

Etude de sensibilité du modèle biogéochimique C-RIVE en vue du développement d'une méthodologie d'assimilation de données

Shuaitao Wang^{1*}, Flipo Nicolas¹, Thomas Romary¹

¹ Centre de Géosciences, Mines ParisTech, 35 rue Saint-Honoré, 77305 Fontainebleau

* shuaitao.wang@mines-paristech.fr

Résumé

Les teneurs en oxygène traduisent le métabolisme des cours d'eau et sont utilisées comme indicateur de la santé de ces milieux. Ce métabolisme des fleuves dépend la production d'oxygène par les producteurs primaires autotrophes, de la respiration bactérienne hétérotrophe, ainsi que du processus physique de réaération à l'interface air-eau. Ce paramètre est suivi en continu dans le milieu depuis longtemps. Cependant, les modèles numériques peinent à parfaitement reproduire la dynamique de cette variable clé car leur paramétrisation est incertaine. L'objectif de la thèse est d'assimiler en continu les mesures d'oxygènes dissous de la colonne d'eau afin de déterminer l'évolution temporelle des propriétés physiologiques des communautés vivantes. La première étape de ce travail consiste en une analyse de sensibilité du module biogéochimique C-RIVE, cœur biogéochimique des modèles du PIREN-Seine. Cette analyse a pour objectif de hiérarchiser l'ensemble des paramètres selon leur influence, en vue de développer un schéma d'assimilation de données le plus pertinent possible.

Dans ce travail, le fonctionnement biogéochimique au sein d'un cas d'étude synthétique, caractéristique de la Seine à la traversée de l'agglomération parisienne, est simulé par C-RIVE sous différents états trophiques (Pré-bloom, Bloom et Post-bloom). L'analyse de sensibilité est réalisée pour une communauté phytoplanctonique avec 25 paramètres, une communauté bactérienne hétérotrophe avec 9 paramètres, et 5 paramètres physiques, en suivant les deux méthodologies de Morris pour restreindre le nombre de paramètres à étudier, puis de Sobol pour quantifier les effets croisés de paramètres.

Pour l'état trophique en bloom, le paramètre le plus sensible est $\text{maint}_{\text{phy}}$ (Taux de respiration liée à la survie de la cellule). Les paramètres photosynthétiques sont aussi sensibles. En revanche, les paramètres bactériens sont peu sensibles lors du bloom des algues.

Pour les états trophiques pré-bloom et post-bloom, le rendement de croissance de la bactérie (Y_{bp}) est le plus sensible. Egalement, le taux de croissance maximal ($\mu_{\text{max, bp}}$) et le taux de mortalité (mort_{bp}) sont identifiés comme sensibles. Les paramètres phytoplanctoniques ne sont pas sensibles.

Cette étude permet de sélectionner 16 paramètres sensibles pour l'assimilation de données. Selon les résultats des scénarios simulés, un traitement différencié sera effectué pour les périodes bloom.

Introduction

Les teneurs en oxygène dans l'eau des cours d'eau résultent de nombreux processus physiques et biologiques. Ces teneurs en oxygène traduisent le métabolisme des cours d'eau, et sont utilisées comme indicateur de la santé de ces derniers (Odum, 1956; Needoba et al., 2012; Escoffier et al., 2016). Ce métabolisme des fleuves dépend de la production d'oxygène par les producteurs primaires autotrophes, de la respiration bactérienne hétérotrophe, ainsi que les processus physiques de réaération à l'interface air-eau et eau-sédiment (Odum, 1956; Vilmin et al., 2016b).

L'apparition de capteurs in-situ en continu et à pas de temps fin permet de suivre très précisément l'évolution des teneurs en oxygène du milieu depuis quelques années comme dans le cas de la Seine (Réseau MeSeine du SIAAP, le projet CARBOSEINE (Groleau et al., 2012; Groleau et al., 2013)). Grâce à ces données haute fréquence et aux modèles numériques de qualité de l'eau développés, de nombreuses études ont pu être réalisées, comme le suivi du métabolisme de la Seine en temps réel (Escoffier et al., 2016), et la quantification des flux de carbone internes à la Seine à la traversée de l'agglomération parisienne (Vilmin et al., 2016a). De plus, il devient possible de simuler finement les processus biogéochimiques sur des durées de plus en plus longues (Vilmin, 2014).

