**Nonlinear Processes
in Geophysics**

Inverse modelling of atmospheric tracers: non-Gaussian methods and second-order sensitivity analysis

M. Bocquet[1,2]

[1]Université Paris-Est, CEREA, Research and Teaching Centre in Atmospheric Environment, Joint laboratory École Nationale des Ponts et Chaussées / EDF R&D, avenue Blaise Pascal, 77455 Champs sur Marne, France
[2]INRIA, Paris-Rocquencourt research centre, France

Abstract. For a start, recent techniques devoted to the reconstruction of sources of an atmospheric tracer at continental scale are introduced. A first method is based on the principle of maximum entropy on the mean and is briefly reviewed here. A second approach, which has not been applied in this field yet, is based on an exact Bayesian approach, through a maximum a posteriori estimator. The methods share common grounds, and both perform equally well in practice. When specific prior hypotheses on the sources are taken into account such as positivity, or boundedness, both methods lead to purposefully devised cost-functions. These cost-functions are not necessarily quadratic because the underlying assumptions are not Gaussian. As a consequence, several mathematical tools developed in data assimilation on the basis of quadratic cost-functions in order to establish a posteriori analysis, need to be extended to this non-Gaussian framework. Concomitantly, the second-order sensitivity analysis needs to be adapted, as well as the computations of the averaging kernels of the source and the errors obtained in the reconstruction. All of these developments are applied to a real case of tracer dispersion: the European Tracer Experiment [ETEX]. Comparisons are made between a least squares cost function (similar to the so-called 4D-Var) approach and a cost-function which is not based on Gaussian hypotheses. Besides, the information content of the observations which is used in the reconstruction is computed and studied on the application case. A connection with the degrees of freedom for signal is also established. As a byproduct of these methodological developments, conclusions are drawn on the information content of the ETEX dataset as seen from the inverse modelling point of view.

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

1 Introduction

1.1 Inverse modelling of tracers

There has been increasing interest in the use of inverse modelling techniques with a view to estimate atmospheric chemical emissions, sources, or fluxes. This is due to the increasing computing performances as well as the advances made in transport and chemistry numerical modelling. Besides there is a real need and appeal for the top-down approach (inferring the emissions from the observations): for example in climate change with estimating the greenhouse gases fluxes, or in industrial risk assessment with the identification of a source of an accidental release into the atmosphere.

This paper focusses on this latter topic, though many of the developments may apply to other situations. The main characteristics of an accidental release of a pollutant in the atmosphere, which make it different from species with spatialised emissions, are:

– In most cases, the source term is point-wise, though an extension in time is often to be considered. Obviously the source is positive, in contrast with situations (greenhouse gases) where sinks of pollutant would exist.

– The numerical modelling of such pollutant involves strong gradients in the concentrations fields, which often leads to strong representativeness errors in the comparison with pointwise observations.

– From the statistical point of view, which is often resorted to in data assimilation, there is (hopefully) only one accident of a type. As a consequence, it is difficult to infer statistics on a large class of dispersion accidents. Occasionally, this can be circumvented when synthetic dispersion events are available, when these can easily be generated and when they are representative of the risk situation. As a consequence the statistics that can be inferred from the data are usually rather limited.

In the following it will be assumed that the physics (advection, diffusion, depletion terms, chemistry) is first order in the concentration field. This rules out species involved in second-order reactions, such as ozone and its precursors. Yet many trace gases abide to this restriction : tracers, trace gases like heavy metals, radionuclides.

One significant technical advantage of the linearity assumption is that the cost function that is used as an optimisation criterion to find out the source, can often be expressed in the space of the observations, of much fewer dimensions than the parameter space. This establishes a dual cost function that is likely to be minimised significantly faster.

1.2 Non-Gaussian approaches

The inverse modelling of tracers in the atmospheric physics community has been a growing topic over the last fifteen years. References to first attempts and the developments of new methods can be found in the introduction of (Bocquet, 2007). The emphasis is put here on the newly proposed methods, which consistently take into account prior hypotheses such as positivity of the source or its boundedness.

In (Bocquet, 2005b,c), it was proposed to use the principle of maximum entropy on the mean as an inference principle to obtain an estimate of the source to be reconstructed. The method was later extended (Bocquet, 2005a; Krysta and Bocquet, 2007) to statistically consistent diagnostic tools to estimate a posteriori the quality of a reconstruction. It has several advantages such as the fact that the cost function that is built is always convex. A dual criterion can often be obtained analytically, and the equivalence with the primal problem is essentially granted by the convexity of the primal and dual cost functions.

The method was then successfully applied to the real data of the Euporean Tracer Experiments [ETEX] (Bocquet, 2007; Krysta et al., 2007) with a view to a high-resolution reconstruction of the source. The source was correctly localised (ETEX-I and ETEX-II), and the total released mass was correctly estimated up to a few percents (ETEX-I). In the case of the Chernobyl accident, it was also applied to the reconstruction of radionuclides available for long-range transport with a strong sanitary impact (Davoine and Bocquet, 2007). In both applications, this approach performed significantly better than the optimisation with a cost function based on a quadratic criterion (that is to say based on Gaussian prior laws for the source and the modelled errors). It was therefore proven that, in this risk assessment context, devising such *non-Gaussian* cost functions as a criterion is not only legitimate but useful. This type of approach will be referred to as non-Gaussian in the following.

This methodology will be recalled briefly in Sect. 2.

Another type of inference used in inverse modelling is to resort to a Bayesian analysis, and to use a maximum a posteriori criterion as an estimator. Note that both types of inference can be designated as probabilistic inverse modelling

techniques (Tarantola and Valette, 1982). Similarly to MEM, relying on mathematical background of linear and non-linear convex analysis ((Borwein and Lewis, 2000) and reference therein), it was noted in (Lagarde, 2000) that it is possible to build a criterion in the dual space, though, without convexity, no equivalence between the primal and the dual problems can be guaranteed. In Sect. 3, we will build on this idea. Section 4 will provide with concrete examples on real data. It will define precisely the context and setup of the reconstructions.

Both approaches yield non necessarily Gaussian cost function. This may be problematic in a posterior study of the analysis. Any statistical methods developed in geophysical data assimilation dealing with second-order sensitivity of the solution are to be affected by the non-quadratic nature of these cost-functions.

The main aim of this paper is to present such second-order analysis, to show how to extend previous available tools. It is also to exhibit the specifics of such a second-order analysis based on a non-Gaussian approach as compared to a quadratic criterion approach. A secondary aim is to apply them on a real regional dispersion case. This will be studied in Sect. 5.

1.3 The reconstruction problem

The inverse problem based on the discretised physics can ultimately be brought to the form

$$\boldsymbol{\mu} = \boldsymbol{H}\boldsymbol{\sigma} + \boldsymbol{\varepsilon}\,, \qquad (1)$$

where $\boldsymbol{\mu}$ is the vector of all d observations (usually concentration measurements), $\boldsymbol{\sigma}$ is the vector of the discretised source components (N in total), $\boldsymbol{\varepsilon}$ is the vector of the errors due to instrumental, representativeness and modelling errors. $\boldsymbol{H}$ is a matrix of size $d \times N$ which both describes the observation operator as well as the transport and chemistry model. When the number of observations is significantly smaller than the number of source components, it is more efficient to compute the entries of $\boldsymbol{H}$ through the use of the adjoint of the model. Rows of $\boldsymbol{H}$ are then adjoint solutions indexed by the observations (Uliasz, 1983; Pudykiewicz, 1998).

The inverse problem consists in estimating $\boldsymbol{\sigma}$ and the errors $\boldsymbol{\varepsilon}$, given the observations $\boldsymbol{\mu}$, the model represented by Eq. (1), and some prior hypotheses on $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$. Because of the physics of dispersion, this inverse problems is known to be ill-conditioned. Because $N \gg d$, it is also ill-posed. The prior hypotheses act as regularisation. They help pick up a single physically-motivated solution among the many solutions allowed by the observations.

1.4 The least squares approach

The classical variational approach to solve Eq. (1), is to resort to a least squares criterion (possibly weighted by an observation error covariance matrix). It amounts to minimising

the l_2-norm of the departure $\boldsymbol{\mu} - \boldsymbol{H\sigma}$. However the number of variables could be far superior to the number of observations. Moreover, observation errors and model errors are reducing the information content of these observations. Therefore one needs a regularisation procedure, which picks up the most a priori plausible solution among those satisfying the observation constraints. To do so, it is usual to associate a regularising Tikhonov term to the error departure so that the complete cost function to minimise is

$$\mathcal{L}(\boldsymbol{\sigma}) = \frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{H\sigma})^{\dagger} \boldsymbol{R}^{-1} (\boldsymbol{\mu} - \boldsymbol{H\sigma}) + \frac{1}{2} \boldsymbol{\sigma}^{\dagger} \boldsymbol{B}^{-1} \boldsymbol{\sigma} , \quad (2)$$

where $\boldsymbol{R}$ is the observation error covariance matrix, and $\boldsymbol{B}$ is the background error covariance matrix which contains the a priori known statistical information on the source. The $\dagger$ symbol denotes vector or matrix transposition. This cost function is similar to usual four dimensional variational cost function [4D-Var]: time and space components are all solved for altogether. More precisely, it can be shown that it is actually equivalent to a weak formulation of 4D-Var, the source field playing the role of model error[1]. Such an optimisation problem can be equivalently solved in the physical space of measurements of lower dimensionality d. The dual cost function (known as the Physical-Space Statistical Assimilation System [PSAS] cost function (Courtier, 1997)) is

$$\widehat{\mathcal{L}}(\boldsymbol{\beta}) = \frac{1}{2} \boldsymbol{\beta}^{\dagger} \left(\boldsymbol{HBH}^{\dagger} + \boldsymbol{R} \right) \boldsymbol{\beta} - \boldsymbol{\beta}^{\dagger} \boldsymbol{\mu} , \quad (3)$$

and should be minimised on $\boldsymbol{\beta} \in \mathbb{R}^d$. The optimal source and diagnosed errors are then given by

$$\overline{\boldsymbol{\sigma}} = \boldsymbol{BH}^{\dagger} \overline{\boldsymbol{\beta}} , \qquad \overline{\boldsymbol{\varepsilon}} = \boldsymbol{R} \overline{\boldsymbol{\beta}} , \quad (4)$$

where $\overline{\boldsymbol{\beta}}$ is the minimiser $\min_{\boldsymbol{\beta}} \widehat{\mathcal{L}}(\boldsymbol{\beta})$ of the dual cost function. This solution identifies with the minimiser which could be obtained from criterion Eq. (2).

It is well known that a Bayesian reasoning underlies this cost-function, on the basis that the prior errors (background and observations) are assumed of Gaussian nature (Talagrand, 1997).

1.5 Application to a real dispersion event: ETEX-I

The methodological developments put forward in this paper will be applied on the data of a real continental atmospheric dispersion experiment: ETEX. The European Tracer Experiment [ETEX] was organised by the Joint Research Centre, Ispra, in 1994. ETEX-I consisted in the release of 340 kg of an inert gaseous tracer. The tracer is insensitive to wet scavenging and dry deposition. The release rate was meant constant over twelve hours. The subsequent plume was monitored over Europe by 168 WMO stations (Nodop et al., 1998). A few years after the Chernobyl accident, its purpose was to

---

[1]Bocquet, M. and Wu, L., unpublished

test the capabilities of emergency centres and numerical dispersion models to forecast and reconstruct the plume. While models compared favourably well with ETEX-I data, it is still not the case with ETEX-II, which is therefore left apart so far for modelling purpose. ETEX-I was later used as an inverse modelling experiment with real data. Given the (purposefully) accidental nature of the release, the aim is to identify the location of the release, and estimate the release rate as well as the total tracer mass released. It is still considered as the best instrumented dispersion experiment at continental scale to date.

2 Maximum entropy on the mean

The maximum entropy on the mean formalism is a Bayes-like approach which allows to take into account prior information, such as the source positivity. Like the Bayes formula, it may be turned into a probabilistic inference principle. In (Bocquet, 2005b), it has been been extended to the framework and language of geophysical data assimilation.

