

Driss KHAYOUB
Master 2
Statistiques et Probabilités Appliquées



Mélange d'eau au forage d'Orléans

Estimation de paramètres d'un modèle cinétique



Projet régional FDTE
Fouille de données temporelles – Application à des données environnementales

Université d'Orléans
Département de Mathématiques
Rue de Chartres - 45100 Orléans

Maître de stage : **M. Stéphane BINET**, Maître de Conférences
M. Richard EMILION, Professeur

Année 2013 –2014

Avant-propos

Ce stage a été réalisé dans le cadre du projet régional FDTE, « Fouille de données temporelles – Application à des données environnementales », au laboratoire MAPMO à l'Université d'Orléans de juin à Septembre 2014.

Il avait pour but d'étudier le mélange d'eau au forage d'Orléans en donnant une estimation des paramètres d'un modèle cinétique.

Remerciements

Je remercie mon professeur encadrant et maître de stage, M. Richard EMILION, pour m'avoir permis d'effectuer ce stage, accordé sa confiance, pour son aide et sa sympathie.

Je remercie M. Stéphane BINET, maître de conférences à l'Observatoire de la Terre d'Orléans, Mme Fatima LAGGOUN-DEFARGE, chargée de recherche CNRS, pour leurs temps consacrés durant cette période de stage ainsi que M. Amor Keziou, maître de conférence à l'université de Reims pour son apport sur la méthode à appliquer qu'il a utilisé dans des études similaires.

Je remercie également la Région Centre pour le financement de ce stage ainsi que tous les participants de ce projet.

Sommaire

AVANT-PROPOS.....	3
REMERCIEMENTS.....	3
1. INTRODUCTION.....	7
1.1. CONTEXTE.....	7
1.2. OBJECTIFS.....	7
1.3. PLAN.....	8
2. HYDROLOGIE DU VAL D'ORLEANS.....	9
3. POLLUTION PAR LES NITRATES.....	11
3.1. ORIGINE.....	11
3.2. PRINCIPE DE POLLUTION [1].....	11
3.3. CONSEQUENCES.....	12
4. MODELE CINETIQUE.....	13
4.1. PROBLEMATIQUE.....	13
4.2. MODELE.....	13
4.3. FORMALISME MATHEMATIQUE.....	14
5. DONNEES DISPONIBLES ET TRAITEMENTS.....	15
5.1. DESCRIPTION DES DONNEES.....	15
5.2. TRAITEMENTS DES DONNEES ET MISE EN FORME.....	16
5.2.1. Filtrage.....	16
5.2.2. Mise en forme.....	18
5.2.3. Choix des données issues de la Loire par interpolation.....	18
5.2.4. Choix des données issues du Loiret pour un Δt petit.....	19
6. METHODE DU GRADIENT.....	20
6.1. METHODE DU GRADIENT SIMPLE.....	21
6.2. METHODE DU GRADIENT CONJUGUE.....	22
6.2.1. Principe de la méthode.....	22
6.2.2. Choix du critère d'arrêt.....	22
6.3. METHODES DE PENALISATIONS.....	22
6.3.1. La méthode des pénalités extérieures.....	23
6.3.2. La méthode des pénalités intérieures.....	23
6.4. METHODE DU GRADIENT CONJUGUE PROJETE.....	25
6.5. METHODE DU GRADIENT CONJUGUE PROJETE STOCHASTIQUE.....	25
7. MISE EN APPLICATION.....	27
7.1. EXEMPLE THEORIQUE.....	27
7.2. CAS REEL.....	28
7.2.1. Existence d'un minimum de la fonction de coût.....	28
7.2.2. Choix des bornes des paramètres.....	28
7.2.3. Choix du nombre d'initialisation de la méthode de Monte Carlo.....	28
7.2.4. Résultats.....	29
7.2.5. Améliorations futures.....	33
7.2.6. Analyse des résidus.....	33

7.2.6.1.	Normalité des résidus.....	33
7.2.6.2.	Indépendance des résidus	35
7.2.6.3.	Homogénéité des variances des résidus.....	36
8.	CONCLUSION	36
9.	ANNEXE	37
9.1.	MISE EN ŒUVRE DU MODELE CINETIQUE [9]	37
9.1.1.	<i>Modèle d'advection dispersion</i>	37
9.1.2.	<i>Modèles conceptuels des écoulements autour du forage.....</i>	37
9.2.	DERIVEES DE LA FONCTION f DU MODELE CINETIQUE	38
9.2.1.	<i>Dérivées premières de la fonction f.....</i>	38
9.2.2.	<i>Dérivées seconde (Hessienne) de la fonction f</i>	40
9.3.	CODES DU TRAITEMENT DES DONNEES	42
9.4.	CODES DE L'EXEMPLE THEORIQUE	44
9.5.	CODES DU CAS REEL	45
REFERENCE.....		49

1. Introduction

1.1. Contexte

Très solubles dans l'eau, les nitrates constituent aujourd'hui la cause majeure de la pollution des grands réservoirs d'eau souterraine du globe qui par ailleurs présentent en général une qualité chimique et bactériologique satisfaisante pour l'alimentation.

En France, la présence de nitrates dans les eaux continentales provient à 66 % de l'agriculture, suite à l'épandage de doses massives d'engrais azotés et de lisier (effluents d'élevage), les zones les plus atteintes étant les plaines alluviales qui récoltent les eaux des grands bassins versants et sont des lieux privilégiés d'agriculture intensive. Le reste est issu des rejets des collectivités locales (22 %) et de l'industrie (12 %) [1].

En région Centre, 177 captages d'eau potable ont été abandonnés en raison d'une pollution trop importante aux nitrates et la zone dite vulnérable c'est-à-dire menacée, tend à s'agrandir chaque année [2].

Ainsi dans le cadre du projet « Fouille de données temporelles – Application à des données environnementales » (FDTE), porté par la région Centre, il est important d'étudier ces phénomènes environnementaux afin de les comprendre et de préconiser des mesures pour préserver ces eaux souterraines contre la pollution des nitrates.

Dans ce projet d'autres travaux en cours de développement et plus conséquents portent sur des données de la tourbières de la guette en région Centre et de Frasnes dans le Jura [3].

Plusieurs partenaires participent à ce programme, LI à l'université de Tours, l'ISTO, LIFO ainsi que le MAPMO à l'Université d'Orléans.

1.2. Objectifs

Notre étude est basée sur des données recueillies sur 3 cours d'eau situés dans le département du Loiret au Val d'Orléans, le Dhuy, le Loiret et la Loire.

Considérant la qualité de l'eau du forage issue du Loiret (concentration en nitrates) comme étant un mélange d'eau provenant de la Loire et du Dhuy, l'idée du projet est d'estimer la part de mélange entre ces deux origines en fonction du temps.

1.3. Plan

Suite à une présentation sur l'hydrologie au Val d'Orléans, sur la pollution des nitrates et sur leurs conséquences, nous exposerons la problématique de ce projet et le modèle cinétique sur lequel nous travaillerons.

Nous reformulerons également ce problème de mélange d'eau au forage d'Orléans en un problème d'optimisation.

Après une présentation des données dont nous disposons, nous traiterons des modifications apportées et de leurs mises en forme.

La partie suivante sera consacrée à la méthode d'optimisation mise en œuvre pour résoudre ce problème d'estimation de paramètre. Nous verrons les différentes étapes qui nous ont conduit à choisir la méthode du gradient conjugué projeté stochastique.

En fin de rapport, avant d'exposer les résultats et l'analyse des résidus permettant d'examiner si les hypothèses de base du modèle linéaire ne sont pas violées, nous verrons un cas pratique qui nous permettra de tester l'algorithme mis en application.

Nous fournissons en annexe la mise en œuvre du modèle cinétique, les différentes dérivées nécessaires à l'application de l'algorithme.

Nous proposons également les codes permettant d'exploiter les données et les fonctions nécessaires à l'exécution de l'algorithme (développement sous R).

2. Hydrologie du Val d'Orléans

Le Val d'Orléans est un bassin allongé d'environ 15 hectares de superficie. Il mesure dans sa plus grande longueur de 35 à 36 km depuis Sigloy, à l'est, jusqu'à la confluence à l'ouest de la Loire. Sa largeur en moyenne de 6 km peut atteindre 8 km à Sandillon. En plus de la Loire qui délimite la partie nord du val, les eaux de surface du Val d'Orléans sont drainées par deux cours d'eau principaux, le Dhuy affluent du Loiret, et ce dernier qui se jette dans la Loire à la pointe de Courpain [4].

La Loire est plus grand fleuve de France avec 1006 kilomètres de longueur. Il prend sa source en Ardèche, au sud-est du Massif Central, au pied du mont Gerbier-de-Jonc, autrefois ancienne cheminée de volcan. Il se jette sur l'océan Atlantique à Saint Nazaire. La présence d'une nappe phréatique sous le mont Gerbier-de-Jonc donne naissance à de multiples sources relativement voisines et trois d'entre elles sont mises en avant comme sources du fleuve :



Figure 1, sources de la Loire, de gauche à droite, « source géographique », « source authentique », « source véritable »

Le Dhuy, long de 34.2 km, s'écoule au pied du plateau de Sologne sur la commune de Sully-sur-Loire et se jette dans le Loiret 1000 m à l'aval de la source du Bouillon. Depuis sa source et jusqu'à Tigy, on le nomme le Leu. Puis à hauteur de Vienne-en-Val, les habitants l'appellent, la Bergeresse.

Ces principaux influents sont L'Ousson et la Marmagne.

Il présente de grandes variations de débits. Durant les périodes de hautes eaux, il peut présenter des débits de l'ordre de 15 m³/sec. Par contre, au cours de la période d'étiage, son débit est presque nul. Dans les années 1970, le cours du Dhuy a été fortement artificialisé, afin de favoriser les activités agricoles. Le résultat fut une baisse importante de la biodiversité et une nette dégradation de la qualité des eaux. Ce qui provoque cependant une incidence non négligeable sur celle du Loiret.

Le Loiret, long de 11.6 km, est l'un des plus petit affluent de la Loire. Il s'agit plus exactement d'une résurgence de la Loire dont l'essentiel du débit est souterrain. Il prend sa source dans l'enceinte du parc floral d'Orléans La Source : le Bouillon. Cette source apparue en 1672 suite au comblement accidentel et partiel de l'Abîme, l'eau s'est frayé un autre chemin pour sortir de terre. Résultant de ce phénomène, le Bouillon est aujourd'hui la source principale du Loiret [5].



Figure 2, Source du Loiret : le Bouillon

Le Bouillon surgit à 7 mètres de profondeur après une circulation dans des galeries souterraines dont certaines font plus de 8 mètres de diamètre. Régulièrement cartographiée par des spéléologues, cette source est une résurgence karstique de la Loire. En effet, suite à la présence de plateaux calcaires, une partie des eaux du fleuve circule dans un réseau de galeries souterraines d'environ 40km. Ces eaux resurgissent au milieu du parc pour ensuite s'en aller de nouveau rejoindre la Loire.

Le Loiret a eu plusieurs sources nées sur le domaine du Parc Floral de la Source [6]:

- L'Abîme : Cette résurgence du Loiret située en contrebas du Château était autrefois une source abondante et limpide, jusqu'à ce que l'un des propriétaires du domaine y fit exécuter divers aménagements qui engendrèrent accidentellement un effondrement partiel des rives.
- Le Gouffre : La troisième source du Loiret n'existe plus aujourd'hui. Située à l'embouchure du Dhuy, elle fut comblée progressivement par la chute d'arbres en 1879 et par des ouvriers qui y jetèrent des sacs de ciment. Le Gouffre tirait son nom d'un phénomène souvent observé : en période de hautes-eaux de la Loire, le Loiret jaillissait, alimentant le Dhuy ; en période de basses-eaux, les eaux du Dhuy s'y engouffraient, se mêlant à celles du Loiret qui refluaient.

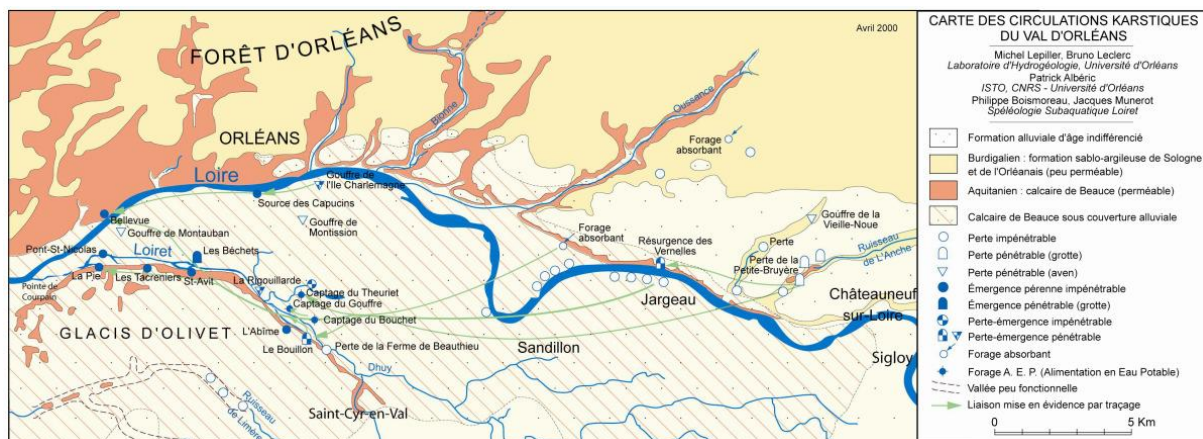


Figure 3, carte hydrogéologique du Val d'Orléans [4]

Le Loiret est alimenté en moyenne à 70 % par les pertes de la Loire, des résurgences qui réapparaissent au niveau de ses sources. Seules 30% des eaux du Loiret proviennent du Dhuy.

Pour assurer l'alimentation en eau potable de 155 000 habitants, La ville d'Orléans dispose de 3 captages dans la zone du Val, le Theuriet, le Gouffre et Le Bouchet. L'eau brute y est puisée dans la nappe alluviale par forage à 20 m de profondeur puis traitée des différentes pollutions à l'usine du Val.

3. Pollution par les nitrates

3.1. Origine

Cette pollution a débuté à la fin des années 1950 et n'a fait qu'augmenter depuis lors. Alors qu'en l'absence de contamination, la teneur en nitrates des eaux souterraines varie de 0,1 à 1 milligramme par litre d'eau, elle dépasse souvent aujourd'hui 50 milligrammes par litre, norme retenue pour les eaux potables par l'Organisation Mondiale de la Santé. Désormais, de telles eaux nécessitent donc un traitement spécifique pour pouvoir être consommées.

L'essentiel de cette pollution est dû à la différence entre les apports en nitrates sous forme d'engrais et ce qui est réellement consommé par les plantes. En France, selon un bilan du ministère de l'Agriculture, cet excédent est passé de 320 000 tonnes en 1995, à 400 000 tonnes en 1997, les régions les plus touchées étant la Bretagne, la Champagne-Ardenne, le Poitou-Charentes, l'Île-de-France et le Centre.

En région Centre, la pollution par les nitrates affecte principalement les nappes d'eau souterraine peu protégées naturellement compte tenu des caractéristiques hydrogéologiques du sous-sol. Ces eaux souterraines constituent la principale ressource en eau utilisée pour l'alimentation humaine. En 2011, 1 087 captages d'eau souterraine et 12 prises d'eau superficielle sont ainsi utilisés.

3.2. Principe de pollution [1]

La pollution par les nitrates est un problème complexe. Outre d'être des nutriments pour les plantes, les nitrates sont également consommés par les microorganismes (bactéries et champignons) présents dans la terre. Ils participent ainsi à la synthèse des matières organiques du sol qui stockent en leur sein l'azote contenu dans les nitrates qui n'ont pas été consommés dans l'année par les plantes, et ce jusqu'à leur mort. Leur décomposition par les bactéries libère alors l'azote qu'elles contiennent sous la forme de nitrates. Mais cette libération peut se produire à tout moment de l'année, notamment lorsque les plantes sont au repos et s'alimentent peu : dans ce cas, ne pouvant être consommés par ces dernières, les nitrates libérés sont lessivés par les eaux de ruissellement et d'infiltration.

