Ensembles of Global Climate Model Variants Designed for the Quantification and Constraint of Uncertainty in Aerosols and Their Radiative Forcing

M. Yoshioka[1] (iD)**, L. A. Regayre[1]** (iD)**, K. J. Pringle[1], J. S. Johnson[1], G. W. Mann[1,2]** (iD)**, D. G. Partridge[3], D. M. H. Sexton[4], G. M. S. Lister[5], N. Schutgens[6,7], P. Stier[6]** (iD)**, Z. Kipling[6,8]** (iD)**, N. Bellouin[9]** (iD)**, J. Browse[1,10], B. B. B. Booth[4], C. E. Johnson[4], B. Johnson[4]** (iD)**, J. D. P. Mollard[7]** (iD)**, L. Lee[1]** (iD)**, and K. S. Carslaw[1]** (iD)

[1]School of Earth and Environment, University of Leeds, Leeds, UK, [2]National Centre for Atmospheric Science, University of Leeds, Leeds, UK, [3]College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, UK, [4]U.K. Hadley Centre Met Office, Exeter, Exeter, UK, [5]National Centre for Atmospheric Science, University of Reading, Reading, UK, [6]Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK, [7]Now at Cluster Earth & Climate, Faculty of Science, VU University Amsterdam, Amsterdam, The Netherlands, [8]Now at European Centre for Medium-Range Weather Forecasts, Reading, UK, [9]Department of Meteorology, University of Reading, Reading, UK, [10]Center for Geography, Environment and Society, University of Exeter, Penryn, UK

Abstract Tropospheric aerosol radiative forcing has persisted for many years as one of the major causes of uncertainty in global climate model simulations. To sample the range of plausible aerosol and atmospheric states and perform robust statistical analyses of the radiative forcing, it is important to account for the combined effects of many sources of model uncertainty, which is rarely done due to the high computational cost. This paper describes the designs of two ensembles of the Met Office Hadley Centre Global Environment Model-U.K. Chemistry and Aerosol global climate model and provides the first analyses of the uncertainties in aerosol radiative forcing and their causes. The first ensemble was designed to comprehensively sample uncertainty in the aerosol state, while the other samples additional uncertainties in the physical model related to clouds, humidity, and radiation, thereby allowing an analysis of uncertainty in the aerosol effective radiative forcing. Each ensemble consists of around 200 simulations of the preindustrial and present-day atmospheres. The uncertainty in aerosol radiative forcing in our ensembles is comparable to the range of estimates from multimodel intercomparison projects. The mean aerosol effective radiative forcing is -1.45 W/m^2 (credible interval of -2.07 to -0.81 W/m^2), which encompasses but is more negative than the -1.17 W/m^2 in the 2013 Atmospheric Chemistry and Climate Model Intercomparison Project and -0.90 W/m^2 in the Intergovernmental Panel on Climate Change Fifth Assessment Report. The ensembles can be used to reduce aerosol radiative forcing uncertainty by challenging them with multiple measurements as well as to isolate potential causes of multimodel differences.

Plain Language Summary Atmospheric aerosol particles such as dust, pollutants, and smoke interfere with light from the Sun and modify the properties of clouds and thereby affect Earth's climate. However, the effect that aerosols have on climate is one of the major causes of uncertainty in global climate model simulations. We performed a large number of climate model simulations (called an ensemble), with many parts of the model slightly varied, in order to understand the complex behavior of the model and to explore the causes of uncertainty in model outputs. This paper describes the designs of two climate model ensembles and provides the first analyses of the causes of model uncertainty. The first ensemble was designed to comprehensively understand the behavior of aerosols in the atmosphere, while the other includes more general uncertainties in atmospheric processes that can affect aerosols. Each ensemble consists of around 200 simulations. The ranges of the aerosol climate effect in our ensembles are comparable to the ranges of previous estimates from studies that analyzed multiple climate models. These ensembles can be used to reduce uncertainty in how aerosols affect climate by comparing with satellite and ground-based measurements.

AGU
100
ADVANCING EARTH
AND SPACE SCIENCE

Journal of Advances in Modeling Earth Systems

10.1029/2019MS001628

1. Introduction

Earth's radiative energy balance is strongly affected by atmospheric aerosol particles that directly scatter and absorb solar and terrestrial radiation and also indirectly modulate the radiative properties of clouds. Over the industrial period anthropogenic emissions have significantly altered the properties of aerosols (i.e., atmospheric loading, distribution, composition, and size) and caused a change in the Earth's radiative energy balance (radiative forcing [RF]; Boucher et al., 2013; Carslaw, Lee, Reddington, Pringle, et al., 2013; Haerter et al., 2009; Lohmann & Ferrachat, 2010; Myhre et al., 2013; Schulz et al., 2006). Changes in aerosols also have important regional effects on climate, atmospheric circulation, clouds, and precipitation (Bollasina et al., 2011; Bollasina et al., 2013; Booth et al., 2012; Kaufman et al., 2005; Philipona et al., 2009; Rosenfeld et al., 2008; Shindell & Faluvegi, 2009; Smith et al., 2016; Stevens & Feingold, 2009; Wang et al., 2014).

Our ability to robustly quantify the effects of aerosols on the climate ultimately rests on being able to reliably simulate aerosol properties and radiative effects on regional and global scales. However, despite major improvements in global observing systems and climate models representing aerosols more realistically, the estimated net aerosol RF over the industrial period ranges from near zero to around -2 W/m^2 (Boucher et al., 2013). This uncertainty has persisted through all intergovernmental panel assessment reports since 1996 and significantly limits our ability to understand the causes of historical climate change and our confidence in climate change projections (Andreae et al., 2005; Collins et al., 2013; Schwartz, 2018; Seinfeld et al., 2016). In other words, while our knowledge about aerosols and the driving processes has improved, this has not yet been translated into more reliable climate model simulations.

Multimodel intercomparisons provide valuable information about the differences between models, often called the model diversity (Bellouin et al., 2016; Mann et al., 2014; Myhre et al., 2013; Randles et al., 2013; Shindell et al., 2013; Stier et al., 2013). The AeroCom intercomparison of aerosol microphysical properties (particle concentrations and size distributions) showed that global aerosol models can capture many of the important spatial and temporal variations in aerosols on a global scale (Mann et al., 2014). Nevertheless, the models have a very large spread (a factor 2 to 30, depending on region) in their simulations of aerosol microphysical properties. Although model intercomparisons provide valuable information about model diversity, it is difficult to attribute these differences to particular processes or model inputs when there are many structural differences among the models (Stier et al., 2013). Most importantly, the small ensemble size used in multimodel comparison studies inhibits drawing statistically robust conclusions (Carslaw et al., 2018).

A perturbed parameter ensemble (PPE) is a set of simulations created using a single model structure with perturbations made to uncertain physical parameters (Collins et al., 2011) or inputs to the model, like aerosol emissions (Lee et al., 2013). PPEs are widely used in many fields of science to understand and quantify model uncertainty, such as in environmental models (Salter & Williamson, 2016), hydrological models (Liu & Gupta, 2007), astrophysics (Rodrigues et al., 2017), and disease transmission (Andrianakis et al., 2017). PPEs have been used to develop an understanding of uncertainties in climate change projections (Collins et al., 2011; Fletcher et al., 2018; Murphy et al., 2004; Sexton et al., 2012) as well as to explore the causes of uncertainty in global aerosol concentrations (Lee et al., 2011, 2012, 2013), aerosol RF (Carslaw, Lee, Reddington, Pringle, et al., 2013; Regayre et al., 2014, 2015), precipitation (Qian et al., 2015), and radiative fluxes (Shiogama et al., 2012). PPEs can also be used to systematically identify model structural deficiencies (McNeall et al., 2016; Yokohata et al., 2010).

Causes of uncertainty in the aerosol component of a global model have previously been identified using ensembles of "offline" models (often called chemical transport models), where the couplings of aerosols to the radiation and clouds were not accounted for (Carslaw, Lee, Reddington, Mann, et al., 2013, Regayre et al., 2014, 2015). Here we go further by constructing PPEs in a global climate model in which aerosols are coupled to physical climate processes. This enables us to analyze the uncertainty in aerosol RF including atmospheric rapid adjustments (i.e., semidirect effect and cloud lifetime effect; Boucher et al., 2013).

We designed two PPEs to address different aspects of the aerosol RF in the Met Office Hadley Centre Global Environment Model (HadGEM)-U.K. Chemistry and Aerosol (UKCA) aerosol-chemistry-climate model. Both ensembles were designed to be comparable to measurements made under particular meteorological

conditions so were therefore nudged toward analyzed meteorology. Both PPEs also included equivalent ensembles for preindustrial emissions so that the uncertainty in aerosol RF can be quantified. The first "aerosol" ensemble (hereafter referred to as the AER PPE) perturbs 26 parameters related to aerosol processes and emissions and was run in such a way that changes in aerosols did not affect the meteorology. The AER PPE therefore allows the aerosol RF to be diagnosed. The second "aerosol-atmosphere" ensemble (hereafter referred to as the AER-ATM PPE) was designed to quantify the relative importance of uncertain parameters in the aerosol model and the host physical atmosphere model. The AER-ATM PPE therefore allows the main causes of uncertainty in aerosol effective RF (ERF, Boucher et al., 2013) to be diagnosed.

Outputs from PPEs are subsequently used to create Gaussian process emulators (Lee et al., 2011; O'Hagan, 2006) to enable several million "model variants" to be generated across the multidimensional parameter spaces. This approach enables a sufficiently high sampling density to perform robust statistical analyses of model parametric uncertainty. In our analysis we represent aerosol forcing uncertainties using standard deviations or 90% credible intervals calculated using samples of one million model variants.

In this article, we describe the design of the two PPEs and examine the major differences between them by comparing the spatial patterns of uncertainty and by contrasting the sensitivities of global mean values to the parameter perturbations. In section 2 we summarize the base model setup, modifications we made to the model structure, the aerosol emissions used in all simulations, the parameters we perturbed, and the design of the PPEs. In section 3 we analyze the major differences between the PPEs in terms of aerosols and aerosol RF. In section 4 we evaluate the potential for these ensembles to inform multiple aspects of climate research, including identifying priorities for model development, model structural deficiencies, and regions where new, targeted observations will better constrain aerosol forcing uncertainty.

2. Methods

2.1. Base Model

This work uses global simulations of the HadGEM3 (Hewitt et al., 2011) with N96 horizontal resolution (1.25° in latitude and 1.875° in longitude) and 85 vertical levels up to 85 km above sea level. The model uses the fourth Global Atmosphere configuration (Walters et al., 2014) of the Met Office Unified Model. Sea surface temperatures and sea ice were prescribed using reanalyzed monthly varying fields based on the methodology of Reynolds et al. (2007) as used in phase 2 of the Atmospheric Model Intercomparison Project.

We used version 8.4 of the UKCA model, which is a whole atmosphere chemistry and aerosol model within the HadGEM3 host model (Abraham, 2014; O'Connor et al., 2014). Within the UKCA model the evolution of particle size distribution and size-resolved chemical composition of aerosols are calculated using the modal version of the GLObal Model of Aerosol Processes model (GLOMAP-mode; Mann et al., 2010). The GLOMAP-mode simulates new particle formation, gas-to-particle transfer, aerosol coagulation, cloud processing of aerosols, and both dry and wet deposition of gases and aerosols. In our model setup GLOMAP resolves five aerosol components (sulfate, organic carbon, black carbon, sea salt, and dust) in seven internally mixed modes with a lognormal number size distribution (soluble modes in nucleation, Aitken, accumulation, and coarse size ranges and insoluble modes in all but the nucleation size range).

Particles form through binary homogeneous nucleation (Vehkamäki et al., 2002) throughout the atmosphere and through organically mediated nucleation (Metzger et al., 2010), which occurs mainly in the planetary boundary layer close to biogenic volatile organic carbon compound emissions. Particles grow following microphysical processes such as condensation of gas species and coagulation between particles. They are moved from one mode to another when the mean modal size becomes larger than a prescribed threshold or when insoluble particles are aged to become soluble. Secondary organic carbon is produced from the first stage oxidation products of biogenic monoterpenes under the assumption of zero vapor pressure. Secondary organic material is then combined as one chemical tracer with other organic carbon after condensing to the particle phase.

Aerosol particles are removed from the atmosphere through dry deposition (gravitational settling and turbulent mixing) via nucleation scavenging (activation to cloud drops followed by autoconversion to rain) and via impaction scavenging (by precipitating rain drops; Kipling et al., 2013).

Hygroscopic growth of soluble particles occurs according to the local relative humidity following Zadanovskii (1948) and Stokes and Robinson (1966) using water activity coefficient data from Jacobson et al. (1996). Activation of aerosols to cloud droplets is calculated applying κ-Köhler theory (Petters & Kreidenweis, 2007) with composition-dependent hygroscopicity factors (κ; kappa) and an updraft velocity probability density function (PDF) representing subgrid variability of vertical velocities (West et al., 2014).

Aerosol optical properties (extinction, absorption, and asymmetry parameter) for each of the GLOMAP internally mixedsize modes are calculated using the RADAER module (Bellouin et al., 2013), and aerosol radiative effects are calculated via the Met Office Unified Model radiative transfer module (Edwards & Slingo, 1996).

Note that to improve the fidelity of predicted aerosol RFs in these ensembles we evolved the GLOMAP codebase in UM-UKCA for a number of key improvements (the updated model GLOMAP v8.1). We document these improvements as part of this article, and the changes described are added as a progression from the v8.0 codebase, which enabled GLOMAP to be applied across both tropospheric and stratospheric conditions (Dhomse et al., 2014).

Tropospheric ozone, HO_x, NO_x, and VOC (volatile organic compound) chemistry are calculated using the mechanism described by O'Connor et al. (2014), which includes reactions of odd oxygen (O_x), nitrogen oxides (NO_y), hydrogen radicals ($HO_x = OH + HO_2$) and CO, as well as methane and other short chain nonmethane VOCs. The scheme has been extended to include additional sulfur (Mann et al., 2010) and monoterpene (Spracklen et al., 2006) chemistry. The Fast-JX scheme is implemented to calculate photolysis rates based on simulated aerosols, liquid and ice cloud water content, and ozone. Fast-JX is a version of the Fast-J scheme (Wild et al., 2000) with an extended wavelength range suitable for stratospheric conditions (Abraham et al., 2012; Neu et al., 2007; Telford et al., 2013). Dry and wet deposition of gas-phase species is described in O'Connor et al. (2014). The UKCA chemistry module is currently one-way coupled to the aerosol module, which means that prognostic aerosols are calculated within the atmospheric chemistry scheme but these prognostic aerosols do not affect atmospheric chemistry (through either atmospheric oxidants, photolysis, or heterogeneous reactions) during a model time step.

2.2. Aerosol Sources

Emission fluxes of sea salt and dust are calculated within the model as a function of surface wind speed and the emissions of other primary particles and aerosol precursor gases are prescribed using emission inventories. Emissions of dimethyl sulfide (DMS), a precursor of sulfate aerosol, are calculated as a function of surface wind speeds following Nightingale et al. (2000), based on prescribed surface seawater DMS concentrations given by Kettle and Andreae (2000). The emission inventories used for aerosols and precursor gases in recent decades (1978, 1998, and 2008) and for the preindustrial atmosphere are shown in Tables S1 and S2 in the supporting information.

Sulfur dioxide (SO_2) emissions from anthropogenic (fossil fuel [FF] and biofuel) and volcanic sources, and carbonaceous (i.e., organic carbon and black carbon) emissions from combustion (biomass burning [BB]) sources are prescribed distinctly for each period (Tables S1 and S2). However, in AER-ATM PPE the surface emission of anthropogenic sulfur dioxide was erroneously duplicated and given as high-level emissions representing the industrial and half of energy sector emissions. As a result, global, annual mean high-level sulfur dioxide emission is 7.67e-13 kg·m^{-2}·s^{-1}, which is smaller by a factor of 3 than it should be (2.31e-12 kg·m^{-2}·s^{-1}) in this PPE. This caused differences in global budgets of sulfate aerosol between two PPEs (Tables S7 and S8) and led to weaker forcing over industrial regions in AER-ATM PPE (section 3). Biogenic VOC emissions of isoprene and monoterpenes are the same for both periods following Guenther et al. (1995). As stated in section 2.4, the meteorology is identical in all AER simulations. Therefore, the emissions of aerosols and precursors (sea salt, dust, and DMS flux) are simulated identically before perturbation. In the AER-ATM simulations, however, rapid adjustments in the boundary layer as well as perturbations to host model physical parameters cause the calculated emission fluxes of sea salt, dust, and DMS to differ between simulations.

