Atmospheric
Chemistry
and Physics

Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods

A. L. Ganesan[1,*], **M. Rigby**[2], **A. Zammit-Mangion**[3], **A. J. Manning**[4], **R. G. Prinn**[1], **P. J. Fraser**[5], **C. M. Harth**[6], **K.-R. Kim**[7], **P. B. Krummel**[5], **S. Li**[8], **J. Mühle**[6], **S. J. O'Doherty**[2], **S. Park**[9], **P. K. Salameh**[6], **L. P. Steele**[5], and **R. F. Weiss**[6]

[1]Center for Global Change Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[2]School of Chemistry, University of Bristol, Bristol, UK
[3]School of Geographical Sciences, University of Bristol, Bristol, UK
[4]Hadley Centre, Met Office, Exeter, UK
[5]Centre for Australian Weather and Climate Research, CSIRO Marine and Atmospheric Research, Aspendale, Victoria, Australia
[6]Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, USA
[7]GIST College, Gwangju Institute of Science and Technology, Kwangju, South Korea
[8]Research Institute of Oceanography, Seoul National University, Seoul, South Korea
[9]Department of Oceanography, Kyungpook National University, Sangju, South Korea
[*]now at: School of Chemistry, University of Bristol, Bristol, UK

Correspondence to: A. L. Ganesan (anita.ganesan@bristol.ac.uk)

Abstract. We present a hierarchical Bayesian method for atmospheric trace gas inversions. This method is used to estimate emissions of trace gases as well as "hyper-parameters" that characterize the probability density functions (PDFs) of the a priori emissions and model-measurement covariances. By exploring the space of "uncertainties in uncertainties", we show that the hierarchical method results in a more complete estimation of emissions and their uncertainties than traditional Bayesian inversions, which rely heavily on expert judgment. We present an analysis that shows the effect of including hyper-parameters, which are themselves informed by the data, and show that this method can serve to reduce the effect of errors in assumptions made about the a priori emissions and model-measurement uncertainties. We then apply this method to the estimation of sulfur hexafluoride (SF_6) emissions over 2012 for the regions surrounding four Advanced Global Atmospheric Gases Experiment (AGAGE) stations. We find that improper accounting of model representation uncertainties, in particular, can lead to the derivation of emissions and associated uncertainties that are unrealistic and show that those derived using the hierarchical method are likely to be more representative of the true uncertainties in the system. We demonstrate through this SF_6 case study that this method is less sensitive to outliers in the data and to subjective assumptions about a priori emissions and model-measurement uncertainties than traditional methods.

1 Introduction

Inverse modeling is widely used to estimate sources and sinks of trace gas fluxes and their distributions using measurements of atmospheric mole fractions and chemical transport models (CTMs). The estimation of surface fluxes has been performed at a variety of spatial and temporal scales, ranging from regional (e.g., Stohl et al., 2009; Manning et al., 2011; Rigby et al., 2011) to global (e.g., Chen and Prinn, 2006; Mühle et al., 2010; Bousquet et al., 2011) and for timescales of hours to years.

Most inversions utilize a Bayesian framework and incorporate a priori information to condition the system, as shown by Eq. (1) (normalizing factors are not shown throughout this

text for brevity) (Enting et al., 1995).

$$\rho(\boldsymbol{x}|\boldsymbol{y}) \propto \rho(\boldsymbol{y}|\boldsymbol{x})\rho(\boldsymbol{x}) \qquad (1)$$

The Bayesian framework with Gaussian probability density functions (PDFs) and linear models used in most trace gas inversions gives rise to the cost function shown in Eq. (2). The deviations between measurements, $\boldsymbol{y}$, and model-simulated mole fractions, $\mathbf{H}\boldsymbol{x}$, where $\mathbf{H}$ is a matrix that contains the sensitivities of atmospheric mole fractions to changes in emissions sources and $\boldsymbol{x}$ is a vector containing the emission sources, are weighted by uncertainty covariance, $\mathbf{R}$. Similarly, deviations between emissions and their a priori values, $\boldsymbol{x}_{\text{prior}}$, are weighted by uncertainty covariance, $\mathbf{P}$. This cost function is minimized with respect to $\boldsymbol{x}$ to find the "optimal" point that minimizes the total mismatch of the two terms (Enting, 2002; Tarantola, 2005).

$$J = (\boldsymbol{y} - \mathbf{H}\boldsymbol{x})^{\mathrm{T}}\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}) + (\boldsymbol{x} - \boldsymbol{x}_{\text{prior}})^{\mathrm{T}}\mathbf{P}^{-1}(\boldsymbol{x} - \boldsymbol{x}_{\text{prior}}) \quad (2)$$

This framework has generally been used because of its simplicity to solve. Several limitations are however present. Bayesian methods rely on knowledge of model-measurement ($\mathbf{R}$) and a priori emissions ($\mathbf{P}$) uncertainties, and the derived fluxes and associated uncertainties strongly depend on these parameters. Model-measurement uncertainties describe uncertainties in the instruments as well as uncertainties associated with the model's simulation of a measurement. The model error can be split into several components: structural errors within the CTM or meteorological model (Peylin et al., 2002; Thompson et al., 2011); model representation error, which describe errors in the representation of a point measurement in representing a grid volume (Chen and Prinn, 2006); aggregation errors, which result from averaging parameters over space and time and assuming fixed distributions within those domains (Kaminski et al., 2001; Thompson et al., 2011). Knowledge of these uncertainties is critical for robustly estimating posterior fluxes and their uncertainties; however, they are largely elicited through "expert judgment". The lack of objective methodology in ascertaining these uncertainties has been identified in many studies as a major limitation of traditional inverse methods (Rayner et al., 1999; Peylin et al., 2002; Law et al., 2002; Kaminski et al., 1999).

Some studies have used atmospheric data to "tune" covariances and then used these optimized parameters to derive fluxes (Michalak et al., 2005; Berchet et al., 2013). Several issues exist in these methods: (1) the uncertainties derived in the tuned $\mathbf{R}$ and $\mathbf{P}$ cannot be propagated through to estimating fluxes and their uncertainties, and (2) the Bayesian statistics used to derive fluxes assume that each term of the cost function in Eq. (2) is independent. Because both the a priori emissions and observations were used to derive the tuned $\mathbf{R}$ and $\mathbf{P}$, independence between the two terms can no longer be assumed and correlations between the two will exist that are not fully accounted for; (3) these methods typically use Gaussian PDFs in setting up the cost function.

We present a hierarchical Bayesian method to estimate trace gas emissions and additional parameters, which we call "hyper-parameters", that describe the a priori emissions PDF and the model-measurement uncertainty PDF. Compared to traditional methods, the a priori information in the system is extended to include a set of hyper-parameters equipped with their own prior distributions, which we call "hyper-priors". Throughout the text, we refer to the emissions PDF that is characterized by these hyper-parameters as the "a priori emissions PDF".

