On the relationship between aerosol model uncertainty and radiative forcing uncertainty

Lindsay A. Lee[a], Carly L. Reddington[a], and Kenneth S. Carslaw[a,1]

[a]School of Earth and Environment, University of Leeds, Leeds LS2 9JT, United Kingdom

The largest uncertainty in the historical radiative forcing of climate is caused by the interaction of aerosols with clouds. Historical forcing is not a directly measurable quantity, so reliable assessments depend on the development of global models of aerosols and clouds that are well constrained by observations. However, there has been no systematic assessment of how reduction in the uncertainty of global aerosol models will feed through to the uncertainty in the predicted forcing. We use a global model perturbed parameter ensemble to show that tight observational constraint of aerosol concentrations in the model has a relatively small effect on the aerosol-related uncertainty in the calculated forcing between preindustrial and present-day periods. One factor is the low sensitivity of present-day aerosol to natural emissions that determine the preindustrial aerosol state. However, the major cause of the weak constraint is that the full uncertainty space of the model generates a large number of model variants that are equally acceptable compared to present-day aerosol observations. The narrow range of aerosol concentrations in the observationally constrained model gives the impression of low aerosol model uncertainty. However, these multiple "equifinal" models predict a wide range of forcings. To make progress, we need to develop a much deeper understanding of model uncertainty and ways to use observations to constrain it. Equifinality in the aerosol model means that tuning of a small number of model processes to achieve model–observation agreement could give a misleading impression of model robustness.

aerosol–cloud interaction | uncertainty analysis | model constraint

All models of environmental systems are uncertain because the model design cannot be a faithful reproduction of reality, and the magnitudes of many model quantities (parameters) are poorly defined by theory or experiment. To make models "accurate," they are therefore evaluated against measurements so that the model state matches the observed state as closely as possible. However, models need to make predictions outside the conditions for which they were evaluated, and they need to be able to predict quantities for which observational constraints do not exist. Because of the difficulty of performing a sufficient number of model simulations that sample all possible settings of uncertain processes, it is rarely demonstrated how observational constraints truly reduce the uncertainty in related predicted quantities. Here we show that, even when extensive measurements are available to constrain a particular model output, other closely related outputs may not be well constrained at all.

We focus on aerosol radiative forcing as an important climate model uncertainty and attempt to reduce the uncertainty based on idealized present-day (PD) aerosol measurements. Changes in aerosols between the preindustrial (PI) and PD periods have caused changes in cloud properties, resulting in a very uncertain radiative forcing of between $-1.33\ \mathrm{W \cdot m^{-2}}$ and $-0.06\ \mathrm{W \cdot m^{-2}}$ (1) calculated across multiple global climate models. This large uncertainty remains despite our increased understanding of aerosol science and the improvement in model fidelity and continues to limit our ability to estimate the climate sensitivity (2–4). Although the modulating effect of aerosols on clouds can be observed in the modern atmosphere (5), the PI to PD aerosol–cloud radiative forcing is not directly measureable because the PI aerosol state is

unknown. The PI aerosol state has been shown to be a major component of the forcing uncertainty (6–8). Satellite measurements of aerosol and cloud properties have been used to estimate the forcing (1, 9, 10), but they are also dependent on unverifiable assumptions about the PI aerosol and cloud state (1, 11, 12), and cloud–aerosol relationships may not extrapolate well to the appropriate low-aerosol conditions (13). We therefore rely on models evaluated against PD conditions to predict how historical changes in aerosols affect cloud radiative forcing.

In developing and evaluating models, there is an implicit and reasonable assumption that, if we can reduce the uncertainty of the aerosol component of a model, we will also reduce the forcing uncertainty. In this paper, we investigate this assumption for aerosol–cloud forcing by quantifying the relationship between the uncertainty in an aerosol model and the uncertainty in aerosol–cloud forcing predicted by that model.

Fig. 1 outlines the nature of the model constraint problem. We simulate aerosol properties A [such as cloud condensation nucleus concentration (CCN) or aerosol optical depth] with uncertainties that are determined by a large set of uncertain processes in the aerosol model. The combined and interacting effects of all of the process uncertainties result in an uncertainty in the magnitude of the aerosol property A, represented by a probability distribution in each global grid cell (Fig. 1A). We use the model to predict the forcing F, which will have an associated uncertainty caused by the uncertainty in the aerosol model. In this paper, we examine the link between the uncertainty in the aerosol component of the global model and the effect it has on the uncertainty in forcing.

The task of aerosol model constraint involves finding models that predict aerosol properties A within the observed range (Fig. 1C) and quantifying the spread of forcings from this smaller set of models (Fig. 1D). Normally, because of computational demands, we do not know the probability distribution of the model outputs. Rather, we have only a small number of structurally different models (14) or a few sensitivity tests in which single parameters are varied within one model. Because these approaches provide a very limited sampling of model uncertainties, it has not previously been possible to quantify the relationship between uncertainty in A and F. However, based on refs. 15 and 6, we have information about the probability distribution of aerosol properties and forcing from a sampling of 140,000 runs exploring all dimensions of 28 uncertain parameters of a global aerosol model (from an emulator),
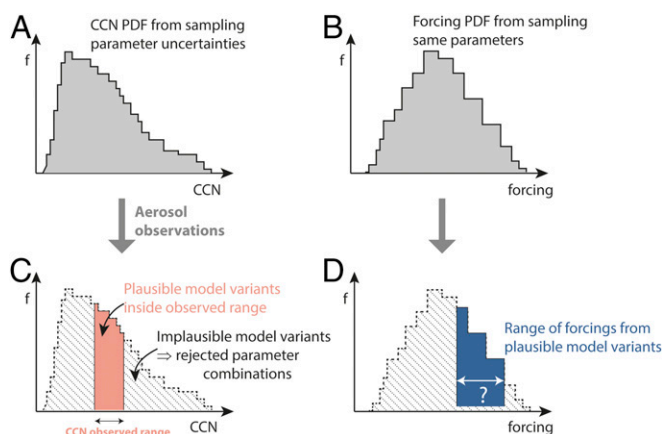
Fig. 1. A schematic outlining the nature of the model constraint problem. (*A* and *B*) The prior CCN (*A*) and forcing (*B*) uncertainty determined by perturbation of 28 uncertain aerosol parameters in a global model. (*C*) We identify a plausible range of CCN values compared with observations with known uncertainty and constrain the 28-dimensional parametric uncertainty space. (*D*) We calculate the forcing constraint associated with the plausible parameter space determined from the CCN constraint.

which allows us to quantify this relationship for the first time, to our knowledge.

The question we pose is: If we could constrain a model using aerosol measurements A with known uncertainty in every grid box of the model (Fig. 1*C*), how well would we be able to constrain the uncertainty in forcing F (Fig. 1*D*) caused by the uncertainty of the aerosol model? There are several factors that may limit the extent to which aerosol observations will help narrow the forcing uncertainty.