En outre, les processus biogéochimiques nécessitent de nombreux paramètres caractéristiques. Les nombreuses recherches menées dans le cadre du PIREN Seine ont permis dans les années 1990 de déterminer des structures moyennes de communautés, et grâce à de multiples expériences de laboratoire, d'en déterminer les paramètres physiologiques (Garnier et al., 1992; Garnier & Billen, 1993; Garnier et al., 1995; Garnier et al., 1998). Ces paramètres physiologiques sont aussi importants que les mesures à haute fréquence pour modéliser correctement les comportements des producteurs primaires, des bactéries hétérotrophes et nitrifiantes (Vilmin, 2014; Raimonet et al., 2015). Les études ont permis d'établir et de quantifier le lien causal entre pressions humaines, du fait des rejets des grandes agglomérations dans les rivières qui les traversent, et le métabolisme des cours d'eau (Even et al., 2007a; Flipo et al., 2007b; Escoffier et al., 2016; Vilmin et al., 2016a).

Or les modèles, aussi sophistiqués soient-ils, peinent à reproduire les teneurs en oxygène car leur paramétrisation est incertaine dans le cas d'un transfert de paramètres de communautés vivantes déterminés dans les conditions contrôlées du laboratoire au milieu naturel. Ceci nous pousse à réfléchir à l'usage que nous pouvons faire de ces nouvelles technologies. Parallèlement, des méthodes numériques permettent aujourd'hui d'assimiler des chroniques observées dans une simulation numérique et de modifier au choix les variables d'état ou les paramètres du modèle afin de mieux déterminer l'évolution des teneurs en oxygène du milieu.

Ces problématiques constituent un des objectifs opérationnels de la phase VII du PIREN Seine et le cœur de ce travail.

1 ProSe et C-RIVE

Le modèle ProSe a été initialement développé par (Even, 1995; Even et al., 1998) à MINES ParisTech. Ce modèle est constitué par trois modules : hydraulique, transport et biogéochimique (Fig. 1) Ensuite, Flipo et al., (2004) ont intégré la simulation du périphyton dans le module biogéochimique RIVE (Billen et al., 1994; Garnier et al., 1995).

Le module hydraulique résout les équations de Saint Venant à une dimension. La résolution numérique se fait par le schéma de PREISSMAN. Le module de transport décrit les mécanismes d'advection et de dispersion.

Le module biogéochimique RIVE permet de simuler les processus biogéochimiques comme la photosynthèse, la respiration, la croissance ainsi que les cycles des nutriments (azote, phosphore) dans plusieurs compartiments (EAU, VASE et PERIPHYTON). L'échange de dioxygène à l'interface eau-air et les diffusions turbulentes entre eau-sédiment sont également intégrés dans le module. Les processus comme l'érosion/sédimentation, l'adsorption du phosphore, et la dynamique de l'azote ont été récemment améliorés (Vilmin et al., 2015a; Raimonet et al., 2015).

2 La méthodologie d'analyse de sensibilité

Dans cette section, la méthodologie suivie pour l'analyse de sensibilité est présentée. L'analyse de sensibilité est conceptualisée avec des nombres scalaires. Cependant nous travaillons sur des données temporelles (Fig. 2.) Afin d'effectuer l'analyse de sensibilité, il convient de transformer les chroniques (des vecteurs temporelles) en nombres scalaires. Une ACP est utilisée pour réduire la dimension des données simulées et calculer un indice (Partie 2.1) pour chaque simulation. Ensuite les indices sont analysés par la méthode de Morris afin de restreindre le nombre de paramètres (Partie 2.2 (Morris, 1991)), puis par celle de Sobol (Partie 2.3, (Sobol, 1993)).

