17th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
9-12 May 2016, Budapest, Hungary

A BAYESIAN APPROACH OF THE SOURCE TERM ESTIMATE COUPLING RETRO-DISPERSION COMPUTATIONS WITH A LAGRANGIAN PARTICLE DISPERSION MODEL AND THE ADAPTIVE MULTIPLE IMPORTANCE SAMPLING

Harizo Rajaona[1], François Septier[2,3], Yves Delignon[2,3], Patrick Armand[1], Laurent Makke[4], Christophe Olry[4], and Armand Albergel[4]

[1]CEA, DAM, DIF, F-91297 Arpajon, France
[2]Institut Mines-Télécom / Télécom Lille 1, F-59650 Villeneuve d'Ascq, France
[3]CRIStAL UMR CNRS 9189, F-59651 Villeneuve d'Ascq CEDEX, France
[4]ARIA Technologies, F-92100 Boulogne-Billancourt, France

Abstract: This paper presents an enhanced version of a STE adaptive algorithm based on probabilistic Bayesian inference (AMIS) that estimates the parameters of an atmospheric pollution source term. After introducing the problem and presenting the initial results obtained with the first version of the algorithm, we describe an efficient way to reduce the computational load in its procedure by shifting the most time-consuming step outside the iterative loop. This is made possible by applying the duality relationship between forward and adjoint advection-diffusion equations, and is proven to work well on a synthetic example using Retro-SPRAY, the backward implementation of SPRAY, the Lagrangian particle dispersion model of the PMSS suite.

Key words: *Source Terme Estimate, Bayesian approach, LPDM, retro-dispersion, AMIS.*

INTRODUCTION

The threat of Chemical, Biological, Radiological, and Nuclear (CBRN) releases in the atmosphere is a key issue. Such incidents may be due to terrorist acts, using non-conventional methods such as dirty bombs in order to create panic. The origin of these events can also be accidental, for example given a leak of hazardous material on an industrial site. Either way, the development of tools to detect the source and assess the parameters of the release is a major concern for the population's safety. Scientifically speaking, the problem of source term estimation (STE) is quite challenging, because obtaining the most accurate estimation within the shortest amount of time is crucial.

There are currently several approaches to solve STE problems, each of them using a specific set of skills. One line of study focuses on adjoint-transport modelling and retro-transport, as further developed in Pudykiewicz (1998) or Issartel and Baverel (2003) where backward simulations are computed using the principle of time-symmetry in atmospheric transport to reconstruct the source. These methods perform well, but most of them do not offer the possibility to properly quantify the uncertainty relative to the given estimation.

Another approach consists in using probabilistic methods. The optimization problem of computing the source term parameters can be treated as a maximum-likelihood estimation scheme (Nehorai, 1995), or by finding the maximum of the posterior distribution of the source parameters given the concentration measurements (Winiarek, 2012). Such methods provide point estimates but are not meant to provide a complete overview of the posterior distribution. To do so, it is possible to use stochastic simulation techniques such as Monte Carlo methods, coupled with Bayesian inference, so that the resulting algorithm delivers a sample of parameters whose distribution aims at approximating the posterior. This kind of method also allows taking into account potential prior information on the source parameters. Several related examples exist in the literature, that use the notorious Markov Chain Monte Carlo (MCMC) algorithm, such as (Delle Monache, 2008), (Chow, 2008), (Keats, 2007) or (Yee, 2014).

Even though MCMC methods are efficient in a number of cases, they are prone to several issues, regarding the inherent *burn-in* time necessary before the convergence, or the choice of how to initialize properly the Markov chain.

In this study, we focus on an alternative Bayesian method called *Adaptive Multiple Importance Sampling (AMIS)*. As presented in (Rajaona, 2015), by adding an advanced adaptive layer to the classical IS scheme, it provides good results for STE problems within a correct amount of time. This paper presents an improvement of the original method, aiming at optimizing the most time-consuming step in the algorithm by using the duality relationship with adjoint models for evaluating concentrations. The enhanced method is illustrated with a synthetic example.

