

# Inférence bayésienne adaptative pour la reconstruction de source en dispersion atmosphérique

Harizo Rajaona

Directeurs de thèse: Yves Delignon, François Septier

Lille

21 novembre 2016

- ① Contexte et problématique
- ② Méthodologie adaptative pour l'inférence bayésienne
- ③ Application au cas expérimental FFT07
- ④ Application avec modèle rétrograde aux cas simulés Beaune et Opéra
- ⑤ Conclusions et perspectives

- 1 Contexte et problématique
- 2 Méthodologie adaptative pour l'inférence bayésienne
- 3 Application au cas expérimental FFT07
- 4 Application avec modèle rétrograde aux cas simulés Beaune et Opéra
- 5 Conclusions et perspectives

Les rejets **NRBC**<sup>1</sup> dans l'atmosphère peuvent être d'origine :

- accidentelle (fuite ou explosion sur un site industriel),
- malveillante (actes terroristes)



Fukushima (2011)



Igualada (2015)



Los Angeles (2015)

Priorités :

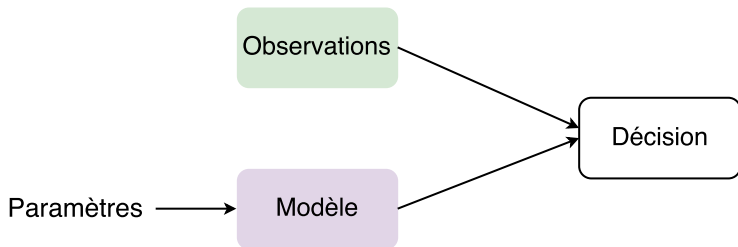
- informer et protéger les populations,
- atténuer/neutraliser le risque.

---

1. Nucléaires, Radiologiques, Biologiques, Chimiques

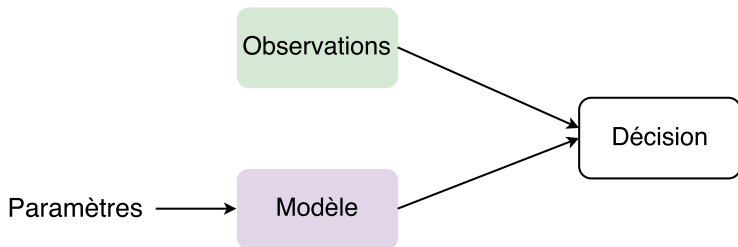
Outils de détection et d'évaluation du risque :

- données d'observation (capteurs mesurant la concentration de polluant)



Outils de détection et d'évaluation du risque :

- données d'observation (capteurs mesurant la concentration de polluant)
- outils de modélisation des phénomènes atmosphériques



## Modèle de dispersion

Outil de calcul numérique permettant de simuler la propagation dans l'atmosphère d'un rejet de polluant.

Typologie des modèles selon :

- l'échelle (locale, régionale, synoptique),
- le degré de simplification des équations de la mécanique des fluides

Paramètres d'entrée :

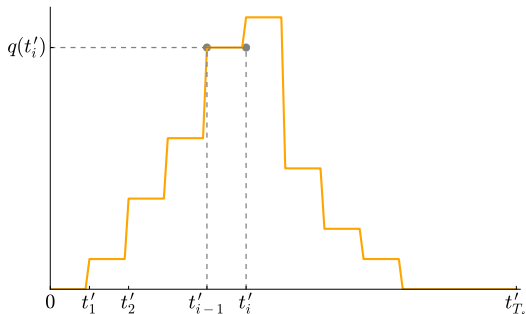
- données météorologiques : vent (direction + vitesse), température, humidité, nébulosité, flux de rayonnement...
- terme source : position, quantités émises, durée, substance émise...

