Response to reviewer comments on "Quantifying uncertainty from aerosol and atmospheric parameters and their impact on climate sensitivity" by Fletcher et al. (2018).

We thank the reviewers for taking the time to read and comment on our manuscript. The reviewer comments are reproduced here in blue, and our response to each comment is immediately below in black. The changes made to the text are described below each response, in italics.

Response to Review Comments RC1:

(1) I can follow your reasoning why you perform your analyses with such a coarse resolution. However, I am not so convinced that it does not have any influence on your results. If you use a coarse resolution gases as well as aerosols will probably not as accurately simulated than using a higher resolution and thus I would expect an effect on radiative forcing. Can't you do a simple test? For example you could use the combination of the input parameters where you found in your analyses the highest uncertainty and repeat the analyses to check if the results are still the same when you apply model simulations with a higher resolution.

We thank the reviewer for mentioning the potentially important effect of model horizontal resolution. This work was really a "proof-of-concept" study, and to minimize the overhead on our computing resources, we made the decision to focus on the low resolution (T31, approx. 3x4-deg) version of CAM4. The only real computing expense of this work is in producing the training simulations with CAM4; once the CAM4 simulations are complete, the emulator effectively runs for "free". The cost of running the training simulations with CAM4 at 2-degree (1-degree) resolution is a factor of 3 (15) higher than at T31, and our computing allocation in 2018 was insufficient to complete these simulations. However, we agree with the reviewer that an evaluation of the impact of resolution on parametric uncertainty is important, and we plan to carry out that study in future.

We considered the idea of running a test simulation at higher resolution with the optimal input parameters taken from the lower-resolution model. However, this may lead to erroneous conclusions, because the default values for the input parameters can actually be resolution-dependent (as mentioned on Page 7 of the manuscript). Therefore, we elect to hold off on testing the impact of resolution until we have sufficient computing resources to re-run the entire emulation procedure from scratch using higher resolution versions of CAM.

(2) The matrix plots are quite difficult to read. Wouldn't it be possible to use colours or frames to emphasize the correlations that are significant or have a high correlation coefficient? Further, the figures would be better readable if the labelling of input and output would be at the x and y-axis. I also would suggest to use different colours to separate between input and output.

We thank the reviewer for these suggestions, and we have attempted to implement them for all matrix-type plots in the revised version.

Changes made:
- *Background colours have been added to Fig.1 to emphasise the sign and magnitude of correlations. INPUT and OUTPUT labels have been added to make clear which panels are which.*

- *The clarity of axis labelling has been improved throughout*
- *Red lines have been replaced by dots to indicate the parameter settings of the default model*
- *Figs 6 and 7 have been combined for brevity, and the information made distinct.*
- *A colour bar has been added to the old Fig.8 (now Fig.7)*

Specific comments:

P1, L12: Shouldn't it be plural? Thus "uncertainties"?

Changed.

P2, L2: There are also uncertainties concerning the emissions of trace gases. This should be mentioned here as well.

We were not sure what the reviewer meant by "trace gases" here. In the next sentence, we already mention uncertainty in emissions of GHGs and aerosols.

P4, L4: The abbreviation AGCMs has not been used yet.

Changed to "atmospheric general circulation models".

P5, L3: Large scale transport is generally poorly simulated when model resolu- tion is low. Of course, resolution does not solve all problems with the circulation problem, but it has been shown that it definitely looks better when a higher resolution is used.

Thank you for this additional information. We have added a comment to reflect this, at the end of the sentence.

P5, L1: O[105]? What is the O standing for?

We were using "Big-O" notation to represent the order of the number of simulations that would be required to properly train the GCM. We have modified this to just say "10^5".

P5, L1: If anyway an emulation is used, why is it then not possible to use a higher model resolution? How would a higher resolution change the computing time needed for the emulation?

See detailed response about resolution above.

P7, L30: I cannot see the correlation between x3 and AOD. Is these really cor- rect? Is it not possible to improve the presentation of the results in the figures? See my specific comments on the figures below and above.

There is a fairly weak positive correlation between AOD and x3 (r ~ 0.2). This is clearer now that we have added background shading to the panels of Fig.1. The text has been amended to reflect this: *"AOD is … a weaker function of BC mass scaling"*.

P8, L13: What is "Cess"? What do you mean with "Cess climate sensitivity"?

This is a standard metric for estimating CS using atmospheric models with prescribed SST, and it is described in Section 2.1. In the revised manuscript, we have modified our notation to be consistent with that of Cess et al. (1989), where this method was first proposed, such that the Cess CS is denoted by the symbol \lambda. Page 8 is the first place where we present results using this quantity, and we wanted to emphasise to readers that it is the "Cess CS", rather than the equilibrium or transient climate sensitivity, that is being presented here.

P9, Figure 1: The figure(s) are quite difficult to read. In the caption some guid- ance how to read this matrix plot should be given.

See detailed response about figures above. But as suggested we have modified the caption to help the reader:

The background shading of each panel indicates the strength and sign of the correlation between a particular pair of variables, with reds indicating positive, and blues indicating negative, correlations, and stronger correlations represented by darker shading. For example, there is a strong positive correlation between x_5 and FNET, and a weak negative correlation between x_9 and FNET.

P10, L9: 10th percentile → 90% percentile (?)

The original article is correct: there are only ~10% of training cases that produce a value of CS lower than the default value.

P10, L14: Either use parenthesis around the reference Kay et al. (2014) or write "as in Kay et al. (2014)".

Wording is changed for clarity:

Relative to modern ensembles with comprehensive ESMs \citep{kay_community_2014}, our sample of $n=350$ training cases provides a large ensemble of cases

P10, L30: The abbreviation RMSE has not been introduced yet.

Definition is added at first use.

P11, Figure 2: There should be more space added between each row of figure panels. Wouldn't it possible to add a figure legend with the variable names or add them in the caption? Alternatively, this information could be given in a table. Additionally, the units at the axis should be given.

Figure has been changed:
- Spacing has been increased
- x-axis titles and units have been added

P12, Figure 3 caption: I would suggest to move the last sentence up, so that it is the second sentence in the caption.

Caption has been changed accordingly.

P14, Figure 4 caption: by the → by

Caption has been changed accordingly.

P14, Figure 4 caption: Comma after "parameter" redundant?

Matter of style. Comma is left as is.

P15, Figure 5: Also here I would suggest to add x und y-axis label for better readability or to give some guidance on how to read the figure in the figure caption. The read lines are not really helpful. Is there another way of presentation? Also here units should be given. Further, I would suggest to somehow mark (emphasize) in the figure the examples mentioned in the text.

Figure has been improved, as discussed above. We have added more description to the caption to make this figure easier to read:

The blue shading in each panel indicates the density of plausible cases for each pair of input parameters (x_1 -- x_9). Plausible cases are defined as the subset of $n=100,000$ emulated samples with SS>0.85, and dAOD<0.08. Darker shading indicates higher density, and white areas indicate zero density (i.e., no plausible cases). The default values for each parameter in CAM4 are indicated by red dots in each panel (except for x_4, which has no default).

P15, L1: Comma after "than" redundant?

Wording is changed for clarity:

compressed toward a central value that is slightly higher than the default value in CAM4

P15, Figure 6: Similar comments on this figure. Add x and y axis labelling and chose another presentation than the red lines. These are rather confusing than helpful.

P19, Figure 7: Add colourbar so that the highest and lowest values (range of values) are easier to differentiate. Also here, the red lines are rather confusing than helpful.

Figures 6 and 7 have been combined, with low (high) CS cases shown in blue (red).

P20, Figure 8: Same comments as for the previous figures.

Figure has been improved (discussed above).

P21, L26: Could you give here a number? How many percent?

We have attempted to quantify the "relatively small degree" to which a model could be moved along the Kiehl curve:

> *However, the relatively modest impact of parametric uncertainty means that a model could move along the curve by, perhaps, only 10-15 \% relative to the spread among CMIP5 models.*

Response to Review Comments RC2:

This work attempts analyzing the relative influence of aerosol and atmospheric pa- rameters uncertainty on climate sensitivity. The methodology used here is to sample combinations of 9 parameters (4 representing the aerosols and 5 representing clouds) in the CAM4 general circulation model. The number of emulations to be performed is of the order of 10^5. The design of the emulations and the methodology are well described but I believe that the conclusions of the paper are too far reaching. Here is why: Although the setup and the methodology appear to be well chosen, I would argue that the authors sample only a small part of the aerosol space and not the most rel- evant parameters for ascertaining which part of the variation in climate sensitivity can be explained by atmospheric parameters and which one is due to aerosol parameters. Therefore part of the sentence included in the abstract and the conclusions: '' The atmospheric parameters explain more than 85% of the variance in climate sensitivity for the ranges of parameters explored here, with two parameters being the most im- portant: one controlling low cloud amount, and one controlling the timescale for deep convection." can mislead readers. I recommend to edit this conclusion for the following reasons: 1) Page 4, line 18-19 the authors state: '' CAM4 does not include aerosol- cloud interactions, yet sulfate aerosols are known to be effective cloud condensation nuclei." This is right, sulfate aerosols are effective CCN (cloud condensation nuclei) as can be organic aerosols, seasalt and large dust particles. But since CAM4 does not include the effects of aerosol on cloud microphysics, trying to mimic it by varying the hygroscopicity (aptitude to uptake water) by sulphate is a shortcut that can hyde a large part of the effect aerosols have on cloud and precipitation. 2) The authors restrict the influence of aerosol parameters to studying the hygroscopicity of sulphate and the abundance of black carbon. To be complete, they should also integrate possible un- certainties of the distribution of organics, seasalt, dust and nitrate, their aging as well as their ability to be CCNs.

If these were taken into account, I doubt that the results of Figure 4 would remain the same as the aerosol influence on cloud microphysics would change the relative impor- tance of aerosol versus atmospheric parameters on climate sensitivity. This is why I propose that the authors

take out from their conclusion that: "atmospheric parameters explain more than 85% of the variance in climate sensitivity". I anticipate that future publications will contradict this result. It is very likely that the authors are correct when they state that: "low level clouds and the time scale for deep convection are the two most important parameters" since this will probably hold true even when the aerosol parameter space is better sampled.

We thank the reviewer for this suggestion, and we agree with the comment and recommendation to amend the conclusions of the study to better represent what can be concluded from our experimental setup.

Changes made:

- Abstract*: we removed the mention of 85% variance explained, and added caveats about model configuration

    In this experimental setup where aerosols do not affect the properties of clouds, the atmospheric parameters explain the majority of variance in climate sensitivity, with two parameters being the most important: one controlling low cloud amount, and one controlling the timescale for deep convection. Although the aerosol parameters strongly affect aerosol optical depth, their impacts on climate sensitivity are substantially weaker than the impacts of the atmospheric parameters, but this result may depend on whether aerosol-cloud interactions are simulated.

- Conclusions: We retain the 85% have added a new paragraph focusing on the limitations of not explicitly representing aerosol-cloud interactions in CAM4, or the role of other aerosol species.

    However, while we attempt to quantify the impact of aerosol-cloud interactions (ACI) through the hygroscopicity parameter for sulfate aerosols x_1, the CAM4 model does not include direct simulation of ACI, which would be expected to substantially increase the importance of the aerosol parameters \citep{regayre_aerosol_2018}. Future work should quantify the importance of uncertainties in parameters related to subgrid-scale activation of cloud droplets by aerosols in newer versions of the CESM-CAM models that include these processes, and quantify their impacts on ACI and λ \citep{golaz_cloud_2013}. In addition, our study focuses entirely on sulfate and black carbon aerosols, but important contributions to aerosol radiative forcing could be expected from uncertainties in the distribution of organic, sea salt, dust and nitrate aerosol, and the representation of their aging properties, and activation of cloud droplets \citep[e.g.,][]{chen_uncertainty_2006}. Therefore, while we expect the overall importance of x_5 and x_9 to be robust, we recommend caution in interpreting the precise numerical details of these results (for example, the 85 \% variance explained by atmospheric parameters), since these figures could be highly sensitive to the details of the model configuration.

