



*Mécanique des Fluides, Énergétique et  
Environnement - MF2E 1A*

## Optimisation TDM : Moindres carrés

*Auteurs :*

TRAN Ugo

██████████ Romain

*Responsable du TDM :*

Pr. Rémi Zamansky

*Mai 2024*

## Sommaire

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Régression linéaire et non linéaire</b>	<b>3</b>
<b>3</b>	<b>Prévision du débit d'un cours d'eau et calage de modèle</b>	<b>5</b>
3.1	Jeu de données . . . . .	5
3.2	Modèle pluies/débits . . . . .	6
3.2.1	Calage des paramètres du modèle . . . . .	7
3.2.2	Etude de la fenêtre d'observation . . . . .	8
3.3	Modification du modèle de base . . . . .	8
3.3.1	Paramètre de délai $\tau_0$ . . . . .	9
3.3.2	Modèle à M réservoir disposés en serie (cascade de Nash) . . . . .	9

## Liste des figures

1	Évolution de la population en fonction du temps pour le 1er modèle . . . . .	3
2	Évolution de la pluie et de débit fonction du temps pour le 1er modèle . . . . .	4
3	Évolution de la pluie et de débit fonction du temps pour le 2ème modèle . . . . .	5
4	Évolution de la pluie et de débit fonction du temps . . . . .	5
5	Évolution de la pluie et de débit fonction du temps sur une semaine . . . . .	6
6	Schéma du modèle Input/Output . . . . .	6
7	Evolution du débit modèle pour 300 valeurs comparé aux données théorique . . . . .	8
8	Schéma du modèle en cascade de Nash . . . . .	9
9	Schéma pour 10 réservoirs comparé aux données théoriques . . . . .	9

## 1 Introduction

L'objectif de notre étude sera de prévoir le débit d'un cours d'eau, pour cela, on pourra étudier les données de pluies et de débits à Aliou (93), on pourra ensuite utiliser la méthode des moindres carrés pour trouver un modèle "optimal" pour prévoir le débit d'un cours d'eau. L'ensemble de l'étude sera faite sur python avec les bibliothèque de base matplotlib pour tracer les graphes, math, numpy ...

## 2 Régression linéaire et non linéaire

Dans cette partie, on va se familiariser avec la méthode des moindres carrés, pour cela on s'intéresse à la croissance d'une population de bactérie, on nous donne les données suivante :

Temps (min)	Population $N_t$ (en milliers)
0	1
100	5
198	48
306	505

On peut donc tracer l'évolution de la population en fonction du temps (Figure 1) :

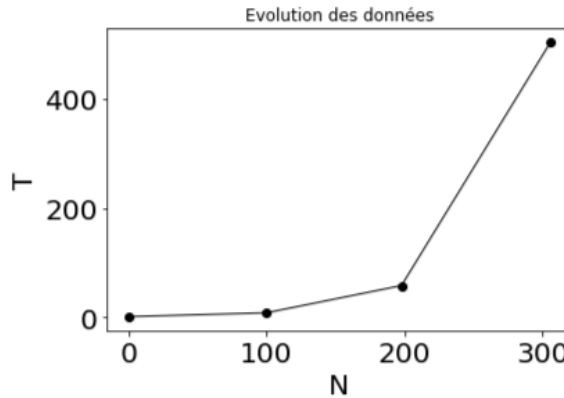


Figure 1: Évolution de la population en fonction du temps pour le 1er modèle

On va définir le "résidu"  $\epsilon$  tel que :

$$\epsilon = \frac{1}{N} \sum_{i=1}^N \epsilon_i^2 \quad \text{avec } \forall i \in \{1, \dots, n\} \quad \epsilon_i = y_i^{obs} - y^{model}(t_i, p_0)$$

Où N représente le nombre d'observation, dans notre cas on a donc N=5

On a donc besoin d'une fonction qui nous servira de modèle, c'est à dire que l'évolution des bactéries sera décrite par notre modèle, qui ici dépendra d'un seul paramètre  $p_0$ .

