

Inverse transport problem of estimating point-like source using a Bayesian parametric method with MCMC



Aurélien Hazart^{a,c,*}, Jean-François Giovannelli^{a,b}, Stéphanie Dubost^c,
Laurence Chatellier^c

^a Lab. des Signaux et Systèmes (CNRS-SUPELEC-UPS), Plateau de Moulon, 91192 Gif-sur-Yvette, France

^b Lab. de l'Intégration du Matériau au Système (University Bordeaux-CNRS-IPB), 33405 Talence, France

^c Electricité De France R&D, 6 quai Watier, 78400 Chatou, France

ARTICLE INFO

Article history:

Received 3 December 2012

Received in revised form

5 August 2013

Accepted 25 August 2013

Available online 6 September 2013

Keywords:

Bayesian parametric estimation

Gibbs sampling

Metropolis-Hastings

Point-like source

Groundwater pollution

ABSTRACT

Recovering the origin of an incident after detection of a polluting substance in the environment is crucial to start the remediation procedures. The lack of observations, the measurement errors and the model uncertainties make the problem of source estimation an ill-posed inverse problem that requires regularization to determine a solution. The two most frequent methods of regularization are source parametrization and penalization of undesirable solutions. In this paper, the proposed approach combines both methods in order to obtain a strong regularization that is efficient in case of few and erroneous observations. Point sources with parametric temporal releases and parameter penalizations are incorporated in a Bayesian framework where observations and prior information are combined in a hierarchical probabilistic model and the posterior law is explored with a Markov Chain Monte Carlo sampling algorithm. Estimation of the source parameters is provided by the posterior mean and uncertainties are provided by the posterior variance. To validate the method, several simulated cases with different emission events are considered. Quality of the estimate as well as impact of source model errors are also investigated. Then, a comparison with two existing least squares methods is conducted, in various configurations of sensors and noise level. Finally, the behavior of the method is described on a strongly underdetermined real case where only one sensor recorded the pollution.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The problem of estimating the source of an emission from measurements is generating extended interest in the data processing community [1]. Among the numerous applications, the most important are soil pollution [2,3], atmospheric pollution [5] and odor emissions [4].

Specifically in groundwater pollution, the advantages of using a source estimation tool are the possibility of identifying responsibilities when the location and period of dumping are known and the control and preservation of the quality of the groundwater, one of the steps toward sustainable development [6]. When a soil pollution plume is detected, identifying the origin consists essentially in estimating the spatial location of the source, its temporal release and the quantity of pollutant leaked. Practical constraints affect the quantity of recorded information: the time delay before the first measurements and the number and configurations of the sensors available. The lack of observations requires the use of data processing

* Corresponding author. Postal address: Inverse problems Group, Lab. of Signal and Systems, Plateau de Moulon, 91192 Gif-sur-Yvette, France.

E-mail addresses: auhazart@hotmail.com (A. Hazart),

Giova@IMS-Bordeaux.fr (J.-F. Giovannelli), stephanie.dubost@edf.fr (S. Dubost), laurence.chatellier@edf.fr (L. Chatellier).

methods to identify the source parameters. In this context, the problem is *ill-posed* for three main reasons.

- The strong and quick diffusion of the pollution plume acts as a low-pass filter on the point source. The source is generally concentrated in space and time; the plume is smoothed so that observations lack high-frequency information about the source.
- The spatial sensor array leads to an important subsampling of the whole pollution plume. In real applications, there are generally few sensors (less than ten) and sometimes only few observations per sensor.
- The observations are made in a difficult physical context. The observation model is necessarily imperfect, implying that errors (measurement errors, modeling errors) must be accounted for.

Mathematically, a problem is considered to be *ill-posed* when the solution is not unique or if it is not a continuous function of the observations. In such problems, a minor perturbation in the data can cause a large perturbation in the solution; this is due to a lack of information. To resolve this difficulty, the general methods consist in introducing additional information in order to compensate for the lack of information contained in the observations. The initial problem is then transformed into a *well-posed* problem whose solution is deemed to be acceptable [7,8] and the problem is said *regularized*.

In the case of the inverse transport problem, there are two approaches to regularization: the penalized non-parametric approach and the parametric approach. The ideas behind these approaches differ and lead to different methods. The method proposed in this paper enters into the class of parametric methods. However, to be able to tackle real applications in which very few observations are available, the values of the parameters are penalized according to the additional information about the source. Accordingly, our approach is also inspired by penalized non-parametric methods.

1.1. Parametric methods

The methods in this category regularize the problem by reducing the number of unknowns [9]. The space of the solutions is structured by the choice of a parametric model describing the source. This type of approach generally makes it possible to estimate both the location and the temporal release, even with a small number of observations. Existing methods include numerical methods based on minimum least squares [10–12,21], maximum likelihood [4,13,14,22], analytical methods based on the explicit expression of concentrations in the environment [15,16], methods based on the adjoint operator [17–19] and more recently Bayesian methods [56,57].

Several source models have been investigated in the literature. They share common points regarding the separation of space and time characteristics and the point location modeling. To model the temporal release, the instantaneous release modeling a brief incident is the focal point of most of the work. Among other types of releases, a constant endless temporal release is estimated by analytical methods

[16,20] or numerical methods [16,21]. Constant release on a fixed period is used in [22] that estimates the source characteristics and several parameters of the physical transport model. However, the method is limited to bringing into competition several possible locations and several release periods. The case of the infinite constant release (steady state) is dealt with by numerical optimization [23], adjoint method [5] and Bayesian method [24]. More complex releases consisting of multiple sources [55,58] or source with several temporal release [25] have also been considered. In the works presented here, a set of functions representing different temporal releases is defined. It includes most of the simple temporal releases.

1.2. Penalized non-parametric methods

The penalized non-parametric approach consists in modeling the source by a function that can vary in space and/or time and in seeking the solution that complies with a number of conditions. To restrict the space of the solutions, these methods provide regularization by penalizing the undesirable solutions. A state of the art of these methods can be found in [26]. In theory, these methods can be applied for any type of source. However, because the number of unknowns is far greater than the number of observations, the problem is highly underdetermined and some characteristics of the source are usually assumed to be known.

The most often encountered problem involves the temporal release estimation of a source with a known location. In this case, penalization concerns the temporal function and different types of penalization have been proposed. In a deterministic framework, Tikhonov's regularization method [27] produces a smooth solution [28–30]. In a stochastic framework, the method of the minimum relative entropy (MRE) [31,32] has been developed. A comparison of Tikhonov's method and MRE's method is made in [33]. Other existing stochastic methods are the geostatistical methods [34,35]; a discussion regarding the positioning of this method within the Bayesian methods can be found in [2,36].

Recently, some authors have extended the previous works to source with unknown location. In [37,38], the authors apply the Tikhonov regularization method while also estimating the localization parameters of a point source. In addition, [39] proposes a minimum relative entropy method to estimate the location and the temporal release of an atmospheric pollution source. Finally, [26] extends the geostatistics method to estimating the space and time contamination history by applying the adjoint operator method [19,33]. However, the method that supplies source location requires prior knowledge of the release period.

1.3. Our penalized parametric method

The approach presented in this paper belongs to the class of the parametric approach: the source is a point source and the temporal release is chosen among a set of parametric functions. It is an extension of our work [40], where only instantaneous release was processed in a simpler manner. Unlike other methods, our approach is not limited to a specific parametric function but is

adaptable to various temporal releases. Unlike our work, most of the existing methods estimate the source characteristics by maximizing parameter likelihood without regularizing the problem any further. In this paper, the main purpose is to estimate the location, the temporal release and the amplitude of a pollution source while taking into account available information on the source. The additional information is introduced by means of prior probability densities. One advantage of the probabilistic approach is to supply both the estimated source and the uncertainties regarding the estimation. This framework also allows to apply the method to difficult cases of pollution, where only few noisy measurements are available.

The paper is presented in the following manner. Section 2 describes the physical model used to formalize the source and the observations. The proposed inverse approach is described in detail in Section 3 (Bayesian strategy) and Section 4 (MCMC algorithm). Section 5 outlines numerical applications and through several simulated cases and one real underdetermined case.

2. Physical model

2.1. Source model

Because of the small spatial extent of the cases of interest, the source is approximated by a *point source*. We consider in this work non-moving point source assumed to occur at the surface in $z=0$, releasing a pollutant quantity q_0 according to a temporal release $\chi(t)$. The source is described by the term s that formalizes a concentration of substance by unit of time (concentration rate):

$$s(x, y, t) = q_0 \delta(x - x_0) \delta(y - y_0) \chi(t) \quad (1)$$

where δ represents the Dirac distribution. With this model, the location is entirely defined by the two coordinates (x_0, y_0) .

For the temporal release $\chi(t)$, we also define a parametric function that depends on few parameters to enforce the explicit structure (see 1.1): a time position parameter t_0 and a time spreading parameter τ_0 . The integral of the temporal release is normalized to one, so the parameter q_0 represents the total quantity of pollutants. Accordingly, the source is characterized by a set of P parameters: the quantity q_0 and the spatio-temporal parameters collected in the row vector $\theta = [x_0, y_0, t_0, \tau_0]$. For $\chi(t)$, various incidents are investigated by the definition of seven parametric functions: Dirac, Step, Window, Decreasing slope, Triangle, Laplacian, and Gaussian. Simple temporal releases like instantaneous release (Dirac) or constant release (Step) can sometimes lead to explicit expressions of the direct model output.

The Dirac distribution models an instantaneous incident. On the contrary, the Step distribution imitates a continuous release with a starting time and no ending (steady state when infinite time is considered). Same release with an ending time is considered through the Window function. The Decreasing slope models a linearly decreasing pollution event. The Triangle function (isosceles triangle) models linearly increasing and decreasing release. Laplacian and Gaussian functions represent symmetrical more or less spiky releases. See Fig. 1 for a representation of the temporal releases with the parameters $t_0 = 10$ days and $\tau_0 = 5$ days for a representation of the Laplacian temporal release with the parameters $t_0 = 10$ th day, $\tau_0 = 5$ days.

Except for the Step function, the temporal releases are normalized. They depend on a single parameter t_0 (Dirac and Step) or on two parameters t_0, τ_0 (Window, Triangle, Decreasing slope, Laplacian, and Gaussian). A common feature of these functions is that they describe a single pollution event. Methodologically, any parametric model

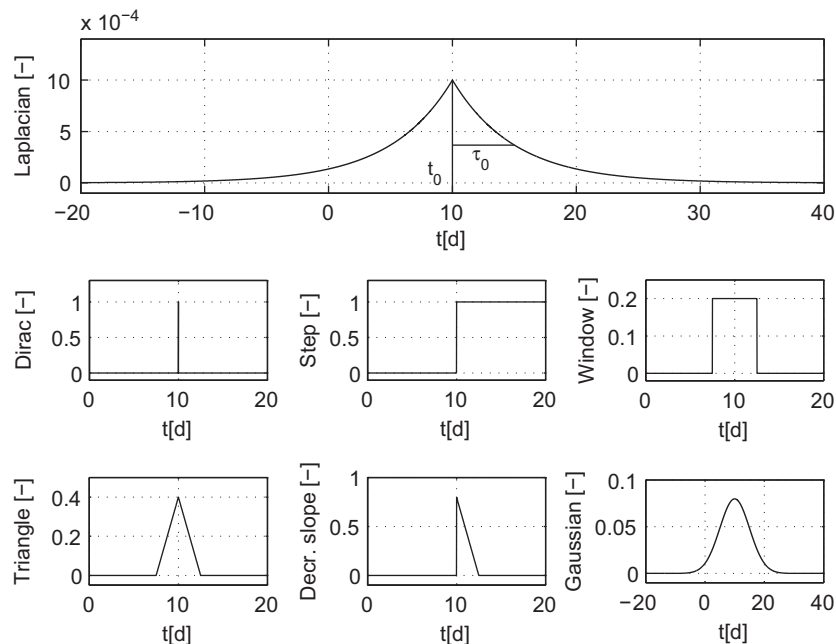


Fig. 1. Example of temporal release with $t_0 = 10$ th days, $\tau_0 = 5$ days. Except for Dirac and Step, the releases are normalized.