Les nitrates emportés par les eaux d'infiltration au cours d'une année ne proviennent donc que pour une faible part des engrais apportés cette même année. L'essentiel provient de la production de nitrates par la matière organique morte des sols, c'est-à-dire des nitrates épandus les années précédentes et stockés. À ceci s'ajoute parfois la lenteur de la progression de l'eau d'infiltration dans les sols.

Chaque épandage contribue donc peu chaque année à la contamination des eaux, mais il y contribue durant de nombreuses années. D'année en année, ces contributions " retardées " s'additionnent les unes aux autres et les quantités de nitrates lessivés atteignant les nappes augmentent. C'est ainsi que la pollution actuelle des nappes souterraines provient de 20 à 30 années d'épandage d'engrais.

Même si l'on arrêta aujourd'hui de fertiliser les sols, il faudrait attendre plusieurs décennies avant de retrouver une situation normale.

3.3. Conséquences

La pollution des eaux par les nitrates présente un double risque. Ingérés en trop grande quantité, les nitrates ont des effets toxiques sur la santé humaine :

- Les nitrates transformés en nitrites dans l'organisme peuvent par la modification des propriétés de l'hémoglobine du sang, empêcher un transport correct de l'oxygène par les globules rouges dans le sang.
- En milieu acide comme celui de l'estomac, les nitrites peuvent également se combiner avec les amines pour produire des nitrosamines. Si des effets cancérigènes ont été observés chez l'animal exposé à de très fortes doses en nitrates, le risque de cancérogenèse lié aux nitrates n'a pas été démontré pour l'homme, de façon absolue. Toutefois le Centre International de Recherche sur le Cancer (CIRC) a classé les nitrates et les nitrites en cancérogènes probables pour l'homme.

Par ailleurs, ils contribuent avec les phosphates à modifier l'équilibre biologique des milieux aquatiques en provoquant des phénomènes tel que :

- Eutrophisation (modification et dégradation d'un milieu aquatique)
- Dystrophisation (état de l'eutrophisation, qui se traduit par la mort des organismes animaux et végétaux supérieurs) [7].

4. Modèle cinétique

4.1. Problématique

Nous disposons des suivis de nitrates (NO_3^-) aux forages du Loiret qui servent à l'alimentation en eau potable de la ville d'Orléans.

Sur ce site, ils peuvent avoir deux origines la Loire qui participe à alimenter 80% de l'eau du forage et les eaux du Dhuy qui s'infiltrent sur le bassin versant à hauteur de 20%. Ces pourcentages sont des moyennes et varient au cours du temps. Ainsi sous terre, portée par le courant et dispersée par les propriétés du milieu, la concentration en nitrate va évoluer. Elle peut atteindre 60% en provenance du Dhuy et 40% de la Loire. Elle peut également provenir entièrement de la Loire.

Cette étude consiste à estimer la part de mélange entre ces deux origines en fonction du temps.

4.2. Modèle

Le problème posé est de trouver les valeurs optimales des paramètres du modèle suivant :

$$C_{f(t+\Delta t)} = C_{L(t)}\alpha \sqrt{\frac{Pe_L}{4\pi(t+\Delta t)}} e^{\frac{-Pe_L(t_L-(t+\Delta t))^2}{4(t+\Delta t)}} + C_{D(t)}(1-\alpha) \sqrt{\frac{Pe_D}{4\pi(t+\Delta t)}} e^{\frac{-Pe_D(t_D-(t+\Delta t))^2}{4(t+\Delta t)}}$$

Concentrations connues :

- C_L : Concentration en NO_3^- mesurée sur la Loire
- C_f : Concentration en NO_3^- mesurée sur le Loiret
- C_D : Concentration en NO_3^- mesurée sur le Dhuy

Paramètres à estimer :

- Pe_L et Pe_D : nombre de Peclet qui caractérise respectivement la Loire et le Dhuy
- t_L et t_D : temps moyen de séjour qui caractérise respectivement la Loire et le Dhuy
- α : constante qui pourrait correspondre à une probabilité que l'eau ait pour origine la Loire, $1-\alpha$ pourrait correspondre à une probabilité que l'eau ait pour origine le Dhuy

Et Δt est le pas de temps.

Le nombre de Peclet (Pe) est un nombre sans dimension utilisé en transfert thermique et en transfert massique. Il représente le rapport de deux temps caractéristiques : celui du transfert par convection et celui du transfert par diffusion dans ce modèle.

Pour la mise en œuvre du modèle voir l'annexe [9].

4.3. Formalisme mathématique

La forme générale du modèle non linéaire statistique est :

$$U = f((X, Y, dt); \beta) + \varepsilon \quad (1)$$

où le triplet (X, Y, dt) est un vecteur de variable explicative, β est un vecteur de dimension 5 de paramètres inconnus et ε est un terme d'erreur de loi normale $N(0, \sigma^2)$.

Nous disposons de n vecteurs (U, X, Y, dt) de points expérimentaux. On connaît la forme explicite et non linéaire de la fonction liant ces points dans laquelle figurent 5 paramètres inconnus $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ à déterminer :

$$f : ((X, Y, dt); \beta) \rightarrow X\beta_1\sqrt{\frac{\beta_2}{4\pi(t+\Delta t)}}e^{\frac{-\beta_2(\beta_3-(t+\Delta t))^2}{4(t+\Delta t)}} + Y(1 - \beta_1)\sqrt{\frac{\beta_4}{4\pi(t+\Delta t)}}e^{\frac{-\beta_4(\beta_5-(t+\Delta t))^2}{4(t+\Delta t)}} \quad (2)$$

Pour estimer ces 5 paramètres, on cherche à minimiser la somme des écarts quadratiques entre les valeurs expérimentales et les valeurs simulées :

$$D = \sum_{i=1}^n (u_i - f(x_i, y_i, dt_i, \beta))^2 \quad (3)$$

La fonction coût (3) peut être considérée comme une mesure de la distance entre les données expérimentales et le modèle théorique (2) qui prédit ces données.

Pour obtenir le minimum de la fonction (3), il serait nécessaire de parcourir tout l'espace des variables, ce qui est impossible vu la grandeur de l'espace et le nombre de variables. Nous allons alors restreindre cet espace par des contraintes liées aux variables, des contraintes liées au modèle et par des contraintes que nous allons imposer pour limiter l'espace de recherche :

- Le paramètre α désigne une probabilité et devra prendre ses valeurs entre 0 et 1 ($0 \leq \alpha \leq 1$).
- Les autres paramètres devront être tous positifs et limités par un vecteur b , à déterminer pour faciliter la recherche du minimum :

$$\begin{cases} 0 \leq Pe_L \leq b \\ 0 \leq Pe_D \leq b \\ 0 \leq t_L \leq b \\ 0 \leq t_D \leq b \end{cases}$$

Cela revient à un problème d'optimisation (P) définie par :

$$\begin{array}{ll} \min D(\beta) & \beta \text{ sur } R^5 \\ \text{Avec} & 0 \leq \beta_1 \leq 1 \\ \text{Et} & 0 \leq \beta_j \leq b \quad \text{pour } j = 2 \dots 5 \end{array}$$

5. Données disponibles et traitements

5.1. Description des données

Nous disposons des données sur le suivi de nitrate (NO₃⁻) mesuré en milligramme par litre aux forages du Loiret qui servent à l'alimentation en eau potable de la ville d'Orléans. D'autres mesures sont également disponibles tel que la station de prélèvement, la date (jour/mois/année), la température (°C), le pH et la conductivité (mg/l).

Station	Date	Temp. (°C)	EC (µS/cm)	pH (-)	NH4+ (mg/L)	Cl- (mg/L)	NO2- (mg/L)	NO3- (mg/L)	PO43- (mg/L)
LOIRE ORLEANS	04/01/00	6,2	220	7,3	0,05	13	< ,05	13,0	0,18
LOIRE ORLEANS	04/01/00	6,2	220	7,3	0,05	13	< ,05	13,0	0,18
LOIRE ORLEANS	11/01/00	5,5	225	7,4	< 0,1			13,7	0,18
LOIRE ORLEANS	11/01/00	5,5	225	7,4	< 0,1			13,7	0,18
LOIRE ORLEANS	18/01/00	4,9	250	7,5	0,05	16	< ,05	14,0	0,16
LOIRE ORLEANS	18/01/00	4,9	250	7,5	0,05	16	< ,05	14,0	0,16
LOIRE ORLEANS	25/01/00	2,8	240	7,6	< 0,1			14,0	0,19
LOIRE ORLEANS	25/01/00	2,8	240	7,6	< 0,1			14,0	0,19
LOIRE ORLEANS	01/02/00	6,8	290	7,5	0,06	18	0,05	14,0	0,2
LOIRE ORLEANS	01/02/00	6,8	290	7,5	0,06	18	0,05	14,0	0,2
LOIRE ORLEANS	08/02/00	7,8	217	7,4	0,09			12,0	0,2
LOIRE ORLEANS	08/02/00	7,8	217	7,4	0,09			12,0	0,2
LOIRE ORLEANS	15/02/00	7,8	240	7,4	0,06	15	< ,05	12,0	0,17
LOIRE ORLEANS	15/02/00	7,8	240	7,4	0,06	15	< ,05	12,0	0,17
LOIRE ORLEANS	22/02/00	6,3	195	7,3	< 0,1			11,0	0,13
LOIRE ORLEANS	22/02/00	6,3	195	7,3	< 0,1			11,0	0,13
LOIRE ORLEANS	01/03/00	8,6	223	7,3	0,06			12,0	0,2
LOIRE ORLEANS	01/03/00	8,6	223	7,3	0,06			12,0	0,2
LOIRE ORLEANS	07/03/00	7,8	210	7,3	< 0,1	11	< 0,05	11,0	0,14
LOIRE ORLEANS	07/03/00	7,8	210	7,3	< 0,1	11	< 0,05	11,0	0,14

Figure 4, Extrait des données disponibles sur Excel

Nous connaissons également le signal d'entrée en nitrates grâce aux mesures de nitrates réalisées en Loire et sur le Dhuy-Bergeresse.

Dans notre cas, nous nous intéresserons seulement aux mesures de nitrate prélevées à différentes dates.

Ces mesures sont effectuées sur plusieurs lieux des cours d'eau Orléanais :

- La Loire à Orléans, Saint Pryvé Saint Mesmin et à Jargeau
- Le Dhuy à Orléans la source au niveau du Bergeresse et à Sandillon
- Le Loiret aux puits Bouchet, Gouffre, Theuriet et à la source Bouillon

Comme on peut le voir sur la carte ci-dessous.

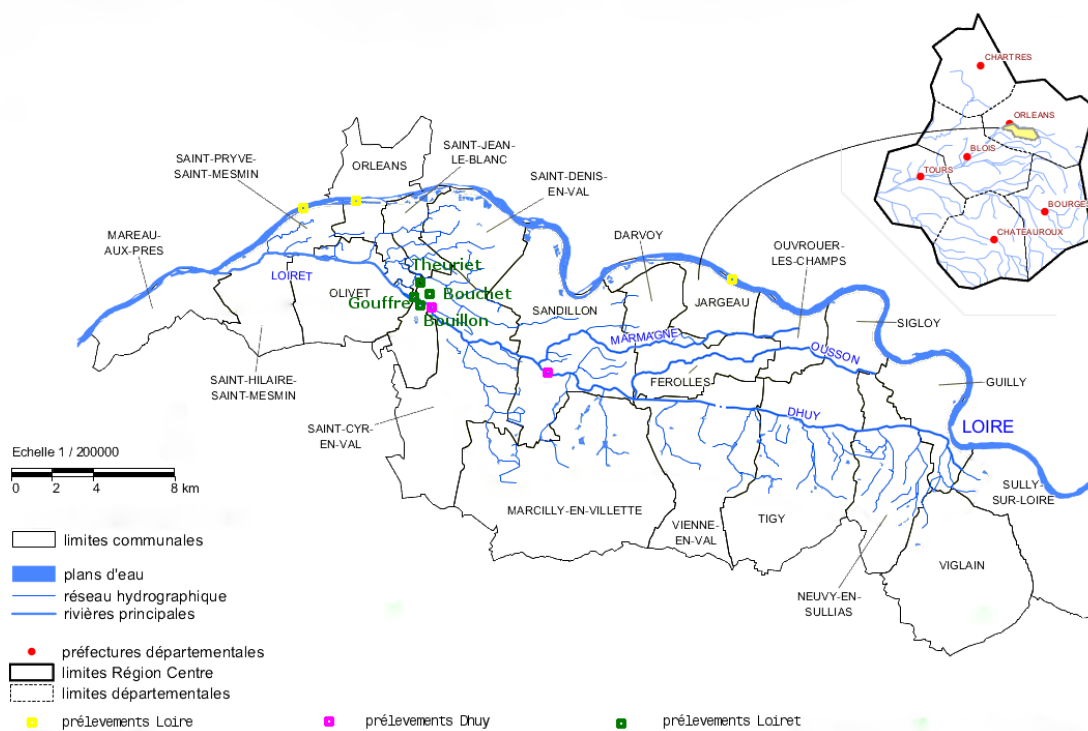


Figure 5, Emplacements des prélèvements de nitrate sur la Loire, le Dhuy et le Loiret

Elles sont effectuées avec un intervalle de temps irrégulier pouvant aller de quelques heures à 1 mois et répartie sur la période de janvier 2000 à août 2011.

La cohérence des données étant fait préalablement, nous allons procéder maintenant à leurs traitements ainsi qu'à leurs mises en forme avant leurs exploitations.

5.2. Traitements des données et mise en forme

5.2.1. Filtrage

Les prélèvements sur le Dhuy sont cohérents et complets. Ils seront utilisés sans modification et serviront de base pour le traitement des prélèvements sur les autres cours d'eau.

Sur les mesures effectuées sur la Loire et le Loiret, on observe plusieurs valeurs manquantes. Sans la valeur des mesures, ces prélèvements ne sont pas exploitables et afin de faciliter les calculs elles seront supprimées.

D'autres mesures indiquent seulement l'existence de nitrate de concentration comprise entre 0 et 1, comme on peut le voir ci-dessous :

LOIRE JARGEAU	09/07/02	22,3	234	8,75	0,05	21	0,07	2,5
LOIRE JARGEAU	09/07/02	22,3	234	8,75	0,05	21	0,07	2,5
LOIRE JARGEAU	17/07/02	20	287	8,7	< 0,1		0,04	2,7
LOIRE JARGEAU	17/07/02	20	287	8,7	< 0,1		0,04	2,7
LOIRE JARGEAU	24/07/02	20,7	263	8,8	< 0,1		0,06	2,4
LOIRE JARGEAU	24/07/02	20,7	263	8,8	< 0,1		0,06	2,4
LOIRE JARGEAU	31/07/02	21,8	255	8,2	0,22		0,04	< 1
LOIRE JARGEAU	31/07/02	21,8	255	8,2	0,22		0,04	< 1
LOIRE JARGEAU	07/08/02	20	268	7,9	0,12		0,04	< 1
LOIRE JARGEAU	07/08/02	20	268	7,9	0,12		0,04	< 1
LOIRE JARGEAU	14/08/02	22	281	8,1	< 0,1		0,05	1,9
LOIRE JARGEAU	14/08/02	22	281	8,1	< 0,1		0,05	1,9
LOIRE JARGEAU	28/08/02	19,2	292	7,67	0,09	20	0,06	3,5
LOIRE JARGEAU	28/08/02	19,2	292	7,67	0,09	20	0,06	3,5
LOIRE JARGEAU	04/09/02	19		7,65	< 0,1		0,03	6,6
LOIRE JARGEAU	04/09/02	19		7,65	< 0,1		0,03	6,6
LOIRE JARGEAU	10/09/02	17,7	270	8,21	0,13	20	0,04	6,5
LOIRE JARGEAU	10/09/02	17,7	270	8,21	0,13	20	0,04	6,5
LOIRE JARGEAU	18/09/02	17,2	257	7,5	< 0,1		0,03	7,0
LOIRE JARGEAU	18/09/02	17,2	257	7,5	< 0,1		0,03	7,0
LOIRE JARGEAU	25/09/02	12,1	268	7,8	< 0,1		< ,03	8,3
LOIRE JARGEAU	25/09/02	12,1	268	7,8	< 0,1		< ,03	8,3
LOIRE JARGEAU	02/10/02	15,6	278	7,8	< 0,1		< ,03	7,0

Figure 6, Manque de précision (encadré rouge)

Ce manque de précision est dû aux analyses chimiques qui possèdent un seuil de détection, la machine ne voit pas les concentrations plus petites que 1.