First, F depends on model processes that do not affect A. This is, of course, the case because F is additionally sensitive to cloud distribution as well as model processes like updraft speed and cloud entrainment that do not greatly affect aerosols. However, here we explore only the effects of uncertainty in the aerosol component of the model and design the experiment such that uncertainties in F and A are fully defined by the same set of parameters (i.e., there are no hidden sources of uncertainty that are unique to F). Although this is only part of the forcing uncertainty problem, the uncertainty range due to aerosol processes alone is comparable to that calculated from multiple models (6).

Second, although uncertainties in F and A depend on the same set of processes, the dependencies can be different. If F is very sensitive to a particular parameter but A is less sensitive, then observations of A can only weakly constrain F. Nevertheless, we might expect that, by making global measurements of A, we can sample environments in which, eventually, all parameter sensitivities become important and can be constrained by the observations.

Third, F depends on the change in aerosols between an historical baseline period and the present day, but we can observe A only in the present day. A large fraction of the uncertainty in forcing can be attributed to natural aerosols that determine PI conditions (6), and natural aerosol sensitivities are suppressed by pollution (16). Nevertheless, the modeled PI and PD aerosol properties are determined by the same model processes, so we expect a model that is constrained using PD measurements to have increased skill in simulating PI conditions also. Furthermore, if we can measure A in environments that have changed little since the baseline period (12), then we might be able to constrain the processes that have the greatest influence on aerosols in PI environments.

Fourth, F is nonlinearly related to A (6). At very low aerosol concentrations, F depends quite steeply on changes in A, but, in polluted high-aerosol environments, F becomes increasingly insensitive to changes in A. Under clean conditions, F is much more sensitive to the constraint of A. This itself is likely to be an irreducible uncertainty in F unless the PI baseline aerosol can be defined exactly.

Fifth, there are many compensating model factors that affect A, so the values of the process parameters will not be uniquely defined by measurements of A. For example, errors in deposition and emission rates could be difficult to separate using measurements of A alone because they have compensating effects. This problem is termed "equifinality" (17) and is well recognized in other fields of environmental science. Equifinality is not a problem if the aim is to develop a model of A for PD conditions. However, equifinality can become a problem when trying to make predictions outside the conditions for which the model was constrained. Nevertheless, the dependence of A on the model parameters will vary substantially across the globe and throughout the year. The question is whether there are enough different environments providing enough relationships between A and the parameters that we can eventually overcome equifinality.

The above considerations lead to the following questions, which we address in this paper.

How much does the uncertainty in modeled forcing fall when we constrain the aerosol model using measurements of known uncertainty? Can we quantify how a constraint on aerosol uncertainty translates to a constraint on forcing uncertainty and identify the best observations to constrain forcing uncertainty? Can we identify the global environments in which different parameter sensitivities are important for aerosol uncertainty to ensure our observations sample these different environments and maximize the possible constraint on forcing?

Methodology

To study the effects of observational constraints, we use a parametric uncertainty analysis of global aerosol (15) and aerosol–cloud forcing (6) (Fig. 2 *C* and *D*).

We define a parameter to be any scalar value in the global aerosol model that can be perturbed to investigate the uncertainty in different model processes, including scaling or absolute perturbations to nucleation, deposition, emission rates, and emitted particle sizes. Both studies were based on the same perturbed parameter ensemble of 28 parameters of the GLObal Model of Aerosol Processes (GLOMAP) global aerosol model, which was run over 1 y driven by meteorological reanalyses. The uncertainties and associated parameter sensitivities were calculated from sampling validated Bayesian emulators (18, 19) of CCN and forcing. The emulators are fast surrogate models of GLOMAP that define 28-dimensional response surfaces of a monthly mean model output in each grid box in terms of variations in model parameters related to the aerosol component of the global model. The emulators were conditioned on a 168-member perturbed parameter ensemble of GLOMAP with the 28 parameter values sampled in such a way that the emulators capture covariations between parameter effects. To test the reliability of the emulators, they were validated against a further 84 GLOMAP simulations.

Three aerosol properties are used to evaluate model constraint, which we choose based on the wide availability of measurements. The CCN concentration at a defined supersaturation is a measured aerosol quantity of direct relevance to cloud droplet concentrations and hence cloud albedo. We also use the total particle concentration (particles larger than 3 nm diameter, N_3) and black carbon (BC) mass concentration (6). Although N_3 and BC appear to be less closely related to cloud droplet concentrations and cloud albedo than CCN, the key consideration is which processes control the aerosol property. If two aerosol properties are controlled by shared parameters (such as wet deposition), then constraint of one aerosol property will influence the constraint of the other. Below we show that N_3 measurements can also constrain forcing, but BC less so. Relationships between aerosol and cloud properties are often used to constrain model processes and forcing (e.g., ref. 9). We do not use such observational constraints here because our aim is to understand how uncertainty in the aerosol model itself affects forcing uncertainty. Most of the uncertainty in aerosol–cloud relations will be related to cloud physics processes, which were not part of the experimental design of the ensembles we use (6, 15). These processes would therefore increase

the forcing uncertainty range reported in Carslaw et al. (6) and would need to be constrained using aerosol–cloud process-based metrics.

The parametric uncertainty of annual global mean aerosol–cloud forcing (figure 1 in ref. 6) has a 95% credible interval of -0.7 W·m^{-2} to -1.6 W·m^{-2}. Our uncertainty estimate is of comparable size to the multimodel ensemble range of -0.06 W·m^{-2} to -1.33 W·m^{-2} estimated by the Intergovernmental Panel on Climate Change (IPCC) (1). The forcing in each grid box is assumed to be determined solely by changes in cloud albedo with the global distribution of other cloud properties (coverage, depth, etc.) the same in each model run—i.e., we explore the first indirect or cloud brightening effect rather than the effective radiative forcing that allows for cloud adjustment to changes in aerosols (1). Therefore, the uncertainty in the forcing is determined by uncertainty in the changes in cloud droplet concentration and hence only by the 28 perturbations applied to the aerosol component of the model. We can therefore determine precisely how aerosol measurements with well-defined uncertainty feed through to forcing. In reality, rapid adjustments to cloud properties in response to changes in aerosol can occur (1), which will introduce further uncertainties in forcing that we do not account for here. Additional observation types, such as cloud liquid water path, will be required to reduce the further uncertainties introduced by rapid adjustments, so we are considering only a subset of all sources of forcing uncertainty, albeit with a sufficiently large uncertainty range.

