

# Estimation du Temps de Résidence Hydrologique: Déconvolution 1D

Alina G. MERESESCU<sup>1,2</sup>, Matthieu KOWALSKI<sup>2</sup>, Frédéric SCHMIDT<sup>1</sup>, François LANDAIS<sup>1</sup>

<sup>1</sup>GEOPS, Univ. Paris-Sud, CNRS, Université Paris-Saclay, Rue du Belvédère, Bât. 504-509, 91405 Orsay, France

<sup>2</sup>L2S, Univ. Paris-Sud, CentraleSupélec, CNRS, Université Paris-Saclay, 3 rue Joliot Curie, 91192 Gif-sur-Yvette, France

alina-georgiana.meresescu@u-psud.fr, matthieu.kowalski@u-psud.fr,

frederic.schmidt@u-psud.fr, francois.landais@u-psud.fr

**Résumé** – Le temps de résidence hydrologique est équivalent à la réponse impulsionnelle du système linéaire de la propagation de l'eau dans un milieu, par exemple dans un massif montagneux, de la pluie en surface jusqu'à l'aquifère en profondeur. Pour son estimation, nous proposons une déconvolution 1D avec des contraintes de positivité et causalité. Comparée aux méthodes usuelles du domaine, basées sur des modèles paramétriques ou la corrélation croisée, nous obtenons des estimations plus précises, de la courbe de temps de résidence hydrologique, même dans le cas où les mesures sont bruitées.

**Abstract** – The water residence time is equivalent to the impulse response of a linear system allowing the propagation of water through a channel, for example the propagation of rain water from the top of the mountain to its base towards the aquifers. For its estimation we propose a 1D deconvolution algorithm with positivity and causality constraints. Compared to the usual methods based on parametric modeling and cross-correlation, our method delivers a more precise water residence time distribution curve, even in the case of noisy measurements.

## 1 Introduction

Le temps de résidence hydrologique est une mesure permettant l'analyse du transit de l'eau à travers un milieu donné. Son estimation est nécessaire pour connaître la composition hydro-bio-géochimique d'une montagne [5]; pour mieux gérer et protéger les sources d'eau potable de la pollution [8]; pour utiliser des zones humides comme dépolluants des eaux [12]; ou pour étudier le transport de nutriments dissous dans l'eau [4].

Le temps de résidence de l'eau peut être modélisé comme la réponse impulsionnelle du système; et peut être estimé par déconvolution [6]. Pour obtenir ce temps de résidence de l'eau, on peut distinguer deux familles de techniques: actives ou passives. Les techniques actives sont réalisées en lançant à l'entrée du système des traceurs à un instant donné, comme des colorants artificiels, puis en mesurant ces traceurs à la sortie du système [10, 12]. Bien que ces méthodes soient très coûteuses en raison des moyens exigés sur le terrain. Les techniques passives consistent à enregistrer des données à l'entrée et à la sortie du canal comme des isotopes particuliers de l'eau [5], la conductivité électrique [8] ou simplement enregistrer le signal de pluie en surface et le niveau moyen de la nappe phréatique [3]. Les méthodes de déconvolution alors utilisées peuvent être paramétriques [5, 7, 12] ou non-paramétriques [3, 4, 6, 8].

Notre méthode fait partie de la deuxième catégorie tout

en prenant en compte certaines limitations des méthodes précédentes. Par rapport à [6], nous étudions l'influence de la longueur des données sur la qualité d'estimation, ainsi que l'aspect non-stationnaire du canal [5, 10]. L'algorithme est issu de la minimisation d'une fonctionnelle sous contrainte, là où deux fonctionnelles sont nécessaires dans [6]. On utilise des mesures strictement passives comme les mesures de la pluie et le niveau d'aquifères, alors que [4] présente une estimation par Filtre Wiener et inversion géostatistique sur la base des mesures de NaCl largué dans le canal en tant que traceur. Enfin, on s'affranchit de l'hypothèse trop forte supposant que le signal de pluie est proche d'un bruit blanc [3], condition nécessaire pour utiliser la méthode de corrélation croisée. Par rapport à [3, 4, 6, 8] on prend en compte la positivité, l'aspect "lisse" ainsi que la causalité du temps de résidence de l'eau.

