



# Robust observational constraint of uncertain aerosol processes and emissions in a climate model and the effect on aerosol radiative forcing

Jill S. Johnson<sup>1</sup>, Leighton A. Regayre<sup>1</sup>, Masaru Yoshioka<sup>1</sup>, Kirsty J. Pringle<sup>1</sup>, Steven T. Turnock<sup>2</sup>, Jo Browse<sup>3</sup>, David M. H. Sexton<sup>2</sup>, John W. Rostron<sup>2</sup>, Nick A. J. Schutgens<sup>4</sup>, Daniel G. Partridge<sup>5</sup>, Dantong Liu<sup>6,a</sup>, James D. Allan<sup>6,7</sup>, Hugh Coe<sup>6</sup>, Aijun Ding<sup>8</sup>, David D. Cohen<sup>9</sup>, Armand Atanacio<sup>9</sup>, Ville Vakkari<sup>10,11</sup>, Eija Asmi<sup>10</sup>, and Ken S. Carslaw<sup>1</sup>

<sup>1</sup>Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, UK

<sup>2</sup>Met Office Hadley Centre, Exeter, UK

<sup>3</sup>Centre for Geography and Environmental Science, University of Exeter, Penryn, UK

<sup>4</sup>Earth Sciences, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

<sup>5</sup>College for Engineering, Mathematics, and Physical Science, University of Exeter, Exeter, UK

<sup>6</sup>Centre for Atmospheric Sciences, School of Earth and Environmental Sciences, University of Manchester, Manchester, UK

<sup>7</sup>National Centre for Atmospheric Science, University of Manchester, Manchester, UK

<sup>8</sup>Joint International Research Laboratory of Atmospheric and Earth System Sciences (JiRLATEST),

School of Atmospheric Sciences, Nanjing University, Nanjing, China

<sup>9</sup>Centre for Accelerator Science, ANSTO, New Illawarra Rd, Lucas Heights, NSW, Australia

<sup>10</sup>Finnish Meteorological Institute, Helsinki, Finland

<sup>11</sup>Atmospheric Chemistry Research Group, Chemical Resource Beneficiation, North-West University, Potchefstroom, South Africa

<sup>a</sup>now at: Department of Atmospheric Sciences, School of Earth Sciences, Zhejiang University, Hangzhou, Zhejiang, China

**Correspondence:** Jill S. Johnson (j.s.johnson@leeds.ac.uk)

Received: 17 September 2019 – Discussion started: 21 November 2019

Revised: 23 June 2020 – Accepted: 28 June 2020 – Published: 13 August 2020

**Abstract.** The effect of observational constraint on the ranges of uncertain physical and chemical process parameters was explored in a global aerosol–climate model. The study uses 1 million variants of the Hadley Centre General Environment Model version 3 (HadGEM3) that sample 26 sources of uncertainty, together with over 9000 monthly aggregated grid-box measurements of aerosol optical depth, PM<sub>2.5</sub>, particle number concentrations, sulfate and organic mass concentrations. Despite many compensating effects in the model, the procedure constrains the probability distributions of parameters related to secondary organic aerosol, anthropogenic SO<sub>2</sub> emissions, residential emissions, sea spray emissions, dry deposition rates of SO<sub>2</sub> and aerosols, new particle formation, cloud droplet pH and the diameter of primary combustion particles. Observational constraint rules

out nearly 98 % of the model variants. On constraint, the  $\pm 1\sigma$  (standard deviation) range of global annual mean direct radiative forcing (RF<sub>ari</sub>) is reduced by 33 % to  $-0.14$  to  $-0.26$  W m<sup>-2</sup>, and the 95 % credible interval (CI) is reduced by 34 % to  $-0.1$  to  $-0.32$  W m<sup>-2</sup>. For the global annual mean aerosol–cloud radiative forcing, RF<sub>aci</sub>, the  $\pm 1\sigma$  range is reduced by 7 % to  $-1.66$  to  $-2.48$  W m<sup>-2</sup>, and the 95 % CI by 6 % to  $-1.28$  to  $-2.88$  W m<sup>-2</sup>. The tightness of the constraint is limited by parameter cancellation effects (model equifinality) as well as the large and poorly defined “representativeness error” associated with comparing point measurements with a global model. The constraint could also be narrowed if model structural errors that prevent simultaneous agreement with different measurement types in multiple locations and seasons could be improved. For example, con-

straints using either sulfate or  $\text{PM}_{2.5}$  measurements individually result in  $\text{RF}_{\text{ari}} \pm 1\sigma$  ranges that only just overlap, which shows that emergent constraints based on one measurement type may be overconfident.

## 1 Introduction

Global model simulations of aerosols and their climatic effects are very uncertain. Different global aerosol models have large spread in their simulations of aerosol microphysics, radiation and forcing (Mann et al., 2014; Myhre et al., 2013; Shindell et al., 2013; Tsigaridis et al., 2014). This multi-model spread can be due to different model structures, missing processes, parameter settings, algorithms or coding errors. Individual climate models are also very uncertain because the values of parameters related to physical processes and emissions are often poorly defined (Johnson et al., 2018; L. A. Lee et al., 2011, 2013; Regayre et al., 2018). The uncertainty in the aerosol effective radiative forcing (ERF) over the industrial period caused by aerosol processes, physical atmosphere model processes and emissions could be as large as the multi-model spread (Johnson et al., 2018; Regayre et al., 2014, 2018).

There are two main methods to reduce model uncertainty, often called bottom-up and top-down approaches. The bottom-up approach involves improving the representation of model processes and refining estimates of the associated parameter values through experiment and theory. This approach is necessary to improve model fidelity, but it does not provide an estimate of the model uncertainty, and the uncertainty may grow if the increase in model complexity requires a large number of new and poorly defined parameters. To reduce model uncertainty, bottom-up model development needs to be combined with top-down approaches in which numerous uncertain process-related parameters and emissions are adjusted to improve the agreement of models with measurements.

The difficulty with top-down model adjustments (in its simplest form, model tuning) is that the uncertainty stems from large combinations of uncertain model input parameters. This means that the adjustment of small sets of parameters to improve model agreement with measurements will not produce robust results (Carslaw et al., 2018). For example, a model simulation of particle concentrations could be improved by adjusting particle formation rates, but many other combinations of parameters related to emissions, chemistry or deposition might be able to achieve similar model skill (Carslaw et al., 2013b). Models that are narrowly tuned in this way can therefore produce a wide range of results when used to make predictions outside the range of conditions under which they were tuned. This is likely to be a cause of the large uncertainty in aerosol radiative forcing, which is a predicted rather than observable quantity.

If other aerosol–climate models are comparable with our own model, then they contain at least 20 important uncertain parameters related to emissions and processes, although fewer than about 10 parameters will dominate the uncertainty in a particular model variable in any one environment and time of year (Lee et al., 2016; Regayre et al., 2014, 2018). Therefore, to define and reduce the model uncertainty, it is necessary to find from within 10 dimensions of parameter space all the parameter combinations that produce plausible agreement with different aerosol properties observed across all seasons and global environments. A single well-configured version of a model produced by parameter tuning tells us nothing about the combinations of parameter values that can achieve consistency with measurements within their uncertainty range, nor does it tell us anything about the model output uncertainty.

In this paper, we address the following question: to what extent do extensive and diverse aerosol measurements enable the plausible range of model parameters to be constrained if the full range of their compensating effects is accounted for? By “constrain”, we mean a narrowing of the probability distribution of a parameter (and potentially the absolute range) compared to the uncertainty range that was assumed when the model was built. We also quantify how the identification of observationally plausible parameter ranges feeds through to a reduction in the uncertainty in predictions of aerosol radiative forcing over the industrial period. The study focuses on model constraint using measurements of aerosol properties rather than cloud properties; therefore, we emphasise the effect on aerosol–radiation interaction forcing rather than aerosol–cloud interaction.

## 2 Methods

### 2.1 The HadGEM3-UKCA climate model

We use the Global Atmosphere 4 (GA 4.0; Walters et al., 2014) configuration of the Hadley Centre General Environment Model version 3 (HadGEM3) (Hewitt et al., 2011), which incorporates the United Kingdom Chemistry and Aerosol (UKCA) model at version 8.4 of the UK Met Office’s Unified Model (UM). UKCA simulates trace gas chemistry and the evolution of the aerosol particle size distribution and chemical composition using the GLObal Model of Aerosol Processes (GLOMAP-mode; Mann et al., 2010) and a whole-atmosphere chemistry scheme (Morgenstern et al., 2009; O’Connor et al., 2014). The model has a horizontal resolution of  $1.25 \times 1.875^\circ$  and 85 vertical levels.

The aerosol size distribution is defined by seven log-normal modes: one soluble nucleation mode as well as soluble and insoluble Aitken, accumulation and coarse modes. The aerosol chemical components are sulfate, sea salt, black carbon (BC), organic carbon (OC) and dust. The model does not include any representation of nitrate aerosols. Secondary

organic aerosol (SOA) material is produced from the first-stage oxidation products of biogenic monoterpenes under the assumption of zero vapour pressure. SOA is combined with primary particulate organic matter after kinetic condensation.

GLOMAP simulates new particle formation, coagulation, gas-to-particle transfer, cloud processing and deposition of gases and aerosols. The activation of aerosols into cloud droplets is calculated using globally prescribed distributions of subgrid vertical velocities (West et al., 2014) and the removal of cloud droplets by autoconversion to rain is calculated by the host model. Aerosols are also removed by impactation scavenging of falling raindrops according to the parameterisation of clouds and precipitation collocation (Boutle et al., 2014; Lebsock et al., 2013). Aerosol water uptake efficiency is determined by  $\kappa$ -Kohler theory (Petters and Kreidenweis, 2007) using composition-dependent hygroscopicity factors.

Anthropogenic emission scenarios prepared for the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP) and prescribed in some of the CMIP phase 5 experiments are used here. Biomass burning emissions for recent decades were prescribed using a 10-year average of 2002 to 2011 monthly mean data from the Global Fire and Emissions Database (GFED3; van der Werf et al., 2010) and according to Lamarque et al. (2010) for 1850. Volcanic SO<sub>2</sub> emissions are prescribed in the model by combining emissions from the Andres and Kasgnoc (1998) dataset for continuously erupting and sporadically erupting volcanoes and the Halmer et al. (2002) dataset for explosive volcanoes.

A full description of the set-up for our model simulations can be found in Yoshioka et al. (2019), which we summarise here. The base model simulation was subject to a multi-year spin-up period. Parameter perturbations were then applied distinctly to individual ensemble members (which branch from the base model) and spun up for a further month. We then ran each simulation for a further 12 months to produce the data used here. Horizontal winds and temperatures in the simulations are nudged towards European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalyses for 2008 between approximately 1.2 and 80 km using a 6 h relaxation timescale. Nudging means that pairs of simulations have identical synoptic-scale features, which enables the effects of perturbations to aerosol and chemical processes to be quantified using single-year simulations, although the magnitude of forcing will vary with the chosen year (Fiedler et al., 2019; Yoshioka et al., 2019).

## 2.2 Creation of perturbed parameter model variants

Our method to determine observational constraint on the model parameters and radiative forcings involves producing a very large set of “model variants”, each with a different combination of parameter values, and then ruling out model variants for which a set of model outputs are judged to be

implausible against measurements (see Sect. 2.4). The model variants were generated using a perturbed parameter ensemble (PPE) of 235 model simulations of HadGEM3-UKCA (the “AER PPE” detailed in Yoshioka et al., 2019) that samples 26 sources of uncertainty in the aerosol model (Carslaw et al., 2017; Yoshioka et al., 2019); see Table A1 in Appendix A.

A set of 235 simulations alone is much too small to allow statistical analysis of model performance across 26 dimensions of parameter space. We therefore built Gaussian process emulators (surrogate models) using the PPE simulations as training data (L. A. Lee et al., 2011), which define how the model outputs vary continuously over the 26-dimensional parameter space and enable dense sampling over parameter uncertainty. Separate emulators were built describing the monthly mean value of each model output in each model grid cell. We then used Monte Carlo sampling from these emulators to produce output for a set of 1 million model variants (parameter input combinations). Uniform distributions were assumed for each parameter in this sampling. The emulator is not a perfect representation of a model output, but its uncertainty can be estimated and accounted for in the model–measurement comparison. In the rest of this paper, we refer to the emulator-derived values of model outputs at each sampled 26 d input combination as a “model variant”.

The AER PPE samples only uncertainties in the aerosol component of the model and the radiative forcing does not account for atmospheric and cloud adjustments; i.e. it is a radiative forcing (RF) rather than an effective radiative forcing, which we analysed in previous papers (Johnson et al., 2018; Regayre et al., 2018). The prior (unconstrained) 95 % credible interval (CI) of global mean aerosol RF is  $-2.23 \pm 0.94 \text{ W m}^{-2}$ . However, because of the way that multiple parameters compensate (Lee et al., 2016; Regayre et al., 2018), the forcing uncertainty in this PPE is similar to the aerosol–atmosphere (AER-ATM) PPE in which additional physical atmosphere model parameters were perturbed and cloud adjustments accounted for (Yoshioka et al., 2019). Because the AER PPE analysed here samples only aerosol uncertainties, we restrict the constraints to measurements of aerosol properties. In future work, we will extend the analysis to radiation, precipitation and cloud measurements that are relevant to the wider range of parameters in the AER-ATM PPE.

The choice of the 26 perturbed parameters and their uncertainty ranges were defined using expert elicitation (Yoshioka et al., 2019). The parameters (Table A1; full descriptions are given in Yoshioka et al., 2019) relate to natural and anthropogenic emission fluxes of aerosol precursor gases and primary particles, the properties of primary particles (size), aerosol processes, aerosol hygroscopicity, removal rates and cloud droplet formation (updraft speed). The list of parameters is not exhaustive, but one-at-a-time parameter perturbation tests were used to show that any other parameters have a smaller effect regionally and globally in our model than

the set we chose. Finally, we note that the evaluated uncertainty in global annual mean RF in this study differs from that shown in Yoshioka et al. (2019) as we have used uniform parameter distributions when sampling over the parameter uncertainty space, while elicited parameter distributions were used in Yoshioka et al. (2019). Our choice to use uniform distributions here means that the constraint can be fully attributed to the model–measurement comparison.

### 2.3 Measurements

We use aerosol measurements from ground stations, ship campaigns and aircraft campaigns covering the following aerosol properties: aerosol optical depth (AOD),  $\text{PM}_{2.5}$  concentrations, sulfate mass concentrations, organic carbon mass concentrations and number concentrations of particles larger than 3 nm dry diameter ( $N_3$ ) and 50 nm dry diameter ( $N_{50}$ ); see Appendix B and Table S1 in the Supplement. All measurements used are from within the boundary layer, which we define to be at an atmospheric pressure greater than 800 hPa. We do not attempt to constrain aerosol properties above the boundary layer.

The measurements were all made at specific locations and times (i.e. they are “point measurements”) in the period from October 1995 to December 2015, and we use measurements from all years within this period regardless of whether the year of the measurement matches the year of the PPE model simulations. (We take account of the interannual differences by incorporating an error term in the constraint process; see Sect. 2.4.) The measurements were aggregated to monthly mean values in grid cells of size  $2.50^\circ$  longitude by  $3.75^\circ$  latitude (four model grid boxes of the N96 model grid). In cases where there is more than one measurement in a model grid cell, the observed values were averaged. This processing resulted in 9464 monthly aggregated grid-box measurements (over six aerosol properties and 12 months). Figure 1 shows the global spatial coverage of the gridded measurements for each aerosol property, along with the monthly temporal coverage for each measurement, which is indicated by the colour scale. Table 1 shows the breakdown of the number of grid-box measurements by variable and month.

The AOD data are level-2.0 (quality assured) monthly mean data at 440 nm wavelength from the AERONET (Aerosol Robotic Network) network (Giles et al., 2019; Holben et al., 1998). Our dataset includes an average of 312 aggregated grid-box measurements for comparison in each month. Figure 1 shows that the measurements are well distributed across all continental regions except Antarctica. The coverage at high northern latitudes is relatively sparse, and there are only a small number of island measurement that are representative of marine aerosol environments. The temporal coverage is very good, with the majority of stations providing measurements in all months of the year.

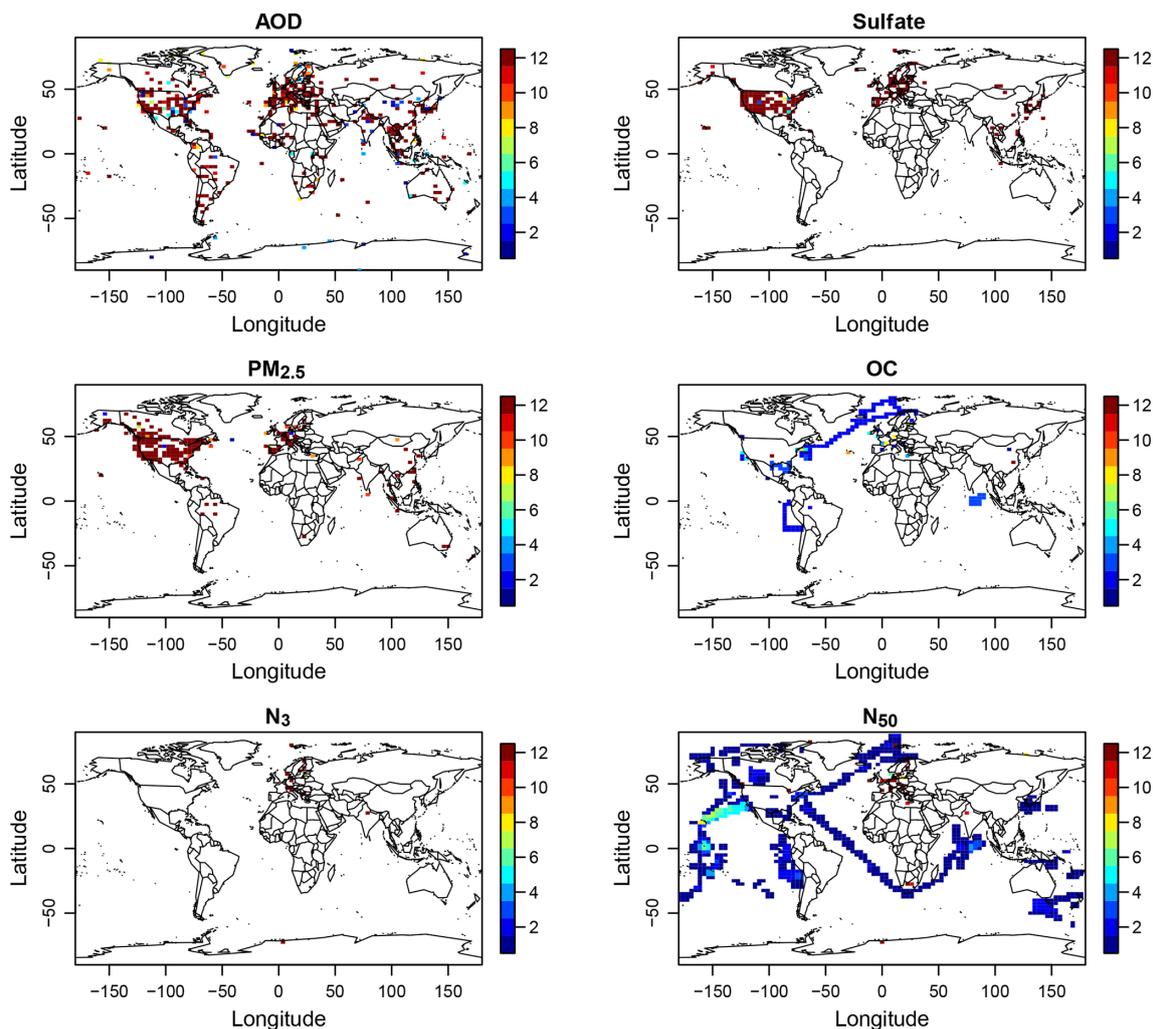
The  $\text{PM}_{2.5}$  and sulfate concentration data come from multiple large networks. The sulfate concentration data are

**Table 1.** The number of monthly aggregated grid-box measurements for each variable in each month. The total number over all months and all variables is 9464.

	AOD	Sulfate	$\text{PM}_{2.5}$	OC	$N_3$	$N_{50}$
Jan	294	149	168	6	13	77
Feb	301	148	168	14	13	90
Mar	309	151	170	82	13	148
Apr	316	151	170	74	12	199
May	322	149	167	23	12	64
Jun	320	150	170	23	12	96
Jul	323	148	172	23	13	115
Aug	326	148	169	23	13	109
Sep	321	147	166	22	13	133
Oct	315	147	165	41	13	119
Nov	309	146	168	37	13	155
Dec	298	147	169	15	12	67
Total	3754	1781	2022	383	152	1372

from the Interagency Monitoring of Protected Visual Environments (IMPROVE) network (USA), the European Monitoring and Evaluation Programme (EMEP) network and the Acid Deposition Monitoring Network in East Asia (EANET). For  $\text{PM}_{2.5}$ , we use data from the IMPROVE network, the World Data Centre for Aerosols (WDCA) (European sites), the Asia-Pacific Aerosol Database (A-PAD) and the Canadian National Air Pollution Surveillance Program (NAPS). Other  $\text{PM}_{2.5}$  measurements are included from smaller networks and individual stations in Australia, South America, Taiwan and South Africa, as well as sulfate and  $\text{PM}_{2.5}$  data recorded at the Station for Observing Regional Processes of the Earth System (SORPES) in Nanjing, East China. The  $\text{PM}_{2.5}$  data (except for the SORPES site) were obtained, processed and gridded to the N96 model grid as described in Browse et al. (2019). Figure 1 shows that these  $\text{PM}_{2.5}$  and sulfate measurements are highly clustered over polluted land areas of the Northern Hemisphere, mostly in North America and Europe with limited coverage elsewhere, especially in remote and marine areas. Nearly all stations in these datasets have full temporal coverage, leading to approx. 150 and 170 aggregated grid-box measurements for comparison in each month for  $\text{PM}_{2.5}$  and sulfate, respectively (Table 1).

For  $N_{50}$  particle concentrations and OC concentrations, we have a mixture of measurements from a small number of land-based ground stations along with measurements taken over marine environments from ship and aircraft campaigns (see Appendix B and Table S1 in the Supplement). The  $N_{50}$  concentration data were mainly derived from size distribution measurements and gridded to the N96 model grid as described in Browse et al. (2019). The amount of campaign data, and hence global spatial coverage in the gridded data, is greater for  $N_{50}$  than for OC (Fig. 1), and the number of aggregated grid-box measurements is variable between months



**Figure 1.** The distribution of measurements used in the constraint. The colours indicate the number of months covered by the measurements (although the data may not cover all days within a month).

(Table 1). Due to the nature of field campaigns, the temporal coverage is much sparser for these variables, with each campaign only measuring for 1–3 months of the year, shown by the blue colours for the data of these variables in Fig. 1.

The measurement data for  $N_3$  particle concentration have the smallest number of grid-box measurements over the year and spatially is the sparsest dataset included here. The data for this aerosol property come from only 13 ground stations (ACTRIS; Asmi et al., 2013), which are mostly located in Europe, with one in the Arctic, one in Antarctica and one in northern India. The  $N_3$  concentrations at each site were derived directly by integrating size distribution measurements. These data were then averaged over multiple years for each month and location by the authors.

## 2.4 Constraint methodology

We apply the statistical methodology of history matching, which has been applied to complex models in a range of fields, including epidemiological modelling of virus transmission (Andrianakis et al., 2017), risk assessment for oil field developments (Craig et al., 1997), modelling galaxy formation (Rodrigues et al., 2017) and climate modelling (Edwards et al., 2011; McNeill et al., 2016; Williamson et al., 2013). The methodology is described in detail in previous papers (Johnson et al., 2018; Regayre et al., 2018), which built upon our earlier study (Lee et al., 2016). We therefore describe the overall methodology only briefly here but present a full description of the new aspects related to using real measurements rather than “synthetic” measurements (Johnson et al., 2018).

In the comparison of the model and measurements, we account for emulator uncertainty, measurement uncertainty (in-

strument error), representativeness uncertainties (caused by spatial and temporal mismatches in resolution and sampling between model and measurements) and potential structural model uncertainty. The model–measurement difference together with these measures of uncertainty is incorporated into an “implausibility measure” and our model constraint procedure in order to identify implausible parts of parameter space (model variants).

