

Modélisation inverse des sources de pollution atmosphérique accidentelle : progrès récents

Inverse modelling of sources of atmospheric accidental release: recent progress

Marc BOCQUET*, **

Résumé

Dans cette revue, nous discutons de progrès récents dans la modélisation inverse des sources de polluants atmosphériques, en particulier d'origine accidentelle, de l'échelle régionale à l'échelle continentale. Il s'agit typiquement de caractériser et d'estimer une source de pollution émanant d'un site industriel au moyen d'un modèle numérique de chimie-transport et de mesures de concentrations de ce polluant distribuées en espace et en temps. Le formalisme mathématique permettant d'unifier ces informations est celui de l'assimilation de données, qui offre un cadre bayésien. Nous montrons comment formuler mathématiquement un tel problème, puis comment le résoudre, notamment à l'aide de techniques non-paramétriques. Parce que la chimie associée à un tel événement physique peut être souvent considérée linéaire à court terme, des méthodes mathématiques plus avancées peuvent être mises en œuvre avec des résultats mieux contraints. Au-delà de l'estimation des sources, la caractérisation des incertitudes liées à ces estimations est ensuite discutée. Des exemples récents issus de la littérature sont donnés sur les cas d'ETEX, de l'accident de Tchernobyl, et de l'incident d'Algésiras. Enfin, nous décrivons l'impact des résolutions spatiale et temporelle de l'espace des paramètres sur la reconstruction de la source. Dans le contexte de la dispersion atmosphérique, l'estimation de la source peut en dépendre significativement.

Mots clés

Modélisation inverse. Dispersion accidentelle. Assimilation de données.

Abstract

This review discusses recent progress of the inverse modelling of sources of atmospheric pollution, especially from accidental origin, from regional to continental scale. Typically, it amounts to estimate and characterise a pollution source from an industrial site, with a chemistry and transport model, and with measurements of concentration of this pollutant distributed in space and time. The mathematical formalism that allows to fuse these sources of information is given by the Bayesian framework of data assimilation. We show how to define mathematically such a problem, then how to solve it, in particular thanks to non-parametrical methods. Because the chemistry of species related to such a dispersion event can often be considered linear in the short term, more advanced mathematical methods can be implemented with better results. Beyond the estimation of these sources, the computation of the uncertainties attached to this estimation is discussed. Recent examples from the literature are then presented: the ETEX case, the Chernobyl case, and the Algeciras case. Lastly, we describe the impact of the spatio-temporal resolution of the parameter space on the retrieval. In the atmospheric dispersion context, the estimation of the source can depend significantly on this choice.

Keywords

Inverse modelling. Accidental dispersion. Data assimilation.

* Université Paris-Est – CEREa – Laboratoire commun École des Ponts ParisTech et EDF R&D – France.

** INRIA – Centre de Recherche Paris-Rocquencourt – France – E-mail : bocquet@cerea.enpc.fr

1. Intérêt de la modélisation inverse pour la pollution atmosphérique

1.1 Approche montante

Encore aujourd'hui, les inventaires d'émissions de polluants atmosphériques sont déterminés par l'approche montante. Il s'agit de recueillir longuement l'ensemble des informations nécessaires à leur construction : caractérisation des sources en nature, en intensité, dans l'espace et dans le temps. Cela s'applique à un grand nombre de types de chimie atmosphérique : inventaire des précurseurs de l'ozone en qualité de l'air, inventaire des sources anthropiques et estimations des sources et puits biogéniques pour le dioxyde de carbone et le méthane, et de façon similaire pour de nombreux gaz traces (polluants organiques persistants, métaux lourds, etc.).

Ces inventaires peuvent être validés par des mesures directes de flux (nécessairement délicates et qui ne permettent qu'un échantillonnage local). Ils peuvent être ensuite validés indirectement par la modélisation numérique de la chimie et du transport, en comparant les résultats de simulations aux mesures de concentrations de polluants. L'approche est indirecte car de nombreuses sources d'incertitude s'ajoutent : erreurs dans la modélisation du transport (météorologie) et de la chimie/physique (paramétrisations des réactions et de la physique), erreurs liées à l'échelle spatiale à laquelle travaille le modèle comparée à l'échelle des mesures.

L'incertitude associée à cette construction est significative. Pour les polluants primaires de la qualité de l'air elle est de l'ordre de plus ou moins 40 % [1].

1.2 Approche descendante

Une approche complémentaire a ensuite été envisagée : la modélisation inverse. Il s'agit d'utiliser l'ensemble des mesures de concentrations disponibles pour remonter aux champs d'émission, à l'aide du modèle numérique de chimie transport [2]. Par opposition à l'approche montante, cette manière de procéder est appelée approche descendante. Elle peut, ou non, faire usage d'informations *a priori* sur les émissions, c'est-à-dire avant l'exploitation des mesures : par exemple des estimations obtenues par l'approche montante. Ces estimations *a priori* forment ce que l'on appelle une ébauche. La méthode inverse est la mise en équations rigoureuse du procédé qui consiste à simuler des concentrations partant d'un inventaire *a priori*, puis en fonction du biais diagnostiqué par comparaison avec les mesures, à corriger l'inventaire d'un ou plusieurs facteurs, puis à resimuler, et ainsi de suite jusqu'à obtention d'un accord satisfaisant entre la simulation et les mesures.

La modélisation inverse peut être comprise comme une branche de l'assimilation de données. À chacune des informations *a priori*, est associée une incertitude : incertitude sur l'inventaire *a priori*, incertitude sur les mesures (qui n'ont pas encore été exploitées), incertitude sur le modèle. Ces incertitudes sont

souvent considérées comme bien établies et connues quantitativement. On considère souvent que c'est un point faible de la méthode dont on ne peut s'abstraire. Les observations exploitées sont reliées à ces *a priori* grâce au modèle numérique décrivant le système géophysique. Puis une inférence, généralement de type bayésien, est menée. Elle permet d'obtenir le meilleur compromis sur l'état réel du système, compte tenu de toutes les incertitudes précédemment décrites. Il est souvent possible et souhaitable d'associer une incertitude *a posteriori* à cette meilleure estimation.

En chimie atmosphérique, les applications de la modélisation inverse se multiplient :

- modélisation inverse des précurseurs de l'ozone [3, 4] ;
- modélisation inverse des sources de dioxyde de carbone, et plus généralement des gaz à effet de serre (dont le méthane) [5, 6] ;
- modélisation des espèces traces : monoxyde de carbone, particules, métaux lourds, etc. [7].