THE BAYESIAN SOLUTION OF THE STE PROBLEM

The main goal of our work is to use the concentrations measured by a sensor network, resulting from the release of hazardous materials from a source point, in order to characterize the parameters of the latter, namely its position $\boldsymbol{\theta} = (x_s, y_s)$ and its temporal emission profile $\boldsymbol{q} = (q_1, q_2, \dots, q_{Ts})$ over T_s time-steps. We assume that the source and the sensors are located on the same vertical level.
The data model we use assumes the following source-receptor relationship:

$$\boldsymbol{d} = C(\boldsymbol{\theta})\boldsymbol{q} + \boldsymbol{\varepsilon} \tag{1}$$

$\boldsymbol{d} = (d_1, \dots, d_m)$ is the vector of observations, and $\boldsymbol{\varepsilon}$ is the vector that gathers representativeness, observation and model error in a single term. We assume that $\boldsymbol{\varepsilon}$ is independent and identically distributed over the sensors, and follows a normal distribution $N(0, \sigma_{obs}^2)$, where σ_{obs}^2 is the error variance. $C(\boldsymbol{\theta})$ is a source-receptor matrix of the concentrations obtained at each sensor from a unitary release of a source at potential location $\boldsymbol{\theta}$.
Our objective is to compute the posterior probability distribution $p(\boldsymbol{\theta}, \boldsymbol{q} \,|\boldsymbol{d})$ of the source parameters given the observations. By using Bayesian inference, the posterior can be rewritten as follows, using a marginalization relationship:

$$p(\boldsymbol{q}, \boldsymbol{\theta}|\boldsymbol{d}) = p(\boldsymbol{q}|\boldsymbol{\theta}, \boldsymbol{d})p(\boldsymbol{\theta}|\boldsymbol{d}) \tag{2}$$

Since we assume that the observation noise is Gaussian and that the random variables $\boldsymbol{q}$ and $\boldsymbol{\theta}$ are independent, then by making a Gaussian assumption over the prior distribution $p(\boldsymbol{q})$ as in (Winiarek et al, 2011), the marginal posterior $p(\boldsymbol{q}|\boldsymbol{\theta}, \boldsymbol{d})$ can be computed analytically, leaving only the posterior distribution of the source location to be estimated numerically. This distribution, denoted by $\pi(\boldsymbol{\theta})$, can be expressed up to the normalizing constant $p(\boldsymbol{d})$ by the following formula, thanks to Bayes rule:

$$\pi(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\boldsymbol{d}) = \frac{p(\boldsymbol{d}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{d})} \propto p(\boldsymbol{d}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{3}$$

$p(\boldsymbol{\theta})$ is the prior distribution of the source location, and is chosen to be uniform over the domain, implying that the source can be anywhere within the domain boundaries. $p(\boldsymbol{d}|\boldsymbol{\theta})$ is the likelihood of having the concentrations y given a source located at $\boldsymbol{\theta}$.

To estimate the so-called target posterior distribution $\pi(\boldsymbol{\theta})$, we resort to the Adaptive Multiple Importance Sampling (AMIS) algorithm (Cornuet, 2012), which iteratively draws sets of samples $\boldsymbol{\theta} = \theta_1, \dots, \theta_{N_p}$ from a proposal distribution $\varphi(\boldsymbol{\theta})$, computes the importance weights $w = \frac{\pi(\boldsymbol{\theta})}{\varphi(\boldsymbol{\theta})}$, and adjusts the parameters of φ so that it tends to fit the posterior distribution we try to approximate. The originality of the AMIS lies in the fact that all the importance weights that were previously computed are recycled at each iteration, and reused, in order to accelerate the convergence towards the posterior estimate. A common choice for φ is a mixture of D multivariate Gaussian distributions (Cappé, 2008) as it is flexible enough to adapt to complex targets, and the update equations for its parameters are analytically available.