We first describe the development of a hierarchical framework. Beginning with Bayes' theorem in Eq. (1), we seek to estimate the joint distribution of two parameters, $\boldsymbol{x}$ and $\boldsymbol{\theta}$, using data $\boldsymbol{y}$ (Eq. 3).

$$\rho(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \rho(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\rho(\boldsymbol{x}, \boldsymbol{\theta}) \qquad (3)$$

A joint distribution can be decomposed using the "probability chain rule" (Eq. 4).

$$\rho(\boldsymbol{x}, \boldsymbol{\theta}) = \rho(\boldsymbol{x}|\boldsymbol{\theta})\rho(\boldsymbol{\theta}) \qquad (4)$$

Substitution of Eq. (4) into Eq. (3) leads to a hierarchical Bayesian model, in which both $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are informed by the data.

$$\rho(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \rho(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\rho(\boldsymbol{x}|\boldsymbol{\theta})\rho(\boldsymbol{\theta}) \qquad (5)$$

Hierarchical methods have been successfully applied in other fields (Riccio et al., 2006; Gelman, A. and Hill, J., 2002; Lehuger et al., 2009). A general summary and the application to uncertainty analysis can be found in Cressie et al. (2009).

In this application, in which we estimate hyper-parameters in addition to fluxes, a hierarchical approach allows us to explore the additional space of "uncertainties in uncertainties". The framework ensures that estimated parameters and their uncertainties and covariances are passed systematically through the inversion. The entire set of fluxes and hyper-parameters are updated in one step, therefore using measurements only once.

Because there often does not exist an analytical solution to maximize the posterior PDF represented in Eq. (5), Markov chain Monte Carlo (MCMC) is used (Tarantola, 2005). MCMC samples the PDFs of a set of parameters by constructing a Markov chain that represents the posterior PDF after a large number of steps. MCMC has the additional advantage that it may be used on a broad class of models, which need not be Gaussian. For example, positive fluxes can be constrained through the use of an a priori emissions PDF with support only on the positive real axis (Rigby et al., 2011).

We show that the hierarchical method results in a more complete uncertainty characterization than traditional inverse methods in which a priori emissions and model-measurement covariances are based largely on expert judgment. We present an application of this method for inversions of trace gas

emissions and explore the ways in which the method can be used to quantify uncertainties in these inversions. Finally, we utilize this method to estimate regional sulfur hexafluoride (SF_6) emissions using measurements from the Advanced Global Atmospheric Gases Experiment (AGAGE) network and the UK Met Office Numerical Atmospheric-dispersion Modelling Environment v3 (NAME) transport model (Jones et al., 2007; Ryall and Maryon, 1998).

2 Application of hierarchical Bayesian modeling to emissions estimation

2.1 Theoretical framework

We are interested in estimating fluxes of trace gases and their uncertainties using measurements of atmospheric mole fractions. Fluxes and hyper-parameters could vary in space and time and are shown in this framework as vectors that could be estimated with spatial and temporal resolution. We apply the hierarchical Bayesian model to use data, y, to estimate x, a vector of emissions and boundary conditions to the inversion domain, as well as a set of hyper-parameters that govern the a priori emissions and model-measurement uncertainty PDFs. The hyper-parameters include vectors μ_x and σ_x, which describe the log mean and log standard deviation of a lognormal a priori emissions PDF, the vector σ_y, which describes the standard deviation of a Gaussian model-measurement uncertainty PDF, and scalar τ, which is a model-measurement autocorrelation timescale. The model-measurement covariance matrix, $\mathbf{R}$, is formed with diagonal terms comprised by the squares of σ_y. Off-diagonal terms are computed through Eq. (6), where r_{ij} is the covariance between measurements i and j, r_{ii} and r_{jj} are the variances of each measurement, $\Delta t_{i,j}$ is the time between measurements and τ is the autocorrelation timescale.

$$r_{ij} = \sqrt{r_{ii}} \cdot \sqrt{r_{jj}} \cdot \exp\left(\frac{-\Delta t_{i,j}}{\tau}\right) \qquad (6)$$

The joint distribution of $x, \mu_x, \sigma_x, \sigma_y$ and τ is expressed through Eq. (7), following the framework developed above.

$$\rho(x, \mu_x, \sigma_x, \sigma_y, \tau | y) \propto \rho(y|x, \sigma_y, \tau) \cdot \rho(x|\mu_x, \sigma_x)$$
$$\cdot \rho(\mu_x) \cdot \rho(\sigma_x) \cdot \rho(\sigma_y) \cdot \rho(\tau) \qquad (7)$$

It is shown in Eq. (7) that each hyper-parameter requires a hyper-prior PDF to be specified. We have chosen to represent the PDF of each parameter by Eqs. (8)–(13). The lognormal distribution (LN) was used for emissions and PDF parameters because the distribution is skewed so that negative values are not defined but large values have small nonzero densities. An exponential PDF (EXP) was used for τ because the mode is zero, and in most inversions it is generally assumed that there is no model-measurement autocorrelation. Model-measurement uncertainties were assumed to be Gaussian (N) because we assume random errors in the instrument

and model to be symmetric. The functional forms used in this application of the hierarchical Bayesian framework represent only one of many possible applications and can be reformulated to represent different assumptions. Model parametric uncertainties have not been explicitly considered here but have the potential to lead to biases in the outcome of the inversion.

$$\rho(\mu_x) = \text{LN}(\mu_{x,\text{prior}}, \sigma_{\mu,x,\text{prior}}) \qquad (8)$$

$$\rho(\sigma_x) = \text{LN}(\sigma_{x,\text{prior}}, \sigma_{\sigma,x,\text{prior}}) \qquad (9)$$

$$\rho(\sigma_y) = \text{LN}(\sigma_{y,\text{prior}}, \sigma_{\sigma,y,\text{prior}}) \qquad (10)$$

$$\rho(\tau) = \text{EXP}\left(\frac{1}{\tau_{\text{prior}}}\right) \qquad (11)$$

$$\rho(x|\mu_x, \sigma_x) = \text{LN}(\mu_x, \sigma_x) \qquad (12)$$

$$\rho(y|x, \mathbf{R}) = \text{N}(\mathbf{H}x, \mathbf{R}) \qquad (13)$$

The posterior joint distribution, $\rho(x, \mu_x, \sigma_x, \sigma_y, \tau | y)$, is estimated using MCMC with a Metropolis–Hastings algorithm (Rigby et al., 2011; Tarantola, 2005). The Metropolis–Hastings algorithm generates states from a proposal distribution and selectively accepts transitions so that the stationary distribution of the resulting chain represents the posterior distribution. A "burn-in" period of 25 000 iterations was discarded to remove any memory of the initial state, followed by 25 000 iterations to form the posterior PDFs. Convergence can be assessed using metrics such as Geweke's diagnostic (Geweke, 1992). One of the main advantages of this algorithm is that the normalization factor implied in Eq. (7) does not need to be computed. The chain is constructed such that each parameter has "knowledge" of the state of the other parameters, and, therefore, uncertainties and correlations between parameters are built into the chain. Using this hierarchical approach, posterior emissions and associated uncertainties, which are of primary interest, fully account for uncertainties in hyper-parameters.