To constrain the aerosol model, we first generate a very large ensemble of model outputs by sampling densely from the emulators across the full uncertainty space of the 28 parameters. This generates a probability distribution of aerosol properties and forcing in each grid box (or an average over any region), shown schematically in Fig. 1. We then reduce the sample by discarding ensemble members that predict aerosol properties that are deemed implausible compared with the measurements. By reducing the range of allowable aerosol properties, the set of uncertain parameter values is also reduced. Approximately 3 million samples were drawn from the emulators, using a Monte Carlo approach. This is a sufficiently large number that, when the 28-dimensional parameter space is reduced, there remain enough points to perform statistical analysis.

This approach, often called history matching (or precalibration), has been used effectively in other fields (20–22). When dealing with a complex multidimensional uncertainty space, history matching is an effective way to constrain a model within observed limits, which is impossible by tuning individual parameters. Similarly, generalized likelihood uncertainty estimation (23) constrains the model by identifying behavioral and nonbehavioral parameter sets (or models), with the addition of assigning probability to the remaining uncertain parameter sets based on its closeness to observations.

Results and Discussion

Defining Optimum Measurement Locations. Our aim is to constrain the aerosol component of the global model in every grid cell. However, we show below that an analysis of the uncertainties in the model allows this to be achieved by making a small number of measurements in locations that represent the dominant parameter sensitivities.

Fig. 2 shows the effect of constraining PD CCN concentrations using synthetic measurements in just two locations—over the Southern Ocean (Fig. 2E) and over Central Europe (Fig. 2G). The observational constraint is actually an idealized CCN measurement (with an assumed very small uncertainty of ±0.5%) taken from a model simulation with all parameters set to the expert-defined median value. This approach guarantees that a reduced part of parameter space can be found by history matching (see *Methodology*).

As hypothesized, each measurement constrains the model far beyond the location of the measurement. Measurements in the Southern Hemisphere marine region, for example, lead to a reduction in CCN uncertainty over most marine regions, showing that, in this model, the processes and emissions that account for CCN uncertainty are fairly common to all marine environments. Likewise, a CCN measurement over Central Europe (Fig. 2G) leads to very large reductions in uncertainty over similarly polluted regions of North America, Europe, and Asia, and extending over much of the North Atlantic. There is an assumption in our approach that the parameter uncertainties can be applied uniformly across the globe (*SI Methodology*). Therefore, for example, if CCN measurements over Europe allow a part of parameter space to be eliminated that has high deposition rates, then this part of parameter space will be eliminated globally. Whether local constraints actually work globally

in reality could easily be confirmed by using multiple real measurements. As a result, our approach is likely to overestimate the feasible uncertainty reduction for both CCN and forcing.

The spatial patterns of uncertainty reduction in Fig. 2 E and G suggest that uncertainty could be reduced globally by making a small number of measurements in regions that are representative of the model's sources of uncertainty. We defined 11 measurement locations that are representative of larger model environments using k-means cluster analysis (24) to group locations in which similar parameter combinations account for the CCN uncertainty (Fig. S1). The two measurement locations used in Fig. 2 were placed at points within their clusters that are most representative of the cluster uncertainties. Although we have sampled parameter values across 28-dimensional space, typically about 90% of the CCN and forcing variance in any one location is accounted for by less than 10 parameters (Fig. S2). Examination of the average parameter contribution to uncertainty in each cluster (Fig. S2) shows that physical meaning can be applied to the cluster environments despite being defined through purely statistical methodology.

As we show in *Constraint of CCN and Forcing*, further measurements within the same cluster have a progressively weaker effect on model constraint, because the same parameters are repeatedly constrained. Based on this result, we propose that, for the purposes of model uncertainty reduction, it is not necessary to make measurements with global coverage. Rather, the key objective should be to make measurements in regions that are representative of the model uncertainty to be constrained.

Constraint of CCN and Forcing. For a single measurement made at a geographical location that is most representative of a model uncertainty cluster, the CCN variance is reduced in individual grid boxes of the model by between 0% and greater than 90% (compare the uncertainty reduction maps in Fig. 2 E and G with the cluster maps in Fig. S1). The constraint becomes weaker toward the geographical margins of the cluster where the sources of uncertainty differ most from where the measurement was made. Uncertainty is reduced elsewhere (outside the cluster geographical margins) because most clusters share at least some common parameter relationships that are important for CCN uncertainty.

Fig. 2 shows that there is much less constraint on the forcing uncertainty than on the CCN uncertainty. Also, the regions of forcing constraint are different than those for CCN, even accounting for the spatial pattern of cloudiness. For example, a measurement in the Southern Hemisphere marine region reduces the forcing uncertainty over parts of the continental Northern Hemisphere (Fig. 2F) because of shared but spatially different combinations of parameters that control CCN and forcing variance.

Fig. 3 shows the probability distribution of the prior and constrained CCN and forcing in the North Atlantic as a result of making 10 CCN measurements throughout the northern hemispheric marine region. The measurements have an assumed uncertainty of 30%, leading to constraint of the CCN range as expected. However, the forcing is hardly constrained at all, with almost the full range between 0 W·m^{-2} and -0.9 W·m^{-2} being plausible for a model with a much narrower plausible range of PD CCN. This implies that eliminating model variants that predict low ($\sim$0–200 cm^{-3}) or high ($\sim$400–800 cm^{-3}) CCN concentrations in the PD North Atlantic has little effect on the range of forcings predicted by the model.

Fig. 4 summarizes the constraint on CCN and forcing in terms of the quantity

$$\text{global relative constraint} = 1 - \overline{\left(\frac{var_{constrained}}{var_{unconstrained}} \right)},$$