Vu la gamme de variation de la concentration de nitrate dans le Loiret et la Loire qui est respectivement de 0 mg/l à 27,1 mg/l et de 0 mg/l à 21 mg/l, nous négligerons cette faible concentration en la réduisant à 0.

On obtient le décomptage des données brutes et des données après traitement sur le tableau suivant :

Cours d'eau	Prélèvements (données brutes)	Valeurs manquantes	Valeur comprise entre 0 et 1	Doublons	Données exploitables
Loiret	1396	564	0	0	832
Loire	866	1	18	439	426
Dhuy	185	0	0	0	185

Figure 7, Décomptage des données brutes et des données traitées

On obtient aussi une représentation graphique des points exploitables :

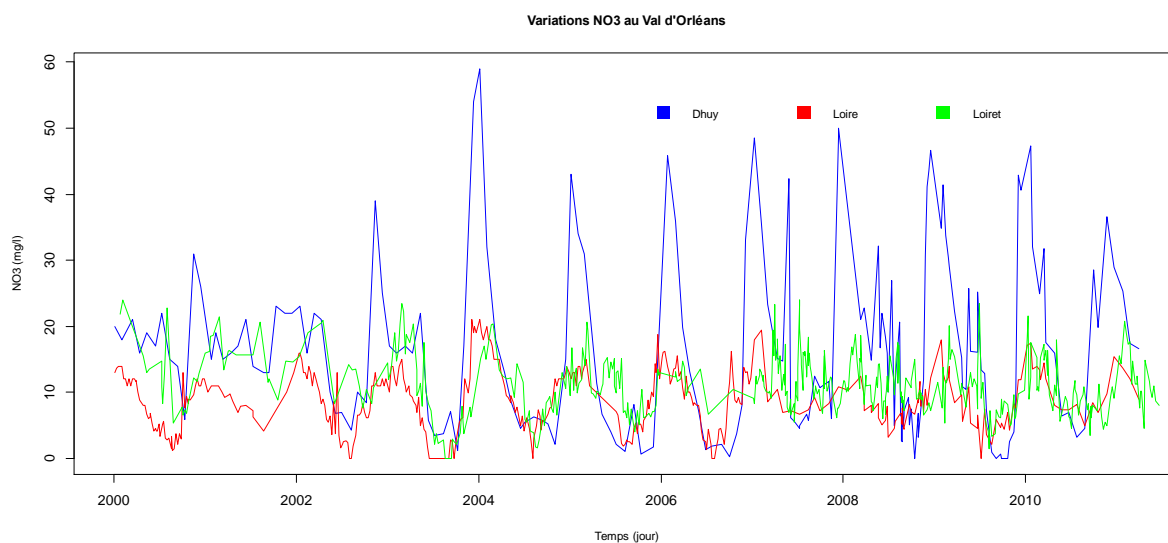


Figure 8, Evolution de la concentration de nitrate, de 2000 à 2011 sur les trois cours d'eau

5.2.2. Mise en forme

Sachant que sur l'ensemble des 3 cours d'eau, le Dhuy est celui qui représente le moins de prélèvements, soit 185 prélèvements.

Les prélèvements sur le cours d'eau du Dhuy sont effectués généralement une fois par mois.

Nous allons adapter les données issues de la Loire et du Loiret en fonction de celles effectuées sur le Dhuy. Nous aurons ainsi au maximum 185 données exploitables.

5.2.3. Choix des données issues de la Loire par interpolation

D'après le modèle cinétique les mesures issues de la Loire et du Dhuy doivent être prélevées au même temps ce qui n'est pas toujours le cas.

Nous avons 426 mesures concernant la Loire (environ 2 fois plus que le Dhuy).

Sachant que les prélèvements sur le Dhuy sont effectués une fois par mois, de plus l'écart temporel maximal de prélèvement entre 2 dates de ces 2 cours d'eau ne dépasse pas 3 jours, nous allons procéder à une interpolation.

Nous obtenons maintenant après interpolation 184 mesures au lieu de 185. La dernière mesure n'a pas été prise car le dernier prélèvement a été effectué sur le Dhuy.

Les conséquences d'une interpolation des mesures de la Loire sur le Dhuy sont minimales comme on peut le voir sur le résultat graphique suivant :

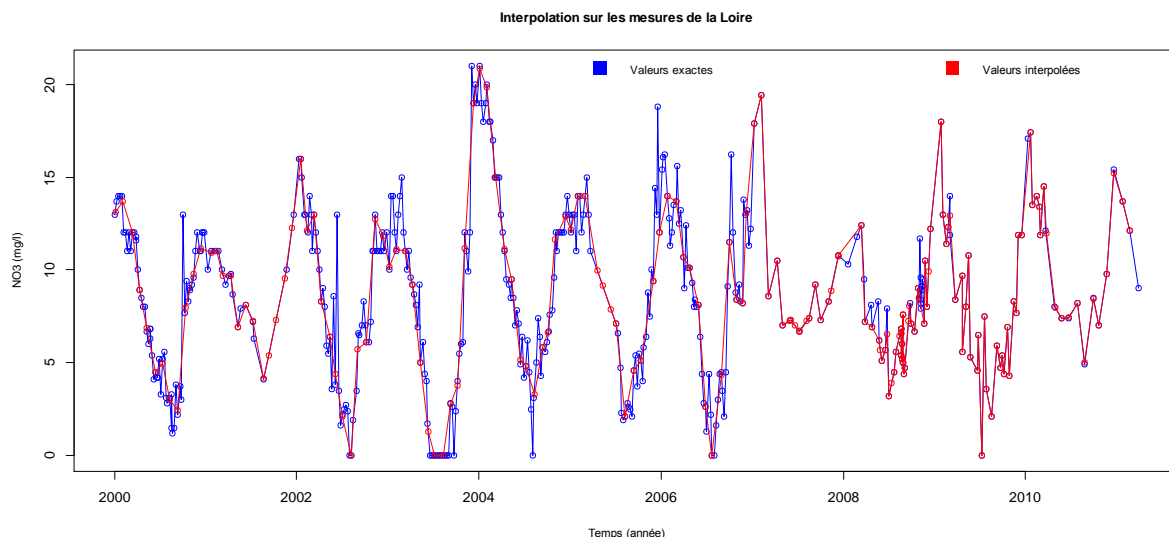


Figure 9, Evolution des mesures en concentration de nitrate réel de la Loire (bleu) et mesures interpolées de la Loire par rapport à celle du Dhuy (rouge)

5.2.4. Choix des données issues du Loiret pour un Δt petit

Les données concernant le Loiret sont prélevées sur 4 lieux au même temps (sur le Loiret aux puits Bouchet, Gouffren, Theuriet et à la source Bouillon). Afin d'obtenir une seule mesure pour une date donnée, nous avons décidé de moyenner les mesures.

On obtient 482 mesures.

Nous allons prendre 184 mesures dont la date de prélèvement est la plus proche de celle des mesures du Dhuy. On aura ainsi un Δt le plus petit possible.

Cependant on obtient des doublons sur les mesures du Loiret :

	DateDhuy	DhuyNO3	LoireNO3	LoiretNO3	DateLoiret	dt
1	2000-01-05	20.0	13.100000	21.83333	2000-01-24	19
2	2000-02-02	18.0	13.714286	24.00000	2000-02-03	1
3	2000-03-14	21.0	12.000000	15.50000	2000-04-25	42
4	2000-04-12	16.0	8.900000	15.50000	2000-04-25	13
5	2000-05-09	19.0	6.862500	13.00000	2000-05-10	1
6	2000-06-13	17.0	4.500000	14.70000	2000-07-10	27
7	2000-07-11	22.0	4.962500	8.30000	2000-07-13	2
8	2000-08-09	15.0	3.012500	5.30000	2000-08-24	15
9	2000-09-11	14.0	2.428571	7.75000	2000-09-28	17
10	2000-10-11	5.9	7.942857	6.80000	2000-10-16	5

Figure 10, Doublet (encadré rouge)

Nous allons alors supprimer celles qui possèdent un Δt (dt) le plus élevé.

Après suppression des doublons, on obtient au final 166 mesures pour chaque cours d'eau représentées dans une matrice de taille 166×4 .

	DhuyNO3	LoireNO3	LoiretNO3	dt
1	20.0	13.100000	21.83333	19
2	18.0	13.714286	24.00000	1
4	16.0	8.900000	15.50000	13
5	19.0	6.862500	13.00000	1
6	17.0	4.500000	14.70000	27
7	22.0	4.962500	8.30000	2
8	15.0	3.012500	5.30000	15
9	14.0	2.428571	7.75000	17
10	5.9	7.942857	6.80000	5
11	31.0	9.800000	11.70000	12

Figure 11, Extrait de la matrice de données finale qui sera utilisé pour l'estimation des paramètres

Voici les 166 mesures pour chaque cours d'eau tracées sur le graphe suivant :

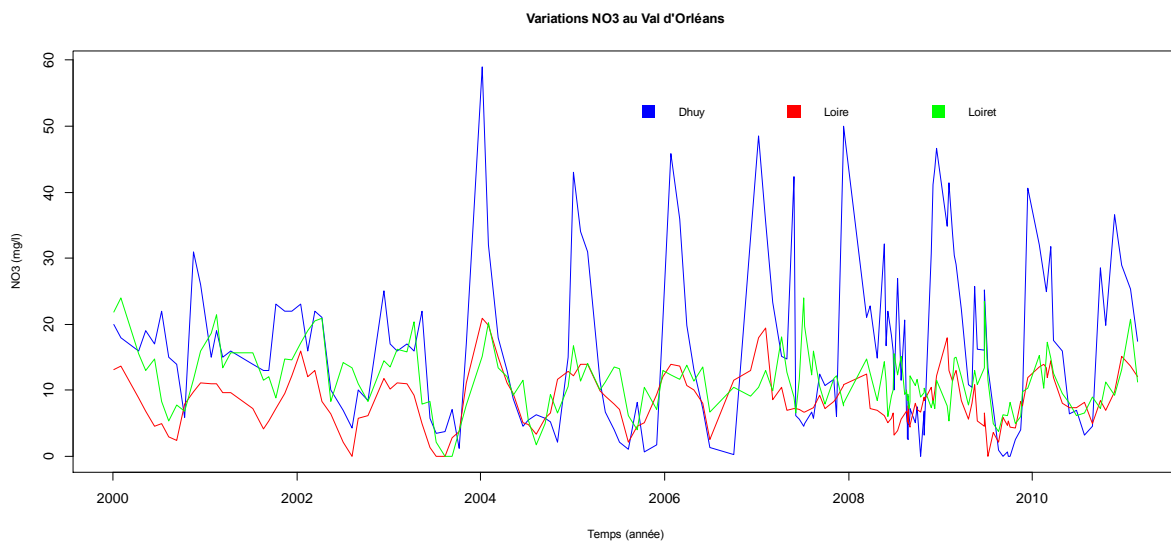


Figure 12, Evolution des mesures en concentration de nitrates qui seront pris en compte dans l'ajustement du modèle

Sur la figure 12, on peut observer le décalage des concentrations en nitrate entre celle de la Loire et du Dhuy et celle du Loiret. Le décalage représente le temps nécessaire permettant à la pollution en nitrate dans la Loire et le Dhuy d'arriver au Loiret.

Ces dernières données seront utilisées pour l'ajustement des paramètres $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ sur le modèle (2) afin que la courbe $U = f((X, Y, dt); \beta)$ représente au mieux les points expérimentaux (U, X, Y, dt) .

Pour cela Il n'y a pas de formule explicite pour estimer ces paramètres inconnus, nous avons alors recourt dans notre étude à un processus itératif, la méthode du gradient.

6. Méthode du gradient

Le principe de la méthode du gradient repose sur le fait qu'un minimum de la fonction de coût est atteint si sa dérivée (son gradient) est nul.

Il existe plusieurs méthodes du gradient. Les différences entre elles résident dans le choix de la direction de descente et dans le choix du pas. Nous allons voir la méthode du gradient simple, facile à mettre en œuvre mais ayant ses limites.

6.1. Méthode du gradient simple

Cette méthode calcule de manière itérative le vecteur β des paramètres inconnus par la relation :

$$\beta^{i+1} = \beta^i - t * \nabla D(\beta^i)$$

$$\text{Où } D(\beta) = \begin{pmatrix} \frac{\partial D}{\partial \beta_1} \\ \frac{\partial D}{\partial \beta_2} \\ \frac{\partial D}{\partial \beta_3} \\ \frac{\partial D}{\partial \beta_4} \\ \frac{\partial D}{\partial \beta_5} \end{pmatrix} \text{ et } \frac{\partial D}{\partial \beta} = -2 \sum_{i=1}^n (u_i - f(x_i, y_i, dt_i, \beta)) \frac{\partial f}{\partial \beta}(x_i, y_i, dt_i, \beta)$$

Pour le calcul du gradient, on utilisera la forme matricielle suivante :

$$D(\beta) = -2 * A * B$$

Où

$$A = (u_1 - f(x_1, y_1, dt_1, \beta) \quad u_2 - f(x_2, y_2, dt_2, \beta) \quad \cdots \quad u_{n-1} - f(x_{n-1}, y_{n-1}, dt_{n-1}, \beta) \quad u_n - f(x_n, y_n, dt_n, \beta))$$

Et

$$B = \begin{pmatrix} \frac{\partial f}{\partial \beta_1}(x_1, y_1, dt_1, \beta) & \frac{\partial f}{\partial \beta_2}(x_1, y_1, dt_1, \beta) & \frac{\partial f}{\partial \beta_3}(x_1, y_1, dt_1, \beta) & \frac{\partial f}{\partial \beta_4}(x_1, y_1, dt_1, \beta) & \frac{\partial f}{\partial \beta_5}(x_1, y_1, dt_1, \beta) \\ \frac{\partial f}{\partial \beta_1}(x_2, y_2, dt_2, \beta) & \frac{\partial f}{\partial \beta_2}(x_2, y_2, dt_2, \beta) & \frac{\partial f}{\partial \beta_3}(x_2, y_2, dt_2, \beta) & \frac{\partial f}{\partial \beta_4}(x_2, y_2, dt_2, \beta) & \frac{\partial f}{\partial \beta_5}(x_2, y_2, dt_2, \beta) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f}{\partial \beta_1}(x_{n-1}, y_{n-1}, dt_{n-1}, \beta) & \frac{\partial f}{\partial \beta_2}(x_{n-1}, y_{n-1}, dt_{n-1}, \beta) & \frac{\partial f}{\partial \beta_3}(x_{n-1}, y_{n-1}, dt_{n-1}, \beta) & \frac{\partial f}{\partial \beta_4}(x_{n-1}, y_{n-1}, dt_{n-1}, \beta) & \frac{\partial f}{\partial \beta_5}(x_{n-1}, y_{n-1}, dt_{n-1}, \beta) \\ \frac{\partial f}{\partial \beta_1}(x_n, y_n, dt_n, \beta) & \frac{\partial f}{\partial \beta_2}(x_n, y_n, dt_n, \beta) & \frac{\partial f}{\partial \beta_3}(x_n, y_n, dt_n, \beta) & \frac{\partial f}{\partial \beta_4}(x_n, y_n, dt_n, \beta) & \frac{\partial f}{\partial \beta_5}(x_n, y_n, dt_n, \beta) \end{pmatrix}$$

n , désigne le nombre de points expérimentaux pris en compte. La constante t désigne le pas du gradient et dans la méthode du gradient simple, il est fixe.

La direction de descente est donc simplement l'opposé de celle du gradient, c'est en effet la direction suivant laquelle la fonction de coût diminue le plus rapidement.

En pratique, la méthode du gradient simple peut être efficace lorsque l'on est loin du minimum de la fonction de coût. Quand on s'en approche, la norme du gradient diminue et donc l'algorithme progresse plus lentement.