(even with multiple events) can be used but the present paper is devoted to a single event.

2.2. Direct model

The direct model or forward model is made up of the transport model and the observation model.

The transport model formalizes the physical process involved in the propagation of the release from the source to the sensors. No specific transport model is required by the method proposed in this paper, the only requirement being to compute the model output for a given source. Nevertheless, it is noteworthy that the computation time of the transport model will impact the computation time of the inverse method. Most of the transport models are based on partial differential equations named advection diffusion equations (ADE). In these equations, the coefficients are the physical parameters that characterize the soil properties. The transport model of interest is described in the Appendix A. Experts in geology built it from the characteristics of several real sites.

The observation model describes the way the measurements are collected by the sensors in the groundwater. Various observation models can be defined, as long as it formalizes the concentrations at the outputs for a given source at the input. In investigated case here, the number of sensors is small as compared to the spatial extent of the site and measurements are made as a function of time at each sensor. The proposed approach naturally applies to other observation schemes where measurements are made at a unique time e.g. [28] or onto an irregular grid. We refer to the Appendix B for the description of the observation model used in this paper.

Transport and observation model lead to the following expression of the concentrations:

$$c(x, y, t) = 2v_{z,1}h_1(t) * \bar{h}_2(x, y, t) * s(x, y, t) \quad (2)$$

where $*$ is the convolution operator (see Appendices A and B for the explanation of the different terms). After replacing Eq. (1) in Eq. (2) and introducing the notation $q_0h_{kj}(\theta)$ for the model output at sensor k and at time j to reveal the dependency with respect to the source parameters (q_0, θ) , the data can be written as

$$\mathbf{d} = q_0\mathbf{h}(\theta) + \epsilon \quad (3)$$

where the vector $\mathbf{h}(\theta)$ of size $M = \sum_{k=1}^K J_k$ contains the elements $h_{kj}(\theta)$, the vector \mathbf{d} groups all the available observations and ϵ includes both modeling errors due to the imperfection of the direct model and errors due to the measurement system. In Eq. (3), we can see that the direct model is a linear function of q_0 and a non-linear function of θ . Here, the parametric model for the source ensures the non-negativity of the concentrations at the output: as the source cannot be negative, the concentrations are also non-negative.

By construction, the model output reproduces the observed concentration measurements. Assuming additive errors, the difference between the observations and the model output represents the reconstruction error at the output. In the real-life application considered here, there is no evidence on how the errors distribution is. When no

specific knowledge about the errors distribution is available, it is usual to consider that the errors ϵ are independent and identically distributed according to the normal density with zero mean and a given variance matrix. In the following, the variance matrix is written $r_\epsilon I_M$ with I_M the identity matrix of size M (number of observations) and r_ϵ the unknown variance parameter. The choice of a Gaussian distribution for the error will also ease the analytical calculation, as it will be seen later.

3. Penalized parametric approach to inverse problem

3.1. Bayesian approach

A Bayesian formulation of the problem is interesting because prior knowledge is taken into consideration naturally and allows to obtain uncertainties about the estimated parameter. Based on the likelihood and the prior probability, the joint posterior density for q_0, θ and r_ϵ built by the Bayes formula takes the form:

$$p(\theta, q_0, r_\epsilon | \mathbf{d}) = \frac{p(\mathbf{d} | \theta, q_0, r_\epsilon) p(\theta, q_0) p(r_\epsilon)}{\int p(\mathbf{d} | \theta, q_0, r_\epsilon) p(\theta, q_0) p(r_\epsilon) d\theta dq_0 dr_\epsilon} \quad (4)$$

where $[\theta, q_0]$ and r_ϵ are supposed to be a priori independent.

The posterior probability contains the available information about the unknown parameters θ . Insofar as the denominator of Eq. (4) does not depend on the parameters, the posterior density is proportional to the numerator where $p(\mathbf{d} | \theta, q_0, r_\epsilon)$ is the likelihood and $p(\theta, q_0) p(r_\epsilon)$ is the prior density.

3.2. Parameter likelihood

Since the error is additive and normally distributed with zeros mean and variance matrix $r_\epsilon I_M$ (see Eq. (3)), \mathbf{d} is distributed according to a normal density with mean $q_0\mathbf{h}(\theta)$ and same variance $r_\epsilon I_M$. Then, the likelihood takes the form:

$$p(\mathbf{d} | \theta, q_0, r_\epsilon) = (2\pi r_\epsilon)^{-M/2} \exp \left\{ -\frac{\|\mathbf{d} - q_0\mathbf{h}(\theta)\|^2}{2r_\epsilon} \right\} \quad (5)$$

where $\|\cdot\|$ designs the Euclidean norm. So the likelihood depends on the error variance r_ϵ , which is sometimes called hyperparameter, and is generally unknown. In this paper, r_ϵ is estimated as well as the source parameters. Such model is usually referred to as hierarchical Bayesian model.

3.3. Prior density

The parametric source models represent the first level of prior information. As the inverse problem is particularly ill-posed, we consider a second level of prior information about the source by parameter penalization. The purpose is to ensure that each parameter belongs to its definition domain (for instance $\tau_0 > 0$) and to integrate additional information when any such knowledge exists. There are many ways of incorporating prior information. The Bayesian framework used here introduces this information through the prior density. The prior that favors regions of high probability thus structures the parameter space.

As information about spatial location mainly comes from the site characteristics (building layout, etc.), we consider that the two parameters x_0, y_0 are not independent. For the other parameters, because no information is available about eventual dependencies, the prior law is separable. Consequently, the prior probability takes the form of a product:

$$p(\theta, q_0) = p(x_0, y_0)p(t_0)p(\tau_0)p(q_0). \quad (6)$$

One advantage of the parametric Bayesian approach over non-parametric approaches is that it offers an intuitive way of taking into account prior knowledge about the source. For each component of θ , we investigate two different options: a uniform density on a given domain and a Gaussian density distributed around a nominal value. These laws model the two most usual ways of expressing prior knowledge. A uniform density requires a lower and an upper bound on each parameter, it implies hard constraints on the parameter values, for instance by authorizing or inhibiting location. Conversely, Gaussian density allows to integrate flexible constraints, for instance by attributing to each point in space a prior presence probability of the source. For the parameter q_0 , both options are possible. We choose here a Gaussian prior density for q_0 , $p(q_0) = \mathcal{N}(\mu_{q_0}, r_{q_0})$, where μ_{q_0} and r_{q_0} are respectively the mean and the variance of $p(q_0)$, with a high value for r_{q_0} to set a non-informative prior. This choice eases the computation as it will be seen below.

The hierarchical Bayesian framework also requires to define a prior density for the error variance r_e . If it exists, the classic choice is a conjugate density [41], because it allows to easily generate samples from the posterior density. Here, the likelihood is normally distributed and the conjugate density is the Inverse Gamma density $p(r_e) = \mathcal{IG}(r_e | \alpha_e, \beta_e)$ defined by

$$\mathcal{IG}(r_e | \alpha_e, \beta_e) \propto \frac{\exp(-\beta_e/r_e)}{r_e^{\alpha_e+1}} \mathbb{1}_{\mathbb{R}^+}(r_e) \quad (7)$$

where $\mathbb{1}_{\mathbb{R}^+}$ is the indicator function on \mathbb{R}^+ . The parameters (α_e, β_e) of the prior density for r_e are sometimes referred to as hyperparameters. Assuming no prior knowledge on r_e , α_e and β_e are chosen close to zero. Thus, we obtain the prior law $p(r_e) = 1/r_e$, which is the non-informative Jeffreys prior [42].

3.4. Posterior density

If the prior densities on θ are uniform on the domain \mathcal{D} , the posterior densities take the following form:

$$p(\theta, q_0, r_e | \mathbf{d}) \propto \exp\left\{-\frac{\|\mathbf{d} - q_0 \mathbf{h}(\theta)\|^2}{2r_e}\right\} \mathbb{1}_{\mathcal{D}}(\theta) \exp\left\{-\frac{\beta_e}{r_e}\right\} r_e^{-\alpha_e-1} \mathbb{1}_{\mathbb{R}^+}(r_e). \quad (8)$$

Given r_e and q_0 , this option is equivalent to truncating the likelihood of the parameters in \mathcal{D} . In this case, the prior information consists in integrating constraints on the source parameters.

If the prior densities are Gaussian, the posterior law is written:

$$p(\theta, q_0, r_e | \mathbf{d}) \propto \exp\left\{-\frac{\|\mathbf{d} - q_0 \mathbf{h}(\theta)\|^2}{2r_e}\right\} \exp\left\{-\frac{\beta_e}{r_e}\right\} r_e^{-\alpha_e-1} \mathbb{1}_{\mathbb{R}^+}(r_e) \exp\left\{-\frac{1}{2}[\theta - \mu_\theta]^T R_\theta^{-1} [\theta - \mu_\theta] - \frac{(q_0 - \mu_{q_0})^2}{2r_{q_0}}\right\} \quad (9)$$

where μ_θ and R_θ are respectively the mean and the covariance matrix of the prior density $p(\theta)$. This time, the prior density penalizes the solution: the further θ is from its prior mean, the faster the posterior density decreases. The covariance matrix influences the structure of decrease.

The choice between the formulations (8) and (9) should be guided by the kind of prior information the user wants to introduce. The first case is useful to set the range of a parameter and can be seen as a *hard penalization*, as it prohibits some values. The second case can be seen as a *soft penalization* that smoothly disadvantages some values. Moreover, in both cases, the prior can be tuned to allow more or less large ranges depending on available information. In this paper, we focus on the uniform prior density, leading to the posterior density given by Eq. (8). The Gaussian choice or even a combination of uniform and Gaussian densities is naturally possible.

4. Exploration by MCMC algorithm and estimation

The posterior probability density $p(\theta, q_0, r_e | \mathbf{d})$ integrates all the available information on the source. It provides information about high probability zones in the parameter space. Visualization of such zones in appropriate subspaces is a relevant way of understanding the inverse problem difficulties. It can also give an idea of the source components.

To explore the posterior density, basic methods rely on discretization: the posterior density is simply evaluated on a regular grid of parameter values. As the parameter space dimension is higher than three, such an approach requires high computational resources (for a P -dimensional parameter and a discretization on L points in each dimension, L^P evaluations are required). In addition, the number of parameters to be estimated may increase in future developments. More powerful approaches are given by Monte-Carlo sampling methods which compute values essentially in regions with high probability. Instead of splitting up the parameter space into a uniform grid, the idea is to use a set of points representative of the probability density. Here, the posterior probability cannot be directly sampled because of the likelihood complexity. We propose therefore Markov Chain Monte Carlo (MCMC) algorithms to explore the posterior density.