### 2.4.1 Implausibility measure

The implausibility metric  $I(\mathbf{x})$  is calculated for each of the 1 million model variants  $\mathbf{x}$ , for each gridded measurement.  $I(\mathbf{x})$  weights the difference between the model and measurements by the uncertainties associated with the comparison (Craig et al., 1996; Williamson et al., 2013)

$$I(\mathbf{x}) = \frac{|M - O|}{\sqrt{[\text{Var}(M) + \text{Var}(O) + \text{Var}(R) + \text{Var}(S)]}}, \quad (1)$$

where  $M$  is the estimate of model output calculated using the emulator and  $O$  is the observed value (the measurement). In the denominator,  $\text{Var}(M)$  is the variance in the model estimate (associated with replacing the model with the emulator),  $\text{Var}(O)$  is the variance in the measurement (i.e. instrument or retrieval uncertainty),  $\text{Var}(R)$  is the variance associated with the comparison of the model with the measurements, called the representativeness error (Schutgens et al., 2017, 2016a, b), and  $\text{Var}(S)$  is a model structural error term.

A low value of the implausibility metric indicates either the model–measurement difference is small (i.e. the model is skilful) or that the uncertainty in the denominator is large (i.e. we cannot tell whether the model is skilful because the uncertainties are too large). Therefore, the implausibility metric allows model variants to be ruled out if the model–measurement difference is large and we can be confident that it is large.

The representativeness error  $\text{Var}(R)$  has three components.  $\text{Var}(R_{\text{sp}})$  (sp: spatial) accounts for uncertainty associated with spatial variability below the grid scale of the model, which means that a point measurement may not be representative of the grid-box mean (Schutgens et al., 2016b).  $\text{Var}(R_{\text{temp}})$  (temp: temporal) accounts for the temporal sampling of a measurement, which may not match the temporal sampling of the model (e.g. a ship track through the grid box over a short time period which is compared with a monthly mean model value (Schutgens et al., 2016a).  $\text{Var}(R_{\text{iav}})$  (iav: interannual variability) accounts for the fact that we sometimes match measurements and the model for the correct calendar month but not for the correct year. This is necessary in cases where we use measurements from years for which we have not run the model. We assume that

$$\text{Var}(R) = \text{Var}(R_{\text{sp}}) + \text{Var}(R_{\text{temp}}) + \text{Var}(R_{\text{iav}}). \quad (2)$$

The magnitude of these errors is discussed in Sect. 2.4.2.

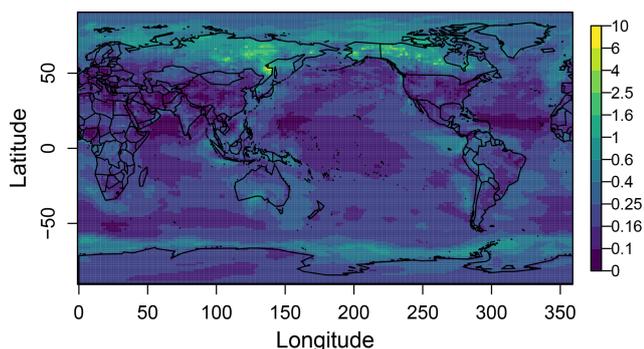
The structural error term  $\text{Var}(S)$  has been included in previous studies using the implausibility metric. It is intended to represent an estimate of the potential structural error in the model. Practically, however, we have no way to estimate this term for all variables at all times and geographical locations. We therefore set it to zero and instead use very large values of implausibility to point us towards potential structural errors in the model–observation comparison and constraint procedure, as described in Sect. 2.4.3.

### 2.4.2 Estimation of uncertainty terms

Our estimates of the uncertainty terms in Eq. (1) are preliminary and are designed to test our approach. We discuss in the conclusions the need to refine our understanding of these uncertainty terms.

For all aerosol properties, we assume an instrument uncertainty of 10 %, a spatial co-location uncertainty of 20 % and a temporal sampling uncertainty of 10 % on the measured value. The spatial sampling uncertainty for monthly mean aerosol properties is estimated based on Schutgens et al. (2017, 2016b). These studies examined a typical spatially heterogeneous continental environment where the sampling error is dominated mainly by local aerosol sources that are not resolved by the global model. The magnitude of uncertainty is likely to vary globally (especially between land and ocean), with surface measurements typically having larger errors than column measurements and the magnitude of error also depending on the location of a ground site with respect to the grid-box centre, but we do not account for these variations. We base our estimate of the temporal sampling uncertainty on Schutgens et al. (2016a), who quantified the error associated with the different temporal sampling of models and measurements (daily measurements or temporally sporadic measurements versus monthly mean model, etc.). The emulator uncertainty is taken from the Gaussian error on the emulator mean prediction, which is known for every parameter combination (i.e. each of the 1 million model variants).

The interannual uncertainty was defined to be the standard deviation of monthly mean aerosol properties in each grid cell over a 30-year period. We take information from an analysis of the trend and variation of gridded aerosol properties in a HadGEM3-UKCA hindcast simulation over the period of 1980–2009 (Turnock et al., 2015). For each month and grid box, the monthly mean output of the aerosol variable of interest for each year of the simulation was obtained. These values were de-trended using linear regression and the resulting residuals were then analysed. We use a relative measure of monthly mean uncertainty defined by the standard deviation of these residuals divided by the de-trended mean. As an example, Fig. 2 shows the relative standard deviation for the surface-level  $N_{50}$  concentration in July.



**Figure 2.** The relative standard deviation for surface-level  $N_{50}$  concentration in July, used in the estimation of the interannual variability component of representativeness error  $R_{\text{iaV}}$ .

### 2.4.3 Methodology for ruling out observationally implausible parts of parameter space

There is an element of subjectivity involved in comparing a model with point measurements and reaching a conclusion about the fidelity of the model. The comparison may indicate either (a) the model seems to be structurally adequate, but the parameters need to be adjusted to optimise agreement, or (b) the model is structurally deficient (i.e. there are missing or incorrect process representations in the model). Structural deficiencies may be apparent, for example, because the model skill is particularly poor in one region or at one time of year, or it is not possible to obtain good skill across multiple variables simultaneously.

Our use of 1 million model variants and more than 9000 monthly aggregated grid-box measurements means that we need to automate the model–measurement comparison processes and detection of potential structural errors while also using the measurements to rule out implausible parts of parameter space. The difficulties for us in detecting structural errors are as follows: (a) we cannot inspect each of the 1 million model variants individually, so we need to rely on summary statistics; (b) many of the aerosol point measurements are spatially and temporally sparse, so we cannot easily detect spatial and temporal changes in model skill that might indicate structural error; (c) the measurements do not have the same spatial distribution in all months (because of brief, localised field campaigns) so spatial–temporal biases are hard to detect; (d) the uncertainty in each measurement (particularly the representativeness error, Sect. 2.4.2) is spatially and temporally heterogeneous and often very poorly defined.

Our approach is summarised in Fig. 3. It is designed to rule out implausible parts of parameter space while avoiding doing so in cases where the biases shared by many model variants could be caused by structural errors in the model.

The steps are as follows:

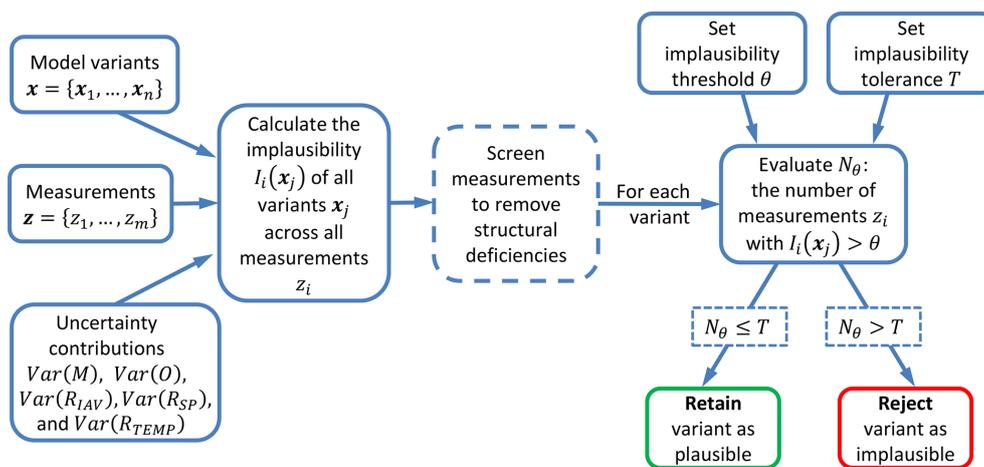
1. The implausibility is quantified for each of the 1 million variants across all measurements of a single type

in a particular month. Figure 4 shows an example for the measurements of  $N_{50}$  in July. For each measurement (numbered on the horizontal axis in Fig. 4a), the distribution of the implausibility over the variants is shown by the bar representing the 95 % credible interval.

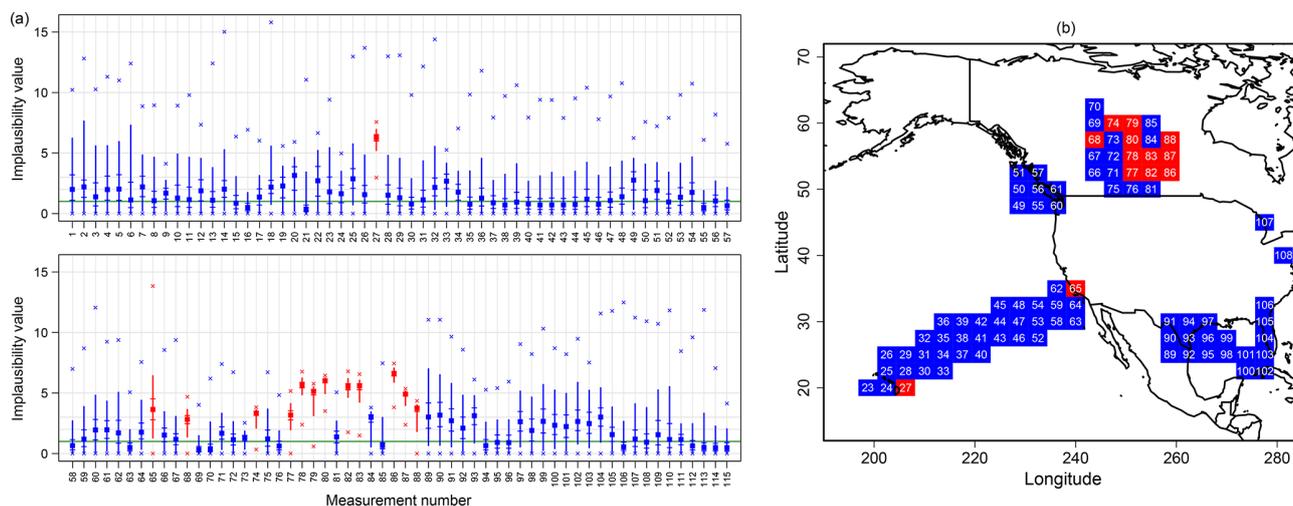
2. Measurements are identified for which 97.5 % of the model variants have an implausibility  $I > 1$ . These measurements are excluded from the constraint procedure (shown in red in Fig. 4). We assume that this large implausibility for the significant majority of variants indicates that either there is a structural error in the model or that the model is unable to represent these point measurements because of its low spatial and temporal resolution. An alternative explanation is a mismatch in the model’s meteorological year to the year of the measurement (Sect. 2.4.1). We flag these measurements for further investigation of potential structural errors or underestimated error terms (these are not examined further in this study).
3. Using all other measurements (where more model variants have lower implausibility, shown in blue in Fig. 4), we use the implausibility metric values to decide whether to rule each variant out as implausible or retain it as plausible. If we ruled out all model variants with high implausibility for each measurement in turn (treating the measurements independently, as in many emergent constraint studies), we could end up ruling out all parts of parameter space. Our criterion is therefore to rule out a model variant if more than a defined fraction (or number) of the measurements (tolerance  $T$ ) exceeds a defined implausibility threshold ( $\theta$ ). For example, we might rule out a model variant if more than 20 % of measurements exceed an implausibility of 3.5 (i.e. bias is 3.5 times the expected error).

We apply this approach to the set of measurements for each variable (measurement type) in each month and then combine the constraints to a joint constraint over months and/or over variables such that if a variant is ruled out for any single month/variable combination, then it is also ruled out in the joint constraint. This method allows us to identify the set of model variants that capably represent measurements of a range of variables and across multiple locations and seasons. We extensively explored various choices of the tolerance and threshold values in each variable/month case and found that the final constrained parameter ranges were reasonably robust, except when the number of measurements was small.

Our choices of the threshold and tolerance for each measurement type are given in Table A2 in Appendix A. A wide range of values were tested in each case, starting with a set threshold of  $\theta = 3.5$  and iterating through increasing tolerances  $T$  up to a maximum of  $T = 33\%$  (1/3 of the measurements), before further increasing  $\theta$  by 0.5 (to a maximum of  $\theta = 4.5$ ) and re-iterating over  $T$  in order to retain (approx-



**Figure 3.** Flowchart detailing the process followed for each model variant  $x_j$ , in using the calculated implausibility over a set of  $m$  measurements  $z = \{z_1, z_2, \dots, z_m\}$  (for a single output variable  $y$ ) simultaneously to constrain the model uncertainty.



**Figure 4.** (a) The distribution of implausibility calculated over the 1 million model variants for each measurement in the July  $N_{50}$  concentration set, shown vertically. For each measurement, numbered along the  $x$  axis, the range of the implausibility distribution is shown by the outer crosses, the bar corresponds to the 95 % credible interval (2.5 % to 97.5 % empirical quantiles), the horizontal markers through the bar show the interquartile range, and the square point is the median implausibility. Here, we assume no structural error term in the implausibility calculations and use the implausibility distribution to identify potential structural errors. Measurements coloured red are ruled out as potential structural error cases (as the lower 95 % credible interval bound is  $> 1$ ), and those coloured blue are retained and used in our constraint procedure. (b) Corresponding map to show the locations of the rejected July  $N_{50}$  measurements (red) and those retained for constraint (blue), over the North Pacific and North American region (outside this region, all measurements were retained). We hypothesise that the red points over the Pacific correspond to ports with localised pollution sources, while the red points over Canada correspond to localised fire emissions that are not represented at the resolution of the model.

imately) a chosen percentage of model variants. Approximately the same percentage of variants was attained for all months of a variable type and combined for an “all-months” constraint. Our final choices for each variable type on its own (left column in Table A2) were relaxed for the joint “all-variables-months” constraint (retaining a larger percentage of variants in each month for each variable, so a weaker constraint; right column in Table A2), in order to retain a reason-

able number of model variants and avoid overconstraining on any one observational type.

Our assumption of zero structural error ( $\text{Var}(S) = 0$ ) in the implausibility calculations means that structural errors in the model can easily come to light in our constraint process. This occurs either in the calculated implausibility values for a measurement (where large values are consistently produced over the 1 million variants covering the model uncertainty, indicating a large model–measurement discrep-

ancy, e.g. Fig. 4) or when bringing together the constraint effects of different sets/types of measurements (where very few, if any, model variants that lead to plausible model output in all cases/measurements types simultaneously can be identified and retained). Even though we do not directly account for structural errors in the implausibility measure itself, our constraint approach offsets the effects of such errors on the achieved constraint as much as possible. This is accomplished by screening out observations with large model–measurement discrepancies from the constraint process (step 2; Fig. 4) and by relaxing the constraint criteria for the joint all-variables-months constraint. Through this approach, we are able to produce as robust a constraint as possible, given the limitations we have in specifying structural and representational errors.

## 2.5 Interpretation of constrained parameter probability distributions

Observationally plausible parts of parameter space exist in 26 dimensions. We show the results as one-dimensional marginal probability distributions, which are one-dimensional projections of the 26-dimensional parameter probability distribution. Figure 5 shows an idealised representation for a two-dimensional parameter constraint. The white parts of the joint distribution are ruled out, leaving the shaded region of joint parameter space as observationally plausible. The effect on the marginal probability distribution of parameter 1 is to entirely rule out the lowest and highest values (i.e. there is no combination of these values of parameter 1 with parameter 2 that produces an observationally plausible model). Where some values of parameter 1 are ruled out over the range of parameter 2, the likelihood of parameter 1 having those values is reduced.

In the results below, the parameter probability distributions therefore reflect the relative likelihood of the parameter having particular values, with lower probabilities indicating that there are fewer ways in which the parameter can be combined with the other 25 parameters to produce a plausible model. For conciseness in the results section, we say, for example, that “a measurement constrains the parameter to low values”, which means that we retain a larger proportion of model variants with low values.

Figure 5 also shows the separate and joint effects of two observational constraints. We show this conceptually because it arises in the results. Measurement 1 rules out the lowest values of parameter 1 and suggests that parameter 1 is likely to be at the high end of the sampled range. Conversely, measurement 2 suggests that parameter 1 is likely to be at the low end of the range. However, the correct interpretation of this situation is that intermediate values of the parameter are consistent with both measurements (measurement 1 is consistent with the model for all but the lowest values of the parameter and measurement 2 is consistent for all but the highest values). Only in cases where the two sepa-

rate constrained parameter PDFs do not overlap can we conclude with certainty that there is likely to be a structural deficiency in the model. However, to obtain multivariate constraint, we prevent this happening by screening out measurements with large model–measurement discrepancies and relaxing the constraint criteria with each measurement type.

## 3 Results

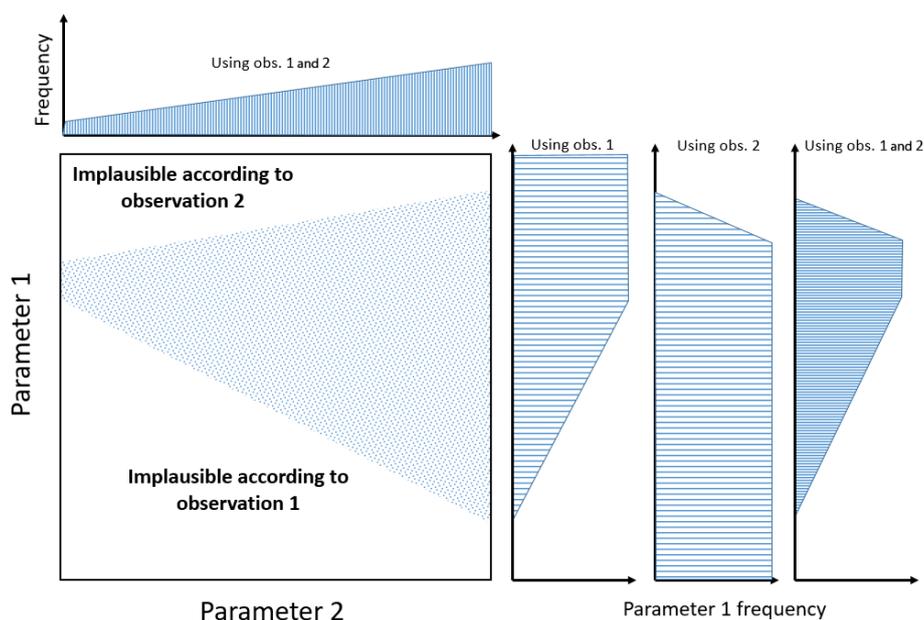
### 3.1 Constraint using individual measurement types

Figure 6 shows the constrained marginal parameter distributions for all parameters based on using individual measurement types (each column on left) and all measurement types together (right column).

*AOD measurements* constrain aerosol and precursor emissions to low values and removal rates to high values. These constraints imply that the PPE produces generally too-high AODs across the sampled parameter space, which is the case (Sect. 3.4). In particular, sea spray emissions higher than about 3.6 times the baseline emissions are ruled out, but emissions down to as low as 0.125 times the baseline emissions are plausible. For anthropogenic SO<sub>2</sub> emissions, the likelihood of the emissions scale factor being below 1 (corresponding to the default value from the inventory) increases from 55 % to 70 % on constraint. For biogenic volatile organic compound (BVOC) emissions, the likelihood of the emissions (or effectively the production of SOA) being more than 3 times the default emission of 46 Tg yr<sup>-1</sup> (= 138 Tg yr<sup>-1</sup>) is reduced from 31 % to 13 %, but all lower values from the default emission down to our lower bound of 37 Tg yr<sup>-1</sup> are equally plausible.

The AOD measurements also constrain the model to low values of other parameters: more variants with higher cloud droplet pH values are ruled out (judged implausible) and as a result cloud droplet pH is nearly 3 times as likely to be below the central value of its range of 5.8 as above it, which is consistent with a higher likelihood of slower production of sulfate aerosol from in-cloud SO<sub>2</sub> oxidation. The hygroscopicity of OC in the particles ( $\kappa_{OC}$ ) is also weakly constrained to low values, which reduces the water content of aerosol and reduces AOD. The rate of aerosol scavenging by precipitating raindrops (the Rain\_Frac parameter) is weakly constrained to high values.

*Sulfate measurements* strongly constrain SO<sub>2</sub> emissions to low values, which is consistent with the AOD constraint. Given this constraint, the SO<sub>2</sub> emissions have a 78 % likelihood of being below the default value from the inventory and the median emission is reduced to 0.78 times the default. Also consistent with the AOD constraint, the deposition rate of accumulation-mode particles is constrained strongly to high values, with an 87 % likelihood of the rate being above the default value. Likewise, the SO<sub>2</sub> dry deposition rate is constrained fairly weakly to higher values, with a 60 % likeli-



**Figure 5.** Schematic of parameter constraint in two dimensions using two measurements.

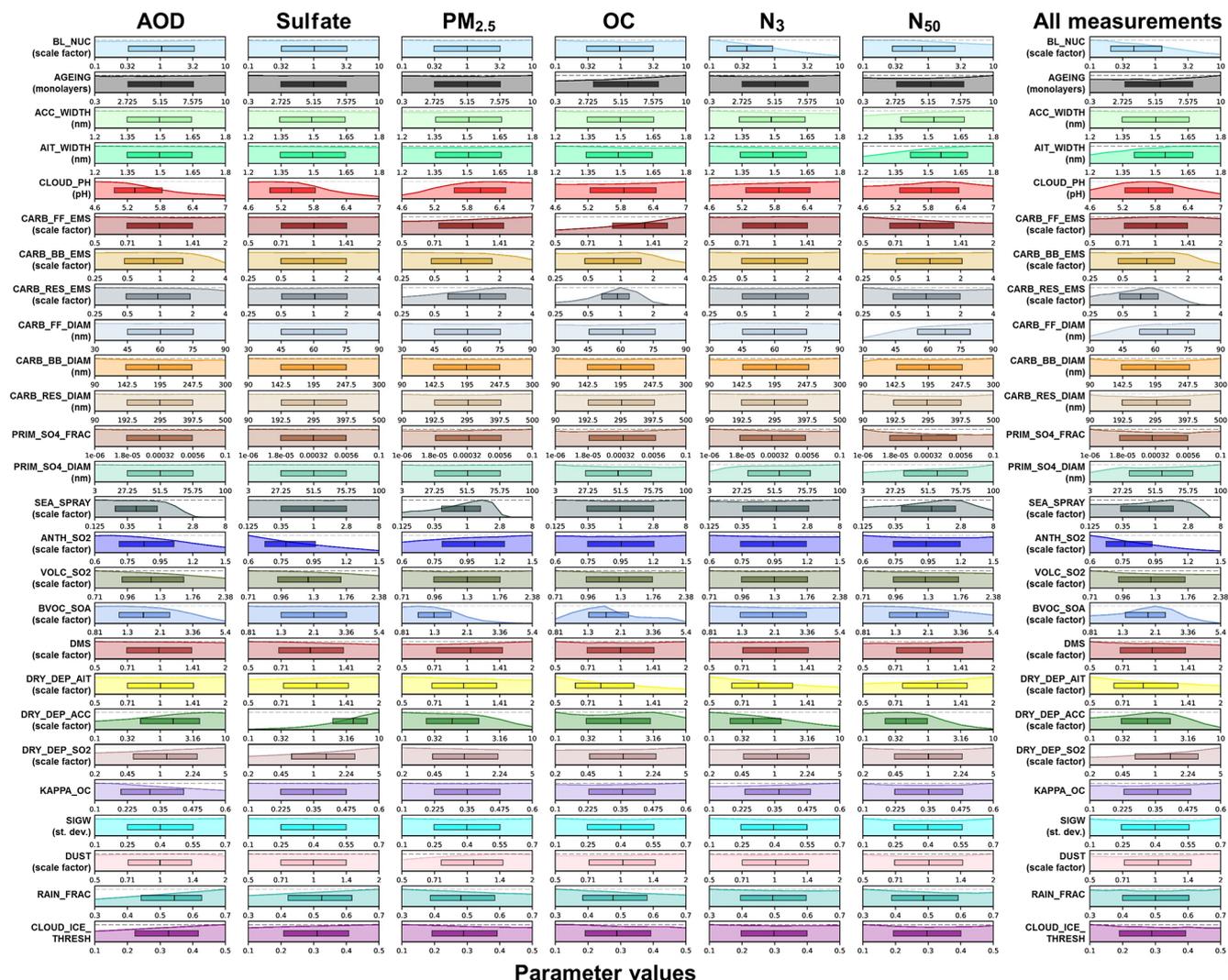
hood of the scaled value being above the default value. Each of these constraints is consistent with too-high sulfate concentrations in many of the sampled model variants across the parameter uncertainty space (Sect. 3.4).

