**Atmospheric
Chemistry
and Physics**

Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters

L. A. Lee, K. S. Carslaw, K. J. Pringle, G. W. Mann, and D. V. Spracklen

Institute for Climate and Atmospheric Science, University of Leeds, UK

Abstract. Sensitivity analysis of atmospheric models is necessary to identify the processes that lead to uncertainty in model predictions, to help understand model diversity through comparison of driving processes, and to prioritise research. Assessing the effect of parameter uncertainty in complex models is challenging and often limited by CPU constraints. Here we present a cost-effective application of variance-based sensitivity analysis to quantify the sensitivity of a 3-D global aerosol model to uncertain parameters. A Gaussian process emulator is used to estimate the model output across multi-dimensional parameter space, using information from a small number of model runs at points chosen using a Latin hypercube space-filling design. Gaussian process emulation is a Bayesian approach that uses information from the model runs along with some prior assumptions about the model behaviour to predict model output everywhere in the uncertainty space. We use the Gaussian process emulator to calculate the percentage of expected output variance explained by uncertainty in global aerosol model parameters and their interactions. To demonstrate the technique, we show examples of cloud condensation nuclei (CCN) sensitivity to 8 model parameters in polluted and remote marine environments as a function of altitude. In the polluted environment 95 % of the variance of CCN concentration is described by uncertainty in the 8 parameters (excluding their interaction effects) and is dominated by the uncertainty in the sulphur emissions, which explains 80 % of the variance. However, in the remote region parameter interaction effects become important, accounting for up to 40 % of the total variance. Some parameters are shown to have a negligible individual effect but a substantial interac-

tion effect. Such sensitivities would not be detected in the commonly used single parameter perturbation experiments, which would therefore underpredict total uncertainty. Gaussian process emulation is shown to be an efficient and useful technique for quantifying parameter sensitivity in complex global atmospheric models.

1 Introduction

Aerosols have an important but very uncertain impact on climate (Forster et al., 2007). The uncertainty has many sources, but inter-model differences, as well as uncertainties and limitations in the driving aerosol processes, are key factors. Until recently, climate models used simple representations of aerosol, which were based mostly on just particle mass. But the recognition that simplification of physical processes limits model predictive capability has led to the development of more complex "second generation" aerosol microphysics schemes that are intended to enhance model realism and improve the reliability of predictions (Binkowski and Shankar, 1995; Jacobson, 1997; Whitby and McMurry, 1997; Ackermann et al., 1998; Ghan et al., 2001; Adams and Seinfeld, 2002; Lauer et al., 2005; Liu et al., 2005; Stier et al., 2005; Spracklen et al., 2005a; Debry et al., 2007; Spracklen et al., 2008). Model realism has undoubtedly improved, but the diversity in model aerosol radiative forcing estimates has remained high in successive IPCC assessments (Schimel et al., 1996; Penner et al., 2001; Forster et al., 2007).

There are three reasons why an understanding of model sensitivity to uncertain inputs is important. Firstly, we need to attribute the uncertainty in model predictions to various processes and the poorly constrained model parameters that describe these processes. At present, most of our

understanding about model uncertainty derives from the diversity of predictions of several structurally different models (Textor et al., 2006, 2007; Meehl et al., 2007). The computational demands of complex global atmospheric models mean that the sources of uncertainty at the process level have not been rigorously quantified. Secondly, the sensitivity analysis of individual models will help to quantify how much diversity is due to different parameter settings and how much is due to differences in model formulation and complexity. The third reason is that sensitivity analysis can guide future model development. It is recognised that aerosol models include only a fraction of the important processes, and that the models need to develop further in the future. We therefore need procedures to assess the necessary level of model complexity objectively. At present, our limited understanding of the main sources of model uncertainty means that it is difficult to justify an increase in aerosol model complexity in favour of deploying computational resources to better effect in other parts of climate models.

The most commonly used sensitivity analysis approach used in complex global atmospheric models is single parameter perturbation or one-at-a-time (OAT) tests. The OAT test quantifies the departure of the model output from some baseline case to a perturbation in a single model input. We have used that approach previously in our own global aerosol model (Spracklen et al., 2005b). The OAT approach is appealing because it always calculates the change in the model away from a well known baseline calculated using "default" parameters. However, there are two significant disadvantages of OAT tests: firstly, the fraction of parameter space sampled quickly tends to zero as the number of model inputs increases (Saltelli and Annonia, 2010). Secondly, the approach ignores interactions between parameters (for example, whether sensitivity to aerosol nucleation varies as emissions change); essentially all sensitivity information is calculated at one point in parameter space. For these reasons, it is well recognised in policy applications that the OAT approach is inadequate (Gaber et al., 2009).

Other methods of sensitivity analysis have been developed that cover the space of the uncertain parameters and their interactions. For example, factorial analysis (Fisher, 1926) uses a more effective experimental design than OAT because it is based on setting the different parameters (or factors) to several values and testing all possible combinations of the different parameter values. However, the number of experiments required grows rapidly with the number of parameters examined; for example, when testing only the highest and lowest plausible value for each of k parameters there will be 2^k experiments necessary. Factorial designs provide information about parameter interactions, but the number of experiments quickly becomes prohibitive for complex atmospheric models.

The Met Office Hadley Centre quantifying uncertainty in model predictions (QUMP) project has resulted in several sensitivity studies undertaken using climate models attempt-

ing to improve on the OAT approach (Murphy et al., 2004). Sensitivity analysis experiments with a single model are often referred to as perturbed physics ensembles (PPEs) and a good review of those carried out in the Hadley Centre can be found in Collins et al. (2010). Another important PPE study called climateprediction.net used the home computers of many users to repeatedly run the Hadley Centre HadCM3 climate model with different parameter settings. The climateprediction.net ensemble was used in Ackerley et al. (2009) to study the climate responses to changes in atmospheric aerosol, albeit with a simpler aerosol scheme than we use here. In Sanderson et al. (2008) an emulator was used together with the many climateprediction.net runs to carry out sensitivity analysis. The number of ensembles produced by climateprediction.net is seldom possible in practice. PPEs have been carried out with other climate models including Niehörster et al. (2006) and Annan et al. (2005). Yokohata et al. (2010) compared two different climate models using the information from the PPE studies on each model. Rougier et al. (2009) discussed the idea of emulating the climate model so that every point in the output space is estimated in order to carry out an uncertainty analysis where the uncertainty in the model output due to the uncertain inputs is quantified.

The first uncertainty analysis of the aerosol indirect effect was carried out by Pan et al. (1997). They used the probabilistic collocation method to produce an approximation to their computer model in order to make uncertainty analysis feasible. Liu et al. (2007) isolated the uncertainty in global aerosol models due to meteorology by running the same model with different meteorological datasets. More recently, Haerter et al. (2009) studied the parametric uncertainty in aerosol indirect radiative forcing based on 7 cloud-related parameters with the ECHAM5 model using both OAT tests and multi-parameter perturbation tests. A Latin hypercube design was used to define multiple parameter perturbation experiments which are compared to single perturbation experiments to identify the interaction effects. Vignati et al. (2010) used two models to assess parameter uncertainty. They compared a simple bulk model and a more detailed chemistry transport model to look at the effect of the wet deposition parameters on black carbon. Lohmann and Ferrachat (2010) examined the parametric uncertainty effects on the climate by systematically varying 4 cloud parameters at specified values following a factorial design with 168 model runs.

Here we introduce the use of variance-based sensitivity analysis (Saltelli et al., 2000) to understand the sensitivity of a global aerosol model at the process level. The aim of the sensitivity analysis is to quantify the relative contribution of different model parameters and their interactions to the overall uncertainty in the model prediction. Two measures of sensitivity are computed for each model input (Saltelli et al., 2000): the "main effect" measures the reduction in the output variance when the model input can be learnt exactly, and the "total effect" measures the remaining variance

in the model output when everything except the input under investigation is learnt. The total effect sensitivity compared to the main effect sensitivity gives an indication of how each input interacts with others, which can then be further investigated. Variance-based methods require complete specification of the model output throughout the space of the parameter uncertainty. In many applications (Saltelli et al., 2000) these outputs are generated in a Monte Carlo simulation using a very large number (usually many thousands) of model runs. Here we use Gaussian process emulation, which generates the same level of information required by variance-based sensitivity analysis but requires considerably fewer model runs than Monte Carlo (see Sect. 2).

The aim of this paper is to demonstrate the potential of the emulation approach applied to a complex global aerosol model. We use the Global Model of Aerosol Processes, GLOMAP (Spracklen et al., 2005a; Mann et al., 2010) and follow a previous sensitivity study using the OAT technique (Spracklen et al., 2005b). The model predicts a wide range of aerosol properties relevant to climate and air quality. Here we focus on cloud condensation nuclei (CCN), which is the subset of aerosol particles that can form cloud drops. The concentration of CCN is a key quantity in the prediction of the very uncertain aerosol indirect effect. It is also a quantity where an understanding of model uncertainty will greatly benefit the analysis of newly compiled global datasets (Spracklen et al., 2011).

This paper is set out as follows. In Sect. 2 emulation is introduced and compared with other approaches. In Sect. 3 we describe the global aerosol model and specify the uncertain parameters. In Sect. 4 the application of the sensitivity analysis on the global aerosol model using emulation is presented.

2 Emulation of the global aerosol model GLOMAP

The basic procedure for an emulation study is shown in Fig. 1. No screening or formal elicitation is carried out as part of the initial study.

2.1 Why is emulation necessary?

Emulation is the process by which the computer model is replaced by a statistical surrogate model that can be run more efficiently. The global aerosol model used here is a complex computer code so it is practically impossible to explore the entire parameter uncertainty space. Haerter et al. (2009) and Lohmann and Ferrachat (2010) study various combinations of parameter values but the amount of information generated is not sufficient for a full variance-based analysis. When a simple computer model with very short run time is available emulation is redundant since the actual computer model can be used to provide output throughout the parameter uncertainty space; this is a Monte Carlo simulation.
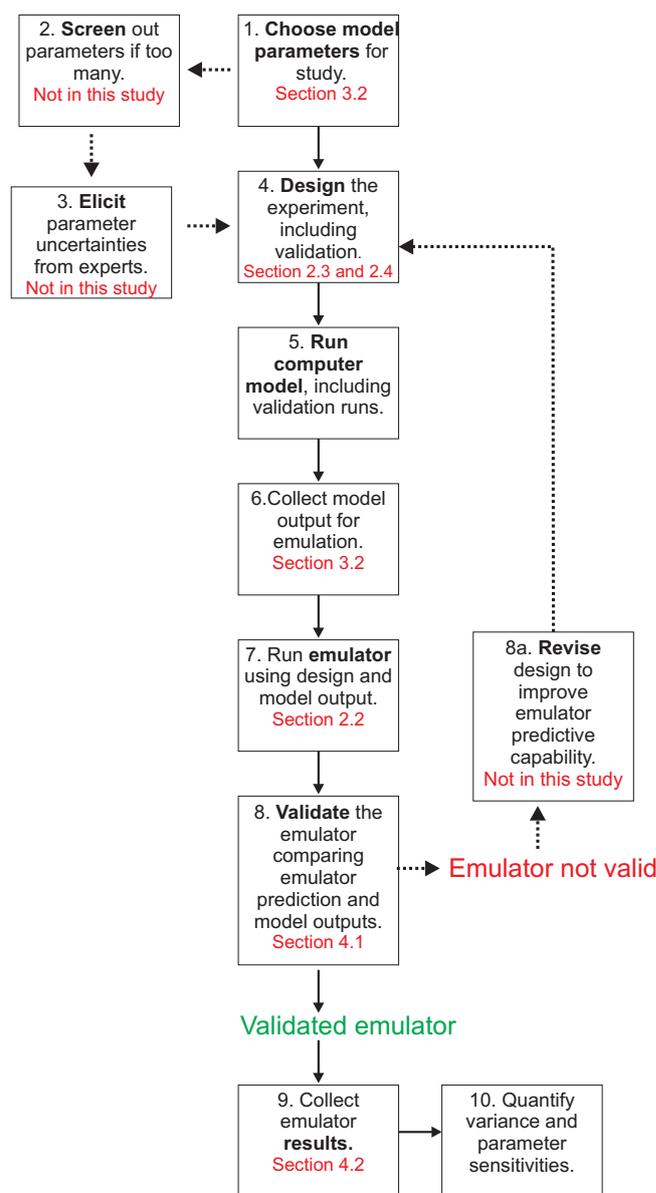
Fig. 1. The basic procedure to follow in an emulation study.