# Terme source : définitions

Hypothèses sur la nature de la source :

- localisée (représentée par un point géographique  $\mathbf{x}_s \in \mathbb{R}^3$ ),
- unique (un seul point d'émission),
- non-instantanée, avec un profil temporel d'émission :

$$\mathbf{q} = (q(t'_1), q(t'_2), \dots, q(t'_{T_s}))$$

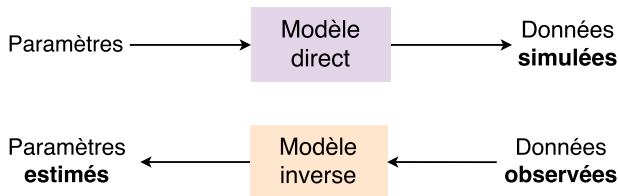


$\Rightarrow$  émission  
constante sur le  
palier  $[t'_{i-1}, t'_i]$



# Terme source : estimation

Reconstruire les paramètres d'un terme source ( $STE^2$ ) à partir des observations est un problème inverse.



Plusieurs approches de résolution possibles :

- rétro-transport,
- résolution d'un système linéaire,
- algorithmes évolutionnaires,
- méthodes bayésiennes et simulation stochastique.

# Problématique de recherche

On se concentre sur les méthodes bayésiennes :

- formalisme rigoureux pour estimation et quantification de l'incertitude,
- exploitation d'un nombre limité de mesures (régularisation),
- temps de calcul potentiellement élevés,
- estimation disjointe de la position et du profil d'émission.

## Problématique :

- ▶ Développer une méthode bayésienne pour estimer la localisation **et** le profil d'émission d'une source.
- ▶ Coupler cette méthode avec un modèle de dispersion atmosphérique dans une chaîne de calcul opérationnelle.

- 1 Contexte et problématique
- 2 Méthodologie adaptative pour l'inférence bayésienne**
- 3 Application au cas expérimental FFT07
- 4 Application avec modèle rétrograde aux cas simulés Beaune et Opéra
- 5 Conclusions et perspectives

**Principe** : estimation probabiliste des paramètres  $\theta$  d'un système ayant généré un ensemble d'observations  $\eta$ .

- $\theta \Rightarrow$  paramètres du terme source
- $\eta \Rightarrow$  mesures de concentration observées

## Règle de Bayes

$$\pi(\theta) = p(\theta|\eta) = \frac{p(\theta)p(\eta|\theta)}{p(\eta)} \propto p(\theta)p(\eta|\theta)$$

- loi a posteriori  $\pi(\theta)$  : information sur  $\theta$  connaissant  $\eta$ ,
- loi a priori  $p(\theta)$  : information préalable sur  $\theta$ ,
- vraisemblance  $p(\eta|\theta)$  : probabilité d'observer  $\eta$  pour  $\theta$  fixé.

- **Problème** :  $p(\eta|\theta)$  trop coûteuse (ou impossible) à calculer  
⇒ pas d'expression analytique pour  $\pi(\theta)$  !  
⇒ recours à des méthodes d'approximation numérique

## Méthodes de Monte-Carlo

Permettent d'approximer l'espérance de toute fonction d'une variable aléatoire de loi  $\pi$  en échantillonnant depuis cette loi :

$$\mathbb{E}_{\pi}[f(\theta)] = \int f(\theta)\pi(\theta)d\theta \simeq \frac{1}{N} \sum_{i=1}^N f(\theta^{(i)}), \quad \theta^{(i)} \sim \pi$$

- Obtention d'estimateurs bayésiens par simulation (ex : poser  $f(\theta) = \theta$  pour le MMSE<sup>3</sup>).

