we also use changes in aerosol and cloud properties and relationships between these properties along transects from stratocumulus cloud dominated regions to cumulus-dominated regions (Methods; SI Fig. S12 and table S3) and refer to these as transect variables. Relationships between aerosol and cloud properties can affect the magnitude of ΔF_{aci} (37). We use the strength of mean relationships along transects rather than a set of point observations, so we can evaluate the effect of these relationships on ΔF_{aer} under similar meteorological conditions. Although meteorological covariability (changes induced in both variables by shared meteorological drivers) may limit the direct interpretability of such relationships (38), the transect variables were chosen to inform ΔF_{aer}, and for that purpose it is only the shared causes of uncertainty and parameter dependencies that matter (Fig. 2 and SI Fig. S8-11).

In total, we have more than 450 constraint variables spanning monthly mean, annual mean and seasonal amplitudes, for global mean H_d and 6 observation types (F_{SW}, N_d, f_c, LWP, τ_c and r_e) across 5 stratocumulus-

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

dominated regions (state variables), along with 9 constraint variables for changes in aerosol and cloud properties, and their relationships, each along 4 transects from stratocumulus- to cumulus-dominated regions (transect variables).

First, we evaluate the causes of uncertainty in ΔF_{aer} and its components (Fig. 2). There is substantial overlap between the parametric causes of uncertainty in ΔF_{aer} and ΔF_{aci}, with the parameter that controls the diameter of newly formed accumulation mode sulfate particles ('prim_so4_diam') causing the largest amount of uncertainty. Increasing the value of this parameter increases the diameter of newly emitted sulfate particles and thus decreases the number of particles emitted (for fixed emission mass flux), which makes ΔF_{aci} more negative (stronger) on average since larger particles are more likely to act as cloud condensation nuclei. Any constraint that rules out the most positive ΔF_{aci} values will likely constrain newly formed sulfate particles towards higher diameters.

Other key causes of ΔF_{aer} uncertainty include the parameters controlling sub-grid updraft velocities (sig_w), emission fluxes of sea spray aerosol (sea_spray) and dimethyl-sulfide (dms), the dry deposition removal rate of accumulation mode aerosol (dry_dep_acc), the refractive index controlling carbonaceous aerosol radiative properties (bc_ri) and anthropogenic SO_2 emission fluxes (anth_so2_r). The physical atmosphere parameter controlling cloud top entrainment (a_ent_1_rp) affects ΔF_{aci} uncertainty and the parameter controlling sub-grid cloud heterogeneity (two_d_fsd_factor) affects ΔF_{ari}. However, in contrast with previous PPE analyses of this kind (15, 20), no physical atmosphere parameters feed through to causes of global mean ΔF_{aer}, likely due to model structural developments related to clouds and radiation (39, 40).

We understand how the key causes of uncertainty affect ΔF_{aci} in the model. Increasing the value of the updraft parameter increases N_d, particularly in the present-day atmosphere with relatively high cloud condensation nuclei concentrations where droplet activation is limited by vertical velocity. Thus, increasing the value of the updraft parameter makes median ΔF_{aci} more negative (stronger) by increasing cloud albedo, particularly in the relatively polluted present-day atmosphere. The influence of natural emission flux parameters on ΔF_{aci} uncertainty is well established (12). Increasing sea spray or dimethyl-sulfide emission fluxes makes global mean ΔF_{aci} less negative (weaker) on average by increasing the background aerosol concentration and thus reducing the sensitivity of cloud albedo to anthropogenic aerosol. The removal rate of accumulation mode aerosol similarly affects background aerosol concentrations. These three parameters also influence present-day N_d in relatively low anthropogenic aerosol environments such as the Southern Ocean (41) so can be collectively constrained using appropriate observations (25). However, compensating errors in aerosol emission fluxes and removal rates moderate our ability to constrain these parameters individually (25).

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

The constraint variable that shares most causes of uncertainty with ΔF_{aer} and ΔF_{aci}, is H_d, the hemispheric difference in marine N_d. The key parameters that cause uncertainty in ΔF_{aci} (related to vertical velocities and sea spray emissions) also cause most of the uncertainty in H_d in all months (Fig. 2 and SI Fig. S1). This suggests we may extract much of the potential constraint from this type of observation using a single representative month (with dependencies on key parameters most closely aligned to ΔF_{aci} parameter dependencies). Other important parameters (newly formed sulfate diameters, DMS emissions and dry deposition velocities) also cause H_d uncertainty in some months. Seasonal differences in causes of H_d uncertainty can be traced to regional causes of N_d uncertainty (not shown), so based on shared causes of uncertainty, both H_d and regional N_d observations have potential to constrain ΔF_{aci}.

Several other observable state variables share key causes of uncertainty with ΔF_{aci}. Vertical velocities cause uncertainty in r_e and τ_c (around 20 to 30%), as do dry deposition velocities and, to a lesser extent, newly formed sulfate diameters (around 5 to 10%). Some transect variables also share causes of uncertainty with ΔF_{aci} (SI Fig. S8-11). For example, along the North Atlantic transect, the diameter of newly formed sulfate particles causes up to 50% of the uncertainty in many transect variables, including variables associated the cloud albedo response (gradient of the relationship between r_e and N_d for given LWP) and cloud adjustments (LWP and f_c vs. N_d). Vertical velocities cause up to 35% of the uncertainty in these cloud-related variables in the South Atlantic, whilst dry deposition causes up to 30% in the South Pacific. These regionally distinct causes of uncertainty suggest that observations from transects in several regions may constrain the model when combined, even if each transect variable constrains just one source of parametric uncertainty. Additionally, the radiative properties of carbonaceous aerosol (an important cause of ΔF_{ari} uncertainty) causes around 20% of the uncertainty in several variables along transects in the North Pacific. Thus, North Pacific transect variables have potential to constrain a process related to ΔF_{aer} through ΔF_{ari} that is otherwise unconstrained by our set of satellite-derived observations. In contrast with model constraint efforts designed to improve model skill more generally (26), our evaluation of shared causes of uncertainty provides process-based insight into the nature of any constraint we may achieve.

