



# Estimating bacteria emissions from inversion of atmospheric transport: sensitivity to modelled particle characteristics

S. M. Burrows<sup>1,\*</sup>, P. J. Rayner<sup>2</sup>, T. Butler<sup>1,\*\*</sup>, and M. G. Lawrence<sup>1,\*\*</sup>

<sup>1</sup>Max Planck Institute for Chemistry, Mainz, Germany

<sup>2</sup>University of Melbourne, School of Earth Sciences, Melbourne, Australia

\* now at: Pacific Northwest National Laboratory, Richland, Washington, USA

\*\* now at: Institute for Advanced Sustainability Studies e.V., Potsdam, Germany

Correspondence to: S. M. Burrows (susannah.burrows@pnnl.gov)

Received: 12 November 2012 – Published in Atmos. Chem. Phys. Discuss.: 15 February 2013

Revised: 16 April 2013 – Accepted: 3 May 2013 – Published: 4 June 2013

**Abstract.** Model-simulated transport of atmospheric trace components can be combined with observed concentrations to obtain estimates of ground-based sources using various inversion techniques. These approaches have been applied in the past primarily to obtain source estimates for long-lived trace gases such as CO<sub>2</sub>. We consider the application of similar techniques to source estimation for atmospheric aerosols, using as a case study the estimation of bacteria emissions from different ecosystem regions in the global atmospheric chemistry and climate model ECHAM5/MESSy-Atmospheric Chemistry (EMAC).

Source estimation via Markov Chain Monte Carlo is applied to a suite of sensitivity simulations, and the global mean emissions are estimated for the example problem of bacteria-containing aerosol particles. We present an analysis of the uncertainties in the global mean emissions, and a partitioning of the uncertainties that are attributable to particle size, activity as cloud condensation nuclei (CCN), the ice nucleation scavenging ratios for mixed-phase and cold clouds, and measurement error.

For this example, uncertainty due to CCN activity or to a 1 μm error in particle size is typically between 10% and 40% of the uncertainty due to observation uncertainty, as measured by the 5–95th percentile range of the Monte Carlo ensemble. Uncertainty attributable to the ice nucleation scavenging ratio in mixed-phase clouds is as high as 10–20% of that attributable to observation uncertainty. Taken together, the four model parameters examined contribute about half as much to the uncertainty in the estimated emissions as do the observations. This was a surprisingly large contribution from

model uncertainty in light of the substantial observation uncertainty, which ranges from 81–870% of the mean for each of ten ecosystems for this case study. The effects of these and other model parameters in contributing to the uncertainties in the transport of atmospheric aerosol particles should be treated explicitly and systematically in both forward and inverse modelling studies.

## 1 Introduction

Atmospheric aerosol particles are recognized as a critical part of the climate system, with their direct and indirect effects on climate identified as one of the key uncertainties in current understanding of climate change (Solomon et al., 2007). In order to advance prediction of aerosol climate impacts, their representation in global models must be improved, including their composition and distribution in the atmosphere. In particular, while much current research focuses on characterizing anthropogenic aerosols, naturally occurring aerosols are still poorly characterized in global models due to a variety of issues including sparsity of observational data, low model resolution, and inherent uncertainties in model parameterizations (Kinne et al., 2006; Huneeus et al., 2011). While global models typically account for several classes of naturally occurring aerosol (sea spray, dust) and man-made aerosol (sulfate, soot from industry and biomass burning), almost no global atmospheric model currently includes any explicit representation of primary biological aerosol particles (PBAP) – material such as bacteria, pollen, fungal spores,

and leaf fragments – despite the fact that these particles make up a large fraction of the observed aerosol at many locations (Jaenicke, 2005). For example, observations of particles with radius greater than  $0.2\ \mu\text{m}$  between 2000 and 2008 in Mainz, Germany, found that between 5 and 50 % of particle volume was composed of primary biological aerosol particles; at Lake Baikal, Russia, an average of ca. 20 % of particles was observed to be PBAP (by number and volume; particles with radius  $> 0.2\ \mu\text{m}$ ) (Jaenicke et al., 2007). Measurements in the Amazon rainforest found that 40 % of submicron and up to 80 % of supermicron particles were primary biological particles (Graham et al., 2003). Although the biological fraction reported depends on the definition and on the measurement technique used, on which there has been little consensus (Després et al., 2012), it is clear that biological particles can be important contributors to the atmospheric aerosol population, particularly in the supermicron size range.

Two broad groups of efforts to quantify emissions can be distinguished: forward modelling and inverse modelling approaches. Briefly, forward modelling approaches use empirical emissions data, derived e.g. from field experiments or industry data, and use ancillary information (e.g. maps, climate variables) to upscale to a global emissions map (Rayner et al., 2010; Jung et al., 2011). In contrast, inverse modelling approaches use observed concentrations in conjunction with a model of atmospheric transport to estimate emissions (Enting, 2000, 2002; Tarantola, 2005). This is achieved by applying mathematical techniques to infer the necessary emissions required for the model to optimally match observations, accounting for estimates of the uncertainties in observations and for prior information about the emission. Inverse problems typically are underconstrained, so their solution often requires the use of known emissions as prior information in order to obtain a stable solution. In addition, the observed variables typically differ from the modelled variables, e.g. in their spatial and temporal distribution and representativeness (Kaminski et al., 2001).

An inherent challenge of inverse modelling is the appropriate estimation of uncertainties, which can arise from many sources, including errors in the observational data, differences between model and observations in sampling location and representativeness, forecasting errors, and errors in modelling (Enting, 2002). This study considers some of the uncertainties inherent in the estimation of aerosol emissions by inversion. In particular, we examine the uncertainty contributed by the following model parameters: particle size, activity as cloud condensation nuclei (CCN), and the ice nucleation scavenging ratio. The uncertainty arising from these model parameters is compared to uncertainties arising from observations.

In this study, we use the estimation of bacteria emissions from different ecosystems as a case study. In Sect. 2, we briefly describe the observational data and the emissions model. In Sect. 3, we describe the transport model, including a discussion of key processes affecting aerosol removal and

the impact of selected model parameters on aerosol residence time. In Sect. 4, we briefly present the method used for inversion of atmospheric transport and estimation of sources, and present the main results from this inversion for different sensitivity cases. These results show how the estimated global mean emissions respond to changes in key model parameters. In Sect. 4.4, we introduce the term “normalized model uncertainties” and present the normalized model uncertainties for a set of model parameters affecting atmospheric transport. We discuss our findings further in Sect. 6. In Sect. 7 we summarize and discuss key findings.

## 2 Bacteria observations

Published observations of bacteria concentrations in the atmosphere are scarce, and many suffer from methodological limitations. The concentrations of bacteria-containing particles in the atmosphere can be measured by a variety of methods with varying degrees of accuracy (Burrows et al., 2009b). Most commonly, atmospheric particulates are collected onto either a filter or an impaction plate, which is then analysed for microbial content. One common method of analysis is the cultivation of bacterial cells on a culture medium, followed by counting of colonies. Because many bacteria are not readily cultivated, this method results in serious undercounting and a large uncertainty. A more robust, but also more labour-intensive method, is the use of microscopy: for instance, optical microscopy combined with protein staining or the use of autofluorescence to identify microorganisms. Burrows et al. (2009b) reviewed a large number of published measurements of bacteria concentrations and recommended low, best, and high estimates of the mean boundary-layer number concentrations of bacteria-containing particles in each of ten ecosystems, which are reproduced here in Table 1.

The high uncertainty in these estimates arises from uncertainties inherent in the observational methods discussed above, from the scarcity of observations, particularly long-term observations, and also from the high spatial and temporal variability in concentrations. The estimated uncertainty in the best-estimate mean concentrations derived from a review of the literature ranges from 81–870 % of the mean for the ten individual ecosystems (Burrows et al., 2009a). Contrast this with an atmospheric component that can be measured far more precisely, such as  $\text{CO}_2$ , which due to its long atmospheric residence time has a much smaller degree of spatial and temporal variability, and the concentration of which can be measured with a high degree of precision, the relative uncertainty in each measurement being far less than 1 %.

**Table 1.** Estimates of total mean number concentration of bacteria-containing particles [ $\text{m}^{-3}$ ] in near-surface air of various ecosystem types, reproduced with corrections from Burrows et al. (2009b,a).

Ecosystem	Low estimate <sup>b</sup>	Best estimate <sup>a</sup>	High estimate	Percent uncertainty
coastal <sup>c</sup>	$2.3 \times 10^4$	$7.6 \times 10^4$	$1.3 \times 10^5$	300
crops <sup>c</sup>	$4.1 \times 10^4$	$1.1 \times 10^5$	$1.7 \times 10^5$	81
deserts <sup>d,e</sup>	$1.6 \times 10^2$	<i>(<math>1 \times 10^4</math>)</i>	$3.8 \times 10^4$	380
forests <sup>f</sup>	$3.3 \times 10^4$	$5.6 \times 10^4$	$8.8 \times 10^4$	100
grasslands <sup>c,g</sup>	$2.5 \times 10^4$	$1.1 \times 10^5$	$8.4 \times 10^5$	290
land ice <sup>h,i</sup>	<i>(<math>1 \times 10^1</math>)</i>	<i>(<math>5 \times 10^3</math>)</i>	$1 \times 10^4$	200
seas <sup>c,g,j</sup>	$1 \times 10^1$	$1 \times 10^4$	$8 \times 10^4$	800
shrubs <sup>f,g</sup>	$1.2 \times 10^4$	$3.5 \times 10^5$	$8.4 \times 10^5$	240
tundra <sup>g,k</sup>	<i>(<math>1 \times 10^1</math>)</i>	$1.2 \times 10^4$	$5.6 \times 10^4$	470
wetlands <sup>l</sup>	$2 \times 10^4$	$9 \times 10^4$	$8 \times 10^5$	870

<sup>a</sup> Additional values have been assumed for fields left blank by Burrows et al. (2009a); these are denoted by parentheses and italic font.

<sup>b</sup> Percent uncertainties are calculated as  $\text{best} = (\text{high} - \text{low}) \times 100$ .

<sup>c</sup> Harrison et al. (2005).

<sup>d</sup> Lighthart and Shaffer (1994).

<sup>e</sup> Assumed the same as best estimate for seas.

<sup>f</sup> Shaffer and Lighthart (1997)

<sup>g</sup> Tong and Lighthart (1999); Tilley et al. (2001).