*PM<sub>2.5</sub> measurements* have a similar effect to AOD on some parameters, but for others there are differences. Emissions of sea spray and BVOC emissions are constrained similarly (to low values). However, SO<sub>2</sub> emissions and cloud droplet pH are weakly constrained to higher values and the dry deposition rate of accumulation-mode particles is weakly constrained to low values, opposite to the AOD and sulfate constraints for these parameters. The PM<sub>2.5</sub> measurements also weakly constrain the residential combustion emissions to high values. PM<sub>2.5</sub> and AOD are strongly correlated in the PPE (Johnson et al., 2018), so differences in the constrained parameters most likely reflect differences in the spatial distribution of the measurements (Fig. 1) and how that maps onto the spatial variations in sensitive parameters. As described in Sect. 2.5, these apparently opposing constraints are not necessarily inconsistent: for AOD and PM<sub>2.5</sub>, there may be other parameter settings that can be combined with low SO<sub>2</sub> emissions to achieve agreement with the measurements (so the space is not ruled out).

*OC measurements* strongly constrain the scaled magnitude of residential carbonaceous emissions to a narrower credible interval of about 0.3–1.8 centred near the default value specified in the emissions. Emissions above 2.0 times the default value are effectively ruled out and there is only a 13 % likelihood of the emissions being below half the default value. Fossil fuel emissions have a 70 % likelihood of being above the default emission value. The OC measurements also constrain the scaled BVOC emissions in a similar

way to PM<sub>2.5</sub> and AOD, with scaled emissions above about a factor of 2.1 (97 Tg yr<sup>-1</sup>) having only a 31 % likelihood (compared to 50 % prior to constraint). OC measurements also constrain the lowest values of BVOC emissions, which was not achieved with PM<sub>2.5</sub> and AOD. The likelihood of the scaled emissions being below 1 (46 Tg yr<sup>-1</sup>) is 6 % (compared with 11 %). The dry deposition rate of Aitken-mode particles is constrained to the low part of the range, which will tend to increase OC concentrations in the atmosphere consistent with the constraint of fossil fuel emissions to high values. There is also a weak constraint of the ageing rate towards higher values, which has a 55 % likelihood of being in the upper half of the range. The rate of aerosol scavenging by precipitating raindrops (Rain\_Frac parameter) is constrained similarly but to lower values. Again, although weak, these two constraints imply slower ageing, slower removal rates, longer OC lifetime and higher atmospheric concentrations. Biomass burning emissions are only very weakly constrained towards lower emissions. The lack of constraint on the biomass burning emissions from OC measurements here is likely a result of the limited coverage, if any, of the OC measurements in regions important for biomass burning such as Africa and southeast Asia (Fig. 1).

*Particle concentration (N<sub>3</sub> and N<sub>50</sub>) measurements* constrain a wider range of parameters than the measurements of mass-related properties. The rate of boundary layer nucleation is strongly constrained to the low part of the sampled range by the N<sub>3</sub> measurements (a 77 % likelihood of being below the default rate), suggesting N<sub>3</sub> concentrations are generally too high across the PPE. N<sub>3</sub> also weakly constrains the dry deposition of Aitken- and accumulation-mode particles to low values. Low deposition rates of accumulation-



**Figure 6.** Marginal parameter distributions after constraint using individual measurement types over all months (six columns on the left) and after using all measurement types over all months together (right column). The 25th, 50th and 75th percentiles of each constrained distribution are shown in the central boxes, and the parameter values on the x axes correspond to values as they are used in the model (parameters that are multiplicative scaling factors are shown on the  $\log_{10}$  scale), covering the full parameter ranges (Yoshioka et al., 2019). The corresponding choices of threshold  $\theta$  and tolerance  $T$  that were applied in the constraint process to generate these results are given in Table A2 (left column for each individual measurement type; right column for the joint measurement-type constraint), along with the percentage of model variants that is retained in the constrained sample in each case. See Sect. 2.5 for a definition of marginal parameter distributions.

mode particles (hence higher atmospheric concentrations) will result in a higher condensation sink and more removal of sulfuric acid that participates in particle nucleation, so this is consistent with the constraint of nucleation rates to low values. The constraint of Aitken-mode deposition to low values is less obvious. Aitken-mode particles can contribute substantially to  $N_3$ , so low deposition rates would enhance  $N_3$  (opposite to the constraint on nucleation rates). However, nucleation rates are constrained to very low values, so in such a situation Aitken particles can begin to act as a sink term for nucleation by affecting the condensation sink and by growing into accumulation-mode particles. BVOC emissions are not

constrained by  $N_3$  measurements, even though SOA enters the nucleation rate expression. This is most likely because high BVOC emissions also enhance total SOA, which acts as a condensation sink for nucleation, so the two effects cancel (Carslaw et al., 2013b).

For  $N_{50}$ , the constraints are consistent with shifting the  $N_{50}$  concentrations in the ensemble towards lower values (Sect. 3.4).  $N_{50}$  has very little effect on the range of boundary layer nucleation rate. In contrast, a previous study found that boundary layer nucleation made a statistically significant difference to model skill at about half of the ground sites they analysed (Reddington et al., 2011) – although that

study tested the effect of including or not including boundary layer nucleation rather than perturbing the rate as we do here. Without boundary layer particle formation, the model was structurally deficient and had poor skill at around half the sites analysed. However, our results show that uncertainty in the parameter value itself is unimportant when other parameter uncertainties are considered. This parameter is unconstrained by  $N_{50}$  measurements because there are many alternative ways of achieving model–measurement agreement.

$N_{50}$  measurements also tend to constrain primary particle emissions to the lower end of the range (fossil fuel and primary sulfate emissions), albeit weakly. Residential particle emissions are not constrained, but the measurements we used are not well located to achieve this. It also constrains the emitted particle diameters to the high end of their ranges (fossil fuel, primary sulfate), which is again consistent with low number concentrations (since we perturb emission diameter independently of the mass, so number concentration is affected). The constraint of particle emission sizes is consistent with a previous study that showed cloud condensation nuclei (CCN) concentrations are sensitive to the assumed size (Reddington et al., 2011). Our results show that  $N_{50}$  measurements allow the emission size to be constrained, even though there are many other compensating factors that can affect CCN concentrations.  $N_{50}$  weakly constrains cloud pH to higher values, consistent with greater production of sulfate aerosol and a higher sink for nucleation. BVOC emissions are constrained to the low end, which is consistent with reduced growth of nucleation-mode particles into the Aitken and accumulation modes.  $N_{50}$  also constrains deposition rates: accumulation-mode deposition is constrained to low values and Aitken-mode deposition to high values, suggesting a shift in the aerosol size distribution towards larger aerosols is consistent with  $N_{50}$  measurements.

### 3.2 Seasonal variations in constraint

Many of the parameter constraints vary seasonally, which can be linked to seasonal variations in emissions and parameter sensitivity. Some examples are shown in Fig. 7. Cloud pH is constrained more by AOD in Northern Hemisphere winter (Fig. 7a) when in-cloud oxidation of  $\text{SO}_2$  by ozone dominates sulfate production. BVOCs are constrained by AOD only in Northern Hemisphere summer when the emissions are strong (Fig. 7b). There are several other seasonal variations in the constraint effect from AOD measurements that we do not show. For example, anthropogenic  $\text{SO}_2$  emissions are constrained by AOD more in winter because the AOD uncertainty in summer is dominated by the uncertainty in SOA. The hygroscopicity of OC is also constrained more in summer when OC is a larger component of the aerosol. Biomass burning emissions are constrained in NH summer as expected from wildfire emission seasonality and the Northern Hemisphere bias of our measurements dataset. Residential emissions are only constrained in winter when emissions

are high. Microphysical process rates (dry deposition of accumulation mode and wet scavenging rates) are consistently constrained throughout the year.

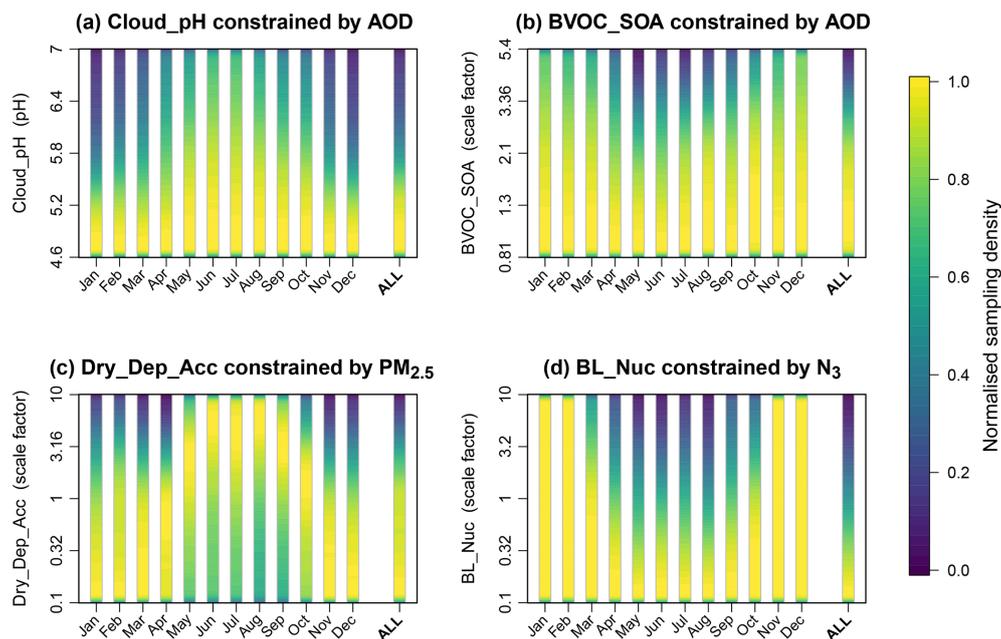
For  $\text{PM}_{2.5}$ , the seasonality of constraint is very similar to AOD with one notable exception. The dry deposition rate of accumulation-mode particles is constrained to high values in summer (consistent with AOD and sulfate) but to low values in the winter (Fig. 7c). This may occur just because of the way in which the combinations of parameters control  $\text{PM}_{2.5}$ ; for example, BVOCs can account for  $\text{PM}_{2.5}$  in summer so high dry deposition rates cannot be ruled out. However, it may also indicate a structural deficiency. The low deposition rates in winter imply that  $\text{PM}_{2.5}$  has missing sources in winter but not in the summer, such as nitrate. Our model does not include aerosol nitrate, which (if included) would increase Northern Hemisphere winter  $\text{PM}_{2.5}$  concentrations and weaken the constraint on dry deposition towards lower values in Northern Hemisphere winter.

For  $N_3$ , we find that the boundary layer nucleation rate is constrained only in summer when photochemical production of the nucleating vapours is fast (Fig. 7d). This is consistent with previous studies that have examined the seasonal cycle of organic-mediated nucleation (Riccobono et al., 2014). Similarly,  $N_3$  measurements constrain  $\text{SO}_2$  emissions and cloud droplet pH in summer when nucleation is most active. This is in contrast to the AOD and sulfate measurements, which constrained these two parameters in winter when their relative contribution to aerosol mass is greater.

For  $N_{50}$ , we find that parameter constraints do not vary smoothly throughout the year (not shown). This is because the  $N_{50}$  measurements we have used are primarily from campaigns, which move around the globe, resulting in constraint of regionally important parameters. This is one indication that we need to add long-term network measurements of  $N_{50}$  to the dataset.

### 3.3 Constraint using all measurement types

The multivariate constraint is shown as the right-hand column of PDFs in Fig. 6 and Table 2 shows corresponding parameter distribution statistics from this constraint. For each individual variable/month constraint that feeds into this multivariate constraint, the implausibility threshold and tolerance criteria ( $\theta$  and  $T$ ) were relaxed from the individual measurement constraints to retain approximately 75 % of the 1 million model variants (Table A2). This relaxed criterion leads to measurements that provide stronger constraint being downweighted and individual parameter constraints becoming weaker, but it means that we are able to avoid overconstraining on any one measurement type. Using all measurement types together leads to retention of only 2.1 % of the original 1 million model variants as plausible models (nearly 98 % rejected; Table A2). In most cases, the marginal parameter distributions from this constraint can be understood in



**Figure 7.** Seasonal variation in the constraint of parameter marginal probability distributions. The examples are (a) constraint of the pH of cloud droplets (Cloud\_pH) parameter using global AOD measurements, (b) constraint of SOA production from BVOCs (BVOC\_SOA) using AOD measurements, (c) constraint of the dry deposition rate of accumulation-mode particles (Dry\_Dep\_Acc) using global PM<sub>2.5</sub> measurements and (d) constraint of boundary layer nucleation rates (BL\_Nuc) using  $N_3$  measurements mainly over Europe.

terms of the combination of individual constraints described above.

*Boundary layer nucleation rates* are constrained to the low end of the range, which can be attributed almost entirely to the  $N_3$  measurements. However, the constraint is slightly weaker than when just  $N_3$  measurements are used because of the need to relax the tolerances and thresholds applied when ruling out model variants using multiple measurement types (Sect. 2.4.3). The nucleation rate is constrained such that the likelihood of it being in the lower half of the range (0.1–1 times the default value) is 70% – more than twice the likelihood of it being in the upper half of the range (1–10 times the default value).

*The pH of cloud droplets*, which controls aqueous-phase oxidation of SO<sub>2</sub> to form sulfate aerosol, is constrained to be more likely in the middle of our elicited range. This results from a combination of AOD and sulfate measurements constraining it to the lower end of the range and PM<sub>2.5</sub> measurements constraining it to the higher end. Observational constraint is unable to rule out any of the pH values between 4.6 and 7.0, although there is a reduction of 0.13 in the 95% credible interval to 4.69–6.84 (from 4.66–6.94 before constraint) and a larger reduction of 0.32 in the interquartile range to 5.24–6.12 (from 5.2–6.4 before constraint).

*Biomass burning emissions* are weakly constrained. The likelihood of emissions being more than a factor of 2 above the default value is reduced to 14% (from 25%), but all val-

ues below this down to 0.25 times the default value are still equally likely, as they were before constraint.

*Residential carbonaceous emissions* are constrained primarily through a combination of PM<sub>2.5</sub> and OC measurements. This emissions scaling parameter is constrained to be most likely near the middle of its range around the default setting, with emissions higher than about 2.7 times the default emission rate ruled out completely and also some weaker constraint at the lower end of the range. The 95% credible interval has significantly shifted towards lower values, from 0.27–3.73 times the default value before constraint to 0.27–1.85 times the default value after constraint, with the constrained interquartile range being 0.46–1.06 (Table 2).

*The diameter of fossil fuel particles* is constrained mainly through the  $N_{50}$  measurements towards larger diameters, with a likelihood of being in the upper half of our elicited range (60–90 nm diameter) of 61% and the median of this parameter distribution shifting to a larger diameter on constraint, increasing from 60 to 65.63 nm.

*Sea spray emissions* are constrained through a combination of AOD and PM<sub>2.5</sub> measurements, and to a lesser extent by  $N_{50}$ . The multivariate constraint is slightly weaker than was achieved by AOD and PM<sub>2.5</sub> individually, although we are still able to rule out emissions in the range 4.7–8 times the default value. Emissions in the range 0.125–2.8 times the default value are not strongly constrained by any of the measurements.

**Table 2.** Marginal parameter distribution statistics (median and 95 % credible interval) for the unconstrained sample of 1 million model variants in normal font and the constrained sample of model variants from the constraint using all measurement types simultaneously in bold. The final column shows the ratio of the constrained to the unconstrained 95 % credible interval range, accounting for the nature of the parameter (absolute or multiplicative) by using the  $\log_{10}$  scale for the calculation when the parameter is a multiplicative scaling.

Parameter	Median		95 % credible interval		95 % CI range ratio (constrained/unconstrained)
BL_Nuc <sup>a</sup>	1.00	<b>0.47</b>	(0.11, 8.91)	<b>(0.11, 6.79)</b>	0.94
Ageing	5.15	<b>5.51</b>	(0.54, 9.76)	<b>(0.55, 9.81)</b>	1.00
Acc_Width	1.50	<b>1.50</b>	(1.21, 1.79)	<b>(1.21, 1.79)</b>	1.00
Ait_Width	1.50	<b>1.55</b>	(1.21, 1.79)	<b>(1.23, 1.79)</b>	0.97
Cloud_pH	5.80	<b>5.67</b>	(4.66, 6.94)	<b>(4.69, 6.84)</b>	0.94
Carb_FF_Ems <sup>a</sup>	1.00	<b>1.01</b>	(0.52, 1.93)	<b>(0.52, 1.93)</b>	1.00
Carb_BB_Ems <sup>a</sup>	1.00	<b>0.83</b>	(0.27, 3.73)	<b>(0.26, 3.28)</b>	0.97
Carb_Res_Ems <sup>a</sup>	1.00	<b>0.73</b>	(0.27, 3.73)	<b>(0.27, 1.85)</b>	0.73
Carb_FF_Diam	60.00	<b>65.63</b>	(31.50, 88.50)	<b>(35.16, 88.80)</b>	0.94
Carb_BB_Diam	195.00	<b>194.97</b>	(95.25, 294.75)	<b>(94.89, 295.27)</b>	1.00
Carb_Res_Diam	295.00	<b>299.73</b>	(100.25, 489.75)	<b>(99.26, 492.03)</b>	1.01
Prim_SO4_Frac <sup>a</sup>	$3.16 \times 10^{-4}$	<b><math>2.41 \times 10^{-4}</math></b>	$(1.33 \times 10^{-6}, 7.50 \times 10^{-2})$	<b><math>(1.26 \times 10^{-6}, 7.46 \times 10^{-2})</math></b>	1.00
Prim_SO4_Diam	51.50	<b>56.43</b>	(5.43, 97.58)	<b>(7.06, 98.04)</b>	0.99
Sea_Spray <sup>a</sup>	1.00	<b>0.82</b>	(0.14, 7.21)	<b>(0.14, 3.69)</b>	0.83
Anth_SO2 <sup>a</sup>	0.95	<b>0.77</b>	(0.61, 1.47)	<b>(0.61, 1.35)</b>	0.90
Volc_SO2 <sup>a</sup>	1.30	<b>1.25</b>	(0.73, 2.31)	<b>(0.73, 2.30)</b>	1.00
BVOC_SOA <sup>a</sup>	2.09	<b>1.88</b>	(0.85, 5.15)	<b>(0.86, 3.74)</b>	0.82
DMS <sup>a</sup>	1.00	<b>0.97</b>	(0.52, 1.93)	<b>(0.52, 1.92)</b>	1.00
Dry_Dep_Ait <sup>a</sup>	1.00	<b>0.88</b>	(0.52, 1.93)	<b>(0.51, 1.90)</b>	1.00
Dry_Dep_Acc <sup>a</sup>	1.00	<b>0.76</b>	(0.11, 8.91)	<b>(0.11, 5.73)</b>	0.90
Dry_Dep_SO2 <sup>a</sup>	1.00	<b>1.45</b>	(0.22, 4.61)	<b>(0.23, 4.76)</b>	1.00
Kappa_OC	0.35	<b>0.36</b>	(0.11, 0.59)	<b>(0.11, 0.59)</b>	1.00
Sig_W	0.40	<b>0.40</b>	(0.12, 0.68)	<b>(0.11, 0.69)</b>	1.04
Dust <sup>a</sup>	1.00	<b>1.03</b>	(0.52, 1.93)	<b>(0.52, 1.94)</b>	1.00
Rain_Frac	0.50	<b>0.50</b>	(0.31, 0.69)	<b>(0.31, 0.69)</b>	1.00
Cloud_Ice_Thresh	0.30	<b>0.29</b>	(0.11, 0.49)	<b>(0.11, 0.49)</b>	1.00

<sup>a</sup> Parameter values given as a multiplicative scaling.

*Anthropogenic SO<sub>2</sub> emissions* are strongly constrained to the lower part of the elicited range by a combination of AOD and sulfate measurements. The emissions are most likely to be at the lower end of our elicited range (0.6 times the default value) and the likelihood of the emissions being in the range 0.6–1 times the default value is 82 %. Our interquartile range after constraint is 0.67–0.93 times the baseline emission value of 98 Tg yr<sup>-1</sup> from the MACC/CityZEN EU project (MACCity) inventory, so 65–91 Tg yr<sup>-1</sup>. Our constrained range therefore lies largely below the baseline value, with only an 18 % probability of it being above the baseline value. Liu et al. (2018) have developed a new SO<sub>2</sub> emission inventory based on Ozone Monitoring Instrument (OMI) measurements. They did not provide a global estimate of SO<sub>2</sub> emissions, but over the US and Europe, where most of our sulfate measurements are located, their revised emissions are 40 % lower than in the Hemispheric Transport of Air Pollution (HTAP) inventory, which is in the same direction as our constraint. In their inverse model study, C. Lee et al. (2011)

estimated global land SO<sub>2</sub> emissions of 100–105 Tg yr<sup>-1</sup> (with an estimated uncertainty of 20 %), in agreement with MACCity emissions, but their central value is around our 85th percentile.

*BVOC emissions* are constrained to a central value that corresponds to a global annual SOA production of about 86.5 Tg yr<sup>-1</sup>. No values in the parameter range (corresponding to an emissions range of 37–250 Tg yr<sup>-1</sup>) are ruled out, although the likelihood of SOA production being in either the upper (above 150 Tg yr<sup>-1</sup>) or lower (below 60 Tg yr<sup>-1</sup>) quadrants of the scaling range is significantly reduced and the interquartile range of the parameter distribution has reduced from 60–155 to 62–111 Tg yr<sup>-1</sup>. BVOCs were constrained in Spracklen et al. (2011) using global aerosol mass spectrometer measurements (which we also used) and a set of model runs that perturbed combinations of biogenic monoterpene and isoprene emissions as well as an anthropogenic VOC. Here, we have used a combination of AOD, PM<sub>2.5</sub>, OC, N<sub>50</sub> and N<sub>3</sub> measurements, all of which

are influenced by SOA. Their best estimate of the global SOA source was  $140 \text{ Tg yr}^{-1}$  with an uncertainty range of  $50\text{--}380 \text{ Tg yr}^{-1}$ . This included  $100 \text{ Tg yr}^{-1}$  from anthropogenic sources (which they called anthropogenically controlled SOA), which we do not include in our set of perturbed parameters. When we use just global OC from AMS measurements, we find a 95 % range on BVOC SOA of  $42\text{--}195 \text{ Tg yr}^{-1}$ . Measurements of  $\text{PM}_{2.5}$ , AOD and, to a lesser extent  $N_{50}$ , provide additional constraint, resulting in a 95 % interval of  $40\text{--}172 \text{ Tg yr}^{-1}$  and a median of  $86.5 \text{ Tg yr}^{-1}$ . This range accounts for potential compensating effects of uncertainty in deposition rates and other parameters that were not considered in Spracklen et al. (2011).

The dry deposition rate of Aitken-mode particles is weakly constrained to low values, which comes mainly from the OC and  $N_3$  observational constraint. The likelihood of the deposition rate being in the range 0.5–1.0 times the default value (1.0) is increased from 50 % to 60 % on constraint.