Not all state variables share causes of uncertainty with global mean ΔF_{aer} and ΔF_{aci} (SI Fig. S2-7). Outgoing radiative flux (F_{SW}) shares a few causes of uncertainty with ΔF_{aci} but is largely controlled by physical atmosphere parameters (in agreement with (20)). Similarly, global mean f_c and LWP share only a few minor causes of uncertainty with ΔF_{aci}, with values for these variables predominantly controlled by physical atmosphere parameters and uncertainties in the autoconversion scheme (that converts cloud drops to rain drops). Yet at the regional level (not shown), key parameters like the updraft parameter contribute between 5 to 10% of the f_c and LWP uncertainty

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

in most months, so f_c and LWP observations may still influence the ΔF_{aci} constraint. Thus, although, based on shared causes of uncertainty, some observation types have far greater potential for ΔF_{aci} constraint than others we do not exclude any from our constraint process at this stage.

Structural model inadequacies

Our goal is to constrain parametric uncertainty in ΔF_{aci}, ideally using all of the available observations, but in practice using a subset of observations for which the model-observation comparison is not affected by structural model inadequacies. We use two key indicators to identify potential structural model inadequacies. Firstly, some observations lie outside the range of the 1 million model variants, or are amongst the most extreme values. This indicates a discrepancy between the model and the observations that adjustments to model parameters cannot overcome (even by adjusting multiple parameter values simultaneously). That is, the discrepancy is more likely caused by a structural model deficiency than by parametric uncertainty. In practice, the discrepancy between model values and observations may be caused by very large, unquantified observational uncertainties or their lack of spatiotemporal representativeness (42). In such cases, either the model is incorrect due to some structural error, or the observation is unreliable. So, variables associated with this type of indicator are not useful for model constraint. Secondly, when we constrain the model using observations related to some constraint variables, the model skill at simulating other variables falls (24, 28, 43). In such cases, we can constrain the model towards one set of constraint variables or another, but not both simultaneously without systematically weakening the constraint. This suggests structural inadequacies prevent the model from consistently representing all processes associated with these constraint variables.

We begin by analysing potential structural errors in just one stratocumulus-dominated region. Fig. 3 shows the seasonal cycles of cloud physical and radiative properties in the North Atlantic (for other regions see SI Fig. S14-17). The distribution of the 1 million model F_{SW} values is centred on observed values, which is expected since extensive evaluations of F_{SW} across multiple model configurations feed into the model development process. Similarly, the distribution of f_c values is centred on the observations, with the exception of the April observation. This suggests that the f_c observation for April may be corrupted, or affected by some atypical event the model did not simulate, so should probably not inform our constraint. Model variants generally overestimate N_d, although N_d observations are well within the model's parametric uncertainty range. For LWP, τ_c and to a lesser extent r_e, observed values are near the edge of the model parametric uncertainty range or outside the range by a small margin. We have accounted for a very wide range of parameter uncertainties, but cannot adequately reproduce observed LWP and τ_c values in this region (more extreme in other regions; SI Fig. S14-17), which suggests the model bias

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

is caused by some structural deficiency. We cannot rule out satellite retrieval biases as an explanation for the model-observation bias with this first type of indicator, but the distinction between model structural error and observation error is not important in terms of model constraint. We therefore refer to such biases as *potential* structural inadequacies and remove the associated constraint variables from our process. Figure 3 exemplifies how we can compare observations to a broad range of model output to identify potential structural inadequacies where only extreme model behaviour aligns with observations.

We identified instances of the second indicator of potential structural inadequacy, associated with inconsistent model process representations, by evaluating how constraint of each model variable affects all other variables. We call this 'pairwise' comparison. Figure 3 shows two contrasting sets of pairwise comparisons, highlighting both consistency and inconsistency. We constrained the model to match the mean N_d observation for November in the North Atlantic and, separately, to match November LWP observations in this region. We chose November to exemplify the effect of the second indicator of potential structural inadequacy, because the parametric uncertainty in N_d and LWP peaks in this month. In each case, we ruled out model variants with relatively large model-observation differences (quantified using normalised root mean squared errors; NRMSEs) as observationally implausible and retained the subset of model variants with values closest to observations (Methods). Individual constraint variables have a large effect on the uncertainty in other variables because they share common causes of uncertainty in the model. For example, constraint to November N_d consistently reduces N_d uncertainty in all other months and brings the remaining model variants into close agreement with measured N_d values. This set of model variants also closely match F_{SW} and f_c observations, with the exception of April f_c, which we have already identified as problematic. These pairwise comparisons suggest that representations of N_d, F_{SW} and f_c are internally consistent in the model and we may only need a subset of these constraint variables to reduce uncertainty in ΔF_{aer}. However, the set of N_d-constrained model variants do not span the LWP, τ_c or r_e observations in most months, suggesting that model N_d is inconsistent with LWP, τ_c and r_e over the North Atlantic. In the other constraint shown in Fig. 3, the model variants that are consistent with November LWP in the North Atlantic do not span F_{SW}, f_c or r_e observations. Retrievals of cloud properties are consistent by design due to dependencies in their calculation. That is, multiple retrieved cloud properties from the same instrument share causes of observation bias. Thus, our results suggest structural deficiencies in the model related to internal inconsistencies in the representations of macro- and micro-physical cloud properties, which may prevent us using all available data for constraint of ΔF_{aci}.