The set of parameters being estimated $(x, \mu_x, \sigma_x, \sigma_y, \tau)$ is denoted as X. The criterion for acceptance at iteration n for the proposed set of parameters, X', is

$$X_n = \begin{cases} X', & \text{with probability } a \\ X_{n-1}, & \text{with probability } 1-a \end{cases} \qquad (14)$$

$$a = \min\left[1, \frac{\rho(X'|y)}{\rho(X|y)}\right]. \qquad (15)$$

The sizes of the proposal distributions were adjusted so that the resulting acceptance ratio for each parameter was between 0.25 and 0.5 to achieve optimal mixing (Roberts et al., 1997).

The main computational differences between the hierarchical inversion and a non-hierarchical inversion performed using MCMC are as follows (in order of importance): (1) the hierarchical inversion solving for uncertainty parameters requires that the inverse and determinant of covariance matrices (e.g., $\mathbf{R}$) be computed in every iteration, while a traditional inversion with a fixed uncertainty structure would

require that this operation be performed only once; (2) the hierarchical inversion solves for a larger number of parameters, inducing a small additional computational cost. For our applications, the main computational burden was 1. Based on traditional computational methods, as the size of the observation and/or parameter space grows, the cost of computing the inverse and determinant will be approximately n^3. Therefore, for very large problems, alternative computational approaches may be necessary. Several methods exist to dramatically reduce the computational cost of inverting large (covariance) matrices, if required, for higher-dimensional applications (see for example, Sun et al., 2012). For an inverse problem similar to the one presented in this study, observational and/or parameter spaces of the order of a thousand elements can be readily solved with minimal code modification, on the order of minutes on a moderately powerful workstation. We anticipate that MCMC algorithms could be developed that could extend this to tens of thousands of elements, if required.

2.2 Pseudo-data experiment

We are interested in investigating the effect of the hierarchical method on posterior emissions uncertainties and use a pseudo-data simulation to demonstrate the concept. In this simulation, 1000 ensemble members were randomly generated from a known emissions distribution, $\rho(\boldsymbol{x}|\boldsymbol{\mu}_x^*, \boldsymbol{\sigma}_x^*)$, where $\boldsymbol{\mu}_x^*$ and $\boldsymbol{\sigma}_x^*$ were known and fixed.

Each emissions ensemble member was used to simulate mole fraction pseudo-data, which were then applied in both a hierarchical Bayesian inversion (HB) and a non-hierarchical Bayesian inversion (NHB) to infer emissions for cases in which the a priori emissions uncertainties were incorrectly specified. In this work we adopt the stance that uncertainty quantification is correct if, on average, a given realization, $\boldsymbol{x}^*$, sampled from $\rho(\boldsymbol{x}|\boldsymbol{\mu}^*, \boldsymbol{\sigma}^*)$, is consistent with the marginal distribution $\rho(\boldsymbol{x}|\boldsymbol{y})$. If uncertainty is correctly captured, then 5 % of the time $\boldsymbol{x}^*$ should lie in the 5th percentile of the posterior distribution, 10 % within the 10th percentile and so on. We can thus plot a quantile–quantile (Q–Q) plot to compare the theoretical with the empirical quantile. If the empirical and true quantiles lie on the 1 : 1 line, we can conclude that the inversion is correctly capturing system uncertainties. If the uncertainties used in the inversion are too tightly constrained, ensemble members will tend to consistently lie in the tails of the estimated posterior distribution so that the Q–Q plot resembles an inverted "S-curve" around the diagonal. If the uncertainties are too lax, then the posterior judgments are under-confident and the Q–Q plot follows an S-curve around the diagonal.

Gaussian distributions were chosen for this pseudo-experiment for their simplicity and symmetry, but these results can be extrapolated to any distribution. Pseudo-data were generated from each realization, $\boldsymbol{x}^*$, using the NAME model for 1 month at Mace Head, Ireland, and include ran-

dom Gaussian noise with standard deviation, $\boldsymbol{\sigma}_y^*$. For each inversion, with parameter and hyper-parameter PDFs shown through Eq. (16), we fixed the mean of the a priori emissions distribution to be $\boldsymbol{\mu}_x^*$ and tested the effect of making incorrect assumptions in the inversion about the a priori emissions uncertainty. Two cases were investigated: one in which the a priori emissions uncertainty ($\boldsymbol{\sigma}_{x,\mathrm{prior}}$) was one-half and one in which it was twice $\boldsymbol{\sigma}_x^*$, respectively. This is equivalent to an inversion where we incorrectly assumed the a priori emissions uncertainty to be smaller or larger than the "truth". In the NHB case, these values were fixed (i.e., there is no uncertainty in $\boldsymbol{\sigma}_{x,\mathrm{prior}}$); in the HB case, some flexibility was allowed for the inversion to adjust these values. The uncertainty in the uncertainty, $\boldsymbol{\sigma}_{\sigma,x}$, was assumed to be 100 % of $\boldsymbol{\sigma}_{x,\mathrm{prior}}$, and model-measurement uncertainty was specified exactly as $\boldsymbol{\sigma}_y^*$. To generate the estimated quantiles, we tracked the quantile of $\boldsymbol{x}^*$ in the posterior distribution of each inversion and determined how often each quantile was being sampled (a perfectly characterized system would result in uniform sampling of all quantiles, as explained above).

$$\begin{aligned} \rho(\boldsymbol{x}|\boldsymbol{\sigma}_x) &= N(\boldsymbol{\mu}_x^*, \boldsymbol{\sigma}_x) \\ \rho(\boldsymbol{\sigma}_x) &= N(\boldsymbol{\sigma}_{x,\mathrm{prior}}, \boldsymbol{\sigma}_{\sigma,x}) \end{aligned} \tag{16}$$