2.3. Structural Modifications to the Model

The baseline GLOMAP-mode aerosol model in UKCA is essentially as described in Mann et al. (2010) with small adjustments to the widths of the particle modes (Mann et al., 2012). In this section we describe the improvements made to the release version of the model and how they benefit the model setup used to create the ensembles. Structural improvements in model release version 8.4 were identified during the design phase of these ensembles. The main changes to the model for this study were the following.

2.3.1. New Particle Formation

The release version of the model has multiple options for new particle formation (particle nucleation) in the boundary layer. Here we use an organically mediated parametrization (Metzger et al., 2010). In the original formulation it was assumed that 13% of α-pinene first-stage oxidation products produce secondary organic aerosol (SOA) and also enter into the nucleation rate expression, which gave realistic results (Metzger et al., 2010). In the PPEs we now allow this 13% SOA yield to vary over a wide range to represent a realistic uncertainty in total SOA production. To span realistic ranges of observed particle concentrations we therefore found it necessary to allow only one tenth of this SOA to participate in nucleation (see section 2.5.2).

2.3.2. Organic Aerosol Water Uptake

We previously assumed that water uptake per mole of the organic component of the particles was 65% of that of sulfate (Mann et al., 2010). In this study water uptake is calculated using the κ-Köhler approach (Petters & Kreidenweis, 2007) to be consistent with the use of κ in cloud droplet activation (West et al., 2014).

2.3.3. Cloud Droplet Activation

Cloud droplet activation is calculated using the parameterization of Abdul-Razzak and Ghan (2000; West et al., 2014), with a varying updraft velocity, including Gaussian PDF for subgrid variability whose width is linked to the turbulent kinetic energy in the model. In the PPEs we made this width independent of turbulent kinetic energy so that its value could be defined more precisely. Above the boundary layer a fixed value of 0.1 is assumed as in the base model.

2.3.4. In-Cloud Aerosol Scavenging

In our previous studies we perturbed the size of aerosol particles at which scavenging occurs (Lee et al., 2013; Carslaw, Lee, Reddington, Mann, et al., 2013) and in later studies the drizzle rate (Regayre et al., 2014, 2015). Within UM-UKCA, the host general circulation model has large-scale precipitation and cloud schemes, so we elected to perturb the parameter controlling the raining fraction of the grid box (Rain_Frac). This is applied to large-scale precipitating cloud but not to convective precipitation. We did not perturb the autoconversion rate which is a potentially important parameter (see section 4).

2.3.5. Aerosol Scavenging in Mixed-Phase Clouds

The scavenging of aerosol material in dynamic rain systems is controlled partly by the rain formation process. Our previous work has shown that if aerosol scavenging scales with total water removal (as was originally the case in the base HadGEM3-UKCA model), then too little aerosol is transported into the Arctic (Browse et al., 2012). In our previous studies in the GLOMAP model without any cloud microphysics we therefore introduced an adjustable in-cloud temperature threshold below which scavenging was switched off (to account for less efficient scavenging in mixed-phase clouds). To exploit the HadGEM3-UKCA cloud microphysics we instead define an ice mass fraction threshold above which no nucleation scavenging occurs. We assume that the wet scavenging of all aerosol particles (soluble and insoluble) is set to 0 in large-scale raining clouds if the simulated ice to total water mass fraction is higher than a fixed value (determined by the new parameter Cloud_Ice_Thresh).

In our preliminary experiment, the newly implemented Rain_Frac and Cloud_Ice_Thresh parameters were found to improve the agreements of simulated aerosol concentrations with station observations in the Arctic (see Figure S1).

2.3.6. Dust in the Modal Aerosol Scheme

In the release version of HadGEM3 dust is simulated in the CLASSIC bin scheme (Bellouin et al., 2007; Woodward, 2001) separately from all other aerosols (which are treated in GLOMAP) so that interactions between dust and other aerosol species are not simulated. However, recent research shows that dust emissions can have important regional influences on cloud drop concentrations (Karydis et al., 2017). Therefore, in our ensembles we emit dust into the GLOMAP modal scheme where the particles can be aged through condensation of soluble materials (such as sulfate and organic carbon). Dust particles are transferred into soluble modes when they acquire sufficient soluble materials to be considered hygroscopic (as

defined by the "Aging" parameter; sections 2.4 and 2.5). Emissions in bins 1 (0.0316–0.1 μm in radius) and 2 (0.1–0.316 μm) and half of the emission in bin 3 (0.316–1 μm) are transferred into accumulation insoluble mode. The remaining emissions in bin 3 and emissions in bins 4 (1–3.16 μm) and 5 (3.16–10 μm) are transferred into the coarse insoluble mode. Dust in bin 6 (10–31.6 μm) is not currently transferred to the GLOMAP modal scheme because this size of particles is not properly represented the GLOMAP modal scheme. These particles have very short atmospheric life time and hence are not affected by microphysical processes.

2.3.7. Aerosol Optical Properties

We replaced the lookup tables of aerosol optical properties (Bellouin et al., 2016) that include optical properties for mineral dust (Balkanski et al., 2007) with a higher resolution in the increments of the imaginary part of the refractive index to better resolve the specific absorption coefficient of aerosols, especially at the low-absorption end. This development was necessary to obtain aerosol optical depth (AOD) in accumulation and coarse insoluble modes where dust is the only aerosol component. However, the optical properties for mineral dust used in AER-ATM are based on Woodward (2001), leading to increased absorption at visible wavelengths. This results in higher-absorption AOD shown in Table S10. Since the absorption by dust applies to both present-day and preindustrial simulations, this has a minimal effect on aerosol forcing.

2.4. Model Setup

As described in section 1 we have created two complementary PPEs designed to tackle multiple aspects of aerosol forcing uncertainty quantification and reduction. The HadGEM-UKCA model was set up differently for each ensemble, as summarized in Table 1.

2.4.1. The AER Aerosol PPE

The AER PPE was designed to understand and constrain the uncertainty in model output resulting from perturbations to uncertain aerosol parameters using high spatial and temporal resolution observations (Reddington et al., 2017). For this purpose, we wanted to have identical meteorology for all ensemble members to remove the noise caused by differences in dynamics and cloud properties (internal model variability).

In AER the double call configuration (Bellouin et al., 2013) is used to calculate radiative fluxes. In this configuration the radiation scheme is called twice to calculate radiative fluxes with and without simulated aerosols. The model dynamics and physics in the following time step is calculated using the "prognostic call" with no radiative perturbation by aerosols. Therefore, the changes in aerosol concentrations and radiative properties occurs only in the "diagnostic call" and do not affect temperature, clouds, humidity, or precipitation during the simulations. Although atmospheric oxidants are calculated interactively within UKCA, the radiative effects of these oxidants in the diagnostic call do not affect the simulation in the next time step. Thereby, meteorological fields are completely decoupled from aerosols and chemistry and are identical across all simulations in AER. In addition, we use the same trace gas concentrations such as CO_2, CH_4, N_2O, and chlorofluorocarbons and other boundary conditions for the preindustrial (1850) and present-day (2008) simulations. Since these pairs of simulations also share all parameter values, the sole difference between each pair of simulations is anthropogenic and volcanic aerosol emissions (section 2.2). Radiative forcing is calculated as the difference in shortwave and longwave radiation at the top of the atmosphere in the diagnostic call for pairs of simulations with 2008 and 1850 anthropogenic emissions. Clear-sky RF has been diagnosed in the same way using clear-sky radiative fluxes that have been calculated within the model ignoring radiative effects of clouds.

Horizontal wind and temperature fields are relaxed (Telford et al., 2008) toward European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA)-Interim values between model levels 12 (around 1 km) and 80 (around 60 km). This makes the model behave like GLOMAP-mode coupled to a global chemistry-transport model driven by meteorological fields from reanalyses. The benefit of coupling GLOMAP to HadGEM-UKCA is that, among others such as better convective transport, the atmospheric model simulates cloud fields explicitly rather than being prescribed according to observed climatologies.

2.4.2. The AER-ATM Aerosol-Atmospheric PPE

The AER-ATM PPE has two main purposes. First, it is designed to allow the investigation of uncertainty in aerosol ERF. Therefore, rapid atmospheric adjustments to clouds, water vapor, and temperature resulting from the radiative effects of aerosols (e.g., the cloud lifetime effect and the semidirect effect) are active in these simulations. Second, it is designed to quantify the relative importance of aerosol and physical

Table 1
Major Differences Between Model Setup for AER and AER-ATM

Ensemble	AER	AER-ATM
Perturbed parameters	26 aerosol parameters	27 aerosol, physical, and radiative parameters
Main outputs of interest	Aerosol state and radiative forcing	Effective radiative forcing
Radiation scheme	Double-call	Double-call with clear- and clean-air diagnostics
Nudging	Nudged to horizontal winds and temperature at and above level 12 (around 1,000 m)	Nudged to horizontal winds at and above level 17 (around 2,000 m)
Rapid atmospheric adjustments to aerosols	Completely turned off	Allowed but effect evident below meteorological nudging level
Anthropogenic aerosol emission years	1850 and 2008	1850, 1978, 1998, and 2008
ERA-Interim meteorological data (year)	2008	2006

atmosphere model parameters as causes of aerosol ERF uncertainty. Therefore, the nudging is applied in a manner that retains the effects of physical atmosphere parameters within the boundary layer, while ensuring all ensemble members have identical synoptic-scale features. Specifically, atmospheric temperatures evolve freely in these simulations and horizontal wind fields are relaxed toward ERA-Interim values at and above model vertical level 17 (around 2 km).

In AER-ATM the radiative flux double-call method is configured so that changes in aerosols interact with meteorology and the above-cloud aerosol radiative effects can be accounted for (Ghan, 2013). Aerosol ERF is calculated as the difference in outgoing shortwave and longwave radiation at the top of the atmosphere between pairs of simulations with distinct anthropogenic emissions (1850, 1978, 1998, and 2008). Similarly, the components of aerosol ERF (the aerosol-cloud interaction ERF_{ACI} and aerosol-radiation interaction ERF_{ARI}) were calculated as the difference between clear-sky, clean-sky, and clear-clean-sky radiative fluxes (following Ghan, 2013) from simulations with different anthropogenic emissions.

2.5. Parameter Uncertainty

2.5.1. Expert Elicitation of Uncertain Parameters

We followed the recommendations for expert elicitation of uncertainty outlined in the Sheffield Elicitation Framework (Oakley & O'Hagan, 2010) to determine the choice of uncertain parameters to be included in the ensembles and the probability distributions of their values. Multiple expert elicitation meetings were conducted over the course of several months. During these meetings statisticians facilitated discussions between small panels of experts to elicit their understanding of the uncertainties in the model parameters as evidenced by the current state of research in the relevant field. Probability distributions based on the discussions were created as the discussion progressed and were used to elicit further information from the experts. We predominantly use trapezoidal probability density functions in "R" (Hetzel, 2012) to represent parameter uncertainty, in a similar way to Sexton et al. (2018). Using univariate trapezoidal distributions avoids oversampling the central region of multidimensional parameter space. Furthermore, experts more readily converged on agreement about important aspects of parameter uncertainty when using trapezoidal distributions (see Figure S2) than they did with other types of distribution.

Probability distributions of aerosol parameter values perturbed in AER are shown in Figure S3. Probability distributions for parameters included in previous ensembles (e.g., Lee et al., 2013; Regayre et al., 2014) were reelicited for this study to account for improved process understanding from recent research activity.

2.5.2. The AER Parameters

The 26 aerosol parameters and their perturbation ranges in AER are summarized in Table S3. Thirteen aerosol parameters are related to microphysical aerosol processing, and another 13 are related to aerosol emission and deposition fluxes. Twenty-one parameters have previously been identified as important causes of model uncertainty in our earlier studies (Carslaw, Lee, Reddington, Mann, et al., 2013; Lee et al., 2013; Regayre et al., 2014; Regayre et al., 2015). These parameters are described briefly here and more fully in Lee et al. (2013).

1. **BL_Nuc**: The boundary layer nucleation rate. We use the parameterization of Metzger et al. (2010) in which the nucleation rate depends on concentrations of gaseous sulfuric acid and condensable organic molecules. We perturb the nucleation rate coefficient.

2. **Particle aging**: The aging rate. Black carbon, organic carbon, and dust are nonhygroscopic when freshly emitted but assumed to become soluble following condensation of sulfuric acid and condensable organic matter. In UKCA particles in the insoluble modes are transferred to soluble modes after acquiring soluble materials. This parameter defines the amount of soluble material required for this transfer to occur as the number of monolayers.

3. **Acc_Width**, **Ait_Width**: Width of accumulation and Aitken modes. GLOMAP uses fixed geometric widths of the lognormal size distribution modes defined by the standard deviation of the size distribution. The uncertainty range was chosen based mainly on Heintzenberg et al. (2004) and Birmili et al. (2001). Changing the mode width modifies the size distribution for particles in that mode, which in turn affects aerosol properties and processes such as dry and wet deposition rates, cloud drop activation, and aerosol optical properties as well as aerosol diagnostics such as aerosol number concentrations in given size ranges.

4. **Cloud_pH**: The acidity of cloud droplets. The rate of in-cloud sulfate formation through oxidation of SO_2 by ozone is highly sensitive to the pH of cloud droplets, thereby affecting the sulfate aerosol mass and aerosol radiative effects (Turnock, 2016). In contrast to the previous studies where cloud pH parameters were set separately for clean and polluted environments, here a single parameter is specified globally since we perturb cloud pH in different periods.

5. **Carb_FF_Ems**, **Carb_BB_Ems**, **Carb_Res_Ems**: The carbonaceous particle emission fluxes from FF, BB, and residential fuel (Res) consumption. The primary emission fluxes of black carbon and organic carbon are scaled by the same amounts.

6. **Carb_FF_Diam**, **Carb_BB_Diam**, **Carb_Res_Diam**: The carbonaceous particle emission diameters from FF, BB, and Res consumption. These parameters directly control the number of emitted particles for a given mass flux and therefore affect various aerosol microphysical processes. Although recent measurements provide some information about emitted particle number concentrations (Janhäll et al., 2010), the particle size remains highly uncertain. The size of FF combustion particles depends on the source. BB particle size depends on burning efficiency (Janhäll et al., 2010) among other factors that are encompassed within our ensemble range.

7. **Prim_SO4_Frac**, **Prim_SO4_Diam**: Fraction and diameter of subgrid-scale formation of sulfate particles. These parameters describe the formation of particles in subgrid-scale plumes, such as power plants (Luo & Yu, 2011; Mather et al., 2003; Stevens et al., 2012). Prim_SO4_Frac defines the fraction of the SO_2 mass that enters the model grid square as new sulfate particles, and Prim_SO4_Diam defines the size of these particles. Previous studies have shown this to be an important source of global cloud condensation nuclei (CCN; Spracklen et al., 2005, Pierce & Adams, 2006, Luo & Yu, 2011), but other studies suggest a more limited effect (Stier et al., 2006), likely because of very different assumptions about the emission size distributions used. We base our ranges on the plume-scale study of Stevens et al. (2012).

8. **Sea_Spray**: Sea spray emission flux. Uncertainties in the wind-driven mass flux of sea spray particles are accounted for by scaling the interactively computed flux within the model with this parameter. This parameter conflates multiple sources of uncertainty: the function describing the wind-speed dependence of the flux, processes that are unaccounted for in the existing parameterizations (such as fetch and sea state), the wind speed itself, and the effect of spatial resolution of the wind fields used by the model.

9. **Anth_SO2**: Anthropogenic SO_2 emission flux. Uncertainty in anthropogenic emission mass flux of SO_2 is accounted for by scaling the baseline emission from Granier et al. (2011) by this parameter.

10. **Volc_SO2**: Volcanic SO_2 emissions. Volcanic SO_2 emissions from continuously degassing volcanoes and time-averaged sporadic eruptions are scaled by this parameter. The baseline emissions are based on Andres and Kasgnoc (1998) and Halmer et al. (2002) data sets.

11. **BVOC_SOA**: Biogenic emission for SOA production. In our model, SOA material is produced through multistep oxidation reactions of the precursor gases (i.e., biogenic monoterpenes) by atmospheric oxidants. The elicited quantity was the global annual SOA production (teragrams per year) and in the model we scale the percentage yield, which therefore conflates the uncertainties in the precursor gas emissions and in the yield of the SOA itself. The baseline emission is 47.5 Tg per year.

12. **DMS**: Dimethyl sulfide emission flux. DMS emissions are calculated interactively based on the seawater concentration of DMS from Kettle and Andreae (2000) and the wind-driven sea-air exchange parameterization (Nightingale et al., 2000). We conflate uncertainties in these two factors by varying the calculated sea-air transfer flux by a given factor.

13. **Dry_Dep_Ait**, **Dry_Dep_Acc**: Dry deposition of Aitken (Ait) and accumulation (acc) mode particles. GLOMAP calculates the dry deposition velocities due to Brownian diffusion, impaction, and interception depending on wind speeds, particle sizes, and surface types following Slinn (1982). In the perturbed runs, the calculated dry deposition velocities in a time step over different surface type are scaled by parameters for particle size ranges.

