Sequential Monte Carlo sampler applied to source term estimation in complex atmospheric environments

François Septier, Patrick Armand, Christophe Duchenne

Version of Record: https://www.sciencedirect.com/science/article/pii/S1352231021006440
Manuscript_27ceb0b87f0ec0ea36b4a295224808b2

Sequential Monte Carlo Sampler Applied to Source Term Estimation in Complex Atmospheric Environments

François Septier[a,*], Patrick Armand[b], Christophe Duchenne[b]

[a] *Université Bretagne Sud, LMBA UMR CNRS 6205, F-56000 Vannes, France*
[b] *CEA, DAM, DIF, F-91297 Arpajon, France*

Abstract

The accurate and rapid reconstruction of a pollution source represents an important but challenging problem. Several strategies have been proposed to tackle this issue among which we find the Bayesian solutions that have the interesting ability to provide a complete characterization of the source parameters through their posterior probability density function. However, these existing techniques have certain limitations such as their computational complexity, the required model assumptions, their difficulty to converge, the sensitive choice of model/algorithm parameters which clearly limit their easy use in practical scenarios. In this paper, to overcome these limitations, we propose a novel Bayesian solution based on a general and flexible population-based Monte Carlo algorithm, namely the sequential Monte Carlo sampler. Owing to its full adaptivity through the learning process, the main advantage of such an algorithm lies in its capability to be used without requiring any specific assumptions on the underlying statistical model and also without requiring from the user any difficult choices of certain parameter values. The performance of the proposed inference strategy is assessed using twin experiments in complex built-up environments.

Keywords: Source term estimation, Bayesian approach, dispersion model, adaptivity, sequential importance sampling, synthetic example.

1. Introduction

The threat of chemical, radiological, biological, and nuclear (CRBN) releases raises some complex and challenging scientific issues. In the event of a CBRN incident, it is of great importance to have as soon as possible an accurate assessment of the damage likely to be caused by the release which is usually undertaken using an atmospheric dispersion model of the contaminant. However, such spatial and temporal forecast, especially in complex built-up environments, is only possible by providing the source term parameters to the dispersion model. As a consequence, the main objective consists in rapidly obtaining an accurate estimation of the source parameters from noisy measurements of the concentration levels observed by several sensors deployed in a specific surveillance area.

*Corresponding author
Email address: francois.septier@univ-ubs.fr (François Septier)

Due to its ill-posed nature, this challenging inverse problem has been tackled using the proposition of different algorithms which can be grouped into two categories. The first one aims at obtaining a single point estimate of the unknown parameters by solving an optimization problem where a cost function has to be minimized using least squares or genetic algorithms, e.g (Winiarek et al., 2012).

Unfortunately, such approaches do not allow us to quantify the uncertainty associated to the unknowns, which could be really problematic in such an important context. To overcome this limitation, Bayesian algorithms have been designed to solve an inference problem by providing the complete probability density function of the parameters of interest given the measurements observed by a network of sensors. Owing to the complex and nonlinear nature of the source term estimation (STE) model, the exact computation of such a distribution is not feasible analytically, and one has to resort to some approximation techniques such as stochastic simulation algorithms.

More specifically, these techniques consists in obtaining samples from the distribution of interest by using either some Markov Chain Monte Carlo (MCMC) kernel (Delle Monache et al., 2008; Chow et al., 2008; Keats et al., 2007a; Yee et al., 2014) or using the principle of importance sampling (Septier, 2019; Septier et al., 2020). As recently discussed in (Septier et al., 2020), the application of an Adaptive Multiple Importance Sampling (AMIS) technique on the challenging STE problem allows us to obtain significant gain compared to state-of-the-art MCMC algorithms in both synthetic and real data experiments.

Unfortunately, this AMIS algorithm has been designed for a specific observation model and in particular, it is assumed that both the measurement errors and the emission rate levels follow a normal distribution which is not fully appropriate to real situations since negative values can appear in the concentration levels and have to be removed to keep the physical meaning of the data. However, relaxing these model assumptions will result in an increasing dimension of the state to estimate, thus leading generally to poorer estimation results as the performance of this AMIS algorithm is quite sensitive to the dimension related to the unknown parameters. Moreover, the parameters of these normal distributions have to be set a priori by the user which could clearly be difficult in practice.

In this work, we propose a novel Bayesian STE strategy based on a more general and flexible population-based Monte Carlo framework. Originally proposed in (Del Moral et al., 2006; Peters, 2005), this technique called Sequential Monte-Carlo (SMC) sampler is a class of sampling algorithms which combine importance sampling and resampling. They have been primarily used as "particle filter" to solve optimal filtering problems (Doucet et al., 2000). In this context, SMC methods/particle filters have benefited from wide-spread use in various applications (tracking, computer vision, digital communications) due to the fact that they provide a simple way of approximating complex filtering distribution sequentially in time. But in (Del Moral et al., 2006), the authors developed a general framework that allows SMC to be used to simulate from a single and static target distribution, thus becoming a promising alternative to standard MCMC methods. The SMC sampler framework involves the construction of a sequence of artificial distributions on spaces of increasing dimensions which admit the distributions of interests as particular marginals. The mechanism is similar to sequential importance sampling (resampling) (Doucet et al., 2001; Liu, 2001), with one of the crucial differences being the framework under

which the random samples, also called particles, are allowed to move, resulting in differences in the calculation of the weights of the particles.

These methods have several advantages over traditional and population-based MCMC methods. Firstly, unlike MCMC, SMC methods do not require any burn-in period and do not face the sometimes contentious issue of diagnosing convergence of a Markov chain. Secondly, as discussed in (Jasra et al., 2007), compared to population-based MCMC, SMC sampler is a richer method since there is substantially more freedom in specifying the mutation kernels in SMC: kernels do not need to be reversible or even Markov. As a consequence, adaptive proposal distributions can be easily used, thus giving a lot more of opportunities to improve its efficiency. Moreover, unlike MCMC, SMC samplers provide an unbiased estimate of the normalizing constant of the posterior distribution which can be one quantity of interest in the inference problem to deal with. Indeed, this normalizing constant is the marginal likelihood and therefore could be used for selecting some model assumptions (e.g. choice of distribution for the measurement noise or for the prior of parameters). Let us finally denote that many other inference technique based on importance sampling such as Annealed Importance sampling (Neal, 2001), population Monte Carlo (Cappé et al., 2004) and its more advanced variants like the Adaptive Multiple Importance Sampling (AMIS) (Cornuet et al., 2012) can all be considered as special cases of the SMC sampler.

Although this approach presents many advantages over traditional MCMC methods, the potential of these emergent techniques is however largely underexploited in practice. In (Nguyen et al., 2016), some strategies have been described and proposed in order to make easier its efficient implementation in practical problems. In this work, we propose to design an STE algorithm based on this sequential Monte Carlo sampler. The rest of this paper is organized as follows. Section 2 describes the statistical model and describes the Bayesian framework used for this STE problem. In Section 3, the general principle of the SMC sampler is firstly presented, then we describe the proposed SMC sampler applied to the STE problem. Numerical simulations are performed in Section 4 to assess the performances of the proposed approach. Conclusions are finally given in Section 5.

2. Statistical Model of the STE Problem

This section firstly describes the statistical model of the source term estimation problem, and then the Bayesian framework for estimating the characteristics of the source is discussed.

2.1. Atmospheric dispersion model

In this paper, a point-wise and static source fully characterized by the parameter $\boldsymbol{\beta} = [\boldsymbol{x}_s, \boldsymbol{q}]$ is considered where $\boldsymbol{x}_s = [x_s, y_s]$ stands for the spatial position of the source and $\boldsymbol{q}$ is the release rate vector resulting from the discretization of the plausible emission time window into T_s time intervals.

A network of N_c sensors is deployed over a 2-dimensional monitoring region to measure the concentration levels. By using an atmospheric dispersion model, the output simulated concentration at the location of the i-th sensor at time t_j

is defined as

$$\widetilde{y}_{i,j} = \sum_{n=1}^{T_s} q_n C_{i,j}(\boldsymbol{x}_s, \Delta t_n), \tag{1}$$

where $j = 1, \ldots, T_c$ with T_c the number of time samples collected by each sensor. Each value results from the superposition of the T_s releases at the different time steps $\{\Delta t_n\}_{n=1}^{T_s}$ weighted by their associated emission rates $\{q_n\}_{n=1}^{T_s}$ of the source. $C_{i,j}(\boldsymbol{x}_s, \Delta t_n)$ corresponds therefore to the simulated concentration obtained at the i-th sensor at time t_j if a unitary release is made during the time step Δt_n from a source that is located at $\boldsymbol{x}_s$. Let us note that the proposed method can be used for any specific choice of atmospheric dispersion model as long as we are able to obtain $C_{i,j}(\boldsymbol{x}_s, \Delta t_n)$. All these simulated concentration values obtained at the different time samples of all sensors can be written in the following matrix form:

$$\widetilde{\boldsymbol{y}} = \boldsymbol{C}(\boldsymbol{x}_s)\boldsymbol{q}, \tag{2}$$

where $\widetilde{\boldsymbol{y}} = \begin{bmatrix} \widetilde{y}_{1,1} & \cdots & \widetilde{y}_{1,T_c} & \cdots & \widetilde{y}_{N_c,1} & \cdots & \widetilde{y}_{N_c,T_c} \end{bmatrix}^T$ is the vector of simulated concentration values from the used atmospheric dispersion model and $\boldsymbol{C}(\boldsymbol{x}_s)$, generally called *source-receptor* matrix (Seibert and Frank, 2004), takes the following matrix form

$$\boldsymbol{C}(\boldsymbol{x}_s) = \begin{bmatrix} C_{1,1}(\boldsymbol{x}_s, \Delta t_1) & \cdots & C_{1,1}(\boldsymbol{x}_s, \Delta t_{T_s}) \\ \vdots & \ddots & \vdots \\ C_{1,T_c}(\boldsymbol{x}_s, \Delta t_1) & \cdots & C_{1,T_c}(\boldsymbol{x}_s, \Delta t_{T_s}) \\ \vdots & \ddots & \vdots \\ C_{N_c,1}(\boldsymbol{x}_s, \Delta t_1) & \cdots & C_{N_c,1}(\boldsymbol{x}_s, \Delta t_{T_s}) \\ \vdots & \ddots & \vdots \\ C_{N_c,T_c}(\boldsymbol{x}_s, \Delta t_1) & \cdots & C_{N_c,T_c}(\boldsymbol{x}_s, \Delta t_{T_s}) \end{bmatrix}. \tag{3}$$

The computation of this source-receptor matrix is an important part in an STE procedure as it links the source characteristics with the measurements and quantifies the predicted concentration value at some location and time from a dispersion model for a given source. As a consequence, in stochastic simulation based inference techniques, this matrix has to be computed for each generated sample which are quite numerous (at least several thousands) for a satisfactory estimation accuracy level. The computation of this matrix with a Lagrangian particle dispersion model (LPDM) in a forward mode constitutes the most time-consuming step of the algorithm proposed in (Rajaona et al., 2015). In this study as proposed in (Keats et al., 2007a; Yee et al., 2008; Septier et al., 2020), we use an alternative strategy which consists in using instead the backward mode of a LPDM. Using this backward mode is computationally advantageous if the number of receptors is less than the number of sources, which is generally the case in practice.