Figure 1 shows Q–Q plots for the HB and NHB cases. When the assumed a priori emissions uncertainty was too constrained and not allowed to adjust in the inversion, the result was a greater sampling of the tails of the distribution. When the hyper-parameter was included, the estimated distribution shifted towards the 1 : 1 line, indicating a better representation of the true distribution. A similar situation was observed when the assumed uncertainty was too large, reflecting a posterior distribution that sampled the middle of the distribution more frequently than expected. Inclusion of the hyper-parameter again resulted in a shift toward the 1 : 1 line and a better characterization of the true distribution. The results of this pseudo-data experiment illustrate the ways in which the hierarchical method can reduce the effect of errors in our assumptions about the uncertainties governing the system. Similar results were found in experiments testing the effect of incorrect assumptions about model-measurement uncertainty. An important feature of the HB framework is that the posterior emissions PDF is less sensitive to assumptions about the hyper-parameters governing the a priori emissions PDF than if direct assumptions were made about this PDF, as is the case in a NHB framework. This is because, in a HB framework, the parameters governing the a priori emissions PDF are themselves informed by the data.
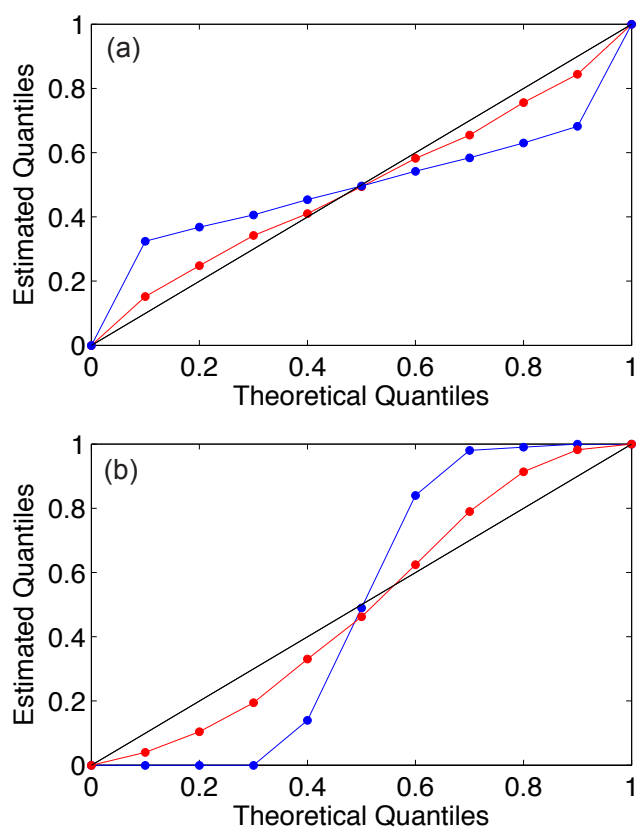
Fig. 1. Quantile–quantile (Q–Q) plots of a hierarchical Bayesian inversion (red) and a non-hierarchical Bayesian inversion (blue) in which a priori emissions uncertainties used in the inversion were **(a)** smaller than the true uncertainty (over-confident) and **(b)** larger than the true uncertainty (under-confident).

3 Case study: regional SF₆ emissions

3.1 Inversion setup

We use the above methodology to derive monthly regional SF$_6$ emissions and boundary conditions for the regions around four AGAGE stations, monthly model-measurement uncertainties, monthly means and standard deviations of the a priori emissions PDFs and monthly autocorrelation timescales. Model-measurement uncertainties were calculated for two time periods each month: daytime (approx. 06:00–18:00 local time of each station) and nighttime (approx. 18:00–06:00), in order to investigate the common assumption that uncertainties at night are larger than those during the day. We have chosen a monthly estimation timescale, but in principle any resolution can be used.

High-frequency measurements used in this study are dry air SF$_6$ mole fractions from the AGAGE stations at Mace Head, Ireland; Trinidad Head, USA; Cape Grim, Australia; and Gosan, South Korea, for the period of January–December 2012 (Prinn et al., 2000; Rigby et al., 2010). Gosan measurements from the summer period were excluded

due to the complexities induced by frequently shifting sampling of both northern and southern hemispheric background air. All measurements were made on the Medusa GC-MS system and were calibrated on the Scripps Institution of Oceanography (SIO) – 2005 scale (Miller et al., 2008). Measurements were three hourly averages, which is the temporal resolution of the model sensitivity output.

The UK Met Office's Lagrangian Particle Dispersion Model (LPDM), NAME, simulates atmospheric transport by following particles backwards in time from the measurement station. NAME has been used in previous studies for modeling trace gas transport at various sites (O'Doherty et al., 2004; Manning et al., 2011; Ganesan et al., 2013). NAME directly outputs the sensitivity of measurements to surface emissions by tracking the mass of particles and time spent in the lower 100 m of the model over the duration of the simulation. These model-derived sensitivities are referred to as "air histories". For each site modeled here, the Met Office's Unified Model (UM) was used at $0.352° \times 0.234°$ horizontal resolution and at 3 h temporal resolution. Particles were followed backwards in time for 20 days. Computational domains were chosen to be large enough to model the transport of pollutants from important source regions to each site as well as to allow for the assumption that boundary conditions to the LPDM domain do not vary significantly over the 20-day period. All air histories were generated with the release of particles from a 100 m column over the model surface. Grid cells were aggregated into approximately 20 regions over each domain following the methodology of Rigby et al. (2011). Regions were aggregated from grid cells as a function of a priori emissions and average sensitivity so that grid cells with high a priori emissions and/or high sensitivity would be estimated at higher resolution than those with low a priori emissions and/or low sensitivity.

Various methods have been used to determine boundary conditions to LPDM domains (Stohl et al., 2009; Manning et al., 2011; Rigby et al., 2011). In this application, boundary conditions were assumed to be constant over each month and were solved in the inversion as the part of the simulated mole fractions that were not accounted for by the 20-day air histories (i.e., emissions from farther back in time).

We follow the hierarchical system outlined by Eqs. (7)–(13). In this setup, $\boldsymbol{\mu}_{x,\text{prior}}$ was chosen to be the natural logarithm of 2008 values from the EDGAR v4.2 (henceforth referred to as EDGAR) database extrapolated to 2012 based on the linear trend for the inversion domain from 2004 to 2008 (JRC/PBL, 2011). These trends were assumed to be 2.8 % per year growth in Europe, 0.6 % per year growth in North America, 3.4 % per year growth in Oceania and 11 % per year growth in East Asia. Prior a priori emissions uncertainty, $\boldsymbol{\sigma}_{x,\text{prior}}$, was chosen as the value that resulted in 68 % of the lognormal PDF contained between 50 and 150 % of $\boldsymbol{\mu}_{x,\text{prior}}$ (similar to regional uncertainties used in Rigby et al., 2010). Prior model-measurement uncertainty, $\boldsymbol{\sigma}_{y,\text{prior}}$, was chosen to be the natural logarithm of the sum
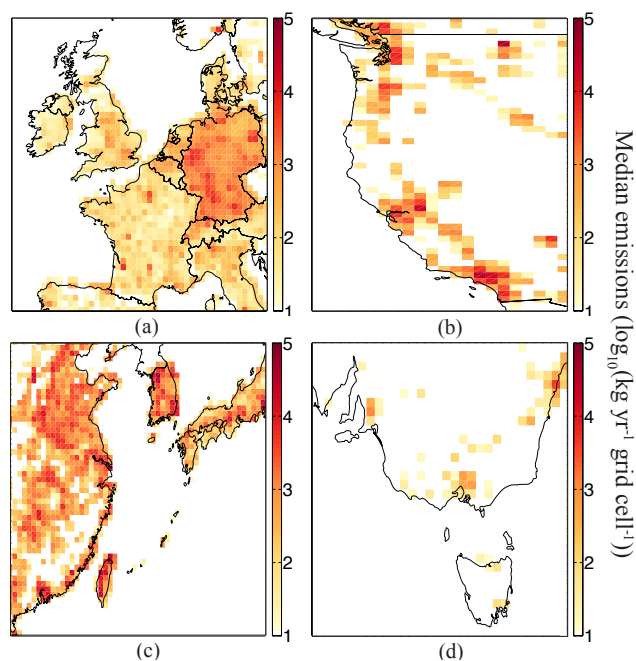
Fig. 2. Median 2012 SF$_6$ emissions for regions around **(a)** Mace Head, Ireland; **(b)** Trinidad Head, USA; **(c)** Gosan, South Korea; and **(d)** Cape Grim, Australia. Derived emissions were redistributed from the aggregated regions solved for in the inversion by assuming the distribution of EDGAR emissions.

