Atmos. Chem. Phys., 13, 7115–7132, 2013
www.atmos-chem-phys.net/13/7115/2013/
doi:10.5194/acp-13-7115-2013

Atmospheric Chemistry and Physics
Open Access

Towards better error statistics for atmospheric inversions of methane surface fluxes

A. Berchet[1], I. Pison[1], F. Chevallier[1], P. Bousquet[1], S. Conil[2], M. Geever[3], T. Laurila[4], J. Lavrič[5], M. Lopez[1], J. Moncrieff[6], J. Necki[7], M. Ramonet[1], M. Schmidt[1], M. Steinbacher[8], and J. Tarniewicz[1]

[1]Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, UMR8212, IPSL, Gif-sur-Yvette, France
[2]Andra, DRD/OS, Observatoire Pérenne de l'Environnement, France
[3]National University of Ireland, Galway, Ireland
[4]FMI, Finnish Meteorological Institute, Helsinki, Finland
[5]Max Planck Institute for Biogeochemistry, Jena, Germany
[6]University of Edinburgh, Edinburgh, UK
[7]AGH University of Science and Technology, Krakow, Poland
[8]Empa, Laboratory for Air Pollution/Environmental Technology, Duebendorf, Switzerland

Correspondence to: A. Berchet (aberchet@lsce.ipsl.fr)

Received: 11 December 2012 – Published in Atmos. Chem. Phys. Discuss.: 11 February 2013
Revised: 10 June 2013 – Accepted: 18 June 2013 – Published: 29 July 2013

Abstract. We adapt general statistical methods to estimate the optimal error covariance matrices in a regional inversion system inferring methane surface emissions from atmospheric concentrations. Using a minimal set of physical hypotheses on the patterns of errors, we compute a guess of the error statistics that is optimal in regard to objective statistical criteria for the specific inversion system. With this very general approach applied to a real-data case, we recover sources of errors in the observations and in the prior state of the system that are consistent with expert knowledge while inferred from objective criteria and with affordable computation costs. By not assuming any specific error patterns, our results depict the variability and the inter-dependency of errors induced by complex factors such as the misrepresentation of the observations in the transport model or the inability of the model to reproduce well the situations of steep gradients of concentrations. Situations with probable significant biases (e.g., during the night when vertical mixing is ill-represented by the transport model) can also be diagnosed by our methods in order to point at necessary improvement in a model. By additionally analysing the sensitivity of the inversion to each observation, guidelines to enhance data selection in regional inversions are also proposed. We applied our method to a recent significant accidental methane release from an offshore platform in the North Sea and found methane fluxes of the same magnitude than what was officially declared.

1 Introduction

Quantifying the methane (CH_4) fluxes between the surface and the atmosphere, establishing their temporal variability and spatial distribution, and estimating the anthropogenic and natural contributions to these fluxes is critical for closing the present-day methane budget. One of the approaches used for this purpose, called the atmospheric inversion, assimilates information about atmospheric composition to infer surface fluxes. This type of top-down estimation relies on the assimilation of in-situ observations of atmospheric concentrations (Houweling et al., 1999, 2006; Hein et al., 1997; Pison et al., 2009; Bousquet et al., 2011, 2006; Bergamaschi et al., 2005) and/or of remote-sensing data from satellite-based instruments (e.g., Bergamaschi et al., 2009). Using observations for inversions at the global scale reduces the uncertainties on the mean CH_4 flux balances on large regions (typically a few millions of km^2 large). At the regional and mesoscales, high-resolution inversions potentially provide the spatial distribution of the fluxes, so that the characterisation of the processes involved can be improved (Bergamaschi et al., 2010).

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

Inversions at any scale depend on simulations of the atmospheric mixing and advection by Chemistry-Transport Models (CTMs) to estimate the influence of emissions and sinks on the atmospheric concentrations where they are measured. Whether they are based on coarse (Chen and Prinn, 2006; Hein et al., 1997), varying (Peylin et al., 2005) or high (Lauvaux et al., 2008; Sarrat et al., 2007) resolutions, all the CTMs suffer to a certain extent from uncertainties in reproducing the atmospheric concentrations. The uncertainties are due to the transport errors (Baker et al., 2006; Geels et al., 2007; Peylin et al., 2002; Ahmadov et al., 2007; Prather et al., 2008), to the assumption that a point observation can be compared to the mean simulated concentration on the corresponding grid box, i.e., the representation errors (Gerbig et al., 2003; Tolk et al., 2008), or to the errors from aggregating the fluxes on large regions (Kaminski et al., 2001).

In the framework of Bayesian atmospheric inversion (Enting et al., 1993; Tarantola, 1987), the implementation of a system requires obtaining an advanced understanding of the statistics of the observational and instrumental errors, the transport errors, the representation errors, and the errors of the prior distribution and magnitude of the fluxes prescribed in the system. Most of the cited works empirically assigned these error statistics. Objective methods of tuning the errors in the system also exist (Wahba et al., 1994; Dee, 1995; Desroziers and Ivanov, 2001) and have been applied to get the general structure of the errors (Michalak et al., 2005; Winiarek et al., 2012). But these methods rely on subjective prior knowledge on the error structure (e.g., isotropic spatial correlation or temporal decay in the correlations), which can limit the generality of the results.

In this study, we apply three different methods based on the statistical and algebraic properties of the errors, but with a minimum of additional physical assumptions on the error patterns. In inversion systems typically solving fluxes at the model resolution (e.g., $\sim 0.5^\circ \times 0.5^\circ$ each week during a season or a year in regional scale studies), this approach would require the handling of matrices of error covariances the total size of which exceeds billions or even trillions of components. To embrace memory limitations and reduce the computation costs, we have chosen a short window of inversion and have aggregated the surface fluxes on synoptic-scale regions. This simplification allows applying powerful generic methods, but induces limitations (Kaminski et al., 2001; Bocquet et al., 2011) that have to be taken into account when moving to a full-resolution inversion system.

Our study exploits a recent unexpected release of CH_4 in the North Sea in spring 2012 to apply this statistical approach and test the ability of a European network of atmospheric observations to detect the leak. On 25 March 2012, an offshore oil platform on the Elgin field, located 200 km east of Scotland shores (57° N, 1.53° E), was evacuated due to a gas leak. The company operating the platform gave a rough evaluation of the flux reaching 200 000 $m^3\,d^{-1}$ or 140 metric tons per day ($t\,d^{-1}$) for CH_4, which accounts for less than 1 % of the daily regional emissions (within a radius of $\sim$ 750 km around the leak point) according to the Emission Database for Global Atmospheric Research (EDGAR v4.2; http://edgar.jrc.ec.europa.eu) for the year 2008. The leak was stopped two months after. The methane plume emitted by this point source is difficult to extract from the observation noise and from the variability of the other sources, which makes the assignment of error statistics particularly critical (Winiarek et al., 2012). We develop and apply a regional inversion framework based on CHIMERE CTM simulations (Vautard et al., 2001) on a domain covering the European continent (Fig. 1). Relying on objective statistical criteria, we optimise the covariance matrices of the errors of the observations and of the prior state vector (surface fluxes, initial and lateral boundary conditions) for two independent time windows of inversion: the 2 weeks before the beginning of the leak and the 2 weeks after. The three methods of optimisation are implemented with acceptable computation times. They managed to produce error covariance matrices, which are specifically suited to the system and the inversion window, and that are a best guess of the optimum in regard to the objective criteria of each method. Complex error structures are then retrieved. And within the framework of the underlying assumptions, every piece of information provided by the observations and the prescribed fluxes is entirely used. We then use the computed matrices to invert the European fluxes before and after the leak start and test whether or not the atmospheric network detected this methane plume.

In Sect. 2, we describe the inversion methods and the dataset used in the study. We also present the algorithms that we implemented following the literature to specify the inversion system configuration. In Sect. 3, the results of these tuning methods are presented. The inversion results from these sets are analysed in Sect. 4 and their limitations and possible adaptation to larger systems are discussed in Sect. 5.

2 Methods

2.1 Inversion system

2.1.1 Theory: analytical framework

We apply classical data assimilation methods based on the Bayesian formalism (Courtier et al., 1994; Enting et al., 1993, 1995; Tarantola, 1987). In the following we use the unified notation by Ide et al. (1997). Assuming a Gaussian nature for all the errors, the method basically relies on the minimisation of the cost function:

$$\begin{aligned} J(\boldsymbol{x}) &= \tfrac{1}{2}(\boldsymbol{y}^0 - H(\boldsymbol{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\boldsymbol{y}^0 - H(\boldsymbol{x})) \\ &\quad + \tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{b}}) \\ &= J^{\mathrm{o}}(\boldsymbol{x}) + J^{\mathrm{b}}(\boldsymbol{x}) \end{aligned} \tag{1}$$

J^{o} (resp. J^{b}) is the contribution of the observations (resp. the background) to the total cost function. $\boldsymbol{y}^0$ accounts

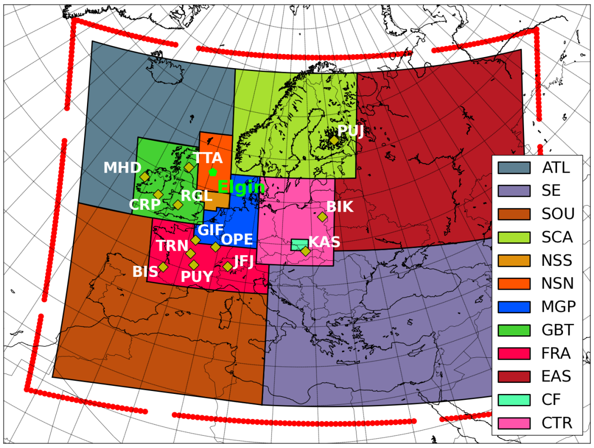


Fig. 1. Spatial extension of the simulations with the CTM CHIMERE: observation sites marked by yellow diamonds and white legends; the leak position in the northern part of the North Sea is pointed by the green pentagon. The fluxes, the spatial distribution of which is interpolated from EDGAR v4.2 inventory, are aggregated by regions figured by the coloured area. The red edges of the domain denote the spatial distribution of the lateral boundary condition components.

for the observation vector (dim $\boldsymbol{y}^0 \sim 5000$; description in Sect. 2.1.4); $\boldsymbol{x}$ is the state vector (i.e., the series of the variables to optimise), $\boldsymbol{x}^\mathrm{b}$ the background vector including the prior knowledge on the state of the system (dim $\boldsymbol{x} \sim 100$; description in Sect. 2.1.2).

H is the observation operator converting the information in the state vector to the observation space. The atmosphere advects and mixes the emissions and the initial and lateral boundary conditions. Since the time of residence of air masses within our domain is of typically two weeks, to be compared to the mean CH_4 lifetime in the atmosphere of about nine years (e.g., Dentener et al., 2003), the chemistry along the transport is neglected in all the following. For each observation $(\boldsymbol{y}^0)_i$ at a specified place and time, an equivalent $H(\boldsymbol{x})_i$ is induced from the state vector $\boldsymbol{x}$ with a CTM (description in Sect. 2.1.3). Throughout the study, H is supposed linear: the operator is represented by its Jacobian matrix $\mathbf{H}$ and $H(\boldsymbol{x})$ is identified to $\mathbf{H}\boldsymbol{x}$. With our relatively low-dimensional system, it is possible to explicitly estimate the observation operator with so-called "response functions". Response functions (Bousquet et al., 1999) are calculated for each component of the state vector by running the CTM in forward mode and extracting the simulated concentration at each point where a corresponding observation is available.

The covariance matrix $\mathbf{R}$ describes the errors $\boldsymbol{\epsilon} = \boldsymbol{y}^0 - \mathbf{H}\boldsymbol{x}$ between the observations and their reconstruction from the state vector with the model. We assume that the errors are unbiased, i.e., $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. $\mathbf{R}$ then encompasses the errors directly related to the measurement process, but also to the transport model (Ahmadov et al., 2007; Peylin et al., 2002), to the model representation, i.e., its inability to represent the local variability within the grid cells (Tolk et al., 2008; Geels et al., 2007) and to the aggregation process (Kaminski et al., 2001; Bocquet et al., 2011). $\mathbf{B}$ is the covariance matrix of the errors on the background vector $\boldsymbol{x}^\mathrm{b}$ (details in Sect. 2.1.2).