Elles font désormais appel non seulement aux observations ponctuelles mais aussi aux données satellitaires, à des modèles régionaux de chimie-transport mais aussi à des modèles globaux. Les applications des premier et troisième types relèvent de la pollution chronique (par opposition aux événements exceptionnels liés à des accidents de dispersion). L'application de la modélisation inverse aux sources de pollution chronique a été notamment décrite dans [31].

1.3 Modélisation inverse des sources accidentelles de pollution

Une application particulière concerne l'estimation des sources accidentelles en situation de crise, notamment d'un rejet accidentel d'un polluant dans l'atmosphère. Parce que le cadre conceptuel lié à ce type de problème peut paraître assez simple, les sources d'incertitudes qui lui sont associées sont assez bien identifiées. Le rejet étant généralement localisé en espace (beaucoup moins souvent en temps), des simplifications peuvent être envisagées dans la construction d'un inventaire *a priori*. Comme une ou quelques espèces sont généralement considérées, la chimie et les transformations physiques qui en découlent sont bien circonscrites. La modélisation physique et chimique est donc généralement plus simple. Compte tenu des échelles de temps, il en résulte souvent que la chimie est effectivement linéaire dans une bonne approximation. Comme le transport est linéaire, c'est l'ensemble de la chimie transport qui, en première approximation, peut être vu comme linéaire. C'est une source de simplification mathématique qui est mise à profit dans la mise en œuvre de l'inférence.

La source étant localisée, il est souvent suffisant de prendre en compte des conditions d'air pur pour les conditions aux limites d'un modèle de transport régional (à aire limitée), ce qui est impossible pour la majorité des autres problèmes de modélisation inverse atmosphérique.

La première source d'incertitude sur un événement de ce type est la source elle-même, puisqu'en l'absence de conditions aux limites et de condition initiale, c'est le seul facteur qui détermine la masse de polluant introduite dans le domaine. Le deuxième facteur d'incertitude est la météorologie qui est à l'origine d'au moins deux sources d'erreur importantes : les champs de vents d'une part et la caractérisation du transport vertical turbulent d'autre part.

1.4 Modélisation inverse et assimilation de données pour la dispersion accidentelle

L'identification (et la caractérisation) d'une source accidentelle est une fin en soi. Mais elle ne constitue pas le seul objectif digne d'intérêt. En cas de rejet atmosphérique accidentel, une des premières préoccupations du décideur est de prévoir l'évolution temporelle du panache et son étendue spatiale. En utilisant conjointement le modèle numérique de dispersion et les observations, on a recours à l'assimilation de données pour obtenir la meilleure prévision. Dans ce contexte particulier, puisqu'on a vu que la connaissance de la source est déterminante, l'assimilation de données pour prévoir l'évolution du panache est directement dépendante de la modélisation inverse de la source.

2. Approche bayésienne pour la modélisation inverse

2.1 Les approches statistiques

En l'absence de connaissance précise sur le modèle de dispersion, une approche possible consiste à établir une relation statistique entre les sources et les mesures. On peut avoir recours aux techniques très répandues de factorisation matricielle non-négative [8] pour établir un tel lien. Plusieurs contributions de ce numéro spécial de *Pollution Atmosphérique* utilisent et mettent en valeur cette approche.

2.2. L'approche bayésienne

L'approche bayésienne présuppose que le système est incertain et qu'on peut attribuer aux variables qui le décrivent des lois de probabilité. L'information *a priori* sur la source se modélise par une distribution $p(x)$. La probabilité de mesurer le vecteur d'observations y (composé de d observations) connaissant l'état du système est donnée par la loi de vraisemblance conditionnelle $p(y | x)$. La loi de Bayes permet alors de trouver la distribution *a posteriori* de la source x connaissant le jeu de mesures y :

$$p(x | y) = \frac{p(y | x) p(x)}{p(y)} \quad (1)$$

La distribution $p(y)$ peut être vue comme un simple facteur de normalisation de la distribution

a posteriori lorsque seuls les paramètres des sources sont recherchés. L'équation (1) constitue donc une manière élégante de poser un problème inverse [9], dans l'hypothèse où l'on est capable de spécifier (plus ou moins bien) les incertitudes dans le système.

La majorité des inférences de l'assimilation de données tire, *in fine*, leur justification de cette relation très simple. L'implémentation pratique de cette relation n'est toutefois pas évidente dans des espaces d'état de grande taille. Une large majorité d'études repose sur un choix gaussien pour les statistiques des variables d'état ou de contrôle. Dans le contexte des inversions de source, cela suppose que les erreurs *a priori* sur la source suivent une distribution gaussienne, tandis que la distribution des erreurs *a priori* est également une loi gaussienne. Il s'ensuit que la distribution *a posteriori* est elle-même gaussienne. Comme estimateur de la source, on recherche souvent le mode, c'est-à-dire la source maximisant la distribution *a posteriori*. Il est alors élégant de définir l'opposé du logarithme de la loi *a posteriori* comme une fonctionnelle à minimiser.

$$L(x) = -\ln(p(x | y)) = -\ln(p(y | x)) - \ln(p(x)) + \text{cste} \quad (2)$$

Dans les cas d'hypothèses gaussiennes *a priori*, et pour un modèle linéaire, cette fonctionnelle est quadratique. Il s'agit de la formulation variationnelle du problème de l'estimation de la source.

3. Modèles paramétriques et non-paramétriques

Une partie de la difficulté consiste à choisir les variables à inverser, encore appelées *variables de contrôle*. On peut distinguer deux types d'approches.

3.1. Intérêt et limitations de l'approche paramétrique

L'approche paramétrique consiste à ne choisir qu'un nombre limité de paramètres, seuls degrés de liberté de la source ou des émissions. Par exemple une source ponctuelle et instantanée au sol peut être complètement caractérisée par les coordonnées spatiales du point et par l'intensité du rejet. Cette approche est très efficace puisqu'elle réduit considérablement le nombre de degrés de liberté. Elle permet notamment l'exploitation complète de la formule de Bayes, au-delà de l'approximation gaussienne, lorsque les paramètres ne dépassent pas la dizaine [10, 11]. Mais la méthode a le défaut de ses qualités : en réduisant fortement le nombre de degrés de liberté, la structure de la source est très contrainte. Une erreur dans cette modélisation ou bien des erreurs dans le modèle de dispersion, auront pour conséquence une incohérence dans le système d'inversion.

3.2. L'approche générale non-paramétrique

L'approche non-paramétrique consiste à introduire autant de variables de contrôle que de degrés

de liberté exprimés dans la modélisation directe du problème de dispersion. Naturellement, cela conduit souvent à une multiplication des *variables de contrôle* (les variables caractérisant les émissions de façon typique), qu'un jeu d'observations en nombre plus réduit ne permettra pas de caractériser complètement.