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

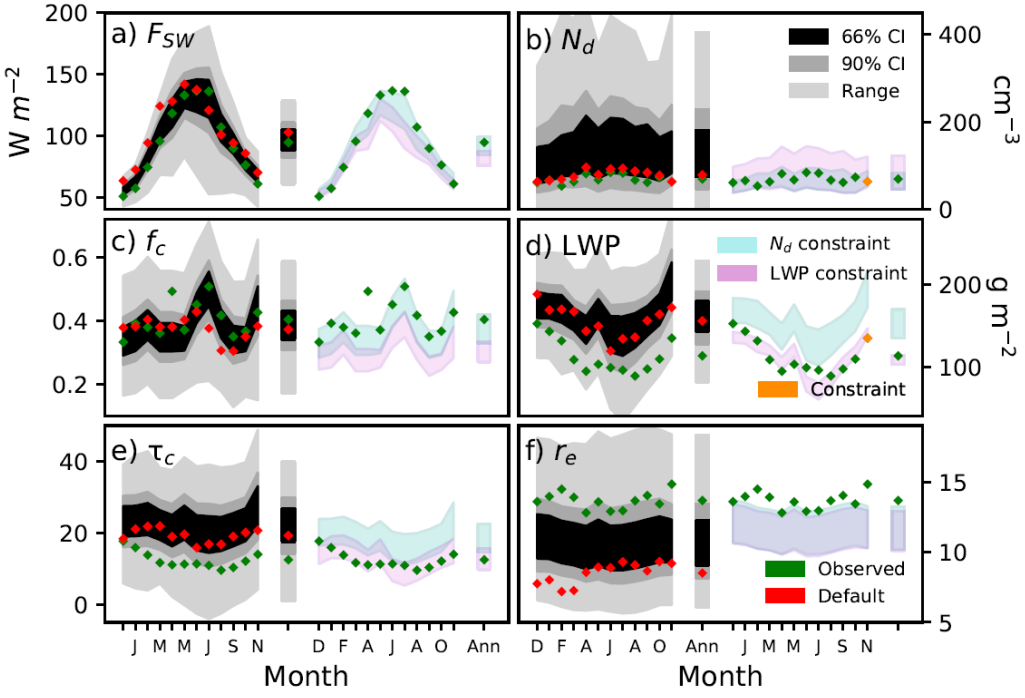

Figure 3. Seasonal cycles of North Atlantic mean radiative fluxes and cloud properties before and after observational constraint. Model output for individual months spanning December 2016 to November 2017 are followed by the annual mean. Credible intervals for the full set of model variants are shown (grey shading) along with satellite-derived observations and the default UKESM1-A model values. Each panel also shows the range (shading) of values from the November N_d constraint (green) and the November LWP constraint (pink; arrows show constraint variable).

To reveal the full extent of internal model consistencies and inconsistencies, we extended the analysis in Fig. 3 to all pairwise comparisons of the 88 North Atlantic constraint variables and with H_d (Fig. 4, other regions are shown in SI Fig. S18-21). Calculation of each of these pairwise effects across the full parameter space requires one million model-to-observation comparisons for each variable that is constrained and another million for each variable being compared. Two constraint variables are judged to be pairwise consistent if constraint to one variable improves the model–observation comparison for the other variable and vice versa. We quantify the impact on model-observation comparison as the percentage change in the average NRMSE, when moving from the unconstrained set of 1 million model variants to the set of model variants retained by the constraint (Methods). For each pairwise comparison, green shading in Fig. 4 indicates that observational constraint of the variable on the y-axis improves the model-observation agreement (reduces average NRMSE) for the variable on the x-axis. Pink shading indicates that the average NRMSE increases, which suggests that the two variables are inconsistent – that

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

is, the set of model variants that best match the variable on the y-axis are on average further from observations related to the variable on the x-axis than in the original (unconstrained) set. For example, model skill at simulating April f_c declines after constraint of any other variable, even f_c in most other months (vertical pink stripe). This supports our hypothesis that an observational error is the cause of the April f_c discrepancy and rules out using this constraint variable. These pairwise comparisons of constraint effects reveal inconsistencies between model variables LWP, τ_c and r_e, and other variables related to cloud properties (F_{SW}, N_d and f_c) in the North Atlantic (top right quadrant and bottom right panel of Fig. 4) and other regions (SI).

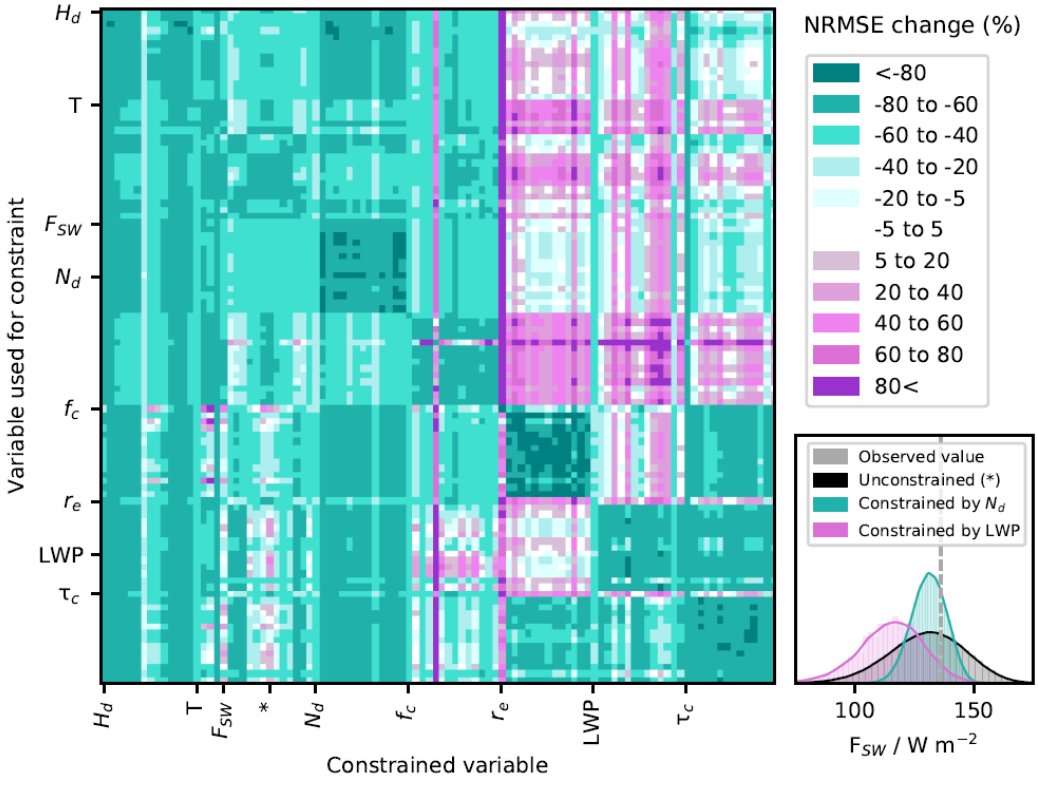

Figure 4. Pairwise comparisons of North Atlantic constraint variables (and H_d) showing how constraint of one variables affects all others. For each state variable, the first row/column is the seasonal amplitude, followed by individual months from January to December and the annual mean. Transect variables are within the section labelled 'T'. The y-axis labels refer to variables used to constrain the model output and the x-axis labels refer to variables whose values have been consequently constrained. Shading indicated the percentage change in average NRMSE after constraint. In the bottom-right panel we exemplify the effect of constraint on average NRMSE in one pixel in terms of probability density functions of July F_{SW} (*) in the unconstrained set of model variants along with two sets of constrained model variants. The first set is constrained to match July N_d observations and the second is constrained to match July LWP.