14. **Dry_Dep_SO2**: The dry deposition velocity of gaseous SO_2. SO_2 is a precursor for aerosol sulfate (SO_4) and new particle formation. Therefore, the removal of atmospheric SO_2 can limit SO_4 formation and the growth of existing aerosols.

15. Five parameters were either newly introduced because they were shown to be important in recent studies or have been included as new parameters corresponding to the structural advances in the modeling framework. These five parameters are described below. The probability distributions of parameter values are presented in Figure S3.

16. **Kappa_OC**: Hygroscopicity parameter of organic carbon. In the model aerosol water uptake efficiency is determined by κ-Köhler theory using size- and composition-dependent hygroscopicity factors (κ; Petters & Kreidenweis, 2007). Perturbations to this parameter will change CCN concentration and cloud drop activation and hence affect the indirect RF (Liu & Wang, 2010; Rastak et al., 2017).

17. **Sig_W**: Updraft vertical velocity standard deviation. This parameter controls the width of the probability distribution of subgrid vertical velocities used to calculate the activation of aerosols into cloud drops (West et al., 2014). A higher value of Sig_W enables higher maximum subgrid updraft velocity and hence higher maximum supersaturation, which allows smaller particles to activate. This in turn leads to changes in cloud properties and precipitation such as higher cloud droplet concentration, lower precipitation efficiency through lower autoconversion rates, higher liquid water path, and higher cloud albedo. Sig_W has greater impacts on cloud properties in regions of relatively high aerosol concentrations where droplet activation is updraft limited rather than aerosol limited.

18. **Dust**: Dust emission flux scale factor. The emission flux of mineral dust is calculated interactively within the model using the CLASSIC bin scheme then transferred into the UKCA modal scheme (as described in section 2.3). The dust emission scheme relies on a tuning factor to account for emission dependence on model resolution and the representation of microenvironment (Woodward, 2001).

19. **Rain_Frac:** The fraction of the cloud covered area where rain forms. As described in section 2.3 we introduced this parameter as the fraction of the cloudy part of the grid box where rain sees the aerosol and thereby aerosol scavenging is allowed to take place. In the default version of the model the raining fraction of the grid box is assumed to be the same as the cloud fraction, and therefore, it corresponds to the case with Rain_Frac = 1. An overestimation of aerosol scavenging in the base model leading to insufficient aerosol loadings over Arctic regions (e.g., Figure S1) stimulated the inclusion of this parameter in our ensembles.

20. **Cloud_Ice_Thresh**: Threshold of cloud ice-phase fraction for scavenging. We introduced this parameter in conjunction with the Rain_Frac parameter. It is designed to suppress aerosol scavenging by both cloud liquid and ice water in dynamic clouds when the cloud ice water fraction is higher than the threshold defined by the parameter value.

2.5.3. The AER-ATM Parameters

In AER-ATM, 27 uncertain parameters are perturbed. Sixteen of the 18 aerosol parameters included here are also included in AER (as indicated in Table S3). Two additional aerosol parameters and nine physical atmosphere parameters were also included in AER-ATM. The effect of these parameter perturbations would be suppressed by the prescribed meteorological fields in AER. We did not include a comprehensive set of physical atmosphere parameters (e.g., Sexton et al., 2018) in AER-ATM because the prescribed meteorological fields force synoptic-scale features to be very similar across the ensemble. Additional parameters included in AER-ATM are summarized in Table S4, and the corresponding probability distributions are presented in Figure S4. The climatic effects of AER-ATM parameters are described in Regayre et al. (2018), and further detail on the elicitation of physical atmosphere parameters can be found in Sexton et al. (2018). We briefly describe the additional aerosol and physical atmosphere parameters below.

1. **Rad_Mcica_Sigma**: The fractional standard deviation of the subgrid cloud condensate as seen by radiation. This parameter controls the inhomogeneity of cloud condensate within vertically overlapping subgrid clouds, which is used to calculate cloud radiative fluxes. This parameter affects the cloud albedo calculated in the radiation module and thus alters atmospheric temperatures in AER-ATM, which can induce changes in precipitation rates and cloud amount.

2. **C_R_Correl**: Cloud-rain subgrid variability correlation coefficient. Different PDFs are used to represent the subgrid variability of cloud and rain, as detailed in Boutle et al. (2014). This parameter determines the degree of correlation between those PDFs, which is typically less than 1 due to differences in how turbulence evolves the PDFs, and the fact that rain can be falling through cloud from which it did not originate. The correlation determines the accretion rate of cloud droplets by rain droplets, which in turn affects the scavenging of aerosol, in-cloud aerosol concentrations, and cloud amount.

3. **Niter_Bs:** The number of cloud microphysics substeps. This discrete parameter controls the number of substeps over which the cloud microphysics is called within each model time step. Increasing the frequency of calls allows more accurate treatment of any nonlinear interactions. An example of this is the interaction of drizzle drops with the environmental air; with more frequent calculations there is more opportunity for evaporation of these drops. Although this is a discrete model parameter it is treated as continuous in the emulation process.

4. **Ent_Fac_Dp**: The entrainment amplitude scale factor. This convection scheme parameter controls the shape of the convective mass flux and the sensitivity of convection to relative humidity. This parameter affects convective depth, precipitation, cloud amount, and hence the spatial distribution of aerosols and clouds. Increased values lead to the reduction of convection depth and to some extent suppression of active precipitating convection.

5. **Amdet_Fac**: The mixing detrainment rate scale factor. This parameter controls the rate of humidification of the atmosphere and the shape of the convective heating profile. Increasing this parameter increases large-scale atmospheric humidity and affects temperatures and thus convective processes.

6. **Dbsdtbs_Turb_0**: The cloud erosion rate. This parameter defines the rate at which unresolved subgrid motions mix clear and cloudy air. Therefore, this parameter affects cloud liquid water content, the autoconversion of cloud droplets to rain drops and cloud amount. Increasing this parameter value tends to reduce the atmospheric lifetimes of aerosols and precursor gases.

7. **Mparwtr**: The maximum value of the function controlling convective parcel maximum condensate. In the model precipitation is triggered in convective parcels near the Earth's surface when the amount of moisture reaches a threshold set by this parameter. This parameter affects cloud amount and lifetime by altering convective precipitation.

8. **Dec_Thres_Cld**: The threshold for cloudy boundary layer decoupling. This parameter defines a threshold of buoyancy flux at which the boundary layer decouples from the rest of the atmosphere. It therefore affects the extent of boundary layer mixing, cloudiness, and ultimately the time available for in-cloud aerosol processing.

9. **Fac_Qsat**: The rate of change of convective parcel maximum condensate with altitude. The maximum amount of moisture a convective parcel can hold transitions from the threshold set by the parameter *Mparwtr* at the surface to a much smaller threshold at high altitudes. This parameter affects precipitation efficiency and cloud amount by altering the rate at which the moisture threshold decreases with altitude.

10. **BC_RI**: The imaginary part of the black carbon refractive index. This additional aerosol parameter controls the absorption of radiation as it passes through black carbon-containing aerosol particles. The real part of the refractive index was adjusted according to the imaginary part assuming a linear relationship between real and imaginary parts (Bond & Bergstrom, 2006). Therefore, this parameter affects the absorption, reemission, refraction, and reflection of incoming radiation by black carbon aerosols.

11. **OC_RI**: The imaginary part of the organic carbon refractive index. This parameter controls the absorption of radiation by organic carbon. The real part of the organic carbon refractive index is not perturbed.

12. Some of the AER parameters (BL_Nuc, Acc_Width, Ait_Width, Carb_FF_Ems, Carb_Res_Ems, Carb_FF_Diam, Carb_Res_Diam, Prim_SO$_4$_Frac, Prim_SO$_4$_Diam, and Dry_Dep_Ait) were omitted from the AER-ATM design during the expert elicitation process. Uncertainty in these parameters was neglected (by setting them to their median parameter values in Table S3) based on evidence

indicating they are only important for aerosols and/or aerosol forcing in a small number of regions and/or seasons (e.g., Lee et al., 2013; Regayre et al., 2014, and one-at-a-time screening experiments using the current model setup). The subset of AER aerosol parameters included in AER-ATM were identified by experts as being the most important for achieving the research goals of AER-ATM (section 2.4.2). Reducing the dimensionality of the ensemble means that the remaining parameter combinations more densely fill the multidimensional parameter space (with the same computational cost) and potentially provide better quality information for statistical model emulator construction (as described in sections 2.6 and 2.10). However, the effect of removing the parameters from the design cannot be known a priori. It is possible that the robustness of the statistical model emulation approach is sufficient to accommodate more parameters and a less dense coverage of the parameter space than we chose to implement here.

2.6. PPE Design

For each ensemble, Maximin Latin Hypercube sampling (Lee et al., 2013; McKay et al., 1979) was used to create a parameter combination design, of $k * N$ points, that spans the N-dimensional uncertain parameter space (where $N = 26$, $k = 7$ for AER and $N = 27$, $k = 6$ for AER-ATM). Experience with the GLOMAP model suggests that with six ensemble members per uncertain parameter (k) a design that sufficiently spans the N-dimensional parameter space (for emulation and sensitivity analysis purposes; section 2.10) can be constructed. Twenty-six members in AER-ATM ensemble failed to complete the 1-year simulation in at least one of three periods 1850, 1978, and 2008 because parameter combinations produced values that exceeded model performance thresholds. Therefore, k was increased to 7 for AER. A simulation with all parameters set to their median values (Tables S3 and S4) was additionally included in each ensemble to ensure coverage of an important region of parameter space. A further set of $2 * N$ simulations with parameter combinations that augment the original design were created to validate the emulators (Bastos & O'Hagan, 2009). In total $(k + 2) * N + 1$ full-year simulations (ensemble members) were created for each anthropogenic emission period. This equates to 235 simulations for AER and 217 for AER-ATM. Preliminary parameter combination screening tests revealed that combined high values of two physical atmosphere parameters (Ent_Fac_Dp > 1.8 and Amdet_Fac > 8.0) cause AER-ATM simulations to fail. This part of the 27-dimensional parameter space (a corner of a 2-D plane) was removed from the AER-ATM design. Some AER-ATM ensemble members (usually with parameter values close to the omitted part of parameter space) also failed to complete a full simulation year for all of the 1850, 1978, and 2008 anthropogenic emission periods. Therefore, the AER-ATM ensemble is made up of the remaining 191 simulations. Additional simulations failed in the AER-ATM for 1998, which were performed after simulations for 1850, 1978, and 2008 periods had been completed and some analyses had been made for them. As a result, there are only 174 simulations for the 1998 period.

2.7. Simulations

For each ensemble, the effect of parameter perturbations was allowed to spin up before output was collected. In AER, all simulations for 2008 and 1850 were started from a spun-up base job previously run for multiple years and branched at December of the previous year. Parameter perturbations were applied to all AER ensemble members including the median simulations and then these were run for a further 13 months. In AER-ATM the median simulation was run for 4 months before parameter perturbations were applied and then run for a further 15 months. The final 12 months of outputs for each emission period were used in each PPE to create the PPE data and analyzed in section 3.

2.8. Model Output Variables

Monthly, daily, and 3-hourly mean outputs were produced for AER and vertical profiles were produced using hourly means. For AER-ATM only monthly mean values were output (see Table S5). The data storage saving allowed the AER-ATM ensemble to be run for four anthropogenic emission years (1850, 1978, 1998, and 2008), whereas output from AER is available for only two anthropogenic emission years (1850 and 2008). The additional temporal resolution in AER is crucial for comparison to a large suite of observations.

Monthly mean outputs include around 100 global two-dimensional fields (e.g., AOD, top-of-atmosphere radiative fluxes, aerosol surface emission, and deposition fluxes) and around 350 global three-dimensional fields (e.g., aerosol number and mass mixing ratios by mode and component, aerosol optical properties,

and aerosol microphysical processes such as nucleation and condensation fluxes). Daily mean outputs include around 10 two-dimensional and around 60 three-dimensional fields with global coverage (including AODs and mixing ratios). The same two-dimensional fields and surface values of three-dimensional fields were also output at the 3-hourly temporal resolution. Hourly vertical profiles of around 40 fields (such as aerosol mixing ratios and optical properties) were output at 24 locations coinciding with long-term aerosol observation sites. Daily, 3-hourly and hourly outputs are intended to be used to diagnose aerosol indirect effects that can readily be compared to observations. Because all AER simulations share the same meteorology, meteorological fields were only output in the median simulation. However, in AER-ATM temperatures evolve freely throughout the atmosphere, as do horizontal winds in the boundary layer. Therefore, multiple meteorological fields of interest were additionally output for all AER-ATM ensemble members.

2.9. Derived Fields

Some variables of scientific interest (such as aerosol mass concentrations and aerosol column loadings) need to be derived from model output. Major derived fields with their time resolutions and spatial domains are listed in Table S6. Model outputs such as aerosol number mixing ratio (ratio of particle number to number of air molecules) for all seven aerosol modes as well as mass mixing ratios of five aerosol components in each mode are used in the derivation of other important fields.

N3 and N50 are number concentrations of particles whose dry diameters are larger than 3 and 50 nm, respectively. These values were derived using the number mixing ratios within each lognormally distributed aerosol mode. Since the widths (geometric standard deviations) of the Aitken and accumulation modes are treated as uncertain parameters in AER (as AIT_WIDTH and ACC_WIDTH), mode widths used to derive N3 and N50 from each ensemble member needed to be varied accordingly.

Similarly, PM2.5 has been derived by calculating the mass fractions of dry particle size larger than 2.5 μm of all aerosol modes. Different widths of Aitken and accumulation modes were used to match the values used in the simulations.

CCN is number concentration of CCN at given altitude and supersaturation. In this study we use the cloud base height (assumed to be around 650-m elevation) and 0.2% supersaturation. We derive CCN using the concentration of soluble particles and mass mixing ratios of internally mixed aerosol components. We used the Köhler theory with hygroscopicity factors (κ) specific to each aerosol component. Since the hygroscopicity factor for OC (KAPPA_OC) was treated as an uncertain parameter in both ensembles, member-specific values of this parameter were used in the CCN derivation. The perturbed values of AIT_WIDTH and ACC_WIDTH were also used in the CCN derivation for AER.

The derivation of aerosol ERF and its components as well as the equivalent values without rapid atmospheric adjustments are described in sections 2.4.1 and 2.4.2.

The raw simulation outputs and derived fields from both ensembles comprise about 160 TB of data. See section 5 for details.

2.10. Emulation and Sensitivity Analysis

We use Gaussian process emulation (Lee et al., 2011; O'Hagan, 2006) to statistically represent model output across the N-dimensional parameter space bounded by the upper and lower limits of the uncertain model parameters. Here emulators were created for multiple fields in order to compare uncertainty in the two ensembles. Emulators were created for monthly and annual mean aerosol mass concentrations, AOD at 550 nm, CCN, as well as for 1850 to 2008 aerosol ERF and its components in AER-ATM and their equivalents without rapid atmospheric adjustments in AER. In each case the emulators were created using the $k * N$ ensemble members then validated using the $2 * N$ additional model runs designed for this purpose (see section 2.6) ensuring that the emulator uncertainty around its mean was low compared with the parametric uncertainty (see Figure S5). All ensemble members were combined postvalidation to produce new emulators with better space-filling properties, under the assumption that emulators created using denser parameter spaces provide an equally good or better representation of the model output.

To compare the model sensitivities within these ensembles we use the Extended-FAST sampling method (Saltelli et al., 1999) to generate very large samples ($N * 10,000$) of model variants for every emulated variable. These large samples allow us to perform sensitivity analyses (Lee et al., 2011; Saltelli et al., 2000) and

decompose the variance into its components. We create emulators of every model output in individual model grid boxes. To make this approach tractable, the horizontal resolution has been reduced to N48 (2.5° in latitudes and 3.75° in longitudes). The use of reduced spatial resolution will not seriously affect the results of analyses of global or regional mean RF. It may, however, lessen the quality of data comparisons and parameter constraints using point measurements. In this case it is an option to extract and use simulation data at given location from the original resolution data.