2.2. Definition of the observation model

The likelihood function expresses the probability of observing a specific set of concentrations from the N_c sensors sensors given all the parameters of the source $\boldsymbol{\beta}$. In other words, this likelihood function provides a probabilistic information about the discrepancy between the measured concentration values,

denoted by the vector $\boldsymbol{y}$, and the simulated concentration values, $\widetilde{\boldsymbol{y}}$, obtained from the dispersion model and defined in Eq. (2). The chosen parametric distribution for this likelihood should therefore characterize the three classical types of error: the dispersion modeling error, the observation error and the representativeness error due to the interpolation in both time and space of the dispersion model (Koohkan and Bocquet, 2012). Since the $\widetilde{\boldsymbol{y}}$ is a complex but deterministic mapping of the source parameters $\boldsymbol{\beta}$, let us denote the likelihood function by

$$p(\boldsymbol{y}|\widetilde{\boldsymbol{y}};\boldsymbol{\Psi}_l) = p(\boldsymbol{y}|\boldsymbol{\beta};\boldsymbol{\Psi}_l), \tag{4}$$

where $\boldsymbol{\Psi}_l$ represents all the parameters of the specific parametric likelihood chosen in the study. The parameters $\boldsymbol{\Psi}_l$ are assumed either known by a calibration from historical data or unknown and therefore need to be jointly estimated with the source parameters $\boldsymbol{\beta}$. For the rest of the paper, we will keep this general formulation of the likelihood since the estimation method we propose in this work does not depend on the choice of this likelihood - the only assumption we need is the ability of evaluating this function point-wise.

As an illustration of a possible plausible choice of likelihood function which has been used in (Lewellen and Sykes, 1986; Septier et al., 2009) in order to ensure the positivity of measurements, the noisy concentration of pollutant at a given location and time can be considered as conditionally independent and can be modeled by a clipped normal distribution, i.e.

$$p(\boldsymbol{y}|\widetilde{\boldsymbol{y}};\boldsymbol{\Psi}_l) = \prod_{i=1}^{N_c}\prod_{n=1}^{T_c} \mathcal{CN}\left(y_{i,n};\widetilde{y}_{i,n},\sigma_\epsilon^2\right), \tag{5}$$

where

$$\mathcal{CN}\left(y;\mu,\sigma_\epsilon^2\right) = \begin{cases} 0 & \text{if } y<0 \\ \Phi(-\frac{\mu}{\sigma_\epsilon}) & \text{if } y=0 \\ \mathcal{CN}\left(y_{i,n};\widetilde{y}_{i,n},\sigma_\epsilon^2\right) & \text{otherwise} \end{cases} \tag{6}$$

where $\Phi(y)$ is the cumulative distribution function (CDF) of the standard normal distribution. The delta function at zero corresponds to the intermittency (periods of zero concentration) in observed concentrations. With such a choice, the (possibly unknown) parameter of the distribution is $\boldsymbol{\Psi}_l = \sigma_\epsilon^2$ which corresponds to the variance (level) of the noise in the measurements. This likelihood is clearly more adapted to the problem under study than the traditional normal distribution used for example in (Yee et al., 2008; Rajaona et al., 2015; Septier et al., 2020).

2.3. A priori knowledge about the source parameters

Our belief regarding the characteristics of the unknown state of interest, $\boldsymbol{\beta}$, is encapsulated within the prior probability distributions of the proposed Bayesian model. As pointed out above with the likelihood, the method we propose in this work could be applied for any choice of prior probability distribution for the source parameters $\boldsymbol{\beta}$ as soon as sampling from such distribution and its point-wise evaluation are possible.

As an illustration, if we just know that the release could appear anywhere uniformly in the region of surveillance denoted here by $\Omega \subseteq \mathbb{R}^2$, the following uniform prior distribution could be chosen for the position of the source:

$$p(\boldsymbol{x}_s) = \mathcal{U}_\Omega(\boldsymbol{x}_s). \tag{7}$$

Of course, in some scenarios of interest, it could be more appropriate to incorporate a more informative distribution to represent our initial guess about this source location (nuclear plants, industrial sites, etc).

Regarding now the emission rate vector, $\boldsymbol{q}$, as in (Keats et al., 2007b), we consider a point source which releases material at a steady rate of α and whose turn-on and turn-off times are t_{on} and t_{off}, i.e.:

$$q_n = \alpha \mathbb{1}_{[t_{\text{on}};t_{\text{off}}]}(n), \tag{8}$$

with $\mathbb{1}(\cdot)$ denoting the indicator function. As a consequence, the emission rate vector $\boldsymbol{q}$ is fully characterized by the three parameters: $\alpha \in \mathbb{R}^+$, $(t_{\text{on}}, t_{\text{off}}) \in \{1, \ldots, T_s\}$ with $t_{\text{on}} \leq t_{\text{off}}$. However, in practice, the emission rate α is assumed to be bounded such as $Q_{\min} \leq \alpha \leq Q_{\max}$. If nothing else is known a priori about these unknown parameters, the following uniform distribution for these parameters can be chosen as prior distribution:

$$\begin{aligned} p(\alpha) &= \mathcal{U}_{[Q_{\min};Q_{\max}]}(\alpha), \\ p(t_{\text{on}}, t_{\text{off}}) &\propto \mathbb{1}_{\{1,\ldots,T_s\}}(t_{\text{on}}) \times \mathbb{1}_{\{t_{\text{on}},\ldots,T_s\}}(t_{\text{off}}). \end{aligned} \tag{9}$$

By using such simple prior distributions, the unknown parameters that fully characterize the source is given by $\boldsymbol{\beta} = [\boldsymbol{x_s}, \alpha, t_{\text{on}}, t_{\text{off}}]$ and moreover there does not exist any unknown hyperparameters, so $\boldsymbol{\Psi}_p = \varnothing$.

2.4. Source term estimation in a Bayesian framework

In this work, a Bayesian solution is considered in order to solve efficiently the challenging STE problem. Instead of just a point-wise estimation of the source characteristics, $\boldsymbol{\beta}$, we are therefore interested in obtaining the full posterior distribution of the unknown parameters, $p(\boldsymbol{\beta}|\boldsymbol{y})$, which completely characterizes the available information on $\boldsymbol{\beta}$ given the measurements $\boldsymbol{y}$ obtained from all the sensors deployed in the field. With such a quantity, one can obtain all possible quantities of interest about the parameters such as, for example, point estimates and more importantly confidence intervals. Additionally in this problem, we want to jointly estimate the unknown parameters of both the prior and the likelihood probability functions, denoted respectively by $\boldsymbol{\Psi}_p$ and $\boldsymbol{\Psi}_l$. As a consequence, the complete Bayesian solution is to consider the posterior distribution on the extended space $(\boldsymbol{\beta}, \boldsymbol{\Psi}_l, \boldsymbol{\Psi}_p)$ which can be decomposed as follows

$$p(\boldsymbol{\beta}, \boldsymbol{\Psi}_l, \boldsymbol{\Psi}_p|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\Psi}_l)p(\boldsymbol{\beta}|\boldsymbol{\Psi}_p)p(\boldsymbol{\Psi}_p, \boldsymbol{\Psi}_l)}{p(\boldsymbol{y})}, \tag{10}$$

where $p(\boldsymbol{\Psi}_p, \boldsymbol{\Psi}_l)$ represents the prior distribution on the unknown parameters $(\boldsymbol{\Psi}_p, \boldsymbol{\Psi}_l)$.

Unfortunately, even with some particular choice for the likelihood and prior distributions, this joint posterior distribution of interest in (10) is analytically intractable. Indeed, the dependence of the position of the source in the measurements is highly nonlinear due to the complex structure of the source-receptor matrix $\boldsymbol{C}(\boldsymbol{x}_s)$ as discussed in Section 2.1. In this work, we will develop an efficient stochastic simulation based algorithm to approximate this complex posterior distribution $p(\boldsymbol{\beta}, \boldsymbol{\Psi}_l, \boldsymbol{\Psi}_p|\boldsymbol{y})$.

3. Proposed SMC Sampler to STE

In this section, we first present the general principle of the SMC sampler then we describe how to apply the SMC sampler to efficiently solve the STE problem.

3.1. General Idea of the SMC Sampler

The SMC Sampler methodology is a generic approach to approximate a sequence of probability distributions $\{\pi_t\}_{t=1}^{T}$ regarding some parameters of interest $\boldsymbol{\theta}$ and which are defined upon a common measurable space E (Del Moral et al., 2006), where the final distribution π_T is the distribution of interest. The main ideas of the SMC sampler can be summarized as:

a) Instead of sampling directly from the complex target posterior distribution, a sequence of intermediate $\{\pi_t\}_{t=1}^{T}$ is designed such that the transition from a simpler distribution, π_0 to the one of interest, π_T, is smooth.
b) The problem is then solved by transforming this problem in the standard SMC filtering framework, where the sequence of target distributions on the path-space denoted by $\{\tilde{\pi}_t\}_{t=1}^{T}$, which admits $\{\pi_t\}_{t=1}^{T}$ as marginals, is defined on the product space, i.e. $\text{supp}(\tilde{\pi}_t) = E \times \cdots \times E = E^t$. This novel sequence of target distributions $\tilde{\pi}_t$ is defined as follows:

$$\tilde{\pi}_t(\boldsymbol{\theta}_{1:t}) = \frac{\tilde{\gamma}_t(\boldsymbol{\theta}_{1:t})}{Z_t} \tag{11}$$

where

$$\tilde{\gamma}_t(\boldsymbol{\theta}_{1:t}) = \gamma_t(\boldsymbol{\theta}_t) \prod_{k=1}^{t-1} \mathcal{L}_k(\boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k) \tag{12}$$

in which the artificial kernels $\{\mathcal{L}_k\}_{k=1}^{t-1}$ are called *backward* Markov kernels since $\mathcal{L}_t(\boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}_t)$ denotes the probability density of moving back from $\boldsymbol{\theta}_{t+1}$ to $\boldsymbol{\theta}_t$. γ_t and $\tilde{\gamma}_t$ represent the unnormalized version of the target distribution π_t and $\tilde{\pi}_t$, respectively. Z_t thus corresponds to the normalizing constant of the target distribution $\tilde{\pi}_t$. By using such a sequence of extended target distributions $\{\tilde{\pi}_t\}_{t=1}^{T}$ based on the introduction of backward kernels $\{\mathcal{L}_k\}_{k=1}^{t-1}$, sequential importance sampling can be used in the same manner as standard SMC filtering algorithms.

Within this framework, one may then work with the constructed sequence of distributions, $\tilde{\pi}_t$, under the standard SMC algorithm (Doucet et al., 2001). In summary, the SMC sampler algorithm involves three stages:

1. *mutation*, where the particles are moved from $\boldsymbol{\theta}_{t-1}$ to $\boldsymbol{\theta}_t$ via a mutation kernel $\mathcal{K}_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$;
2. *correction*, where the particles are reweighted with respect to π_t via the incremental importance weight (Equation (16)); and
3. *selection*, where according to some measure of particle diversity, such as effective sample size, the weighted particles may be resampled in order to reduce the variability of the importance weights.