D'autre part un réglage du pas du gradient est nécessaire. En effet, une petite valeur de ce paramètre ralentit la progression de l'algorithme ; en revanche une grande valeur aboutit généralement à un phénomène d'oscillation autour de la solution.

Et descendre le long du gradient (la plus grande pente) n'est pas toujours optimal. Il peut s'avérer inefficace notamment lorsque la fonction de coût forme des vallées très étroites. L'algorithme oscille entre les deux bords de la vallée et la convergence se trouve ralentie. Pour pallier à cet inconvénient, nous avons recours à l'utilisation de méthodes plus performantes. Elles sont dites du second ordre parce qu'elles prennent en considération la dérivée seconde de la fonction de coût. Nous présentons ci-dessous celles que nous avons mises en œuvre dans notre travail.

6.2. Méthode du gradient conjugué

6.2.1. Principe de la méthode

La méthode du gradient conjugué est également basée sur le même principe que celle du gradient simple, elle calcule de façon itérative le vecteur β des paramètres inconnus par la formule :

$$\beta^{i+1} = \beta^i - t^i * v^i$$

Cependant le pas de temps et la direction de descente varie à chaque itération :

- Le pas de temps varie selon le gradient et la Hessienne de la fonction de coût.

$$t^i = \frac{-\nabla D(\beta^i) * v^i}{(Hv^i) * v^i}$$

- La direction de descente varie également tel que les vecteur v^{i+1} et v^i soient orthogonaux :

$$v^{i+1} = \nabla D(\beta^i) + \frac{\|\nabla D(\beta^i)\|}{\|\nabla D(\beta^{i-1})\|} v^i$$

On choisira la norme euclidienne.

6.2.2. Choix du critère d'arrêt

Nous avons choisi un critère d'arrêt sur la norme du gradient de la fonction de coût D . L'algorithme du gradient s'arrête lorsque la norme du gradient est suffisamment petite, c'est-à-dire lorsque :

$$\|\nabla D(\beta)\| \leq eps$$

Cette méthode permet de parcourir efficacement une vallée étroite, chose que ne peut pas faire la méthode du gradient simple.

Cette méthode permet aussi de converger très vite vers un minimum, cependant nous sommes confrontés aux contraintes sur les paramètres inconnus. Nous proposons de voir deux méthodes, la méthode de pénalisation et la méthode du gradient projeté, celle qui nous permettra de résoudre ce problème de contraintes d'optimisation.

6.3. Méthodes de pénalisations

L'intérêt de ces méthodes est la simplicité de leurs principes et leurs relatives efficacités pratiques. Le concept de base consiste à remplacer un problème avec contraintes par un problème sans contrainte dont la fonction de coût « décourage » les points non admissibles.

La fonction de coût (3) du problème (P) est alors remplacée par la fonction suivante à minimiser :

$$\varphi(\beta, r) = D(\beta) + r \times h(\beta)$$

Où $h(\beta)$ est la fonction pénalité dépendant des contraintes et r est un coefficient de pénalité. La fonction de pénalité est choisie de telle sorte que les contraintes soient garanties.

Suivant le type de contraintes et le type de fonction $h(\beta)$, on distingue la méthode des pénalités intérieures et la méthode des pénalités extérieures.

6.3.1. La méthode des pénalités extérieures

La fonction $h(\beta)$ est utilisée afin de défavoriser les positions non admissibles. Elle doit être continue et dérivable.

On choisira :

$$h(\beta) = \max(0, -\beta)^2 + \max\left(0, \frac{\beta}{\text{borne}} - 1\right)^2$$

Cette méthode est résolue de manière itérative, une suite de valeurs croissantes de r est générée et le problème d'optimisation suivant est résolu :

$$\varphi(\beta, r) = D(\beta) + r \times \max(0, -\beta)^2 + r \times \max\left(0, \frac{\beta}{\text{borne}} - 1\right)^2$$

6.3.2. La méthode des pénalités intérieures

Dans la méthode des pénalités intérieures, on cherche à définir la fonction $h(\beta)$ de telle sorte que plus le point se rapproche de la frontière du domaine admissible, plus la fonction de pénalisation $h(\beta)$ croît et tend vers l'infini et par conséquent, moins on a de chance de trouver le minimum proche de la frontière du domaine admissible.

On choisira :

$$h(\beta) = -\frac{1}{\beta}$$

Cette méthode est aussi résolue de manière itérative, une suite de valeurs décroissantes de r est générée et le problème d'optimisation suivant est résolu :

$$\varphi(\beta, r) = D(\beta) - r \times \frac{1}{\beta}$$

Pour l'algorithme on appliquera la méthode des pénalités intérieures pour les paramètres β_2 et β_4 sur la borne inférieure, qui doivent être toujours supérieurs à 0. La méthode des pénalités extérieures sera utilisée pour le reste des paramètres sur la borne inférieure et supérieure.

Voici un algorithme qui utilise la méthode du gradient simple couplée aux deux méthodes de pénalisations (algorithme à améliorer) :

Etape 1 : Initialisation

- ϵ : précision de l'algorithme
- t : pas fixe de la méthode du gradient simple
- $itmax$: nombre d'itération maximal du gradient simple
- borne : vecteur des valeurs maximales acceptables des paramètres
- β : choix des paramètres à l'intérieur du compact

Etape 2 :

- $i = 1$ et $\nabla D(\beta) = \text{eps} + 1$
- **Tant que** $\text{norm}(\nabla D(\beta) > \text{eps})$ et $i < \text{itmax}$
 - fonction **penalInt** au point β_2 et β_4
 - **Si** $\beta_1 < 0$ ou $\beta_3 < 0$ ou $\beta_5 < 0$ ou $\beta_1 > \text{borne}_1$ ou $\beta_2 > \text{borne}_2$ ou $\beta_3 > \text{borne}_3$ ou $\beta_4 > \text{borne}_4$ ou $\beta_5 > \text{borne}_5$
 - Si $\beta_2 > \text{borne}_2$ alors fonction **penalInt** au point β_2
 - Si $\beta_4 > \text{borne}_4$ alors fonction **penalInt** au point β_4
 - Si $\beta_1 < 0$ ou $\beta_1 > \text{borne}_1$ alors fonction **penalInt** au point β_1
 - Si $\beta_3 < 0$ ou $\beta_3 > \text{borne}_3$ alors fonction **penalInt** au point β_3
 - Si $\beta_5 < 0$ ou $\beta_5 > \text{borne}_5$ alors fonction **penalInt** au point β_5
 - **Sinon**
 - $\beta^{i+1} = \beta^i - t * \nabla D(\beta)$
 - **Fin si**
 - $i = i + 1$
- **Fin tant que**
- On retourne β

Algorithme Pénalisation extérieure : penalExt

- Choix de r et eps1
- $\nabla D(\beta) = \text{eps} + 1$
- **Tant que** $\beta_i < 0$ ou $\beta_i > \text{borne}_i$
 - **Tant que** $\text{norm}(\nabla D(\beta) > \text{eps1})$
 - g est la i -ième composante de $\nabla D(\beta)$:
$$g(\beta_i) = g(\beta_i) + r * (1/\beta_i^2)$$
 - $\beta^{i+1} = \beta^i - t * \nabla D(\beta)$
 - **Fin tant que**
 - $r = r + 5$
 - $\text{eps1} = \text{eps1}/5$
- **Fin tant que**
- On retourne β

Algorithme Pénalisation intérieure : penalInt

- Choix de r
- $c = 0$
- **Tant que** $\text{abs}(c - \beta_i) > 0$ et $\beta_i > 0$
- $c = \beta_i$
 - g est la i -ième composante de $\nabla D(\beta)$:

$$g(\beta_i) = g(\beta_i) - 2 * r * \max(0, -\beta_i) + 2 * r * \max(0, (\beta_i/\text{borne}_i) - 1)$$

- $\beta^{i+1} = \beta^i - t * \nabla D(\beta)$
- **Fin tant que**
- $r = r - 3$
- **Fin tant que**
- On retourne β

6.4. Méthode du gradient conjugué projeté

Pour résoudre le problème d'optimisation sous contraintes, nous allons adapter l'algorithme du gradient conjugué vu précédemment de façon à respecter les contraintes à chaque pas de l'itération.

Comme les contraintes sont de la forme :

$$0 \leq \beta_1 \leq 1 \quad \text{et} \quad 0 \leq \beta_j \leq b \quad \text{pour} \quad j = 2 \dots 5$$

Nous pouvons trouver un ensemble K de la forme d'un pavé qui contiendra la fonction de coût :

$$K = [0, 1] * \prod_{j=1}^4 [0, b_j]$$

Dans ce cas, il est facile d'exprimer le projeté d'un vecteur sur K .

Si on prend P_k le projeté sur K tel que :

$$\beta^{i+1} = P_k(\beta^i - t^i * v^i)$$

Alors le vecteur β^{i+1} a pour composantes :

$$\beta_1^{i+1} = \min(\max(0, \beta_1^i - t^i * v_1^i), 1)$$

$$\text{Et } \beta_j^{i+1} = \min(\max(0, \beta_j^i - t^i * v_j^i), b_j) \quad \text{pour } j = 2 \dots 5$$

Cependant la méthode du gradient conjugué projeté ne permet de trouver qu'un minimum local. Pour remédier à ce problème, nous allons coupler cette technique de résolution à une méthode stochastique. Nous allons proposer d'initialiser les paramètres avec une méthode de Monte Carlo.

6.5. Méthode du gradient conjugué projeté stochastique

La méthode de Monte Carlo consiste à choisir des points aléatoirement selon une distribution (en général une loi uniforme ou normale) à chaque itération. La fonction de coût est évaluée en ce point. Si elle est meilleure que le minimum courant, cette valeur est enregistrée, ainsi que la solution correspondante et le processus continue jusqu'à que les conditions d'arrêt soient vérifiées.

Pour comparer deux minimums, le critère utilisé est la somme des carrés des résidus :

$$\text{critère} = \sum_{i=1}^n \varepsilon_i^2$$

Cependant cette méthode demande un nombre important d'évaluations de la fonction de coût. La qualité du minimum trouvé dépendra du nombre d'initialisation effectué.

Plus le nombre d'initialisation suivant une loi uniforme sera grande, meilleur sera le minimum. Toutefois cette méthode ne garantit pas la localisation du minimum global.

Pour notre étude nous avons choisie d'initialiser les paramètres avec une loi uniforme.

Voici l'algorithme du gradient conjugué projeté stochastique :

Etape 1 : Initialisation

- eps : précision de l'algorithme
- $itmax$: nombre d'itération maximal du gradient conjugué projeté
- $binf$: vecteur des valeurs minimales acceptables des paramètres
- $bsup$: vecteur des valeurs maximales acceptables des paramètres
- n : nombre d'itération de la méthode de Monte Carlo

Etape 2 :

- On fixe $p > grande\ valeur$
- **Pour** $j = 1 \dots n$
 - $i = 1$
 - On initialise β^i suivant la loi uniforme $U([0, 1]) \times bsup$
 - calcul de $v^i = -\nabla D(\beta^i)$
 - **Tant que** $norm(\nabla D(\beta^i)) > eps$ et $i < itmax$
 - Calcul du nouveau pas de temps :
 -

$$t^i = \frac{-\nabla D(\beta^i) * v^i}{(Hv^i) * v^i}$$

- Calcul du nouveau β :

$$\beta^{i+1} = \min(\max(bin f, \beta^i - t^i * v^i), bsup)$$

- Calcul de la nouvelle direction :

$$v^{i+1} = -\nabla D(\beta^i) + \frac{norme(\nabla D(\beta^{i+1}))^2}{norme(\nabla D(\beta^i))^2} v^i$$

- $i = i + 1$
- **Fin tant que**
- Si $D(\beta^{i+1}) < p$ alors $minimum = \beta^{i+1}$
- **Fin pour**
- On retourne le *minimum*

7. Mise en Application

7.1. Exemple théorique

Afin de tester l'algorithme (voir annexe), nous allons chercher le minimum d'une fonction de coût qui est plus simple et dans un périmètre plus restreint.

Nous avons choisi la fonction de coût suivante :

$$f(x, y) = \sin(5.5x)\cos(5y) + x^2 + 1$$

La fonction f est continue et indéfiniment dérivable sur \mathbb{R}^2 , nous effectuerons notre recherche de minimum sur le compact $[0, 2] \times [0, 2]$.

Pour mieux visualiser cette fonction, nous la représentons :

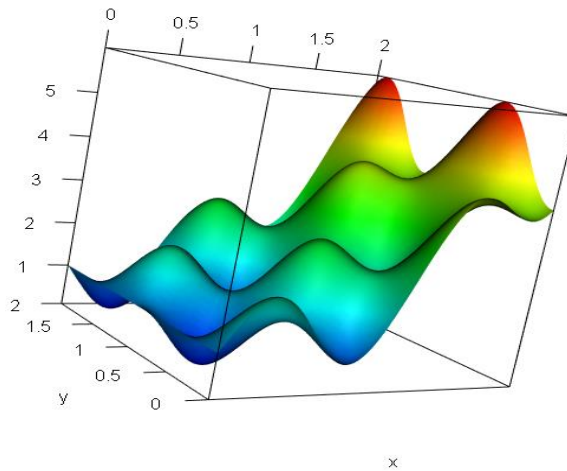


Figure 13, Représentation 3D de la fonction f

Nous utiliserons le gradient et la matrice Hessienne de la fonction de coût :

- gradient de f :

$$\nabla f = \left(\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y} \right) = (5.5\cos(5.5x)\cos(5y) + 2x \quad -5\sin(5.5x)\sin(5y))$$

- Hessienne de f :

$$H = \begin{pmatrix} -30.25\sin(5.5x)\cos(5y) + 2 & -27.5\sin(5y)\cos(5.5x) \\ -27.5\sin(5y)\cos(5.5x) & -25\sin(5.5x)\cos(5y) \end{pmatrix}$$

Cette fonction possède plusieurs minimums locaux :

- $(1.34 \quad 0.63)$ et $(1.34 \quad 1.88)$ qui valent 1.91
- $(0.27 \quad 1.88)$ et $(0.27 \quad 0.63)$ qui valent 0.08 et correspondent aux minimums globaux
- $(1.86 \quad 0)$ et $(1.86 \quad 1.26)$ qui valent 3.74
- $(0.8 \quad 0)$ et $(0.8 \quad 1.26)$ qui valent 0.69

En appliquant la méthode du gradient conjugué projeté vu précédemment, on tombe généralement sur l'un des minimums locaux de la fonction f .

Sachant que f est une fonction analytique, pour être sûr que le point trouvé est un minimum local on vérifie que la matrice Hessienne est bien définie positif. Pour cela, on vérifie que les déterminants des sous-matrices principales de H sont strictement positifs.

Ainsi en initialisant les paramètres avec une loi uniforme et en itérant l'algorithme un nombre de fois suffisant ($n \approx 20$), on retrouve toujours un des minimums globaux.

7.2. Cas réel

7.2.1. Existence d'un minimum de la fonction de coût

La fonction de coût $D : R^5 \rightarrow R$ est continue.

K , l'ensemble choisi sur l'espace des variables est un compact. Il s'agit d'un ensemble fermé borné non vide. Alors le problème (P) admet au moins une solution.

La fonction de coût D est dérivable de classe C^2 .

Nous pouvons utiliser la méthode du gradient conjugué pour trouver un minimum de la fonction de coût D .

7.2.2. Choix des bornes des paramètres

Pour la borne inférieure on prendra un nombre positif proche de 0 pour éviter les divisions par zéro. Dans notre étude nous prendrons 0.001 pour tous les paramètres.

Sachant que le temps moyen de séjour pour le Dhuy et la Loire est connu. Il est d'environ 3 jours pour la Loire et d'environ une année pour le Dhuy. De plus n'ayant pas d'approximation sur le nombre de Peclet sur ces deux cours d'eau, nous prendrons comme borne supérieure 400 pour les paramètres ($\beta_2, \beta_3, \beta_4, \beta_5$) et 1 pour β_1 .