-

Quantifying uncertainty from aerosol and atmospheric parameters and their impact on climate sensitivity

Christopher G. Fletcher[1], Ben Kravitz[2], and Bakr Badawy[1,*]

[1]Department of Geography and Environmental Management, University of Waterloo, Waterloo ON, Canada
[2]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA
[*]now at: Environment and Climate Change Canada, Dorval, Québec, Canada

Correspondence: Christopher G. Fletcher, 200 University Ave. West, Waterloo, ON N2L 3G1, Canada.
(chris.fletcher@uwaterloo.ca)

Abstract. Climate sensitivity in Earth System Models (ESMs) is an emergent property that is affected by structural (missing or inaccurate model physics) and parametric (variations in model parameters) uncertainty. ~~Here we explore the roles of~~ This work provides the first quantitative assessment of the role of compensation between uncertainties in aerosol forcing and ~~parametric uncertainty in affecting~~ atmospheric parameters, and their impact on the climate sensitivity of the Community Atmosphere Model, Version 4 (CAM4). Running the model with prescribed ocean and ice conditions, we perturb four parameters related to sulfate and black carbon aerosol radiative forcing and distribution, as well as five atmospheric parameters related to clouds, convection, and radiative flux. ~~The~~ In this experimental setup where aerosols do not affect the properties of clouds, the atmospheric parameters explain ~~more than 85% of the~~ the majority of variance in climate sensitivity ~~for the ranges of parameters explored here~~, with two parameters being the most important: one controlling low cloud amount, and one controlling the timescale for deep convection. Although the aerosol parameters strongly affect aerosol optical depth, ~~the effects of these aerosol parameters~~ their impacts on climate sensitivity are substantially weaker than the ~~effects~~ impacts of the atmospheric parameters, but this result may depend on whether aerosol-cloud interactions are simulated. Based on comparisons to inter-model spread of other ~~Earth System Models~~ESMs, we conclude that structural uncertainties in this configuration of CAM4 likely contribute three times more to uncertainty in climate sensitivity than parametric ~~uncertainty~~uncertainties. We provide several parameter sets that could provide plausible (measured by a skill score) configurations of CAM4, but with different sulfate aerosol radiative forcing, black carbon radiative forcing, and climate sensitivity.

1 Introduction

Climate models are exceptional tools for understanding contributions of different forcing agents to climate change. When forced with observed climate forcings (e.g., greenhouse gases and aerosols), they are skillful at reproducing past climate change (Flato et al., 2013). Their ability to do so builds confidence that climate models can provide plausible projections of future climate change under various assumptions about ranges of changes in forcing agents (i.e., emissions scenarios).

At the core of these projections is the climate sensitivity (CS) of the model, which is an emergent property of a model that describes how much global warming will be simulated by the model for a prescribed change in the carbon dioxide (CO_2)

concentration. Variations in CS arise due to *structural uncertainty* (how models represent the underlying physics of the climate system) and *parametric uncertainty* (how the simulation varies with model parameters, typically those associated with sub-grid-scale parameterizations). Each parameter in a model parameterization has an associated uncertainty due to deficiencies in representing model physics with these parameterizations, or a lack of available observational data to constrain the physics. A third source of uncertainty ~~is~~ comes from the emissions of climate forcings—particularly anthropogenic and natural aerosols—which are also uncertain to some degree, and have plausible ranges (e.g., Bond et al., 2013; Samset and Myhre, 2015; Ban-Weiss et al., 2011)~~, even if the models skillfully reproduce the climate of the 20th century.~~.

Reducing the uncertainty in CS is a key focus of climate science research. Fortunately, there are tools available to provide constraints on the estimates. Kiehl (2007), and more recently Forster et al. (2013), observed that there is a statistical relationship among climate models (the "Kiehl curve"), where models that are the most sensitive to CO_2 also tend to show largest historical radiative forcing by aerosols. As such, by exploring uncertainties in key model parameters and/or aerosol emissions rates, we expect to be able to alter a model's climate sensitivity, effectively moving the model along the Kiehl curve. With a sufficiently large number of such parameter sets, we can identify which versions of the model (i.e., sets of parameters or emissions rates) produce plausible climates. We hypothesize that, through this procedure, it may be possible to create multiple versions of a single climate model that plausibly reproduce observations but have discernibly different climate sensitivities.

The idea of obtaining multiple plausible versions of a model with different climate sensitivities was also explored by Mauritsen et al. (2012) for the Max Planck Institute for Meteorology's climate model. These authors found that the CS was not as sensitive to parameters as expected; however, they did not attempt to explore uncertainty in the forcing. Similarly, Golaz et al. (2013) altered the settings of a cloud detrainment parameter in the deep convection scheme, and produced three versions of the GFDL CM3 model with very different climate sensitivities, but again did not vary the forcing, and so two of the candidate versions were not able to reproduce the transient evolution of historical global mean surface air temperature. A recent, and highly relevant, contribution is that of Regayre et al. (2018), who show that uncertainties in historical aerosol radiative forcing, and top-of-atmosphere shortwave radiative flux, in a comprehensive chemistry-climate model are controlled by a combination of aerosol parameters *and emissions*, as well as uncertain atmospheric parameters. Their results show that, particularly in recent decades, constraining aerosol and atmospheric parameters allows regional climate impacts of aerosols to be more faithfully reproduced.

This area of research is closely related to Uncertainty Quantification (UQ), and to the more applied topic of model tuning (see also: Watanabe et al. (2010); Gent et al. (2011); Hourdin et al. (2016, 2013); Zhao M. et al. (2018)). Modern climate and ~~earth-system models~~ Earth System Models (ESMs) are so complex, and with so many tunable parameters, that running comprehensive calibration schemes with the full dynamical model is impractical. The typical approach is to conduct one-at-a-time (OAT) tests, where a single uncertain parameter is varied while holding all other parameters at their default value (e.g., Covey et al., 2013). While a lot can be learned about a model in OAT mode, this approach neglects potentially important parameter interactions, which can only be studied in vastly more expensive all-at-a-time (AAT) sampling designs. A very useful approach, therefore, is to construct a statistical emulator trained on output from the dynamical model, which is used to predict unsampled outputs from the dynamical model. Many recent studies have used Gaussian Process (GP) emulators to

sample an almost infinitely-large number of parameter combinations in climate and atmospheric models (Regayre et al., 2018; McNeall et al., 2016; Lee et al., 2011; Carslaw et al., 2013). The GP emulator is a Bayesian statistical technique that fits a smooth nonlinear function to a set of training data based on some prior information/assumptions, and provides an estimate of its own posterior uncertainty for each prediction (Lee et al., 2011).

5   The main goal of this paper is to ~~use a GP emulator to~~ quantify how much variation in climate sensitivity ~~exists within a~~ within a single climate model (CAM4, the atmospheric component of CCSM4/CESM-CAM4; see Section 2.1 below) ~~,~~ ~~as~~ is a function of ~~a range of~~ uncertainties in aerosol forcing and ~~uncertain~~ atmospheric parameters. We will use statistical emulation to test ~~a very large (O10^6) number of parameter~~ 10^5 combinations of parameters/forcing~~combinations, and use this sample~~ , to identify those combinations that yield plausible climates, but with different climate sensitivities. In other words, we

10 seek to quantify the degree of equifinality in CAM4's climate sensitivity. This work differs in focus from the recent study of Regayre et al. (2018), because here we explicitly assess the combined impact of parametric uncertainty and forcing uncertainty on climate sensitivity. While the conceptual idea of the Kiehl curve has existed for more than a decade, this work is, to our knowledge, the first explicit attempt to move a single climate model to different parts of the curve, and to quantify the impact on climate sensitivity. We stress that this contribution does not tackle the problem of calibration to observations: all simulations are

15 conducted as perturbations to preindustrial (1850) conditions, and are compared to the default version of CAM4 as a reference. The idea ~~is that we will end up with~~ from this proof-of-concept study, is to configure a series of ~~High~~candidate models with high/~~Low CS candidate models that could~~ low CS, which could then be run with interactive ocean components, to produce transient RCP-type simulations (historical and future scenarios) beginning in the preindustrial era.

2   Data, Models and Methods

20 ## 2.1   CAM4 model and atmospheric parameters

We use the National Center for Atmospheric Research (NCAR) Community Atmosphere Model Version 4 (CAM4), the atmospheric component of the Community Climate System Model Version 4 (CCSM4) and the Community Earth System Model Version 1.0.4 (CESM1), fully documented by Gent et al. (2011) and Collins et al. (2006). Hereafter, we refer to this model simply as "CAM4". In this proof-of-concept study, CAM4 is run at coarse horizontal resolution (3.75° longitude × 3.75°

25 latitude) with 26 vertical layers extending from the surface to 3 hPa (∼40 km). The coarse horizontal resolution is selected to increase computational efficiency, and is appropriate to represent the broad-scale features of the climate response described in this study (Shields et al., 2012). In addition, we note that this model configuration includes a crude representation of the stratosphere (only 4 layers are located above 100 hPa); however, since our primary focus is on radiation, and not circulation, we consider this resolution to be sufficient for the purpose of separating the radiative effects of aerosols in the troposphere and

30 stratosphere. This is also the same vertical resolution that was employed in a similar manner by Ban-Weiss et al. (2011). All simulations are performed in preindustrial (1850) mode: the CO_2 concentration is fixed at 284 ppmv, while other atmospheric constituents are prescribed from a monthly-varying climatology. For natural and anthropogenic aerosols the climatology is taken from a simulation using interactive (i.e., prognostic) chemistry (Lamarque et al., 2010), and the output is taken as the

mean of all realizations. The uncertainty due to atmospheric and oceanic initial conditions is estimated using the standard deviation of the realizations.

We assess climate sensitivity (CS) in CCSM4/CESM-CAM4 using the method proposed by ~~Cess et al. (1991)~~Cess et al. (1989), where a uniform 2 K warming is applied to global sea surface temperatures (SSTs). The Cess CS is then ~~given by $CS = \Delta F / \Delta T$,~~

5 ~~where ΔT~~ $\lambda = \Delta T_s / \Delta F$, where ΔT_s is the change in global annual mean near-surface air temperature (K), and ΔF is the change in global annual mean top-of-atmosphere net radiative flux (ΔF, W m^{-2}). The changes are evaluated as the difference between the simulation with warmed SSTs minus the simulation with preindustrial SSTs. The ~~CS~~ λ diagnostic is useful for model calibration and tuning because it can be run with ~~AGCMs~~ atmospheric general circulation models using prescribed SSTs (Golaz et al., 2013; Zhao M. et al., 2018), and it is highly correlated with the transient climate response and the equilibrium

10 climate sensitivity of the same models run with coupled interactive ocean (Medeiros et al., 2014).

2.2 Perturbations to aerosol forcing and atmospheric parameters

We focus on two aerosol species, sulfate and black carbon (BC), and we perturb the radiative forcing from both species simultaneously. Current best estimates of aerosol radiative forcing (ARF) are -1.9 to 0.1 W m^{-2} with medium confidence (Boucher et al., 2013), with by far the dominant source of uncertainty arising from interactions with clouds (Stevens, 2015;

15 Regayre et al., 2018). Sulfate exerts a well-known and strong negative radiative forcing on climate due to direct scattering of solar radiation, and through interactions with cloud properties (Boucher et al., 2013). By contrast, BC is an absorbing aerosol that generally exerts a positive ARF, although the sign depends on whether the BC layers are above (negative) or below (positive) cloud layers (Kim et al., 2015; Regayre et al., 2018; Ban-Weiss et al., 2011). The largest uncertainties in BC ARF are due to an incomplete inventory of emissions, as well as poor understanding of aging/scavenging rates, vertical and horizontal

20 transport, and deposition (e.g., Bond et al., 2013).

To examine uncertainty due to sulfate and BC forcing in CAM4, we introduced four new parameters to the model (x_1– x_4, see Table 1), with x_1 controlling sulfate and x_2–x_4 controlling BC. CAM4 does not include aerosol-cloud interactions, yet sulfate aerosols are known to be effective cloud condensation nuclei. As a proxy, parameter x_1 attempts to mimic these interactions by specifying the fraction of the sulfate mass that uses optical properties for sulfate in "hygroscopic mode," i.e., a

25 pure SO$_4$ molecule grown hygroscopically (following Köhler theory) corresponding to a relative humidity of 99%. The optical properties for the resulting sulfate aerosol (single scattering albedo, asymmetry parameter, and extinction coefficient) are given by $k_{\text{default}}(1-x_1)+k_{\text{hygro}}x_1$, where k_{default} indicates the default sulfate aerosol parameters in CAM4, and k_{hygro} indicates the parameters corresponding to this hygroscopic aerosol. This procedure ensures that the total mass of sulfate, and its geographic location, are the same in all cases, and the only perturbation is the fraction of sulfate that is water-coated (regardless of the

30 atmospheric humidity profile in the vicinity of the sulfate molecules).