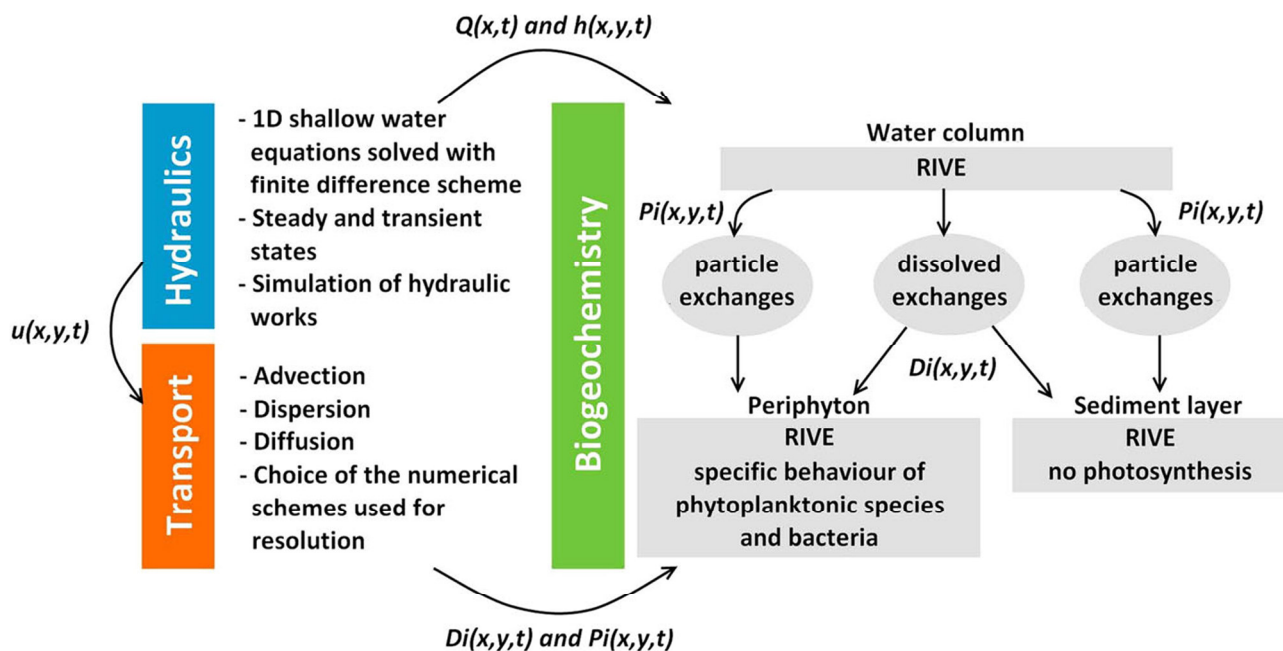


Figure 1. Schéma simplifié du logiciel Prose Schéma simplifié du logiciel Prose (d'après Vilmin et al., (2015b))

