

Pollution de milieux poreux : identifiabilité et identification de modèles paramétriques de sources

Aurélien HAZART^{1,2}, Jean-François GIOVANNELLI², Stéphanie DUBOST¹, Laurence CHATELLIER¹

¹Electricité De France - R&D
6, quai Watier, 78401 Chatou Cedex, France

²Laboratoire des Signaux et Systèmes (CNRS-Supélec-UPS)
Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France
aurelien.hazart@edf.fr, giova@lss.supelec.fr
stephanie.dubost@edf.fr, laurence.chatellier@edf.fr

Résumé – Cette communication concerne le problème d’identification de sources de pollution à partir d’un faible nombre de mesures de concentrations. Il s’agit d’un problème inverse. Plusieurs facteurs le rendant fortement mal posé, nous étudions l’identifiabilité et la stabilité de sa solution avant d’inverser. Pour cela, la source est modélisée par une fonction paramétrique de type impulsionnelle. Notre objectif est de caractériser l’incertitude sur l’estimation des paramètres de la source.

Abstract – This communication deals with the problem of identification of pollution sources based on few concentration measurements. As several factors make the problem ill-posed, we study the identifiability and the stability of its solution before inversion. The source is modeled by a parametric function. Our goal is to characterize uncertainty on the estimate of the source parameters.

1 Introduction

Dans le cadre de la maîtrise de l’impact environnemental de ses installations de production, EDF a développé de nombreux moyens de mesure et de calcul permettant la détection et l’évaluation d’une éventuelle pollution. Dans le cas supposé d’une pollution industrielle, les phénomènes physiques qui interviennent dans le transport de polluant sont modélisés par une équation aux dérivées partielles à coefficients constants (appelée équation de *convection-dispersion*). Pour un milieu bidimensionnel infini sans profondeur, la concentration $c(x, y, t)$ et la source sous la forme de son débit $f(x, y, t)$ vérifient l’équation:

$$\frac{\partial c}{\partial t} + v \frac{\partial c}{\partial x} - D_1 \frac{\partial^2 c}{\partial x^2} - D_t \frac{\partial^2 c}{\partial y^2} = f(x, y, t) \quad (1)$$

où v, D_1, D_t désignent les constantes physiques (respectivement la vitesse de convection, les coefficients de dispersion longitudinal et transverse). La solution du système de transport avec des conditions initiales et aux limites nulles s’exprime comme un produit de convolution

$$c = h * f \quad (2)$$

où h est la réponse impulsionnelle du système donnée par [1]:

$$h(x, y, t) = \frac{1}{4\pi t \sqrt{D_1 D_t}} \exp\left(-\frac{[x - vt]^2}{4D_1 t} - \frac{y^2}{4D_t t}\right). \quad (3)$$

La concentration n’est mesurée en réalité qu’en un nombre limité de points P correspondant au nombre de capteurs disponibles, localisés en (x_p, y_p) . Chaque capteur effectue N_p mesures temporelles conduisant aux concentrations effectivement mesurées c_{p,n_p} . L’enjeu est donc d’estimer la fonction f à partir du système de transport (2) et des concentrations mesurées.

La difficulté du problème inverse résulte des trois aspects suivants :

- le système convolutif (2) se comporte comme un filtre passe-bas qui retire une partie de l’information haute fréquence contenue dans la fonction source f ,
- le système de mesures effectue un sous-échantillonnage spatial et temporel des données,
- les incertitudes sur les mesures (erreurs de mesures ou erreurs de modélisation) peuvent être importantes.

Le problème inverse est mal posé : l’existence et l’unicité de solutions ne sont pas garanties. L’information sur la source contenue dans les données n’étant pas suffisante pour inverser, le problème est «régularisé» par la prise en compte d’informations supplémentaires. Il s’agit de restreindre l’espace des solutions afin de rendre la solution unique. On recense globalement deux approches de régularisation.

- *L’approche non-paramétrique pénalisée* consiste à rechercher la solution qui respecte certaines conditions (continuité,...). [2] et [3] proposent la régularisation de Tikhonov.
- *L’approche paramétrique* consiste à modéliser la distribution temporelle par une fonction paramétrique : l’impulsion [4], l’échelon [5], la somme d’impulsions ou la somme de gaussiennes [6], une fonction quelconque supposée connue [7].