7.2.3. Choix du nombre d'initialisation de la méthode de Monte Carlo

La fonction de l'exemple théorique admet 3 minimums globaux sur un ensemble compact $[0, 2] \times [0, 2]$. Grâce à l'algorithme on trouve toujours un des minimums globaux au bout d'environ 20 itérations. Sachant cette approximation sur l'exemple théorique, on peut la transposer sur le cas réel pour obtenir une approximation du nombre d'itération au bout duquel on pourrait probablement toujours tomber sur le minimum global.

Comme l'espace de recherche est de taille $[0, 400] \times [0, 400] \times [0, 400] \times [0, 400]$, on peut donner une approximation du nombre d'itération, soit environ $128 \cdot 10^7$.

Cependant nous effectuerons seulement 10^7 itérations, pour cause le nombre de calcul qui peut prendre jusqu'à environ 20 heures pour obtenir les résultats.

7.2.4. Résultats

Pour chaque année, à partir de l'année 2000 jusqu'à l'année 2010, nous possédons en moyenne 12 points. Nous ne tiendrons pas compte de l'année 2011 car elle représente seulement 2 points. Dans notre étude nous effectuerons une estimation des paramètres selon l'année.

Les résultats sont regroupés sur les tableaux suivants :

Année	Nombre d'observation	Durée d'exécution (heures)	Nombre d'itération (gradient conjugué)
2000	11	9	500
2001	10	31.2	1000
2002	10	13.7	1000
2003	11	21.5	2000
2004	12	39.5	2000
2005	12	8.6	1000
2006	8	10	1000
2007	18	11.7	1000
2008	34	21	1000
2009	23	29.4	1000
2010	15	24	1000

Figure 14, Sorties de l'algorithme

Année	α	Pe_L	t_L	Pe_D	t_D
2000	0.331925572	151.5137	0.9910726	152.1958	18.27820
2001	0.7758762	262.5732	1.933719	344.3928	3.076463
2002	0.199444319	371.2797	2.003214	223.9710	9.170581
2003	0.870294832	301.7652	4.963710	273.3622	6.547701
2004	0.950486061	195.3589	17.46259	202.7169	3.128063
2005	0.5133363	339.1374	5.956757	4.279034	209.2755
2006	0.8051100	399.6124	5.817969	28.41932	2.888086
2007	0.66390674	60.65074	0.8345354	241.5386	5.399036
2008	0.29952632	197.4399	1.0229093	34.87205	3.235315
2009	0.627639040	176.6125	1.133761	2.525827	1.3292112
2010	0.879334493	103.2985	4.1910961	289.3494	14.31161

Figure 15, Estimation des paramètres selon l'année

Selon l'année, nous avons représenté sur le graphe ci-dessous, la courbe des mesures réelles en vert et le meilleur ajustement déterminé par la méthode du gradient conjugué projeté stochastique représenté en bleu :

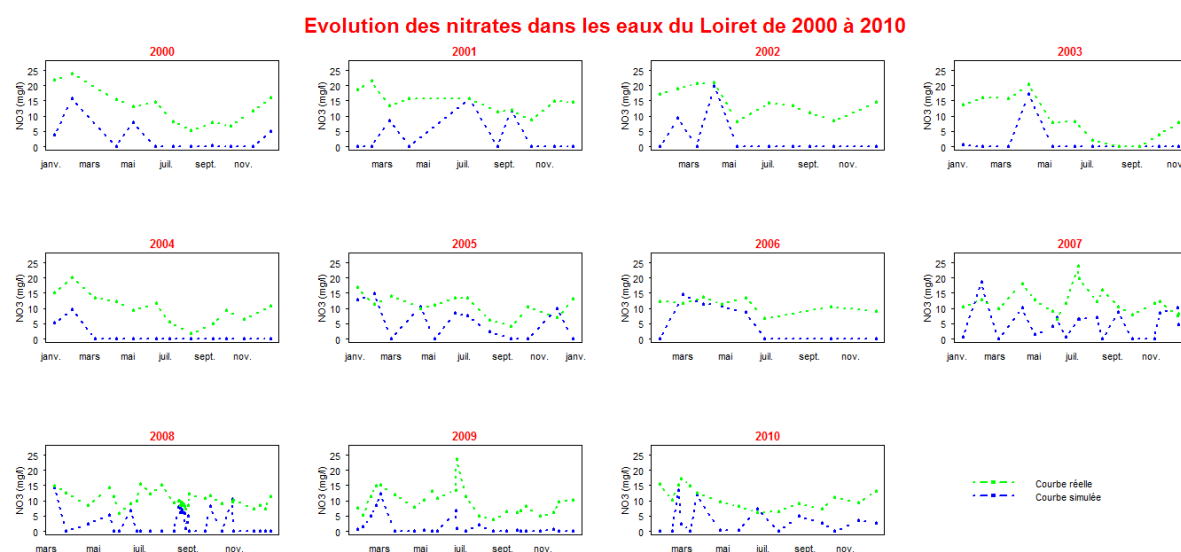


Figure 16, Evolution de la concentration des nitrates, courbes simulées et courbes réelles selon l'année

Pour la simulation du modèle en 2008, on observe :

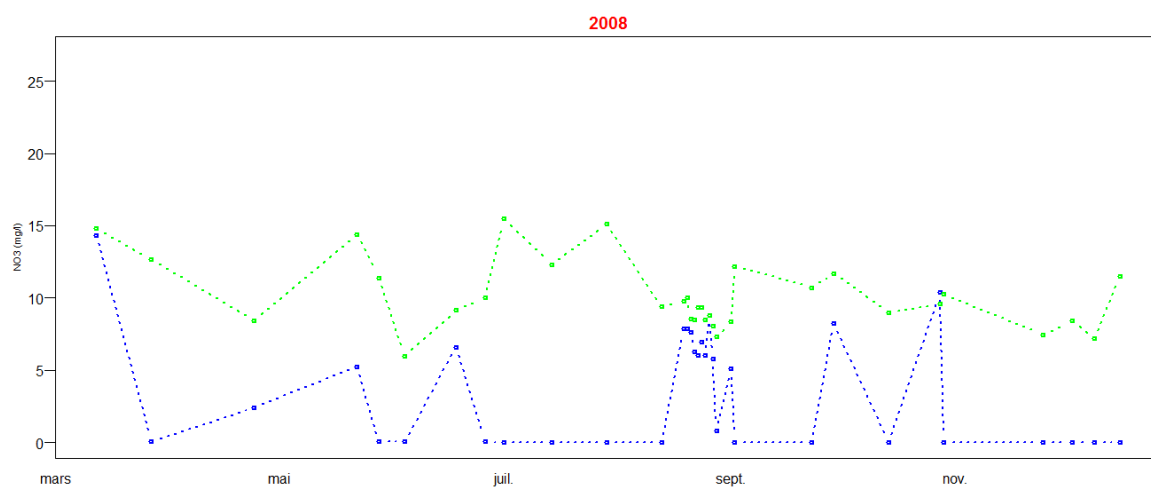


Figure 17, Evolution de la concentration des nitrates, courbe simulée et courbe réelle pour l'année 2008

Et enfin pour la simulation du modèle en 2009, on a :



Figure 18, Evolution de la concentration des nitrates, courbe simulée et courbe réelle pour l'année 2009

On observe parfois une tendance comme par exemple sur le graphe de l'année 2000 et 2005. On arrive à obtenir un bon ajustement de certains points mais en général on observe une courbe simulée en dents de scies. L'erreur entre le modèle estimé et la réalité est assez grande.

Plusieurs causes peuvent probablement justifier cet ajustement :

- Le modèle cinétique ne prend pas en compte tous les phénomènes naturels complexes d'où des erreurs assez importantes.
- Il n'y a pas assez de points dans chaque échantillon.
- Pour chaque cours d'eau tous les prélèvements ne sont pas effectués sur le même lieu.
- Le nombre d'itérations de la méthode de Monte Carlo n'est pas encore suffisant pour parcourir tout le périmètre de recherche.

On a également tracé l'évolution des différents paramètres, ainsi que la répartition de la concentration en nitrate provenant du Dhuy et de la Loire sur les graphes suivants de l'année 2000 à 2010 :

Evolution des paramètres estimés de 2000 à 2010

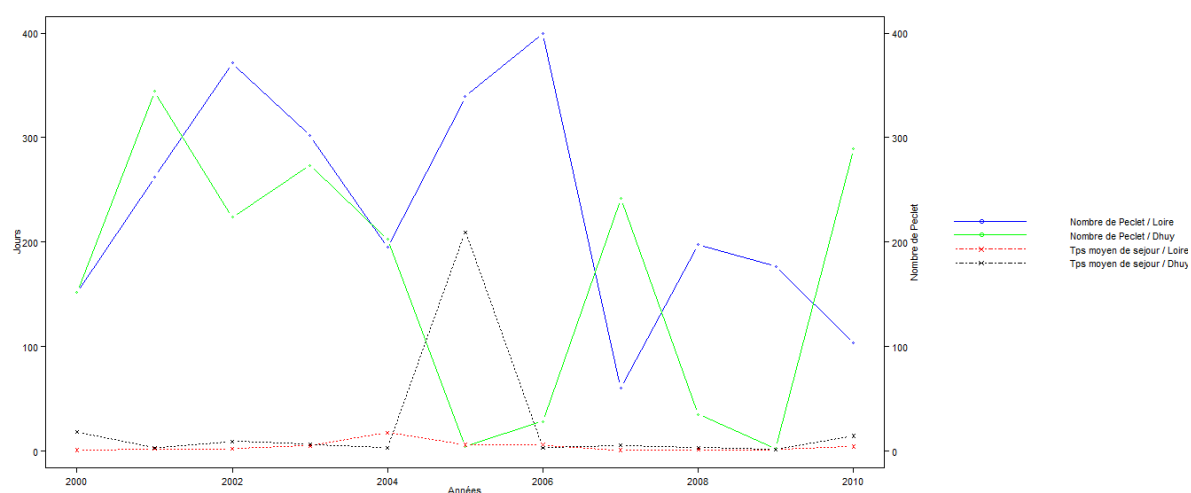


Figure 19, Evolution des paramètres de 2000 à 2010

Il est intéressant d'observer que le temps moyen de séjour de la Loire varie entre 1 jour et 17 jours avec une moyenne d'environ 4 jours. Ce qui correspond au temps moyen de séjour déjà connu, c'est-à-dire de 3 jours environ.

Cependant le temps moyen de séjour du Dhuy connu est d'environ 1 an, ce qui ne correspond pas aux résultats puisque le temps moyen de séjour du Dhuy obtenu varie entre 1 jour et demi et 18 jours avec une valeur maximale qui atteint 210 jours.

Nous pouvons aussi observer l'estimation de la répartition de la concentration en nitrate provenant de la Loire et du Dhuy :

Estimation de la répartition de la concentration en nitrate provenant de la Loire et du Dhuy entre 2000 et 2010

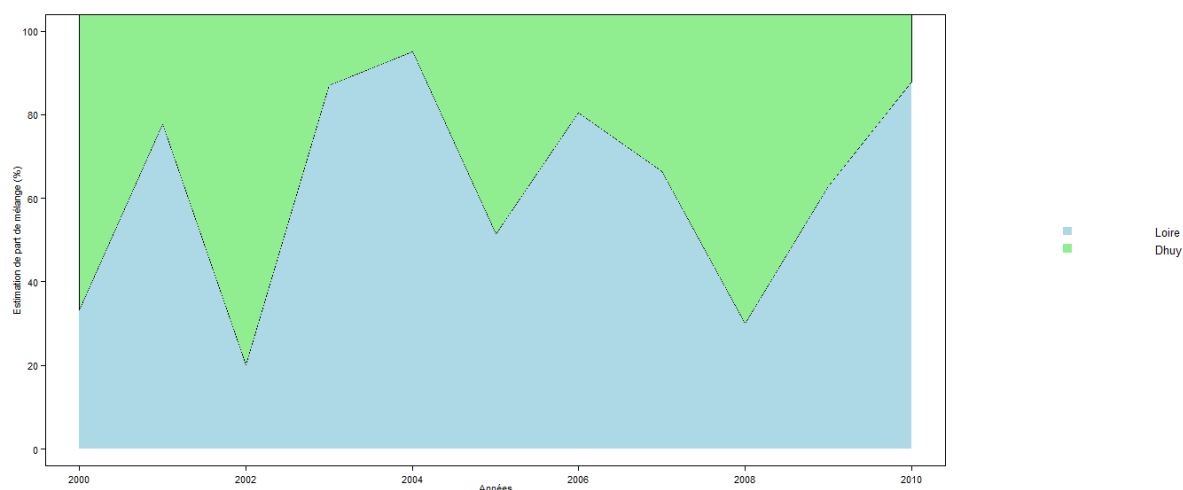


Figure 20, Evolution de l'estimation de la répartition de la concentration en nitrate provenant de la Loire et du Dhuy

La contribution en nitrate provenant du Dhuy varie d'environ de 10% à 80% et celle de la Loire varie d'environ de 20% à 90%. En général la contribution en nitrate provenant de la Loire est plus importante que celle du Dhuy. Hormis la contribution en nitrate provenant du Dhuy élevée en 2002

avec 80% et en 2008 avec 70%, les résultats de l'estimation de la répartition de la concentration en nitrate entre la Loire et le Dhuy reflètent assez bien la réalité.

7.2.5. Améliorations futures

Nous avons obtenu des estimations des paramètres du modèle qui ne nous semble pas suffisamment précis. Pour y remédier, nous pouvons procéder à d'autres simulations avec quelques modifications :

- simuler des données en plus de celles déjà existantes.
- Nous pouvons également augmenter le nombre d'itérations de la méthode de Monte Carlo afin d'augmenter la probabilité de tomber sur d'autres minimums, s'ils existent et de meilleure qualité que ceux déjà trouvés.
- Nous pouvons aussi augmenter les bornes supérieures du nombre de Peclet pour la Loire et le Dhuy. Et simultanément réduire la borne supérieure du temps moyen de séjour à la Loire, puisque nous avons trouvé que ce paramètre varie de 1 jour à 17 jours et que cela correspond bien à la réalité.
- Nous pouvons aussi augmenter le nombre d'itérations de la méthode du gradient conjugué pour améliorer la précision de l'estimation des paramètres.

7.2.6. Analyse des résidus

L'étude de l'analyse des résidus consiste à vérifier les hypothèses du modèle (1) formulées pour l'erreur ε :

$$\varepsilon = U - f((X, Y, dt); \beta)$$

Rappelons ces hypothèses : les erreurs sont des variables aléatoires indépendantes et identiquement distribuées selon une loi normale de moyenne nulle et de variance constante σ^2 .

Si le modèle postulé correspond bien aux données observées, alors les résidus observés devraient refléter les hypothèses de base que nous avons spécifiées pour ε .

7.2.6.1. Normalité des résidus

Pour tester la normalité des résidus, il existe plusieurs outils statistiques. Nous utiliserons le test de Shapiro-Wilk pour tester l'hypothèse nulle selon laquelle les échantillons choisis sont issus d'une population normalement distribuée, nous tracerons les « Q-Q plot » et également les histogrammes lorsque les observations des différents échantillons sont suffisantes.

Nous testons l'hypothèse nulle

$$H_0 = \text{l'échantillon est normalement distribué}$$

Contre l'hypothèse alternative

$$H_1 = \text{l'échantillon n'est normalement distribué}$$

On récapitule sur le tableau suivant, les résultats du test Shapiro-Wilk sur les différents échantillons :

Année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
p-value	0.3447	0.5018	0.9183	0.2529	0.4286	0.4455	0.9231	0.9511	0.05985	0.1007	0.6954

Figure 21, Résultats du test de Shapiro-Wilk

L'échantillon de l'année 2008 possède une p – value comprise entre 0.05 et 0.1, il y a une faible présomption contre l'hypothèse nulle.

Pour le reste des échantillons, on obtient une toujours p – value > 0.1 , il n'y a pas de présomption contre l'hypothèse nulle.