2.1 Level-two probabilistic inference

The implementation of the principle relies on the minimisation of the information content that separates the a posteriori probability density function [pdf] of the source and the errors $p(\boldsymbol{\sigma}, \boldsymbol{\varepsilon})$ from the a priori density $\nu(\boldsymbol{\sigma}, \boldsymbol{\varepsilon})$, with the exception of the information provided by the observations. The discrepancy is measured by the relative entropy

$$\mathcal{K}(p, \nu) = \sum_{\boldsymbol{\sigma}, \boldsymbol{\varepsilon}} p(\boldsymbol{\sigma}, \boldsymbol{\varepsilon}) \ln \left(\frac{p(\boldsymbol{\sigma}, \boldsymbol{\varepsilon})}{\nu(\boldsymbol{\sigma}, \boldsymbol{\varepsilon})} \right) , \quad (5)$$

where the sum (possibly an integral) runs on all potential sources and errors. With the observation constraints properly taken care of by a vector of Lagrange multipliers $\boldsymbol{\beta}$, one has to minimise on p the functional

$$\widetilde{\mathcal{L}}(p) = \mathcal{K}(p, \nu) + \boldsymbol{\beta}^{\dagger} \left\{ \boldsymbol{\mu} - \sum_{\boldsymbol{\sigma}, \boldsymbol{\varepsilon}} p(\boldsymbol{\sigma}, \boldsymbol{\varepsilon}) \left(\boldsymbol{H\sigma} + \boldsymbol{\varepsilon} \right) \right\} , \quad (6)$$

which leads to the effective cost function in the dual space of the $\boldsymbol{\beta}$ (isomorphic to the observation space $\mathbb{R}^d$),

$$\widehat{\mathcal{L}}(\boldsymbol{\beta}) = \widehat{\nu}(\boldsymbol{H}^{\dagger} \boldsymbol{\beta}, \boldsymbol{\beta}) - \boldsymbol{\beta}^{\dagger} \boldsymbol{\mu} , \quad (7)$$

provided the definition of the log-Laplace transform

$$\widehat{\nu}(\boldsymbol{\alpha}, \boldsymbol{\delta}) = \ln \left\{ \sum_{\boldsymbol{\sigma}, \boldsymbol{\varepsilon}} \nu(\boldsymbol{\sigma}, \boldsymbol{\varepsilon}) \exp \left(\boldsymbol{\alpha}^{\dagger} \boldsymbol{\sigma} + \boldsymbol{\delta}^{\dagger} \boldsymbol{\varepsilon} \right) \right\} . \quad (8)$$

The minimising vector $\overline{\boldsymbol{\beta}}$ of this effective cost function $\widehat{\mathcal{L}}(\boldsymbol{\beta})$ determines completely the optimal law p. An estimator of the source and the errors can then be obtained from p.

The most convenient estimator turns out to be the average. For any prior law ν, it is given by

$$\overline{\sigma}_{\text{MEM}} = \nabla_{\alpha} \widehat{\nu}(\alpha, \delta)_{|H^{\dagger}\overline{\beta}, \overline{\beta}}, \quad \overline{\varepsilon}_{\text{MEM}} = \nabla_{\delta} \widehat{\nu}(\alpha, \delta)_{|H^{\dagger}\overline{\beta}, \overline{\beta}}. \quad (9)$$

These estimates satisfy the observation equation: $\mu = H\overline{\sigma}_{\text{MEM}} + \overline{\varepsilon}_{\text{MEM}}$. As is clear from Eq. (7), the actual form of the cost function that results from the technique depends on the a priori laws. When the source prior is Gaussian, and when the observation errors are also Gaussian, then the cost function Eq. (7) turns out to be the least squares cost function Eq. (3).

The error prior statistics will be assumed Gaussian from now on, while the source prior statistics will be variable.

2.2   Example: Bernoulli law

A typical and easy to implement prior assumption is based on the Bernoulli law. At first, it is assumed a priori that a grid cell k emits a mass m_k of tracer, with a probability γ_k. $1 - \gamma_k$ is the probability that it emits nothing. It is a bounded law. The formalism then leads to the cost-function

$$\widehat{\mathcal{L}}(\beta) = \sum_{k=1}^{N} \ln \left\{ 1 - \gamma_k + \gamma_k \exp \left(m_k [\beta^{\dagger} H]_k \right) \right\}$$
$$+ \frac{1}{2} \beta^{\dagger} R \beta - \beta^{\dagger} \mu. \quad (10)$$

The optimal source is then given, component-wise, by

$$\overline{\sigma}_k^{\text{MEM}} = m_k \frac{\gamma_k \exp \left(m_k \left[\overline{\beta}^{\dagger} H \right]_k \right)}{1 - \gamma_k + \gamma_k \exp \left(m_k \left[\overline{\beta}^{\dagger} H \right]_k \right)}, \quad (11)$$

and the analysed errors are still given formally by $\overline{\varepsilon}_{\text{MEM}} = R\overline{\beta}$ of Eq. (4), though $\overline{\beta}$ is now the minimum of the cost function Eq. (10). An application will be presented in Sect. 4.

2.3   Level-one primal cost function

It is often convenient to introduce the level-one primal cost function

$$\mathcal{L}(\sigma) = \widehat{\nu}^{*}(\sigma, \mu - H\sigma), \quad (12)$$

where the Fenchel conjugate of $\widehat{\nu}$ is defined by

$$\widehat{\nu}^{*}(\sigma, \varepsilon) = \sup_{\alpha, \delta} \left(\alpha^{\dagger} \sigma + \delta^{\dagger} \varepsilon - \widehat{\nu}(\alpha, \delta) \right). \quad (13)$$

It can be shown by Fenchel duality that minimising the dual cost function Eq. (7) is equivalent to minimising the level-one primal cost function Eq. (12). This can be achieved because the criterion yielded by the maximum entropy inference is always convex (for a set of linear constraints), so that the Fenchel duality gap is vanishing (see Borwein and Lewis (2000) for a recent review on the topic). That is to say

$$\inf_{\sigma \in \mathbb{R}^N} \mathcal{L}(\sigma) = - \inf_{\beta \in \mathbb{R}^d} \widehat{\mathcal{L}}(\beta). \quad (14)$$

Besides $\overline{\sigma}_{\text{MEM}}$ and $\overline{\beta}$ are linked by Eq. (9). So that the dual criterion is equivalent to the primal one, as far as source estimations are concerned.

Note that the minimiser $\overline{\sigma}_{\text{MEM}}$ can be interpreted as the Maximum A Posteriori [MAP] of the unnormalised probability density function $\exp(-\mathcal{L}(\sigma))$. This is supported by the following chain of reasoning:

$$\nu(\sigma) \xrightarrow[\mu]{\text{MEM}} p(\sigma | \mu) = \frac{\nu(\sigma) e^{\beta^{\dagger} H\sigma}}{\sum_{\sigma} \nu(\sigma) e^{\beta^{\dagger} H\sigma}}.$$

$$\text{expectation}$$

$$\overline{\sigma}_{\text{MAP}} = \min_{\sigma} \widehat{\nu}^{*}(\sigma) \longleftrightarrow \overline{\sigma}_{\text{MEM}} = \partial_{\alpha} \widehat{\nu} \left(H^{\dagger} \overline{\beta} \right)$$

$$\text{MAP}$$

$$p(\sigma | \mu) \propto \exp(-\widehat{\nu}^{*}(\sigma)) \xleftarrow[\text{Bayes}]{\mu} \exp(-\widehat{\nu}^{*}(\sigma))$$

Here, the errors are not being considered for the sake of simplicity. The upper arch corresponds to the MEM inference. It starts with the prior ν and leads (by choice) to the average estimator $\overline{\sigma}_{\text{MEM}}$. The lower arch corresponds to the Bayesian inference with the choice of the MAP estimator $\overline{\sigma}_{\text{MAP}}$. It starts with the prior $\exp(-\widehat{\nu}^{*})$. This latter prior is chosen so that the two estimators $\overline{\sigma}_{\text{MEM}}$ and $\overline{\sigma}_{\text{MAP}}$ coincide.

Any monotonic embedding function f (leading to $f(-\widehat{\nu}^{*})$), not only the exponential function, can perform the task of matching the two estimators. However, large deviation results (Ellis, 1985; Le Besnerais et al., 1999) of pdfs strongly support this particular form, on theoretical grounds. Indeed, the exponential of minus the relative entropy is known to be the a measure of the deviation around the prior law, in the large deviation limit. The relevance of large deviation limit can be hinted at by the fact that any solution that represents a source of accidental nature, and which is understood as rare by the prior, is in the tail of the multidimensional prior pdf $\nu(\sigma)$.

3   Exact Bayesian approach

An alternative route to MEM consists in a strictly speaking Bayesian inference. It is assumed again that $\nu(\sigma, \varepsilon)$ represents the prior pdf of σ and ε. The Bayesian inference on the basis of this prior and the observation $\mu = H\sigma + \varepsilon$ yields the posterior pdf $p(\sigma, \varepsilon) \equiv p(\sigma, \varepsilon | \mu)$ according to

$$p(\sigma, \varepsilon) = \frac{\delta(\mu - H\sigma - \varepsilon) \nu(\sigma, \varepsilon)}{\sum_{\sigma, \varepsilon} \delta(\mu - H\sigma - \varepsilon) \nu(\sigma, \varepsilon)}. \quad (15)$$

The Dirac distribution $\delta(\mu - H\sigma - \varepsilon)$ enforces the observation constraints. An estimator is given by the MAP estimator of this pdf (or alternatively the logarithm of the pdf):

$$\overline{\sigma}_{\text{MAP}}, \overline{\varepsilon}_{\text{MAP}} = \underset{\sigma, \varepsilon}{\text{argmax}} \ \ln p(\sigma, \varepsilon)$$

$$= \underset{\sigma, \varepsilon : \mu = H\sigma + \varepsilon}{\text{argmax}} \ \ln \nu(\sigma, \varepsilon). \quad (16)$$

One easy way of lifting the constraints is to minimise the effective cost-function $\mathcal{L}(\boldsymbol{\sigma}) = -\ln \nu(\boldsymbol{\sigma}, \boldsymbol{\mu} - \boldsymbol{H}\boldsymbol{\sigma})$, thanks to the substitution $\boldsymbol{\varepsilon} = \boldsymbol{\mu} - \boldsymbol{H}\boldsymbol{\sigma}$.

3.1 Possible Fenchel duality

Invoking the Fenchel duality it is then possible to show that on very general grounds:

$$\inf_{\boldsymbol{\sigma} \in \mathbb{R}^N} \mathcal{L}(\boldsymbol{\sigma}) = \inf_{\boldsymbol{\sigma} \in \mathbb{R}^N} (-\ln \nu(\boldsymbol{\sigma}, \boldsymbol{\mu} - \boldsymbol{H}\boldsymbol{\sigma})) \qquad (17)$$

$$\geq - \inf_{\boldsymbol{\beta} \in \mathbb{R}^d} \left((-\ln \nu)^*(\boldsymbol{H}^\dagger \boldsymbol{\beta}, \boldsymbol{\beta}) - \boldsymbol{\beta}^\dagger \boldsymbol{\mu} \right) \qquad (18)$$

$$= - \inf_{\boldsymbol{\beta} \in \mathbb{R}^d} \widehat{\mathcal{L}}(\boldsymbol{\beta}), \qquad (19)$$

where $(-\ln \nu)^*$ is the Fenchel conjugate of $(-\ln \nu)$. $\widehat{\mathcal{L}}$ is the dual cost function. In the case where $-\ln \nu$ and $(-\ln \nu)^*$ are convex the reverse inequality also stands, and the useful duality result is

$$\inf_{\boldsymbol{\sigma} \in \mathbb{R}^N} \mathcal{L}(\boldsymbol{\sigma}) = - \inf_{\boldsymbol{\beta} \in \mathbb{R}^d} \widehat{\mathcal{L}}(\boldsymbol{\beta}). \qquad (20)$$

Besides $\overline{\boldsymbol{\sigma}}_{\text{MAP}}$, $\overline{\boldsymbol{\varepsilon}}_{\text{MAP}}$ and $\overline{\boldsymbol{\beta}}$ are linked by

$$\begin{aligned} \overline{\boldsymbol{\sigma}}_{\text{MAP}} &= \nabla_{\boldsymbol{\alpha}}(-\ln \nu)^*(\boldsymbol{\alpha}, \boldsymbol{\delta})_{|\boldsymbol{H}^\dagger \overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\beta}}}, \\ \overline{\boldsymbol{\varepsilon}}_{\text{MAP}} &= \nabla_{\boldsymbol{\delta}}(-\ln \nu)^*(\boldsymbol{\alpha}, \boldsymbol{\delta})_{|\boldsymbol{H}^\dagger \overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\beta}}}. \end{aligned} \qquad (21)$$

By construction, these estimates satisfy the observation equation: $\boldsymbol{\mu} = \boldsymbol{H}\overline{\boldsymbol{\sigma}}_{\text{MAP}} + \overline{\boldsymbol{\varepsilon}}_{\text{MAP}}$. The generalised dual cost function one should then optimise is given by

$$\widehat{\mathcal{L}}(\boldsymbol{\beta}) = (-\ln \nu)^*(\boldsymbol{H}^\dagger \boldsymbol{\beta}, \boldsymbol{\beta}) - \boldsymbol{\beta}^\dagger \boldsymbol{\mu}, \qquad (22)$$

but only provided that $-\ln \nu$ and $(-\ln \nu)^*$ are convex. This is a major restriction when comparing with the MEM approach, where the convexity is guaranteed.

It is easy to check, as is well known, that under Gaussian hypotheses, the dual cost function is that of PSAS again Eq. (3), up to a constant.