O'Hagan (2006) compares Monte Carlo and emulation techniques in the sensitivity analysis of computer models. A comprehensive variance-based sensitivity analysis may require millions of model runs, and even for a model that takes just one second to run just one million runs takes 11.5 days of continuous CPU time. With a complex computer code such as a global aerosol model a Monte Carlo simulation is not feasible. The aim of the emulator is to estimate the output of the model at a large number of untried parameter combinations so that variance-based sensitivity analysis (Saltelli et al., 2000) becomes feasible. In this work the Gaussian process is used for emulation (O'Hagan, 2006), but other emulation methods are available and have been applied
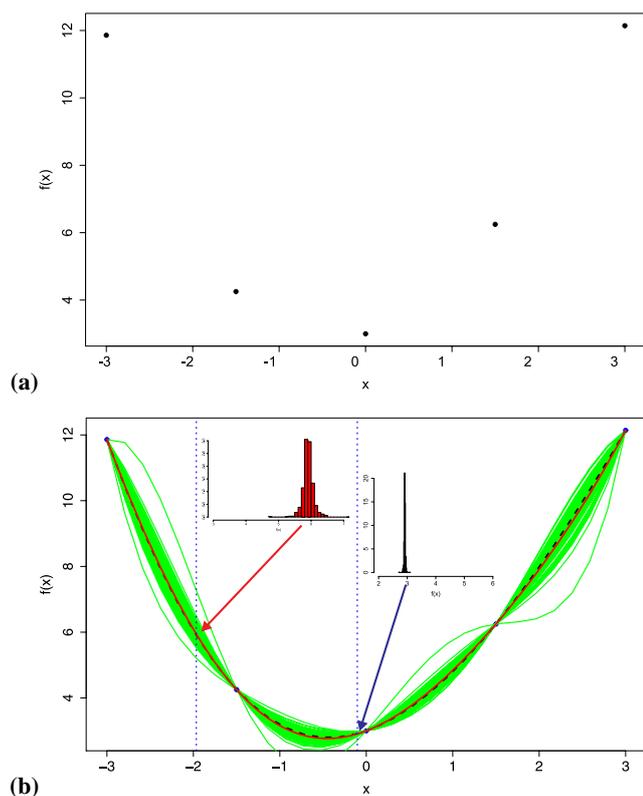
Fig. 2. (a) Five known points from some unknown function $f(x)$. These five points are used to train the emulator. **(b)** 600 realisations (green lines) from a one-dimensional emulator based on the five points in **(a)**. The mean of the emulator (dashed) is used to estimate the true curve (the solid red line) and the spread of the realisations gives visual indication of the emulator uncertainty. The histograms inset show the posterior Gaussian distribution at two points on the curve shown by the arrows, one point is near the training data so the Gaussian distribution is much narrower (the blue histogram).

to climate and ocean models (Sanderson et al., 2008; Goldstein and Rougier, 2006). The mathematics behind the Gaussian process emulator is explained in Appendix A1 and in Sect. 2.2 the emulator is described with a one-dimensional toy example followed in Sects. 2.3 and 2.4 by a discussion of the design of the model runs used to inform and validate the emulator.

2.2 The Gaussian process emulator

The application of the Gaussian process to understanding parametric uncertainty in global models is quite new, so we start by explaining the mechanism by which the emulator becomes an estimator for the global model. The Gaussian process is illustrated by a simple example of emulating with just one parameter, which can be viewed as a form of non-parametric curve fitting. Figure 2 shows such a function $f(x)$, for which we know the solution at five "training points". Here the five points have been drawn from the true

curve shown in red, but in reality the points would be generated through simulations of the complex model.

The Bayesian paradigm is used to combine prior beliefs about model behaviour with results from some model runs to produce a posterior distribution for the model which is then used to estimate the model output across the parameter space, and to quantify the uncertainty and carry out sensitivity analysis. In this work the prior is the Gaussian process. The Gaussian process is a statistical model that exploits the theory of conditional probability to estimate model output throughout the input uncertainty space using some known model output. Using conditional probability allows probabilistic statements to be made about model output and so uncertainty can be quantified. The Gaussian process represents a smooth function (here the global aerosol model output versus parameter value) such that each unknown output point has a normal distribution and any collection of outputs has a multivariate normal distribution. The functional form of the model is not assumed and so it is a non-parametric method. The posterior distribution is also a Gaussian process (the functional analogue of the conditional multivariate normal distribution) conditioned on the training data (model runs). The result is a posterior mean function and a posterior covariance function from which the model output can be estimated and all the sensitivity measures can be derived analytically. The uncertainty from using the emulator for sensitivity analysis rather than the simulator can also be derived analytically. The green curves in Fig. 2 show different estimates of the aerosol model based on the posterior Gaussian process, and the red dashed curve shows the mean function estimated by the emulator. The mean curve is used to carry out the parameter sensitivity analysis while the spread of green curves indicates the uncertainty in using the emulator rather than the simulator. Figure 3 shows the same curve but with five badly spaced training points. Here the uncertainty outside the range of the known points is so large that the mean cannot be used to estimate the true curve with any confidence. This example highlights the importance of using sufficient and well distributed training data.
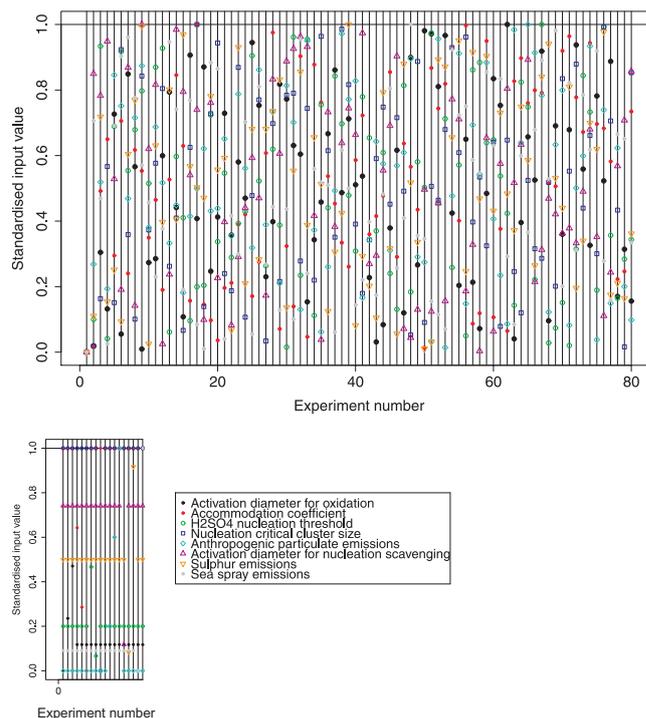
The Gaussian process has the desirable properties that the curve fits through the known points (each of the green lines passes exactly through all five training points) and a measure of uncertainty is calculated for every estimated point. In two-dimensions the Gaussian process would fit a surface with uncertainty calculated for each estimated point in both dimensions. The same is true in higher dimensions, so the Gaussian process can be used to build an emulator with any number of input variables given a suitable number of model runs. More mathematical descriptions of the Gaussian process emulator can be found in Appendix A1 and in Sacks et al. (1989), Currin et al. (1991), O'Hagan (1994), Neal (1999) and Santner et al. (2003). A discussion of different specifications of the prior beliefs can be found in Oakley (1999). A tutorial on Gaussian process emulation for non-mathematicians can be found in O'Hagan (2006).
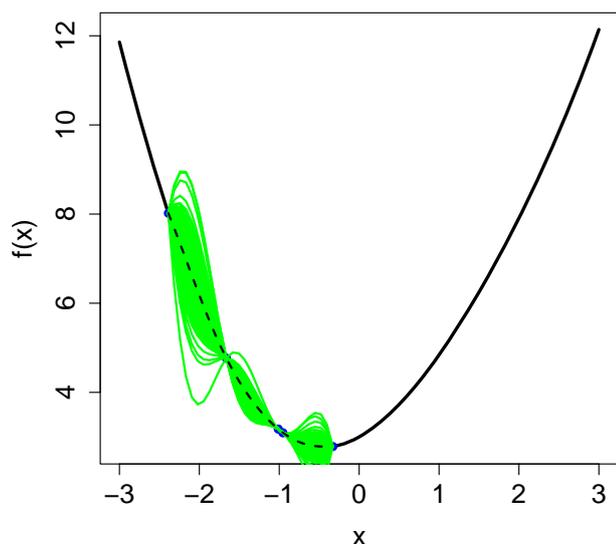
Fig. 3. 600 realisations from a one-dimensional emulator of the same curve in Fig. 2 but trained by five poorly spaced points. The uncertainty outside the training data is so great that the mean can not be considered representative of the true curve. Extrapolation should be avoided where possible.

Fig. 4. Upper: the design used here. Lower: the design used in Spracklen (2005).

The Gaussian process has been used to carry out uncertainty analysis (Haylock and O'Hagan, 1996; O'Hagan and Haylock, 1997; Oakley and O'Hagan, 2002) including methods for estimating the percentiles of the output uncertainty distribution. Oakley and O'Hagan (2004) extend their previous work to include sensitivity analysis in order to apportion the uncertainty in the output to the inputs and their interactions. The effect of the individual inputs and their interactions on the output is found by integrating the posterior multivariate mean with respect to various subsets of inputs and the expected variances are found similarly. The details of the integrations and the formulas involved in performing the sensitivity analysis can be found in Oakley and O'Hagan (2004). Morris et al. (2008) show a practical application of Gaussian process emulation for sensitivity analysis using a radiative transfer model.

Here we used readily available software, the Gaussian Emulation Machine for Sensitivity Analysis, http://ctcd.group. shef.ac.uk/gem.html. GEM-SA produces the main effect and total effect sensitivity measures for each input variable and the relationship between the model output and each of the uncertain parameters can be plotted. The spread of the lines in the plots produced compared to the range covered on the y-axis gives an indication of the emulator uncertainty compared to the effect of the parametric uncertainty. The first-order interaction sensitivity measures can be requested and their relationship with the model output plotted. Kennedy et al. (2008) use GEM-SA for sensitivity analysis of a dynamic vegetation model.

2.2.1 Important assumptions for the Gaussian process emulator for sensitivity analysis

There are two important assumptions relating to the use of the Gaussian process emulator for sensitivity analysis. These are:

The computer model is smooth and continuous with respect to its inputs The increased efficiency of the emulator over the computer model is based upon being able to use the information from a few runs to predict the output at untried points. This information comes from the output covariance between pairs of points and depends on the distance between the two points. When the output is smooth and continuous with respect to the inputs there is higher correlation between points, allowing a lower uncertainty in predictions far from the training points. If the computer model is not smooth then the increased efficiency is lost since too many runs would be required to build the emulator. The smoothness assumption is tested using validation data.

Separately identifiable emulator inputs The emulator inputs (the model parameters under investigation) should be separately identifiable. The identifiability of the inputs may not be known before the emulator is built but when there is some prior knowledge of an identifiability issue between parameters then only one or some function of them should be varied. Using separately identifiable inputs also keeps the necessary model runs to a minimum.

2.3 Designing to inform the emulator

To minimise the required number of model runs, it is important that the model runs generate information about as much of the uncertainty space as possible. The design of such experiments has been, and continues to be, a huge area of statistical research (Sacks et al., 1989; Bates et al., 1995).

Here we use a maximin Latin hypercube (McKay et al., 1979) to fill the uncertainty space of the parameters, which has been shown to be the best design for Gaussian process emulation (Jones and Johnson, 2009). The maximin Latin hypercube is based on the Latin square, a common example of which is the sudoku puzzle, which consists of nine overlying Latin squares; that is there is precisely one of each number in each row and column. The maximin Latin square maximises the minimum distance between points in the square in order to ensure optimum space filling. The maximin Latin hypercube has the same properties as the Latin square but in higher dimensions.

The number of points in the Latin hypercube used here is 10 times the number of parameters investigated, as recommended by Loeppky et al. (2009). The maximin Latin hypercube can be augmented with further points if diagnostics suggest there are not enough runs to build a suitably accurate emulator. In this analysis there are 8 uncertain parameters and therefore we configure 80 initial model runs with the parameter values based on Latin hypercube sampling in the same ranges as in Spracklen et al. (2005b). The space filling properties of the 80 runs used here are shown in Fig. 4 next to the OAT design used in Spracklen et al. (2005b).

2.4 Designing to validate the emulator

It cannot be guaranteed that the design used to build the emulator is sufficient to describe the model behaviour at the untried points. The emulator is therefore validated using further runs of the model. For the validation design we follow the recommendations of Bastos and O'Hagan (2009) and set the number of additional runs equal to three times the number of parameters studied. A third of these runs are deliberately close to some of the points in the original design and two-thirds are placed further away, chosen by a separate Latin hypercube design. Choosing specific runs to validate the emulator helps to identify specific failures with the statistical assumptions used to build the emulator. When the 95 % probability bound is constructed around the emulator prediction it should contain the GLOMAP prediction for 95 % of the validation points. In this experiment there are 24 validation runs, 8 of which have input settings close to those in the original 80 GLOMAP runs. The validation of the emulator is shown in Sect. 4.