Alors on ne rejette pas l'hypothèse H_0 au niveau de signification 0.05. Les échantillons sont normalement distribués avec un doute sur les échantillons de 2008 et 2009

Nous avons tracé les histogrammes des résidus pour l'année 2008 et 2009 :

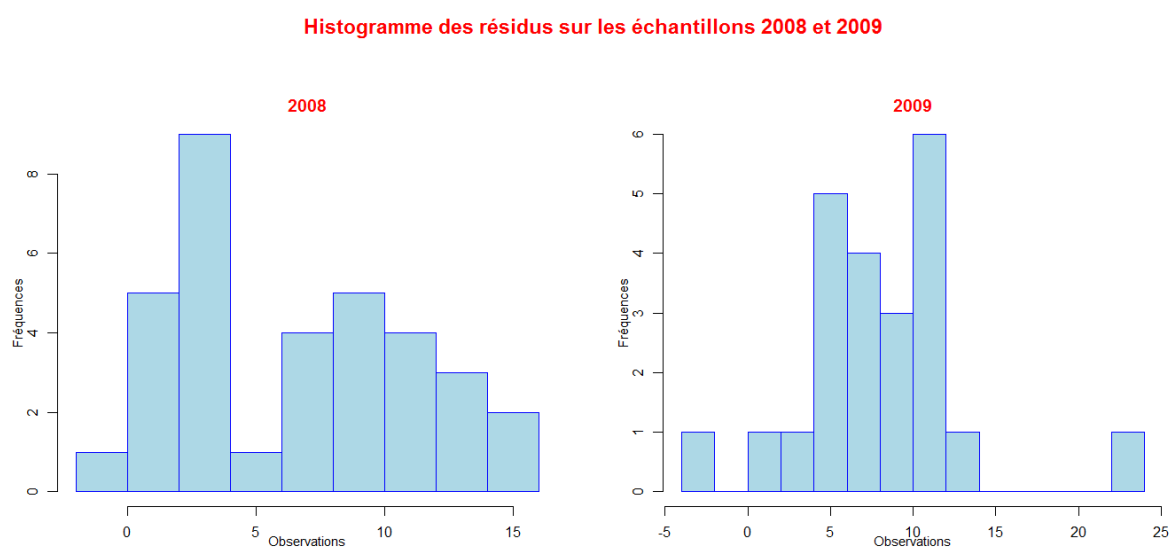


Figure 22, Histogrammes des résidus de l'année 2008 et 2009

Les échantillons des années 2008 et 2009 contiennent le plus d'observations, avec 34 observations en 2008 et 23 observations en 2009. On observe ainsi un début de loi normale pour l'échantillon de 2009 cependant sur l'histogramme de l'échantillon 2008, on observe un histogramme en forme de deux lois normales.

Toutefois, en prenant en compte les 166 observations de l'année 2000 à 2011, on obtient un histogramme qui décrit bien une loi normale.

Afin d'appuyer ce test nous pouvons tracer les « Q-Q plot » pour comparer les distributions observées avec une loi normale. Ils sont regroupés sur le graphe suivant :

Diagramme Q-Q plot

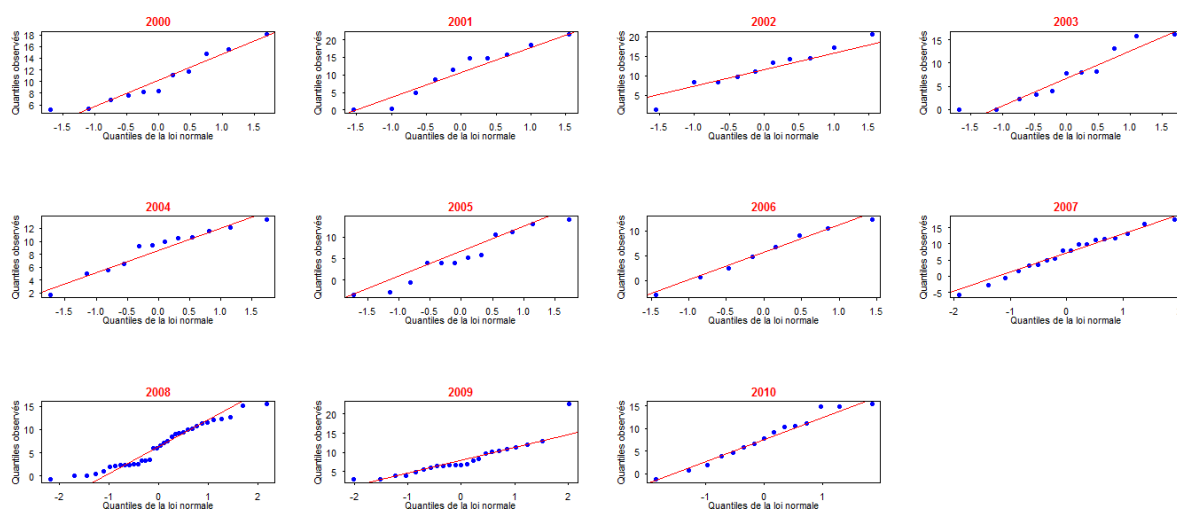


Figure 23, Diagrammes Q-Q plot

Sur les différents graphiques, les points sont généralement alignés sur la droite en rouge, avec parfois un léger biais apparaissant aux valeurs extrêmes.

On observe également des points suivant une forme spirale autour de la droite en rouge. Le résultat est assez satisfaisant, la distribution des résidus suit une loi normale.

7.2.6.2. Indépendance des résidus

Le graphique suivant représente les résidus estimés contre les valeurs prédites par le modèle estimé.

Résidus estimés contre les valeurs prédites

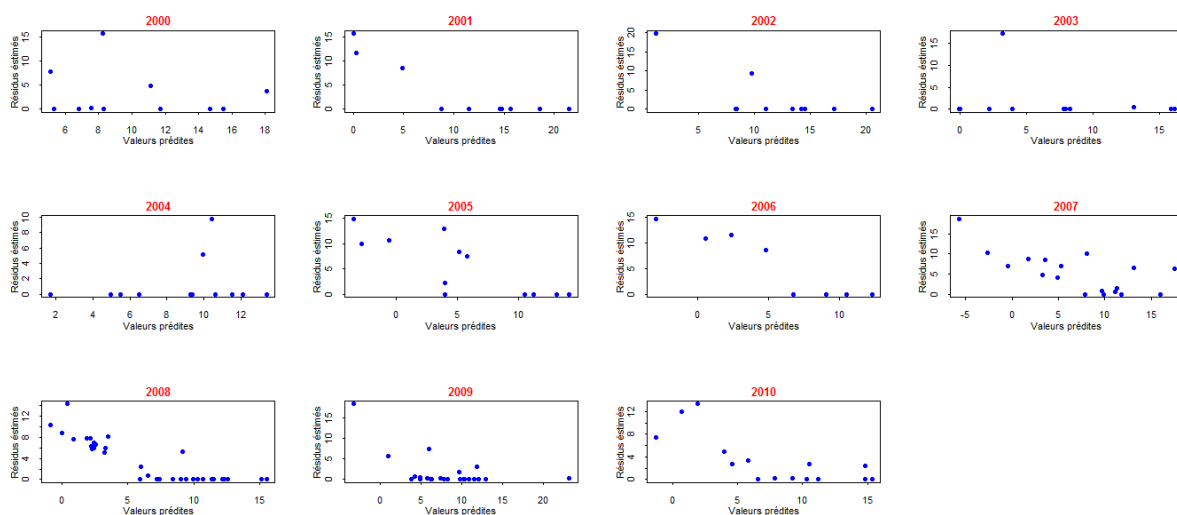


Figure 24, Résidus estimés contre les valeurs prédites

Malgré le manque d'observation dans certains échantillons, on observe un nuage de points plus ou moins dispersés ne présentant aucune structure le long de l'axe des ordonnées ou de forme particulière.

Cependant sur les échantillons de 2008 et 2009, on remarque un regroupement de points, on pourrait penser que les observations ne sont pas indépendantes.

7.2.6.3. Homogénéité des variances des résidus

Pour vérifier l'homogénéité des variances des résidus, nous utilisons également le graphe précédent, des résidus estimés contre les valeurs prédites.

La variance des résidus a parfois tendance à croître avec les valeurs prédites, elle doit toujours rester du même ordre de grandeur pour garantir l'homogénéité des variances des résidus. On peut observer sur le graphe précédent que les écarts ne croient pas en fonction des valeurs prédites, on peut dire que les résidus sont homogènes.

8. Conclusion

Le modèle cinétique étudié permet de trouver la concentration en nitrate du Loiret sachant celle de la Loire et du Dhuy à un certain temps. Nous avons mis en œuvre la méthode du gradient conjugué projeté couplé à la méthode de Monte Carlo afin d'estimer les paramètres du modèle cinétique. Malgré l'erreur qui est assez grande, nous observons des tendances entre les mesures réelles et les mesures simulées. Nous avons également remarqué que le temps moyen de séjour de la Loire simulé correspond assez bien à la réalité et la part de mélange en concentration de nitrate entre le Dhuy et la Loire correspond également.

9. Annexe

9.1. Mise en œuvre du modèle cinétique [9]

9.1.1. Modèle d'advection dispersion

Le bilan de masse réalisé sur un volume élémentaire montre que la concentration $c(x, t)$ répond à l'équation de transport diffusion suivante :

$$\frac{dC}{dt} + u \frac{dC}{dx} = D \frac{d^2C}{dx^2} \quad (1)$$

Où u est la vitesse d'écoulement de l'eau (m/s) et D (m²/s) la capacité de dispersion hydrodynamique. Si l'on considère le transport au travers d'un réservoir en une dimension de longueur L (m) et de surface A , il existe des solutions analytiques de l'équation (1). En résolvant l'équation du transport diffusion pour un réservoir ouvert des deux côtés, les variations temporelles de la masse ($M(t)$) de soluté ou de sa concentration ($c(t)$) à la distance l , causée par l'injection brève d'une masse M_0 peut-être décrite ainsi :

$$M(t) = \frac{c(t)}{A} = \frac{M_0}{\sqrt{4\pi Dt}} e^{-\frac{(L-ut)^2}{4Dt}} \quad \text{Wilson (1978)}$$

La concentration est fonction des propriétés u , L et D du milieu.

Si l'on définit le nombre de Peclet comme :

$$Pe = \frac{uL}{D} \quad (2)$$

Et le temps moyen de séjour comme :

$$\bar{t} = \frac{L}{u}$$

La concentration en fonction du temps pour une injection brève de masse M_0 dans un réservoir peut être écrite sous la forme suivante :

$$C(t) = \frac{M_0}{A} \sqrt{\frac{Pe}{4\pi t}} e^{-\frac{Pe(\bar{t}-t)^2}{4t}} \quad (3)$$

La masse de soluté est :

$$M(t) = M_0 \sqrt{\frac{Pe}{4\pi t}} e^{-\frac{Pe(\bar{t}-t)^2}{4t}} \quad (4)$$

Il est donc possible de décrire la concentration ou la masse sortant d'un réservoir en fonction du temps en connaissant simplement son nombre de Peclet et son temps moyen de séjour.

9.1.2. Modèles conceptuels des écoulements autour du forage

Pour décrire l'évolution des concentrations en fonction du temps dans le forage d'Orléans, nous allons découper l'aquifère (= réserve d'eau souterraine) qui alimente ce forage en 2 réservoirs distincts, un alimenté par la Loire, l'autre par le Dhuy. Les indices L et D feront respectivement

appellent au réservoir provenant de la Loire et Dhuy. L'eau qui s'infiltre va percoler à travers ces deux réservoirs dont les propriétés hydro-dispersives (Pe_L et Pe_D) et advectives (t_L et t_D) sont fixées.

La concentration au forage est un mélange, la masse de soluté au forage est la somme de celle provenant de la Loire et de celle provenant des eaux de surface (Dhuy).

$$M_{f(t)} = M_{L(t)} + M_{D(t)} \quad (5)$$

Ainsi en réutilisant l'équation 4 pour un réservoir venant de la Loire et un second venant du Dhuy, on écrit avec les indices L et D:

$$M(t) = M_{L_0} \sqrt{\frac{Pe_L}{4\pi t}} e^{\frac{-Pe_L(t_L-t)^2}{4t}} + M_{D_0} \sqrt{\frac{Pe_D}{4\pi t}} e^{\frac{-Pe_D(t_D-t)^2}{4t}}$$

La masse est produite du flux d'eau ($Q(t)$) avec la concentration $C(t)$ que nous sommes capable de mesurer, on peut donc dire que :

$$C_{f(t)} Q_{f(t)} = C_{L(0)} Q_{L(t)} \sqrt{\frac{Pe_L}{4\pi t}} e^{\frac{-Pe_L(t_L-t)^2}{4t}} + C_{D(0)} Q_{D(t)} \sqrt{\frac{Pe_D}{4\pi t}} e^{\frac{-Pe_D(t_D-t)^2}{4t}}$$

Et $Q_{f(t)} = Q_{L(t)} + Q_{D(t)}$ (7)

Si l'on pose que $\alpha = \frac{Q_{L(t)}}{Q_{D(t)}}$ et $1 - \alpha = \frac{Q_{D(t)}}{Q_{L(t)}}$ (ce qui pourrait correspondre à une probabilité que l'eau ait pour origine la Loire), on écrit :

$$C_{f(t)} = C_{L(0)} \alpha \sqrt{\frac{Pe_L}{4\pi t}} e^{\frac{-Pe_L(t_L-t)^2}{4t}} + C_{D(0)} (1 - \alpha) \sqrt{\frac{Pe_D}{4\pi t}} e^{\frac{-Pe_D(t_D-t)^2}{4t}}$$

La fonction $C_{f(t)}$ décrit les concentrations observées au forage au temps t engendrées par une injection brève à $t=0$ d'eau ayant des concentrations $C_{L(0)}$ et $C_{D(0)}$ et suivant le mélange α . Ces concentrations, observées sur le terrain, sont données. Nous connaissons donc la concentration des eaux qui pénètre dans le système, $C_{L(t)}$ et $C_{D(t)}$, ainsi que celle qui en sort $C_{f(t)}$. Si au temps t , les deux rivières ont une concentration, $C_{L(t)}$ et $C_{D(t)}$, la part de cette masse entrée au temps t qui sortira à $t+\Delta t$ au forage peut s'écrire :

$$C_{f(t+\Delta t)} = C_{L(t)} \alpha \sqrt{\frac{Pe_L}{4\pi(t+\Delta t)}} e^{\frac{-Pe_L(t_L-(t+\Delta t))^2}{4(t+\Delta t)}} + C_{D(t)} (1 - \alpha) \sqrt{\frac{Pe_D}{4\pi(t+\Delta t)}} e^{\frac{-Pe_D(t_D-(t+\Delta t))^2}{4(t+\Delta t)}} \quad (9)$$

9.2. Dérivées de la fonction f du modèle cinétique

9.2.1. Dérivées premières de la fonction f

Le vecteur des dérivées partielles s'écrit :

$$D(\beta) = \left(\frac{\partial D}{\partial \beta_1} \quad \frac{\partial D}{\partial \beta_2} \quad \frac{\partial D}{\partial \beta_3} \quad \frac{\partial D}{\partial \beta_4} \quad \frac{\partial D}{\partial \beta_5} \right)$$

Les dérivées partielles se dérivent de la façon suivante :

$$\frac{\partial D}{\partial \beta} = -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \frac{\partial f}{\partial \beta}(x_i, y_i, \beta)$$

On pose $dt = t + \Delta t$

On obtient les dérivées partielles suivantes :

$$\frac{\partial D}{\partial \beta_1} = -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(x_i \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2(\beta_3-dt)^2}{4dt}} - y_i \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4(\beta_5-dt)^2}{4dt}} \right)$$

$$\frac{\partial D}{\partial \beta_2} = -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(x_i \beta_1 \frac{1}{2} \frac{1}{4\pi dt} \left(\frac{\beta_2}{4\pi dt} \right)^{-\frac{1}{2}} e^{\frac{-\beta_2(\beta_3-dt)^2}{4dt}} - \frac{(\beta_3-dt)^2}{4dt} x_i \beta_1 \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2(\beta_3-dt)^2}{4dt}} \right)$$

$$\frac{\partial D}{\partial \beta_3} = -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(-2 \frac{\beta_2(\beta_3-dt)}{4dt} x_i \beta_1 \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2(\beta_3-dt)^2}{4dt}} \right)$$

$$\frac{\partial D}{\partial \beta_4} = -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(y_i (1 - \beta_1) \frac{1}{8\pi dt} \sqrt{\frac{4\pi dt}{\beta_4}} e^{\frac{-\beta_4(\beta_5-dt)^2}{4dt}} - \frac{(\beta_5-dt)^2}{4dt} y_i (1 - \beta_1) \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4(\beta_5-dt)^2}{4dt}} \right)$$

$$\frac{\partial D}{\partial \beta_5} = -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(-2 \frac{\beta_4(\beta_5-dt)}{4dt} y_i (1 - \beta_1) \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4(\beta_5-dt)^2}{4dt}} \right)$$