3.2 Examples: semi-Gaussian prior law and Bernoulli law

To illustrate the Bayesian method, let us define for the source prior ν a semi-Gaussian law, that is for any cell k of the domain, the local prior pdf ν_k is defined by

$$\begin{cases} \nu_k(\sigma_k) = 0 \quad \text{if} \quad \sigma_k < 0 \\ \nu_k(\sigma_k) = \sqrt{\dfrac{2}{\pi m_k^2}} \exp\left(-\dfrac{\sigma_k^2}{2m_k^2}\right) \quad \text{if} \quad \sigma_k \geq 0, \end{cases}$$

whereas the errors are still modelled by an unbiased normal law of covariance matrix $\boldsymbol{R}$. It can then be shown that, up to a constant,

$$\widehat{\mathcal{L}}(\boldsymbol{\beta}) = \sum_{k=1}^N \frac{1}{2} m_k^2 \left[\boldsymbol{H}^\dagger \boldsymbol{\beta}\right]_k^2 \Theta\left(\left[\boldsymbol{H}^\dagger \boldsymbol{\beta}\right]_k\right)$$
$$+ \frac{1}{2}\boldsymbol{\beta}^\dagger \boldsymbol{R} \boldsymbol{\beta} - \boldsymbol{\beta}^\dagger \boldsymbol{\mu}. \qquad (23)$$

where Θ is the Heaviside function (defined by 1 on $\mathbb{R}_+$, and 0 elsewhere). Both the primal and the dual criteria are convex (it is easy to check that their respective Hessian exists and is positive everywhere), and the duality gap vanishes. The corresponding MAP estimator is

$$\overline{\sigma}_k^{\text{MAP}} = m_k^2 \left[\boldsymbol{H}^\dagger \overline{\boldsymbol{\beta}}\right]_k \Theta\left(\left[\boldsymbol{H}^\dagger \overline{\boldsymbol{\beta}}\right]_k\right). \qquad (24)$$

An application will be presented in Sect. 4.

It is now assumed that the prior law is the same Bernoulli law that was originally used in the MEM framework. Yet, the resulting cost function will be shown to be quite different. Then, for cell k, the local prior pdf is defined by

$$\ln \nu_k(\sigma_k) = \ln \gamma_k \delta_{\sigma_k, m_k} + \ln(1 - \gamma_k)\delta_{\sigma_k, 0}. \qquad (25)$$

To compute the dual cost function, one establishes first

$$(-\ln \nu)^*(\boldsymbol{\alpha}) = \sum_{k=1}^N (m_k \alpha_k + g_k) \Theta(m_k \alpha_k + g_k)$$
$$+ \sum_{k=1}^N \ln(1 - \gamma_k), \qquad (26)$$

where Θ is the Heaviside function again, and $g_k = \ln \gamma_k/(1 - \gamma_k)$. As a consequence, thanks to Eq. (22), the dual cost function is given, up to a constant, by

$$\widehat{\mathcal{L}}(\boldsymbol{\beta}) = \sum_{k=1}^N \left(m_k \left[\boldsymbol{H}^\dagger \boldsymbol{\beta}\right]_k + g_k\right) \Theta\left(m_k \left[\boldsymbol{H}^\dagger \boldsymbol{\beta}\right]_k + g_k\right)$$
$$+ \frac{1}{2}\boldsymbol{\beta}^\dagger \boldsymbol{R} \boldsymbol{\beta} - \boldsymbol{\beta}^\dagger \boldsymbol{\mu}. \qquad (27)$$

This function is continuous on $\mathbb{R}^d$. It is differentiable, except for the $\boldsymbol{\beta}$'s such that there exists a cell index k_0 with $\left[\boldsymbol{H}^\dagger \boldsymbol{\beta}\right]_{k_0} = -g_{k_0}/m$. This leads to numerical hardships for a minimisation by a descent algorithm since the gradient is not a continuous vector field, and may trigger erratic jumps in the course of the iterative optimisation. To circumvent this problem, one can slightly modify locally the cost function in such a way that the gradient is continuous. The function $\alpha\Theta(\alpha)$ is to be replaced by

$$\begin{cases} 0 & \text{if } \alpha < -\epsilon \\ \dfrac{1}{4\epsilon}(\alpha + \epsilon)^2 & \text{if } -\epsilon \leq \alpha \leq \epsilon, \\ \alpha & \text{if } \alpha > \epsilon \end{cases} \qquad (28)$$

with ϵ a deformation parameter ($\epsilon = 0$ means no deformation). The solution will then be *almost* binary: the solution in a few cells may yield an intermediary value, between 0 and m_k, in the case where $\epsilon > 0$.

The MAP estimator is then

$$\overline{\sigma}_k^{\text{MAP}} = \begin{cases} m_k & \text{if } \left[\boldsymbol{H}^\dagger \overline{\boldsymbol{\beta}}\right]_k \geq -g_k/m_k \\ 0 & \text{otherwise} \end{cases} \qquad (29)$$

with a deformation following

$$\overline{\sigma}_k^{\text{MAP}} = \frac{m}{2} \left(1 + \epsilon^{-1} \left(m \left[\boldsymbol{H}^\dagger \overline{\boldsymbol{\beta}} \right]_k + g_k \right) \right)$$

$$\text{if} - \epsilon \le m \left[\boldsymbol{H}^\dagger \overline{\boldsymbol{\beta}} \right]_k + g_k \le \epsilon. \tag{30}$$

As a strict Bayesian approach the solution is binary in the limit ϵ goes to zero: $\overline{\sigma}_k^{\text{MAP}} \in \{0, m_k\}$ whereas in the MEM approach also based on the same Bernoulli prior, the estimate is only constrained to lie in the interval $[0, m_k]$, though 0 and m_k are more likely a priori.

In the following, MEM and MAP estimators will indistinctly be noted $\overline{\boldsymbol{\sigma}}$ and $\overline{\boldsymbol{\varepsilon}}$, since several results will apply to both inferences.

3.3 Comparison of the two inferences

A comparison can be established between the two inference schemes (Mohammad-Djafari, 1996) on a mathematical basis that goes beyond the need of this paper.

However the second order analysis that applies to the MEM inference is to rely partially on this equivalence, as will be explained in the following.

3.3.1 Correspondence for the first-order moments

It can be observed from the results obtained so far that, on general grounds, the two inference schemes yield distinct cost functions for the same source and errors priors. However, in the particular case of Gaussian hypotheses, the two schemes lead, up to a constant, to the same cost-function, that is to say to the least squares criterion. When the priors are not Gaussian, the resulting criterion is not quadratic, possibly of higher order in the control variables.

Because we were able for both inferences to recast them in the form of a MAP criterion (primal cost functions), it is observed that passing from one criterion to the other involves the formal transformation

$$\widehat{\nu}^* \Longleftrightarrow (-\ln \nu) \quad \text{or} \quad \widehat{\nu} \Longleftrightarrow (-\ln \nu)^*. \tag{31}$$

which can be read off from the chain of inferences of Sect. 2.3. In the Gaussian case, both terms of these equivalences are equal, up to a constant independent on $\boldsymbol{\sigma}$. For any prior, using this translation leads through the chain of inferences to the same cost function and therefore to the same estimator.

3.3.2 Re-defining the MEM analysed second-order moments

However this correspondence fails for second-order moments if the MEM approach is applied blindly. It was stressed in (Bocquet, 2007) that this could be ascribed to the "on the mean" type of approach because errors and source

fluctuations are considered independent in the MEM approach whereas they are linked by the measurement equations in the exact Bayesian approach. In particular it was shown in the appendix of (Bocquet, 2007), that the analysis covariance matrix obtained from the MEM a posteriori pdf is different from the mathematically sound Best Linear Unbiased Estimator [BLUE] result (the former actually identifies with the background covariance matrix).

This drawback can be circumvented by inferring directly on the primal cost function Eq. (12), because the errors $\boldsymbol{\varepsilon}$ are replaced there with $\boldsymbol{\mu} - \boldsymbol{H}\boldsymbol{\sigma}$. Using the chain of inferences of Sect. 2.3, the primal cost function is then re-defined as resulting from a Bayesian MAP estimator. This way, a proper second-order analysis that mimics the strict Bayesian inference is defined. In the Gaussian case, the BLUE analysis covariance matrix is then predicted. This re-definition of the MEM second-order moments will be used throughout the rest of the paper.

In the general case, this allows practically for a unified treatment in the second-order analysis of these two inference mechanisms. Most results ahead, though written down in the MEM context, can be transposed to a Bayesian MAP context using the transformation Eq. (31).

4 Application setup

The aim of this section is to illustrate the inferences described formally earlier. The retrievals will all abide to the following setup. So will the second-order analysis that will be carried out in Sect. 5.

The reconstruction performed on ETEX-I is based on the results of (Bocquet, 2007) and (Krysta et al., 2007). It essentially makes use of a 911-measurement dataset, which represents most of the positive observations of the tracer (that is to say above the background level). The space resolution of the inversion grid is $2.25° \times 2.25°$ while the time resolution is one hour. This represents about 20×10^3 cells for the inversion grid domain which encompasses all Western Europe, while the simulation domain encompasses all Western and Central Europe. A finer (inversion grid) resolution could have been used. However the quality of the retrieval starts degrading then (Bocquet, 2007), so that it is safer to choose the coarser one for the demonstrations ahead. The Jacobian matrix has been built using the chemistry and transport model POLAIR3D, which has been validated for such tracer in (Quélo et al., 2007). The model was driven by the 40-year re-analysis ECMWF meteorological data (wind fields, and fields necessary to the diagnostic of the vertical turbulent diffusion coefficient K_z). Since the number of observations is much lower than the number of variables to invert, the rows of $\boldsymbol{H}$ are computed using POLAIR3D in an (approximate) adjoint mode using a technique that has now been vastly advertised in this context (Issartel and Baverel, 2003; Hourdin and Talagrand, 2006; Davoine and Bocquet, 2007).
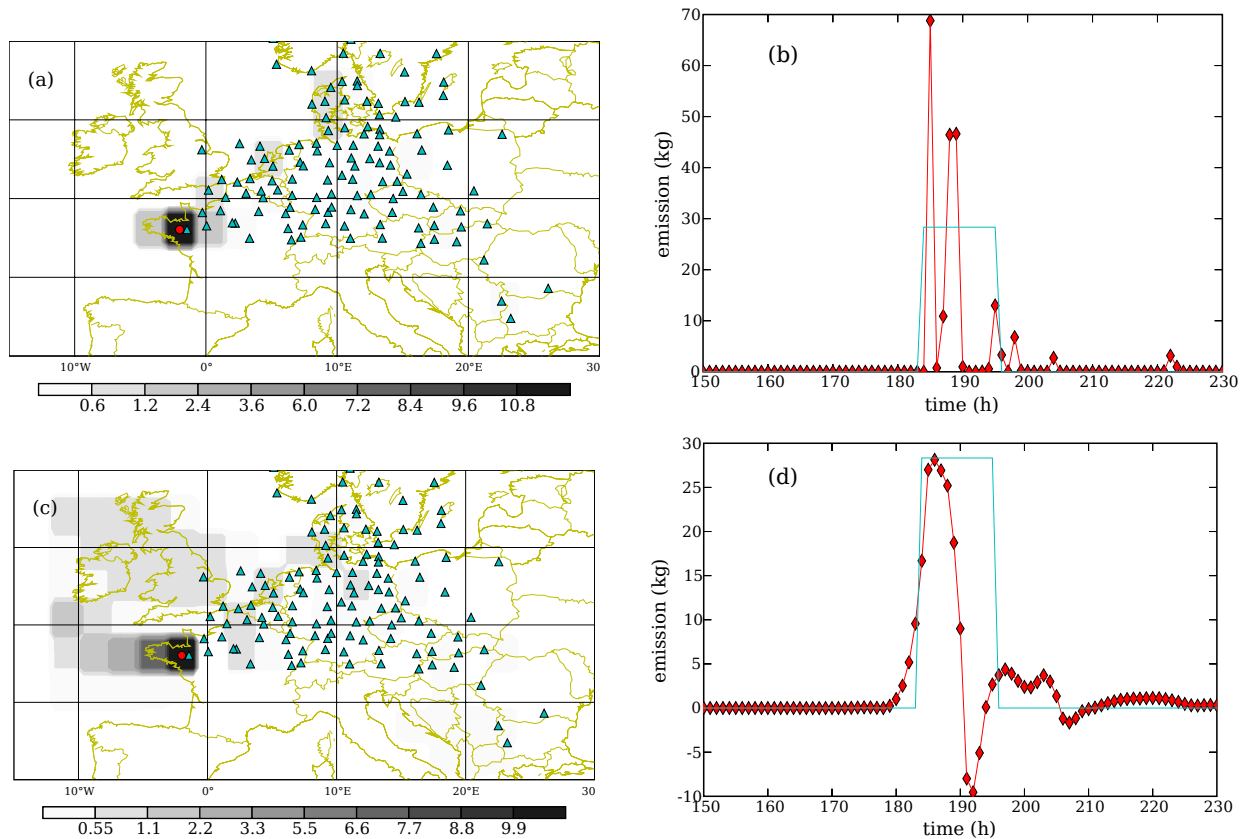
Fig. 1. Reconstruction through a Bernoulli prior on a basis of a MEM inference (top row), and through a Gaussian prior (MEM and Bayes inferences) (bottom row). Panels **(a)** and **(c)** represents density plot of the integrated released mass in time for every space grid cell. The triangles represent the stations involved in the observations. Units are in kg. Panels **(b)** and **(d)** represent the profile of the retrieved source in true release grid cell. The step function indicates the true emission rate profile.

Two types of prior and therefore two types of reconstruction will serve as illustrations throughout the second-order analysis of Sect. 5.

Both use a Gaussian prior for the errors, with a diagonal covariance matrix $R=\chi I$, of homoscedastic diagonal coefficients. $\sqrt{\chi}$ is chosen optimally by an L-curve a posteriori analysis and confirmed by a maximum likelihood estimation, in a sense defined elsewhere (Davoine and Bocquet, 2007; Krysta et al., 2007) to be $\sqrt{\chi}=0.3 \, \text{ng m}^{-3}$.