where the mean is over all model grid boxes at approximately cloud base (915 hPa) for CCN and over all top of the atmosphere
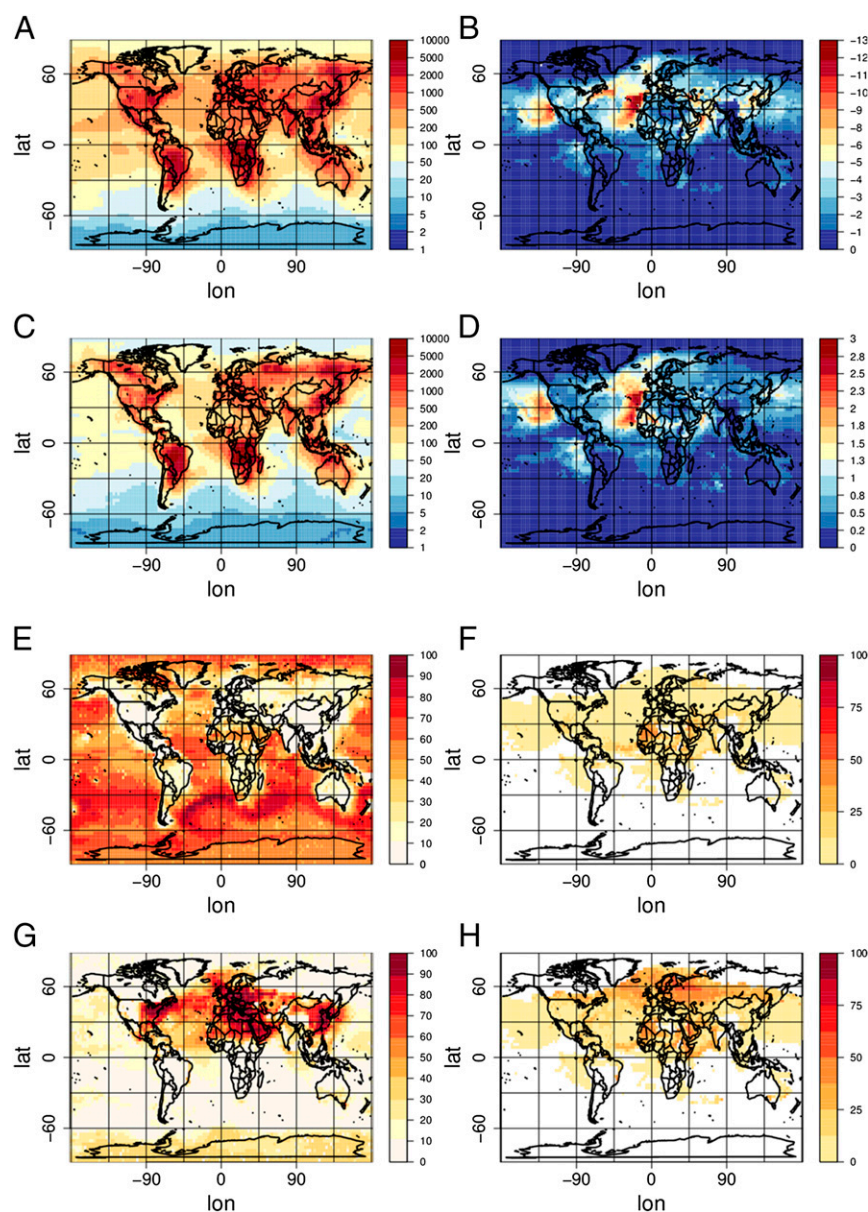
Fig. 2. Model CCN and forcing, uncertainty, and the constraint offered by CCN measurements in July. (*A*) Present-day CCN concentrations (cm^{-3}). (*B*) Top of the atmosphere aerosol–cloud radiative forcing between 1750 and 2000 (W·m^{-2}). (*C*) Prior (unconstrained) present-day CCN standard deviation. (*D*) Prior (unconstrained) forcing standard deviation. (*E*) Constraint on CCN using a single representative CCN measurement in the Southern Ocean (70° E, 50° S; see Fig. S1). The constraint on the model is the reduction in CCN variance as a percentage of the initial unconstrained variance. (*F*) Constraint on forcing when present-day model CCN is constrained as in *E*. (*G*) Constraint on CCN using a single representative CCN measurement over Europe (10° E, 50° N; see Fig. S1). (*H*) Constraint on forcing when present-day CCN is constrained as in *G*.

grid boxes for forcing. The single representative measurements made in each of the 11 clusters reduce the global CCN variance by about 5–40% (Fig. 4, small symbols with dots). However, the forcing is constrained only by 0–12%.

Fig. 4 is an ideal way to directly compare the magnitudes of CCN and forcing uncertainty reduction, but it might present a misleading picture of how much the forcing is constrained if the global mean values are dominated by large relative constraints of small absolute forcings. We tested this by varying the threshold value of forcing above which each grid box is included in the calculation of the global mean relative constraint, which shows that the threshold of 1 W·m^{-2} used in Fig. 4 provides a useful measure of mean forcing uncertainty reduction representative of regions where forcing is important (Fig. S3). Other choices of the threshold would not change our conclusions.

Multiple measurements at 10 random locations within a single cluster provide further constraint of CCN (Fig. 4, small open symbols), but the additional constraint from successive measurements decreases because similar parameters are repeatedly constrained and the unique information content of further observations

is reduced. This result shows that the value of multiple measurements lies more in determining a representative aerosol state than in obtaining additional unique information for uncertainty reduction. For these multiple measurements, we assumed a CCN measurement uncertainty of 30% to account for the likely real CCN variability within the clusters, although the choice of the assumed CCN uncertainty is not important because we are comparing CCN and forcing relative constraints. The CCN was also constrained successively by measurements representative of each of the 11 clusters assuming thresholds of 40%, 50%, and 90% on CCN variability. These globally distributed measurements constrain CCN more than multiple measurements from within a single cluster. This occurs because the sensitive parameters are different in each cluster, so different regions of the 28-dimensional uncertainty space can be constrained.

The global mean CCN relative constraint from single and multiple measurements ranges from about 5% to 85% reduction in variance relative to the prior range, whereas the forcing uncertainty constraint is in the range 3–40%. The ratios of forcing constraint to CCN constraint (dashed lines in Fig. 4) lie mostly in
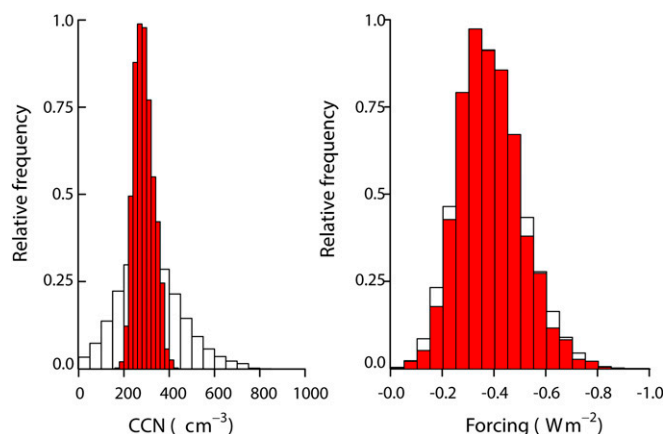
Fig. 3. CCN and forcing probability density functions over the North Atlantic in January before (white) and after (red) the model parameters are constrained using ten North Atlantic CCN measurements with 30% uncertainty. The CCN is well constrained, as expected, but the range of forcing due to the same 28 constrained parameters has hardly reduced.

the range 0.125–0.75, with a mean of about 0.25—that is, forcing is constrained on average only about one-quarter as much as CCN, even though the same set of parameters are being constrained and there are no additional sources of uncertainty unique to forcing in this study.

Causes of Weak Forcing Constraint. By analyzing the dependence of CCN and forcing on the parameters we can explore the reasons for weak forcing constraint. CCN and forcing uncertainty depends on the same 28 parameters, so the region of plausible model performance defined by the CCN measurements will be a complex 28-dimensional parameter space. We therefore illustrate the problem in just two dimensions in Fig. 5 (a schematic is shown in Fig. S4). The model results in Fig. 5 come from the representative grid box in the anthropogenic source region (Fig. S1). These diagrams were generated by densely sampling from the emulator while keeping 26 of the 28 parameters at their median settings.