Après simplification on obtient les 5 dérivées partielles :

$$\frac{\partial D}{\partial \beta_1} = \frac{1}{\sqrt{\pi dt}} \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(-x_i \sqrt{\beta_2} e^{\frac{-\beta_2(\beta_3-dt)^2}{4dt}} + y_i \sqrt{\beta_4} e^{\frac{-\beta_4(\beta_5-dt)^2}{4dt}} \right)$$

$$\frac{\partial D}{\partial \beta_2} = \frac{1}{2\sqrt{\pi dt}} \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) x_i \beta_1 e^{\frac{-\beta_2(\beta_3-dt)^2}{4dt}} \left(-\sqrt{\frac{1}{\beta_2}} + \frac{\sqrt{\beta_2}(\beta_3-dt)^2}{2dt} \right)$$

$$\frac{\partial D}{\partial \beta_3} = \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(x_i \beta_1 \frac{\beta_2(\beta_3-dt)}{dt} \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2(\beta_3-dt)^2}{4dt}} \right)$$

$$\frac{\partial D}{\partial \beta_4} = \frac{1}{2\sqrt{\pi dt}} \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) y_i (1 - \beta_1) e^{\frac{-\beta_4(\beta_5-dt)^2}{4dt}} \left(-\sqrt{\frac{1}{\beta_4}} + \frac{\sqrt{\beta_4}(\beta_5-dt)^2}{2dt} \right)$$

$$\frac{\partial D}{\partial \beta_5} = \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(y_i (1 - \beta_1) \frac{\beta_4 (\beta_5 - dt)}{dt} \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4 (\beta_5 - dt)^2}{4dt}} \right)$$

9.2.2. Dérivées seconde (Hessienne) de la fonction f

La matrice Hessienne de la fonction f s'écrit :

$$H = \begin{pmatrix} 0 & \frac{\partial D}{\partial \beta_1 \partial \beta_2} & \frac{\partial D}{\partial \beta_1 \partial \beta_3} & \frac{\partial D}{\partial \beta_1 \partial \beta_4} & \frac{\partial D}{\partial \beta_1 \partial \beta_5} \\ \frac{\partial D}{\partial \beta_2 \partial \beta_1} & \frac{\partial D}{\partial \beta_2 \partial \beta_2} & \frac{\partial D}{\partial \beta_2 \partial \beta_3} & 0 & 0 \\ \frac{\partial D}{\partial \beta_3 \partial \beta_1} & \frac{\partial D}{\partial \beta_3 \partial \beta_2} & \frac{\partial D}{\partial \beta_3 \partial \beta_3} & 0 & 0 \\ \frac{\partial D}{\partial \beta_4 \partial \beta_1} & 0 & 0 & \frac{\partial D}{\partial \beta_4 \partial \beta_4} & \frac{\partial D}{\partial \beta_4 \partial \beta_5} \\ \frac{\partial D}{\partial \beta_5 \partial \beta_1} & 0 & 0 & \frac{\partial D}{\partial \beta_5 \partial \beta_4} & \frac{\partial D}{\partial \beta_5 \partial \beta_5} \end{pmatrix}$$

On obtient les dérivées secondes suivantes :

$$\frac{\partial D}{\partial \beta_1 \partial \beta_2} = -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(x_i \frac{1}{2} \frac{1}{4\pi dt} \sqrt{\frac{4\pi dt}{\beta_2}} e^{\frac{-\beta_2 (\beta_3 - dt)^2}{4dt}} - \frac{(\beta_3 - dt)^2}{4dt} x_i \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2 (\beta_3 - dt)^2}{4dt}} \right)$$

$$\frac{\partial D}{\partial \beta_1 \partial \beta_3} = -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(-2 \frac{\beta_2 (\beta_3 - dt)}{4dt} x_i \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2 (\beta_3 - dt)^2}{4dt}} \right)$$

$$\frac{\partial D}{\partial \beta_1 \partial \beta_4} = -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(-y_i \frac{1}{2} \frac{1}{4\pi dt} \sqrt{\frac{4\pi dt}{\beta_4}} e^{\frac{-\beta_4 (\beta_5 - dt)^2}{4dt}} + \frac{(\beta_5 - dt)^2}{4dt} y_i \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4 (\beta_5 - dt)^2}{4dt}} \right)$$

$$\frac{\partial D}{\partial \beta_1 \partial \beta_5} = -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(2 \frac{\beta_4 (\beta_5 - dt)}{4dt} y_i \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4 (\beta_5 - dt)^2}{4dt}} \right)$$

$$\begin{aligned} \frac{\partial D}{\partial \beta_2 \partial \beta_2} = & -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(-x_i \beta_1 \frac{1}{2} \frac{1}{8\pi dt} \frac{4\pi dt}{\beta_2^2} \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2 (\beta_3 - dt)^2}{4dt}} \right. \\ & - \frac{(\beta_3 - dt)^2}{4dt} x_i \beta_1 \frac{1}{8\pi dt} \sqrt{\frac{4\pi dt}{\beta_2}} e^{\frac{-\beta_2 (\beta_3 - dt)^2}{4dt}} \\ & \left. - \frac{(\beta_3 - dt)^2}{4dt} x_i \beta_1 \frac{1}{2} \frac{1}{4\pi dt} \sqrt{\frac{4\pi dt}{\beta_2}} e^{\frac{-\beta_2 (\beta_3 - dt)^2}{4dt}} + \frac{(\beta_3 - dt)^4}{16dt^2} x_i \beta_1 \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2 (\beta_3 - dt)^2}{4dt}} \right) \end{aligned}$$

$$\begin{aligned}
\frac{\partial D}{\partial \beta_2 \partial \beta_3} &= -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(-2 \frac{\beta_2(\beta_3 - dt)}{4dt} x_i \beta_1 \frac{1}{8\pi dt} \sqrt{\frac{4\pi dt}{\beta_2}} e^{\frac{-\beta_2(\beta_3 - dt)^2}{4dt}} \right. \\
&\quad \left. + 2 \frac{\beta_2(\beta_3 - dt)^3}{16dt^2} x_i \beta_1 \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2(\beta_3 - dt)^2}{4dt}} - 2 \frac{(\beta_3 - dt)}{4dt} x_i \beta_1 \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2(\beta_3 - dt)^2}{4dt}} \right) \\
\frac{\partial D}{\partial \beta_3 \partial \beta_3} &= \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(\frac{\beta_2}{dt} x_i \beta_1 \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2(\beta_3 - dt)^2}{4dt}} - \frac{2\beta_2^2(\beta_3 - dt)^2}{4dt^2} x_i \beta_1 \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2(\beta_3 - dt)^2}{4dt}} \right) \\
\frac{\partial D}{\partial \beta_4 \partial \beta_4} &= -2 \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(-y_i(1 - \beta_1) \frac{1}{2} \frac{1}{8\pi dt} \frac{4\pi dt}{\beta_4^2} \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4(\beta_5 - dt)^2}{4dt}} \right. \\
&\quad - \frac{(\beta_5 - dt)^2}{4dt} y_i(1 - \beta_1) \frac{1}{8\pi dt} \sqrt{\frac{4\pi dt}{\beta_4}} e^{\frac{-\beta_4(\beta_5 - dt)^2}{4dt}} \\
&\quad \left. - \frac{(\beta_5 - dt)^2}{4dt} y_i(1 - \beta_1) \frac{1}{2} \frac{1}{4\pi dt} \sqrt{\frac{4\pi dt}{\beta_4}} e^{\frac{-\beta_4(\beta_5 - dt)^2}{4dt}} + \frac{(\beta_5 - dt)^4}{16dt^2} y_i(1 - \beta_1) \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4(\beta_5 - dt)^2}{4dt}} \right) \\
\frac{\partial D}{\partial \beta_5 \partial \beta_4} &= \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(\frac{(\beta_5 - dt)}{dt} y_i(1 - \beta_1) \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4(\beta_5 - dt)^2}{4dt}} \right. \\
&\quad \left. + \frac{\beta_4(\beta_5 - dt)}{dt} y_i(1 - \beta_1) \frac{1}{2} \frac{1}{4\pi dt} \sqrt{\frac{4\pi dt}{\beta_4}} e^{\frac{-\beta_4(\beta_5 - dt)^2}{4dt}} - \frac{\beta_4(\beta_5 - dt)^3}{4dt^2} y_i(1 - \beta_1) \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4(\beta_5 - dt)^2}{4dt}} \right) \\
\frac{\partial D}{\partial \beta_5 \partial \beta_5} &= \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(\frac{\beta_4}{dt} y_i(1 - \beta_1) \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4(\beta_5 - dt)^2}{4dt}} \right. \\
&\quad \left. - \frac{2\beta_4^2(\beta_5 - dt)^2}{4dt^2} y_i(1 - \beta_1) \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4(\beta_5 - dt)^2}{4dt}} \right)
\end{aligned}$$

Après simplification on obtient les dérivées de la Hessienne :

$$\begin{aligned}
\frac{\partial D}{\partial \beta_1 \partial \beta_2} &= \frac{1}{2\sqrt{\pi dt}} \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) x_i e^{\frac{-\beta_2(\beta_3 - dt)^2}{4dt}} \left(-\sqrt{\frac{1}{\beta_2}} + \frac{\sqrt{\beta_2}(\beta_3 - dt)^2}{2dt} \right) \\
\frac{\partial D}{\partial \beta_1 \partial \beta_3} &= \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(x_i \frac{\beta_2(\beta_3 - dt)}{dt} \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2(\beta_3 - dt)^2}{4dt}} \right) \\
\frac{\partial D}{\partial \beta_1 \partial \beta_4} &= \frac{1}{2\sqrt{\pi dt}} \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) y_i e^{\frac{-\beta_4(\beta_5 - dt)^2}{4dt}} \left(\sqrt{\frac{1}{\beta_4}} - \frac{\sqrt{\beta_4}(\beta_5 - dt)^2}{2dt} \right) \\
\frac{\partial D}{\partial \beta_1 \partial \beta_5} &= \sum_{i=1}^n (u_i - f(x_i, y_i, \beta)) \left(-y_i \frac{\beta_4(\beta_5 - dt)}{dt} \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4(\beta_5 - dt)^2}{4dt}} \right)
\end{aligned}$$

$$\frac{\partial D}{\partial \beta_2 \partial \beta_2} = \frac{1}{4\sqrt{\pi}dt} \sum_{i=1}^n \left(u_i - f(x_i, y_i, \beta) \right) \frac{x_i \beta_1}{\sqrt{\beta_2}} e^{\frac{-\beta_2(\beta_3-dt)^2}{4dt}} \left(\frac{1}{\beta_2} + \frac{(\beta_3-dt)^2}{dt} - \frac{\beta_2(\beta_3-dt)^4}{4dt^2} \right)$$

$$\frac{\partial D}{\partial \beta_2 \partial \beta_3} = \frac{1}{2\sqrt{\pi}dt} \sum_{i=1}^n \left(u_i - f(x_i, y_i, \beta) \right) \frac{x_i \beta_1 \sqrt{\beta_2} (\beta_3-dt)}{dt} e^{\frac{-\beta_2(\beta_3-dt)^2}{4dt}} \left(\frac{3}{2} - \frac{\beta_2(\beta_3-dt)^2}{4dt} \right)$$

$$\frac{\partial D}{\partial \beta_3 \partial \beta_3} = \sum_{i=1}^n \left(u_i - f(x_i, y_i, \beta) \right) \frac{x_i \beta_1 \beta_2}{dt} \sqrt{\frac{\beta_2}{4\pi dt}} e^{\frac{-\beta_2(\beta_3-dt)^2}{4dt}} \left(1 - \frac{2\beta_2(\beta_3-dt)^2}{4dt} \right)$$

$$\frac{\partial D}{\partial \beta_4 \partial \beta_4} = \sum_{i=1}^n \left(u_i - f(x_i, y_i, \beta) \right) \frac{y_i(1-\beta_1)}{4dt\sqrt{\beta_4\pi}dt} e^{\frac{-\beta_4(\beta_5-dt)^2}{4dt}} \left(\frac{2}{\beta_4} + \frac{2(\beta_5-dt)^2}{2dt} - \frac{\beta_4(\beta_5-dt)^4}{2dt} \right)$$

$$\frac{\partial D}{\partial \beta_5 \partial \beta_4} = \sum_{i=1}^n \left(u_i - f(x_i, y_i, \beta) \right) \frac{y_i \sqrt{\beta_4} (1-\beta_1) (\beta_5-dt)}{2\sqrt{\pi}dt dt} e^{\frac{-\beta_4(\beta_5-dt)^2}{4dt}} \left(\frac{3}{2} - \frac{\beta_4(\beta_5-dt)^2}{4dt} \right)$$

$$\frac{\partial D}{\partial \beta_5 \partial \beta_5} = \sum_{i=1}^n \left(u_i - f(x_i, y_i, \beta) \right) \frac{y_i(1-\beta_1)\beta_4}{dt} \sqrt{\frac{\beta_4}{4\pi dt}} e^{\frac{-\beta_4(\beta_5-dt)^2}{4dt}} \left(1 - \frac{\beta_4(\beta_5-dt)^2}{2dt} \right)$$

9.3. Codes du traitement des données

```
#####Dhuy      #pas de problème de données
dhuy=read.csv(file='Dhuy.csv',header=T,sep=';')
names(dhuy)=c("Station","Date","Temp","EC","pH","NH4","Cl","NO2","NO3","PO4
3","O2.diss","MES","DBO5","DCO","nkj.bis","pheo.bis")
attach(dhuy)
dhuy=data.frame(Station, Date, NO3)
dhuy$Date=as.Date(dhuy$Date,format='%d/%m/%Y') #changement au format date

#####Loire
loire=read.csv(file='Loire.csv',header=T,sep=';')
names(loire)=c("Station","Date","Temp","EC","pH","NH4","Cl","NO2","NO3","PO
43","O2.diss","MES","DBO5","DCO","nkj.bis","pheo.bis")
attach(loire)
loire=data.frame(Station, Date, NO3)
loire[loire$NO3=='< 1',3]=0 #remplacement '< 1' par 0
loire=loire[loire$NO3 !='',] #suppression des valeurs manquantes
loire$Date=as.Date(loire$Date,format='%d/%m/%Y')
loire$NO3=as.numeric(as.character(loire$NO3))

#####Loiret
loiret=read.csv(file='Loiret.csv',header=T,sep=';')
names(loiret)=c("Station","Date","Temp","EC","pH","NH4","Cl","NO2","NO3","P
O43","O2.diss","MES","DBO5","DCO","nkj.bis","pheo.bis")
attach(loiret)
loiret=data.frame(Station, Date, NO3)
loiret$Date=as.Date(loiret$Date,format='%d/%m/%Y')
loiret=loiret[!is.na(loiret$NO3),] #suppression des valeurs manquantes
loiret=loiret[order(loiret$Date),] #trie Date ordre croissant

levels(loiret$Station)=c("Bt","Gn","Tt","Bn") #renommer Station
attach(loiret)

#moyenne NO3 selon Date
loiret_moy <- aggregate( data.frame(NO3=loiret$NO3),
by=list(Date=loiret$Date), mean)
```

```

tab=data.frame(Station,Date)
loiret_stat=aggregate(Station ~ Date, data=tab, paste, sep=",") #regroupe
station selon date

loiret=merge(loiret_stat,loiret_moy)

#####représentation des données traitées
plot(x=dhuy$Date, y=dhuy$NO3, type="l",
xlab="Temps (jour)", ylab="NO3 (mg/l)",
col = "blue", main = "Variations NO3 au Val d'Orléans",
cex.main = 0.9, cex.lab = 0.8)
lines(x=loire$Date, y=loire$NO3, col = "red")
lines(x=loiret$Date, y=loiret$NO3, col = "green")

legend("topright", legend = c("Dhuy", "Loire", "Loiret"),
col = c("blue", "red", "green"),
pch = 15, bty = "n", pt.cex = 2,
cex = 0.8, text.col = "black", horiz = TRUE, inset = c(0.1, 0.1))

#####interpolation Loire par rapport au Dhuy
app=approx (x=loire$Date, y=loire$NO3, xout=dhuy$Date, method="linear",
rule=1, f=0)

t=1:184
M=data.frame(DateDhuy=app$x[t], DhuyNO3=dhuy$NO3[t], LoireNO3=app$y[t])

plot(x=loire$Date, y=loire$NO3, type="o",
xlab="Temps (jour)", ylab="NO3 (mg/l)",
col = "blue", main = "Interpolation sur les mesures de la Loire",
cex.main = 0.9, cex.lab = 0.8)
lines(x=M$DateDhuy, y=M$LoireNO3, col = "red", type="o")

legend("topright", legend = c("Valeurs exactes", "Valeurs interpolées"),
col = c("blue", "red"),
pch = 15, bty = "n", pt.cex = 2,
cex = 0.8, text.col = "black", horiz = TRUE, inset = c(0.1, 0.1))

#####Prise d'un delta t petit
dt=NULL
LoiretNO3=NULL
DateLoiret=NULL
for (i in 1:length(M$DateDhuy)){
  x=difftime(loiret$Date, M$DateDhuy[i], units = "days")
  g=min(x[x>0])
  dt=c( dt, g )
  LoiretNO3= c( LoiretNO3, loiret$NO3[g==x] )
  DateLoiret=c( DateLoiret, loiret$Date[g==x] )
  DateLoiret=as.Date(DateLoiret, origin = "1970-01-01")
}

M=data.frame(M, LoiretNO3, DateLoiret, dt)

#suppression de doublon en gardant le plus petit delta t
m=data.frame()
i=1
while (i<=dim(M)[1] ){
  if ( sum(M$DateLoiret[i]==M$DateLoiret)>1 ){