4.1. Exploration of the source space

The adopted algorithm to generate samples from the posterior probability belongs to MCMC family. Generally speaking, these algorithms generate a Markov chain with the target density as stable distribution. The two main MCMC methods involve Gibbs sampler and the

Metropolis-Hasting (MH) algorithms [43,44]. In this study, the proposed algorithm is based on both methods (namely hybrid Gibbs algorithms [44]). The sampling problem is split into three Gibbs steps: at each iteration, we generate samples from the conditional posterior densities $p(q_0|\mathbf{d}, \theta, r_\epsilon)$, $p(r_\epsilon|\mathbf{d}, \theta, q_0)$ and $p(\theta|\mathbf{d}, q_0, r_\epsilon)$. One of the advantages of such algorithms is to provide simulations even for intricate posterior law, even if the normalization constant is unknown. Another nice property is to separate non-linear source parameters from the linear parameter and the error variance and to allow specific processing for q_0 and r_ϵ . A drawback is that it can be computationally expensive. However, in our implementations, the computation time is of the order of 10 min,¹ which is compatible with the practical demands.

From the joint density for θ , q_0 and r_ϵ , the three conditional laws exist and can be derived from the Bayes formula, taking into account that θ , q_0 and r_ϵ are assumed to be independent (so $p(q_0|\theta, r_\epsilon) = p(q_0)$ for instance). Each of them is proportional to the product of the likelihood and the corresponding prior.

1. The conditional law of q_0 :

$$p(q_0|\mathbf{d}, \theta, r_\epsilon) = \frac{1}{p(\mathbf{d}|\theta, r_\epsilon)} p(\mathbf{d}|\theta, q_0, r_\epsilon) p(q_0). \quad (10)$$

2. The conditional law of r_ϵ :

$$p(r_\epsilon|\mathbf{d}, \theta, q_0) = \frac{1}{p(\mathbf{d}|\theta, q_0)} p(\mathbf{d}|\theta, q_0, r_\epsilon) p(r_\epsilon). \quad (11)$$

3. The conditional law of θ :

$$p(\theta|\mathbf{d}, q_0, r_\epsilon) = \frac{1}{p(\mathbf{d}|q_0, r_\epsilon)} p(\mathbf{d}|\theta, q_0, r_\epsilon) p(\theta). \quad (12)$$

For the first step, we have to sample $p(q_0|\mathbf{d}, \theta, r_\epsilon)$. It can be seen that the Gaussian density for q_0 is a conjugate density for the Gaussian likelihood (see 3.3 and Appendix C). Therefore, the conditional posterior for q_0 is also Gaussian with parameters:

$$p(q_0|\mathbf{d}, \theta, r_\epsilon) \propto \mathcal{N}\left(\frac{r_{q_0} \mathbf{d}^t \mathbf{h} + r_\epsilon \mu_{q_0}}{r_{q_0} \mathbf{h}^t \mathbf{h} + r_\epsilon}, \frac{r_\epsilon r_{q_0}}{r_{q_0} \mathbf{h}^t \mathbf{h} + r_\epsilon}\right). \quad (13)$$

At this point, the algorithm shows a similarity with the deterministic maximum likelihood approach proposed in [4]. Indeed, the mean of the conditional posterior $p(q_0|\mathbf{d}, \theta, r_\epsilon)$ becomes the maximum likelihood estimate of q_0 when no prior on q_0 is considered ($r_{q_0} = +\infty$). The method can be seen as an extension of the maximum likelihood approach that includes prior information and explores the parameter space.

For the second step, the density to be sampled is $p(r_\epsilon|\mathbf{d}, \theta, q_0)$. Due to the conjugacy (see 3.3 and Appendix C) it has the interesting property to be an Inverse Gamma

density, with density:

$$p(r_\epsilon|\mathbf{d}, \theta, q_0) \propto \mathcal{IG}\left(\alpha_\epsilon + \frac{M}{2}, \beta_\epsilon + \frac{\|\mathbf{d} - q_0 \mathbf{h}(\theta)^{(n)}\|^2}{2}\right). \quad (14)$$

To set a non-informative prior on ϵ , the parameters α_ϵ and β_ϵ are chosen close to zero.

For the third step, the density to be sampled is $p(\theta|\mathbf{d}, q_0, r_\epsilon)$ which does not have a simple form. To generate these samples, we use a random walk MH algorithm (sometimes simply called Metropolis algorithm) with the Gaussian proposal density $\mathcal{N}(\theta^{(n)}, \alpha R_l)$ where

- θ^p is the proposal sample,
- $\theta^{(n)}$ is the current sample (iteration n),
- R_l is the prior covariance matrix,
- α is a scale factor to be practically set.

Thus, the proposal sample is obtained by adding a random perturbation tuned by α to the current sample. Hence, the algorithm is practically tuned by a unique coefficient α which affects the distance between the current and the proposal sample. In theory, α does not affect the result of the algorithm but only the computation time. In practice, a compromise should be made between a fast exploration of the posterior density and a high acceptance rate: propositions from a wider neighborhood can produce a better parameter space exploration but a greater chance of rejection. Here, α is chosen in order to reach a 25% acceptance rate, as recommended by [45] for more than two-dimensional problems. To take into account the different scales of each parameter, we define R_l according to the prior, which at least limits the space to the plausible values. In the uniform prior case of Eq. (8), R_l is set to a diagonal matrix with i -th diagonal element equal to $|\theta_i^{\max} - \theta_i^{\min}|/2$. In the Gaussian prior case of Eq. (9), R_l is simply set to the prior covariance matrix R_θ . The acceptance step relies on the classical Metropolis acceptance probability:

$$\rho(\theta^p, \theta^{(n)}) = \min\left\{1, \frac{p(\theta^p|\mathbf{d}, q_0, r_\epsilon)}{p(\theta^{(n)}|\mathbf{d}, q_0, r_\epsilon)}\right\}. \quad (15)$$

The three steps of the proposed algorithm are summed up in Table 1 with a pseudo-code notation. The algorithm is named unsupervised because it automatically estimates the error variance, which is relevant to avoid a difficult human decision.

Table 1
Unsupervised MCMC algorithm.

```

initialize  $\theta^{(0)}, r_\epsilon^{(0)}$ 
for  $n = 0, \dots, N-1$  do
  (1) generate  $q_0^{(n+1)} \sim p(q_0|\mathbf{d}, \theta^{(n)}, r_\epsilon^{(n)})$  Eq. (13)
  (2) generate  $r_\epsilon^{(n+1)} \sim p(r_\epsilon|\mathbf{d}, \theta^{(n)}, q_0^{(n+1)})$  Eq. (14)
  (3) generate  $\theta^{(n+1)} \sim p(\theta|\mathbf{d}, r_\epsilon^{(n+1)}, q_0^{(n+1)})$  with the steps:
    - propose a new sample  $\theta^p \sim \mathcal{N}(\theta^{(n)}, \alpha R_l)$ 
    - compute the acceptance probability  $\rho(\theta^p, \theta^{(n)})$  Eq. (15)
    - draw  $u \sim \mathcal{U}_{[0,1]}$ 
    - if  $u < \rho(\theta^p, \theta^{(n)})$  do  $\theta^{(n+1)} \leftarrow \theta^p$  (accept the proposed sample)
    - else do  $\theta^{(n+1)} \leftarrow \theta^{(n)}$  (keep the current sample)
end for

```

¹ The proposed algorithm has been implemented with the computing environment Matlab on a PC, with a 2 GHz AMD-Athlon CPU, and 1 GB of RAM.

Remark 1. The error variance affects the behavior of the MH algorithm, as it has a significant influence on the posterior density shape. Indeed, the higher the variance, the flatter the posterior density and the higher the acceptance rate of the MH algorithm.

The obtained samples can be used to depict the zones of high probability of the source localization in the parameter space. However, they do not supply a unique solution to the inverse problem. To estimate the source parameters, the two remaining questions concern the choice of a point estimate and its computation.

4.2. Point estimate and confidence intervals

The estimation of the source parameters from the posterior density requires the choice of a point estimate. In our context, numerous point estimates are possible: the posterior maximizer, the posterior mean, the posterior median, etc. Here, we focus on the posterior mean (PM) estimate which supplies the average of the probable sources:

$$\begin{aligned} [\hat{\theta}, \hat{q}_0, \hat{r}_\epsilon]_{\text{PM}} &= \mathbb{E}_{\theta, q_0, r_\epsilon | \mathbf{d}} [\theta, q_0, r_\epsilon | \mathbf{d}] \\ &= \int [\theta, q_0, r_\epsilon] p(\theta, q_0, r_\epsilon | \mathbf{d}) d\theta dq_0 dr_\epsilon. \end{aligned} \quad (16)$$

Remark 2. In statistical terms, the PM is an optimal estimator [42]: of all the possible functions of the data (be they Bayesian or not, empirical or not, a computation code, etc.), it is the one that yields the Minimum Mean Square Error (MMSE). Note that the MSE is the expected value of the squared norm of the difference between estimated and true values under the joint distribution of the observation and the unknown. Regarding first order statistics, this estimator has moreover a zero mean bias.

The calculation of the PM requires numerical evaluation of a multidimensional integral. In our problem, the expression of the posterior density does not allow the analytical calculation of Eq. (16), in particular because of the complexity of the likelihood expression. The developed solution is given by the Monte Carlo method. If $[\theta^{(n)}, q_0^{(n)}, r_\epsilon^{(n)}]$, $n = 0, \dots, N-1$ are N realizations of the posterior density by means of the MCMC algorithm described above, one can approximate the posterior mean (16) by the empirical mean:

$$[\hat{\theta}, \hat{q}_0, \hat{r}_\epsilon]_{\text{PM}} \simeq \frac{1}{N} \sum_{n=0}^{N-1} [\theta^{(n)}, q_0^{(n)}, r_\epsilon^{(n)}]. \quad (17)$$

According to the law of large numbers and the ergodicity, the expression (17) converges to the posterior mean when N tends to infinity.

Uncertainties can be obtained by the posterior covariance matrix $\hat{R}_{\theta, q_0, r_\epsilon | \mathbf{d}}$ that measures the dispersion of the posterior density around its mean. In the same manner as for the posterior mean, we approximate the posterior covariance matrix with the empirical one:

$$\begin{aligned} \hat{R}_{\theta, q_0, r_\epsilon | \mathbf{d}} &= \frac{1}{N} \sum_{n=0}^{N-1} \left[[\theta^{(n)}, q_0^{(n)}, r_\epsilon^{(n)}] - [\hat{\theta}, \hat{q}_0, \hat{r}_\epsilon]_{\text{PM}} \right]^t \\ &\quad \left[[\theta^{(n)}, q_0^{(n)}, r_\epsilon^{(n)}] - [\hat{\theta}, \hat{q}_0, \hat{r}_\epsilon]_{\text{PM}} \right]. \end{aligned} \quad (18)$$

Then, the confidence intervals at 95% are computed with the approximation

$$[[\hat{\theta}, \hat{q}_0, \hat{r}_\epsilon]_{\text{PM}} - 2\hat{\mathbf{r}}_R^{1/2}, [\hat{\theta}, \hat{q}_0, \hat{r}_\epsilon]_{\text{PM}} + 2\hat{\mathbf{r}}_R^{1/2}]$$

where $\hat{\mathbf{r}}_R$ is a row vector with diagonal elements of $\hat{R}_{\theta, q_0, r_\epsilon | \mathbf{d}}$.