In more detail, suppose that at time $t-1$, we have a set of weighted particles $\left\{\boldsymbol{\theta}_{1:t-1}^{(m)}, \widetilde{W}_{t-1}^{(m)}\right\}_{m=1}^{N}$ that approximates $\tilde{\pi}_{t-1}$ via the empirical measure

$$\tilde{\pi}_{t-1}^{N}(d\boldsymbol{\theta}_{1:t-1}) = \sum_{m=1}^{N} \widetilde{W}_{t-1}^{(m)} \delta_{\boldsymbol{\theta}_{1:t-1}^{(m)}}(d\boldsymbol{\theta}_{1:t-1}) \tag{13}$$

where $\delta(\cdot)$ is the Dirac delta function. These particles are first propagated to the next distribution $\tilde{\pi}_t$ using a Markov kernel $\mathcal{K}_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$ to obtain the set of particles $\left\{\boldsymbol{\theta}_{1:t}^{(m)}\right\}_{m=1}^{N}$. IS is then used to correct for the discrepancy between the sampling distribution $\eta_t(\boldsymbol{\theta}_{1:t})$ defined as

$$\eta_t(\boldsymbol{\theta}_{1:t}^{(m)}) = \eta_1(\boldsymbol{\theta}_1^{(m)}) \prod_{k=2}^{t} \mathcal{K}_k(\boldsymbol{\theta}_{t-1}^{(m)}, \boldsymbol{\theta}_t^{(m)}) \tag{14}$$

and $\tilde{\pi}_t(\boldsymbol{\theta}_{1:t})$. In this case the new expression for the unnormalized importance weights is given by

$$W_t^{(m)} \propto \frac{\tilde{\pi}_t(\boldsymbol{\theta}_{1:t}^{(m)})}{\eta_t(\boldsymbol{\theta}_{1:t}^{(m)})} = \frac{\pi_t(\boldsymbol{\theta}_t^{(m)}) \prod_{s=1}^{t-1} \mathcal{L}_s(\boldsymbol{\theta}_{s+1}^{(m)}, \boldsymbol{\theta}_s^{(m)})}{\eta_1(\boldsymbol{\theta}_1^{(m)}) \prod_{k=2}^{t} \mathcal{K}_k(\boldsymbol{\theta}_{t-1}^{(m)}, \boldsymbol{\theta}_t^{(m)})} \propto w_t(\boldsymbol{\theta}_{t-1}^{(m)}, \boldsymbol{\theta}_t^{(m)}) W_{t-1}^{(m)} \tag{15}$$

where w_t, termed the (unnormalized) *incremental weights*, are calculated as,

$$w_t(\boldsymbol{\theta}_{t-1}^{(m)}, \boldsymbol{\theta}_t^{(m)}) = \frac{\gamma_t(\boldsymbol{\theta}_t^{(m)}) \mathcal{L}_{t-1}(\boldsymbol{\theta}_t^{(m)}, \boldsymbol{\theta}_{t-1}^{(m)})}{\gamma_{t-1}(\boldsymbol{\theta}_{t-1}^{(m)}) \mathcal{K}_t(\boldsymbol{\theta}_{t-1}^{(m)}, \boldsymbol{\theta}_t^{(m)})} \tag{16}$$

However, as in the particle filter, since the discrepancy between the target distribution $\tilde{\pi}_t$ and the proposal η_t increases with t, the variance of the unnormalized importance weights tends therefore to increase as well, leading to a degeneracy of the particle approximation. A common criterion used in practice to check this problem is the effective sample size $\mathbb{ESS}$, which is given by:

$$\mathbb{ESS}_t = \left[\sum_{m=1}^{N} (\widetilde{W}_t^{(m)})^2\right]^{-1} = \frac{\left(\sum_{m=1}^{N} \widetilde{W}_{t-1}^{(m)} w_t(\boldsymbol{\theta}_{t-1}^{(m)}, \boldsymbol{\theta}_t^{(m)})\right)^2}{\sum_{j=1}^{N} \left(\widetilde{W}_{t-1}^{(j)}\right)^2 \left(w_t(\boldsymbol{\theta}_{t-1}^{(j)}, \boldsymbol{\theta}_t^{(j)})\right)^2} \tag{17}$$

If the degeneracy is too high, i.e., the $\mathbb{ESS}_t$ is below a prespecified threshold, $\overline{\mathbb{ESS}}$, then a resampling step is performed. The particles with low weights are discarded whereas particles with high weights are duplicated. After resampling, the particles are equally weighted.

The final weighted particles at distribution π_T are considered weighted samples from the target distribution π of interest. As a consequence, the SMC sampler provides an estimate of this distribution

$$\pi_T^N(d\boldsymbol{\theta}) = \sum_{m=1}^{N} \widetilde{W}_T^{(m)} \delta_{\boldsymbol{\theta}_T^{(m)}}(d\boldsymbol{\theta}) \tag{18}$$

and the estimation of expectation of some function $\varphi(\cdot)$ with respect to the target distribution of interest is given by

$$\mathbb{E}_{\pi_T^N}\left[\varphi(\boldsymbol{\theta})\right] = \sum_{m=1}^{N} \widetilde{W}_T^{(m)} \varphi(\boldsymbol{\theta}_T^{(m)}) \tag{19}$$

3.2. Algorithm settings

As described in Section 2.4 - Eq. (10), the target distribution of interest in our case is defined as the posterior distribution over the unknown parameters, i.e.

$$\pi(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\boldsymbol{y}), \tag{20}$$

where all the parameters of interest that have to be jointly estimated is denoted by $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\Psi}_l, \boldsymbol{\Psi}_p\}$. The algorithm presented in the previous subsection is very general. There is a wide range of possible choices to consider when designing an SMC sampler algorithm, the appropriate sequence of distributions $\{\pi_t\}_{1\leq t\leq T}$, the choice of both the mutation kernel $\{\mathcal{K}_t\}_{2\leq t\leq T}$ and the backward mutation kernel $\{\mathcal{L}_{t-1}\}_{t=2}^{T}$ (for a given mutation kernels), see details in (Del Moral et al., 2006). In this subsection, we provide a discussion on how to choose these parameters of the algorithm for our estimation problem.

3.2.1. Sequence of distributions π_t

There are many potential choices for $\{\pi_t\}$ leading to various integration and optimization algorithms. As a special case, we can set $\pi_t = \pi$ for all $t \in \mathcal{N}$. Alternatively, to maximize $\pi(\boldsymbol{\theta})$, we could consider $\pi_t(\boldsymbol{\theta}_t) = [\pi(\boldsymbol{\theta}_t)]^{\xi_t}$ for an increasing schedule $\{\xi_t\}_{t\in\mathcal{N}}$ to ensure $\pi_T(\boldsymbol{\theta})$ is concentrated around the set of global maxima of $\pi(\boldsymbol{\theta})$. In the context of Bayesian inference for static parameters which is the main focus of this paper, we consider the *likelihood tempered* target sequence (Neal, 2001)

$$\pi_t(\boldsymbol{\theta}) \propto p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})^{\phi_t} \tag{21}$$

where $\{\phi_t\}$ is a non-decreasing temperature schedule with $\phi_0 = 0$ and $\phi_T = 1$. We thus sample initially from the prior distribution $\pi_0 = p(\boldsymbol{\theta})$ directly and introduce the effect of the likelihood gradually in order to obtain at the end $t = T$ an approximation of the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$. Tempering the likelihood could significantly improve the exploration of the state space in complex multimodal posterior distribution. The approximation of an expectation with respect to the posterior distribution of interest is therefore given by:

$$\mathbb{E}_{\pi^N}\left[\varphi(\boldsymbol{\theta})\right] = \sum_{i=1}^{N} \widetilde{W}_T^{(i)} \varphi(\boldsymbol{\theta}_T^{(i)}) \tag{22}$$

As discussed in (Nguyen et al., 2016), several statistical approaches have been proposed in order to automatically obtain such a tempering schedule via the optimization of some criteria, which are known as *on-line* schemes (Jasra et al., 2011; Zhou et al., 2016). In particular, Zhou et al. (2016) propose to set the temperature at the t-th iteration of the algorithm such that the *conditional* $\mathbb{ESS}$ ($\mathbb{CESS}$), decays in a regular predefined way. This $\mathbb{CESS}$ given by

$$\mathbb{CESS}_t = \frac{N\left(\sum_{i=1}^{N} \widetilde{W}_{t-1}^{(i)} w_t(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t^{(i)})\right)^2}{\sum_{j=1}^{N} \widetilde{W}_{t-1}^{(j)} \left(w_t(\boldsymbol{\theta}_{t-1}^{(j)}, \boldsymbol{\theta}_t^{(j)})\right)^2}. \tag{23}$$

is a slight modification of the $\mathbb{ESS}$ defined in Eq. (17). As shown in (Zhou et al., 2016), by characterizing how good an importance sampling proposal π_{t-1} would be for the estimation of expectation under π_t, the use of this quantity rather than the $\mathbb{ESS}$ leads to a reduction in estimator variance.

3.2.2. Sequence of mutation kernels $\mathcal{K}_t$

The performance of SMC samplers depends heavily upon the selection of the transition kernels $\{\mathcal{K}_t\}_{t=2}^{T}$ and the auxiliary backward kernels $\{\mathcal{L}_{t-1}\}_{t=2}^{T}$. There are many possible choices for $\mathcal{K}_t$ which have been discussed in (Del Moral et al., 2006). In this study, we propose to employ MCMC kernels of invariant distribution π_t for $\mathcal{K}_t$. This is an attractive strategy since we can use the vast literature on the design of efficient MCMC algorithms to build a good importance distributions (See (Robert and Casella, 2004)).

More precisely, since we are interested in complex models with potentially multimodal posterior distribution, a series Metropolis-within-Gibbs kernels with local moves (Robert and Casella, 2004) will be employed in order to successively move:

- the position of the source $\boldsymbol{x}_s$,
- the level of emission α,
- the time of emission $(t_{\text{on}}, t_{\text{off}})$,
- the hyperparameters of the likelihood distribution $\boldsymbol{\Psi}_l = \sigma_\epsilon^2$.
- and the hyperparameters of the prior distributions $\boldsymbol{\Psi}_p$, if any.

More specifically, after transforming the parameters σ_ϵ^2 and α to take value in the real line by using the $\log(\cdot)$ transform, an adaptive random walk proposal is used for each parameter except for the time of emission which values are discrete. This proposal consists, at the t-iteration of the SMC sampler, in adding a random perturbation, which follows generally a centered Normal distribution with covariance $\boldsymbol{\Sigma}_{:,t}^{\text{rw}}$ which is computed adaptively from the past simulations, to the current value of the parameter. As an example, for the position of the source, the proposed sample is obtained at the i-th move of the MCMC kernel of the t-th iteration of the SMC sampler as:

$$\boldsymbol{x}_{s,t}^{*} = \boldsymbol{x}_{s,t}^{i} + \boldsymbol{b}_{x_s,t} \tag{24}$$

in which $\boldsymbol{b}_{x_s,t}$ is a Gaussian random variable with zero mean and covariance matrix $\boldsymbol{\Sigma}_{x_s,t}^{\text{rw}}$. As with any sampling algorithm, faster mixing does not harm performance and in some cases will considerably improve it. In the particular case of Metropolis-Hastings kernels, the mixing speed relies on adequate proposal scales. As a consequence, we adopt the strategy proposed in (Jasra et al., 2011). The authors applied an idea used within adaptive MCMC methods (Andrieu and Moulines, 2006) to SMC samplers by using variance of parameters estimated from its particle system approximation as the proposal scale for the next iteration, i.e., the covariance matrix of the random-walk move for the

position of the source is thus given by:

$$\begin{aligned} \boldsymbol{\Sigma}^{\mathrm{rw}}_{x_s,t} &= \sum_{m=1}^{N} \widetilde{W}^{(m)}_{t-1} \left(\boldsymbol{x}^{(m)}_{s,t-1} - \boldsymbol{\mu}_{x_s,t-1}\right)\left(\boldsymbol{x}^{(m)}_{s,t-1} - \boldsymbol{\mu}_{x_s,t-1}\right)^T \\ \text{with} \qquad \boldsymbol{\mu}_{x_s,t-1} &= \sum_{m=1}^{N} \widetilde{W}^{(m)}_{t-1} \boldsymbol{x}^{(m)}_{s,t-1} \end{aligned} \tag{25}$$

The motivation is that if π_{t-1} is close to π_t (which is recommended for an efficient SMC sampler algorithm), then the variance estimated at iteration $t-1$ will provide a sensible scaling at time t. In difficult problems such as the one addressed in this paper, other approaches could be added in order to have appropriate scaling adaptation; one approach demonstrated in (Jasra et al., 2011) is to simply employ a pair of acceptance rate thresholds and to alter the proposal scale from the simply estimated value whenever the acceptance rate falls outside those threshold values. This scheme is to ensure that the acceptance rates in the Metropolis-Hastings steps did not get too large or small. Through all this work, we use this procedure which consists for example to multiply the covariance matrix by 5 (resp. 1/5) if the rate exceeded 0.7 (resp. fell below 0.2). The same procedure is used respectively to sample respectively the log transform of the level of emission $\log(\alpha)$ and of the likelihood variance $\log(\sigma^2_\epsilon)$.