In (Rajaona, 2015), the AMIS methodology is applied to the Fusion Fields Trial 2007 (FFT07) experiment, using real concentration data to estimate the source location and emission profile for the trial #7. The AMIS algorithm is coupled with a Gaussian puff dispersion model for the calculation of the source-receptor matrix, and provides good results for locating the source as shown in Figure 1. Regarding the emission rate estimation, despite the fact that the original profile is not completely reconstructed, the estimation provided by the algorithm manages to approximate the start and stop times of the emission, and the estimated release rates remain well within the confidence interval.

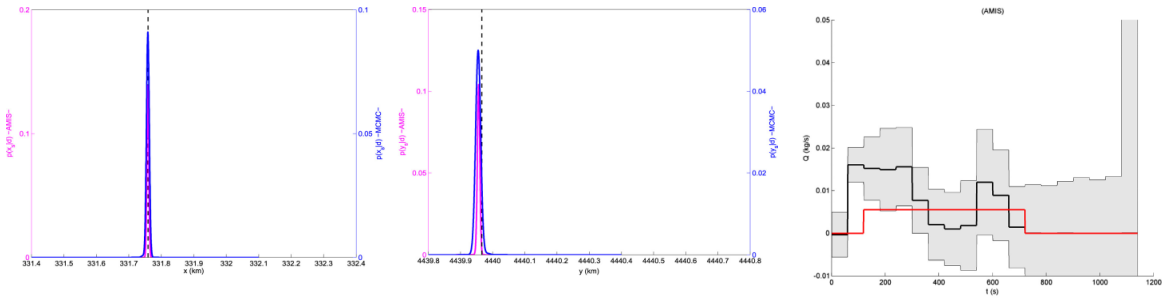

Figure 1. Results of the AMIS algorithm for the estimation of FFT07-trial#7's source. Left and middle: estimation of $p(x_s|d)$ and $p(y_s|d)$ for AMIS (blue) and MCMC (magenta) compared to the true value (dashed black). Right: estimation of $p(q|\theta,d)$ (black) and $\pm 2\sigma$ confidence interval (grey) compared to the true value (red).

THE ALTERNATIVE SOLUTION FOR COMPUTING THE SOURCE-RECEPTOR RELATIONSHIP

One of the most important feedbacks that emerge from the FFT07 case study was the high number of calls to the dispersion model, which had to be summoned for each sample at each iteration. The Gaussian model we used allowed a reasonable total computational time, but in its original form, the overall algorithmic scheme would certainly scale badly if we were to switch for a more accurate model.

That is why we came up with a more efficient way of optimizing the building of the source-receptor matrix, which is the most costly operation in computation time. In its current version, the source-receptor matrix $C(\boldsymbol{\theta})$ is obtained by solving a forward advection-diffusion equation and retrieving its solutions at the sensor locations. However, the duality relationship mentioned in (Keats, 2007) also stipulates that the same values can be obtained by solving the adjoint advection-diffusion equation, provided that the considered release is unitary and instantaneous. In other words, by using a backward dispersion model, it is possible to fill the source-receptor matrix with values of conjugate concentrations to build an operator $C^*(\boldsymbol{\theta})$ that can substitute to $C(\boldsymbol{\theta})$ in Equation (1).

More precisely, if $C(R_i, t_j|\theta, t'_n)$ denotes the concentration given by a forward dispersion model at sensor R_i and observation time t_j, due to a release at emission time t'_n of a source located at position $\boldsymbol{\theta}$, then the equivalent value in the backward framework is $C^*(\boldsymbol{\theta}, t'_n|R_i, t_j)$, which represents the conjugate concentration measured at point $\boldsymbol{\theta}$ at time t'_n due to a backward release from the sensor R_i at time t_j (Figure 2).

