A Bayesian approach of the source estimate coupling retro-dispersion computations with a Lagrangian particle dispersion model and the Adaptive Multiple Importance Sampling

Rajaona Harizo, Septier F., Delignon Y., Armand P., Makke L., Olry C., Albergel A.

17th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes

9-12 May 2016, Budapest, Hungary

CRIStAL Centre de Recherche en Informatique, Signal et Automatique de Lille

ARIA TECHNOLOGIES

DE LA RECHERCHE À L'INDUSTRIE cea

Direct modeling : the "murderer" point of view

(*) an almost perfect ~~crime~~ model

Inverse modeling: the "investigator" point of view

- A lot of solutions
- What are the clues ?
- More clues, more efficient !

Outline

1- Introduction

2- The Bayesian framework and the AMIS algorithm

3- Optimizing the AMIS by using a retro-dispersion model

4- Two-use cases : open-country and city downtown

5- Conclusion

Introduction

- Goal: estimate the parameters (location and release rate) of a hazardous release, in order to:
  - Give the best term source (where and how much)
  - Enhance the simulation of the resulting plume by giving better input data to the dispersion model ➔ provide the best impact assessment map of the current situation
- Two main approaches are possible and have been investigated in the literature:
  - Optimizatic
    - Variation
    - Genetic a


Inverse modelling of methane fugitive emissions from waste water treatment plant

Y. Roustan[1], M. Bocquet[1], E. Eriksson[2], D. Buty[2]
A. Finlayson[3], F. Innocenti[3], R. Robinson[3], D. Butterfield[3]
[1] CEREA, joint laboratory École des Ponts ParisTech - EDF R&D, Université Paris-Est
[2] ARIA Technologies, [3] National Physical Laboratory

yelva.roustan@enpc.fr

  - Bayesian inference and stochastic sampling (find the highest probability)

The Bayesian framework

- Why a probabilistic approach :
  - Taking into account the various uncertainties in the observations and in the dispersion model
  - Dealing with the presence (or also absence) of prior information on the source term parameters
  - Estimating the uncertainty related to the estimation results
- Generative model for the observation data:

$$d = q\, C(\theta) + \varepsilon$$

  - q (t) temporal release profile
  - $C(\theta)$ is a source-receptor matrix obtained by running a dispersion model for a unitary release from a source located at position θ
  - ε is an aggregation of all error sources (model, observation, representativeness) into a single vector
- In the Bayesian context, the objective is to estimate the posterior distribution $p(\theta, q \mid d)$ of the source parameters (location θ and temporal release profile q) given the concentration measurements d provided by a sensor network:

$$p(\theta, q \mid d) = \frac{p(d \mid \theta, q)\, p(\theta, q)}{p(d)}$$

The Bayesian framework

- The posterior distribution of the source parameters can be expressed using its marginal components:

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

- Estimating the marginal posterior $p(q|\theta, d)$ of q can be done analytically :
  - prior distribution $p(q)$ is Gaussian
  - the observation error is also Gaussian centered on measurements

- Estimating the posterior distribution $p(\theta|d)$ of θ :
  - Calls for the application of Bayes rule :

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

  - Requires the use of simulation-based methods (Monte Carlo) because it has no closed form due a highly non-linear likelihood $p(d|\theta)$ that relies on an atmospheric dispersion model run.
  - Can be done using a sequential-based approach **(Markov Chain Monte Carlo -MCMC-)** or a population-sampling approach **(Importance Sampling).**

The AMIS algorithm

- The *Adaptive Multiple Importance Sampling* (AMIS) algorithm is based on an consequent sampling scheme, where a target distribution (namely the posterior distribution) is approximated by weighted samples from a proposition distribution

- The *"Adaptive"* algorithm improves the standard importance sampling procedure by:
  - Allowing the update of the proposal distribution, which can be chosen as a flexible combination of well-known kernels (e.g. a multivariate Gaussian mixture)
  - Optimally recycling the importance weights at each iteration to fully exploit the full available information and accelerate the convergence

First Application restriction

- One release (time dependent / accidental release)
- The source and measurements are at ground level
- No plume rise
- The time of release starting and ending are unknown
- Meteorological data are time dependent

