Atmos. Chem. Phys. Discuss., 12, 33219–33263, 2012
www.atmos-chem-phys-discuss.net/12/33219/2012/
doi:10.5194/acpd-12-33219-2012
© Author(s) 2012. CC Attribution 3.0 License.

Atmospheric Chemistry and Physics Discussions

This discussion paper is/has been under review for the journal Atmospheric Chemistry and Physics (ACP). Please refer to the corresponding final paper in ACP if available.

Estimation of volatile organic compound emissions for Europe using data assimilation

M. R. Koohkan[1,2], M. Bocquet[1,2], Y. Roustan[1], Y. Kim[1], and C. Seigneur[1]

[1]Université Paris-Est, CEREA, joint laboratory École des Ponts ParisTech and EDF R&D, Champs-sur-Marne, France
[2]INRIA, Paris Rocquencourt research centre, France

Received: 6 October 2012 – Accepted: 28 November 2012 – Published: 20 December 2012

Correspondence to: M. Bocquet (bocquet@cerea.enpc.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The emission of volatile organic compounds (VOCs) over western Europe for the year 2005 are estimated via inverse modelling, by assimilation of in situ observations of concentration and compared to a standard emission inventory. The study focuses on fifteen VOC species: five aromatics, six alkanes, two alkenes, one alkyne and one biogenic diene. The inversion relies on a validated fast adjoint of the chemistry transport model used to simulate the fate and transport of these VOCs. The assimilated ground-based measurements over Europe are provided by the European Monitoring and Evaluation Programme (EMEP) network. The background emissions errors and the prior observational errors are estimated by maximum likelihood approaches. The positivity assumptions on the VOC emission fluxes is pivotal for a successful inversion and this maximum likelihood approach consistently accounts for the positivity of the fluxes. For most species, the retrieval leads to a significant reduction of the bias, which underlines the misfit between the standard inventories and the observed concentrations. The results are validated through a forecast test and a cross-validation test. It is shown that the statistically consistent non-Gaussian approach based on a reliable estimation of the errors offers the best performance. The efficiency in correcting the inventory depends on the lifetime of the VOCs. In particular, it is shown that the use of in-situ observations using a sparse monitoring network to estimate emissions of isoprene is inadequate because its short chemical lifetime significantly limits the spatial radius of influence of the monitoring data. For species with longer lifetime (a few days), successful, albeit partial, emission corrections can reach regions hundreds of kilometres away from the stations. Domainwide corrections of the emissions inventories of some VOCs are significant, with underestimations on order of a factor of two of propane, ethane, ethylene and acetylene.

cc BY

1 Introduction

Volatile organic compounds (VOCs) are of particular environmental concern because they are precursors of secondary pollutants, such as ozone and fine particulate matter ($PM_{2.5}$), and some VOCs are pollutants in their own right due to their adverse carcinogenic and/or non-carcinogenic health effects. Therefore, it is essential to have accurate emission inventories of VOCs to conduct air quality modelling studies for the development of optimal emission control strategies and for air quality forecasting as well as to follow their temporal emission trends over the years as emission control strategies get implemented. The large number of emission sources, both anthropogenic and biogenic, and processes leading to those emissions (combustion, evaporation, vegetation metabolism) make the development of accurate VOC emission inventories difficult. Furthermore, VOC emissions cannot be derived from mass balances and they must be obtained from experimental measurements conducted at the source of the emissions. Also, emission testing is costly and some emission factors are developed for total VOCs rather than for individual VOCs. Therefore, a chemical speciation must be applied to total VOC emission factors using speciation data that are limited and uncertain. Although uncertainties in anthropogenic emissions have been reduced over the years as a result of better characterisation of major emission sources, uncertainties still remain. Furthermore, large uncertainties are associated with biogenic emissions due to the difficulty of estimating the meteorology-dependent emission rates for a large number of vegetation species as well as characterising the land-use/land-cover of the area of interest. Therefore, several approaches have been used to evaluate the accuracy of VOC emission inventories and, if appropriate, apply some correction.

Uncertainties in emission inventories have been estimated, for example, by comparing ambient air measurements in tunnels with vehicle exhaust emission estimates (e.g., Staehelin et al., 1998; Sawyer et al., 2000; Touaty and Bonsang, 2000; Stemmler et al., 2005; Ho et al., 2007). However, such experiments characterise only one source category (on-road traffic) and focus on a single location and time period.

Satellite measurements have been used to assess VOC emissions, but such techniques are limited by the number of VOCs that can be measured via satellite-borne instruments (Vijayaraghavan et al., 2008) and to areas that are specific to a major source category, e.g., the use of formaldehyde (an oxidation product of isoprene) to estimate isoprene emissions in remote areas where biogenic emissions dominate (e.g., Shim et al., 2005; Fu et al., 2007; Millet et al., 2008; Dufour et al., 2009). Measurements of VOC concentrations aloft have also been used to estimate fluxes of VOCs originating from an area (e.g., Hopkins et al., 2009); however, there are uncertainties associated with the mass balance method used to estimate the atmospheric transport flux and relate it to an emission flux. Furthermore, such an approach is limited to the estimation of an emission flux for a given region over a given period. Comparisons of the output concentrations of air quality model simulations with observations aloft (e.g., Xiao et al., 2008) or at ground level (e.g., Harley and Cass, 1995) provide some estimates of uncertainties in VOC emissions; however, such information is typically also limited to a specific region and period for which those measurements are available. Inverse modelling has been conducted using ground-level ambient concentrations to estimate emission inventories for some air pollutants, but such studies (e.g., Quélo et al., 2005; Elbern et al., 2007; Koohkan and Bocquet, 2012) have focused so far on regulated air pollutants with ambient concentration data available from routine monitoring networks and have not yet included VOCs.

It is, therefore, of great interest to investigate the current status of VOC emissions using an approach that provides information for several major VOCs with spatial distribution over a large domain and for a long period of time. To that end, we present here the first assessment of the emissions of several VOCs measured routinely at several remote sites, that covers all VOC sources over western and central Europe for an entire year.

In Sect. 2, the chemical transport model (CTM) used for modelling the VOCs is introduced and its reduced counterpart is described. The source-receptor relationship is built using an approximate adjoint model which is validated. The control variables (i.e., the emission fluxes), and the inversion modelling method are described. The method to estimate the so-called hyperparameters that parametrise the prior error statistics is introduced. In Sect. 3, the setup of the numerical experiments is described. Details about the observations set and the first guess inventory are provided. The optimal values of the hyperparameters to be used in the inversions are computed. In Sect. 4, the results of the inversions are presented and discussed. A forecast test and a cross-validation test are provided to validate the corrected emission fluxes using independent observations. Conclusions are presented in Sect. 5.

2 Methodology

2.1 Full chemical transport model and reduced VOC model

The chemical transport model (CTM) POLAIR3D (Sartelet et al., 2007) of the POLYPHEMUS air quality modelling system (Mallet et al., 2007) is chosen to model the atmospheric concentrations of chemical species. The numerical discretisation of the model for chemistry and transport, based on a first order time splitting algorithm, can be summarised as follows:

$$\boldsymbol{c}^{k+1} = \mathcal{X}_k\left(\boldsymbol{M}_k(\boldsymbol{c}^k)\right) + \boldsymbol{\Delta} t\, \boldsymbol{e}_{k+1} \qquad (1)$$

where,

- $\boldsymbol{c}^k$ is the field of the concentrations of all simulated species at time step k.
- $\boldsymbol{M}_k$ is the linear advection-diffusion operator. It also includes the deposition processes.
- $\mathcal{X}_k$ is the chemical reaction operator
- $\boldsymbol{e}_k$ is the emission field at time step k.

Table 1 lists the fifteen VOC species for which experimental measurements are available from the European Monitoring and Evaluation Programme (EMEP) database. In order to simulate the concentrations of these species, the RACM 2 (Regional Atmospheric Chemistry Mechanism, version 2) chemical kinetic mechanism (Goliff and Stockwell, 2008) is used within the CTM (Kim et al., 2009). The chemical reactions considered for these species and their typical lifetime are presented in Table 1. After undergoing oxidation reactions, these fifteen primary species result in secondary species. The latter are not presented in Table 1 because they are not relevant to our study: they are not measured in the EMEP network and, therefore, cannot be assimilated; they are, however, included in RACM 2 either explicitly or via surrogate species. Some of the primary VOC species (isoprene, acetylene, ethane, ethylene, benzene) are treated explicitly in RACM 2. The others are represented through a lumped molecule approach and, therefore, need a specific treatment to be followed separately. They are added as explicit species with their own oxidation reactions written in a way that does not affect RACM 2.

This chemical mechanism involves more than three hundred reactions that result in nonlinear interactions among the chemical species. As a result, computational burden of inverse modelling studies is very large. To address this issue, we developed a reduced chemical mechanism, denoted $\boldsymbol{X}_k$, which uses the concentration fields of the oxidants, hydroxyl radicals (OH), ozone (O_3) and nitrate radicals (NO_3), provided as external data. The oxidant concentration fields are pre-computed with RACM 2 and used later in the reduced mechanism. This approximation makes sense if $\delta\mathcal{X}_k(c^k) = \mathcal{X}_k(c^k) - \boldsymbol{X}_k(c^k)$ is small with respect to $\boldsymbol{X}_k(c^k)$. The validation of this approximations is checked a posteriori in Sect. 4.1.1. When replacing $\mathcal{X}_k$ by $\boldsymbol{X}_k$, Eq. (1) becomes linear with respect to the emission fields and the computational cost of inverse modelling becomes manageable.

2.2 The source receptor model

The source-receptor model provides the relationship between the emissions and the observations. For species s, this can be written as follows:

$$\boldsymbol{\mu}^s = \mathcal{H}^s \boldsymbol{e}^s + \boldsymbol{\lambda}^s + \boldsymbol{\epsilon}^s \tag{2}$$

where $\boldsymbol{\mu}^s \in \mathbb{R}^{d_s}$ represents the vector of the observations (d_s is the number of observations for species s). $\mathcal{H}^s$ is the Jacobian operator with respect to $\boldsymbol{e}^s$ and $\boldsymbol{e}^s = (\boldsymbol{e}_0^s, \boldsymbol{e}_1^s, \ldots, \boldsymbol{e}_k^s, \ldots \boldsymbol{e}_{N_t}^s) \in \mathbb{R}^E$ defines the hourly and spatially discretised emission vector, where $E = N_t \times N_\mathrm{x} \times N_\mathrm{y} \times N_z$. N_t is the total number of time steps, and N_x, N_y, N_z are the total number of elements (grid cells) in the x, y and z directions. $\boldsymbol{\lambda}^s \in \mathbb{R}^{d_s}$ is the vector of the concentrations induced by the initial and the boundary conditions for species s. μ^{s,i_k} is the observation of species s at time t_k at station i, $\boldsymbol{\lambda}_{i_k}^s$ is the concentration at the same time and location, computed with the full CTM, i.e., Eq. (1), with $\boldsymbol{e}^s = \mathbf{0}$. The vector $\boldsymbol{\epsilon}^s$ represents the errors: representativeness error, model error and instrumental error of the observations.

The Jacobian operator $\mathcal{H}^s$ can be built following two different methods. The first method consists in using the CTM. Let us assume that $\boldsymbol{e}^s = \delta_{l,h,k'}$ is the unity source at the surface coordinate $l \in [1; N_\mathrm{x} \times N_\mathrm{y}]$, altitude $h \in [1; N_z]$ and time $k' \in [1; N_t]$, and equals zero anywhere else. The CTM simulated concentration for clean air boundary and initial conditions at time t_k and station i with this source term is stored in $\mathcal{H}_{l,h}^{s,i_{k'},k}$. In order to compute the $\mathcal{H}^s$ operator with this method, the CTM model needs to be run E times, which is computationally intensive. That is why this method is usually restricted to point-wise emission sources.