The major improvement resulting from this modification is that there are no more iterative calls to the dispersion model in the iterative scheme of the AMIS as the values of C^* for each point of the domain are pre-computed and stored in external files. This way, while computing the likelihood during the execution of the AMIS, the algorithm just pulls the C^* data corresponding to the sampled values of $\boldsymbol{\theta}$, rather that launching a forward dispersion calculation as it originally did.

$$C(\theta) = \begin{bmatrix} C(R_1,t_1|\theta,t'_1) & C(R_1,t_1|\theta,t'_2) & \cdots & C(R_1,t_1|\theta,t'_{T_s}) \\ C(R_1,t_2|\theta,t'_1) & C(R_1,t_2|\theta,t'_2) & \cdots & C(R_1,t_2|\theta,t'_{T_s}) \\ \vdots & \vdots & & \vdots \\ C(R_1,t_{T_c}|\theta,t'_1) & C(R_1,t_{T_c}|\theta,t'_2) & \cdots & C(R_1,t_{T_c}|\theta,t'_{T_s}) \\ C(R_2,t_1|\theta,t'_1) & C(R_2,t_1|\theta,t'_2) & \cdots & C(R_2,t_1|\theta,t'_{T_s}) \\ C(R_2,t_2|\theta,t'_1) & C(R_2,t_2|\theta,t'_2) & \cdots & C(R_2,t_2|\theta,t'_{T_s}) \\ \vdots & \vdots & & \vdots \\ C(R_2,t_{T_c}|\theta,t'_1) & C(R_2,t_{T_c}|\theta,t'_2) & \cdots & C(R_2,t_{T_c}|\theta,t'_{T_s}) \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ C(R_{N_c},t_{T_c}|\theta,t'_1) & C(R_{N_c},t_{T_c}|\theta,t'_2) & \cdots & C(R_{N_c},t_{T_c}|\theta,t'_{T_s}) \end{bmatrix} \Leftrightarrow C^*(\theta) = \begin{bmatrix} C^*(\theta,t'_1|R_1,t_1) & C^*(\theta,t'_2|R_1,t_1) & \cdots & C^*(\theta,t'_{T_s}|R_1,t_1) \\ C^*(\theta,t'_1|R_1,t_2) & C^*(\theta,t'_2|R_1,t_2) & \cdots & C^*(\theta,t'_{T_s}|R_1,t_2) \\ \vdots & \vdots & & \vdots \\ C^*(\theta,t'_1|R_1,t_{T_c}) & C^*(\theta,t'_2|R_1,t_{T_c}) & \cdots & C^*(\theta,t'_{T_s}|R_1,t_{T_c}) \\ C^*(\theta,t'_1|R_2,t_1) & C^*(\theta,t'_2|R_2,t_1) & \cdots & C^*(\theta,t'_{T_s}|R_2,t_1) \\ C^*(\theta,t'_1|R_2,t_2) & C^*(\theta,t'_2|R_2,t_2) & \cdots & C^*(\theta,t'_{T_s}|R_2,t_2) \\ \vdots & \vdots & & \vdots \\ C^*(\theta,t'_1|R_2,t_{T_c}) & C^*(\theta,t'_2|R_2,t_{T_c}) & \cdots & C^*(\theta,t'_{T_s}|R_2,t_{T_c}) \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ C^*(\theta,t'_1|R_{N_c},t_{T_c}) & C^*(\theta,t'_2|R_{N_c},t_{T_c}) & \cdots & C^*(\theta,t'_{T_s}|R_{N_c},t_{T_c}) \end{bmatrix}$$

Figure 2. Equivalence relationship between $C(\theta)$ and $C^*(\theta)$ over a range of N_c sensors, $(1, \dots, T_c)$ observation time steps and $(1, \dots, T_s)$ emission time steps.