To perturb BC forcing we assume that the optical properties are known, but that the total mass and horizontal and vertical distributions are uncertain, broadly consistent with the conclusions of Bond et al. (2013). We define three parameters x_2 (horizontal distribution), x_3 (mass scaling) and x_4 (vertical distribution) that control the distribution and amount of BC mass in the model. More specifically, x_2 describes linear interpolation between the default horizontal distribution for BC mass ($x_2 = 0$)

4

and BC being uniformly distributed throughout the globe ($x_2 = 1$). This parameter addresses the uncertainty associated with BC aging and transport: the longer that BC particles are able to survive in the atmosphere without being scavenged, the further the distribution should spread into pristine marine and polar environments. x_3 takes values between 0 and 40 and simply serves as a multiplier on the BC distribution, indicating uncertainties in total emissions and hence total mass loading. x_4 corresponds to an altitude (0–40 km), indicating where a "layer" of BC (with mass equal to the total default mass) is added to the model, and then the total mass is rescaled to be the appropriate value per parameter x_3. This parameter addresses uncertainties associated with large-scale transport of BC by the atmospheric circulation, which is known to be poorly simulated, especially when model spatial resolution is low (Lamarque et al., 2010).

We also investigate the sensitivity to five uncertain atmospheric parameters in CESM1-CAM4 that are related to clouds and convection. The parameters (denoted x_5–x_9) were identified as highly important in a previous one-at-a-time sensitivity analysis (Covey et al., 2013), and are described in Table 1. Two parameters (x_5 and x_8) control the threshold of atmospheric relative humidity that must be achieved before low and high clouds form, respectively; increasing either of these parameters will reduce the amount of low, or high, cloud in the model. Parameter x_6 changes the radius of liquid cloud droplets over the ocean, with smaller radii associated with brighter marine clouds, that are known to be highly important for climate sensitivity (Stevens, 2015; Sherwood et al., 2014). Parameters x_7 and x_9 are the timescales for shallow and deep convection, respectively; increasing either parameter will result in longer-duration convective precipitation. These parameters exert a large control on the mean climate in CAM4, but they are also expected to influence the climate sensitivity (Gent et al., 2011; Bony et al., 2015; Sherwood et al., 2014). We therefore vary these five atmospheric parameters in tandem with changes to the aerosol forcing to identify plausible climates with different climate sensitivities.

2.3 Emulation

For the nine parameters described in Section 2.2, we would need to perform at least ~~O~~10^5 simulations with CAM4 to adequately sample the parameter space in a typical all-at-a-time mode. Even running CAM4 at relatively low resolution, this would be impractical. The solution is to train an efficient statistical emulator of the dynamical model, which can be used to predict the climate output for any combination of parameter values, provided that the parameters lie within the range over which the emulator has been trained.

Following Lee et al. (2012) and McNeall et al. (2016), we construct a Gaussian Process (GP) emulator of CAM4 using the R package diceKriging (Roustant et al., 2012), which fits an N-dimensional nonlinear regression model to predict an output y based on a series of k predictors (input parameters x_1–x_9). Alternative methods of emulation have been used for climate modeling applications, including generalized linear models (e.g., Yang et al., 2017) and artificial neural networks (Sanderson et al., 2008). However, the GP model has two attractive properties that make it highly applicable to this type of problem: it can capture nonlinear interactions between the output and multiple inputs, and it provides an estimate of its posterior uncertainty.

We begin by defining $n = 350$ combinations of parameter values (x_1–x_9) for the training points in the 9-parameter space, using a Latin Hypercube design that ensures good distribution of cases, even in the corners of the hypercube (McKay et al., 1979). For each training point, we produce three one-year realizations of CAM4, each using different atmospheric and oceanic

Table 1. List of parameters that are perturbed in this study, including for each parameter a description, the range of perturbed values, and the default value in CAM4 (where applicable).

parameter	description (CAM4 parameter name)	min	default	max	notes
x_1	Fraction of hygroscopic SO_4	0.0	0.0	1.0	Proxy for sulfate indirect effect (no units).
x_2	Spatial uniformity of BC (1 = globally uniform)	0.0	0.0	1.0	Proxy for BC aging and scavenging (no units).
x_3	Scaling factor for global BC mass	0.0	1.0	40.0	Proxy for uncertainty in BC emissions (no units)
x_4	Altitude for insertion of uniform BC layer	0.0	–	39.0	Proxy for vertical transport of BC (units km). Note: new parameter, no default.
x_5	RH threshold for low cloud formation (cldfrc_rhminl)	0.80	0.88	0.99	Value grid box RH must exceed before low cloud forms (no units)
x_6	Effective radius of liquid cloud droplets over ocean (cldopt_rliqocean)	8.4	14.0	19.6	(units microns)
x_7	Timescale for consumption rate of shallow CAPE (hkconv_cmftau)	900	1800	14440	(units seconds)
x_8	RH threshold for high cloud formation (cldfrc_rhminh)	0.50	0.50	0.85	Value grid box RH must exceed before high cloud forms (no units)
x_9	Timescale for consumption rate of deep CAPE (zmconv_tau)	1800	3600	28800	(units seconds)

initial conditions drawn from a 500-year control integration of the coupled ocean-atmosphere version of CAM4 at the same resolution (see Section 2.1). The mean of the resulting outputs from the three realizations is used to train the emulator, which reduces the noise arising from internal climate variability. To quantify the impact of the parameters on climate sensitivity, the training process must be repeated twice: once using prescribed preindustrial SSTs, and then again using warmed SSTs (see Section 2.1). For each training point, the necessary simulations take approximately six hours on a single 8-core node of a high performance computing cluster, giving a total computing time of $350 \times 6 \times 8 = 17,000$ core hours. Applying the emulator to predict an output variable at 10^5 or 10^6 uniformly sampled points in parameter space takes less than 30 seconds on a single core of a modern desktop computer. This is many orders of magnitude faster than it would take to run a set of 10^5 or 10^6 simulations with the dynamical model.

The output from the training simulations are used to construct the emulator. We make standard assumptions to configure the GP model, using a linear prior and the default Matérn covariance function (Roustant et al., 2012; Lee et al., 2011), although we have verified that our conclusions are insensitive to these choices. Before the emulator can be used for prediction, its performance is validated using leave-one-out cross-validation (LOOCV): each case from the $n = 350$ training set is left out in

turn, and the emulator is rebuilt using the remaining $n = 349$ cases. The resulting model is then used to predict the output for the case that was left out, and so on until each case has been predicted. The performance metric used here is the correlation coefficient between the $n = 350$ outputs simulated by CAM4 and the $n = 350$ predictions from the emulator run in LOOCV mode. The validation results are presented in Section 4.1.

2.4 Quantifying the plausibility of candidate models

The plausibility of the climate produced by a particular combination of input parameters is assessed using a multivariate skill score (SS), based on Pierce et al. (2009):

$$SS_X = r_{p,d}^2 - \left[r_{p,d} - (\sigma_p/\sigma_d)\right]^2 - \left[(\bar{p} - \bar{d})/\sigma_d\right]^2 \tag{1}$$

where for a spatial grid of particular output variable X (e.g., precipitation, low cloud amount, etc.), p denotes a perturbed model, and d denotes the default (reference) model, $r_{p,d}$ is the anomaly (pattern) correlation between X in p and d, σ is the spatial standard deviation of X, and overbars denote the spatial mean of X.

The SS quantifies the mean bias, spatial correlation, and spatial variance of six key simulated variables for each perturbed model relative to the default version of CAM4. The variables included in SS are low cloud fraction (CLDL), total precipitation (PRECT), net TOA radiative flux (FNET), shortwave cloud forcing (SWCF), longwave cloud forcing (LWCF), and global vertically-integrated longwave heating rate (QRL). We calculate SS for each variable separately, and then average the SS values to obtain the final SS for each perturbed model. To obtain a high value of SS (SS~1), a parameter combination must produce a simulated climate that is simultaneously close to that of the default model in all of these fields. We apply a stringent threshold of SS > 0.85 to determine whether a particular perturbed model is plausible, which equates to approximately the 85th percentile of the SS distribution. We compute SS for the n=350 training cases, and then use the emulator to predict SS for all possible parameter combinations (see Section 4.1).

3 ~~Relationship between inputs and outputs~~ Controls on climate sensitivity in the CAM4 training simulations

3.1 ~~Marginal relationships~~Relationship between inputs and outputs

Figure 1 presents the relationships between all inputs (perturbed parameters x_1–x_9) and all outputs, for the $n = 350$ training simulations run with CAM4. The Latin Hypercube sampling of input parameters (Section 2.3) ensures an even sampling of values across each input parameter's full range (see Table 1). Correlations between parameter values are very weak, which provides confidence that each training case is an independent event drawn from the parameter population. The default values in CAM4 of parameters x_5 and x_6 are located within the center of their distributions, while the values for aerosol parameters x_1–x_3 ~~(a scaling factor for BC mass that is unlikely to be less than 1) and~~, and atmospheric parameters x_7–x_9, are at the lower end. For the sensitivity parameter for high clouds (x_8), the default is also the minimum value (0.5); however, we note that the default value varies considerably with horizontal resolution, and at the 2-degree resolution used, for example, by Covey et al. (2013), ~~it~~ its value is 0.80. No default exists in CAM4 for the new BC altitude parameter x_4.

The marginal relationships between inputs and outputs can provide an indication of which parameters may be important for emulating the outputs. Intuitively, the aerosol optical depth output (AOD) is a strong function of the sulfate hygroscopic fraction (x_1)~~and~~, and a weaker function of BC mass scaling (x_3), but it shows no obvious relationships with the other aerosol-related parameters (x_2 or x_4), and most other atmospheric outputs are not strongly correlated with any of the aerosol parameters. One exception is total precipitation, which appears to decrease as a function of BC mass scaling (x_3), presumably because of a reduction in precipitating clouds induced by so-called semi-direct effects (Bond et al., 2013). For the atmospheric parameters, there are clear relationships between inputs and the output variable(s) directly related to the parameter being perturbed. For example, there is a strong negative relationship between low cloud amount (CLDL) and the sensitivity parameter for low clouds (x_5), and longwave cloud forcing (LWCF) is negatively correlated with the sensitivity parameter for high clouds (x_8). Most apparent from Fig. 1 is the interconnectedness of the outputs; for example, low cloud, net radiation and shortwave cloud forcing are highly correlated. Parameter x_5 strongly affects CLDL, and this produces knock-on effects to all other outputs: positive correlations with FNET, QRL and SWCF, and negative correlations with LWCF and PRECT.

The multivariate skill score (SS, see Section 2.4) shows no obvious relationships with any aerosol parameters, nor with the majority of the atmospheric parameters. Exceptions are the nonlinear relationship with x_5 (higher values of x_5 are inconsistent with low SS), and the deep convective timescale (x_9), which is negatively correlated with, and sets an upper bound on, SS. This suggests that it is not feasible to achieve a climate that is consistent with the default model when the timescale for deep convection is longer than 3–4 hours (the default value of x_9 in CAM4 is 1 hour). Similar results are seen for the Cess climate sensitivity (~~CS~~λ), which shows no relationship to the aerosol parameters or atmospheric parameters x_6–x_8. A positive correlation is found between ~~CS~~ λ and x_5: higher x_5 produces less low cloud, and high-x_5 cases are, on average, slightly more sensitive to warming, which is consistent with expectations based on comprehensive ~~earth system models~~ ESMs (e.g., Siler et al., 2018). Lastly, we note an interesting nonlinear impact of x_9 on ~~CS, where CS~~ λ, where λ increases linearly with x_9 up to about a 4-hour timescale, then becomes much more variable and begins to decrease again.