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

Optimal constraint of aerosol forcing

The pairwise comparisons in Fig. 4 show that it is not appropriate to use all observed variables to constrain the model because, due to potential structural model inconsistencies, different variables are consistent with different combinations of model parameters. The set of variants that simultaneously encompasses as many observed variables as possible is essentially the full initial set of one million. However, a smaller set of variants could be identified that agree with those observed variables that are represented consistently in the model. This is an approach taken either deliberately or inadvertently in model tuning in which some variables are deprioritised or neglected altogether. For example, F_{SW} is almost always treated as a high-priority target when tuning climate models because of its importance for energy balance, while N_d is more commonly treated as an adjustment term to achieve greater agreement with target values, and many other cloud variables are often neglected completely (18). Model tuning approaches attempt to minimise the effect of biases in a well-configured model version rather than seeking to identify structural systematic biases across a large number of model variants as we do here. There is no agreed best practice for identifying which combinations of model variables are structurally consistent. To explore the potential for constraint of ΔF_{aci}, we take the approach of constraining to the *most consistent* set of observed variables across our selected regions, then add more variables to understand the effect of accounting for inconsistencies.

We first identify constraint variables that are pairwise consistent with N_d at the regional level (see Methods). We choose the 225 constraint variables that are N_d-pairwise consistent because N_d is one of the most uncertain variables we evaluate here (Fig. 3), and N_d is a common adjustment variable for ΔF_{aer} constraint (18) due to its sensitivity to aerosol and its importance for ΔF_{aci}. In practice, we could use other constraint variables to define an internally consistent set (top left corner of Fig. 4). We evaluate the 25,200 combinations of these 225 constraint variables to reveal structural inconsistencies (Methods). First, we identify the constraint variable with the greatest individual effect on reducing ΔF_{aci} uncertainty, then progressively add constraint variables that are consistent with the existing set of variables (and N_d at the regional level) and contribute most to the ΔF_{aci} constraint. Figure 5 shows the effect of progressively adding constraint variables in this way (orange points).

The hemispheric contrast in N_d (H_d) in the Northern Hemisphere summer (August) provides the strongest individual constraint on ΔF_{aci}. The constraint towards lower values of H_d in August reduces the credible ΔF_{aci} uncertainty range in the unconstrained set of model variants by around 44%. August H_d shares causes of uncertainty

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

with ΔF_{aci}, and with H_d in all other months, but the nature of the relationships between the associated parameters (parameter dependencies) may be more clearly defined in August, since in most other months H_d is sensitive to additional parameters (SI Fig. S1). In combination with August H_d, additional constraint comes from next including South Pacific N_d in September (dependencies on natural emission flux parameters and dry deposition velocity), followed by March H_d (carbonaceous aerosol properties). Further constraint comes from North Pacific f_c in August (updraft velocity, autoconversion and physical atmosphere parameters) and changes in LWP along the North Pacific transect (carbonaceous aerosol radiative properties, autoconversion and physical atmosphere parameters). Southern Ocean N_d in December (natural emission fluxes and dry deposition velocities) and changes in LWP and N_d along the North Atlantic transect (updraft velocity and primary sulfate diameter) additionally constrain ΔF_{aci}.

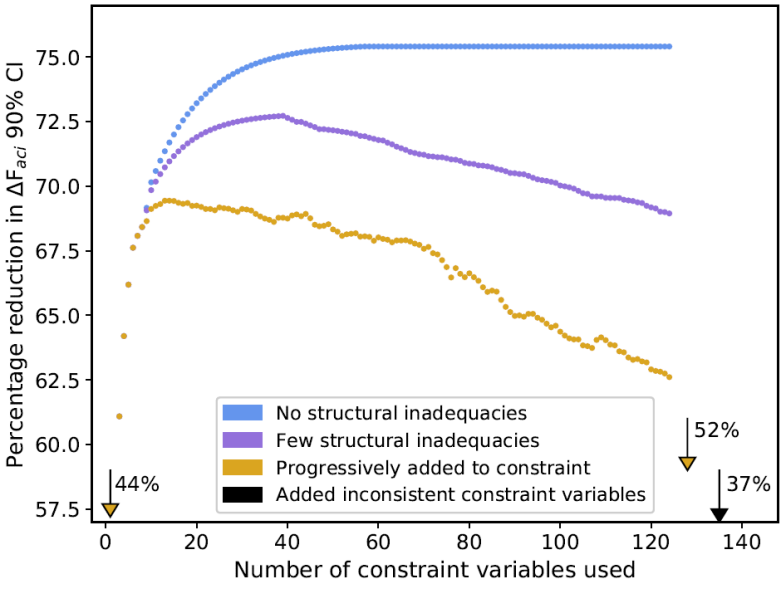

Figure 5. Constraint of ΔF_{aci} and the effect of varying the number of constraint variables used. We show the effect of progressively adding constraint variables with the greatest influence on ΔF_{aci} uncertainty (orange), alongside synthetic examples of how the constraint might improve with very few, or no structural model inadequacies (purple and blue respectively). In each case, only the first 125 of 225 N_d-pairwise consistent constraint variables are shown. Arrows indicate the constraint of ΔF_{aci} using the single strongest constraint (44%), all 225 N_d-pairwise consistent constraint variables (52%) and all 450 constraint variables (37%), including those associated with identified structural inadequacies at the regional level (e.g. Fig. 3 and 4).