On peut envisager des approches hybrides. Par exemple dans le cas des sources accidentelles, la source est localisée alors qu'elle a une durée, et se prolonge donc dans le temps. La position de la source pourrait donc être caractérisée de façon paramétrique (coordonnées), alors que la source serait discrétisée en temps et ainsi divisée en plusieurs taux d'émission.

Dans le cas non-paramétrique, on peut requérir que la source soit préférentiellement ponctuelle, sans l'imposer strictement, grâce à une fonction de coût judicieusement construite [12]. Cela conduit dans le cas de sources accidentelles ponctuelles à une meilleure estimation des sources.

Dans la suite, nous nous restreignons aux méthodes non-paramétriques, en discrétisant les champs à estimer.

4. Méthodes non-paramétriques pour la caractérisation des sources en dispersion accidentelle

4.1 Exploiter l'hypothèse de linéarité

La dispersion accidentelle peut être souvent modélisée comme un processus linéaire. Dans ce cas la source et les mesures sont liées par une matrice H qui représente à la fois l'opérateur d'observation et le modèle numérique de dispersion. Cet opérateur peut être pré-calculé. Dans un cas non-linéaire il dépendrait des variables de contrôle (le vecteur x), et rendrait alors réhibitoire tout calcul préalable puisqu'il faudrait le recalculer pour chaque nouvelle estimation de l'état du système. Observations et source sont donc liées par :

$$y = Hx + e \quad (3)$$

L'erreur est caractérisée par le vecteur e (à d composantes scalaires). La matrice H se calcule de deux manières. Lorsque le nombre d'observations dépasse le nombre de variables à inverser, il est préférable de la calculer colonne par colonne. Chacune des colonnes se calcule par une unique simulation directe, à partir d'une source élémentaire. Dans le cas contraire, lorsqu'il y a plus de variables à inverser que d'observations, il est préférable de calculer H ligne par ligne. Une ligne se calcule à l'aide d'une simulation adjointe. C'est-à-dire qu'elle est le résultat de la simulation du modèle décrit par l'équation différentielle adjointe de l'équation différentielle de la dispersion. Dans le cas d'une dispersion linéaire, cette équation adjointe prend une forme similaire à l'équation de la dispersion, en renversant le sens du vent et du temps. On

peut ainsi, de façon approchée, utiliser le modèle numérique original de dispersion pour réaliser ce calcul.

Lorsque l'on a directement accès à H , notamment grâce à la linéarité du problème, il n'est pas nécessaire *a priori* de recourir à des techniques du type filtrage séquentiel ou assimilation variationnelle où le temps est explicité (4D-Var). Dans le cas où le nombre de variables et le nombre d'observations sont très conséquents, la seule approche numériquement réalisable est le 4D-Var, qui minimise itérativement la fonction de coût (voir par exemple [13]).

4.2 Modélisation inverse Gaussienne, méthodes de projection, et simplexe

On suppose maintenant connue l'équation (3), c'est-à-dire que l'on a calculé H . Sous des hypothèses gaussiennes sur la source *a priori* et sur les erreurs, on peut alors minimiser

$$L(x) = \frac{1}{2} (y - Hx)^T R^{-1} (y - Hx) + \frac{1}{2} x^T B^{-1} x \quad (4)$$

pour reconstruire la source. R est la matrice de covariance d'erreur d'observation qui caractérise les moments d'ordre deux des erreurs d'observation, alors que B est la matrice de covariance d'erreur d'ébauche lié à l'incertitude *a priori* sur la source. Cette méthode est équivalente à la projection orthogonale de l'ébauche sur la base des solutions adjointes (lignes de H) en utilisant le produit scalaire de matrice B^{-1} . En l'absence d'informations sur la source, on préfère supposer l'ébauche de la source nulle. Une méthode élégante proposée dans [14] pour résoudre l'équation (3) en la transformant en une série d'inégalités linéaires repose sur l'algorithme du simplexe.

4.3 Modélisation inverse non-Gaussienne

Dans bien des cas en géophysique, la modélisation statistique gaussienne des variables est plus que suffisante. C'est notamment le cas lorsqu'on connaît bien le premier moment (la moyenne), c'est-à-dire l'ébauche de l'état du système. Les moments du second ordre (matrices de covariance d'erreur) caractérisent alors les fluctuations autour de cette ébauche. Pourtant dans le cas d'une source accidentelle, cette ébauche est très mal caractérisée : on ne sait pas la situer. Même si l'on connaît ses coordonnées, l'amplitude de l'émission n'est certainement pas connue. Dans un cadre gaussien, il est alors préférable de supposer *a priori* que cette ébauche est nulle, avec de fortes fluctuations potentielles.

Un autre argument plaide également en faveur d'une modélisation statistique dépassant le cadre gaussien dans ce contexte. Les sources correspondent à des quantités positives. Même lorsque l'on considère aussi bien des flux positifs (sources) que des flux négatifs (puits), l'*a priori* est localement (en espace et en temps) majoritairement positif ou négatif. Or la positivité ne s'exprime pas bien dans l'espace des moments : la spécification des moyennes et des

variances/covariances ne permet pas d'assurer la positivité. Il est possible de contraindre cette positivité en se reposant sur la technique des minimiseurs. Par exemple, L-BFGS-B ou MN1QN3, qui reposent sur l'algorithme quasi-Newton de Broyden, Fletcher, Goldfarb et Shanno (BFGS), sont des minimiseurs permettant d'imposer des bornes aux variables. Il est toutefois plus souple et plus élégant d'imposer cette contrainte dès le départ.

On peut ainsi choisir pour distribution des sources *a priori* des lois de probabilité positives, comme les lois de Poisson, Bernoulli, exponentielle, etc. Les possibilités sont innombrables mais il est souhaitable de choisir une loi permettant de mener des calculs analytiques, car les fonctions de coût effectives construites sur ces lois font souvent intervenir des transformations mathématiques de ces distributions (logarithme de ces lois, transformation log-Laplace, transformation de Legendre-Fenchel).

4.4 Inférence non-Gaussienne

L'application de la formule de Bayes sur le cas non-Gaussien peut être direct (par exemple on peut choisir pour $p(x)$ une loi positive et appliquer la formule Eq.(1)). On notera toutefois que, même dans le cas d'un modèle linéaire, aucune convexité de la fonction de coût ne peut être garantie *a priori*. Il est alors difficile de s'assurer de l'existence d'un seul minimum, ce qui complique considérablement l'interprétation de la minimisation de la fonction de coût par des méthodes déterministes de type descente (comme BFGS) [15].