3.2 Probability distribution of output variables

The black lines in Fig. 2 show the distribution of global mean outputs based on the $n = 350$ training cases from CAM4, for the six variables that comprise the SS, and ~~the Cess CS~~λ. The six variables are expressed as biases relative to the default model, so that a value of zero represents a case that perfectly reproduces the default model's global mean for that variable. Their distributions are unimodal and straddle zero, indicating that the majority of cases produce climates reasonably close to the default CAM4. However, for most variables the tails of the distribution are much longer than a Gaussian, indicating a large number of climates that are far from the default model. Indeed, the full range of climates produced in this ensemble is dramatic: the spread in top of atmosphere (TOA) net radiative flux (FNET) is between -30 and $+20$ W m^{-2}, which is an order of magnitude larger than the FNET response expected due to a doubling of the atmospheric CO_2 concentration (Andrews et al., 2012). The distributions of CLDL and FNET are also shifted toward positive and negative values, respectively, indicating a tendency for the perturbations to increase low cloud in the model, increasing shortwave flux to space. Fig. 1 shows that this effect is controlled almost entirely by parameter x_5, the sensitivity parameter for low clouds, which has previously been
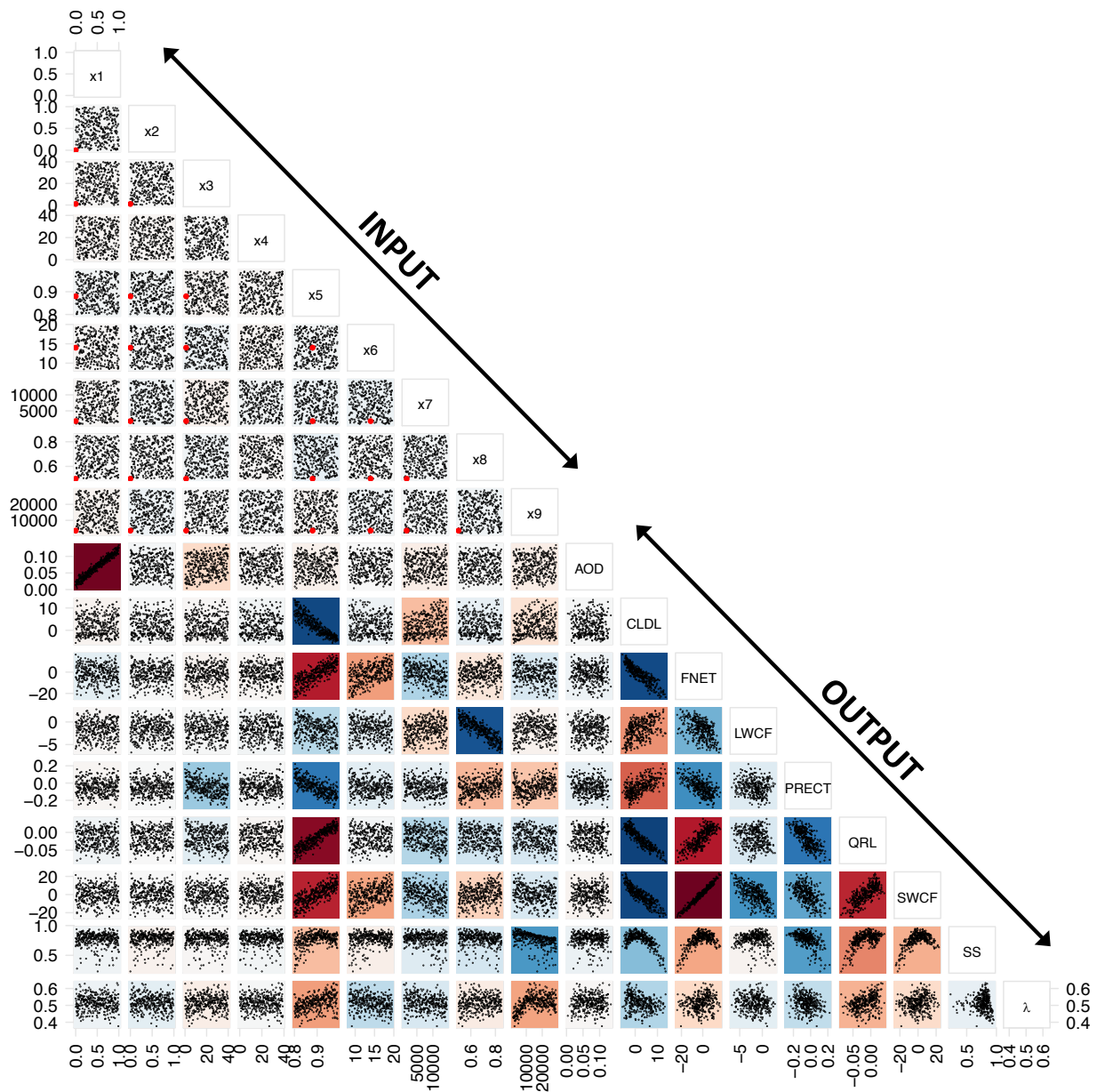
Figure 1. Scatterplot matrix showing the relationship between all input parameters (x_1–x_9), and all output variables (names defined in the text) in the $n = 350$ training cases run with CAM4. Red points show the default values for all input parameters except x_4, which has no default. The background shading of each panel indicates the strength and sign of the correlation between a particular pair of variables, with reds indicating positive, and blues indicating negative, correlations, and stronger correlations indicated by darker shading. For example, there is a strong positive correlation between x_5 and FNET, and a weak negative correlation between x_9 and FNET.

identified as very important in CAM4 (Covey et al., 2013). The impact of x_5 may also be asymmetric: reducing x_5 by 0.01 from its default value of 0.88 is likely to have a greater impact on cloudiness than increasing x_5 by the same amount, because x_5 is a relative humidity (RH) threshold, and the distribution of RH is heavily skewed toward lower values.

The distribution of SS is bounded by 0 and 1, by construction, with a single peak at around ~~0.75~~0.8, a maximum SS of 0.953, and a long left tail. The peak in the distribution of SS at ~~0.75~~0.8, and the fact that $\max(\mathrm{SS}) < 1.0$, implies that all the cases within our ensemble are—to a greater or lesser extent—imperfect representations of the default model. ~~This is largely related to our inclusion of the new aerosol parameters x_1, x_2, and~~ This is mostly explained by the parameter x_4 ~~in the experimental design: there can be no case~~ (the altitude of injection of a uniform layer of BC): no perturbed case can produce a climate exactly like the default model, because ~~that~~ the default model does not include ~~these new parameters. Our analysis (not shown) suggests that the x_4 parameter (the altitude of injection of a uniform layer of BC) is most important in this regard, because~~, and the experimental design specifies that a BC layer is always ~~be~~ injected *somewhere* ~~in the atmosphere~~ between 0–40 km. We reiterate that the emphasis here is on identifying plausible candidate models with altered aerosol and atmospheric parameters, not on tuning/calibration to make the default model more *realistic* (relative to observations). The distribution of λ shows a range between $0.35 - 0.65$ K/Wm^{-2}, meaning that around 90 % of all candidate models produce a higher climate sensitivity than the default value of 0.45 K/Wm^{-2}. The input parameters driving these changes are explored in Section 5.

4 Exploring the parameter space through emulation

4.1 Validating the emulator

~~The~~ Relative to modern ensembles with comprehensive ESMs (Kay et al., 2014), our sample of $n = 350$ training cases provides a large ensemble of cases ~~(relative to modern ensembles with comprehensive ESMs Kay et al. (2014)) with~~ with which to study the effects on CS from aerosol forcing and atmospheric parameters (Figs. 1–2). However, the nine parameters x_1–x_9 map onto a vast parameter space that is computationally impractical to sample adequately using CAM4 itself. A more practical way to explore the response space (i.e., to fill in the unsampled regions for the output variables in Fig. 1) is by using a statistical emulator.

Using the emulator of CAM4 described in Section 2.3, we make predictions for each output variable shown in Fig. 2 using fine-resolution uniform sampling over the full range of each parameter x_1–x_9. The emulated results are shown as the gray shaded regions in Fig. 2, and it is immediately apparent that the emulator does a very good job at reproducing the simulated distribution for each variable. The close agreement in all cases indicates that the uncertainty in the emulator is small, which provides confidence that the emulator is a useful tool to explore the parameter space of CAM4. However, we first conduct a more quantitative validation of the emulator, by performing leave-one-out cross-validation (LOOCV) to sequentially leave out, and then predict, each individual case from the $n = 350$ training sample. The results of this LOOCV procedure are shown in Fig. 3, and reveal that the emulator is, in general, highly successful at predicting model outputs for unseen parameter combinations.
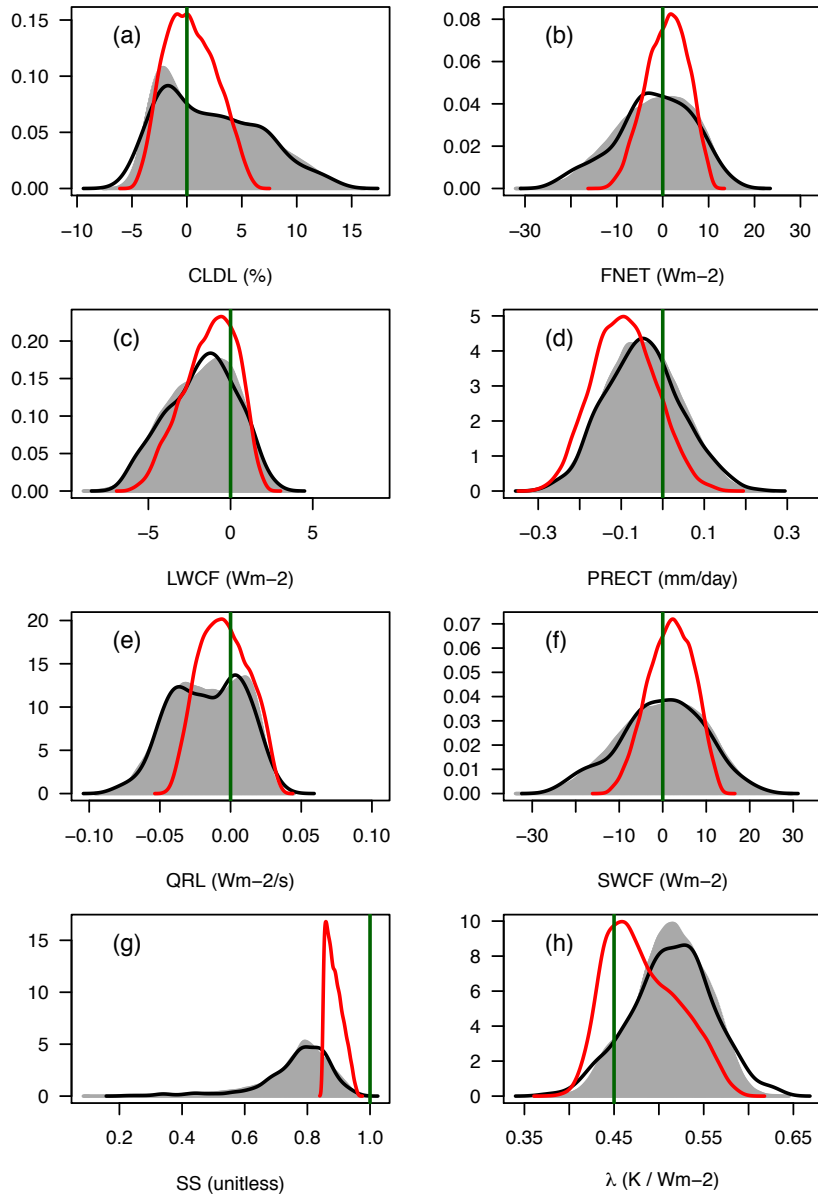
Figure 2. The black line ~~in each panel~~ shows the density distribution of simulated values from CAM4 for ~~each output variable in~~ (a) CLDL, (b) FNET, (c) LWCF, (d) PRECT, (e) QRL, (f) SWCF, (g) multivariate skill score (SS), and (h) Cess climate sensitivity (λ), from the $n = 350$ training cases (variable names are defined in the text). The gray shaded region shows the density distribution from 100,000 emulated samples of each ~~output~~ variable, and the red line shows the density distribution for the subset of emulated samples that is most plausible (see Section 4.3). Note the different x-axis scales in each panel~~, and (for the upper six panels only) a value of zero corresponds to a perfect representation of the global mean in the default model~~. The dark green vertical line in each panel indicates the ~~default model value~~ location of ~~CS~~CAM4 with its default parameter settings.