The AMIS algorithm

Iterative scheme of the AMIS algorithm applied to the STE problem:

Sample N_p elements $\theta_k^{(1)}, \theta_k^{(2)}, \dots, \theta_k^{(N_p)}$ from the current proposal distribution

Compute the likelihood $p(d|\theta_k^{(i)})$ and $p(q|\theta_k^{(i)}, d)$ by calling a dispersion model

Compute the current importance weight $w_k^{(i)}$

Loop over all the sampled elements

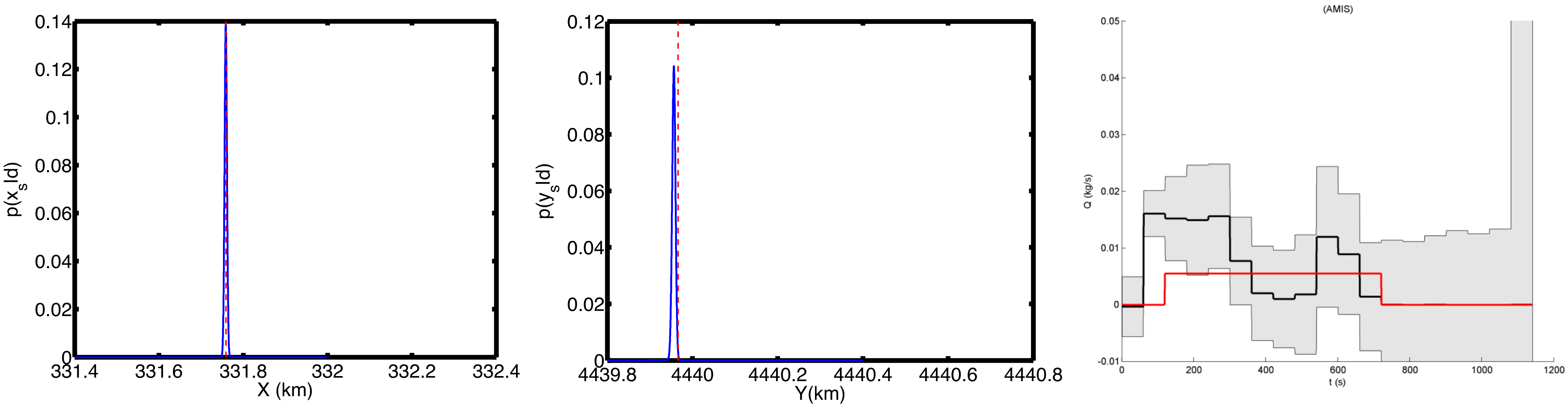
Compute the recycled importance weights over all previous iterations $w_{0:k-1}$

Update the proposal parameters

Need of a very quick dispersion model that does not allow to take benefit of all the 3D available info

The AMIS algorithm

- Preliminary tests applied on an experimental case (FFT07 experiment) showed a good estimation of the source location, but the release rate reconstruction is not as accurate as expected [Rajaona et al, 2015]



Estimation of the source parameters for the trial 7 of FFT07: position in x (left) and y (center) compared to the true location (red), and reconstructed release rate (right) with 95% confidence interval compared to the true emission profile (red). The dispersion model used to compute the likelihood is a simple Gaussian puff model.

[Rajaona et al., 2015] : 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.

The AMIS algorithm

- One of the downsides of the stochastic Bayesian approach is the high number of calls to a CPU-time consuming forward model when it comes to iteratively compute the likelihood for each sampled element.

- The model cannot scale efficiently if a more complex dispersion model is needed (e.g. in an urban scenario).

- The AMIS scheme needs to be optimized in order to deal with a more elaborate dispersion model.