For this linear problem, the cost function J admits a global minimum reached for the optimum state vector $\hat{\boldsymbol{x}}^\mathrm{a}$ such that:

$$\hat{\boldsymbol{x}}^\mathrm{a} = \boldsymbol{x}^\mathrm{b} + \mathbf{K}(\boldsymbol{y}^0 - \mathbf{H}\boldsymbol{x}^\mathrm{b}) \tag{2}$$

where $\mathbf{K} = \mathbf{B}\mathbf{H}^\mathrm{T}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^\mathrm{T})^{-1}$ is the Kalman gain matrix.

The associated covariance matrix of errors, representing the a posteriori uncertainties in Gaussian assumptions, is given by:

$$\mathbf{P}^\mathrm{a} = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B}. \tag{3}$$

2.1.2 The state vector

In Eq. (1), $\boldsymbol{x}$ stands for the state vector which is optimised by the inversion and $\boldsymbol{x}^\mathrm{b}$ figures the assumed state (called the "background") before the assimilation of the observations. The vector $\boldsymbol{x}$ contains all the information on all the degrees of freedom of the system, e.g., on the emissions (spatial and temporal distribution), boundary conditions (all CH_4 concentration fields at the edges of the domain) and initial conditions (3-dimensional distribution of the CH_4 concentrations at the first step of the period of interest). In most realistic systems, dealing with the complete state vector implies prohibitive computational costs and/or unaffordable memory needs. Solving a high dimension system needs the implementation of variational algorithms (Chevallier et al., 2005). An alternative is to reduce the size of the inverse problem.

We drastically simplify the state and background vectors in order to allow the manipulation of the matrix $\mathbf{B}$. For each of the two windows of inversion (the two weeks before and the two after the leak start), computed independently, the simplified vector has a dimension of 99 and contains coefficients of linear corrections on: (1) the aggregated emissions on 12 regions (see Fig. 1); the spatial and temporal distributions of the emissions in each region are interpolated from the EDGARv4.2 database (http://edgar.jrc.ec.europa.eu); (2) the concentrations on the boundaries of the domain; boundaries are divided into 17 sub-parts: one for the top side, 16 for the lateral sides (8 horizontal parts $\times$ 2 vertical components per part; spatial horizontal distribution in Fig. 1; the vertical partition is situated at the half of the domain in pressure coordinates, i.e., at ~ 700 hPa); one coefficient is attributed per sub-part per period of 3 days; for each sub-part, boundary concentrations are supposed constant and uniform; the lateral boundary conditions (LBC) are then fixed by $(8 \times 2 + 1) \times 5$ periods $= 85$ coefficients; (3) the initial condition 3-dimensional concentration field (IC); the model is first vented with background boundary concentrations (extracted from global simulations with the CTM LMDz; Bousquet et al., 2011) and EDGAR emissions during 10 days before the period of inversion; and (4) an offset, constant and uniform along the whole domain; the prior offset was calculated from the available observations as an estimation of the background concentrations during the window of inversion; the initial and boundary concentrations are expressed as perturbations from this offset.

This simplification implicitly implies the hypothesis of pure correlation of the information within each aggregated region of the state vector (see Sect. 5.1). We chose an extended domain compared to the network coverage in order to cope with the spatial and temporal ill-representation of the LBC (Lauvaux et al., 2012).

2.1.3 Atmospheric transport model

We use the Eulerian mesoscale non-hydrostatic chemistry transport model CHIMERE for this study (Vautard et al., 2001). This model was developed in a framework of pollution simulations (Schmidt et al., 2001; Pison et al., 2007), but is also used for greenhouse gas studies (Broquet et al., 2011). We use here a regular horizontal grid of 50 km-side cells with 25 layers geometrically spaced from the surface to 450 hPa (~ 6000 m). The model time step varies dynamically from 4 to 6 min depending on the maximum wind speed in the domain. The model is an off-line model which needs meteorological fields as forcing. The forcing fields are deduced from interpolated meteorological fields from the European Centre for Medium-range Weather Forecast (ECMWF) with a horizontal resolution of $0.5^\circ \times 0.5^\circ$ every 3 h. The model is operated in a domain of limited area spanning over the whole continental Europe (roughly 24×10^6 km^2; see Fig. 1).

2.1.4 Observations

The study is based on the assimilation of measurements of the atmospheric composition. Concentrations of CH_4 are measured in-situ in 13 European sites (see Fig. 1; details in Table 1) at altitudes from sea level up to 3580 m a.s.l. and with different instrumentation and time resolution. Some stations are equipped with CRDS analysers (frequency of up to 1 Hz magnitude), whereas others are Gas Chromatographs. Hourly aggregates were used as input in $\boldsymbol{y}^0$ for the inversion. The observation sites can be split into three categories: (1) mountain sites which monitor almost all the time the free troposphere; they are scarcely influenced by the local emissions and are representative of the continental and global budget; (2) coastal sites with primary influence from the ocean when the air flows towards lands; as mountain sites, they are representative of global patterns; and (3) rural sites, inland but remote from anthropogenic emissions hot spots. All instruments are calibrated by tanks traceable to the NOAA 2004 CH_4 scale (Dlugokencky et al., 2005) with a calibration precision of ± 2 ppb.

2.2 Error configuration: description of the algorithms

In order to apply the Bayesian inversion framework, a perfect knowledge of the background and observation error statistics is needed. The tuple of covariance matrices ($\mathbf{R}$, $\mathbf{B}$) must then be established. Tuning and calculating optimal covariance matrices has long been of interest in data assimilation (e.g., Talagrand, 1998; Desroziers and Ivanov, 2001; Chapnik et al., 2004). Statistical studies on large sets of data are required to reach a sufficient threshold of information to get a reliable approximation of $\mathbf{R}$ and $\mathbf{B}$. In most cases, the sets of data are not available and the covariance matrices are built relying on physical considerations and an expertise on the observation and model behaviours (Bergamaschi et al., 2010). In this section, we describe different objective methods to infer the best tuple of $\mathbf{R}$ and $\mathbf{B}$ matrices: first, the Desroziers' scheme, second, the maximisation of the likelihood, third, observation space diagnostics. The Desroziers' scheme and the maximisation of the likelihood are computed on the subspace of the diagonal matrices for both $\mathbf{R}$ and $\mathbf{B}$, while the observation space diagnostics allow the recovery of full matrices.

Before further discussion in Sects. 3 and 4, the relevance of the 3 proposed methods is validated by a χ^2 test (see Sect. 2.2.1), used in many studies (e.g., Lauvaux et al., 2012; Winiarek et al., 2012; Peylin et al., 2002; Rayner et al., 1999).

2.2.1 Validation test: χ^2 distribution

It can be shown, within Gaussian assumptions, that for the state vector $\hat{\boldsymbol{x}}^a$ minimising the cost function J, $J(\hat{\boldsymbol{x}}^a) = J^o(\hat{\boldsymbol{x}}^a) + J^b(\hat{\boldsymbol{x}}^a)$ has the statistics of a χ^2 distribution with a mean equal to $d/2$, d being the total number of available observations. We then define a χ^2 index $\frac{2J(\hat{\boldsymbol{x}}^a)}{d}$ that shall be close to 1.

The index can be written:

$$\chi^2(\tilde{\mathbf{R}}, \tilde{\mathbf{B}}) = \frac{1}{d}\Big[(\boldsymbol{y}^0 - H(\hat{\boldsymbol{x}}^a))^T \tilde{\mathbf{R}}^{-1} (\boldsymbol{y}^0 - H(\hat{\boldsymbol{x}}^a)) + (\hat{\boldsymbol{x}}^a - \boldsymbol{x}^b)^T \tilde{\mathbf{B}}^{-1} (\hat{\boldsymbol{x}}^a - \boldsymbol{x}^b) \Big] \quad (4)$$

Though insufficient, this test provides a low-cost insurance that we got a well-defined tuple of covariance matrices. The three methods presented below are tested and validated in regard to this test. A deeper analysis of the algorithms' results is presented in Sect. 3.

2.2.2 Desroziers' scheme: subsets application

We describe here a method to roughly infer the shapes of $\mathbf{R}$ and $\mathbf{B}$ covariance matrices with a very low computation cost. It has been shown by Talagrand (1998) and Desroziers and Ivanov (2001) that for a given ($\boldsymbol{x}^b$, $\boldsymbol{y}^0$), for any subspace j

Table 1. Site characteristics. The altitudes of the sites are given as m above sea level (a.s.l.) and the inlet height is in m above ground level (a.g.l.). The sites are grouped into three categories relatively to the topography and their close environment: rural (R), mountain (M), coastal (C). * These sites are recent and still do not have related publications. [1] Observatoire Pérenne de l'Environnement.

Station	ID	Location			Inlet height (m)	Site type	Reference
		Lon (° E)	Lat (° N)	Alt (m)			
Bialystok	BIK	23.0	53.2	183	300	R	Popa et al. (2010)
Biscarosse	BIS	−1.2	44.4	120	47	C	Ahmadov et al. (2009)
Carnsore Point	CRP	−6.4	52.2	9	4	C	*
Gif-sur-Yvette	GIF	2.2	48.7	160	7	R	Lopez et al. (2012)
Jungfraujoch	JFJ	8.0	46.6	3580	5	M	Reimann et al. (2008)
Kasprowy	KAS	19.6	49.1	1989	0	M	Necki et al. (2003)
Mace Head	MHD	−9.9	53.3	25	15	C	Ramonet et al. (2010)
Obs. Pér. Environ.[1]	OPE	5.5	48.6	390	10 50 120	R	Ramonet et al.
Puijo	PUJ	27.0	62.0	232	84	R	Ramonet et al.
Puy-de-Dôme	PUY	3.0	45.8	1465	10	M	Lopez (2012)
Ridge Hill	RGL	−2.5	52.0	199	45 90	R	*
Trainou	TRN	2.11	48.0	131	5 50 100 180	R	Lopez et al. (2012)
Angus	TTA	−3.0	56.6	254	222	R	Vermeulen et al. (2005)

independent of the complementary space in the observation or state space, the optimal tuple of matrices ($\mathbf{R}$, $\mathbf{B}$) follows the expression below:

$$\begin{aligned} \mathrm{E}[J_j^{\mathrm{o}}] &= \frac{1}{2}\left[p_j - \mathrm{tr}(\mathbf{P}_j(\mathbf{HK})\mathbf{P}_j^{\mathrm{T}})\right] \\ \mathrm{E}[J_j^{\mathrm{b}}] &= \frac{1}{2}\left[n_j - \mathrm{tr}(\mathbf{P}_j(\mathbf{I}_n - \mathbf{KH})\mathbf{P}_j^{\mathrm{T}})\right] \end{aligned} \tag{5}$$

where tr(.) stands for the trace operator, $\mathrm{E}[J_j^{\mathrm{o}}]$ (resp. $\mathrm{E}[J_j^{\mathrm{b}}]$) is the expectation of the contribution at the maximum $\hat{\boldsymbol{x}}^{\mathrm{a}}$ to the cost function J of the independent subspace j of the observation (resp. state) space, $\mathbf{P}_j$ is a projector from the whole observation (resp. state) space to the subspace j and p_j (resp. n_j) the dimension of the subspace j. $\mathbf{I}_n$ stands for the identity matrix in the background space.

One cannot ascertain statistical independence between subspaces of the observation or of the state space before running the algorithm. To apply the method, we then make the following assumptions to divide the observation and the state spaces into 41 independent subspaces: (1) each measurement site is independent from all others, (2) at each site, all the observations during the afternoon (12:00–17:00) are gathered in one subspace (named "day" period in the following) and all the remaining observations during the morning and the night in a second subspace (named "night"); the planetary boundary layer (PBL) during the night is ill-represented by models; ensuing erroneous vertical mixing is expected to deteriorate the ability of the CTM to simulate realistic concentrations during the night; the shared cause of the enhanced errors justifies the assumed dependency of night-time observations; for similar reasons, we group the remaining observations during the afternoon with well-mixed PBL. (3) the LBC are independent from other state dimensions, (4) same with the IC, (5) same with the offset, (6) every aggregated region of emissions is independent from the others.