## 2 Modèle

### 2.1 Problème direct

Le modèle direct de la propagation de l'eau est le suivant :

$$\mathbf{y} = \mathbb{1}c + \mathbf{x} * \mathbf{k} + \mathbf{n} \quad (1)$$

avec :

- $\mathbf{y} \in \mathbb{R}_+^T$ ,  $\mathbf{y} = (y_1, \dots, y_T)$  la sortie du système (ex : le niveau de la nappe phréatique, qui est connu),  $\mathbb{R}_+^T$  signal réel positif de taille  $T$
- $\mathbb{1}$  un vecteur de 1 de taille  $T$
- $c \geq 0$  le niveau moyen de la nappe phréatique, qu'on doit estimer.
- $\mathbf{x} \in \mathbb{R}_+^T$ ,  $\mathbf{x} = (x_1, \dots, x_T)$  l'entrée du système (ex : les précipitations, qui sont connues).
- $*$  le produit de convolution
- $\mathbf{k} \in \mathbb{R}_+^{K+1}$ ,  $\mathbf{k} = (k_{-K/2}, \dots, k_{-1}, k_0, k_1, \dots, k_{K/2})$  la réponse impulsionnelle à estimer.
- $\mathbf{n} \in \mathbb{R}^T$  le bruit de mesure qu'on suppose blanc gaussien.

Le but est d'estimer le niveau moyen  $c$  de la nappe phréatique ainsi que la réponse impulsionnelle  $\mathbf{k}$  supposé positif, causal et lisse.

## 2.2 Problème inverse

Pour estimer  $\mathbf{k}$ , nous proposons de résoudre le problème d'optimisation sous contrainte suivant :

$$\hat{\mathbf{k}}, \hat{c} = \underset{\mathbf{k} \in \mathbb{R}_+^{K+1}, c}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{x} * \mathbf{k} - c\mathbb{1}\|_2^2 + \lambda \|\nabla \mathbf{k}\|_2^2 \quad (2)$$

$$s.t. \text{ causalité : } \forall i \in \{-K/2, \dots, -1\} k_i = 0$$

Cette fonctionnelle fait apparaître classiquement un terme d'attache aux données  $\ell_2$  correspondant à un bruit blanc gaussien, ainsi qu'un terme de régularisation  $\ell_2$  sur le gradient de  $\mathbf{k}$  afin de favoriser les solutions "lisses". Dans la suite, on ré-écrit la fonctionnelle (2) à l'aide d'opérateurs matriciels

$$J(\mathbf{k}, c) = \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{k} - c\mathbb{1}\|_2^2 + \lambda \|\mathbf{D}\mathbf{k}\|_2^2 \quad (3)$$

$$s.t. \forall i \in \{-K/2, \dots, -1\} k_i = 0 \text{ et } \forall i k_i \geq 0$$

où  $\mathbf{X}$  est la matrice circulaire correspondant à la convolution par le signal  $\mathbf{x}$  et  $\mathbf{D}$  la matrice aux différences finis correspondant au gradient.

Nous proposons d'estimer  $\mathbf{k}$  et  $c$  par un algorithme de minimisation alterné. À  $c$  fixé, le problème est une simple optimisation quadratique sous contrainte que l'on résout à l'aide d'un Newton Projetée [2]. À  $\mathbf{k}$  fixé, l'estimation de  $c$  est donnée par une formule analytique. Ce cas particulier de minimisation alternée convexe par rapport au couple et quadratique permet d'assurer la convergence du schéma de minimisation [1].

## 2.3 Algorithme de Minimisation Alternée pour la Deconvolution 1D

### 2.3.1 Itération de Newton Projetée

La mise à jour de  $\mathbf{k}$  à  $c$  fixé par la méthode de Newton projeté est donnée par :

$$\begin{aligned} \mathbf{k}_{t+1} &= P(\mathbf{k}_t + \alpha_t \cdot (-\nabla^2 J(\mathbf{k}_t, c))^{-1} \cdot \nabla J(\mathbf{k}_t, c)) \\ &= P((1 - \alpha_t)\mathbf{k}_t + \alpha_t \cdot (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \cdot \mathbf{X}^T (\mathbf{y} - c\mathbb{1})) \end{aligned} \quad (4)$$

où  $\alpha_t > 0$  est le pas descente et  $P$  la projection orthogonale sur les contraintes. Pour

$\mathbf{k} = \{k_{-K/2}, \dots, k_{-1}, k_0, k_1, \dots, k_{K/2}\}$ , on a  $P(\mathbf{k}) = \{0, \dots, 0, (k_0)^+, \dots, (k_{K/2})^+\}$ , où  $(x)^+ = \max(0, x)$ .