11

Taking into account emulator uncertainty (shown by the 95% confidence interval on each prediction), the agreement between the emulated and simulated values of the multivariate skill score (SS) is close to perfect ($r > 0.98$). For the Cess climate sensitivity (~~CS~~λ), while there is greater uncertainty on each prediction (margin of error ~ 0.05 K/Wm^{-2}), and the emulated values show a general tendency to be slightly underpredicted, ~~RMSE~~ root-mean-square error (RMSE) is very low (0.023 K/Wm^{-2}) and the correlation skill is high ($r = 0.86$). A final demonstration of the value of using the Gaussian Process emulator is shown by repeating the emulation using a multiple least-squares linear regression (MLR) model. The righthand panels of Fig. 3 show that the MLR emulator performs substantially worse than the GP emulator in terms of both RMSE and correlation skill.

4.2 Parameter sensitivity / importance

Parameter sensitivity is quantified following Carslaw et al. (2013) and McNeall et al. (2016) using the so-called FAST methodology (Saltelli et al., 1999), in the R package *sensitivity* (Pujol et al., 2017). This method separates the contribution to the total response from "main effects", that are directly attributable to variations in each (normalized) parameter, and interactions between parameters, which are calculated as the residual: interaction = total − main effect.

Figure 4 reveals that atmospheric parameters x_5 and x_9 are most influential for the outputs SS and ~~CS~~λ, explaining a combined total of $\sim 75\%$ of the total variance in each output. The variation in output AOD is explained almost entirely by x_1 (sulfate hygroscopic fraction), with a small residual contribution by x_3 (BC mass scaling). No other aerosol parameters are influential for any of the output variables shown, and there are small ($< 10\%$) contributions from atmospheric parameters x_6–x_8 for SS and ~~CS~~λ. In general, the main effects are dominant; however, non-negligible parameter interactions are found for SS, where they make up almost half of the total variance explained by x_5. While difficult to directly interpret, we hypothesize that this emphasis on the interaction terms is due to the interrelated nature of the parameters x_5–x_9, all of which influence clouds, precipitation, and radiative flux (i.e., all of the variables assessed in computing SS) in some form. These results agree closely with previous work examining the sensitivity to atmospheric parameters in CAM4. For example, Covey et al. (2013) show that x_5 and x_9 are highly influential parameters for top-of-atmosphere radiative flux, although their one-at-a-time methodology did not permit an examination of parameter interactions.

4.3 Identifying a plausible set of input parameters

In this section, we use the statistical emulator to identify regions of the 9-dimensional parameter space that produce plausible climates, which are defined as those similar to the climate of the default CAM4. We begin by using the emulator to construct a set of $n = 100,000$ cases for output variables SS, AOD and ~~Cess CS~~λ, based on a uniform sample of the distributions of each parameter (x_1–x_9). Applying first the threshold SS>0.85 eliminates $\sim 85\%$ of cases, and adding a second constraint (AOD<0.08) eliminates a further 6% of cases whose AOD is too far from that of default CAM4. The threshold for AOD represents a trade-off between finding a sufficiently large sample of cases, and their fidelity to the default model. Since present-day AOD in default CAM4 tends to be biased low [satellite observations from MODIS+MISR for present day show AOD~ 0.16, compared to 0.11 for CAM4; Remer et al. (2008)], and the aerosol perturbations x_1–x_4 tend to increase AOD, the threshold
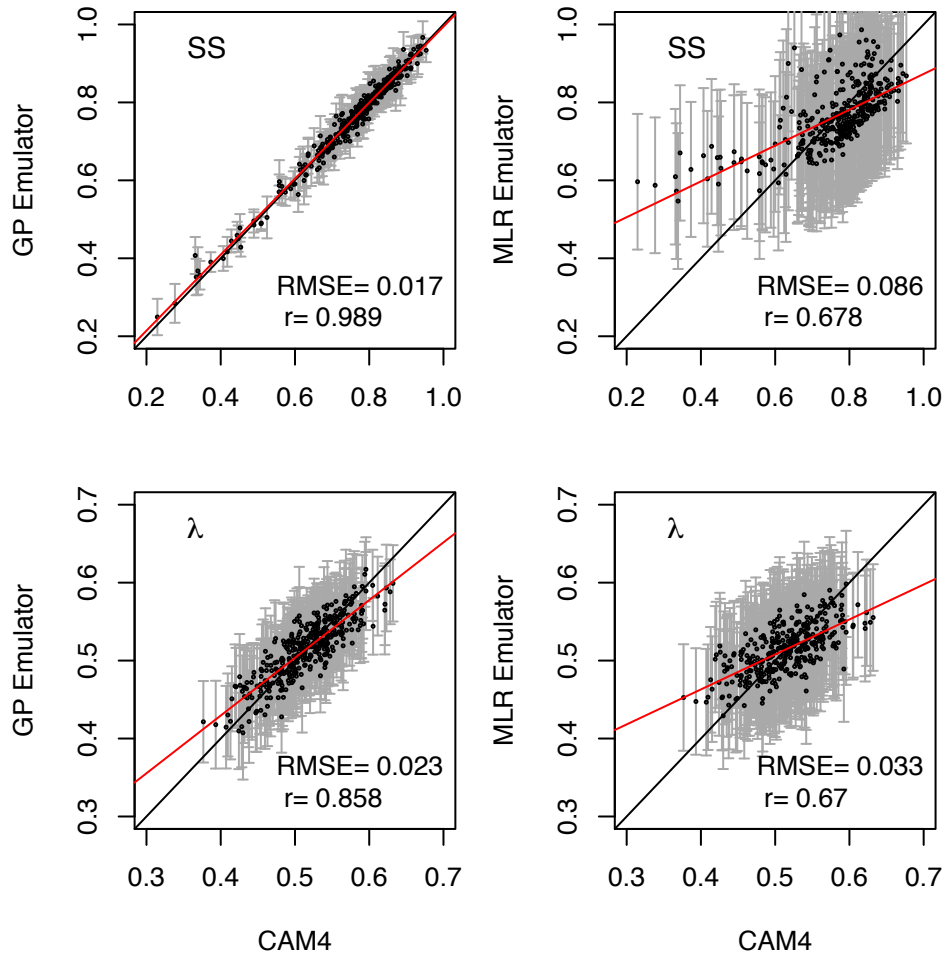
Figure 3. Points show predictions from the emulator (y-axis) against CAM4 simulated values (x-axis) for SS (top) and CS (bottom) for each case in the $n = 350$ training set, using the statistical emulator in LOOCV mode (see Section 4.1). The lefthand panels show predictions made using the GP emulator, and the righthand panels show predictions made using the MLR emulator. Gray error bars show the 95% prediction interval for each case. The red line is a least-squares linear fit of the predicted against simulated values, and the black solid line shows the 1:1 line. ~~The lefthand panels show predictions made using the GP emulator, and the righthand panels show predictions made using the MLR emulator.~~
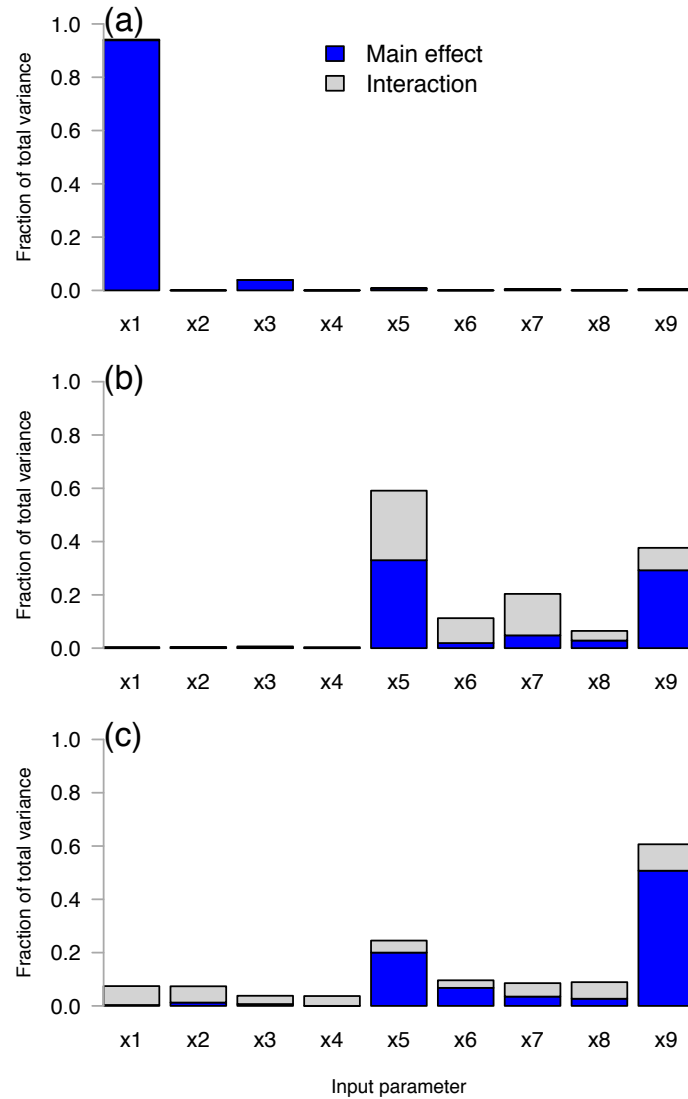
13

Figure 4. Fraction of total variance explained by ~~the~~ each input parameter, for ~~AOD (top~~a) AOD, ~~SS (middle~~b) SS, and ~~Cess CS (bottom~~c) λ. The light gray portion of each bar represents the contribution of interaction terms, while the blue portion represents the main effect. The full height of each bar represents the total variance explained by each input parameter, with more influential parameters explaining a higher fraction of variance.

ensures that plausible cases maintain a global mean AOD that is within 50% uncertainty of the default CAM4. After applying both thresholds, only ∼9% of the original parameter space remains plausible, and the density distributions of parameters for this remaining space are shown in Fig. 5.

Next we attempt to constrain the parameter ranges by examining the regions of parameter space in Fig. 5 that produce higher/lower densities of plausible outputs. The only aerosol parameter that can be constrained is x_1, where all values above 0.6 are implausible. The BC mass scaling parameter x_3 shows a slightly reduced density of plausible cases for very high BC mass; however, even very high BC mass cannot be ruled out completely, because 15% of cases remain plausible with $x_3 > 32$. For the atmospheric parameters (x_5–x_9), the range of x_5 is compressed toward a central value that is ~~not far from, but with a tendency to be higher than ,~~ slightly higher than the default value in CAM4 (denoted by the red ~~lines~~ points in Fig. 5). For x_8 (high cloud sensitivity) there is a tendency for most plausible cases to have higher parameter values than default, but very high values are ruled out. The parameter x_9 (deep CAPE timescale) is dramatically constrained from its original range: all values above ∼4 hours are ruled out, and the default value is located well within the plausible range. Parameters x_6 (liquid drop radius over ocean) and x_7 (shallow CAPE timescale) do not show any obvious reduction in their plausible ranges.

~~Lastly~~ Finally, we examine the emulated outputs associated with the subset of plausible cases. The red line in each panel of Fig. 2 shows the distribution of output values only for the plausible cases, which provides an estimate of how much our candidate model versions differ from the default CAM4. For all variables the spread for the plausible subset is ~~smaller—often considerably—than~~ often considerably smaller than the spread for all cases, and tends to be shifted toward a mean of zero (a perfect representation of the global mean from the default model). This indicates that our threshold-based approach to plausibility is working as desired: the outputs from the plausible models *should* be closer to the default model, by construction. Interestingly, high values of ~~CS~~ λ become much less likely after imposing the thresholds. It could be argued that ~~CS~~ λ is directly influenced by the variables in the SS, and so it is not independent of our thresholding approach; however, we consider ~~CS~~ λ to be an emergent property of the model and, therefore, this result could not have been predicted a priori.