4.3. Application procedure

Although being unsupervised (automatic estimation of the hyperparameters), the proposed method required some attention in order to produce accurate results. The procedure is listed below. And a flow chart is given in Fig. 2.

1. Select the source temporal model (see Section 2.1).
2. Formalize the prior knowledge on the source parameters as a prior density (see Section 3.3).
3. Initialize the source parameters.

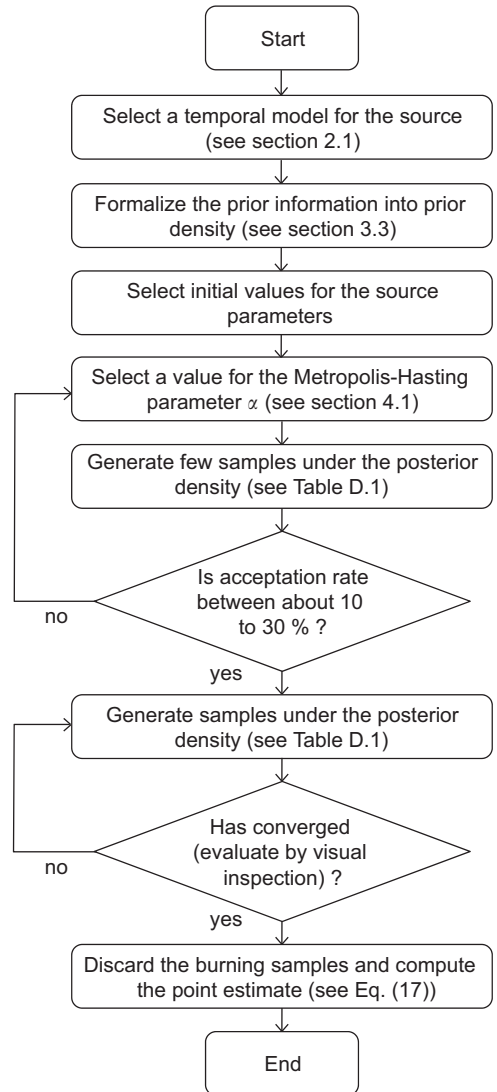


Fig. 2. Flow chart of the proposed method.

4. Select the Metropolis-Hastings sampler parameter α (see Section 4.1).
5. Execute the algorithm on few iterations.
6. Restart from step 4 until the acceptance rate is about 10–30%.
7. Evaluate the convergence by visual inspection of the posterior samples as function of the iterations.
8. Discard the burning samples and compute the point estimate (see Eq. (17)).
9. If needed, restart from 1 with another temporal model.

Step 1 would generally be done by visual inspection of the observations at some points as functions of time or by prior knowledge from the experts. Regarding the step 2, the choice between soft and hard constraints has been found to be not critical. A typical procedure would be to define admissible ranges of values for each parameter (possibly large if little knowledge is available) and to convert the minimum and maximum values of hard constraints into the mean and variance of soft constraints with $x_{\text{mean}} = (x_{\text{min}} + x_{\text{max}})/2$ and $x_{\text{var}} = (x_{\text{max}} - x_{\text{min}})^2/16$, respectively. In the proposed method, the choice of the initial value of the source parameter is arbitrary but can have an influence on the convergence time. In practice, the center of the prior range or the mean of the prior Gaussian density would make an acceptable starting point. As for

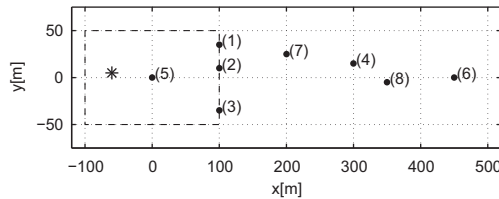


Fig. 3. Position of the sensors (points) on the hypothetical site. Depending on the case, the sensors 1–4, 1–6 or 1–8 are used. The star shows the location of the source in Sections 5.1.1 and 5.1.2. The rectangle (dash line) shows the possible location of the source in Section 5.2. The advection flow is hold by the x axis.

the initialization, the Metropolis-Hastings parameter only affects the convergence time. The proposed trial and error procedure to fix this parameter is common in the literature. Finally the evaluation of the convergence of the algorithm is let to the practitioner evaluation. This would usually be done by representing the samples and the point estimate as a function of the iterations: smooth evolution of the point estimate and good mixing of the posterior density are indicators of convergence.

5. Numerical applications

To evaluate the proposed method, various pollution events have been simulated on a hypothetical site derived from a real site. The site has four to eight sensors located as shown in Fig. 3. At first, the method's ability to recover the source is analyzed when correct temporal release shapes are selected, with a focus on the Laplacian temporal release. Then, an impact study of incorrect temporal models is presented. This validation stage is followed by a quantitative comparison with respect to two existing methods widely used in the literature ; nine configurations of sensors and noise that reproduce a large range of situations are considered. Finally, the behavior of the method is described on a real case where only one sensor measured the pollution. The transport model and the values of the physical parameters are detailed in Appendix A. Note that the advection phenomenon is carried by the increasing x -value, so the emission tends to move in this direction.

5.1. Validation of the proposed method

In this section, several pollution events are simulated from different source configurations. In each event, the contaminant plume is recorded by six sensors, the sensors 1–6 of Fig. 3. The simulated set of observations is generated by adding zero-mean Gaussian noise to the model output. In order to reproduce real measurement

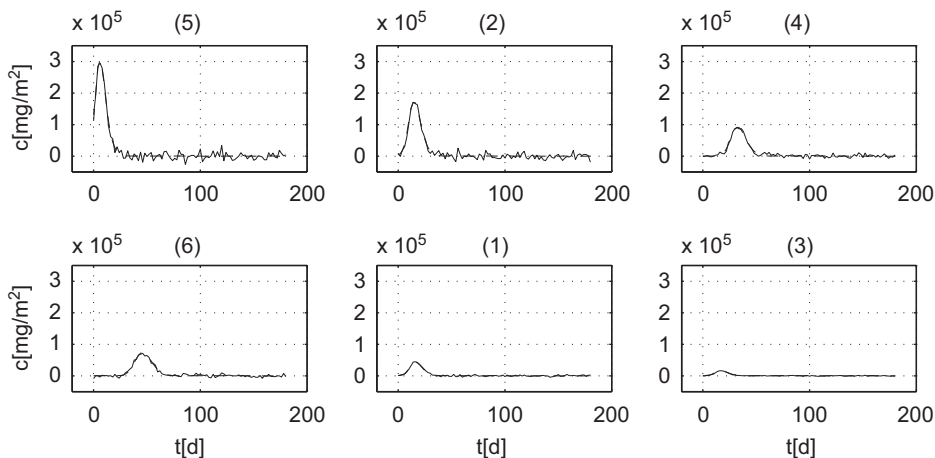


Fig. 4. Simulated (lines) and reconstructed (dash lines) observations on the hypothetical site. The abscissa represent the time in day, the ordinates represent the concentrations in milligrams per square meter. One observation per 2 days and per sensor is made (6×91 measurements). The simulated observations and the reconstructed output from the estimated parameters (Table 2) are here indistinguishable.

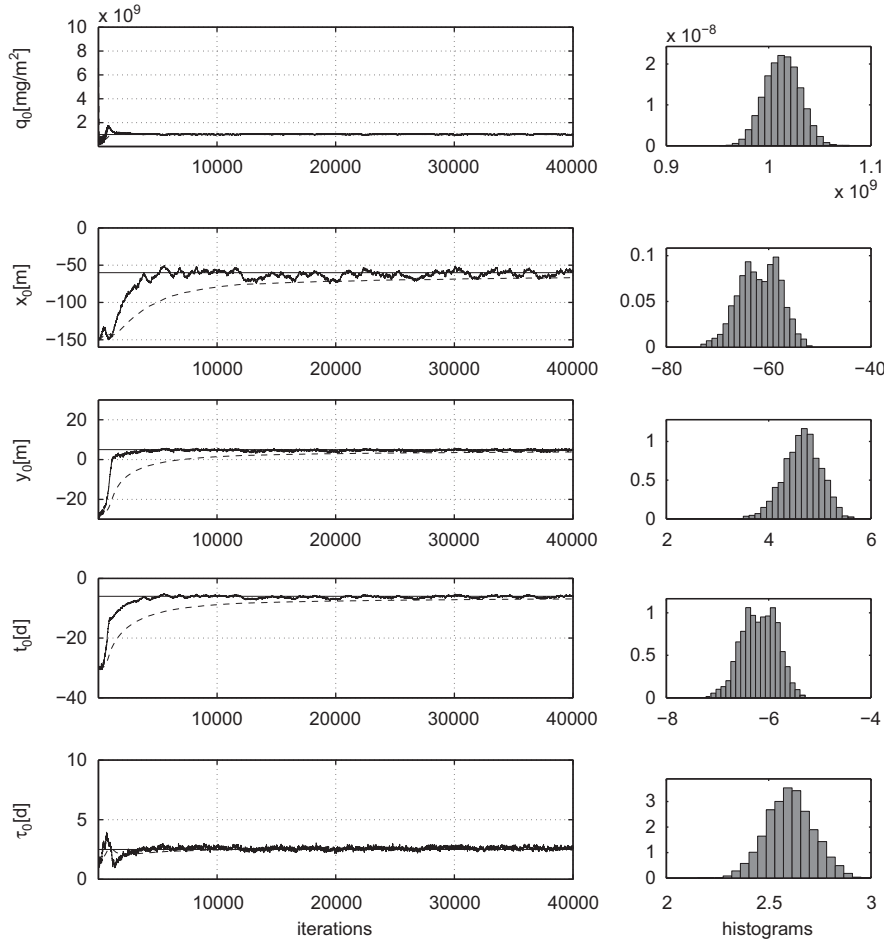


Fig. 5. Left-hand column: evolution of the parameters (solid lines) and the posterior mean (dashed lines) during the MCMC algorithm computation. The straight lines represent the true values of the parameters. Right-hand column: samples histograms of the marginal posterior density (last 2×10^4 samples).