### 2.3.2 Estimation de $c$

À  $\mathbf{k}$  fixé, l'estimation de  $c$  est simplement donné par la moyenne empirique du résidu entre  $\mathbf{y}$  et  $\mathbf{X}\mathbf{k}$  :

$$c = \overline{\mathbf{y} - \mathbf{X}\mathbf{k}} \quad (5)$$

L'algorithme de minimisation alterné permettant d'estimer  $\mathbf{k}$  et  $c$  est présenté dans l'algorithme 1.

---

#### Algorithm 1 Minimisation Alternée

---

**Input :**  $\mathbf{x}, \mathbf{y}$ ,  $s_{max} = 1000$ ,  $t_{max} = 1000$

**Output :**  $\mathbf{k}_{est}, c_{est}, \mathbf{y}_{rec}$

Initialisation :

- 1:  $c = \bar{\mathbf{y}}, \tilde{\mathbf{y}} = \mathbf{y} - c$
  - 2:  $\mathbf{k}_{est} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \cdot \mathbf{X}^T \tilde{\mathbf{y}}$
  - 3:  $\alpha = 1, err_{rel} = 1$
  - 4: **while**  $s \neq s_{max}$  and  $err_{rel} > eps$  **do**
  - 5:  $J_{ref} = \frac{1}{2} \|\tilde{\mathbf{y}}\|_2^2, \mathbf{k}_{old} = \mathbf{k}_{est}$
  - 6: **while**  $t \neq t_{max}$  and  $\alpha > eps$  **do**
  - 7:  $\mathbf{k}_{est}$  est estimé par (4)
  - 8:  $J(k)^{t+1} = \frac{1}{2} \|\mathbf{y} - \mathbf{y}_{rec}\|_2^2$
  - 9: **if**  $(J(k)^{t+1} > J_{ref})$  **then**
  - 10:  $\mathbf{k}_{est} = \mathbf{k}_{old}, \alpha = 0.9 \cdot \alpha$
  - 11: **else**
  - 12:  $J_{ref} = J(k)^{t+1}$ , **break**;
  - 13: **end if**
  - 14:  $t = t + 1$
  - 15: **end while**
  - 16:  $\tilde{\mathbf{y}}_{rec} = \mathbf{x} * \mathbf{k}_{est}, c_{est} = \overline{\mathbf{y} - \tilde{\mathbf{y}}_{rec}}, \mathbf{y}_{rec} = \tilde{\mathbf{y}}_{rec} + c_{est}$
  - 17: recalculer  $err_{rel}, s = s + 1$
  - 18: **end while**
  - 19: **return**  $\mathbf{k}_{est}, \mathbf{y}_{rec}, c_{est}$
- 

## 3 Résultats sur données synthétiques

Dans le cadre d'une validation synthétique réaliste nous avons généré les signaux de pluie  $\mathbf{x}$  avec une simulation multi-fractales basée sur [11]. D'autre part  $\mathbf{k}$ , peut être modélisé avec une distribution Beta,  $B(x, \alpha, \beta)$  où  $\alpha = 2$  et  $\beta = 6$ . Nous avons choisi arbitrairement  $c = 100$ .