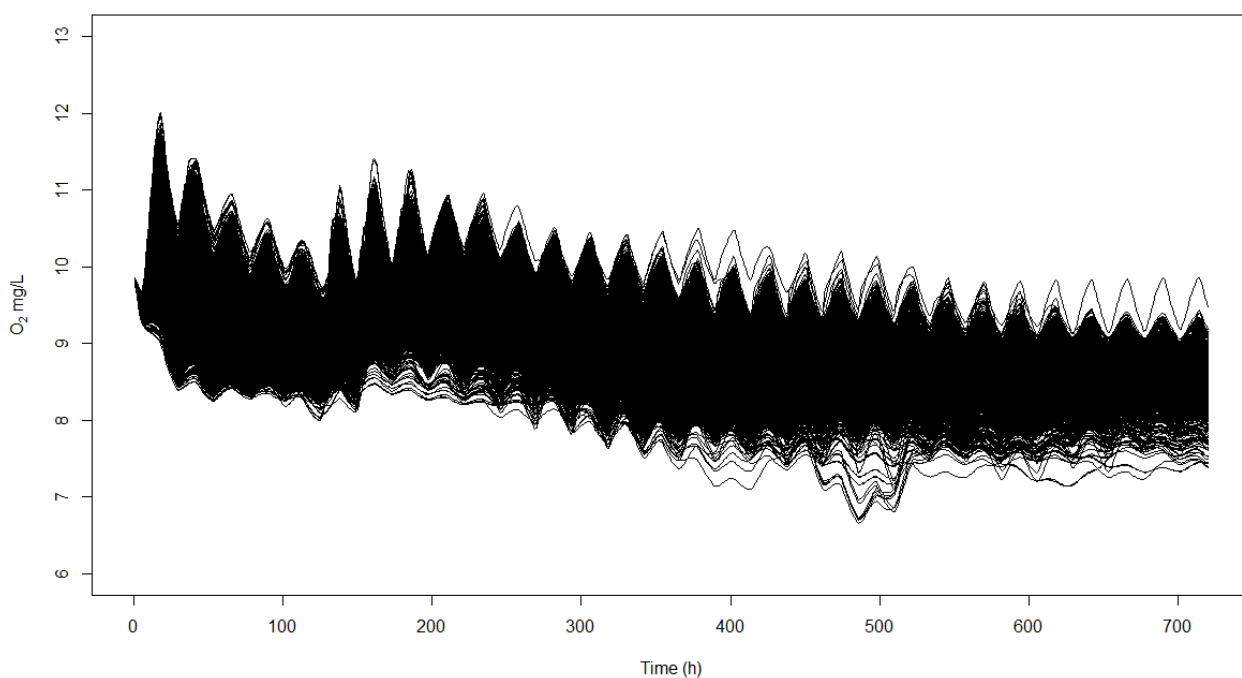


Figure 2. Profils ponctuels d'oxygène dissous simulés en bloom (8000 simulations)

2.1 Réduction de la dimension des données (ACP)

L'ACP est une méthode factorielle de réduction de dimension des données multivariées en conservant leurs informations (Wold et al., 1987). Elle se base sur la décomposition des variances. Dans cette analyse, la dimension des données n correspond à la longueur de la chronique simulée (le nombre de pas de temps analysés).