Table 1. National SF$_6$ emissions for 2012 derived using the hierarchical Bayesian method for the regions surrounding four AGAGE stations. The fraction of the country for which emissions are derived (i.e., fraction of EDGAR emissions contained within the domains shown in Fig. 2) is tabulated along with 50th (median), 16th and 84th percentiles of the posterior emissions PDF.

Country	Country fraction (%)	Emissions (Mg yr^{-1})		
		Q_{50}	Q_{16}	Q_{84}
China	75	1576	1259	2107
Germany	100	348	264	455
South Korea	100	278	195	418
USA	21	269	193	407
Japan	96	242	154	416
France	100	61	40	90
UK	100	21	13	32
Australia	80	14	8	23

of the instrument uncertainty and the uncertainty associated with propagating the calibration scale, each assumed to be 0.05 pmol mol^{-1}. We did not include a prior estimate of the model representation error in the hierarchical inversions and allowed the inversion the flexibility to deduce this uncertainty. In the comparable inversions performed without the hierarchical method, we investigated the effect of including or not including a model representation error equal to the standard deviation of daily measurements. The mean a priori autocorrelation timescale, τ_{prior}, was assumed to be 7 days, which is an approximate timescale of synoptic-scale meteorological events. Prior baseline values were assumed to be the minimum measured value during the month with an uncertainty of 3 %.

Uncertainties in these hyper-parameters, $\sigma_{\mu,x,prior}$, $\sigma_{\sigma,x,prior}$ and $\sigma_{\sigma,y,prior}$, were calculated as the values that resulted in 68 % of the PDF contained between 50 and 150 % of $\mu_{x,prior}$, $\sigma_{x,prior}$ and $\sigma_{y,prior}$, respectively.

3.2   Results and discussion

We present median 2012 SF$_6$ emissions for the regions around four AGAGE stations, monthly boundary conditions, diurnal model-measurement uncertainty and autocorrelation timescales. Median posterior emissions are shown in Fig. 2 and are tabulated in Table 1. Deviations of these emissions from the prior are presented in Fig. 3. The domain for which

emissions are presented is smaller than the inversion domain but represents the region that the measurement station is most sensitive toward. Figure 4 presents national emissions derived from the HB inversion along with emissions derived from two NHB inversions that use the same PDFs and either include or exclude a model representation error. Uncertainties reflect the 16th to 84th percentiles of the posterior emissions PDFs to show consistency with previous studies citing 1σ uncertainties of Gaussian distributions. The results of the HB inversion show that (1) SF$_6$ emissions from the UK, France and Germany have deceased from the scaled EDGAR emissions and are also smaller than 2008 EDGAR emissions; (2) East Asian SF$_6$ emissions have decreased from scaled EDGAR values but have increased compared to 2008 EDGAR emissions; (3) emissions from the western coast of North America have largely decreased from 2008 EDGAR emissions; (4) Australian emissions have approximately remained the same as the scaled EDGAR emissions. In comparison to the 2007–2009 estimates made in Rigby et al. (2011), there are some statistically significant differences. In particular, the emissions from South Korea derived in this study are lower; however, emissions from Asia were shown through the sensitivity studies performed in Rigby et al. (2011) to be highly sensitive to inversion parameters, such as measurement averaging period.

To compare derived emissions and uncertainties between the HB and NHB methods, two NHB inversions were run for each site: one in which a model representation error was included and assumed to be the standard deviation of measurements each day and one in which no model representation error was used. Without including a model representation error, emissions become unrealistically large for East Asian countries due to the significantly elevated measurements that are not captured by the model and prior. Uncertainties on
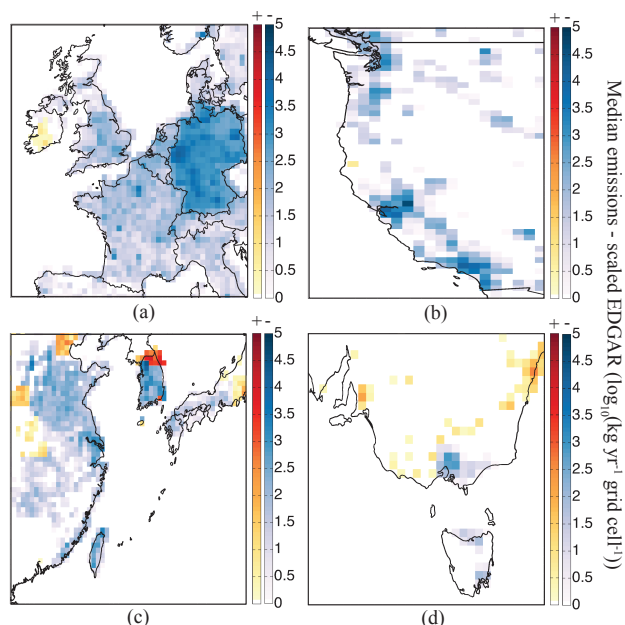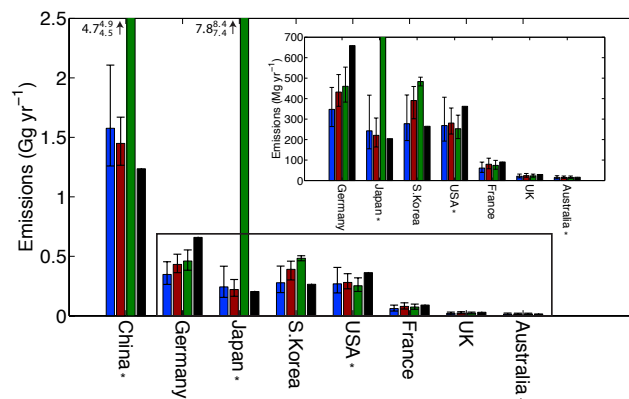
Fig. 3. Difference between derived 2012 SF$_6$ emissions and the scaled EDGAR emissions for regions around **(a)** Mace Head, Ireland; **(b)** Trinidad Head, USA; **(c)** Gosan, South Korea; and **(d)** Cape Grim, Australia. Positive differences are shown by the red logarithmic color map and negative differences by the blue logarithmic color map.