L'algorithme a été testé pour différentes valeurs de rapport signal à bruit (SNR) en entrée, sur une gamme  $\lambda$

choisie entre  $10^{12}$  et  $10^{-5}$ . Pour chaque  $\lambda$ , une moyenne sur 30 paires de signaux  $\mathbf{x}$  et  $\mathbf{y}$  différents a été effectuée. À chaque essai, un signal de pluie différent de  $\mathbf{x}$  a été choisi au hasard, et pour chaque  $\mathbf{y}$  ainsi simulé, un signal de bruit gaussien a été généré pour obtenir différents niveaux de SNR en entrée. Afin de choisir le meilleur  $\lambda$ , trois stratégies sont comparées : on choisit le

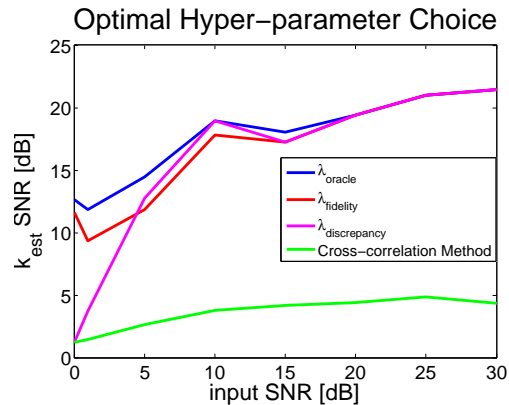
1.  $\lambda$  qui correspond au meilleur SNR en sortie sur  $\mathbf{k}$  ( $\lambda_{oracle}$ ). Cette stratégie ne fonctionne que si l'on a connaissance de la solution. Il s'agit de la valeur maximum "oracle" atteignable.
2.  $\lambda$  qui donne un résidu de variance proche de celle du bruit. Cette méthode est connue sous le nom de "Morozov's discrepancy principle" [9] ( $\lambda_{discrepancy}$ ).
3.  $\lambda$  correspondant au meilleur SNR sur le signal reconstruit  $\mathbf{y}_{rec} = \mathbf{x} * \mathbf{k}_{est}$  ( $\lambda_{fidelity}$ ). Il s'agit de la valeur de l'optimum de reconstruction. Cette méthode totalement heuristique donne une sélection automatique de l'hyper-paramètre avec une performance proche de la méthode de Morozov, sans connaissance *a priori* de la variance du bruit, qui peut être difficile à estimer en pratique sur ce problème.

Nous comparons les trois estimations du noyau de convolution  $\mathbf{k}$  données par ces trois stratégies ( $\lambda_{oracle}$ ,  $\lambda_{fidelity}$ ,  $\lambda_{discrepancy}$ ) en fonction du SNR d'entrée sur la figure 1(a). Ces résultats sont de plus comparés avec la méthode de corrélation croisée de l'état de l'art. L'évolution de l'hyperparamètre  $\lambda$  est tracée pour les trois stratégies sur la figure 1(b).

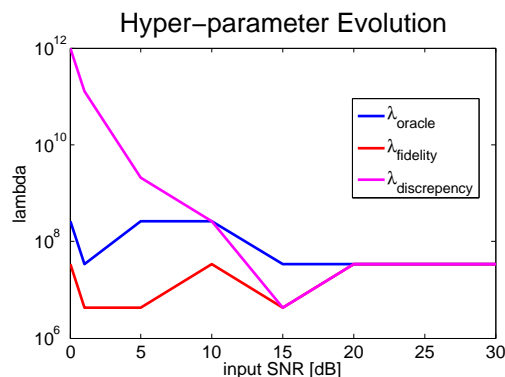
Sur la figure 1(a), nous voyons que le  $\lambda_{oracle}$  donne les meilleurs résultats comme attendu. On remarque que les stratégies  $\lambda_{fidelity}$  et  $\lambda_{discrepancy}$  donnent des résultats comparables pour des SNR allant de 10dB à 30dB. Pour les niveaux de bruit les plus élevés avec un  $SNR < 10$ dB,  $\lambda_{discrepancy}$  donne les meilleurs résultats. La méthode par corrélation croisée donne les plus mauvais résultats quelque soit le SNR. La figure 1(b) reporte l'évolution de  $\lambda$  en fonction du SNR en entrée. On note que malgré une décroissance "globale" de l'hyperparamètre, l'évolution des courbes montrent que le choix de cet hyperparamètre semble très sensible aux données en entrées.