3. Results

3.1. Magnitude and Uncertainty in Global, Annual Mean Aerosol Radiative Forcing

Figure 1 and Table 2 summarize the magnitude and uncertainty in global, annual mean 1850–2008 aerosol RF (i.e., direct effect plus cloud albedo effect; see section 2.4.1) in the AER PPE and the aerosol ERF (RF plus rapid adjustment) in the AER-ATM compared to equivalent ranges in Forster et al. (2007), Shindell et al. (2013), Myhre et al. (2013), and Boucher et al. (2013). We also show ERF_{ARI} and ERF_{ACI} as well as clear-sky components of RF and ERF_{ARI}. As summarized in Table 2, means and 90% credible intervals are −2.12 (−2.76, −1.47) W/m² for the all-sky aerosol RF and −0.81 (−1.21, −0.46) W/m² for clear-sky forcing . Means and credible intervals are −1.45 (−2.07, −0.81) for the ERF, −0.03 (−0.16, 0.11) for the ERF_{ARI}, and −1.42 (−2.08, −0.74) W/m² for the ERF_{ACI}. The global, annual mean 1850–2008 aerosol RF in AER is stronger than the mean aerosol ERF in AER-ATM. This will be discussed in section 3.2.2. The credible (from 5% to 95%) ranges are 1.29 W/m² for aerosol RF in AER and 1.26 W/m² for aerosol ERF in the AER-ATM. The comparable magnitude of aerosol RF and ERF uncertainty in the two samples suggests that in the absence of rapid atmospheric adjustments the aerosol parameters included only in AER (Table S3) cause as much additional uncertainty as the physical atmosphere parameters included in AER-ATM (Table S4).

The clear-sky component of RF (i.e., RF_{Clr}) in AER is −0.81 (−1.21, −0.46) W/m², whereas that of ERF_{ARI} (i.e., $ERF_{ARI-Clr}$) in AER-ATM is −0.45 (−0.65, −0.27) W/m². The causes of the weaker clear-sky forcing component in AER-ATM include lower anthropogenic SO_2 emissions (section 2.2), higher dust and sea salt loadings (Tables S7–S9), optical properties of carbonaceous aerosols (perturbed in AER-ATM only), and rapid atmospheric adjustments (in response to the radiative effects of these absorbing aerosols). The contribution of clear sky to the all-sky forcing is clear-sky forcing multiplied by clear-sky fraction (1–cloud fraction). Therefore, in both PPEs, the clear-sky forcing components make smaller contributions to net aerosol forcing (RF and ERF) than the cloudy-sky components. The negative clear-sky ERF_{ARI} is nearly canceled by accounting for the positive above-cloud absorption (Ghan, 2013) to make global average net ERF_{ARI} only −0.03 W/m². The uncertainty in clear-sky forcing is also much larger in AER (90% credible range of 0.75 W/m²) than the $ERF_{ARI-Clr}$ uncertainty in AER-ATM (credible range is 0.38 W/m² for clear sky only). However, the small magnitude and uncertainty in global, annual mean clear-sky forcing is the result of regions of positive and negative ERF_{ARI} canceling each other (Regayre et al., 2018). At the regional level ERF_{ARI} is likely to be climatically important, so the causes of important regional ERF_{ARI} uncertainties need to be examined and understood.

The global, annual mean 1850–2008 aerosol forcing is stronger (more negative) in our models than the mean forcing assessments from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP; Shindell et al., 2013), the Aerosol Comparisons between Observations and Models phase II (AeroCom II; Myhre et al., 2013) and the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR5; Boucher et al., 2013; see Figure 1 and Table 2). Where direct comparisons can be made, our mean clear-sky aerosol RF is −0.81 versus −0.67 W/m² in AeroCom II, and our mean aerosol ERF is −1.45 versus −1.17 W/m² for ACCMIP and −0.90 W/m² for AR5. Our ERF_{ARI} of −0.03 W/m² is much weaker than the −0.45 W/m² in AR5, while our ERF_{ACI} of −1.42 W/m² is much stronger than the −0.45 W/m² in AR5. The weakest clear-sky RF value in the AeroCom II model intercomparison (−0.35 W/m²; around 50% weaker than most model values; Myhre et al., 2013) was produced by the modeling group (CAM5.1) who calculated ERF in the same method as we did here (Ghan et al., 2012). Their all-sky RF is nearly 0, which is in agreement with our result. In this method absorption by above-cloud aerosols is included in ERF_{ARI}.
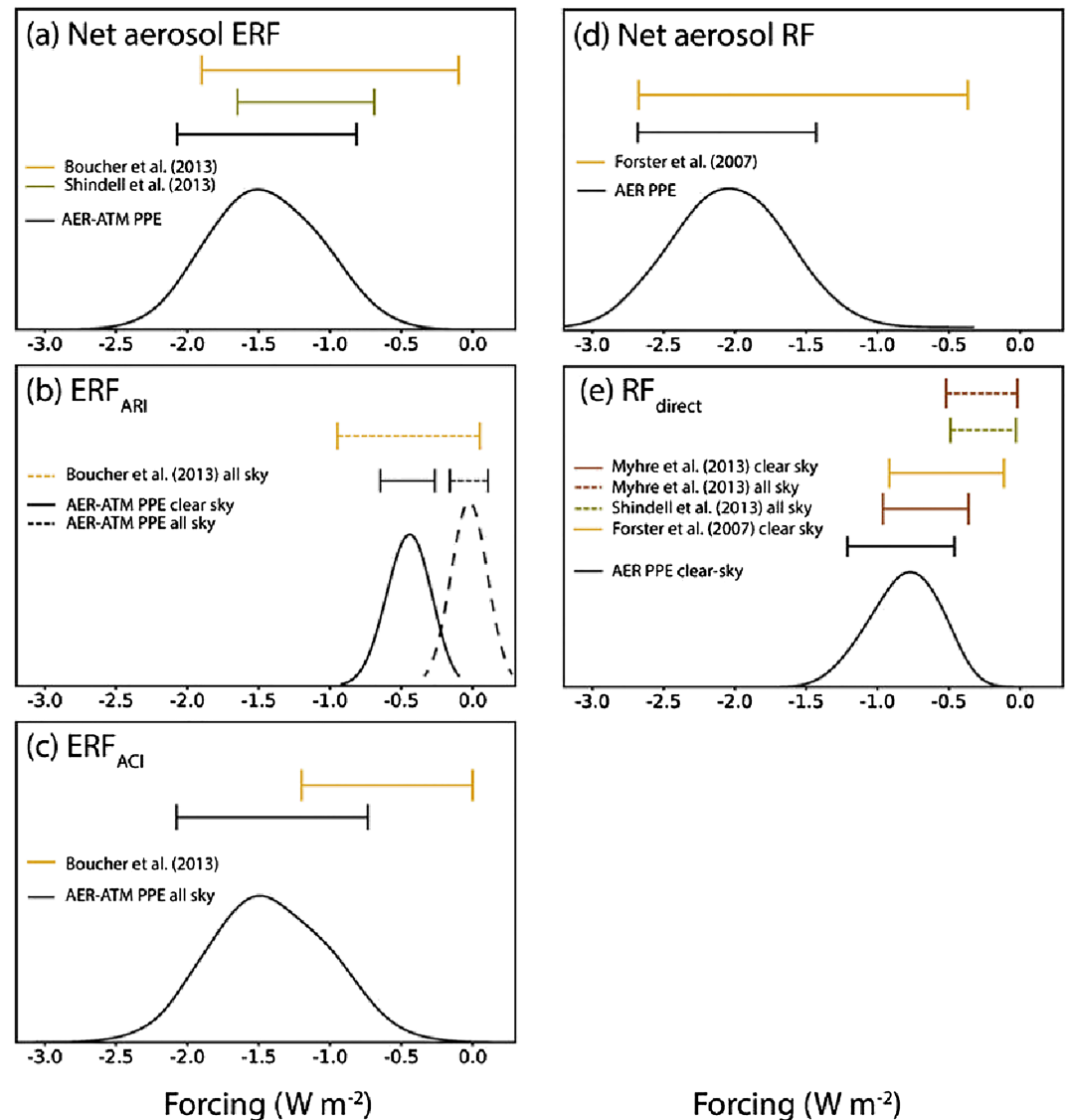
Figure 1. Relative frequencies of global, annual mean aerosol radiative forcing in AER and AER-ATM. Horizontal bars show 90% ranges of these compared with 90% ranges in Shindell et al. (2013), Myhre et al. (2013), and Boucher et al. (2013) shown in Table 2 where applicable. (a) Net ERF, (b) ERF due to aerosol-radiation interaction and clear-sky component of it, (c) ERF due to aerosol-cloud interaction, (d) net aerosol RF, and (e) aerosol direct radiative forcing and clear-sky RF.

Table 2 shows that the aerosol forcing uncertainty in our samples is comparable to the uncertainty from model intercomparisons. These results suggest that our approach has sufficiently sampled the uncertainty in a single model to characterize the model's range of behaviors.

3.2. Spatial Distributions
3.2.1. Aerosol Spatial Distributions
Monthly means and 90% credible intervals of CCN number concentrations calculated using AER ensemble member output at locations coinciding with observation stations (Schmale et al., 2018) are compared to observations in Figure 2. Stations are located across Europe, Asia, and Alaska and represent coastal, rural, alpine, and urban environments. These results show that the ensemble output overlaps the observed values and ranges at most of the stations. However, the observed and ensemble mean values differ markedly in some locations. For example, from winter to spring the ensemble mean CCN is much smaller than the

Table 2
Global, Annual Mean RF and ERF and Their 90% Credible Intervals (W/m^2)

	This study 1850–2008	IPCC AR4; Forster et al. (2007) 1750–2005	ACCMIP; Shindell et al. (2013) 1850–2000s	AeroCom II; Myhre et al. (2013) 1850–2010	IPCC AR5; Boucher et al. (2013) 1750–2010
Aerosol RF	–2.12 (–2.76, –1.47)	–1.2 (–2.7, –0.4)			
Aerosol DRF		–0.5 (–0.9, –0.1)	–0.26 (–0.49, –0.03)	–0.27 (–0.52, –0.02)	
Clear-sky RF	–0.81 (–1.21, –0.46)			–0.67 (–0.97, –0.37)	
Aerosol ERF	–1.45 (–2.07, –0.81)		–1.17 (–1.65, –0.69)		–0.90 (–1.90, –0.10)
ERF$_{ARI}$	–0.03 (–0.16, 0.11)				–0.45 (–0.95, +0.05)
Clear-sky ERF$_{ARI}$	–0.45 (–0.65, –0.27)				
ERF$_{ACI}$	–1.42 (–2.08, –0.74)				–0.45 (–1.20, 0.00)

Note. For this study, emulator mean and the 5th and 95th percentile values are shown. For Forster et al. (2007), best estimate and 90% interval values are as reported for 1750–2005. Values for aerosol RF were read from Figure (2.20). For Shindell et al. (2013) and Myhre et al. (2013), the 5th and 95th percentile values were calculated as multimodel mean plus and minus standard deviation times 1.645, respectively. For Boucher et al. (2013) best estimates and 90% interval values are as reported for 1750–2010.

observed values at many sites (e.g., Cabauw, Mace Head, and Melpitz). However, in Finokalia the ensemble mean CCN is much larger than the observed value especially in early summer. The lack of model-observation agreement in these locations can be used to identify model structural deficiencies. For example, it may provide strong evidence that our factor 2 uncertainty in primary aerosol from FF sources is conservative in parts of Europe, although there are likely to be other compensating factors.

Figures 3 and 4 show the means and standard deviations of annual mean CCN at 650-m altitude, AOD at 550 nm, and aerosol RF for the 1850–2008 period. The means and standards deviations were taken from the large (number of parameters times 10,000) samples of output from individual statistical model emulators (as described in section 2.10).

Annual emulator-mean AOD is typically in the range 0.1–0.3 over the North Atlantic, 0.1–0.2 over the Arctic, 0.15–0.3 over the Southern Ocean, 0.2–0.3 over Europe, 0.15–0.4 over the North Pacific, and up to 1.0 over China and India. These emulator-mean AODs compare well with satellite observations (Shindell et al., 2013) and satellite-assimilated simulations (Boucher et al., 2013) especially when we account for the emulator standard deviation, which is about 0.1–0.15 over oceans and as high as 0.5 over China. A comparison with observations will enable implausible parts of the sampled parameter space to be identified, thereby helping to reduce the standard deviation of our AOD estimates and in forcing estimates (Johnson et al., 2018).

The spatial distributions of CCN at cloud base height with the 2008 emissions are similar in the two ensembles. However, peak CCN concentrations over the largest anthropogenic emission sources (China and India) are higher in AER than in AER-ATM. Similarly, the mean AOD is higher near anthropogenic emission sources in AER. These are mainly due to the low anthropogenic SO_2 emission in AER-ATM (section 2.2) with possible contributions of other factors such as primary sulfate aerosol emission sizes and numbers perturbed only in AER PPE. However, aerosols are also advected further from anthropogenic sources over marine environments in AER-ATM. This is partly due to the longer atmospheric lifetime of sulfate aerosol in AER-ATM that arises from less effective condensation and coagulation (see Table S7–S9). Also, surface wind speeds and precipitation rates are noticeably different in the two PPEs over many marine areas. The wind speed differences largely arise from whether or not physical parameters are perturbed and whether or not rapid adjustment of boundary layer is allowed. Larger surface wind speeds in AER-ATM produce larger emission fluxes of sea spray aerosols including regions near anthropogenic emission sources (Figure 5). Use of different years of ERA-Interim wind speed for nudging causes much smaller differences in emissions of sea salt and dust than different model setups in AER and AER-ATM (see Tables S7–S9). Higher aerosol concentrations reduce precipitation rates in AER-ATM, which includes aerosol feedbacks on clouds, radiation, and dynamics. Perturbations to the optical properties of carbonaceous aerosols additionally have the potential to produce nonlinear aerosol and cloud lifetime effects, and physical atmosphere parameters can additionally affect surface wind speeds. In summary, the means and standard deviations of CCN and AOD are higher over marine regions in AER-ATM because of higher surface wind speeds and aerosol
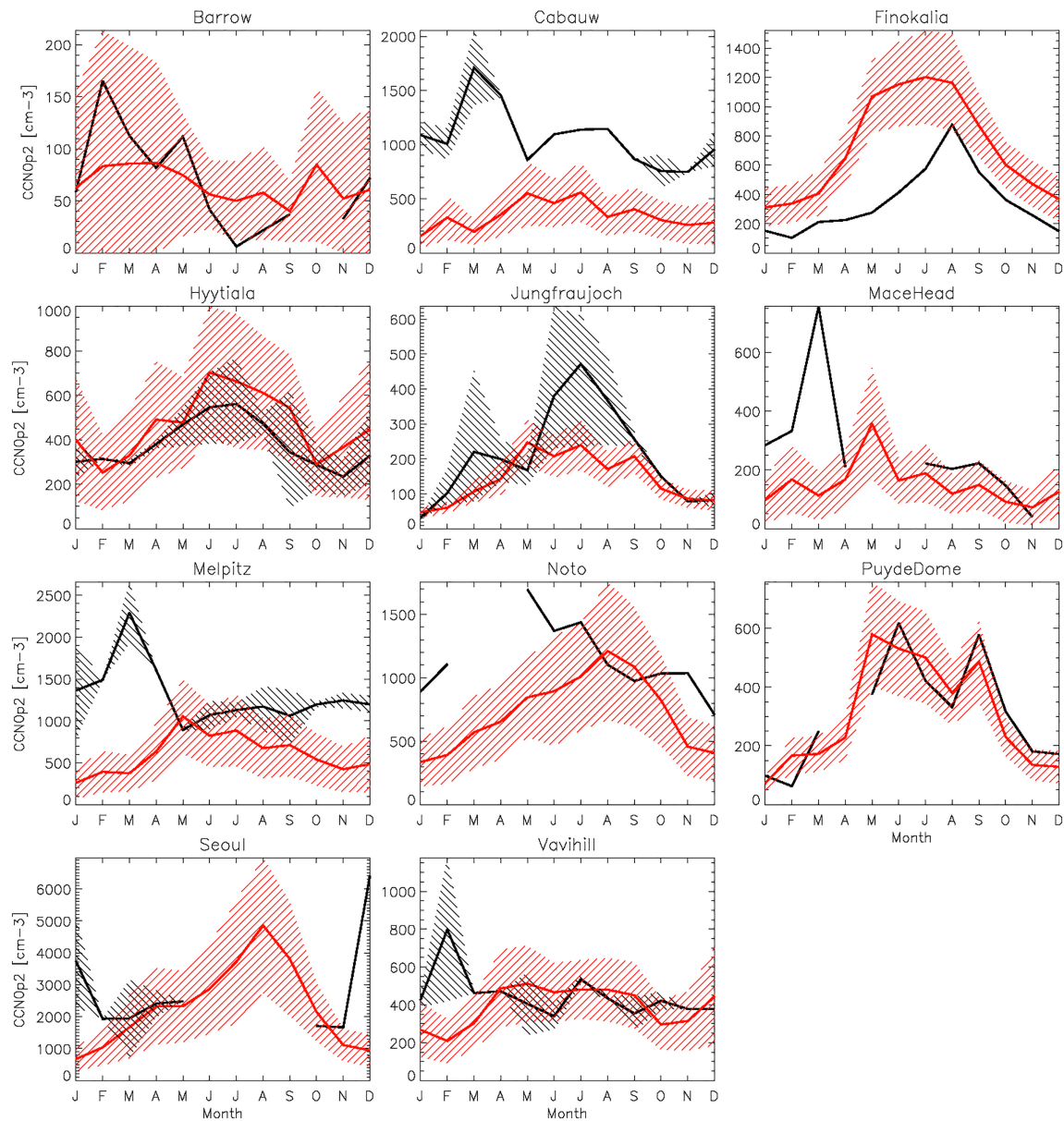
Figure 2. Ensemble monthly means of CCN at supersaturation of 0.2% at observational station locations calculated for AER are shown in red lines with 90% credible intervals in shadings. Monthly means of measurements at stations are shown in black lines. Where observations are available in more than 1 year, interannual means and ranges of smallest and largest monthly means are shown in black lines and black shading.

feedbacks, while continental CCN and AOD are higher in AER due to higher SO_2 emission and nonlinear responses to perturbations in sulfate aerosol emission size and number. Concentrations of natural aerosols are higher in AER-ATM in both PD and PI atmospheres, whereas anthropogenic aerosol concentrations are lower in AER-ATM in the PD. Both of these lead to weaker aerosol ERF in AER-ATM than aerosol RF in AER.