# Méthodes d'échantillonnage

- Algorithmes **MCMC**<sup>4</sup> :  $\pi$  est la distribution stationnaire d'une chaîne de Markov construite par itérations successives.
  - Metropolis-Hastings
  - échantillonneur de Gibbs

Bons résultats obtenus dans la littérature STE :

- en milieu urbain : Keats (2007), Chow (2008)
- en multi-source : Yee (2008)

Inconvénients :

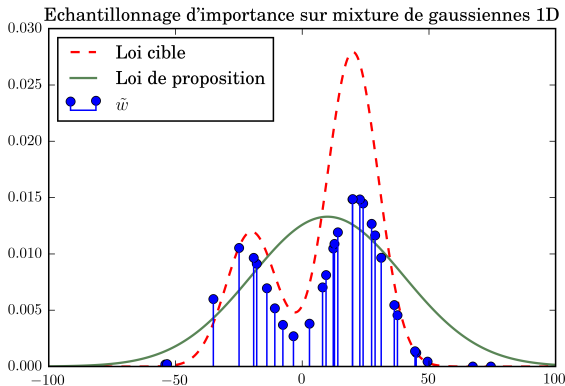
- perte d'une partie des échantillons générés (*burn-in*)
  - états corrélés (non-parallélisable)
  - MH : performances liées au choix du noyau et de l'initialisation
  - Gibbs : requiert les lois conditionnelles de  $\theta$
- Algorithmes **d'échantillonnage d'importance (IS)**<sup>5</sup> : tirage d'une population d'échantillons pondérés (ou particules) à partir d'une loi de proposition.

---

4. Markov Chain Monte Carlo

5. Importance Sampling

# Echantillonnage d'importance



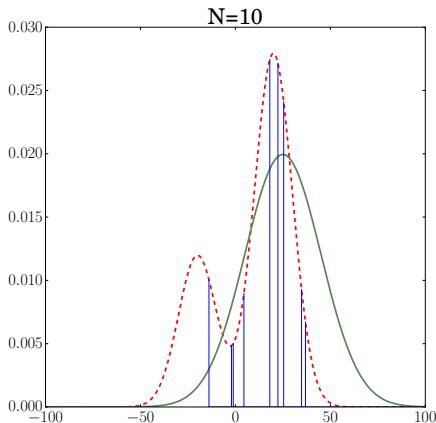
Avantages :

- échantillons i.i.d. : traitement parallélisable
- exploitation de tous les échantillons générés

# Echantillonnage d'importance

Inconvénients :

- performances fortement conditionnées par le choix de la loi de proposition !

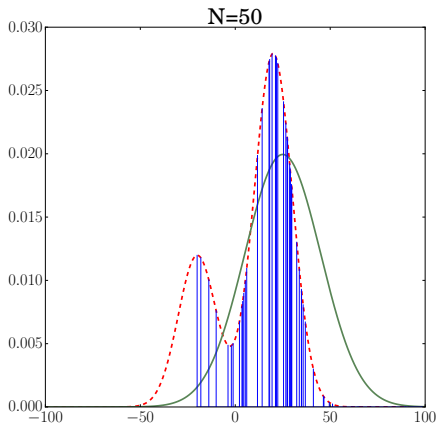




# Echantillonnage d'importance

Inconvénients :

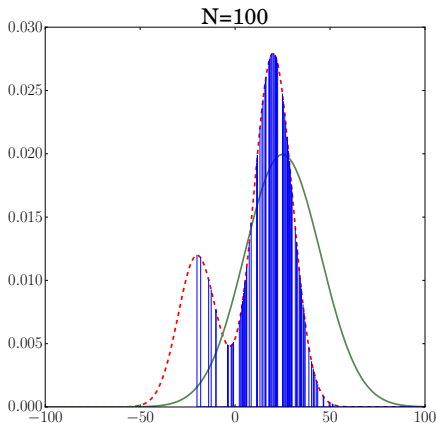
- performances fortement conditionnées par le choix de la loi de proposition !



# Echantillonnage d'importance

Inconvénients :

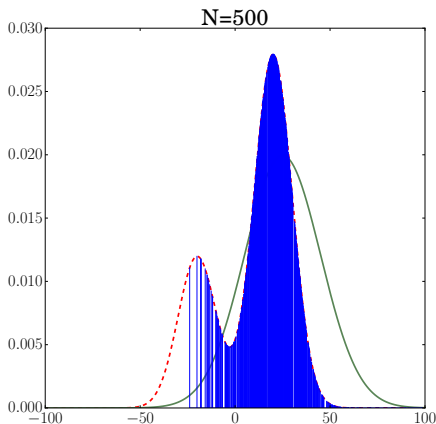
- performances fortement conditionnées par le choix de la loi de proposition !