## 4 Résultats sur données réelles

L'approche proposée a été comparée à la méthode de corrélation croisée sur les données réelles  $\mathbf{x}$  et  $\mathbf{y}$ , issues de [3]. L'estimation a été effectuée en fonction d'une gamme de  $\lambda$  choisie autour des valeurs optimales trouvées avec l'ensemble de données synthétiques :  $\lambda \in [10^2, 10^8]$ . Sur la figure 2, on peut voir les estimations des courbes de temps de résidence obtenue. L'algorithme de déconvolution a réussi à reconstruire le  $\mathbf{y}_{rec}$  avec un SNR qui varie entre 5 et 14 dB, avec le  $\lambda$  correspondante au meilleur  $\mathbf{y}_{rec}$  SNR. L'esti-



(a)



(b)

FIGURE 1 – Résultats moyens pour 30 simulations synthétiques en fonction du SNR de  $\mathbf{y}$ . (a) comparaison de l'estimation de  $\mathbf{k}$  pour les 3 stratégies de choix d'hyper-paramètre, en fonction du niveau de bruit. (b) valeur de l'hyper-paramètre, en fonction du niveau de bruit.

mation du noyau  $\mathbf{k}$  est meilleure dans les cas où l'information est concentrée au début des séries  $\mathbf{x}$ . Dans tous les cas, les courbes estimées respectent les contraintes données et l'apparition des lobes multiples correspond aux sections des chutes des pluies de  $\mathbf{x}$ . Pour la corrélation croisée, la courbe de temps de distribution de l'eau estimée a l'inconvénient de ne respecter ni la contrainte de positivité, ni la contrainte de causalité.

## 5 Conclusion

Nous avons proposé une nouvelle approche pour l'estimation de la distribution du temps de résidence hydrologique, en tenant compte des *a priori* de douceur et des contraintes de positivité et de causalité. L'estimation du noyau  $\mathbf{k}$  se fait grâce à un algorithme rapide d'estimation alternée entre déconvolution 1D et estimation d'une constante correspondant au niveau moyen de la nappe phréatique. Nous avons validé l'approche sur des données