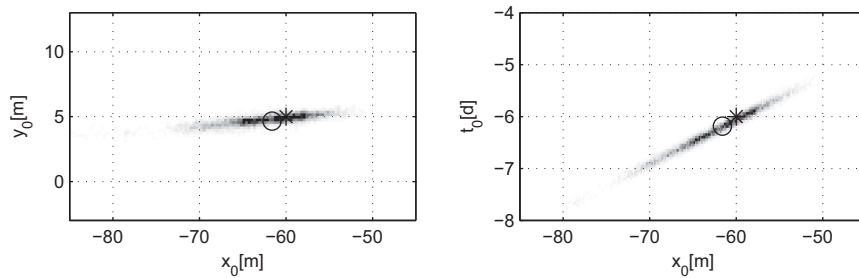


Fig. 6. Samples histograms of the marginal posterior density (last 2×10^4 samples). Left: in the x_0, y_0 space. Right: in the x_0, t_0 space. The star represents the true value of the parameters and the circle represents the mean of the samples.

conditions, the error variances are adjusted in order to get a similar signal to noise ratio (SNR) at each sensor:

$$r_e^{(k)} = \frac{10^{-\text{SNR}/10}}{J_k} \sum_{j=1}^{J_k} (q_0 h_{kj})^2 \quad (19)$$

where k indices the sensor, J_k is the number of measurements from the sensor k and h_{kj} is the model output at sensor k and time j . The SNR is computed in decibel with $\text{SNR} = 10 \log_{10}(\sum_{i=1}^M d_i^2 / \sum_{i=1}^M e_i^2)$ and has been set to 15 dB

in order to obtain observations similar to a real data set (the average of the error standard deviations is 0.12×10^5 mg/m²). Each sensor provides an observation every 2 days during 180 days, i.e. 6×91 concentration measurements. Simulated observations are represented in Fig. 4. The three main propagation phenomena (attenuation, advection and dispersion) can be seen: the further the sensors are from the source, the more attenuated, delayed and moved are the measurements.

5.1.1. Correct temporal model

Here, the same temporal model is used for generating the observations and for inverting. First, a point source with a Laplacian temporal release is considered. The true values of the source parameters are $\tau_0 = 2.5$ d, $q_0 = 10^9$ mg/m², $x_0 = -60$ m, $y_0 = 5$ m and $t_0 = -6$ d. The unsupervised MCMC algorithm is applied to generate samples from the posterior density of Eq. (8). Fig. 5 shows the

Table 2

Parameter estimation – Laplacian release.

	q_0 (mg/m ²)	x_0 (m)	y_0 (m)	t_0 (d)	τ_0 (d)
True	1.00×10^9	–60	5	–6	2.5
Estimation	1.01×10^9	–61.6	4.65	–6.16	2.61
Confidence	$\pm 0.04 \times 10^9$	± 10.7	± 0.85	± 0.91	± 0.26

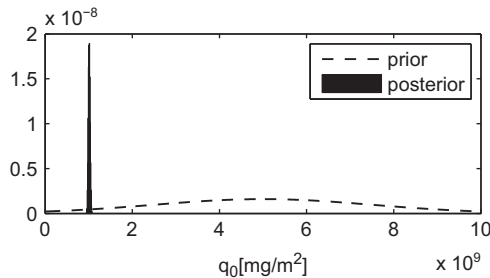


Fig. 7. Prior density and marginal posterior density for q_0 . To make the figure more clear, the prior has been multiplied by a factor 10.

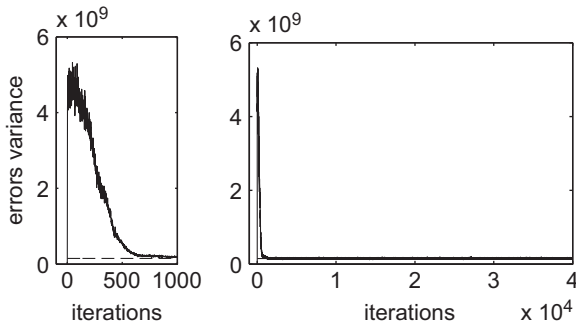


Fig. 8. Evolution of the error variance during the MCMC algorithm computation. Left-hand figure: the first 1000 iterations. Right-hand figure: all the iterations. The dash line represents the mean of the true error variances.

Table 3

Parameter estimation – different temporal releases (correct shapes).

	q_0 (mg/m ²) ($\times 10^9$)	x_0 (m)	y_0 (m)	t_0 (d)	τ_0 (d)
Dirac	1.03 ± 0.03	-66.5 ± 6.19	4.43 ± 0.69	-6.55 ± 0.53	No value
Step	1.01 ± 0.04	-62.4 ± 11.8	4.81 ± 0.92	-6.20 ± 1.28	No value
Window	1.06 ± 0.21	-66.0 ± 5.6	4.44 ± 0.66	-6.51 ± 0.48	2.60 ± 0.52
Decr. slope	1.03 ± 0.20	-65.7 ± 5.8	4.49 ± 0.71	-6.48 ± 0.43	2.62 ± 1.40
Triangle	1.04 ± 0.04	-58.9 ± 8.2	4.85 ± 0.77	-5.92 ± 0.65	5.23 ± 1.54
Laplacian	1.01 ± 0.04	-62.1 ± 9.8	4.63 ± 0.77	-6.21 ± 0.84	2.61 ± 0.26
Gaussian	1.00 ± 0.04	-58.2 ± 11.7	4.88 ± 0.90	-5.88 ± 0.99	2.68 ± 0.26
True values	1.00	–60.0	5.00	–6.00	2.50

samples as a function of iteration index (left-hand column). The first samples of the MCMC run (empirically evaluated to the first 2×10^4 samples) constitutes the so-called burn-in period: it is a current practice to throw it away, since the chain is not in its stationary state. Further development of the method could include automatic detection of convergence, for instance based on multi-chains criterion [46]. The dashed lines show the evolution of the iterative empirical mean of the samples and the straight lines represent the true values of the parameters. In this figure, the limits of the vertical axis are the prior bounds on the parameters. Evaluation of the uncertainties can be seen with the sample oscillations around the mean and with the histograms width (right-hand column).

The algorithm performed in about 3 min (on a standard desktop computer). Note that several hours of computation are usually accepted for such inverse problem, as the relatively low velocity of the pollutant propagation prevents from rapid change in the observations. The proposed method is therefore in the fast range.

In Fig. 6, histogram representations of marginal posterior density in the x_0, y_0 (left) and x_0, t_0 (right) space shows interesting shapes. First, the posterior density is particularly spiky (note the axis ranges). Second, the extension of the density is greater in the x direction, which means that small variations of the x_0 parameter still give acceptable source. Moreover in the x_0, t_0 space, the correlation between the two parameters, shown via the histogram orientation with a slope of about 11 m/d (close to the convection velocity $v_{x,2}$) expresses a space-time uncertainty (a source located in $x_0 = -70$ m and at $t_0 = -7$ d makes still an acceptable candidate).

The PM estimate is computed from the 2×10^4 last samples. Table 2 reports the true values, the PM estimates and 95% confidence intervals in the first, second and third line respectively. The estimated parameters are close to their true values, with small errors on q_0, τ_0, t_0 and y_0 . Relatively, the estimation x_0 is more inaccurate though it is still included in the confidence interval. This result is due to the importance of the advection phenomenon causing fast and strong carrying of the pollution into the groundwater. In the Fig. 4, the reconstructed concentrations obtained from the estimated source are also plotted (dash lines). It shows that the concentrations are correctly reconstructed and that the method naturally ensures the non-negativity of the data.

In addition, Fig. 7 plots the prior and the marginal posterior densities for q_0 . The integration of all available

information about the source leads to a very spiky posterior density compared with the prior.

The method also gives an estimation of the error variance r_e . As shown in Appendix D, the proposed algorithm gives an estimation the mean of the variances used to generate the observations on each sensor. Here, the variances mean is equal to 14.8×10^7 and the estimation is 14.3×10^7 with a confidence interval of $\pm 1.8 \times 10^7$. Fig. 8 shows the evolution of the error variance as a function of iterations with an initialization arbitrarily set to one. It can be seen on the right plot of Fig. 8 that the variance increases rapidly to a high value then converges around the mean of the true variances. This behavior is due to the poor data fitting during the first iterations. When the source parameters are strongly erroneous, the data adequation term $\|\mathbf{d} - q_0 \mathbf{h}(\theta)\|^2$ is high and the Inverse Gamma density of Eq. (14) will more likely generate high value of r_e . Besides, this is a nice behavior: high error variances lead to flat posterior densities and allow a wide exploration of the parameter space during the burning period.

Validation of the proposed approach for the other temporal releases listed in Section 2.1 is briefly presented below. As for the Laplacian release, results are summarized in a table (Table 3). It can be seen that the quantity q_0 is close to its true value, only slightly over estimated. Regarding the parameter τ_0 , the estimation error is less than 0.20 day (except for the Triangle). As the uncertainty is close to 3 days for the Decreasing slope and the Triangle, accurate estimation of τ_0 is not ensured for these releases. For the location x_0 , the greater error is for the Dirac distribution and the greater uncertainties are for the Step, the Laplacian and the Gaussian function. Note that in all the cases, the uncertainty on x_0 is important, as already pointed out. The difficulty to estimate the localization in the direction of the flow is therefore one of the specificities of the problem. For y_0 and t_0 , the errors are small with acceptable confidence intervals. Note that the uncertainty on t_0 is the most important for the Step release (± 1.28 d), which is the only release that never ends.

In conclusion, whatever the shape of the release, the parameters are accurately estimated. In most cases, the relative errors do not exceed 10% and are often under 5%. Moreover, the true value is almost always included in the 95% confidence interval.

5.1.2. Incorrect temporal model

In the present section, impact of erroneous temporal model on the source estimation is investigated. Indeed, the choice of the model for the temporal release can be incorrect,

as it would be made typically by visual inspection of the observations or by prior knowledge. To assess the robustness with respect to the source model error, two different models are used to generate the observations and to solve the inverse problem. Six situations are considered, summarized in Table 4, that leads to three main conclusions.

1. Using Laplacian (Gaussian) releases to estimate Dirac (Window) releases is efficient. The true releases are well fitted and the localization and amplitude parameters are estimated with good accuracy. For the Window/Gaussian case, the model error does not affect the quality of the estimation at all. For the Dirac/Laplacian case, only the quantity of pollutant is mainly under evaluation (factor 2).
2. Using Dirac (Window) releases to estimate Laplacian (Gaussian) releases do not allow to recover the source. The estimated sources are indeed located further in x (40–50 m before) and happened earlier (3–4 days). It can be explained by the insufficient degrees of freedom of the Dirac and Window distributions.
3. The difference between Laplacian and Gaussian functions appears to be too slight to be discriminated, as the crossed results are very good. The parameter τ_0 is adapted to fit the true release and the other parameters are well estimated. Note that the confidence intervals are generally larger than without source model error. For the parameter x_0 , the estimated value is close to the true one but the uncertainty is high, as the confidence intervals are greater than 20 m.

Following the results on the selected cases, the proposed method is quite robust to an erroneous source model. Without serious prior information about the shape of the release, the use of Laplacian and Gaussian releases as source models are strongly recommended, for they are more flexible and lead to the best estimates.

5.2. Comparison with existing methods

As mentioned in Remark 2, from a statistical standpoint, the chosen point estimate (the posterior mean) is an optimal one: of all the possible functions of the data, it yields the minimum mean square error [41]. The proposed simulated study reinforces this theoretical result. A large amount of 1000 sources has been drawn (uniformly distributed in the prior interval). For each source, a data set has been simulated

Table 4
Parameter estimation – different temporal releases (incorrect shapes).