<sup>h</sup> Bauer et al. (2002).

<sup>i</sup> Estimated low value for seas taken as lower bound; average of high and low values taken as best estimate.

<sup>j</sup> Griffin et al. (2006).

<sup>k</sup> Estimated low value for seas taken as lower bound.

<sup>l</sup> Assumed to be within bounds of estimates in coastal and grassland/crops regions.

### 3 Global atmospheric model description, key processes, and impacts of selected parameters

#### 3.1 Model description

Following Burrows et al. (2009a), bacteria tracers are emitted homogeneously from the ten ecosystems classes presented in Table 1. The ecosystem classification is based on the Olson World Ecosystems data set (Olson, 1992), with ecosystems lumped into broader classes as described in Burrows et al. (2009a).

All model simulations were conducted using a modified version of the global chemistry-climate model ECHAM5/MESy-Atmospheric Chemistry (EMAC), version 1.9. EMAC consists of a climate model that simulates the underlying meteorological parameters such as winds, combined with a number of submodels representing various physical and chemical processes in the atmosphere. The processes related to particulate emissions and loss processes are encapsulated in the submodels ONLEM (online emissions, Kerkweg et al., 2006b), DRYDEP (dry deposition onto land, water and plant surfaces, Kerkweg et al., 2006a), SEDI (sedimentation, Kerkweg et al., 2006a), and SCAV (precipitation scavenging, Tost et al., 2006a). The submodel CVTRANS calculates particulate transport analogously to gas-phase transport as a sum of large-scale advection and parameterized small-scale convective transport (Lawrence and Rasch, 2005; Tost et al., 2010).

Modifications to the model comprised updates to the SCAV submodel as described in Tost et al. (2010). All simulations were conducted at T63 horizontal resolution ( $1.9^\circ \times 1.9^\circ$  or about  $140 \times 210$  km at midlatitudes) with 31 vertical levels up to 10 hPa, for 5 simulated years (plus an additional year of spin-up). Initial meteorological fields were derived from the ECMWF ERA-15 reanalysis for 1 January 1990, following which meteorology was simulated online. Monthly prescribed sea surface temperatures were applied from the AMIP-II data set (available from <http://www-pcmdi.llnl.gov/>).

EMAC simulates a realistic climate similar to that produced by other global chemistry-climate models (Lamarque et al., 2013), and the climate is similar in free-running simulations and simulations nudged with meteorological data (Klinger, 2011). The EMAC model, in a configuration similar to the one used here, has been shown to produce satisfactory simulations of the observed deposition patterns of radioactive particles following the Chernobyl nuclear meltdown (Lelieveld et al., 2012). Other model versions and configurations of the EMAC model (e.g. including simulation of aerosol microphysics) have been shown to be capable of producing satisfactory simulations of the atmospheric dust life cycle (Astitha et al., 2012; Gläser et al., 2012), of dust and black carbon simultaneously (Aquila et al., 2011) and of multiple interacting natural and anthropogenic aerosol species (Pringle et al., 2010). The ability of the model to predict aerosol optical depth has been evaluated in detail for the

Mediterranean region (de Meij et al., 2012) and globally (Pozzer et al., 2012).

We assume that to a first approximation, bacteria-containing aerosol particles are transported similarly to other insoluble aerosol particles such as dust – i.e. they have the same size-dependent rate of removal in the model as other particles.

### 3.2 Simulations and sensitivity studies

Simulations were performed for monodisperse passive aerosol tracers, with aerodynamic diameters between 1  $\mu\text{m}$  and 10  $\mu\text{m}$ , and CCN-ACTIVE vs CCN-INACTIVE particles. In all cases, particles are removed by sedimentation, dry deposition, impaction and interception scavenging, uptake into cloud droplets by diffusion during droplet growth, and ice-phase scavenging during ice nucleation and precipitation. CCN-ACTIVE particles are additionally removed by cloud droplet nucleation on the particles and subsequent precipitation. In the model setup used, cloud formation is independent of simulated aerosol concentrations – i.e. clouds affect aerosols (via wet removal), but aerosols do not affect clouds. The derived sensitivities thus refer only to the linear, one-directional effect of loss processes on concentrations, not to any potential nonlinear aerosol–cloud feedbacks.

We consider the base case (unmodified model) and four ice scavenging sensitivity cases described in Table 2, ten particle sizes in 1  $\mu\text{m}$  increments from 1 to 10  $\mu\text{m}$ , and CCN-ACTIVE vs CCN-INACTIVE particles.

In the cases SENS\_COLD and SENS\_MIXED, small perturbations were made to test the quasi-linear response of the model to a 1 % change in the respective ice scavenging parameter. The SENS\_MIXED case resulted in an appreciable perturbation of aerosol concentrations in the midlatitude tropopause region and the surface atmosphere at high latitudes (not shown). However, the effect on estimated emissions was too small to be detected within the random noise of the Monte Carlo simulation. These cases will not be discussed in further detail, but are included in some of the presented results.

In the cases LIM\_COLD and LIM\_MIXED, we use larger perturbations in order to gauge the response of the Monte Carlo emissions estimate to a larger perturbation to the respective ice scavenging coefficients, respectively changing the coefficient for cold clouds from 0.05 to 0.1 and the coefficient for warm clouds from 0.1 to 0.7.

For each combination of values of these parameters, we derive a separate inter-ecosystem transport matrix  $\mathbf{G}$ , for a total of 100 different cases (10 sizes  $\times$  2 CCN activities  $\times$  5 IN sensitivity cases). For each case, we performed a full atmospheric model simulation to generate a realization of  $\mathbf{G}$ . For each realization of  $\mathbf{G}$ , we performed a Markov Chain Monte Carlo (MCMC) inversion with one million trial solutions, where each trial solution was either accepted or rejected as a member of the posterior ensemble according to

**Table 2.** Ice nucleation scavenging ratio in the base case and four sensitivity cases.

	$T > -35^\circ\text{C}$	$T \leq -35^\circ\text{C}$
BASE	0.1	0.05
SENS_COLD	0.1	0.0505
SENS_MIXED	0.101	0.05
LIM_COLD	0.1	0.1
LIM_MIXED	0.7	0.05

the Metropolis rule (Sect. 4.3). The lowest acceptance rate was ca. 15 %, producing an ensemble with more than one hundred fifty thousand members.

Ice nucleation scavenging in EMAC is calculated using a constant scavenging ratio. The ice and liquid water contents of clouds are each represented in EMAC by a single bulk variable. In the unmodified model (base case), ice nucleation is treated as follows: for mixed-phase clouds warmer than  $-35^\circ\text{C}$ , the ice nucleation scavenging ratio is set to 0.1; otherwise it is set to 0.05. This ratio describes the fraction of the aerosol particles within the cloud that are incorporated into cloud ice crystals. The removal of these particles from a model grid box (by scavenging) further depends on the rate at which frozen precipitation falls from the grid box, relative to the amount of cloud ice within the grid box.

This parameterization is broadly consistent with field studies of aerosol partitioning in clouds at the Jungfraujoch (Swiss Alps; Henning et al., 2004; Verheggen et al., 2007), which show that the fraction of aerosol particles that are incorporated into the cloud liquid water and cloud ice decreases very rapidly at low temperatures and high ice mass fractions. This is likely due primarily to the Bergeron–Findeisen effect, which leads to the growth of a small number of ice crystals at the expense of the evaporation of a larger number of cloud droplets, which upon evaporation release the particles they contained back into the aerosol phase (Schwarzenböck et al., 2001). In those studies, the activated fraction ranges from 0.05 or less at temperatures below  $-15^\circ\text{C}$  to about 0.7 at near-zero temperatures (Henning et al., 2004; Verheggen et al., 2007). The activated aerosol represents an upper limit on the aerosol that may be removed due to nucleation scavenging.

### 3.3 Aerosol loss processes and their dependence on particle characteristics

Aerosols are removed from the atmosphere by both dry and wet deposition processes. For small particles, the most efficient removal is by coagulation with hydrometeors and dry surfaces associated with Brownian diffusion. For large particles, gravitational settling becomes increasingly efficient, and particles are more likely to be collected by hydrometeors via inertial impaction and interception, and subsequently removed via precipitation. For aerosol particles with

aerodynamic diameters close to 1  $\mu\text{m}$ , precipitation scavenging is the dominant atmospheric loss process, but particles of this size fall into the so-called “scavenging gap” and thus have comparatively long atmospheric residence times (Pruppacher and Klett, 1997).

Particle size influences the rate of dry removal, and both particle size and chemical composition influence the rate of wet removal. In particular, particle chemical composition influences the likelihood that particles will act as CCN and as heterogeneous nuclei for the formation of ice crystals, and thus it influences the likelihood of scavenging in mixed-phase clouds. However, the effect of IN activity on aerosol lifetime is presently not accounted for in EMAC or most other global atmospheric models. Thus, uncertainties related to particle characteristics also impact transport and loss processes (Pruppacher and Klett, 1997; Seinfeld and Pandis, 2006).

## 4 Source inversion and error analysis

### 4.1 Statistical model

The forward model is described by the following elements:

1. The observable quantities  $\mathbf{d}$ : a vector representing the mean concentration of bacteria in each ecosystem.
2. The model (parameters)  $\mathbf{m}$ : a vector representing the surface fluxes in each ecosystem.
3. The model function  $\mathbf{m} \mapsto g(\mathbf{m})$ : The results from global atmospheric model simulations (as described in Sect. 3) are used to derive a simplified statistical model of inter-ecosystem transport. The modelled concentrations are linear in source strengths so the model can be represented by a  $10 \times 10$  matrix describing the relationship between homogeneous, constant emissions from each source ecosystem and mean boundary-layer concentrations in each destination ecosystem. The simulated concentrations are given by the product of the inter-ecosystem transfer matrix  $\mathbf{G}$  and the flux vector  $\mathbf{m}$ :

$$g(\mathbf{m}) = \mathbf{G}\mathbf{m}. \quad (1)$$

The linearity of  $g$  is guaranteed by design in the EMAC model setup used here (neglecting small numerical errors) as all simulated removal rates are proportional to aerosol mixing ratios, and there is no feedback of aerosol concentrations onto other model variables such as cloud microphysics or meteorological variables.

4. The *prior probability density*  $\rho(\mathbf{d}, \mathbf{m})$  represents prior information constraining both the observable quantities and the model parameters.