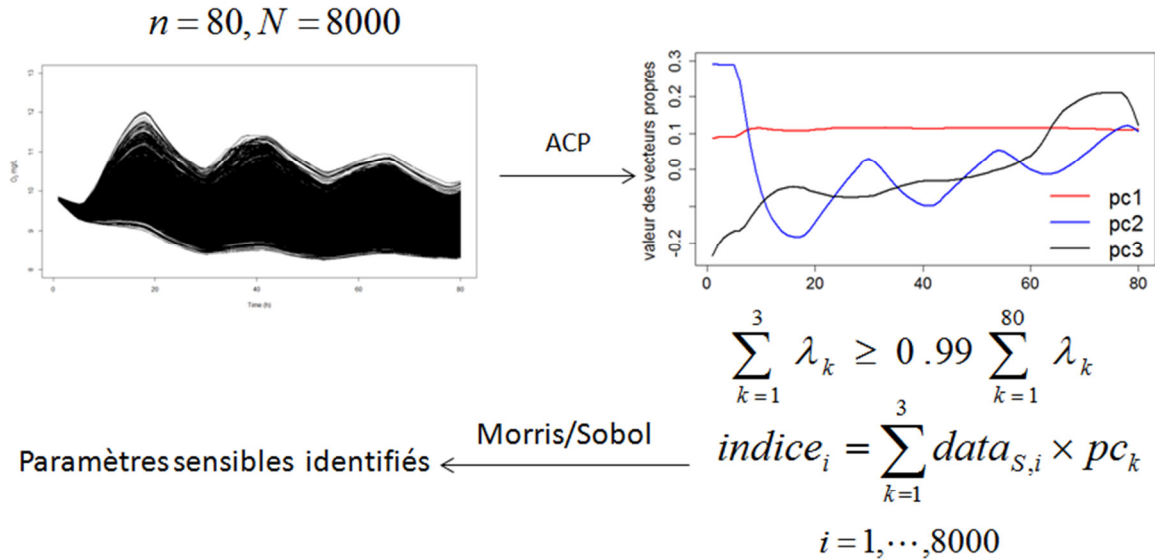


Figure 3. Schéma synthétique de la méthodologie suivie pour l'analyse de sensibilité.
Avec n : le nombre de pas de temps (colonnes), N : le nombre de simulations (lignes), $data_{S,i}$: les données de la i ème simulation, $indice$: indices utilisés pour l'analyse de sensibilité, pc : vecteurs propres, λ : valeurs propres associées

Nous avons utilisé la fonction *prcomp* de *R* pour calculer les composantes principales *pc*. On peut sélectionner p composantes en conservant plus de 99% de la variance de la sortie du modèle (Fig. 3, $p = 3$). Les données simulées sont notées $data_S$. Comme les p composantes sont des vecteurs orthogonaux, on peut définir l'indice pour la i ème simulation comme la somme des produits scalaires entre $data_{S,i}$ ($i = 1, \dots, N$) et pc_k ($k = 1, \dots, p$) (Eq. 1), c'est-à-dire la somme des coefficients de la représentation de chaque chronique dans la base des vecteurs propres. Cet indice permet de résumer la variabilité de la chronique en un unique scalaire et permet de caractériser plus de 99% de la variance de $data_S$. Il peut ainsi être employé pour l'analyse de sensibilité (Lamboni et al., 2011).

$$indice_i = \sum_{k=1}^p data_{S,i} \times pc_k \quad (1)$$

Puisque les petits métabolites (M) et les produits de réserve (R) s'épuisent au bout d'une centaine de pas de temps, nous n'avons mené l'analyse sur les 80 premiers pas de temps (80 heures, Fig. 3).

2.2 Méthode de Morris

La méthode de Morris est l'une des méthodes de criblage (« screening ») se basant sur la discrétisation des paramètres d'entrée en n niveaux. La technique de criblage la plus utilisée est le plan **OAT** « One At a Time ». Il s'agit de modifier un paramètre en fixant les autres pour chaque simulation. Un plan **OAT** consiste à faire varier successivement tous les paramètres analysés.

La méthode de Morris (Morris, 1991) préconise d'effectuer r fois un plan **OAT** aléatoirement en balayant dans l'ensemble de l'espace des paramètres analysés (range). « Aléatoirement » signifie que l'on tire aléatoirement la valeur de départ des paramètres et la suite de variations (négatif/positif) des paramètres. Cette méthode nécessite $r(p + 1)^1$ simulations.

¹ p paramètres + un état de référence

Supposons que l'on répète r plans **OAT**, chaque répétition i ($i = 1, \dots, r$) permet de calculer un effet élémentaire E_j^i pour le paramètre X_j . L'ensemble des répétitions donnent ainsi r effets élémentaires pour chaque paramètre analysé.

L'effet élémentaire E_j^i du $j^{\text{ème}}$ paramètre pour la $i^{\text{ème}}$ répétition est défini comme suit :

$$E_j^i = \frac{f(X(j) + \Delta e_i) - f(X(j))}{\Delta e_i} \quad (2)$$

Les indices de sensibilité sont notés :

$$\begin{aligned} - \quad \mu_j^* &= \frac{1}{r} \sum_{i=1}^r |E_j^i| && \text{(la moyenne des valeurs absolues des effets élémentaires)} \\ - \quad \sigma_j &= \sqrt{\frac{1}{r} \sum_{i=1}^r (E_j^i - \frac{1}{r} \sum_{i=1}^r E_j^i)^2} && \text{(écart-type des effets élémentaires)} \end{aligned}$$

L'interprétation des indices :

μ_j^* indique l'influence de paramètre $X(j)$ sur la dispersion de la sortie du modèle. Plus μ_j^* est important, plus $X(j)$ est sensible.