The second method consists in using the adjoint model of the CTM (e.g., Roustan and Bocquet, 2006). For a monitoring site i at time t_k, using the linearity of the CTM for the VOCs, the adjoint solution can be written as follows:

$$\pi_i^k = \boldsymbol{M}_k^\dagger \left(\boldsymbol{X}_k^\dagger (\boldsymbol{\phi}_i^{k+1}) \right) + \boldsymbol{\Delta} t \pi_k^i \tag{3}$$

where, $\pi^i_k = \delta_{i,k}$ is the *sampling function* that represents the concentration measurement at station i and time t_k, $\boldsymbol{X}^{\dagger}_k$ is the adjoint of $\boldsymbol{X}_k$, and $\boldsymbol{M}^{\dagger}_k$ is the adjoint of $\boldsymbol{M}_k$. At the final time N_t, $\boldsymbol{\phi}^{N_t}_i$ is chosen to be $\mathbf{0}$. The adjoint model is also computed for clean air boundary conditions. Then, the Jacobian matrix is given by $\mathcal{H}^{s,i_{k'},k}_{l,h}$. The adjoint model is run d times ($d \leq \sum_s d_s$). This method is of great interest to reduce the computational time when the problem is ill-posed ($d \ll E$). It is appropriate for estimating the emissions originating from spatially distributed sources.

The Jacobian matrix of the present study is computed row by row, that is using adjoint solutions.

2.3 Control space

In order to reduce the dimension of the control space, that is the space of the fluxes to be estimated via inverse modelling, we introduce a relation between the effective control variables $\boldsymbol{\alpha}^s$ and the emission $\boldsymbol{e}^s$:

$$[\boldsymbol{e}^s_k]_{l,h} = [\boldsymbol{\alpha}^s]_l [\boldsymbol{e}^{s,b}_k]_{l,h}\,. \tag{4}$$

In this equation, $\boldsymbol{e}^{s,b}$ is the a priori (first guess or background) vector of emission for species s. Obviously the first guess value of the scaling factors $\boldsymbol{\alpha}^s$, is $\boldsymbol{\alpha}^s_b = \mathbf{1}$, where $\mathbf{1} = (1,\ldots,1)^{\mathrm{T}}$. Indices k, l, h are respectively related to the time sequence, the horizontal space grid, and the vertical grid. The $\boldsymbol{\alpha}^s$ factors, rather than the full 3-D time-dependent emission fields of Eq. (1), will be optimised. This choice of control parameters, which is only a function of l, implies that the correction of the emission fluxes is spatially distributed in the horizontal directions but that the vertical and the temporal distribution of the emission fluxes are not modified by the data assimilation analysis.

2.4 Inversion method

Combining Eq. (2) with Eq. (4), one obtains

$$\boldsymbol{\mu}^s = \mathbf{H}^s\boldsymbol{\alpha}^s + \boldsymbol{\lambda}^s + \boldsymbol{\epsilon}^s \tag{5}$$

where $\mathbf{H}^s$ is the Jacobian matrix that relates $\boldsymbol{\mu}^s$ to $\boldsymbol{\alpha}^s$:

$$\mathbf{H}^s_{i_k,l} = \sum_k [\boldsymbol{\phi}^k_{s,i}]_l [\boldsymbol{e}^{s,b}_k]_l \,. \tag{6}$$

In order to optimise the $\boldsymbol{\alpha}$ parameters, the following objective function with a regularisation term is used:

$$\begin{aligned}\mathcal{L}_s(\boldsymbol{\alpha}^s) = {} & \frac{1}{2}\left(\boldsymbol{\mu}^s - \mathbf{H}^s\boldsymbol{\alpha}^s - \boldsymbol{\lambda}^s\right)^{\mathrm{T}}\mathbf{R}_s^{-1}\left(\boldsymbol{\mu}_s - \mathbf{H}^s\boldsymbol{\alpha}^s - \boldsymbol{\lambda}^s\right)\\ & + \frac{1}{2}\left(\boldsymbol{\alpha}^s - \mathbf{1}\right)^{\mathrm{T}}\mathbf{B}_s^{-1}\left(\boldsymbol{\alpha}^s - \mathbf{1}\right).\end{aligned} \tag{7}$$

The vector $\boldsymbol{\alpha}^s_b = \mathbf{1} = (1,\ldots,1)^{\mathrm{T}}$ is the first guess of $\boldsymbol{\alpha}^s$. $\mathbf{R}_s$ is the observation error covariance matrix. $\mathbf{B}_s$ is the background error covariance matrix. For each species, these two matrices are both chosen to be diagonal with uniform variances, that is, $\mathbf{B}_s = m_s^2\mathbf{I}_{N_x \times N_y}$, $\mathbf{R}_s = r_s^2\mathbf{I}_{d_s}$. These statistical assumptions imply that we neglect any spatial and temporal correlations between grid cells in the errors. The anthropogenic emissions of VOCs are not expected to induce long-range correlation in the errors. However, potential important reasons for this hypothesis to fail are when the biogenic VOC emissions have correlated errors due for some VOCs common sources and similar emission model formulations, or when transport model errors induce temporal correlation in the error covariance matrix. That is why our diagonal assumption is an approximation.

In the following, two solutions of Eq. (7) are considered and compared. The first one assumes that the errors are Gaussian-distributed and the analysed parameters $\boldsymbol{\alpha}^s_a$ are given by the Best Linear Unbiased Estimator (BLUE):

$$\boldsymbol{\alpha}^s_a = \mathbf{1} + \mathbf{K}^s\left(\boldsymbol{\mu}^s - \mathbf{H}^s\mathbf{1} - \boldsymbol{\lambda}^s\right) \tag{8}$$

where $\mathbf{K}^s$ is the gain matrix:

$$\mathbf{K}^s = \mathbf{B}_s \mathbf{H}^{s\mathrm{T}} \left(\mathbf{R}_s + \mathbf{H}^s \mathbf{B}_s \mathbf{H}^{s\mathrm{T}} \right)^{-1} \tag{9}$$

The second solution of Eq. (7) is obtained assuming a truncated Gaussian distribution for the background error statistics, so that $\boldsymbol{\alpha}$ is optimised under a positivity constraint of each one of its entry $[\boldsymbol{\alpha}^s]_l$. As opposed to the Gaussian case, the retrieved scaling factors $[\boldsymbol{\alpha}^s]_l$ cannot be negative.

2.5 Estimation of hyperparameters

The parameters of the prior statistics, such as r_s and m_s, usually coined *hyperparameters*, often need to be estimated because their first guess is usually inaccurate, while the dependence of the retrieval on the hyperparameters can be dramatic (Davoine and Bocquet, 2007).

The estimation method for the hyperparameters depends on the statistical assumptions underlying Eq. (7). In the first case, the error $\boldsymbol{\epsilon}^s$ (in Eq. 5) is assumed to be Gaussian-distributed, $\boldsymbol{\epsilon}^s \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_s)$. The same assumption applies to the control parameters: $\boldsymbol{\alpha}^s \sim \mathcal{N}(\mathbf{1}, \mathbf{B}_s)$. The probability density function (pdf) of the observation vector can be computed as follows:

$$p(\boldsymbol{\mu}^s | r_s, m_s) = \int p(\boldsymbol{\mu}^s | \boldsymbol{\alpha}^s, r_s, m_s) p(\boldsymbol{\alpha}^s | r_s, m_s) \mathrm{d}\boldsymbol{\alpha}^s = \int p(\boldsymbol{\epsilon}^s | r_s) p(\boldsymbol{\alpha}^s | m_s) \mathrm{d}\boldsymbol{\alpha}^s \tag{10}$$

or analytically,

$$p(\boldsymbol{\mu}^s | r_s, m_s) = \frac{e^{-\frac{1}{2}(\boldsymbol{\mu}^s - \mathbf{H}^s \mathbf{1} - \boldsymbol{\lambda}^s)^{\mathrm{T}} (\mathbf{R}_s + \mathbf{H}^s \mathbf{B}_s \mathbf{H}^{s\mathrm{T}})^{-1} (\boldsymbol{\mu}^s - \mathbf{H}^s \mathbf{1} - \boldsymbol{\lambda}^s)}}{\sqrt{(2\pi)^{d_s} |\mathbf{R}_s + \mathbf{H}^s \mathbf{B}_s \mathbf{H}^{s\mathrm{T}}|}} . \tag{11}$$

This pdf is also proportional to the likelihood of r_s and m_s. In order to estimate the hyperparameters r_s and m_s, Desroziers and Ivanov (2001) suggested an iterative method to converge towards a fixed point. Chapnik et al. (2006) showed that this approach converges to one local maximum of the likelihood. The maximisation of $\log\left(p(\boldsymbol{\mu}^s|r_s, m_s)\right)$ with respect to the two hyperparameters gives the stationary conditions:

$$r_s^2 = \frac{\| \boldsymbol{\mu}^s - \mathbf{H}^s \boldsymbol{\alpha}_a^s \|^2}{\mathrm{Tr}\left(\mathbf{I}_{d_s} - \mathbf{H}^s\mathbf{K}^s\right)}, \qquad m_s^2 = \frac{\| \boldsymbol{\alpha}_a^s - \mathbf{1} \|^2}{\mathrm{Tr}\left(\mathbf{H}^s\mathbf{K}^s\right)} \tag{12}$$

where $\|\cdot\|$ is the Euclidean norm.

However, the Desroziers method relies on Gaussian assumptions, and, for the sake of consistency, one needs another approach to compute the likelihood under the truncated Gaussian assumption (Winiarek et al., 2012). In this case, the prior on the scaling factors is:

$$p(\boldsymbol{\alpha}^s) = \frac{e^{-\frac{1}{2}(\boldsymbol{\alpha}^s-\mathbf{1})^{\mathrm{T}}\mathbf{B}_s^{-1}(\boldsymbol{\alpha}^s-\mathbf{1})}}{\sqrt{(2\pi)^{N_x\times N_y}|\mathbf{B}_s|}\left(1-\Phi_{\mathbf{1},\mathbf{B}_s}(\mathbf{0})\right)} \mathbb{1}_{\boldsymbol{\alpha}^s\geq\mathbf{0}} \tag{13}$$

where $\Phi_{\mathbf{1},\mathbf{B}_s}(\mathbf{0})$ is the Gaussian cumulative density function (cdf) of $\mathcal{N}(\mathbf{1},\mathbf{B}_s)$. $\mathbb{1}_{\boldsymbol{\alpha}^s\geq\mathbf{0}}$ is a function equal to unity when $[\boldsymbol{\alpha}^s]_l \geq 0$ for each l, and equal to zero otherwise. This pdf is referred to as $\mathcal{N}(\mathbf{1},\mathbf{B}_s,\mathbf{0})$ or the truncated Gaussian distribution. From Eq. (10) and Eq. (13), one can derive:

$$p(\boldsymbol{\mu}^s|r_s^+, m_s^+) = \frac{e^{-\frac{1}{2}(\boldsymbol{\mu}^s-\mathbf{H}^s\mathbf{1}-\boldsymbol{\lambda}^s)^{\mathrm{T}}(\mathbf{R}_s+\mathbf{H}^s\mathbf{B}_s\mathbf{H}^{s\mathrm{T}})^{-1}(\boldsymbol{\mu}^s-\mathbf{H}^s\mathbf{1}-\boldsymbol{\lambda}^s)}}{\sqrt{(2\pi)^{d_s}|\mathbf{R}_s+\mathbf{H}^s\mathbf{B}_s\mathbf{H}^{s\mathrm{T}}|}} \times \frac{\int_{\boldsymbol{\alpha}^s\geq\mathbf{0}} e^{-\frac{1}{2}(\boldsymbol{\alpha}^s-\mathbf{1})^{\mathrm{T}}(\mathbf{P}_s^a)^{-1}(\boldsymbol{\alpha}^s-\mathbf{1})}}{\sqrt{(2\pi)^{N_x\times N_y}|\mathbf{B}_s|}\left(1-\Phi_{\mathbf{1},\mathbf{B}_s}(\mathbf{0})\right)} \tag{14}$$

where r_s^+ and m_s^+ refer respectively to the standard deviation of the error and of the emission noise (departure of the surface fluxes from their a priori values) according to the truncated Gaussian distribution. $\mathbf{P}_s^a$ is the a posteriori error covariance matrix of the control variables of the Gaussian case,

$$\mathbf{P}_s^a = \mathbf{B}_s \left(\mathbf{I} - \mathbf{K}^s \mathbf{H}^s\right) . \tag{15}$$

Even though it is of formal use in the truncated Gaussian case, $\mathbf{P}_s^a$ *is not* the error covariance matrix of the truncated Gaussian case. A mathematical hardship is that Eq. (15) requires the computation of a 1,768-dimensional ($N_\mathrm{x} \times N_\mathrm{y}$) integral over the positive cone. To overcome this difficulty, we resort to the sampling technique used by Winiarek et al. (2012).