To validate our method, we substituted the Gaussian model with a more elaborate Lagrangian Particle Dispersion Model, namely the Parallel Micro-Swift-Spray (PMSS) suite (Tinarelli et al., 2013). We take advantage of Retro-SPRAY (Armand, 2013), which is the implementation of the backward version of PMSS, in order to prepare the repository of conjugate concentration data. We considered a countryside landscape over a 6km $\times$ 6km square domain based on real topographical data for our experiment. A single release is simulated, that emits at a constant rate for one hour, and which resulting concentrations are given by a network of 11 sensors scattered over the domain (Figure 3).

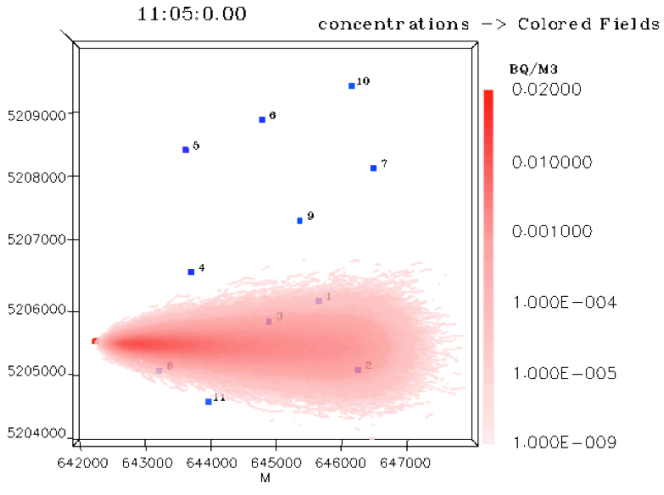

Figure 3. Snapshot of the release over the domain, with the location of the source (in red) and the sensors (in blue)

As seen in Figure 4, the results illustrate a good estimation of the source location obtained after 10 iterations of the AMIS algorithm. Furthermore, the total amount of computation time is quite lower than the estimated projection for the forward dispersion configuration in the same use case: it takes around 7.5 seconds to perform an AMIS iteration using backward pre-calculated concentration, whereas an iteration with on-the-fly computation of forward concentrations takes approximatively 15 minutes in the same context.

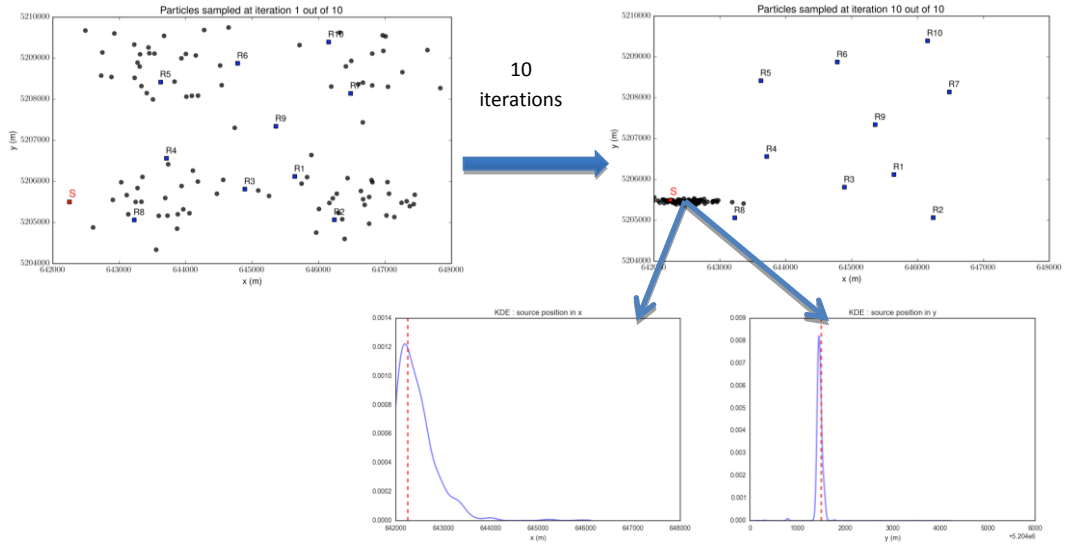

Figure 4: Scatterplot of the samples generated by the AMIS algorithm using Retro-SPRAY, with any initial knowledge at the initialization (left), after 10 iterations (right) with the corresponding estimated spatial posterior distributions of the source location (bottom right).