Doing so, we have 26 independent subspaces within the observation space (13 sites x day/night) and 15 for the state space (12 regions + LBC + IC + offset). Desroziers and Ivanov (2001) proposed an iterative tuning procedure that converges to a tuple that satisfies Eq. (5); we refer to this procedure as Desroziers' scheme (DS). Let us rewrite the cost function:

$$J_k(\boldsymbol{x}) = \sum_j \frac{1}{(s_{j,k}^{\mathrm{b}})^2} J_{j,k}^{\mathrm{b}}(\boldsymbol{x}) + \frac{1}{(s_{j,k}^{\mathrm{o}})^2} J_{j,k}^{\mathrm{o}}(\boldsymbol{x}) \tag{6}$$

where $s_{j,k}^{\mathrm{b}}$ and $s_{j,k}^{\mathrm{o}}$ denote the adapted weights for the subspace j at step k of the iterative procedure to balance the observations and the background in the cost function.

Desroziers' scheme is described by the following system of equations for every step k:

$$\begin{cases} (s_{j,k+1}^{\mathrm{b}})^2 = \frac{2J_{j,k}^{\mathrm{b}}(\hat{\boldsymbol{x}}^{\mathrm{a}})}{n_j - \mathrm{tr}(\mathbf{P}_j(\mathbf{I}_n - \mathbf{K}_k\mathbf{H})\mathbf{P}_j^{\mathrm{T}})} \\ (s_{j,k+1}^{\mathrm{o}})^2 = \frac{2J_{j,k}^{\mathrm{o}}(\hat{\boldsymbol{x}}^{\mathrm{a}})}{p_j - \mathrm{tr}(\mathbf{P}_j(\mathbf{HK}_k)\mathbf{P}_j^{\mathrm{T}})} \\ \mathbf{B}_{j,k+1} \quad = s_{j,k+1}^{\mathrm{b}}\mathbf{B}_{j,0} \\ \mathbf{R}_{j,k+1} \quad = s_{j,k+1}^{\mathrm{o}}\mathbf{R}_{j,0} \end{cases} \tag{7}$$

$\mathbf{B}_j$ and $\mathbf{R}_j$ stand for the diagonal sub-matrices associated to the subset j. $\mathbf{K}_k$ is the Kalman gain matrix calculated with $(\mathbf{R}_k, \mathbf{B}_k)$ (see Eq. 2). The method implicitly relies on the χ^2 distribution of the cost function. The computed tuples then converges to a tuple filling the χ^2 criterion.

We start this algorithm from a tuple $(\mathbf{R}_0, \mathbf{B}_0)$ with no physical assumption, i.e., $\mathbf{R}_0 = \mathbf{I}_d$ and $\mathbf{B}_0 = \mathbf{I}_n$. We then stop the iterative scheme when every subset contribution to the cost function $E[J_j^{\mathrm{o}}]$ or $E[J_j^{\mathrm{b}}]$ is less than 1 % away from its theoretical expected value (right member of Eq. 5); this is equivalent to get a χ^2 test greater than 99 %. The algorithm converges in no more than 15 steps (i.e., a couple of minutes on a standard office computer).

2.2.3 Maximum of likelihood

Desroziers' scheme (DS) relies on coarse approximations and, for example, cannot extract the variability of the observational errors day-by-day and hour-by-hour. The following method allows the computation of a tuple ($\mathbf{R}$, $\mathbf{B}$) which is tuned not by block but by component individually. This improvement implies computational cost drastically higher than for DS method, but still affordable (less than a day) in our case study. In Gaussian assumptions, the likelihood of the observations $\boldsymbol{y}^0$ for given $\mathbf{R}$ and $\mathbf{B}$ can be written as follows (Michalak et al., 2005):

$$p(\boldsymbol{y}^0|\mathbf{R},\mathbf{B}) = \frac{e^{-\frac{1}{2}(\boldsymbol{y}^0-\mathbf{H}\boldsymbol{x}^{\mathrm{b}})^{\mathrm{T}}(\mathbf{R}+\mathbf{HBH}^{\mathrm{T}})^{-1}(\boldsymbol{y}^0-\mathbf{H}\boldsymbol{x}^{\mathrm{b}})}}{\sqrt{(2\pi)^d|\mathbf{R}+\mathbf{HBH}^{\mathrm{T}}|}}$$

The function diverges to infinity when $\mathbf{R}+\mathbf{HBH}^{\mathrm{T}} \sim (\boldsymbol{y}^0 - \mathbf{H}\boldsymbol{x}^{\mathrm{b}})(\boldsymbol{y}^0 - \boldsymbol{H}\boldsymbol{x}^{\mathrm{b}})^{\mathrm{T}}$. But when supposing that $\mathbf{R}$ and $\mathbf{B}$ are diagonal definite positive, $\mathbf{S}$ cannot be written as a matrix $\mathbf{R}+\mathbf{HBH}^{\mathrm{T}}$. Hence, in these assumptions, one can prove that the function is bounded and admits a computable maximum (Burg et al., 1982). A proper $(\mathbf{R}, \mathbf{B})$ tuple for the inversion system is necessarily a maximum of the function (Dee, 1995).

For memory limitation reasons, we do not maximise the function itself, but equivalently its logarithm:

$$\ln\ p(\boldsymbol{y}^0|\mathbf{R},\mathbf{B}) = -\frac{1}{2}\mathrm{tr}(\mathbf{S}_{\mathbf{R},\mathbf{B}}^{-1}\mathbf{S}) - \frac{1}{2}\ln|\mathbf{S}_{\mathbf{R},\mathbf{B}}| + C$$

with $\mathbf{S}_{\mathbf{R},\mathbf{B}} = \mathbf{R}+\mathbf{HBH}^{\mathrm{T}}$, $\mathbf{S} = (\boldsymbol{y}^0-\mathbf{H}\boldsymbol{x}^{\mathrm{b}})(\boldsymbol{y}^0-\boldsymbol{H}\boldsymbol{x}^{\mathrm{b}})^{\mathrm{T}}$ and C a constant not relevant for computing the maximum of the function. $|.|$ stands for the determinant operator.

The function maximum cannot be easily computed analytically in general. Hence, we use an ascending pseudo-Newtonian method based on the calculation of the gradient of the log-likelihood. The algorithm converges to a local maximum (Chapnik et al., 2004), but we have no insurance of converging to the global maximum. The result of the algorithm can be very dependent from the $(\mathbf{R}, \mathbf{B})$ tuple chosen as a starting point. To ensure the robustness of the result, we test this method with 2 different starting tuples: (1) one constructed relying on expert knowledge (e.g., diagonal background matrix with variances consistent with inventories specification), (2) the other with uniform errors of 50 ppb for observations, the LBC and the offset and 10 % for the emissions. The results are very similar: the difference between the two errors related to an observation j does not exceed 5 % and is less than 1 % in average.

Additionally, to accelerate the convergence of the algorithm, one can notice (details in Burg et al., 1982) that the log-likelihood is maximum only if $(\mathbf{R}, \mathbf{B})$ satisfies:

$$\alpha = \frac{\mathrm{tr}(\mathbf{S}_{\mathbf{R},\mathbf{B}}^{-1}\mathbf{S})}{d} = 1$$

Consequently, we force α to stay close to 1 at each step of the algorithm by rescaling $(\mathbf{R}, \mathbf{B})$ by α. This normalisation is equivalent to the χ^2 test; hence the constrained maximum of likelihood algorithm necessarily fulfills the χ^2 test.

This method makes the observation and background error variances statistically consistent with the prior difference between $\boldsymbol{y}^0$ and $\mathbf{H}\boldsymbol{x}^{\mathrm{b}}$. When assuming that $\mathbf{R}$ and $\mathbf{B}$ are diagonal, a unique set of variances matches this criterion on $\boldsymbol{y}^0 - \mathbf{H}\boldsymbol{x}^{\mathrm{b}}$. Hence, a maximum total amount of d pieces of information is used in the algorithm, while more are available in the background and the observation operator for the subsequent computation of $\hat{\boldsymbol{x}}^{\mathrm{a}}$. A more precise and complete quantification of the balance between the information used for the optimisation of the matrices and for the inversion itself is difficult in a real case study. A dedicated OSSE (Observing System Simulation Experiment) could improve our knowledge in this direction.

2.2.4 Observation space diagnostics

With the two previous methods, $\mathbf{R}$ and $\mathbf{B}$ are confined to the sub-space of diagonal matrices. But the errors on the observations are known to be correlated through the $\mathbf{H}$ operator errors amongst others. Errors on the background are also correlated, mainly because of shared errors in the inventory methods and in flux process modelling. We carry out in this section an algorithm to produce a setup of non-diagonal matrices $\mathbf{R}$ and $\mathbf{B}$.

Inquiring into error correlations requires a huge amount of information. The available information is not sufficient to characterise deterministically the full non-diagonal matrices $\mathbf{R}$ and $\mathbf{B}$. This section must then be seen as a way to infer guidelines for covariance building.

Design on R

Desroziers et al. (2005) showed that the innovation vectors $\hat{\boldsymbol{d}}_{\mathrm{a}}^{\mathrm{o}} = \hat{\boldsymbol{y}}^0 - \mathbf{H}\hat{\boldsymbol{x}}^{\mathrm{a}}$ and $\hat{\boldsymbol{d}}_{\mathrm{b}}^{\mathrm{o}} = \hat{\boldsymbol{y}}^0 - \mathbf{H}\hat{\boldsymbol{x}}^{\mathrm{b}}$ should fulfill the following equation:

$$\mathrm{E}[\hat{\boldsymbol{d}}_{\mathrm{a}}^{\mathrm{o}}(\hat{\boldsymbol{d}}_{\mathrm{b}}^{\mathrm{o}})^{\mathrm{T}}] = \mathbf{R} \tag{8}$$

Equation (8) is valid if and only if (1) the observation operator is linear, (2) the errors are Gaussian and unbiased, and (3) the matrix $\mathbf{HBH}^{\mathrm{T}}(\mathbf{R}+\mathbf{HBH}^{\mathrm{T}})^{-1}$ is consistent with the statistics of the background and observation errors.

In order to build a matrix $\mathbf{R}$ consistent with Eq. (8), we test an iterative algorithm similar to Desroziers' scheme using the following instruction for every k:

$$\begin{cases} \mathbf{R}'_{k+1} &= \mathrm{E}[\hat{\boldsymbol{d}}^{\mathrm{o}}_{\mathrm{a}}(\hat{\boldsymbol{d}}^{\mathrm{o}}_{\mathrm{b}})^{\mathrm{T}}] \\ \mathbf{R}_k+1 &= \chi^2(\mathbf{R}'_{k+1},\mathbf{B}_k)\cdot\mathbf{R}'_{k+1} \\ \mathbf{B}_k+1 &= \chi^2(\mathbf{R}'_{k+1},\mathbf{B}_k)\cdot\mathbf{B}_k \end{cases} \tag{9}$$

The tuple is normalised by the associated χ^2 test in order to constrain better the algorithm.

The expectation could be calculated explicitly as a combination of $\mathbf{R}_k$, $\mathbf{B}_k$ and $\mathbf{H}$. However, as suggested by Desroziers et al. (2005), the exact iterative scheme is not expected to converge. In our case, we indeed found no convergence. Therefore, we use a Monte Carlo evaluation of the expectation with 50 000 perturbations of $\boldsymbol{y}^0$ and $\boldsymbol{x}^{\mathrm{b}}$ with Gaussian distribution of covariances $\mathbf{R}_k$ and $\mathbf{B}_k$. The tuples of covariance matrices generated with the Monte-Carlo algorithm have likelihood values bigger (hence closer to the maximum of likelihood) that the one inferred from the explicit algorithm. Their likelihood (-480) is also significantly higher than with the diagonal tuple calculated in Sect. 2.2.3 ($\sim -15\,000$). The non-diagonal tuple is then more 'likely' in the sense of Sect. 2.2.3 compared with the diagonal one.