3 Setup of the numerical experiments

3.1 Observations

The in situ observations used in this study are extracted from the EMEP database (http://www.nilu.no/projects/ccc/emepdata.html). The EMEP monitoring network covers most of Europe. Eleven stations of western Europe are used in this study, resulting in a rather sparse network. These stations measure the concentrations of fourteen different VOCs. Note that the m-xylene and p-xylene are combined in a lumped mp-xylene category. The observations are from 11 January 2005 to 29 December 2005.

Table 2 gives the number of observations used in the inversion per species and per station, for a total of 18 675 observations from 11 stations. A forecast test will also be performed using 19 746 observations of the year 2006. The VOC station Kollumerwaard, in the Netherlands (code NL0009R), does not provide any observation in 2005. However, this station provides 26 732 observations in 2006, and they will be used for cross-validation. The locations of the EMEP sampling stations for VOCs are shown in Fig. 1.

3.2 Inversion and validation setup

Two simulation periods are considered. The first one is the the assimilation time window of the study, from January 11 to December 29, 2005. The second one is the subsequent validation of the inversion results and corresponds to the whole year 2006.

The simulation domain over western Europe extends between [40° N;8° W] (the left bottom corner) and [57° N;18° E] (the right top corner). The grid resolution is 0.5° × 0.5°. Nine levels are considered above the surface: 50, 250, 600, 1000, 1500, 2100, 2800, 3600 and 4500 m a.g.l. The control space discretisation follows the simulation grid, with $N_x = 52$ and $N_y = 34$. As a result, the number of control variable $[\boldsymbol{\alpha}^s]_l$ used in the inversion for each species s, is 1.768.

The meteorological data are generated from the re-analysis fields of the European Centre for Medium Range Weather Forecast (ECMWF), delivered in 60 vertical levels and every 3 h, with a horizontal resolution of 0.36° × 0.36°.

For anthropogenic emissions, the background emissions over the whole domain are provided by the EMEP inventory for the years 2005 and 2006 (Tarrasón et al., 2007; Fagerli et al., 2008). The disaggregation of the EMEP NMVOC emissions among the many VOC species simulated in the CTM is performed following Passant (2002).

The anthropogenic emissions of EMEP have a resolution of 0.5° × 0.5°. These emissions are modulated in time with the help of the hourly, weekly and monthly distribution coefficients, provided by the GENEMIS project (GENEMIS, 1994). The biogenic emissions of isoprene are also taken into account using the model proposed by Simpson et al. (1999). All these emissions are used as a first guess, $\boldsymbol{e}^b$, in the data assimilation experiments.

The initial and boundary conditions concentration fields are obtained from the global chemistry transport model MOZART 2 (Horowitz et al., 2003). Since the species we are interested in are not all explicitly present in MOZART 2, the values for the VOCs not included in MOZART 2 were inferred from the concentration fields of some species present in MOZART 2. The factors applied for this inference are given in Table 3 (see Rudolph and Ehhalt, 1981; Rudolph and Johnen, 1990; Penkett et al., 1993).

3.3 Verification of the adjoint solutions

In order to generate the adjoint solutions at a low computational cost, we have used an approximate adjoint model, following the construction of Bocquet (2005). Moreover, we have assumed the lifetime of the species within the domain to be less than 10 days. After 10 days, the VOCs are assumed to be out of the domain or consumed by chemical reactions, so that the sensitivity of the concentrations within the domain to the emissions is negligible.

In order to check these approximations, the concentration fields from the adjoint model, obtained from the contribution of the source ($\mathbf{H}\boldsymbol{\alpha}$ when $\boldsymbol{\alpha} = \mathbf{1}$), were compared with the concentration fields from the direct simulation (for $\boldsymbol{e} = \boldsymbol{e}^b$, the EMEP inventory first-guess, with clean air boundary and initial conditions). This is the so-called duality test (Davoine and Bocquet, 2007). The correlation between both computations, averaged over the 14 species, is 0.994. The normalised root mean square difference between both computations, averaged over the 14 species, is 3×10^{-3}.

Figure 2 shows the comparison between the source contribution estimated with the forward model and with the adjoint model. These results indicate that the adjoint model is accurate enough for our inverse modelling purpose.

3.4 Values of the hyperparameters

In the Gaussian case, the optimal values of the hyperparameters r_s and m_s are obtained by value screening of the pdf Eq. (11). Their optimal values are reported in Table 4 for each species. As an example, Fig. 3 displays $p(\boldsymbol{\mu}^s|r_s, m_s)$ for the species NBUT. The coordinates are normalised with respect to the values of the hyperparameters obtained from the fixed-point solutions of Eq. (12). It can be seen that the optimal value of the likelihood Eq. (11) is equal to the result of the fixed-point method.

A comparison between the likelihoods Eq. (11) and Eq. (14) is also shown in Table 4. The optimal value of r_s^+ and m_s^+ computed with Eq. (14) are obviously different from r_s and m_s. Remarkably r_s^+ is always larger than r_s. Indeed, the Gaussian assumption inversion incorrectly interprets part of the noise within the observations as useful information, while the positivity constraint of the truncated Gaussian assumption offers less flexibility. Therefore, the inversion based on Gaussian assumptions underestimates the errors' magnitude. The comparison between m_s^+ and m_s is of less relevance because the two parameters do not represent the same statistical information within the Gaussian or truncated Gaussian assumptions.

As an example, Fig. 4 displays the likelihood of the hyperparameters for ethane in the truncated Gaussian case. Again, the variables are normalised with respect to the values obtained from the Desroziers method.

4 Inversion results

Four types of simulations conducted for the year 2005 are reported here:

- In case A, the VOC concentrations are simulated using the EMEP inventories, $\boldsymbol{e}_k = \boldsymbol{e}_k^b$ in Eq. (1). This is a free run serving as a reference since the observations are not assimilated.

- In cases B1 and B2, the concentrations are simulated using the emissions obtained from Eq. (4). The scaling factors $\boldsymbol{\alpha}^s$ used in this equation, are obtained by the minimisation of Eq. (7). The L-BFGS-B optimisation tool (Byrd et al., 1995) is used under the constraints that $\boldsymbol{\alpha}^s \geq \mathbf{0}$ (case B2). In case B1, we merely use the BLUE matrix formula, Eq. (8), for the estimator of $\boldsymbol{\alpha}^s$. However, it is assumed in both cases that the errors are essentially Gaussian, so that the hyperparameters used in the inversion are computed with the Gaussian likelihood following Sect. 2.5. Because the statistical assumptions are different in the estimation of the fluxes and the estimation of the hyperparameters, this may lead to inconsistencies. However, the fact that the r^+ and r hyperparameters are not too different proves that these inconsistencies are small. Yet, by construction, case B1 can lead to negative emission fluxes. This may be considered unphysical but it might have some potential use for air quality forecast.

We have also considered a third case B3, that takes the results of B2 and artificially sets all negative values to zero. However, we found that because it is not a minimum of Eq. (7), it leads to a poorly performing estimation. Therefore, case B3 uses ruled out and is not reported here.

- Case C is similar to case B2 except that the hyperparameters are obtained using the truncated Gaussian likelihood, following Sect. 2.5. In this case the estimation of the emission fluxes and the estimation of the hyperparameters are statistically fully consistent.

4.1 Analysis of the inversion results

4.1.1 A posteriori verification of the model linearisation

In the full CTM, which is used in the present study, the chemical kinetics of the reactions can be written as follows:

$$\mathcal{X}(c) = \left(\frac{\delta\mathcal{X}(c)}{\boldsymbol{X}(c)} + 1\right)\boldsymbol{X}(c) \tag{16}$$

where $\delta\mathcal{X}(c)$ denotes the variation of the VOC concentrations field with respect to the variation of the oxidant concentration field. Until now, we assumed that $\delta\mathcal{X}(c) \ll \boldsymbol{X}(c)$, and that the reduced model was linear. In order to check that hypothesis, the a priori and the a posteriori oxidant concentration fields are compared. Figure 5 shows the comparison of the concentration fields of OH before and after data assimilation (case B2). Each point in the figure denotes the average concentration of OH over the spatial domain for a 2-hour period. The mean value of the concentrations is of

5.51×10^{-5} µg m^{-3} for the a priori fields and 5.02×10^{-5} µg m^{-3} for the a posteriori fields. For the species NO_3 and O_3, the average values of the a priori concentrations are 0.0147 µg m^{-3} and 87.73 µg m^{-3} respectively. They are 0.0131 µg m^{-3} and 88.73 μg m^{-3} for the a posteriori concentrations. The Pearson correlation between the a priori and a posteriori concentration fields is about 1.00. Furthermore, an examination of the a posteriori VOC concentration fields shows that the results obtained with the reduced linear model are very close to those obtained with the complete model. The relative bias between the two sets of concentrations for all of the VOC species and over the entire spatial domain is 1 % for the year 2005 and the correlation is 1.00. Therefore, since the oxidant concentration fields are little affected by the VOC data assimilation, we consider that the hypothesis that the reduced model is about linear is verified.

4.1.2 Comparison to the observations

The four runs A, B1, B2 and C are compared with the observations of the analysis period (year 2005). Statistical indicators for this comparison are reported in Table 5. For most species the bias between the concentrations and the observations decreases with data assimilation, except when the mean of the simulation is already close to the measurement mean. The root mean square errors (RMSEs) and the normalised mean square errors (NMSEs) are systematically improved in the re-analysis runs, which is consistent with the fact that our inversion scheme minimises the quadratic error. For all species, except ACE, the Pearson correlation coefficients R, FA_2 and FA_5 are remarkably improved in the re-analyses. Note that FA_x is the fraction of the simulated concentrations within a factor x of the corresponding observations. In the very few cases where an indicator is not improved, other indicators are improved. The decrease of the correlation in the ACE case is due to a very large bias, which is compensated by a very significant improvement of the other indicators, starting with the bias.

Considering all species (14) and all statistical indicators (6) together, we have counted how often the forecast runs B1, B2 and C beat the free run A: 81 times out of 84 in case B1; 80 times out of 84 in case B2; and 79 times out of 84 in case C.

The fact that run B1 is slightly closer to the observations than B2 and C is consistent with the fact that optimisation on which B1 relies is less constrained than that of B2 or C (emission fluxes can be negative in case B1). Yet, it does not prove that method B1 is better than method B2 or C, since a comparison with the (already assimilated) observations is merely a check of consistency, not a validation.