The dry deposition rate of accumulation-mode particles is constrained to the middle of the range. This is likely because sulfate measurements constrain the deposition rate to be towards the high end while the other measurements constrain it towards the low end. The multivariate constraint is weaker than when individual measurement types are used (AOD, sulfate,  $\text{PM}_{2.5}$ ,  $N_{50}$ ,  $N_3$ ), which results from relaxing the individual constraints in order to retain a reasonable number of model variants when multiple variables do not agree on the best value of the deposition rate.

The dry deposition rate of  $\text{SO}_2$  is constrained to the upper part of the elicited range, with the likelihood of it being in the range 1–5 times the default value (i.e. an increase in  $\text{SO}_2$  emissions) now 62 % after constraint.

### 3.4 Model–measurement comparison

Figure 8 compares the unconstrained (black) and constrained distributions of model outputs with the measurements (green). We show the results when single measurement types are used for constraint (blue) and when all measurement types are used together (red). The constraint procedure clearly rules out wide ranges of model outputs that are inconsistent with the measured values, shown by the vertical green lines. For example, the unconstrained distribution of mean global sulfate concentration (at the measurement sites) extends up to about  $6 \mu\text{g m}^{-3}$  in January, but the tail of the distribution is limited to  $3 \mu\text{g m}^{-3}$  after constraint.

The constrained model distribution sometimes agrees much better with the measurements when only a single measurement type is used compared to when all measurements are used. The weaker multivariate constraint is because we relax the constraint on individual variables so as not to rule out all model variants. This effect is most apparent for sulfate and  $\text{PM}_{2.5}$ . The mean of the constrained  $\text{PM}_{2.5}$  distribution using all measurements is about 40 % lower than the mean of the measurements in January but the mean of the

sulfate distribution is about 50 % higher than the mean of the measurements. This is likely to indicate a structural error in the model that prevents good model–measurement agreement with both quantities in the same parts of model parameter space. One explanation could be that the model is missing sources of  $\text{PM}_{2.5}$  mass (e.g. nitrate aerosols in winter), which forces a compromise in which the constraint methodology rules in sulfate concentrations that are at the upper end of the uncertainty range to minimise the error for  $\text{PM}_{2.5}$ . Although relaxing our constraint criteria offsets many effects of such structural errors, the shifting of these all-measurement constraint distributions away from the measurements indicates some structural error is still not fully accounted for. It is possible that our constraint would adjust better to account for this structural deficiency if we could directly specify a structural error term in the implausibility measure through  $\text{Var}(S)$ .

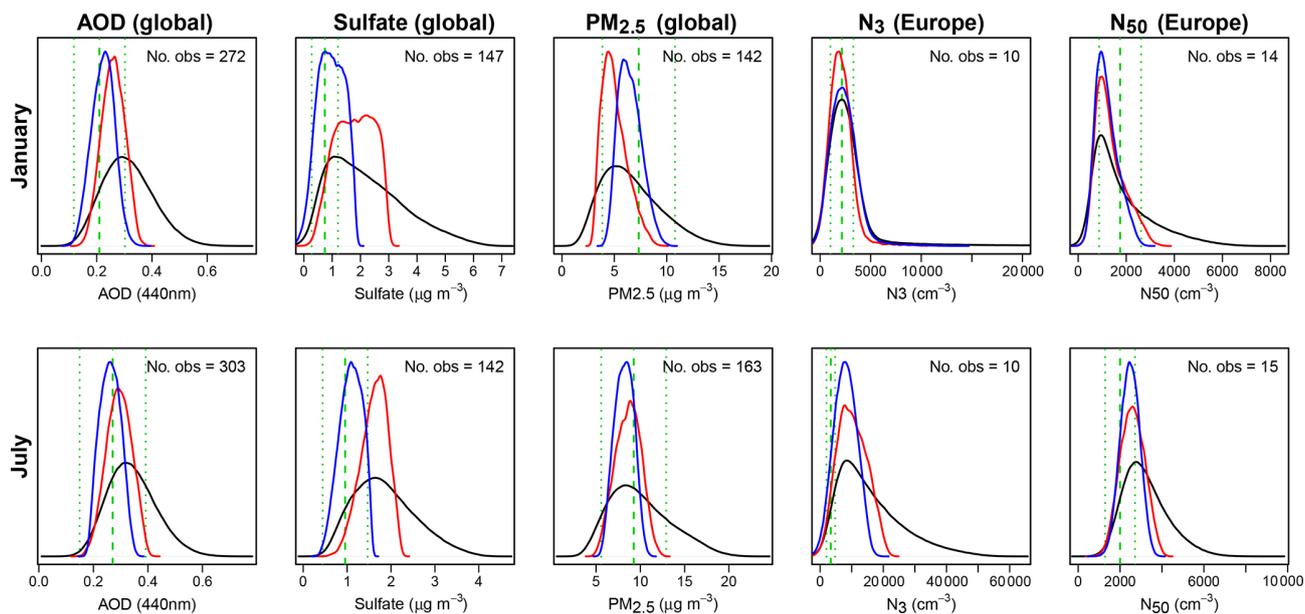
### 3.5 Unconstrained parameters

Several parameters are barely constrained or not constrained at all using all the measurements. Unconstrained microphysical processes or assumptions are the ageing rate of insoluble into soluble particles, the width of the lognormal accumulation mode, the hygroscopicity of organic material ( $\kappa_{\text{OC}}$ ), the updraft speed and wet deposition rates. Among the emissions, unconstrained parameters are the emission rates of fossil fuel particles, degassing volcanic  $\text{SO}_2$ , dimethyl sulfide (DMS) and dust emissions.

There are several potential reasons for the lack of constraint. It is possible that parts of the joint parameter space are ruled out but with a negligible effect on the marginal parameter distribution (i.e. the ruled-out parameter space is uniform across the parameter of interest). For example, wet deposition rates are directly compensated by emission rates and the ageing rate affects the wet removal rate. Another reason is that we did not include measurements in regions where the six variables are sensitive to these parameters. This is likely to be the case for DMS, volcanic and dust emissions given the relative lack of measurements over remote ocean regions and downwind of dust sources, which means these regions are not strongly weighted in the overall constraint process. Furthermore, some parameters may be more related to other aerosol properties that we have not used for constraint. For example, ageing rates in the model are not constrained, likely because the ageing process predominantly affects the black carbon concentration which is not included as a measurement type in this study.

### 3.6 Implications for constraint of aerosol forcing

Figure 9 shows the nine most important parameters for the uncertainty in global mean aerosol forcing in the PPE in terms of the forcing uncertainty they account for (Yoshioka et al., 2019). Some of these parameters are fairly strongly con-



**Figure 8.** Comparison of the constrained model with the measurements for January (top) and July (bottom). The distributions were calculated as a mean over model grid boxes containing measurements. AOD, sulfate and  $\text{PM}_{2.5}$  are global comparisons;  $N_3$  and  $N_{50}$  are Europe-only comparisons due to the limited global coverage of these measurements in each month. The black line shows the prior (unconstrained) probability distribution of the model. The blue line shows the constrained model distribution when only measurements of each type are used in the constraint. The red line shows the constrained model distribution when all measurement types are used. The green dashed line shows the mean of the measurements and the dotted lines show the approximate 95 % uncertainty range on an average observation that was accounted for in the constraint.

strained by the measurements, but others are unconstrained. Within the joint parameter space of just these nine parameters there is considerable potential for model variants that compensate, thereby reducing the effectiveness of the constrained parameters on the forcing. It also needs to be borne in mind that global mean forcing is the sum of regional forcings, and in each region a different set of parameters is being constrained and may be constraining the same parameters to different parts of their range (Lee et al., 2016; Regayre et al., 2015).

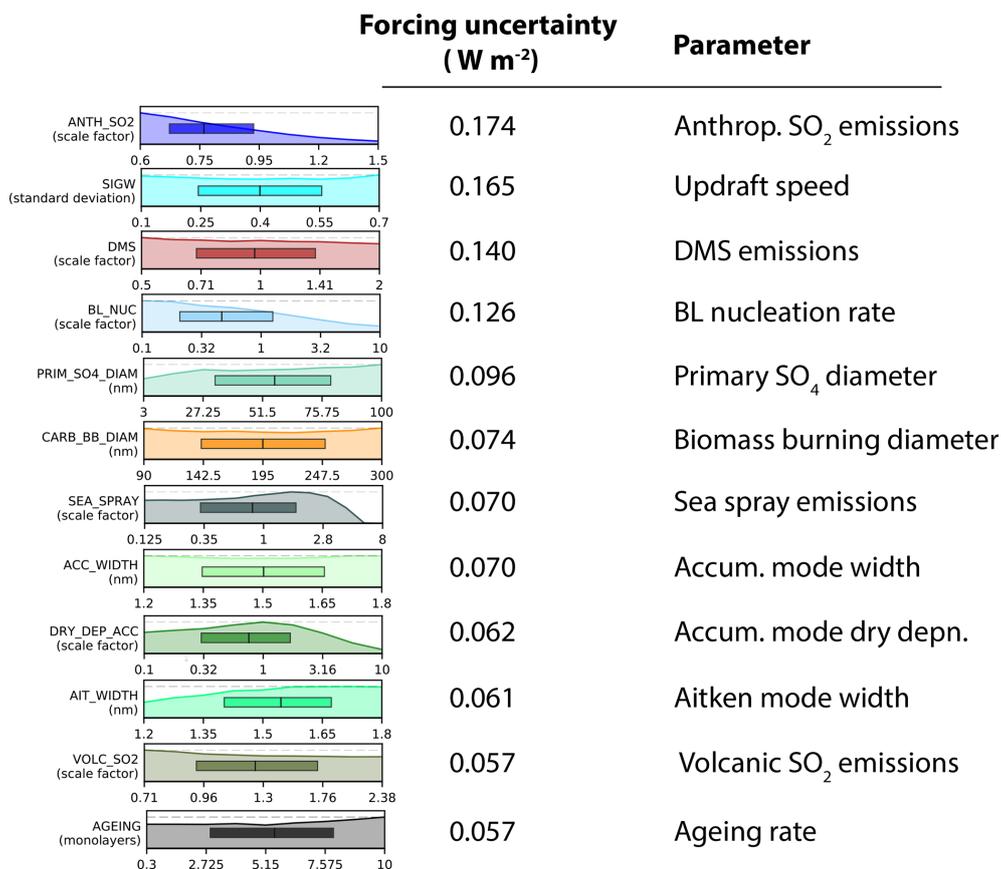
Figure 10 shows how the constrained parameters affect the uncertainty in predicted global annual mean net RF and its component parts due to aerosol–cloud interactions ( $\text{RF}_{\text{aci}}$ ) and aerosol radiation interactions ( $\text{RF}_{\text{ari}}$ ). (Note that this calculation of RF differs from that shown in Yoshioka et al. 2019, which used elicited parameter distributions when sampling over the parameter uncertainty space, while we use uniform distributions for the sampling here.) Table 3 shows the corresponding parameter distribution statistics (median, interquartile range,  $\pm 1\sigma$  range (on mean value) and 95 % credible interval) for these forcing constraints.

The net RF is dominated by  $\text{RF}_{\text{aci}}$ , which is only weakly constrained (Fig. 10b) by 6 %, in line with the net RF (Fig. 10a). This occurs because our constraint uses measurements of aerosol properties rather than cloud properties. Although the overall reduction in the  $\text{RF}_{\text{aci}}$  uncertainty is weak, the PDFs in Fig. 10b show a slight shift in  $\text{RF}_{\text{aci}}$  to stronger

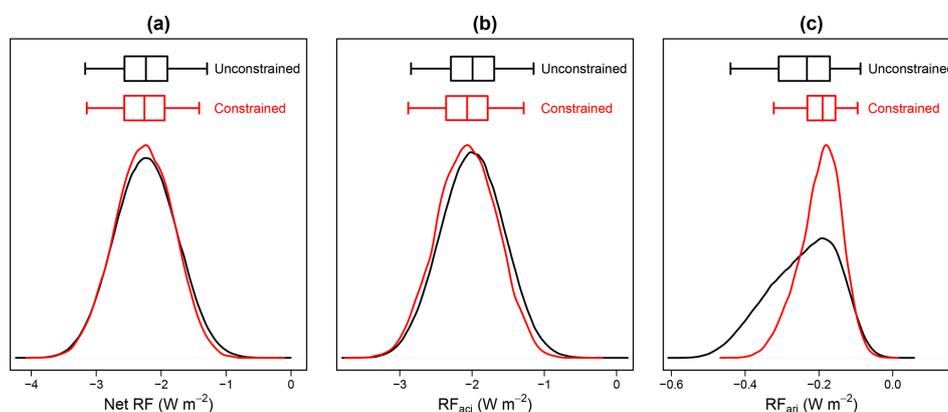
forcings, with the median  $\text{RF}_{\text{aci}}$  shifting from  $-1.99 \text{ W m}^{-2}$  in the prior (unconstrained) distribution to  $-2.07 \text{ W m}^{-2}$  after constraint. The likelihood of the strength of  $\text{RF}_{\text{aci}}$  being weaker than  $-1.5 \text{ W m}^{-2}$  (less negative) is reduced by 38 % and the likelihood of it being stronger (more negative) than  $-2.5 \text{ W m}^{-2}$  is increased by 20 %. In general, although the anthropogenic emissions were constrained to lower values (which should weaken the forcing), the sea spray emissions were constrained to lower values, which acts to strengthen the forcing (Carslaw et al., 2013a; Regayre et al., 2014). The 95 % credible interval for the net RF is reduced by 8 %.

The 95 % credible interval of direct forcing,  $\text{RF}_{\text{ari}}$ , is reduced by 34 % (Fig. 10c) and the  $\pm 1\sigma$  range is reduced by 33 %.  $\text{RF}_{\text{ari}}$  is constrained most strongly by the  $\text{PM}_{2.5}$ , AOD and sulfate measurements (not shown) but insignificantly by the OC,  $N_3$  and  $N_{50}$  measurements. The inconsistency between the constraints on  $\text{PM}_{2.5}$ , AOD and sulfate (Fig. 8) leads to inconsistency in the constraint on  $\text{RF}_{\text{ari}}$ . In particular, using just sulfate measurements results in an  $\text{RF}_{\text{ari}} \pm 1\sigma$  range of  $-0.10$  to  $-0.22 \text{ W m}^{-2}$ , but using just  $\text{PM}_{2.5}$  measurements results in a range  $-0.20$  to  $-0.36 \text{ W m}^{-2}$ . This highlights the importance of detecting and fixing structural deficiencies in the model as well as the limitations of using single-variable emergent constraints.

Furthermore, we have found that the observational constraint on present-day (2008) global annual mean aerosol radiative effects (RE; relative to no aerosol radiative effects)



**Figure 9.** Ranked list of parameters that dominate the uncertainty in global mean aerosol radiative forcing in the ensemble (Yoshioka et al., 2019).



**Figure 10.** Effect of observational constraint using all measurement types on the probability distribution of global annual mean aerosol radiative forcing: (a) net RF, (b) RF<sub>aci</sub> (aerosol–cloud interaction) and (c) RF<sub>ari</sub> (aerosol–radiation interaction). The black line shows the prior (unconstrained) distribution and the red line shows the constrained distribution. For each box-and-whisker plot, the box represents the interquartile range split at the median forcing (the line inside the box), and the whiskers extend to the lower and upper bounds of the 95 % credible interval of the distribution.

**Table 3.** Uncertainty distribution statistics (median, interquartile range,  $\pm 1\sigma$  range (on mean value) and 95 % credible interval) for the global annual mean aerosol radiative forcing (net RF, RF<sub>aci</sub> and RF<sub>ari</sub>) from the unconstrained sample of 1 million model variants in normal font and the constrained sample of model variants from the all-variables constraint in bold. The final two columns show the ratio of the constrained to the unconstrained 95 % credible interval and corresponding percentage reduction in this interval on constraint.

	Median	Interquartile range	$\pm 1\sigma$ range	95 % credible interval	95 % CI range ratio (constrained/unconstrained)	95 % CI reduction
Net RF	-2.23 <b>-2.26</b>	(-2.56, -1.90) <b>(-2.57, -1.95)</b>	(-2.71, -1.75) <b>(-2.71, -1.81)</b>	(-3.17, -1.29) <b>(-3.15, -1.41)</b>	0.92	8 %
RF <sub>aci</sub>	-1.99 <b>-2.07</b>	(-2.29, -1.69) <b>(-2.36, -1.78)</b>	(-2.43, -1.55) <b>(-2.48, -1.66)</b>	(-2.84, -1.15) <b>(-2.88, -1.28)</b>	0.94	6 %
RF <sub>ari</sub>	-0.23 <b>-0.19</b>	(-0.31, -0.17) <b>(-0.23, -0.16)</b>	(-0.33, -0.15) <b>(-0.26, -0.14)</b>	(-0.44, -0.09) <b>(-0.32, -0.10)</b>	0.66	34 %

is stronger than the constraint on aerosol radiative forcing. The uncertainty in the clear-sky aerosol RE is reduced by 47 %. Industrial-period aerosol radiative forcing has distinct sources of uncertainty from aerosol radiative effects in the present-day atmosphere (Regayre et al., 2018), so a stronger constraint on present-day radiative effects is in line with expectations.

It is important to note that the probability distribution of net aerosol RF includes values that, with current knowledge, would produce a net negative (greenhouse gas plus aerosol) forcing over the industrial period. The forcing is dominated by aerosol–cloud processes, which we have not attempted to constrain here. Nevertheless, the lack of constraint shows, as in Lee et al. (2016), a well-configured global aerosol model has little bearing on the uncertainty in RF<sub>aci</sub>. In contrast, the constraint of RF<sub>ari</sub>, which has not been attempted in our previous studies, is significant and encouraging.

#### 4 Conclusions

We have used extensive point measurements of AOD, PM<sub>2.5</sub>, sulfate mass, organic carbon mass and the concentrations of particles larger than 50 nm dry diameter ( $N_{50}$ ) and 3 nm ( $N_3$ ) from surface sites, aircraft and ships to constrain uncertain aerosol parameters in a global aerosol–climate model. A total of 26 parameters related to aerosol emissions and processes were varied in a perturbed parameter ensemble and statistical emulators were used to generate a set of 1 million model variants that represent the model outputs for combinations of parameter values across the 26-dimensional uncertainty space. The plausibility of each model variant was tested against each measurement type in turn and then in combination using a history-matching procedure based on an implausibility metric (Craig et al., 1996; Williamson et al., 2013). The resulting probability distributions of aerosol forcings can be considered as the “observationally plausible” ranges for the HadGEM3-UKCA model.

Observational constraint ruled out almost 98 % of the 26-dimensional parameter space and the probability distributions of many parameter values were effectively constrained. Overall, 14 of the parameters were constrained to some extent, despite the fact that there are many ways in which parameter values can be combined to produce plausible results within the uncertainties. Constraint of a parameter means that the probability distribution of a parameter (and potentially its absolute range) is narrowed, and hence the likelihood of the parameter taking a particular range of values within its absolute range is increased; i.e. there are more ways to combine these parameter values with values of the other 25 parameters to produce a plausible model. For two parameters, some of the individual prior elicited parameter ranges were ruled out entirely: very high sea spray emissions and very high residential carbonaceous emissions. The very highest BVOC emissions were nearly ruled out. However, for the remaining parameters, it was not possible to entirely rule out any part of the prior range.

Parameter constraints are mostly, but not always, consistent across multiple measurement types, even though the different measurement types were made at very different locations and sometimes at different times of the year. For example, we often found consistent constraint of parameters related to the production of aerosol mass or number (emissions, nucleation, secondary aerosol mass) and parameters related to removal (condensation sink, deposition rates of gases and aerosols). There is also a very clear seasonal variation in the magnitude of constraint related to variations in the dominant processes.

The multivariate constraint has the following effect on the parameter probability distributions, which were assumed before constraint to be equally plausible between lower and upper bounds defined by expert elicitation:

1. Boundary layer nucleation rates based on a sulfuric acid–organic mechanism (Metzger et al., 2010) are con-

strained to the low end of the elicited range, mainly from the  $N_3$  measurements.

2. The pH of cloud droplets, which controls aqueous-phase oxidation of  $\text{SO}_2$  to form sulfate aerosol, is constrained to be more likely in the middle of our elicited range but we were unable to rule out any of the pH values between 4.6 and 7.0. The constraint led to a reduction of 0.32 in the interquartile range to 5.24–6.12 (from 5.2–6.4 before constraint) and a reduction of 0.13 in the 95 % credible interval to 4.69–6.84 (from 4.66–6.94 before constraint).
3. Biomass burning emissions are weakly constrained by the  $\text{PM}_{2.5}$ , AOD and OC measurements, with a reduced likelihood of the emissions exceeding a factor of 2 above the default value. All lower emissions down to 0.25 times the default value are equally likely as they were before constraint.
4. Residential carbonaceous emissions are constrained primarily through a combination of  $\text{PM}_{2.5}$  and OC measurements. Emissions higher than about 2.7 times the default emission rate were effectively ruled out by the OC measurements as observationally implausible.
5. The diameter of fossil fuel particles is constrained by  $N_{50}$  measurements, with a reduced likelihood of being in the range 30–45 nm, but diameters in the range of 60–90 nm are unconstrained by any measurements.
6. Sea spray emissions are constrained through a combination of AOD,  $\text{PM}_{2.5}$  and  $N_{50}$  measurements. Emissions in the range 4.7–8.0 times the default value are ruled out but emissions in the range 0.125–2.8 times the default value are not strongly constrained by any of the measurements.
7. Anthropogenic  $\text{SO}_2$  emissions are strongly constrained to low values by AOD and sulfate measurements, with an 82 % likelihood of being below the default value from the emission inventory.
8. BVOC emissions are constrained by AOD,  $\text{PM}_{2.5}$ , OC,  $N_{50}$  and  $N_3$  measurements. The likelihood of either high or low emissions is reduced. On constraint, our median estimate corresponds to  $86.5 \text{ Tg yr}^{-1}$  SOA production, with a 95 % credible interval of 40–172  $\text{Tg yr}^{-1}$ . However, no values in the range 37–250  $\text{Tg yr}^{-1}$  are ruled out entirely.
9. The dry deposition rate of Aitken-mode particles is weakly constrained to low values using OC and  $N_3$  measurements and the dry deposition rate of accumulation-mode particles is constrained to the middle of the range by AOD, sulfate,  $\text{PM}_{2.5}$ ,  $N_{50}$  and  $N_3$  measurements.

The dry deposition rate of  $\text{SO}_2$  is constrained to the upper part of the elicited range, with a 62 % likelihood of it being above the default value.

Several parameters of importance to aerosol forcing were not well constrained, in particular parameters related to micro-physical processes (primary sulfate particle diameter, the diameter of biomass burning particles, the width of the accumulation mode and the ageing rate). Dimethyl sulfide emissions were also not strongly constrained.

The prior (unconstrained) uncertainty (95 % credible interval) in the pre-industrial to present-day net aerosol RF is reduced by 8 %. The radiative forcing in the ensemble accounts for direct and indirect (cloud albedo) effects but not cloud adjustments.  $\text{RF}_{\text{ari}}$  uncertainty (95 % credible interval) is reduced by 34 %, but the net RF uncertainty is dominated by the  $\text{RF}_{\text{aci}}$  uncertainty, which is reduced by only 6 %. The recent assessment of aerosol forcing (Bellouin et al., 2019) adopted  $\pm 1\sigma$  ranges to define the uncertainty. Our equivalent  $\pm 1\sigma$  ranges are  $-0.14$  to  $-0.26 \text{ W m}^{-2}$  for  $\text{RF}_{\text{ari}}$  (reduction of 33 % due to constraint) and  $-1.66$  to  $-2.48 \text{ W m}^{-2}$  for  $\text{RF}_{\text{aci}}$  (reduction of 7 % due to constraint). The reduction in uncertainty is much larger for  $\text{RF}_{\text{ari}}$  than  $\text{RF}_{\text{aci}}$  because our constraints focus on aerosol properties rather than cloud properties.

Our results highlight the importance of using multiple measurement types to constrain aerosol–climate models. We have shown that use of a single measurement type, as is done in emergent constraint studies, would lead to an overconfident constraint. This is because potential structural deficiencies in our model prevent consistently good constraint across several measurement types. In particular, we showed that constraint using  $\text{PM}_{2.5}$  or sulfate aerosol measurements led to probability distributions of  $\text{RF}_{\text{ari}}$  that barely overlap. The final multivariate constraint on forcing is therefore a compromise that achieves reasonable agreement with all observations rather than being overconfidently constrained by one metric.