Fig. 4. National emissions derived using three methods: (1) hierarchical Bayesian (HB) inversion (blue); (2) non-hierarchical Bayesian inversion (NHB) with model-measurement uncertainties that include a model error (red); and (3) non-hierarchical Bayesian inversion in which no model error was included (green). A priori emissions are shown as black bars, and uncertainties reflect the 16th to 84th percentiles of the posterior national emissions. The asterisk (*) refers to countries in which emissions were derived for only a fraction of the country. For clarity, the inset shows a magnified view of countries with relatively smaller emissions.

these emissions are likely too small, owing to the underestimated model-measurement uncertainty. In previous studies, methods such as statistical filtering have been used to remove measurements that cannot be resolved by the model prior to the inversion to prevent unrealistic emissions from being derived (Manning et al., 2011). When a model representation error was included in the NHB inversion, emissions substantially decreased in East Asia and became more consistent with those derived in previous East Asian studies, stressing the importance of properly accounting for the model-measurement uncertainty (Vollmer et al., 2009; Kim et al., 2010; Li et al., 2011; Rigby et al., 2011; Fang et al., 2013).

In the HB inversion, emissions uncertainties are generally larger than those derived in the NHB inversions. These results suggest that the HB method is able to account for a larger space of uncertainties in the inversion and derive emissions uncertainties that are likely more representative of the true uncertainties in the system. To investigate this claim, we examine Chinese emissions and uncertainties derived in five previous studies between 2006 and 2012. These studies derived 0.8 (0.5–1.1) Gg yr^{-1} in Vollmer et al. (2009), 1.3 (0.9–1.7) Gg yr^{-1} in Kim et al. (2010) and 1.2 (0.9–1.7) Gg yr^{-1} in Li et al. (2011) for 2006–2008, 2.3 (2.1–2.5) in Rigby et al. (2011) for 2007–2009 and 2.8 (2.3–3.3) Gg yr^{-1} in Fang et al. (2013) for 2012. For the re-
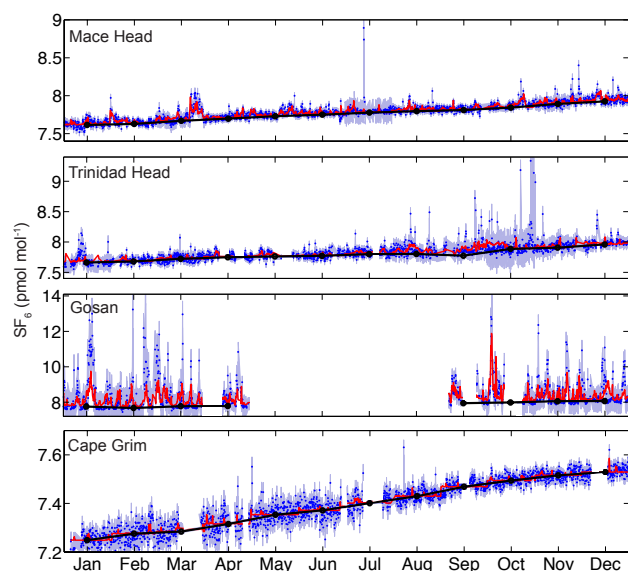
sults obtained for the period 2006–2009, these studies produced statistically different emissions, likely indicating that the uncertainty analyses are not robust. These five studies used various methods by which they assessed uncertainties in emissions, making it difficult to compare uncertainties between each study. Vollmer et al. (2009), Kim et al. (2010) and Fang et al. (2013) derived emissions uncertainties by examining an ensemble of inversions with varying a priori emissions and a priori emissions uncertainties. These sensitivity tests were used to quantify the effect of incorrect assumptions about a priori emissions and a priori emissions uncertainties, as our study aims to do; however each of these inversion ensemble members cannot be considered independent, and the resulting uncertainties may not be statistically robust or fully propagated through to emissions. Rigby et al. (2011) presented uncertainties derived solely from the Bayesian inversion, which is the likely the cause for the much smaller uncertainties than the other studies. The results of the HB inversion presented in this study show 2012 Chinese emissions to be 2.1 (1.7–2.8) Gg yr^{-1} (derived value scaled by country fraction), which is statistically consistent with the recent results of Fang et al. (2013). While our derived uncertainties are similar to some of these studies, we propose that the uncertainty quantification outlined in this work is more statistically justifiable, complete and traceable than those presented elsewhere.

Derived model-measurement uncertainties for all stations are shown in Fig. 5 and simulated mole fractions and boundary conditions in Fig. 6. Uncertainties at Mace Head are generally lower than the a priori value used, suggesting that the
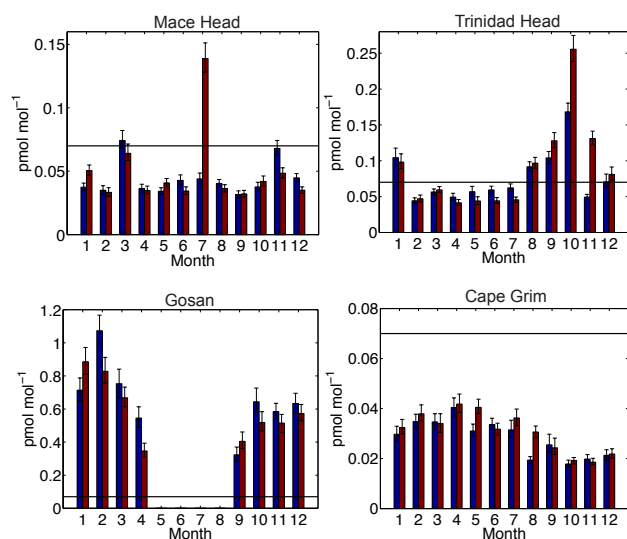
Fig. 5. Monthly 2012 SF$_6$ model-measurement uncertainties derived for daytime observations (blue) and nighttime observations (red) with error bars corresponding to 16th to 84th percentiles of the posterior distributions. The a priori value is shown by the black line.

Fig. 6. Simulated 2012 SF$_6$ mole fractions (red line) and observations (blue dot) at each AGAGE station. Note that each station is plotted with a different y axis range. Shading represents the 16th to 84th percentiles of the posterior model-measurement uncertainty distributions derived in the inversion. Black circles indicate the baseline values derived at each month.

model does well at representing this site. During July, a large pollution event is observed but is not captured by the model. The result is that measurements during this month have larger derived uncertainties and are therefore weighted less prominently in the inversion. Though a monthly timescale was used in this case study for deriving model-measurement uncertainties, solving for weekly or daily values would result in fewer observations being strongly de-weighted in the inversion. While some previous studies have filtered outlier measurements prior to the inversion through various methods, the HB method is less sensitive to these outliers. At Trinidad Head there is less consistency between the model and measurements than at Mace Head, leading to model-measurement uncertainties that are almost twice as large as those derived for Mace Head. These uncertainties are significantly higher in the latter part of the year (August–December), when more regional pollution is intercepted. At this time, the nighttime uncertainties are considerably larger than the daytime uncertainties, and one possible cause could be from errors in the model that have diurnal characteristics (e.g., sea breezes that are not captured). Uncertainties at Gosan are an order of magnitude larger than those derived at any other site. While the model captures the timing of many of the pollution events at this site, the sizes of the pollution events are considerably larger than predicted by the model and prior. This suggests that there could be emissions in close proximity to the station that are not captured by the model at the resolution used or that the model is under-representing the sensitivity to surface emissions. Uncertainties at Cape Grim are the smallest of all of the stations modeled here, owing to the measurement of

mostly baseline air and very small pollution events. Notably, the derived model-measurement uncertainties at Cape Grim decreased in the second half of the year when improved instrumentation was installed, which resulted in better instrumental precision. For most of the sites, we have found that there was not a clear advantage to using daytime observations only. Derived autocorrelation timescales (not shown) were between 6 h and 1 day for each month and station and are similar to those presented in other studies (Berchet et al., 2013).