Si σ_j est négligeable par rapport à μ_j^* , alors une relation linéaire entre le paramètre $X(j)$ et la sortie du modèle existe.

Cette méthode nous permet ainsi de classer les paramètres en trois groupes :

- Les paramètres peu sensibles, μ_j^* et σ_j négligeables.
- Les paramètres ayant une influence linéaire, σ_j négligeable et μ_j^* non négligeable.
- Les paramètres ayant une interaction/une influence non linéaire, μ_j^* et σ_j non négligeables.

Dans cette étude, la méthode de Morris est utilisée pour réduire le nombre de paramètres. On répète 200 (r) plans **OAT** pour la méthode de Morris. Ceci conduit à réaliser $8000 = 200(39 + 1)^2$ simulations. Les paramètres sont discrétisés en 25 niveaux pour bien balayer l'ensemble de l'espace des paramètres. Les paramètres sensibles identifiés sont ensuite analysés par les indices de Sobol.

2.3 Indice de Sobol

L'indice de Sobol (Sobol, 1993) se base sur la décomposition de variance de la sortie du modèle. Elle peut être utilisée pour analyser les modèles non linéaires. Ce qui semble a priori très probable pour un modèle biogéochimique complexe de simulation de l'état des eaux de surface (Nossent et al., 2011).

Supposons que le modèle soit une fonction comme :

$$Y = f(X) = f(X_1, \dots, X_p) \quad (3)$$

Avec :

Y : la sortie du modèle, ici l'indice calculé à partir de la concentration en oxygène dissous [O_2] la colonne d'eau

X : les paramètres du modèle

Selon Sobol (1993), il est possible de représenter cette fonction par les fonctions élémentaires :

$$f(X) = f_0 + \sum_{i=1}^p f_i(X_i) + \sum_{i=1}^{p-1} \sum_{j=i+1}^p f_{ij}(X_i, X_j) + \dots + f_{1,\dots,p}(X_1, \dots, X_p) \quad (4)$$

² $r(p + 1)$

Où :

$f(\cdot)$ est intégrable sur $\Omega = [0,1]^{p3}$, f_0^4 est une constante et les autres fonctions vérifient les conditions suivantes :

$$\int_0^1 f_{i_1, \dots, i_s}(X_{i_1}, \dots, X_{i_k}) dX_{i_k} = 0 \quad \forall k = 1, \dots, s, \quad \forall \{i_1, \dots, i_s\} \subseteq \{1, \dots, p\} \quad (5)$$

Si les X_i sont aléatoires et indépendantes, on peut déduire la décomposition de la variance fonctionnelle. De plus, cette décomposition est unique.

$$V(Y) = \sum_{i=1}^p V_i(X_i) + \sum_{i=1}^{p-1} \sum_{j=i+1}^p V_{ij}(X_i, X_j) + \dots + V_{1, \dots, p}(X_1, \dots, X_p) \quad (6)$$

La variance totale de la sortie du modèle est définie comme suit :

$$V(Y) = \int_{\Omega^p} f^2(X) dX - f_0^2 \quad (7)$$

Les variances partielles (Eq. 6) sont calculées :

$$V_{i_1, \dots, i_s} = \int_0^1 \dots \int_0^1 f_{i_1, \dots, i_s}^2(X_{i_1}, \dots, X_{i_s}) dX_{i_1}, \dots, dX_{i_s} \quad (8)$$

Où :

$1 \leq i_1 \leq \dots \leq p$ et $s = 1, \dots, p$.