Any inconsistency in the requirements of Eq. (8) will make the expectation non-symmetric positive definite. In particular, biases in the observations carry information that is considered by the algorithm as random errors. The built matrices are then partly influenced by the intrinsic biases in the system. However, a simple diagnostic on the spectrum shows that after 50 000 perturbations, only less than 0.1 % of the Eigen values are negative. The biases that create asymmetry in the matrix then do not seem to significantly impact our case. This could not always apply to other frameworks. The limitation of the algorithm is to build up statistics from a single occurrence of $\hat{\boldsymbol{y}}^0$ and $\hat{\boldsymbol{x}}^{\mathrm{b}}$ and therefore is not fully generic.

But since asymmetry does not seem to be critical in our case, we rebuild a symmetric semi-definite positive matrix by correcting $\mathbf{R}'_{k+1}$ spectrum. We start the algorithm from two different tuples of matrices: the one which maximises the log-likelihood in the diagonal assumption, and another with the same $\mathbf{B}$ but with $\mathbf{R}=\mathbf{I}_{\mathrm{d}}$. The convergence is slower with the second starting tuple but the two optimum tuples are similar.

Calculating a matrix expectation of dimension 5000×5000 with only 50 000 Monte-Carlo perturbations is very unstable. Additionally, the perturbations on the observations are generated from $\boldsymbol{y}^0$, whereas it should be computed from the unknown perfect unperturbed observation vector $\tilde{\boldsymbol{y}}$. Hence, the closer $\mathbf{R}_k$ gets to the optimal value, the weaker is the approximation based on $\boldsymbol{y}^0$. For this reason, after a few steps of improvement of the log-likelihood value, the algorithm quickly diverges. We then keep the last step before the beginning of the divergence as a guess for the true covariance matrices tuple.

Design on B

Desroziers et al. (2005) also reported an equation constraining $\mathbf{B}$ in the observation space:

$$\mathrm{E}[\hat{\boldsymbol{d}}^{\mathrm{a}}_{\mathrm{b}}(\hat{\boldsymbol{d}}^{\mathrm{o}}_{\mathrm{b}})^{\mathrm{T}}]=\mathbf{HBH}^{\mathrm{T}} \tag{10}$$

with the innovation vector $\hat{\boldsymbol{d}}^{\mathrm{a}}_{\mathrm{b}}=\mathbf{H}(\hat{\boldsymbol{x}^{\mathrm{a}}}-\hat{\boldsymbol{x}^{\mathrm{b}}})$. The expectation is again based on a Monte-Carlo estimation. We compute an iterative scheme similar to that developed in Eq. (9) with also a normalisation by the χ^2 test. We use the tuple calculated above as a starting point.

Equation (10) relies on the same assumptions as Eq. (8). In particular, bias and other mismatches between $\mathbf{HBH}^{\mathrm{T}}(\mathbf{R}+\mathbf{HBH}^{\mathrm{T}})^{-1}$ and the correct statistics can induce inconsistent asymmetries that must be corrected. Another critical point in building $\mathbf{B}$ from constraints in the observation space is that the computed expectation could be outside the ensemble $\{\mathbf{HBH}^{\mathrm{T}}\ /\ \mathbf{B}$ symmetric definite positive$\}$. We then project the expectation onto this structure to recover a compatible $\mathbf{B}$ matrix.

Because of the Monte Carlo and the indirect estimation of $\mathbf{B}$, the instability is even sharper than for the computation of $\mathbf{R}$. We then do a unique iteration of the algorithm to get an evaluation of the potential optimal correlation coefficients in $\mathbf{B}$.

3 Results

We run the three algorithms described in Sect. 2.2 to infer a best guess for the optimum tuple ($\mathbf{R}$, $\mathbf{B}$) of the error covariance matrices. In the following section, we describe the shape of the calculated matrices in regard to known physical patterns of errors. We will refer to the diagonal tuple computed from the Desroziers' Scheme (resp. the Maximum of log-likelihood) as ($\mathbf{R}_{\mathrm{DS}}$, $\mathbf{B}_{\mathrm{DS}}$) (resp. ($\mathbf{R}_{\mathrm{ML}}$, $\mathbf{B}_{\mathrm{ML}}$)); the non-Diagonal tuple from the observation space algorithm will be referred to as ($\mathbf{R}_{\mathrm{ND}}$, $\mathbf{B}_{\mathrm{ND}}$).

3.1 Patterns in the error variances for the 3 methods

In Fig. 2, the variances of the observation errors in the 3 methods are compared for the period after the leak start. For the ML and ND algorithms, the variances of the observation errors are averaged along the same subspaces as in DS method (see Sect. 2.2.2) to be comparable. Most of the observation errors remain within the same interval (5–20 ppb) with the 3 methods. Their magnitudes are comparable with other

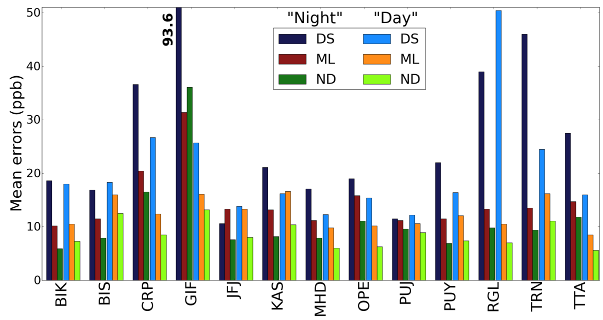


Fig. 2. Comparison of the observation errors calculated by the 3 methods presented in Sect. 2.2 for the inversion window after the leak start (25 March 2012). The errors are averaged on two periods each day: "day" (12:00 to 17:00) and "night" (the remaining hours of the day) for comparison to DS method. Values out of the graph window are printed on the matching bar.

studies which build the errors from physical considerations (e.g., Bergamaschi et al., 2010). At most sites (GIF, JFJ and KAS excepted), the 3 methods attribute averaged observation errors that follow the same order: DS > ML > ND. The errors from $\mathbf{R}_{\mathrm{ML}}$ are in average 34 % smaller than the errors in $\mathbf{R}_{\mathrm{DS}}$. The error variances in $\mathbf{R}_{\mathrm{ND}}$ are calculated to be even smaller (54 % less than in $\mathbf{R}_{\mathrm{DS}}$ in average). In the Bayesian unbiased framework, the inversion with the tuple from the maximum of likelihood is then expected to be more constrained by the observations than the one from Desroziers' scheme. The non-diagonal tuple seems more constrained by the observations than the other two, but the covariances make it difficult to precisely compare only the variances. The 3 methods share the same day/night patterns at all sites apart from the 3 mountain sites (JFJ, KAS and PUY) and the two sites BIS and BIK: compared to the errors during the "night" (17:00–00:00 plus 00:00–12:00), the errors during the "day" (12:00–17:00) are 25 % (resp. 23 % and 31 %) smaller for the DS (resp. ML and ND) method. The errors are consistently smaller when the PBL is well developed, i.e., when the local emissions are quickly mixed in the atmosphere, and hence when the CTM more realistically simulates the atmospheric concentrations. At the mountain sites (JFJ, KAS and PUY; see Table 1), the rough DS method does not calculate the same patterns as the other two; this primarily suggests that, for the sites mostly located in the free troposphere in spring (characterised by synoptic variability), the averaging on "day" and "night" intervals is less relevant than for the sites influenced by the PBL. Additionally, in mountain sites, the low-precision DS method disagrees with the other two because it cannot compute the errors that occur when the PBL height is close to the site altitude and when polluted air masses can be locally uplifted to the site. This phenomenon occurs at time scales that are smaller than and not synchronized with the partitioning made in Desroziers' scheme, that makes the DS method unable to accurately detect this source of error. On the opposite side, the two other methods which are handled individually each observation can detect this phenomenon and take it into account in the error modelling, giving averaged errors that are consistent with each other.

Regarding the background, the attributed errors partly balance the confidence attributed to the observations. For the two comparable diagonal algorithms (ML and DS), the ML method better optimises the use of the observations than DS does on large subspaces. To avoid corrections dominated by the observations, background errors 7 times smaller than for the DS algorithm are then computed by the ML algorithm. On the other hand, no clear general pattern appears with the tuple ($\mathbf{R}_{\mathrm{ND}}$, $\mathbf{B}_{\mathrm{ND}}$), compared with the other two. While the errors on the observations are on average smaller than the ones calculated with the ML method, the background errors with the ND method are of the same magnitude as the DS ones, i.e., higher than the ML ones. With non-diagonal covariance matrices, the variances by themselves are insufficient to identify all the properties of the errors. The non-diagonal elements in $\mathbf{R}_{\mathrm{ND}}$ and $\mathbf{B}_{\mathrm{ND}}$, characterising correlations of errors, must be taken into account to understand the error patterns. Strong negative (resp. positive) correlations are expected to increase (resp. decrease) the average confidence in the background. We analyse the effects of these correlations more precisely in Sect. 3.3.

3.2 Temporal variability and diurnal cycle of the observation errors

We focus here on the hour-by-hour variability of the observation errors; as a consequence, the DS method is not commented in this section. The variability in the variances is comparable in $\mathbf{R}_{\mathrm{ML}}$ and $\mathbf{R}_{\mathrm{ND}}$ ($r = 0.89$; linear regression coefficient: $\lambda = 0.97$). While no physical assumptions have been added to the algorithms, we notice that the observation errors follow known physical patterns for both algorithms. One of these is the source of errors related to the PBL height mis-estimation and the vertical mixing parameterization in CTMs. In Table 2, we show for each station the linear correlations between the modelled PBL heights and the calculated error variances for ML method. We also compute a logarithm transformation that shows that some PBL errors follow an exponential decay: $\exp(-h_{\mathrm{PBL}}/h_0)$, with h_{PBL} the modelled PBL height and h_0 a reference height. The magnitude of these calculated reference heights is 1000 m for the sites with significant correlations; above this reference threshold, the PBL can be approximately considered as well mixed. Logarithmic correlations are of the same magnitude in ND method at the sites with small p values ($p < 10^{-2}$; e.g., logarithm correlation: at CRP $r = -0.64$, at GIF $r = -0.57$, at MHD $r = -0.37$). The logarithm correlations are stronger than the linear ones in the sites where the p value is small ($< 10^{-2}$), except at RGL where the linear fit is better. For these sites (including RGL), we approximate the errors by the exponential

Table 2. Linear and logarithmic correlations (r) between calculated error standard deviation and the simulated PBL height (given by ECMWF). At each site, the calculation of r and the p value are carried out with **R** from ML method. Some p values are very small because calculated on large samples of observations. Error contribution (μ) for log-regression is the mean ratio between the fit and the computed errors, when the correlation is significant.

Stat	Linear regression		Log-regression		
	r	p	r	p	μ
BIK	0.10	0.06	0.8	0.16	–
BIS	0.13	0.01	0.11	0.04	–
CRP	−0.33	2.6×10^{-10}	−0.60	1.7×10^{-35}	60 %
GIF	−0.34	1.3×10^{-8}	−0.50	4.3×10^{-18}	50 %
JFJ	−0.04	0.5	−0.05	0.36	–
KAS	−0.01	0.8	−0.10	0.08	–
MHD	−0.41	2.9×10^{-15}	−0.47	2.1×10^{-19}	52 %
OPE	−0.31	5.6×10^{-22}	−0.36	1.5×10^{-28}	54 %
PUJ	−0.04	0.5	0.05	0.33	–
PUY	0.06	0.5	0.01	0.84	–
RGL	−0.17	9.1×10^{-6}	−0.12	0.002	58 %
TRN	0.21	7.9×10^{-11}	0.02	0.59	–
TTA	−0.64	2.1×10^{-22}	−0.72	2.3×10^{-29}	41 %

fit and estimate the relative contribution of the PBL error as the ratio between the fit error and the total error. The PBL error then seems to account for about half of the total error.

The strong correlations between the PBL height and the errors on the observations can also be partly related to systematic errors (especially during the night) in the CTM, which are difficult to distinguish from random ones based on the available data. The systematic nature of some errors can conflict with the hypothesis of random unbiased errors in the inversion framework: $\boldsymbol{\epsilon} = \mathbf{y}^0 - \mathbf{H}\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ in the unbiased framework would become $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\eta}, \mathbf{R}')$ with a systematic bias $\boldsymbol{\eta}$. Our methods will then try to find $\mathbf{R}$ such as $\mathbf{R} \sim \mathbf{R}' + \boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}$. Though statistically consistent with all the available pieces of information, such a biased formulation of the problems will have critical effect on the inversion results (e.g., Dee, 1995). Therefore, the biases have to be taken into account and our method seems to provide an efficient tool to detect them (e.g., high redundant errors associated to night-time vertical mixing) in order to potentially fix them before the inversion.