We only need to include 7 additional constraint variables, in combination with the constraints identified above, (13 in total) to optimally constrain ΔF_{aci} (i.e. greatest reduction in the ΔF_{aci} 90% credible interval, orange points in Fig. 6). We define the 'optimal constraint' to be the greatest reduction in ΔF_{aci} achievable using our

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

specified set of observations and structurally imperfect model. This optimal set of constraint variables spans the observation types, regions and seasons, and provides information about the key uncertain parameters associated with these observations (and ΔF_{aci} dependencies on key model parameters). The optimally constrained set of model variants reduces ΔF_{aci} uncertainty by nearly 70% (90% credible interval -0.9 to -0.1 W m^{-2}) and ΔF_{aer} uncertainty by more than 50% (-1.3 to -0.1 W m^{-2}; Fig. 1). This constrained ΔF_{aer} range is narrower than previous best estimates (6) and purely process-based constraints (20, 24, 25) even though the ΔF_{ari} component of forcing is effectively unconstrained here. Additionally, the optimally constrained lower negative bound is now in close agreement with energy-balance constraints (table 1).

When applied in combination with the set of the 13 optimal constraint variables, any additional variables weaken the constraint (Fig. 5). This is because the additional variables are inconsistent with those already used. We retain at least five thousand model variants for each combined constraint (see Methods), so the result of adding further observations is a compromise in the sense that the existing constraints of ΔF_{aci} dependencies on key parameters need to be relaxed to accommodate conflicting information introduced by inconsistent variables. Practically, this means that some of the model variants with low NRMSE values are no longer retained. Instead, these model variants are replaced with other variants that have tolerable (not low) NRMSE values for the existing set of constraint variables *and* tolerable NRMSE values in relation to the new variable. Thus, the constraint is no longer optimal (for our model and these observations). Including all 225 N_d-pairwise consistent observations (identified independently in each region) only reduces the ΔF_{aci} uncertainty by just over 50% and adding regionally inconsistent observations to the constraint reduces the uncertainty by less than 40%. We anticipated the value of including additional constraint variables to level off, since many variables share key parameter sensitivities (SI Fig. S1-11). We also expected a decline in constraint efficacy (levelling off when progressively adding constraint variables; Fig. 5) once hidden structural inconsistencies started to mitigate the benefits of including additional constraint variables. However, we did not anticipate the optimal constraint to include so few constraint variables. These results suggest across one million variants, the model is structurally incapable of matching more than 13 of our chosen observations simultaneously.

Constraint of uncertain model parameters

Our approach consistently constrains the values of model parameters (SI Fig. S23, 24). Most parameters that cause ΔF_{aci} uncertainty (Fig. 2) are constrained, as are numerous other parameters that cause uncertainty in variables

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

associated with our set of optimal observations, that are not shared with (but not ΔF_{aci}). We entirely rule out some values as observationally implausible for parameters related to vertical velocity and newly formed sulfate particle diameters. Vertical velocities are constrained towards lower values, which are consistent with lower N_d concentrations in the relatively polluted Northern Hemisphere, a lower hemispheric contrast in N_d and weaker (less negative) median ΔF_{aci}. Conversely, newly formed sulfate particle diameters are constrained towards higher values, consistent with higher concentrations of cloud active aerosol concentrations and stronger (more negative) median ΔF_{aci}. Low sulfate emission diameters likely contributed to the spurious positive ΔF_{aci} values in Fig. 1. Dry deposition removal rates are also constrained towards higher values. This constraint reduces background aerosol concentration (consistent with lower N_d) and causes stronger (more negative) median ΔF_{aci} (increased sensitivity to anthropogenic aerosol). These key parameters are constrained concurrently, so have the effect of ruling out the strongest *and* weakest ΔF_{aci} (and ΔF_{aer}) values in our original set of model variants.

There is little evidence to support altering the current model representations of natural emission fluxes. Two key causes of ΔF_{aci} uncertainty, the emission fluxes of sea spray aerosol and DMS are constrained towards central values. However, the constraints on these parameters are relatively modest given their importance as causes of uncertainty. Additional constraint using in situ observations in relatively unpolluted regions (41, 44) could further constrain these parameters and the ΔF_{aci} uncertainty (25). Also, additional ΔF_{aer} constraint could be achieved using in-situ observations that target processes related to the ΔF_{ari} component of ΔF_{aer} (24, 30), which is effectively unconstrained by the satellite-derived observations used here.

Table 1. 90% credible intervals for ΔF_{aer}, ΔF_{aci} and ΔF_{ari} from the original one million model variants, and after constraint using process-based and energy-balance methods. We also include plausible bounds (90% credible interval) from energy-balance constraints.

	ΔF_{aer} (W m^{-2})	ΔF_{aci} (W m^{-2})	ΔF_{ari} (W m^{-2})
Unconstrained	-1.8 to 0.9	-1.5 to 1.0	-0.6 to 0.3
All variables	-1.5 to 0.2	-1.2 to 0.2	-0.6 to 0.2
Variables consistent with regional N_d	-1.4 to 0.0	-1.1 to 0.1	-0.6 to 0.3
Optimal constraint	-1.3 to -0.1	-0.9 to -0.1	-0.6 to 0.3
Smith et al. 2021; 1750 to 2019	-1.5 to -0.4	-1.2 to -0.1	-0.6 to -0.1
Albright et al. 2021	-1.2 to -0.33	N/A	N/A

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

Discussion

We illustrate some of the benefits of climate model evaluation that accounts for parametric uncertainty. In addition to constraining the lower bound on ΔF_{aer} to -1.3 W m^{-2}, a value in close agreement with energy-balance constraints, we have shown how this type of model evaluation can reveal structural model inadequacies. In our case, prioritising structural improvements to address model inconsistencies related to the representations of cloud variables, would increase the number and type of observations that could be used to further reduce ΔF_{aci} and ΔF_{aer} uncertainty in the model.