5   Impact on climate sensitivity

Running the CAM4 model with its default (unperturbed) settings for parameters x_5–x_9, and without perturbations to the additional aerosol parameters x_1–x_4, we find ~~a Cess CS value of 0.45~~ $\lambda = 0.45$ K/Wm^{-2}. The median value of ~~CS~~ λ for all $n = 100,000$ emulated cases is 0.51 K/Wm^{-2}, which implies that the net effect of the perturbations to the aerosol parameters (x_1–x_4) is to increase ~~CS~~ λ. The 95% interval of ~~CS values~~ λ for the 9% of emulated cases that are plausible is 0.418 K/W m^{-2} (7% lower than default) to 0.538 K/W m^{-2} (20% higher than default). This shows that, to some extent, ~~CS~~ λ is a tunable quantity; however, this range is only about 25% of the range in ~~CS~~ λ across an ensemble of CMIP5 models by Medeiros et al. (2014), ~~suggesting~~ and only about 16% of the range found for a much earlier set of models in a slightly different experimental configuration (Cess et al., 1989). This suggests that structural uncertainty (not sampled here) ~~also plays a major role in contributing to the spread of simulated CS~~ is probably more important than parameteric uncertainty in explaining the intermodel spread of λ.
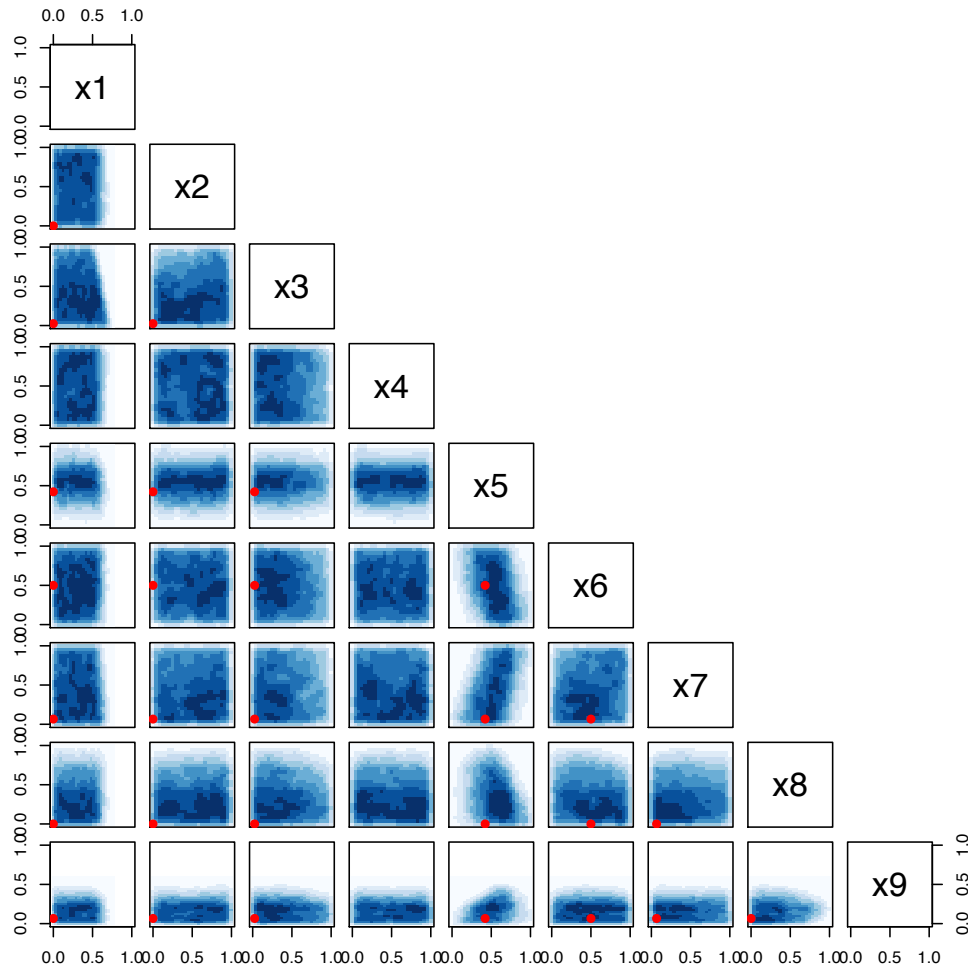
Figure 5. ~~Pairs density plots for input parameters, showing~~ The blue shading in each panel indicates the density of plausible cases ~~,~~ for each pair of input parameters $(x_1 - x_9)$. Plausible cases are defined as the subset of $n = 100,000$ emulated samples ~~that have~~ with SS>0.85, and ~~AOD~~dAOD<0.08. Darker shading indicates higher density, and white areas indicate very low density. The default values for each parameter in CAM4 are indicated by red ~~lines~~ dots in ~~the panels~~each panel (except for x_4, which has no default).
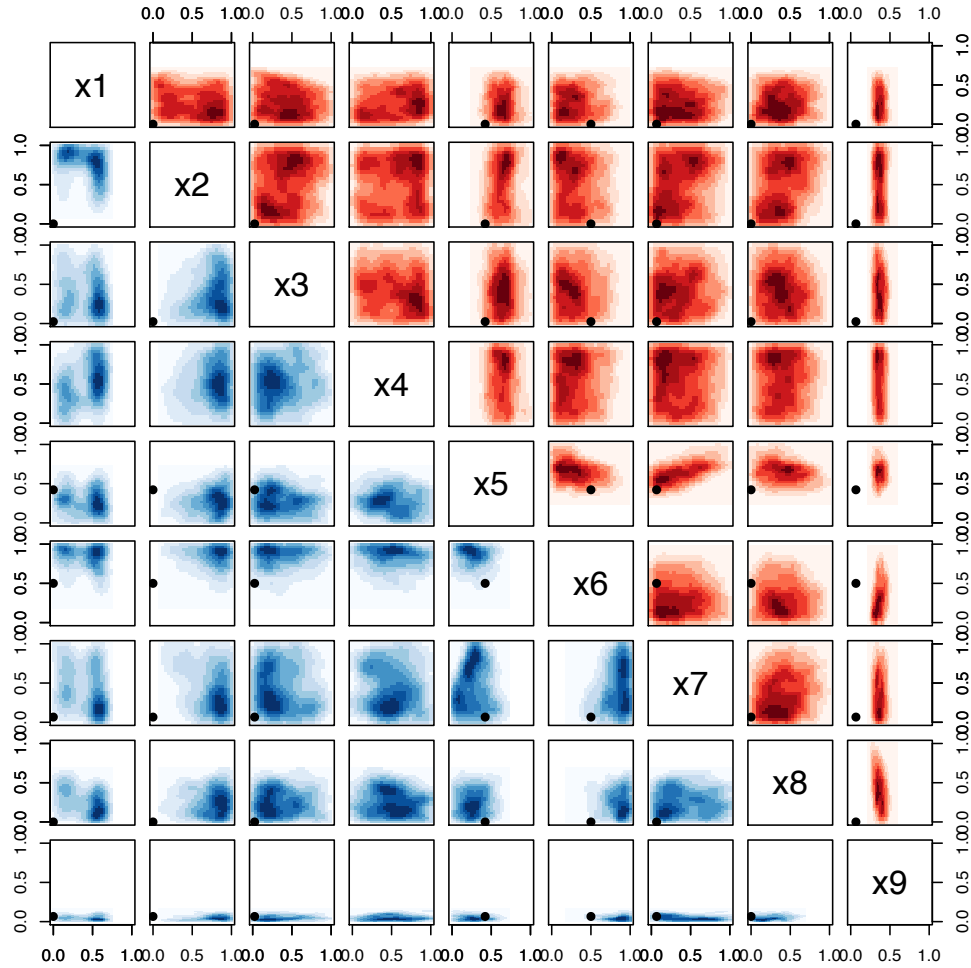
Figure 6. As Fig. 5, except ~~only~~ that panels with blue shading (lower triangle) are for the subset of cases ~~in the lowest 2.5%~~ with low values of ~~the distribution of~~ Cess climate sensitivity ($\lambda < 0.418$ K/Wm^{-2}), and panels with red shading (upper triangle) are for the subset of cases with high values of λ ($\lambda > 0.538$ K/Wm^{-2}). The default values for each parameter in CAM4 are indicated by black dots in each panel (except for x_4, which has no default).

We next examine the distribution of aerosol and atmospheric parameters that are associated with high (~~CS > 0.538~~ $\lambda > 0.538$ K/Wm^{-2}) and low (~~CS < 0.418~~ $\lambda < 0.418$ K/Wm^{-2}) sensitivity cases; i.e., ~~cases with CS~~ plausible cases with λ in the upper and lower 2.5% of all ~~plausible cases. Figures ??-7 show~~ cases. Figure 6 shows little difference in the distribution of the aerosol parameters (x_1–x_4) for high or low sensitivity, suggesting that neither the hygroscopicity of sulfate, nor the mass and spatial distribution of BC, are important for determining ~~CS~~ λ in this model. Much larger differences between the high and low sensitivity cases are found for the atmospheric parameters. As suggested from Fig. 1, high sensitivity is associated with higher values of x_5 (less low cloud), x_8 (less high cloud) and x_9 (longer deep convection), and lower values of x_6 (smaller liquid cloud droplets). The parameter x_7 (shallow CAPE timescale) appears to have little influence on ~~CS~~λ. The situation for low sensitivity is broadly the inverse, and the narrow ranges for some parameters (e.g., x_5 and x_9) provide clear constraints on radiative-convective processes that control climate sensitivity in this model. The speckled blue-yellow-red nature of the panels for x_1–x_4 in Fig. 7 shows that the spread of ~~CS~~ λ is very similar for all values of the aerosol parameters. This suggests that, in tandem with the right combination of atmospheric parameters, any value of ~~CS~~ λ within the model's range can be achieved for any strength of aerosol forcing. Figure 7 also reveals multiple effects of x_9 on ~~CS~~λ: the main effect (c.f. Fig. 4) is a strong linear gradient in ~~CS~~ λ from low to high parameter values, in addition to clear nonlinear interactions with parameters x_6 and x_8.

Collectively, these results suggest that our overall objective of configuring different, but equally plausible, versions of CAM4 with varying strengths of aerosol forcing and ~~Cess CS~~ climate sensitivity (λ), is eminently achievable. To this end, we conclude this Section by presenting examples in Table 2 of parameters with very different aerosol forcings selected from the emulated cases, which produce ~~CS~~ λ values across the full range ~~from~~ for this model. To extract these cases we apply joint thresholds to parameters x_1 (sulfate hygroscopicity) and x_3 (BC mass scaling) to identify combinations with high ("h") and low ("l") sulfate ("S") and BC ("B") forcing; for example, Sh.Bl denotes cases with high sulfate, and low BC, forcing. After extracting a distribution of cases for each combination of sulfate and BC forcing, we record the parameters that produce the minimum (low sensitivity), and maximum (high sensitivity), value of ~~CS~~λ.

As expected from Fig. 7 and the discussion above, we are able to identify both high and low sensitivity cases for all combinations of aerosol forcing. Comparing pairs of rows for the same aerosol forcing at high and low sensitivity reveals that, by construction, they tend to have similar aerosol parameters. However, clear differences emerge in the atmospheric parameters: the high sensitivity cases tend to have higher x_5, x_8 and x_9, and lower x_6. Only x_7 appears unrelated to ~~CS~~λ, perhaps due to its influence on shallow cumulus clouds, which are more likely to be overlain by higher clouds and, therefore, have limited influence on the top-of-atmosphere energy budget. This table provides a prototype for a future study to test a suite of cases in CAM4 with a fully interactive ocean, to determine the relationship between ARF, ~~Cess CS~~ λ and the transient climate response (e.g., Golaz et al., 2013; Zhao M. et al., 2018).