In Fig. 3, the statistics (median and interquartile range) of the errors are displayed by layers of 250 m for h_{PBL} for the sites with significant correlations as calculated in Table 2 on the one side and for the mountain sites on the other side. The anti-correlation between h_{PBL} and the errors is confirmed and a more complex behaviour appears for mountain sites. For example, the relation to the PBL is inverted for the Kasprowy mountain site (KAS; see Fig. 3): the site is set on the summit of a mountain ridge in vicinity of a region with high CH_4 emissions due to coal mining. The resolution of

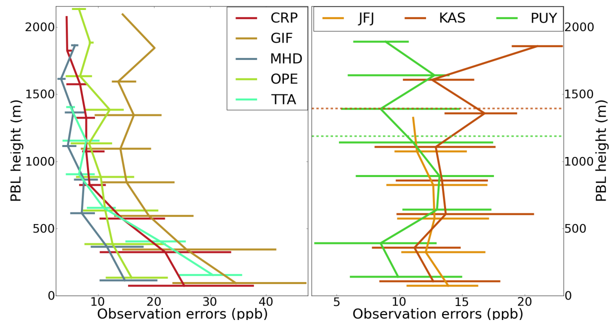


Fig. 3. Statistics of the errors projected along ECMWF-simulated PBL height for ML method: median and inter-quartile gap per 250 m-high layer. (Left) sites with strong correlations as calculated in Table 2; not displayed site RGL exhibits the same patterns but with higher errors. (Right) Mountain sites with influence from the PBL less prevailing. Dashed lines refer to the station altitude in the model; JFJ is above the maximum simulated PBL height.

the model makes the simulation of the upward winds from the valley impossible. Then, an observation can be noticeably influenced by polluted air masses from the PBL while the model considers the observation to be in the free troposphere, and inversely; that explains the peak in errors when the modelled h_{PBL} is close to the altitude of the site in the CHIMERE model: ~ 1400 m above the local pixel orography level. The mountain site at Puy-de-Dôme (PUY) appears to pose the same problem in other studies (Broquet et al., 2011); but in the particular meteorological situation of our window of inversion (2 weeks in spring), the issue does not clearly appear.

Another identified source of errors (not shown in the figures) is the temporal and spatial mismatches which can occur in situations of steep gradients of concentrations when the air masses are changing, like in frontal systems. Air mass changes occur 0 to 3 times during each inversion window of two weeks at the sites. The quantification of these errors is uncertain during night-time since it is difficult to separate the errors related to the PBL and the errors due to the gradients (e.g., mismatches or numerical diffusion). So, we focus on air mass changes occurring during daytime. In these cases, the computed errors exceed 4 times the mean error during daytime; the increased errors are then directly attributable to the inability of the model to simulate high-resolution phenomena. The relationship between the observation errors and the temporal evolution of the 3-D meteorological fields can be complex; our method allows a relevant estimation of the errors specifical to the meteorological conditions at each site.

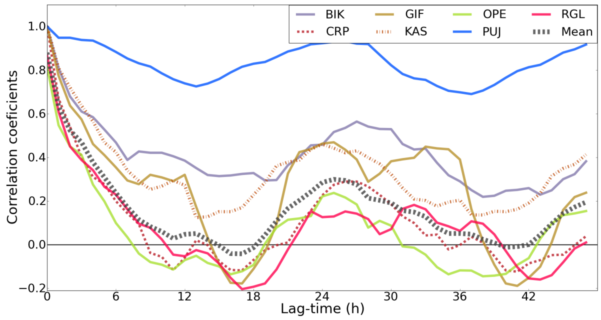


Fig. 4. Mean temporal auto-correlation. For each site, are figured the average correlations of all the observations from 12 to 7 p.m. with the following 48 h. Only the sites with a local maximum at 24 h are figured: the 6 sites with strong correlation with the PBL (in Table 2) and KAS mountain site. Solid lines for rural sites; dashed (resp. dotted) lines for coastal (resp. mountain) sites.

3.3 Correlations in the observation errors

If we focus on the non-diagonal terms of $\mathbf{R}_{\mathrm{ND}}$, a large part (64 %) of the correlation coefficients are very large ($|r| > 0.9$ with even positive and negative repartition). Strong positive and negative correlations mark a redundancy of the information provided by the observations. We then group the available observations into classes of redundant information. Amongst each class, we chose $|r_{i,j}| > 0.9$ between every tuple (i, j) of observations. Following this criterion, we divide the set of observations into 625 balanced classes, figuring 625 independent pieces of information given by the observations. This figure can be compared to other studies which filter out part of the data before the assimilation. For example, Bergamaschi et al. (2010) proposed to keep one observation per day and per site only, to avoid an over-constraining due to spatial and temporal correlations; in our case, it would have meant keeping 195 observations (i.e., 3 times less than the result of ND algorithm). Our method suggests keeping more independent pieces of information but the amount remains low compared to the total set of observations ($\sim$ 15 % of the $\sim$ 5000 observations). Then, in our system with a drastically reduced state space dimension, the network over-constrains the fluxes; but this result is very dependent of the analytical framework with aggregated regions and is unlikely to apply to full-resolution configurations. In systems with state vectors larger than the observation vector, one would expect that the operator H will not reduce the amount of independent pieces of information that can be assimilated so much.

The temporal structure of the error correlations is shown in Fig. 4. At each site, the mean time auto-correlations of the errors at lags 0 to 48 h are calculated. The average auto-correlations are computed with starting hours from 12:00 to 19:00 only; the patterns for the other hours of the day are the same, but with mean correlations that are closer to 0. It appears that the absolute correlations quickly decrease below $|r| < 0.25$ in about 5 h at each site. In Fig. 4, we display the sites with a day-to-day significant correlation. For these sites, we observe a maximum of correlation at 24 h, related to processes with a diurnal cycle. But this is not necessarily the PBL diurnal cycle since every site with strong correlations between the PBL height and the errors does not exhibit the 24 h peak. The 24 h periodic correlation could be related to the surface temperature diurnal cycle for instance or any other diurnal cycle in the atmospheric state.

A spatial structure of the observation covariances could also have been expected (Lauvaux et al., 2009) at sites close to each other such as GIF and TRN, which are about 100 km distant. But the calculated $\mathbf{R}_{\mathrm{ND}}$ covariance matrix does not exhibit any global spatial patterns. Distant observations can be strongly correlated, but the dominant underlying process is likely related to the PBL height.

3.4 Correlations in the background errors

3.4.1 The lateral boundary conditions (LBC)

In $\mathbf{B}_{\mathrm{ND}}$, the components related to the LBC are found to be independent ($|r| < 0.1$) from the ones of the aggregated fluxes. Amongst the LBC, two independent groups of regions appear. Within each group, the components are very strongly correlated or anti-correlated ($|r| > 0.9$) with the other components of the group and are independent from the ones not in the group. About 10 % of the LBC cannot be attributed to any of the two groups. In Fig. 5, the boundary regions have been sorted accordingly to these groups. The colours in the figure denotes this classification: blue and red for the two groups and green for the few remaining boundaries with mild correlations with other regions. The boundaries have also been sorted according to the influence of the assimilation of the observations on them. We quantified this influence by using the diagonal elements of the matrix $\mathbf{KH}$ (see Eqs. 2 and 3 in Sect. 2.1.1) which are necessarily in the interval [0, 1]. We highlight the regions strongly influenced ($\mathbf{KH}_{i,i} > 0.9$; "+" sign) against the others ("o" sign). Then it comes that the regions unseen (resp. constrained) by the inversion are strongly correlated with each other and not correlated with the constrained (resp. unseen) regions.

Drawing conclusions about the signs of the correlations between the boundary components is more difficult. Within the "unseen" group, none of the three algorithms can retrieve information about these components of the state vector since the corresponding elements in $\mathbf{H}$ are negligible compared to the others. Thus, the positive and negative correlations are likely to be numerical artifacts in the algorithms. For the constrained group, a large negative correlation between two regions means that the prior mean contribution of the two regions is well constrained. Equivalently, the two LBC components are well known on average but the two individual contributions are not separated from each other; this

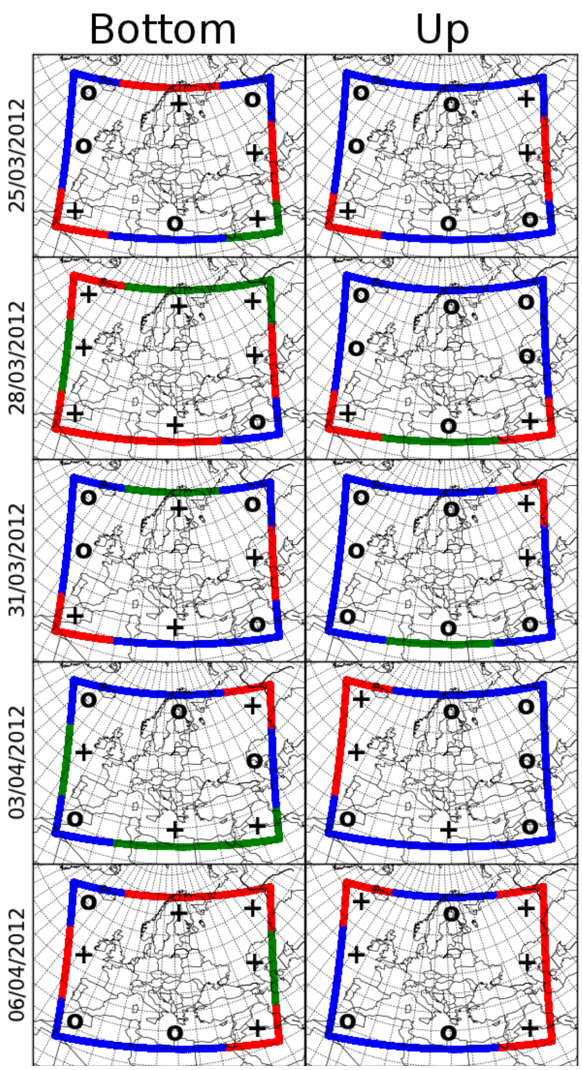


Fig. 5. Correlations between the errors on the LBC elements of the background vector for the two weeks after the leak start, calculated from $\mathbf{B}_{\mathrm{ND}}$. Dates are the starting dates of every 3 day window of uniform constant boundary concentrations. The bottom boundaries span from the surface to $\sim$700 hPa; up is the remaining part of the vertical direction. Red and blue highlight 2 classes of boundaries very strongly correlated ($|r| > 0.9$). Green boundaries are not significantly correlated to any other regions. "+" (resp. "o") signs corresponds to boundaries strongly (resp. lightly) influenced by the inversion (criterion of selection described in Sect. 3.4.1).

may indicate that the contributions are similar in magnitude with simultaneous transitions observed at the sites, hence not discernible. On the opposite, a positive correlation reveals a constraint on the difference of the contributions, hence on the spatial and/or temporal gradients in the LBC. For example, for two LBC regions positively correlated and upwind the observations, if a clear transition between the contribution of one region and the other is observable from the network,

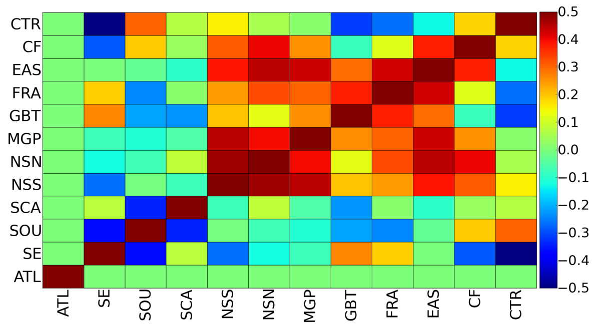


Fig. 6. Correlations between the errors of the background aggregated regions for the non-diagonal matrix $\mathbf{B}_{\mathrm{ND}}$ for the two weeks after the leak start. Refer to Fig. 1 for the region names, location and extension.

the gradient between the two will be well fixed, whereas the total balance will be unclear and biased by all the other components of the state vector. The sign of the correlation then mainly depends on the meteorological conditions at the site and of the air mass history when it is observed. It can be very variable and virtually unpredictable with a general formula. Our method gives an objective estimation of the issue.