3.2.2. Forcing Spatial Distributions

Annual mean total forcing (RF and ERF) is more negative than -2 W/m^2 over large areas of the North Atlantic and Europe and -5 W/m^2 over large areas of East Asia and North Pacific in both ensembles. The regional peaks of aerosol RF and ERF are much stronger than the multimodel mean in Shindell et al. (2013, Figure 11), consistent with the comparisons of global means shown in Table 2. For example, negative forcing just exceeds -1 W/m^2 over east and south Asia, Europe, and eastern North America in the
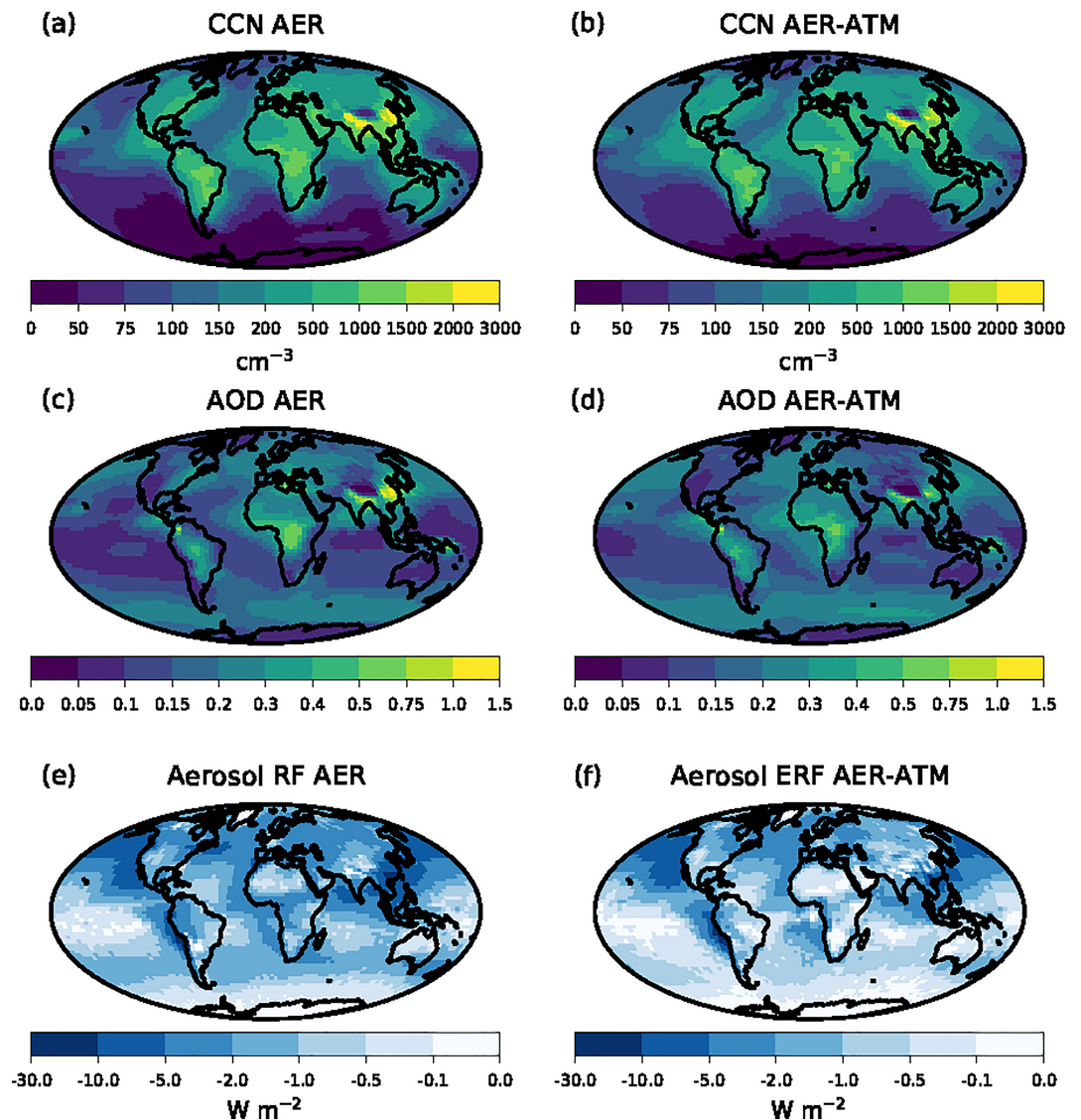
Figure 3. Annual emulator mean CCN (a, b), AOD (c, d), aerosol RF (e), and ERF (f) in AER (a, c, and e) and AER-ATM (b, d, and f). Samples of $N * 10,000$ model variants taken from emulators of individual model variables are used to calculate the means.

multimodel mean, whereas negative RF and ERF here are up to −10 and −5 W/m^2 in the same regions. Over the North Atlantic and the North Pacific, our PPEs estimates stronger forcing than over these continental regions, while the multimodel mean RF is weaker over these ocean basins. Off the west coast of South America is one of the regions with strongest negative forcing in both of our PPEs, but the forcing in this region is not particularly strong in Shindell et al. (2013). The strength of the aerosol forcing in our PPEs compared to the multimodel mean suggests that our base model may be more sensitive to changes in aerosols than other global climate models.

The cleanest comparison between ensemble members can be made using the simulations where all parameters are set to their median values. However, there are multiple differences between these two "median" simulations: the degree of atmospheric nudging, the meteorological years used in the nudging process, the inclusion of rapid atmospheric adjustments, and the inclusion of nonzero imaginary indices for carbonaceous aerosols, on top of the difference in high-level SO$_2$ emissions. These two median model simulations show similar spatial patterns to the emulator mean distributions (not shown). Both CCN and AOD are consistently lower over most of the continents and higher over most of the oceans and the Antarctica in AER-
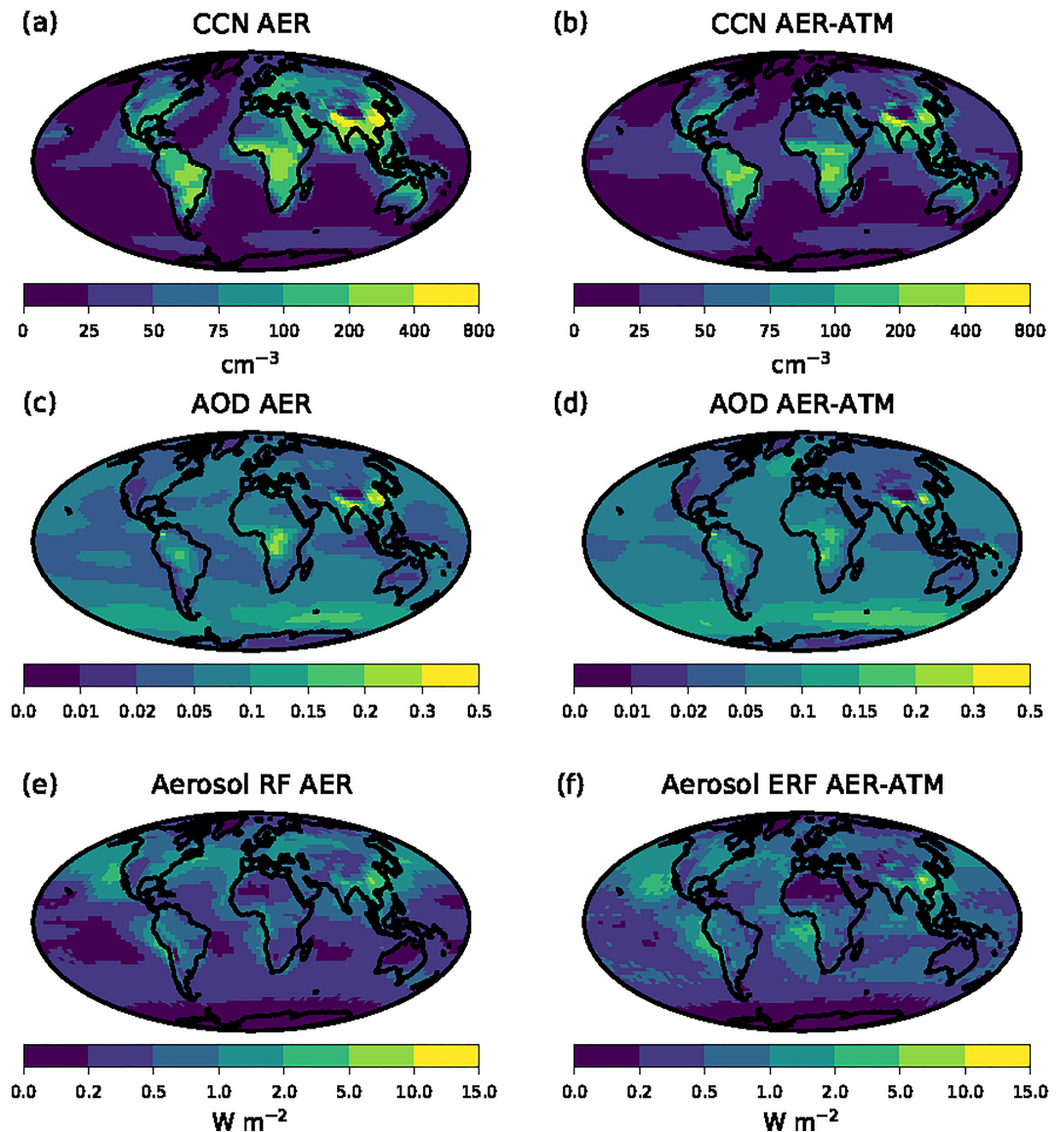
Figure 4. Emulator standard deviations corresponding to the quantities presented in Figure 3.

ATM compared to AER. Lower SO_2 emission in AER-ATM tends to lead to longer atmospheric lifetime of anthropogenic aerosols through less efficient condensation of sulfate gas onto existing particles leading to slower particle growth and lower particle hygroscopicity. The meteorological difference caused by rapid atmospheric adjustments with contribution of the radiative properties of carbonaceous aerosols (which affect aerosol and cloud lifetimes; see sections 2.4 and 2.5 and Table 1) are additional factors contributing the different aerosol distributions in two PPEs. Ensemble mean aerosol distributions may also be affected by the skewness of the probability distribution functions of parameter values, nonlinearity in the model sensitivity to the parameter perturbations, and interactions between multiple parameters.

As we saw in section 3.1, the global, annual mean 1850–2008 aerosol RF in AER is stronger (-2.12 W/m^2) than the mean aerosol ERF in AER-ATM (-1.45 W/m^2). Inspection of the forcing maps (Figures 3e and 3f) shows that the AER aerosol RF is stronger than the AER-ATM aerosol ERF over anthropogenic source regions (in line with higher CCN concentrations and AOD) as well as over remote marine regions (such as oceans in the Southern Hemisphere, North Atlantic Ocean, and much of Africa and Siberia). This stronger remote aerosol RF in AER occurs despite the fact that anthropogenic aerosols are advected further from source in AER-ATM.
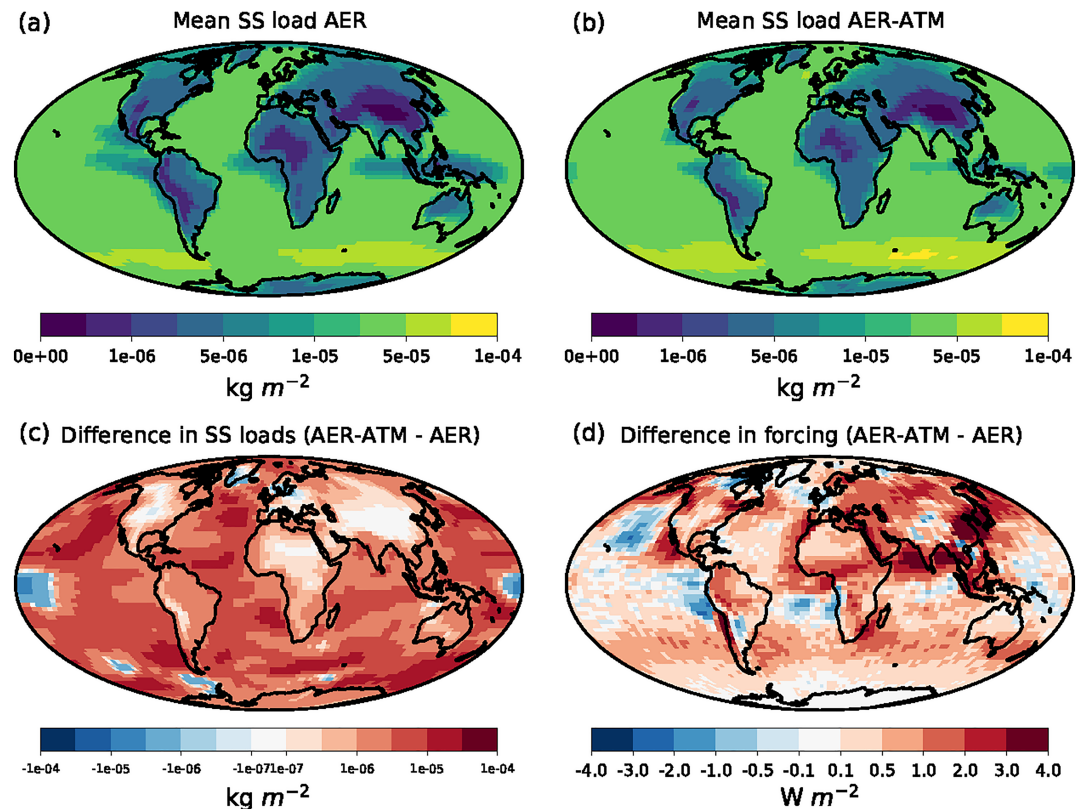
Figure 5. Annual emulator mean sea salt loadings in AER (a) and AER-ATM (b), their difference (c), and difference in radiative forcing between two ensemble (d).

Remote aerosol forcing is weaker in AER-ATM because that ensemble has higher (by 16%; see Tables S7 and S8) natural aerosol (sea salt and dust) concentrations than AER. Figure 5 shows that sea salt loadings are higher in AER-ATM (especially over southern high-latitude regions). The spatial pattern of increased sea salt loadings roughly corresponds to the spatial pattern of weaker aerosol forcing in AER-ATM (Figures 5c and 5d). Surface DMS concentration is also higher in AER-ATM than in AER over most of the oceans in the Northern Hemisphere (not shown). Natural aerosol emissions are some of the most important causes of 1850–2008 aerosol forcing uncertainty in both AER and AER-ATM (see Figure 6 and Table 3). These results are consistent with the fact that high natural aerosol emissions in both 1850 and 2008 atmospheres weaken the aerosol-cloud-interaction component of forcing because of the highly nonlinear relationship between emissions and cloud albedo (Carslaw, Lee, Reddington, Pringle, et al., 2013).

The forcing uncertainty is larger over anthropogenic emission regions in AER (where the mean forcing is also stronger) but larger in other part of the globe especially over remote marine regions in AER-ATM (Figures 4e and 4f). Both aerosol and physical atmosphere parameters are important causes of aerosol ERF uncertainty (Regayre et al., 2018), and the combined effect of these uncertainties (which determine the magnitude of rapid atmospheric adjustments to aerosols) makes aerosol ERF more uncertain in marine regions, despite having weaker (less negative) global and regional mean forcing values.