In terms of future directions and requirements to achieve better constraint, we make the following recommendations:

1. We need to understand and quantify model–measurement representativeness errors. The biggest challenge (and the factor that most limits the constraint, other than model structural error) is quantification of the representativeness error associated with comparing point measurements with a global model (Reddington et al., 2017; Schutgens et al., 2017, 2016a, b). The ambiguity in deciding whether a model–measurement bias is related to structural error in the model or underestimation of the uncertainty terms was the main limiting factor in our constraint procedure. The representativeness uncertainties have been estimated based on model simulations at a few locations (Schutgens et al., 2016b), but they have not been measured, and we have no information about these uncertainties at

other locations. Small campaigns to characterise the space–time variability around existing sites and for potential new sites should be considered. Alternatively, representation errors could be reduced by increasing the temporal and spatial resolution of our model, and the effect of interannual variability could be reduced if we always compare the model to measurements for the correct year. In our calculations of model variant implausibility, interannual variability accounts for around 30 to 90 % of the model–measurement comparison representation uncertainty. With smaller representation errors in the model–measurement comparison, we could more confidently reject implausible model variants. However, this approach is not currently computationally feasible.

2. We need to understand and quantify structural deficiencies in the model. Our current approach, to assume zero structural error in our implausibility calculations, enables some structural errors in the model to be detected by comparing single and multi-measurement constraints. Ideally, to produce the most robust constraint possible, these structural errors should be accounted for directly using  $\text{Var}(S)$  in the implausibility measure – or, better still, corrected in the model.
3. We should prioritise expansion of measurements to cover more long-term well-characterised sites. Such measurements will have much lower (or at least well-defined) representativeness error when used for model evaluation and constraint. Sites should be characterised in terms of how they represent a typical model grid box over a long period of time. These sites should be in diverse locations so that they help to constrain a wide range of model uncertainties (Reddington et al., 2017).
4. We should aim to dedicate part of field campaigns to routine, unbiased (or effectively random) sampling of aerosols across the scale of model grid boxes. Such measurements will also have much lower representativeness error than measurements that target specific processes or aerosol environments. Many field campaigns (particularly using aircraft) often prioritise measurements to explore aerosol processes or to characterise particular aerosol environments (e.g. pollution plumes). Others, like the Atmospheric Tomography Mission, emphasised the effective sampling of air masses in an unbiased way. A greater emphasis on such tomography missions, even just as part of a larger field campaign, would benefit model uncertainty reduction.

5. Additional aerosol measurements could be used to further constrain the parameters and forcing. We expect the following (already available) measurements would provide further constraint: (i) black carbon measurements to help constrain the aerosol absorption component of radiative forcing but also as a measure of aerosol removal rates; (ii)  $\text{SO}_2$  concentrations to avoid ambiguity between sulfate and  $\text{PM}_{2.5}$  constraints; (iii) measurements from biomass burning regions to help constrain both the emissions and the size distribution of the particles; (iv) vertical profiles of aerosol (Watson-Parris et al., 2019); (v) more particle number size distribution information instead of just  $N_3$  and  $N_{50}$ , which have not constrained the size distribution well enough.
6. Measured or derived process rates would be very useful because they would help to constrain model parameters directly, rather than relying on indirect constraint through measured state variables. For example, dry and wet deposition rates (Emerson et al., 2018) are required on the scale of model grid boxes (tens to hundreds of kilometres). Similarly, direct estimates of particle formation and growth rates (Kerminen et al., 2018) would be useful, rather than just relying on integral particle number concentrations averaged over long periods.

It is very challenging to constrain model uncertainty using a large set of perturbed parameter model variants and extensive measurements of different types. However, the resulting ranges of model parameters and outputs (like radiative forcings) estimated in this way are much more robust than those based on a very small number of models. Although our study is incomplete (not all parameters were perturbed and not all measurements were used), the outcome is an estimate of the “observationally plausible” range of aerosol forcings for the HadGEM3-UKCA aerosol–climate model. The ranges might be wider if we accounted for more sources of uncertainty and directly accounted for structural errors in the implausibility calculations, but they could also be narrowed if we improved model structural deficiencies, reduced model–measurement representativeness errors and used a wider set of measurement types.

## Appendix A

**Table A1.** The 26 aerosol parameters included in the AER PPE. Further details are provided in a separate publication (Yoshioka et al., 2019).

Index	Parameter name	Description
1	BL_Nuc	Boundary layer nucleation rate
2	Ageing	Ageing of hydrophobic aerosols (no. of monolayers of soluble material)
3	Acc_Width	Modal width of accumulation modes (nm)
4	Ait_Width	Modal width of Aitken modes (nm)
5	Cloud_pH	pH of cloud droplets (used to calculate the conversion of SO <sub>2</sub> into sulfate)
6	Carb_FF_Ems	Carbonaceous fossil fuel emissions scale factor
7	Carb_BB_Ems	Carbonaceous biomass burning emissions scale factor
8	Carb_Res_Ems	Carbonaceous residential (biofuel) emissions scale factor
9	Carb_FF_Diam	Carbonaceous fossil fuel emission diameter (nm)
10	Carb_BB_Diam	Carbonaceous biomass burning emission diameter (nm)
11	Carb_Res_Diam	Carbonaceous residential (biofuel) emission diameter (nm)
12	Prim_SO4_Frac	Mass fraction of SO <sub>2</sub> converted to new sulfate particles in power plant plumes
13	Prim_SO4_Diam	Mode diameter of new subgrid sulfate particles (nm)
14	Sea_Spray	Sea spray aerosol scale factor
15	Anth_SO2	Anthropogenic SO <sub>2</sub> emission scale factor
16	Volc_SO2	Volcanic SO <sub>2</sub> emission scale factor
17	BVOC_SOA	Biogenic secondary aerosol formation from volatile organic compounds scale factor
18	DMS	Dimethyl sulfide surface ocean concentration scale factor
19	Dry_Dep_Ait	Aitken-mode dry deposition velocity scale factor
20	Dry_Dep_Acc	Accumulation-mode dry deposition velocity scale factor
21	Dry_Dep_SO2	SO <sub>2</sub> dry deposition velocity scale factor
22	Kappa_OC	$\kappa$ -Köhler coefficient of organic carbon
23	Sig_W	Updraft vertical velocity standard deviation (used to calculate the activation of aerosols into cloud drops)
24	Dust	Dust emission scale factor
25	Rain_Frac	Fraction of cloud-covered area in large-scale clouds where aerosol scavenging by raindrops occurs
26	Cloud_Ice_Thresh	Threshold of cloud ice fraction above which nucleation scavenging is suppressed (restricting further activation of aerosols into cloud drops)

**Table A2.** The choices of the number of measurements (tolerance,  $T$ ) allowed to exceed a threshold implausibility ( $\theta$ ) for constraint on each of the monthly observed aerosol properties, along with the corresponding percentage of the large sample of 1 million model variants (that covers the PPE parametric uncertainty) that is retained on constraint for each choice. The choices on the left correspond to those used for the all-months constraints shown on the left of Fig. 6. The choices on the right correspond to those used for the joint all-variables-months constraint shown on the right in Fig. 6. The percentage of variants retained is also shown for the combined all-months constraint for each variable (ALL) as well as for the joint all-variables-months constraint (final row). LHS indicates the left-hand side and RHS the right-hand side.

AOD		All-months constraint choices (LHS Fig. 6)			Joint all-variables-months constraint choices (RHS Fig. 6)		
Month	No. obs used in constraint	Threshold implausibility	Tolerance (no. obs)	% Variants retain	Threshold implausibility	Tolerance (no. obs)	% Variants retain
Jan	272/294	3.5	33	40.03	4	76	75.60
Feb	274/301	3.5	27	42.37	3.5	82	75.71
Mar	284/309	3.5	23	40.54	3.5	80	75.78
Apr	289/316	3.5	17	39.70	3.5	52	76.47
May	292/322	3.5	12	34.55	3.5	47	76.17
Jun	295/320	3.5	24	43.10	3.5	53	74.13
Jul	303/323	3.5	24	43.38	3.5	48	75.46
Aug	310/326	3.5	19	35.51	3.5	50	74.06
Sep	303/321	3.5	24	43.09	3.5	61	74.65
Oct	285/315	3.5	23	39.57	3.5	80	75.97
Nov	273/309	3.5	27	41.91	3.5	93	74.75
Dec	267/298	3.5	43	41.44	4	91	75.82
ALL	3447/3754			19.01			60.42
Sulfate		All-months constraint choices (LHS Fig. 6)			Joint all-variables-months constraint choices (RHS Fig. 6)		
Month	No. obs used in constraint	Threshold implausibility	Tolerance (no. obs)	% Variants retain	Threshold implausibility	Tolerance (no. obs)	% Variants retain
Jan	147/149	3.5	3	38.76	4	50	75.42
Feb	164/148	3.5	3	38.79	4	38	74.64
Mar	144/151	3.5	6	42.70	4	46	74.99
Apr	149/151	3.5	6	38.66	4	51	75.22
May	146/149	3.5	9	41.38	4.5	35	74.60
Jun	142/150	3.5	6	38.19	3.5	48	74.40
Jul	142/148	3.5	9	41.75	4	40	75.19
Aug	143/148	3.5	9	38.97	4	46	75.00
Sep	144/147	3.5	3	39.13	4	46	74.77
Oct	143/147	3.5	3	38.03	4	46	75.02
Nov	143/146	3.5	3	40.21	4	46	75.52
Dec	144/147	3.5	3	41.52	4	49	72.42
ALL	1733/1781			27.73			64.90
PM <sub>2.5</sub>		All-months constraint choices (LHS Fig. 6)			Joint all-variables-months constraint choices (RHS Fig. 6)		
Month	No. obs used in constraint	Threshold implausibility	Tolerance (no. obs)	% Variants retain	Threshold implausibility	Tolerance (no. obs)	% Variants retain
Jan	142/168	3.5	20	41.52	3.5	40	74.02
Feb	145/168	3.5	20	36.27	3.5	38	73.51
Mar	154/170	3.5	15	39.14	3.5	34	75.09
Apr	152/170	3.5	9	43.70	3.5	24	72.93
May	154/167	3.5	9	43.16	3.5	22	75.83
Jun	158/170	3.5	13	34.89	3.5	32	74.91
Jul	163/172	3.5	16	40.27	3.5	36	74.08
Aug	161/169	3.5	13	36.70	3.5	39	75.68
Sep	154/166	3.5	9	42.71	3.5	31	75.37
Oct	151/165	3.5	12	45.71	3.5	24	75.59
Nov	142/168	3.5	14	38.76	3.5	31	73.69
Dec	148/169	3.5	24	40.51	3.5	44	75.52
ALL	1824/2022			10.80			48.89

Table A2. Continued.

OC		All-months constraint choices (LHS Fig. 6)			Joint all-variables-months constraint choices (RHS Fig. 6)		
Month	No. obs used in constraint	Threshold implausibility	Tolerance (no. obs)	% Variants retain	Threshold implausibility	Tolerance (no. obs)	% Variants retain
Jan	3/6	3.5	0	63.60	4.5	0	76.25
Feb	10/14	3.5	3	58.59	3.5	4	86.56
Mar	41/82	3.5	2	57.69	3.5	3	72.06
Apr	49/74	3.5	2	68.29	3.5	3	76.20
May	20/23	3.5	1	56.91	3.5	2	76.69
Jun	22/23	3.5	3	71.67	3.5	4	76.64
Jul	18/23	3.5	2	57.40	3.5	4	73.36
Aug	21/23	3.5	2	65.05	3.5	3	72.19
Sep	17/22	4	4	67.45	3.5	5	77.65
Oct	24/41	3.5	4	72.35	3.5	4	72.35
Nov	15/37	4.5	0	58.57	3.5	1	81.81
Dec	11/15	3.5	1	71.62	3.5	1	71.62
ALL	251/383			6.94			20.00
$N_3$		All-months constraint choices (LHS Fig. 6)			Joint all-variables-months constraint choices (RHS Fig. 6)		
Month	No. obs used in constraint	Threshold implausibility	Tolerance (no. obs)	% Variants retain	Threshold implausibility	Tolerance (no. obs)	% Variants retain
Jan	12/13	3.5	0	95.74	3.5	0	95.74
Feb	12/13	3.5	0	96.60	3.5	0	96.60
Mar	12/13	3.5	0	85.80	3.5	0	85.80
Apr	11/12	3.5	0	68.78	3.5	1	75.35
May	11/12	3.5	0	56.05	3.5	4	77.67
Jun	12/12	3.5	1	58.61	3.5	4	75.70
Jul	13/13	3.5	1	57.15	3.5	4	73.41
Aug	13/13	3.5	1	53.12	3.5	4	73.59
Sep	13/13	3.5	0	67.29	3.5	1	74.97
Oct	12/13	3.5	0	76.94	3.5	0	76.94
Nov	12/13	3.5	1	97.44	3.5	1	97.44
Dec	10/12	3.5	0	94.03	3.5	0	94.03
ALL	143/152			44.03			61.58
$N_{50}$		All-months constraint choices (LHS Fig. 6)			Joint all-variables-months constraint choices (RHS Fig. 6)		
Month	No. obs used in constraint	Threshold implausibility	Tolerance (no. obs)	% Variants retain	Threshold implausibility	Tolerance (no. obs)	% Variants retain
Jan	74/77	3.5	3	55.41	3.5	9	76.09
Feb	82/90	3.5	7	53.78	3.5	13	75.28
Mar	132/148	3.5	8	50.40	3.5	16	74.89
Apr	177/199	3.5	25	52.59	3.5	32	71.65
May	62/64	3.5	2	53.39	3.5	7	74.25
Jun	88/96	3.5	4	48.43	3.5	9	73.21
Jul	102/115	3.5	4	48.30	3.5	10	74.10
Aug	97/109	3.5	4	58.29	3.5	8	73.51
Sep	114/133	3.5	7	52.34	3.5	14	76.95
Oct	112/119	3.5	9	45.93	3.5	16	72.77
Nov	113/155	3.5	20	49.80	3.5	27	77.18
Dec	62/67	3.5	9	46.17	3.5	14	77.70
ALL	1215/1372			9.99			40.45
Final % variants retain for the joint all-variables-months constraint:							2.085

## Appendix B: Description of the datasets used in this study

For the AERONET AOD data used in this study, we thank the PIs of the AERONET sites used for maintaining their instruments and providing their data to the community. We also acknowledge AERONET for their continuous efforts in providing high-quality measurements and derivative products. All data used in this work can be accessed through the AERONET web page: <http://aeronet.gsfc.nasa.gov/> (last access: 17 September 2019).

For the sulfate data used in this study, we acknowledge the EMEP (<http://ebas.nilu.no/>, last access: 17 September 2019; Tørseth et al., 2012), IMPROVE (<http://views.cira.colostate.edu/fed/>, last access: 17 September 2019) and EANET (<https://www.eanet.asia/>, last access: 17 September 2019) measurement networks for making their measurement data available, along with all data managers involved in data collection. Additional ground station observations from the SORPES (Station for Observing Regional Processes of the Earth System) monitoring station in Nanjing, China (Ding et al., 2016), are also included. Data on the acid deposition in the east Asian region were provided from the Network Center for EANET, <https://monitoring.eanet.asia/document/public/index> (last access: 7 June 2018). IMPROVE is a collaborative association of state, tribal and federal agencies and international partners. The US Environmental Protection Agency is the primary funding source, with contracting and research support from the National Park Service. The Air Quality Group at the University of California, Davis, is the central analytical laboratory, with ion analysis provided by Research Triangle Institute and carbon analysis provided by Desert Research Institute.

For the PM<sub>2.5</sub> data used in this study, we acknowledge the IMPROVE (<http://views.cira.colostate.edu/fed/>, last access: 17 September 2019), WMO GAW-WDCA (<https://www.gaw-wdca.org>, last access: 17 September 2019; <http://ebas.nilu.no/>, last access: 17 September 2019; Tørseth et al., 2012), A-PAD (Atanacio et al., 2016) and NAPS (Galarnau et al., 2016; <http://maps-cartes.ec.gc.ca/rmspa-naps/data.aspx?lang=en>, last access: 17 September 2019) measurement networks for making their measurement data available, along with all data managers involved in data collection. Further ground station measurements are included from sites in Australia (ANSTO stations: Cohen and Atanacio, 2015), South America (Artaxo et al., 2013), Taiwan (Fang and Chang, 2010), South Africa (Vakkari et al., 2013) and Nanjing, China (SORPES station; Ding et al., 2016). The PM<sub>2.5</sub> data for Europe were obtained from the World Data Centre for Aerosols (WDCA), and we thank the following data providers to this network: Adamos Adamides (Cyprus), Jacobus P. J. Berkhout (the Netherlands), Elke Bieber (Germany), Tanja Bolte (Slovenia), Geoff Broughton (United Kingdom), Darius Ceburnis (Ireland), Anna Degorska (Poland), Iveta Dubakova

(Latvia), Fermin Elizaga (Spain), Marina Froehlich (Austria), Marina Frolova (Latvia), Robert Gehrig (Switzerland), Alberto Gonzalez (Spain), Carsten Gruening (Italy), Savvas Kleanthous (Cyprus), Manuel Lambas (Spain), Maj Britt Larka Abellan (Spain), Marijana Murovec (Slovenia), Jaroslav Pekarek (Czech Republic), Noemi Perez (Spain), Cinzia Perrino (Italy), Jean-Philippe Putaud (Italy), Xavier Querol (Spain), Stephane Sauvage (France), Karin Sjöberg (Sweden), Andre Sonntag (Germany), Gerald Spindler (Germany), Daan P. J. Swart (the Netherlands), Karin Uhse (Germany), Milan Vana (Czech Republic), Keith Vincent (UK), Pasi Aalto (Finland), Markku Kulmala (Finland), Anne-Gunn Hjellbrekk, (Norway) and Aas Wenche (Norway).

The N<sub>3</sub> data used in this study were obtained from the EBAS ACTRIS database (Asmi et al., 2013; <https://www.actris.eu/>, last access: 17 September 2019; <http://ebas.nilu.no/>, last access: 17 September 2019), collated via the Global Aerosol Synthesis and Science Project, GASSP (Reddington et al., 2017, <http://gassp.org.uk/>, last access: 17 September 2019), and public data on the EBAS database. The EBAS database has largely been funded by the UNECE CLRTAP (EMEP) and AMAP and through NILU internal resources. Specific developments have been possible due to projects like EUSAAR (EU-FP5; EBAS web interface), EBAS Online (Norwegian Research Council INFRA; upgrading the database platform) and HTAP (European Commission DG-ENV; import and export routines to build a secondary repository in support of <http://www.htap.org>; last access: 4 April 2019). A large number of specific projects have supported development of data and metadata reporting schemes in dialogue with data providers (EU; CREATE, ACTRIS and others). Through ACTRIS, the research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 654109. For a complete list of programmes and projects for which EBAS serves as a database, please consult the information box in the Framework filter of the web interface. These are all highly acknowledged for their support.

The observations of OC used in this study are collated from seven observational campaigns and supplemented by additional ground station observations. The campaign data were collated via GASSP and are derived from size distribution measurements taken during the following campaigns: VOCALS (NERC grant NE/F019874/1; Allen et al., 2011; Hawkins et al., 2010; Wood et al., 2011), CalNex (Ryerson et al., 2013), WACS (Quinn et al., 2014), ICEALOT (Frossard et al., 2011), DYNAMO (DeWitt et al., 2013), NEAQS-2004 (Quinn et al., 2006; Wang et al., 2007), TEXAQS06 (Bates et al., 2008), RHAMBLE (NERC grants NE/D006570/1, NE/E011454/1; Allan et al., 2009) and ACCACIA (NERC grant NE/I028696/1; Allan et al., 2015). The OC ground station observations used are from the AMS Global Database which has a worldwide cover-

age (Zhang et al., 2017), along with data for further European sites from ACTRIS (<https://www.actris.eu/>, last access: 17 September 2019; <http://ebas.nilu.no/>, last access: 17 September 2019; collated via GASSP) and data from individual stations including Chilbolton (England; Crippa et al., 2014), COPS (NERC grant NE/E016200/1; Hornisgrinde, Germany; Irwin et al., 2010; Jones et al., 2011), Holme Moss (England; Liu et al., 2011), OP3 (NERC grant NE/D004624/1; southeast Asia; Hewitt et al., 2010) and MC4 (NERC grant NE/H008136/1; Weybourne, England; Liu et al., 2013) collated via GASSP, the SORPES site in Nanjing, China (Ding et al., 2016), and AMF stations in the US and northeast Atlantic (Atmospheric Radiation Measurement (ARM) user facility, 2014a, b). The OC data at the AMF stations (US and northeast Atlantic) were obtained from the ARM programme sponsored by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, Climate and Environmental Sciences Division. The AMS Global Database ground station data cover the following sites: Barcelona (Minguillón et al., 2011; Mohr et al., 2012), Beijing (Sun et al., 2010), Blodgett Forest (Farmer et al., 2011), Boulder (Nemitz et al., 2008), Cape Hedo, Chebogue Point, Cheju Island (Topping et al., 2004), Chilbolton, Cool (Setyan et al., 2012), Duke Forest (Stroud et al., 2007), Edinburgh (Allan et al., 2003a), Finokalia (Hildebrandt et al., 2010), Fukue Island (Takami et al., 2005), Helsinki (Timonen et al., 2013), Houston (Canagaratna et al., 2007), Hyytiälä (Allan et al., 2006), Jungfraujoch (Ng et al., 2010), Komaba (Takegawa et al., 2006), K-Puszta, Mace Head (Dall’Osto et al., 2010), Mainz (Ng et al., 2010), Manaus (Chen et al., 2009), Manchester (Allan et al., 2003b), Melpitz (Poulain et al., 2011), Mexico City (Aiken et al., 2009), Montseny, New York City (Drewnick et al., 2004; Sun et al., 2011; Weimer et al., 2006), Pasadena (Hayes et al., 2013), Pinnacle State Park (Bae et al., 2007), Pittsburgh (Zhang et al., 2005), Point Reyes National Seashore – ARM Mobile Facility (AMF) (Ervens et al., 2010), Puy de Dome (Freney et al., 2011), Riverside – SOAR field site (Docherty et al., 2011; Williams et al., 2010), San Pietro Capofume, Storm Peak, Trinidad Head (Millet et al., 2004), Vancouver (Alfarra et al., 2004; Boudries et al., 2004), Weybourne Atmospheric Observatory, Whistler Mountain (Sun et al., 2009), Whiteface Mountain (Hogrefe et al., 2004) and Writtle Agricultural College.

The observations of  $N_{50}$  used in this study are collated from 19 observational campaigns and supplemented by additional ground station observations. The campaign data were collated via GASSP and derived from size distribution measurements taken during the following campaigns: ACE1 (Bates et al., 1998; Clarke et al., 1998), VOCALS (NERC grant NE/F019874/1; Allen et al., 2011; Hawkins et al., 2010; Wood et al., 2011), DOE ARM MAGIC (Lewis and Teixeira, 2015), CalNex (Ryerson et al., 2013), WACS (Quinn et al., 2014), NEAQS-2002 (Bates et al., 2005; Quinn and Bates, 2005), ARCTAS (McNaughton et al., 2011), ASCOS (Heintzenberg and Leck, 2012; Tjernström et al., 2014), ICEALOT (Frossard et al., 2011), AEROSOL99 (Bates et al., 2001), DYNAMO (DeWitt et al., 2013), INDOEX (Quinn and Bates, 2005; Ramanathan et al., 2001), PEM-Tropics-A (Fenn et al., 1999), PEM-Tropics-B (Raper et al., 2001), PASE (Hudson and Noble, 2009), NAURU99 (Long and McFarlane, 2012), ACE-ASIA (Bates et al., 2004; Huebert et al., 2003), NEAQS-2004 (Quinn et al., 2006; Wang et al., 2007) and TEXAQS06 (Bates et al., 2008). The  $N_{50}$  ground station observations used, collated via GASSP and public data on the EBAS database, are from Canada (Jeong et al., 2010; Leaitch et al., 2013; Takahama et al., 2011), South Africa (Vakkari et al., 2013), the Russian Arctic (Asmi et al., 2016), India (Hyvärinen et al., 2010), Antarctica (Fiebig et al., 2009) and European sites (Asmi et al., 2011). ASCOS (the Arctic Summer Cloud Ocean Study) was funded by the Knut and Alice Wallenberg Foundation and DAMOCLES (EU Sixth Framework Programme). The Swedish Polar Research Secretariat provided access to the icebreaker Oden and logistical support.