On s'intéresse naturellement à minimiser les résidus, on va donc chercher la valeur optimal de  $p_0$ , lorsque  $\frac{\partial \epsilon}{\partial p_0} = 0$ . Après calcul on obtient la condition suivante :

$$\frac{1}{N} \sum_{i=1}^N \left[ -2y_i^{obs} \frac{\partial y^{model}}{\partial p_0}(t_i, p_0) + 2 \frac{\partial y^{model}}{\partial p_0}(t_i, p_0) y^{model}(t_i, p_0) \right] = 0 \quad (1)$$

On va dans un premier temps considérer un modèle de croissance exponentielle :

$$y^{model}(t_i, p_0) = p_0 \exp\left(\frac{t_i}{50}\right)$$

Où  $p_0$  est le paramètre à optimiser. On a donc  $\frac{\partial y^{model}}{\partial p_0}(t_i, p_0) = \exp\left(\frac{t_i}{50}\right)$ .

L'équation (1) devient alors :

$$\frac{1}{N} \sum_{i=1}^N \left[ -2y_i^{obs} \exp\left(\frac{t_i}{50}\right) + 2 \exp\left(\frac{2t_i}{50}\right) \right] = 0$$

On obtient finalement :

$$p_0 = \frac{\sum_{i=1}^N y_i^{obs}}{\sum_{i=1}^N \exp\left(\frac{t_i}{50}\right)} = 1,084 \quad (2)$$

On peut alors comparer l'évolution du modèle avec celui des données :

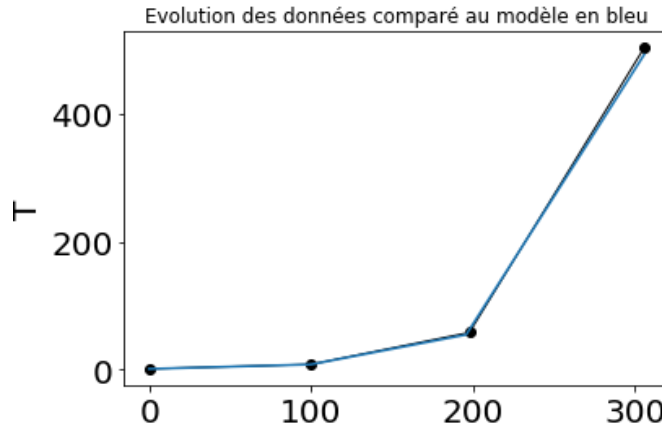


Figure 2: Évolution de la pluie et de débit fonction du temps pour le 1er modèle

Malgré quelques différences, le modèle suit bien les données des observations.

On va maintenant chercher à optimiser le temps caractéristique, on va donc utiliser le modèle suivant :

$$y^{model}(t_i, p_0, \tau) = p_0 \exp\left(\frac{t_i}{\tau}\right)$$

Il va donc falloir chercher à optimiser les paramètres  $p_0$  et  $\tau$ .

On va donc chercher à résoudre le système d'équation suivant :

$$\begin{cases} \frac{\partial \epsilon}{\partial p_0}(t_i, p_0, \tau) = 0 \iff \frac{1}{N} \sum_{i=1}^N \left[ -2y_i^{obs} \exp\left(\frac{t_i}{\tau}\right) + 2 \exp\left(\frac{2t_i}{\tau}\right) \right] = 0 \\ \frac{\partial \epsilon}{\partial \tau}(t_i, p_0, \tau) = 0 \iff \sum_{i=1}^N \left[ \frac{2y_i^{obs} p_0 t_i}{\tau^2} \exp\left(\frac{t_i}{\tau}\right) + \frac{2p_0^2 t_i}{\tau^2} \exp\left(\frac{2t_i}{\tau}\right) \right] = 0 \end{cases}$$

On obtient alors le système d'équation non linéaire suivant :

$$\begin{cases} p_0 = \frac{\sum_{i=1}^N y_i^{obs}}{\sum_{i=1}^N \exp\left(\frac{t_i}{\tau}\right)} \\ \sum_{i=1}^N y_i^{obs} + p_0 \exp\left(\frac{t_i}{\tau}\right) \end{cases} \quad (3)$$

Ce système n'étant pas linéaire, nous ne sommes donc pas en mesure de le résoudre analytiquement. Nous allons donc utiliser la fonction "curve\_fit" sur python qui va trouver les paramètres optimaux  $p_0$  et  $\tau$  numériquement. On peut donc comparer sur la Figure (3) les données obtenus grâce au modèle et aux observations.