Une approche alternative est une approximation de la loi de Bayes (asymptotiquement identique dans certaines conditions raisonnables), appelé *maximum d'entropie sur la moyenne*. Bien qu'elle ne mette en œuvre la loi de Bayes que de façon approchée, elle possède de nombreux atouts. Par exemple, on peut montrer qu'elle n'engendre que des fonctions de coût convexes. Bien sur, si le problème physique est tel qu'il existe physiquement plusieurs minima, alors l'inférence par le maximum d'entropie n'est qu'une simplification (que l'on peut voir comme une sorte d'enveloppe convexe de la fonction de coût originale). De plus, la méthode est souple et permet la construction de fonctions de coût adaptées à de nombreux problèmes sous contraintes [16, 17]. Prenons l'exemple d'une loi *a priori* de Poisson. Elle stipule qu'*a priori*, dans la maille k (parmi N) de l'espace de contrôle, une masse de polluant $x_k = mu_k$ est émise avec la probabilité

$$P(u_k) = e^{-\theta_k} \frac{\theta_k^{u_k}}{u_k!} \quad (5)$$

m est l'échelle de masse *a priori* par composante et θ_k le paramètre caractérisant dans la maille k la loi de Poisson. Alors on substitue à la fonction de coût quadratique Eq.(4) une fonction de coût, construite avec le maximum d'entropie sur la moyenne

$$L(x) = \frac{1}{2} (y - Hx)^T R^{-1} (y - Hx) + \sum_{k=1}^N \frac{x_k}{m} \ln \frac{x_k}{m\theta_k} + \theta_k - \frac{x_k}{m} \quad (6)$$

On observe que, dans ce cas précis, les erreurs sur les observations sont toujours considérées gaussiennes. Parce qu'elles exploitent mieux l'information disponible (positivité de la source par exemple), les méthodes non-gaussiennes engendrent des estimateurs de la source plus fins.

4.5. Linéarité et modélisation inverse non-Gaussienne : construction duale exacte

Souvent, les fonctions de coût sont exprimées dans l'espace primal, c'est-à-dire directement en fonction des variables de contrôle (source, émissions) comme pour les équations (4) et (6). Lorsque le nombre d'observations est inférieur au nombre des variables de contrôle (ce qui reste souvent le cas généralement en géophysique), il est possible de construire dans le cas gaussien une fonction de coût alternative, dite duale, équivalente à la fonction de coût primale, qui s'exprime en fonction de variables qui sont au nombre des observations. La méthode porte le nom de PSAS (*Physical space Statistical Analysis System*).

Dans le cas non-Gaussien, lorsque la fonction de coût primale n'est pas quadratique, il n'y a pas de méthode générale exacte permettant de construire une fonction de coût duale équivalente. Toutefois, lorsque le modèle est linéaire, comme on l'a supposé ici, l'analyse convexe non-linéaire autorise une construction élégante et efficace [15].

Par exemple la fonction de coût duale de (6) est

$$\Lambda(\beta) = \frac{1}{2} \beta^T R \beta + \sum_{k=1}^N \theta(\exp(m[\beta^T H]_k) - 1) - \beta^T y \quad (7)$$

où β est un vecteur de d multiplicateurs de Lagrange (autant que d'observations). L'estimateur de la source est reliée à β par

$$\bar{x}_k = m\theta_k \exp(m[\beta^T H]_k) \quad (8)$$

4.6. La caractérisation des erreurs d'estimation

Un avantage des méthodes paramétriques à faible nombre de variables est que, grâce à l'exploitation systématique de la formule de Bayes, il est aisé de caractériser l'incertitude liée à la source reconstruite. Toutefois, même dans le cas des méthodes déterministes (axées sur une fonction de coût), il est possible de mener une analyse dite du second ordre. Il s'agit de déterminer les incertitudes liées à la source reconstruite. La matrice covariance d'analyse qui caractérise ces incertitudes est proportionnelle à la matrice Hessienne calculée à partir de la fonction de coût. Par exemple dans le cas gaussien, il s'agit simplement de la matrice de covariance

$$P_a = (B^{-1} + H^T R^{-1} H)^{-1} \quad (9)$$

Il est possible de généraliser cette analyse aux fonctions de coût non-gaussiennes [15].

5. Application à des cas de dispersion réels

5.1. ETEX

En 1994, les sorties de plusieurs modèles numériques de dispersion de traceur ont été comparées aux observations de l'expérience ETEX (*European Tracer Experiment*). Celle-ci a été conduite par le *Joint Research Centre* de la commission européenne (ETEX symposium 1997; ETEX report 1998). Dans la première phase de l'expérience, 340 kilogrammes d'un traceur complètement inerte (c'est-à-dire sans chimie, sans dépôt sec ni humide) ont été rejetés en Bretagne près de Rennes. Le panache a été suivi par des mesures de surface à travers toute l'Europe aux stations synoptiques de l'Organisation Météorologique Mondiale [OMM].

Cette expérience a permis de comparer avec succès les mérites des modèles, au travers de l'exercice d'intercomparaison originel et de l'exercice ATMES-II. Un deuxième rejet, prévu dans la campagne (ETEX-II), a conduit à de fortes incohérences entre mesures et simulations, sans qu'il soit clairement établi s'il s'agit d'erreurs instrumentales (comme une dilution des échantillons due aux précipitations) ou d'erreurs dues aux faiblesses de la modélisation (comme la convection). ETEX-I constitue encore aujourd'hui l'expérience de dispersion à l'échelle continentale la mieux instrumentée et documentée. Données et caractéristiques sont accessibles à l'adresse : <http://rem.jrc.cec.eu.int/etex/>

À la fin des années 1990, les premières reconstructions de sources furent tentées, en utilisant les mesures d'ETEX (et spécifiquement ETEX-I). L'utilisation d'une approche variationnelle fut suggérée très tôt sur la base d'arguments théoriques par Robertson et Persson [18]. Elle fut appliquée aux données d'ETEX-I dans [19] mais dans le but de reconstruire le profil temporel de la source connaissant le point d'émission.

Comme alternative à l'approche variationnelle, l'utilisation du transport adjoint et de son interprétation particulière ainsi que son application à l'estimation des paramètres de la source ont été avancées par Pudykiewicz en 1998 [20]. Cela faisait suite aux travaux d'Uliasz sur l'utilisation de solutions adjointes du modèle [21]. Cet effort portant sur les solutions adjointes a été poursuivi par Hourdin et Issartel [22], suite à l'impulsion donnée par les chercheurs associés au traité d'interdiction des essais nucléaires (CTBT).