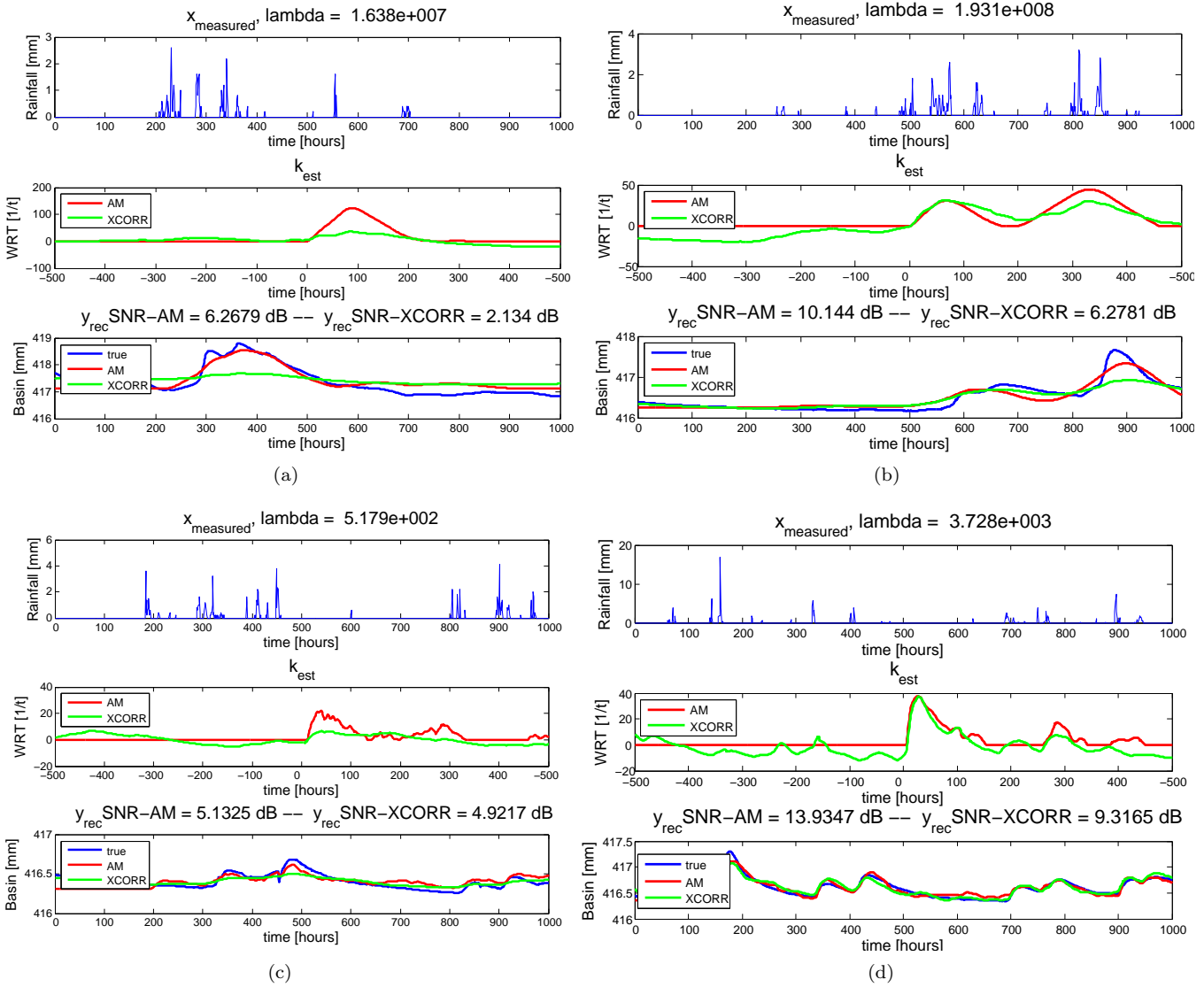


FIGURE 2 – Résultats sur données réels.

synthétiques ainsi que sur des données réelles, et proposé plusieurs stratégies pour estimer l'hyperparamètre.

This work is supported by the Center for Data Science, funded by the IDEX Paris-Saclay, ANR-11-IDEX-0003-02.

## Références

- [1] A. Beck. On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes. *SIAM J OPTIMIZ*, 25(1) :185–209, 2015.
- [2] D. P. Bertsekas. Projected newton methods for optimization problems with simple constraints. *SIAM J CONTROL OPTIM*, 20(2) :221–246, 1982.
- [3] C. Delbart, D. Valdes, F. Barbecot, A. Tognelli, P. Richon, and L. Couchoux. Temporal variability of karst aquifer response time established by the sliding-windows cross-correlation method. *J Hydrol*, 511 :580–588, 2014.
- [4] M. N. Gooseff, D. A. Benson, M. A. Briggs, M. Weaver, W. Wollheim, B. Peterson, and C. S. Hopkinson. Residence time distributions in surface transient storage zones in streams : Estimation via signal deconvolution. *Water Resour Res*, 47(5), 2011.
- [5] K. J. McGuire and J. J. McDonnell. A review and evaluation of catchment transit time modeling. *J Hydrol*, 330(3) :543–563, 2006.
- [6] S. Neuman, S. Resnick, R. Reebles, and D. B. Dunbar. Developing a new deconvolution technique to model rainfall-runoff in arid environments. 1982.
- [7] S. P. Neuman and G. De Marsily. Identification of linear systems response by parametric programing. *Water Resour Res*, 12(2) :253–262, 1976.
- [8] O. A. Cirpka et al. Analyzing bank filtration by deconvoluting time series of electric conductivity. *Groundwater*, 45(3) :318–328, 2007.
- [9] S. Pereverzev and E. Schock. Morozov's discrepancy principle for tikhonov. *Numer Funct Anal Optim*, 21(7-8) :901–916, 2000.
- [10] R. A. Payn et al. Comparison of instantaneous and constant-rate stream tracer experiments through non-parametric analysis of residence time distributions. *Water Resour Res*, 44(6), 2008.
- [11] Y. Tessier, S. Lovejoy, P. Hubert, D. Schertzer, and S. Pecknold. Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions. *J Geophys Res Atmos*, 101(D21) :26427–26440, 1996.
- [12] T. M. Werner and R. H. Kadlec. Wetland residence time distribution modeling. *Ecol Eng*, 15(1) :77–90, 2000.