CONCLUSION

In this paper we have presented an enhanced version of the AMIS algorithm for STE problems. By using the adjoint version of the PMSS model, we demonstrate that it is possible to scale our algorithm by shifting most of the computational load outside the iterative scheme. Provided that a set of conjugate concentrations was pre-calculated, the execution time of our STE algorithm is considerably reduced compared to the initial scheme which resorted to forward dispersion computations.

Even if additional work is still required to ensure the stability of the algorithm, especially on the estimation of the emission profile, such improvements pave the way for future work on more complex situations such as urban scenarios, where applying elaborate models such as PMSS becomes essential.

REFERENCES

Armand, P., Olry, C., Albergel, A., Duchenne, C., & Moussafir, J. (2013). Development and application of Retro-spray, a backward atmospheric transport and dispersion model at the regional and urban scal, 15th International conference on Harmonisation within Atmospheric dispersion Modelling fir Regulatory Purposes, Harmo'15, pp. 789-793.

Cappé, O., Douc, R., Guillin, A., Marin, J. M., & Robert, C. P. (2008). Adaptive importance sampling in general mixture classes. *Statistics and Computing*, **18**, 447-459.

Chow, F. K., Kosovic, B., & Chan, S. (2008). Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations. *Journal of applied meteorology and climatology*, **47**, 1553-1572.

Cornuet, J.M., Mira, A., & Robert, C. P. (2012). Adaptive multiple importance sampling. Scandinavian *Journal of Statistics*, **39**, 798-812.

Delle Monache, L., Lundquist, J. K., Kosovic, B., Johannesson, G., Dyer, K. M., Aines, R. D., ... & Loosmore, G. A. (2008). Bayesian inference and Markov chain Monte Carlo sampling to reconstruct a contaminant source on a continental scale. *Journal of Applied Meteorology and Climatology*, **47**, 2600-2613.

Issartel, J. P., & Baverel, J. (2003). Inverse transport for the verification of the Comprehensive Nuclear Test Ban Treaty. *Atmospheric Chemistry and Physics*, **3**, 475-486.

Keats, A., Yee, E., & Lien, F. S. (2007). Bayesian inference for source determination with applications to a complex urban environment. *Atmospheric environment*, **41**, 465-479.

Nehorai, A., Porat, B., & Paldi, E. (1995). Detection and localization of vapor-emitting sources. *IEEE Transactions on Signal Processing*, **43**, 243-253.

Pudykiewicz, J. A. (1998). Application of adjoint tracer transport equations for evaluating source parameters. *Atmospheric Environment*, **32**, 3039-3050.

Rajaona, H., Septier, F., Armand, P., Delignon, Y., Olry, C., Albergel, A., & Moussafir, J. (2015). An adaptive Bayesian inference algorithm to estimate the parameters of a hazardous atmospheric release. *Atmospheric Environment*, **122**, 748-762.

Tinarelli, G. Mortarini, L., Trini-Castelli, S., Carlino, G., Moussafir, J., Olry C., Armand, P. and Anfossi D. (2013). Review and validation of Micro-Spray, a Lagrangian particle model of turbulent dispersion. Lagrangian Modeling of the Atmosphere, Geophysical Monograph, 200, 311-327

Winiarek, V., Bocquet, M., Saunier, O., & Mathieu, A. (2012). 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: Atmospheres*, **117**(D5).

Yee, E., Hoffman, I., & Ungar, K. (2014). Bayesian inference for source reconstruction: A real-world application. International Scholarly Research Notices, 2014.