Les méthodes de rétro-trajectoires et d'attribution statistique de source, la modélisation Lagrangienne orientée récepteur de la dispersion, ont notamment été les premières techniques publiées applicables aux données réelles d'un événement de dispersion. Elles ont été essayées par Seibert sur ETEX-I [23].

En ce qui concerne les modèles eulériens, les premières tentatives d'inversions non-paramétriques tridimensionnelles (deux dimensions pour la surface et une pour le temps) sont dues à Issartel et Baverel

[14]. En se servant d'une approche par les solutions adjointes (baptisées *rétropanaches* dans ce contexte), une solution reposant sur la programmation linéaire (simplexe) a été développée pour le cas d'une source ponctuelle. Une méthode distincte reposant sur une projection sur une base de rétropanaches a été proposée pour traiter les sources ayant une extension spatiale dans [24]. Cette approche est équivalente à la minimisation de Eq.(4). Cependant, ces techniques ont principalement été testées sur des expériences jumelles où l'erreur modèle est absente.

Les premières reconstructions 3D satisfaisantes avec les données réelles ont été obtenues en utilisant les techniques d'inversion non-gaussiennes décrites précédemment [16, 25, 26]. La Figure 1 donne les résultats d'inversion en utilisant une approche gaussienne et une approche non-gaussienne. Dans le cas gaussien, on reconstruit 241 kg près de la source réelle et 684 kg au total. Dans le cas non-gaussien, on obtient respectivement 264 kg et 279 kg. Cette bonne estimation du terme source a permis une excellente reconstruction du panache d'ETEX-I [25]. Avec une estimation des paramètres qui entre dans la fonction de coût (1 ou 2 généralement), comme le rapport entre l'incertitude sur la source et celle sur les mesures, on peut approcher des 340 kg théoriques [26].

5.2. Algésiras

Après ETEX-I, une nouvelle application des méthodes d'inversion non-gaussiennes a été testée. L'incident d'Algésiras (Algeciras en Anglais) est la conséquence de la fusion accidentelle d'une source de radiothérapie dans un haut-fourneau à Algésiras près de Gibraltar dans le Sud de l'Espagne. Il est à l'origine de la dispersion de césium 137 dans l'atmosphère au printemps 1998. Quelques jours après l'accident, l'emplacement précis de la source a été établi, après que la direction du site Acerinox d'Algésiras ait signalé l'incident. On estime que le rejet a eu lieu dans un intervalle de deux heures dans la matinée du 30 mai 1998. La quantité relâchée est connue à un ordre de grandeur près, entre 8 et 80 Ci ou, de façon équivalente, entre $2,96 \cdot 10^{11}$ et $2,96 \cdot 10^{12}$ Bq. Ce sont des niveaux d'activité radiologique bien inférieurs à ceux résultant d'un accident de centrale nucléaire [fusion de cœur ou rupture d'un tube de générateur de vapeur]. On ne peut donc pas restreindre *a priori* l'ensemble des sites possibles aux seules centrales nucléaires [12].

Le nombre très faible de mesures significatives (118), qui correspondent de plus à des mesures sur filtre souvent intégrées sur plusieurs journées ne permet pas une reconstruction 3D. En revanche une reconstruction 1D (profil temporel) peut être conduite avec succès : bonne estimation de l'ordre de grandeur de l'activité relâchée et du moment du rejet (Figure 2).

Plus récemment, la même expérience a été tentée à l'aide de techniques paramétriques [10].

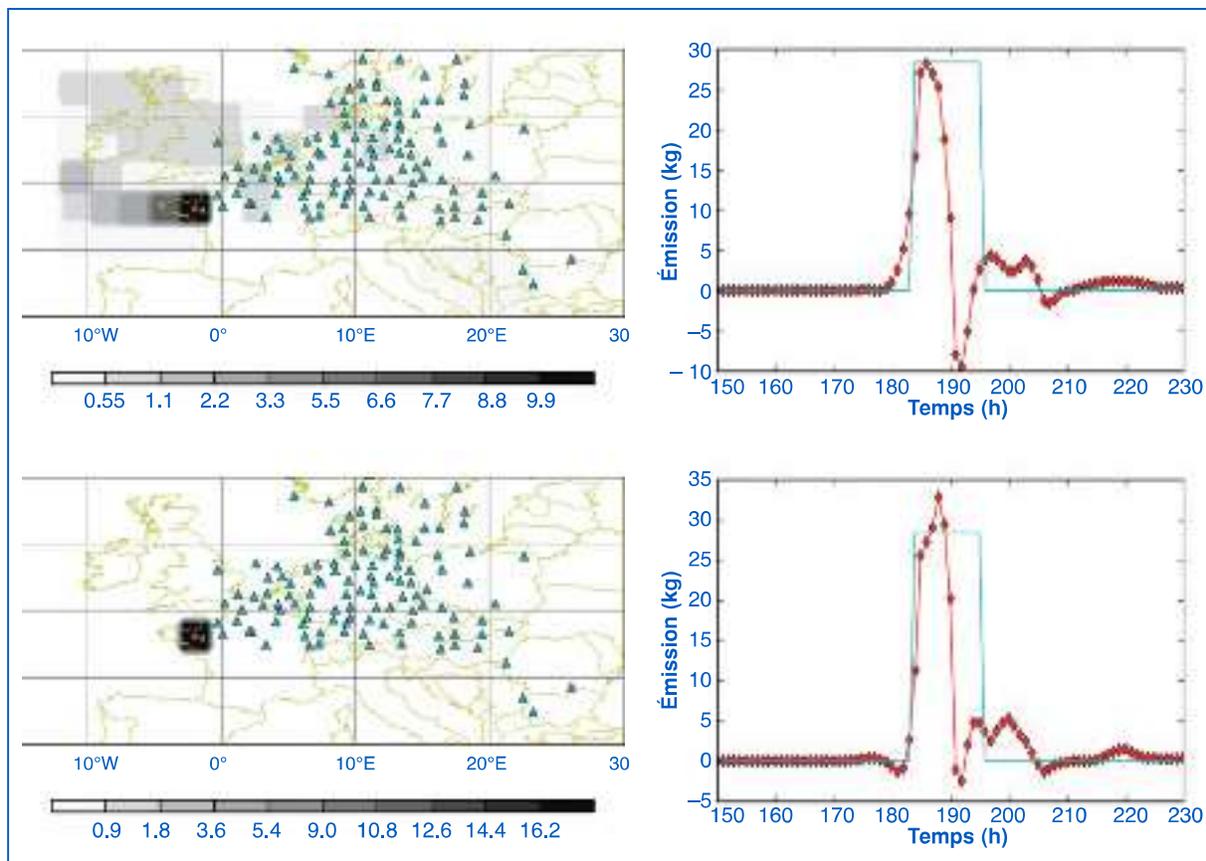


Figure 1.