**Data availability.** Measurement data can be sourced from the databases and data providers as outlined in Appendix B and Table S1 in the Supplement. Model data and analysis code can be made available upon request to the corresponding author. The authors welcome use of the perturbed parameter ensemble for advancing climate research.

**Supplement.** The supplement related to this article is available online at: <https://doi.org/10.5194/acp-20-9491-2020-supplement>.

**Author contributions.** JSJ applied the statistical methodology and generated and analysed the results. JSJ, LAR and KSC wrote the article. JSJ, LAR, KSC, KJP, STT, JB, DMHS, JWR and NAJS contributed to the analysis and interpretation of results. KJP, MY, LAR, NAJS, DGP, KSC and JSJ helped prepare the model configuration that served as the template for the PPE. LAR and JSJ designed the experiments. All PPE simulations were created by MY. LAR and JSJ elicited probability density functions of all aerosol parameters, and KSC, KJP, MY and STT participated (alongside many other experts) in the formal elicitation process. DL, JDA, HC, AD, DDC, AA, VV and EA contributed measurement data to the study from individual measurement campaigns.

**Competing interests.** The authors declare that they have no conflict of interest.

**Acknowledgements.** We made use of the N8 HPC facility funded from the N8 consortium, an Engineering and Physical Sciences Research Council Grant to use ARCHER (EP/K000225/1) and the JASMIN facility (<http://www.jasmin.ac.uk/>; last access: 17 September 2019) via the Centre for Environmental Data Analysis funded by NERC and the UK Space Agency and delivered by the Science and Technology Facilities Council. We acknowledge the following additional funding: the Royal Society Wolfson Merit Award (Carslaw); a doctoral training grant from NERC and a CASE studentship with the Met Office Hadley Centre (Regayre).

**Financial support.** This research has been supported by the Natural Environment Research Council (grant nos. NE/J024252/1, NE/I020059/1, NE/P013406/1, NE/F019874/1, NE/D004624/1, NE/H008136/1, NE/D006570/1, NE/E011454/1, NE/E016200/1 and NE/I028696/1), the Newton Fund (grant no. UK–China Research and Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China), the National Centre for Atmospheric Science, the Engineering and Physical Sciences Research Council (grant no. EP/K000225/1), the European Union (grant no. ACTRIS-2 (262254)) and the Horizon 2020 (grant no. INTAROS (727890)).

**Review statement.** This paper was edited by Pedro Jimenez-Guerrero and reviewed by two anonymous referees.

## References

- Aiken, A. C., Salcedo, D., Cubison, M. J., Huffman, J. A., DeCarlo, P. F., Ulbrich, I. M., Docherty, K. S., Sueper, D., Kimmel, J. R., Worsnop, D. R., Trimborn, A., Northway, M., Stone, E. A., Schauer, J. J., Volkamer, R. M., Fortner, E., de Foy, B., Wang, J., Laskin, A., Shutthanandan, V., Zheng, J., Zhang, R., Gaffney, J., Marley, N. A., Paredes-Miranda, G., Arnott, W. P., Molina, L. T., Sosa, G., and Jimenez, J. L.: Mexico City aerosol analysis during MILAGRO using high resolution aerosol mass spectrometry at the urban supersite (T0) – Part 1: Fine particle composition and organic source apportionment, *Atmos. Chem. Phys.*, 9, 6633–6653, <https://doi.org/10.5194/acp-9-6633-2009>, 2009.
- Alfarra, M. R., Coe, H., Allan, J. D., Bower, K. N., Boudries, H., Canagaratna, M. R., Jimenez, J. L., Jayne, J. T., Garforth, A. A., Li, S.-M., and Worsnop, D. R.: Characterization of urban and rural organic particulate in the Lower Fraser Valley using two Aerodyne Aerosol Mass Spectrometers, *Atmos. Environ.*, 38, 5745–5758, <https://doi.org/10.1016/j.atmosenv.2004.01.054>, 2004.
- Allan, J. D., Jimenez, J. L., Williams, P. I., Alfarra, M. R., Bower, K. N., Jayne, J. T., Coe, H., and Worsnop, D. R.: Quantitative sampling using an Aerodyne aerosol mass spectrometer 1. Techniques of data interpretation and error analysis, *J. Geophys. Res.-Atmos.*, 108, 4090, <https://doi.org/10.1029/2002JD002358>, 2003a.
- Allan, J. D., Alfarra, M. R., Bower, K. N., Williams, P. I., Gallagher, M. W., Jimenez, J. L., McDonald, A. G., Nemitz, E., Canagaratna, M. R., Jayne, J. T., Coe, H., and Worsnop, D. R.: Quantitative sampling using an Aerodyne aerosol mass spectrometer 2. Measurements of fine particulate chemical composition in two U.K. cities, *J. Geophys. Res.-Atmos.*, 108, 4091, <https://doi.org/10.1029/2002JD002359>, 2003b.
- Allan, J. D., Alfarra, M. R., Bower, K. N., Coe, H., Jayne, J. T., Worsnop, D. R., Aalto, P. P., Kulmala, M., Hyötyläinen, T., Cavalli, F., and Laaksonen, A.: Size and composition measurements of background aerosol and new particle growth in a Finnish forest during QUEST 2 using an Aerodyne Aerosol Mass Spectrometer, *Atmos. Chem. Phys.*, 6, 315–327, <https://doi.org/10.5194/acp-6-315-2006>, 2006.
- Allan, J. D., Topping, D. O., Good, N., Irwin, M., Flynn, M., Williams, P. I., Coe, H., Baker, A. R., Martino, M., Niedermeier, N., Wiedensohler, A., Lehmann, S., Müller, K., Herrmann, H., and McFiggans, G.: Composition and properties of atmospheric particles in the eastern Atlantic and impacts on gas phase uptake rates, *Atmos. Chem. Phys.*, 9, 9299–9314, <https://doi.org/10.5194/acp-9-9299-2009>, 2009.
- Allan, J. D., Williams, P. I., Najera, J., Whitehead, J. D., Flynn, M. J., Taylor, J. W., Liu, D., Darbyshire, E., Carpenter, L. J., Chance, R., Andrews, S. J., Hackenberg, S. C., and McFiggans, G.: Iodine observed in new particle formation events in the Arctic atmosphere during ACCACIA, *Atmos. Chem. Phys.*, 15, 5599–5609, <https://doi.org/10.5194/acp-15-5599-2015>, 2015.
- Allen, G., Coe, H., Clarke, A., Bretherton, C., Wood, R., Abel, S. J., Barrett, P., Brown, P., George, R., Freitag, S., McNaughton, C., Howell, S., Shank, L., Kapustin, V., Brekhovskikh, V., Kleinman, L., Lee, Y.-N., Springston, S., Toniazzo, T., Krejci, R., Fochesatto, J., Shaw, G., Krecl, P., Brooks, B., McMeeking, G., Bower, K. N., Williams, P. I., Crosier, J., Crawford, I., Connolly, P., Allan, J. D., Covert, D., Bandy, A. R., Russell, L. M., Trem-

- bath, J., Bart, M., McQuaid, J. B., Wang, J., and Chand, D.: South East Pacific atmospheric composition and variability sampled along 20° S during VOCALS-REx, *Atmos. Chem. Phys.*, 11, 5237–5262, <https://doi.org/10.5194/acp-11-5237-2011>, 2011.
- Andres, R. J. and Kasgnoc, A. D.: A time-averaged inventory of subaerial volcanic sulfur emissions, *J. Geophys. Res.-Atmos.*, 103, 25251–25261, <https://doi.org/10.1029/98JD02091>, 1998.
- Andrianakis, I., Vernon, I., McCreech, N., McKinley, T. J., Oakley, J. E., Nsubuga, R. N., Goldstein, M., and White, R. G.: History matching of a complex epidemiological model of human immunodeficiency virus transmission by using variance emulation, *J. R. Stat. Soc. Ser. C Appl. Stat.*, 66, 717–740, <https://doi.org/10.1111/rssc.12198>, 2017.
- Artaxo, P., Rizzo, L. V., Brito, J. F., Barbosa, H. M. J., Arana, A., Sena, E. T., Cirino, G. G., Bastos, W., Martin, S. T., and Andreae, M. O.: Atmospheric aerosols in Amazonia and land use change: from natural biogenic to biomass burning conditions, *Faraday Discuss.*, 165, 203–235, <https://doi.org/10.1039/C3FD00052D>, 2013.
- Asmi, A., Wiedensohler, A., Laj, P., Fjaeraa, A.-M., Sellegri, K., Birmili, W., Weingartner, E., Baltensperger, U., Zdimal, V., Zikova, N., Putaud, J.-P., Marinoni, A., Tunved, P., Hansson, H.-C., Fiebig, M., Kivekäs, N., Lihavainen, H., Asmi, E., Ulevicius, V., Aalto, P. P., Swietlicki, E., Kristensson, A., Mihalopoulos, N., Kalivitis, N., Kalapov, I., Kiss, G., de Leeuw, G., Henzing, B., Harrison, R. M., Beddows, D., O'Dowd, C., Jennings, S. G., Flentje, H., Weinhold, K., Meinhardt, F., Ries, L., and Kulmala, M.: Number size distributions and seasonality of submicron particles in Europe 2008–2009, *Atmos. Chem. Phys.*, 11, 5505–5538, <https://doi.org/10.5194/acp-11-5505-2011>, 2011.
- Asmi, A., Collaud Coen, M., Ogren, J. A., Andrews, E., Sheridan, P., Jefferson, A., Weingartner, E., Baltensperger, U., Bukowiecki, N., Lihavainen, H., Kivekäs, N., Asmi, E., Aalto, P. P., Kulmala, M., Wiedensohler, A., Birmili, W., Hamed, A., O'Dowd, C., G Jennings, S., Weller, R., Flentje, H., Fjaeraa, A. M., Fiebig, M., Myhre, C. L., Hallar, A. G., Swietlicki, E., Kristensson, A., and Laj, P.: Aerosol decadal trends – Part 2: In-situ aerosol particle number concentrations at GAW and ACTRIS stations, *Atmos. Chem. Phys.*, 13, 895–916, <https://doi.org/10.5194/acp-13-895-2013>, 2013.
- Asmi, E., Kondratyev, V., Brus, D., Laurila, T., Lihavainen, H., Backman, J., Vakkari, V., Aurela, M., Hatakka, J., Viisanen, Y., Uttal, T., Ivakhov, V., and Makshtas, A.: Aerosol size distribution seasonal characteristics measured in Tiksi, Russian Arctic, *Atmos. Chem. Phys.*, 16, 1271–1287, <https://doi.org/10.5194/acp-16-1271-2016>, 2016.
- Atanacio, A. J., Cohen, D. D., Begum, B. A., Ni, B., Pandit, G. G., Sahu, S. K., Santoso, M., Lestiani, D. D., Lim, J. M., Rahman, S. A., Elias, M. S., Shagjjamba, D., Markwitz, A., Siddique, N., Pabroa, P. C., Santos, F. L., Seneviratne, M. C. S., Handagiri-pathira, L., Wimolwattanapun, W., Vuong, T. B., and Karydas, A.: The APAD and ASFID?: Long-term fine and coarse ambient particulate matter and source fingerprint databases for the Asia-Pacific Region, *Air Qual. Clim. Chang.*, 50, 41–49, 2016.
- Atmospheric Radiation Measurement (ARM) user facility: Updated hourly. Aerosol Chemical Speciation Monitor (AOSACSM), 2014-07-10 to 2015-08-11, Eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal (C1), compiled by: Behrens, B., Salwen, C., Springston, S., and Watson, T., ARM Data Center, available at: <https://www.arm.gov/> (last access: 17 September 2019), 2014a.
- Atmospheric Radiation Measurement (ARM) user facility: Updated hourly. Aerosol Chemical Speciation Monitor (AOSACSM), 2014-10-01 to 2015-08-11, Southern Great Plains (SGP) Central Facility, Lamont, OK (C1), compiled by: Behrens, B., Salwen, C., Springston, S., and Watson, T., ARM Data Center, available at: <https://www.arm.gov/> (last access: 17 September 2019), 2014b.
- Bae, M.-S., Schwab, J. J., Zhang, Q., Hogrefe, O., Demerjian, K. L., Weimer, S., Rhoads, K., Orsini, D., Venkatachari, P., and Hopke, P. K.: Interference of organic signals in highly time resolved nitrate measurements by low mass resolution aerosol mass spectrometry, *J. Geophys. Res.*, 112, D22305, <https://doi.org/10.1029/2007JD008614>, 2007.
- Bates, T. S., Huebert, B. J., Gras, J. L., Griffiths, F. B., and Durkee, P. A.: International Global Atmospheric Chemistry (IGAC) Project's First Aerosol Characterization Experiment (ACE 1): Overview, *J. Geophys. Res.-Atmos.*, 103, 16297–16318, <https://doi.org/10.1029/97JD03741>, 1998.
- Bates, T. S., Quinn, P. K., Coffman, D. J., Johnson, J. E., Miller, T. L., Covert, D. S., Wiedensohler, A., Leinert, S., Nowak, A., and Neusüss, C.: Regional physical and chemical properties of the marine boundary layer aerosol across the Atlantic during Aerosols99: An overview, *J. Geophys. Res.-Atmos.*, 106, 20767–20782, <https://doi.org/10.1029/2000JD900578>, 2001.
- Bates, T. S., Quinn, P. K., Coffman, D. J., Covert, D. S., Miller, T. L., Johnson, J. E., Carmichael, G. R., Uno, I., Guazzotti, S. A., Sodeman, D. A., Prather, K. A., Rivera, M., Russell, L. M., and Merrill, J. T.: Marine boundary layer dust and pollutant transport associated with the passage of a frontal system over eastern Asia, *J. Geophys. Res.*, 109, D19S19, <https://doi.org/10.1029/2003JD004094>, 2004.
- Bates, T. S., Quinn, P. K., Coffman, D. J., Johnson, J. E., and Middlebrook, A. M.: Dominance of organic aerosols in the marine boundary layer over the Gulf of Maine during NEAQS 2002 and their role in aerosol light scattering, *J. Geophys. Res.*, 110, D18202, <https://doi.org/10.1029/2005JD005797>, 2005.
- Bates, T. S., Quinn, P. K., Coffman, D., Schulz, K., Covert, D. S., Johnson, J. E., Williams, E. J., Lerner, B. M., Angevine, W. M., Tucker, S. C., Brewer, W. A., and Stohl, A.: Boundary layer aerosol chemistry during TexAQS/GoMACCS 2006: Insights into aerosol sources and transformation processes, *J. Geophys. Res.*, 113, D00F01, <https://doi.org/10.1029/2008JD010023>, 2008.
- Boudries, H., Canagaratna, M. R., Jayne, J. T., Alfarra, M. R., Allan, J., Bower, K. N., Coe, H., Pryor, S. C., Jimenez, J. L., Brook, J. R., Li, S., and Worsnop, D. R.: Chemical and physical processes controlling the distribution of aerosols in the Lower Fraser Valley, Canada, during the Pacific 2001 field campaign, *Atmos. Environ.*, 38, 5759–5774, <https://doi.org/10.1016/j.atmosenv.2004.01.057>, 2004.
- Boutle, I. A., Abel, S. J., Hill, P. G., and Morcrette, C. J.: Spatial variability of liquid cloud and rain: observations and microphysical effects, *Q. J. R. Meteorol. Soc.*, 140, 583–594, <https://doi.org/10.1002/qj.2140>, 2014.
- Browse, J., Carslaw, K. S., Lee, L. A., Reddington, C. R., Regayre, L. A., Pringle, K. J., Johnson, J. S., Sangeeta, S., Ogren, J., Sherman, J., Labuschagne, C., Mayol-Bracero, O. L., Mihalopoulos,

- N., Kouvarakis, G., Hillamo, R., Sang-Woo, K., Harrison, R., Kaminski, U., Kulmala, M., Rodriguez, S., Baltensperger, U., Sherrer, P., Monahan, C., Marinoni, A., Bonasoni, P., Weller, R., Kivekas, N., Andriejauskienė, J., Sellegri, K., Gannett-Hallar, A., Tunved, P., Cohen, D., Vakkari, V., Leaitch, R., Laurila, T., Hyvarien, A., Lunder, C., Fiebig, M., Asmi, A., Quinn, P., Johnson, J., Clarke, A., Bates, T. S., Lewis, E., and Leck, C.: Evaluating a Perturbed Parameter Ensemble against global aerosol observations to constrain parametric uncertainty: Is there an ideal aerosol model?, *Atmos. Chem. Phys. Discuss.*, in preparation, 2020.
- Canagaratna, M. R., Jayne, J. T., Jimenez, J. L., Allan, J. D., Alfarra, M. R., Zhang, Q., Onasch, T. B., Drewnick, F., Coe, H., Middlebrook, A., Delia, A., Williams, L. R., Trimborn, A. M., Northway, M. J., DeCarlo, P. F., Kolb, C. E., Davidovits, P., and Worsnop, D. R.: Chemical and microphysical characterization of ambient aerosols with the aerodyne aerosol mass spectrometer, *Mass Spectrom. Rev.*, 26, 185–222, <https://doi.org/10.1002/mas.20115>, 2007.
- Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., Mann, G. W., Spracklen, D. V., Woodhouse, M. T., Regayre, L. A., and Pierce, J. R.: Large contribution of natural aerosols to uncertainty in indirect forcing, *Nature*, 503, 67–71, <https://doi.org/10.1038/nature12674>, 2013a.
- Carslaw, K. S., Lee, L. A., Reddington, C. L., Mann, G. W., and Pringle, K. J.: The magnitude and sources of uncertainty in global aerosol, *Faraday Discuss.*, 165, 495–512, <https://doi.org/10.1039/c3fd00043e>, 2013b.
- Carslaw, K. S., Gordon, H., Hamilton, D. S., Johnson, J. S., Regayre, L. A., Yoshioka, M., and Pringle, K. J.: Aerosols in the Pre-industrial Atmosphere, *Curr. Clim. Chang. Reports*, 3, 1–15, <https://doi.org/10.1007/s40641-017-0061-2>, 2017.
- Carslaw, K. S., Lee, L. A., Regayre, L. A., and Johnson, J. S.: Climate Models Are Uncertain, but We Can Do Something About It, *Eos (Washington, DC)*, 99, <https://doi.org/10.1029/2018EO093757>, 2018.
- Chen, Q., Farmer, D. K., Schneider, J., Zorn, S. R., Heald, C. L., Karl, T. G., Guenther, A., Allan, J. D., Robinson, N., Coe, H., Kimmel, J. R., Pauliquevis, T., Borrmann, S., Pöschl, U., Andreae, M. O., Artaxo, P., Jimenez, J. L., and Martin, S. T.: Mass spectral characterization of submicron biogenic organic particles in the Amazon Basin, *Geophys. Res. Lett.*, 36, L20806, <https://doi.org/10.1029/2009GL039880>, 2009.
- Clarke, A. D., Varner, J. L., Eisele, F., Mauldin, R. L., Tanner, D., and Litchy, M.: Particle production in the remote marine atmosphere: Cloud outflow and subsidence during ACE 1, *J. Geophys. Res.-Atmos.*, 103, 16397–16409, <https://doi.org/10.1029/97JD02987>, 1998.
- Cohen, D. D. and Atanacio, A. J.: The IAEA/RCA Fine and Coarse Particle Ambient Air Database, ANSTO report/E-784, ISBN: 1921268247, 2015.
- Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A.: Bayes linear strategies for history matching of hydrocarbon reservoirs, in: *Bayesian Statistics 5*, edited by: J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 69–95, Clarendon Press, Oxford, UK, 1996.
- Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A.: Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments, in: *Case Studies in Bayesian Statistics, Lecture Notes in Statistics*, edited by: Gatsonis, C., Hodges, J. S., Kass, R. E., McCulloch, R., Rossi, P., and Singpurwalla, N. D., 121, 37–93, Springer, New York, NY, 1997.
- Crippa, M., Canonaco, F., Lanz, V. A., Äijälä, M., Allan, J. D., Carbone, S., Capes, G., Ceburnis, D., Dall’Osto, M., Day, D. A., DeCarlo, P. F., Ehn, M., Eriksson, A., Freney, E., Hildebrandt Ruiz, L., Hillamo, R., Jimenez, J. L., Junninen, H., Kiendler-Scharr, A., Kortelainen, A.-M., Kulmala, M., Laaksonen, A., Mensah, A. A., Mohr, C., Nemitz, E., O’Dowd, C., Ovadnevaite, J., Pandis, S. N., Petäjä, T., Poulain, L., Saarikoski, S., Sellegri, K., Swietlicki, E., Tiitta, P., Worsnop, D. R., Baltensperger, U., and Prévôt, A. S. H.: Organic aerosol components derived from 25 AMS data sets across Europe using a consistent ME-2 based source apportionment approach, *Atmos. Chem. Phys.*, 14, 6159–6176, <https://doi.org/10.5194/acp-14-6159-2014>, 2014.
- Dall’Osto, M., Ceburnis, D., Martucci, G., Bialek, J., Dupuy, R., Jennings, S. G., Berresheim, H., Wenger, J., Healy, R., Facchini, M. C., Rinaldi, M., Giulianelli, L., Finessi, E., Worsnop, D., Ehn, M., Mikkilä, J., Kulmala, M., and O’Dowd, C. D.: Aerosol properties associated with air masses arriving into the North East Atlantic during the 2008 Mace Head EUCAARI intensive observing period: an overview, *Atmos. Chem. Phys.*, 10, 8413–8435, <https://doi.org/10.5194/acp-10-8413-2010>, 2010.
- DeWitt, H. L., Coffman, D. J., Schulz, K. J., Brewer, W. A., Bates, T. S., and Quinn, P. K.: Atmospheric properties over the equatorial Indian Ocean and the impact of the Madden-Julian Oscillation, *J. Geophys. Res.-Atmos.*, 118, 5736–5749, <https://doi.org/10.1002/jgrd.50419>, 2013.
- Ding, A., Nie, W., Huang, X., Chi, X., Sun, J., Kerminen, V.-M., Xu, Z., Guo, W., Petäjä, T., Yang, X., Kulmala, M., and Fu, C.: Long-term observation of air pollution-weather/climate interactions at the SORPES station: a review and outlook, *Front. Environ. Sci. Eng.*, 10, 15, <https://doi.org/10.1007/s11783-016-0877-3>, 2016.
- Docherty, K. S., Aiken, A. C., Huffman, J. A., Ulbrich, I. M., DeCarlo, P. F., Sueper, D., Worsnop, D. R., Snyder, D. C., Peltier, R. E., Weber, R. J., Grover, B. D., Eatough, D. J., Williams, B. J., Goldstein, A. H., Ziemann, P. J., and Jimenez, J. L.: The 2005 Study of Organic Aerosols at Riverside (SOAR-1): instrumental intercomparisons and fine particle composition, *Atmos. Chem. Phys.*, 11, 12387–12420, <https://doi.org/10.5194/acp-11-12387-2011>, 2011.
- Drewnick, F., Schwab, J. J., Jayne, J. T., Canagaratna, M., Worsnop, D. R., and Demerjian, K. L.: Measurement of Ambient Aerosol Composition During the PMTACS-NY 2001 Using an Aerosol Mass Spectrometer. Part I: Mass Concentrations Special Issue of Aerosol Science and Technology on Findings from the Fine Particulate Matter Supersites Program, *Aerosol Sci. Tech.*, 38, 92–103, <https://doi.org/10.1080/02786820390229507>, 2004.
- Edwards, N. R., Cameron, D., and Rougier, J.: Precalibrating an intermediate complexity climate model, *Clim. Dynam.*, 37, 1469–1482, <https://doi.org/10.1007/s00382-010-0921-0>, 2011.
- Emerson, E. W., Katich, J. M., Schwarz, J. P., McMeeking, G. R., and Farmer, D. K.: Direct Measurements of Dry and Wet Deposition of Black Carbon Over a Grassland, *J. Geophys. Res.-Atmos.*, 123, 12277–12290, <https://doi.org/10.1029/2018JD028954>, 2018.
- Ervens, B., Cubison, M. J., Andrews, E., Feingold, G., Ogren, J. A., Jimenez, J. L., Quinn, P. K., Bates, T. S., Wang, J., Zhang, Q., Coe, H., Flynn, M., and Allan, J. D.: CCN predic-