If the data (observations of concentrations) are independent of the prior information about model parameters, we can write

$$\rho(\mathbf{d}, \mathbf{m}) = \rho_D(\mathbf{d})\rho_M(\mathbf{m}), \quad (2)$$

where  $\rho_D(\mathbf{d})$  represents prior information about the observable quantities and  $\rho_M(\mathbf{m})$  represents prior information about the model.

Here we assume that no information is available about the values of the observable parameters prior to the inversion – i.e. the measurements are the only source of information about the observable parameters:

$$\rho_D(\mathbf{d}) = \mu_D(\mathbf{d}). \quad (3)$$

5. The *conditional probability*  $\theta(\mathbf{d}|\mathbf{m})$  is the probability that the result of the model (the simulated data vector  $\mathbf{G}\mathbf{m}$ ) is correct, given the data.

In this study, we assume that the observations are independent and their uncertainty is Gaussian-distributed. The discretized formulation of  $\theta(\mathbf{d}|\mathbf{m})$  is then equal to the product of the normalized Gaussian probability density functions for each data point:

$$\theta(\mathbf{d}|\mathbf{m}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi s_{D,i}^2}} \exp \frac{-(g^i(\mathbf{m}) - \mathbf{d}_{\text{obs}}^i)^2}{2s_{D,i}^2}, \quad (4)$$

where  $\mathbf{d}_{\text{obs}}^i$  is the observed value of the  $i$ -th component of  $\mathbf{d}$ ,  $s_{D,i}$  is the standard deviation of the  $i$ -th observation (measurement uncertainty), and  $g^i(\mathbf{m})$  is the  $i$ -th component of  $g(\mathbf{m})$ ,

$$g^i(\mathbf{m}) = \sum_{j=1}^M \mathbf{G}_{ij} \mathbf{m}_j. \quad (5)$$

#### 4.1.1 Observable parameters

For the observations we assume Gaussian uncertainty, with coefficients taken from Table 1. The mean value is given by the best estimate. We estimate the standard deviation following the “range rule”, by which the standard deviation is approximately one-quarter of the range of observed values, which we calculate as (high estimate – low estimate)/4. We assume no prior information about the value of  $\mathbf{d}$ .

#### 4.1.2 Model prior and model error

For the prior information about the model we treat two cases. In the case NO-PRIOR,

$$\rho_M(\mathbf{m}) = \mu_M(\mathbf{m}), \quad (6)$$

while in the case PRIOR-POS,

$$\rho_M(\mathbf{m}) = \begin{cases} \mu_M(\mathbf{m}), & \mathbf{m} \geq 0 \\ 0, & \mathbf{m} < 0 \end{cases}. \quad (7)$$

We tested the effects of a prior positivity constraint on the model, i.e. a constraint that disallows negative emissions. Inversions with this constraint are designated PRIOR-POS, and the model prior  $\rho_M(\mathbf{m})$  is given by

$$\rho_M(\mathbf{m}) = \begin{cases} \mu_M(\mathbf{m}), & \mathbf{m} \geq 0 \\ 0, & \mathbf{m} < 0 \end{cases}, \quad (8)$$

where  $\mu_M(\mathbf{m})$  represents the homogeneous probability distribution for  $\mathbf{m}$  (Tarantola, 2005, Ch. 1). A positivity constraint can be justified by assuming that the atmospheric model accurately represents the removal of aerosols from the atmosphere, or at least does not underestimate removal. However, this constraint may not be justified if the atmospheric model underestimates removal. For this reason, and to illustrate the effect of the prior positivity constraint, we also present some results from inversions for which no prior constraint is applied, designated NO-PRIOR. Our reasoning for introducing the prior constraint is discussed further in Sect. 5.2.

For the purposes of the inversion, we do not explicitly include the model error. The Eq. (1.74) of Eq. (1.74) of Tarantola (2005) shows that that, for Gaussian error statistics, one can add the variances associated with model error and observational error. Given the very large observational errors we treat the model error as negligible (for each realization of the model).

## 4.2 Solution of the inverse problem

The solution of the inverse problem, the posterior probability distribution,  $\sigma_M(\mathbf{d}, \mathbf{m})$ , is given by

$$\sigma_M(\mathbf{d}, \mathbf{m}) = k\rho_M(\mathbf{m})L(\mathbf{m}), \quad (9)$$

where  $k$  is a normalization constant and  $\rho_M(\mathbf{m})$  is the prior probability density in the model space. The likelihood function  $L(\mathbf{m})$  is a measure of how well a candidate model  $\mathbf{m}$  explains the observations:

$$L(\mathbf{m}) = \int_{\mathbb{D}} \left[ \frac{\rho_D(\mathbf{d})\theta(\mathbf{d}|\mathbf{m})}{\mu_D(\mathbf{d})} \right] d\mathbf{d}, \quad (10)$$

where  $\int_{\mathbb{D}}$  is the integral over the data space  $\mathbb{D}$ .

Using Eqs. (3) and (4), we can directly calculate the likelihood function:

$$\begin{aligned} L(\mathbf{m}) &= \int_{\mathbb{D}} [\theta(\mathbf{d}|\mathbf{m})] d\mathbf{d} \\ &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi s_{D,i}^2}} \exp \frac{-(g^i(\mathbf{m}) - d_{\text{obs}}^i)^2}{2s_{D,i}^2}, \quad (11) \end{aligned}$$

where  $g^i(\mathbf{m})$  is the  $i$ -th component of  $g(\mathbf{m})$ ,  $d_{\text{obs}}^i$  is the observed value of the  $i$ -th component of  $\mathbf{d}$ , and  $s_{D,i}$  is the standard deviation of the  $i$ -th observation (measurement uncertainty).

## 4.3 The Metropolis algorithm

To solve the inversion problem, we apply a Markov Chain Monte Carlo algorithm (Metropolis et al., 1953), generating a large ensemble of solutions via a random walk algorithm, and applying a probabilistic rule to preferentially select solutions that have a higher likelihood.

1. Select initial guess for model parameters  $\mathbf{m}_0$ .
2. Add a small random vector  $\boldsymbol{\epsilon}$  to select next candidate solution,

$$\mathbf{m}_{i+1} = \mathbf{m}_i + \boldsymbol{\epsilon}. \quad (12)$$

3. Obtain a sample of the prior distribution: apply the Metropolis rule to determine whether to accept or reject the candidate solution based on the prior probability density  $\rho_M(\mathbf{m})$ :

- a. If  $\rho_{\mathbf{m}_{i+1}} \geq \rho_{\mathbf{m}_i}$ , accept the proposed transition to  $\mathbf{m}_{i+1}$ .
- b. If  $\rho_{\mathbf{m}_{i+1}} < \rho_{\mathbf{m}_i}$ , then accept the proposed move with probability

$$P_{i \rightarrow i+1} = \frac{\rho_{\mathbf{m}_{i+1}}}{\rho_{\mathbf{m}_i}}. \quad (13)$$

For PRIOR-POS, all candidate solutions  $\mathbf{m}_i$  in which all elements (individual model fluxes) are non-negative are retained, while all candidates with negative elements are rejected (Eq. 7). The retained  $\mathbf{m}_i$  are a sample of the prior ensemble.

4. For each member of the prior ensemble, calculate the likelihood function and apply the Metropolis rule to determine whether to accept or reject the candidate solution:

- a. If  $L(\mathbf{m}_{i+1}) \geq L(\mathbf{m}_i)$ , accept the proposed transition to  $\mathbf{m}_{i+1}$ .
- b. If  $L(\mathbf{m}_{i+1}) < L(\mathbf{m}_i)$ , then accept the proposed move with probability

$$P_{i \rightarrow j} = \frac{L(\mathbf{m}_{i+1})}{L(\mathbf{m}_i)}. \quad (14)$$

5. Repeat steps (2)–(4) for the desired number of trials.

Note that because only the ratios of the likelihoods associated with each model are of interest, there is no need to know the normalization parameter  $k$ , which remains constant.

Given a sufficiently large number of iterations, the ensemble of  $\mathbf{m}_i$  is a sample of the posterior probability distribution  $\sigma(\mathbf{d}, \mathbf{m})$ , i.e. the solution to the inverse problem (Eq. 9; Metropolis et al., 1953; Mosegaard and Tarantola, 1995). The centre and spread of the ensemble can be used as summary statistics of the posterior PDF.

**Table 3.** Uncertainty ranges used for calculation of model uncertainty, observation uncertainty, and normalized model uncertainty.

Source of uncertainty	Uncertainty range
Observations	5–95th percentile of ensemble
Particle diameter	$\pm 1 \mu\text{m}$
CCN activity	never CCN – always CCN
Mixed-phase ice scavenging	0.1–0.7
Cold ice scavenging	0.0–0.1

The MCMC method provides great flexibility at the cost of computational expense. For the current problem, the forward model is small and cheap (multiplication of a vector with a  $10 \times 10$  matrix), so the computational expense required to produce and analyse a large Monte Carlo ensemble is acceptable. While a full EMAC model run in the setup used for this study requires approximately 70 CPU hours of computation, the matrix multiplication can be performed one million times in less than 0.01 s of CPU time.

The ensemble generated is interpreted as an estimate of  $\sigma(\mathbf{m})$ . The centre of the ensemble is interpreted as an estimate of  $\mathbf{m}$ , and the spread of the ensemble as an estimate of the uncertainty in  $\mathbf{m}$ . As a measure of the centre, we use the mode (from the half-range mode estimator of Bickel, 2002) as testing revealed that the mode gave a more robust indication of the location of the ensemble peak than the median or mean, particularly in the PRIOR-POS simulations (compare Fig. 2). As a measure of the ensemble spread, we use the 5–95th percentile range.

#### 4.4 Definition of normalized model uncertainty

Comparing the uncertainties in the inversion that arise from the observation uncertainty with those that arise from model parameters (particle size, ice scavenging parameters, and CCN activity) can aid in developing an understanding of the relative utility of investing research efforts in various aspects of model development or in further observations.

For continuous parameters (particle size and ice nucleation scavenging ratio), we define the “linear model uncertainty” as the shift in the centre of the posterior distribution resulting from a small change in the value of the parameter:

$$\text{Linear model uncertainty} = \frac{\text{Parameter uncertainty} \times \text{Difference in centre of posterior distribution}}{\text{Difference in parameter}} \quad (15)$$

For discontinuous parameters (CCN-ACTIVE vs CCN-INACTIVE), we define the associated “model uncertainty” as the shift in the centre of the posterior distribution resulting from using different values of the parameter:

$$\text{Model uncertainty} = \text{Difference in centre of posterior distribution.} \quad (16)$$

To directly compare the uncertainty from model parameters with the uncertainty from observations, we define the “normalized model uncertainty” as the contribution of model uncertainty to the uncertainty in the posterior distribution, normalized relative to the observation uncertainty – i.e.