Reconstruction du terme source d'ETEX-I à partir des mesures réelles. La première ligne correspond à une inversion reposant sur une ébauche gaussienne, alors que la seconde ligne détaille une inversion reposant sur une ébauche non-gaussienne. La première colonne visualise la distribution spatiale de la masse émise reconstruite, intégrée en temps. La seconde colonne compare le profil temporel de l'émission exacte (en bleu) au profil reconstruit (en rouge) dans la maille où se situe le rejet du polluant (perfluorométhylcyclohexane).

Reconstruction of the ETEX-I source term using real measurements. The first row corresponds to an inversion based on a Gaussian background, whereas the second row corresponds to an inversion based on a non-Gaussian background. The first column displays the spatial distribution of the emitted tracer mass, integrated on time. The second column compares the true temporal profile (in blue) to the retrieved profile (in red) in the grid-cell where the pollutant (perfluoromethylcyclohexane) was released.

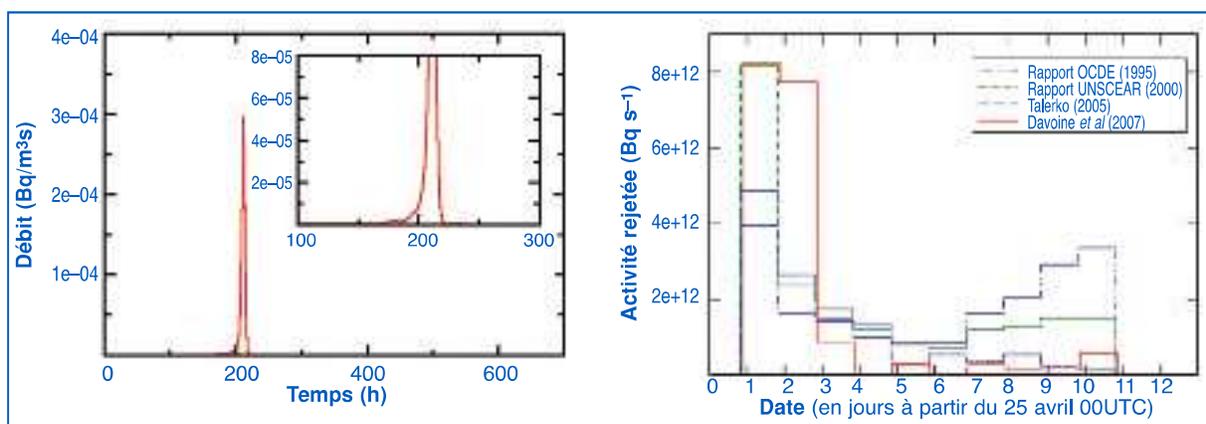


Figure 2.

À gauche : reconstruction du profil temporel du terme source du rejet accidentel d'Algésiras à l'aide d'une méthode non-Gaussienne. Le rejet apparaît comme un pic qui n'est décalé que de quelques heures par rapport à la date estimée du rejet réel. À droite : profil journalier du terme source de Tchernobyl reconstruit, pour l'iode 131 (en rouge) et comparaison aux estimations antérieures.

On the left: retrieved temporal profile of the Algeciras source term thanks to a non-Gaussian approach. The release appears as a peak which is shifted in time by only a few hours compared to the estimated true release date. On the right: daily profile of the retrieved Chernobyl source term for iodine 131 (in red), and comparison with previous estimations.

5.3 Tchernobyl

Le 25 avril 1986 à 21 h 23 UTC, a débuté à Tchernobyl en Ukraine, la plus grande catastrophe de l'histoire du nucléaire civil. L'un des enjeux scientifiques dès lors est de caractériser le terme source, c'est-à-dire d'estimer les radioéléments et les radionucléides qui ont été engendrés lors de l'accident. Des reconstructions du terme source ont été préalablement menées utilisant des inventaires de cœur, estimés par des modèles de cinétique des radioéléments, qui simulent la dynamique du cœur. Une autre source d'information est l'ensemble des

concentrations d'activité dans l'air mesurées en Europe (mais aussi en Asie et aux États-Unis), ainsi que les activités déposées, particulièrement en Ukraine et en Biélorussie. L'utilisation de ces données impliquait jusqu'alors le recours à des méthodes purement statistiques ou des allers et retours entre les données et des simulations de modèles dépendants de champs météorologiques.

Ces méthodes ont permis d'établir un consensus sur le terme source : celui-ci décrit la ventilation de la source en chacun des nombreux radionucléides qui la composent, avec une marge d'erreur importante. Nos reconstructions [27] ont mis en jeu des méthodes