Structural inconsistencies weaken model observational constraint because to achieve tolerable agreement with more variables than in the optimal set, the inconsistencies demand a compromise in the tightness of constraint achieved. In UKESM1-A, the set of optimal constraint variables is surprisingly small, containing only around 3% of the constraint variables. At present, the remaining 97% of variables degrade the constraint. If we could make these variables consistent, for example by altering the structure of the model, then they would instead add to the constraint by further defining parameter relationships that were not constrained by the 3%. The hypothetical lines in Fig. 5 (purple and blue) describe what might be achieved if some or all of the structural model inadequacies were identified and improved – moving the peak to the right (more constraint variables used consistently in a structurally different model) and raising the peak (tighter parametric constraint of ΔF_{aci} and ΔF_{aer}). Ultimately, in a model without any structural inadequacies, the constraint versus number of variables would asymptote – additional variables would further constrain parameter relationships that were already partially constrained. The magnitude of constraint at this hypothetical asymptote is currently unknown. It will be determined in part by the effects of observational uncertainty and model-observation representation errors (42). Thus, we consider our optimal constraint the *minimum level* of process-based constraint that we might achieve, with this set of observations, if we could eliminate structural model inadequacies.

We suggest modelling groups may benefit from replacing existing model tuning strategies with a new approach to model evaluation and development that accounts for parametric uncertainty and strategically identifies the causes of model inconsistencies as well as ways to overcome their effects. In practice, the magnitude and distribution of observationally constrained ΔF_{aer} values in a structurally improved model may differ from the original model values (even with an identical set of parameter combinations). Thus, coherent progress at improving model skill at simulating aerosol-cloud interactions may require several cycles of uncertainty quantification, constraint, structural error identification and model development. Open source tools and code can help simplify some aspects of model evaluation within an uncertainty framework (45) and thus streamline some aspects of this

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

cycle. Furthermore, identifying optimal replacements for inconsistent process representations will require additional insight into causes of uncertainty within and across climate models, although knowledge of inconsistencies between variables provided by our approach will be a significant step in that direction. This valuable insight could be achieved by extending model intercomparisons, such as the 6th Coupled Model Intercomparison Project (CMIP6) (46), to include a cross-model perturbed parameter component. The breadth of model behaviour sampled in enhanced inter-comparisons would help to identify optimal combinations of process representations and parameter values that minimise key biases in climate models. Additionally, data from such ensembles would be invaluable for training relatively simple climate models (10, 11) and would contribute to efforts to identify robust emergent constraints (47). Experiments that sample parametric uncertainty *and* structural model differences could help deliver a step change in model skill at making climate projections beyond the advances we have achieved here using a single climate model.

Methods

We used the atmosphere-only configuration of version 1 of the UK Earth System Model (UKESM1) (48), to create our PPE. Here, we refer to this model version as UKESM1-A. UKESM1 was the model version submitted to CMIP6 (46). The model was forced using anthropogenic SO_2 and carbonaceous aerosol emissions, for the years 2014 and 1850 and some natural aerosol and gas species concentrations were derived from a fully-coupled UKESM1 simulation (SI Model version).

We created a new PPE of 221 model simulations for this study, over two stages. Each member of the PPE has a distinct combination of 37 aerosol and physical atmosphere parameter values (SI table S1). In the first stage, we created a PPE of several hundred members that spanned the 37-dimensional parameter space. To extend our sample of model simulations from 221 to one million model variants, we created statistical Gaussian process emulators (49) of multiple model outputs that densely sample model parameter uncertainty. We ruled out model variants (parameter combinations) that compared poorly to multiple global mean observations (see SI Multi-stage experimental design), then created a second PPE that spanned the not-ruled-out parameter space. We again extended our sample to one million model variants using statistical emulators of each variable based on output from the second stage PPE.

Our analysis started with 14 H_d constraint variables, 70 regional constraint variables (12 months, annual mean and seasonal amplitude over 5 regions) for each of the top-of-the-atmosphere outgoing shortwave radiative flux (F_{SW}), cloud drop number concentration (N_d), liquid cloud fraction (f_c), liquid water path (LWP), cloud optical

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

depth (τ_c) and cloud drop effective radius (r_e) state variables and 36 transect variables, all derived from satellite retrievals (SI Observations). Of these, we excluded 20 of the 36 transect variables where observations were outside of the model variant range, or at the extreme of the range (outside the 90% credible interval). We constrained our model using the remaining 450 individual constraint variables and tens of thousands of combinations of constraint variables (SI Constraint process). In each case, we calculated normalised root mean square (NRMSE) values for each of the one million model variants and retained the larger of either a) all model variants with errors smaller than the emulator uncertainty, or b) the 5000 model variants with the lowest NRMSE values (or lowest mean NRMSE values across variables included in a combined constraint). The choice to retain 5000 model variants has little influence on the results (SI Fig. S25 and table S4) Our results are insensitive to the number of model variants retained (SI Constraint process). This approach ensures we retain a similar sample size as previous studies of this kind (24, 25) and avoids over-constraint. We identified 225 of the original much larger set of constraint variables that are pairwise consistent with N_d in each region (Fig. 4 and SI Fig. S18-21). We define a variable as being 'consistent' with N_d when the constraint to match N_d does not increase the mean NRMSE calculated across the remaining model variants in the associated region and vice versa. To identify an optimal constraint, we quantified and ranked the constraint efficacy (as a percentage reduction in 90% credible interval of ΔF_{aci}) of the 225 N_d-consistent variables. We identified the most effective constraint variable, then quantified the constraint efficacy of the remaining 224 variables, in combination with the first. We progressively added constraint variables to achieve the greatest possible consistent constraint on ΔF_{aci}. Thirteen combined constraint variables optimally constrain ΔF_{aci}. We evaluated multiple alternative approaches to identifying an optimal constraint, which all produced very similar outcomes (SI Constraint process).

Data availability

Output from the A-CURE PPE is available on the CEDA archive (Regayre *et al.*, 2022). Python code used in this research is available here: https://github.com/Leighton-Regayre/article_code_constraint_aerosol_ERF.git

Author Contributions

LR and KC wrote the article. LR designed the experiments, with input from JJ, KC, PS, LD, DWP, KP and JM. LR, KP and CS implemented code changes for parameter perturbations. JM advised on specific aspects of model

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

set-up associated with aerosols and couplings between model components and LR implemented these changes. DS and JR advised on choice of physical atmosphere parameters to perturb in our ensemble. LR created the first stage ensemble. LR, LD, TL, CS and MR created the much larger second stage ensemble. GL, MD and MR provided technical expertise that assisted with creation of the PPE. LR processed the data, with contributions from DG and HG. LR analysed the data in collaboration with KC, DG, JJ, LD, PS, DWP and DS. LR, KC and JJ developed aspects of the methodology for identifying model structural inadequacies. All co-authors commented on the article.