Fig. 5 *A* and *B* shows a case in which uncertainty in forcing is caused by parameters that PD CCN concentrations are insensitive to. This problem can occur because the forcing is sensitive to

parameters that determine the PI baseline aerosol state (6), but the effects of these parameters on PD CCN can be masked by air pollution (12, 16). In Fig. 5*A*, a CCN measurement of 600–700 cm^{-3} strongly constrains the Aitken mode width but leaves the dimethyl sulfide (DMS) flux essentially unconstrained because CCN is not strongly sensitive to DMS in the anthropogenic source region in the model. However, the forcing here depends strongly on DMS flux (6) (Fig. 5*B*), which leaves the modeled forcing unconstrained in the range −0.72 $W \cdot m^{-2}$ to −1.0 $W \cdot m^{-2}$.

Fig. 5 *C* and *D* shows a case in which two parameters contribute similarly to the uncertainty in forcing and CCN but the dependencies are very different. In this case, both parameters would contribute approximately equally to the variance in CCN and forcing, which would suggest that uncertainty in both would fall if CCN concentrations were constrained by measurements. In this case, a CCN measurement would tightly constrain the relationship between the two parameters, but neither of the individual parameters is well constrained, thereby leaving the prior forcing range essentially unconstrained. To further illustrate this problem, Fig. S5 shows the contributions of parameters to uncertainty in CCN and forcing in the Northern Hemisphere marine location shown in Figs. 3 and 6 and Fig. S1. The sources of uncertainty in CCN and forcing are similar. However, our results show that 10 CCN measurements with a 30% uncertainty across the Northern Hemisphere marine uncertainty cluster reduce the CCN variance in a North Atlantic grid box by ∼75% as expected, but the forcing uncertainty remains very close to the unconstrained range of 0 $W \cdot m^{-2}$ to −0.9 $W \cdot m^{-2}$. Therefore, even when the sources of uncertainty in CCN and forcing are very similar, the reductions in the magnitudes of the uncertainties are very different.

The examples in Fig. 5 show fundamentally why a tightly constrained model of PD aerosol concentrations does not necessarily imply that the modeled forcing will be any better constrained as a result, even when the same set of uncertainty sources is considered. The key problem is equifinality (17): There are multiple parameter combinations (or model variants) that produce an equally good (or equifinal) model of CCN, but these models predict a wide range of forcings. In the case shown in Fig. 5, the dark shaded region shows that CCN is well constrained, but the forcing is essentially unconstrained. Although we have shown only two dimensions, the problem extends to all of the dimensions of the important parameters, which is typically about 10 at any one location. Therefore, when aerosol measurements are used to
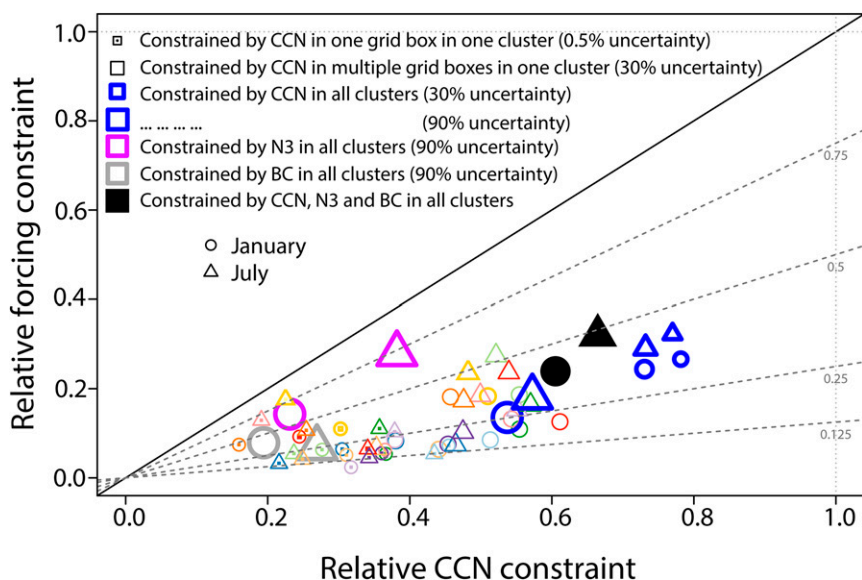

Fig. 4. Comparison of the constrained global mean CCN and forcing for various constraint experiments in January and July. The symbols containing dots represent constraint of model parameters using a CCN measurement (with 0.5% uncertainty) in a single representative grid box in one cluster from Fig. S1. Each color represents a different cluster according to the map in Fig. S1. The small open symbols represent constraint from CCN measurements (with 30% uncertainty) in ten random grid boxes in one cluster. The large bold open symbols show the constraint using either CCN, N3, or BC (with 90% uncertainty) from all eleven clusters (results are shown for CCN measurements also for 30% uncertainty). The solid black symbols show the constraint using all measurements (CCN, N3, and BC, 90% uncertainty) simultaneously from all of the clusters.
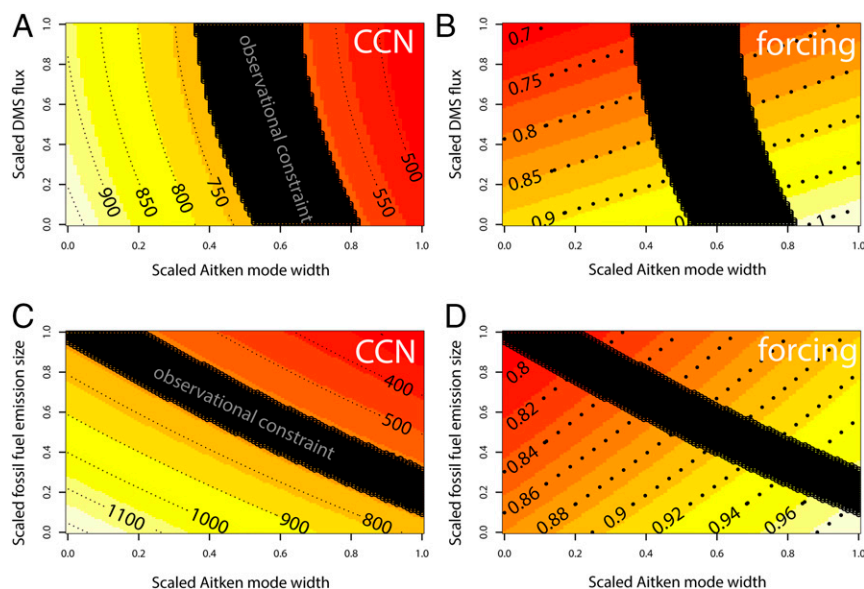
Fig. 5. Dependence of CCN and forcing on two of the 28 parameters, showing why constraining CCN does not necessarily constrain forcing in a two-parameter setting. Results are shown for the representative model grid box within the anthropogenic source cluster in Fig. S1. (*A*) CCN is only sensitive to a single parameter, but (*B*) forcing is sensitive to both. CCN measurements constrain the Aitken mode width almost completely, but DMS emissions are unconstrained. Both CCN (*C*) and forcing (*D*) are sensitive to both parameters, but the dependencies are different.