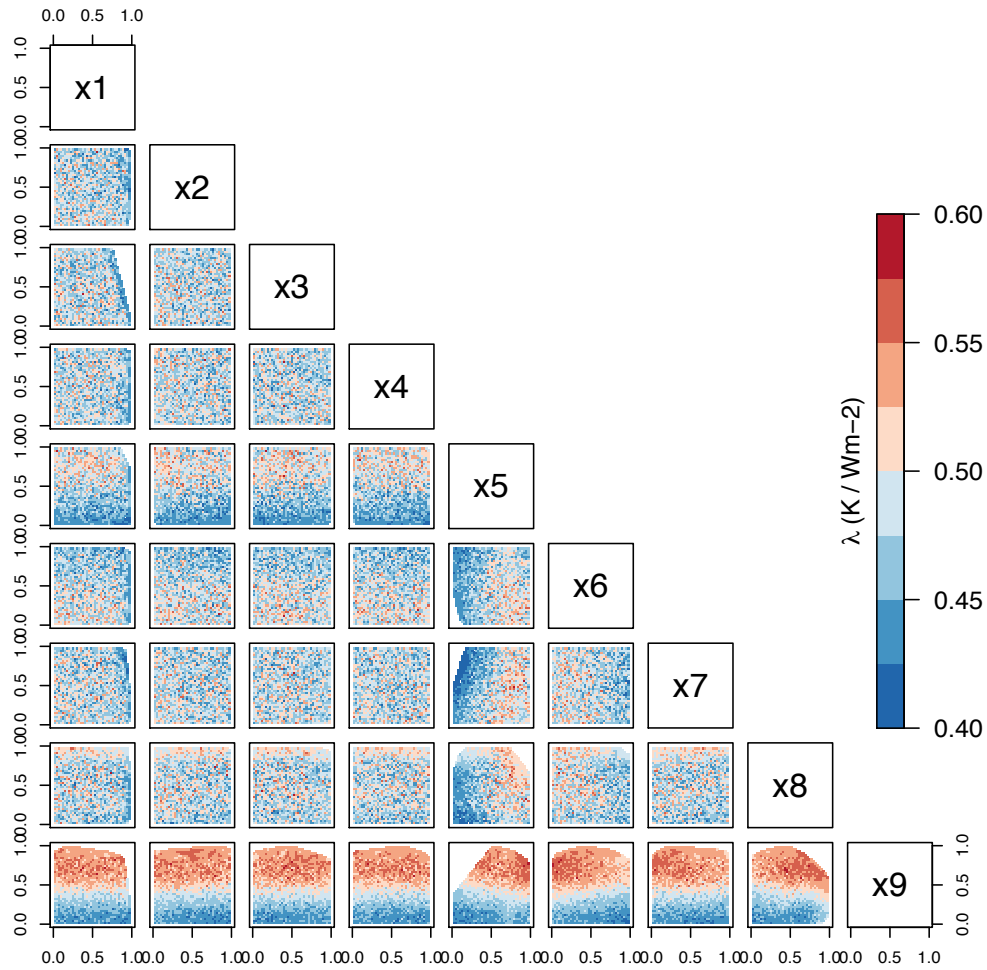~~As Fig. 5, except only for the cases in the highest 2.5% of the distribution of Cess climate sensitivity (>0.538 K/Wm^{-2}).~~

Figure 7. Color shading shows Cess climate sensitivity (λ, units K/Wm^{-2}) for plausible cases as a function of each pair of input parameters ($x_1 - x_9$). ~~Red shades indicate higher climate sensitivity in the range 0.25-0.65 K/Wm^{-2}, and blue shades indicate lower climate sensitivity.~~ White areas indicate regions of parameter space where there are no plausible cases.

Table 2. Each row shows the input parameter values (columns 2-10), and the Cess climate sensitivity (~~CS~~λ, units K/Wm^{-2}; column 11) predicted by the emulator for that combination of parameters. The first column denotes the strength of aerosol radiative forcing (ARF) in a case from sulfate (S) and Black Carbon (B): h=high, m=medium and l=low (defined in the text). For reference, the default parameter values in CAM4, and the associated ~~CS~~λ, are shown ~~near~~ toward the top.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	~~CS~~λ
				Default case						
	0.0	0.0	1.0	–	0.88	14.0	1800	0.50	3600	0.45
				High Sensitivity Cases						
Sh.Bl	0.50	0.05	0.9	37.9	0.98	9.3	9870	0.63	13066	0.61
Sh.Bh	0.52	0.48	23.7	28.3	0.93	10.5	1860	0.63	13202	0.58
Sm.Bm	0.20	0.08	11.7	38.0	0.97	9.1	12279	0.65	12759	0.60
Sl.Bl	0.02	0.24	3.9	0.7	0.95	11.3	4547	0.60	13753	0.57
Sl.Bh	0.04	0.62	20.7	4.2	0.99	10.3	7016	0.55	12297	0.59
				Low Sensitivity Cases						
Sh.Bl	0.55	0.47	3.1	21.0	0.82	14.7	4570	0.52	2524	0.39
Sh.Bh	0.58	0.74	26.2	31.4	0.84	18.0	8667	0.51	3741	0.40
Sm.Bm	0.22	0.89	10.2	4.1	0.90	18.2	12548	0.70	1896	0.40
Sl.Bl	0.08	0.48	2.9	38.1	0.86	15.4	2635	0.51	1824	0.41
Sl.Bh	0.08	0.54	22.1	36.7	0.83	18.1	6066	0.51	2185	0.39

6 Conclusions

We employ a statistical emulation procedure to sample the parameter/response space of the atmospheric general circulation model NCAR CESM-CAM4. The influence of four aerosol parameters controlling the aerosol radiative forcing (ARF) from sulfate and black carbon, and five atmospheric parameters controlling clouds and convection, are assessed in combination across their full range of uncertainty. A multivariate skill score is used to determine the plausibility of each combination of parameters, and thus to constrain plausible parameter ranges, and the spread of an important emergent property of the model: its climate sensitivity (~~CS~~λ). We find that atmospheric parameters explain more than 85% of the variance in ~~CS~~λ, and two parameters are most important: x_5 controls the amount of low cloud in the model, and x_9 controls the time scale for deep convection. The aerosol parameters have little impact on ~~CS~~ λ in our model configuration, making it equally possible to identify cases with high/low ARF that have high, or low, ~~CS~~ λ (Table 2).

However, while we attempt to quantify the impact of aerosol-cloud interactions (ACI) through the hygroscopicity parameter for sulfate aerosols x_1, the CAM4 model does not include direct simulation of ACI, which would be expected to substantially increase the importance of the aerosol parameters (Regayre et al., 2018). Future work should quantify the importance of uncertainties in parameters related to subgrid-scale activation of cloud droplets by aerosols in newer versions of the CESM-CAM models that include these processes, and quantify their impacts on ACI and λ (Golaz et al., 2013). In addition, our study focuses

entirely on sulfate and black carbon aerosols, but important contributions to aerosol radiative forcing could be expected from uncertainties in the distribution of organic, sea salt, dust and nitrate aerosol, and the representation of their aging properties, and activation of cloud droplets (e.g., Chen and Penner, 2005). Therefore, while we expect the overall importance of parameters x_5 and x_9 to be robust, we recommend caution in interpreting the precise numerical details of these results (for example, the 85 % variance explained by atmospheric parameters), since these figures could be highly sensitive to the details of the model configuration.

Our results indicate that the climate sensitivity of CAM4 can be modified, and possibly constrained, through adjustments to select uncertain ~~input~~ atmospheric parameters, primarily x_5 and x_9. ~~This procedure is closely related to model tuning, which is typically applied to comprehensive Earth System Models (ESMs) to ensure fidelity relative to available observational data (Gent et al., 2011; Zhao M. et al., 2018; Hourdin et al., 2016). The results presented here~~ These results can be compared with previous studies that examined the impact of tuning parameters on climate sensitivity (~~CS~~λ) in ESMs. We find a plausible spread of ~~Cess CS values~~ λ between 0.418 K/W m^{-2} and 0.538 K/W m^{-2}, which spans approximately 25% of the range derived from a suite of CMIP5 models that performed a similar experiment (Medeiros et al., 2014). This is in good agreement with previous studies that found that the spread in ~~CS~~ λ for a single ESM (either with interactive or prescribed ocean components) due to uncertain tuning parameters related to clouds and convection was smaller than the spread among the ensemble of CMIP models (Mauritsen et al., 2012; Golaz et al., 2013). This body of work, therefore, suggests that structural deficiencies in the configuration of ESMs ~~contributes~~ contribute more to the uncertainty in ~~CS~~ λ than parametric uncertainty. Of the ~25% of the spread in ~~CS~~ λ due to parametric uncertainty, our study indicates that atmospheric parameters explain the vast majority, with only a minor role for aerosol parameters. ~~However, a major caveat in the present study is the absence in CAM4 of aerosol-cloud interactions (ACI), which would be expected to increase the importance of aerosol parameters (Regayre et al., 2018). Future work should examine the importance of uncertainties in parameters related to subgrid scale activation of cloud droplets in newer versions of CESM-CAM models that include these processes, and quantify their impact on ACI and CS (Golaz et al., 2013).~~ The major new finding from this work is that a given model's position on the Kiehl curve *can* be varied through compensating adjustments to atmospheric parameters and radiative forcing, although perhaps only be a relatively modest amount compared to structural uncertainty.

This study explores only the question of whether plausible alternative versions of CAM4 can be configured (through uncertain aerosol and atmospheric parameters) to have different climate sensitivities, relative to CAM4 at the same horizontal resolution with its default parameter settings. Using the default model to determine plausibility explicitly avoids the question of plausibility relative to observations, or finding parameter combinations to "improve" CAM4. That would be an exercise in model tuning/calibration, which is beyond the scope of this study. However, our opinion is that the range of plausible solutions that have been revealed through the emulation procedure makes it highly likely that parameter combinations exist within the sampled parameter/response space that provide better matches to the observed climate than the default settings. This hypothesis will be examined in a future study.

References

Andrews, T., Gregory, J. M., Webb, M. J., and Taylor, K. E.: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models, Geophysical Research Letters, 39, https://doi.org/10.1029/2012GL051607, https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2012GL051607, 2012.

5   Ban-Weiss, G. A., Cao, L., Bala, G., and Caldeira, K.: Dependence of climate forcing and response on the altitude of black carbon aerosols, Climate Dynamics, 38, 897–911, https://doi.org/10.1007/s00382-011-1052-y, http://dge.stanford.edu/labs/caldeiralab/Caldeira_research/BanWeiss_Cao.html, 2011.

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N.,

10   Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, Journal of Geophysical Research: Atmospheres, 118, 5380–5552, https://doi.org/10.1002/jgrd.50171, http://doi.wiley.com/10.1002/jgrd.50171, 2013.

Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T. G., Sherwood, S. C., Siebesma, A. P., Sobel, A. H., Watanabe, M., and Webb, M. J.: Clouds, circulation and climate sensitivity, Nature Geoscience, 8, 261–268,

15   https://doi.org/10.1038/ngeo2398, https://www.nature.com/articles/ngeo2398, 2015.

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S., Sherwood, S., Stevens, B., and Zhang, X.: Clouds and Aerosols, book section 7, p. 571–658, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.016, www.climatechange2013.org, 2013.

20   Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., Mann, G. W., Spracklen, D. V., Woodhouse, M. T., Regayre, L. A., and Pierce, J. R.: Large contribution of natural aerosols to uncertainty in indirect forcing, Nature, 503, 67–71, https://doi.org/10.1038/nature12674, http://www.nature.com/nature/journal/v503/n7474/full/nature12674.html, 2013.

Cess, R. D., Potter, G. L., Blanchet, J. P., Boer, G. J., Ghan, S. J., Kiehl, J. T., Treut, H. L., Li, Z.-X., Liang, X.-Z., Mitchell, J. F. B., Morcrette, J.-J., Randall, D. A., Riches, M. R., Roeckner, E., Schlese, U., Slingo, A., Taylor, K. E., Washington, W. M., Wetherald,

25   R. T., and Yagai, I.: Interpretation of Cloud-Climate Feedback as Produced by 14 Atmospheric General Circulation Models, Science, 245, 513–516, https://doi.org/10.1126/science.245.4917.513, http://science.sciencemag.org/content/245/4917/513, 1989.