The first source prior is Gaussian, without bias and with a background diagonal covariance matrix given by $B=m^2 I$. m is estimated optimally to be $m=0.025 \, m_0/\lambda$ for our setup. m_0 is a reference mass that is defined to be the total mass released in one hour of the ETEX-I release (more than 28 kg) and λ is a scale parameter that is conventionally set to 4 for the present inversion grid resolution. Note that from the cost function perspective, $\sqrt{\chi}$ and m represent the same degree of freedom, so that $\sqrt{\chi}$ is practically fixed a priori, while m is later estimated.

The second source prior is a Bernoulli prior which leads to the cost function Eq. (10) and the estimator Eq. (11). γ_k is uniformly set to a very low value of 10^{-15}, expressing the improbable nature of an accident in a given cell. m_k is uniformly and optimally set to $m_k \equiv 5 \, m_0/\lambda^2$. (This conventional λ^2 factor differs from the not less conventional λ Gaussian factor because of scaling arguments on prior statistics.) Note that this time, m and $\sqrt{\chi}$ do not represent the same degree of freedom, so that they have been jointly estimated by iterative L-curve techniques following (Davoine and Bocquet, 2007; Krysta et al., 2007).

Both priors have been chosen because they illustrate the main differences between a reconstruction based on Gaussian and non-Gaussian hypotheses.

The main characteristics of the reconstructions are given in Table 1: mass retrieved in the true source grid cell $\mathcal{M}_m$, total retrieved mass $\mathcal{M}$, total relative entropy of the solution $\mathcal{K}_{\sigma,\varepsilon}$ or $\mathcal{L}_{\sigma,\varepsilon}$, relative entropy of the retrieved source $\mathcal{K}_\sigma$ or $\mathcal{L}_\sigma$, and relative entropy of the retrieved errors $\mathcal{K}_\varepsilon$ or $\mathcal{L}_\varepsilon$. The results of the retrievals are illustrated in Fig. 1.

Table 1. Main scalar results for four reference inversions. $\mathcal{M}_m$ is the mass found at the ETEX-I true release grid cell. $\mathcal{M}$ is the total mass retrieved. For the Bayesian inversions, the dual cost functions values are given up to constants (depending only on prior parameters).

	Method	$\mathcal{M}_m$ (kg)	$\mathcal{M}$ (kg)	$\mathcal{L}_{\sigma,\varepsilon}$	$\mathcal{L}_\sigma$	$\mathcal{L}_\varepsilon$
Gaussian	MEM average	241	684	1168	347	821
Bernoulli	MEM average	222	342	1299	69	1230
	Method	$\mathcal{M}_m$ (kg)	$\mathcal{M}$ (kg)	$\widehat{\mathcal{L}}_{\sigma,\varepsilon}$	$\widehat{\mathcal{L}}_\sigma$	$\widehat{\mathcal{L}}_\varepsilon$
Gaussian	Bayes MAP	241	684	1168	347	821
Semi-Gaussian	Bayes MAP	208	458	1450	152	1298
Bernoulli	Bayes MAP	177	282	1568	273	1295

They stress the difference between a positive and an unbiased law. In the Gaussian prior case, negative values for the source are recovered. The localisation is also weaker because the Gaussian is less constraining than the Bernoulli prior (35% of the mass in the correct grid-cell as compared to 65%): it brings in less information in the assimilation system.

The exact Bayesian inferences given as examples earlier have also been implemented. The semi-Gaussian approach is illustrated in Fig. 2. The reconstruction makes use again of an optimally determined m through an L-curve analysis. It does not significantly differ from the Gaussian case, i.e. $m \simeq 0.025 m_0/\lambda$. To illustrate the Bernoulli prior Bayesian inference, $m = m_0$ has been chosen, while ϵ is very small (it depends on the machine precision). The main scalar results are also given in Table 1. The results are displayed in Fig. 2. The emission rates in the profile are 0 or m_0, according to this inference constraints. One value is comprised in between and gives away the freedom granted by the ϵ deformation parameter.

The semi-Gaussian prior leads, as it should, to positive values for the source, but it does not constrain the location of the source as well as the MEM Bernoulli inference (45% of the mass is recovered in the correct grid-cell). The Bernoulli prior leads to a very good reconstruction, but the actual release rate was given as prior information.

Note that all four inversions presented here suffer from a strong depletion in the source reconstruction at the end of the release true window. We believe it can be explained on physical grounds because of the modelling of the vertical turbulence diffusion parameter K_z that is strongly affected by convective rain in our model. This is precisely the case at the end of the release interval in the true emission grid-cell.

5   Second-order sensitivity analysis

Two non-Gaussian inference schemes are now available, that can often be unified into the same formalism. Besides, an application setup on ETEX-I has been described to test them

with. We are now ready to proceed to the second-order analysis discussion.

In order to be explicit in the following mathematical expressions, it is assumed in the following that the prior density $\nu(\boldsymbol{\sigma}, \boldsymbol{\varepsilon})$ is separable into $\nu(\boldsymbol{\sigma}, \boldsymbol{\varepsilon}) = \nu(\boldsymbol{\sigma})\zeta(\boldsymbol{\varepsilon})$, since it is reasonable to assume that the prior knowledge is to a first approximation independent from the knowledge on the errors. Then the dual cost function splits according to

$$\widehat{\mathcal{L}}(\boldsymbol{\beta}) = \widehat{\nu}(\boldsymbol{H}^\dagger \boldsymbol{\beta}) + \widehat{\zeta}(\boldsymbol{\beta}) - \boldsymbol{\beta}^\dagger \boldsymbol{\mu}, \qquad (32)$$

where $\widehat{\nu}$ and $\widehat{\zeta}$ definitions are particular cases of Eq. (8). The related primal cost function is

$$\mathcal{L}(\boldsymbol{\sigma}) = \widehat{\nu}^*(\boldsymbol{\sigma}) + \widehat{\zeta}^*(\boldsymbol{\mu} - \boldsymbol{H}\boldsymbol{\sigma}), \qquad (33)$$

with obvious definitions of $\widehat{\nu}^*$ and $\widehat{\zeta}^*$ stemming from Eq. (13).

5.1   Analysis covariance matrix

Remember that, in the MEM case, in order to carry out a proper second-order analysis, one needs to deduce it from the primal cost function. Defining the Hessians matrices

$$\mathcal{H}_{k,l}^{\boldsymbol{\sigma},\boldsymbol{\sigma}} = \partial_{\sigma_k} \partial_{\sigma_l} \widehat{\nu}^*_{|\overline{\boldsymbol{\sigma}}}, \quad \mathcal{H}_{i,j}^{\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}} = \partial_{\varepsilon_i} \partial_{\varepsilon_j} \widehat{\zeta}^*_{|\boldsymbol{\mu} - \boldsymbol{H}\overline{\boldsymbol{\sigma}}}. \qquad (34)$$

one obtains

$$\mathcal{P}_a^{-1} = \mathcal{H}^{\boldsymbol{\sigma},\boldsymbol{\sigma}} + \boldsymbol{H}^\dagger \mathcal{H}^{\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}} \boldsymbol{H}. \qquad (35)$$

The dimensionality of spaces plays an important role and justifies the use of the dual space. That is why it is useful to express $\mathcal{P}_a$ through a dual formulation. Let us define

$$\mathcal{H}_{k,l}^{\boldsymbol{\alpha},\boldsymbol{\alpha}} = \partial_{\alpha_k} \partial_{\alpha_l} \widehat{\nu}_{|\boldsymbol{H}^\dagger \overline{\boldsymbol{\beta}}}, \quad \mathcal{H}_{i,j}^{\boldsymbol{\delta},\boldsymbol{\delta}} = \partial_{\delta_k} \partial_{\delta_l} \widehat{\zeta}_{|\overline{\boldsymbol{\beta}}}. \qquad (36)$$

Because the Hessians of two Fenchel-conjugated functionals are reciprocal inverse matrices, one has

$$\mathcal{H}^{\boldsymbol{\alpha},\boldsymbol{\alpha}} = \left(\mathcal{H}^{\boldsymbol{\sigma},\boldsymbol{\sigma}}\right)^{-1}, \qquad \mathcal{H}^{\boldsymbol{\delta},\boldsymbol{\delta}} = \left(\mathcal{H}^{\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}}\right)^{-1}. \qquad (37)$$

Thanks to this property and to the Sherman-Morrison-Woodbury [SMW] formula, one obtains

$$\mathcal{P}_a = \mathcal{H}^{\boldsymbol{\alpha},\boldsymbol{\alpha}} - \mathcal{H}^{\boldsymbol{\alpha},\boldsymbol{\alpha}} \boldsymbol{H}^\dagger \left(\mathcal{H}^{\boldsymbol{\delta},\boldsymbol{\delta}} + \boldsymbol{H}\mathcal{H}^{\boldsymbol{\alpha},\boldsymbol{\alpha}} \boldsymbol{H}^\dagger\right)^{-1} \boldsymbol{H}\mathcal{H}^{\boldsymbol{\alpha},\boldsymbol{\alpha}}. \qquad (38)$$

It is straightforward to check that in the Gaussian case, the Hessians identify with the background and error covariance matrices, as they should,

$$\mathcal{H}_{k,l}^{\boldsymbol{\alpha},\boldsymbol{\alpha}} = [\boldsymbol{B}]_{kl}, \qquad \mathcal{H}_{i,j}^{\boldsymbol{\delta},\boldsymbol{\delta}} = [\boldsymbol{R}]_{ij}, \qquad (39)$$

which are independent of the solution. Otherwise, these Hessians do depend on the solution. For example, within the MEM framework, the Hessian related to the Bernoulli prior law is

$$\mathcal{H}_{k,l}^{\boldsymbol{\alpha},\boldsymbol{\alpha}} = \delta_{k,l} m_k^2 \frac{\gamma_k(1 - \gamma_k) e^{m_k [\boldsymbol{H}^\dagger \overline{\boldsymbol{\beta}}]_k}}{\left(1 - \gamma_k + \gamma_k e^{m_k [\boldsymbol{H}^\dagger \overline{\boldsymbol{\beta}}]_k}\right)^2}. \qquad (40)$$
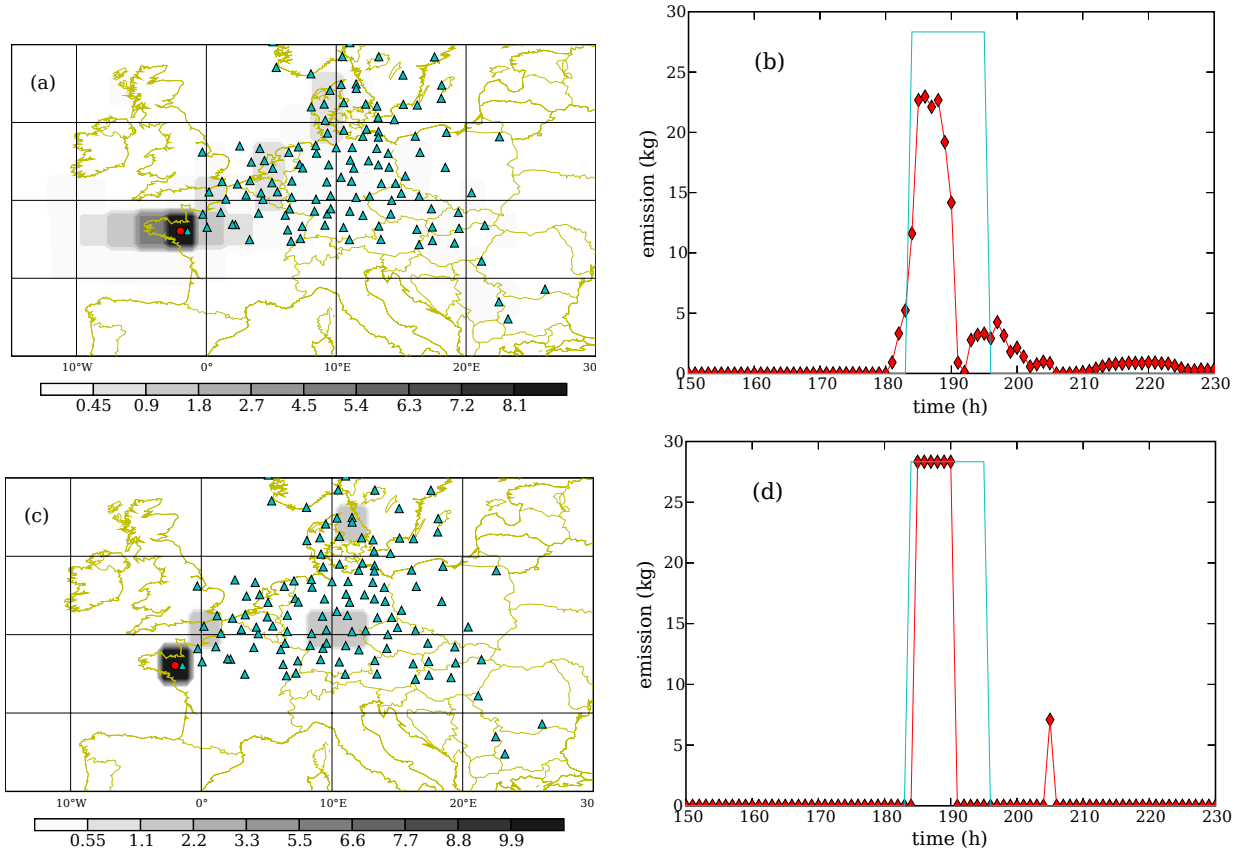
Fig. 2. Top row: reconstruction through a semi-Gaussian prior on the basis of a Bayesian inference and a MAP estimator. Bottom row: reconstruction through a Bernoulli prior on the basis of a Bayesian inference and a MAP estimator. Panels **(a)** and **(c)** represent a density plot of the integrated released mass in time for every space grid cell. The triangles represent the stations involved in the observations. Units are in kg. Panels **(b)** and **(d)** represent the profile of the retrieved source in the true release grid cell. The step function indicates the true emission rate profile.