3.4.2 The aggregated emissions

The error correlations for the aggregated regions of emissions are displayed in Fig. 6. Negative correlations occur only between regions that are not upwind any sites (i.e., in Fig. 6, for the period after the leak start: "ATL", "SCA", "SE", and "SOU"; see Fig. 1; e.g., $r = -0.36$ between "SOU" and "SE"). As for LBC, interpreting the correlations between unconstrained regions is hazardous. On the opposite, the positive error correlations between the regions close and upwind the sites (for example, $r = 0.45$ between the neighbouring regions NSS and MGP) denote the confidence in the background flux gradients amongst these regions but not necessarily in the total flux balance. The confidence in the prescribed gradients in emissions is consistent with the methodology used to build the inventory maps. Activity maps by sectors are convolved with emission factors. In Europe, the declared activity is considered reliable. Then, the emission factors, hence the overall magnitude of the emission, is a more critical source of uncertainties than the spatial distribution.

4 Flux inversion: the Elgin leak case

From here, we use the optimised tuples of matrices to actually compute the inversion with Eqs. (2) and (3).

4.1 Physical relevance of the inverted fluxes

In this study we have selected a domain for the simulations which spans over a region much bigger than the network coverage area. This choice is expected to decrease the errors due to the coarse approximations made on the LBC. But, in return, we do not expect the results to be accurate on the outer regions because of the inability of the inversion to either constrain unobserved regions, or distinguish the contributions from these regions and from the LBC when they are all upwind the observations.

The three methods rely on the assumption of the Gaussianity of all the errors. In particular, this assumption theoretically allows surface fluxes to be negative. With the optimised state vector $\boldsymbol{x}^{\mathrm{a}}$ calculated with ($\mathbf{R}_{\mathrm{DS}}$, $\mathbf{B}_{\mathrm{DS}}$), increments (defined as $\frac{(\hat{\boldsymbol{x}^{\mathrm{a}}})_j-(\boldsymbol{x}^{\mathrm{b}})_j}{(\boldsymbol{x}^{\mathrm{b}})_j}$) of $-150\,\%$ and below appear on the emissions in regions close to the sides of the domain. Hence the inversion generates strongly negative surface fluxes, due to inaccurate separations between emissions and boundary condition signals. Net surface uptake is physically not acceptable for CH_4 in Europe where anthropogenic emissions are largely prevailing (Bergamaschi et al., 2010) compared to soil uptake. Then, despite its very low-computation cost, a simple DS can not be applied unless one uses it on a larger number of subsets and with additional physical constraints.

In regard of the LBC issue, the other two algorithms (ML and ND) seem reliable. The ML algorithm does not compute absolute increments over 35 % for these regions and keeps posterior uncertainties compatible with a 0 % increment (i.e., no change from the prior flux). The non-diagonal tuple ($\mathbf{R}_{\mathrm{ND}}$, $\mathbf{B}_{\mathrm{ND}}$) leads to increments incompatible with the positivity of the fluxes in some of the outer regions but in an acceptable range ($> -10\,\%$) considering the posterior errors in the matrix $\mathbf{P}^{\mathrm{a}}$ for these regions.

Moreover, the diagonal algorithm DS and ML do not compute any significant posterior error correlation ($r^2 < 0.1$) between outer regions and the LBC; inversely, the ND tuple explicitly estimates strong posterior error correlations for these components (numerous correlation coefficients $r^2 > 0.5$). Hence, ND method can account for the erroneous separation from the outer emissions and the LBC.

4.2 Using the optimised tuples ($\mathbf{R}_{\mathrm{ND}}$, $\mathbf{B}_{\mathrm{ND}}$) for flux inversion

The non-diagonal tuple reliably takes the LBC ill-separation into account. We then use the inversion carried out with this tuple to analyse the posterior fluxes on the regions close to the network that are better constrained than the outer ones. The increments for the two periods (15-day long each) before and after the leak start are shown in Fig. 7. The posterior errors are not displayed because they do not exceed 1 % for most regions (maximum of 1.1 % for "NSN" region before the leak start).

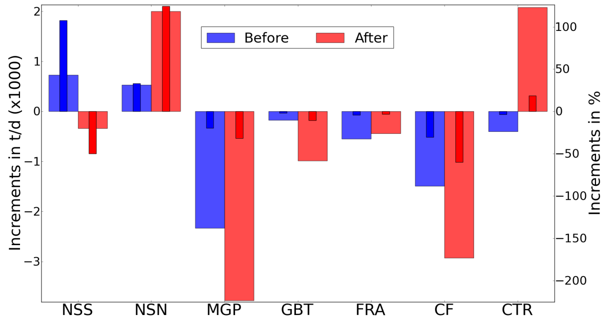


Fig. 7. Increments on the emissions for the regions not contiguous to the boundaries in $\mathrm{t\,d^{-1}}$ (large bars) and in % (thin bars) of the initial total with the non-diagonal tuple ($\mathbf{R}_{\mathrm{ND}}$, $\mathbf{B}_{\mathrm{ND}}$). The two temporal windows of inversion are the two weeks before (blue) the leak start and the two weeks after (red). Acronyms in Fig. 1.

These surprisingly small posterior uncertainties can seem physically inconsistent. But the figures apply to very large regions. Hence, uncertainties on pixels are expected to be smoothed by the aggregation. Moreover, as CH_4 emissions in spring in Europe are mainly anthropogenic and well documented, even the pixel errors could be relatively low.

Before the leak start, the total emissions from these inner regions ("NSS", "NSN", "MGB", "FRA" and "CTR") are corrected from $52\,455 \pm 16\,838\,\mathrm{t\,d^{-1}}$ to $51262 \pm 1999\,\mathrm{t\,d^{-1}}$, that is to say an increment of $-2.3 \pm 3.9\,\%$. After the leak start, the correction to the inner budget is from $52\,455 \pm 12\,431\,\mathrm{t\,d^{-1}}$ to $48\,062 \pm 48\,\mathrm{t\,d^{-1}}$, i.e. an increment of $-8.4 \pm 0.1\,\%$. The inversion of the CH_4 emissions over the inner regions suggests an over-estimation in the inventories which actually were not designed for year 2012 but for 2008. The decrease in the emissions by $3200 \pm 2000\,\mathrm{t\,d^{-1}}$ (i.e., $-6.1 \pm 3.8\,\%$ reported to the prior total balance) between the two inversion windows remains consistent with the uncertainties of the inventories and can be explained by the typical variability of the emissions. But one should recall that the results are averaged on aggregated regions, whereas the areas of influence of the sites do not necessary overlay the whole region (see Sect. 5.1). The inversion corrects simulated concentrations, not considering the implications in regards to absolute emissions. The increments can be amplified and suffer from aggregation errors and sampling heterogeneity (Kaminski et al., 2001). More critically, the aggregation errors on the areas that are not within the footprint of the network cannot be recovered from our methods since the operator H and the covariance matrix $\mathbf{B}$ are aggregated before the algorithms are run.

Focusing on the estimation of the CH_4 release from the Elgin platform, the inversion suggests an increase in the regional emission where the leak occurred ("NSS" region) of $+1472 \pm 30\,\mathrm{t\,d^{-1}}$ (the errors on the difference is calculated assuming that the two inversion windows are independent).

But the two parts of the North Sea are expected to be ill-distinguished by the system, as confirmed by a posterior correlation coefficient r of -0.78 between the errors in the two regions after the leak start. Then the flux that can be attributed to the leak is defined as the difference of the emissions after and before the leak start over the whole North Sea area. Our inversion computes a flux of $+406 \pm 33\,\mathrm{t\,d^{-1}}$. The inversion detects an increase of emissions from the direction of the leak, but fails to unambiguously affect the increased flux to the proper region. The figure we compute is of the same magnitude (3 times higher) than the estimation given by the operator and does not exceed 18 % of the background emissions related to oil and gas extraction in the North Sea.

Our results also reveal a high dependency to the meteorological situation during an inversion window. The ratio between the increment and the posterior errors on the emission budget in the inner regions is very different for the two periods (~ 1 before and $\gg 1$ after the leak start). The reconstructed error on the total budget largely depends on the correlation coefficients in $\mathbf{P}^{\mathrm{a}}$. For the period before the leak start, most posterior correlations are large and positive ($r > 0.9$). The gradients are then well constrained while the total budget stays uncertain. On the opposite, after the leak start, neighbouring regions exhibit negative correlations by pair (e.g., "NSN" – "NSS", "MGP" –"GBT"). The assimilation of the observations cannot separate the contribution from these close regions, but it leads to a good reduction of the error on the total balance. These two different behaviours may be related to different synoptic regimes during each inversion window: before the leak start, an anticyclone was laying on central and western Europe; after the leak start, air masses coming from North Europe vented the domain.

5 Discussion

5.1 Limitations and hypothesis probation

All the results depend on strong statistical and physical hypotheses, which may not all be robust. First, we show in Sect. 3.1 that the assumed Gaussian errors of the background can produce physically inconsistent inverted fluxes. Adding Lagrangian correcting factors (e.g., Göckede et al., 2010) to the cost function (Eq. 1) can ensure physically consistent fluxes. But that would alter the algebraic properties of the problem and make the implementation of our methods more complex. With regions that act as buffers against the uncertainties on the LBC, the ND methods proved to acceptably deal with the issue.

Second, all CTMs suffer from weaknesses and errors in their parameterizations and numerical scheme. The induced errors can be systematic and not only random, as suggested in Sect. 3.2. They should then be considered as a bias $\boldsymbol{\eta}$ in the observational errors $\boldsymbol{\epsilon} = \boldsymbol{y}^0 - \mathbf{H}\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\eta}, \mathbf{R}')$. Further investigations on the effects of the parameterizations, the resolution and the inputs to the CTM shall be carried out to quantify and fix as much as possible the bias $\boldsymbol{\eta}$. More specifically, in Sect. 3.2, we showed that very high diagnosed error variances during the night could be related to systematic state-dependent biases in the CTM vertical mixing and in PBL modelling.

Third, aggregating fluxes within bigger regions implicitly implies full correlations of the errors on the background in each region. Despite this strong assumption, our methods are supposed to diagnose the error on the aggregated fluxes that are within the footprint of the network. On the opposite, the aggregated regions that are partly within and partly outside the network footprint will exhibit strongly biased diagnosed errors and increments.

Kaminski et al. (2001) studied the issue and found potential errors of the same magnitude as the fluxes themselves. A better choice of the resolution and of aggregated regions considering the prior fluxes and the transport patterns (e.g., Wu et al., 2011) during the window of inversion should significantly improve the results of the methods.

Despite these weaknesses in our methods, the optimal tuple of covariance matrices gives better results than a tuple built on expert considerations: either these expert-built tuples, which are most of the time diagonal, are similar to $(\mathbf{R}_{\mathrm{DS}}, \mathbf{B}_{\mathrm{DS}})$ that causes inconsistent negative CH_4 fluxes, or the observation errors are enhanced to reduce their impact on the inversion; but in this latter configuration applied to our inversion windows, the corresponding inverted fluxes remain close to the prior ones and the flux uncertainties are not noticeably reduced. Our objectively calculated tuple gives better inversion results, with reduced posterior uncertainties. Moreover, some of the computed error patterns are generic and are transferable to other larger systems. In this study, we chose a particular representation of the complete full-resolution state vector. Most errors represented by the covariance matrix $\mathbf{R}$ are independent of this representation (Bocquet et al., 2011). As a consequence most results on $\mathbf{R}$ will remain valid in the framework of a full-resolution state vector. The recovery of the errors of the non-aggregated background vector are more ambiguous and only large patterns could be inferred for finer resolutions. Additional hypotheses must be made on the shape of the full-resolution background errors to deduce their values from the aggregated matrix. Our methods can then be seen as a way to simplify and project problems with large state vectors in order to infer the patterns of the errors with relatively small computation costs.