Finally, for the time of emission, the proposal distribution used is defined as:

$$q(t_{\mathrm{on}}, t_{\mathrm{off}} | t^i_{\mathrm{on},t}, t^i_{\mathrm{off},t}) = q(t_{\mathrm{on}} | t^i_{\mathrm{on},t}) q(t_{\mathrm{off}} | t_{\mathrm{on}}, t^i_{\mathrm{off},t}), \tag{26}$$

where

$$\begin{aligned} q(t_{\mathrm{on}} | t^i_{\mathrm{on},t}) &\propto \mathbb{1}_{\{\max(1, t^i_{\mathrm{on},t} - \Delta^{\mathrm{rw}}_{t,\mathrm{on}}), \ldots, \min(T_s, t^i_{\mathrm{on},t} + \Delta^{\mathrm{rw}}_{t,\mathrm{on}})\}}(t_{\mathrm{on}}), \\ q(t_{\mathrm{off}} | t_{\mathrm{on}}, t^i_{\mathrm{off},t}) &\propto \mathbb{1}_{\{\max(t_{\mathrm{on}}, t^i_{\mathrm{off},t} - \Delta^{\mathrm{rw}}_{t,\mathrm{off}}), \ldots, \min(T_s, t^i_{\mathrm{off},t} + \Delta^{\mathrm{rw}}_{t,\mathrm{off}})\}}(t_{\mathrm{off}}). \end{aligned} \tag{27}$$

where $\Delta^{\mathrm{rw}}_{t,\mathrm{on}}$ and $\Delta^{\mathrm{rw}}_{t,\mathrm{off}}$ are adapted using the same strategy as previously described for the other proposals, i.e. for $\Delta^{\mathrm{rw}}_{t,\mathrm{on}}$:

$$\begin{aligned} \Delta^{\mathrm{rw}}_{t,\mathrm{on}} &= \left[\sum_{m=1}^{N} \widetilde{W}^{(m)}_{t-1} \left(t^{(m)}_{\mathrm{on},t-1} - \mu_{t_{\mathrm{on}},t-1}\right)^2\right]^{\frac{1}{2}} \\ \text{with} \qquad \mu_{t_{\mathrm{on}},t-1} &= \sum_{m=1}^{N} \widetilde{W}^{(m)}_{t-1} t^{(m)}_{\mathrm{on},t-1} \end{aligned} \tag{28}$$

$\Delta^{\mathrm{rw}}_{t,\mathrm{off}}$ is obtained by just replacing $t^{(m)}_{\mathrm{on},t-1}$ by $t^{(m)}_{\mathrm{off},t-1}$. This adaptive Metropolis within Gibbs used in the implementation of the SMC sampler through this work is summarized in Algorithm 2.

3.2.3. Sequence of backward kernels $\mathcal{L}_t$

The backward kernel $\mathcal{L}_t$ is arbitrary, however as discussed in (Del Moral et al., 2006), it should be optimized with respect to mutation kernel $\mathcal{K}_t$ to obtain good performance. (Del Moral et al., 2006) establish that the backward kernel which minimizes the variance of the unnormalized importance weights, W_t, are given by

$$\mathcal{L}^{\mathrm{opt}}_t(\boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}_t) = \frac{\eta_t(\boldsymbol{\theta}_t)\mathcal{K}_{t+1}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1})}{\eta_{t+1}(\boldsymbol{\theta}_{t+1})} \tag{29}$$

However, it is typically impossible to use these optimal kernels as they rely on the marginalization of the joint distribution defined in Eq. (14) which do not admit any closed form expression, especially if an MCMC kernel is used as $\mathcal{K}_t$ which is π_t-invariant distribution. Thus we can either choose to approximate $\mathcal{L}_t^{\text{opt}}$ or choose kernels $\mathcal{L}_t$ so that the importance weights are easily calculated or have a familiar form. As discussed in (Del Moral et al., 2006), if an MCMC kernel is used as forward kernel, the following $\mathcal{L}_t$ is employed

$$\mathcal{L}_{t-1}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}) = \frac{\pi_t(\boldsymbol{\theta}_{t-1})\mathcal{K}_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)}{\pi_t(\boldsymbol{\theta}_t)} \tag{30}$$

which is a good approximation of the optimal backward if the discrepancy between π_t and π_{t-1} is small; note that (30) is the reversal Markov kernel associated with $\mathcal{K}_t$. In this case, the unnormalized incremental weights becomes

$$w_t^{(m)}(\boldsymbol{\theta}_{t-1}^{(m)}, \boldsymbol{\theta}_t^{(m)}) = \frac{\gamma_t(\boldsymbol{\theta}_{t-1}^{(m)})}{\gamma_{t-1}(\boldsymbol{\theta}_{t-1}^{(m)})} = p(\boldsymbol{y}|\boldsymbol{\theta}_{t-1}^{(m)})^{(\phi_t - \phi_{t-1})} \tag{31}$$

This expression (31) is remarkably easy to compute and valid regardless of the MCMC kernel adopted. Note that $\phi_t - \phi_{t-1}$ is the step length of the cooling schedule of the likelihood at time t. As we choose this step larger, the discrepancy between π_t and π_{t-1} increases, leading to increase the variance of the importance approximation when it deteriorates. Thus, it is important to construct a smooth sequence of distributions $\{\pi_t\}_{0 \leq t \leq T}$ by judicious choice of an associated real sequence $\{\phi_t\}_{t=0}^T$.

Let us remark that when such backward kernel is used, the unnormalized incremental weights in Eq. (31) at time t does not depend on the particle value at time t but just on the previous particle set. As suggested in (Del Moral et al., 2006), in such case, the particles $\left\{\boldsymbol{\theta}_t^{(m)}\right\}$ should be sampled after the weights $\left\{W_t^{(m)}\right\}$ have been computed and after the particle approximation $\left\{W_t^{(m)}, \boldsymbol{\theta}_{t-1}^{(m)}\right\}$ has possibly been resampled.

Based on this discussion regarding the different choices, the SMC sampler that will be used for Bayesian inference in the STE model is summarized in Algorithm 1.

Algorithm 1 SMC Sampler Algorithm for STE

1: Initialize particle system

2: Sample $\left\{\boldsymbol{\theta}_1^{(m)}\right\}_{m=1}^{N} \sim \eta_1(\cdot)$ and compute $\widetilde{W}_1^{(m)} = \left(\frac{\gamma_1(\boldsymbol{\theta}_1^{(m)})}{\eta_1(\boldsymbol{\theta}_1^{(m)})}\right)\left[\sum_{j=1}^{N}\frac{\gamma_1(\boldsymbol{\theta}_1^{(j)})}{\eta_1(\boldsymbol{\theta}_1^{(j)})}\right]^{-1}$ and do resampling if $\mathbb{ESS} < \overline{\mathbb{ESS}}$

3: **for** $t = 2, \ldots, T$ **do**

4: Set the temperature schedule such that the CESS defined in Eq. (23) decays in a regular predefined way.

5: Computation of the weights: for each $m = 1, \ldots, N$

$$W_t^{(m)} = \widetilde{W}_{t-1}^{(m)} p(\boldsymbol{y}|\boldsymbol{\theta}_{t-1}^{(m)})^{(\phi_t - \phi_{t-1})}$$

Normalization of the weights : $\widetilde{W}_t^{(m)} = W_t^{(m)}\left[\sum_{j=1}^{N} W_t^{(j)}\right]^{-1}$

6: Selection: if $ESS < \overline{\mathbb{ESS}}$ then Resample

7: Mutation: for each $m = 1, \ldots, N$: Sample $\boldsymbol{\theta}_t^{(m)} \sim \mathcal{K}_t(\boldsymbol{\theta}_{t-1}^{(m)}; \cdot)$ where $\mathcal{K}_t(\cdot;\cdot)$ is a $\pi_t(\cdot)$ invariant Markov kernel described in more details in Algo. 2.

8: **end for**

Algorithm 2 Adaptive Metropolis-within-Gibbs Kernel $\mathcal{K}_t(\cdot;\cdot)$ for the m-th particle

Input: Set $\boldsymbol{\theta}^0 = \left\{\boldsymbol{x}_s^0, \log(\alpha)^0, (t_{\text{on}}^0, t_{\text{off}}^0), \log(\sigma_\epsilon^2)^0\right\} = \boldsymbol{\theta}_{t-1}^{(m)}$
for $i = 1, \ldots, N_{\text{MCMC}}$ **do**
  MHwGibbs for the position of the source:
  Sample $\boldsymbol{x}_s^* \sim \mathcal{N}\left(\boldsymbol{x}_s^{i-1}, \boldsymbol{\Sigma}_{x_s,t}^{\text{rw}}\right)$ with $\boldsymbol{\Sigma}_{x_s,t}^{\text{rw}}$ defined in Eq. 25
  Compute the Acceptance ratio:

$$\rho_x = \min\left\{1, \frac{p(\boldsymbol{y}|\boldsymbol{x}_s^*, \boldsymbol{\theta}_{-\boldsymbol{x}_s}^{i-1})^{\phi_t} p(\boldsymbol{x}_s^*)}{p(\boldsymbol{y}|\boldsymbol{x}_s^{i-1}, \boldsymbol{\theta}_{-\boldsymbol{x}_s}^{i-1})^{\phi_t} p(\boldsymbol{x}_s^{i-1})}\right\}$$

  with $\boldsymbol{\theta}_{-\boldsymbol{x}_s}^{i-1} = \{\log(\alpha)^{i-1}, (t_{\text{on}}^{i-1}, t_{\text{off}}^{i-1}), \log(\sigma_\epsilon^2)^{i-1}\}$ corresponding to the set containing all the unknown parameters of interest excepting $\boldsymbol{x}_s$.
  Set $\boldsymbol{x}_s^i = \boldsymbol{x}_s^*$ with probability ρ_x, otherwise set $\boldsymbol{x}_s^i = \boldsymbol{x}_s^{i-1}$
  MHwGibbs for the level of emission:
  Sample $\log(\alpha)^* \sim \mathcal{N}\left(\log(\alpha)^{i-1}, \boldsymbol{\Sigma}_{\alpha,t}^{\text{rw}}\right)$ by following the same procedure as for $\boldsymbol{x}_s$
  Compute the Acceptance ratio:

$$\rho_\alpha = \min\left\{1, \frac{p(\boldsymbol{y}|\log(\alpha)^*, \boldsymbol{\theta}_{-\alpha}^{i-1})^{\phi_t} p(\log(\alpha)^*)}{p(\boldsymbol{y}|\log(\alpha)^{i-1}, \boldsymbol{\theta}_{-\alpha}^{i-1})^{\phi_t} p(\log(\alpha)^{i-1})}\right\}$$

  with $\boldsymbol{\theta}_{-\alpha}^{i-1} = \{\boldsymbol{x}_s^i, (t_{\text{on}}^{i-1}, t_{\text{off}}^{i-1}), \log(\sigma_\epsilon^2)^{i-1}\}$
  Set $\log(\alpha)^i = \log(\alpha)^*$ with probability ρ_α, otherwise set $\log(\alpha)^i = \log(\alpha)^{i-1}$
  MHwGibbs for the time of emission:
  Sample $(t_{\text{on}}^*, t_{\text{off}}^*) \sim q(t_{\text{on}}, t_{\text{off}}|t_{\text{on}}^{i-1}, t_{\text{off}}^{i-1})$ defined in Eq. 26
  Compute the Acceptance ratio:

$$\rho_t = \min\left\{1, \frac{p(\boldsymbol{y}|t_{\text{on}}^*, t_{\text{off}}^*, \boldsymbol{\theta}_{-t}^{i-1})^{\phi_t} p(t_{\text{on}}^*, t_{\text{off}}^*) q(t_{\text{on}}^{i-1}, t_{\text{off}}^{i-1}|t_{\text{on}}^*, t_{\text{off}}^*)}{q(t_{\text{on}}^*, t_{\text{off}}^*|t_{\text{on}}^{i-1}, t_{\text{off}}^{i-1}) p(\boldsymbol{y}|\boldsymbol{x}_s^{i-1}, \boldsymbol{\theta}_{-t}^{i-1})^{\phi_t} p(t_{\text{on}}^{i-1}, t_{\text{off}}^{i-1})}\right\}$$

  with $\boldsymbol{\theta}_{-t}^{i-1} = \{\boldsymbol{x}_s^i, \log(\alpha)^i, \log(\sigma_\epsilon^2)^{i-1}\}$
  Set $(t_{\text{on}}^i, t_{\text{off}}^i) = (t_{\text{on}}^*, t_{\text{off}}^*)$ with probability ρ_t, otherwise set $(t_{\text{on}}^i, t_{\text{off}}^i) = (t_{\text{on}}^{i-1}, t_{\text{off}}^{i-1})$
  MHwGibbs for the variance of measurements' noise:
  Sample $\log(\sigma_\epsilon^2)^* \sim \mathcal{N}\left(\log(\sigma_\epsilon^2)^{i-1}, \boldsymbol{\Sigma}_{\sigma_\epsilon^2,t}^{\text{rw}}\right)$ by following the same procedure as for $\boldsymbol{x}_s$
  Compute the Acceptance ratio:

$$\rho_{\sigma_\epsilon^2} = \min\left\{1, \frac{p(\boldsymbol{y}|\log(\sigma_\epsilon^2)^*, \boldsymbol{\theta}_{-\sigma_\epsilon^2}^{i-1})^{\phi_t} p(\log(\sigma_\epsilon^2)^*)}{p(\boldsymbol{y}|\log(\sigma_\epsilon^2)^{i-1}, \boldsymbol{\theta}_{-\sigma_\epsilon^2}^{i-1})^{\phi_t} p(\log(\sigma_\epsilon^2)^{i-1})}\right\}$$

  with $\boldsymbol{\theta}_{-\sigma_\epsilon^2}^{i-1} = \{\boldsymbol{x}_s^i, \log(\alpha)^i, (t_{\text{on}}^i, t_{\text{off}}^i)\}$
  Set $\log(\sigma_\epsilon^2)^i = \log(\sigma_\epsilon^2)^*$ with probability $\rho_{\sigma_\epsilon^2}$, otherwise set $\log(\sigma_\epsilon^2)^i = \log(\sigma_\epsilon^2)^{i-1}$
end for
Set the new particle value at time t as
$\boldsymbol{\theta}_t^{(m)} = \left\{\boldsymbol{x}_s^{N_{\text{MCMC}}}, \log(\alpha)^{N_{\text{MCMC}}}, (t_{\text{on}}^{N_{\text{MCMC}}}, t_{\text{off}}^{N_{\text{MCMC}}}), \log(\sigma_\epsilon^2)^{N_{\text{MCMC}}}\right\}$

4. Numerical Simulations

The performances of the source term estimate method presented in this paper are assessed using twin experiments assuming a fictitious malevolent release in a complex atmospheric environment. The exact nature of the release is not at stakes for this hypothetical event. It could be constituted of chemical products, biological pathogenic agents, radioactive-nuclear materials or explosives (CBRN-E) In order to extend the study done by Septier et al. (2020), we consider the same actual urban area corresponding to the district of the Opera in Paris (France), where all buildings are explicitely accounted for. The 3D simulation domain has dimensions of 1.1km $\times$ 0.9km $\times$ 1.6km. It is meshed at a horizontal regular resolution of 2 meters and a vertical resolution of 2 meters near the ground decreasing with the elevation above the ground. Figure 1a depicts the simulated source and 20 virtual sensors set up in the urban district, which are represented respectively by a green asterisk and red crosses. The location of the source is chosen here as an example and it could be anywhere in the simulation domain. The sensors are displayed quite regularly in the street network. For the sake of simplicity, both the simulated source and sensors are assumed to be located at the same level close to the ground, immersed in the urban canopy. Even if the source term localization is a two-dimensional problem in these twin experiments, the flow and dispersion simulations are performed in three-dimensional space across the time dimension. The meteorological conditions considered during and after the fake noxious release correspond to a real weather sequence with the wind initially blowing from the west-northwest, then gradually rotating and, finally, blowing from north-northeast. Dispersion simulations are carried out with the model defined in Eq. (2) in backward mode from each virtual sensor emitting unit releases each minute over a 45-minute period. Owing to the reciprocity between direct and retrograde atmospheric transport, the backward computations allow us to obtain the source-receptor matrix $\boldsymbol{C}(\cdot)$ in Eq. (3). As discussed in Section 2.1, backward atmospheric transport has the advantage of drastically reducing the number of computations necessary to obtain this matrix as there are many less virtual sensors than potential ($N \times N_{\text{MCMC}} \times T$) positions of the emitting source generated by the SMC sampler. More precisely, using the dispersion model in backward mode, $N_c \times T_c$ computations are needed where N_c is the number of sensors and T_c the number of time samples collected by each sensor. In the test-case presented in the paper, N_c and T_c are equal respectively to 20 and 45. One can notice that in a practical situation, these numbers would even be lower. For example, results are given for the source represented by the green asterisk in Figure 1a. The simulated concentration values on all sensors affected by the plume are shown as a function of time in Figure 1b with a different color for each sensor. The concentration histories with additional noise are assimilated to pseudo-measurements, which in turn are used by the source term estimate method. For example, realizations of the noisy concentration pseudo-measurements are presented in Figure (1c-1d), where the clipped normal distribution, defined in Eq. (5), is used with two different values of the variance for all types of errors (measurements, model, etc) denoted by σ_ϵ^2. The different colors represent different sensors.

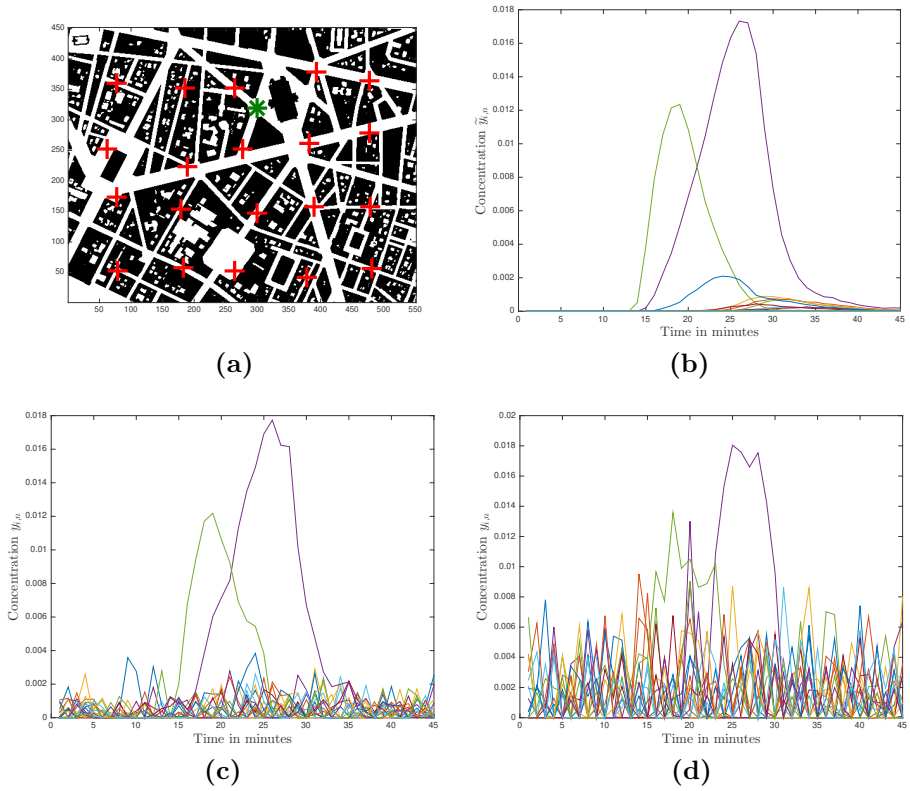
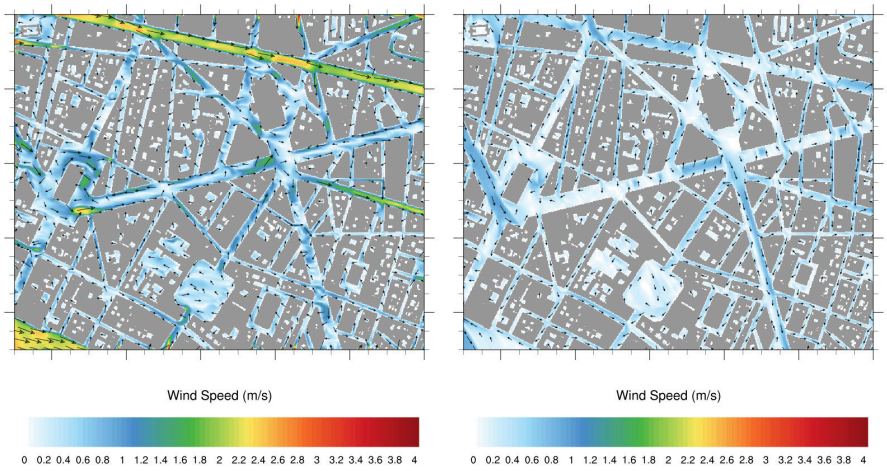
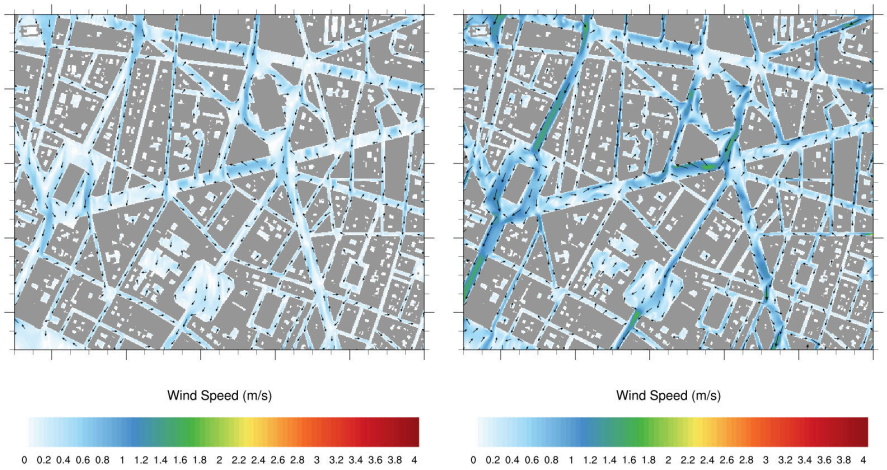