On peut ainsi introduire les indices de sensibilité (Sobol, 1993; Saltelli et al., 2010) :

- Indice du premier ordre $S_i = \frac{V_i}{V}$
- Indice du second ordre $S_{ij} = \frac{V_{ij}}{V}$
- Indice total $S_{T_i} = S_i + \sum_{j \neq i} S_{ij} + \dots + S_{1, \dots, p}$

L'indice du premier ordre S_i représente la contribution individuelle du paramètre X_i sur la variance de la sortie du modèle, appelé aussi « main effect ». L'indice du second ordre S_{ij} caractérise une influence due à l'interaction entre les paramètres X_i et X_j sur la sortie du modèle. L'indice total S_{T_i} indique l'impact global du paramètre X_i y compris toutes les interactions avec d'autres paramètres, sur la réponse du modèle. Sauf si le modèle est linéaire, la somme des S_i est inférieure à 1.

Les indices de Sobol permettent de quantifier la sensibilité des paramètres, et de les classer. Cependant, elle est très couteuse au niveau du temps de calcul. Elle a besoin de milliers de simulations pour évaluer S_i , S_{ij} et S_{T_i} . Dans cette étude, on calcule seulement S_i et S_{T_i} à l'aide d'un package de Python SALib (Herman & Usher, 2017). Saltelli (2002) a proposé une méthode qui a besoin d'effectuer $n(p+2)$ simulations pour calculer S_i et S_{T_i} . Cette méthode est intégrée dans SALib et est utilisée dans cette étude.

Les paramètres sensibles identifiés par la méthode de Morris (Morris, 1991) et les indices de Sobol (Sobol, 1993) seront à ajuster lors de l'assimilation des données.

³ Ω : Une espace de p dimensions, l'espace des paramètres (range) sont normalisés entre 0 et 1

⁴ Si les conditions (Eq. 5) sont vérifiées, f_0 est l'espérance de la sortie du modèle $f(X)$

3 Modèle et conditions initiales

Dans cette section, le cas test et les conditions initiales sont d'abord présentés (Partie 3.1 et 3.2). Ensuite, différents scénarios sont décrits pour identifier les paramètres sensibles selon la situation hydro-biologique (Partie 3.3). Enfin, une liste de paramètres à analyser est établie (Partie 3.4).

3.1 Modèle simulé

On s'intéresse à l'oxygène dissous simulé ($[O_2]$) dans la colonne d'eau. Nous avons construit un cas test (Fig. 4) pour être en situation contrôlée. Le fonctionnement biogéochimique au sein d'un tronçon de 1000 m, de 100 m de large et à section rectangulaire est simulé par C-RIVE.

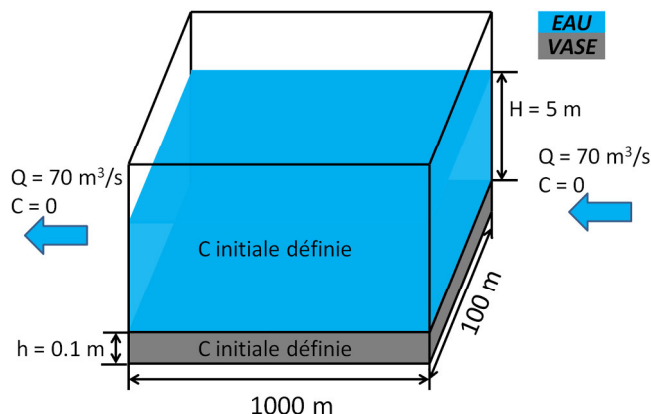


Figure 4. Schéma synthétique du modèle simulé (Q : débit ; C : concentration des espèces simulés)

Ce tronçon est constitué de deux compartiments : EAU (en bleu) et VASE (en gris). La hauteur d'eau dans ce tronçon est de 5 m. Le débit est fixé à $70 \text{ m}^3/\text{s}$ tout au long de la simulation. Ce débit correspond à un étiage avec le temps de retour de 4 ans à la station Austerlitz. Ce cas test simplifié représente des conditions idéalisées en Seine en été, s'inspirant d'un cas test mis en place par (Vilmin, 2014). Il est important de noter que **ce cas test s'apparente au suivi d'une masse d'eau** au fil de l'eau pendant toute la durée de la simulation.