5.2 Implications for data selection

The framework we chose allows the explicit computation of the sensitivity of the inversion to each observation. We follow Cardinali et al. (2004) and calculate the influence matrix, which gives the effect of a small change of $\boldsymbol{y}^0$ on $\mathbf{H}\hat{\boldsymbol{x}}^{\mathrm{a}}$: $\mathbf{S} = (\mathbf{R} + \mathbf{HBH}^{\mathrm{T}})^{-1}\mathbf{HBH}^{\mathrm{T}}$. For every observation $\boldsymbol{y}^0_j$, high observation errors reduce the contribution of the observation

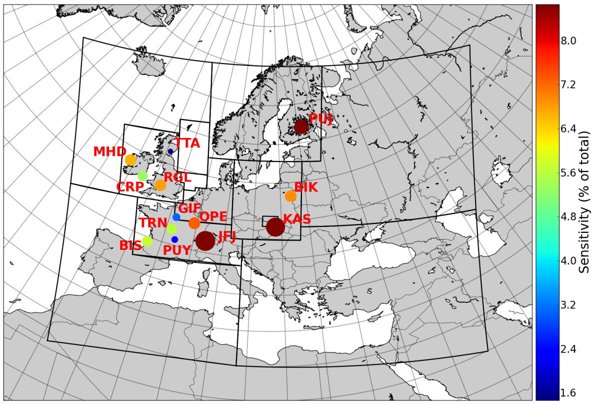
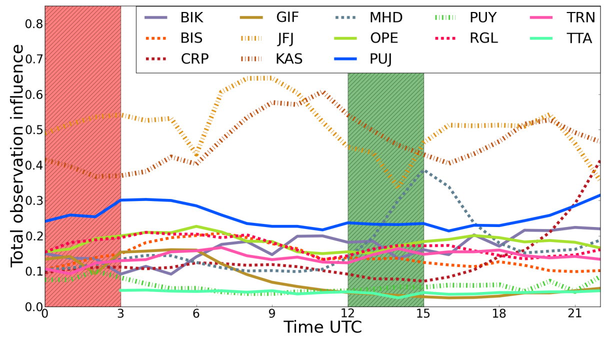
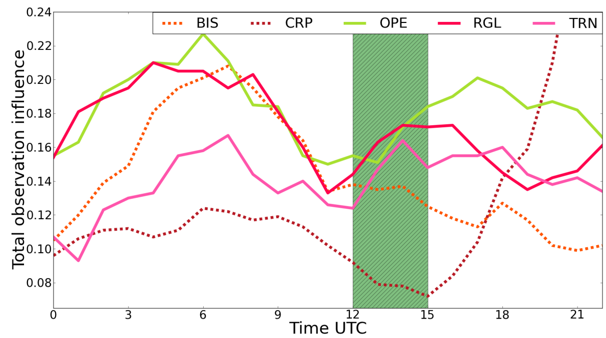


Fig. 8. Sensitivity of the inversion to each site as the sum of the diagonal elements of the sensitivity matrix **S** (details in Sect. 5.2) associated to the site. The figures are normalised by the total influence of the observations, i.e., the trace of **S**.



(a) Diurnal cycle of sensitivity for all sites.



(b) The same as Fig. 9a for 5 plain sites.

Fig. 9. Total sensitivity per site computed for each hour of the day (time UTC). Details on the sensitivity computation are described in Sect. 5.2. A sensitivity of 1 roughly corresponds to the constraint on 1 degree of freedom of the system. The green (resp. red) band highlights the interval of data selection generally used in most global inversion systems for the plain (resp. mountain) observation sites. Solid lines for rural sites; dashed (resp. dotted) lines for coastal (resp. mountain) sites.

j to the total inversion. On the opposite, high local contributions tend to enhance the amount of information the inversion can extract from a single observation. The sensitivity matrix **S** encompasses both these compensating effects. In Fig. 8, each site is colored according to its contribution to the inversion, calculated by adding all the diagonal terms of **S** associated to the site. KAS and JFJ have contributions 2.5 times higher than the average contribution, related to their situations in the free troposphere and the constraints they give on the LBC. As a site filling a gap in the observations, PUJ also has a strong contribution (1.5 times the average). On the other hand, GIF, TTA and PUY are very close to the core of the network. The algorithm attributes negligible contributions to these sites in favour of the other sites. Totaling the influence of the whole set of observations, we get the figure of about 50 % of the posterior state vector fixed by the observation; the other half comes from the background.

Usually, inversion systems assimilate only a few hours of observations per day, while efforts are made to monitor the atmospheric composition continuously. For example, Bergamaschi et al. (2010) chose to average 3 h of observation per day and per site (bands in Fig. 9a). This choice is justified by the confidence given to the model during the afternoon when the vertical mixing in the PBL is strong. Flagging out the other data may be acceptable when inquiring into patterns at the continental scale, but one needs more information at the local and regional scales. In Sect. 3.1, we show that the observation errors during the day are only 30 % lower than during the night, though this figure is mitigated by non-negligible systematic biases during the night. This small difference suggests that night observations could reasonably be assimilated if the biases had a smaller impact.

In Fig. 9a, the sensitivity by site of each hour of the day is shown. Apart from MHD, the usual intervals of selected data (green band for the plain site, blue for the mountain ones) do not exhibit significantly higher sensitivity in the inversion system than the other hours of the day. This selection then leads to a global loss of more than 85 % of information: the sum of the diagonal elements of the sensitivity matrix **S** related to the observations in the selection band reaches 15 % of the total trace. The mountain stations are known to suffer from the issue of ascendant polluted streams from the PBL as developed in Sect. 3.2. Our method implicitly automatically filters out these air masses. Then, the usable observations are not confined to the middle of the night. Moreover, concerning the plain sites (see Fig. 9b), the selected band corresponds to the minimum of sensitivity, that is to say the data that least constrain the local and regional fluxes we inquire into. The most influential observations are situated just before the beginning of the day, when the high errors of the night start decreasing and the local contributions are still significant. The end of the afternoon (at about 18:00) is also more influent that the middle of the day, but less than early morning. One should expect a better confidence in late afternoon results from the CTM, when the vertical mixing is still active (though reduced), than early morning.

We emphasised in Sect. 3.2 the diurnal patterns of the errors on the observations. Further efforts have to be made in modelling the PBL height and the vertical mixing to ensure better quality and reduced bias in the simulations at the end of the night. Our method would then allow a better use of the observations for local and regional inversions.

6 Conclusions

We inquired into the possibility of precisely and objectively estimating the covariance matrices of the errors on the observations and the background (**R** and **B**) that best fit inversion system requirements. A best guess of these matrices with regard to objective criteria is needed in the Bayesian inversion framework, especially for regional studies. To do so, we used algorithms developed in a theoretical framework, but too complex to be tested in full-resolution systems. The translation to a regional configuration was carried out by simplifying the system and reducing the total size of the covariance matrices to allow an algorithmic tuning of **R** and **B** that estimates optimum tuples (**R**, **B**) in terms of statistical properties.

We tested 3 algorithms of growing complexity (and computational costs) to estimate the optimal tuple (**R**, **B**). Unlike other studies, which make strong physical assumptions, such as isotropic spatial correlations in the observation errors or temporal decay of the correlations, we minimised the number of assumptions to keep a better objectivity in our results. In principle, all the patterns of errors can then be recovered specifically to the system during the window of the inversion we focused on.

Amongst other noteworthy patterns, our algorithms retrieved the errors due to the mis-estimation of the planetary boundary layer height in global circulation models, i.e., large errors during the night and lower during the day when the CTM reproduces the atmospheric transport better. This source of errors contribute in a large part (50 % in most plain sites) to the diurnal variability of the observation errors and also causes significant temporal correlations within a 24 h period.

Additionally, our approach does not require a prior filtering of the observations we could consider as ill-simulated by the model. In theory, all the available observations can then be assimilated in the inversion system and not only the ones during early afternoon. However, our study points at probable significant systematic bias in the CTM during the night. The night observations should then be excluded. But the algorithms give objective tools to diagnose the need for efforts to better simulate the atmospheric behaviour during the late night when the observations seem to have the biggest impact on the inversion results. Late afternoon observations were also computed to have a significant influence on the inversion results. A cautious implementation of these observations into an inversion system is expected to enhance the efficiency of the system.

The prospects from this work will be to quantify the uncertainties in our methods and their impact on the optimised fluxes. A dedicated Observing System Simulation Experiment could be carried out in that sense. The computational costs should also be reduced by running our scripts in parallel. In the framework of an inversion system with full temporal and spatial resolution, when variational algorithms are necessary to compute the optimal fluxes, our general method may overburden the computer and memory capacity. Indeed, the limiting factor in our algorithms comes from the diagonal maximisation of the log-likelihood needed to compute the non-diagonal optimal tuple. The maximising algorithm induces computational costs limited by the size of the 'background' vector $\boldsymbol{x}^{\mathrm{b}}$. The computational complexity is at least $\mathrm{O}\left([\dim(\boldsymbol{x}^{b})]^3\right)$ while full-resolution state space dimension is of several orders of magnitude larger than our reduced state space. Nevertheless, the method could be tested in systems of intermediate complexity to infer additional knowledge on the statistics of the errors.

Acknowledgements. We thank all the PIs and supporting staff from the sites we used for making available the validated observations quickly. We wish to thank Bertrand Bonan (INRIA), Lin Wu (Laboratoire des Sciences du Climat et de l'Environnement) and Victor Winiarek (Université Paris-Est, CEREA) for the fruitful discussions and advices. We are grateful to the referees for their valuable comments which led to a substantial improvement in the presentation of the paper. This research was supported by the Commissariat à l'Énergie Atomique et aux Énergies Renouvelables.

Edited by: C. Gerbig

The publication of this article is financed by CNRS-INSU.

References

Ahmadov, R., Gerbig, C., Kretschmer, R., Koerner, S., Neininger, B., Dolman, A. J., and Sarrat, C.: Mesoscale covariance of transport and CO_2 fluxes: Evidence from observations and simulations using the WRF-VPRM coupled atmosphere-biosphere model, J. Geophys. Res., 112, D22107, doi:10.1029/2007JD008552, 2007.

Ahmadov, R., Gerbig, C., Kretschmer, R., Körner, S., Rödenbeck, C., Bousquet, P., and Ramonet, M.: Comparing high resolution WRF-VPRM simulations and two global CO2 transport models with coastal tower measurements of CO_2, Biogeosciences, 6, 807–817, doi:10.5194/bg-6-807-2009, 2009.

Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B., Taguchi, S., and Zhu, Z.: TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988–2003, Global Biogeochem. Cy., 20, GB1002, doi:10.1029/2004GB002439, 2006.

Bergamaschi, P., Krol, M., Dentener, F., Vermeulen, A., Meinhardt, F., Graul, R., Ramonet, M., Peters, W., and Dlugokencky, E. J.: Inverse modelling of national and European CH_4 emissions using the atmospheric zoom model TM5, Atmos. Chem. Phys., 5, 2431–2460, doi:10.5194/acp-5-2431-2005, 2005.

Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Villani, M. G., Houweling, S., Dentener, F., Dlugokencky, E. J., Miller, J. B., Gatti, L. V., Engel, A., and Levin, I.: Inverse modeling of global and regional CH_4 emissions using SCIAMACHY satellite retrievals, J. Geophys. Res., 114, D22301, doi:10.1029/2009JD012287, 2009.

Bergamaschi, P., Krol, M. C., Meirink, J. F., Dentener, F., Segers, A., van Aardenne, J., Monni, S., Vermeulen, A. T., Schmidt, M., Ramonet, M., Yver, C., Meinhardt, F., Nisbet, E. G., Fisher, R. E., O'Doherty, S., and Dlugokencky, E. J.: Inverse modeling of European CH_4 emissions 2001–2006, J. Geophys. Res., 115, D22309, doi:10.1029/2010JD014180, 2010.

Bocquet, M., Wu, L., and Chevallier, F.: Bayesian design of control space for optimal assimilation of observations. Part I: Consistent multiscale formalism, Q. J. Roy. Meteorol. Soc., 137, 1340–1356, 2011.