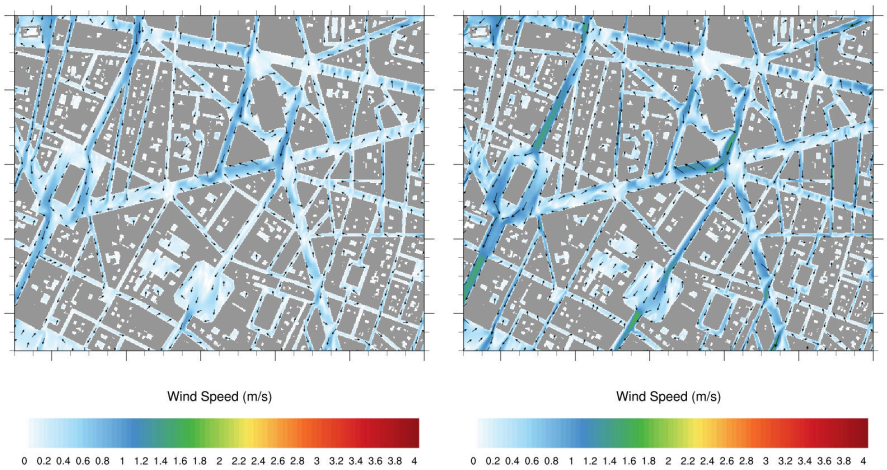
Figure 1: Scenario under study – (a) 20 sensors (red) & 1 source (green). (b) Output concentration levels, $\widetilde{y}_{i,n}$ at the sensor location from the dispersion model obtained every minute by the 20 sensors from 9:00 to 9:45. A realization of the noisy measurements, $y_{i,n}$ obtained using the clipped normal likelihood defined in Eq. (5) with $\sigma_\epsilon^2 = 10^{-6}$ (c) and $\sigma_\epsilon^2 = 10^{-5}$ (d).



(a) 2:00 p.m. **(b) 2:09 p.m.**



(c) 2:18 p.m. **(d) 2:27 p.m.**



(e) 2:36 p.m. **(f) 2:45 p.m.**

Figure 2: Wind Velocity fields of the considered scenario at different time instants and 2 meters from the ground.

In this study, the 3D flow illustrated and dispersion simulations have been performed with the Parallel-Micro-SWIFT-SPRAY (PMSS) system presented in (Oldrini et al., 2017) and (Oldrini et al., 2019). PMSS is the efficient parallel version in time, space, and Lagrangian particles of Micro-SWIFT-SPRAY (MSS), which was originally developed to provide simplified CFD solutions for the flow and dispersion, respectively with SWIFT and SPRAY, in built-up environments in a limited amount of time (Tinarelli et al., 2013). On the one hand, SWIFT is a 3D terrain-following mass-consistent diagnostic model taking account of the buildings and providing the 3D fields of wind, turbulence, temperature, and humidity. The resulting velocity fields of the wind used in all the simulations are depicted in Fig. 2. On the other hand, SPRAY is a 3D Lagrangian Particle Dispersion Model able to account for the presence of buildings. PMSS has been systematically validated against numerous wind tunnel and in-field experimental campaigns for short and prolonged releases (Trini Castelli et al., 2018; Oldrini and Armand, 2019). Besides the compliance of PMSS results with the statistical acceptance criteria of dispersion models in built-up environments, SPRAY dispersion model can be run in both direct and retrograde mode what is a great advantage to evaluate the source-receptor matrix $\boldsymbol{C}(\cdot)$ in Eq. (3).

The performances of the proposed SMC sampler are assessed with $N = 200$ particles and $N_{\text{MCMC}} = 10$ MCMC moves within the Markov kernel $\mathcal{K}_t(\cdot;\cdot)$. Figures 3 and 4 give the results obtained by the proposed SMC sampler for a source (represented respectively by a green asterisk in Fig. 1a) located at (300;320) with the variance of the model and measurements errors set to $\sigma_\epsilon^2 = 10^{-6}$ and $\sigma_\epsilon^2 = 10^{-5}$, respectively. Firstly from Figs. 3a and 4a, we can remark that in both cases, the position of the source is correctly estimated since the mean of the posterior distribution of the source position $p(x_s, y_s|\boldsymbol{y})$ approximated by the algorithm is close to the ground truth (≤ 10 meters for the noisiest scenario in Fig. 4a). As expected, the uncertainty given by the empirical measure of the posterior distribution for both $p(x_s|\boldsymbol{y})$ and $p(y_s|\boldsymbol{y})$ is higher when the errors are larger (i.e. $\sigma_\epsilon^2 = 10^{-5}$). The remarks holds regarding the estimation of the posterior distributions of σ_ϵ^2 and $\boldsymbol{q}$ depicted in Figs (3b,4b) and (3c,4c), respectively. Let us remark that the SMC sampler learns only from the the measurements that the errors are larger in Fig. 4 than in Fig. 3 since the variance of the noise is unknown and estimated by the algorithm quite accurately. Having an estimation of such posterior distributions for all the unknown parameters is of high interest for users in order to be able to quantify the associated uncertainty which is not possible when an optimization technique is used to solve this STE problem. Moreover, the resulting posterior distributions can be easily obtained without requiring a complex tuning of parameters by a non-expert user thanks to the proposed complete automatic and adaptive Bayesian solution.

In Figure 5, the proposed SMC sampler is compared to a more traditional MCMC algorithm. The MCMC approach is using the same Adaptive Metropolis-within-Gibbs Kernel described in Algo. 2 as in the proposed SMC sampler. The MCMC estimator obtained to compute the source position is obtained by taking the last 60 % of the simulated Markov chain values (the first samples are considered to be the burn-in period). Concerning the SMC sampler, the estimator is based only on the last population of $N = 200$ particles generated at a given iteration. The median of the Squared Error (SE) has been obtained by running 50 independent runs of both algorithms. From the results, the SMC

sampler significantly outperforms the corresponding MCMC algorithm based on the same MCMC kernel. Firstly, the SMC sampler has the benefit of using a population of N samples compared to the MCMC in which the samples are generated successively. Additionally, the use of likelihood tempered target sequence clearly facilitates the exploration of the state-space and thus limits the problem of having the inference algorithm to get stuck in one region of the space, which could easily happen in our problem due to the presence of many buildings in the surveillance area. Over the 50 trials, the squared errors of the estimates are lower than 100 (which means that the estimate of the position of the source lies in a circle of 10 meters around the ground truth) only 10 times for the MCMC algorithm but 50 times for the proposed SMC sampler. These results clearly highlight the benefit of using the SMC sampler in this complex environment. This fact also confirms in our context the remark mentioned in Jasra et al. (2011) stating that the SMC samplers, when using a given MCMC kernel, often out-perform the corresponding MCMC algorithm.

Figures 6 and 7 show the results obtained using the SMC sampler for a source located at (300;320) but with two different emission profile, a continuous and a prompt release, which may be due to an insidious long emission of a deleterious substance and an explosion, respectively. In both scenarios, the algorithm is able to provide accurate estimates for all the parameters of interest. By comparing the two scenarios, the results empirically show that the level of posterior uncertainties in the parameters decreases with the length of duration of the release. The larger amount of pollutant in the atmosphere in the case of Fig. 6 due to a longer release allows us to have more useful information from the sensor measurements and therefore more confidence about the parameters of interest.

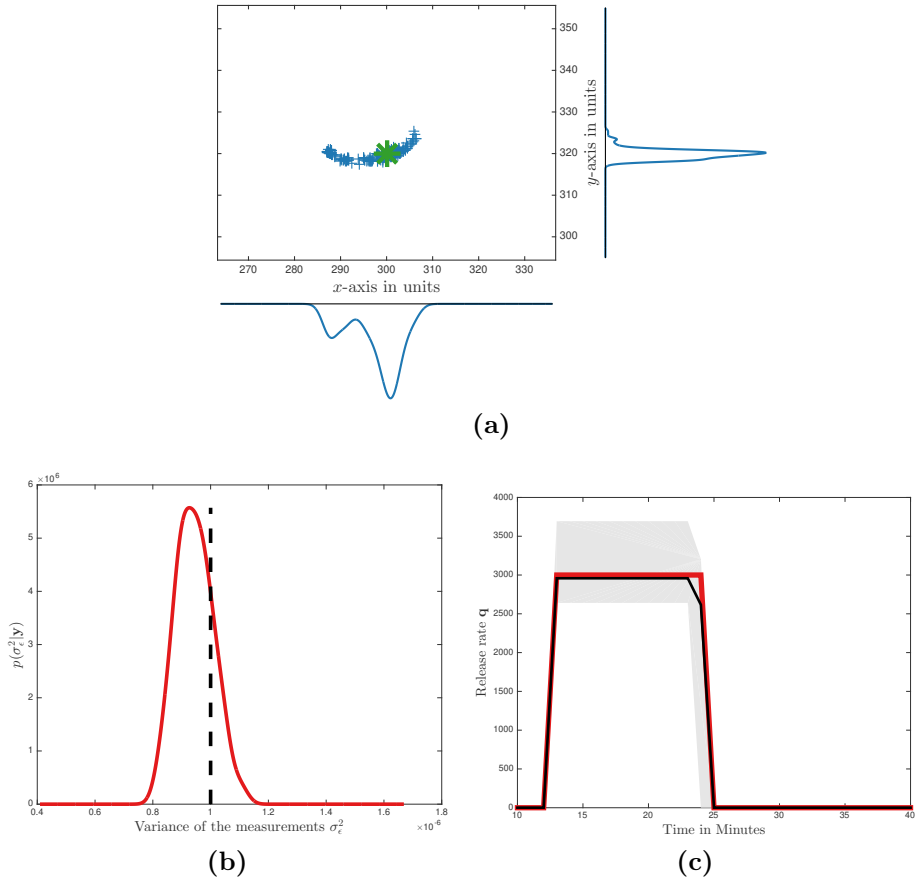


Figure 3: Results of the SMC sampler for STE with a release at location (300;320) and noise variance $\sigma_\epsilon^2 = 10^{-6}$ in Eq. (5). (a) Representation of the $N = 200$ particles regarding the source position (blue) along with the ground truth (green) and the estimation of $p(x_s|\boldsymbol{y})/p(y_s|\boldsymbol{y})$ on the bottom/right. (b) Estimation in red of $p(\sigma_\epsilon^2|\boldsymbol{y})$ and the true value in dashed black. (c): Mean of $p(\boldsymbol{q}|\boldsymbol{y})$ (black) and the 90% confidence interval (grey) compared to the ground truth (red).