```

```

        s=M[M$DateLoiret[i]==M$DateLoiret,]
        m=rbind(m,s[ s$dt==min(s$dt),])
        i=i+dim(s)[1]
    }
    else {
        m=rbind(m,M[i,])
        i=i+1
    }
}
m=m[,c(2,3,4,6)]

#####représentation des données traitées
plot(x=m$DateDhuy, y=m$DhuyNO3, type="l",
xlab="Temps (jour)", ylab="NO3 (mg/l)",
col = "blue", main = "Variations NO3 au Val d'Orléans",
cex.main = 0.9, cex.lab = 0.8)
lines(x=m$DateDhuy, y=m$LoireNO3, col = "red")
lines(x=m$DateDhuy, y=m$LoiretNO3, col = "green")

legend("topright", legend = c("Dhuy", "Loire", "Loiret"),
col = c("blue", "red", "green"),
pch = 15, bty = "n", pt.cex = 2,
cex = 0.8, text.col = "black", horiz = TRUE, inset = c(0.1, 0.1))

```

9.4. Codes de l'exemple théorique

```

#Représentation 3D de f
install.packages("rgl")
require(rgl)
x=seq(0, 2, length=100)
y=seq(0, 2, length=100)
z=outer(x,y,f)
nbcol = 100
color = rev ( rainbow ( nbcol , start = 0 / 6 , end = 4 / 6 ))
zcol = cut ( z , nbcol )
persp3d ( x , y , z , theta = 50 , phi = 25 , expand = 0.75 , col = color [
zcol ], ticktype = "detailed" , xlab = "x" , ylab = "y" , zlab = "" , axes
= TRUE)

#fonction f
f <- function(x,y) {
    return ( sin(5.5*x)*cos(5*y)+x*x+1 )
}

#Norme euclidienne
norm <- function(b0) sum(sqrt(b0*b0))

#Gradient de f
grad1 <- function(x,y){
    return ( 5.5*cos(5.5*x)*cos(5*y)+2*x )
}

grad2 <- function(x,y){

```

```

    return ( -5*sin(5.5*x)*sin(5*y) )
}

gradf <- function(x,y){
  return ( c(grad1(x,y), grad2(x,y)) )
}

#Hessienne de f
hessf <- function(x,y){
  k= c(-30.25*sin(5.5*x)*cos(5*y)+2 , -27.5*sin(5*y)*cos(5.5*x) , -
27.5*sin(5*y)*cos(5.5*x) , -25*sin(5.5*x)*cos(5*y))
  return ( matrix(k,2,2) )
}

#Matrice définie positive
defPositif <- function(H){
  return ( H[1,1]>0 && det(H)>0 )
}

#Gradient conjugué projeté, initialisation des paramètre avec loi uniforme
gradPC <- function(x,y,eps,itmax, binf, bsup, n){
  p=300
  for (j in 1:n){
    b = matrix (NA, nrow = itmax, ncol =2)
    b[1,] = c( bsup[1]*runif(1),bsup[2]*runif(1))
    i=1
    b0=b[i,]
    D= -gradf(b0[1],b0[2])
    while(norm( gradf(b0[1],b0[2]) )>eps && i<itmax){
      H=hessf(b0[1],b0[2])
      t=as.numeric( -(gradf(b0[1],b0[2])%*%D)/(D%*%H%*%D) )
      b[i+1,]=pmin(pmax(binf,b0+t*D),bsup)
      i=i+1
      b0=b[i,]
      D=-
gradf(b0[1],b0[2])+(norm(gradf(b0[1],b0[2]))^2)/(norm(gradf(b[i-1,1],b[i-1,2]))^2)*D
    }
    g=f(b0[1],b0[2])
    if (g<p && defPositif(H)==TRUE) {p=g; q=b; d=gradf(b0[1],b0[2]) }
  }
  if (p==300) return ('pas de minimum')
  else return ( list(estimation=q, itération=i-1, min=p, gradient=d) )
}

#Choix des paramètres de la fonction gradPC et exécution
eps=0.01;
itmax=30;
n=10
binf=c(0,0)
bsup=c(2,2)
s=gradPC(x,y,eps,t,itmax,binf, bsup,n)
s

```

9.5. Codes du cas réel

```
##Matrice
```

```

nbrLignDebut=43
nbrLignFin=54
nbrLign = nbrLignFin - nbrLignDebut + 1
h=m[nbrLignDebut:nbrLignFin,]
U=h$LoiretNO3
X=h$LoiretNO3
Y=h$DhuyNO3
dt=h$dt
#b=(alpha, Pel, Tl, Ped, Td)

#fonction f
f <- function(X,Y,b0,dt) {
  return ( X*b0[1]*sqrt( b0[2]/(4*pi*dt) )*exp( -b0[2]*(b0[3]-
dt)^2/(4*dt) )
  + Y*( 1-b0[1] )*sqrt( b0[4]/(4*pi*dt) )*exp( -b0[4]*(b0[5]-dt)^2/(4*dt)
) )
}

#Ecart entre les valeurs expérimentaux et les valeurs simulées
ecart <- function(U,X,Y,b0,dt){
  return ( U-f(X,Y,b0,dt) )
}

#Norme euclidienne
norm <- function(b0) sum(sqrt(b0*b0))

#Gradient et Hessienne de f
grad1 <- function(X,Y,b0,dt,A, B){
  return ( 1/sqrt(pi*dt)*( -X*sqrt(b0[2])*A + Y*sqrt(b0[4])*B ) )
}

grad2 <- function(X,Y,b0,dt, A){
  return ( 1/(2*sqrt(pi*dt))*X*b0[1]*A*( -sqrt(1/b0[2]) +
sqrt(b0[2])*(b0[3]-dt)^2/(2*dt) ) )
}

grad3 <- function(X,Y,b0,dt, A, C){
  return ( X*b0[1]*b0[2]*(b0[3]-dt)/dt*(C*A) )
}

grad4 <- function(X,Y,b0,dt, B){
  return ( 1/(2*sqrt(pi*dt))*Y*(1-b0[1])*B*( -sqrt(1/b0[4]) +
sqrt(b0[4])*(b0[5]-dt)^2/(2*dt) ) )
}

grad5 <- function(X,Y,b0,dt, B, D){
  return ( Y*(1-b0[1])*D*B*( b0[4]*(b0[5]-dt)/dt ) )
}

grad1d2 <- function(X,Y,b0,dt,A){
  return ( 1/(2*sqrt(pi*dt))*X*A*( -sqrt(1/b0[2]) + sqrt(b0[2])*(b0[3]-
dt)^2/(2*dt) ) )
}

grad1d3 <- function(X,Y,b0,dt, A, C){
  return ( X*b0[2]*(b0[3]-dt)/dt*(C*A) )
}

```

```

grad1d4 <- function(X,Y,b0,dt, B){
  return ( 1/(2*sqrt(pi*dt))*Y*B*( sqrt(1/b0[4])-sqrt(b0[4])*(b0[5]-
dt)^2/(2*dt) ) )
}

grad1d5 <- function(X,Y,b0,dt, B, D){
  return ( -Y*b0[4]*(b0[5]-dt)/dt*(D*B) )
}

grad2d2 <- function(X,Y,b0,dt, A){
  return ( 1/(4*sqrt(pi*dt))*X*b0[1]*A/sqrt(b0[2])*( (1/b0[2]) + (b0[3]-
dt)^2/dt - b0[2]*(b0[3]-dt)^4/(4*dt^2) ) )
}

grad2d3 <- function(X,Y,b0,dt, A){
  return ( 1/(2*sqrt(pi*dt))*X*b0[1]*sqrt(b0[2])*(b0[3]-dt)/dt*A*( 3/2 -
b0[2]*(b0[3]-dt)^2/(4*dt) ) )
}

grad3d3 <- function(X,Y,b0,dt, A, C){
  return ( X*b0[1]*b0[2]/dt*C*A*( 1 - 2*b0[2]*(b0[3]-dt)^2/(4*dt) ) )
}

grad4d4 <- function(X,Y,b0,dt, B){
  return ( Y*(1-b0[1])*B/(4*dt*sqrt(b0[4]*pi*dt))*( 2/b0[4] + 2*(b0[5]-
dt)^2 - b0[4]*(b0[5]-dt)^4/(2*dt) ) )
}

grad5d4 <- function(X,Y,b0,dt, B){
  return ( Y*sqrt(b0[4])*(1-b0[1])*(b0[5]-dt)*B/(2*sqrt(pi*dt)*dt)*( 3/2
- b0[4]*(b0[5]-dt)^2/(4*dt) ) )
}

grad5d5 <- function(X,Y,b0,dt, B, D){
  return ( Y*(1-b0[1])*b0[4]/dt*D*B*( 1 - (b0[4]*(b0[5]-dt)^2)/(2*dt) ) )
}

gradD <- function(U, X, Y, b0,dt,nbrLign){
  A=exp( -b0[2]*((b0[3]-dt)^2)/(4*dt) ); B=exp( -b0[4]*((b0[5]-
dt)^2)/(4*dt) )
  C=sqrt( b0[2]/(4*pi*dt) ); D=sqrt( b0[4]/(4*pi*dt) )
  g1=grad1(X,Y,b0,dt, A, B)
  g2=grad2(X,Y,b0,dt, A)
  g3=grad3(X,Y,b0,dt, A, C)
  g4=grad4(X,Y,b0,dt, B)
  g5=grad5(X,Y,b0,dt, B, D)
  K=matrix(c(g1, g2, g3, g4, g5),nrow=nbrLign, ncol=5)
  return ( ecart(U,X,Y,b0,dt) %*% K )
}

hessD <- function(U, X, Y, b0,dt,nbrLign){
  A=exp( -b0[2]*((b0[3]-dt)^2)/(4*dt) ); B=exp( -b0[4]*((b0[5]-
dt)^2)/(4*dt) )
  C=sqrt( b0[2]/(4*pi*dt) ); D=sqrt( b0[4]/(4*pi*dt) )
  g12=grad1d2(X,Y,b0,dt, A)
  g13=grad1d3(X,Y,b0,dt, A, C)
  g14=grad1d4(X,Y,b0,dt, B)
  g15=grad1d5(X,Y,b0,dt, B, D)

```

```

g22=grad2d2(X,Y,b0,dt, A)
g23=grad2d3(X,Y,b0,dt, A)
g33=grad3d3(X,Y,b0,dt, A, C)
g44=grad4d4(X,Y,b0,dt, B)
g45=grad5d4(X,Y,b0,dt, B)
g55=grad5d5(X,Y,b0,dt, B, D)
zero=rep(0,nbrLign)
K=matrix(c(zero, g12, g13, g14, g15, g12, g22, g23, zero, zero,
           g13, g23, g33, zero, zero, g14, zero, zero, g44, g45,
           g15, zero, zero, g45, g55),nrow=nbrLign, ncol=25)
K1=ecart(U,X,Y,b0,dt) %*% K
return ( matrix(K1,nrow=5, ncol=5) )
}

defPositif <- function(H){
  return ( H[1,1]>0 && det(H[1:2,1:2])>0 && det(H[1:3,1:3])>0 &&
det(H[1:4,1:4])>0 && det(H)>0 )
}

#Gradient conjugué projeté, initialisation des param avec loi uniforme
gradPC <- function(b,X,Y,U,dt,eps,itmax, binf, bsup, n){
  time1<-Sys.time() #temps d'exécution
  p=30000
  for (j in 1:n){
    b = matrix (NA, nrow = itmax, ncol = 5)
    b[1,] = c(
runif(1),bsup[2]*runif(1),bsup[3]*runif(1),bsup[4]*runif(1),bsup[5]*runif(1)
) )
    i=1
    b0=b[i,]
    D=-gradD(U,X,Y,b0,dt,nbrLign)
    while (norm( gradD(U,X,Y,b0,dt,nbrLign) )>eps && i<itmax){
      H=hessD(U, X, Y, b0,dt,nbrLign)
      t=as.numeric( -
((gradD(U,X,Y,b0,dt,nbrLign))%*%t(D))/(D%*%H%*%t(D)) )
      b[i+1,]=pmin(pmax(binf,b0+t*D),bsup)
      i=i+1
      b0=b[i,]
      D=-
gradD(U,X,Y,b0,dt,nbrLign)+(norm(gradD(U,X,Y,b0,dt,nbrLign))^2)/(norm(gradD
(U,X,Y,b[i-1,],dt,nbrLign))^2)*D
    }
    g=sum( ( ecart(U,X,Y,b0,dt)^2 ) )
    if (g<p ) {p = g; q=b; r=abs( ecart(U,X,Y,b[i,],dt)); k=i;
v=f(X,Y,b[i,],dt); d=gradD(U,X,Y,b0,dt,nbrLign) }
  }
  time2<-Sys.time()
  tps=difftime(time2, time1)
  if (p==30000) return (print('pas de minimum'))
  else return ( list(estimateurs=q, residus=r, image=p, itération=k,
valsimu=v,temps=tps, gradient=d) )
}

```


Référence

- [1] Dossier scientifique l'eau
http://www.cnrs.fr/cw/dossiers/doseau/decouv/degradation/07_pollution.htm
- [2] Nitrates : la pollution progresse
samedi 16 aout 2014
<http://www.lanouvellerepublique.fr>
- [3] Site de LIFO - Laboratoire d'Informatique Fondamentale d'Orléans
Action de Recherche : FDTE
<http://www.univ-orleans.fr/lifo/action.php?id=42&lang=fr>
- [4] Rapport sur l'impact des changements environnementaux sur la ressource en eau
Ecrit par Stéphane Binet, Chrystelle Auterives, Jean-Baptiste Charlier.
- [5] Wikipédia,
C'est pas sorcier, <https://www.youtube.com/watch?v=JDeaWTTxp84>
- [6] site du Parc Floral à Orléans la Source
http://www.parcfloraldelasource.com/fr/93/La_source_du_Loiret.html
- [7] Bilan 2012 sur le site de l'Agence Régionale de Santé Centre (ARS)
Les teneurs en nitrates dans les eaux distribuées en région Centre en 2012
- [8] SAGE du bassin du LOIRET 2003, diagnostic
- [9] Problématique de notre étude sur pdf