Therefore a second-order analysis on the reconstruction of an atmospheric tracer source, whether it relies on a non-Gaussian cost function of any of the two types considered here, calls for a generalisation to accommodate this dependence.

In Fig. 3 a partial representation of the ETEX-I confidence matrix $\mathcal{P}_a^{-1}$ is given in the Gaussian case. It can also be interpreted as an information matrix. It essentially shows how much confidence in the observations is propagated and diffused by the model ($\boldsymbol{H}^{\dagger}\mathcal{H}^{\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}}\boldsymbol{H}$) plus a uniform background confidence term ($\mathcal{H}^{\boldsymbol{\sigma},\boldsymbol{\sigma}}$). In the Gaussian case, it does not depend on the solution. It is remarkable that the ETEX-I release (in Brittany, France) is far from the most trustworthy region (Central Europe). Incidentally, it shows that even though ETEX-I has a well-suited setup for dispersion modelling validation of numerical transport models, it is not the easiest setup for an inverse modelling experiment.

In Fig. 4, a partial representation of the ETEX-I covariance matrix $\mathcal{P}_a$ is given in the non-Gaussian case.

The dependence on the solution is very significant, for at least two reasons. First the true release site area corresponds to a much stronger uncertainty, likely because it is there that the optimisation puts most of the tracer mass. Second, in cells where the solution is close to zero, the uncertainty is also very small. This reflects the structure of the Bernoulli law: negative estimates are forbidden, so that fluctuations (hence variances) around a near zero solution are strongly damped. This is the reason why, in this case, it was chosen to represent the covariance matrix instead of the confidence matrix.

In (Bocquet, 2005a), a performance indicator

$$\rho = \widehat{v}^*(\overline{\boldsymbol{\sigma}}_{\mathrm{MEM}}, \overline{\boldsymbol{\varepsilon}}_{\mathrm{MEM}})/\widehat{v}^*(\boldsymbol{\sigma}_t, \boldsymbol{\varepsilon}_t) \qquad (41)$$

fully consistent with the MEM approach was proposed: ρ close to 1 indicates an excellent reconstruction, whereas ρ close to 0 indicates a failure of the reconstruction. An expansion of ρ in the discrepancy between the estimate $\overline{\boldsymbol{\sigma}}_{\mathrm{MEM}}$ and
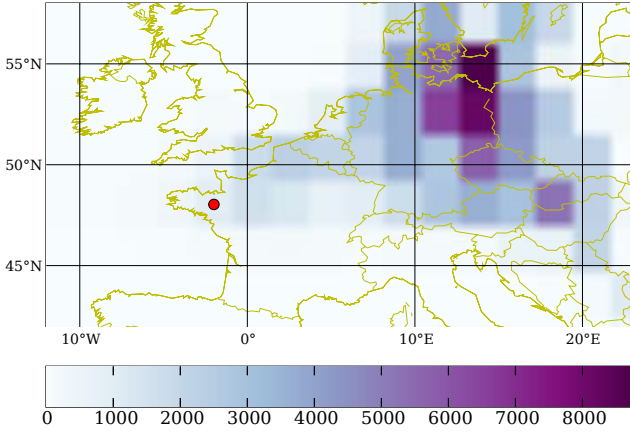
Fig. 3. Partial graphical representation of the analysis confidence matrix in the Gaussian prior case. The density plot represents the the diagonal of the matrix, with a sum over time steps in each space grid cell. This is only an approximation as off-diagonal elements connecting different times in the same space grid cell are not taken into account. The units are in kg^{-2}. For clarity, the uniform background term contribution has been subtracted.

the true source $\boldsymbol{\sigma}_t$ leads to

$$\rho = 1 - \frac{1}{2\,\widehat{\nu}^*(\overline{\boldsymbol{\sigma}}_{\mathrm{MEM}},\,\overline{\boldsymbol{\varepsilon}}_{\mathrm{MEM}})}\,(\overline{\boldsymbol{\sigma}}_{\mathrm{MEM}} - \boldsymbol{\sigma}_t)^\dagger\,\mathcal{P}_a^{-1}\,(\overline{\boldsymbol{\sigma}}_{\mathrm{MEM}} - \boldsymbol{\sigma}_t)$$
$$+ \mathrm{O}(||\overline{\boldsymbol{\sigma}}_{\mathrm{MEM}} - \boldsymbol{\sigma}_t||^2)\,. \tag{42}$$

This makes the connection between ρ and the confidence matrix $\mathcal{P}_a^{-1}$ explicit.

5.2 Averaging kernels of the source and the errors

The second-order analysis is to be built on the equation satisfied by the solution $\overline{\boldsymbol{\sigma}}$ and $\overline{\boldsymbol{\varepsilon}}$. Since in most cases considered here the cost functions are more than quadratic, the analysis will only be local, in the vicinity of the solution, even though all derivations are without approximations.

5.2.1 Sensitivity of the retrieved source to the real source

Consider the optimality condition on the primal cost function $\mathcal{L}(\boldsymbol{\sigma})$:

$$0 = \partial_{\sigma_k}\mathcal{L} = \partial_{\sigma_k}\widehat{\nu}^* - H_{ik}\partial_{\varepsilon_i}\widehat{\zeta}^*\,. \tag{43}$$

Here, and below, the Einstein convention on the repetition of indices is used. To extract the sensitivity of the retrieved source to the real source (averaging kernel of the source), one differentiates Eq. (43) with respect to the coefficients of $\boldsymbol{\sigma}_t$:

$$\frac{\partial\overline{\sigma}_h}{\partial\sigma_l^t}\left\{\partial_{\sigma_h}\partial_{\sigma_k}\widehat{\nu}^* + H_{jh}\partial_{\varepsilon_i}\partial_{\varepsilon_j}\widehat{\zeta}^*H_{ik}\right\} = H_{ik}\frac{\partial\mu_j}{\partial\sigma_l^t}\partial_{\varepsilon_i}\partial_{\varepsilon_j}\widehat{\zeta}^*\,. \tag{44}$$

Knowing that $\boldsymbol{\mu} = \boldsymbol{H}\overline{\boldsymbol{\sigma}} + \overline{\boldsymbol{\varepsilon}} = \boldsymbol{H}\boldsymbol{\sigma}_t + \boldsymbol{\varepsilon}_t$, one gets

$$\frac{\partial\overline{\boldsymbol{\sigma}}}{\partial\boldsymbol{\sigma}_t} = \left(\mathcal{H}^{\sigma,\sigma} + \boldsymbol{H}^\dagger\mathcal{H}^{\varepsilon,\varepsilon}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^\dagger\mathcal{H}^{\varepsilon,\varepsilon}\boldsymbol{H}\,. \tag{45}$$

The following matrix indices convention has been chosen

$$\left[\frac{\partial\boldsymbol{A}}{\partial\boldsymbol{B}}\right]_{ij} = \frac{\partial\boldsymbol{A}_i}{\partial\boldsymbol{B}_j}\,. \tag{46}$$

An analog result can be obtained using a dual route. The Lagrange parameters are solutions of

$$0 = \partial_{\beta_i}\widehat{\mathcal{L}} = H_{ik}\partial_{\alpha_k}\widehat{\nu} + \partial_{\delta_i}\widehat{\zeta} - \mu_i\,. \tag{47}$$

A differentiation with respect to the measurements $\boldsymbol{\mu}$ yields (in matrix form)

$$\frac{\partial\overline{\boldsymbol{\beta}}}{\partial\boldsymbol{\mu}} = \left(\boldsymbol{H}\mathcal{H}^{\alpha,\alpha}\boldsymbol{H}^\dagger + \mathcal{H}^{\delta,\delta}\right)^{-1}\,. \tag{48}$$

To obtain the sensitivity of the retrieved source to the measurements, one applies $\overline{\boldsymbol{\sigma}} = (\partial_{\boldsymbol{\alpha}}\widehat{\nu})_{|\boldsymbol{H}^\dagger\overline{\boldsymbol{\beta}}}$. This gives

$$\frac{\partial\overline{\boldsymbol{\sigma}}}{\partial\boldsymbol{\mu}} = \mathcal{H}^{\alpha,\alpha}\boldsymbol{H}^\dagger\left(\boldsymbol{H}\mathcal{H}^{\alpha,\alpha}\boldsymbol{H}^\dagger + \mathcal{H}^{\delta,\delta}\right)^{-1}\,. \tag{49}$$

This, in turn, yields the averaging kernel:

$$\frac{\partial\overline{\sigma}_k}{\partial\sigma_l^t} = \frac{\partial\overline{\sigma}_k}{\partial\mu_i}\frac{\partial\mu_i}{\partial\sigma_l^t} = \frac{\partial\overline{\sigma}_k}{\partial\mu_i}H_{il} = H_{jm}\frac{\partial\overline{\beta}_j}{\partial\mu_i}\partial_{\alpha_k}\partial_{\alpha_m}H_{il}\,, \tag{50}$$

so that

$$\frac{\partial\overline{\boldsymbol{\sigma}}}{\partial\boldsymbol{\sigma}_t} = \mathcal{H}^{\alpha,\alpha}\boldsymbol{H}^\dagger\left(\boldsymbol{H}\mathcal{H}^{\alpha,\alpha}\boldsymbol{H}^\dagger + \mathcal{H}^{\delta,\delta}\right)^{-1}\boldsymbol{H}\,. \tag{51}$$

This formula could have also been obtained from the SMW formula and from Eq. (45).

5.2.2 Sensitivity of the retrieved errors to the real errors

Very similar results can be obtained for the errors. For example the errors averaging kernel is

$$\frac{\partial\overline{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\varepsilon}_t} = \frac{\partial\overline{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\mu}} = \boldsymbol{I} - \boldsymbol{H}\left(\mathcal{H}^{\sigma,\sigma} + \boldsymbol{H}^\dagger\mathcal{H}^{\varepsilon,\varepsilon}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^\dagger\mathcal{H}^{\varepsilon,\varepsilon}$$
$$= \mathcal{H}^{\delta,\delta}\left(\boldsymbol{H}\mathcal{H}^{\alpha,\alpha}\boldsymbol{H}^\dagger + \mathcal{H}^{\delta,\delta}\right)^{-1} \tag{52}$$

$\partial_{\boldsymbol{\varepsilon}_t}\overline{\boldsymbol{\varepsilon}}$ can be shown to be symmetric positive definite matrix with eigenvalues in the range [0,1]. The closest to 1, the more truthful the reconstruction is. In Fig. 5, we have chosen to represent the diagonal of $\boldsymbol{I}-\partial_{\boldsymbol{\varepsilon}_t}\overline{\boldsymbol{\varepsilon}}$. The difference between the two types of reconstructions is significant. It emphasises how differently the same information (observations) is interpreted by the optimality system.
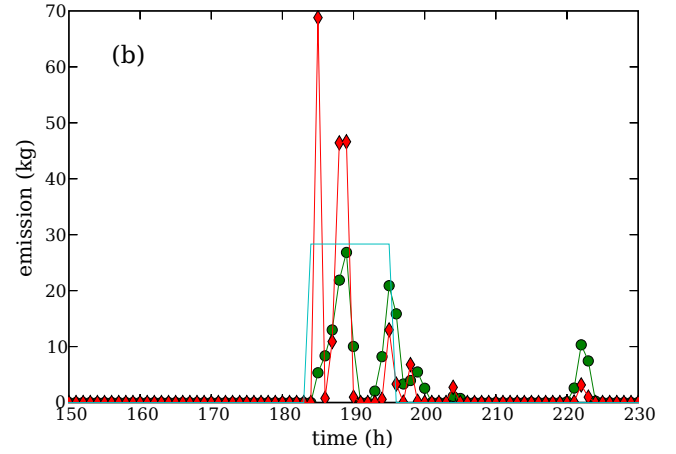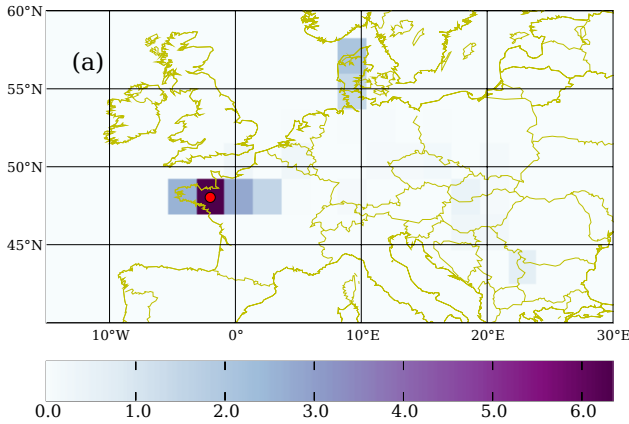
Fig. 4. Partial graphical representation of the analysis covariance matrix in the Bernoulli prior case. Panel **(a)** represents the diagonal of the matrix, with a sum over time steps in each space grid cell. This is only an approximation as off-diagonal elements connecting different times in the same space grid cell are not taken into account. The units are in kg×kg. Panel **(b)** displays the profile of the reconstructed source (◆), and the profile of the square root of covariance matrix elements in the true release grid cell (●) (no approximation is used). The units are in kg.