Pour identifier une source de pollution, toutes ces méthodes prennent en compte le caractère mal posé du problème, mais sans en étudier les caractéristiques. Notre objectif est ici de caractériser l’identifiabilité du problème inverse. Pour cela, la seconde classe de méthodes est intéressante car elle permet d’étudier séparément l’influence de chacun des paramètres de la

source. La spécificité de l'étude réside dans la prise en compte d'un faible nombre de données.

Dans ce document, l'intérêt est plus particulièrement porté sur la paramétrisation de type impulsionnelle. Le terme source f de l'équation (1) prend la forme paramétrique

$$f(x, y, t) = q_0 \delta(x - x_0) \delta(y - y_0) \delta(t - t_0) \quad (4)$$

où q_0 représente la quantité de polluant injectée, (x_0, y_0) la localisation de la source et t_0 l'instant d'émission. Une telle source correspond à un éventuel scénario de rejet instantané et entraîne le panache de pollution de concentration

$$c(x, y, t) = q_0 h(x - x_0, y - y_0, t - t_0). \quad (5)$$

Pour la suite, en posant $M_{p,N_p}(x_0, y_0, t_0) = h(x_p - x_0, y_p - y_0, t_p - t_0)$, les concentrations mesurées sont exprimées sous forme discrète :

$$c_{p,n_p} = q_0 M_{p,n_p}(x_0, y_0, t_0) \quad (6)$$

pour $p = 1 \dots P$, pour $n_p = 1 \dots N_p$. Dans le cas expérimental traité à la section suivante, deux puits ($P = 2$) effectuant une mesure par jour pendant 28 jours ($N_1 = N_2 = 28$) sont pris en compte.

2 Estimations et incertitudes

Dans le but d'identifier la source impulsionnelle à partir des données, notre approche repose sur le critère des moindres carrés :

$$J(q_0, x_0, y_0, t_0) = \sum_{p=1}^P \sum_{n_p=1}^{N_p} [c_{p,n_p} - q_0 M_{p,n_p}(x_0, y_0, t_0)]^2. \quad (7)$$

Même en présence de données bruitées, la minimisation de ce critère peut être suffisante pour estimer les paramètres de la source (et donc résoudre le problème inverse). La condition est que le minimum global du critère soit unique, le modèle paramétrique est alors identifiable. Cependant deux facteurs peuvent perturber l'estimation : l'incertitude sur l'estimation peut être grande si le critère est plat autour de son minimum et la présence de minima locaux peut tromper un algorithme de minimisation. Dans [5], des minima locaux sont constatés dans le cas d'une source indicielle et une technique d'obtention du minimum global du critère est proposée. L'objectif n'étant pas ici de développer une méthode d'identification, le critère (7) est simplement calculé sur une grille de valeurs des paramètres $[q_0, x_0, y_0, t_0]$. La visualisation du critère est alors utilisée comme un premier élément pertinent pour caractériser les incertitudes sur l'estimation des paramètres du modèle.

2.1 Analyse de données expérimentales

Les mesures expérimentales dont nous disposons, présentées à la figure 1, sont effectuées par deux capteurs alignés dans le sens d'écoulement x du milieu et distants de 55 mètres. Les coordonnées du capteur 1 sont fixées à $(x, y) = (0, 0)$, le capteur 2 est localisé en $(55, 0)$. Enfin l'instant de la première mesure est choisie comme origine de l'axe des temps.

Le tableau 1 indique les paramètres physiques du modèle de transport considéré. Dans ce cas expérimental de pollution, la