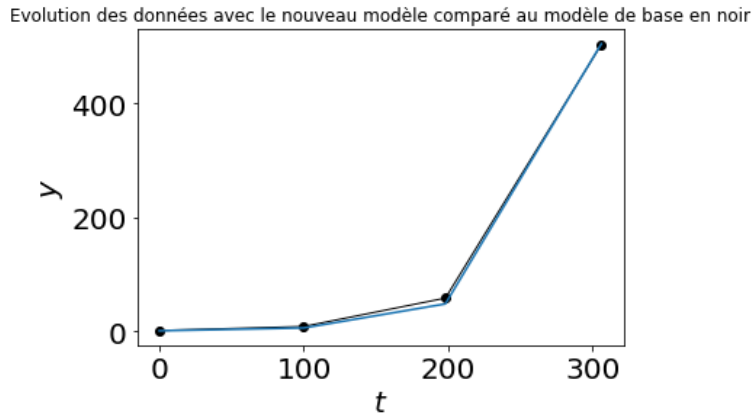


Figure 3: Évolution de la pluie et de débit fonction du temps pour le 2ème modèle

On trouve finalement comme paramètre optimaux :

$$p_0 = 0,63$$

$$\tau = 45$$

Il faut cependant noter que les conditions initiales choisis pour notre programme peuvent influencer sur les paramètres optimaux finales. En effet si la valeurs initiale de  $p_0$  n'a aucune influence sur les valeurs finales, il faut noter que si la valeur initiale de  $\tau$  est trop faible, alors les valeurs obtenues deviennent incohérentes.

### 3 Prévision du débit d'un cours d'eau et calage de modèle

Dans cette partie nous chercherons à prévoir l'évolution du débit d'un cours d'eau afin de prévoir certains phénomènes pour l'agriculture, la prévention ou encore les installations hydroélectrique. Ainsi il est important de quantifier la précipitation qui est un paramètre important pour prévoir le débit.

#### 3.1 Jeu de données

A partir d'un jeu de données mise à notre disposition nous avons pu tracer l'évolution de débit (en  $m^3/s$ ) et de pluie (en  $mm/h$ ) sur 1 semestre (hydrogrammes pour le débit et hyétogramme pour la pluie).

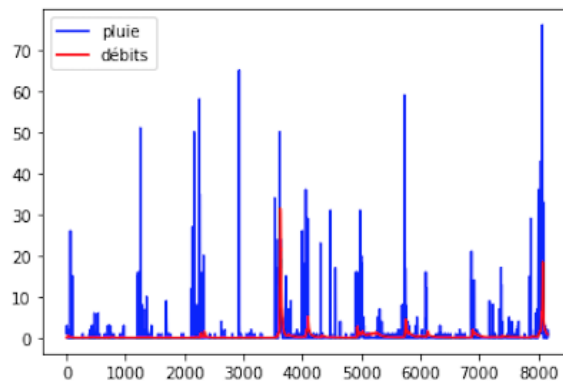


Figure 4: Évolution de la pluie et de débit fonction du temps

Afin de pouvoir mieux identifier certains phénomènes, nous allons nous intéresser à l'évolution de ces paramètres sur une semaine, entre les jours 73 et 80.