5.3 Lagrange multipliers, errors and sensitivities

Before any further investigation of second-order properties, one should consider the Lagrange multipliers, that are readily available as the vector minimising the dual criterion. It is known in convex analysis that they are prone to a physical interpretation (see for example (Nocedal and Wright, 2006)). The primal problem on $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ consists in optimising $\mathcal{L}(\boldsymbol{\sigma}, \boldsymbol{\varepsilon}) = \widehat{v}^*(\boldsymbol{\sigma}, \boldsymbol{\varepsilon}) + \boldsymbol{\beta}^\dagger (\boldsymbol{\mu} - \boldsymbol{H}\boldsymbol{\sigma} - \boldsymbol{\varepsilon})$. At the saddle point, one has $0 = \partial_{\varepsilon_i}\mathcal{L} = \partial_{\varepsilon_i}\widehat{v}^* - \beta_i$. The alternate formulation of the primal criterion is $\mathcal{L}(\boldsymbol{\sigma}) = \widehat{v}^*(\boldsymbol{\sigma}, \boldsymbol{\mu} - \boldsymbol{H}\boldsymbol{\sigma})$. It can be considered at its minimum point $\overline{\boldsymbol{\sigma}}$, understood as a function of $\boldsymbol{\mu}$. Then

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\mu_i}(\overline{\boldsymbol{\sigma}}(\boldsymbol{\mu})) = \frac{\mathrm{d}\overline{\sigma}_k}{\mathrm{d}\mu_i}\partial_{\sigma_k}\mathcal{L} + \frac{\partial\widehat{v}^*}{\partial\varepsilon_i} = \overline{\beta}_i ,$$

because at the minimum $\partial_{\sigma_k}\mathcal{L} = 0$. As a consequence, the multipliers $\overline{\boldsymbol{\beta}}$ are a measure of the sensitivity of the cost-function to the measurements. The larger the $\overline{\beta}_i$, the more constraining is the related measurement μ_i onto the optimality system. A similar derivation can be performed on the dual criterion. Ultimately

$$\overline{\beta}_i = -\frac{\mathrm{d}\widehat{\mathcal{L}}}{\mathrm{d}\mu_i} = \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\mu_i} . \qquad (53)$$

When the errors are of Gaussian nature (it was chosen so here), the retrieved errors are of the form $\overline{\boldsymbol{\varepsilon}} = \boldsymbol{R}\overline{\boldsymbol{\beta}}$. Provided that the errors are homocedastic $\boldsymbol{R} = \chi\boldsymbol{I}$, then the errors are given by the Lagrange multipliers up to a scale factor. It seems consistent that the stronger the constraint due to μ_i, the stronger the diagnosed error related to μ_i.

The Lagrange multipliers $\overline{\boldsymbol{\beta}}$ are plot in Fig. 6 for the reconstructions based on the two typical priors.

Although different, these plots exhibit very similar magnitude for the $\overline{\beta}_i$. From Eq. (53), the $\overline{\boldsymbol{\beta}}$ represent the marginal information contribution of the observations to the total relative entropy (or total gain of information). What makes a significant difference between the two types of reconstruction is the way the optimality system dispatches the total gain of information between the source and the noise, not really the total gain itsef which is mainly driven by the measurements information content. This point is the foccuss of the next section.

5.4 Information content and observations

5.4.1 Marginal contribution of the observations to the source and errors relative entropies

One way of estimating the influence of observations on the reconstruction is to compute the sensitivities $\partial_{\boldsymbol{\mu}}\overline{\boldsymbol{\sigma}}$ or $\partial_{\boldsymbol{\mu}}\overline{\boldsymbol{\varepsilon}}$, which turn out to be by-products of the averaging kernels derivations. The graphical representation and interpretation of such sensitivities can be awkward since they are represented by very large matrices ($N \times d$ and $d \times d$, respectively). A more concise approach would be to define a wisely chosen sensitivity function that depends on the observations. Such an approach has been fostered for instance in (Le Dimet et al., 1997; Ngodock, 1996). Here, we choose for such a function the information content that a given obervation brings to the inversion. Obviously, this contribution depends on all other observations, so that only a marginal contribution to the information used in the source inversion can be computed from information theory.
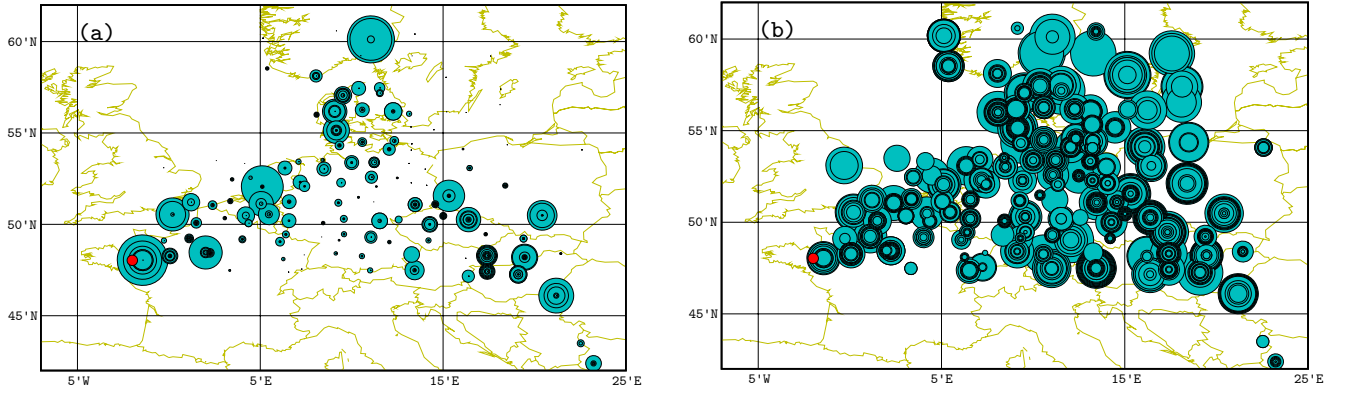
Fig. 5. Plot of the diagonal of $I - \partial_{\varepsilon_t}\overline{\varepsilon}$. The disks radii of the disks are proportional to the diagonal elements $[I - \partial_{\varepsilon_t}\overline{\varepsilon}]_{ii}$. Panel **(a)** is related to the use of the Bernoulli prior, whereas panel **(b)** is related to the use of the Gaussian prior. The same scale radius applies to both plot.

The information is rigorously measured by the relative entropy of the source. To avoid any confusion, note this is not specific to the MEM inference, but also applies to the Bayesian inference.

5.4.2 MEM inference

However in the specific case of the MEM approach, the relative entropy can be obtained as a by-product of the optimisation, which makes it simpler to estimate numerically, but also analytically. By definition, from both dual (second line, first member) and primal (second line, second member) points of view, the source relative entropy is:

$$
\begin{aligned}
\mathcal{K}_{\boldsymbol{\sigma}} &= \sum_{\boldsymbol{\sigma}} p(\boldsymbol{\sigma}) \ln\left(\frac{p(\boldsymbol{\sigma})}{\nu(\boldsymbol{\sigma})}\right) \\
&= \overline{\boldsymbol{\sigma}}^{\dagger} \boldsymbol{H}^{\dagger}\overline{\boldsymbol{\beta}} - \widehat{\nu}(\boldsymbol{H}^{\dagger}\overline{\boldsymbol{\beta}}) = \widehat{\nu}^{*}(\overline{\boldsymbol{\sigma}}),
\end{aligned}
\tag{54}
$$

at the solution point. So that the sensitivity of $\mathcal{K}_{\boldsymbol{\sigma}}$ to μ_i is

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\mu_i}\left\{\overline{\boldsymbol{\sigma}}^{\dagger}\boldsymbol{H}^{\dagger}\overline{\boldsymbol{\beta}} - \widehat{\nu}(\boldsymbol{H}^{\dagger}\overline{\boldsymbol{\beta}})\right\} \\
&= \overline{\beta}_j H_{jk}\frac{\partial\overline{\sigma}_k}{\partial\mu_i} + \frac{\partial\overline{\beta}_j}{\partial\mu_i}H_{jk}\left(\overline{\sigma}_k - \frac{\partial\widehat{\nu}}{\partial\alpha_k}\right) \\
&= \overline{\beta}_j H_{jk}\frac{\partial\overline{\sigma}_k}{\partial\mu_i},
\end{aligned}
\tag{55}
$$

because, at the optimality, one has $\partial_{\boldsymbol{\alpha}}\widehat{\nu}_{|\boldsymbol{H}^{\dagger}\boldsymbol{\beta}} = \overline{\boldsymbol{\sigma}}$. Because, we know $\partial_{\mu_i}\overline{\sigma}_k$ from Eq. (49), one gets

$$
\begin{aligned}
\frac{\mathrm{d}\mathcal{K}_{\boldsymbol{\sigma}}}{\mathrm{d}\boldsymbol{\mu}} &= \left(\boldsymbol{H}\mathcal{H}^{\alpha,\alpha}\boldsymbol{H}^{\dagger} + \mathcal{H}^{\delta,\delta}\right)^{-1}\boldsymbol{H}\mathcal{H}^{\alpha,\alpha}\boldsymbol{H}^{\dagger}\overline{\boldsymbol{\beta}} \\
&= \mathcal{H}^{\varepsilon,\varepsilon}\boldsymbol{H}\left(\mathcal{H}^{\sigma,\sigma} + \boldsymbol{H}^{\dagger}\mathcal{H}^{\varepsilon,\varepsilon}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\dagger}\overline{\boldsymbol{\beta}}.
\end{aligned}
\tag{56}
$$

The second line has been obtained through a similar derivation on $\widehat{\nu}^{*}(\overline{\boldsymbol{\sigma}})$ directly, or using the SMW formula.

The marginal information that is brought in by an observation μ_i onto the errors, which is often considered as lost in noise, can be obtained in the same manner. One gets

$$
\begin{aligned}
\frac{\mathrm{d}\mathcal{K}_{\boldsymbol{\varepsilon}}}{\mathrm{d}\boldsymbol{\mu}} &= \left(\boldsymbol{H}\mathcal{H}^{\alpha,\alpha}\boldsymbol{H}^{\dagger} + \mathcal{H}^{\delta,\delta}\right)^{-1}\boldsymbol{H}\mathcal{H}^{\delta,\delta}\overline{\boldsymbol{\beta}} \\
&= \overline{\boldsymbol{\beta}} - \mathcal{H}^{\varepsilon,\varepsilon}\boldsymbol{H}\left(\mathcal{H}^{\sigma,\sigma} + \boldsymbol{H}^{\dagger}\mathcal{H}^{\varepsilon,\varepsilon}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\dagger}\overline{\boldsymbol{\beta}}.
\end{aligned}
\tag{57}
$$

As a consequence of Eq. (56) and Eq. (57), one has

$$
\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\mu}}\left(\mathcal{K}_{\boldsymbol{\sigma}} + \mathcal{K}_{\boldsymbol{\varepsilon}}\right) = \overline{\boldsymbol{\beta}},
\tag{58}
$$

which has already been derived from the total entropy (cost function in the MEM formalism) Eq. (53).

These sensitivities are plot in Fig. 7.

The results for ETEX-I clearly emphasise the difference between the two reference schemes. The Gaussian source retrieval is mostly influenced by the closest observations (those obtained in France), whereas in the non-Gaussian inversion it is also influenced by farther measurements as far as the Netherlands, with a non-negligible contribution from the Nordic observations. It demonstrates that the positive law scheme is able to accommodate farther observations. This result incidentally raises questions on the relevance of a Gaussian ETEX-I inversion, using more than the nearby (French for the matter) observations, irrespective of their intrinsic quality. This could have been understood by trial and error by changing the dataset, but it was worth a rigorous demonstration.