# Echantillonnage d'importance

Inconvénients :

- performances fortement conditionnées par le choix de la loi de proposition !



# Echantillonnage d'importance adaptatif

Solution : adapter itérativement la loi de proposition  $\varphi$

## Population Monte Carlo [Cappé et al., 2004]

Introduction du concept d'adaptation par minimisation de :

$$KL(\pi, \varphi) = \int \log \left( \frac{\pi(\boldsymbol{\theta})}{\varphi(\boldsymbol{\theta})} \right) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (\text{divergence KL})$$

## D-kernel PMC [Douc et al., 2007]

- Loi de proposition  $\varphi_{\alpha} \Rightarrow$  mélange de noyaux fixes pondérés  $\{(\alpha_d, \varphi_d)\}_{1 \leq d \leq D}$  :

$$\varphi_{\alpha}(\boldsymbol{\theta}) = \sum_{d=1}^D \alpha_d \varphi_d(\boldsymbol{\theta})$$

- Optimisation des  $\alpha_d$  par minimisation KL.

## M-PMC [Cappé et al., 2008]

- Loi de proposition  $\varphi_{(\alpha, \nu)} \Rightarrow$  mélange de noyaux paramétriques pondérés  $\left\{ \left( \alpha_d, \varphi_d(\cdot | \nu_d) \right) \right\}_{1 \leq d \leq D}$  :

$$\varphi_{(\alpha, \nu)}(\theta) = \sum_{d=1}^D \alpha_d \varphi_d(\theta | \nu_d)$$

- Optimisation des  $\alpha_d$  et  $\nu_d$  par minimisation KL (algorithme EM).

Jusqu'ici : optimisation itérative seulement en fonction de l'itération précédente !

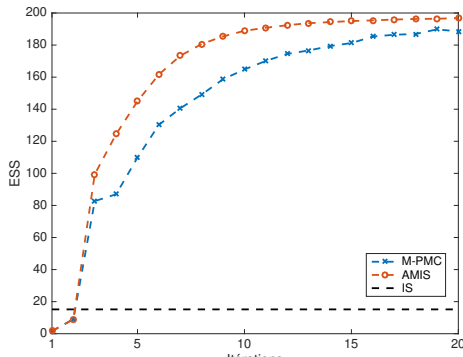
# Echantillonnage d'importance adaptatif

## Adaptive Multiple Importance Sampling (AMIS) [Cornuet et al., 2012]

- Loi de proposition identique à celle du M-PMC
- Ré-utilisation des particules de toutes les itérations pour :
  - le calcul et recyclage de tous les poids d'importance,
  - l'optimisation des  $\alpha_d$  et  $\nu_d$ .

### Avantages :

- utilisation efficace de tous les échantillons disponibles
- convergence plus rapide vers la loi cible
- variance d'erreur d'estimation réduite

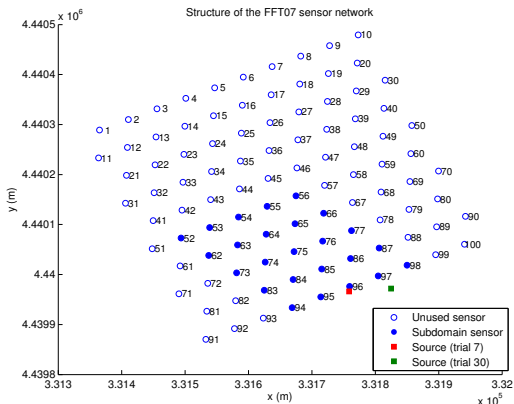


- 1 Contexte et problématique
- 2 Méthodologie adaptative pour l'inférence bayésienne
- 3 Application au cas expérimental FFT07**
- 4 Application avec modèle rétrograde aux cas simulés Beaune et Opéra
- 5 Conclusions et perspectives