3.3. Sources of Uncertainty in Aerosol Radiative Forcing

Here we analyze the importance of individual parameters as causes of uncertainty in aerosol RF in the two ensembles by using variance decomposition (Lee et al., 2011). Parameters that cause a similar magnitude of uncertainty in both ensembles can be considered to have effects on the model that are insensitive to the inclusion of rapid atmospheric adjustments and a dynamically responsive boundary layer, which were included in AER-ATM but not AER.
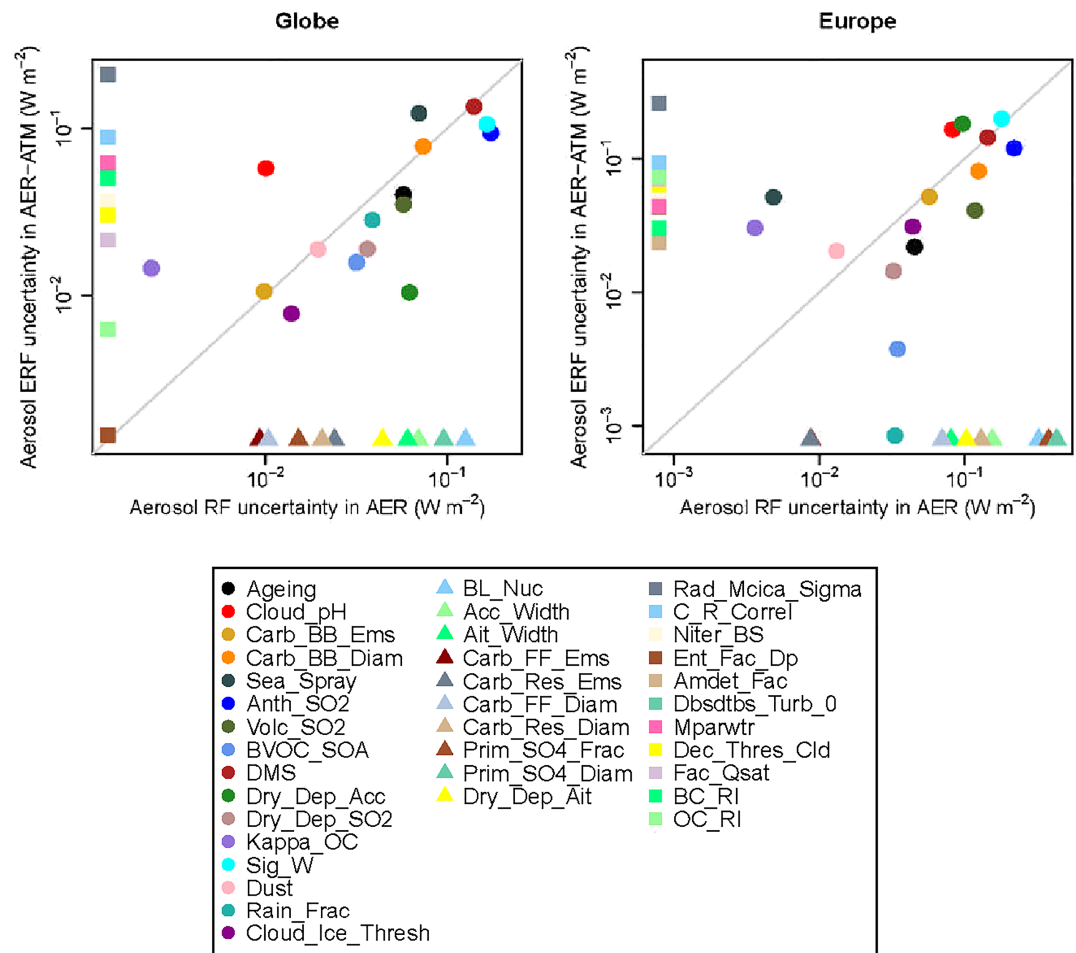
Figure 6. Standard deviations in annual mean global and European aerosol RF and aerosol ERF (W/m²) caused by individual parameters in AER and AER-ATM, respectively. Circles show parameters that are perturbed in both ensembles. Triangles show aerosol parameters that are perturbed only in AER, and squares show physical atmosphere parameters perturbed only in AER-ATM.

The main causes of aerosol ERF uncertainty in AER-ATM were analyzed in Regayre et al. (2018). In that study emissions of natural (Sea_Spray and DMS) and anthropogenic (Anth_SO2) aerosols, cloud process (Sig_W) and radiation (Rad_Mcica_Sigma) parameters were found to cause around 70% of total ERF variance. However, the percentage of uncertainty depends on total uncertainty in each PPE and which other parameters were perturbed. In Figure 6 and Table 3 the absolute uncertainties caused by individual parameters are summarized as standard deviations of annual mean aerosol RF and ERF in the AER and AER-ATM, respectively. The causes of global and European mean variance were calculated by creating emulators for global and regional averages (taken from individual ensemble members) and then sampling from the emulators and performing sensitivity analyses as outlined in section 2.10.

Many parameters included in both ensembles (circles in Figure 6a) cause larger magnitudes of uncertainty in global, annual mean aerosol forcing in AER than AER-ATM. For example, anthropogenic emission parameter (Anth_SO2) and the updraft velocity parameter (Sig_W) are among the most important parameters causing forcing uncertainty in both ensembles but cause more than twice as much uncertainty in forcing in AER than in AER-ATM despite being perturbed within the same uncertainty ranges. In contrast, the sea spray emission parameter (Sea_Spray) causes around half as much uncertainty in forcing in AER as it does in AER-ATM. Sea spray emissions cause more forcing uncertainty in AER-ATM through the higher aerosol loading as well as their effect on rapid adjustments and interaction with physical atmosphere parameters, which affect the properties of low-level clouds in marine environments. Among parameters causing

Table 3

Standard Deviations in Annual Mean Global and European Aerosol RF and Aerosol ERF (W/m^2) Caused by Individual Parameters in AER and AER-ATM, Respectively

Parameter	Uncertainty (W/m^2)			
	Globe		Europe	
	AER	AER-ATM	AER	AER-ATM
Acc_Width	**0.070**		**0.154**	
Aging	0.057	0.040	0.045	0.022
Ait_Width	0.061		0.080	
Anth_SO2	**0.174**	**0.094**	**0.218**	**0.119**
BL_Nuc	**0.126**		**0.321**	
BVOC_SOA	0.032	0.016	0.034	0.004
Carb_BB_Diam	**0.074**	**0.078**	0.124	**0.080**
Carb_BB_Ems	0.010	0.011	0.057	0.052
Carb_FF_Diam	0.010		0.070	
Carb_FF_Ems	0.009		0.009	
Carb_Res_Diam	0.020		**0.129**	
Carb_Res_Ems	0.024		0.009	
Cloud_Ice_Thresh	0.014	0.008	0.044	0.031
Cloud_pH	0.010	0.058	0.082	**0.165**
DMS	**0.140**	**0.135**	**0.144**	**0.144**
Dry_Dep_Acc	0.062	0.010	0.096	**0.182**
Dry_Dep_Ait	0.044		0.102	
Dry_Dep_SO2	0.036	0.019	0.032	0.014
Dust	0.019	0.019	0.013	0.020
Kappa_OC	0.002	0.014	0.004	0.030
Prim_SO4_Diam	**0.095**		**0.428**	
Prim_SO4_Frac	0.015		**0.376**	
Rain_Frac	0.039	0.028	0.033	0.001
Sea_Spray	**0.070**	**0.123**	0.005	0.051
Sig_W	**0.165**	**0.106**	**0.179**	**0.199**
Volc_SO2	0.057	0.035	0.117	0.041
Rad_Mcica_Sigma		**0.211**		**0.260**
C_R_Correl		**0.089**		**0.092**
Niter_BS		0.036		0.046
Ent_Fac_Dp		0.001		0.043
Amdet_Fac		0.006		0.023
Dbsdtbs_Turb_0		0.056		0.068
Mparwtr		**0.062**		0.044
Dec_Thres_Cld		0.030		0.063
Fac_Qsat		0.021		0.070
BC_RI		0.050		0.030
OC_RI		0.006		0.072

Note. The eight largest values within each column are shown in bold and double underlined, and the next eight are single underlined.

smaller forcing uncertainties, cloud droplet acidity (Cloud_pH) and organic carbon hygroscopicity (Kappa_OC) cause many times larger uncertainty, whereas accumulation mode dry deposition (Dry_Dep_Acc) causes many times smaller uncertainty in AER-ATM than in AER. Causes of these differences include the lower high-level anthropogenic SO_2 emissions within AER-ATM and higher natural aerosols in AER-ATM described in section 3.2.2. Furthermore, the emission size and amount of primary sulfate aerosols are perturbed in AER. These effects are likely to be combined to make anthropogenic emissions a more important source of uncertainty in AER. Other parameters such as DMS and the diameter of emitted BB particles (Carb_BB_Diam) cause a similar amount of uncertainty in the two ensembles. This suggests that the aerosol forcing uncertainty to these parameter perturbations is relatively insensitive to rapid atmospheric adjustments, meteorological factors, and other differences in ensemble setup (sections 2.4.1 and 2.4.2).

Aerosol parameters included in AER but not in AER-ATM are relatively unimportant for global, annual mean aerosol forcing. However, at least two important causes of forcing uncertainty (BL_Nuc and Ait_Width) were neglected in AER-ATM. These results suggest the global, annual mean aerosol ERF uncertainty would be larger if all uncertain aerosol parameters were included in AER-ATM. However, interactions between the parameters and dependencies on model setup (e.g., including rapid adjustments) mean that an increase in uncertainty is not assured. The importance of aerosol parameters included in AER but not in AER-ATM is more pronounced for European aerosol RF than in the global mean case. Because aerosols have a relatively short atmospheric lifetime, aerosol parameters have a highly region-dependent influence on aerosol RF uncertainty (Regayre et al., 2015). Over Europe parameters that alter the size and emission flux of primary sulfate particles (Prim_SO4_Diam and Prim_SO4_Frac) are the largest causes of aerosol RF uncertainty. In response to the constant primary sulfate emissions in AER-ATM (the associated parameters are held constant), parameters related to aerosol removal (such as Dry_Dep_Acc) and production (Cloud_pH) processes cause more aerosol ERF uncertainty.

Physical atmosphere parameters were not perturbed in AER because their effects would be suppressed by the model setup (section 2.4.1). However, Figure 6 and Table 3 show that physical atmosphere parameters are among the largest causes of aerosol ERF uncertainty. Indeed, the cloud condensate parameter (Rad_Mcica_Sigma) is the largest cause of aerosol ERF uncertainty in AER-ATM. Regayre et al. (2018) showed that aerosols collectively cause more of the global mean

aerosol ERF uncertainty than physical atmosphere parameters, and in fact Figure 6 and Table 3 show that physical atmosphere parameters are less important for aerosol ERF over Europe. Although we have previously shown that the uncertainty in European aerosol ERF could be reduced by as much as 30% using multiple observations (Johnson et al., 2018), the underlying uncertainty would be larger if additional aerosol parameters were included in AER-ATM.

In summary, neither ensemble comprehensively samples model uncertainty. Because physical atmosphere parameters were perturbed in AER-ATM and rapid atmospheric adjustments were included within that ensemble, it can be used to inform our understanding of which processes have the greatest influence on aerosol ERF. However, because AER contains additional aerosol parameters that are important causes of uncertainty, a more detailed understanding of the effect of aerosol processes on aerosol RF can be gained by analyzing output from this ensemble.

4. Summary and Discussions

We have presented two PPEs of the HadGEM atmosphere-only global climate model coupled to the UKCA chemistry and aerosol model. These ensembles were designed to be complementary and address different aspects of the aerosol RF uncertainty reduction challenge. The AER (aerosol) ensemble was designed to be compared extensively to atmospheric measurements mainly on high time resolutions. Model internal variability was completely suppressed in this ensemble by prescribing wind and temperature fields and neglecting aerosol radiative feedbacks on meteorology, that is, the model was run essentially as a chemical transport model as in previous studies (Carslaw, Lee, Reddington, Mann, et al., 2013; Lee et al., 2011, 2012, 2013; Regayre et al., 2014, 2015). As such, the effects of perturbing 26 aerosol emission, microphysical and deposition parameters on aerosol concentrations, and aerosol RF can be quantified in the absence of additional uncertainties from meteorological adjustments. The AER-ATM (aerosol/atmosphere) ensemble was designed to quantify the relative importance of 27 aerosol and physical atmosphere parameters for annual and monthly mean aerosol ERF (including rapid atmospheric adjustments) at the global and regional level. Therefore, in this ensemble the boundary layer (containing low-level clouds important for aerosol RF) was free to respond to changes caused by parameter perturbations.

Uncertainty in output from both ensembles can potentially be constrained by observations (Johnson et al., 2018; Regayre et al., 2018). Here we performed a preliminary comparison of model-derived CCN to observe values at regionally representative stations (Schmale et al., 2018). We find that the modeled CCN uncertainty (in large samples of model variants based on statistical emulators) spans the observed CCN at most locations. However, at some stations, for some seasons we note a lack of model-observation agreement, which could be used to identify areas for model structural development.

The global mean 1850–2008 aerosol RFs are -2.12 (credible interval -2.76, -1.47) W/m^2 (RF) based on the AER PPE and -1.45 (-2.07, -0.81) W/m^2 (ERF) based on the AER-ATM PPE. There are several factors that contribute to the weaker forcing in AER-ATM. First, more natural aerosols (sea spray, DMS, and dust) are typically emitted in AER-ATM than in AER because of their dependence on surface wind speeds that are larger over most marine environments in AER-ATM, largely due to perturbed physical atmosphere parameters and rapid adjustments in the boundary layer. Larger natural aerosol emissions in both 1850 and 2008 reduce the magnitude of the aerosol-cloud interaction (Carslaw, Lee, Reddington, Pringle, et al., 2013). Second, uncertainty in the radiative properties of carbonaceous aerosols are included in AER-ATM. Rapid atmospheric adjustments to carbonaceous aerosols produce regions of positive and negative forcing that cancel and cause a near-zero global mean aerosol-radiation interaction component of forcing across the ensemble (Figure 1, see also Regayre et al., 2018). Third, anthropogenic emissions from industrial and energy sectors were too low in AER-ATM simulations. This weakened the aerosol forcing over anthropogenic sources in AER-ATM.

The global, annual mean aerosol forcings are stronger in both of our ensembles than in model intercomparison studies (Boucher et al., 2013; Myhre et al., 2013; Shindell et al., 2013). The mean and uncertainty of forcing in our samples will change substantially by identifying all model variants (parts of parameter space) that are plausible compared to aerosol, cloud, and radiation observations; a task begun in Regayre et al. (2018) and Johnson et al. (2018). We are also aware that we do not include some potentially important marine sources of organic aerosols (Gantt et al., 2015) and additional sources of preindustrial fire emissions (Hamilton et al., 2018) which, if included, would increase preindustrial aerosol concentrations and weaken the aerosol forcing (Carslaw, Lee, Reddington, Pringle, et al., 2013). The autoconversion rate (the rate at which cloud droplets convert into rainfall) is also a potentially important parameter that we did not perturb in our ensembles. The relationship between the autoconversion rate and the droplet number (which is controlled by the updraft-driven CCN activation), for example, the exponent in the Khairoutdinov and Kogan (2000) parameterization, would be another interesting parameter to perturb in future studies.

The credible aerosol forcing ranges in our ensembles are larger than or comparable to those obtained from model intercomparison projects ACCMIP (Shindell et al., 2013), AeroCom II (Myhre et al., 2013), and IPCC AR5 (Boucher et al., 2013; Figure 1 and Table 2), even though inclusion of observationally implausible model variants somewhat contributes to additional uncertainty. Multimodel intercomparison studies are useful to identity structural issues in the representation of aerosol cloud interactions in global models that PPEs cannot address. However, the large uncertainty within a single model suggests that PPEs can be

used to identify the key processes causing model uncertainty and differences among multimodel ensemble members. Additionally, the large parametric uncertainty sampled here suggests a need for caution when interpreting conclusions drawn from multimodel intercomparison studies that represent unknown combination of structural and parametric uncertainties from small size of ensembles (Carslaw et al., 2018). Both structural and parametric uncertainties need to be understood and interpreted in order to properly develop a strategy for reducing the uncertainty and improve confidence in climate projections.

Regionally, uncertainty in aerosol forcing is larger over most marine regions in AER-ATM (Figure 4) with exceptions of regions near anthropogenic sources. However, global, annual mean aerosol forcing uncertainties are roughly the same in two PPEs (Figure 1 and Table 2), as a result of the cancelation between regional effects of opposite signs in AER-ATM. These results suggest that uncertainty in the spatial redistribution of aerosols between regions is important for understanding the climate effects of aerosols at the regional level. We also neglected several aerosol parameters (related to the size and number of primary sulfate emissions) in AER-ATM that are important causes of global and regional uncertainties of aerosol RF in AER (Figure 6 and Table 3). The regional aerosol ERF uncertainty in AER-ATM would be even larger if these additional aerosol parameters were included.