On remarque une crue importante pendant le temps [3600-3650] soit entre le jour 75 et 76. Cette crue nous sera importante pour poursuivre l'étude plus tard. On remarque que lorsqu'il pleut, le débit dans

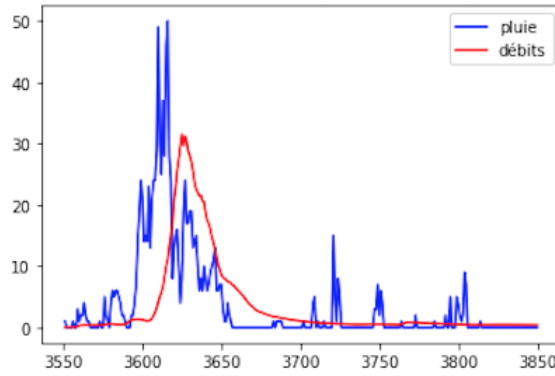


Figure 5: Évolution de la pluie et de débit fonction du temps sur une semaine

le cours d'eau augmente un peu après, ce qui s'explique par le fait que quand l'eau tombe sur les terres il faut un peu de temps pour qu'elle coule jusqu'au cours d'eau et donc augmenter le débit au sein de celui-ci.

### 3.2 Modèle pluies/débits

Dans cette partie nous allons modéliser l'évolution de ce débit par un type Input/Output.  $P(t)$  correspond à la précipitation en entrée et  $Q_{model}(t)$  au débit en sortie d'eau.

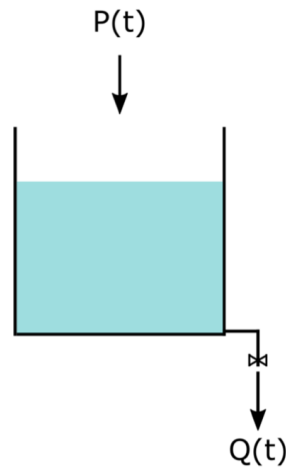


Figure 6: Schéma du modèle Input/Output

On considère  $A$  la section du réservoir,  $h(t)$  la hauteur d'eau. À partir de la conservation de la masse, on montre que:

$$\dot{h}(t) = P(t) - \frac{Q(t)}{A} \quad (4)$$

De plus on sait que:

$$Q_{model}(t) = kAh(t) \quad (5)$$

Avec  $k$  la conductance spécifique,  $[k]=s^{-1}$ . Ainsi à partir des équations (2) et (3) on peut trouver une équation différentielle sur  $Q_{model}$ , soit:

$$\dot{Q}(t) + kQ(t) = kAP(t) \quad (6)$$

On résout dans un premier temps l'équation homogène:  $\dot{Q} + kQ = 0 \iff Q(t) = C(t)e^{-kt}$ , avec  $C$  une constante réelle. On remplace l'expression de  $Q(t)$  dans l'équation (4), on trouve:  $\dot{C}(t)e^{-kt} -$

$kC(t)e^{-kt} + kC(t)e^{-kt} = AkP(t)$ , soit:

$$C(t) = \int_0^t AkP(s)e^{ks} ds + C_1 \quad (7)$$

Avec  $C_1$  une constante réelle. En remplaçant l'expression de  $C(t)$  dans celle de  $Q(t)$ , on a:  $Q(t) = \int_0^t AkP(s)e^{ks}e^{-kt} ds + C_1e^{-kt}$ . En se plaçant à  $t=0$ , en posant  $Q(t=0) = Q_0$  et en posant  $a=kA$ , on trouve:

$$Q(t) = \int_0^t aP(s)e^{-k(t-s)} ds + Q_0e^{-kt} \quad (8)$$

Le terme de gauche avec l'intégrale présente la quantité d'eau qui entre dans le réservoir à un moment  $s$  pendant le temps  $t$ . Le terme de droite avec le  $Q_0$  représente le débit d'eau au sein du réservoir au début de l'étude c'est à dire au temps 0.

Pour pouvoir calculer ce modèle numériquement il est plus simple d'approximer l'intégrale par une somme discrète et de discrétiser le temps  $t$  par  $t = i\Delta t$ , ainsi:

$$\int_0^t aP(s)e^{-k(t-s)} ds = \sum_{i=0}^N aP(i\Delta t)e^{-k(t-i\Delta t)} \quad (9)$$

Avec  $N=8158$

On peut en déduire la formule finale de l'évolution du débit:

$$Q_j(t) = Q_0e^{-kj\Delta t} + \sum_{i=0}^N aP(i\Delta t)e^{-k(j\Delta t-i\Delta t)} \quad (10)$$

Ainsi on connaît tout les paramètres sauf  $Q_0$ ,  $k$  et  $A$ , donc ce seront les paramètres à ajuster du modèle pour la suite de l'étude numérique.