# L'expérience FFT07

## Campagne expérimentale :

- rejets de gaz traceur sur terrain instrumenté dans diverses configurations (période, météo, nombre de sources...)
- création de données de référence pour validation d'algorithmes STE



- Restriction à  $N_C = 25$  capteurs
- Plage d'observations par capteur  $T_C$  réduite (moyennage)



# Formalisation du problème STE

Objectif : estimer les paramètres de position  $\mathbf{x}_s$  et d'émission  $\mathbf{q}$  de la source pour une configuration donnée (*trial*) de l'expérience FFT07.

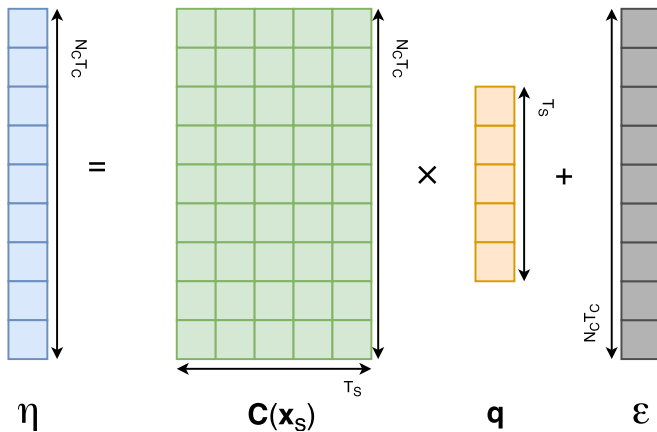
Modèle de données :

$$\boldsymbol{\eta} = \mathbf{C}(\mathbf{x}_s)\mathbf{q} + \boldsymbol{\varepsilon}$$

où :

- $\boldsymbol{\eta} \in \mathbb{R}^{N_C T_C}$  : observations concaténées par capteur
- $\mathbf{C}(\mathbf{x}_s) \in \mathbb{R}^{N_C T_C \times T_s}$  : matrice source-récepteur construite avec un modèle de dispersion
- $\mathbf{q} \in \mathbb{R}^{T_s}$  : profil d'émission
- $\boldsymbol{\varepsilon} \in \mathbb{R}^{N_C T_C}$  : erreurs (observation, modèle)

# Formalisation du problème STE



# Formalisation du problème STE

$$\begin{matrix} R_1 \\ \eta \end{matrix} = \begin{matrix} 1_{t'_1} \\ C(x_s) \end{matrix} \times \begin{matrix} q \end{matrix} + \begin{matrix} \varepsilon \end{matrix}$$

# Formalisation du problème STE

The diagram illustrates the formalization of the STE problem as a matrix equation. It shows the following components:

- $R_1$ : A vertical column vector of 10 light blue squares, with the top square highlighted in a darker blue. Labeled  $R_1$  in blue text to its left.
- $\eta$ : A label in black text centered below the  $R_1$  vector.
- $=$ : An equals sign indicating the equation.
- $1_{t'_2}$ : A horizontal row vector of 5 green squares, labeled  $1_{t'_2}$  in green text above it.
- $C(x_s)$ : A 10x5 grid of light green squares, labeled  $C(x_s)$  in black text below it.
- $\times$ : A multiplication symbol.
- $q$ : A vertical column vector of 5 orange squares, labeled  $q$  in black text below it.
- $+$ : A plus sign indicating addition.
- $\varepsilon$ : A vertical column vector of 10 squares, with the top square being dark gray and the remaining 9 squares being light gray. Labeled  $\varepsilon$  in black text below it.

The overall equation represented is:

$$\eta = C(x_s) \times q + \varepsilon$$

# Formalisation du problème STE

The diagram illustrates the formalization of the STE problem as a matrix equation. It shows the relationship between the observed data  $\eta$ , the model matrix  $C(x_s)$ , the parameters  $q$ , and the noise  $\varepsilon$ .