3 The global aerosol model GLOMAP

3.1 Aerosol model description

The GLObal Model of Aerosol Processes (GLOMAP) (Spracklen et al., 2005a; Mann et al., 2010) is an aerosol microphysics module simulating the evolution of the size distribution and composition of a population of aerosol particles via processes such as new particle formation, coagulation, gas-to-particle transfer and cloud processing. The original version of the model (Spracklen et al., 2005a) uses a bin-resolved aerosol dynamics approach (GLOMAP-bin) but more recently a computationally cheaper version has been developed which uses modal aerosol dynamics called GLOMAP-mode (Mann et al., 2010). Both GLOMAP-bin and GLOMAP-mode are implemented within the TOMCAT global 3-D offline chemistry transport model (Stockwell and Chipperfield, 1996; Chipperfield, 2006) and GLOMAP-mode is also implemented within the HadGEM-UKCA composition-climate model (Morgenstern et al., 2009). We use GLOMAP-mode (from here on referred to as GLOMAP), which represents the aerosol by a particle number concentration and several component masses in a series of log-normal modes. These modes are split between two distributions (hydrophillic and hydrophobic) and four aerosol size categories (nucleation, Aitken, accummulation, coarse). The component mass and number concentrations of the log-normal modes are prognostic variables on the model grid, but the geometric standard deviation is fixed. The modal structure is similar to that used by Stier et al. (2005) and Pringle et al. (2010).

The model is run with the same setup as described in detail by Mann et al. (2010). It includes the treatment of sea spray, black carbon, organic carbon and dust and has been shown to compare well to ground based observations of aerosol mass and number (Mann et al., 2010; Spracklen et al., 2010). The model resolution is $2.81 \times 2.81°$ with 31 vertical levels. The outputs are requested monthly for all model levels across the globe and daily at the surface. An OAT parameter uncertainty study was carried out in GLOMAP-bin by Spracklen et al. (2005b). As well as the difference in aerosol dynamics, the modal GLOMAP version used in this present study differs from that used in Spracklen et al. (2005b) in two important ways: (i) Spracklen et al. (2005b) used a single-component version with only sulphate and sea-salt aerosol, and (ii) the models are separated by five years of model development (in particular emissions have been updated and boundary layer nucleation has been implemented (Spracklen et al., 2006; Merikanto et al., 2009)). Nevertheless, the model treatment of the core microphysical processes has remained similar. For detailed information on the GLOMAP-bin model used in Spracklen et al. (2005b) see Spracklen et al. (2005a).

At the resolution used here GLOMAP-mode takes about 5200 s to run per month on 32 cores on the HECTOR XT4 supercomputer and requires a spin-up period of at least 3

months. The perturbations are then applied with all model runs having identical initial conditions. After the perturbation, a further spin-up of 2 months is then carried out for each model run to ensure the perturbations take effect and the model output is assessed over the month after that. To carry out 80 model runs with which to train the emulator therefore takes 1 263 600 s on 32 cores, or nearly 15×32 core-days.

3.2 GLOMAP parameter uncertainties

3.2.1 Choice of model output and region

The GLOMAP model simulates the global distribution of many aerosol properties describing particle mass, number concentration, size distribution, chemical composition, etc. Each build of the emulator calculates the relationship between the parameter values and one of these many outputs in one grid cell or, with appropriate aggregation, over a larger domain of the model. Emulation of the output averaged over multiple grid points (not done here) needs to take into account whether the chosen output depends on the parameters in a similar way, which could be identified using cluster analysis, for example. Here, we focus on the sensitivity of simulated concentration of cloud condensation nuclei (CCN) at two representative locations; one polluted centred over London (UK, 51.0° N, 0.0° E) and one remote centred over the Pacific Ocean (4.2° S, 165.9° E). The emulation is carried out for June 2000 using monthly mean model output.

CCN are the subset of the aerosol particles that activate to form cloud droplets at a given supersaturation (here 0.2 %). They play a pivotal role in the interaction of aerosols and clouds. Prediction of CCN within a global model has only recently become possible with the development of microphysical models, and the processes and model parameterisations controlling their abundance and distribution remain uncertain.

3.2.2 Choice of model parameters

Spracklen et al. (2005b) examined the sensitivity of global mean condensation nuclei (CN) and CCN concentrations to 8 model parameters. The parameters include factors that scale the precursor emissions as well as microphysical process parameters. The parameters are briefly described below, and a full description of their handling in GLOMAP-mode is given in Mann et al. (2010). The chosen parameters may not be generic to all models and may also not represent the optimum selection, but our aim is to illustrate the method following a previous OAT approach. No formal elicitation is carried out as part of this study. Some uncertainty ranges here are different to those in Spracklen et al. (2005b) where the uncertainty is thought to be better understood compared to five years ago.

- X_1: *oxidation activation diameter* (**OX_DIAM**) In GLOMAP the aqueous phase oxidation activation diameter defines the diameter above which aerosol particles

activate into cloud droplets in stratiform clouds. Droplet formation is an important process in the global CCN budget because it enables SO_2 oxidation chemistry to grow the activated aerosol particles through addition of sulphate mass. The activation diameter varies greatly between clouds and regions depending on the particle size distribution, chemistry and updraught speed but is given a constant global value in these simulations. The sensitivity of CCN to all these processes could be investigated separately, but following Spracklen et al. (2005b) we quantify the sensitivity of CCN to the uncertainty in the aqueous phase oxidation activation diameter in the range $(0.04, 0.125)\,\mu m$.

- X_2: *mass accommodation coefficient* (**ACC_COEF**) In GLOMAP the mass accommodation coefficient defines the probability that a molecule of H_2SO_4 becomes bound to an aerosol particle upon collision. Changes in the accommodation coefficient affect particle growth rates as well as the amount of H_2SO_4 available for nucleation, which are important (but sometimes competing) processes in the production of CCN (Woodhouse et al., 2008). This is one of the interaction effects that may be highlighted through the sensitivity analysis. The uncertainty in the accommodation coefficient is set in the range $(0.02, 1.00)$.

- X_3: H_2SO_4 *nucleation threshold* (**NUC_THRESH**) In the Kulmala et al. (1998) mechanism the formation of new particles through binary nucleation occurs only when the atmospheric H_2SO_4 concentration is greater than a defined threshold value. Reducing the nucleation threshold causes more frequent nucleation and higher aerosol concentrations. The uncertainty in the H_2SO_4 nucleation threshold is set in the range $(0.25, 4) \times$ the baseline value.

- X_4: *nucleation critical cluster size* (**NUCRIT_SIZE**) The nucleation critical cluster size defines the smallest size above which a cluster of H_2SO_4 molecules is stable. Smaller critical cluster sizes take longer to grow and so are subject to coagulational scavenging for a longer period of time. The uncertainty in the nucleation critical cluster is set in the range $(50, 100)$ molecules.

- X_5: *particulate emissions associated with anthropogenic* SO_2 (**SO2_PART**) This parameter defines the fraction of the total sulphur emissions in a grid box emitted as particulate sulphate (rather than SO_2). The large grid scale of the model means particle formation in power plant plumes cannot be resolved, so a fraction of the anthropogenic SO_2 in global models is often emitted directly as particles (Adams and Seinfeld, 2003). The uncertainty in the particulate emissions of anthropogenic SO_2 is set in the range $(0, 5)\%$ of anthropogenic SO_2.

- X_6: *cloud nucleation scavenging diameter* (**SCAV_DIAM**) The nucleation scavenging diameter defines the diameter of aerosol particles above which they form cloud drops that subsequently undergo efficient collision-coalescence to produce raindrops, and are therefore wet scavenged. In reality SCAV_DIAM ought to be related to the aqueous activation diameter. However, global models do not currently resolve the in-cloud raindrop formation processes, so there are likely to be many additional uncertainties in the scavenging diameter than in the activation diameter alone. In GLOMAP we therefore treat the activation diameter and scavenging diameter as independent parameters with their own uncertainties. The uncertainty in the cloud scavenging activation diameter is set in the range $(0.08, 0.25)\,\mu m$.

- X_7: *sulphur emissions* (**SO2_EMS**) Emissions inventories used to drive global models are known to be highly uncertain, although uncertainties are smaller in areas with more information. The uncertainty in the AERO-COM emissions used here (Dentener et al., 2006) is believed to be no more that 30 %, so the uncertainty range here is (70, 130) % of the baseline emissions and is considered to be the same everywhere in this study.

  X_8: *sea spray emissions* (**SS_EMS**) GLOMAP uses the sea spray function of Gong (2003). Under clean marine conditions sea spray particles may dominate the accumulation mode and hence contribute significantly to CCN. Uncertainty in the sea spray emissions have been reviewed by de Leeuw et al. (2011). Here we set an uncertainty in the particle flux in the range $(0.1, 10)$ times the baseline value and apply it uniformly over the ocean.

The uncertainty distributions of the parameters are assumed to be uniform (giving equal weight to any point in a bounded range of uncertainty). The experimental design depends primarily on the range of uncertainty given to individual parameters rather than the specific uncertainty distribution, and the same models runs can be used to perform sensitivity analysis if the uncertainty distribution on any parameter is changed. The robustness of the statistical assumptions can therefore be tested by building emulators based on different input uncertainty distributions as shown in Rougier and Sexton (2007). In contrast, Monte Carlo methods require completely new experimental designs and thus further model runs when the distribution is changed. The range of the uncertainty given to the parameters however should remain the same to avoid extrapolation beyond the training data.


(a)

(b)
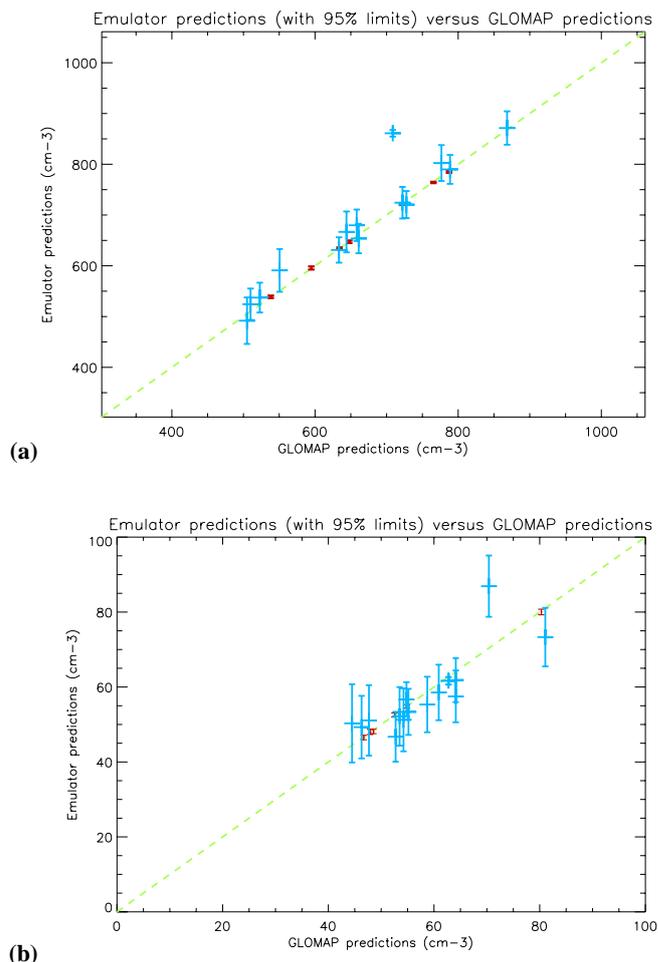
Fig. 5. GLOMAP estimates of June 2000 surface CCN concentrations in **(a)** London and **(b)** the Pacific Ocean versus emulator predictions of June 2000 surface CCN with 95 % probability (2σ) limits. The uncertainty here represents the emulator uncertainty rather than uncertainty due to the parameters. The red points are those from the experimental design that were purposely placed close to the original training data.

4 The application of Gaussian process emulation for sensitivity analysis of GLOMAP

4.1 Evaluation of the emulator

Figure 5 shows the evaluation of the emulator at the polluted and remote marine sites. With the emulator the variance (and hence uncertainty) due to emulation versus the simulator can be calculated, so it is possible to construct 95 % probability bounds for the emulator predictions, which are shown as the error bars in Fig. 5. The 95 % probability bounds around the validation points should cover the GLOMAP simulations for 95 % of the validation points.
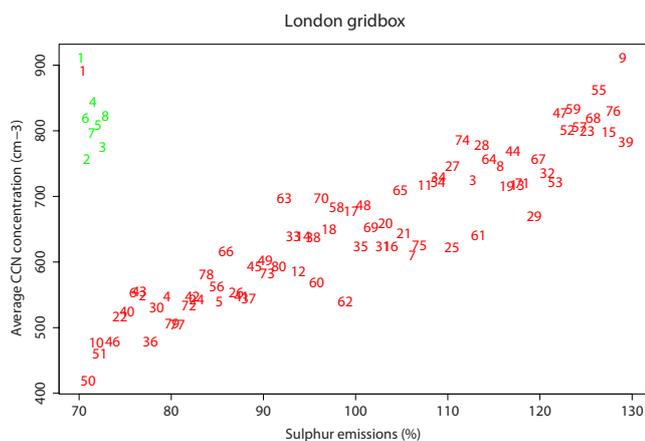
Fig. 6. June 2000 CCN concentrations for London versus the sulphur emissions parameter values in the 80 original GLOMAP runs (red) and the new GLOMAP runs where all parameters were set low (green). The numbers show the experiment number in each of the designs: original and low.