constrain the aerosol state, a commensurate constraint on forcing will be applied only when the dependence of the aerosol state on the parameters is the same as the dependence of forcing on the same parameters. In Fig. 5, this means that the CCN and forcing contour lines would need to be parallel to each other, and the more they deviate from parallel, the less the forcing will be constrained by making CCN measurements. According to our results, such similarity of parameter sensitivities appears to be uncommon, even when sampling across all global aerosol environments; otherwise, the CCN measurements would have constrained more of the forcing uncertainty (Fig. 4). Any mismatch of the parameter dependencies results in unconstrained forcing parameter space. There are many examples in the atmosphere where such compensation of parameter effects occurs, such as erroneously high emissions being compensated by high aerosol removal rates. There are also likely to be many such cancellation effects in other components of climate models.

Multiple Measurement Constraints. It might be argued that we cannot expect the observational constraint of one aerosol variable to fully constrain forcing. Nevertheless, as we argued in the Introduction, because uncertainties in CCN and forcing are caused by exactly the same 28 parameters, we might expect global measurements to eventually reduce forcing uncertainty. Equifinality means that this does not occur. We now explore the effect of additional constraints on the same 28 parameters.

The constraint provided by other state variable measurements will depend on how the state variable depends on the uncertain parameters. Surprisingly, measurements of the total particle concentration (N_3) with an uncertainty threshold of 90% constrain forcing more than CCN measurements (Fig. 4) in July, even though most of these particles do not form cloud droplets. This result occurs because, in many locations, N_3 depends on more of the natural parameters that are important for the forcing uncertainty. Measurements of N_3 also provide additional constraint for CCN beyond that provided by making just CCN measurements, even when the measurements are made at the same location. The additional constraint occurs because the N_3 parameter relationships are similar to those for CCN at locations where we have not previously taken CCN measurements.

BC mass concentration measurements with an uncertainty threshold of 90% barely constrain forcing at all. Although BC as a state variable is essentially unrelated to cloud drop concentrations and forcing, its uncertainty is controlled by many of the

same uncertain parameters. For example, carbonaceous particle emission fluxes and dry deposition rates both affect CCN and BC, so constraint of BC could help to constrain forcing. The very weak constraint has not been investigated further but is likely caused by a high degree of equifinality.

Multiple measured variables can be used together. However, measurements of CCN, N_3 and BC in all 11 clusters constrain forcing only about one-third to one-half as much as they constrain CCN. Again, the ability of these measurements to further constrain forcing is limited by equifinality, so the posterior uncertainty will not converge to zero unless the parameter relationships are the same as for forcing.

Conclusions

Although it is well recognized that there are compensating processes in environmental models (21, 25, 26), the effect on our
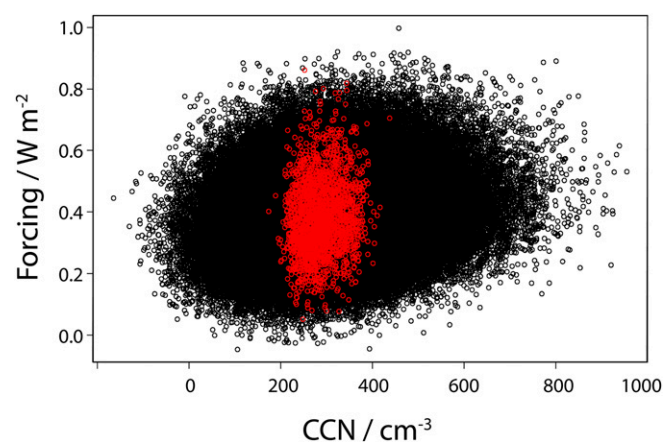

Fig. 6. Scatter plot of the modeled forcing against the CCN in the North Atlantic. Each black dot represents a model variant with a specific setting of each of the 28 perturbed parameters. The red dots are the models remaining after constraint of CCN with measurements from 10 grid boxes from the Northern Hemisphere marine region (Fig. 3 shows histograms of the same data). The scatter shows some dependence of forcing on CCN when only the unconstrained parameter set is considered. The apparent relationship does not exist in the constrained parameter set, as the equifinal CCN values lead to a large range of forcing values.

ability to constrain unobservable model quantities through state variable measurements has not previously been quantified. We posed the question: If we could make aerosol measurements with known uncertainty in every grid box of the model, how well would we be able to constrain the uncertainty in forcing caused by these aerosol uncertainties? We first showed that it is possible to define regions of the world according to the "information content" of measurements for CCN constraint. This could be valuable information for defining observational strategies aimed at reducing model uncertainty—a few representative measurements are as effective as widely distributed measurements. We then used CCN measurements from these representative locations to constrain the 28-dimensional parametric uncertainty space and calculated the resulting constraint on modeled aerosol−cloud forcing. We found that, although globally CCN could be constrained well by measurements made in representative environments, the 28-dimensional parametric uncertainty space was not constrained in such a way that resulted in a similarly constrained forcing uncertainty.

The reasons for the limited forcing constraint relate back to our original hypotheses. We first hypothesized that the effectiveness of aerosol measurements to constrain uncertainty in forcing would be reduced if the two quantities depended on different processes. However, we eliminated this possibility by ensuring that the same 28 parameters in a single global aerosol model were the only sources of uncertainty in both aerosol and forcing. We further hypothesized that if aerosols and forcing had different sensitivities to the parameters, then this would reduce the effectiveness of aerosol measurements for constraining forcing. Nevertheless, we hoped to find enough environments globally to constrain most of the 28 uncertain dimensions and therefore eventually constrain forcing uncertainty. In fact, we did identify regions in which the uncertainties in aerosol and forcing were controlled by similar parameters, but we found that, even in these environments, the constraint on the forcing was limited. The cause of the weak constraint is related to the fact that uncertain parameters have compensating effects in the model, which leads to many ways in which combinations of different parameters lead to plausible aerosol models compared with measurements. This is termed equifinality. Our results show that equifinality in the global aerosol system significantly limits our ability to constrain the uncertainty in forcing caused by uncertainty in the aerosol model, even when using apparently closely related aerosol measurements that have the same sources of uncertainty. The result is that these equifinal aerosol models predict a narrow range of aerosols but a wide range of forcings.