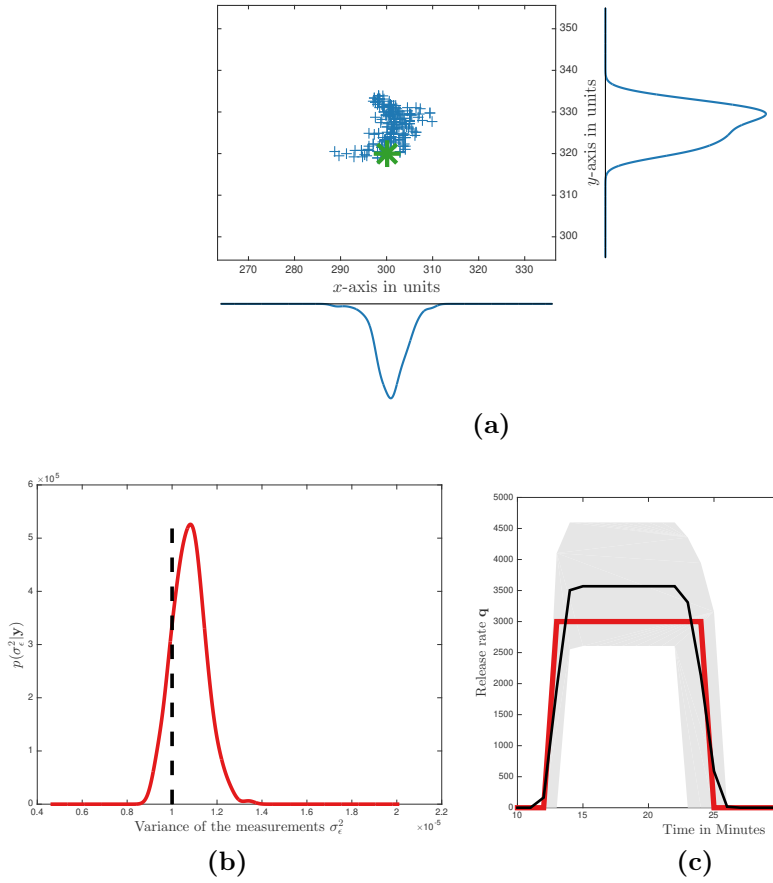


Figure 4: Results of the SMC sampler for STE with a release at location (300;320) and noise variance $\sigma_\epsilon^2 = 10^{-5}$ in Eq. (5). (a) Representation of the $N = 200$ particles regarding the source position (blue) along with the ground truth (green) and the estimation of $p(x_s|\boldsymbol{y})/p(y_s|\boldsymbol{y})$ on the bottom/right. (b) Estimation in red of $p(\sigma_\epsilon^2|\boldsymbol{y})$ and the true value in dashed black. (c): Mean of $p(\boldsymbol{q}|\boldsymbol{y})$ (black) and the 90% confidence interval (grey) compared to the ground truth (red).

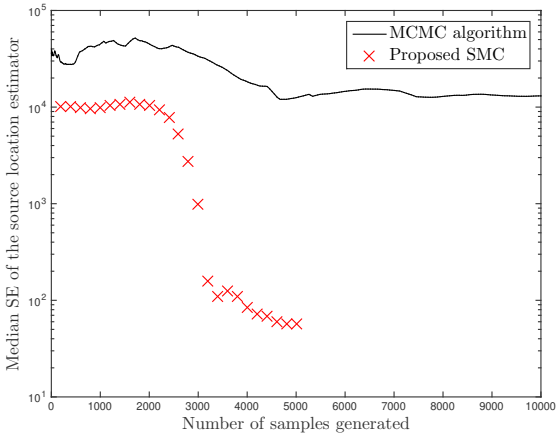


Figure 5: Median of the squared error between the ground truth source position (300;320) and the estimator provided by the proposed SMC sampler and an MCMC algorithm as a function of the number of samples that have been generated in the algorithm. The parameters of the source term are similar to the ones of Figure 4.

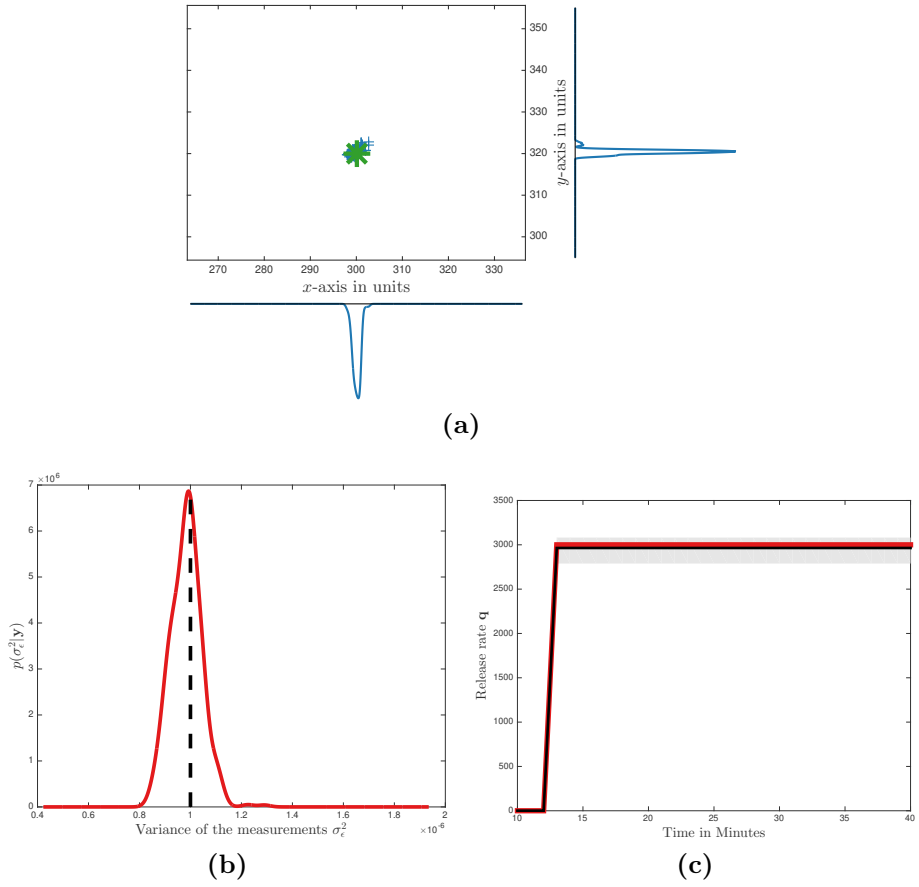


Figure 6: Results of the SMC sampler for STE with a continuous release [variance of the observation uncertainty $\sigma_\epsilon^2 = 10^{-6}$ in Eq. (5)]. (a) Representation of the $N = 200$ particles regarding the source position (blue) along with the ground truth (green) and the estimation of $p(x_s|\boldsymbol{y})/p(y_s|\boldsymbol{y})$ on the bottom/right (b) Estimation in red of (a) $p(\sigma_\epsilon^2|\boldsymbol{y})$ and the true value in dashed black. (c): Mean of $p(\boldsymbol{q}|\boldsymbol{y})$ (black) and the 90% confidence interval (grey) compared to the ground truth (red).

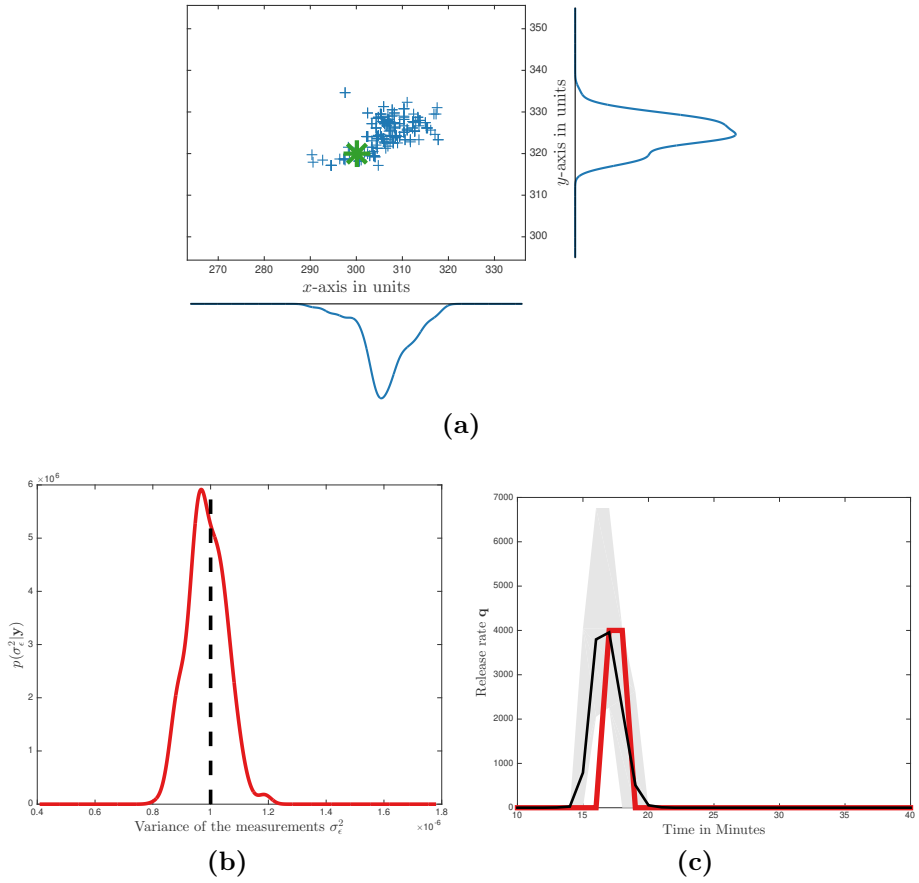


Figure 7: Results of the SMC sampler for STE with a prompt release [variance of the observation uncertainty $\sigma_\epsilon^2 = 10^{-6}$ in Eq. (5)]. (a) Representation of the $N = 200$ particles regarding the source position (blue) along with the ground truth (green) and the estimation of $p(x_s|\boldsymbol{y})/p(y_s|\boldsymbol{y})$ on the bottom/right (b) Estimation in red of (a) $p(\sigma_\epsilon^2|\boldsymbol{y})$ and the true value in dashed black. (c): Mean of $p(\boldsymbol{q}|\boldsymbol{y})$ (black) and the 90% confidence interval (grey) compared to the ground truth (red).

Finally, Fig. (8) shows the output of a single run of the proposed SMC sampler in the case of source located at (410;80). As can be seen in Fig (8a), it is difficult to distinguish any smooth useful output concentration levels as in Fig. 1b from the measurements due to the relatively large level of noise. Nevertheless, even in this challenging the STE problem, the algorithm is still able to accurately estimate the source location as well as both the release rate and the level of observation noise.