The 8 validation points placed close to the training data (shown in red) have small 95 % confidence intervals that cover the GLOMAP simulations showing that the emulator is estimating well close to the training data. With the exception of one, the other 16 emulated points have 95 % confidence intervals that cover the GLOMAP simulations showing the emulator is estimating well even at points far away from the training data.

Normally, one outlying point would not indicate that the emulator is invalid, but in Fig. 5a (the London grid box) the 95 % confidence interval is very small considering the distance of the point from the GLOMAP simulated value. We therefore investigated more closely the model predictions corresponding to this point. The outlying point in Fig. 5a is shown to have high CCN in the original GLOMAP simulation and it is necessary to evaluate the realism of this model prediction by comparison with observations. The outlying point corresponds to all the parameters set to their lowest values. The high CCN concentrations are surprising because a low value of some parameters (especially SO2_EMS) should favour low CCN. To explore both the model and the emulator behaviour when all parameters are set low a further 8 GLOMAP runs were performed with all parameter values in the bottom 5 % of the parameter range, defined using Latin hypercube sampling. Figure 6 shows the relationship between CCN and SO2_EMS in London from all 88 GLOMAP simulations. As expected CCN concentrations generally increase with SO2_EMS, but the additional 8 simulations behave differently. Figure 6 shows that the model is behaving oddly in this region of the parameter space, the emulator cannot capture this behaviour.
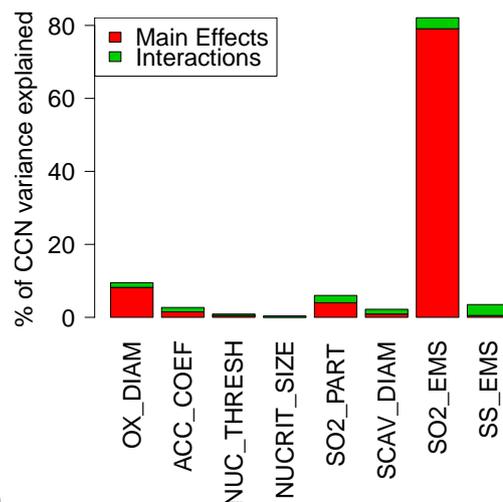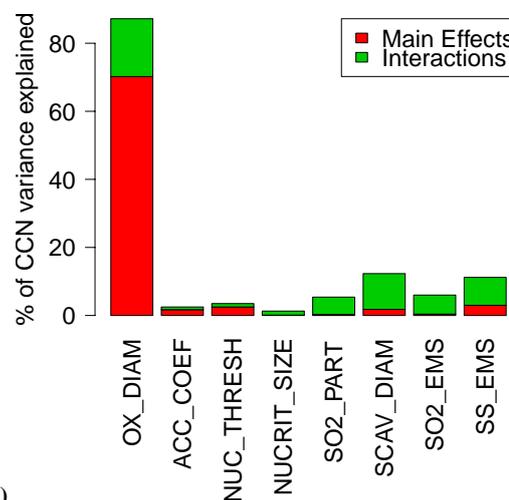
Fig. 7. A bar chart displaying the parameter sensitivities in **(a)** the polluted grid cell and **(b)** the remote grid cell. The red bars show the main effect sensitivities and the green bars show how much each parameter interacts with the others to contribute to the CCN variance.

There are two reasons to reject points from this tiny corner of parameter space. Firstly, the global aerosol fields show that total particle concentrations lie well outside observed ranges (Spracklen et al., 2010). Secondly, the behaviour of the aerosol system appears to be unphysical and not consistent with observed behaviour. The high CCN concentrations are created by extremely high number concentrations of nucleation mode aerosol, which grow mainly by coagulation to CCN sizes. Rapid nucleation throughout the atmosphere is sustained by a low vapour condensation sink (low particle surface areas) caused by efficient aerosol scavenging (low SCAV_DIAM) and a low nucleation threshold (low NUC_THRESH). In this environment, lower sulphur emissions act to exacerbate the low condensation sink more than they reduce the nucleation rate, so nucleation is enhanced further.

This trial emulation of a complex global aerosol model shows how odd behaviour can occur in very small fractions of the entire parameter space sampled by the space-filling design. In future applications, such behaviour needs to be detected by evaluating multiple diagnostics against observations and by studying the various microphysical budgets. It is possible to build the emulator with the original design and to remove implausible regions of the joint probability space (calibration) before carrying out the sensitivity analysis but this will be a topic of further research and will not be part of this paper. The robustness of the emulator results in this study and the leverage of the outlier in the corner of the parameter space are tested by building the emulator without training point one and comparing to the original emulator. The two emulators show similar results so in this study the point is included.

Figure 5b shows the emulator performance for the Pacific grid cell. It can be seen that one point in the validation design is not well predicted by the emulator. This point is not the same one that was highlighted in the London validation procedure since that has already been removed from the plot. The outlier in Fig. 5b is not an extreme outlier, there is nothing unusual about the GLOMAP simulation, and 95 % of all the points cover the GLOMAP simulations in their 95 % confidence interval, so the emulator is considered valid.

4.2 Sensitivity analysis

#### 4.2.1 Variance contributions at the surface

The CCN concentration at the surface for London (in June 2000) is estimated by the emulator to be $647 \, cm^{-3}$ with an emulator standard deviation of $2.1 \, cm^{-3}$. Uncertainty in the 8 parameters leads to an estimated standard deviation of $106 \, cm^{-3}$ around the expected CCN concentration. A strength of the emulator is that it predicts not only the CCN value and the parametric uncertainty, but also the 95 % probability bounds as a measure of the "confidence" of the emulated prediction. The small emulator standard deviation here along with the successful validation in Fig. 5a shows that using the emulator has had a very small effect on the accuracy with which the parameter sensitivities are estimated and hence we have an accurate emulator.

The sensitivity analysis partitions the variance due to the 8 parameters into the variance due to each parameter and their interactions. The results are summarised in Table 1. The *main effect variance* is the percentage of the total variance due to the perturbation of each parameter individually (this is the variance captured by the OAT tests) and the *total effect variance* is the percentage of the total variance calculated due to the main effect plus the interaction between different parameters. In London, 95 % of the variance due to the 8 parameters is described by the main effect variance terms and there are only weak interactions between the uncertain parameters (only a further 3 % of the variance in CCN is de-

scribed by the first-order interaction effects and the remaining 2 % by higher order interactions). The interaction variance contributions and the total variance contributions are also added together from the 8 parameters in the table. The total variance contributions will typically not add to 100 % because the variance contributions are shared between parameters in the presence of interactions, the further it is away from 100 % gives a quick indication of the interaction effects. In Table 1 the summed total effect variance is 107 % showing that the interaction effects are small.

The parameter sensitivities are plotted in Fig. 7 where the main effects and interactions can be seen clearly. The relationship between CCN concentration and each parameter are plotted in Fig. 8. Figure 7a shows the parameter sensitivities and Fig. 8a shows the CCN versus parameter relationships in the London grid box. There is a clear positive linear relationship between the sulphur emissions (SO2_EMS) and the estimated CCN. There is also a negative correlation between the oxidation activation diameter (OX_DIAM) and CCN and a positive correlation between the particulate sulphur emissions (SO2_PART) and CCN, but these are not very strong relationships compared to that with sulphur emissions, and the other parameters show little variation with CCN. Overall, the emulator predicts that 79 % of the variance in June 2000 surface-level CCN is due to the uncertainty in the sulphur emissions, 8 % is due to the uncertainty in the oxidation activation diameter, and 4 % due to the uncertainty in particulate sulphur emissions. The green sections in Fig. 7a are relatively small compared to the red sections showing that the main effects are most important in the London grid cell.

The estimated CCN concentration at the surface of the remote marine site in the Pacific is $57 \, cm^{-3}$, with an emulator standard deviation of $0.5 \, cm^{-3}$ and a parametric standard deviation of $14 \, cm^{-3}$. The sensitivity results are summarised in Table 2 and plotted in Fig. 7b. The total effect sensitivities indicate much stronger interaction between the model parameters than at the polluted site, shown by the fact that the summed total effect is 126 %. The interactions between the model parameters are also shown clearly by the green regions in Fig. 7b. Overall, 80 % of the variance is due to the individual parameters and 20 % due to the interactions. The interaction effects are separated by parameter. For example, the total effect of the oxidation activation diameter is 87 % compared to its main effect of 70 %, and the total effect of the nucleation scavenging diameter is 12 % compared to its main effect of only 2 %. The main effect relationships between the parameters and CCN concentration for the Pacific grid box can be seen in Fig. 8b where the dominance of oxidation activation diameter is clear. Figure 8b shows that non-linear relationships between the parameters and CCN concentration can be captured by the emulator. The first order interaction between the oxidation activation diameter and nucleation scavenging diameter is the most important of the interaction effects and accounts for 6 % of the total variance in CCN concentration. The joint effect of the
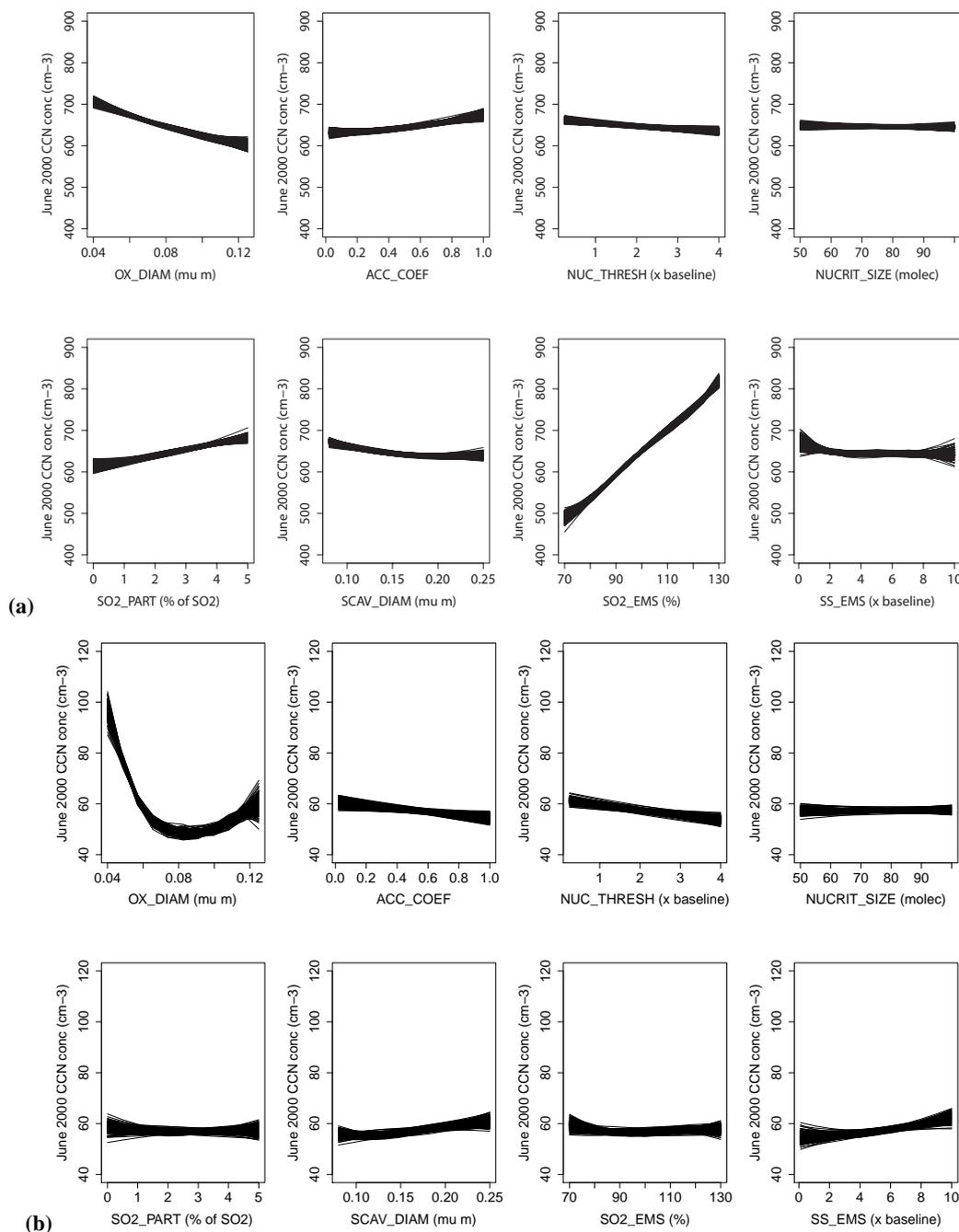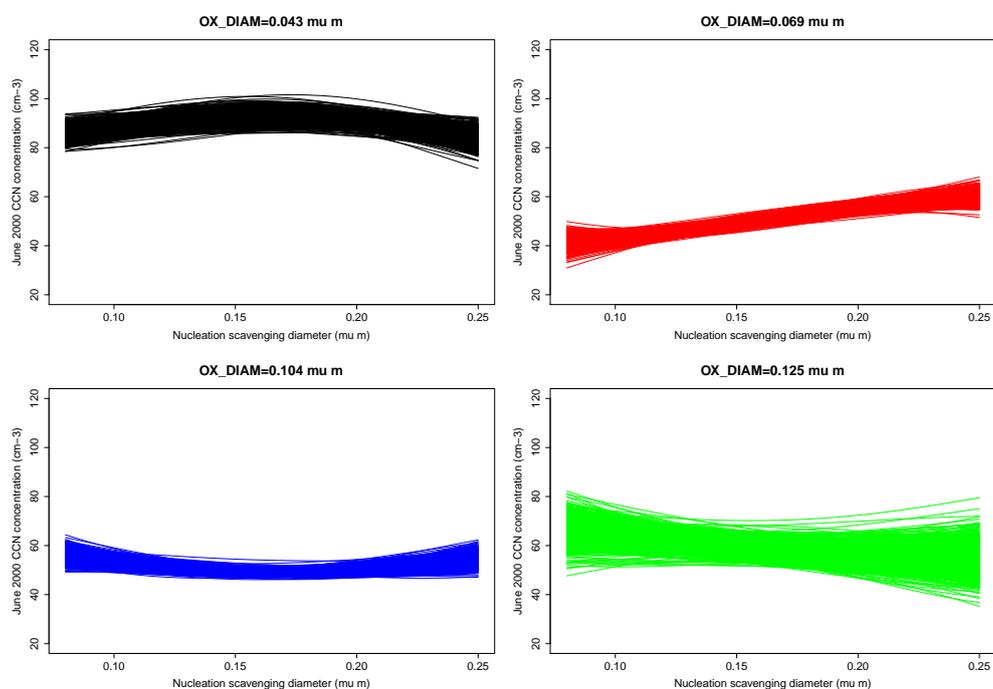
Fig. 8. The relationship between the CCN concentration and each of the parameters listed in Sect. 3 for **(a)** London grid cell and **(b)** Pacific grid cell. In each case the thickness of the line represents the emulator uncertainty whilst the spread on the y axis represents the uncertainty due to the uncertainty in the parameter. Each plot here is based on all other parameters being fixed. It is shown that uncertainty in sulphur emissions dominates uncertainty in CCN concentration.