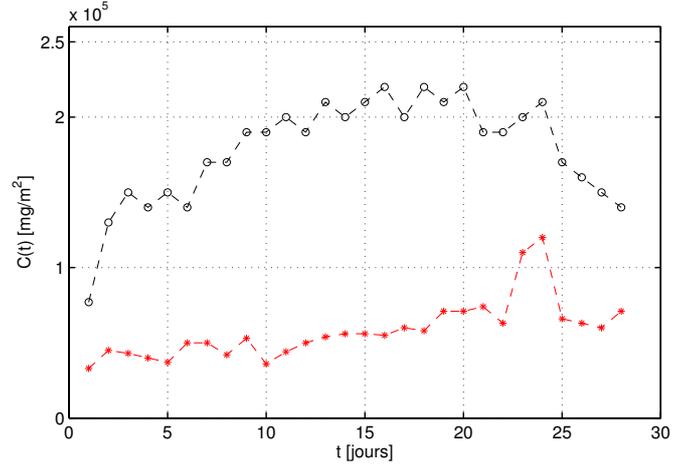


FIG. 1: Données expérimentales : capteur 1 (o), capteur 2(*).

source n'est pas connue et le modèle impulsionnel n'est pas nécessairement le meilleur modèle. Nous sommes donc en présence de données bruitées (erreurs de mesures, erreurs de modélisation). L'intérêt de l'approche réside dans l'estimation des paramètres de la source impulsionnelle mais surtout dans l'incertitude sur cette estimation. Les paramètres estimés, minimisant le critère (7), sont indiqués au tableau 2.

TAB. 1: constantes physiques du modèle de transport

v	D_1	D_t
2.3 m/j	11 m ² /j	2 m ² /j

TAB. 2: paramètres estimés de la source impulsionnelle.

\hat{q}	\hat{x}	\hat{y}	\hat{t}
10 ⁹ mg	-210 m	0 m	-69 j

Deux tracés du critère autour de son minimum sont représentés.

- La figure 2 montre les coupes du critère en fonction de chacun des paramètres, les autres paramètres étant fixés à leur valeur estimée. On observe que le minimum est toujours bien défini.
- La figure 3 représentant le critère de moindres carrés en fonction des paramètres x_0 et t_0 (q_0 et y_0 sont fixés à leur valeurs estimées) montre en revanche que ces deux paramètres sont liés : le critère comporte une direction privilégiée pour laquelle la valeur du critère est relativement constante. Bien que le minimum du critère soit unique, l'estimation conjointe de la localisation dans le sens d'écoulement et de l'instant d'émission est incertaine.

2.2 Analyse de données simulées

Afin de confirmer les résultats obtenus sur des données expérimentales, l'approche est appliquée de manière similaire sur un jeu de données simulées à partir d'une source impulsionnelle connue. Pour la simulation, la source a pour paramètres

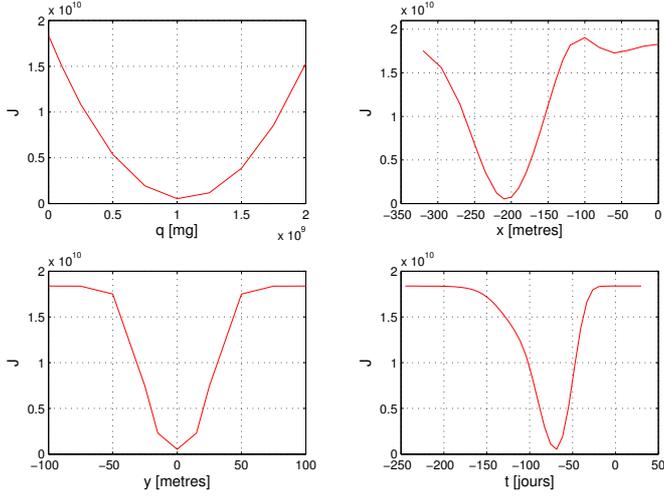


FIG. 2: Coupes monodimensionnelle du critère des moindres carrés. Les trois paramètres restants sont fixés aux valeurs estimées (tableau 2).

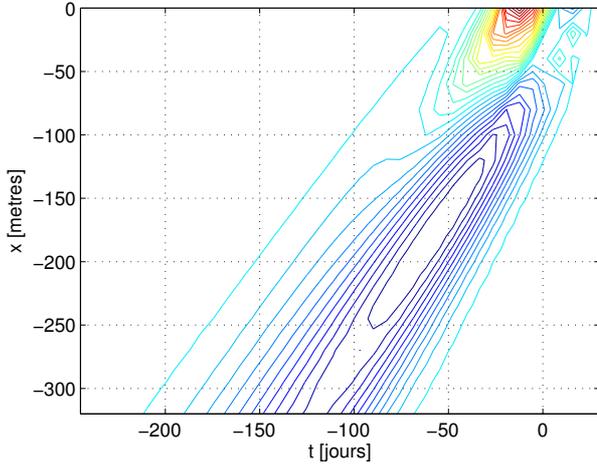


FIG. 3: Critère de moindres carrés dans le plan des paramètres xt . Les paramètres q et y sont fixés aux valeurs estimées (tableau 2).