4.1.3 Estimated inventories

The total emitted mass of the EMEP inventories for the fifteen species are compared with the a posteriori emissions obtained from data assimilation following Gaussian assumption (B1/B2) and truncated Gaussian assumption (C). The results are reported in Table 6. Caution must be used on the interpretation of the results of the statistically consistent Gaussian case B1 since fluxes are allowed to be negative in this case (e.g. isoprene emissions).

The results of case B2 and case C indicate that the EMEP inventories may underestimate the true emissions for C_3H_8, IBUT, IPEN, C_3H_6, ACE, C_2H_6, C_2H_4, and BEN. They may overestimate the true emissions for the other VOC species. For all the species, the total mass obtained from the inversion based on the truncated Gaussian assumption (C case) is between the EMEP inventory and the total mass estimated from the Gaussian assumption (case B2). Comparison between cases A and C shows a strong correction for C_3H_8, ACE, C_2H_6, and C_2H_4. It is less than 20 % for NBUT, NPEN, IPEN, TOLU, OXYL and MXYL. Figure 6 presents the ratio between the correction of the emission (i.e., the total posterior emission minus the total prior EMEP emissions) and the total prior emission for the Gaussian cases (B1 and B2) and the truncated Gaussian case (C).

CC BY

4.1.4 Spatial distribution

The spatial extent of the corrections from the EMEP network depends on the nature of the species. As an example, Fig. 7 displays the spatial ratio between the posterior emissions and the prior EMEP emissions for the species C_2H_6 (with an average lifetime of about 60 days), IPEN (with an average lifetime of 4 days), OXYL (with a lifetime of about 1.1 days), and ISO (with an average lifetime of 1.7 h), respectively. Obviously, the corrections extend much farther from the monitoring stations for the long-lived species, such as C_2H_6 and IPEN than for the short-lived species.

As can be seen, with the B2 approach based on a Gaussian estimation of the prior errors, the magnitude of the corrections is significantly higher than when using the fully consistent non-Gaussian approach C. One way to put it is that, with the Gaussian assumption, part of the error (noise) is mis-interpreted as valuable information (signal), so that the Gaussian assumption leads to over-corrections. A similar phenomenon was put forward by Koohkan and Bocquet (2012) in the inversion of carbon monoxide emission fluxes: the proper identification of representativeness errors leads to smaller corrections of the emission fluxes.

Several studies have performed inverse modelling study of isoprene emissions over Europe. To do so, they assimilate satellite observations of formaldehyde (e.g. Curci et al., 2010; Dufour et al., 2009). Specifically, they exploit the observations of SCIAMACHY (instrument operating onboard the sunsynchronous Envisat satellite) with a resolution of 30 × 60 km^2. The study of Curci et al. (2010) shows an increase of about 5 % for the MEGAN (Guenther et al., 2006) emissions. Our results show that the emission of isoprene decreases by about 24 % in case B2 and by 13 % in case C for the emissions obtained from Simpson et al. (1999). This discrepancy is explained in part by the fact that the emission inventories of isoprene by Simpson et al. (1999) and MEGAN differ significantly (Bessagnet et al., 2008; Sartelet et al., 2012) with the former leading to greater isoprene emissions on average over Europe by a factor of about 2.5 (Sartelet et al., 2012). This discrepancy is also explained by the very short lifetime of isoprene

(1–2 h). Assimilation of in-situ observations are inoperative because the information is not spread far enough by the model, because its transport representation becomes of minor interest. Our results are only valid near the five stations measuring isoprene. On the contrary, satellite observations are well-suited for this short-lived species as remote sensing offers an (indirect) spatially well-resolved snapshot of the concentrations. In addition, the errors made by approximate adjoint model (or automatic adjoint up to some numerical precision) are larger for short-lived species (Bocquet, 2012). As a consequence, the case of isoprene in this study is somehow singular because its lifetime is too short. Nevertheless, most results of isoprene are given in this study for the sake of comparison, and to document the inadequacy of the in situ observations for regional inverse modelling in such a case.

4.2 Forecast test

In order to test and possibly validate the corrected inventory obtained from inverse modelling, one needs observations that have not been assimilated. One stringent test is to perform a forecast using the corrected inventory over a period of time different from the data assimilation window, assuming some time persistence of the VOC inventories.

Four inventories are generated with Eq. (4), using the background EMEP emission $\boldsymbol{e}_b^s$, or $\boldsymbol{\alpha}_b^s = \mathbf{1}$ (case A), and the scaling factors of cases B1, B2 and C. In each case, a forecast is performed for the year 2006. These simulations are then compared with the independent observations of year 2006.

Note that because some of the fluxes retrieved in case B1 are negative, and because numerical schemes of CTMs often rely on the positivity of the concentrations, we had to circumvent the difficulty. One solution consists in decomposing the scaling factors into a positive part and a negative part: $\boldsymbol{\alpha}^s = \boldsymbol{\alpha}_+^s + \boldsymbol{\alpha}_-^s$, so that, invoking the linearity of physics, the concentrations are given by $\mathbf{H}^s(\boldsymbol{\alpha}^s) = \mathbf{H}^s(\boldsymbol{\alpha}_+^s) - \mathbf{H}^s(-\boldsymbol{\alpha}_-^s)$.

The statistical indicators are reported in Table 7. The results indicate that, for most species, the scores are improved using the retrieved scaling factors of case B2 and case C, whereas they are degraded in case B1. Considering all species and all statistical indicators together, we have counted how often the forecast runs B1, B2 and C beat the free run A: 41 times out of 84 in case B1; 69 times out of 84 in case B2; and 77 times out of 84 in case C.

The fully-consistent truncated Gaussian approach C performs best and beats the Gaussian-based but positively constrained approach B2. Both positively constrained approaches beats the fully consistently Gaussian approach B1. Since B1 partially leads to unphysical negative fluxes, this could have been expected. However, as was shown by Bocquet (2012), an unconstrained optimisation of parameters that are allowed to take unphysical values may sometimes lead to valuable better forecasts because it compensates for other sources of model error. Obviously, this is not the case here.

According to Table 7, an improvement of all the scores can be seen for C_3H_8, NBUT, NPEN, C_3H_6, C_2H_6, C_2H_4 and BEN using the optimal scaling factors. However, the bias is increased for OXYL (in case B2), MPXYL (in cases B2 and C) and ISO (in case B2), using the corrected emissions. Despite that degraded bias, the other indicators are improved. For IBUT, the RMSE is increased in case B2, and the correlation is decreased in cases B2 and C. The statistical indicators show that the RMSE increases for IPEN using the corrected emissions. For that species, the Pearson correlation coefficient is also deteriorated in case B2. For TOLU, the bias, NMSE and FA_5 are degraded using scaling factors. The RMSE is also increased in case B2. The scores also show that a decrease of the Pearson correlation coefficient for ACE, using the corrected emissions.

4.3 Cross-validation test

Since the EMEP Netherlands VOC station Kollumerwaard does not provide any observation for the year 2005, that station is not included in the list of stations used for the analysis of 2005. Yet, for the year 2006, eleven out of fourteen VOCs were measured at Kollumerwaard, excluding NBUT, ACE and C_2H_6. Kollumerwaard is located far from the other stations whose observations lead to the corrections in the emission inventories. Therefore, we do not expect the simulated concentrations of the VOC with a short lifetime, at this station, to be very sensitive to the correction of the emissions. Indeed,

CC BY

as shown in Table 8, the scores obtained from the comparison of the simulated concentrations in case A of ISO (lifetime of about 1.7 h) with the measurements are similar to those obtained from the comparison of the concentrations in case B and C with the observations.

For species with a longer lifetime (about one day or two), for C_3H_6 (lifetime of 15 h), OXYL (lifetime of about 25-26 h), MPXYL (lifetime of about 14-25 h) and TOLU (lifetime of 2.4 days), the corrections of the inventories are more spread over the domain. The scores indicate that the results in case C are in a better agreement with the observations. The results also show that for the species with a lifetime between 4 to 7 days (IBUT, NPEN, and IPEN), the scores are deteriorated. It is likely that the corrections performed close to the Kollumerwaard station, are not reliable enough. For the species with a lifetime longer than 10 days (C_3H_8 and BEN), the scores are remarkably improved. The emission inventories of C_3H_8 are corrected almost all over the domain.

Considering all species and all statistical indicators together, we have counted how often the forecast runs B2 and C beat the free run A at the Kollumerwaard station: 34 times out of 66 in case B2; and 43 times out of 66 in case C. The use of the corrected emission from case C has a positive impact on the forecast at the station, even though the station is far away from the other stations whose observations were assimilated.

4.4 Information content and DFS

The degrees of freedom for the signal (DFS) is a metric that is representative of the fraction of the observations effectively used in the inversion to retrieve the source (Koohkan et al., 2012). A better data assimilation system can either lead to an increase of the DFS when the observations are better used by the system, or lead to a decrease of the DFS if the system better diagnoses the errors and correctly identifies more noise in the observations (Koohkan and Bocquet, 2012). Therefore, a comparison of DFS from two data assimilation systems relying on different assumptions is not straightforward. However, a comparison between DFS of different species for the same data assimilation system is of simpler interpretation. The DFS for each species is given

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀| |▶
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

by:

$$\mathrm{DFS}_s = \mathbb{E}\left[\left(\boldsymbol{\alpha}^s - \mathbf{1}\right)^{\mathrm{T}} \mathbf{B}_s^{-1} \left(\boldsymbol{\alpha}^s - \mathbf{1}\right)\right]. \tag{17}$$

The ratios between DFS_s and $\mathcal{L}_s$ (see Eq.(7)) are reported in Table 9 for cases B1, B2, and C. For all species, the ratio is greater for B1. Indeed, by permitting negative fluxes, the data assimilation system is incorrectly interpreting degrees of freedom for the noise as DFS. As for the positively constrained inversions, for almost all species except BEN, the ratio is greater in the statistically consistent truncated Gaussian system. The DFS, 4 ± 2 % in the B2 case, and 7 ± 2 %, is consistent with the figures usually met in air pollution source inverse modelling systems (Koohkan et al., 2012, argue that it is usually between 5 % and 15 % for dispersion problems). Isoprene is clearly identified as the species with the lower ratio in the B2 case.

5 Conclusions

The goal of this study was to estimate the emission inventories of fifteen VOC species using ground-based in situ measurements. The concentration observations at eleven stations from the EMEP network over western Europe were assimilated to perform inverse modelling of the emission field for each one of the fifteen species, for the year 2005.

For that purpose, the Jacobian matrix, i.e., the source-receptor relationship, was built using the POLAIR3D CTM. To compute that matrix, a fast version of this CTM, as well as its validated approximate adjoint model have been developed. The chemistry module of this fast version only includes the chemical reactions between the VOC species and three oxidants (OH, NO_3 and O_3), the concentrations of which are pre-computed with the full CTM.

For each species and each grid cell, a scaling factor that multiplies the local EMEP emission flux, is computed. The uncertainty attached to the prior scaling factors and the covariance matrix of the observation errors, which are crucial statistical components of the inversion, are obtained using the maximum likelihood principle. The principle was implemented using two different assumptions: (1) the errors attached to the scaling factors follow a Gaussian pdf or (2) the scaling factor follows a truncated Gaussian pdf.

In the Gaussian case, the simulated concentrations for the year 2005 using the corrected emissions lead to a significant improvement in most statistical indicators. However the fact that the VOC fluxes are positive is not statistically accounted for, and this is shown to lead to a probable over-fitting to the observations, and to over-corrections of the EMEP emissions. Using a fully consistent truncated Gaussian assumption for the emission fluxes, including the use of a non-Gaussian likelihood for the estimation of the hyperparameters, the corrections are significantly smaller.