- tions using simplified assumptions of organic aerosol composition and mixing state: a synthesis from six different locations, *Atmos. Chem. Phys.*, 10, 4795–4807, <https://doi.org/10.5194/acp-10-4795-2010>, 2010.
- Fang, G.-C. and Chang, S.-C.: Atmospheric particulate (PM<sub>10</sub> and PM<sub>2.5</sub>) mass concentration and seasonal variation study in the Taiwan area during 2000–2008, *Atmos. Res.*, 98, 368–377, <https://doi.org/10.1016/J.ATMOSRES.2010.07.005>, 2010.
- Farmer, D. K., Kimmel, J. R., Phillips, G., Docherty, K. S., Worsnop, D. R., Sueper, D., Nemitz, E., and Jimenez, J. L.: Eddy covariance measurements with high-resolution time-of-flight aerosol mass spectrometry: a new approach to chemically resolved aerosol fluxes, *Atmos. Meas. Tech.*, 4, 1275–1289, <https://doi.org/10.5194/amt-4-1275-2011>, 2011.
- Fenn, M. A., Browell, E. V., Butler, C. F., Grant, W. B., Kooi, S. A., Clayton, M. B., Gregory, G. L., Newell, R. E., Zhu, Y., Dibb, J. E., Fuelberg, H. E., Anderson, B. E., Bandy, A. R., Blake, D. R., Bradshaw, J. D., Heikes, B. G., Sachse, G. W., Sandholm, S. T., Singh, H. B., Talbot, R. W., and Thornton, D. C.: Ozone and aerosol distributions and air mass characteristics over the South Pacific during the burning season, *J. Geophys. Res.-Atmos.*, 104, 16197–16212, <https://doi.org/10.1029/1999JD900065>, 1999.
- Fiebig, M., Lunder, C. R., and Stohl, A.: Tracing biomass burning aerosol from South America to Troll Research Station, Antarctica, *Geophys. Res. Lett.*, 36, L14815, <https://doi.org/10.1029/2009GL038531>, 2009.
- Fiedler, S., Kinne, S., Huang, W. T. K., Räisänen, P., O'Donnell, D., Bellouin, N., Stier, P., Merikanto, J., van Noije, T., Makkonen, R., and Lohmann, U.: Anthropogenic aerosol forcing – insights from multiple estimates from aerosol-climate models with reduced complexity, *Atmos. Chem. Phys.*, 19, 6821–6841, <https://doi.org/10.5194/acp-19-6821-2019>, 2019.
- Frenay, E. J., Sellegri, K., Canonaco, F., Boulon, J., Hervo, M., Weigel, R., Pichon, J. M., Colomb, A., Prévôt, A. S. H., and Laj, P.: Seasonal variations in aerosol particle composition at the puy-de-Dôme research station in France, *Atmos. Chem. Phys.*, 11, 13047–13059, <https://doi.org/10.5194/acp-11-13047-2011>, 2011.
- Frossard, A. A., Shaw, P. M., Russell, L. M., Kroll, J. H., Canagaratna, M. R., Worsnop, D. R., Quinn, P. K., and Bates, T. S.: Springtime Arctic haze contributions of submicron organic particles from European and Asian combustion sources, *J. Geophys. Res.*, 116, D05205, <https://doi.org/10.1029/2010JD015178>, 2011.
- Galarneau, E., Wang, D., Dabek-Zlotorzynska, E., Siu, M., Celo, V., Tardif, M., Harnish, D., and Jiang, Y.: Air toxics in Canada measured by the National Air Pollution Surveillance (NAPS) program and their relation to ambient air quality guidelines, *J. Air Waste Manage. Assoc.*, 66, 184–200, <https://doi.org/10.1080/10962247.2015.1096863>, 2016.
- Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements, *Atmos. Meas. Tech.*, 12, 169–209, <https://doi.org/10.5194/amt-12-169-2019>, 2019.
- Halmer, M. M., Schmincke, H.-U., and Graf, H.-F.: The annual volcanic gas input into the atmosphere, in particular into the stratosphere: a global data set for the past 100 years, *J. Volcanol. Geotherm. Res.*, 115, 511–528, [https://doi.org/10.1016/S0377-0273\(01\)00318-3](https://doi.org/10.1016/S0377-0273(01)00318-3), 2002.
- Hawkins, L. N., Russell, L. M., Covert, D. S., Quinn, P. K., and Bates, T. S.: Carboxylic acids, sulfates, and organosulfates in processed continental organic aerosol over the southeast Pacific Ocean during VOCALS-REx 2008, *J. Geophys. Res.*, 115, D13201, <https://doi.org/10.1029/2009JD013276>, 2010.
- Hayes, P. L., Ortega, A. M., Cubison, M. J., Froyd, K. D., Zhao, Y., Cliff, S. S., Hu, W. W., Toohey, D. W., Flynn, J. H., Lefer, B. L., Grossberg, N., Alvarez, S., Rappenglück, B., Taylor, J. W., Allan, J. D., Holloway, J. S., Gilman, J. B., Kuster, W. C., de Gouw, J. A., Massoli, P., Zhang, X., Liu, J., Weber, R. J., Corrigan, A. L., Russell, L. M., Isaacman, G., Worton, D. R., Kreisberg, N. M., Goldstein, A. H., Thalman, R., Waxman, E. M., Volkamer, R., Lin, Y. H., Surratt, J. D., Kleindienst, T. E., Offenberg, J. H., Dusanter, S., Griffith, S., Stevens, P. S., Brioude, J., Angevine, W. M., and Jimenez, J. L.: Organic aerosol composition and sources in Pasadena, California, during the 2010 CalNex campaign, *J. Geophys. Res.-Atmos.*, 118, 9233–9257, <https://doi.org/10.1002/jgrd.50530>, 2013.
- Heintzenberg, J. and Leck, C.: The summer aerosol in the central Arctic 1991–2008: did it change or not?, *Atmos. Chem. Phys.*, 12, 3969–3983, <https://doi.org/10.5194/acp-12-3969-2012>, 2012.
- Hewitt, C. N., Lee, J. D., MacKenzie, A. R., Barkley, M. P., Carslaw, N., Carver, G. D., Chappell, N. A., Coe, H., Collier, C., Commane, R., Davies, F., Davison, B., DiCarlo, P., Di Marco, C. F., Dorsey, J. R., Edwards, P. M., Evans, M. J., Fowler, D., Furneaux, K. L., Gallagher, M., Guenther, A., Heard, D. E., Helfter, C., Hopkins, J., Ingham, T., Irwin, M., Jones, C., Karunaharan, A., Langford, B., Lewis, A. C., Lim, S. F., MacDonald, S. M., Mahajan, A. S., Malpass, S., McFiggans, G., Mills, G., Misztal, P., Moller, S., Monks, P. S., Nemitz, E., Nicolas-Perea, V., Oetjen, H., Oram, D. E., Palmer, P. I., Phillips, G. J., Pike, R., Plane, J. M. C., Pugh, T., Pyle, J. A., Reeves, C. E., Robinson, N. H., Stewart, D., Stone, D., Whalley, L. K., and Yin, X.: Overview: oxidant and particle photochemical processes above a south-east Asian tropical rainforest (the OP3 project): introduction, rationale, location characteristics and tools, *Atmos. Chem. Phys.*, 10, 169–199, <https://doi.org/10.5194/acp-10-169-2010>, 2010.
- Hewitt, H. T., Copley, D., Culverwell, I. D., Harris, C. M., Hill, R. S. R., Keen, A. B., McLaren, A. J., and Hunke, E. C.: Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system, *Geosci. Model Dev.*, 4, 223–253, <https://doi.org/10.5194/gmd-4-223-2011>, 2011.
- Hildebrandt, L., Kostenidou, E., Mihalopoulos, N., Worsnop, D. R., Donahue, N. M., and Pandis, S. N.: Formation of highly oxygenated organic aerosol in the atmosphere: Insights from the Finokalia Aerosol Measurement Experiments, *Geophys. Res. Lett.*, 37, L23801, <https://doi.org/10.1029/2010GL045193>, 2010.
- Hogrefe, O., Schwab, J. J., Drewnick, F., Lala, G. G., Peters, S., Demerjian, K. L., Rhoads, K., Felton, H. D., Rattigan, O. V., Husain, L., and Dutkiewicz, V. A.: Semicontinuous PM<sub>2.5</sub> Sulfate and Nitrate Measurements at an Urban and a Rural Location in New York: PMTACS-NY Summer 2001 and

- 2002 Campaigns, *J. Air Waste Manage. Assoc.*, 54, 1040–1060, <https://doi.org/10.1080/10473289.2004.10470972>, 2004.
- Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – A Federated Instrument Network and Data Archive for Aerosol Characterization, *Remote Sens. Environ.*, 66, 1–16, [https://doi.org/10.1016/S0034-4257\(98\)00031-5](https://doi.org/10.1016/S0034-4257(98)00031-5), 1998.
- Hudson, J. G. and Noble, S.: CCN and cloud droplet concentrations at a remote ocean site, *Geophys. Res. Lett.*, 36(13), L13812, <https://doi.org/10.1029/2009GL038465>, 2009.
- Huebert, B. J., Bates, T., Russell, P. B., Shi, G., Kim, Y. J., Kawamura, K., Carmichael, G., and Nakajima, T.: An overview of ACE-Asia: Strategies for quantifying the relationships between Asian aerosols and their climatic impacts, *J. Geophys. Res.*, 108, 8633, <https://doi.org/10.1029/2003JD003550>, 2003.
- Hyvärinen, A.-P., Lihavainen, H., Komppula, M., Panwar, T. S., Sharma, V. P., Hooda, R. K., and Viisanen, Y.: Aerosol measurements at the Gual Pahari EUCAARI station: preliminary results from in-situ measurements, *Atmos. Chem. Phys.*, 10, 7241–7252, <https://doi.org/10.5194/acp-10-7241-2010>, 2010.
- Irwin, M., Good, N., Crosier, J., Choularton, T. W., and McFiggans, G.: Reconciliation of measurements of hygroscopic growth and critical supersaturation of aerosol particles in central Germany, *Atmos. Chem. Phys.*, 10, 11737–11752, <https://doi.org/10.5194/acp-10-11737-2010>, 2010.
- Jeong, C.-H., Evans, G. J., McGuire, M. L., Chang, R. Y.-W., Abbatt, J. P. D., Zeromskiene, K., Mozurkewich, M., Li, S.-M., and Leaitch, W. R.: Particle formation and growth at five rural and urban sites, *Atmos. Chem. Phys.*, 10, 7979–7995, <https://doi.org/10.5194/acp-10-7979-2010>, 2010.
- Johnson, J. S., Regayre, L. A., Yoshioka, M., Pringle, K. J., Lee, L. A., Sexton, D. M. H., Rostron, J. W., Booth, B. B. B., and Carslaw, K. S.: The importance of comprehensive parameter sampling and multiple observations for robust constraint of aerosol radiative forcing, *Atmos. Chem. Phys.*, 18, 13031–13053, <https://doi.org/10.5194/acp-18-13031-2018>, 2018.
- Jones, H. M., Crosier, J., Russell, A., Flynn, M. J., Irwin, M., Choularton, T. W., Coe, H., and McFiggans, G.: In situ aerosol measurements taken during the 2007 COPS field campaign at the Hornsgrinde ground site, *Q. J. Roy. Meteor. Soc.*, 137, 252–266, <https://doi.org/10.1002/qj.727>, 2011.
- Kerminen, V.-M., Chen, X., Vakkari, V., Petäjä, T., Kulmala, M., and Bianchi, F.: Atmospheric new particle formation and growth: review of field observations, *Environ. Res. Lett.*, 13, 103003, <https://doi.org/10.1088/1748-9326/aadf3c>, 2018.
- Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, *Atmos. Chem. Phys.*, 10, 7017–7039, <https://doi.org/10.5194/acp-10-7017-2010>, 2010.
- Leaitch, W. R., Sharma, S., Huang, L., Toom-Sauntry, D., Chivulescu, A., Macdonald, A. M., von Salzen, K., Pierce, J. R., Bertram, A. K., Schroder, J. C., Shantz, N. C., Chang, R. Y.-W., and Norman, A.-L.: Dimethyl sulfide control of the clean summertime Arctic aerosol and cloud, *Elem. Sci. Anthr.*, 1, 000017, <https://doi.org/10.12952/journal.elementa.000017>, 2013.
- Lebsack, M., Morrison, H., and Gettelman, A.: Microphysical implications of cloud-precipitation covariance derived from satellite remote sensing, *J. Geophys. Res.-Atmos.*, 118, 6521–6533, <https://doi.org/10.1002/jgrd.50347>, 2013.
- Lee, C., Martin, R. V., van Donkelaar, A., Lee, H., Dickerson, R. R., Hains, J. C., Krotkov, N., Richter, A., Vinnikov, K., and Schwab, J. J.: SO<sub>2</sub> emissions and lifetimes: Estimates from inverse modeling using in situ and global, space-based (SCIAMACHY and OMI) observations, *J. Geophys. Res.*, 116, D06304, <https://doi.org/10.1029/2010JD014758>, 2011.
- Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., and Spracklen, D. V.: Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters, *Atmos. Chem. Phys.*, 11, 12253–12273, <https://doi.org/10.5194/acp-11-12253-2011>, 2011.
- Lee, L. A., Pringle, K. J., Reddington, C. L., Mann, G. W., Stier, P., Spracklen, D. V., Pierce, J. R., and Carslaw, K. S.: The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei, *Atmos. Chem. Phys.*, 13, 8879–8914, <https://doi.org/10.5194/acp-13-8879-2013>, 2013.
- Lee, L. A., Reddington, C. L., and Carslaw, K. S.: On the relationship between aerosol model uncertainty and radiative forcing uncertainty, *P. Natl. Acad. Sci. USA*, 113, 5820–5827, <https://doi.org/10.1073/pnas.1507050113>, 2016.
- Lewis, E. and Teixeira, J.: *Dispelling Clouds of Uncertainty*, Eos (Washington, DC), 96, <https://doi.org/10.1029/2015EO031303>, 2015.
- Liu, D., Allan, J., Corris, B., Flynn, M., Andrews, E., Ogren, J., Beswick, K., Bower, K., Burgess, R., Choularton, T., Dorsey, J., Morgan, W., Williams, P. I., and Coe, H.: Carbonaceous aerosols contributed by traffic and solid fuel burning at a polluted rural site in Northwestern England, *Atmos. Chem. Phys.*, 11, 1603–1619, <https://doi.org/10.5194/acp-11-1603-2011>, 2011.
- Liu, D., Allan, J., Whitehead, J., Young, D., Flynn, M., Coe, H., McFiggans, G., Fleming, Z. L., and Bandy, B.: Ambient black carbon particle hygroscopic properties controlled by mixing state and composition, *Atmos. Chem. Phys.*, 13, 2015–2029, <https://doi.org/10.5194/acp-13-2015-2013>, 2013.
- Liu, F., Choi, S., Li, C., Fioletov, V. E., McLinden, C. A., Joiner, J., Krotkov, N. A., Bian, H., Janssens-Maenhout, G., Darmenov, A. S., and da Silva, A. M.: A new global anthropogenic SO<sub>2</sub> emission inventory for the last decade: a mosaic of satellite-derived and bottom-up emissions, *Atmos. Chem. Phys.*, 18, 16571–16586, <https://doi.org/10.5194/acp-18-16571-2018>, 2018.
- Long, C. N. and McFarlane, S. A.: Quantification of the Impact of Nauru Island on ARM Measurements, *J. Appl. Meteorol. Clim.*, 51, 628–636, <https://doi.org/10.1175/JAMC-D-11-0174.1>, 2012.
- 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, <https://doi.org/10.5194/gmd-3-519-2010>, 2010.
- Mann, G. W., Carslaw, K. S., Reddington, C. L., Pringle, K. J., Schulz, M., Asmi, A., Spracklen, D. V., Ridley, D. A., Woodhouse, M. T., Lee, L. A., Zhang, K., Ghan, S. J., Easter, R. C., Liu, X., Stier, P., Lee, Y. H., Adams, P. J., Tost, H., Lelieveld,

- J., Bauer, S. E., Tsigaridis, K., van Noije, T. P. C., Strunk, A., Vignati, E., Bellouin, N., Dalvi, M., Johnson, C. E., Bergman, T., Kokkola, H., von Salzen, K., Yu, F., Luo, G., Petzold, A., Heintzenberg, J., Clarke, A., Ogren, J. A., Gras, J., Baltensperger, U., Kaminski, U., Jennings, S. G., O'Dowd, C. D., Harrison, R. M., Beddows, D. C. S., Kulmala, M., Viisanen, Y., Ulevicius, V., Mihalopoulos, N., Zdimal, V., Fiebig, M., Hansson, H.-C., Swietlicki, E., and Henzing, J. S.: Intercomparison and evaluation of global aerosol microphysical properties among AeroCom models of a range of complexity, *Atmos. Chem. Phys.*, 14, 4679–4713, <https://doi.org/10.5194/acp-14-4679-2014>, 2014.
- McNaughton, C. S., Clarke, A. D., Freitag, S., Kapustin, V. N., Kondo, Y., Moteki, N., Sahu, L., Takegawa, N., Schwarz, J. P., Spackman, J. R., Watts, L., Diskin, G., Podolske, J., Holloway, J. S., Wisthaler, A., Mikoviny, T., de Gouw, J., Warneke, C., Jimenez, J., Cubison, M., Howell, S. G., Middlebrook, A., Bahreini, R., Anderson, B. E., Winstead, E., Thornhill, K. L., Lack, D., Cozic, J., and Brock, C. A.: Absorbing aerosol in the troposphere of the Western Arctic during the 2008 ARCTAS/ARCPAC airborne field campaigns, *Atmos. Chem. Phys.*, 11, 7561–7582, <https://doi.org/10.5194/acp-11-7561-2011>, 2011.
- McNeill, D., Williams, J., Booth, B., Betts, R., Challenor, P., Wiltshire, A., and Sexton, D.: The impact of structural error on parameter constraint in a climate model, *Earth Syst. Dynam.*, 7, 917–935, <https://doi.org/10.5194/esd-7-917-2016>, 2016.
- Metzger, A., Verheggen, B., Dommen, J., Duplissy, J., Prevot, A. S. H., Weingartner, E., Riipinen, I., Kulmala, M., Spracklen, D. V., Carslaw, K. S., and Baltensperger, U.: Evidence for the role of organics in aerosol particle formation under atmospheric conditions, *P. Natl. Acad. Sci. USA*, 107, 6646–6651, <https://doi.org/10.1073/pnas.0911330107>, 2010.
- Millet, D. B., Goldstein, A. H., Allan, J. D., Bates, T. S., Boudries, H., Bower, K. N., Coe, H., Ma, Y., McKay, M., Quinn, P. K., Sullivan, A., Weber, R. J., and Worsnop, D. R.: Volatile organic compound measurements at Trinidad Head, California, during ITCT 2K2: Analysis of sources, atmospheric composition, and aerosol residence times, *J. Geophys. Res.-Atmos.*, 109, D23S16, <https://doi.org/10.1029/2003JD004026>, 2004.
- Minguillón, M. C., Perron, N., Querol, X., Szidat, S., Fahrni, S. M., Alastuey, A., Jimenez, J. L., Mohr, C., Ortega, A. M., Day, D. A., Lanz, V. A., Wacker, L., Reche, C., Cusack, M., Amato, F., Kiss, G., Hoffer, A., Decesari, S., Moretti, F., Hillamo, R., Teinilä, K., Seco, R., Peñuelas, J., Metzger, A., Schallhart, S., Müller, M., Hansel, A., Burkhardt, J. F., Baltensperger, U., and Prévôt, A. S. H.: Fossil versus contemporary sources of fine elemental and organic carbonaceous particulate matter during the DAURE campaign in Northeast Spain, *Atmos. Chem. Phys.*, 11, 12067–12084, <https://doi.org/10.5194/acp-11-12067-2011>, 2011.
- Mohr, C., DeCarlo, P. F., Heringa, M. F., Chirico, R., Slowik, J. G., Richter, R., Reche, C., Alastuey, A., Querol, X., Seco, R., Peñuelas, J., Jiménez, J. L., Crippa, M., Zimmermann, R., Baltensperger, U., and Prévôt, A. S. H.: Identification and quantification of organic aerosol from cooking and other sources in Barcelona using aerosol mass spectrometer data, *Atmos. Chem. Phys.*, 12, 1649–1665, <https://doi.org/10.5194/acp-12-1649-2012>, 2012.
- 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, <https://doi.org/10.5194/gmd-2-43-2009>, 2009.
- Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Bernsten, T. K., Bian, H., Bellouin, N., Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D., Iversen, T., Kinne, S., Kirkevåg, A., Lamarque, J.-F., Lin, G., Liu, X., Lund, M. T., Luo, G., Ma, X., van Noije, T., Penner, J. E., Rasch, P. J., Ruiz, A., Seland, Ø., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., Wang, P., Wang, Z., Xu, L., Yu, H., Yu, F., Yoon, J.-H., Zhang, K., Zhang, H., and Zhou, C.: Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations, *Atmos. Chem. Phys.*, 13, 1853–1877, <https://doi.org/10.5194/acp-13-1853-2013>, 2013.
- Nemitz, E., Jimenez, J. L., Huffman, J. A., Ulbrich, I. M., Canagaratna, M. R., Worsnop, D. R., and Guenther, A. B.: An Eddy-Covariance System for the Measurement of Surface/Atmosphere Exchange Fluxes of Submicron Aerosol Chemical Species – First Application Above an Urban Area, *Aerosol Sci. Tech.*, 42, 636–657, <https://doi.org/10.1080/02786820802227352>, 2008.
- Ng, N. L., Canagaratna, M. R., Zhang, Q., Jimenez, J. L., Tian, J., Ulbrich, I. M., Kroll, J. H., Docherty, K. S., Chhabra, P. S., Bahreini, R., Murphy, S. M., Seinfeld, J. H., Hildebrandt, L., Donahue, N. M., DeCarlo, P. F., Lanz, V. A., Prévôt, A. S. H., Dinar, E., Rudich, Y., and Worsnop, D. R.: Organic aerosol components observed in Northern Hemispheric datasets from Aerosol Mass Spectrometry, *Atmos. Chem. Phys.*, 10, 4625–4641, <https://doi.org/10.5194/acp-10-4625-2010>, 2010.
- O'Connor, F. M., Johnson, C. E., Morgenstern, O., Abraham, N. L., Braesicke, P., Dalvi, M., Folberth, G. A., Sanderson, M. G., Telford, P. J., Voulgarakis, A., Young, P. J., Zeng, G., Collins, W. J., and Pyle, J. A.: Evaluation of the new UKCA climate-composition model – Part 2: The Troposphere, *Geosci. Model Dev.*, 7, 41–91, <https://doi.org/10.5194/gmd-7-41-2014>, 2014.
- Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, *Atmos. Chem. Phys.*, 7, 1961–1971, <https://doi.org/10.5194/acp-7-1961-2007>, 2007.
- Poulain, L., Spindler, G., Birmili, W., Plass-Dülmer, C., Wiedensohler, A., and Herrmann, H.: Seasonal and diurnal variations of particulate nitrate and organic matter at the IfT research station Melpitz, *Atmos. Chem. Phys.*, 11, 12579–12599, <https://doi.org/10.5194/acp-11-12579-2011>, 2011.
- Quinn, P. K. and Bates, T. S.: Regional aerosol properties: Comparisons of boundary layer measurements from ACE 1, ACE 2, Aerosols99, INDOEX, ACE Asia, TARFOX, and NEAQS, *J. Geophys. Res.-Atmos.*, 110, D14202, <https://doi.org/10.1029/2004JD004755>, 2005.
- Quinn, P. K., Bates, T. S., Coffman, D., Onasch, T. B., Worsnop, D., Baynard, T., de Gouw, J. A., Goldan, P. D., Kuster, W. C., Williams, E., Roberts, J. M., Lerner, B., Stohl, A., Pettersson, A., and Lovejoy, E. R.: Impacts of sources and aging on submicrometer aerosol properties in the marine boundary layer across the Gulf of Maine, *J. Geophys. Res.-Atmos.*, 111, D23S36, <https://doi.org/10.1029/2006JD007582>, 2006.
- Quinn, P. K., Bates, T. S., Schulz, K. S., Coffman, D. J., Frossard, A. A., Russell, L. M., Keene, W. C., and Kieber, D. J.: Contribution of sea surface carbon pool to organic mat-