The AER and AER-ATM ensembles were designed to help understand and reduce the persistent uncertainty in aerosol RF. These ensembles are the first step toward this goal because they densely sample the uncertainty in our model. Our aim now is to challenge the model uncertainty using multiple observations related to key processes driving model uncertainty (e.g., Johnson et al., 2018; Regayre et al., 2018). Comparison of the ensemble output to observations will also help to identify model structural deficiencies (e.g., McNeall et al., 2016) and prioritize areas of model development. Because we have sampled the full parametric uncertainty range, to our best knowledge, we can be more confident in the attribution of biases to model structural error than if we had only a handful of sensitivity tests to rely on.

Once observationally constrained, these ensembles will provide multiple opportunities for analysis in other aspects of climate research. For example, the constrained parameter space and sensitivity analysis results will inform field campaigns by highlighting the most useful new measurements for reducing the remaining forcing uncertainty. Additionally, observationally constrained aerosol ERF over, say, the North Pacific in recent decades, would provide values to feed into simulations designed to examine the full dynamical response to those regional forcings. Finally, understanding the causes of uncertainty in AER and AER-ATM will inform the development of climate models in general because other climate models will have parametrizations that correspond to the processes that cause uncertainty in our ensembles.

Data Availability Statement

Raw simulation output data in both AER and AER-ATM ensembles are available as Met Office postprocessing data format (.pp; Met Office, 2013) from the JASMIN data infrastructure (http://www.jasmin.ac.uk). Some of the climate-relevant fields are derived and stored in netCDF files (.nc) containing data for all ensemble members and made available as a community research tool. Variables contained in raw outputs and derived data as well as their spatial domains and temporal resolutions are summarized in Tables S5 and S6. The total data sizes are around 85 TB for AER and 75 TB for the AER-ATM.

References

Abdul-Razzak, H., & Ghan, S. J. (2000). A parameterization of aerosol activation 2. multiple aerosol types. *Journal of Geophysical Research*, *105*(D5), 6837–6844. https://doi.org/10.1029/1999JD901161

Abraham, N. L. (2014). UM-UKCA release job 4.0, available from http://www.ukca.ac.uk/wiki/index.php/Release_Job_RJ4.0 ().

Abraham, N. L., Archibald, A. T., Bellouin, N., Boucher, O., Braesicke, P., Bushell, A., et al. (2012). Unified model documentation paper no. 84: United Kingdom Chemistry and Aerosol (UKCA) Technical Description MetUM Version 8.4. UK Met Office, Exeter (UK), pp. 74.

Andreae, M. O., Jones, C. D., & Cox, P. M. (2005). Strong present-day aerosol cooling implies a hot future. *Nature*, *435*(7046), 1187–1190. https://doi.org/10.1038/nature03671

Andres, R. J., & Kasgnoc, A. D. (1998). A time-averaged inventory of subaerial volcanic sulfur emissions. *Journal of Geophysical Research*, *103*(D19), 25,251–25,261. https://doi.org/10.1029/98JD02091

Andrianakis, I., Vernon, I., McCreesh, N., McKinley, T. J., Oakley, J. E., Nsubuga, R. N., et al. (2017). History matching of a complex epidemiological model of human immunodeficiency virus transmission by using variance emulation. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, *66*(4), 717–740. https://doi.org/10.1111/rssc.12198

Asmi, E., Kivekäs, N., Kerminen, V.-M., Komppula, M., Hyvärinen, A.-P., Hatakka, J., et al. (2011). Secondary new particle formation in Northern Finland Pallas site between the years 2000 and 2010. *Atmospheric Chemistry and Physics*, *11*(24), 12,959–12,972. https://doi.org/10.5194/acp-11-12959-2011

Asmi, E., Kondratyev, V., Brus, D., Laurila, T., Lihavainen, H., Backman, J., et al. (2016). Aerosol size distribution seasonal characteristics measured in Tiksi, Russian Arctic. *Atmospheric Chemistry and Physics*, *16*(3), 1271–1287. https://doi.org/10.5194/acp-16-1271-2016

Balkanski, Y., Schulz, M., Claquin, T., & Guibert, S. (2007). Reevaluation of mineral aerosol radiative forcings suggests a better agreement with satellite and AERONET data. *Atmospheric Chemistry and Physics*, *7*(1), 81–95. https://doi.org/10.5194/acp-7-81-2007

Bastos, L. S., & O'Hagan, A. (2009). Diagnostics for Gaussian process emulators. *Technometrics*, *51*(4), 425–438. https://doi.org/10.1198/TECH.2009.08019

Bellouin, N., Baker, L., Hodnebrog, Ø., Olivié, D., Cherian, R., Macintosh, C., et al. (2016). Regional and seasonal radiative forcing by perturbations to aerosol and ozone precursor emissions. *Atmospheric Chemistry and Physics*, *16*(21), 13,885–13,910. https://doi.org/10.5194/acp-16-13885-2016

Bellouin, N., Boucher, O., Haywood, J., Johnson, C., Jones, A., Rae, J., & Woodward, S. (2007). Improved representation of aerosols for HadGEM2, Hadley Centre Technical Note 73, 43 pp., Met Off. Hadley Cent., Exeter, U. K.

Bellouin, N., Mann, G. W., Woodhouse, M. T., Johnson, C., Carslaw, K. S., & Dalvi, M. (2013). Impact of the modal aerosol scheme GLOMAP-mode on aerosol forcing in the Hadley Centre Global Environmental Model. *Atmospheric Chemistry and Physics*, *13*(6), 3027–3044. https://doi.org/10.5194/acp-13-3027-2013

Birmili, W., Wiedensohler, A., Heintzenberg, J., & Lehmann, K. (2001). Atmospheric particle number size distribution in central Europe: Statistical relations to air masses and meteorology. *Journal of Geophysical Research*, *106*(D23), 32,005–32,018.

Bollasina, M., Ming, Y., & Ramaswamy, V. (2011). Anthropogenic aerosols and the weakening of the South Asian summer monsoon. *Science*, *28*, 502–505. https://doi.org/10.1126/science.1204994

Bollasina, M. A., Ming, Y., & Ramaswamy, V. (2013). Earlier onset of the Indian monsoon in the late twentieth century: The role of anthropogenic aerosols. *Geophysical Research Letters*, *40*, 3715–3720. https://doi.org/10.1002/grl.50719

Bond, T. C., & Bergstrom, R. W. (2006). Light absorption by carbonaceous particles: An investigative review. *Aerosol Science and Technology*, *40*(1), 27–67. https://doi.org/10.1080/02786820500421521

Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., & Bellouin, N. (2012). Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, *484*(7393), 228–232. https://doi.org/10.1038/nature10946

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., et al. (2013). Clouds and Aerosols. In Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M. Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P.M. (Eds.) *ClimateChange 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Vol. *7*). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Boutle, I. A., Abel, S. J., Hill, P. G., & Morcrette, C. J. (2014). Spatial variability of liquid cloud and rain: Observations and microphysical effects. *Quarterly Journal of the Royal Meteorological Society*, *140*(679), 583–594. https://doi.org/10.1002/qj.2140

Browse, J., Carslaw, K. S., Arnold, S. R., Pringle, K. J., & Boucher, O. (2012). The scavenging processes controlling the seasonal cycle in Arctic sulphate and black carbon aerosol. *Atmospheric Chemistry and Physics*, *12*(15), 6775–6798. https://doi.org/10.5194/acp-12-6775-2012

Carslaw, K. S., Lee, L. A., Reddington, C. L., Mann, G. W., & Pringle, K. J. (2013). The magnitude of uncertainty in global aerosol. *Faraday Discussions*, *165*, 495–512. https://doi.org/10.1039/C3FD00043E

Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., et al. (2013). Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature*, *503*(7474), 67–71. https://doi.org/10.1038/nature12674

Carslaw, K. S., Lee, L. A., Regayre, L. A., & Johnson, J. S. (2018). Climate models are uncertain, but we can do something about it. *Eos*, *99*. https://doi.org/10.1029/2018EO093757

Collins, M., Booth, B. B. B., Bhaskaran, B., Harris, G. R., Murphy, J. M., Sexton, D. M. H., & Webb, M. J. (2011). Climate model errors, feedbacks and forcings: A comparison of perturbed physics and multi-model ensembles. *Climate Dynamics*, *36*(9-10), 1737–1766. https://doi.org/10.1007/s00382-010-0808-0

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., et al. (2013). In Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P.M. (Eds.) *12. Long-term Climate Change: Projections, Commitmentsand Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working GroupI to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Dhomse, S. S., Emmerson, K. M., Mann, G. W., Bellouin, N., Carslaw, K. S., Chipperfield, M. P., et al. (2014). Aerosol microphysics simulations of the Mt. Pinatubo eruption with the UM-UKCA composition-climate model. *Atmospheric Chemistry and Physics*, *14*(20), 11221–11246. https://doi.org/10.5194/acp-14-11221-2014

Edwards, J. M., & Slingo, A. (1996). Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model. *Quarterly Journal of the Royal Meteorological Society*, *122*(531), 689–719. https://doi.org/10.1002/qj.49712253107

Fletcher, C. G., Kravitz, B., & Badawy, B. (2018). Quantifying uncertainty from aerosol and atmospheric parameters and their impact on climate sensitivity. *Atmospheric Chemistry and Physics*, *18*(23), 17,529–17,543. https://doi.org/10.5194/acp-18-17529-2018

Forster, P., Ramaswamy, V., Berntsen, T., Betts, R., Fahey, D.W., Haywood, J., et al. (2007). Changes in Atmospheric Constituents and in Radiative Forcing. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. and Miller, H.L. (Eds.) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Vol. *7*). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Gantt, B., Johnson, M. S., Crippa, M., Prèvôt, A. S. H., & Meskhidze, N. (2015). Implementing marine organic aerosols into the GEOS-Chem model. *Geoscientific Model Development*, *8*(3), 619–629. https://doi.org/10.5194/gmd-8-619-2015

Ghan, S. J. (2013). Technical note: Estimating aerosol effects on cloud radiative forcing. *Atmospheric Chemistry and Physics*, *13*(19), 9971–9974. https://doi.org/10.5194/acp-13-9971-2013

Ghan, S. J., Liu, X., Easter, R. C., Zaveri, R., Rasch, P. J., Yoon, J.-H., & Eaton, B. (2012). Toward a minimal representation of aerosols in climate models: Comparative decomposition of aerosol direct, semi-direct and indirect radiative forcing. *Journal of Climate*, *25*(19), 6461–6476. https://doi.org/10.1175/JCLI-D-11-00650.1

Granier, C., Bessagnet, B., Bond, T., D'Angiola, A., Denier van der Gon, H., Frost, G. J., et al. (2011). Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period. *Climate Change*, *109*(1-2), 163–190. https://doi.org/10.1007/s10584-011-0154-1

Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., et al. (1995). A global model of natural volatile organic compound emissions. *Journal of Geophysical Research*, *100*(D5), 8873–8892. https://doi.org/10.1029/94JD02950

Haerter, J. O., Roeckner, E., Tomassini, L., & von Storch, J. S. (2009). Parametric uncertainty effects on aerosol radiative forcing. *Geophysical Research Letters*, *36*, L15707. https://doi.org/10.1029/2009GL039050

Halmer, M., Schmincke, H.-U., & Graf, H. F. (2002). The annual volcanic gas input into the atmosphere, in particular into the stratosphere: A global data-set for the past 100 years. *Journal of Volcanology and Geothermal Research*, *115*(3-4), 511–528. https://doi.org/10.1016/S0377-0273(01)00318-3

Hamilton, D. S., Hantson, S., Scott, C. E., Kaplan, J. O., Pringle, K. J., Nieradzik, L. P., et al. (2018). Reassessment of pre-industrial fire emissions strongly affects anthropogenic aerosol forcing. *Nature Communications*, *9*, 3182. https://doi.org/10.1038/s41467-018-05592-9

Heintzenberg, J., Birmili, W., Wiedensohler, A., Nowak, A., & Tuch, T. (2004). Structure, variability and persistence of the submicrometer marine aerosol. *Tellus*, *56B*, 357–367.

Hetzel J. (2012). Package 'Trapezoid', https://cran.r-project.org/web/packages/trapezoid/trapezoid.pdf (last access: 9 January 2019).

Hewitt, H. T., Copsey, D., Culverwell, I. D., Harris, C. M., Hill, R. S. R., Keen, A. B., et al. (2011). Design and implementation of the infrastructure of HadGEM3: The next-generation Met Office climate modelling system. *Geoscientific Model Development*, *4*, 223–253. https://doi.org/10.5194/gmd-4-223-2011

Jacobson, M. Z., Tabazadeh, A., & Turco, R. P. (1996). Simulating equilibrium within aerosols and nonequilibrium between gases and aerosols. *Journal of Geophysical Research*, *101*(D4), 9079–9091. https://doi.org/10.1029/96JD00348

Janhäll, S., Andreae, M. O., & Pöschl, U. (2010). Biomass burning aerosol emissions from vegetation fires: Particle number and mass emission factors and size distributions. *Atmospheric Chemistry and Physics*, *10*, 1427–1439. https://doi.org/10.5194/acp-10-1427-2010

Johnson, J. S., Regayre, L. A., Yoshioka, M., Pringle, K. J., Lee, L. A., Sexton, D. M. H., et al. (2018). The importance of comprehensive parameter sampling and multiple observations for robust constraint of aerosol radiative forcing. *Atmospheric Chemistry and Physics*, *18*, 13,031–13,053. https://doi.org/10.5194/acp-18-13031-2018

Karydis, V. A., Tsimpidi, A. P., Bacer, S., Pozzer, A., Nenes, A., & Lelieveld, J. (2017). Global impact of mineral dust on cloud droplet number concentration. *Atmospheric Chemistry and Physics*, *17*, 5601–5621. https://doi.org/10.5194/acp-17-5601-2017

Kaufman, Y. J., Boucher, O., Tanré, D., Chin, M., Remer, L. A., & Takemura, T. (2005). Aerosol anthropogenic component estimated from satellite data. *Geophysical Research Letters*, *32*, L17804. https://doi.org/10.1029/2005GL023125

Kettle, A. J., & Andreae, M. O. (2000). Flux of dimethylsulfide from the oceans. A comparison of updated data sets and flux models. *Journal of Geophysical Research*, *105*(D22), 26,793–26,808.

Khairoutdinov, M., & Kogan, Y. (2000). A new cloud physics parameterization in a large-eddy simulation model of marine stratocumulus. *Monthly Weather Review*, *128*, 229–243. https://doi.org/10.1175/1520-0493(2000)128<0229:ANCPPI>2.0.CO;2

Kipling, Z., Stier, P., Schwarz, J. P., Perring, A. E., Spackman, J. R., Mann, G. W., et al. (2013). Constraints on aerosol processes in climate models from vertically-resolved aircraft observations of black carbon. *Atmospheric Chemistry and Physics*, *13*, 5969–5986. https://doi.org/10.5194/acp-13-5969-2013

Leaitch, W. R., Sharma, S., Huang, L., Macdonald, A. M., Toom-Sauntry, D., Chivulescu, A., et al. (2013). Dimethyl sulphide control of the clean summertime arctic aerosol and cloud. *Elementa: Science of the Anthropocene*, *1*, 000017. https://doi.org/10.12952/journal.elementa.000017

Lee, L. A., Carslaw, K. S., Pringle, K. J., & Mann, G. W. (2012). Mapping the uncertainty in global CCN using emulation. *Atmospheric Chemistry and Physics*, *12*, 9739–9751. https://doi.org/10.5194/acp-12-9739-2012

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., & Spracklen, D. V. (2011). Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters. *Atmospheric Chemistry and Physics*, *11*, 12,253–12,273. https://doi.org/10.5194/acp-11-12253-2011

Lee, L. A., Pringle, K. J., Reddington, C. L., Mann, G. W., Stier, P., Spracklen, D. V., et al. (2013). The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei. *Atmospheric Chemistry and Physics*, *13*, 8879–8914. https://doi.org/10.5194/acp-13-8879-2013

Liu, X., & Wang, J. (2010). How important is organic aerosol hygroscopicity to aerosol indirect forcing? United States: N. p.. Web. https://doi.org/10.1088/1748-9326/5/4/044010

Liu, Y., & Gupta, H. V. (2007). Uncertainties in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, *43*, W07401. https://doi.org/10.1029/2006WR005756

Lohmann, U., & Ferrachat, S. (2010). Impact of parametric uncertainties on the present-day climate and on the anthropogenic aerosol effect. *Atmospheric Chemistry and Physics*, *10*, 11,373–11,383. https://doi.org/10.5194/acp-10-11373-2010

Luo, G., & Yu, F. (2011). Sensitivity of global cloud condensation nuclei concentrations to primary sulfate emission parameterizations. *Atmospheric Chemistry and Physics*, *11*, 1949–1959. https://doi.org/10.5194/acp-11-1949-2011

Mann, G. W., Carslaw, K. S., Reddington, C. L., Pringle, K. J., Schulz, M., Asmi, A., et al. (2014). Intercomparison and evaluation of global aerosol microphysical properties among AeroCom models of a range of complexity. *Atmospheric Chemistry and Physics*, *14*, 4679–4713. https://doi.org/10.5194/acp-14-4679-2014

Mann, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manktelow, P. T., Chipperfield, M. P., et al. (2010). Description and evaluation of GLOMAP-mode: A modal global aerosol microphysics model for the UKCA composition-climate model. *Geoscientific Model Development*, *3*, 519–551. https://doi.org/10.5194/gmd-3-519-2010

Mann, G. W., Carslaw, K. S., Ridley, D. A., Spracklen, D. V., Pringle, K. J., Merikanto, J., et al. (2012). Intercomparison of modal and sectional aerosol microphysics representations within the same 3-D global chemical transport model. *Atmospheric Chemistry and Physics*, *12*, 4449–4476. https://doi.org/10.5194/acp-12-4449-2012

Mather, T. A., Allen, A. G., Oppenheimer, C., Pyle, D. M., & McGonigle, A. J. S. (2003). Size-resolved characterisation of soluble ions in the particles in the tropospheric plume of Masaya volcano, Nicaragua: Origins and plume processing. *Journal of Atmospheric Chemistry*, *46*, 207–237.