There are several important implications of equifinality in the aerosol system. First, because forcing must be predicted and cannot be measured directly, and because the impact of equifinality can only be estimated by dense sampling of model parameter space, it is likely to be a hidden source of uncertainty in most model studies. A model that is tuned to match an observed aerosol state will predict a particular forcing, but there will exist a range of equally plausible forcings that will not be detected unless the model's full parameter space is sampled. Therefore, obtaining an accurate model for aerosol state variables is a necessary but not sufficient condition for obtaining an accurate model of radiative forcing, even if we account only for uncertainties in the aerosol component of the model. Second, several models, each achieving very good agreement with aerosol measurements, could calculate very different forcings just as we have found in a single model. Current models do not achieve universal good agreement with aerosol measurements (14), which suggests we are a long way from reducing the persistent diversity in aerosol−cloud forcing (1).

The results shown here also have implications for the use of emergent constraints to constrain climate models if parametric uncertainty is not accounted for. Emergent constraints are used when a relationship between two model outputs emerges from collections of climate model simulations, which allows the observational constraint of one output to be used to constrain the other, of relevance to long-term climate. In our model, a relationship between CCN and cloud albedo forcing emerges from multiple model simulations (Fig. 6), as expected. These results are based on a sample from the forcing and CCN emulators of 3 million parameter combinations (based on the same data as Fig. 3). Although a clear positive correlation of forcing with CCN emerges from these model runs, there is very large spread due to the 28-dimensional parametric uncertainty. Clearly, even high-precision CCN measurements will have a relatively small effect on forcing uncertainty when the emergent CCN−forcing relationship is so uncertain. Proper account needs to be taken of equifinality before robust conclusions can be drawn about the usefulness of emergent behavior among multiple models.

The difficulty of aerosol model constraint is related to model complexity. In earlier climate models, only the aerosol mass concentrations were simulated, and cloud drop concentrations were predicted in terms of empirical relationships (27). Although observational constraint of such models would still suffer from equifinality (e.g., compensating emission and removal rates), there were fewer processes to constrain. In an aerosol microphysics model, there are tens of important parameters (6, 15) and multidimensional equifinality. Therefore, although model fidelity may have improved, the uncertainty may now be more difficult to constrain.

Our results are based on the assumption that the parametric uncertainties apply globally. In some places, this will lead to an overestimate of the uncertainty range given to some parameters locally. We have also applied the history matching technique globally so that a constraint on the parameter set from CCN anywhere applies globally, which is likely to lead to an overestimate of the constraint from CCN measurements. We believe that, together with the fact that we have applied tight constraints, we are still likely to be overestimating the constraint on CCN and therefore forcing.

Additional steps could be taken to reduce uncertainty further. First, measurements could be used to constrain relationships between modeled outputs (such as cloud albedo versus aerosol load), which has been shown to be a valuable approach (9). Given what we have learned about compensating model effects (equifinality), we suggest that the effectiveness of aerosol−cloud relations for constraining forcing should be assessed in a way that accounts for model uncertainty. Second, given the problem of compensating processes, it may be more effective to design atmospheric experiments in ways that enable single processes to be isolated. To do this, experiments will need to be guided by models that can point to where the effects of single processes can be isolated, for example, by examining the results of comprehensive sensitivity analyses (e.g., ref. 15). Laboratory measurements can quantify individual processes, but we need to ensure that we understand the uncertainties introduced by parameterizing the processes in global models. Third, given the uncertainty due to the PI aerosol state, we could consider quantifying forcings over recent decades where we have observational constraints. Over more recent periods, we might also expect changes in forcing to scale with aerosol emissions and concentrations (28), thereby reducing some of the uncertainty compensation issues highlighted in this paper. Fourth, and most importantly, we need to treat uncertainty as one of the main scientific problems to be understood and solved. Neglect of important issues like equifinality may render any apparent improvements in model performance illusory.

There are many reasons why the large PI to PD aerosol−cloud forcing uncertainty has persisted through all IPCC reports. Observational constraints on models are essential for reducing the uncertainty. Our results suggest that we need to carefully evaluate how the available observations constrain models, and how this feeds through to forcing uncertainty, which cannot be directly constrained by measurements. Here we have tackled only the uncertainty in the aerosol model. Other cloud properties of importance to Earth's radiative balance not included here, such as cloud cover and thickness, also respond to changes in aerosols

(7) through radiative and dynamical couplings. Some of the uncertainties in these cloud properties are likely to be correlated with the uncertainty in albedo studied here, but there will be additional sources of uncertainty related to cloud and atmospheric physics responses. By neglecting these interactions, we have presented a best-case scenario of how improvements in aerosol models will result in improved simulations of forcing. We can expect equifinality to be a significant issue in models of the fully coupled aerosol−cloud−atmosphere system just as it is in the aerosol−albedo relation. However, the relationship between the uncertainty in the aerosol model and the uncertainty in the coupled climate model response will be even more challenging to demonstrate.