4 Conclusions

We present an application of a hierarchical Bayesian method for trace gas inversions. We show that the inclusion of "hyper-parameters" to represent the a priori emissions PDF, model-measurement uncertainty and measurement autocorrelation timescale results in a more complete quantification of emissions uncertainties over traditional inverse methods that rely heavily on expert judgment.

Using the hierarchical method, we have estimated emissions of SF$_6$ for regions around four AGAGE stations and hyper-parameters for each site. The emissions uncertainties derived using the hierarchical method are generally larger than those derived in traditional inversions as they account for a broader space of uncertainties in the system, including random aggregation, representation and structural errors.

We show that model error is a significant contribution to model-measurement uncertainties at some sites, for example Gosan, without which, unrealistically large emissions would be derived with small uncertainties. The large discrepancy between the model and observations at this site results in the derivation of large model-measurement uncertainties and accordingly larger emissions uncertainties than those resulting from a standard Bayesian inversion. Similarly at Trinidad Head, derived uncertainties are larger than expected, owing to poor model fit, which results in larger emissions uncertainties than the comparable non-hierarchical inversion. In contrast, the generally good agreement between observations and model at Mace Head and Cape Grim results in a model-measurement uncertainty being derived that is smaller than the initial a priori value. Each of these findings is consistent with our expectations about the uncertainty characteristics of model performance at these sites but have been derived using minimal expert judgment.

References

Berchet, A., Pison, I., Chevallier, F., Bousquet, P., Conil, S., Geever, M., Laurila, T., Lavric, J., Lopez, M., Moncrieff, J., Necki, J., Ramonet, M., Schmidt, M., Steinbacher, M., and Tarniewicz, J.: Towards better error statistics for atmospheric inversions of methane surface fluxes, Atmos. Chem. Phys., 13, 7115–7132, doi:10.5194/acp-13-7115-2013, 2013.

Bousquet, P., Ringeval, B., Pison, I., Dlugokencky, E. J., Brunke, E.-G., Carouge, C., Chevallier, F., Fortems-Cheiney, A., Frankenberg, C., Hauglustaine, D. A., Krummel, P. B., Langenfelds, R. L., Ramonet, M., Schmidt, M., Steele, L. P., Szopa, S., Yver, C., Viovy, N., and Ciais, P.: Source attribution of the changes in atmospheric methane for 2006–2008, Atmos. Chem. Phys., 11, 3689–3700, doi:10.5194/acp-11-3689-2011, 2011.

Chen, Y.-H. and Prinn, R. G.: Estimation of atmospheric methane emissions between 1996 and 2001 using a three-dimensional global chemical transport model, J. Geophys. Res., 111, D10307, doi:10.1029/2005JD006058, 2006.

Cressie, N., Calder, C. A., Clark, J. S., Ver Hoef, J. M., and Wikle, C. K.: Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling, Ecol. Appl., 19, 553–570, doi:10.1890/07-0744.1, 2009.

Enting, I.: Inverse problems in atmospheric constituent transport, Cambridge University Press, New York, 2002.

Enting, I. G., Trudinger, C. M., and Francey, R. J.: A synthesis inversion of the concentration and δ13 C of atmospheric CO_2, Tellus B, 47, 35–52, doi:10.1034/j.1600-0889.47.issue1.5.x, 1995.

Fang, X., Thompson, R. L., Saito, T., Yokouchi, Y., Kim, J., Li, S., Kim, K. R., Park, S., Graziosi, F., and Stohl, A.: Sulfur hexafluoride (SF6) emissions in East Asia determined by inverse modeling, Atmos. Chem. Phys. Discuss., 13, 21003–21040, doi:10.5194/acpd-13-21003-2013, 2013.

Ganesan, A. L., Chatterjee, A., Prinn, R. G., Harth, C. M., Salameh, P. K., Manning, A. J., Hall, B. D., Mühle, J., Meredith, L. K., Weiss, R. F., O'Doherty, S., and Young, D.: The variability of methane, nitrous oxide and sulfur hexafluoride in Northeast India, Atmos. Chem. Phys., 13, 10633–10644, doi:10.5194/acp-13-10633-2013, 2013.

Gelman, A. and Hill, J.: Data analysis using regression and multilevel/hierarchical models, Cambridge University Press, New York, 2002.

Geweke, J.: Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments,, in: Bayesian Statistics 4, edited by: Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., Oxford University Press, Oxford, United Kingdom, 1992.

Jones, A., Thomson, D. J., Hort, M. C., and Devenish, B.: The UK Met Office's next-generation atmospheric dispersion model, NAME III, in Air Pollution Modeling and Its Application XVII, edited by: Borrego, C. and Norman, A.-L., Springer, New York, USA., 580–589, 2007.

JRC/PBL: Joint Research Centre of the European Commission (JRC)/Netherlands Environmental Assessment Agency (PBL), Emission Database for Global Atmospheric Research (EDGAR), release version 4.2, 2011.

Kaminski, T., Heimann, M., and Giering, R.: A coarse grid three-dimensional global inverse model of the atmospheric transport: 2. Inversion of the transport of CO_2 in the 1980s, J. Geophys. Res., 104, 18555–18581, doi:10.1029/1999JD900146, 1999.

Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G.: On aggregation errors in atmospheric transport inversions, J. Geophys. Res., 106, 4703–4715, doi:10.1029/2000JD900581, 2001.

Kim, J., Li, S., Kim, K.-R., Stohl, A., Mühle, J., Kim, S.-K., Park, M.-K., Kang, D.-J., Lee, G., Harth, C. M., Salameh, P. K., and Weiss, R. F.: Regional atmospheric emissions determined from measurements at Jeju Island, Korea: Halogenated compounds from China, Geophys. Res. Lett., 37, 2007–2011, doi:10.1029/2010GL043263, 2010.

Law, R. M., Rayner, P. J., Steele, L. P., and Enting, I. G.: Using high temporal frequency data for CO_2 inversions, Global Biogeochem. Cy., 16, 1053, doi:10.1029/2001GB001593, 2002.

Lehuger, S., Gabrielle, B., Van Oijen, M., Makowski, D., Germon, J.-C., Morvan, T., and Hénault, C.: Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model, Agr. Ecosyst. Environ., 133, 208–222, doi:10.1016/j.agee.2009.04.022, 2009.