oxidation activation diameter and nucleation scavenging diameter is shown in Fig. 9. The CCN concentration is more sensitive to the nucleation scavenging diameter when the oxidation activation diameter is between 0.06 and 0.1 μm and the stronger relationship can be seen when oxidation diameter is 0.074 μm in Fig. 9. A possible explanation for the increased sensitivity to nucleation scavenging is that the size and number of CCN available are optimised for wet deposition when the oxidation diameter is in the range of 0.06 and 0.1 μm. Emulator uncertainty is increased outside of the oxidation diameter range of 0.06 and 0.1 μm which can be seen by the spread of the lines showing the relationship between CCN concentration and nucleation scavenging diameter for given oxidation diameter in Fig. 9.

Table 1. Model parameters and their effect on the variance of CCN concentration at the surface in a polluted grid cell covering London (UK).

Expected CCN = 647 cm^{-3}		Emulator st. dev, σ_{emulator} = 2.1 cm^{-3}		Parametric st. dev, $\sigma_{\text{parametric}}$ = 106 cm^{-3}	
Number	Parameter	Range	Main effect variance contribution (%)	Parameter inter-action variance contribution (%)	Total effect variance contribution (%)
X1	Oxidation activation diameter	0.04–0.125 μm	8.2	1.3	9.5
X2	Mass accommodation coefficient	0.02–1.0	1.5	1.2	2.7
X3	H$_2$SO$_4$ nucleation threshold	0.25–4.0×baseline	0.5	0.4	0.9
X4	Nucleation critical cluster size	50–100 molec	0.0	0.4	0.4
X5	Sulphate particulate emissions	0–5 % of SO$_2$	4.0	2.0	6.1
X6	Cloud nucleation scavenging diameter	0.08–0.25 μm	0.9	1.3	2.2
X7	Sulphur emissions	70–130 % baseline	79.1	3.0	82.1
X8	Sea spray emissions	0.1–10×baseline	0.5	3.0	3.5
–	All 8 parameters	–	94.7	12.7	107.4

Fig. 9. The joint effect of oxidation activation diameter and nucleation scavenging diameter in the Pacific Ocean grid cell. For four different values of oxidation activation diameter the effect of nucleation scavenging is shown. The sensitivity of CCN concentration to nucleation scavenging is shown to differ (by the shape of the line) when oxidation activation diameter changes. The emulator is less certain in its nucleation scavenging estimate given a high oxidation diameter, as indicated by the spread of the lines.

4.2.2 Vertical profile of variance contributions

The vertical profile of CCN concentration and the associated variance contributions are shown in Fig. 10. Figure 10a and c show that the absolute CCN concentration and the parametric uncertainty (measured by the 95 % confidence intervals) decrease with altitude in the London grid box but that the parametric uncertainty remains relatively high with altitude in the Pacific grid box. The CCN sensitivity to each uncertain model parameter (Fig. 10b and d) is calculated using separate emulators for each model level. The solid lines in Fig. 10b and d show the main effect of the different parameters and the corresponding dashed lines show the total effect of each parameter.
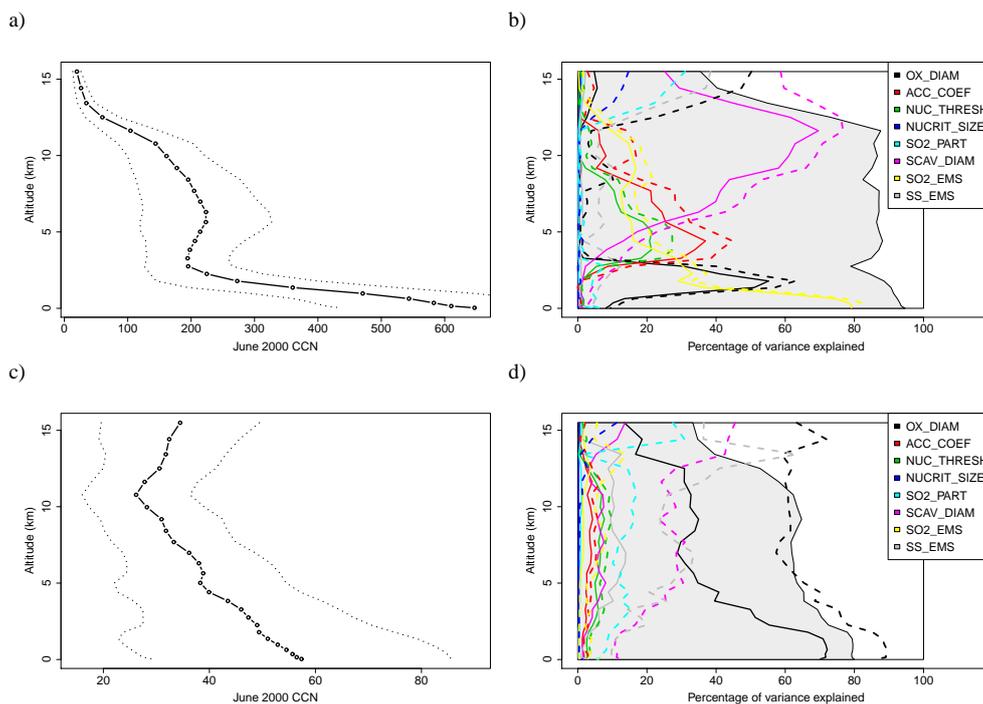
Fig. 10. Emulator estimated CCN concentration (solid line) and its 95 % confidence interval due to uncertainty in the model parameters (dotted line) in **(a)** London and **(c)** the Pacific Ocean as estimated by the emulator at 26 model levels. The results from the 26 emulators are shown as circles but have been linearly interpolated to show the vertical profile. The variance contribution (%) or sensitivity of each parameter is shown in **(b)** for London and **(d)** for the Pacific Ocean. The solid lines show the main effect sensitivity and the dashed lines show the total effect sensitivity. The shaded background shows the total uncertainty explained by the main effects of each parameter; this is the maximum uncertainty that could be measured using one-at-a-time tests.

Table 2. Model parameters and their effect on the variance of CCN concentration at the surface in a remote grid cell over the Pacific Ocean.

	Expected CCN = 57 cm^{-3}		Emulator st. dev, $\sigma_{emulator} = 0.5$ cm^{-3}		Parametric st. dev, $\sigma_{parametric} = 14$ cm^{-3}
Number	Parameter	Range	Main effect variance contribution (%)	Parameter inter- action variance contribution (%)	Total effect variance contribution (%)
X1	Oxidation activation diameter	0.04–0.125 μm	70.2	17.0	87.2
X2	Mass accommodation coefficient	0.02–1.0	1.7	0.8	2.5
X3	H$_2$SO$_4$ nucleation threshold	0.25–4.0×baseline	2.5	1.0	3.5
X4	Nucleation critical cluster size	50–100 molec	0.1	1.2	1.3
X5	Sulphate particulate emissions	0–5 % of SO$_2$	0.3	5.1	5.4
X6	Cloud nucleation scavenging diameter	0.08–0.25 μm	1.8	10.5	12.3
X7	Sulphur emissions	70–130 % baseline	0.4	5.6	6.0
X8	Sea spray emissions	0.1–10×baseline	3.0	8.2	11.2
–	All 8 parameters	–	80.0	46.2	126.2

In the London grid cell the sulphur emissions are the dominant source of uncertainty in the CCN near the surface, but higher in the atmosphere the uncertainty in the model parameters becomes more important. Between 1 and 3 km altitude the uncertainty in the oxidation activation diameter explains most of the variance in CCN concentration, which is consistent with the altitude of low-level clouds. From 3 to 6 km

the uncertainty in the accommodation coefficient contributes most of the variance in the CCN concentration and from 6 km to 12 km the uncertainty in the nucleation scavenging diameter is the dominant source of CCN variance. However, as at the surface, most of the variance at higher altitudes is explained by the individual parameters, with interactions between parameters accounting for no more than 20 % of the

total uncertainty. Above 12 km at this location the interaction effects become more important; however, as shown in Fig. 10a both the CCN and its variance are very low and and so the interactions are not investigated further here.

It is clear that to improve the estimate of the June 2000 CCN concentration near the surface in London it is most important to reduce the uncertainty in the sulphur emissions used to force GLOMAP. Since the uncertainty near the surface is very strongly dependent on emissions rather than processes, models might be expected to be in closer agreement over more polluted regions when the same emissions databases are used. To improve the estimates of CCN higher in the atmosphere it is most important to reduce the uncertainty in the model process parameters.

Figure 10c shows the estimated CCN concentration and its 95 % confidence interval due to the uncertain parameters through the vertical profile for the Pacific Ocean location. In general, the CCN concentration decreases through the vertical profile as does its uncertainty. In Fig. 10d CCN concentration can be seen to be most sensitive to the oxidation activation diameter throughout the atmosphere. However, interactions between the model parameters become increasingly important in the free and upper troposphere. From about 4 km upwards, only about 60 % of the total variance is explained by main effects. The interaction of the scavenging diameter and the sea spray emission flux with other parameters dominate the mid-tropospheric CCN variance. The strong interaction effects of these two parameters means that they exert an indirect control on CCN concentration in the remote marine free troposphere. This interaction effect is plausible since nucleation is an important source of CCN in this part of the atmosphere (Merikanto et al., 2009), and the nucleation rate is strongly affected by the surface area of existing particles, which itself is affected by large sea spray particles and by the particles removed by precipitation. These interaction effects could not be quantified using the traditional OAT tests, thus the total effect of scavenging and sea spray on the CCN uncertainty would be significantly underestimated.

5 Discussion and future work

We have presented a statistically rigorous but computationally efficient approach to quantifying the parametric uncertainty of a complex global aerosol model. The approach is equally applicable to other models that require long computation times. The combination of a good parameter sampling design (here, Latin hypercube) and Gaussian process emulation enables the results from a relatively small number of model simulations to be used to perform a full variance-based sensitivity analysis, which would otherwise require many thousands of models runs in a Monte Carlo-type approach. Through variance decomposition the variance-based sensitivity analysis calculates the contribution of each uncertain parameter to the overall prediction uncertainty. However, the

advantage of our approach over the widely used one-at-a-time tests is that it also enables the interactions between all combinations of parameters to be calculated, and the parametric uncertainty refers to the entire parameter space, rather than being restricted to one point (the baseline model) defined by "default" parameters. As such, the variance-based approach produces a more realistic estimate of uncertainty and provides more useful information to the model developer.