Competing Interests

We have no competing interests to declare.

Acknowledgements

LR, KS and PS acknowledge funding from the FORCeS project under the European Union's Horizon 2020 research programme with grant agreement 821205. We acknowledge funding from NERC under grants AEROS, ACID-PRUF, GASSP and A-CURE (NE/G006172/1, NE/I020059/1, NE/J024252/1 and NE/P013406/1) and the European Union ACTRIS-2 project under grant 262254. DG and KS were supported by the National Environmental Research Council (NERC) national capability grant for The North Atlantic Climate System Integrated Study (ACSIS) program (grant NE/N018001/1) via NCAS and by the NERC ADVANCE Standard Grant project (NE/T006897/1). DWP acknowledges funding from the European Union's Horizon 2020 research and innovation programme iMIRACLI under Marie Skłodowska-Curie grant agreement No 860100. P.S additionally acknowledges funding from the European Research Council (ERC) project constRaining the EffeCts of Aerosols on Precipitation (RECAP) under the European Union's Horizon 2020 research and innovation programme with grant agreement no. 724602. DS was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra. JR was supported by the UK-China Research & Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund. JPM was supported by the Met Office Hadley Centre Climate Programme funded by BEIS. HG acknowledges support from the NASA Roses program under grant number 80NSSC21K1344. This work used the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk) under project allocation n02-NEP013406 to create the ensemble. KC was a Royal Society Wolfson Merit Award holder during this research. We appreciate expert advice provided by Masaru Yoshioka, Ben Johnson, Alex Archibald, Carly Reddington, Steven Turnock and Cat Scott, which allowed us to adjust uncertain model parameter ranges in light of recent research and model developments. We thank Ananth Ranjithkumar who shared code changes specific to high time-resolution output and Haochi Che who provided carbonaceous aerosol emission files in a model-ready format.

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

References

1. P. Forster, *et al.*, "The Earth's energy budget, climate feedbacks, and climate sensitivity" in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, (V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Pean, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Aterfield, O. Yelekci, R. Yu and B. Zhou, Cambridge University Press, 2021).

2. M. O. Andreae, C. D. Jones, P. M. Cox, Strong present-day aerosol cooling implies a hot future. *Nature* **435**, 1187–1190 (2005).

3. J. H. Seinfeld, *et al.*, Improving our fundamental understanding of the role of aerosol−cloud interactions in the climate system. *Proc. Natl. Acad. Sci.* **113**, 5781–5790 (2016).

4. A. H. Peace, *et al.*, Effect of aerosol radiative forcing uncertainty on projected exceedance year of a 1.5 °C global temperature rise. *Environ. Res. Lett.* **15**, 0940a6 (2020).

5. J. C. Fyfe, V. V. Kharin, B. D. Santer, J. N. S. Cole, N. P. Gillett, Significant impact of forcing uncertainty in a large ensemble of climate model simulations. *Proc. Natl. Acad. Sci.* **118**, e2016549118 (2021).

6. N. Bellouin, *et al.*, Bounding Global Aerosol Radiative Forcing of Climate Change. *Rev. Geophys.* **58** (2020).

7. M. Aldrin, *et al.*, Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content: ESTIMATING CLIMATE SENSITIVITY. *Environmetrics* **23**, 253–271 (2012).

8. R. B. Skeie, T. Berntsen, M. Aldrin, M. Holden, G. Myhre, A lower and more constrained estimate of climate sensitivity using updated observations and detailed radiative forcing time series. *Earth Syst. Dyn.* **5**, 139–175 (2014).

9. R. B. Skeie, T. Berntsen, M. Aldrin, M. Holden, G. Myhre, Climate sensitivity estimates – sensitivity to radiative forcing time series and observational data. *Earth Syst. Dyn.* **9**, 879–894 (2018).

10. C. J. Smith, *et al.*, Energy Budget Constraints on the Time History of Aerosol Forcing and Climate Sensitivity. *J. Geophys. Res. Atmospheres* **126** (2021).

11. A. L. Albright, C. Proistosescu, P. Huybers, Origins of a Relatively Tight Lower Bound on Anthropogenic Aerosol Radiative Forcing from Bayesian Analysis of Historical Observations. *J. Clim.* **34**, 8777–8792 (2021).

12. K. S. Carslaw, *et al.*, Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature* **503**, 67–71 (2013).

13. L. A. Regayre, *et al.*, The Climatic Importance of Uncertainties in Regional Aerosol–Cloud Radiative Forcings over Recent Decades. *J. Clim.* **28**, 6589–6607 (2015).

14. Y. Qian, *et al.*, Parametric Sensitivity and Uncertainty Quantification in the Version 1 of E3SM Atmosphere Model Based on Short Perturbed Parameter Ensemble Simulations. *J. Geophys. Res. Atmospheres* **123** (2018).

15. M. Yoshioka, *et al.*, Ensembles of Global Climate Model Variants Designed for the Quantification and Constraint of Uncertainty in Aerosols and Their Radiative Forcing. *J. Adv. Model. Earth Syst.* **11**, 3728–3754 (2019).

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

16. J. Gliß, *et al.*, AeroCom phase III multi-model evaluation of the aerosol life cycle and optical properties using ground- and space-based remote sensing as well as surface in situ observations. *Atmospheric Chem. Phys.* **21**, 87–128 (2021).

17. G. Thornhill, *et al.*, Climate-driven chemistry and aerosol feedbacks in CMIP6 Earth system models. *Atmospheric Chem. Phys.* **21**, 1105–1126 (2021).

18. F. Hourdin, *et al.*, The Art and Science of Climate Model Tuning. *Bull. Am. Meteorol. Soc.* **98**, 589–602 (2017).