Cess, R. D., Potter, G. L., Zhang, M.-H., Blanchet, J.-P., Chalita, S., Colman, R., Dazlich, D. A., Genio, A. D. D., Dymnikov, V., Galin, V., Jerrett, D., Keup, E., Lacis, A. A., Treut, H. L., Liang, X.-Z., Mahfouf, J.-F., Mcavaney, B. J., Meleshko, V. P., Mitchell, J. F. B., Morcrette, J.-J., Norris, P. M., Randall, D. A., Rikus, L., Roeckner, E., Royer, J.-F., Schlese, U., Sheinin, D. A., Slingo, J. M., Sokolov,

30   A. S., Taylor, K. E., Washington, W. M., Wetherald, R. T., and Yagai, I.: Interpretation of Snow-Climate Feedback as Produced by 17 General Circulation Models, Science, 253, 888–892, https://doi.org/10.1126/science.253.5022.888, http://science.sciencemag.org/content/253/5022/888, 1991.

Chen, Y. and Penner, J. E.: Uncertainty analysis for estimates of the first indirect aerosol effect, Atmospheric Chemistry and Physics, 5, 2935–2948, https://doi.org/10.5194/acp-5-2935-2005, http://www.atmos-chem-phys.net/5/2935/2005/, 2005.

35   Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., Briegleb, B. P., Bitz, C. M., Lin, S.-J., and Zhang, M.: The Formulation and Atmospheric Simulation of the Community Atmosphere Model Version 3 (CAM3), Journal of Climate, 19, 2144–2161, https://doi.org/10.1175/JCLI3760.1, http://journals.ametsoc.org/doi/abs/10.1175/JCLI3760.1, 2006.

Covey, C., Lucas, D. D., Tannahill, J., Garaizar, X., and Klein, R.: Efficient screening of climate model sensitivity to a large number of perturbed input parameters, Journal of Advances in Modeling Earth Systems, 5, 598–610, https://doi.org/10.1002/jame.20040, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jame.20040, 2013.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of climate models, pp. 741–882, Cambridge University Press, 2013.

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., and Zelinka, M.: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models, Journal of Geophysical Research: Atmospheres, pp. n/a–n/a, https://doi.org/10.1002/jgrd.50174, http://onlinelibrary.wiley.com/doi/10.1002/jgrd.50174/abstract, 2013.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z.-L., and Zhang, M.: The Community Climate System Model Version 4, Journal of Climate, 24, 4973–4991, https://doi.org/10.1175/2011JCLI4083.1, http://journals.ametsoc.org/doi/abs/10.1175/2011JCLI4083.1, 2011.

Golaz, J.-C., Golaz, J.-C., and Levy, H.: Cloud tuning in a coupled climate model: Impact on 20th century warming, Geophysical Research Letters, 40, 2246–2251, https://doi.org/10.1002/grl.50232, http://onlinelibrary.wiley.com/doi/abs/10.1002/grl.50232/abstract, 2013.

Hourdin, F., Foujols, M.-A., Codron, F., Guemas, V., Dufresne, J.-L., Bony, S., Denvil, S., Guez, L., Lott, F., Ghattas, J., Braconnot, P., Marti, O., Meurdesoif, Y., and Bopp, L.: Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model, Climate Dynamics, 40, 2167–2192, https://doi.org/10.1007/s00382-012-1411-3, https://link.springer.com/article/10.1007/s00382-012-1411-3, 2013.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M., and Williamson, D.: The art and science of climate model tuning, Bulletin of the American Meteorological Society, https://doi.org/10.1175/BAMS-D-15-00135.1, http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-15-00135.1, 2016.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., and Vertenstein, M.: The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability, Bulletin of the American Meteorological Society, 96, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1, http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-13-00255.1, 2014.

Kiehl, J. T.: Twentieth century climate model response and climate sensitivity, Geophysical Research Letters, 34, L22 710, https://doi.org/10.1029/2007GL031383, http://www.agu.org.proxy.lib.uwaterloo.ca/pubs/crossref/2007/2007GL031383.shtml, 2007.

Kim, H., Kang, S. M., Hwang, Y.-T., and Yang, Y.-M.: Sensitivity of the Climate Response to the Altitude of Black Carbon in the Northern Subtropics in an Aquaplanet GCM, Journal of Climate, 28, 6351–6359, https://doi.org/10.1175/JCLI-D-15-0037.1, http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-15-0037.1, 2015.

Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, Atmos. Chem. Phys., 10, 7017–7039, https://doi.org/10.5194/acp-10-7017-2010, http://www.atmos-chem-phys.net/10/7017/2010/, 2010.

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., and Spracklen, D. V.: Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters, Atmos. Chem. Phys., 11, 12 253–12 273, https://doi.org/10.5194/acp-11-12253-2011, https://www.atmos-chem-phys.net/11/12253/2011/, 2011.

Lee, L. A., Carslaw, K. S., Pringle, K. J., and Mann, G. W.: Mapping the uncertainty in global CCN using emulation, Atmos. Chem. Phys., 12, 9739–9751, https://doi.org/10.5194/acp-12-9739-2012, https://www.atmos-chem-phys.net/12/9739/2012/, 2012.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., and Tomassini, L.: Tuning the climate of a global model, Journal of Advances in Modeling Earth Systems, 4, n/a–n/a, https://doi.org/10.1029/2012MS000154, http://doi.wiley.com/10.1029/2012MS000154, 2012.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, Technometrics, 21, 239–245, https://doi.org/10.2307/1268522, http://www.jstor.org/stable/1268522, 1979.

McNeall, D., Williams, J., Booth, B., Betts, R., Challenor, P., Wiltshire, A., and Sexton, D.: The impact of structural error on parameter constraint in a climate model, Earth Syst. Dynam., 7, 917–935, https://doi.org/10.5194/esd-7-917-2016, https://www.earth-syst-dynam.net/7/917/2016/, 2016.

Medeiros, B., Stevens, B., and Bony, S.: Using aquaplanets to understand the robust responses of comprehensive climate models to forcing, Climate Dynamics, 44, 1957–1977, https://doi.org/10.1007/s00382-014-2138-0, http://link.springer.com/article/10.1007/s00382-014-2138-0, 2014.

Pierce, D. W., Barnett, T. P., Santer, B. D., and Gleckler, P. J.: Selecting global climate models for regional climate change studies, Proceedings of the National Academy of Sciences of the United States of America, 106, 8441–8446, https://doi.org/10.1073/pnas.0900094106, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2689003/, 2009.

Pujol, G., Iooss, B., Boumhaout, A. J. w. c. f. K., Veiga, S. D., Delage, T., Fruth, J., Gilquin, L., Guillaume, J., Gratiet, L. L., Lemaitre, P., Nelson, B. L., Monari, F., Oomen, R., Ramos, B., Roustant, O., Song, E., Staum, J., Touati, T., and Weber, F.: sensitivity: Global Sensitivity Analysis of Model Outputs, https://CRAN.R-project.org/package=sensitivity, 2017.

Regayre, L., Johnson, J., Yoshioka, M., Pringle, K., Sexton, D., Booth, B., Lee, L., Bellouin, N., and Carslaw, K.: Aerosol and physical atmosphere model parameters are both important sources of uncertainty in aerosol ERF, Atmos. Chem. Phys. Discuss., 2018, 1–54, https://doi.org/10.5194/acp-2018-175, https://www.atmos-chem-phys-discuss.net/acp-2018-175/, 2018.

Remer, L. A., Kleidman, R. G., Levy, R. C., Kaufman, Y. J., Tanré, D., Mattoo, S., Martins, J. V., Ichoku, C., Koren, I., Yu, H., and Holben, B. N.: Global aerosol climatology from the MODIS satellite sensors, Journal of Geophysical Research, 113, https://doi.org/10.1029/2007JD009661, http://doi.wiley.com/10.1029/2007JD009661, 2008.

Roustant, O., Ginsbourger, D., and Deville, Y.: DiceKriging , DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization, Journal of Statistical Software, 51, https://doi.org/10.18637/jss.v051.i01, http://www.jstatsoft.org/v51/i01/, 2012.

Saltelli, A., Tarantola, S., and Chan, K. P.-S.: A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output, Technometrics, 41, 39–56, https://doi.org/10.1080/00401706.1999.10485594, https://www.tandfonline.com/doi/abs/10.1080/00401706.1999.10485594, 1999.

Samset, B. H. and Myhre, G.: Climate response to externally mixed black carbon as a function of altitude: BC climate response vs altitude, Journal of Geophysical Research: Atmospheres, pp. n/a–n/a, https://doi.org/10.1002/2014JD022849, http://doi.wiley.com/10.1002/2014JD022849, 2015.

Sanderson, B. M., Knutti, R., Aina, T., Christensen, C., Faull, N., Frame, D. J., Ingram, W. J., Piani, C., Stainforth, D. A., Stone, D. A., and Allen, M. R.: Constraints on Model Response to Greenhouse Gas Forcing and the Role of Subgrid-Scale Processes, Journal of Climate, 21, 2384–2400, https://doi.org/10.1175/2008JCLI1869.1, https://journals.ametsoc.org/doi/full/10.1175/2008JCLI1869.1, 2008.

Sherwood, S. C., Bony, S., and Dufresne, J.-L.: Spread in model climate sensitivity traced to atmospheric convective mixing, Nature, 505, 37–42, https://doi.org/10.1038/nature12829, https://www.nature.com/articles/nature12829, 2014.

Shields, C. A., Bailey, D. A., Danabasoglu, G., Jochum, M., Kiehl, J. T., Levis, S., and Park, S.: The Low-Resolution CCSM4, Journal of Climate, 25, 3993–4014, https://doi.org/10.1175/JCLI-D-11-00260.1, https://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-11-00260.1, 2012.

Siler, N., Po-Chedley, S., and Bretherton, C. S.: Variability in modeled cloud feedback tied to differences in the climatological spatial pattern of clouds, Climate Dynamics, 50, 1209–1220, https://doi.org/10.1007/s00382-017-3673-2, https://link.springer.com/article/10.1007/s00382-017-3673-2, 2018.

Stevens, B.: Rethinking the Lower Bound on Aerosol Radiative Forcing, Journal of Climate, 28, 4794–4819, https://doi.org/10.1175/JCLI-D-14-00656.1, http://journals.ametsoc.org/doi/10.1175/JCLI-D-14-00656.1, 2015.

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T., Chikira, M., Ogura, T., Sekiguchi, M., Takata, K., Yamazaki, D., Yokohata, T., Nozawa, T., Hasumi, H., Tatebe, H., and Kimoto, M.: Improved Climate Simulation by MIROC5: Mean States, Variability, and Climate Sensitivity, Journal of Climate, 23, 6312–6335, https://doi.org/10.1175/2010JCLI3679.1, https://journals.ametsoc.org/doi/abs/10.1175/2010JCLI3679.1, 2010.

Yang, B., Qian, Y., Berg, L. K., Ma, P.-L., Wharton, S., Bulaevskaya, V., Yan, H., Hou, Z., and Shaw, W. J.: Sensitivity of Turbine-Height Wind Speeds to Parameters in Planetary Boundary-Layer and Surface-Layer Schemes in the Weather Research and Forecasting Model, Boundary-Layer Meteorology, 162, 117–142, https://doi.org/10.1007/s10546-016-0185-2, https://link.springer.com/article/10.1007/s10546-016-0185-2, 2017.

Zhao M., Golaz J.-C., Held I. M., Guo H., Balaji V., Benson R., Chen J.-H., Chen X., Donner L. J., Dunne J. P., Dunne K., Durachta J., Fan S.-M., Freidenreich S. M., Garner S. T., Ginoux P., Harris L. M., Horowitz L. W., Krasting J. P., Langenhorst A. R., Liang Z., Lin P., Lin S.-J., Malyshev S. L., Mason E., Milly P. C. D., Ming Y., Naik V., Paulot F., Paynter D., Phillipps P., Radhakrishnan A., Ramaswamy V., Robinson T., Schwarzkopf D., Seman C. J., Shevliakova E., Shen Z., Shin H., Silvers L. G., Wilson J. R., Winton M., Wittenberg A. T., Wyman B., and Xiang B.: The GFDL Global Atmosphere and Land Model AM4.0/LM4.0: 2. Model Description, Sensitivity Studies, and Tuning Strategies, Journal of Advances in Modeling Earth Systems, 10, 735–769, https://doi.org/10.1002/2017MS001209, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001209, 2018.