- ter enrichment in sea spray aerosol, *Nat. Geosci.*, 7, 228–232, <https://doi.org/10.1038/ngeo2092>, 2014.
- Ramanathan, V., Crutzen, P. J., Lelieveld, J., Mitra, A. P., Althausen, D., Anderson, J., Andreae, M. O., Cantrell, W., Cass, G. R., Chung, C. E., Clarke, A. D., Coakley, J. A., Collins, W. D., Conant, W. C., Dulac, F., Heintzenberg, J., Heymsfield, A. J., Holben, B., Howell, S., Hudson, J., Jayaraman, A., Kiehl, J. T., Krishnamurti, T. N., Lubin, D., McFarquhar, G., Novakov, T., Ogren, J. A., Podgorny, I. A., Prather, K., Priestley, K., Prospero, J. M., Quinn, P. K., Rajeev, K., Rasch, P., Rupert, S., Sadourny, R., Satheesh, S. K., Shaw, G. E., Sheridan, P., and Valero, F. P. J.: Indian Ocean Experiment: An integrated analysis of the climate forcing and effects of the great Indo-Asian haze, *J. Geophys. Res.-Atmos.*, 106, 28371–28398, <https://doi.org/10.1029/2001JD900133>, 2001.
- Raper, J. L., Kleb, M. M., Jacob, D. J., Davis, D. D., Newell, R. E., Fuelberg, H. E., Bendura, R. J., Hoell, J. M., and McNeal, R. J.: Pacific Exploratory Mission in the Tropical Pacific: PEM-Tropics B, March–April 1999, *J. Geophys. Res.-Atmos.*, 106, 32401–32425, <https://doi.org/10.1029/2000JD900833>, 2001.
- Reddington, C. L., Carslaw, K. S., Spracklen, D. V., Frontoso, M. G., Collins, L., Merikanto, J., Minikin, A., Hamburger, T., Coe, H., Kulmala, M., Aalto, P., Flentje, H., Plass-Dülmer, C., Birmili, W., Wiedensohler, A., Wehner, B., Tuch, T., Sonntag, A., O’Dowd, C. D., Jennings, S. G., Dupuy, R., Baltensperger, U., Weingartner, E., Hansson, H.-C., Tunved, P., Laj, P., Sellegri, K., Boulon, J., Putaud, J.-P., Gruening, C., Swietlicki, E., Roldin, P., Henzing, J. S., Moerman, M., Mihalopoulos, N., Kouvarakis, G., Ždímal, V., Zíková, N., Marinoni, A., Bonasoni, P., and Duchi, R.: Primary versus secondary contributions to particle number concentrations in the European boundary layer, *Atmos. Chem. Phys.*, 11, 12007–12036, <https://doi.org/10.5194/acp-11-12007-2011>, 2011.
- Reddington, C. L., Carslaw, K. S., Stier, P., Schutgens, N., Coe, H., Liu, D., Allan, J., Browse, J., Pringle, K. J., Lee, L. A., Yoshioka, M., Johnson, J. S., Regayre, L. A., Spracklen, D. V., Mann, G. W., Clarke, A., Hermann, M., Henning, S., Wex, H., Kristensen, T. B., Leaitch, W. R., Pöschl, U., Rose, D., Andreae, M. O., Schmale, J., Kondo, Y., Oshima, N., Schwarz, J. P., Nenes, A., Anderson, B., Roberts, G. C., Snider, J. R., Leck, C., Quinn, P. K., Chi, X., Ding, A., Jimenez, J. L., and Zhang, Q.: The Global Aerosol Synthesis and Science Project (GASSP): Measurements and Modeling to Reduce Uncertainty, *B. Am. Meteorol. Soc.*, 98, 1857–1877, <https://doi.org/10.1175/BAMS-D-15-00317.1>, 2017.
- Reddington, C. L., Morgan, W. T., Darbyshire, E., Brito, J., Coe, H., Artaxo, P., Scott, C. E., Marsham, J., and Spracklen, D. V.: Biomass burning aerosol over the Amazon: analysis of aircraft, surface and satellite observations using a global aerosol model, *Atmos. Chem. Phys.*, 19, 9125–9152, <https://doi.org/10.5194/acp-19-9125-2019>, 2019.
- Regayre, L. A., Pringle, K. J., Booth, B. B. B., Lee, L. A., Mann, G. W., Browse, J., Woodhouse, M. T., Rap, A., Reddington, C. L., and Carslaw, K. S.: Uncertainty in the magnitude of aerosol-cloud radiative forcing over recent decades, *Geophys. Res. Lett.*, 41, 9040–9049, <https://doi.org/10.1002/2014GL062029>, 2014.
- Regayre, L. A., Pringle, K. J., Lee, L. A., Rap, A., Browse, J., Mann, G. W., Reddington, C. L., Carslaw, K. S., Booth, B. B. B., and Woodhouse, M. T.: The Climatic Importance of Uncertainties in Regional Aerosol–Cloud Radiative Forcings over Recent Decades, *J. Climate*, 28, 6589–6607, <https://doi.org/10.1175/JCLI-D-15-0127.1>, 2015.
- Regayre, L. A., Johnson, J. S., Yoshioka, M., Pringle, K. J., Sexton, D. M. H., Booth, B. B. B., Lee, L. A., Bellouin, N., and Carslaw, K. S.: Aerosol and physical atmosphere model parameters are both important sources of uncertainty in aerosol ERF, *Atmos. Chem. Phys.*, 18, 9975–10006, <https://doi.org/10.5194/acp-18-9975-2018>, 2018.
- Riccobono, F., Schobesberger, S., Scott, C. E., Dommen, J., Ortega, I. K., Rondo, L., Almeida, J., Amorim, A., Bianchi, F., Breitenlechner, M., David, A., Downard, A., Dunne, E. M., Duplissy, J., Ehrhart, S., Flagan, R. C., Franchin, A., Hansel, A., Junninen, H., Kajos, M., Keskinen, H., Kupc, A., Kürten, A., Kvashin, A. N., Laaksonen, A., Lehtipalo, K., Makhmutov, V., Mathot, S., Nieminen, T., Onnela, A., Petäjä, T., Praplan, A. P., Santos, F. D., Schallhart, S., Seinfeld, J. H., Sipilä, M., Spracklen, D. V., Stozhkov, Y., Stratmann, F., Tomé, A., Tsagkogeorgas, G., Vaattovaara, P., Viisanen, Y., Vrtala, A., Wagner, P. E., Weingartner, E., Wex, H., Wimmer, D., Carslaw, K. S., Curtius, J., Donahue, N. M., Kirkby, J., Kulmala, M., Worsnop, D. R., and Baltensperger, U.: Oxidation products of biogenic emissions contribute to nucleation of atmospheric particles, *Science*, 344, 717–721, <https://doi.org/10.1126/science.1243527>, 2014.
- Rodrigues, L. F. S., Vernon, I., and Bower, R. G.: Constraints on galaxy formation models from the galaxy stellar mass function and its evolution, *Mon. Not. R. Astron. Soc.*, 466, 2418–2435, <https://doi.org/10.1093/mnras/stw3269>, 2017.
- Ryerson, T. B., Andrews, A. E., Angevine, W. M., Bates, T. S., Brock, C. A., Cairns, B., Cohen, R. C., Cooper, O. R., de Gouw, J. A., Fehsenfeld, F. C., Ferrare, R. A., Fischer, M. L., Flagan, R. C., Goldstein, A. H., Hair, J. W., Hardesty, R. M., Hostetler, C. A., Jimenez, J. L., Langford, A. O., McCauley, E., McKeen, S. A., Molina, L. T., Nenes, A., Oltmans, S. J., Parrish, D. D., Pederson, J. R., Pierce, R. B., Prather, K., Quinn, P. K., Seinfeld, J. H., Senff, C. J., Sorooshian, A., Stutz, J., Surratt, J. D., Trainer, M., Volkamer, R., Williams, E. J., and Wofsy, S. C.: The 2010 California Research at the Nexus of Air Quality and Climate Change (CalNex) field study, *J. Geophys. Res.-Atmos.*, 118, 5830–5866, <https://doi.org/10.1002/jgrd.50331>, 2013.
- Schutgens, N., Tsyro, S., Gryspeerdt, E., Goto, D., Weigum, N., Schulz, M., and Stier, P.: On the spatio-temporal representativeness of observations, *Atmos. Chem. Phys.*, 17, 9761–9780, <https://doi.org/10.5194/acp-17-9761-2017>, 2017.
- Schutgens, N. A. J., Partridge, D. G., and Stier, P.: The importance of temporal collocation for the evaluation of aerosol models with observations, *Atmos. Chem. Phys.*, 16, 1065–1079, <https://doi.org/10.5194/acp-16-1065-2016>, 2016a.
- Schutgens, N. A. J., Gryspeerdt, E., Weigum, N., Tsyro, S., Goto, D., Schulz, M., and Stier, P.: Will a perfect model agree with perfect observations? The impact of spatial sampling, *Atmos. Chem. Phys.*, 16, 6335–6353, <https://doi.org/10.5194/acp-16-6335-2016>, 2016b.
- Setyan, A., Zhang, Q., Merkel, M., Knighton, W. B., Sun, Y., Song, C., Shilling, J. E., Onasch, T. B., Herndon, S. C., Worsnop, D. R., Fast, J. D., Zaveri, R. A., Berg, L. K., Wiedensohler, A., Flowers, B. A., Dubey, M. K., and Subramanian, R.: Characterization of submicron particles influenced by mixed biogenic and anthropogenic emissions using high-resolution aerosol mass spectrom-

- etry: results from CARES, *Atmos. Chem. Phys.*, 12, 8131–8156, <https://doi.org/10.5194/acp-12-8131-2012>, 2012.
- Shindell, D. T., Lamarque, J.-F., Schulz, M., Flanner, M., Jiao, C., Chin, M., Young, P. J., Lee, Y. H., Rotstayn, L., Mahowald, N., Milly, G., Faluvegi, G., Balkanski, Y., Collins, W. J., Conley, A. J., Dalsoren, S., Easter, R., Ghan, S., Horowitz, L., Liu, X., Myhre, G., Nagashima, T., Naik, V., Rumbold, S. T., Skeie, R., Sudo, K., Szopa, S., Takemura, T., Voulgarakis, A., Yoon, J.-H., and Lo, F.: Radiative forcing in the ACCMIP historical and future climate simulations, *Atmos. Chem. Phys.*, 13, 2939–2974, <https://doi.org/10.5194/acp-13-2939-2013>, 2013.
- Spracklen, D. V., Jimenez, J. L., Carslaw, K. S., Worsnop, D. R., Evans, M. J., Mann, G. W., Zhang, Q., Canagaratna, M. R., Allan, J., Coe, H., McFiggans, G., Rap, A., and Forster, P.: Aerosol mass spectrometer constraint on the global secondary organic aerosol budget, *Atmos. Chem. Phys.*, 11, 12109–12136, <https://doi.org/10.5194/acp-11-12109-2011>, 2011.
- Stroud, C. A., Nenes, A., Jimenez, J. L., DeCarlo, P. F., Huffman, J. A., Bruintjes, R., Nemitz, E., Delia, A. E., Toohey, D. W., Guenther, A. B., and Nandi, S.: Cloud Activating Properties of Aerosol Observed during CELTIC, *J. Atmos. Sci.*, 64, 441–459, <https://doi.org/10.1175/JAS3843.1>, 2007.
- Sun, J., Zhang, Q., Canagaratna, M. R., Zhang, Y., Ng, N. L., Sun, Y., Jayne, J. T., Zhang, X., Zhang, X., and Worsnop, D. R.: Highly time- and size-resolved characterization of submicron aerosol particles in Beijing using an Aerodyne Aerosol Mass Spectrometer, *Atmos. Environ.*, 44, 131–140, <https://doi.org/10.1016/j.atmosenv.2009.03.020>, 2010.
- Sun, Y.-L., Zhang, Q., Schwab, J. J., Demerjian, K. L., Chen, W.-N., Bae, M.-S., Hung, H.-M., Hogrefe, O., Frank, B., Rattigan, O. V., and Lin, Y.-C.: Characterization of the sources and processes of organic and inorganic aerosols in New York city with a high-resolution time-of-flight aerosol mass spectrometer, *Atmos. Chem. Phys.*, 11, 1581–1602, <https://doi.org/10.5194/acp-11-1581-2011>, 2011.
- Sun, Y., Zhang, Q., Macdonald, A. M., Hayden, K., Li, S. M., Liggio, J., Liu, P. S. K., Anlauf, K. G., Leaitch, W. R., Steffen, A., Cubison, M., Worsnop, D. R., van Donkelaar, A., and Martin, R. V.: Size-resolved aerosol chemistry on Whistler Mountain, Canada with a high-resolution aerosol mass spectrometer during INTEX-B, *Atmos. Chem. Phys.*, 9, 3095–3111, <https://doi.org/10.5194/acp-9-3095-2009>, 2009.
- Takahama, S., Schwartz, R. E., Russell, L. M., Macdonald, A. M., Sharma, S., and Leaitch, W. R.: Organic functional groups in aerosol particles from burning and non-burning forest emissions at a high-elevation mountain site, *Atmos. Chem. Phys.*, 11, 6367–6386, <https://doi.org/10.5194/acp-11-6367-2011>, 2011.
- Takami, A., Miyoshi, T., Shimono, A., and Hatakeyama, S.: Chemical composition of fine aerosol measured by AMS at Fukue Island, Japan during APEX period, *Atmos. Environ.*, 39, 4913–4924, <https://doi.org/10.1016/j.atmosenv.2005.04.038>, 2005.
- Takegawa, N., Miyakawa, T., Kondo, Y., Jimenez, J. L., Zhang, Q., Worsnop, D. R., and Fukuda, M.: Seasonal and diurnal variations of submicron organic aerosol in Tokyo observed using the Aerodyne aerosol mass spectrometer, *J. Geophys. Res.*, 111, D11206, <https://doi.org/10.1029/2005JD006515>, 2006.
- Timonen, H., Carbone, S., Aurela, M., Saarnio, K., Saarikoski, S., Ng, N. L., Canagaratna, M. R., Kulmala, M., Kerminen, V.-M., Worsnop, D. R., and Hillamo, R.: Characteristics, sources and water-solubility of ambient submicron organic aerosol in springtime in Helsinki, *J. Aerosol Sci.*, 56, 31–77, <https://doi.org/10.1016/j.jaerosci.2012.06.005>, 2013.
- Tjernström, M., Leck, C., Birch, C. E., Bottenheim, J. W., Brooks, B. J., Brooks, I. M., Bäcklin, L., Chang, R. Y.-W., de Leeuw, G., Di Liberto, L., de la Rosa, S., Granath, E., Graus, M., Hansel, A., Heintzenberg, J., Held, A., Hind, A., Johnston, P., Knulst, J., Martin, M., Matrai, P. A., Mauritsen, T., Müller, M., Norris, S. J., Orellana, M. V., Orsini, D. A., Paatero, J., Persson, P. O. G., Gao, Q., Rauschenberg, C., Ristovski, Z., Sedlar, J., Shupe, M. D., Sierau, B., Sirevaag, A., Sjogren, S., Stetzer, O., Swietlicki, E., Szczodrak, M., Vaattovaara, P., Wahlberg, N., Westberg, M., and Wheeler, C. R.: The Arctic Summer Cloud Ocean Study (ASCOS): overview and experimental design, *Atmos. Chem. Phys.*, 14, 2823–2869, <https://doi.org/10.5194/acp-14-2823-2014>, 2014.
- Topping, D., Coe, H., McFiggans, G., Burgess, R., Allan, J., Alfarra, M. R., Bower, K., Choularton, T. W., Decesari, S., and Facchini, M. C.: Aerosol chemical characteristics from sampling conducted on the Island of Jeju, Korea during ACE Asia, *Atmos. Environ.*, 38, 2111–2123, <https://doi.org/10.1016/j.atmosenv.2004.01.022>, 2004.
- Tørseth, K., Aas, W., Breivik, K., Fjæraa, A. M., Fiebig, M., Hjellbrekke, A. G., Lund Myhre, C., Solberg, S., and Yttri, K. E.: Introduction to the European Monitoring and Evaluation Programme (EMEP) and observed atmospheric composition change during 1972–2009, *Atmos. Chem. Phys.*, 12, 5447–5481, <https://doi.org/10.5194/acp-12-5447-2012>, 2012.
- Tsigradis, K., Daskalakis, N., Kanakidou, M., Adams, P. J., Artaxo, P., Bahadur, R., Balkanski, Y., Bauer, S. E., Bellouin, N., Benedetti, A., Bergman, T., Berntsen, T. K., Beukes, J. P., Bian, H., Carslaw, K. S., Chin, M., Curci, G., Diehl, T., Easter, R. C., Ghan, S. J., Gong, S. L., Hodzic, A., Hoyle, C. R., Iversen, T., Jathar, S., Jimenez, J. L., Kaiser, J. W., Kirkevåg, A., Koch, D., Kokkola, H., Lee, Y. H., Lin, G., Liu, X., Luo, G., Ma, X., Mann, G. W., Mihalopoulos, N., Morcrette, J.-J., Müller, J.-F., Myhre, G., Myriokefalitakis, S., Ng, N. L., O'Donnell, D., Penner, J. E., Pozzoli, L., Pringle, K. J., Russell, L. M., Schulz, M., Sciare, J., Seland, Ø., Shindell, D. T., Sillman, S., Skeie, R. B., Spracklen, D., Stavrou, T., Steenrod, S. D., Takemura, T., Titt, P., Tilmes, S., Tost, H., van Noije, T., van Zyl, P. G., von Salzen, K., Yu, F., Wang, Z., Wang, Z., Zaveri, R. A., Zhang, H., Zhang, K., Zhang, Q., and Zhang, X.: The AeroCom evaluation and intercomparison of organic aerosol in global models, *Atmos. Chem. Phys.*, 14, 10845–10895, <https://doi.org/10.5194/acp-14-10845-2014>, 2014.
- Turnock, S. T., Spracklen, D. V., Carslaw, K. S., Mann, G. W., Woodhouse, M. T., Forster, P. M., Haywood, J., Johnson, C. E., Dalvi, M., Bellouin, N., and Sanchez-Lorenzo, A.: Modelled and observed changes in aerosols and surface solar radiation over Europe between 1960 and 2009, *Atmos. Chem. Phys.*, 15, 9477–9500, <https://doi.org/10.5194/acp-15-9477-2015>, 2015.
- Vakkari, V., Beukes, J. P., Laakso, H., Mabaso, D., Pienaar, J. J., Kulmala, M., and Laakso, L.: Long-term observations of aerosol size distributions in semi-clean and polluted savannah in South Africa, *Atmos. Chem. Phys.*, 13, 1751–1770, <https://doi.org/10.5194/acp-13-1751-2013>, 2013.
- Walters, D. N., Williams, K. D., Boutle, I. A., Bushell, A. C., Edwards, J. M., Field, P. R., Lock, A. P., Morcrette, C. J., Strat-

- ton, R. A., Wilkinson, J. M., Willett, M. R., Bellouin, N., Bodas-Salcedo, A., Brooks, M. E., Copley, D., Earnshaw, P. D., Hardiman, S. C., Harris, C. M., Levine, R. C., MacLachlan, C., Manners, J. C., Martin, G. M., Milton, S. F., Palmer, M. D., Roberts, M. J., Rodríguez, J. M., Tennant, W. J., and Vidale, P. L.: The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations, *Geosci. Model Dev.*, 7, 361–386, <https://doi.org/10.5194/gmd-7-361-2014>, 2014.
- Wang, W., Rood, M. J., Carrico, C. M., Covert, D. S., Quinn, P. K., and Bates, T. S.: Aerosol optical properties along the northeast coast of North America during the New England Air Quality Study-Intercontinental Transport and Chemical Transformation 2004 campaign and the influence of aerosol composition, *J. Geophys. Res.-Atmos.*, 112, D10S23, <https://doi.org/10.1029/2006JD007579>, 2007.
- Watson-Parris, D., Schutgens, N., Reddington, C., Pringle, K. J., Liu, D., Allan, J. D., Coe, H., Carslaw, K. S., and Stier, P.: In situ constraints on the vertical distribution of global aerosol, *Atmos. Chem. Phys.*, 19, 11765–11790, <https://doi.org/10.5194/acp-19-11765-2019>, 2019.
- Weimer, S., Drewnick, F., Högrefe, O., Schwab, J. J., Rhoads, K., Orsini, D., Canagaratna, M., Worsnop, D. R., and Demerjian, K. L.: Size-selective nonrefractory ambient aerosol measurements during the Particulate Matter Technology Assessment and Characterization Study–New York 2004 Winter Intensive in New York City, *J. Geophys. Res.*, 111, D18305, <https://doi.org/10.1029/2006JD007215>, 2006.
- van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), *Atmos. Chem. Phys.*, 10, 11707–11735, <https://doi.org/10.5194/acp-10-11707-2010>, 2010.
- West, R. E. L., Stier, P., Jones, A., Johnson, C. E., Mann, G. W., Bellouin, N., Partridge, D. G., and Kipling, Z.: The importance of vertical velocity variability for estimates of the indirect aerosol effects, *Atmos. Chem. Phys.*, 14, 6369–6393, <https://doi.org/10.5194/acp-14-6369-2014>, 2014.
- Williams, B. J., Goldstein, A. H., Kreisberg, N. M., Hering, S. V., Worsnop, D. R., Ulbrich, I. M., Docherty, K. S., and Jimenez, J. L.: Major components of atmospheric organic aerosol in southern California as determined by hourly measurements of source marker compounds, *Atmos. Chem. Phys.*, 10, 11577–11603, <https://doi.org/10.5194/acp-10-11577-2010>, 2010.
- Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., and Yamazaki, K.: History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble, *Clim. Dynam.*, 41, 1703–1729, <https://doi.org/10.1007/s00382-013-1896-4>, 2013.
- Wood, R., Mechoso, C. R., Bretherton, C. S., Weller, R. A., Huebert, B., Straneo, F., Albrecht, B. A., Coe, H., Allen, G., Vaughan, G., Daum, P., Fairall, C., Chand, D., Gallardo Klenner, L., Garreaud, R., Grados, C., Covert, D. S., Bates, T. S., Krejci, R., Russell, L. M., de Szoeko, S., Brewer, A., Yuter, S. E., Springston, S. R., Chaigneau, A., Toniazzo, T., Minnis, P., Palikonda, R., Abel, S. J., Brown, W. O. J., Williams, S., Fochesatto, J., Brioude, J., and Bower, K. N.: The VAMOS Ocean-Cloud-Atmosphere-Land Study Regional Experiment (VOCALS-REx): goals, platforms, and field operations, *Atmos. Chem. Phys.*, 11, 627–654, <https://doi.org/10.5194/acp-11-627-2011>, 2011.
- Yoshioka, M., Regayre, L. A., Pringle, K. J., Johnson, J. S., Mann, G. W., Partridge, D. G., Sexton, D. M. H., Lister, G. M. S., Schutgens, N., Stier, P., Kipling, Z., Bellouin, N., Browse, J., Booth, B. B., Johnson, C. E., Johnson, B., Mollard, J. D. P., Lee, L. A., and Carslaw, K. S.: Ensembles of Global Climate Model Variants Designed for the Quantification and Constraint of Uncertainty in Aerosols and their Radiative Forcing, *J. Adv. Model. Earth Syst.*, 11, 3728–3754, <https://doi.org/10.1029/2019MS001628>, 2019.
- Zhang, Q., Canagaratna, M. R., Jayne, J. T., Worsnop, D. R., and Jimenez, J.: Time- and size-resolved chemical composition of submicron particles in Pittsburgh: Implications for aerosol sources and processes, *J. Geophys. Res.*, 110, D07S09, <https://doi.org/10.1029/2004JD004649>, 2005.
- Zhang, Q., Parworth, C., Lechner, M., and Jimenez, J. L.: Aerosol Mass Spectrometer Global Database, available at: <https://sites.google.com/site/amsglobaldatabase> (last access: 17 September 2019), <https://doi.org/10.6084/m9.figshare.3486719>, 2017.