For short-lived species, it is shown that information cannot propagate far from the monitoring stations, so that the corrections are rather local to the stations. That is why we deem the isoprene inversion to be unreliable. That is a typical case where remote sensing assimilation is necessary to offer a satisfying coverage.

The corrected emissions have been partly validated thanks to a forecast conducted for the year 2006 using independent observations. The simulations using the corrected emissions often led to significant improvements in the statistical indicators. Considering all statistical indicators, the fully consistent truncated Gaussian approach emerged as the best approach from this test.

The 2006 forecasts have also been compared to the observations at the Kollumerwaard station, the Netherlands. The Kollumerwaard station is not part of the 11 stations used in the analysis of 2005. Even though this station is far away from the 2005 network, and its surroundings fluxes little affected by the 2005 analysis, some improvements are noticed for several long-lived VOC species using the statistically consistent positively constrained inversion.

For anthropogenic VOC (alkanes, alkenes, acetylene, and aromatics), the inversion modelling suggests that the current EMEP emission inventories are correct within 33 % on average for n-butane, isobutane, n-pentane, isopentane, propylene and aromatics

CC BY

(benzene, toluene and xylenes). However, there are large discrepancies (about an order of magnitude) in specific regions. The emission inventories appear to be significantly underestimated (about a factor of two) for ethane, n-propane, ethylene and acetylene. Revisions to VOC emission inventories appear to be warranted and experimental investigations of speciated VOC emissions from major source categories (e.g., on-road traffic, petrochemical industry) are recommended.

It would be of interest to assess the method potential efficiency in more specific situations. The methodology developed here could be applied for specific source categories using tracer measurements such as levoglucosan for biomass burning. Its efficiency could also be investigated for short-lived species but for local emission inventories, at the urban scale for instance. Of course, this extension of the method implies the availability of a large and consistent set of measurements of specific species.

Acknowledgements. This paper is a contribution to the CARBOSOR project, which is sponsored by the Primequal research program of the French ministry of Ecology and of ADEME, the French Agency of Environmental Protection and Energy Management, and a contribution to the ADOMOCA-2 project, which is sponsored by the INSU/LEFE council. The authors acknowledge the useful pieces of advice from Stéphane Sauvage on the EMEP data and inventory.

References

Bessagnet, B., Menut, L., Curci, G., Hodzic, A., Guillaume, B., Liousse, C., Moukhtar, S., Pun, B., C., S., and Schulz, M.: Regional modeling of carbonaceous aerosols over Europe-focus on secondary organic aerosols, J. Atmos. Chem., 61, 175–202, 2008. 33237

Bocquet, M.: Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Theory, Q. J. Roy. Meteor. Soc., 131, 2191–2208, 2005. 33232

Bocquet, M.: Parameter field estimation for atmospheric dispersion: Application to the Chernobyl accident using 4D-Var, Q. J. Roy. Meteor. Soc., 138, 664–681, doi:10.1002/qj.961, 2012. 33238, 33239

Byrd, R. H., Lu, P., and Nocedal, J.: A Limited Memory Algorithm for Bound Constrained Optimization, SIAM J. Sci. Stat. Comp., 16, 1190–1208, 1995. 33233

cc BY

Chapnik, B., Desroziers, G., Rabier, F., and Talagrand, O.: Properties and first application of an error-statistics tuning method in variational assimilation, Q. J. Roy. Meteor. Soc., 130, 2253–2275, 2006. 33229

Curci, G., Palmer, P. I., Kurosu, T. P., Chance, K., and Visconti, G.: Estimating European volatile organic compound emissions using satellite observations of formaldehyde from the Ozone Monitoring Instrument, Atmos. Chem. Phys., 10, 11501–11517, doi:10.5194/acp-10-11501-2010, 2010. 33237

Davoine, X. and Bocquet, M.: Inverse modelling-based reconstruction of the Chernobyl source term available for long-range transport, Atmos. Chem. Phys., 7, 1549–1564, doi:10.5194/acp-7-1549-2007, 2007. 33228, 33232

Desroziers, G. and Ivanov, S.: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation, Q. J. Roy. Meteor. Soc., 127, 1433–1452, 2001. 33228

Dufour, G., Wittrock, F., Camredon, M., Beekmann, M., Richter, A., Aumont, B., and Burrows, J. P.: SCIAMACHY formaldehyde observations: constraint for isoprene emission estimates over Europe?, Atmos. Chem. Phys., 9, 1647–1664, doi:10.5194/acp-9-1647-2009, 2009. 33222, 33237

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749–3769, doi:10.5194/acp-7-3749-2007, 2007. 33222

Fagerli, H., Gauss, M., Jonson, J., Nyíri, Á., Simpson, D., Tarrasón, L., Tsyro, S., and Wind, P.: Transboundary acidification, euthrophication and ground level ozone in Europe. Part I: Unified EMEP model description, Tech. rep., EMEP, 2008. 33231

Fu, T.-M., Jacob, D., Palmer, P., Chance, K., Wang, Y., Barletta, B., Blake, D., Stanton, J., and Pilling, M.: Space-based formaldehyde measurements as constraints on volatile organic compound emissions in east and south Asia and implications for ozone, J. Geophys. Res., 112, D06312, doi:10.1029/2006JD007853, 2007. 33222

GENEMIS: Generation of European Emission Data for Episodes (GENEMIS) project, Tech. rep., EUROTRAC annual report 1993, Garmisch-Partenkirchen, Germany, 1994. 33231

Goliff, W. S. and Stockwell, W. R.: The Regional Atmospheric Chemistry Mechanism, version 2, an update, international con- ference on Atmospheric Chemical Mechanisms, University of California, Davis, 2008. 33224

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and

cc BY

Aerosols from Nature), Atmos. Chem. Phys., 6, 3181–3210, doi:10.5194/acp-6-3181-2006, 2006. 33237

Harley, R. and Cass, G.: Modeling the atmospheric concentrations of individual volatile organic compounds, Atmos. Environ., 29, 905–922, 1995. 33222

Ho, K., Ho, S., Cheng, Y., Lee, S., and Yu, J.: Real-world emission factors of fifteen carbonyl compounds measured in a Hong Kong tunnel, Atmos. Environ., 41, 1747–1758, 2007. 33221

Hopkins, J. R., Evans, M. J., Lee, J. D., Lewis, A. C., H Marsham, J., McQuaid, J. B., Parker, D. J., Stewart, D., Reeves, C. E., and Purvis, R. M.: Direct estimates of emissions from the meacity of Lagos, Atmos. Chem. Phys., 9, 8471–8477, doi:10.5194/acp-9-8471-2009, 2009. 33222

Horowitz, L. W., Walters, S., Mauzerall, D. L., Emmons, L. K., Rasch, P. J., Granier, C., Tie, X., Lamarque, J. F., Schultz, M. G., Tyndall, G. S., Orlando, J. J., and Brasseur, G. P.: A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2, J. Geophys. Res., 108, D24, doi:10.1029/2002JD002853, 2003. 33231

Kim, Y., Sartelet, K., and Seigneur, C.: Comparison of two gas-phase chemical kinetic mechanisms of ozone formation over Europe, J. Atmos. Chem., 62, 89–119, 2009. 33224

Koohkan, M. R. and Bocquet, M.: Accounting for representativeness errors in the inversion of atmospheric constituent emissions: Application to the retrieval of regional carbon monoxide fluxes., Tellus B, 64, 19 047, 2012. 33222, 33237, 33240

Koohkan, M. R., Bocquet, M., Wu, L., and Krysta, M.: Potential of the International Monitoring System radionuclide network for inverse modelling, Atmos. Environ., 54, 557–567, 2012. 33240, 33241

Mallet, V., Quélo, D., Sportisse, B., Ahmed de Biasi, M., Debry, É., Korsakissok, I., Wu, L., Roustan, Y., Sartelet, K., Tombette, M., and Foudhil, H.: Technical Note: The air quality modeling system Polyphemus, Atmos. Chem. Phys., 7, 5479–5487, doi:10.5194/acp-7-5479-2007, 2007. 33223

Millet, D., Jacob, D., Boersma, K., Fu, T.-M. Kurosu, T., Chance, K., Heald, C., and Guenther, A.: Spatial distribution of isoprene emissions from North America derived from formaldehyde column measurements by the OMI satellite sensor, J. Geophys. Res., 113, D02307, doi:10.1029/2007JD008950, 2008. 33222

Passant, N. R.: Speciation of UK Emissions of NMVOC, Tech. Rep. AEAT/ENV/0545, AEA Technology, 2002. 33231

cc BY

Penkett, S. A., J., B. N., Lightman, P., Marsh, A. R. W., Anwyl, P., and Butcher, G.: The Seasonal Variation of Nonmethane Hydrocarbons in the Free Troposphere Over the North Atlantic Ocean: Possible Evidence for Extensive Reaction of Hydrocarbons With the Nitrate Radical, J. Geophys. Res., 98, 2865–2885, 1993. 33232

Quélo, D., Mallet, V., and Sportisse, B.: Inverse mmodeling of NOx emissions at regional scale over northern France: Preliminary investigation of the second-order sensitivity, J. Geophys. Res., 110, D24 310, doi:10.1029/2005JD006151, 2005. 33222

Roustan, Y. and Bocquet, M.: Sensitivity analysis for mercury over Europe, J. Geophys. Res., 111, D14 304, doi:10.1029/2005JD006616, 2006. 33225

Rudolph, J. and Ehhalt, D.: Measurements of C2-C5 hydrocarbons over the North Atlantic, J. Geophys. Res., 86, 11 959–11 964, 1981. 33232

Rudolph, J. and Johnen, F. J.: Measurements of Light Atmospheric Hydrocarbons Over the Atlantic In Regions of Low Biological Activity, J. Geophys. Res., 95, 20 583–20 591, 1990. 33232

Sartelet, K., Debry, E., Fahey, K., Roustan, Y., Tombette, M., and Sportisse, B.: Simulation of aerosols and gas-phase species over Europe with the Polyphemus system, Part I: model-to-data comparison for 2001, Atmos. Environ., 41, 6116–6131, 2007. 33223

Sartelet, K., Couvidat, F., Seigneur, C., and Roustan, Y.: Impact of biogenic emissions on air quality over Europe and North America, Atmos. Environ., 53, 131–141, 2012. 33237

Sawyer, R., Harley, R., Cadle, S., Norbeck, J., Slott, R., and Bravo, H.: Mobile sources critical review: 1998 NARSTO assessment, Atmos. Environ., 34, 2161–2181, 2000. 33221

Shim, C., Wang, Y., Choi, Y., Palmer, P., Abbot, D., and Chance, K.: Constraining global isoprene emissions with Global Ozone Monitoring Experiment (GOME) formaldehyde column measurements, J. Geophys. Res., 110, D24 301, doi:10.1029/2004JD005629, 2005. 33222

Simpson, D., Winiwarter, W., Börjesson, G., Cinderby, S., Ferreiro, A., Guenther, A., Hewitt, C., Janson, R., Khalil, M., Owen, S., Pierce, T., Puxbaum, H., Shearer, M., Skiba, U., Steinbrecher, R., Tarrason, L., and Oquist, M.: Inventorying emissions from nature in Europe, J. Geophys. Res., 104, 8113–8152, 1999. 33231, 33237