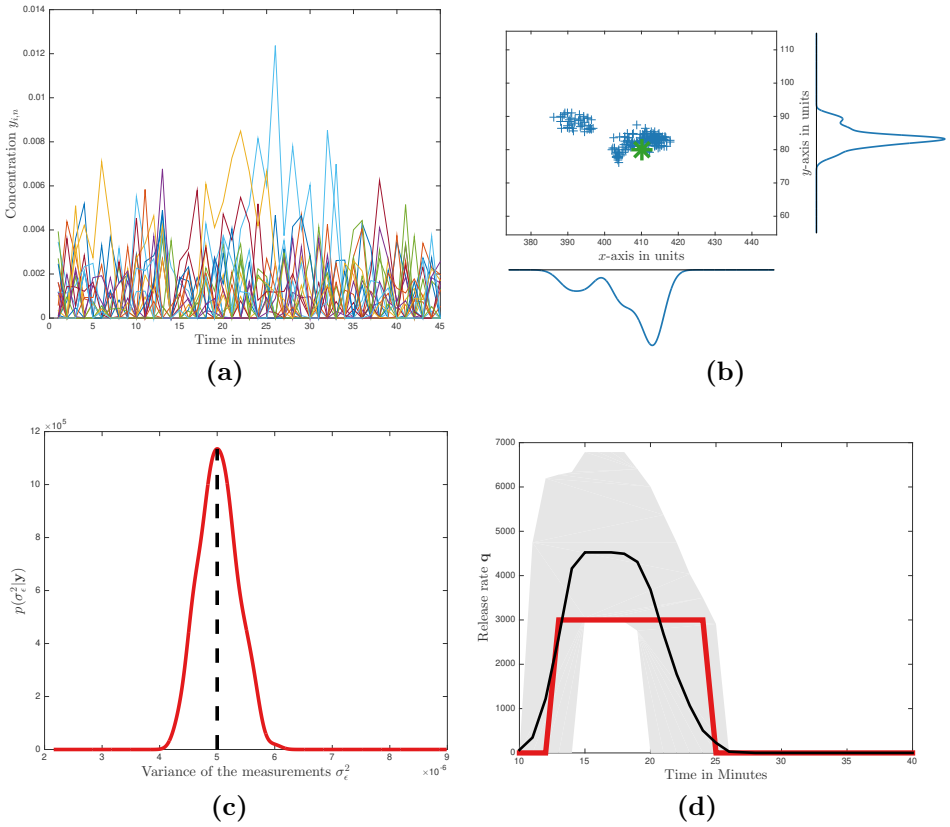


Figure 8: Results of the SMC sampler for STE with a release at another location (410;80). (a) Measurements obtained with a variance of $\sigma_\epsilon^2 = 5.10^{-5}$ in Eq. (5)]. (b) Representation of the $N = 200$ particles regarding the source position (blue) along with the ground truth (green) and the estimation of $p(x_s|\boldsymbol{y})/p(y_s|\boldsymbol{y})$ on the bottom/right. (c) Estimation in red of $p(\sigma_\epsilon^2|\boldsymbol{y})$ and the true value in dashed black. (d): Mean of $p(\boldsymbol{q}|\boldsymbol{y})$ (black) and the 90% confidence interval (grey) compared to the ground truth (red).

5. Conclusion

In this paper, a complete automatic Bayesian solution is proposed to solve the source term estimation problem. In a previous study (Septier et al., 2020), an adaptive Bayesian solution based on importance sampling was proposed. Unfortunately despite its performance, the use of such algorithm was conditioned on quite severe and not realistic assumptions regarding the model. Indeed in the previous work, both the likelihood and the prior function of the emission rate vector had to be assumed to be normal distributions, thus leading to the possibility of having negative values for some parameters such as the emission rate. Moreover, the normal prior distribution necessitated to be tuned quite finely by the user which could clearly be difficult in practice.

In this paper, we provide a novel framework that overcomes these limitations. More realistic and more general distributions of all physical quantities and parameters can therefore be used and moreover the model parameters are now assumed to be completely unknown and are included in the estimation pro-

cedure. Compared to previous algorithms, the proposed strategy provides a full probabilistic solution without having to choose, sometimes arbitrarily, values for certain model parameters, which therefore makes its use easier for practitioners.

The proposed solution is based on recent advances in the field of stochastic simulation algorithms and more precisely on the Sequential Monte Carlo sampler. Although this approach presents many advantages over traditional Monte-Carlo methods, the potential of this emergent technique is still however largely underexploited in practice. Moreover, the SMC sampler has been efficiently designed for this STE problem by integrating some adaptive schemes in order to use a sequence of intermediate distributions using appropriate cooling temperatures and also to explore the state in the MCMC moves using adaptive covariance for each parameter. Numerical twin experiments clearly demonstrate the ability of the proposed solution to accurately estimate the posterior distribution of the parameters of interest in complex atmospheric environments such as industrial and urban ones. As future research directions, we plan to study the impact of the scenario's characteristics such as the environmental variables, source's characteristics, placement and noise of the sensors on the proposed Bayesian solution. Some real datasets, such as the FUSION Fields Trials 2007 (FFT07) Platt and Deriggi (2010) or the MUST campaign Biltoft (2001), will also be used to assess the ability of the proposed SMC sampler to solve the STE problem in a large number of different and various scenarios.

References

Andrieu C, Moulines É. On the ergodicity properties of some adaptive mcmc algorithms. The Annals of Applied Probability 2006;16(3):1462–505.

Biltoft CA. Customer report for mock urban setting test. Technical Report; DPG Document Number 8-CO-160-000-052. Prep. for Def. Threat. Reduct. Agency; 2001.

Cappé O, Guillin A, Marin JM, Robert CP. Population monte carlo. Journal of Computational and Graphical Statistics 2004;13(4):907–29.

Chow FK, Kosovic B, Chan ST. Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations. Journal of Applied Meteorology and Climatology 2008;47(6):1553–72.

Cornuet JM, Marin JM, Mira A, Robert CP. Adaptive Multiple Importance Sampling. Scandinavian Journal of Statistics 2012;.

Del Moral P, Doucet A, Jasra A. Sequential Monte Carlo samplers. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2006;68(3):411–36.

Delle Monache L, Lundquist JK, Kosović B, Johannesson G, Dyer KM, Aines RD, Chow FK, Belles RD, Hanley WG, Larsen SC, Loosmore Ga, Nitao JJ, Sugiyama Ga, Vogt PJ. Bayesian Inference and Markov Chain Monte Carlo Sampling to Reconstruct a Contaminant Source on a Continental Scale. Journal of Applied Meteorology and Climatology 2008;47(10):2600–13.

Doucet A, De Freitas N, Gordon N, editors. Sequential Monte Carlo Methods in Practice. Springer-Verlag, 2001.

Doucet A, Godsill S, Andrieu C. On sequential Monte-Carlo sampling methods for Bayesian filtering. Statistics and Computing 2000;10:197–208.

Jasra A, Stephens DA, Doucet A, Tsagaris T. Inference for lévy-driven stochastic volatility models via adaptive sequential monte carlo. Scandinavian Journal of Statistics 2011;38(1):1–22.

Jasra A, Stephens DA, Holmes CC. On population-based simulation for static inference. Statistics and Computing 2007;17(3):263–79.

Keats A, Yee E, Lien FS. Bayesian inference for source determination with applications to a complex urban environment. Atmospheric Environment 2007a;41(3):465–79.

Keats A, Yee E, Lien FS. Bayesian inference for source determination with applications to a complex urban environment. Atmospheric Environment 2007b;41(3):465–79.

Koohkan MR, Bocquet M. Accounting for representativeness errors in the inversion of atmospheric constituent emissions: application to the retrieval of regional carbon monoxide fluxes. Tellus 2012;64:1–17.

Lewellen WS, Sykes RI. Analysis of concentration fluctuations from lidar observations of atmospheric plumes. Journal of Climatology and Applied Meteorology 1986;5(2):1145–54.

Liu J. Monte Carlo Strategies in Scientific Computing. New York: Springer, 2001.

Neal R. Annealed Importance Sampling. Statistics and Computing 2001;:125–39.

Nguyen TLT, Septier F, Peters GW, Delignon Y. Efficient Sequential Monte-Carlo Samplers for Bayesian Inference. IEEE Transactions on Signal Processing 2016;64(5):1305–19.

Oldrini O, Armand P. Validation and Sensitivity Study of the PMSS Modelling System for Puff Releases in the Joint Urban 2003 Field Experiment. Boundary-Layer Meteorology 2019;171:513–35.

Oldrini O, Armand P, Duchenne C, Olry C, Tinarelli G. Description and preliminary validation of the PMSS fast response parallel atmospheric flow and dispersion solver in complex built-up areas. J of Environmental Fluid Mechanics 2017;17(3):1–18.

Oldrini O, Armand P, Perdriel S. Parallelization performances of PMSS flow and dispersion modelling system over a huge urban area. Atmosphere 2019;10(7).

Peters GW. Topics in Sequential Monte Carlo Samplers. Master's thesis; University of Cambridge; 2005.

Platt N, Deriggi D. Comparative investigation of source term estimation algorithms using FFT07 data. In: 8th Conference on Artificial Intelligence Applications to Environmental Sciences, AMS Annual Meeting. 2010. .

Rajaona H, Septier F, Armand P, Delignon Y, Olry C, Albergel A, Moussafir J. An adaptive Bayesian inference algorithm to estimate the parameters of a hazardous atmospheric release. Atmospheric Environment 2015;122:748–62. URL: https://halinstitut-mines-telecom.archives-ouvertes.fr/hal01238921. doi:10.1016/j.atmosenv.2015.10.026.

Robert CP, Casella G. Monte Carlo statistical methods. Springer, 2004.

Seibert P, Frank A. Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode. Atmospheric Chemistry and Physics 2004;4(1):51–63.

Septier F. Report No. 2 - Source Term estimation. Technical Report; IMT Lille Douai, Univ. Bretagne Sud, CEA; 2019.

Septier F, Armand P, Duchenne C. A Bayesian inference procedure based on inverse dispersion modelling for source term estimation in built-up environments. Atmospheric Environment 2020;242.

Septier F, Carmi A, Godsill S. Tracking of Multiple Contaminant Clouds. In: 12th International Conference on Information Fusion, 2009. FUSION '09. Seattle, WA, United States; 2009. p. 1280–7. URL: https://halinstitut-mines-telecom.archives-ouvertes.fr/hal00566637.

Tinarelli G, Mortarini L, Trini Castelli S, Carlino G, Moussafir J, Olry C, Armand P, Anfossi D. Description and preliminary validation of the PMSS fast response parallel atmospheric flow and dispersion solver in complex built-up areas. American Geophysical Union (AGU) 2013;200(A):311–27.

Trini Castelli S, Armand P, Tinarelli G, Duchenne C, Nibart M. Validation of a Lagrangian particle dispersion model with wind tunnel and field experiments in urban environment. Atmos Env 2018;193:273–89.

Winiarek V, Bocquet M, Saunier O, Mathieu A. Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: Application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant. Journal of Geophysical Research 2012;117(D5):D05122.

Yee E, Hoffman I, Ungar K. Bayesian Inference for Source Reconstruction: A Real-World Application. International Scholarly Research Notices 2014;2014(1):1–12.

Yee E, Lien FS, Keats A, D'Amours R. Bayesian inversion of concentration data: Source reconstruction in the adjoint representation of atmospheric diffusion. Journal of Wind Engineering and Industrial Aerodynamics 2008;96(10-11):1805–16.

Zhou Y, Johansen AM, Aston JA. Toward Automatic Model Comparison: An Adaptive Sequential Monte Carlo Approach. Journal of Computational and Graphical Statistics 2016;25(3):701–26. doi:10.1080/10618600.2015.1060885.