1. Boucher O, et al. (2013) Clouds and aerosols. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, eds Stocker TF, et al. (Cambridge Univ Press, Cambridge, UK), pp 571–658.
2. Andreae MO, Jones CD, Cox PM (2005) Strong present-day aerosol cooling implies a hot future. *Nature* 435(7046):1187–1190.
3. Otto A, et al. (2013) Energy budget constraints on climate response. *Nat Geosci* 6(6):415–416.
4. Myhre G, Myhre CEL, Samset BH, Storelvmo T (2013) Aerosols and their relation to global climate and climate sensitivity. *Nat Educ Knowl* 4(5):7.
5. Rosenfeld D, Sherwood S, Wood R, Donner L (2014) Climate effects of aerosol–cloud interactions. *Science* 343(6169):379–380.
6. Carslaw KS, et al. (2013) Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature* 503(7474):67–71.
7. Koren I, Dagan G, Altaratz O (2014) From aerosol-limited to invigoration of warm convective clouds. *Science* 344(6188):1143–1146.
8. Wilcox LJ, Highwood EJ, Booth BBB, Carslaw KS (2015) Quantifying sources of intermodel diversity in the cloud albedo effect. *Geophys Res Lett* 42(5):1568–1575.
9. Quaas J, Boucher O, Bellouin N, Kinne S (2008) Satellite-based estimate of the direct and indirect aerosol climate forcing. *J Geophys Res* 113(D5):D05204.
10. Ma X, Yu F, Quaas J (2014) Reassessment of satellite-based estimate of aerosol climate forcing. *J Geophys Res* 119(17):10,394–10,409.
11. Bellouin N, Boucher O, Haywood J, Reddy MS (2005) Global estimate of aerosol direct radiative forcing from satellite measurements. *Nature* 438(7071):1138–1141.
12. Hamilton DS, et al. (2014) Occurrence of pristine aerosol environments on a polluted planet. *Proc Natl Acad Sci USA* 111(52):18466–18471.
13. Penner JE, Xu L, Wang M (2011) Satellite methods underestimate indirect climate forcing by aerosols. *Proc Natl Acad Sci USA* 108(33):13404–13408.
14. Mann GW, et al. (2014) Intercomparison and evaluation of global aerosol microphysical properties among AeroCom models of a range of complexity. *Atmos Chem Phys* 14(9):4679–4713.
15. Lee LA, et al. (2013) The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei. *Atmos Chem Phys* 13(3):8879–8914.
16. Spracklen DV, Rap A (2013) Natural aerosol–climate feedbacks suppressed by anthropogenic aerosol. *Geophys Res Lett* 40(19):5316–5319.
17. Beven K (2006) A manifesto for the equifinality thesis. *J Hydrol* 320(1-2):18–36.
18. O'Hagan A (2006) Bayesian analysis of computer code outputs: A tutorial. *Reliab Eng Syst Saf* (91(10-11):1290–1300.
19. Lee LA, Carslaw KS, Pringle KJ, Mann GW, Spracklen DV (2011) Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters. *Atmos Chem Phys* 11(23):12253–12273.
20. Craig PS, Goldstein M, Seheult AH, Smith JA (1996) Bayes linear strategies for matching hydrocarbon reservoir history. *Bayesian Statistics 5*, eds Bernado JM, Berger JO, Dawid AP, Smith AFM (Oxford Univ Press, Oxford), pp 69–95.
21. Edwards NR, Cameron D, Rougier J (2011) Precalibrating an intermediate complexity climate model. *Clim Dyn* 37(7-8):1469–1482.
22. Williamson D, et al. (2013) History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Clim Dyn* 41(7-8):1703–1729.
23. Beven K, Binley A (1992) The future of distributed models: Model calibration and uncertainty prediction. *Hydrol Processes* 6(3):279–298.
24. Hartigan JA, Wong MA (1979) Algorithm AS 136: A K-means clustering algorithm. *Appl Stat* 28(1):100–108.
25. Posselt DJ, Vukicevic T (2010) Robust characterization of model physics uncertainty for simulations of deep moist convection. *Mon Weather Rev* 138(5):1513–1535.
26. Beven K, Binley A (2014) GLUE: 20 years on. *Hydrol Process* 28(24):5897–5918.
27. Jones A, Roberts DL, Slingo A (1994) A climate model study of indirect radiative forcing by anthropogenic sulphate aerosols. *Nature* 370:450–453.
28. Regayre LA, et al. (2014) Uncertainty in the magnitude of aerosol-cloud radiative forcing over recent decades. *Geophys Res Lett* 41(24):9040–9049.
29. Spracklen DV, et al. (2005) A global off-line model of size-resolved aerosol microphysics. I. Model development and prediction of aerosol properties. *Atmos Chem Phys* 5(8):2227–2252.
30. Mann GW, et al. (2010) Description and evaluation of GLOMAP-mode: A modal global aerosol microphysics model for the UKCA composition-climate model. *Geosci Model Dev* 3(2):519–551.
31. Stockwell D, Chipperfield M (1999) A tropospheric chemical transport model: Development and validation of the model transport schemes. *Q J R Meteorol Soc* 125(557):1747–1783.
32. Chipperfield MP (2006) New version of the TOMCAT/SLIMCAT off-line chemical transport model: Intercomparison of stratospheric tracer experiments. *Q J R Meteorol Soc* 132(617):1179–1203.
33. Rossow WB, Schiffer RA (1999) Advances in understanding clouds from ISCCP. *Bull Am Meteorol Soc* 80(11):2261–2287.
34. Spracklen DV, et al. (2010) Explaining global surface aerosol number concentrations in terms of primary emissions and particle formation. *Atmos Chem Phys* 10(10):4775–4793.
35. Reddington CL, et al. (2011) Primary versus secondary contributions to particle number concentrations in the European boundary layer. *Atmos Chem Phys* 11(23):12007–12036.
36. Korhonen H, et al. (2008) Influence of oceanic dimethyl sulfide emissions on cloud condensation nuclei concentrations and seasonality over the remote Southern Hemisphere oceans: A global model study. *J Geophys Res* 113(D15):D15204.
37. Spracklen DV, Carslaw KS, Poschl U, Rap A, Forster PM (2011) Global cloud condensation nuclei influenced by carbonaceous combustion aerosol. *Atmos Chem Phys* 11(17):9067–9087.
38. Spracklen DV, et al. (2011) Aerosol mass spectrometer constraint on the global secondary organic aerosol budget. *Atmos Chem Phys* 11(23):12109–12136.
39. Browse J, Carslaw KS, Arnold SR, Pringle K, Boucher O (2012) The scavenging processes controlling the seasonal cycle in Arctic sulphate and black carbon aerosol. *Atmos Chem Phys* 12(15):6775–6798.
40. Schmidt A, et al. (2011) Excess mortality in Europe following a future Laki-style Icelandic eruption. *Proc Natl Acad Sci USA* 108(38):15710–15715.
41. Pringle KJ, et al. (2012) A multi-model assessment of the impact of sea spray geoengineering on cloud droplet number. *Atmos Chem Phys* 12(23):11647–11663.
42. Fountoukis C, Nenes A (2005) Continued development of a cloud droplet formation parameterization for global climate models. *J Geophys Res* 110(D11):D11212.
43. Edwards JM, Slingo A (1996) Studies with a flexible new radiation code. I. Choosing a configuration for a large-scale model. *Q J R Meteorol Soc* 122(513):689–719.
44. Rap A, et al. (2013) Natural aerosol direct and indirect radiative effects. *Geophys Res Lett* 40(12):3297–3301.
45. Dentener F, et al. (2006) Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom. *Atmos Chem Phys* 6(12):4321–4344.
46. Cofala J, Amann M, Klimont Z, Schopp W (2005) Scenarios of World Anthropogenic Emissions of SO2, NOx and CO up to 2030. Internal report of the Transboundary Air Pollution Programme (International Institute for Applied Systems Analysis, Laxenburg).
47. Bond TC, et al. (2004) A technology-based global inventory of black and organic carbon emissions from combustion. *J Geophys Res* 109:D14203.
48. van der Werf GR, Randerson JT, Collatz GJ, Giglio L (2003) Carbon emissions from fires in tropical and subtropical ecosystems. *Glob Change Biol* 9(4):547–562.
49. Andres RJ, Kasgnoc AD (1998) A time-averaged inventory of subaerial volcanic sulfur emissions. *J Geophys Res* 103(D19):25251–25262.
50. Kettle AJ, Andreae MO (2000) Flux of dimethylsulfide from the oceans: A comparison of updated data sets and flux models. *J Geophys Res* 105(D22):26793–26808.
51. Gong S (2003) A parameterization of sea-salt aerosol source function for sub and super-micron particles. *Glob Biogeochem Cycles* 17(4):1097.
52. Guenther A, et al. (1995) A global model of natural volatile organic compound emissions. *J Geophys Res* 100(D5):8873–8892.