Li, S., Kim, J., Kim, K.-R., Mühle, J., Kim, S.-K., Park, M.-K., Stohl, A., Kang, D.-J., Arnold, T., Harth, C. M., Salameh, P. K., and Weiss, R. F.: Emissions of halogenated compounds in East Asia determined from measurements at Jeju Island, Korea., Environ. Sci. Technol., 45, 5668–5675, doi:10.1021/es104124k, 2011.

Manning, A. J., O'Doherty, S., Jones, A. R., Simmonds, P. G., and Derwent, R. G.: Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach, J. Geophys. Res., 116, D02305, doi:10.1029/2010JD014763, 2011.

Michalak, A. M., Hirsch, A., Bruhwiler, L., Gurney, K. R., Peters, W., and Tans, P. P.: Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, J. Geophys. Res., 110, D24107, doi:10.1029/2005JD005970, 2005.

Miller, B. R., Weiss, R. F., Salameh, P. K., Tanhua, T., Greally, B. R., Mühle, J., and Simmonds, P. G.: Medusa: a sample preconcentration and GC/MS detector system for in situ measurements of atmospheric trace halocarbons, hydrocarbons, and sulfur compounds., Anal. Chem., 80, 1536–1545, doi:10.1021/ac702084k, 2008.

Mühle, J., Ganesan, A. L., Miller, B. R., Salameh, P. K., Harth, C. M., Greally, B. R., Rigby, M., Porter, L. W., Steele, L. P., Trudinger, C. M., Krummel, P. B., O'Doherty, S., Fraser, P. J., Simmonds, P. G., Prinn, R. G., and Weiss, R. F.: Perfluorocarbons in the global atmosphere: tetrafluoromethane, hexafluoroethane, and octafluoropropane, Atmos. Chem. Phys., 10, 5145–5164, doi:10.5194/acp-10-5145-2010, 2010.

O'Doherty, S., Cunnold, D. M., Manning, A., Miller, B. R., Wang, R. H. J., Krummel, P. B., Fraser, P. J., Simmonds, P. G., McCulloch, A., Weiss, R. F., Salameh, P., Porter, L. W., Prinn, R. G., Huang, J., Sturrock, G., Ryall, D., Derwent, R. G., and Montzka, S. A.: Rapid growth of hydrofluorocarbon 134a and hydrochlorofluorocarbons 141b, 142b, and 22 from Advanced Global Atmospheric Gases Experiment (AGAGE) observations at Cape Grim, Tasmania, and Mace Head, Ireland, J. Geophys. Res., 109, D06310, doi:10.1029/2003JD004277, 2004.

Peylin, P., Baker, D., Sarmiento, J., Ciais, P., and Bousquet, P.: Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO_2 data, J. Geophys. Res., 107, 4385, doi:10.1029/2001JD000857, 2002.

Prinn, R. G., Weiss, R. F., Fraser, P. J., Simmonds, P. G., Cunnold, D. M., Alyea, F. N., O'Doherty, S., Salameh, P., Miller, B. R., Huang, J., Wang, R. H. J., Hartley, D. E., Harth, C., Steele, L. P., Sturrock, G., Midgley, P. M., and McCulloch, A.: A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE, J. Geophys. Res., 105, 17751–17792, doi:10.1029/2000JD900141, 2000.

Rayner, P. J., Enting, I. G., Francey, R. J., and Langenfelds, R.: Reconstructing the recent carbon cycle from atmospheric CO_2, $\delta 13C$ and O2/N2 observations, Tellus B, 51, 213–232, doi:10.1034/j.1600-0889.1999.t01-1-00008.x, 1999.

Riccio, A., Barone, G., Chianese, E., and Giunta, G.: A hierarchical Bayesian approach to the spatio-temporal modeling of air quality data, Atmos. Environ., 40, 554–566, doi:10.1016/j.atmosenv.2005.09.070, 2006.

Rigby, M., Mühle, J., Miller, B. R., Prinn, R. G., Krummel, P. B., Steele, L. P., Fraser, P. J., Salameh, P. K., Harth, C. M., Weiss, R. F., Greally, B. R., O'Doherty, S., Simmonds, P. G., Vollmer, M. K., Reimann, S., Kim, J., Kim, K.-R., Wang, H. J., Olivier, J. G. J., Dlugokencky, E. J., Dutton, G. S., Hall, B. D., and Elkins, J. W.: History of atmospheric SF6 from 1973 to 2008, Atmos. Chem. Phys., 10, 10305–10320, doi:10.5194/acp-10-10305-2010, 2010.

Rigby, M., Manning, A. J., and Prinn, R. G.: Inversion of long-lived trace gas emissions using combined Eulerian and Lagrangian chemical transport models, Atmos. Chem. Phys., 11, 9887–9898, doi:10.5194/acp-11-9887-2011, 2011.

Roberts, G. O., Gelman, A., and Gilks, W. R.: Weak convergence and optimal scaling of random walk Metropolis algorithms, Ann. Appl. Probab., 7, 110–120, doi:10.1214/aoap/1034625254, 1997.

Ryall, D. B. and Maryon, R. H.: Validation of the UK Met. Office's name model against the ETEX dataset, Atmos. Environ., 32, 4265–4276, doi:10.1016/S1352-2310(98)00177-0, 1998.

Stohl, A., Seibert, P., Arduini, J., Eckhardt, S., Fraser, P., Greally, B. R., Lunder, C., Maione, M., Mühle, J., O'Doherty, S., Prinn, R. G., Reimann, S., Saito, T., Schmidbauer, N., Simmonds, P. G., Vollmer, M. K., Weiss, R. F., and Yokouchi, Y.: An analytical inversion method for determining regional and global emissions of greenhouse gases: Sensitivity studies and application to halocarbons, Atmos. Chem. Phys., 9, 1597–1620, doi:10.5194/acp-9-1597-2009, 2009.

Sun, Y., Li, B., and Genton, M.: Geostatistics for Large Datasets, in Advances and Challenges in Space-time Modelling of Natural Events, Lecture Notes in Statistics 207, edited by: Porcu, E., Montero, J.-M., and Schlather, M., doi:10.1007/978-3-642-17086-7_3, 2012.

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 2005.

Thompson, R. L., Gerbig, C., and Rödenbeck, C.: A Bayesian inversion estimate of N_2O emissions for western and central Europe and the assessment of aggregation errors, Atmos. Chem. Phys., 11, 3443–3458, doi:10.5194/acp-11-3443-2011, 2011.

Vollmer, M. K., Zhou, L. X., Greally, B. R., Henne, S., Yao, B., Reimann, S., Stordal, F., Cunnold, D. M., Zhang, X. C., Maione, M., Zhang, F., Huang, J., and Simmonds, P. G.: Emissions of ozone-depleting halocarbons from China, Geophys. Res. Lett.,, 36, L15823 doi:10.1029/2009GL038659, 2009.