5.4.3 Example: Gaussian case

In the Gaussian case, the source relative entropy is

$$
\mathcal{K}_{\boldsymbol{\sigma}} = \frac{1}{2}\boldsymbol{\mu}^{\dagger}(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\dagger})^{-1}\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\dagger}(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\dagger})^{-1}\boldsymbol{\mu},
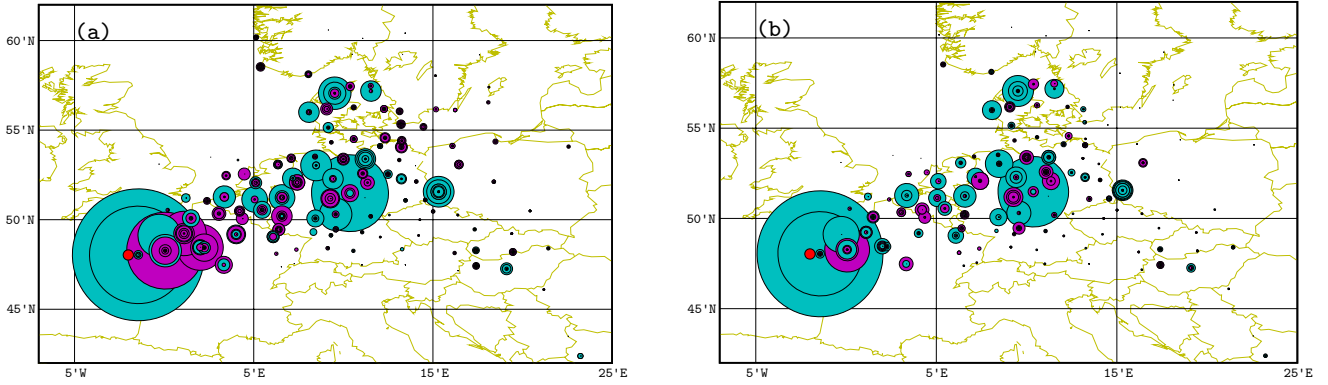\tag{59}
$$

Fig. 6. Plot of the Lagrange multipliers $\overline{\beta}$ which are also proportional to the diagnosed errors. The radii of the disks are proportional to the sensitivities. Panel **(a)** is related to the use of the Bernoulli prior, whereas panel **(b)** is related to the use of the Gaussian prior. The same scale has been used for the two plots.

while the error relative entropy is

$$\mathcal{K}_{\boldsymbol{\varepsilon}} = \frac{1}{2}\boldsymbol{\mu}^{\dagger}(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\dagger})^{-1}\boldsymbol{R}(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\dagger})^{-1}\boldsymbol{\mu}, \quad (60)$$

It does depend on the observation vector $\boldsymbol{\mu}$. When statistics on $\boldsymbol{\mu}$ are available, it is possible to provide with an average information gain on the source. Then the average gain on the signal and the noise are

$$\mathrm{E}_{\boldsymbol{\mu}}\left[\mathcal{K}_{\boldsymbol{\sigma}}\right] = \frac{1}{2}\mathrm{Tr}\left[(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\dagger})^{-1}\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\dagger}\right] \quad (61)$$

$$\mathrm{E}_{\boldsymbol{\mu}}\left[\mathcal{K}_{\boldsymbol{\varepsilon}}\right] = \frac{1}{2}\mathrm{Tr}\left[(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\dagger})^{-1}\boldsymbol{R}\right], \quad (62)$$

whereas the average total entropy $\mathrm{E}_{\boldsymbol{\mu}}\left[\mathcal{K}_{\boldsymbol{\sigma}} + \mathcal{K}_{\boldsymbol{\varepsilon}}\right] = \frac{d}{2}$ identifies with half the degrees of freedom for signal. This is not the conventional way the degrees of freedom for signal are recovered. The following section is dedicated to the route usually taken to recover them.

5.4.4 Degrees of freedom for signal

As was emphasised earlier, in an accidental context, there is only one set of observations. Before averaging, the information analysis of the previous section was carried out on a single observation set $\boldsymbol{\mu}$, and is therefore consistent with the approach. The number of degrees of freedom for signal [dfs], is however an averaged information obtained for a statistical set of $\boldsymbol{\mu}$. It is therefore less relevant in the specific context of accidental release, but could be useful in other circumstances. The usual definition of the dfs is (Rodgers, 1996)

$$\mathrm{dfs} = \mathrm{Tr}\left[\frac{\partial \boldsymbol{\mu}^{\mathrm{a}}}{\partial \boldsymbol{\mu}}\right], \quad (63)$$

Table 2. Information used in the retrieval of the source and the errors, for the Gaussian prior case, for the two inference schemes.

	$\mathcal{K}_{\boldsymbol{\sigma}}(\boldsymbol{\mu})$	$\mathrm{E}_{\boldsymbol{\mu}}\left[\mathcal{K}_{\boldsymbol{\sigma}}(\boldsymbol{\mu})\right]$	$\mathcal{K}_{\boldsymbol{\varepsilon}}(\boldsymbol{\mu})$	$\mathrm{E}_{\boldsymbol{\mu}}\left[\mathcal{K}_{\boldsymbol{\varepsilon}}(\boldsymbol{\mu})\right]$
MEM	876	164	1034	291
Bayes	1070	358	2077	1334

where $\boldsymbol{\mu}^{a}$ is the diagnosed measurements, i.e. $\boldsymbol{\mu}^{a} = \boldsymbol{H}\overline{\boldsymbol{\sigma}}$, therefore (for any prior):

$$\begin{aligned}\mathrm{dfs} &= \mathrm{Tr}\left[\boldsymbol{H}\frac{\partial \overline{\boldsymbol{\sigma}}}{\partial \boldsymbol{\mu}}\right] \\ &= \mathrm{Tr}\left[\boldsymbol{I} - \mathcal{H}^{\delta,\delta}\left(\boldsymbol{H}\mathcal{H}^{\alpha,\alpha}\boldsymbol{H}^{\dagger} + \mathcal{H}^{\delta,\delta}\right)^{-1}\right] \\ &= \mathrm{Tr}\left[\left(\mathcal{H}^{\sigma,\sigma} + \boldsymbol{H}^{\dagger}\mathcal{H}^{\varepsilon,\varepsilon}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\dagger}\mathcal{H}^{\varepsilon,\varepsilon}\boldsymbol{H}\right].\end{aligned} \quad (64)$$

This calculation is exact. However, it only offers a local analysis (in the vicinity of the minimum). These expressions parallel those obtained in (Chapnik et al., 2006) where the assimilation is of Gaussian nature whereas the model is possibly non-linear. In the Gaussian case, this formula identifies with Eq. (61) up to a scaling factor of two. In other words, in the MEM inference based on the fully Gaussian assumption, the dfs and the entropy expressions coincide (up to a factor of two).

5.4.5 Bayesian inference

Such an analysis can also be carried out using the Bayesian inference. The information content used in the reconstruction of the source is again given by the relative entropy

$$\mathcal{K}_{\boldsymbol{\sigma}} = \sum_{\boldsymbol{\sigma}} p(\boldsymbol{\sigma}|\boldsymbol{\mu})\ln\left(\frac{p(\boldsymbol{\sigma}|\boldsymbol{\mu})}{\nu(\boldsymbol{\sigma})}\right). \quad (65)$$
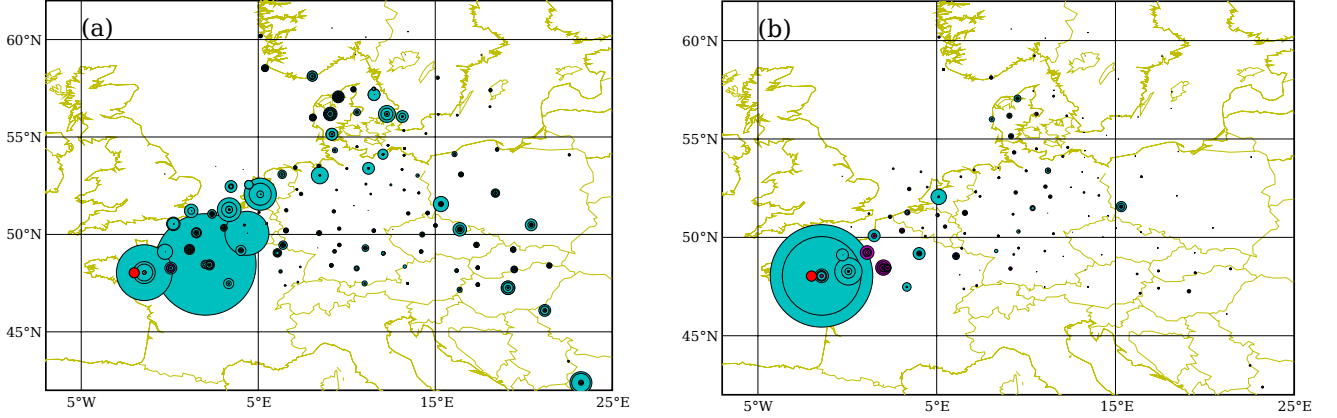
Fig. 7. Marginal contribution to the source relative entropy $\partial_{\mu_i} \mathcal{K}_\sigma$ for each of the observation μ_i. The radii of the disks are proportional to the sensitivities. Panel **(a)** is related to the use of the Bernoulli prior, whereas panel **(b)** is related to the use of the Gaussian prior. Compared to Fig. 6, a scaling factor of 10 has been applied to the radii on panel (a) whereas a factor of 2 has been applied to the radii of panel (b).

However, such a quantity is not as easily available as in the MEM framework. To illustrate the difficulty, let us note that on the Gaussian hypothesis case, the two inference schemes even give different answers. Indeed, for the MEM approach $\mathcal{K}_\sigma$ is given by Eq. (59) whereas in the Bayesian context:

$$
\begin{aligned}
\mathcal{K}_\sigma = &\frac{1}{2} \ln \left| \boldsymbol{B} \left(\boldsymbol{B}^{-1} + \boldsymbol{H}^\dagger \boldsymbol{R}^{-1} \boldsymbol{H} \right) \right| \\
&+ \frac{1}{2} \mathrm{Tr} \left[\left(\boldsymbol{B}^{-1} + \boldsymbol{H}^\dagger \boldsymbol{R}^{-1} \boldsymbol{H} \right)^{-1} \boldsymbol{B}^{-1} - \boldsymbol{I} \right] \\
&+ \frac{1}{2} \boldsymbol{\mu}^\dagger (\boldsymbol{R} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^\dagger)^{-1} \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^\dagger (\boldsymbol{R} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^\dagger)^{-1} \boldsymbol{\mu} \,.
\end{aligned}
\tag{66}
$$

When considering that the MEM inference only uses knowledge at the level of the first moment, whereas the Bayesian inference uses knowledge on all moments (actually two for a Gaussian law), these two expressions are in fact consistent: the Bayesian expression contains the first-moment $\boldsymbol{\mu}$-dependent term which equals the MEM-derived $\mathcal{K}_\sigma$. Yet the Bayesian expression is clearly more demanding, because of the second-order moments induced extra terms.

The connection with the usual definition of entropy of a data set in geophysical data assimilation is as follows.

The expression for $\mathcal{K}_\sigma$ yields the correct information content for a given set of observations $\boldsymbol{\mu}$. However the average gain in information for the source using statistics on $\boldsymbol{\mu}$ is given by

$$
\begin{aligned}
\mathrm{E}_{\boldsymbol{\mu}} [\mathcal{K}_\sigma] &= \sum_{\boldsymbol{\mu}} p(\boldsymbol{\mu}) \sum_{\boldsymbol{\sigma}} p(\boldsymbol{\sigma}|\boldsymbol{\mu}) \ln \left(\frac{p(\boldsymbol{\sigma}|\boldsymbol{\mu})}{v(\boldsymbol{\sigma})} \right) \\
&= \sum_{\boldsymbol{\sigma}, \boldsymbol{\mu}} p(\boldsymbol{\sigma}, \boldsymbol{\mu}) \ln \left(\frac{p(\boldsymbol{\sigma}, \boldsymbol{\mu})}{v(\boldsymbol{\sigma}) p(\boldsymbol{\mu})} \right)
\end{aligned}
\tag{67}
$$

The last term is known as mutual information (Cover and Thomas, 1991).

It is measuring the *reduction in entropy* because of the straightforward splitting $\mathrm{E}_{\boldsymbol{\mu}} [\mathcal{K}_\sigma] = H(\boldsymbol{\sigma}) - H(\boldsymbol{\sigma}|\boldsymbol{\mu})$, where $H(\boldsymbol{\sigma})$ is the entropy of the pdf $v(\boldsymbol{\sigma})$, whereas $H(\boldsymbol{\sigma}|\boldsymbol{\mu})$ is the average (on $\boldsymbol{\mu}$) entropy of the conditional pdf $p(\boldsymbol{\sigma}|\boldsymbol{\mu})$. This expression is often used in geophysical data assimilation (Rodgers, 2000; Fisher, 2003). The average of Eq. (66) is then (Gaussian prior)

$$
\mathrm{E}_{\boldsymbol{\mu}} [\mathcal{K}_\sigma] = \frac{1}{2} \ln \left| \boldsymbol{I} + \boldsymbol{B}^{-1} \boldsymbol{H}^\dagger \boldsymbol{R}^{-1} \boldsymbol{H} \right| \,,
\tag{68}
$$

which is also equal to $-\frac{1}{2} \ln |\boldsymbol{I} - \boldsymbol{A}|$, where

$$
\boldsymbol{A} = \left(\boldsymbol{B}^{-1} + \boldsymbol{H}^\dagger \boldsymbol{R}^{-1} \boldsymbol{H} \right)^{-1} \boldsymbol{H}^\dagger \boldsymbol{R}^{-1} \boldsymbol{H} \,,
\tag{69}
$$

is the averaging kernel ((Rodgers, 2000) and references therein).