Optimizing using a retro-dispersion model

- Solution: use the duality relationship as mentioned in [Keats et al., 2007] to switch to a retro-dispersion model

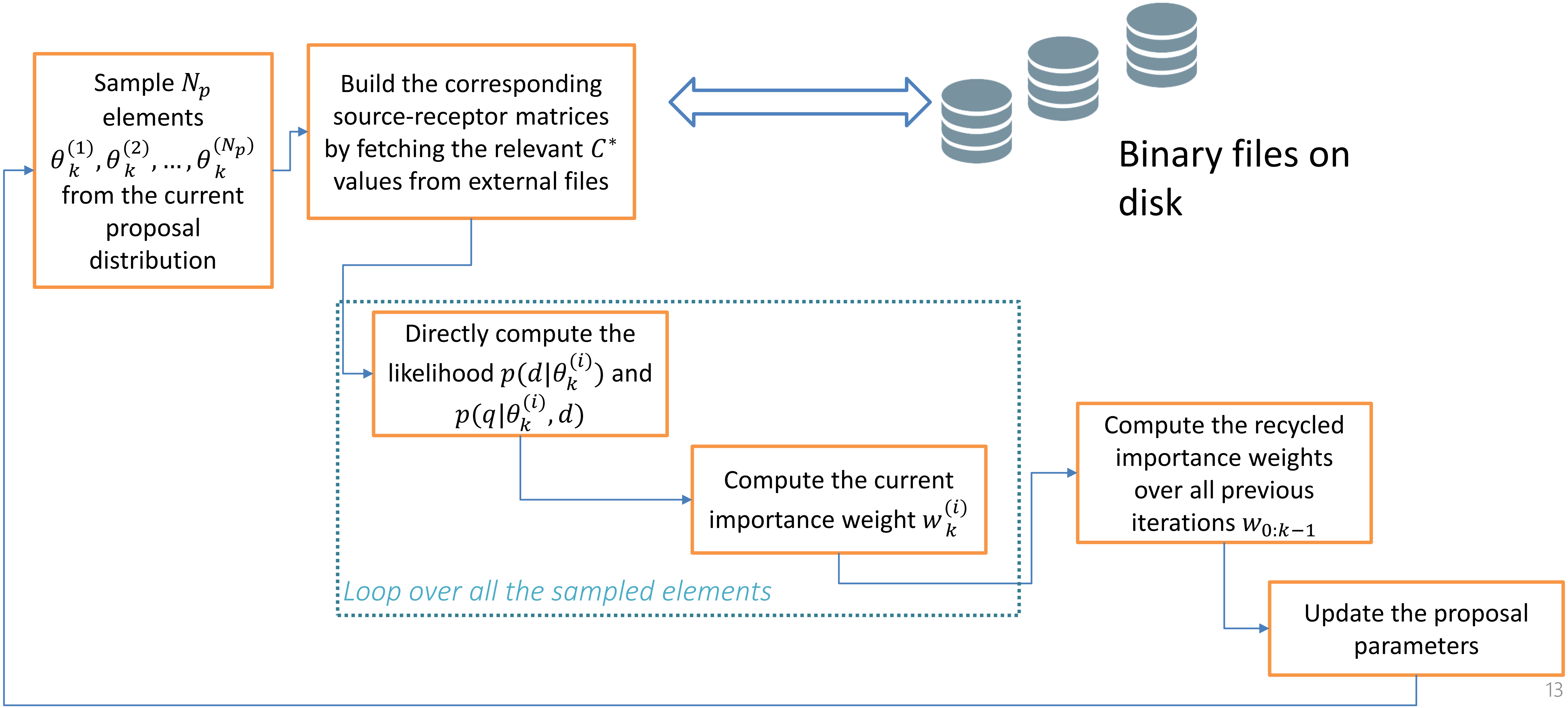
$$C(\theta)=\begin{bmatrix} C(R_1,t_1|\theta,t'_1) & C(R_1,t_1|\theta,t'_2) & \cdots & C(R_1,t_1|\theta,t'_{T_s}) \\ C(R_1,t_2|\theta,t'_1) & C(R_1,t_2|\theta,t'_2) & \cdots & C(R_1,t_2|\theta,t'_{T_s}) \\ \vdots & \vdots & & \vdots \\ C(R_1,t_{T_c}|\theta,t'_1) & C(R_1,t_{T_c}|\theta,t'_2) & \cdots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

- The conjugate concentrations C^* is obtained from the retro-model to build the source-receptor matrices.

- By pre-computing the C^* matrix before the AMIS estimation process and store the results on disk, we remove the multiple calls to the forward dispersion model in the loop.