Staehelin, J., Keller, C., Stahel, W., Schläpfer, K., and Wunderli, S.: Emission factors from road traffic from a tunnel study (Gubris tunnel, Switzerland), Part III: results of organic compounds, SO2 and speciation of organic exhaust emission, Atmos. Environ., 32, 999–1009, 1998. 33221

cc BY

Stemmler, K., Bugmann, S., Buchmann, B., Reimann, S., and Staehelin, J.: Large decrease of VOC emissions of Switzerland's car fleet during the past decade: results from a highway tunnel study, Atmos. Environ., 39, 1009–1018, doi:10.1016/j.atmosenv.2004.10.010, 2005. 33221

Tarrasón, L., Fagerli, H., Jonson, J., Simpson, D., Benedictow, A., Klein, H., and Vestreng, V.: Transboundary acidification, euthrophication and ground level ozone in Europe in 2005, Tech. rep., EMEP, 2007. 33231

Touaty, M. and Bonsang, B.: Hydrocarbon emissions in a highway tunnel in the Paris area, Atmos. Environ., 34, 985–996, doi:10.1016/S1352-2310(99)00195-8, 2000. 33221

Vijayaraghavan, K., Snell, H., and Seigneur, C.: Practical aspects of using satellite data in air quality modeling, Environ. Sci. Technol., 42, 8187–8192, 2008. 33222

Winiarek, V., M. Bocquet, M., Saunier, O., and Mathieu, A.: Estimation of Errors in the Inverse Modeling of Accidental Release of Atmospheric Pollutant: Application to the Reconstruction of the Cesium-137 and Iodine-131 Source Terms from the Fukushima Daiichi Power Plant, J. Geophys. Res., in press, 117, D05122, doi:10.1029/2011JD016932, 2012. 33229, 33230

Xiao, Y., Logan, J., Jacob, D., Hudman, R., Yantosca, R., and Blake, D.: Global budget of ethane and regional constraints on U.S. sources, J. Geophys. Res., 113, D21306, doi:10.1029/2007JD009415, 2008. 33222

Table 1. The volatile organic compounds, their (indicative) lifetime and reactions.

	species	symbol	lifetime (days)	reactions
RACM 2 reactions	isoprene	ISO	0.07	ISO + OH → ISO+ O_3 → ISO + NO_3 →
	acetylene	ACE	110	ACE + OH →
	ethane	C_2H_6	60	C_2H_6 + OH →
	ethylene	C_2H_4	1.45	C_2H_4 + OH → C_2H_4 + O_3 → C_2H_4 + NO_3 →
	benzene	BEN	11	BEN + OH →
additional reactions to RACM2	propane	C_3H_8	14	C_3H_8 + OH →
	n-butane	NBUT	7	NBUT + OH →
	isobutane	IBUT	7.5	IBUT + OH →
	n-pentane	NPEN	5	NPEN + OH →
	isopentane	IPEN	4	IPEN + OH →
	propene	C_3H_6	0.625	C_3H_6 + OH → C_3H_6 + O_3 → C_3H_6 + NO_3 →
	toluene	TOLU	2.4	TOLU + OH →
	o-xylene	OXYL	1.1	OXYL + OH →
	m-xylene*	MXYL	0.625	MXYL + OH →
	p-xylene*	PXYL	1.05	PXYL + OH →

* measured jointly in the EMEP monitoring network and represented with the symbol MPXYL.

Table 2. Number of observations for each species ($N^{obs}_{species}$) and number of observation dates for each station ($N^{obs}_{station}$). The numbers 1–12 are given to help locate the stations on the map of Fig. 1.

species	$N^{obs}_{species}$			station code	$N^{obs}_{station}$	
	2005	2006			2005	2006
C_3H_8	1505	1415	1	CZ0003R	1397	1297
NBUT	1503	1414	2	DE0002R	948	1023
IBUT	1033	1453	3	DE0003R	837	990
NPEN	1501	1409	4	DE0007R	942	1029
IPEN	1610	1553	5	DE0008R	883	1054
C_3H_6	1497	1411	6	DE0009R	938	1020
TOLU	1441	1408	7	DE0043G	9206	9283
OXYL	868	1273	8	ES0009R	259	199
MPXYL	1318	1459	9	FR0008R	1215	1285
ISO	936	1309	10	FR0013R	769	1214
ACE	1034	1407	11	FR0015R	1281	1352
C_2H_6	1494	1409	12	NL0009R	0	26732
C_2H_4	1434	1416				
BEN	1503	1410				

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
|◀ | ▶|
◀ | ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Table 3. Factors applied to MOZART 2 explicit species concentrations to determine the initial and boundary conditions of the model species.

species	Factors and MOZART 2 species
C_3H_8	C_3H_8
NBUT	0.44 C_3H_8
IBUT	0.22 C_3H_8
NPEN	0.05 C_3H_8
IPEN	0.1 C_3H_8
C_3H_6	C_3H_6
TOLU	0.26 C_3H_8
OXYL	0.03 C_3H_8
MPXYL	0.03 C_3H_8
ISO	0.03 C_3H_8
ACE	0.35 C_3H_8
C_2H_6	C_2H_6
C_2H_4	C_2H_4
BEN	0.44 C_3H_8

Table 4. Estimated standard deviations of the observation error and background error, under the Gaussian likelihood and truncated-Gaussian likelihood. The units of r_s and r_s^+ are µg m^{-3}.

species	r_s	m_s	r_s^+	m_s^+
C_3H_8	0.48	16.04	0.50	4.03
NBUT	0.36	2.02	0.37	0.64
IBUT	0.24	35.17	0.28	2.72
NPEN	0.19	6.01	0.20	0.76
IPEN	0.30	8.16	0.32	1.63
C_3H_6	0.14	14.22	0.15	2.84
TOLU	0.34	3.16	0.35	0.63
OXYL	0.08	16.75	0.09	1.33
PXYL	0.16	6.52	0.17	0.52
MXYL	0.16	6.52	0.17	0.52
ISO	0.46	82.65	0.60	2.61
ACE	0.35	41.26	0.36	8.23
C_2H_6	0.55	17.40	0.57	4.37
C_2H_4	0.35	31.19	0.40	3.93
BEN	0.24	11.12	0.27	1.11

Table 5. Scores from the comparison between the observations and the simulated concentrations for four simulations. For each species, the first line represents the scores for the simulations with the a priori fluxes (case A). The scores of the second line and third lines are related to the simulations with the a posteriori emissions from Gaussian hyper parameters estimation (case B1 and case B2 respectively). The scores of the fourth line are related to the simulations performed with the a posteriori emissions under the truncated Gaussian assumption (case C). The means and the RMSE are in $\mu g\, m^{-3}$. Bold numbers compared to the best agreement with the observations.

species	$\overline{O}$	case	$\overline{C}$	RMSE	NMSE	R	FA_2	FA_5
C_3H_8	1.13	A	0.54	0.83	1.14	0.77	0.40	0.946
		B1	**1.08**	**0.46**	**0.18**	**0.85**	**0.91**	**0.999**
		B2	1.07	0.49	0.20	0.83	**0.91**	0.998
		C	1.04	0.51	0.22	0.82	0.90	0.998
NBUT	0.62	A	**0.64**	0.41	0.43	0.68	0.80	0.989
		B1	0.59	**0.36**	**0.35**	**0.75**	**0.85**	**0.996**
		B2	0.59	**0.36**	0.36	0.74	**0.85**	0.995
		C	0.58	0.37	0.38	0.72	0.84	0.995
IBUT	0.33	A	0.18	0.37	2.21	0.55	0.59	0.960
		B1	**0.33**	**0.23**	**0.46**	**0.82**	**0.79**	0.972
		B2	**0.33**	0.28	0.69	0.71	0.78	0.978
		C	0.31	0.30	0.86	0.65	0.77	**0.984**
NPEN	0.30	A	0.25	0.25	0.86	0.48	0.57	0.939
		B1	**0.26**	**0.18**	**0.41**	**0.70**	**0.77**	**0.987**
		B2	0.25	0.20	0.51	0.65	0.71	0.974
		C	0.24	0.21	0.63	0.60	0.66	0.972
IPEN	0.51	A	0.45	0.42	1.39	0.42	0.44	0.875
		B1	**0.46**	**0.29**	**0.36**	**0.65**	**0.82**	**0.993**
		B2	0.45	0.31	0.42	0.60	0.78	0.989
		C	0.43	0.32	0.47	0.56	0.77	0.989
C_3H_6	0.18	A	0.06	0.21	4.37	0.45	0.27	0.653
		B1	**0.16**	**0.13**	**0.61**	**0.72**	**0.78**	**0.970**
		B2	0.15	0.15	0.77	0.66	0.71	0.964
		C	0.15	0.18	0.85	0.64	0.70	0.963
TOLU	0.46	A	**0.46**	0.39	0.73	0.47	0.65	0.953
		B1	0.43	**0.34**	**0.56**	**0.58**	**0.72**	0.957
		B2	0.43	0.35	0.60	0.56	0.70	0.959
		C	0.43	0.36	0.64	0.53	0.68	**0.960**
OXYL	0.09	A	0.055	0.11	2.21	0.37	0.49	0.879
		B1	**0.086**	**0.07**	**0.65**	**0.74**	**0.68**	**0.962**
		B1	0.083	0.09	0.97	0.60	0.63	0.950
		C	0.080	0.09	1.06	0.57	0.62	0.946
MPXYL	0.19	A	**0.17**	0.20	1.25	0.33	0.53	0.909
		B1	**0.17**	**0.15**	**0.71**	**0.59**	**0.65**	**0.956**
		B2	0.16	0.17	0.89	0.52	0.62	0.940
		C	0.16	0.17	1.01	0.47	0.57	0.919
ISO	0.31	A	0.18	0.86	13.47	0.63	0.35	0.708
		B1	**0.31**	**0.46**	**2.22**	**0.91**	0.45	0.716
		B2	0.36	0.59	3.18	0.82	**0.46**	0.789
		C	0.35	0.59	3.22	0.82	**0.46**	**0.795**
ACE	0.52	A	0.12	0.57	4.95	**0.69**	0.11	0.723
		B1	**0.49**	**0.34**	**0.45**	0.67	**0.72**	**0.995**
		B2	**0.49**	0.35	0.49	0.64	0.71	0.992
		C	0.47	0.38	0.59	0.57	0.69	0.989
C_2H_6	1.93	A	1.17	0.99	0.43	0.78	0.76	**1.000**
		B1	**1.83**	**0.53**	**0.08**	**0.86**	**0.99**	**1.000**
		B2	1.81	0.55	0.09	0.85	**0.99**	**1.000**
		C	1.76	0.59	0.10	0.83	**0.99**	**1.000**
C_2H_4	0.64	A	0.20	0.68	3.50	0.63	0.22	0.677
		B1	**0.61**	**0.33**	**0.28**	**0.85**	**0.81**	**0.987**
		B2	0.59	0.39	0.40	0.79	0.77	0.984
		C	0.58	0.43	0.49	0.74	0.76	0.983
BEN	0.47	A	0.42	0.31	0.47	0.69	0.74	**0.994**
		B1	**0.46**	**0.23**	**0.25**	**0.81**	**0.84**	0.977
		B2	**0.46**	0.26	0.31	0.77	0.82	0.992
		C	**0.46**	0.28	0.36	0.73	0.82	0.993

Table 6. For all species, the total emitted mass (in Gg) for the EMEP inventory run (case A), the a posteriori emissions under Gaussian assumption (cases B1 and B2) and the a posteriori emissions under the truncated Gaussian assumption (case C).