These single event and event ensemble estimations of the information used in the retrieval through MEM and Bayes inference schemes are reported in Table 2 in the case of the Gaussian source prior using formula Eq. (59), Eq. (60), Eq. (61), Eq. (66), and Eq. (68). Note that, in the MEM case, the average pieces of information $\mathrm{E}_{\boldsymbol{\mu}} [\mathcal{K}_\sigma(\boldsymbol{\mu})] + \mathrm{E}_{\boldsymbol{\mu}} [\mathcal{K}_\varepsilon(\boldsymbol{\mu})]$ add up to $d/2$, with $d = 911$ here. This agrees with the equivalence with the dfs proven earlier.

As it was established before, the differences between the MEM and the Bayesian inferences are due to the additional gain in information obtained by the Bayes inference on the second-order moments, which is not taken into account by the MEM inference. Clearly, from Table 2 this contribution can be significant.

5.5 Sensitivity of the retrieved mass

It was shown that the reconstructions of an atmospheric source may depend significantly on the prior hypothesis on
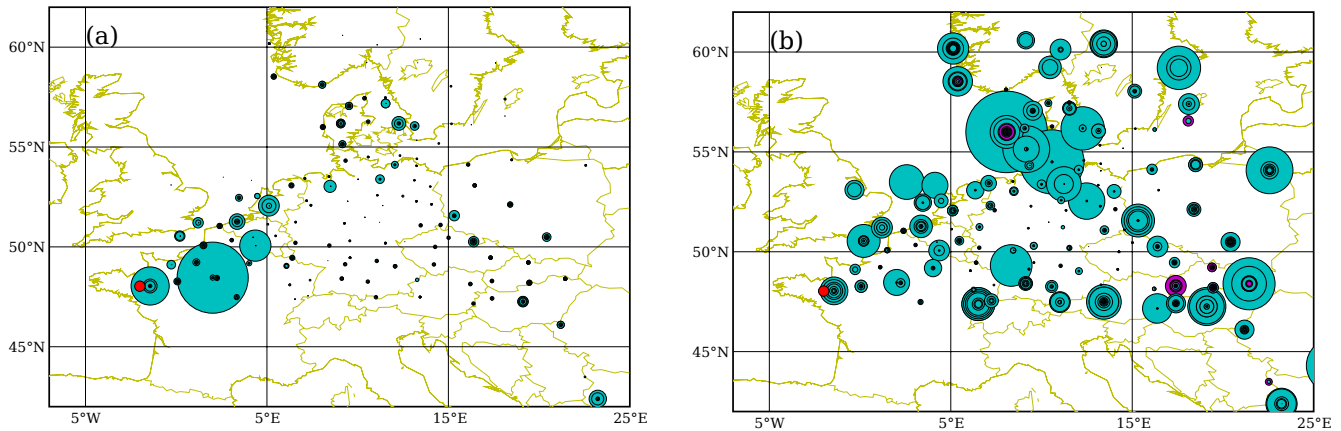
Fig. 8. Normalised sensitivity of the total retrieved mass to the observations $\mathcal{M}^{-1}\partial_{\mu_i}\mathcal{M}$. The radii are proportional to the sensitivities. Panel **(a)** is related to the use of the Bernoulli prior, whereas panel **(b)** is related to the use of the Gaussian prior. the same sensitivity scale has been used for the two plots.

the source, and especially whether it is positive or not. For example, in the ETEX-I case, a Gaussian prior based reconstruction delivers a total mass $\mathcal{M}$ not so significantly different from a positive law reconstruction. However, the fluctuations of the source are very different, with possibly significantly negative values in the Gaussian-based case. It is therefore useful to study the second-order sensitivity of the total retrieved mass to the observations. This sensitivity can easily be obtained from Eq. (49). As for the information gain, as a $d-$rank vector, it is easier to interpret than an averaging kernel. It reads

$$\frac{\partial \mathcal{M}}{\partial \mu_i} = \sum_{k=1}^{N} \frac{\partial \overline{\sigma}_k}{\partial \mu_i}, \qquad (70)$$

where $\overline{\sigma}_k$ is the reconstructed mass released from cell k in the space-time domain. Therefore

$$\frac{\partial \mathcal{M}}{\partial \mu_i} = \sum_{k=1}^{N} \left[\mathcal{H}^{\alpha,\alpha} \boldsymbol{H}^\dagger \left(\boldsymbol{H} \mathcal{H}^{\alpha,\alpha} \boldsymbol{H}^\dagger + \mathcal{H}^{\delta,\delta} \right)^{-1} \right]_{ki} . \quad (71)$$

Such a sensitivity is illustrated by the Gaussian prior/Bernoulli prior comparison in Fig. 8.

The two types of reconstruction lead to very different sensitivities. In the positive law case, the total mass is mainly influenced by observations in a radius of 800 km, and significantly lesser farther away. On the contrary, in the Gaussian hypothesis case, there is remarkably no such trend due to the proximity of the source.

6 Conclusions

Two inference schemes for inverse modelling of atmospheric tracer sources have been proposed. They are meant to handle prior information that cannot always translate onto mean and second-order moments, such as the positivity of the pollutant source.

They lead to cost-functions that differ from the usual least squares criterion. They have been qualified as non-Gaussian. The inference scheme based on the MEM principle has been summarised following earlier references while the exact Bayesian inference used with a maximum a posteriori estimator has been further developed. They both offer similar performances on real case studies. In one hand, the exact Bayesian approach implements faithfully the prior, but does not guarantee the convexity of the resulting criterion. In the other hand, the MEM principle interprets the prior in its large deviation limit, but does guarantee the convexity of any criterion. Both inferences deliver an efficient dual formulation (though convexity is required for the Bayesian one).

Because the resulting cost-functions are non-Gaussian the usual second-order sensitivity analysis needs to be extended beyond the BLUE posterior analysis. In particular, most second-order results depend on the solution of the reconstruction.

Most derivations have been illustrated by two typical inversions: the first one is based on a Gaussian source prior, while the second is based on a Bernoulli source prior (that essentially enforces the positivity of the source) following the MEM scheme. Most results transpose easily to the alternative Bayesian scheme, thanks to a formal translation on the prior pdf.

The analysis covariance and confidence matrices, but also the averaging kernels of the source and the errors have been derived. They are uneasy to interpret since they are represented by matrices defined in a large vector space, built on the underlying physical space with two space dimensions (ground sources) plus time.

Scalar and vector quantities are easier to interpret. In particular the total retrieved mass sensitivity to the observations has been derived.

The difference between the Gaussian and non-Gaussian inference is significant, the sensitivity of the mass in the Gaussian case being strong for almost all observations.

The Lagrange multipliers, which are the minimisers of the dual cost function, have been interpreted as the strength of the constraints imposed by each measurement.

Building on the analytical results of the the second-order analysis, an information content study has been carried out. The purpose was, for a single dispersion dispersion event, to objectively measure the information that is used in the reconstruction of the source (and possibly the errors). By averaging on the possible events, more classical results on entropy reduction and degrees of freedom for signal were recovered.

As by-products of this methodological study, several results have been obtained on ETEX-I and its study as a inverse modelling case. For instance, it has been shown that:

– When considering ETEX-I as an inverse modelling case study (it was not designed to), it was shown that the meteorological conditions make the true source location a not well informed point, while areas in Central Europe would have served as much easier source locations for identification.

– It was shown that, in the Gaussian framework, essentially only the observations in France provide with useful information (irrespective of their qualities). Such a scheme therefore makes the ETEX-I retrieval a very regional exercise. Whereas a more efficient non-Gaussian scheme is sensitive to significantly farther observations, and makes the exercise much more meaningful as a continental inverse modelling case study.

In the case of non-Gaussian cost-functions, the sensitivities that have been derived are depending on the solution. These sensitivities are given by exact formula for a linear model. However, they only represent local fluctuations of the solution, and do not give away these fluctuations at a larger scale. When the retrieved source is close to a bound, which is often the case here where most of the solution is close to zero for a positive law prior, these large scale fluctuations may be significant, whereas the variances are close to zero.

To study larger deviations of the solution, a Monte-Carlo study should be contemplated. Once the H matrix has been computed, the inversions are fast enough on a up-to-date four-core PC (about one minute for ETEX-I, three times less with less stringent but reasonable convergence criteria) so that such a Monte-Carlo study would be practical.

Edited by: O. Talagrand
Reviewed by: two anonymous referees

References

Bocquet, M.: Grid resolution dependence in the reconstruction of an atmospheric tracer source, Nonlin. Processes Geophys., 12, 219–234, 2005a.

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

Bocquet, M.: Reconstruction of an atmospheric tracer source using the principle of maximum entropy, II: Applications, Q. J. Roy. Meteor. Soc., 131, 2209–2223, 2005c.

Bocquet, M.: High resolution reconstruction of a tracer dispersion event, Q. J. Roy. Meteor. Soc., 133, 1013–1026, 2007.

Borwein, J. M. and Lewis, A. S.: Convex analysis and nonlinear optimization: theory and examples, Springer, 2000.

Chapnik, B., Desroziers, G., Rabier, F., and Talagrand, O.: Diagnosis and tuning of observational error in a quasi-operational data assimilation setting, Q. J. Roy. Meteor. Soc., 543–565, 2006.

Courtier, P.: Dual formulation of four-dimensional variational assimilation, Q. J. Roy. Meteor. Soc., 2449–2461, 1997.

Cover, T. M. and Thomas, J. A.: Elements of information theory, Wiley series in telecommunications, Wiley-Interscience, 1991.

Davoine, X. and Bocquet, M.: Inverse modelling-based reconstruction of the Chernobyl source term available for long-range transport, Atmos. Chem. Phys., 7, 1549–1564, 2007, http://www.atmos-chem-phys.net/7/1549/2007/.

Ellis, R. S.: Entropy, Large Deviations, and Statistical Mechanics, Springer, 1985.

Fisher, M.: Estimation of Entropy Reduction and Degrees of Freedom for Signal for Large Variational Analysis Systems, Tech. Rep. 397, ECMWF, 2003.

Hourdin, F. and Talagrand, O.: Eulerian backtracking of atmospheric tracers. I: Adjoint derivation and parametrization of subgid-scale transport., Q. J. Roy. Meteor. Soc., 132, 567–583, 2006.

Issartel, J.-P. and Baverel, J.: Inverse transport for the verification of the Comprehensive Nuclear Test Ban Treaty, Atmos. Chem. Phys., 3, 475–486, 2003, http://www.atmos-chem-phys.net/3/475/2003/.

Krysta, M. and Bocquet, M.: Source reconstruction of an accidental radionuclide release at European scale, Q. J. Roy. Meteor. Soc., 133, 529–544, 2007.

Krysta, M., Bocquet, M., and Brandt, J.: Probing ETEX-II data set with inverse modelling, Atmos. Chem. Phys. Discuss., 2795–2819, 2008.

Lagarde, T.: Nouvelle approche des méthodes d'assimilation de données: les algorithmes de point selle, Ph.D. thesis, Université Paul Sabatier, Toulouse III, France, 2000.

Le Besnerais, G., Bercher, J.-F., and Demoment, G.: A new look at entropy for solving linear inverse problems, Information Theory, IEEE T. Inform. Theory, 45, 1565–1578, 1999.

Le Dimet, F., Ngodock, H.-E., Luong, B., and Verron, J.: Sensitivity analysis in variational data assimilation, J. Meteorol. Soc. Jpn., 75, 245–255, 1997.

Mohammad-Djafari, A.: A comparison of two approaches: Maximum enropy on the mean (MEM) and Bayesian (BAYES) for inverse problems, Kluwer Academic, berg-en-Dal, South Africa, 1996.

Ngodock, H.-E.: Assimilation de données et Analyse de sensibilité: Une application à la circulation océanique, Ph.D. thesis, Univer-

sité Joseph-Fourier – Grenoble I, France, 1996.

Nocedal, J. and Wright, S. J.: Numerical Optimization, Springer Series in Operations Research, Springer, 2006.

Nodop, K., Connolly, R., and Girardi, F.: The field campaigns of The European tracer experiment (ETEX): overview and results, Atmos. Environ., 32, 4095–4108, 1998.

Pudykiewicz, J. A.: Application of adjoint transport tracer equations for evaluating source parameters, Atmos. Environ., 32, 3039–3050, 1998.

Quélo, D., Krysta, M., Bocquet, M., Isnard, O., Minier, Y., and Sportisse, B.: Validation of the Polyphemus platform on the ETEX, Chernobyl and Algeciras cases, Atmos. Environ., 41, 5300–5315, 2007.

Rodgers, C. D.: Information content and optimization of high spectral resolution measurements, Proc. SPIE, 2830, 136–147, 1996.

Rodgers, C. D.: Inverse Methods for atmospheric sounding, World Scientific, Series on Atmospheric, Oceanic and Planetary Physics, 2000.

Talagrand, O.: Assimilation of Observations, an Introduction, J. Meteorol. Soc. Jpn., 75, 191–209, 1997.

Tarantola, A. and Valette, B.: Generalized nonlinear inverse problems solved using the least square criterion, Rev. Geophys. Space Ge., 20, 219–232, 1982.

Uliasz, M.: Application of the Perturbation Theory to the Sensitivity Analysis of an Air Pollution Model., Z. Meteorol., 33, 355–362, 1983.