19. L. A. Lee, C. L. Reddington, K. S. Carslaw, On the relationship between aerosol model uncertainty and radiative forcing uncertainty. *Proc. Natl. Acad. Sci.* **113**, 5820–5827 (2016).

20. L. A. Regayre, *et al.*, Aerosol and physical atmosphere model parameters are both important sources of uncertainty in aerosol ERF. *Atmospheric Chem. Phys.* **18**, 9975–10006 (2018).

21. P. S. Craig, M. Goldstein, A. H. Seheult, J. A. Smith, "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*, Lecture Notes in Statistics., C. Gatsonis, *et al.*, Eds. (Springer New York, 1997), pp. 37–93.

22. D. Williamson, *et al.*, History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Clim. Dyn.* **41**, 1703–1729 (2013).

23. I. Vernon, M. Goldstein, R. Bower, Galaxy Formation: Bayesian History Matching for the Observable Universe. *Stat. Sci.* **29** (2014).

24. J. S. Johnson, *et al.*, Robust observational constraint of uncertain aerosol processes and emissions in a climate model and the effect on aerosol radiative forcing. *Atmospheric Chem. Phys.* **20**, 9491–9524 (2020).

25. L. A. Regayre, *et al.*, The value of remote marine aerosol measurements for constraining radiative forcing uncertainty. *Atmospheric Chem. Phys.* **20**, 10063–10072 (2020).

26. D. M. H. Sexton, J. M. Murphy, M. Collins, M. J. Webb, Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. *Clim. Dyn.* **38**, 2513–2542 (2012).

27. J. Brynjarsdóttir, A. O'Hagan, Learning about physical parameters: the importance of model discrepancy. *Inverse Probl.* **30**, 114007 (2014).

28. D. McNeall, *et al.*, The impact of structural error on parameter constraint in a climate model. *Earth Syst. Dyn.* **7**, 917–935 (2016).

29. J. W. Rostron, *et al.*, The impact of performance filtering on climate feedbacks in a perturbed parameter ensemble. *Clim. Dyn.* **55**, 521–551 (2020).

30. D. Watson-Parris, *et al.*, Constraining Uncertainty in Aerosol Direct Forcing. *Geophys. Res. Lett.* **47** (2020).

31. P. Vignesh, *et al.*, Assessment of CMIP6 Cloud Fraction and Comparison with Satellite Observations. *Earth Space Sci.* **7** (2020).

32. T. Langton, P. Stier, D. Watson-Parris, J. P. Mulcahy, Decomposing Effective Radiative Forcing Due to Aerosol Cloud Interactions by Global Cloud Regimes. *Geophys. Res. Lett.* **48** (2021).

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

33. MODIS, MODIS Cloud Product, https://modis.gsfc.nasa.gov/data/dataprod/mod06.php, (2022).

34. A. Bodas-Salcedo, *et al.*, COSP: Satellite simulation software for model assessment. *Bull. Am. Meteorol. Soc.* **92**, 1023–1043 (2011).

35. G. Saponaro, *et al.*, Evaluation of aerosol and cloud properties in three climate models using MODIS observations and its corresponding COSP simulator, as well as their application in aerosol–cloud interactions. *Atmospheric Chem. Phys.* **20**, 1607–1626 (2020).

36. I. L. McCoy, *et al.*, The hemispheric contrast in cloud microphysical properties constrains aerosol forcing. *Proc. Natl. Acad. Sci.* **117**, 189980–19006 (2020).

37. M. W. Christensen, *et al.*, Opportunistic experiments to constrain aerosol effective radiative forcing. *Atmospheric Chem. Phys.* **22**, 641–674 (2022).

38. E. Gryspeerdt, J. Quaas, N. Bellouin, Constraining the aerosol influence on cloud fraction: AEROSOLS AND CLOUD FRACTION. *J. Geophys. Res. Atmospheres* **121**, 3566–3583 (2016).

39. D. Walters, *et al.*, The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geosci. Model Dev.* **12**, 1909–1963 (2019).

40. K. D. Williams, *et al.*, The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *J. Adv. Model. Earth Syst.* **10**, 357–380 (2018).

41. D. S. Hamilton, *et al.*, Occurrence of pristine aerosol environments on a polluted planet. *Proc. Natl. Acad. Sci.* **111**, 18466–18471 (2014).

42. N. Schutgens, *et al.*, On the spatio-temporal representativeness of observations. *Atmospheric Chem. Phys.* **17**, 9761–9780 (2017).

43. K. Sengupta, *et al.*, A global model perturbed parameter ensemble study of secondary organic aerosol formation. *Atmospheric Chem. Phys.* **21**, 2693–2723 (2021).

44. J. Schmale, *et al.*, Overview of the Antarctic Circumnavigation Expedition: Study of Preindustrial-like Aerosols and Their Climate Effects (ACE-SPACE). *Bull. Am. Meteorol. Soc.* **100**, 2260–2283 (2019).

45. D. Watson-Parris, A. Williams, L. Deaconu, P. Stier, Model calibration using ESEm v1.1.0 – an open, scalable Earth system emulator. *Geosci. Model Dev.* **14**, 7659–7672 (2021).

46. V. Eyring, *et al.*, Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).

47. K. S. Carslaw, L. Lee, L. Regayre, J. Johnson, Climate Models Are Uncertain, but We Can Do Something About It. *Eos* **99** (2018).

48. A. A. Sellar, *et al.*, UKESM1: Description and Evaluation of the U.K. Earth System Model. *J. Adv. Model. Earth Syst.* **11**, 4513–4558 (2019).

49. A. O'Hagan, Bayesian analysis of computer code outputs: A tutorial. *Reliab. Eng. Syst. Saf.* **91**, 1290–1300 (2006).

https://doi.org/10.5194/egusphere-2022-1330
Preprint. Discussion started: 24 November 2022
© Author(s) 2022. CC BY 4.0 License.

660 50. L. A. Regayre, *et al.*, A-CURE: Monthly mean perturbed parameter ensemble data, https://catalogue.ceda.ac.uk/uuid/b735718d66c1403fbf6b93ba3bd3b1a9, (November 23, 2022).