symbols	Case A	Case B1	Case B2	Case C
C_3H_8	7.4	18.3	18.4	15.1
NBUT	19.0	16.5	16.8	18.0
IBUT	4.7	10.3	9.1	6.1
NPEN	9.7	7.1	7.8	8.9
IPEN	9.2	11.4	10.8	10.1
C_3H_6	4.1	5.4	5.9	5.4
TOLU	12.7	11.1	11.4	12.1
OXYL	2.6	1.9	2.4	2.5
PXYL	5.7	3.5	4.4	5.4
MXYL	2.3	2.4	2.1	2.2
ISO	165.0	-925.5	124.9	143.4
ACE	2.1	10.6	9.9	4.8
C_2H_6	7.2	22.9	22.0	17.2
C_2H_4	7.9	19.1	20.1	14.5
BEN	5.6	7.1	6.7	5.8

cc BY

Table 7. Scores of the forecast test (year 2006) from the comparison of the observations and the simulated concentrations for four simulations: case A, case B1, case B2 and case C.

species	$\overline{O}$	case	$\overline{C}$	RMSE	NMSE	R	FA_2	FA_5
C_3H_8	1.10	A	0.55	0.85	1.19	0.76	0.46	0.958
		B1	**1.14**	0.63	**0.32**	**0.80**	0.85	0.992
		B2	**1.06**	0.61	**0.32**	0.78	**0.89**	**0.998**
		C	1.00	**0.59**	**0.32**	0.79	0.88	**0.998**
NBUT	0.56	A	0.75	0.45	0.49	0.72	0.70	0.977
		B1	0.38	0.39	0.40	0.78	**0.83**	0.994
		B2	**0.62**	**0.36**	**0.38**	**0.76**	0.80	**0.994**
		C	0.65	0.37	**0.38**	**0.76**	0.79	**0.994**
IBUT	0.34	A	0.21	**0.33**	1.46	**0.61**	0.67	0.967
		B1	**0.36**	0.46	1.73	0.49	0.56	0.825
		B2	0.37	0.35	**0.93**	0.55	**0.75**	**0.975**
		C	**0.32**	**0.33**	0.98	0.54	0.74	**0.975**
NPEN	0.29	A	**0.31**	0.31	1.08	0.45	0.56	0.919
		B1	0.18	0.30	1.72	0.48	0.39	0.692
		B2	0.26	0.28	1.03	0.51	0.61	0.949
		C	**0.27**	**0.27**	**0.94**	**0.52**	**0.62**	**0.950**
IPEN	0.47	A	0.29	**0.37**	0.99	0.49	0.54	0.906
		B1	0.37	0.62	2.17	0.31	0.43	0.733
		B2	**0.46**	0.46	0.99	0.42	0.64	0.963
		C	0.42	0.38	**0.73**	**0.49**	**0.65**	**0.969**
C_3H_6	0.19	A	0.08	0.24	3.86	0.44	0.39	0.674
		B1	0.07	0.24	4.40	0.51	0.24	0.521
		B2	**0.13**	**0.20**	**1.58**	**0.60**	**0.50**	**0.877**
		C	**0.13**	**0.20**	1.62	**0.60**	0.49	0.873
TOLU	0.66	A	**0.57**	0.69	**1.28**	0.28	0.55	**0.876**
		B1	0.47	0.73	1.65	0.25	0.56	0.852
		B2	0.48	0.70	1.55	0.27	0.58	0.864
		C	0.50	**0.68**	1.42	**0.29**	**0.60**	0.875
OXYL	0.07	A	**0.07**	**0.07**	1.05	0.42	0.54	0.885
		B1	0.04	0.08	2.35	0.42	0.36	0.693
		B2	0.06	**0.07**	0.98	**0.52**	0.54	**0.918**
		C	**0.07**	**0.07**	**0.95**	0.50	**0.56**	0.910
MPXYL	0.21	A	**0.21**	0.25	1.45	0.32	0.51	0.856
		B1	0.10	0.23	2.73	0.49	0.38	0.720
		B2	0.16	**0.21**	1.38	**0.50**	**0.53**	**0.885**
		C	0.18	0.22	**1.33**	0.44	**0.53**	0.863
ISO	0.31	A	**0.35**	1.04	10.25	0.21	0.26	0.571
		B1	1.39	19.95	939.9	-0.18	0.16	0.287
		B2	0.25	**0.81**	8.59	**0.46**	**0.34**	**0.682**
		C	**0.27**	**0.81**	**8.05**	0.44	0.33	0.656
ACE	0.68	A	0.14	0.83	7.17	0.57	0.17	0.649
		B1	**0.49**	**0.58**	**1.00**	**0.60**	**0.63**	**0.966**
		B2	0.47	0.62	1.22	0.51	0.54	0.952
		C	0.36	0.70	1.98	0.39	0.52	0.837
C_2H_6	1.89	A	1.15	1.05	0.51	0.78	0.80	0.998
		B1	**1.96**	0.76	**0.16**	**0.81**	**0.98**	0.999
		B2	1.78	0.76	0.17	0.78	0.97	**1.000**
		C	1.71	**0.75**	0.17	0.79	0.97	0.999
C_2H_4	0.66	A	0.26	0.87	4.43	0.59	0.36	0.751
		B1	0.39	0.74	2.15	0.65	0.39	0.713
		B2	**0.61**	**0.62**	**0.97**	**0.73**	0.64	0.978
		C	0.56	0.66	1.18	0.69	**0.65**	**0.980**
BEN	0.47	A	0.44	0.38	0.69	0.65	0.77	0.995
		B1	0.57	0.42	0.65	0.70	0.71	0.966
		B2	**0.47**	**0.35**	**0.55**	**0.71**	**0.81**	0.994
		C	0.46	0.36	0.59	0.69	0.80	**0.997**

Table 8. Scores at the Kollumerwaard station (for the year 2006) from the comparison of the observations and the simulated concentrations for three simulations: case A, case B2 and case C.

species	case	$\overline{O}$	$\overline{C}$	RMSE	NMSE	R	FA_2	FA_5
C_3H_8	A	1.69	0.82	2.12	3.24	0.05	0.39	0.799
	B2		**0.97**	**1.81**	**2.08**	0.43	**0.63**	**0.936**
	C		0.81	1.87	2.56	**0.47**	0.56	0.932
IBUT	A	0.33	**0.44**	**0.33**	1.32	**0.54**	0.60	0.868
	B2		0.41	0.39	1.19	0.34	0.60	0.927
	C		0.36	0.34	**0.98**	0.37	**0.62**	**0.948**
NPEN	A	0.42	**0.28**	0.42	**1.53**	0.52	0.50	0.899
	B2		0.18	**0.48**	2.99	0.48	0.45	0.859
	C		0.23	0.44	2.00	**0.53**	**0.52**	**0.917**
IPEN	A	0.61	**0.43**	**0.59**	**1.32**	**0.20**	**0.48**	**0.868**
	B2		0.37	0.63	2.04	0.13	0.41	0.770
	C		0.33	0.65	1.75	0.12	0.36	0.814
C_3H_6	A	0.17	0.08	0.30	6.10	0.65	0.36	0.719
	B2		**0.14**	**0.26**	**2.79**	0.63	**0.44**	**0.780**
	C		0.13	**0.26**	2.97	**0.66**	0.39	0.734
TOLU	A	0.53	**0.52**	0.43	**0.68**	0.48	**0.67**	0.886
	B2		0.35	0.35	0.64	0.72	0.61	0.937
	C		0.39	**0.33**	0.52	**0.73**	**0.67**	**0.942**
OXYL	A	0.11	**0.06**	0.10	**1.51**	0.52	**0.39**	**0.847**
	B2		0.05	0.10	1.76	0.59	0.36	0.818
	C		**0.06**	**0.09**	1.58	**0.60**	0.38	0.829
MPXYL	A	0.22	**0.15**	0.22	**1.44**	0.72	**0.39**	0.881
	B2		0.11	0.23	2.26	**0.75**	0.36	0.875
	C		0.14	**0.21**	1.48	**0.75**	0.38	**0.888**
ISO	A	0.18	**0.05**	**0.35**	14.3	**0.33**	0.19	0.445
	B2		**0.05**	**0.35**	**13.29**	0.32	**0.20**	**0.463**
	C		**0.05**	**0.35**	**13.29**	0.32	**0.20**	**0.463**
C_2H_4	A	0.18	0.14	0.44	4.33	**0.57**	0.24	0.609
	B2		**0.17**	0.45	3.58	0.44	**0.33**	**0.691**
	C		0.20	**0.40**	**2.42**	0.55	0.31	0.662
BEN	A	0.44	**0.54**	0.83	**2.92**	0.00	0.48	0.772
	B2		0.24	0.67	4.26	0.23	0.43	0.816
	C		0.30	**0.65**	3.25	**0.27**	**0.52**	**0.869**

cc BY

Table 9. The ratio of DFS to the cost function for each species.

symbols	case B1	case B2	case C
C_3H_8	0.076	0.055	0.108
NBUT	0.040	0.033	0.035
IBUT	0.142	0.052	0.067
NPEN	0.088	0.050	0.067
IPEN	0.072	0.044	0.069
C_3H_6	0.067	0.040	0.091
TOLU	0.045	0.030	0.031
OXYL	0.119	0.037	0.052
MPXYL	0.141	0.034	0.045
ISO	0.081	0.015	0.042
ACE	0.074	0.055	0.069
C_2H_6	0.080	0.060	0.131
C_2H_4	0.126	0.056	0.109
BEN	0.081	0.040	0.033

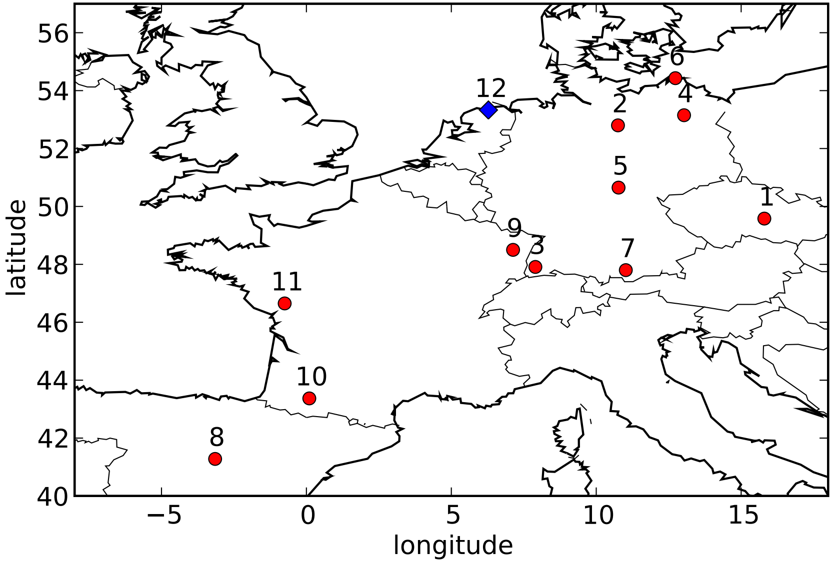

Fig. 1. The 11 monitoring stations of the EMEP network for volatile organic compounds whose observations are assimilated are indicated with a circle. The Kollumerwaard station in the Netherlands used for validation only is indicated by a rhombus.