Bousquet, P., Ciais, P., Peylin, P., Ramonet, M., and Monfray, P.: Inverse modeling of annual atmospheric CO_2 sources and sinks: 1. Method and control inversion, J. Geophys. Res., 104, 26161–26178, 1999.

Bousquet, P., Ciais, P., Miller, J. B., Dlugokencky, E. J., Hauglustaine, D. A., Prigent, C., Werf, G. R. V. d., Peylin, P., Brunke, E.-G., Carouge, C., Langenfelds, R. L., Lathière, J., Papa, F., Ramonet, M., Schmidt, M., Steele, L. P., Tyler, S. C., and White, J.: Contribution of anthropogenic and natural sources to atmospheric methane variability, Nature, 443, 439–443, 2006.

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.

Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T., and Ciais, P.: A European summertime CO_2 biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements, J. Geophys. Res., 116, 23303, doi:10.1029/2011JD016202, 2011.

Burg, J. P., Luenberger, D. G., and Wenger D. L.: Estimation of structured covariance matrices. Proceedings of the IEEE, 30:963974, 1982.

Cardinali, C., Pezzulli, S., and Andersson, E.: Influence-matrix diagnostic of a data assimilation system, Q. J. Roy. Meteor. Soc., 130, 2767–2786, 2004.

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, 2004.

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.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F. M., Chédin, A., and Ciais, P.: Inferring CO_2 sources and sinks from satellite observations: Method and application to TOVS data, J. Geophys. Res., 110, D24309, doi:10.1029/2005JD006390, 2005.

Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, Q. J. Roy. Meteor. Soc., 120, 1367–1387, 1994.

Dee, D. P.: On-line estimation of error covariance parameters for atmospheric data assimilation, Mon. Weather Rev., 123, 1128–1145, 1995.

Dee, D. P.: Bias and data assimilation, Q. J. Roy. Meteor. Soc., 131, 3323–3344, 2005.

Dentener, F., Peters, W., Krol, M., Van Weele, M., Bergamaschi, P., and Lelieveld, J.: Interannual variability and trend of CH_4 lifetime as a measure for OH changes in the 1979–1993 time period, J. Geophys. Res., 108, 4442, doi:10.1029/2002JD002916, 2003.

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.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Q. J. Roy. Meteor. Soc., 131, 3385–3396, 2005.

Dlugokencky, E. J., Myers, R. C., Lang, P. M., Masarie, K. A., Crotwell, A. M., Thoning, K. W., Hall, B. D., Elkins, J. W., and Steele, L. P.: Conversion of NOAA atmospheric dry air CH_4 mole fractions to a gravimetrically prepared standard scale, J. Geophys. Res., 110, D18306, doi:10.1029/2005JD006035, 2005.

Enting, I. G., Trudinger, C. M., Francey, R. J., and Granek, H.: Synthesis inversion of atmospheric CO_2 using the GISS tracer transport model, Tech. Rep. 29, Division of Atmospheric Research Technical Paper, CSIRO, Australia, 1993.

Enting, I. G., Trudinger, C. M., and Francey, R. J.: A synthesis inversion of the concentration and δ^{13} C of atmospheric CO_2, Tellus B, 47, 35–52, 1995.

Geels, C., Gloor, M., Ciais, P., Bousquet, P., Peylin, P., Vermeulen, A. T., Dargaville, R., Aalto, T., Brandt, J., Christensen, J. H., Frohn, L. M., Haszpra, L., Karstens, U., Rödenbeck, C., Ramonet, M., Carboni, G., and Santaguida, R.: Comparing atmospheric transport models for future regional inversions over Europe – Part 1: mapping the atmospheric CO_2 signals, Atmos. Chem. Phys., 7, 3461–3479, doi:10.5194/acp-7-3461-2007, 2007.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, J. Geophys. Res., 108, 4756, doi:10.1029/2002JD003018, 2003.

Göckede, M., Turner, D. P., Michalak, A. M., Vickers, D., and Law, B. E.: Sensitivity of a subregional scale atmospheric inverse CO_2 modeling framework to boundary conditions, J. Geophys. Res., 115, D24112, doi:10.1029/2010JD014443, 2010.

Hein, R., Crutzen, P. J., and Heimann, M.: An inverse modeling approach to investigate the global atmospheric methane cycle, Global Biogeochem. Cy., 11, 43–76, 1997.

Houweling, S., Kaminski, T., Dentener, F., Lelieveld, J., and Heimann, M.: Inverse modeling of methane sources and sinks using the adjoint of a global transport model, J. Geophys. Res., 104, 26137–26160, 1999.

Houweling, S., Röckmann, T., Aben, I., Keppler, F., Krol, M. C., Meirink, J. F., Dlugokencky, E. J., and Frankenberg, C.: Atmospheric constraints on global emissions of methane from plants, Geophys. Res. Lett., 33, L15821, doi:10.1029/2006GL026162, 2006.

Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C.: Unified notation for data assimilation: operational, sequential and variational, J. Meteorol. Soc. Jpn., 181–189, 1997.

Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G.: On aggregation errors in atmospheric transport inversions, J. Geophys. Res., 105, 4703–4715, 2001.

Lauvaux, T., Uliasz, M., Sarrat, C., Chevallier, F., Bousquet, P., Lac, C., Davis, K. J., Ciais, P., Denning, A. S., and Rayner, P. J.: Mesoscale inversion: first results from the CERES campaign with synthetic data, Atmos. Chem. Phys., 8, 3459–3471, doi:10.5194/acp-8-3459-2008, 2008.

Lauvaux, T., Pannekoucke, O., Sarrat, C., Chevallier, F., Ciais, P., Noilhan, J., and Rayner, P. J.: Structure of the transport uncertainty in mesoscale inversions of CO_2 sources and sinks using ensemble model simulations, Biogeosciences, 6, 1089–1102, doi:10.5194/bg-6-1089-2009, 2009.

Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., Sweeney, C., Diaz, L. I., Martins, D., Shepson, P. B., and Davis, K. J.: Constraining the CO_2 budget of the corn belt: exploring uncertainties from the assumptions in a mesoscale inverse system, Atmos. Chem. Phys., 12, 337–354, doi:10.5194/acp-12-337-2012, 2012.

Lopez, M.: Estimation des émissions de gaz à effet de serre à différentes échelles en France à l'aide d'observations de haute précision., Ph.D. thesis, Univ. Paris-Sud, 2012.

Lopez, M., Schmidt, M., Yver, C., Messager, C., Worthy, D., Kazan, V., Ramonet, M., Bousquet, P., and Ciais, P.: Seasonal variation of N_2O emissions in France inferred from atmospheric N_2O and ^{222}Rn measurements, J. Geophys. Res., 117, D14103, doi:10.1029/2012JD017703, 2012.

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.

Necki, J., Schmidt, M., Rozanski, K., Zimnoch, M., Korus, A., Lasa, J., Graul, R., and Levin, I.: Six-year record of atmospheric carbon dioxide and methane at a high-altitude mountain site in Poland, Tellus B, 55, 94–104, 2003.

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.

Peylin, P., Rayner, P. J., Bousquet, P., Carouge, C., Hourdin, F., Heinrich, P., Ciais, P., and AEROCARB contributors: Daily CO_2 flux estimates over Europe from continuous atmospheric measurements: 1, inverse methodology, Atmos. Chem. Phys., 5, 3173–3186, doi:10.5194/acp-5-3173-2005, 2005.

Pison, I., Menut, L., and Bergametti, G.: Inverse modeling of surface NOx anthropogenic emission fluxes in the Paris area during the Air Pollution Over Paris Region (ESQUIF) campaign, J. Geophys. Res., 112, D24302, doi:10.1029/2007JD008871, 2007.

Pison, I., Bousquet, P., Chevallier, F., Szopa, S., and Hauglustaine, D.: Multi-species inversion of CH_4, CO and H_2 emissions from surface measurements, Atmos. Chem. Phys., 9, 5281–5297, doi:10.5194/acp-9-5281-2009, 2009.

Popa, M. E., Gloor, M., Manning, A. C., Jordan, A., Schultz, U., Haensel, F., Seifert, T., and Heimann, M.: Measurements of greenhouse gases and related tracers at Bialystok tall tower station in Poland, Atmos. Meas. Tech., 3, 407–427, doi:10.5194/amt-3-407-2010, 2010.

Prather, M. J., Zhu, X., Strahan, S. E., Steenrod, S. D., and Rodriguez, J. M.: Quantifying errors in trace species transport modeling, P. Natl. Acad. Sci. USA, 105, 19617–19621, 2008.

Ramonet, M., Ciais, P., Aalto, T., Aulagnier, C., Chevallier, F., Cipriano, D., Conway, T. J., Haszpra, L., Kazan, V., Meinhardt, F., Paris, J.-D., Schmidt, M., Simmonds, P., Xueref-Rémy, I., and Necki, J. N.: A recent build-up of atmospheric CO_2 over Europe. Part 1: observed signals and possible explanations, Tellus B, 62, 1–13, 2010.

Ramonet, M., Ciais, P., Rivier, L., Laurila, T., Vermeulen, A., Geever, M. , Jordan, A., Levin, I., Laurent, O., Delmotte, M., Wastine, B., Hazan, L., Schmidt, M., Tarniewicz, J., Vuillemin, C., Pison, I., Spain, G., and Paris, J.-D.: The ICOS Atmospheric Thematic Center (ATC), GAW Report No. 206, 16th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques (GGMT-2011), 2011.

Rayner, P. J., Enting, I. G., Francey, R. J., and Langenfelds, R.: Reconstructing the recent carbon cycle from atmospheric CO_2, $\delta^{13}C$ and O_2/N_2 observations, Tellus B, 51, 213–232, 1999.

Reimann, S., Vollmer, M., Folini, D., Steinbacher, M., Hill, M., Buchmann, B., Zander, R., and Mahieu, E.: Observations of long-lived anthropogenic halocarbons at the high-Alpine site of Jungfraujoch (Switzerland) for assessment of trends and European sources, Sci. Total Environ., 391, 224–231, 2008.

Sarrat, C., Noilhan, J., Dolman, A. J., Gerbig, C., Ahmadov, R., Tolk, L. F., Meesters, A. G. C. A., Hutjes, R. W. A., Ter Maat, H. W., Pérez-Landa, G., and Donier, S.: Atmospheric CO_2 modeling at the regional scale: an intercomparison of 5 mesoscale atmospheric models, Biogeosciences, 4, 1115–1126, doi:10.5194/bg-4-1115-2007, 2007.

Schmidt, H., Derognat, C., Vautard, R., and Beekmann, M.: A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe, Atmos. Environ., 35, 6277–6297, 2001.

Talagrand, O.: A posteriori evaluation and verification of analysis and assimilation algorithms, in: Workshop on Diagnosis of Data Assimilation Systems, 2–4, 1998.

Tarantola, A.: Inverse Problem Theory, Elsevier, New York, USA, 613 pp., 1987.

Tolk, L. F., Meesters, A. G. C. A., Dolman, A. J., and Peters, W.: Modelling representation errors of atmospheric CO_2 mixing ratios at a regional scale, Atmos. Chem. Phys., 8, 6587–6596, doi:10.5194/acp-8-6587-2008, 2008.

Vautard, R., Beekmann, M., Roux, J., and Gombert, D.: Validation of a hybrid forecasting system for the ozone concentrations over the Paris area, Atmos. Environ., 35, 2449–2461, 2001.

Vermeulen, A. T., Team, T. C., and Pieterse, G.: Tall tower observations of greenhouse gases in Europe: possibilities for emission verification, 7th International CO_2 Conference, 25–30 September 2005.

Wahba, G., Johnson, D. R., Gao, F., and Gong, J.: Adaptive tuning of numerical weather prediction models: Part I: randomized GCV and related methods in three and four dimensional data assimilation, Tech. rep., Citeseer, 1994.

Winiarek, V., 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., 117, D05122, doi:10.1029/2011JD016932, 2012.

Wu, L., Bocquet, M., Lauvaux, T., Chevallier, F., Rayner, P., and Davis, K.: Optimal representation of source-sink fluxes for mesoscale carbon dioxide inversion with synthetic data, J. Geophys. Res, 116, D21304, doi:10.1029/2011JD016198, 2011.