True rel./Model rel.	q_0 (mg/m ²) ($\times 10^9$)	x_0 (m)	y_0 (m)	t_0 (d)	τ_0 (d)
Dirac/Laplacian	0.47 ± 0.05	-62.8 ± 9.5	4.61 ± 0.80	-6.25 ± 0.80	0.56 ± 0.35
Laplacian/Dirac	2.20 ± 0.07	-111.0 ± 7.9	2.53 ± 0.80	-10.35 ± 0.69	No value
Window/Gaussian	1.01 ± 0.03	-59.9 ± 7.2	4.78 ± 0.66	-6.01 ± 0.61	1.17 ± 0.22
Gaussian/Rect. win.	1.25 ± 0.23	-100.7 ± 8.14	2.80 ± 0.82	-9.48 ± 0.71	1.70 ± 0.30
Laplacian/Gaussian	1.00 ± 0.04	-61.5 ± 10.6	4.69 ± 0.87	-6.15 ± 0.91	3.10 ± 0.25
Gaussian/Laplacian	1.03 ± 0.04	-64.4 ± 10.9	4.46 ± 0.92	-6.40 ± 0.93	2.09 ± 0.26
True values	1.00	-60	5	-6	2.5

as described in the previous section: using the transport model and an additive Gaussian noise.

Regarding the processing approach, we have compared the proposed method to more generic ones already used to solve similar problems and to compare methods [4,21] (see also [47]). Both are founded on a Least Squares (LS) criterion and regarding the optimization algorithm, the one relies on a gradient approach [4] denoted by LS-G and the other relies on simulated annealing [21], denoted by LS-SA. The Matlab optimization toolbox [48] and the ASA implementation [49] were used respectively for the LS-G and LS-SA methods. Moreover, for ASA algorithm, the parameters of [21] have been used.

Then, each data set is processed by the proposed method and the LS ones. By empirical average of these 1000 results, the MSE is computed for the PM and for the LS estimates. The error is then normalized so that a null estimate yields a 100% error (and the perfect estimate yields a null error).

In order to deepen the study, we have assessed configurations varying the SNR (20 dB, 15 dB, and 0 dB) as well as varying the number of sensors ($K=8, 6$ and 4 , using sensors numbered 1–8, 1–6 and 1–4, respectively, in Fig. 3) and results are given in Table 5. The error varies from 5.46% to 64.19% depending on the configuration and the inverse method. As expected, for each method (each row) the error increases when the number of sensors decreases and when the noise level increases. For the PM method, it is always under 10% in 8 and 6 sensors configuration, and under 15% in the case of 4 sensors. Regarding the comparison between methods, these quantitative results clearly show the superiority of the proposed approach: it yields the smallest error among the three methods in each configuration (about 4–53 percent lower). In addition, methods with global exploration capability of the parameter space such as the proposed method or the LS-SA method seem mandatory, especially in case of very few sensors and high noise level.

Remark 3. It must be mentioned that for high SNR and large number of sensors, the measured data set brings an important amount of information and as a consequence, the posterior density may be very spiky (around the correct value). In such a case, the MCMC algorithm may require a larger number of iterations to explore the parameter space and find the area with substantial density. On the contrary, the gradient algorithm may be more efficient since it is based on directional exploration of the parameter space. Anyway, a combined version based on Langevin approaches [41,42,46] or more advanced approaches [50] could be used to overcome this difficulty if necessary.

5.3. Application on a real case

We propose here the application of the method on an underdetermined case where the source estimation cannot be fully achieved because of the lack of observations. In this case, a unique sensor recorded real observations of a tracer release. One of the advantages of the developed approach is indeed to allow strong regularization and to process situations where only a few noisy observations are available. Even if the source estimation cannot be done because of the non-uniqueness, we show here that the method is still relevant to process the prior information and the data in order to give guidelines for source estimation.

On the experimental site, only one sensor has measured the concentrations in the groundwater over a period of 18 months. Several sensors were displayed on the site, but the quality of their observations was not sufficient. The hydrogeologic conditions on this site are complex. However, the direct model of Appendix A makes an acceptable approximation, due to the relative high velocity of the water in the groundwater that dominates other

Table 5

Comparative study. The relative square root of the Mean Square Error (expressed as a percentage, %) is given for the three compared methods: Posterior Mean (PM), Least Squares with Simulated Annealing (LS-SA) and Least Squares with Gradient (LS-G). Nine configurations are investigated varying the number of sensors (K) and the noise level (SNR). One thousand trials per configuration are averaged.

K	8			6			4		
	20	15	0	20	15	0	20	15	0
PM	5.46	6.32	9.33	9.26	9.85	9.92	12.23	13.58	14.95
LS-SA	11.22	12.84	15.55	13.26	14.00	15.67	16.78	17.97	18.31
LS-G	13.98	24.78	56.28	32.66	49.87	62.90	40.52	56.57	64.19

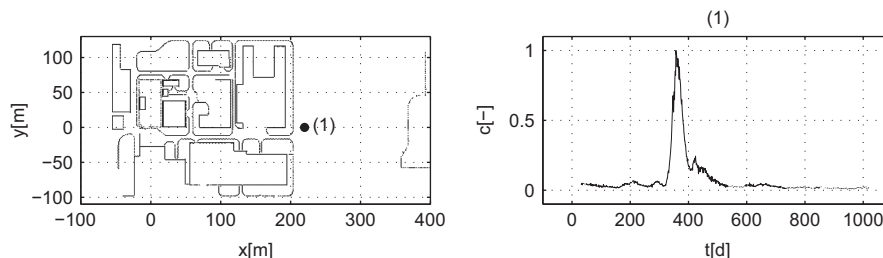


Fig. 9. Left-hand: position of the sensor on the real site (located at (220,0)). The prior source localization is $x_0 \in [-15, 15]$ m, $y_0 \in [5, 25]$ m. Right-hand: normalized observations made by the sensor.

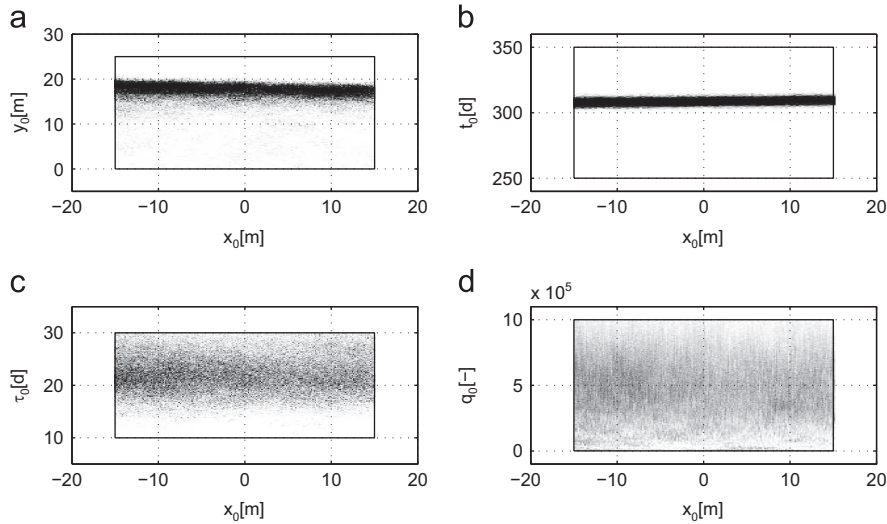


Fig. 10. Samples histograms of the marginal posterior density in the x_0, y_0 space (a), x_0, t_0 space (b), x_0, τ_0 space (c) and space x_0, q_0 (d). The lines denote the prior domain. The samples are clearly localized in the t_0 dimension, also in y_0 and τ_0 , but not at all in x_0 and q_0 .

phenomena. Site and observations are represented in Fig. 9. The maximum of the observations is set to one and the temporal axis starts with the first observation. The observation period is regular with approximately one measurement per day. Measurements indicate a polluting event of greater importance detected around the 380th day and other events of smaller magnitudes that could correspond to measurement noise or to releases of less importance. Our goal here is to characterize the source at the origin of the main observed peak. According to the shape of the peak and to the previous robustness study, the Laplacian function appears suited to model the source.

Due to the lack of spatial information on the pollutant plume, a large uncertainty on the source location is expected. As in the simulated cases, prior information is taken into account by uniform prior for the source parameters. Here, the experts had a prior knowledge of the origin building of the tracer release. The expected source was located in $x_0 \in [-15, 15]$ m, $y_0 \in [5, 25]$ m. Note that it is a relatively short range of the site; thus it supposed that the localization was almost known. On the other parameters, the prior knowledge of the experts was less accurate: the quantity is maximum 10^6 times higher than the maximum measurement, the pollution happened during a maximum of 2 months within 1 year before the measured peak. For the parameters q_0, t_0 and τ_0 , the prior domain is chosen large in a first step ($q_0 \in [10^{-3}, 10^6]$, $t_0 \in [100, 400]$, $\tau_0 \in [1, 30]$) and then it is reduced according to the explored space domain.

Considering the spiky and potentially multimodal nature of the posterior density, the convergence of the Markov chain may be slow and it may be necessary to simulate a great number of samples before convergence. In order to focus this numerical study on the behavior of the method in a very underdetermined situation, the error variance is not estimated. A supervised version of the algorithm is used. It is based on iterations of steps (1) and

(3) of the algorithm given in Table 1, with the step (2) replaced by a fixed value of r_e . The error variance is intentionally fixed to a relatively high value ($r_e = 0.1$) in order to highlight the structure of the posterior density.

We have generated 4.8×10^5 samples; we depicted two-dimensional histograms of the posterior density for (x_0, y_0) , (x_0, t_0) , (x_0, τ_0) and (x_0, q_0) in Fig. 10 (a), (b), (c) and (d), respectively. It can be seen that the width of the density in the y_0, t_0 and τ_0 axis is relatively small. The method gives a good idea of the most probable values of these parameters: y_0 around 19 m, t_0 around 309 days, and τ_0 around 22 days. For the two other parameters, the dispersion of the posterior density is important. It is shown in each figure for the parameters x_0 : no value seems to be more probable than the others and the density is truncated by the prior bounds (it indicates that larger prior ranges would be advisable). Fig. 10 (d) is particularly useful to illustrate the non-uniqueness of the inverse problem solution. The density covers all the x_0, q_0 space, which show that with the available observations and the incorporated prior information, no conclusion on the localization in the x axis and on the quantity of pollutant can be made (the localization in the direction of the flow is not possible). Therefore, the PM estimator (or any point estimate) is not relevant. Instead of giving a point estimate, representations of the posterior density and marginal point estimate are helpful to guide the source identification, for instance by the mean of marginal density histograms of samples. In the example presented here, it could help the experts to confirm the apparition of the contamination around 60 days before the observed abnormal concentrations, with a release duration close to 1 month.

6. Conclusions

In this paper, we have presented a Bayesian parametric method to estimate the location, temporal

release and amplitude of a point source. All available prior information about the source has been incorporated in a hierarchical Bayesian model and posterior law has been explored by means of an MCMC algorithm (Metropolis-Hastings within Gibbs sampler). Compared to the existing parametric methods entirely based on likelihood or least squares, the proposed method naturally incorporates additional prior information. In particular, it is more robust to difficult situations in which the number of available observations is small and the noise is high. Another important feature of the Bayesian approach and of MCMC algorithm is the natural possibility to evaluate uncertainties and to represent the solution space.

Applied on simulated cases built to mimic real contamination events, the relevance of the method has been demonstrated for sources with different temporal releases and robustness to source model error has been shown in various situations. In each case, the source is characterized with good accuracy and the true values are always included in the confidence intervals around the estimated parameters. A default of determination in the space-time plane has been identified to be responsible for larger uncertainty for the longitudinal (x -axis) location. A comparative study varying the signal-to-noise ratio and the number of sensors has been achieved and shows the superiority of the proposed method over the general gradient-based and simulated annealing-based least squares methods. The application of the method to a real case emphasizes its interest when the source is undetermined due to poorly informative data set. Instead of the point estimate, the representation of the posterior density makes it possible to supply clues on the most probable values of the source components.