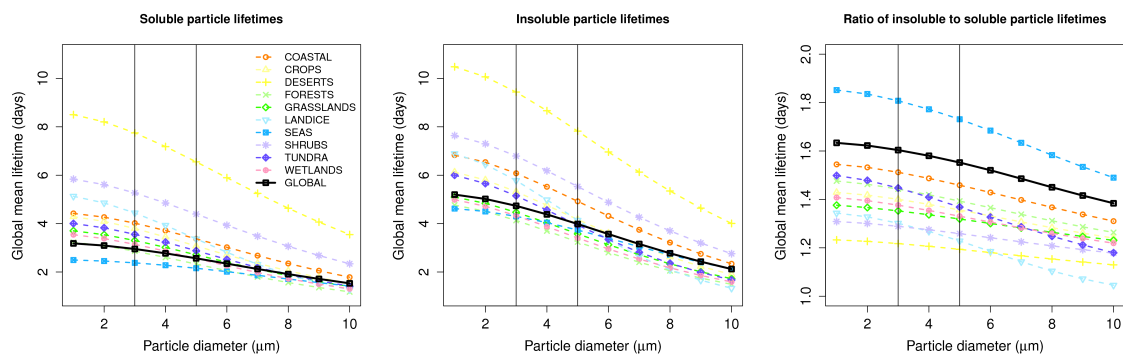
$$\text{Normalized model uncertainty} = \frac{\text{(Linear) model uncertainty}}{\text{Mean spread in posterior distribution}}. \quad (17)$$

The values used for the parameter uncertainty in Eq. (15) and the observation uncertainty are summarized in Table 3. We define the observation uncertainty as the spread in the posterior distribution, given by the middle 90 % range. The denominator in Eq. (17) is the average of the spreads in the two sensitivity cases. The model parameter uncertainties were chosen to approximately span the range of possible values. Limited observations of the size range of bacteria containing particles indicate that they are in the size range of about 1–5  $\mu\text{m}$ , with smaller particles observed at a coastal site (count median diameter 2.4  $\mu\text{m}$  and 95 % confidence interval 3.1–1.6  $\mu\text{m}$ ) and larger particles observed at inland sites (count median diameters near 4  $\mu\text{m}$  and 95 % confidence intervals ranging from 3.2 to 5.0  $\mu\text{m}$ ) (e.g. Shaffer and Lighthart, 1997; see also Tong and Lighthart, 2000; Wang et al., 2007). We therefore consider an uncertainty range of  $\pm 1 \mu\text{m}$  to represent the approximate range of size uncertainty per source. Because it is unclear to what extent bacteria are CCN active in the atmosphere, we consider the limiting cases: particles that are completely inactive and particles that are always active as CCN. The ice scavenging parameters span the ranges of values observed for cold and mixed-phase clouds (Sect. 3.2, Henning et al., 2004).

The normalized model uncertainty allows us to summarize the uncertainties resulting from different model parameters and place them on a single, dimensionless scale. A normalized uncertainty near unity means that the contributions of the uncertainty in the model parameter and of the uncertainty in the observations are roughly equivalent. A normalized uncertainty much less than unity implies that the model uncertainty for this parameter is much smaller than the observation uncertainty, while a normalized uncertainty much greater than unity implies that the model uncertainty is much greater than the observation uncertainty for this parameter.

To calculate the normalized CCN uncertainty, we calculate the difference between the centres of the posterior distributions for each pair of CCN-ACTIVE and CCN-INACTIVE cases where all other factors are held constant (Eq. 16).

To calculate the normalized size uncertainty, we calculate the difference between the centres of each neighbouring pair



**Fig. 1.** Global mean residence times of  $1\ \mu\text{m}$  particles in the BASE model setup as function of aerodynamic diameter, emission ecosystem, and CCN activity. Left panel: CCN-ACTIVE. Middle: CCN-INACTIVE. Right: ratio of CCN-INACTIVE to CCN-ACTIVE residence times. Points indicate the locations of the simulations performed. The colour key is indicated in the left panel.

of particle sizes (i.e.  $1\ \mu\text{m}$  and  $2\ \mu\text{m}$ ;  $2\ \mu\text{m}$  and  $3\ \mu\text{m}$ , and so on), where all other factors are held constant (Eq. 15).

To calculate the cold and mixed ice scavenging uncertainties, we use the differences between the BASE case and the LIM\_COLD and LIM\_MIXED cases, respectively, where the ice nucleation scavenging coefficient is perturbed and all other factors are held constant (Eq. 15).

In each case, we then normalize the result by calculating the spread in each of the pair of distributions and then calculating the mean value of these spreads (Eq. 17).

## 5 Results and discussion

### 5.1 Impact of size and emission region on particle residence times

The global mean atmospheric residence times of particles depend strongly on whether they act as CCN, on the particle diameter, and on the region from which they are emitted (Fig. 1). Atmospheric residence times are longest for particles emitted from deserts, where there is little scavenging and dry convection rapidly transports particles to high altitudes. Residence times of  $1\ \mu\text{m}$  particles are shortest for particles emitted from oceans, where scavenging is strongest. For larger particles and CCN-INACTIVE particles, for which particle lifetime is more strongly driven by dry deposition processes, particle lifetimes are shortest in crops and forest regions, where plants increase friction and provide large surfaces for deposition. The effect of particle size on residence time is somewhat stronger for CCN-INACTIVE particles than for CCN-ACTIVE particles since the residence time for CCN-INACTIVE particles is more strongly determined by dry deposition.

### 5.2 Mean annual flux estimated per ecosystem

The posterior distributions of the estimated fluxes for each particle size and source ecosystem are shown as histograms

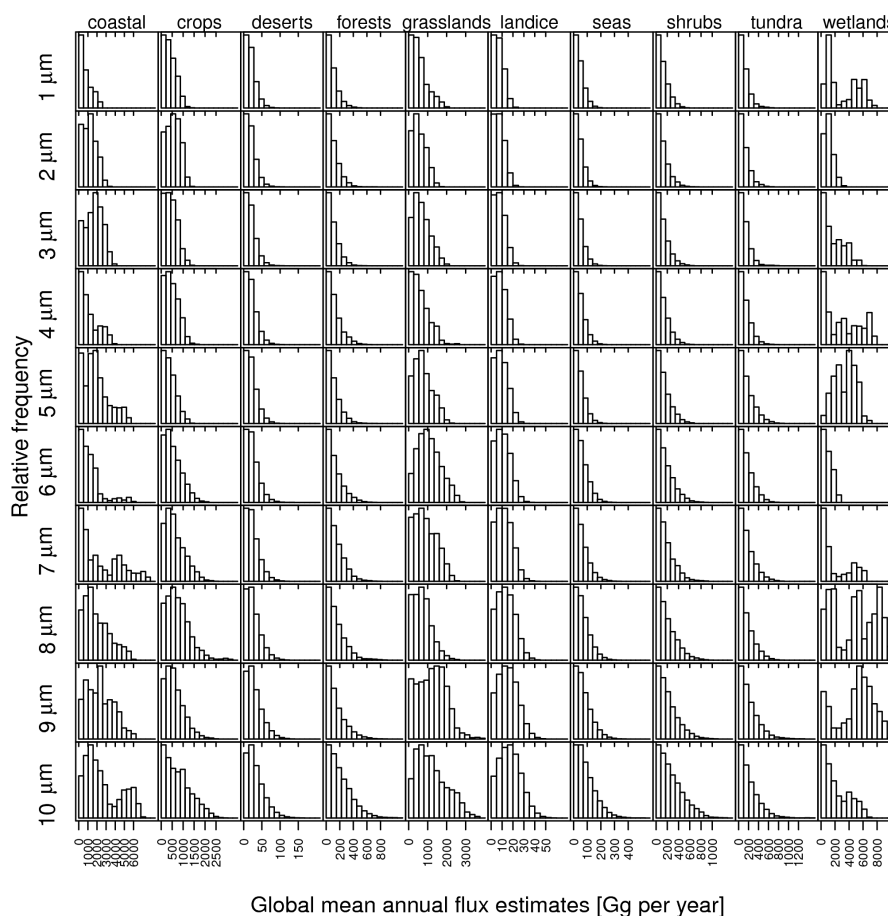
in Fig. 2 and Fig. 3. In each case, the typical posterior distribution of flux estimates for each ecosystem has an approximately Gaussian shape, which results from the assumption that the observation uncertainty has a Gaussian distribution. The histograms appear most irregular in the wetlands region, which is also the most poorly constrained by observations (Table 1).

As particle size increases, not only do estimated emissions increase but also estimated emissions for individual ecosystems typically increase as well. This can be seen in the histograms of the ecosystem emission estimates, especially when the emissions in each ecosystem are constrained to be positive (Fig. 2). When emissions are not constrained to be positive, this pattern is less clear (Fig. 3), especially for wetlands and coastal regions. These regions are poorly constrained by the observations due to their relatively small contribution to simulated concentrations: even large changes in the emissions in these regions have only a small influence on the concentrations in other regions, or on global emissions.

In NO-PRIOR, negative fluxes are allowed, and in some regions the most likely estimate of the flux is negative. Negative fluxes can occur for statistical or physical reasons. In many atmospheric tracer inversions (e.g. Gurney et al., 2002) the data are insufficient to constrain each emission individually. Posterior flux estimates are characterized by dipoles where the combination is well constrained but fluxes from individual regions can take on large values of opposite sign. Physically, negative emissions imply a mis-specification of deposition processes: the model cannot, with the deposition rates assumed in the transport matrix, simulate the concentration gradients observed between ecosystems.

In the NO-PRIOR case, flux estimates in different regions are highly cross-correlated (Fig. 4, left panel). These high correlations suggest that the negative inferred emissions in NO-PRIOR may be statistical and can be better explained by the weakness of the observational constraints on the inversion than by an underestimate of emissions.