### 3.2.1 Calage des paramètres du modèle

On cherchera à l'aide de la méthode des moindres carrés à comparer l'évolution du modèle en fonction de l'évolution des débits observés. La valeur des résidus nous sera utile pour définir si notre modèle se rapproche des observations ou pas. Soit  $\epsilon$  les résidus:

$$\epsilon = \frac{1}{N} \sum_{i=0}^N \left( Q_i^{obs} - Q_0e^{-ki\Delta t} - \Delta t \sum_{j=0}^i aP(j\Delta t)e^{-k(i\Delta t-j\Delta t)} \right)^2 \quad (11)$$

Ainsi pour optimiser ces paramètres, nous allons procéder à l'algorithme suivant:

-Ouvrir les fichiers données pluie et débit et récupérer seulement 300 données entre 3551 et 3850 temps.

Description de la méthode numérique à suivre:

1)-Définir une fonction modèle qui calcul la formule (10). Pour cela on fait 2 boucles, une sur  $i$  et l'autre sur  $j$ . La boucle sur  $j$  calcul la somme et la boucle sur  $i$  rajoute à chaque itération la somme à la valeur du deuxième terme qui dépend de  $Q_0$ .

2)-On créer une liste temps qui prend 300 valeurs pour le moment entre 3351 et 3850, avec la fonction de python linspace par exemple.

3)-Ensuite on créer une liste de 3 valeurs qui comportera les valeurs initiale des paramètres  $A$ ,  $k$  et  $Q_0$ .

4)-On créer une nouvelle liste *param<sub>opt</sub>* qui à l'aide de la fonction *curvefit* donne les paramètres optimales pour  $A$ ,  $K$  et  $Q_0$ . Cette fonction prend comme argument en premier notre fonction modèle,



puis la liste temps, puis la liste débit et enfin la liste des paramètres initiaux.

5)-Ainsi par la suite nous connaissons les paramètres optimaux, donc nous pouvons tracer des figures à l'aide de la bibliothèque subplot de python. On trace sur la même figure pour pouvoir comparer, l'évolution de la fonction modèle avec les paramètres optiamux et l'évolution des données de base.

6)-Enfin nous pouvons comparer, analyser les résultats et conclure.

Ainsi après modification de notre programme sur python pour  $i \in [3351 - 3850]$ , on trouve que  $Q_0 = -0,8m^3/s$ ,  $k = 0,10s^{-1}$  et  $A = 1,1m^2$ . Et on peut tracer une représentation graphique:

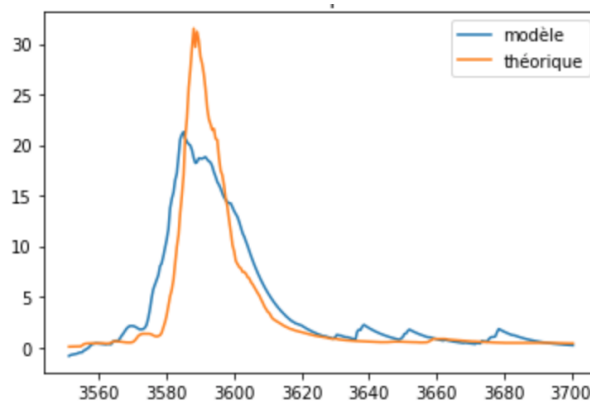


Figure 7: Evolution du débit modèle pour 300 valeurs comparé aux données théorique

On peut remarquer que le modèle à la même forme que la théorie mais atteint un maximum moins important que la théorie. De plus on trouve pour la paramètre  $Q_0$  un débit négatif ce qui n'est pas réaliste, c'est pourquoi nous avons par la suite changé le débit dans la liste des paramètres initiaux, par la suite on trouve un débit positif cette fois-ci. Or encore une fois quand on change ce débit, le programme nous sors un débit négatif optimal, c'est pourquoi par la suite nous avons gardé notre première valeur.