We have shown that the emulator approach is very well suited to a complex global aerosol model. With an appropriate experimental design and number of model runs, the uncertainty in using the emulator instead of the global model is very small compared to the parametric uncertainty of the model itself. The number of model simulations required to train the emulator rises linearly with the number of parameters. Thus, for a very large number of uncertain parameters the approach will become much more efficient (and comprehensive in its coverage of parameter space) than widely used factorial designs.

The primary aim of this paper was to present and test the method for carrying out a sensitivity analysis on a global aerosol model. The next step is to include a more comprehensive study of the model parameters, including for example carbonaceous emissions, size distributions, boundary layer particle formation, secondary organic aerosols, all of which are thought to be important for CCN concentrations (Spracklen et al., 2006, 2008; Merikanto et al., 2009; Spracklen et al., 2011). In the next experiment more effort will be taken to elicit parameter uncertainty distributions from experts rather than simply assigning uncertainty ranges. Another advantage of this approach is that the uncertainty distribution of the inputs can be changed in the emulator without more model runs, which is not possible with the Monte Carlo approach. It is also possible to investigate further model diagnostics without any more experiments provided the emulator is validated.

The results here show that the uncertainty in the CCN concentration due to the model parameters as measured by variance contributions is dependent on the location and altitude. Uncertainty in emissions dominates over the model parameters close to emission sources, but process parameters dominate in the remote region we examined. For the 8 parameters we examined, we also found that their main effects (diagnosed in one-at-a-time tests) dominated the overall variance close to sources. However, in the remote location interactions between parameters become very important, meaning that the overall model uncertainty would be underestimated in one-at-a-time tests.

Little has been discussed in this paper about the statistical assumptions behind the Gaussian process emulator and in particular how to deal with failure of the emulator validation. The work here was shown to be robust to the statistical assumptions by changing the assumptions and carrying out the same sensitivity analysis with little change in the results.

A more comprehensive study with additional model parameters may be more sensitive to the statistical assumptions. In particular, it might be expected that the parameter space leading to unrealistic output will be larger and irregularly shaped (some combinations of parameter values might be unrealistic despite their marginal uncertainties being reasonable) requiring a more sophisticated method of specifying the joint distribution of the input parameters. The emulation is still possible with a more sophisticated joint distribution but the software will need to be advanced to carry out the sensitivity analysis with this new joint distribution since in GEM-SA the joint distribution is a regular shape based on the marginal uncertainty distributions. The basic idea is to remove the area of the parameter space that contains unrealistic parameter value combinations.

A future direction is multivariate emulation which may be useful to help reduce prediction uncertainty in different aspects of aerosol modelling. It may be used to build an emulator for multiple outputs to describe specific aspects of global aerosol distribution determined by multiple model outputs, or to describe the vertical or regional distribution of aerosol by allowing structure between model output in neighbouring grid boxes. Our current work has not allowed for structure between model outputs but we have instead built univariate emulators of individual model outputs. Given the nature of the global aerosol model we would be in an ideal position to compare the univariate emulation study we have completed so far with the multivariate emulation.

The next step in reducing the uncertainty in the global model will be to compare the results more comprehensively to observation data, i.e. calibration. Calibration is used to reduce the uncertainty in the model parameters and thus improve knowledge about them. The study here shows which parameters efforts should be focussed on to improve the model prediction of CCN concentration. The sensitive parameters can be improved using lab experiments or observations when available. The same methods described here can be carried out by any modelling group to quantify the uncertainty in a number of global aerosol models (though these methods are not restricted to aerosol models). The sensitive processes can then be compared in the different models improving our understanding of, and perhaps reducing, model diversity. The implications of the different model structures on the model predictions will be better understood when the uncertainty of each model and its sources is quantified.

Appendix A

Mathematical description of Gaussian process emulation

The June 2000 CCN concentration in a single grid box is defined by Y. In fact Y can be any model quantity of interest. A capital letter is used since the CCN concentration is unknown. The model parameters are defined by X and are uncertain. The GLOMAP model is some function of the parameters that leads to an estimate of the CCN concentration defined by $Y = f(X)$. The GLOMAP model f is unknown in the sense that it is so complicated the output Y is unknown until the model is run on a computer. In this material x refers to the values set for each of the 8 unknown parameters in any single GLOMAP run so here is actually a vector of length 8. Once GLOMAP has been run with inputs x we have a realisation defined as $y = f(x)$ yielding model output y. In this paper y is the monthly mean CCN concentration in a defined grid cell.

The unknown function f is approximated by $\hat{f}$ using the Gaussian process emulator. The aim of the emulator is to produce an accurate approximation for the GLOMAP function f using the minimum number of model runs. The sensitivity of the output Y to the uncertain inputs X is quantified by sensitivity analysis using the emulator. The emulator is required since the model is too CPU intensive to carry out the number of runs needed to perform a full sensitivity analysis using GLOMAP itself.

First, we describe the application of the Gaussian process to produce an emulator for GLOMAP and then the use of the emulator to carry out sensitivity analysis is shown. The work in this Appendix is based on Oakley and O'Hagan (2004).

A1   The Gaussian process emulator

The Gaussian process emulator produces a probability distribution for the GLOMAP model f as a function of the model parameters. The distribution is used to estimate CCN concentrations predicted by GLOMAP and quantify the uncertainty around this estimate. We used the GEM-SA software (Kennedy, 2004) to build the emulator but it is important to understand the choices that can be made in the software and also its limitations.

The emulator is "built" by combining some runs of the model (the training data) and some beliefs about the model behaviour using the Bayesian paradigm. The key assumption is that the model behaves smoothly so that each run of the model gives information about the model output at neighbouring parameter settings. The beliefs about the model behaviour are formulated using the Gaussian process, the prior distribution for f at given values of x. The Gaussian process is the functional analogue of the normal distribution so is described in terms of a mean function and a covariance function. The Gaussian process prior is updated with the training data to produce a posterior distribution, which is therefore a conditional Gaussian process (conditioned on the model data). The mean function is used to estimate GLOMAP and the covariance function used to calculate the uncertainty around this mean function. The mean and covariance function for the prior Gaussian process need to be specified.

The mean function in the Gaussian process prior is given by

$$E\{f(\boldsymbol{x})|\boldsymbol{\beta}\} = \boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{\beta}, \tag{A1}$$

where $\boldsymbol{h}(\cdot)$ is a vector of regression functions with unknown coefficients $\boldsymbol{\beta}$. The mean function is expressed in terms of the expectation (or expected value) since it is technically the population mean of a random variable and often estimated by the sample mean. The $\boldsymbol{\beta}$ parameters are hyperparameters and themselves have a prior distribution, discussed below. The mean is conditional on $\boldsymbol{\beta}$ because these have to be calculated in order to work out the prior mean function. In GEM-SA the vector $\boldsymbol{h}(\cdot)$ can represent a constant mean or a simple linear regression function of the inputs $\boldsymbol{x}$ on output y and hence $\boldsymbol{h}(\cdot) = \boldsymbol{1}$ or $\boldsymbol{h}(\cdot) = (\boldsymbol{1}, \boldsymbol{x}^T)$. For this paper both choices for the mean function were compared with little difference. The results shown in the paper used the simple linear regression mean function.

The covariance function specifies the covariance in the output given any pair of parameter settings $\boldsymbol{x}$ and $\boldsymbol{x}'$ and is defined by

$$\mathrm{cov}\{f(\boldsymbol{x}), f(\boldsymbol{x}')|\sigma^2\} = \sigma^2 c(\boldsymbol{x}, \boldsymbol{x}'), \tag{A2}$$

where $c(\boldsymbol{x}, \boldsymbol{x}')$ is the correlation function and σ^2 is another hyperparameter. As with $\boldsymbol{\beta}$, σ^2 is given a prior distribution, discussed below. The correlation function $c(\boldsymbol{x}, \boldsymbol{x}')$ decreases as the distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$ increases, is equal to 1 when $\boldsymbol{x} = \boldsymbol{x}'$ and ensures that the covariance matrix is positive semidefinite, i.e. invertible. The correlation function usually takes one of a number of specified forms to ensure a valid covariance function. In this paper the Gaussian correlation form is used, defined by

$$c(\boldsymbol{x}, \boldsymbol{x}') = \exp\{-(\boldsymbol{x} - \boldsymbol{x}')^T \mathbf{R}(\boldsymbol{x} - \boldsymbol{x}')\}, \tag{A3}$$

where $\mathbf{R}$ is a diagonal matrix with elements $\mathrm{diag}(r_i)$ where r_i is a hyperparameter describing the smoothness of the function with respect to parameter i.

The hyperparameters $\boldsymbol{\beta}$ are given a multivariate normal distribution (a conjugate prior ensuring that the Gaussian process remains the posterior distribution once $\boldsymbol{\beta}$ are integrated out). The hyperparameter σ^2 is given the inverse gamma distribution. The prior distributions for the hyperparameters $\boldsymbol{\beta}$ and σ^2 are specified such that the joint distribution of the two is the weak normal inverse gamma distribution $p(\boldsymbol{\beta}, \sigma^2) \propto \sigma^2$ so that the prior variance for the function f is infinite. The result is that no real information about GLOMAP is given by the prior distributions on the hyperparameters so that $\boldsymbol{\beta}$ and σ^2 are calculated using the training data. The hyperparameters $\boldsymbol{r}$ are also given weak prior distributions (the weak uniform distribution) and calculated from the training data. The choice of prior distributions in this study are standard uninformative priors for the hyperparameters.

The GLOMAP model is run to produce the 80 sets of training data. The runs are designed using Latin hypercube sampling in 8 dimensions to ensure that the space of the joint input uncertainty, defined by χ, is well represented. In this paper we are studying CCN concentration so the calculated June 2000 CCN is taken from each run to define the training data $\{y_1 = f(\boldsymbol{x}_1), y_2 = f(\boldsymbol{x}_2), ..., y_{80} = f(\boldsymbol{x}_{80})\}$. The Gaussian process is such that any subset of points on the function can be described by the joint multivariate normal distribution with the mean and covariance function specified as above. Using this multivariate distribution for the training data the hyperparameters are calculated. When there is specific information about the model behaviour this can be included through more sophisticated specifications of the hyperparameters; in this work there is no specific information about the hyperparameters and so they are calcuated using the training data via GEM-SA.

The posterior Student t-process (the Gaussian process with estimated variance) is a result of conditioning the prior Gaussian process on the training data. The posterior mean function is

$$m^*(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})^T \hat{\boldsymbol{\beta}} + \boldsymbol{t}(\boldsymbol{x})^T \mathbf{A}^{-1}(\boldsymbol{y} - \mathbf{H}\hat{\boldsymbol{\beta}}), \tag{A4}$$

and the posterior covariance function is

$$\hat{\sigma}^2 c(\boldsymbol{x}, \boldsymbol{x}')^* = \hat{\sigma}^2 (c(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{t}(\boldsymbol{x})^T \mathbf{A}^{-1} \boldsymbol{t}(\boldsymbol{x}') + (\boldsymbol{h}(\boldsymbol{x})^T \\ - \boldsymbol{t}(\boldsymbol{x})^T \mathbf{A}^{-1}\mathbf{H})(\mathbf{H}^T \mathbf{A}^{-1}\mathbf{H})^{-1}(\boldsymbol{h}(\boldsymbol{x}')^T - \boldsymbol{t}(\boldsymbol{x}')^T \mathbf{A}^{-1}\mathbf{H})^T), \tag{A5}$$

where

$$\boldsymbol{y}^T = (f(\boldsymbol{x}_1), ..., f(\boldsymbol{x}_n)), \tag{A6}$$

$$\mathbf{H}^T = (\boldsymbol{h}(\boldsymbol{x}_1), ..., \boldsymbol{h}(\boldsymbol{x}_n)), \tag{A7}$$

$$\mathbf{A} = \begin{pmatrix} 1 & c(\boldsymbol{x}_1, \boldsymbol{x}_2) & \cdots & c(\boldsymbol{x}_1, \boldsymbol{x}_n) \\ c(\boldsymbol{x}_2, \boldsymbol{x}_1) & 1 & & \vdots \\ \vdots & & \ddots & \\ c(\boldsymbol{x}_n, \boldsymbol{x}_1) & \cdots & & 1 \end{pmatrix}, \tag{A8}$$

$$\boldsymbol{t}(\boldsymbol{x})^T = (c(\boldsymbol{x}, \boldsymbol{x}_1), ..., c(\boldsymbol{x}, \boldsymbol{x}_n)), \tag{A9}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^T \mathbf{A}^{-1}\mathbf{H})^{-1}\mathbf{H}^T \mathbf{A}^{-1}\boldsymbol{y} \tag{A10}$$

and

$$\hat{\sigma}^2 = \frac{\boldsymbol{y}^T(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{H}(\mathbf{H}^T \mathbf{A}^{-1}\mathbf{H})^{-1}\mathbf{H}^T \mathbf{A}^{-1})\boldsymbol{y}}{n - q - 2}. \tag{A11}$$

In this experiment $n = 80$ is the number of training runs and $q = 9$ is the number of coefficients in the mean function, i.e. the number of parameters plus one. The derivation of the posterior formulae can be found in O'Hagan (1994).