$(x_0, y_0, t_0, q_0) = (0, 0, 0, 10^{10})$. La figure 4 représente les concentrations mesurées au cours du temps par trois capteurs localisés en $(x_c, y_c) = (120, 10), (90, 30), (80, -40)$. Parmi ces valeurs, il est supposé que chaque capteur effectue deux mesures temporelles en $t_c = 40$ jours et $t_c = 60$ jours, marquées par les cercles sur la figure 4, constituant un jeu de six données.

Dans ce cas, la minimisation du critère $J(q, x, y, t)$ sur une grille de valeurs des paramètres q, x, y, t conduit à l'estimation exacte des paramètres de la source : les données sélectionnées caractérisent la source de manière unique. Comme dans le cas expérimental présenté ci-dessus, l'étude du critère paramètre par paramètre (figure 5) puis dans le plan xt (figure 6) indique :

- un minimum unique et bien défini pour chaque estimée,
- une direction privilégiée dans l'espace xt , selon laquelle les valeurs minimales du critères sont proches.

L'objectif du paragraphe suivant est de caractériser analytiquement l'incertitude de l'estimation des paramètres x et t .

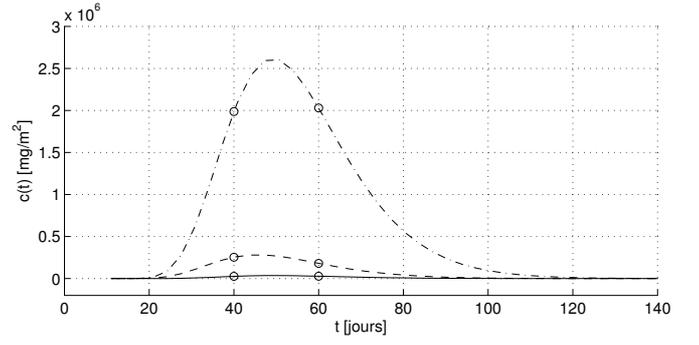


FIG. 4: Données simulées aux capteurs localisés en $(120, 10)$ (—), $(90, 30)$ (---), $(80, -40)$ (-). Les cercles indiquent les mesures retenues aux instants 40 et 60 jours.

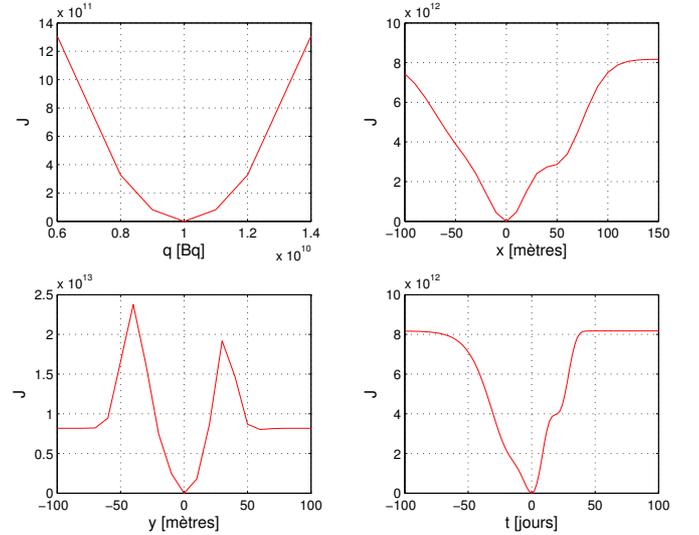


FIG. 5: Coupes monodimensionnelle du critère des moindres carrés. Les trois paramètres restant sont fixés aux valeurs estimées ($\hat{q} = 10^{10}, \hat{x} = 0, \hat{y} = 0, \hat{t} = 0$).