A perspective of this work is to extend the estimation to some parameters of the transport model, such as the velocity or the dispersion coefficients. To this end, our Bayesian parametric approach is relevant: prior information regarding physical parameters is generally available and could be accounted for. In addition, the proposed method as it is, remains limited to relatively simple source and a natural perspective is to consider multiple sources in time. In this way, we plan to apply our method to the source models used by [25]. Moreover, the extension to automatic model selection will also be investigated within our Bayesian parametric approach based on Bayes factor and posterior sampler.

Acknowledgments

This work was partially supported by the French Ministry of Higher Education and Research via the National Association for Research and Technology (ANRT, CIFRE). The authors would like to thank S. Gautier, L. Le Saout and V. Just for being the instigators of the work, T. Kestens and J.-Y. Tourneret for their helpful comments on the paper. The authors are also grateful to Cornelia Vacar, IMS laboratory, for carefully reading the paper and for her helpful comments.

Appendix A. Transport model

The transport model used for the simulation is made up of two zones. Immediately beneath the surface, the first zone models the unsaturated medium in which propagation is solely according to the depth z . In the second zone, modeling the groundwater, propagation is anisotropic in three directions, with an advection velocity carried by the x axis. In each zone, the concentrations are governed by a partial differential equation namely an *advection–dispersion equation* (ADE) [6,51]. Transport models in zones 1 and 2 are given by Eqs. (A.1) and (A.2) respectively:

$$\frac{\partial c_1}{\partial t} + v_{z,1} \frac{\partial c_1}{\partial z} - \frac{D_{z,1}}{4} \frac{\partial^2 c_1}{\partial z^2} = s \delta(z) \quad (\text{A.1})$$

$$\begin{aligned} \frac{\partial c_2}{\partial t} + v_{x,2} \frac{\partial c_2}{\partial x} - \frac{D_{x,2}}{4} \frac{\partial^2 c_2}{\partial x^2} - \frac{D_{y,2}}{4} \frac{\partial^2 c_2}{\partial y^2} - \frac{D_{z,2}}{4} \frac{\partial^2 c_2}{\partial z^2} \\ = s_2 \delta(z-L) \end{aligned} \quad (\text{A.2})$$

where L represents the depth of zone 1, $v_{z,1}$, $v_{x,2}$ denote the advection velocities, $D_{z,1}$, $D_{x,2}$, $D_{y,2}$ denote the dispersion coefficients, c_1 and c_2 represent the concentrations in zones 1 and 2 respectively, s models the source at the surface and s_2 models the transfer from zone 1 to zone 2. s and s_2 represent a concentration of substance by unit of time, for instance a mass concentration rate in mg/d m². The term s_2 is obtained by equaling the incoming flux and outgoing flux at the interface of zones 1 and 2. It depends on the concentration in the first zone with the formula $s_2 = v_{z,1} c_1$.

We assume here that the two zones are homogeneous and invariant: their properties do not depend on space or time variables. Therefore, the coefficients of the equations are uniform and constant. This simplification produces a linear and invariant (convolutive) transport model.

More specifically, concentrations c_1 are defined for $0 \leq z \leq L$. To compute c_1 , it is assumed that the plume spreads out as if the medium was semi-infinite and only the concentrations in the interval $0 \leq z \leq L$ are considered. If we assume zero concentrations level in the environment before the beginning of pollution and at the infinite limits of the zones, the output of the transport model can be written as the following double convolution [52]: $c_2(x, y, z, t) = 2v_{z,1} h_1(t) * h_2(x, y, z, t) * s(x, y, t)$ where $*$ designates the convolution operator, h_1 and h_2 are the impulse responses in zones 1 and 2 respectively (Green functions), the factors 2 and $v_{z,1}$ are respectively modeling the mirror effect and the conservative flux due to the interface between the two zones [4,53]. The expressions of functions h_1 and h_2 are obtained by solving the ADE when the second member is a Dirac [53]:

$$h_1(t) = t^{-1/2} (4\pi D_{z,1})^{-1/2} \exp \left\{ -\frac{(L-v_{z,1}t)^2}{D_{z,1}t} \right\}$$

$$\begin{aligned} h_2(x, y, z, t) = t^{-3/2} ((4\pi)^3 D_{x,2} D_{y,2} D_{z,2})^{-1/2} \\ \exp \left\{ -\frac{(x-v_{x,2}t)^2}{D_{x,2}t} - \frac{y^2}{D_{y,2}t} - \frac{z^2}{D_{z,2}t} \right\}. \end{aligned}$$

At given t , the function h_2 is Gaussian w.r.t. (x, y, z) . At given x, y, z , the function h_2 is neither Gaussian nor symmetrical w.r.t. t .

For the simulated cases of Section 5.1, the physical model parameters are set as follows: $D_{z,1} = 1.33 \text{ m}^2/\text{day}$, $v_{z,1} = 0.66 \text{ m/d}$, $D_{x,2} = 224 \text{ m}^2/\text{d}$, $D_{y,2} = 44.9 \text{ m}^2/\text{d}$ and $v_{x,2} = 11.2 \text{ m/d}$. For the real case of Section 5.3, the parameters are $D_{z,1} = 1.33 \text{ m}^2/\text{d}$, $v_{z,1} = 0.66 \text{ m/d}$, $D_{x,2} = 310 \text{ m}^2/\text{d}$, $D_{y,2} = 4.97 \text{ m}^2/\text{d}$ and $v_{x,2} = 15.5 \text{ m/d}$.

Appendix B. Observation model

We consider that each sensor records a concentration measurement representing the total concentration on the depth of the groundwater, assumed infinite. To formalize the measurement procedure, the concentrations c_2 are integrated w.r.t. z :

$$c(x_k, y_k, t_{kj}) = \int_L^{+\infty} c_2(x_k, y_k, z, t_{kj}) dz \quad (\text{B.1})$$

where (x_k, y_k) with $k = 1, \dots, K$ denotes the coordinates of sensor k and t_{kj} with $j = 1, \dots, J_k$ denotes the j th instant of measurement at sensor k . $c(x_k, y_k, t_{kj})$ represents the concentration at the sensor k and time j after the propagation in zones 1 and 2. Note that c implicitly depends on c_1 : c is the integral of c_2 , c_2 and s_2 are linked by Eq. (A.2) and s_2 is a function of c_1 . Insofar as this integral can be calculated explicitly, the measurement system eliminates the dependency on z . In the expression of the concentrations c_2 , the impulsive response h_2 is replaced by \bar{h}_2 , independent of z :

$$\bar{h}_2(x, y, t) = t^{-1} (4\pi D_{2,x} D_{2,y})^{-1/2} \exp \left\{ -\frac{(x - v_{2,x}t)^2}{D_{2,x}t} - \frac{y^2}{D_{2,y}t} \right\}.$$

and the concentrations becomes $c(x, y, t) = 2v_{z,1}h_1(t) * \bar{h}_2(x, y, t) * s(x, y, t)$.

Appendix C. Conjugate prior densities for Gaussian likelihood

For the sake of completeness, this appendix reminds of the well-known expressions of the conjugate densities in case of Gaussian likelihood (see for instance [42, p. 97] for more results on conjugate priors). Let note $p(d|\mu_{\text{like}}, r_{\text{like}}) = \mathcal{N}(\mu_{\text{like}}, r_{\text{like}})$ the Gaussian likelihood with mean μ_{like} and variance r_{like} . With little calculation, it is not difficult to show the following results:

- if the prior density for the mean μ_{like} is Gaussian with parameter $p(\mu_{\text{like}}) = \mathcal{N}(\mu_{\text{pri}}, r_{\text{pri}})$, then the conditional posterior density for μ_{like} is still Gaussian with parameters

$$p(\mu_{\text{like}}|d, r_{\text{like}}) \propto \mathcal{N}\left(\frac{\mu_{\text{like}}r_{\text{like}} + dr_{\text{pri}}}{r_{\text{like}} + r_{\text{pri}}}, \frac{r_{\text{like}}r_{\text{pri}}}{r_{\text{like}} + r_{\text{pri}}}\right).$$

- if the prior density for the variance r_{like} is inverse gamma with parameter $p(r_{\text{like}}) = \mathcal{IG}(\alpha_{\text{pri}}, \beta_{\text{pri}})$, then the conditional posterior density for r_{like} is still an inverse gamma with parameters

$$p(r_{\text{like}}|d, \mu_{\text{like}}) \propto \mathcal{IG}\left(\alpha_{\text{pri}} + \frac{1}{2}, \beta_{\text{pri}} + \frac{(d - \mu_{\text{like}})^2}{2}\right).$$

Appendix D. Expectation of the error variance estimation

In this appendix, we prove that the variance estimated by our method converges to the mean of the variances on each sensor. First, it is easy to see that the PM estimate and the marginal PM estimate are equal, so $\hat{r}_e = \mathbb{E}_{r_e, \theta, q_0}(\mathbf{d}(r_e)) = \mathbb{E}_{r_e}(\mathbf{d}(r_e))$. Then, according to the Rao–Blackwell theorem [54], we have the following approximation of the marginal density: $p(r_e|\mathbf{d}) = 1/N \sum_{n=1}^N p(r_e|\mathbf{d}, \theta^{(n)}, q_0^{(n)})$ where $\theta^{(n)}, q_0^{(n)}$ are drawn from the joint posterior density. So it leads to the following expression for the estimate:

$$\begin{aligned} \mathbb{E}_{r_e, \theta, q_0}(\mathbf{d}(r_e)) &= \mathbb{E}_{r_e}(\mathbf{d}(r_e)) \\ &= \int r_e p(r_e|\mathbf{d}) dr_e \\ &\approx \int r_e \frac{1}{N} \sum_{n=1}^N p(r_e|\mathbf{d}, \theta^{(n)}, q_0^{(n)}) dr_e \\ &= \frac{1}{N} \sum_{n=1}^N \mathbb{E}_{r_e|\mathbf{d}, \theta^{(n)}, q_0^{(n)}}(r_e). \end{aligned}$$

In our case, the conditional density for r_e , $p(r_e|\mathbf{d}, \theta)$, is an Inverse Gamma density with parameters $\alpha = M/2, \beta = \|\mathbf{d} - q_0 \mathbf{h}(\theta)\|^2/2$ (assuming $\alpha_e = 0$ and $\beta_e = 0$), and its expectation is equal to $\beta/\alpha = \|\mathbf{d} - q_0 \mathbf{h}(\theta)\|^2/M$. Let us write $\mathbf{b}^{(n)} = \mathbf{d} - q_0^{(n)} \mathbf{h}(\theta^{(n)})$, so we have $\mathbb{E}_{r_e|\mathbf{d}, \theta^{(n)}, q_0^{(n)}}(r_e) = \|\mathbf{b}^{(n)}\|^2/M$. Then, by taking the expectation:

$$\mathbb{E}(\hat{r}_e) \approx \frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M \mathbb{E}(\|\mathbf{b}_m^{(n)}\|^2)$$

where M is the observation number. By using the relation $\text{VAR}(\mathbf{b}_m^{(n)}) = \mathbb{E}(\|\mathbf{b}_m^{(n)}\|^2) - \mathbb{E}(\mathbf{b}_m^{(n)})^2$, we obtain

$$\mathbb{E}(\hat{r}_e) \approx \frac{1}{NK} \sum_{n=1}^N \sum_{k=1}^K \frac{1}{J_k} \sum_{j=1}^{J_k} (r_{e,k,j}^{(n)})^2$$

where K is the sensor number, J_k the observation number at sensor k . Finally remind that $r_{e,k,j}$ is constant on each sensor k so the mean for the sensor k is simply $1/J_k \sum_{j=1}^{J_k} (r_{e,k,j}^{(n)})^2 = (r_{e,k}^{(n)})^2$, we have

$$\mathbb{E}(\hat{r}_e) \approx \frac{1}{K} \sum_{k=1}^K \frac{\sum_{n=1}^N (r_{e,k}^{(n)})^2}{N} = \frac{1}{K} \sum_{k=1}^K (\hat{r}_{e,k})^2$$

Thus, the mean estimate value can be approximated by the mean of the empirical variances on the sensors.