On calcul la norme des résidus et on trouve  $\epsilon = 8$ , le but est de trouver un epsilon le plus petit possible, mais petit devant quoi ? C'est pourquoi nous prendrons cette valeur de résidus comme valeur témoin pour la suite de l'étude, le but sera de diminuer au maximum cette valeur.

De plus en fixant  $Q_0 = Q(0)$  ou encore  $Q > 0$ , nous n'observons pas de principal changement graphiquement, ainsi on peut conclure que la principale variation du modèle provient de la somme discrète et par du terme en  $Q_0$ .

### 3.2.2 Etude de la fenêtre d'observation

On peut étudier dans cette partie l'influence de la fenêtre d'observation. Déjà un plus grand nombre de données rend le calcul numérique plus important et donc la simulation prend plus de temps. Nous avons pu remarquer également que si l'on ne choisissait par une bonne fenêtre c'est à dire un plage de temps durant laquelle il y a une crue, et bien le modèle se rapproche moins graphiquement, or la norme de résidus diminue comparé à notre valeur témoin. Ainsi le choix de la fenêtre est primordiale pour une étude optimale.

## 3.3 Modification du modèle de base

Dans cette partie nous chercherons des paramètres à ajouter au modèle pour le rendre plus réaliste.

### 3.3.1 Paramètre de délai $\tau_0$

On ajoute un délai  $\tau_0$  tel que:  $ke^{-k(t-t')}$  devient  $ke^{-k(t-t'-\tau_0)}$ . On n'observe pas de changement graphique ni sur la norme des résidus (comparé à notre  $\epsilon$  témoin), l'ajout de cette constante  $\tau_0$  rend juste les constantes plus réelles et donc le modèle le modèle devient un peu plus juste physiquement.

### 3.3.2 Modèle à M réservoir disposés en serie (cascade de Nash)

Dans cette section nous allons voir l'effet de considérer le modèle suivant:

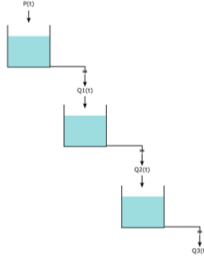


Figure 8: Schéma du modèle en cascade de Nash

L'inclusion de ce dispositif dans notre modèle revient à changer la relation suivante:  $ke^{-k(t-t')}$ , par  $h_1(t-t') = \frac{1}{\Gamma(M)}(k(t-t'))^{M-1}ke^{-k(t-t')}$ , avec  $\Gamma$  la fonction gamma. Cette fonction étant déjà existante sur python nous n'aurons pas besoin de la coder.

Ainsi, nous allons voir l'effet du nombre de cascade.

Dans un premier temps nous fixons  $M=3$  tel que notre modèle numérique ressemble au dispositif de la figure 8.

Ainsi on observe que le modèle suit maintenant mieux la solution des données. Ainsi par la suite nous augmentons le nombre de réservoirs pour observer le nombre optimal. On remarque que pour  $M>13$  le modèle diverge car les fonctions retournent des paramètres aberrants et des résidus énormes comparés à notre résidu témoin  $\epsilon = 8$ .

On trouve que pour  $M=10$  le modèle est optimal car c'est pour cette valeur que nous avons un résidu le plus faible soit  $\epsilon = 1,7$ .

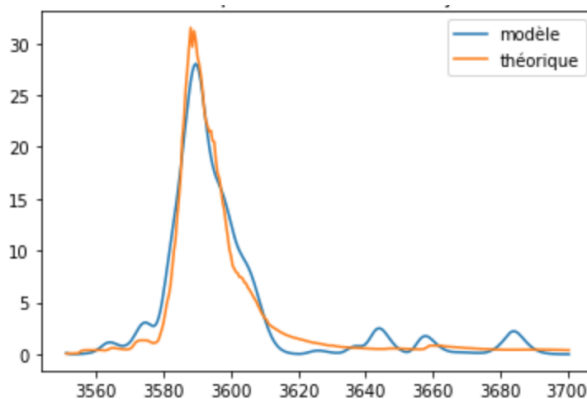


Figure 9: Schéma pour 10 réservoirs comparé aux données théoriques