[Keats et al., 2007]: Keats, A., Yee, E., & Lien, F. S. (2007). *Bayesian inference for source determination with applications to a complex urban environment.* Atmospheric environment, 41(3), 465-479.

Optimizing using a retro-dispersion model

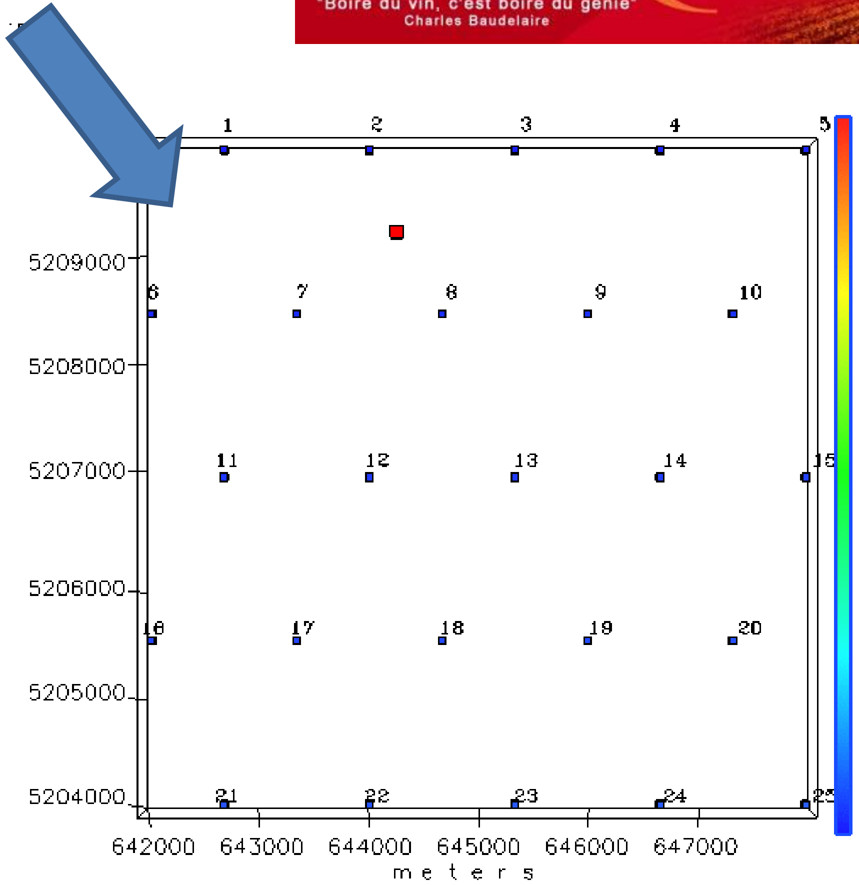


Optimizing using a retro-dispersion model

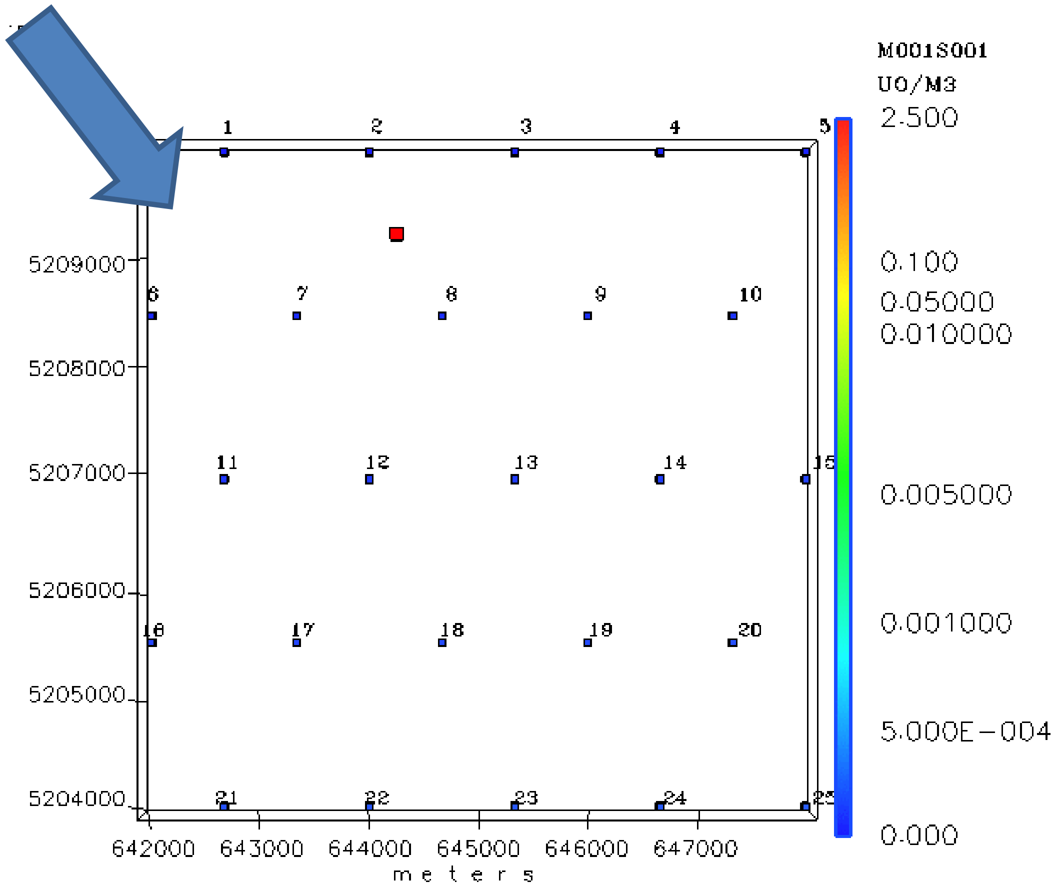
- We use Parallel Micro-SWIFT-SPRAY (PMSS) as dispersion model:
  - SWIFT is a diagnostic model using a mass-conservation principle to build 3D interpolated wind fields
  - SPRAY is a Lagrangian particle dispersion model that is used to generate synthetic concentration observations