References

- [1] A.M. Stuart, Inverse problems: a Bayesian perspective, *Acta Numerica* 19 (2010) 451–559398–409.
- [2] D. McLaughlin, L.R. Townley, A reassessment of the groundwater inverse problem, *Water Resources Research* 32 (5) (1996) 1131–1161.
- [3] J. Atmadja, A.C. Bagtzoglou, State of the art report on mathematical methods for groundwater pollution source identification, *Environmental Forensics* 2 (2001) 205–214.
- [4] A. Nehorai, B. Porat, E. Paldi, Detection and localization of vapor-emitting sources, *IEEE Transactions on Signal Processing* 43 (1) (1995) 243–253.
- [5] I. Dimov, U. Jaekel, H. Vereecken, A numerical approach for determination of sources in transport equation, *Computers and Mathematics with Applications* 32 (5) (1996) 31–42.
- [6] N.R.C. NRC, *Groundwater Models – Scientific and Regulatory Applications*, National Academy Press, Washington, DC, 1990.
- [7] A. Tarantola, *Inverse Problem Theory. Method for Data Fitting and Model Parameter Estimation*, Elsevier Science, Amsterdam, 1994.
- [8] J. Idier, *Bayesian Approach to Inverse Problems*, ISTE Ltd and John Wiley & Sons, Inc, 2008.

- [9] F. Van der Heijden, R.P.W. Duin, D. de Ridder, D.M.J. Tax, Classification, Parameter Estimation and State Estimation. An Engineering Approach Using MatLab, Wiley, Chichester, 2004.
- [10] P. Kathirgamanathan, R. McKibbin, R. McLachlan, Source term estimation of pollution from an instantaneous point source, *Research Letters in the Information and Mathematical Sciences* 3 (2002) 59–67.
- [11] J. Matthes, L. Groll, H.B. Keller, Source localization based on pointwise concentration measurements, *Sensors and Actuator A (Physical)* 115 (2004) 32–37.
- [12] V.N. Christopoulos, S. Roumeliotis, Multi Robot Trajectory Generation for Single Source Explosion Parameter Estimation, Technical Report, University of Minnesota, 2004.
- [13] A. Jeremic, A. Nehorai, Landmine detection and localization using chemical sensor array processing, *IEEE Transactions on Signal Processing* 48 (5) (2000) 1295–1305.
- [14] A. Jeremic, A. Nehorai, Detection and estimation of biochemical sources in arbitrary 2d environments, in: ICASSP, Toulouse, France, 2005.
- [15] P. Sidauruk, A.-D. Cheng, D. Ouazar, Ground water contaminant source and transport parameter identification by correlation coefficient optimization, *Groundwater* 36 (2) (1998) 208–214.
- [16] J. Matthes, L. Groll, H.B. Keller, Source localization by spatially distributed electronic noses for advection and diffusion, *IEEE Transactions on Signal Processing* 53 (5) (2005) 1711–1719.
- [17] J.A. Pudykiewicz, Application of adjoint tracer transport equations for evaluating source parameters, *Atmospheric Environment* 32 (17) (1998) 3039–3050.
- [18] M.E. Alpay, M.H. Shor, Model-based solution techniques for the source localization problem, *IEEE Transactions on Control Systems Technology* 8 (6) (2000) 895–904.
- [19] R.M. Neupauer, J.L. Wilson, Adjoint-derived location and travel time probabilities for a multidimensional groundwater system, *Water Resources Research* 37 (6) (2001) 1657–1668.
- [20] N.K. Ala, P.A. Domenico, Inverse analytical techniques applied to coincident contaminant distributions at Otis Air Force Base, Massachusetts, *Groundwater* 30 (2) (1992) 212–218.
- [21] M. Jha, B. Datta, Three dimensional groundwater contamination source identification using adaptive simulated annealing, *Journal of Hydrologic Engineering* 18 (3) (2012) 307–317.
- [22] B.J. Wagner, Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modelling, *Journal of Hydrology* 135 (1992) 275–303.
- [23] S.M. Gorelick, Identifying source of groundwater pollution: an optimization approach, *Water Resources Research* 19 (3) (1983) 779–790.
- [24] A. Keats, E. Yee, F.-S. Lien, Bayesian inference for source determination with applications to a complex urban environment, *Atmospheric Environment* 41 (3) (2007) 465–479.
- [25] S. Alapati, Z.J. Kabala, Recovering the release history of a groundwater contaminant using a non-linear least-squares method, *Hydrological Processes* 14 (2000) 1003–1016.
- [26] A.M. Michalak, P.K. Kitanidis, Estimation of historical groundwater contamination distribution using the adjoint state method applied to geostatistical inverse modeling, *Water Resources Research* 40 (8) (2004).
- [27] A.N. Tikhonov, V.Y. Arsenin, *Solutions of Ill-Posed Problems*, Winston and Sons, New York, 1977.
- [28] T.H. Skaggs, Z.J. Kabala, Recovering the history of a groundwater contaminant plume, *Water Resources Research* 30 (1) (1994) 71–79.
- [29] C. Liu, P. Ball, Application of inverse methods to contaminant source identification from aquitard diffusion profiles at Dover AFB, Delaware, *Water Resources Research* 35 (7) (1999) 1975–1985.
- [30] P. Kathirgamanathan, R. McKibbin, R. McLachlan, Source release-rate estimation of atmospheric pollution from a non-steady point source. Part 1: source at a known location, *Research Letters in the Information and Mathematical Sciences* 5 (2003) 71–84.
- [31] A. Woodbury, T. Ulrych, Minimum relative entropy inversion: theory and application to the recovery of the release history of a groundwater contaminant, *Water Resources Research* 32 (9) (1996) 2671–2681.
- [32] A. Woodbury, E. Sudicky, T.J. Ulrych, R. Ludwig, Three-dimensional plume source reconstruction using minimum relative entropy inversion, *Journal of Contaminant Hydrology* 32 (1998) 131–158.
- [33] R.M. Neupauer, B. Borchers, A Matlab implementation of the minimum relative entropy method for linear inverse problems, *Computers and Geosciences* 27 (2001) 757–762.
- [34] M.F. Snodgrass, P.K. Kitanidis, A geostatistical approach to contaminant source identification, *Water Resources Research* 33 (4) (1997) 537–546.
- [35] A.M. Michalak, P.K. Kitanidis, Application of geostatistical inverse modeling to contamination source identification at Dover AFB, Delaware, *Journal of Hydraulic Research* 42 (2004) 9–18.
- [36] P.K. Kitanidis, The minimum structure solution to the inverse problem, *Water Resources Research* 33 (10) (1997) 2263–2272.
- [37] P. Kathirgamanathan, R. McKibbin, R. McLachlan, Source release-rate estimation of atmospheric pollution from a non-steady point source part 2: source at an unknown location, *Research Letters in the Information and Mathematical Sciences* 5 (2003) 85–118.
- [38] P. Kathirgamanathan, Source parameter estimation of atmospheric pollution from accidental gas releases, in: *International Environmental Modelling and Software Society (IEMSS)*, 14–17/06/2004, University of Osnabrück, Germany, 2004.
- [39] M. Bocquet, Grid resolution dependence in the reconstruction of an atmospheric tracer source, *Nonlinear Processes in Geophysics* 12 (2005) 219–234.
- [40] A. Hazart, J.-F. Giovannelli, S. Dubost, L. Chatellier, Contaminant source estimation in a two-layers porous environment using a Bayesian approach, in: *IGARSS*, 23–27/07/2007, Barcelona, Spain, 2007.
- [41] C. Robert, G. Casella, *Monte-Carlo Statistical Methods*, Springer, New York, USA, 2004.
- [42] C. Robert, *The Bayesian Choice: From Decision-Theoretic Motivations to Computational Implementation*, Springer Verlag, 2001.
- [43] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (6) (1984) 721–741.
- [44] L. Tierney, Markov chains for exploring posterior distributions, *Annals of Statistics* 22 (1994) 1701–1762.
- [45] G. Roberts, A. Gelman, W. Gilks, Weak convergence and optimal scaling of random walk metropolis algorithms, *Annals of Applied Probability* 7 (1) (1997) 110–120.
- [46] A. Gelman, D.B. Rubin, Inference from iterative simulation using multiple sequences, *Statistical Science* 7 (4) (1992) 457–472.
- [47] M. Redwood, Source Term Estimation and Event Reconstruction: A Survey, Contract Report for ADMLC, 51790, 2011.
- [48] *Optimization Toolbox User's Guide*, The MathWorks, Inc., Natick, MA, USA, 2012.
- [49] L. Ingber, *Adaptive Simulated Annealing (ASA) Global Optimization C-Code*, Caltech Alumni Association, 1993.
- [50] C. Vacar, J.-F. Giovannelli, Y. Berthoumieu, in: *Proceedings of IEEE ICASSP*, Prague, Czech Republic, 2011.
- [51] J. Bear, *Dynamics of Fluids in Porous Media*, Elsevier, New York, 1972.
- [52] D. Bleeker, G. Csordas, *Basic Partial Differential Equations*, International Press, 1996.
- [53] J. Kevorkian, *Partial Differential Equations: Analytical Solution Techniques*, Springer Verlag, 2000.
- [54] A. Gelfand, A. Smith, Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* 85 (1990) 398–409.
- [55] E. Yee, Theory for reconstruction of an unknown number of contaminant sources using probabilistic inference, *Boundary-Layer Meteorology* 127 (3) (2008) 359–394.
- [56] C. Huang, T. Hsing, N. Cressie, A.R. Ganguly, V.A. Protopopescu, N.S. Rao, Bayesian source detection and parameter estimation of a plume model based on sensor network measurements, *Applied Stochastic Models in Business and Industry* 26 (4) (2010) 331–348.
- [57] H. Wang, X. Jin, Characterization of groundwater contaminant source using Bayesian method, *Stochastic Environmental Research and Risk Assessment* 27 (4) (2012) 867–876.
- [58] M. Ortner, A. Nehorai, A. Jeremic, Biochemical transport modeling and Bayesian source estimation in realistic environments, *IEEE Transactions on Signal Processing* 55 (6) (2007) 2520–2532.