**Fig. 2.** Histograms of the Monte Carlo sample of flux estimates for each ecosystem and particle sizes from  $1\ \mu\text{m}$  to  $10\ \mu\text{m}$  in  $1\ \mu\text{m}$  intervals. Here, for CCN-ACTIVE particles in PRIOR-POS, with Gaussian data uncertainty and a prior positivity constraint on fluxes.

For this reason, we introduced a positive prior constraint on each inferred flux (PRIOR-POS). In PRIOR-POS, the typical posterior distribution has the shape of a truncated Gaussian distribution (Fig. 2). This is because negative fluxes are disallowed, and the correlations between flux estimates in the different regions become very small: the additional constraint has the effect of somewhat decoupling the emissions from different regions (Fig. 4, right panel). We will focus on the results of the better-constrained PRIOR-POS inversion in the remainder of this paper.

### 5.3 Global annual mass emissions

Figure 5 shows the distribution of the global annual mass emitted for the Markov Chain Monte Carlo (MCMC) ensemble, for PRIOR-POS (with a positive constraint on the emissions). Interesting features of the posterior ensembles include the following:

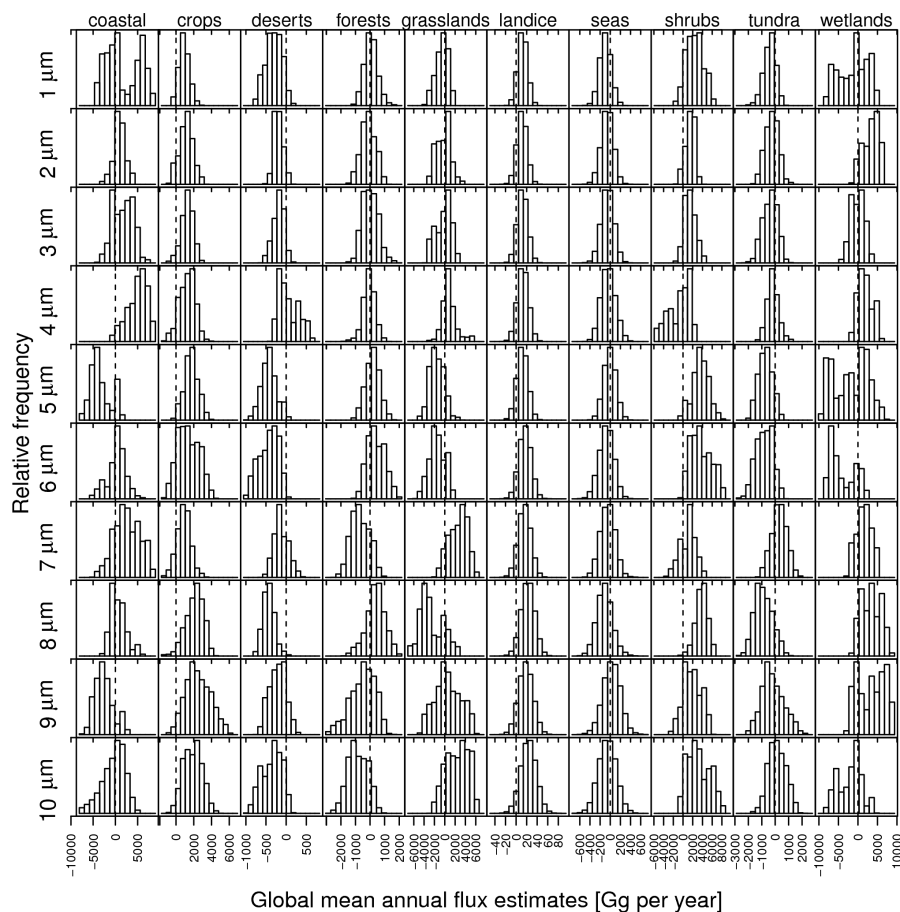
1. The ensembles exhibit a right-skewed distribution, with more high extreme values than low extreme values: this

results directly from the prior positive constraint on the emissions.

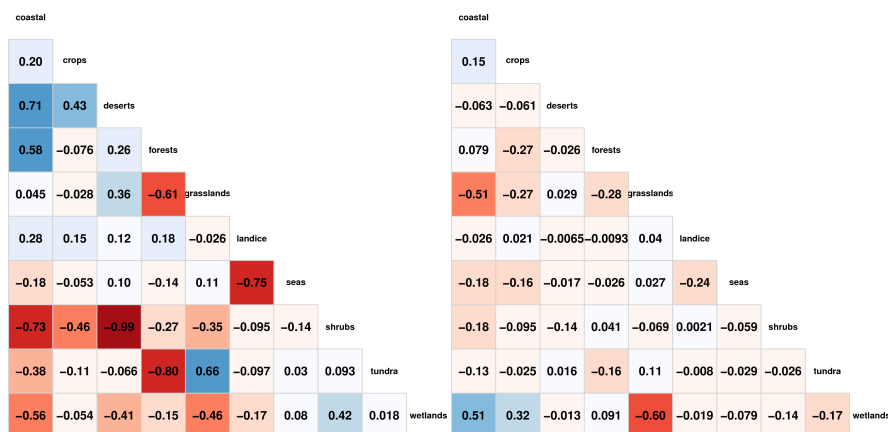
2. Estimated emissions increase with increasing particle size. Since the observed concentrations are defined independently of particle size, and since smaller particles have longer residence times, lower emissions of small particles are required to match the observed number concentrations compared to large particles, which have shorter residence times.
3. Higher emissions are estimated for CCN-ACTIVE particles than for CCN-INACTIVE particles; again, shorter particle lifetimes require larger emissions to match the observed concentrations.

### 5.4 Partitioning of uncertainty in total global emissions

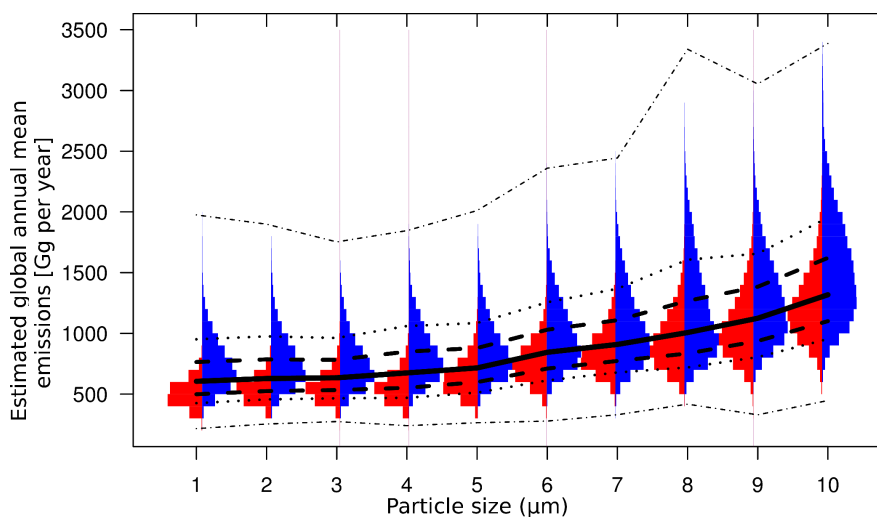
In Fig. 6 we show the magnitude of the uncertainties arising from each of the sources considered: observational uncertainty, particle size, CCN activity, mixed-phase scavenging



**Fig. 3.** Histograms of flux estimates, as in Fig. 2. Here, for CCN-ACTIVE particles in NO-PRIOR, with Gaussian data uncertainty and no prior constraint on fluxes. The vertical dashed lines indicate the position of zero estimated flux in each column.



**Fig. 4.** Correlations between flux estimates for different ecosystems. Here for CCN-ACTIVE, 1  $\mu\text{m}$  particles. Left: NO-PRIOR inversion (with no prior constraint on fluxes). Right: PRIOR-POS inversion (with prior positivity constraint on fluxes). To aid visual interpretation, positive correlations are shaded blue and negative correlations are shaded red, with darker hues corresponding to greater absolute values.



**Fig. 5.** Distributions of global annual mass emissions estimates [Gg per year] shown as histograms. Results are shown from left to right for each particle diameter from 1  $\mu\text{m}$  to 10  $\mu\text{m}$ , and for CCN-INACTIVE (red, on left for each size) and CCN-ACTIVE (blue, on right) particles. Lines demarcate the minimum, 10, 25, 50, 75, and 90th percentiles, and maximum of the total sample (CCN-ACTIVE and CCN-INACTIVE particles combined) for each particle size. Results are for PRIOR-POS.

and cold ice scavenging. As with the emissions themselves, uncertainties increase with particle size.

The uncertainty contribution from CCN activity is approximately constant in absolute terms across the size range of interest. This is expected because particle wet removal as parameterized in SCAV is effectively independent of particle size in this size range.

The uncertainty contribution from CCN activity decreases relative to particle size as particle size increases and median global emissions increase. Because this decrease is compensated by a growth in the relative uncertainty contribution from particle size, the total relative uncertainty remains approximately constant, near 150 % of the median global flux estimate. The contributions from ice scavenging parameters are comparatively small, although mixed-phase ice scavenging contributes more uncertainty than particle size for particle diameters from 1–4  $\mu\text{m}$ .

### 5.5 Normalized model uncertainties

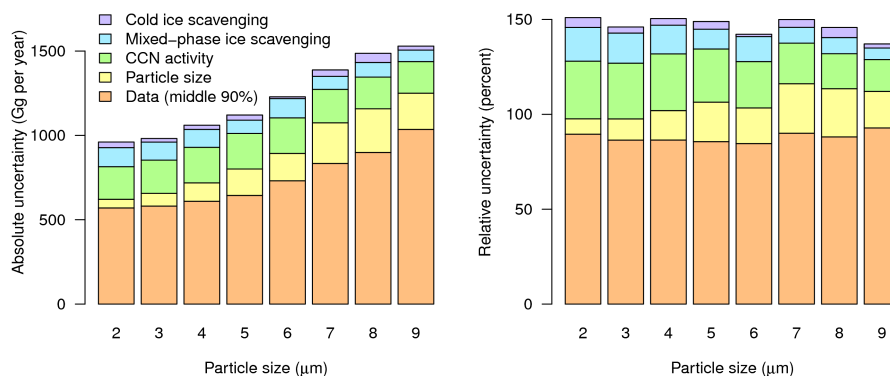
The normalized model uncertainties in the estimated global mean emissions are compared in Fig. 7. On average, the normalized model uncertainty from CCN activity is the largest. The additional effect of changing ice nucleation scavenging coefficients is minimal. The exception is the LIM\_MIXED case, where normalized CCN uncertainty is reduced, at least for smaller particle sizes (Fig. 8, top).