2.3 Caractérisation des indéterminations

Pour expliquer les incertitudes sur les paramètres minimisant le critère, l'identifiabilité est étudiée du point de vue analytique [8]. L'approche consiste à caractériser l'unicité du modèle paramétrique en cherchant l'ensemble des paramètres (q, x, y, t) tels que pour $p = 1 \dots P$, pour $n_p = 1 \dots N_p$

$$q_0 M_{p, n_p}(x_0, y_0, t_0) = q M_{p, n_p}(x, y, t) \quad (8)$$

où q_0, x_0, y_0, t_0 sont les paramètres de la «vraie» source. Cette analyse est possible dans le cas sous-déterminé où une seule donnée est disponible. La position du capteur et l'instant de mesure sont notés respectivement x_c, y_c et t_c . Avec l'expression (3) du modèle de transport, le quadruplet de paramètres (q, x, y, t) permettant de générer la même mesure vérifie l'équation :

$$\frac{q}{(t_c - t)4\pi\sqrt{D_1 D_t}} \exp\left(-\frac{[(x_c - x) - v(t_c - t)]^2}{4D_1(t_c - t)} - \frac{(y_c - y)^2}{4D_t(t_c - t)}\right) = \text{cst.} \quad (9)$$

Afin d'étudier la relation entre les paramètres x et t , l'équation (9) est considérée à q et y fixés (à leur vraie valeur q_0 et y_0) et

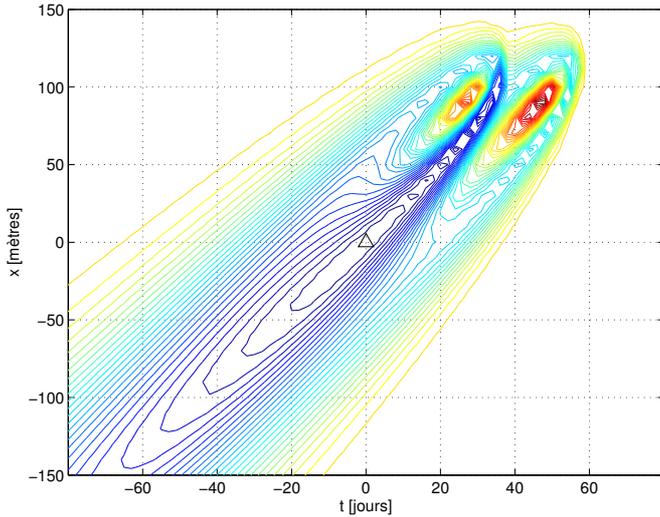


FIG. 6: Contour du critère de moindres carrés dans le plan xt . Les paramètres q et y sont fixés aux valeurs estimées ($\hat{q} = 10^{10}$, $\hat{y} = 0$).

réécrite sous la forme

$$\alpha(x_c - x)^2 + \beta(t_c - t)^2 + \gamma(x_c - x)(t_c - t) + t \ln t + \omega = 0 \quad (10)$$

où les coefficients sont des réels qui dépendent des paramètres physiques et de q , y . La corrélation entre les paramètres x et t est contenue dans le terme $\gamma(x_c - x)(t_c - t)$. Il ne s'agit pas de l'équation d'une conique en raison de la présence du terme en $\ln t$. La figure 7 représente cette équation pour la donnée simulée au capteur localisé en $(120, 10)$ et à l'instant 40 à partir de la source définie au paragraphe précédent $(x_0, y_0, t_0, q_0) = (0, 0, 0, 10^{10})$. Les paramètres q et y sont fixés aux vraies valeurs q_0 et y_0 . On observe que les indéterminations du couple (x, t) sont réparties autour d'un axe dont l'équation est proche de la droite $x = vt + \text{cst}$. L'étude des indéterminations dans le cas d'une seule donnée permet donc de comprendre l'allure des critères obtenus sur les données expérimentales et simulées (figures 3 et 6). Lorsque le nombre de données est supérieur, l'indétermination stricte disparaît mais le critère des moindres carrés conserve une direction $x = vt$ pour les ellipses. Le phénomène de convection transportant la pollution supposée à la vitesse v est à l'origine de l'incertitude sur l'estimation conjointe de la localisation dans le sens d'écoulement et de l'instant d'injection. Dans un milieu où la dispersion est négligeable et où seule la convection intervient, cette incertitude devient une indétermination quelque soit le nombre de données : les sources dont les paramètres vérifient l'équation $x = vt + \text{cst}$ engendrent les mêmes concentrations.