The equation is represented as:

$$\eta = C(x_s) \times q + \varepsilon$$

The components are visualized as follows:

- $\eta$ : A vertical column vector of 10 light blue squares. The top square is labeled  $R_1$  in blue.
- $=$ : An equals sign.
- $C(x_s)$ : A 10x5 grid of light green squares. The top row is highlighted in a darker green and labeled  $1_{t'_3}$  in green.
- $\times$ : A multiplication symbol.
- $q$ : A vertical column vector of 5 orange squares.
- $+$ : A plus sign.
- $\varepsilon$ : A vertical column vector of 10 squares. The top square is dark gray, and the remaining 9 squares are light gray.

# Formalisation du problème STE

Diagram illustrating the formalization of the STE problem:

The equation is represented as:

$$\eta = C(x_s) \times q + \varepsilon$$

Where:

- $\eta$  (blue vector) is a vertical vector of 10 cells, with the top cell labeled  $R_1$ .
- $C(x_s)$  (green grid) is a 10x5 grid. The top row is labeled  $1_{t'_s}$ .
- $q$  (orange vector) is a vertical vector of 5 cells.
- $\varepsilon$  (gray vector) is a vertical vector of 10 cells.

# Formalisation du problème STE

The diagram illustrates the formalization of the STE problem as a matrix equation. It shows the relationship between several variables represented by grids of colored squares:

- $\eta$ : A vertical column of 10 light blue squares. The second square from the top is highlighted with a darker blue border and labeled  $R_2$  to its left.
- $=$ : An equals sign.
- $C(x_s)$ : A 10x5 grid of light green squares. The second row from the top is highlighted with a darker green border.
- $\times$ : A multiplication symbol.
- $q$ : A vertical column of 6 orange squares.
- $+$ : A plus sign.
- $\varepsilon$ : A vertical column of 10 light gray squares. The second square from the top is highlighted with a darker gray border.

The equation is represented as:

$$\eta = C(x_s) \times q + \varepsilon$$

# Formalisation du problème STE

The diagram illustrates the formalization of the STE problem as a matrix equation. It shows four vertical structures representing vectors or matrices, connected by mathematical operators.

- $\eta$** : A vertical vector of 10 light blue squares. The third square from the top is highlighted with a darker blue border and labeled "R3" to its left.
- $=$** : An equals sign.
- $C(x_s)$** : A 10x5 grid of light green squares. The third row from the top is highlighted with a darker green border.
- $\times$** : A multiplication symbol.
- $q$** : A vertical vector of 6 light orange squares.
- $+$** : A plus sign.
- $\varepsilon$** : A vertical vector of 10 light gray squares. The third square from the top is highlighted with a darker gray border.

The equation is represented as:

$$\eta = C(x_s) \times q + \varepsilon$$



# Formalisation du problème STE

The diagram illustrates the formalization of the STE problem. It shows a vertical vector  $\eta$  (blue) equal to a 10x5 grid  $C(x_s)$  (green) multiplied by a vertical vector  $q$  (orange), plus a vertical vector  $\epsilon$  (gray). The bottom row of  $C(x_s)$  and the bottom element of  $\eta$  are highlighted in darker shades.

$\eta = C(x_s) \times q + \epsilon$

$R_{N_C T_C}$

$\eta$   $C(x_s)$   $q$   $\epsilon$

(vide)

- 1 Contexte et problématique
- 2 Méthodologie adaptative pour l'inférence bayésienne
- 3 Application au cas expérimental FFT07
- 4 Application avec modèle rétrograde aux cas simulés Beaune et Opéra**
- 5 Conclusions et perspectives

(vide)

- 1 Contexte et problématique
- 2 Méthodologie adaptative pour l'inférence bayésienne
- 3 Application au cas expérimental FFT07
- 4 Application avec modèle rétrograde aux cas simulés Beaune et Opéra
- 5 Conclusions et perspectives**

(vide)