The normalized model uncertainty from particle size is slightly smaller than that from CCN activity. However, if the uncertainty in the particle size were appreciably larger than the  $\pm 1 \mu\text{m}$  assumed here, then the particle size would be the largest contributor to the uncertainty in source estimation. As

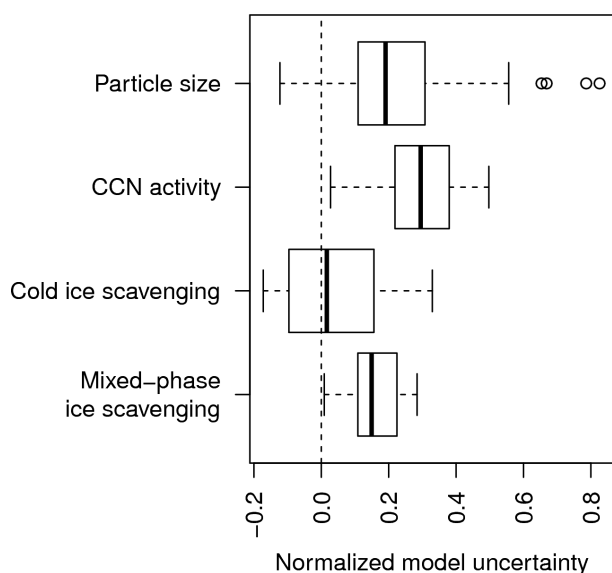
discussed in Sect. 4.4, the few measurements of the size of bacteria-containing particles that are available suggest that particles bearing culturable bacteria have diameters in the range of about 1–5  $\mu\text{m}$ , with ranges of ca.  $\pm 1 \mu\text{m}$  from the median at each individual measurement location (e.g. Shaffer and Lighthart, 1997; Lighthart, 2000; Tong and Lighthart, 2000; Wang et al., 2007), so the uncertainty range in particle size could plausibly be considered to be as large as  $\pm 2 \mu\text{m}$ , rather than the  $\pm 1 \mu\text{m}$  used in this study. Doubling the uncertainty in particle size would approximately double its contribution to the model parameter uncertainty in the inversion.

Figure 7 shows that the normalized model uncertainty for CCN activity is about 20–30 %. The normalized model uncertainty can have a negative value; this indicates that the median of the distribution has shifted in the opposite direction from the expected one, which can occur as the result of randomness in the Monte Carlo method. In Fig. 8, we show that the variability in this value is partly explained by the variation of CCN uncertainty with particle size. For 10  $\mu\text{m}$  particles, the normalized model uncertainty for CCN activity is significantly smaller (close to 20 %) than for 1  $\mu\text{m}$  particles (close to 40 %). This is attributable to the increasing ensemble spread (observation uncertainty) at larger particle sizes (Fig. 5) since the absolute magnitude of the CCN uncertainty remains roughly constant with varying particle size (Fig. 6).

Similarly, the normalized model uncertainty for particle size varies as a function of particle size. Model sensitivity to particle size is greater for larger particles and for CCN-INACTIVE particles. This is also seen in Fig. 1, where the steeper slopes of the lines indicates a higher sensitivity to particle size for larger particles and for CCN-ACTIVE particles.



**Fig. 6.** Top: model and observation uncertainty in the inferred mean global flux (median per uncertainty type for various cases), expressed as absolute uncertainty. Bottom: same, expressed as a relative uncertainty, i.e. (absolute uncertainty in global flux estimate)/(median global flux estimate) · 100. The sum of the individual uncertainties exceeds the value of the median global flux estimate in each case, resulting in relative uncertainties that exceed 100%. The colour key is indicated in the left panel.



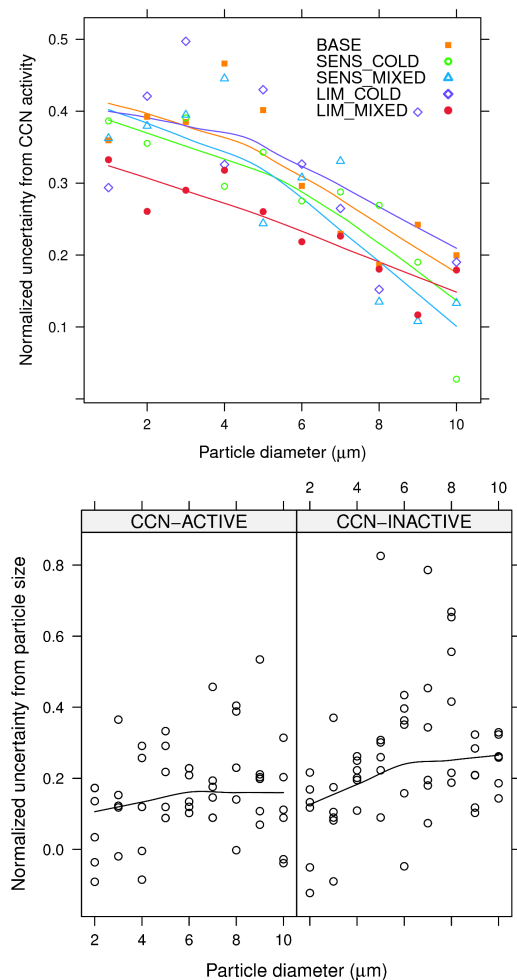
**Fig. 7.** Boxplot of the distributions of normalized model uncertainties for particle size, CCN activity, and the LIM\_MIXED and LIM\_COLD tests of the ice scavenging parameterization. Thick central lines show the median; box extent shows the 25–75th percentile; whiskers extend up to 1.5 times the length of the box, or to the most extreme point; dots show outliers beyond the whiskers.

The sensitivity to the CCN activity of the particles is high for particles of 1  $\mu\text{m}$  diameter, decreasing to only moderate sensitivity for particles around 10  $\mu\text{m}$  diameter (Fig. 8). The sensitivity to particle size is small to moderate for particles around 1  $\mu\text{m}$  diameter, but increases to a large sensitivity for particles around 10  $\mu\text{m}$  diameter. The sensitivity to particle size is higher for CCN-INACTIVE particles than for CCN-ACTIVE particles, particularly at particle sizes closer to 1  $\mu\text{m}$  (Fig. 8).

All effects shown in Fig. 8 and discussed above are statistically significant by the Student's *t* test ( $p < 0.01$ ). Other significant effects include the following: a reduction in the normalized uncertainty from CCN activity in the LIM\_MIXED simulation (after controlling for the effect of particle size;  $p < 0.01$ ) and an increase in the normalized uncertainty from particle size as particle size increases ( $p < 0.01$ ). Three of the statistically significant interactions between model parameters and the normalized model uncertainties are illustrated in the linear effect diagrams shown in Fig. 9. Each diagram shows the marginal effect of one of the three predictor variables (particle size, CCN activity, and IN sensitivity case), while the others are held constant (Fox, 1987, 2003).

## 6 Discussion

In this study, we have focused on the impact of model parameters describing particle characteristics that affect simulated aerosol removal processes within the context of a particular realization of a global chemistry-climate model. However, tracer transport in global chemistry-climate models can also be sensitive to many other aspects of the model that are beyond the scope of this study, such as model resolution, the use of prescribed meteorology to nudge the chemistry climate model towards the observed atmospheric state (Aghedo et al., 2010), numerical formulations of atmospheric dynamics (Rasch et al., 2006), and the formulation of the parameterization of deep convective transport (Mahowald et al., 1995; Luo et al., 2003; Tost et al., 2006b; Lawrence and Salzmann, 2008). The sensitivity of modelled transport to particle characteristics as simulated here could potentially change if these or other aspects of the model were changed. We did not consider changes in the aerosol size distribution due to aerosol microphysics and/or cloud processing (Hoose et al., 2008),



**Fig. 8.** Top: normalized model uncertainty from CCN activity as a function of particle diameter. Bottom: normalized model uncertainty due to particle size as a function of particle size. Lines in both panels are local polynomial regression fits to the points and are provided as a guide to the eye. The colour key is indicated in the top panel.

which could introduce additional uncertainty into the aerosol transport and residence times.

Uncertainty in the estimation of global mean emissions arises from both observation uncertainty and uncertainty in model parameters, including particle size, particle CCN activity, and the rate of ice nucleation scavenging in cold and mixed-phase clouds. Relative to the observation uncertainty (5–95th percentile of the Monte Carlo ensemble), the normalized model uncertainty due to CCN activity or a change in particle size of 1  $\mu\text{m}$  is typically between 10 and 40 %. The model uncertainty from ice nucleation scavenging in cold clouds was negligible, but the normalized model uncertainty from ice nucleation scavenging in mixed-phase clouds was 10–20 %.

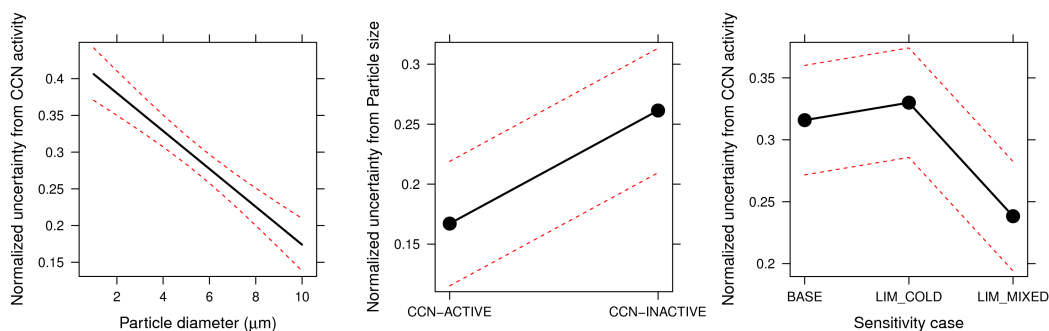
The sensitivity to scavenging in mixed-phase clouds is perhaps surprisingly large – a reflection of the large uncertainty in the process – for which it is difficult to find constraints from observations. Other studies have shown that simulated particle transport to high latitudes and the upper troposphere in particular is highly sensitive to the scavenging rates in mixed-phase clouds, and moderately sensitive to scavenging rates in ice clouds (Burrows, 2011; Bourgeois and Bey, 2011; Zhang et al., 2012). As the relative contributions of different source regions to particulate air pollution in the Arctic has been a matter of significant interest for scientific research and public policy in recent years, this may deserve additional attention.