Any other point $\boldsymbol{x}$ in the function can be estimated with a measure of uncertainty using the above formulae. The whole function and properties of it, such as the variance, can also be estimated from the above formulae for further inference.

A2 Sensitivity analysis

In this work a probabilistic sensitivity analysis is carried out. The term probabilistic implies that the uncertainty in the model parameters is described formally by a probability distribution. That is the uncertainty in the true parameter values X is described by some probability distribution G. The uncertainty in Y due to $G(X)$ is quantified and in particular the sources of uncertainty in Y from the different elements of X (the different parameters) are identified. In practice the probability distribution G_i of the different parameters X_i is specified and then characterised by the joint distribution G. In GEM-SA the parameter distributions G_i are limited to uniform or normal. Different distributions for G_i can be used if the distribution can be transformed to uniform or normal but care must be taken in the interpretation of the results. The emulator software has to be developed further if different probability distributions for G_i are essential. The distribution G_i of X_i can be formally elicited from experts but in this paper the uniform distribution has been used given the parameter ranges in Spracklen et al. (2005b) based on expert advice.

There are two sets of sensitivity results used in this work. The main sensitivity results are the variance-based sensitivity indices but the relationship between the parameters and the output based on conditional expectations are plotted to help interpret the effect on the output from each input and their interactions. The two sets of sensitivity results are described here.

Using the variance to assess uncertainty in the presence of independent model parameters means that variance decomposition can be used to assess the relative importance of each model parameter and its interactions (Cox, 1982) which is important in identifying the major sources of uncertainty in modelling studies. GEM-SA reports two measures of sensitivity for each parameter based on the variance: the main effect sensitivity and the total effect sensitivity (Saltelli et al., 2000). GEM-SA uses variance decomposition to calculate the sensitivities as a percentage of the total variance explained.

Not considering the use of the emulator here, the main effect variance is

$$V_i = \text{var}\{E(Y|X_i)\} = \text{var}\{z_i(X_i)\}, \tag{A12}$$

the expected amount by which the uncertainty is reduced if the true value of X_i was learnt. The total effect variance is

$$V_{Ti} = \text{var}(Y) - \text{var}\{E(Y|X_{-i})\}, \tag{A13}$$

the expected amount of uncertainty left after everything except for X_i is learnt ($-i$ represents all parameters except i). The sensitivities for each parameter are compared by dividing these variances by the total variance of the output Y giving the main effect and total effect sensitivities, S_i and S_{Ti}. The total effect sensitivity compared to the main effect sensitivity is used to highlight interactions between parameter i

and other parameters. The first-order interaction sensitivities are also calculated in GEM-SA using

$$V_{i,j} = \text{var}\{E(Y|X_{i,j})\} = \text{var}\{z_i(X_i) \tag{A14}$$
$$+ z_i(X_j) + z_{i,j}(X_{i,j})\}.$$

The higher order interactions follow similarly but are not calculated in GEM-SA.

When emulation is used to calculate the sensitivities we are in fact finding the expected sensitivities $E^*(V_i)$ and $E^*(V_{Ti})$. This is done by integrating formulae based on Eqs. (A4) and (A5) over different subsets of the parameters and is detailed in Oakley and O'Hagan (2004). The uncertainty in the sensitivity measures (i.e. $V^*(V_i)$) are not calculated. With an accurate emulator the uncertainty in the sensitivity measures is assumed to be small.

The variance-based methods quantify the relative sensitivity of the model output Y to each of the uncertain parameters X_i but they do not give any indication of how the output is actually responding to each of the input parameters. The response of the output Y to each input x_i is plotted in GEM-SA to visualise the effect of the individual parameters. The first-order interaction can also be plotted. This is possible by decomposing the output y into main effects and interactions

$$y = f(x) = E(Y) + \Sigma_{i=1}^d z_i(x_i) + \Sigma_{i<j} z_{i,j}(x_{i,j}) \tag{A15}$$
$$+ ... + z_{1,2,.....p}(x),$$

for p independent parameters. The main effect given the above decomposition is

$$z_i(x_i) = E(Y|x_i) - E(Y) \tag{A16}$$

and the first-order interaction is

$$z_{i,j} = E(Y|x_{i,j}) - z_i(x_i) - z_j(x_j) - E(Y). \tag{A17}$$

The values required to draw these plots are calculated using the emulator.

The main effects and first-order interactions are calculated in GEM-SA by finding the conditional expectations of the output Y given each of the parameters and the subsets of each pair of parameters. In general, the main effect of parameter i is calculated using the conditional expected value of Y,

$$E(Y|x_i) = \int_{\chi_{-i}} f(x) dG_{-i|i}(x_{-i}|x_i), \tag{A18}$$

and the unconditional expected value of Y,

$$E(Y) = \int_{\chi} f(x) dG. \tag{A19}$$

That is the function (GLOMAP) is integrated over the joint distribution of all other parameters $-i$. Again, because we are using the emulator, we are actually calculating the expected values of the main effects and interactions instead. We therefore find the posterior conditional mean of the output Y defined here as $E^*\{E(Y|x_p)\}$ where E^* denotes the

use of the emulator. When Monte Carlo sampling is possible $E(Y|\boldsymbol{x}_p)$ can be derived directly. The specific formulae used to derive the main effect and interactions can be found in Oakley and O'Hagan (2004).

A particular and important consequence of the above formulae is that the emulator can be used to calculate the posterior mean and variance of the expected output Y given the uncertainty in all model parameters. When Eq. (A18) is integrated over all i the posterior mean of the expected output $E^*(E(Y))$ is found; this is equivalent to integrating the posterior mean in Eq. (A4) over all parameters and is the expected CCN taking into account uncertainty in the 8 model parameters. The posterior variance of the expected output is calculated by integrating the posterior covariance in Eq. (A5) over the joint distribution G of all parameters and is used to measure the uncertainty around the mean due to emulation. Integrating the posterior covariance over the different parameter uncertainty distributions allows the emulator uncertainty around the conditional expected values of the output to be calculated. The emulator uncertainty around the model output is shown in Fig. 2 by the spread of the green curves. The posterior variance around the output variance $V^*(\text{Var}(Y))$, the uncertainty in the variance due to the uncertain parameters) is not calculated but is assumed to be small if the emulator is shown to be accurate. The parameter uncertainty $E^*(\text{Var}(Y))$ is calculated and from this we have calculated the standard deviation. In practice because the expectation of the variance has been found it is not trivial to calculate the expected standard deviation. Jensen's Inequality (Jensen, 1906) shows that the expected standard deviation will always be greater than or equal to the square-root of the expected variance

$$\sqrt{(E[\text{Var}(Y)])} \leq E[\sqrt{\text{Var}(Y)}] \qquad (A20)$$

and hence we only have a lower bound for the expected standard deviation. In this case because the emulator variance is so low we assume the emulator is accurate enough for us to believe the expected standard deviation will be close to the square-root of the expected variance. For this same reason in this paper we do not calculate parameter sensitivities in terms of the standard deviation from the variance-based measures; instead we only quote percentage of the expected variance explained by each input.

Edited by: N. Riemer

References

Ackerley, D., Highwood, E., Frame, D., and Booth, B.: Changes in the global sulfate burden due to perturbations in global CO_2 concentrations, J. Climate, 20, 5421–5432, 2009.

Ackermann, I. J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F., and Shankar, U.: Modal aerosol dynamics model for Europe: Development and first applications, Atmos. Environ., 32, 17, 2981–2999, 1998.

Adams, P. J. and Seinfeld, J. H.: Predicting global aerosol size distributions in general circulation models, J. Geophys. Res., 107(D19), 4370, doi:10.1029/2001JD001010, 2002.

Adams, P. J. and Seinfeld, J. H.: Disproportionate impact of particulate emissions on global cloud condensation nuclei concentrations, Geophys. Res. Lett., 30, 1239, doi:10.1029/2002GL016303, 2003.

Annan, J., Hargreaves, J., Ohgaito, R., Abe-Ouchi, A., and Emori, S.: Efficiently constraining climate sensitivity with ensembles of Paleoclimate simulations, SOLA, 1, 181–184, 2005.

Bastos, L. and O'Hagan, A.: Diagnostics for Gaussian Process Emulators, Technometrics, 4, 425–438, 2009.

Bates, R. A., Buck, R. J., Riccomagno, E., and Wynn, H. P.: Experimental design and observation for large systems, J. Roy. Stat. Soc. B, 58, 77–94, 1995.

Binkowski, F. S. and Shankar, U.: The Regional Particulate Matter Model: 1. Model description 30 and preliminary results, J. Geophys. Res., 100(D12), 26191–26209, 1995.

Chipperfield, M.: New version of the TOMCAT/SLIMCAT offline chemical transport model: Intercomparison of stratospheric tracer experiments, Q. J. Roy. Meteorol. Soc., 132, 1179–1203, doi:10.1256/qj.05.51, 2006.

Collins, M., Booth, B., Bhaskaran, B., Harris, G., Murphy, J., Sexton, D., and Webb, M.: A comparison of perturbed physics and multi-model ensembles: Model errors, feedbacks and forcings, Clim. Dynam., doi:10.1007/s00382-010-0808-0, 2010.

Cox, D.: An analytical method for uncertainty analysis of nonlinear output functions, with applications to fault-tree analysis, IEEE Trans. Reliab., 31, 265–268, 1982.

Currin, C., Mitchell, T. J., Morris, M., and Ylvisaker, D.: Bayesian prediction of deterministic functions with applications to the design and analysis of computer experiments, J. Am. Stat. Assoc., 86, 953–963, 1991.

de Leeuw, G., Andreas, E., Anguelova, M., Fairall, C., Lewis, E., O'Dowd, C., Schulz, M., and Schwartz, S.: Production flux of sea spray aerosol, Rev. Geophys., 49, RG2001, doi:10.1029/2010RG000349, 2011.

Debry, E., Fahey, K., Sartelet, K., Sportisse, B., and Tombette, M.: Technical Note: A new SIze REsolved Aerosol Model (SIREAM), Atmos. Chem. Phys., 7, 1537–1547, doi:10.5194/acp-7-1537-2007, 2007.

Dentener, F., Kinne, S., Bond, T., Boucher, O., Cofala, J., Generoso, S., Ginoux, P., Gong, S., Hoelzemann, J. J., Ito, A., Marelli, L., Penner, J. E., Putaud, J.-P., Textor, C., Schulz, M., van der Werf, G. R., and Wilson, J.: Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom, Atmos. Chem. Phys., 6, 4321–4344, doi:10.5194/acp-6-4321-2006, 2006.

Fisher, R.: The arrangement of field experiments, Journal of the Ministry of Agriculture of Great Britain, 33, 503–513, 1926.

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fa-

hey, D., Haywood, J., Lean, J., Lowe, D., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., and Van Dorland, R.: Changes in Atmospheric Constituents and in Radiative Forcing, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Gaber, N., Foley, G., Pascual, P., Stiber, N., Sunderland, E., Cope, B., Nold, A., and Zaleem, Z.: Guidance on the development, evaluation and application of environmental models., Tech. Rep. EPA/100/K-09/003, US Environmental Protection Agency, 2009.

Ghan, S., Laulainen, N., Easter, R., Wagener, R., Nemesure, S., Chapman, E., Zhang, Y., and Leung, R.: Evaluation of aerosol direct radiative forcing in MIRAGE, J. Geophys. Res., 106(D6), 5295–5316, 2001.

Goldstein, M. and Rougier, J.: Bayes linear calibrated prediction for complex systems, J. Am. Stat. Assoc., 101, 1132–1143, 2006.

Gong, S.: A parameterization of sea-salt aerosol source function for sub and super-micron 10 particles, Global Biogeochem. Cy., 17(4), 1097, doi:10.1029/2003GB002079, 2003.

Haerter, J. O., Roeckner, E., Tomassini, L., and von Storch J.-S.: Parametric uncertainty effects on aerosol radiative forcing, Geophys. Res. Lett., 36, L15707, doi:10.1029/2009GL039050, 2009.

Haylock, R. G. and O'Hagan, A.: On inference for outputs of computationally expensive algorithms with uncertainty on the inputs, in: Bayesian Statistics 5, edited by: Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., Oxford: Oxford University Press, 629–637, 1996.

Jacobson, M.: Development and application of a new air pollution modeling system II. Aerosol module structure and design, Atmos. Environ., 31(2), 131–144, 1997.

Jensen, J. L. W. V.: Sur les fonctions convexes et les ingalits entre les valeurs moyennes, Acta Mathematica, 30 (1), 175–193, doi:10.1007/BF02418571, 1906.