3 Conclusion

L'identifiabilité de sources de pollution impulsionnelle est étudiée, d'abord à partir de données expérimentales et simulées puis de manière analytique, en considérant un faible nombre de données. L'allure des critères de moindres carrés obtenus sur des données expérimentales est confirmée et justifiée par l'étude analytique dans le cas où une seule mesure est disponible. Les résultats présentés montrent que la source impulsionnelle est généralement déterminée de manière unique mais

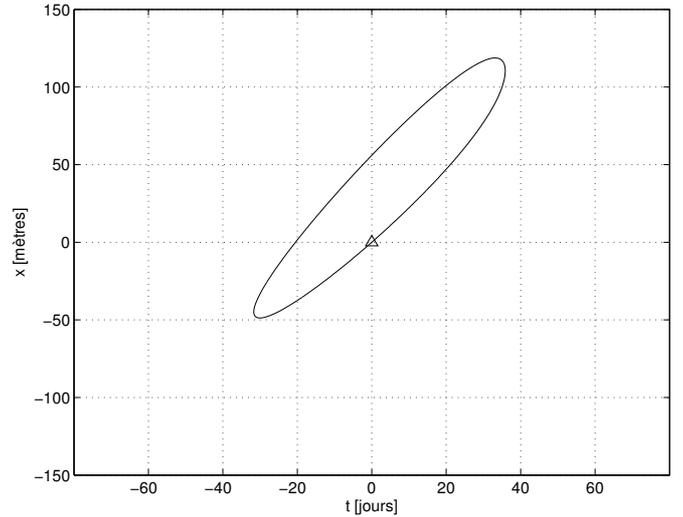


FIG. 7: Indétermination du couple de paramètres (x, t) de la source impulsionnelle pour la mesure en $(x_c, y_c) = (120, 10)$ à l'instant $t_c = 40$. q et y sont fixés à leur vraie valeur q_0 et y_0 . L'ellipse passe par la vraie source $(x_0, t_0) = (0, 0)$.

que l'estimation de ses paramètres est perturbée par une incertitude importante sur l'estimation conjointe de la localisation dans le sens d'écoulement et de l'instant d'émission. Pour la suite, nous envisageons de quantifier l'incertitude et d'estimer les paramètres de la source en nous basant sur le calcul du gradient et du hessien du critère.

Références

- [1] D. Bleecker and G. Csordas. *Basic partial differential equations*. International Press, 1996.
- [2] Todd H. Skaggs and Z. J. Kabala. Recovering the history of a groundwater contaminant plume. *Water Resource Research*, 30(1):71–79, november 1994.
- [3] P. Kathirgamanathan, R. McKibbin, and R.I. McLachlan. Source term estimation of pollution from an instantaneous point source. *Res. Lett. Inf. Math. Sci.*, 3:59–67, 2002.
- [4] P. Sidauruk, A.H-D. Cheng, and D. Ouazar. Ground water contaminant source and transport parameter identification by correlation coefficient optimization. *Ground Water*, 36(2):208–214, Mars-Avril 1998.
- [5] Jörg Matthes, Lutz Gröll, and Hubert B. Keller. Source localization based on pointwise concentration measurements. *Sensors and Actuator A (Physical)*, 115:32–37, 2004.
- [6] S. Alapati and Z. J. Kabala. Recovering the release history of a groundwater contaminant using a non-linear least-squares method. *Hydrological Processes*, 14(1003):1016, 2000.
- [7] Mehmet E. Alpay and Molly H. Shor. Model-based solution techniques for the source localization problem. *IEEE Transactions on Control Systems Technology*, 8(6):895–904, Novembre 2000.
- [8] Eric Walter and Luc Pronzato. *Identification de modèles paramétriques à partir de données expérimentales*. Masson, Paris, FRANCE, 1994.