Since transport and removal processes are the basis for the correct simulation of the distributions of any atmospheric aerosol, some of the results of this study also are broadly applicable to other coarse atmospheric aerosols. The dependence of atmospheric residence time (attributable to transport and removal) on size and emission region is a result that holds generally for particles emitted from those regions (Fig. 1). The important contribution of uncertainty in particle size to uncertainties in simulated particle residence time will also be applicable for other coarse aerosols, particularly as size approaches 10  $\mu\text{m}$  (Fig. 6). The activation of particles as CCN is better constrained for many other atmospheric particle types than for bacterial aerosol, so the contribution of CCN uncertainty may be reduced in other cases. For fine (submicron) aerosol particles, the relative contributions of different uncertainty types are likely to differ from the results presented here as the efficiency of removal processes changes with size.

## 7 Conclusions

In this study, we estimated emissions of bacterial aerosol from inversion of atmospheric transport by a Markov Chain Monte Carlo (MCMC) method. Bacterial aerosols were treated as passive (non-reactive) tracers; thus, we characterize the impacts of model transport and removal processes in isolation from other model processes such as aerosol microphysics or chemistry. The method presented here and applied to a comparatively simple model setup can be adapted in a straightforward way for application to more complex scenarios.

An MCMC inversion was applied to estimate emissions of bacteria-containing particles from different ecosystems given a set of mean observed concentrations and associated uncertainties, analogous to Burrows et al. (2009a). We performed both an unconstrained inversion (NO-PRIOR) and an inversion constrained by a prior assumption that emissions must be positive (PRIOR-POS). The NO-PRIOR inversion leaves open the possibility of an underestimate of deposition that can result in a negative inferred flux. However, the high correlations between the inferred fluxes in different ecosystems



**Fig. 9.** Illustration of the marginal effects of model parameters on the model uncertainty, shown in linear effects plots following Fox (1987, 2003). Each plot shows the values predicted by a generalized linear model across the range of values of a main predictor variable, while other predictor variables are held constant at an average value. Left: effect of particle size on CCN uncertainty, controlling for IN sensitivity case. Middle: effect of CCN activity on size uncertainty, controlling for IN sensitivity case. Right: effect of sensitivity case on CCN uncertainty, controlling for CCN activity and size. Red dashed lines indicate the 95th percentile confidence interval.

indicate that negative inferred fluxes can be better explained by the problem being poorly constrained.

The results of the MCMC estimation agree well with results from a constrained linear optimization for the same problem, as presented in Burrows et al. (2009a). In the MCMC inversion, the range of uncertainty attributed to observations is reduced nearly by half: for 1  $\mu\text{m}$ , CCN-ACTIVE tracers, the 5–95th percentile range is 400–1800  $\text{Gg a}^{-1}$  in Burrows et al. (2009a), compared with 470–1100  $\text{Gg a}^{-1}$  from the MCMC ensemble, a reduction in the uncertainty range of 45 %. The narrower range of the uncertainty in the present study is likely better explained by the fact that we here treat the observation uncertainty here as having a Gaussian probability distribution rather than a homogeneous probability distribution. Also, the ecosystems shown in Burrows et al. (2009a) to be most poorly constrained by the observations – wetlands and coastal regions – were similarly poorly constrained when using the Monte Carlo method. This is apparent in the broad spread and irregular shapes of the distributions in those regions (Fig. 5).

The contribution of the uncertainties attributable to particle size and CCN activity was unexpectedly large compared to observation uncertainty for this case study, implying that these parameters should be better constrained by observations and/or the associated uncertainty should be explicitly considered in the inversion of atmospheric aerosol transport and uncertainty analyses in global atmospheric modelling studies involving aerosols. Ice nucleation scavenging in mixed-phase clouds in particular is rarely considered, but may contribute significantly to overall model uncertainty for problems involving aerosol transport. Even in cases with very large observation uncertainty, the contribution of model parameter uncertainty can be substantial.

The results of this study can be used to help guide future observations of bacterial aerosol that aim to improve the estimation of sources. In particular, they make clear that

the uncertainty in the particle size distribution makes a non-negligible contribution to the uncertainty in source estimates, and future observations of bacterial aerosol should include size information if possible.

*Acknowledgements.* Susannah Burrows was supported in part by the U. S. Department of Energy as part of the Earth System Modeling Program. Peter Rayner is in receipt of an Australian Professorial Fellowship (DP1096309). We gratefully acknowledge the efforts of the EMAC development team to develop and make available the EMAC modelling system.

The service charges for this open access publication have been covered by the Max Planck Society.

Edited by: W. Lahoz

## References

- Aghedo, A. M., Rast, S., and Schultz, M. G.: Sensitivity of tracer transport to model resolution, prescribed meteorology and tracer lifetime in the general circulation model ECHAM5, *Atmos. Chem. Phys.*, 10, 3385–3396, doi:10.5194/acp-10-3385-2010, 2010.
- Aquila, V., Hendricks, J., Lauer, A., Riemer, N., Vogel, H., Baumgardner, D., Minikin, A., Petzold, A., Schwarz, J. P., Spackman, J. R., Weinzierl, B., Righi, M., and Dall’Amico, M.: MADE-in: a new aerosol microphysics submodel for global simulation of insoluble particles and their mixing state, *Geosci. Model Dev.*, 4, 325–355, doi:10.5194/gmd-4-325-2011, 2011.
- Astitha, M., Lelieveld, J., Abdel Kader, M., Pozzer, A., and de Meij, A.: Parameterization of dust emissions in the global atmospheric chemistry-climate model EMAC: impact of nudging and soil properties, *Atmos. Chem. Phys.*, 12, 11057–11083, doi:10.5194/acp-12-11057-2012, 2012.
- Bickel, D.: Robust estimators of the mode and skewness of continuous data, *Computational statistics & data analysis*, 39, 153–163, 2002.



- Bourgeois, Q. and Bey, I.: Pollution transport efficiency toward the Arctic: Sensitivity to aerosol scavenging and source regions, *J. Geophys. Res.*, 116, D08213, doi:10.1029/2010JD015096, 2011.
- Burrows, S. M.: Sources and concentrations of biological particles in the global atmosphere, and their significance as ice nuclei, Ph.D. thesis, Max Planck Graduate Center mit der Johannes Gutenberg-Universität Mainz, 2011.
- Burrows, S. M., Butler, T., Jöckel, P., Tost, H., Kerkweg, A., Pöschl, U., and Lawrence, M. G.: Bacteria in the global atmosphere – Part 2: Modeling of emissions and transport between different ecosystems, *Atmos. Chem. Phys.*, 9, 9281–9297, doi:10.5194/acp-9-9281-2009, 2009a.
- Burrows, S. M., Elbert, W., Lawrence, M. G., and Pöschl, U.: Bacteria in the global atmosphere – Part 1: Review and synthesis of literature data for different ecosystems, *Atmos. Chem. Phys.*, 9, 9263–9280, doi:10.5194/acp-9-9263-2009, 2009b.
- de Meij, A., Pozzer, A., Pringle, K. J., Tost, H., and Lelieveld, J.: EMAC model evaluation and analysis of atmospheric aerosol properties and distribution with a focus on the Mediterranean region, *Atmos. Res.*, 114–115, 38–69, 2012.
- Després, V. R., Huffman, J. A., Burrows, S. M., Hoose, C., Safatov, A. S., Buryak, G., Fröhlich-Nowoisky, J., Elbert, W., Andreae, M. O., Pöschl, U., and Jaenicke, R.: Primary Biological Aerosols in the Atmosphere: A Review of Observations and Relevance, *Tellus B*, 64, 15598, doi:10.3402/tellusb.v64i0.15598, 2012.
- Enting, I.: Inverse Methods in Global Biogeochemical Cycles, chap. Green's function methods of tracer inversion, American Geophysical Union, 2000.
- Enting, I. G.: Inverse problems in atmospheric constituent transport, Cambridge University Press, 2002.
- Fox, J.: Effect displays for generalized linear models, *Sociological methodology*, 17, 347–361, doi:10.2307/271037, 1987.
- Fox, J.: Effect displays in R for generalised linear models, *J. Stat. Softw.*, 8, 1–27, 2003.
- Gläser, G., Kerkweg, A., and Wernli, H.: The Mineral Dust Cycle in EMAC 2.40: sensitivity to the spectral resolution and the dust emission scheme, *Atmos. Chem. Phys.*, 12, 1611–1627, doi:10.5194/acp-12-1611-2012, 2012.
- Graham, B., Guyon, P., Maenhaut, W., Taylor, P. E., Ebert, M., Matthias-Maser, S., Mayol-Bracero, O. L., Godoi, R. H. M., Artaxo, P., Meixner, F. X., Moura, M. A. L., Rocha, C. H. E. D., Grieken, R. V., Glovsky, M. M., Flagan, R. C., and Andreae, M. O.: Composition and diurnal variability of the natural Amazonian aerosol, *J. Geophys. Res.*, 108, 5–1, doi:10.1029/2003JD004049, 2003.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., et al.: Towards robust regional estimates of CO<sub>2</sub> sources and sinks using atmospheric transport models, *Nature*, 415, 626–630, 2002.
- Henning, S., Bojinski, S., Diehl, K., Ghan, S., Nyeki, S., Weingartner, E., Wurzler, S., and Baltensperger, U.: Aerosol partitioning in natural mixed-phase clouds, *Geophys. Res. Lett.*, 31, L06101, doi:10.1029/2003GL019025, 2004.
- Hoose, C., Lohmann, U., Bennartz, R., Croft, B., and Lesins, G.: Global simulations of aerosol processing in clouds, *Atmos. Chem. Phys.*, 8, 6939–6963, doi:10.5194/acp-8-6939-2008, 2008.
- Huneeus, N., Schulz, M., Balkanski, Y., Griesfeller, J., Prospero, J., Kinne, S., Bauer, S., Boucher, O., Chin, M., Dentener, F., Diehl, T., Easter, R., Fillmore, D., Ghan, S., Ginoux, P., Grini, A., Horowitz, L., Koch, D., Krol, M. C., Landing, W., Liu, X., Mahowald, N., Miller, R., Morcrette, J.-J., Myhre, G., Penner, J., Perlwitz, J., Stier, P., Takemura, T., and Zender, C. S.: Global dust model intercomparison in AeroCom phase I, *Atmos. Chem. Phys.*, 11, 7781–7816, doi:10.5194/acp-11-7781-2011, 2011.
- Jaenicke, R.: Abundance of cellular material and proteins in the atmosphere, *Science*, 308, 73, doi:10.1126/science.1106335, 2005.
- Jaenicke, R., Matthias-Maser, S., and Gruber, S.: Omnipresence of biological material in the atmosphere, *Environ. Chem.*, 4, 217–220, doi:10.1071/EN07021, 2007.
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., et al.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, *J. Geophys. Res.*, 116, G00J07, doi:10.1029/2010JG001566, 2011.
- Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G.: On aggregation errors in atmospheric transport inversions, *J. Geophys. Res.-Atmos.* (1984–2012), 106, 4703–4715, 2001.
- Kerkweg, A., Buchholz, J., Ganzeveld, L., Pozzer, A., Tost, H., and Jöckel, P.: Technical Note: An implementation of the dry removal processes DRY DEPosition and SEDimentation in the Modular Earth Submodel System (MESSy), *Atmos. Chem. Phys.*, 6, 4617–4632, doi:10.5194/acp-6-4617-2006, 2006a.
- Kerkweg, A., Sander, R., Tost, H., and Jöckel, P.: Technical note: Implementation of prescribed (OFFLEM), calculated (ONLEM), and pseudo-emissions (TNUDGE) of chemical species in the Modular Earth Submodel System (MESSy), *Atmos. Chem. Phys.*, 6, 3603–3609, doi:10.5194/acp-6-3603-2006, 2006b.
- Kinne, S., Schulz, M., Textor, C., Guibert, S., Balkanski, Y., Bauer, S. E., Bernsten, T., Berglen, T. F., Boucher, O., Chin, M., Collins, W., Dentener, F., Diehl, T., Easter, R., Feichter, J., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Herzog, M., Horowitz, L., Isaksen, I., Iversen, T., Kirkevåg, A., Kloster, S., Koch, D., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Lesins, G., Liu, X., Lohmann, U., Montanaro, V., Myhre, G., Penner, J., Pitari, G., Reddy, S., Seland, O., Stier, P., Takemura, T., and Tie, X.: An AeroCom initial assessment – optical properties in aerosol component modules of global models, *Atmos. Chem. Phys.*, 6, 1815–1834, doi:10.5194/acp-6-1815-2006, 2006.
- Klinger, C.: Quantitative evaluation of ozone and selected climate parameters in the chemistry-climate model EMAC, Ph.D. thesis, Master thesis, Ludwig Maximilian University (LMU), Munich, Germany, 2011.
- Lamarque, J.-F., Shindell, D. T., Josse, B., Young, P. J., Cionni, I., Eyring, V., Bergmann, D., Cameron-Smith, P., Collins, W. J., Doherty, R., Dalsoren, S., Faluvegi, G., Folberth, G., Ghan, S. J., Horowitz, L. W., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Naik, V., Plummer, D., Righi, M., Rumbold, S. T., Schulz, M., Skeie, R. B., Stevenson, D. S., Strode, S., Sudo, K., Szopa, S., Voulgarakis, A., and Zeng, G.: The Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): overview and description of models, simulations and climate diagnostics, *Geosci. Model Dev.*, 6, 179–206, doi:10.5194/gmd-6-179-2013, 2013.
- Lawrence, M. and Rasch, P.: Tracer transport in deep convective updrafts: Plume ensemble versus bulk formulations, *J. Atmos.*