  - Retro-SPRAY is the dual of the SPRAY model, and is used to build the C^* retro-dispersion fields

- Validation tests were performed using simulated observation data over realistic terrain characteristics in two cases:
  - 1st use case ("BEAUNE"): in a countryside landscape with a constant wind
  - 2nd use case ("OPERA"): in an urban context (neighborhood in downtown Paris) with a heterogeneous wind field

First use-case “Beaune”



First use-case “Beaune”

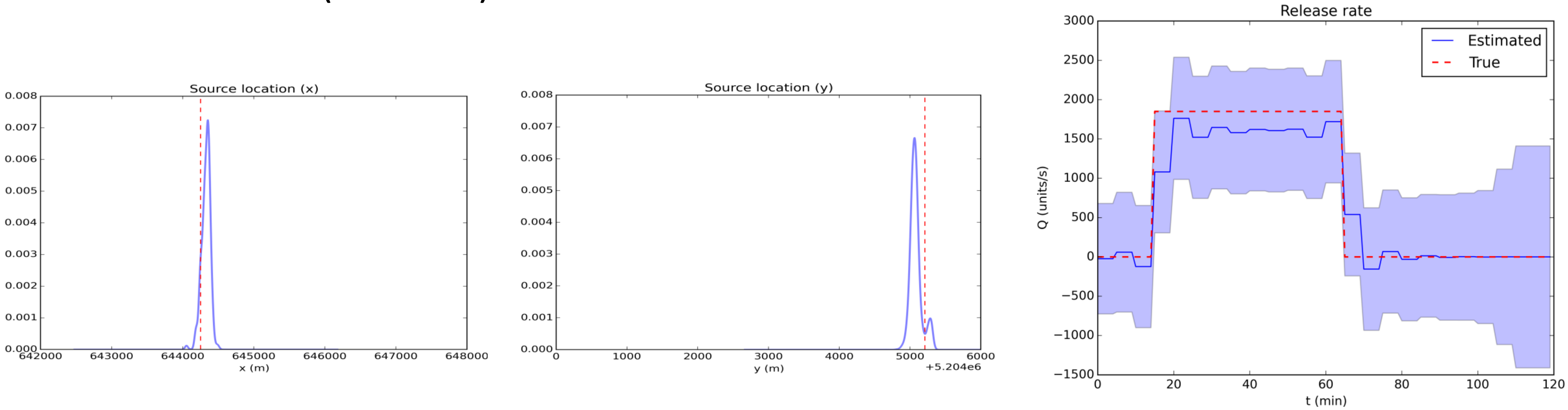


1st use-case (BEAUNE) input data:

SW point coordinates (km)	(642.000; 647.980)
NE point coordinates (km)	(5204.000; 5209.980)
Nb. Of meshes (X,Y)	**(300,300) = 90 000**
Mesh resolution (m)	20
Wind speed	1.5 m/s
Wind direction	330°
Release duration	45 mn (from 10:15 to 11:00 am)
Release rate	1850 units/s
Nb. sensors	**25**
Observation time frame	From 10:05 am to 12:00pm

First use-case "Beaune"

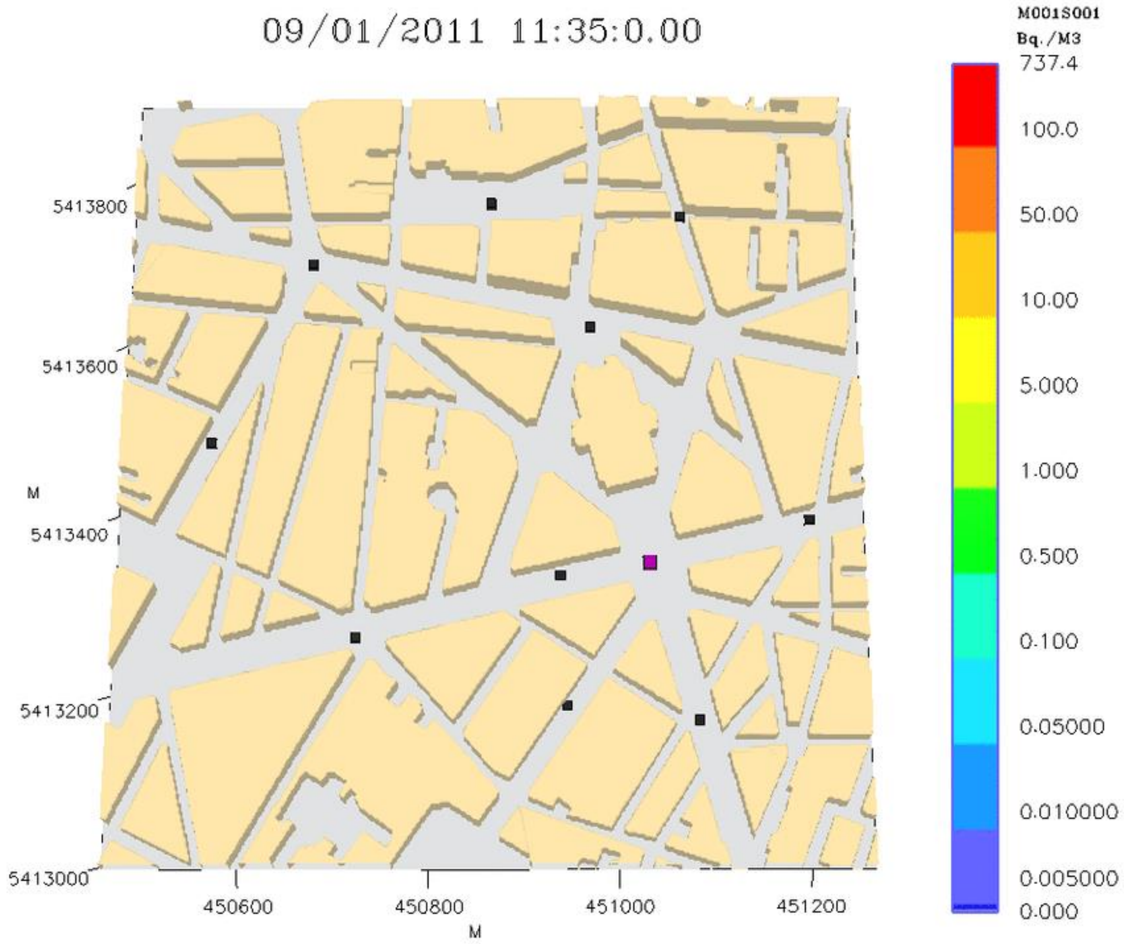
- 1st use case (BEAUNE) results:



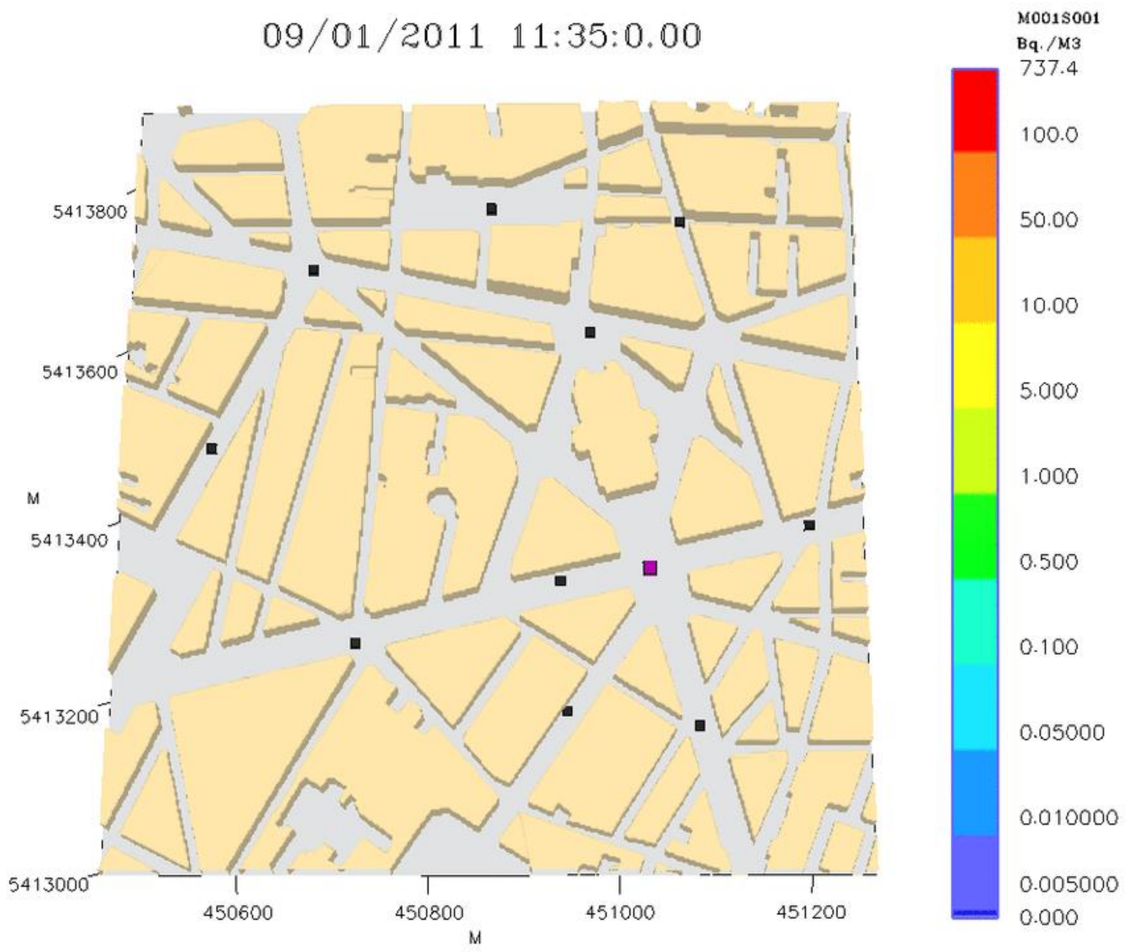
Estimation of the source parameters for the BEAUNE simulation: position in x (left) and y (center) compared to the true location (red), and reconstructed release rate (right) with 95% confidence interval compared to the true emission profile (red).