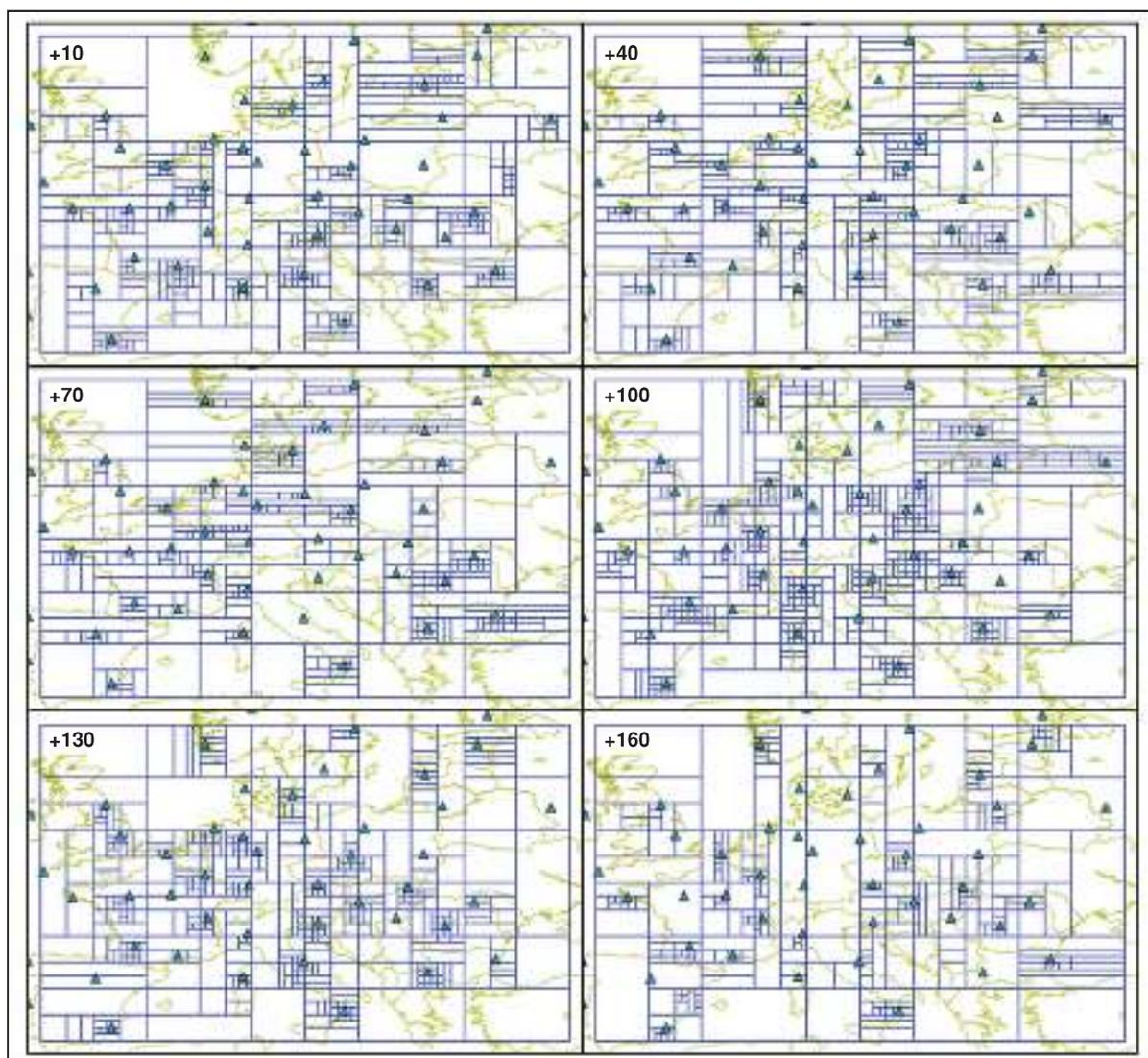


Figure 3.

Maillage optimal de l'espace de contrôle pour la reconstruction d'une source de pollution localisée ou non. Le réseau de stations (triangles) est un réseau fictif. Le maillage étant tridimensionnel, cette figure représente six coupes en temps de ce maillage. Plus le maillage est dense, plus la densité d'information qu'il est possible de retrouver est grande. Le maillage dépend non seulement de la proximité de stations mais aussi de la météorologie.

Optimal tiling of control space for the reconstruction of an accidental source of pollution which could be localised or not. The monitoring network (stations are indicated with triangles) is virtual. Since the adaptive grid is three-dimensional, this figure displays snapshots in time of the tiling. The denser the grid, the higher the information density that could be retrieved through inverse modelling. The adaptive grid depends not only on the proximity of monitoring stations, but also on the meteorology.

inverses non-gaussiennes mais aussi des techniques d'estimation de paramètres (courbe en L). Cela a permis de reconstruire le terme source pour les radionucléides iode 131, césium 137 et césium 134. L'estimation de l'activité totale relâchée pour ces trois espèces est proche des estimations officielles (10 %, 20 % et 25 % respectivement). Le rejet est cependant plus accentué les deux premiers jours de l'accident et moins les derniers jours de rejet. Le profil journalier du terme source reconstruit est donné sur la Figure 2 et comparé aux estimations antérieures obtenues de manière conventionnelle.

6. Impact de la résolution sur la modélisation inverse

6.1. Forte dépendance de l'estimation dans la résolution

Dans le cas non-paramétrique, l'espace de contrôle est discrétisé. Il y a donc un choix de résolution spatiale (et temporelle) arbitraire à faire. Or il a été montré que le résultat de l'inversion est très sensible à ce choix [28]. Lorsque la résolution est accrue, la solution devient de plus en plus piquée sur les détecteurs. En effet la solution qui consiste à placer une petite quantité de traceur en amont du détecteur devient possible à haute résolution et est alors favorisée par les techniques de régularisation traditionnelle (le terme correspondant à l'ébauche dans la fonction coût). Mathématiquement, on peut montrer que plus la résolution croît, plus les observations ne servent à résoudre que les flux de plus en plus proches des récepteurs.

6.2. Maillage adaptatif

Afin de contrecarrer la forte dispersion de l'information associée à une mesure ponctuelle, il est possible de construire des maillages adaptatifs qui partitionnent l'espace de contrôle de manière optimale [29]. Considérons le cas d'un réseau de surveillance de radionucléides d'une cinquantaine de stations utilisé dans [12]. L'espace de contrôle original est tridimensionnel et comporte $164 \times 98 \times 168$ mailles soit 2,7 millions de mailles. La Figure 3 décrit un maillage adaptatif optimal comportant seulement 17 287 mailles (soit 0,6 % du nombre initial de mailles). Cependant le nombre de degrés de liberté sur le signal qu'il est possible d'obtenir pour ce réseau est de 737, soit 99 % des 745 degrés de liberté pour le signal que l'on pourrait obtenir en utilisant le maillage le plus fin. L'utilisation d'un tel maillage optimal permettra un gain de temps considérable pour l'inversion, tout en conservant une capacité à extraire l'information quasi intacte.

7. Conclusions

Qu'elles soient paramétriques ou non-paramétriques, les méthodes de modélisation inverse fondées sur les techniques de l'assimilation de données se sont révélées capables, dans un cadre académique, de reconstruire des sources de polluant dispersé accidentellement, y compris sur des cas réels. Les techniques décrites dans cette revue en attestent. Depuis lors de nouveaux enjeux sont apparus.

D'abord, l'information étant rapidement dispersée et les erreurs sur les modèles et les erreurs de représentativité demeurant très importantes, il est essentiel que les observations soient pertinentes et le plus informatives possibles. Il nous semble donc que les réseaux d'observation devront donc être à l'avenir optimisés pour garantir la pertinence des observations acquises. Dans [30], les auteurs ont ainsi construit un réseau optimal de stations de mesure d'aérosols nucléaires pour la surveillance du parc nucléaire civil français.

En termes de modélisation, la caractérisation des erreurs, souvent mises en avant en assimilation de données, reste cruciale, particulièrement dans les situations de dispersion accidentelle. Il est donc toujours nécessaire d'estimer aussi bien la source et le panache que les incertitudes associées qui peuvent aisément être de l'ordre de grandeur du terme source, ou des concentrations dans le panache.

D'autre part, ces techniques doivent être utilisées dans les centres opérationnels de prévision. Les techniques devront alors être simples à mettre en œuvre. Elles devront être prouvées robustes, ce qui est difficile à garantir car, par définition, chaque situation accidentelle est unique et peut représenter une configuration non ou mal représentée dans l'ébauche du système d'inférence bayésien. Plus le support de l'ébauche sera circonscrit (méthodes paramétriques), plus la méthode sera *a priori* jugée efficace, mais plus grand sera le risque d'un événement non anticipé.