- Sci., 62, 2880–2894, 2005.
- Lawrence, M. G. and Salzmann, M.: On interpreting studies of tracer transport by deep cumulus convection and its effects on atmospheric chemistry, *Atmos. Chem. Phys.*, 8, 6037–6050, doi:10.5194/acp-8-6037-2008, 2008.
- Lelieveld, J., Kunkel, D., and Lawrence, M. G.: Global risk of radioactive fallout after major nuclear reactor accidents, *Atmos. Chem. Phys.*, 12, 4245–4258, doi:10.5194/acp-12-4245-2012, 2012.
- Lighthart, B.: Mini-review of the concentration variations found in the alfresco atmospheric bacterial populations, *Aerobiologia*, 16, 7–16, 2000.
- Luo, C., Mahowald, N. M., and del Corral, J.: Sensitivity study of meteorological parameters on mineral aerosol mobilization, transport, and distribution, *J. Geophys. Res.*, 108, 4447, doi:10.1029/2003JD003483, 2003.
- Mahowald, N. M., Rasch, P. J., and Prinn, R. G.: Cumulus parameterizations in chemical transport models, *J. Geophys. Res.-Atmos.*, 100, 26173–26189, doi:10.1029/95JD02606, 1995.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E.: Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21, 1087, doi:10.1063/1.1699114, 1953.
- Mosegaard, K. and Tarantola, A.: Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, 100, 12431–12447, doi:10.1029/94JB03097, 1995.
- Olson, J.: World ecosystems (WE1.4): Digital raster data on a 10 minute geographic 1080 (2160 grid square), Global Ecosystem Database, Version, 1, <http://www.fao.org/geonetwork/srv/en/metadata.show?id=1007> (last access: February 2013), 1992.
- Pozzer, A., de Meij, A., Pringle, K. J., Tost, H., Doering, U. M., van Aardenne, J., and Lelieveld, J.: Distributions and regional budgets of aerosols and their precursors simulated with the EMAC chemistry-climate model, *Atmos. Chem. Phys.*, 12, 961–987, doi:10.5194/acp-12-961-2012, 2012.
- Pringle, K. J., Tost, H., Message, S., Steil, B., Giannadaki, D., Nenes, A., Fountoukis, C., Stier, P., Vignati, E., and Lelieveld, J.: Description and evaluation of GMXe: a new aerosol submodel for global simulations (v1), *Geosci. Model Dev.*, 3, 391–412, doi:10.5194/gmd-3-391-2010, 2010.
- Pruppacher, H. and Klett, J.: *Microphysics of Clouds and Precipitation*, Kluwer Academic Pub, Dordrecht, The Netherlands, 2nd rev. and enl. ed. edn., 1997.
- Rasch, P. J., Coleman, D. B., Mahowald, N., Williamson, D. L., Lin, S.-J., Boville, B. A., and Hess, P.: Characteristics of atmospheric transport using three numerical formulations for atmospheric dynamics in a single GCM framework, *J. Climate*, 19, 2243–2266, 2006.
- Rayner, P. J., Raupach, M. R., Paget, M., Peylin, P., and Koffi, E.: A new global gridded dataset of CO<sub>2</sub> emissions from fossil fuel combustion: 1: Methodology and evaluation, *J. Geophys. Res.*, 115, D19306, doi:10.1029/2009JD013439, 2010.
- Schwarzenböck, A., Mertes, S., Heintzenberg, J., Wobrock, W., and Laj, P.: Impact of the Bergeron-Findeisen process on the release of aerosol particles during the evolution of cloud ice, *Atmos. Res.*, 58, 295–313, doi:10.1016/S0169-8095(01)00096-5, <http://www.sciencedirect.com/science/article/pii/S0169809501000965>, 2001.
- Seinfeld, J. H. and Pandis, S. N.: *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* (John Wiley), New York, 2006.
- Shaffer, B. T. and Lighthart, B.: Survey of culturable airborne bacteria at four diverse locations in Oregon: Urban, rural, forest, and coastal, *Microb. Ecol.*, 34, 167–177, 1997.
- Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Avery, K. B., Tignor, M., and Miller, H. L., eds.: *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, [http://www.ipcc.ch/publications\\_and\\_data/publications\\_ipcc\\_fourth\\_assessment\\_report\\_wg1\\_report\\_the\\_physical\\_science\\_basis.htm](http://www.ipcc.ch/publications_and_data/publications_ipcc_fourth_assessment_report_wg1_report_the_physical_science_basis.htm) (last access: February 2013), 2007.
- Tarantola, A.: *Inverse problem theory and methods for model parameter estimation*, Society for Industrial and Applied Mathematics, Philadelphia, <http://www.ipgp.fr/~tarantola/Files/Professional/Books/InverseProblemTheory.pdf> (last access: February 2013), 2005.
- Tong, Y. and Lighthart, B.: The annual bacterial particle concentration and size distribution in the ambient atmosphere in a rural area of the Willamette Valley, Oregon, *Aerosol Sci. Tech.*, 32, 393–403, doi:10.1080/027868200303533, 2000.
- Tost, H., Jöckel, P., Kerkweg, A., Sander, R., and Lelieveld, J.: Technical note: A new comprehensive SCAVenging submodel for global atmospheric chemistry modelling, *Atmos. Chem. Phys.*, 6, 565–574, doi:10.5194/acp-6-565-2006, 2006a.
- Tost, H., Jöckel, P., and Lelieveld, J.: Influence of different convection parameterisations in a GCM, *Atmos. Chem. Phys.*, 6, 5475–5493, doi:10.5194/acp-6-5475-2006, 2006b.
- Tost, H., Lawrence, M. G., Brühl, C., Jöckel, P., The GABRIEL Team, and The SCOUT-O3-DARWIN/ACTIVE Team: Uncertainties in atmospheric chemistry modelling due to convection parameterisations and subsequent scavenging, *Atmos. Chem. Phys.*, 10, 1931–1951, doi:10.5194/acp-10-1931-2010, 2010.
- Verheggen, B., Cozic, J., Weingartner, E., Bower, K., Mertes, S., Connolly, P., Gallagher, M., Flynn, M., Choulaton, T., and Baltensperger, U.: Aerosol partitioning between the interstitial and the condensed phase in mixed-phase clouds, *J. Geophys. Res.*, 112, D23202, doi:10.1029/2007JD008714, 2007.
- Wang, C.-C., Fang, G.-C., and Lee, L.-Y.: Bioaerosols study in central Taiwan during summer season, *Toxicol. Ind. Health*, 23, 133–139, doi:10.1177/0748233707078741, 2007.
- Zhang, K., O'Donnell, D., Kazil, J., Stier, P., Kinne, S., Lohmann, U., Ferrachat, S., Croft, B., Quaas, J., Wan, H., Rast, S., and Feichter, J.: The global aerosol-climate model ECHAM-HAM, version 2: sensitivity to improvements in process representations, *Atmos. Chem. Phys.*, 12, 8911–8949, doi:10.5194/acp-12-8911-2012, 2012.