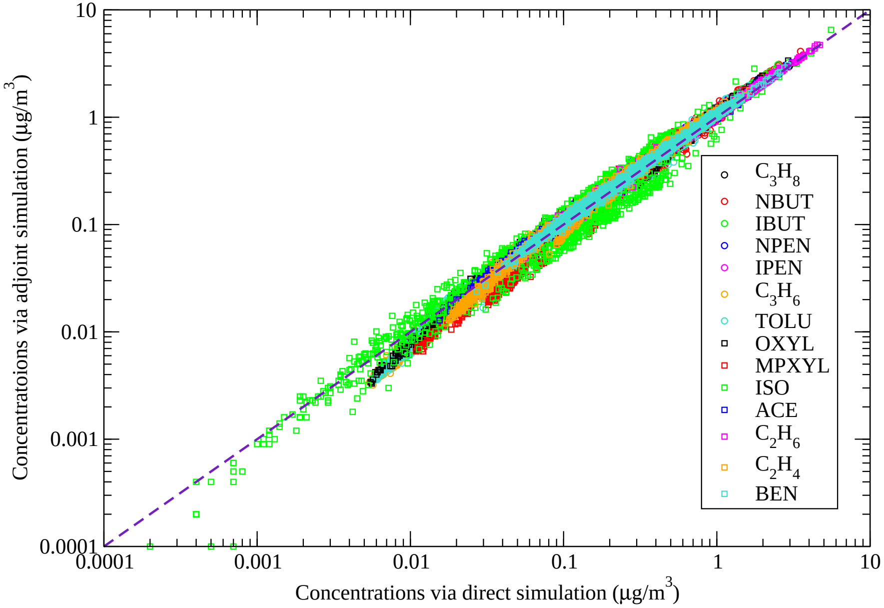

Fig. 2. Comparison of the source contributions estimated with the direct model and the adjoint model.

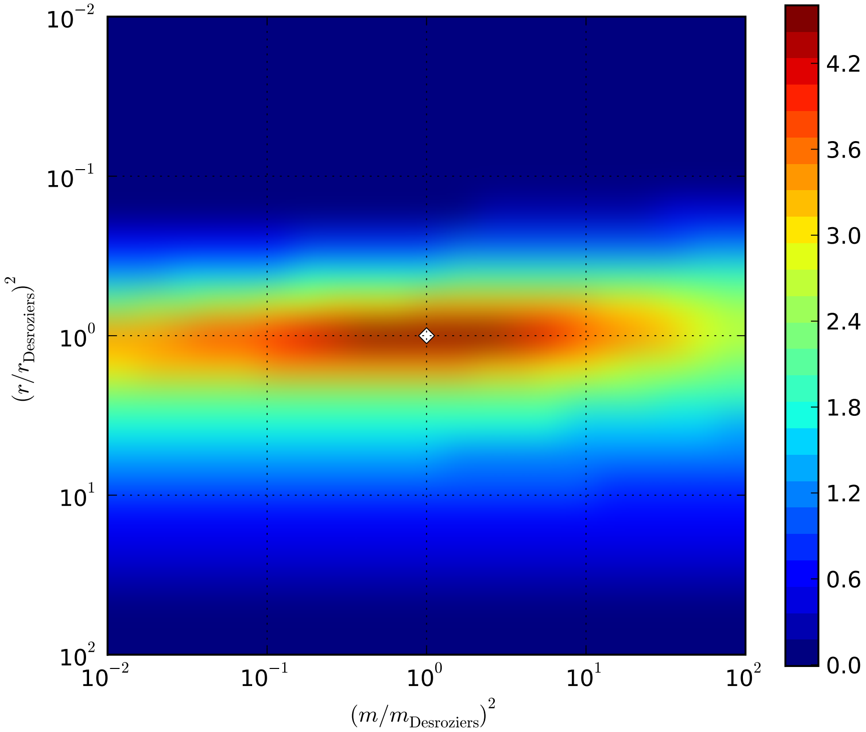

Fig. 3. This density plot displays a monotonic transform of the likelihood of the hyperparameters for NBUT in the Gaussian case. The monotonic transform is used to obtain a better contrast in the density plot. The abscissa and ordinate are normalised according to the optimal parameters obtained from the fixed-point method.

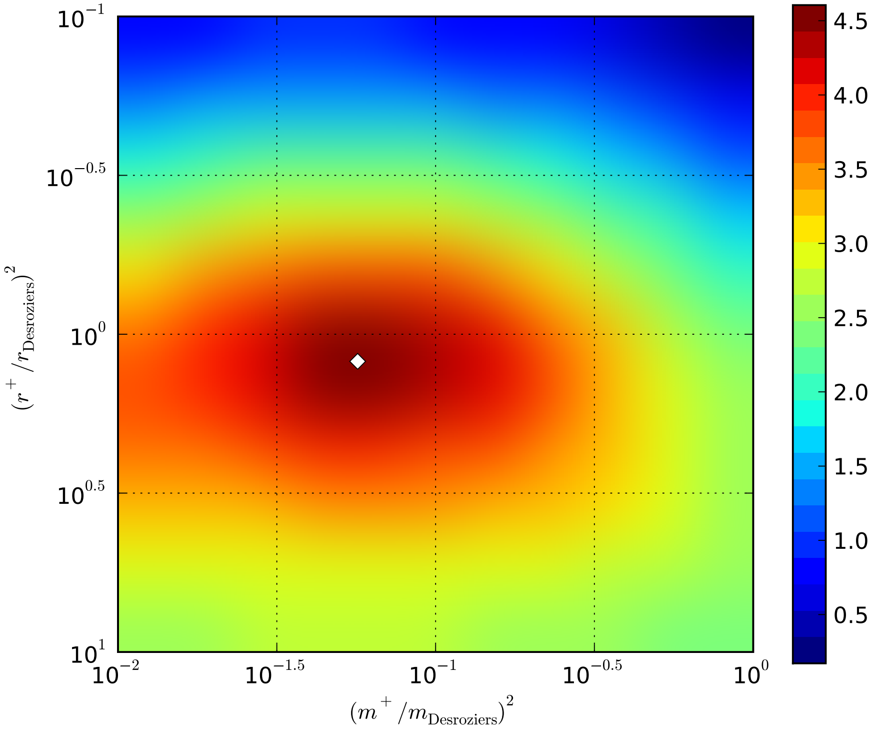

Fig. 4. A monotonic transform of the likelihood of the hyperparameters for C_2H_6 in the truncated Gaussian case. The abscissa and ordinate are normalised according to the fixed-point method.

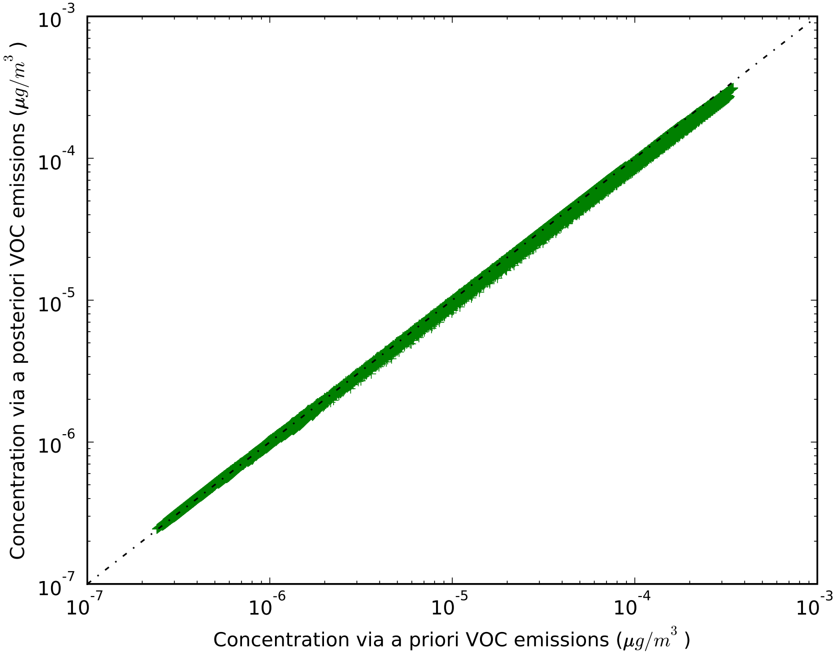

Fig. 5. Comparison between the a priori and a posteriori OH concentration fields.

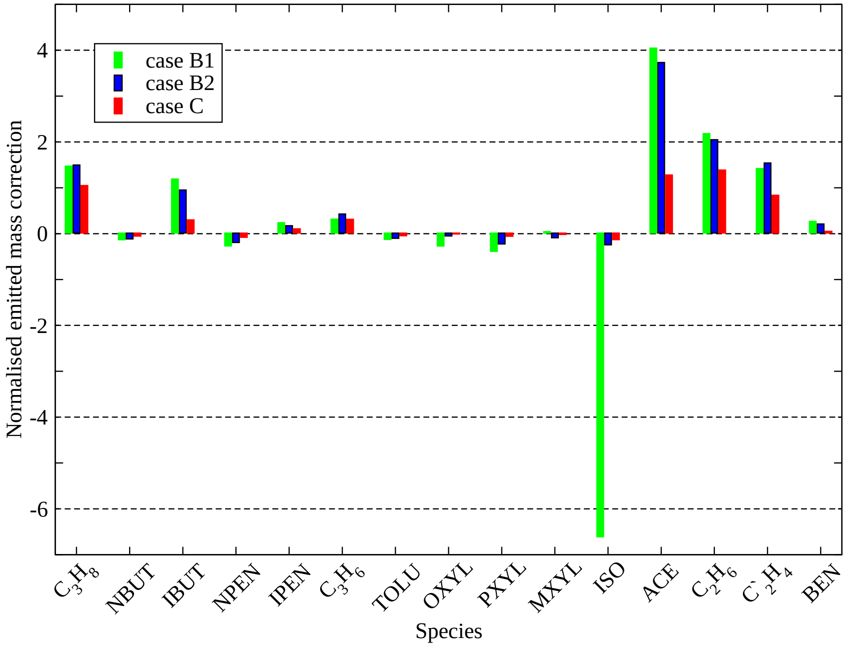

Fig. 6. The total emitted mass correction, normalised with respect to the total emitted mass of the EMEP inventory for cases B1, B2 and C.

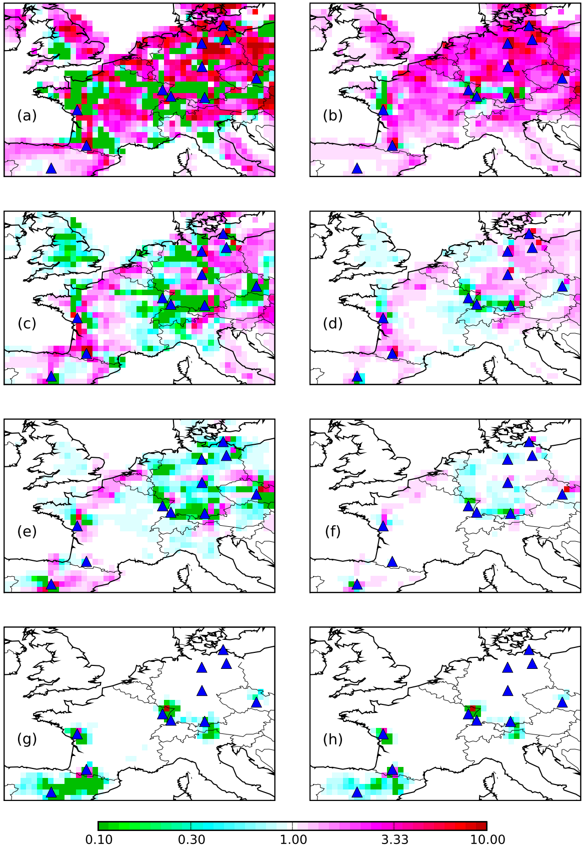

Fig. 7. Gridded ratios of time-integrated retrieved flux to EMEP time-integrated flux for C_2H_6, **(a,b)**, IPEN **(c,d)**, OXYL **(e,f)** and ISO **(g,h)**. Green and blue colours correspond to reduction of the emission fluxes, whereas red and pink colours correspond to increase of the emission fluxes. The left column **(a,c,e,g)** corresponds to case B2 and the right column **(b,d,f,h)** corresponds to case C. The species are ordered by decreasing lifetime.