Jones, B. and Johnson, R.: Design and analysis for the Gaussian process model, Quality and Reliability Engineering International, 25(5), 515–524, 2009.

Kennedy, M.: The GEM software project, Tech. Rep. http://www.ctcd.group.shef.ac.uk/gem.html, Centre for Terrestrial Carbon Dynamics (CTCD), 2004.

Kennedy, M., Anderson, C., O'Hagan, A., Lomas, M., Woodward, I., Gosling, J., and Heinemeyer, A.: Quantifying Uncertainty in the Biospheric Carbon Flux for England and Wales, J. Roy. Stat. Soc. A, 171, Part 1, 109–135, 2008.

Kulmala, M., Laaksonen, A., and Pirjola, L.: Parameterizations for sulfuric acid/water nucleation rates, J. Geophys. Res., 103(D7), 8301–8307, 1998.

Lauer, A., Hendricks, J., Ackermann, I., Schell, B., Hass, H., and Metzger, S.: Simulating aerosol microphysics with the ECHAM/MADE GCM – Part I: Model description and comparison with observations, Atmos. Chem. Phys., 5, 3251–3276, doi:10.5194/acp-5-3251-2005, 2005.

Liu, X., Penner, J. E., and Herzog, M.: Global modeling of aerosol dynamics: Model description, evaluation, and interactions between sulfate and nonsulfate aerosols, J. Geophys. Res., 110, D18206, doi:10.1029/2004JD005674, 2005.

Liu, X., Penner, J., Das, B., Bergmann, D., Rodriguez, J., Stra-

han, S., Wang, M., and Feng, Y.: Uncertainties in global aerosol simulations: Assessment using three meteorological data sets, J. Geophys. Res., 112, D11212, doi:10.1029/2006JD008216, 2007.

Loeppky, J., Sack, J., and Welch, W.: Choosing the Sample Size of a Computer Experiment: A Practical Guide, Technometrics, 51(4), 366–376, 2009.

Lohmann, U. and Ferrachat, S.: Impact of parametric uncertainties on the present-day climate and on the anthropogenic aerosol effect, Atmos. Chem. Phys., 10, 11373–11383, doi:10.5194/acp-10-11373-2010, 2010.

Mann, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manktelow, P. T., Chipperfield, M. P., Pickering, S. J., and Johnson, C. E.: Description and evaluation of GLOMAP-mode: a modal global aerosol microphysics model for the UKCA composition-climate model, Geosci. Model Dev., 3, 519–551, doi:10.5194/gmd-3-519-2010, 2010.

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

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J., Stouffer, R., and Taylor, K.: The WCRP CMIP3 multi-model dataset: A new era in climate change research, B. Am. Meteorol. Soc., 88, 1383–1394, 2007.

Merikanto, J., Spracklen, D. V., Mann, G. W., Pickering, S. J., and Carslaw, K. S.: Impact of nucleation on global CCN, Atmos. Chem. Phys., 9, 8601–8616, doi:10.5194/acp-9-8601-2009, 2009.

Morgenstern, O., Braesicke, P., O'Connor, F. M., Bushell, A. C., Johnson, C. E., Osprey, S. M., and Pyle, J. A.: Evaluation of the new UKCA climate-composition model – Part 1: The stratosphere, Geosci. Model Dev., 2, 43–57, doi:10.5194/gmd-2-43-2009, 2009.

Morris, R., Kottas, A., Taddy, M., Furfaro, R., and Ganapol, B.: A Statistical Framework for the Sensitivity Analysis of Radiative Transfer Models, IEEE T. Geosci. Remote, 46(12), 4062–4074, 2008.

Murphy, J., Sexton, D., Barnett, D., Jones, G., Webb, M., Collins, M., and Stainforth, D.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, Nature, 430, 768–772, 2004.

Neal, R.: Regression and classification using gaussian process priors, in: Bayesian Statistics 6, edited by: Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., Oxford: Oxford University Press, 69–95, 1999.

Niehörster, F., Spangehl, T., Fast, I., and Cubasch, U.: Quantification of model uncertainties: parameter sensitivities of the coupled model ECHO-G with middle atmosphere, Geophys. Res. Abstracts, 8, EGU06-A-08526, 2006.

Oakley, J. E.: Bayesian Uncertainty Analysis for Complex Computer Codes, Ph.D. thesis, Department of Probability and Statistics, University of Sheffield, 1999.

Oakley, J. and O'Hagan, A.: Bayesian inference for the uncertainty distribution of computer model outputs, Biometrika, 89, 769–784, 2002.

Oakley, J. and O'Hagan, A.: Probabilistic sensitivity analysis of complex models: a Bayesian approach, J. Roy. Stat. Soc. B, 66, 751–769, 2004.

O'Hagan, A.: Kendall's Advanced Theory of Statistics, vol. 2B,

Bayesian Inference, London: Arnold, 1994.

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

O'Hagan, A. and Haylock, R. G.: Bayesian uncertainty analysis and radiological protection, in: Statistics for the Environment 3, Pollution Assessment and Control, edited by: Barnett, V. and Turkman, K. F., Wiley, Chichester, 1997.

Pan, W., Tatang, M., McRae, G., and Prinn, R.: Uncertainty analysis of direct radiative forcing by anthropogenic sulfate aerosols, J. Geophys. Res., 102(D18), 21915–21924, 1997.

Penner, J., Andreae, M., Annegarn, H., Barrie, L., Feichter, J., Hegg, D., Jayaraman, A., Leaitch, R., Murphy, D., Nganga, J., and Pitar, G.: Aerosols, their Direct and Indirect Effects. In Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C. A., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2001.

Pringle, K. J., Tost, H., Message, S., Steil, B., Giannadaki, D., Nenes, A., Fountoukis, C., Stier, P., Vignati, E., and Lelieveld, J.: Description and evaluation of GMXe: a new aerosol submodel for global simulations (v1), Geosci. Model Dev., 3, 391–412, doi:10.5194/gmd-3-391-2010, 2010.

Rougier, J.C. and Sexton, D.M.H.: Inference in ensemble experiments, Philosophical Transactions of the Royal Society, Series A, 365, 2133–2143, 2007.

Rougier, J., Sexton, D., Murphy, J., and Stainforth, D.: Analysing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments, J. Climate, 22, 3540–3557, 2009.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P.: Design and analysis of computer experiments, Statistical Science, 4, 409–435, 1989.

Saltelli, A. and Annonia, P.: How to avoid a perfunctory sensitivity analysis, Environ. Modell. Softw., 25(12), 1508–1517, 2010.

Saltelli, A., Chan, K., and Scott, M. E.: Sensitivity Analysis, New York, Wiley, 2000.

Sanderson, B., Knutti, R., Aina, T., Christensen, C., Faull, N., Frame, D., Ingram, W., Piani, C., Stainforth, D., Stone, D., and Allen, M.: Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes, J. Climate, 21, 2384–2400, 2008.

Santner, T. J., Williams, B., and Notz, W.: The Design and Analysis of Computer Experiments, New York, Springer-Verlag, 2003.

Schimel, D., Alves, D., Enting, I., Heimann, M., Joos, F., Raynaud, D., Wigley, T., Prather, M., Derwent, R., Ehhalt, D., Fraser, P., Sanhueza, E., Zhou, X., Jonas, P., Charlson, R., Rodhe, H., Sadasivan, S., Shine, K., Fouquart, Y., Ramaswamy, V., Solomon, S., Srinivasan, J., Albritton, D., Derwent, R., Isaksen, I., Lal, M., and Wuebbles, D.: Radiative forcing of climate change, in: Climate Change 1996, Contribution of Working Group I to the 2nd Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenberg, A., and Maskell, K., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1996.

Spracklen, D. V., Pringle, K. J., Carslaw, K. S., Chipperfield, M. P., and Mann, G. W.: A global off-line model of size-resolved aerosol microphysics: I. Model development and prediction of aerosol properties, Atmos. Chem. Phys., 5, 2227–2252, doi:10.5194/acp-5-2227-2005, 2005a.

Spracklen, D. V., Pringle, K. J., Carslaw, K. S., Chipperfield, M. P., and Mann, G. W.: A global off-line model of size-resolved aerosol microphysics: II. Identification of key uncertainties, Atmos. Chem. Phys., 5, 3233–3250, doi:10.5194/acp-5-3233-2005, 2005b.

Spracklen, D. V., Carslaw, K. S., Kulmala, M., Kerminen, V.-M., Mann, G. W., and Sihto, S.-L.: The contribution of boundary layer nucleation events to total particle concentrations on regional and global scales, Atmos. Chem. Phys., 6, 5631–5648, doi:10.5194/acp-6-5631-2006, 2006.

Spracklen, D. V., Carslaw, K. S., Kulmala, M., Kerminen, V.-M., Sihto, S.-L., Riipinen, I. Merikanto, J., Mann, G. W., Chipperfield, M. P., Wiedensohler, A., Birmili, W., and Lihavainen, H.: Contribution of particle formation to global cloud condensation nuclei concentrations, Geophys. Res. Lett., 35, L06808, doi:10.1029/2007GL033038, 2008.

Spracklen, D. V., Carslaw, K. S., Merikanto, J., Mann, G. W., Reddington, C. L., Pickering, S., Ogren, J. A., Andrews, E., Baltensperger, U., Weingartner, E., Boy, M., Kulmala, M., Laakso, L., Lihavainen, H., Kivekäs, N., Komppula, M., Mihalopoulos, N., Kouvarakis, G., Jennings, S. G., O'Dowd, C., Birmili, W., Wiedensohler, A., Weller, R., Gras, J., Laj, P., Sellegri, K., Bonn, B., Krejci, R., Laaksonen, A., Hamed, A., Minikin, A., Harrison, R. M., Talbot, R., and Sun, J.: Explaining global surface aerosol number concentrations in terms of primary emissions and particle formation, Atmos. Chem. Phys., 10, 4775–4793, doi:10.5194/acp-10-4775-2010, 2010.

Spracklen, D. V., Carslaw, K. S., Pöschl, U., Rap, A., and Forster, P. M.: Global cloud condensation nuclei influenced by carbonaceous combustion aerosol, Atmos. Chem. Phys. Discuss., 11, 6999–7044, doi:10.5194/acpd-11-6999-2011, 2011.

Stier, P., Feichter, J., Kinne, S., Kloster, S., Vignati, E., Wilson, J., Ganzeveld, L., Tegen, I., Werner, M., Balkanski, Y., Schulz, M., Boucher, O., Minikin, A., and Petzold, A.: The aerosol-climate model ECHAM5-HAM, Atmos. Chem. Phys., 5, 1125–1156, doi:10.5194/acp-5-1125-2005, 2005.

Stockwell, D. and Chipperfield, M.: The TOMCAT Offline Transport Model. Part III: Convection and vertical diffusion, Tech. Rep. 44c, UGAMP Internal Report, 1996.

Textor, C., Schulz, M., Guibert, S., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Easter, R., Feichter, H., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Horowitz, L., Huang, P., Isaksen, I., Iversen, I., Kloster, S., Koch, D., Kirkevåg, A., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Liu, X., Montanaro, V., Myhre, G., Penner, J., Pitari, G., Reddy, S., Seland, Ø., Stier, P., Takemura, T., and Tie, X.: Analysis and quantification of the diversities of aerosol life cycles within AeroCom, Atmos. Chem. Phys., 6, 1777–1813, doi:10.5194/acp-6-1777-2006, 2006.

Textor, C., Schulz, M., Guibert, S., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Feichter, J., Fillmore, D., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Horowitz, L., Huang, P., Isaksen, I. S. A., Iversen, T., Kloster, S., Koch, D., Kirkevåg, A., Kristjans-

son, J. E., Krol, M., Lauer, A., Lamarque, J. F., Liu, X., Montanaro, V., Myhre, G., Penner, J. E., Pitari, G., Reddy, M. S., Seland, Ø., Stier, P., Takemura, T., and Tie, X.: The effect of harmonized emissions on aerosol properties in global models – an AeroCom experiment, Atmos. Chem. Phys., 7, 4489–4501, doi:10.5194/acp-7-4489-2007, 2007.

Vignati, E., Karl, M., Krol, M., Wilson, J., Stier, P., and Cavalli, F.: Sources of uncertainties in modelling black carbon at the global scale, Atmos. Chem. Phys., 10, 2595–2611, doi:10.5194/acp-10-2595-2010, 2010.

Whitby, E. and McMurry, P.: Modal aerosol dynamics modeling, Aerosol. Sci. Tech., 27, 673–688, 1997.

Woodhouse, M. T., Mann, G. W., and Carslaw, K. S.: The impact of oceanic iron fertilisation on cloud condensation nuclei, Atmos. Environ., 42, 5728–5730, 2008.

Yokohata, T., Webb, M., Collins, M., Williams, K., Yoshimori, M., Hargreaves, J., and Annan, J.: Structural similarities and differences in climate responses to CO_2 increase between two perturbed physics ensembles, J. Climate, 23, 1392–1410, 2010.