- Good estimation of the release rate
- The estimated source location is not as good. Cause: differences between C and C^* source-receptor matrices ?

Second use-case “OPERA”



Second use-case "OPERA"

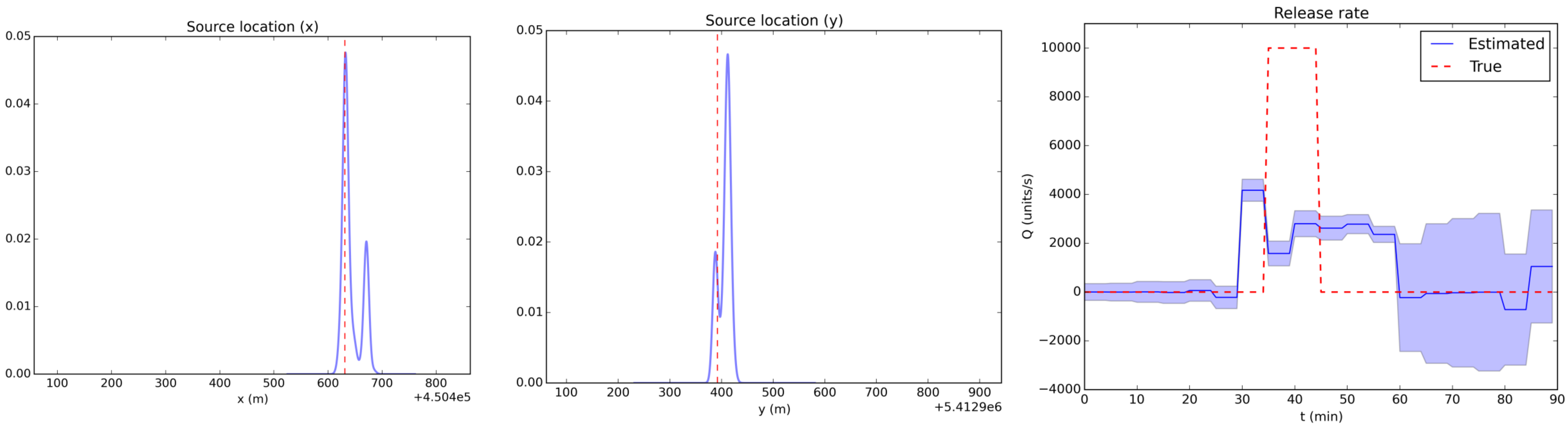


2nd use-case (OPERA) input data :

SW point coordinates (km)	(450.457;451.263)
NE point coordinates (km)	(5412.961;5413.842)
Nb. Of meshes (X,Y)	**(404,441)= 178 164**
Mesh resolution (m)	(2,2)
Wind speed	3m/s
Wind direction	11:00am-12:00 pm: **230°** | 12:00pm-1:00pm: **180°** | 1:00pm-2:00pm: **45°**
Release duration	10 mn (from 12:10 to 12:20 pm)
Release rate	10^4 units/s
Nb. sensors	**10**
Observation time frame	From 11:35 am to 1:00pm

Second use-case "OPERA"

- 2nd use case (OPERA):



Estimation of the source parameters for the OPERA use-case: position in x (left) and y (center) compared to the true location (red), and reconstructed release rate (right) with 95% confidence interval compared to the true emission profile (red).

- Estimation of the source location ➔ OK
- Difficulties to reconstruct the release rate due to the complexity of the use case (wind rotation, obstacles)

Conclusion

- New design of an optimized estimation process relying on Bayesian inference, stochastic modeling and a 3D Lagrangian dispersion model.
- AMIS methods improve MCMC methods efficiently
- The proposed scheme is able to hold the computational load and thus be scalable by using pre-computed retro-dispersion data obtained by the retro-dispersion model.
- The validation tests using this process show that the AMIS algorithm, could be used with any dispersion model and first results in complex situation are encouraging
- Perspectives and improvements:
  - Improve the method and test the estimation process on experimental data in an urban scenario
  - Study the influence of the various parameters in the AMIS algorithm
  - Extension of the method : elevated sources (including plume rises) and multi-sources

Thank you for your attention.

Positive Constraint Procedure