Remerciements

Cet article est une contribution au projet de l'INSU, LEFE-ASSIM2007-21 : « Application de méthodes avancées de modélisation inverse/assimilation de données à la dispersion accidentelle de polluants en situation d'urgence ». Les résultats concernant les maillages adaptatifs ont été obtenus dans le cadre du projet ANR MSDAG : « Assimilation de données multiéchelles en géophysique » de référence ANR-08-SYSC-014. L'auteur remercie Victor Winiarek pour une relecture attentive du manuscrit, ainsi que Monika Krysta et Xavier Davoine, qui ont collaboré à certains des travaux présentés dans cet article.

References

- [1] Hanna SR, Chang JR, Fernau ME. Monte Carlo Estimates of Uncertainties in Predictions by a Photochemical Grid Model (UAM-IV) Due to Uncertainties in Input Variables. *Atmos. Env.* 1998 ; 32 (21) : 3619-28.
- [2] Enting IG. Inverse Problems in Atmospheric Constituent Transport. Cambridge Atmospheric and Space Science Series. Dessler AJ, Houghton JH, Rycroft MJ Eds. 2002.
- [3] Quélo D, Mallet V, Sportisse B. Inverse modelling of NO_x emissions at regional scale over Northern France. Preliminary investigation of the second-order sensitivity. *J. Geophys. Res.* 2005 ; 110(D24310).
- [4] Elbern H, Strunk A, Schmidt H, Talagrand O. Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmos. Chem. Phys.* 2007 ; 7 : 3749-69.
- [5] Gurney KR, Law RM, Denning AS, Rayner PJ, Baker D, Bousquet P, Bruhwiler L, Chen YH, Ciais P, Fan S, Fung IY, Gloor M, Heimann M, Higuchi K, John J, Maki T, Maksyutov S, Masarie K, Peylin P, Prather M, Pak BC, Randerson J, Sarmiento J, Taguchi S, Takahashi T, Yuen CW. Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models. *Nature* 2002 ; 415 : 626-30.
- [6] Bousquet P, Ciais P, Peylin P, Ramonet M, Monfray P. Inverse modeling of annual atmospheric CO₂ sources and sinks. 1. Method and control inversion. *J. Geophys. Res.* 1999 ; 104 (D21) : 26161-78.
- [7] Roustan Y, Bocquet M. Inverse modelling for mercury over Europe. *Atmos. Chemis. Phys.* 2006 ; 6 : 3085-98.
- [8] Paatero P, Tapper U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 1994 ; 5 (2) : 111-26.
- [9] Tarantola A. Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, 2005.
- [10] Delle Monache L, Lundquist JK, Kosovic 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 : 2600-13.
- [11] 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 : 1805-16.
- [12] Krysta M, Bocquet M. Source reconstruction of an accidental radionuclide release at European scale. *Q.J.R. Meteorol.Soc.* 2007 ; 133 : 529-44.
- [13] Talagrand O. Assimilation of Observations, an Introduction. *Journal of the Meteorological Society of Japan* 1997 ; 75 : 191-209.
- [14] Issartel JP, Baverel J. Inverse transport for the verification of the Comprehensive Nuclear Test Ban Treaty. *Atmos. Chem. Phys.* 2003 ; 3 : 475-86.
- [15] Bocquet M. Inverse modelling of atmospheric tracers: Non-Gaussian methods and second-order sensitivity analysis. *Nonlin. Processes Geophys.* 2008 ; 15 : 127-43.
- [16] Bocquet M. Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Theory. *Q.J.R. Meteorol.Soc.* 2005 ; 131 : 2191-208.
- [17] Bocquet M. Reconstruction of an atmospheric tracer source using the principle of maximum entropy. II: Applications. *Q.J.R. Meteorol.Soc.* 2005 ; 131 : 2209-24.
- [18] Robertson L, Persson C. Attempts to apply four dimensional data assimilation of radiological data using the adjoint technique. *Radiation Protection Dosimetry* 1993 ; 50 : 333-7.
- [19] Robertson L, Langner J. Source function estimate by means of adjoint variational data assimilation applied to the ETEX-I tracer experiment. *Atmos. Environ.* 1998 ; 32 (24) : 4219-25.
- [20] Pudykiewicz JA. Application of adjoint transport tracer equations for evaluating source parameters. *Atmospheric Environment* 1998 ; 32 : 3039-50.
- [21] Uliasz M. Application of the Perturbation Theory to the Sensitivity Analysis of an Air Pollution Model. *Z. Meteor.* 1983 ; 33 (6) : 355-62.
- [22] Hourdin F, Issartel JP. Sub-surface nuclear tests monitoring through the CTBT xenon network. *Geophys. Res. Lett.* 2000 ; 27 : 2245-8.
- [23] Seibert P, Stohl A. Inverse modelling of the ETEX-I release with a Lagrangian particle model, 2000. Dans Proceedings of the Third GLOREAM Workshop, Ischia, Italy, September 1999.
- [24] Issartel JP. Rebuilding sources of linear tracers after atmospheric concentration measurements. *Atmos. Chem. Phys.* 2003 ; 3 : 2111-25.
- [25] Bocquet M. High resolution reconstruction of a tracer dispersion event: application to ETEX. *Q. J. R. Meteorol. Soc.* 2007 ; 133 : 1013-26.
- [26] Krysta M, Bocquet M, Brandt J. Probing ETEX-II data set with inverse modelling. *Atmos. Chem. Phys.* 2008 ; 8 : 3963-71.
- [27] Davoine X, Bocquet M. Inverse modelling-based reconstruction of the Chernobyl source term available for long-range transport. *Atmos. Chemis. Phys.* 2007 ; 7 : 1549-64.
- [28] Bocquet M. Grid resolution dependence in the reconstruction of an atmospheric tracer source. *Nonlin. Processes Geophys.* 2005 ; 12 : 219-34.
- [29] Bocquet M. Towards optimal choices of control space representation for geophysical data assimilation. *Mon. Wea. Rev.* 2009 ; 137 : 2331-48.
- [30] Abida R, Bocquet M, Vercauteren N, Isnard O. Design of a monitoring network over France in a case of radiological accidental release. *Atmos. Env.* 2008 ; 42 : 5205-19.
- [31] Bocquet M, Sportisse B. Modélisation inverse pour la qualité de l'air : éléments de méthodologie et exemples. *Pollution Atmosphérique* 2007 ; 196 : 395-404.