McKay, M., Conover, W., & Beckman, R. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*, 239–245.

McNeall, D., Williams, J., Booth, B., Betts, R., Challenor, P., Wiltshire, A., & Sexton, D. (2016). The impact of structural error on parameter constraint in a climate model. *Earth System Dynamics*, *7*(4), 917–935. https://doi.org/10.5194/esd-7-917-2016

Met Office (2013). Unified model documentation paper F3, format of atmospheric dumps, fieldsfiles, ancillary data sets, boundary data sets, observation files and post processed (PP) files for the Unified Forecast/Climate Model, Met Office Data Analysis Applications and Met Office Storage Systems.

Metzger, A., Verheggen, B., Dommen, J., Duplissy, J., Prevot, A. S. H., Weingartner, E., et al. (2010). Evidence for the role of organics in aerosol particle formation under atmospheric conditions. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 6646–6651. https://doi.org/10.1073/pnas.0911330107

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., & Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, *430*, 768–772. https://doi.org/10.1038/nature02771

Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., et al. (2013). Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations. *Atmospheric Chemistry and Physics*, *13*, 1853–1877. https://doi.org/10.5194/acp-13-1853-2013

Neu, J., Prather, M., & Penner, J. (2007). Global atmospheric chemistry: Integrating over fractional cloud cover. *Journal of Geophysical Research*, *112*, D11306. https://doi.org/10.1029/2006JD008007

Nightingale, P. D., Liss, P. S., & Schlosser, P. (2000). Measurements of air-sea gas transfer during an open ocean algal bloom. *Geophysical Research Letters*, *27*, 2117–2120.

Oakley, J. E. & O'Hagan, A. (2010). SHELF: The Sheffield Elicitation Framework (version 2.0), http://tonyohagan.co.uk/shelf, School of Mathematics and Statistics, University of Sheffield, UK.

O'Connor, F. M., Johnson, C. E., Morgenstern, O., Abraham, N. L., Braesicke, P., Dalvi, M., et al. (2014). Evaluation of the new UKCA climate-composition model—Part 2: The troposphere. *Geoscientific Model Development*, *7*, 41–91. https://doi.org/10.5194/gmd-7-41-2014

O'Hagan, A. (2006). Bayesian analysis of computer code outputs—A tutorial. *Reliability Engineering and System Safety*, *91*, 1290–1300.

Petters, M. D., & Kreidenweis, S. M. (2007). A single parameter representation of hygroscopic growth and cloud condensation nucleus activity. *Atmospheric Chemistry and Physics*, *7*(8), 1961–1971.

Philipona, R., Behrens, K., & Ruckstuhl, C. (2009). How declining aerosols and rising greenhouse gases forced rapid warming in Europe since the 1980s. *Geophysical Research Letters*, *36*, L02806. https://doi.org/10.1029/2008GL036350

Pierce, J. R., & Adams, P. J. (2006). Global evaluation of CCN formation by direct emission of sea salt and growth of ultrafine sea salt. *Journal of Geophysical Research*, *111*, D06203. https://doi.org/10.1029/2005JD006186

Qian, Y., Yan, H., Zhangshuan, H., Gardar, J., Klein, S., Lucas, D., et al. (2015). Parametric sensitivity analysis of precipitation at global and local scales in the Comunity Atmosphere Model CAM5. *Journal of Advances in Modeling Earth Systems*, *7*, 382–411. https://doi.org/10.1002/2014MS000354

Randles, C. A., Kinne, S., Myhre, G., Schulz, M., Stier, P., Fisher, J., et al. (2013). Intercomparison of shortwave radiative transfer schemes in global aerosol modeling: Results from the AeroCom radiative transfer experiment. *Atmospheric Chemistry and Physics*, *13*(5), 2347–2379. https://doi.org/10.5194/acp-13-2347-2013

Rastak, N., Pajunoja, A., Acosta Navarro, J. C., Ma, J., Song, M., Partridge, D. G., et al. (2017). Microphysical explanation of the RH-dependent water affinity of biogenic organic aerosol and its importance for climate. *Geophysical Research Letters*, *44*, 5167–5177. http://doi.org/10.1002/2017GL073056

Reddington, C. L., Carslaw, K. S., Stier, P., Schutgens, N., Coe, H., Liu, D., et al. (2017). The Global Aerosol Synthesis and Science Project (GASSP). *Bulletin of the American Meteorological Society*, *8*, 1857–1877. https://doi.org/10.1175/BAMS-D-15-00317.1

Regayre, L. A., Johnson, J. S., Yoshioka, M., Pringle, K. J., Sexton, D. M. H., Booth, B. B. B., et al. (2018). Aerosol and physical atmosphere model parameters are both important sources of uncertainty in aerosol ERF. *Atmospheric Chemistry and Physics*, *18*, 9975–10006. https://doi.org/10.5194/acp-18-9975-2018

Regayre, L. A., Pringle, K. J., Booth, B. B. B., Lee, L. A., Mann, G. W., Browse, J., et al. (2014). Uncertainty in the magnitude of aerosol-cloud radiative forcing over recent decades. *Geophysical Research Letters*, *41*, 9040–9049. https://doi.org/10.1002/2014GL062029

Regayre, L. A., Pringle, K. J., Lee, L. A., Booth, B. B. B., Rap, A., Browse, J., et al. (2015). The climatic importance of uncertainties in regional aerosol-cloud radiative forcings over recent decades. *Journal of Climate*, *28*, 6589–6607. https://doi.org/10.1175/JCLI-D-15-0127.1

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution blended analyses for sea surface temperature. *Journal of Climate*, *20*, 5473–5496.

Rodrigues, L. F. S., Vernon, I., & Bower, R. (2017). Constraints on galaxy formation models from the galaxy stellar mass function and its evolution. *Monthly Notices of the Royal Astronomical Society*, *466*, 2418–2435. https://doi.org/10.1093/mnras/stw3269

Rosenfeld, D., Lohmann, U., Raga, G. B., O'dowd, C. D., Kulmala, M., Fuzzi, S., et al. (2008). Flood or Drought: How do aerosols affect precipitation? *Science*, 1309–1313.

Saltelli, A., Chan, K., & Scott, E. M. (2000). *Sensitivity analysis*. Oxford, UK: Wiley.

Saltelli, A., Tarantola, S., & Chan, K. P. S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, *41*, 39–56.

Salter, J. M., & Williamson, D. (2016). A comparison of statistical emulation methodologies for multi-wave calibration of environmental models. *Environmetrics*, *27*(8), 507–523. https://doi.org/10.1002/env.2405

Schmale, J., Henning, S., Decesari, S., Henzing, B., Keskinen, H., Sellegri, K., et al. (2018). Long-term cloud condensation nuclei number concentration, particle number size distribution and chemical composition measurements at regionally representative observatories. *Atmospheric Chemistry and Physics*, *18*, 2853–2881. https://doi.org/10.5194/acp-18-2853-2018

Schmale, J., Henning, S., Henzing, B., Keskinen, H., Sellegri, K., Ovadnevaite, J., et al. (2017). Collocated observations of cloud condensation nuclei, particle size distributions, and chemical composition. *Scientific Data*, *4*, 170003. https://doi.org/10.1038/sdata.2017.3

Schulz, M., Textor, C., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., et al. (2006). Radiative forcing by aerosols as derived from the AeroCom present-day and pre-industrial simulations. *Atmospheric Chemistry and Physics*, *6*, 5225–5246. https://doi.org/10.5194/acp6-5225-2006

Schwartz, S. E. (2018). Unrealized global temperature increase: Implications of current uncertainties. *Journal of Geophysical Research: Atmospheres*, *123*, 3462–3482. https://doi.org/10.1002/2017JD028121

Seinfeld, J. H., Bretherton, C., Carslaw, K. S., Coe, H., DeMott, P. J., Dunlea, E. J., et al. (2016). Improving our fundamental understanding of the role of aerosol−cloud interactions in the climate system. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(21), 5781–5790. https://doi.org/10.1073/pnas.1514043113

Sexton, D.M.H., Karmalkar, A., Murphy, J., & Booth, B. B. B. (2018). The elicitation of distributions of parameters in HadGEM3 versions GA4 and GA7 for use in perturbed parameter ensembles, Hadley Centre technical note 101, Met Office, U.K.

Sexton, D. M. H., Murphy, J. M., Collins, M., & Webb, M. J. (2012). Multivariate probabilistic projections using imperfect climate models Part I: Outline of methodology. *Climate Dynamics*, *38*, 2513–2542.

Shindell, D., & Faluvegi, G. (2009). Climate response to regional radiative forcing during the twentieth century. *Nature Geoscience*, *2*, 294–300. https://doi.org/10.1038/ngeo473

Shindell, D. T., Lamarque, J.-F., Schulz, M., Flanner, M., Jiao, C., Chin, M., et al. (2013). Radiative forcing in the ACCMIP historical and future climate simulations. *Atmospheric Chemistry and Physics*, *13*, 2939–2974. https://doi.org/10.5194/acp-13-2939-2013

Shiogama, H., Watanabe, M., Yoshimori, M., Yokohata, T., Ogura, T., Annan, J. D., et al. (2012). Perturbed physics ensemble using the MIROC5 coupled atmosphere-ocean GCM without flux corrections: Experimental design and results. *Climate Dynamics*, *39*, 3041–3056. https://doi.org/10.1007/s00382-012-1441-x

Slinn, W. G. N. (1982). Predictions for particle deposition to vegetative canopies. *Atmospheric Environment*, *16*, 1785–1794.

Smith, D. M., Booth, B. B. B., Dunstone, N. J., Eade, R., Hermanson, L., Jones, G. S., et al. (2016). Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, *6*, 936–940. https://doi.org/10.1038/nclimate3058

Spracklen, D. V., Carslaw, K. S., Kulmala, M., Kerminen, V.-M., Mann, G. W., & Sihto, S.-L. (2006). The contribution of boundary layer nucleation events to total particle concentrations on regional and global scales. *Atmospheric Chemistry and Physics*, *6*, 5631–5648. https://doi.org/10.5194/acp-6-5631-2006

Spracklen, D. V., Pringle, K. J., Carslaw, K. S., Chipperfield, M. P., & Mann, G. W. (2005). A global off-line model of size-resolved aerosol microphysics: I. Model development and prediction of aerosol properties. *Atmospheric Chemistry and Physics*, *5*, 2227–2252. https://doi.org/10.5194/acp-5-2227-2005

Stevens, B., & Feingold, G. (2009). Untangling aerosol effects on clouds and precipitation in a buffered system. *Nature*, *461*, 607–613.

Stevens, R. G., Pierce, J. R., Brock, C. A., Reed, M. K., Crawford, J. H., Holloway, J. S., et al. (2012). Nucleation and growth of sulfate aerosol in coal-fired power plant plumes: Sensitivity to background aerosol and meteorology. *Atmospheric Chemistry and Physics*, *12*, 189–206. https://doi.org/10.5194/acp-12-189-2012

Stier, P., Feichter, J., Roeckner, E., Kloster, S., & Esch, M. (2006). The evolution of the global aerosol system in a transient climate simulation from 1860 to 2100. *Atmospheric Chemistry and Physics*, *6*, 3059–3076. https://doi.org/10.5194/acp-6-3059-2006

Stier, P., Schutgens, N. A. J., Bian, H., Boucher, O., Chin, M., Ghan, S., et al. (2013). Host model uncertainties in aerosol radiative forcing estimates: Results from the AeroCom prescribed intercomparison study. *Atmospheric Chemistry and Physics*, *13*, 3245–3270. https://doi.org/10.5194/acp-13-3245-2013

Stokes, R. H., & Robinson, R. A. (1966). Interactions in aqueous nonelectrolyte solutions. I. Solute-solvent equilibria. *Journal of Physical Chemistry*, *70*, 2126–2130.

Telford, P. J., Abraham, N. L., Archibald, A. T., Braesicke, P., Dalvi, M., Morgenstern, O., et al. (2013). Implementation of the Fast-JX photolysis scheme (v6.4) into the UKCA component of the MetUM chemistry-climate model (v7.3). *Geoscientific Model Development*, *6*, 161–177. https://doi.org/10.5194/gmd-6-161-2013

Telford, P. J., Braesicke, P., Morgenstern, O., & Pyle, J. A. (2008). Technical note: Description and assessment of a nudged version of the new dynamics Unified Model. *Atmospheric Chemistry and Physics*, *8*, 1701–1712. https://doi.org/10.5194/acp-8-1701-2008

Tunved, P., Ström, J., & Krejci, R. (2013). Arctic aerosol life cycle: Linking aerosol size distributions observed between 2000 and 2010 with air mass transport and precipitation at Zeppelin station, Ny-Ålesund, Svalbard. *Atmospheric Chemistry and Physics*, *13*, 3643–3660. https://doi.org/10.5194/acp-13-3643-2013

Turnock, S. T. (2016). The impact of changing anthropogenic emissions on European atmospheric aerosols over the second half of the 20th Century. PhD thesis, University of Leeds.

Vehkamäki, H., Kulmala, M., Napari, I., Lehtinen, K. E. J., Timmreck, C., Noppel, M., & Laaksonen, A. (2002). An improved parameterization for sulfuric acid-water nucleation rates for tropospheric and stratospheric conditions. *Journal of Geophysical Research*, *107*(D22), 4622. https://doi.org/10.1029/2002JD002184

Walters, D. N., Williams, K. D., Boutle, I. A., Bushell, A. C., Edwards, J. M., Field, P. R., et al. (2014). The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations. *Geoscientific Model Development*, *7*, 361–386. https://doi.org/10.5194/gmd-7-361-2014

Wang, Q., Jacob, D. J., Spackman, J. R., Perring, A. E., Schwarz, J. P., Moteki, N., et al. (2014). Global budget and radiative forcing of black carbon aerosol: Constraints from pole-to-pole (HIPPO) observations across the Pacific. *Journal of Geophysical Research: Atmospheres*, *119*, 195–206. https://doi.org/10.1002/2013JD020824

West, R. E. L., Stier, P., Jones, A., Johnson, C. E., Mann, G. W., Bellouin, N., et al. (2014). The importance of vertical velocity variability for estimates of the indirect aerosol effects. *Atmospheric Chemistry and Physics*, *14*, 6369–6393. https://doi.org/10.5194/acp-14-6369-2014

Wild, O., Zhu, X., & Prather, M. J. (2000). Fast-J: Accurate simulation of in- and below-cloud photolysis in tropospheric chemical models. *Journal of Atmospheric Chemistry*, *37*, 245. https://doi.org/10.1023/A:1006415919030

Woodward, S. (2001). Modeling the atmospheric life cycle and radiative impact of mineral dust in the Hadley Centre climate model. *Journal of Geophysical Research*, *106*(D16), 18,155–18,166. https://doi.org/10.1029/2000JD900795

Yokohata, T., Webb, M. J., Collins, M. J., Williams, K. D., Yoshimori, M., Hargreaves, J. C., & Annan, J. D. (2010). Structural similarities and differences in climate responses to CO_2 increase between two perturbed physics ensembles by general circulation models. *Journal of Climate*, *23*, 1392–1410. https://doi.org/10.1175/2009JCLI2917.1

Zadanovskii, A. B. (1948). New methods for calculating solubilities of electrolytes in multicomponent systems. *Zhurnal Fizicheskoi Khimii*, *22*, 1475–1485.