3.2 Concentrations initiales

Les concentrations des espèces biogéochimiques amont sont nulles. Afin d'estimer les concentrations initiales, on a effectué une simulation de ProSe sur la période 2007-2012 (Fig. 5, Vilmin, (2014)). Les concentrations initiales des espèces biogéochimiques sont déterminées par la moyenne de la période 2007-2012 (Tab.1) sauf la biomasse bactérienne hétérotrophe et phytoplanctonique, MOD1 et MOD2 qui sont choisies selon la période simulée (Partie 3.3).

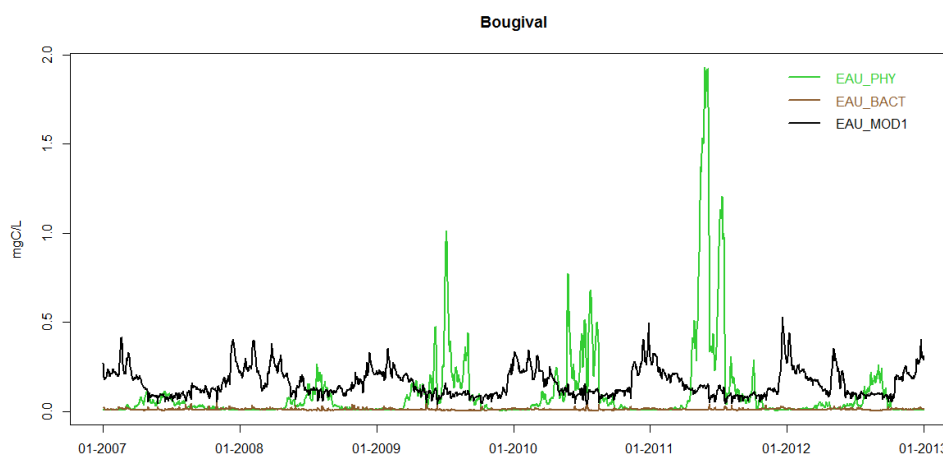


Figure 5. Chronique de biomasse phytoplanctonique et bactérienne à Bougival sur la période 2007-2012 dans EAU

3.3 Scénarios simulés

Le cas test (Fig. 4) est utilisé pour simuler trois états trophiques du système.

- **Pré-bloom**, où il existe peu de bactéries et peu de phytoplanctons.
- **Bloom**, où le phytoplancton est dominant.
- **Post-bloom**, où il y a beaucoup de bactéries et peu de phytoplancton

Pour chaque situation, on a testé deux cas : avec stock de VASE et sans stock de VASE initial. Le cas avec stock de VASE caractérise une année sèche où les sédiments s'accumulent. Tandis que le cas sans stock de VASE signifie une année humide où les sédiments sont évacués par fort débit. Afin de simuler la période d'étiage, un débit faible et une température relativement élevée sont utilisés pour toutes les simulations (Escoffier et al., 2016). Les états initiaux sont estimés sur la base de la simulation pluri-annuelle de Vilmin, (2016) autour du bloom de mars 2011.

Tableau 1. Les concentrations initiales des espèces simulées

Espèces	C _{init,EAU}	C _{init,VASE}	Unité
<i>Bactn1</i>	0.003	0.105	mgC/L
<i>Bactn2</i>	0.001	0.027	mgC/L
NH_4^+	0.182	0.325	mgN/L
NO_2^-	0.041	0.043	mgN/L
NO_3^-	4.769	4.514	mgN/L
<i>MES</i>	16.820	95010.433	mg/L
<i>MOD3</i>	1.948	1.944	mgC/L
<i>MOP1</i>	0.125	43.780	mgC/L
<i>MOP2</i>	0.232	695.784	mgC/L
<i>MOP3</i>	0.403	2555.144	mgC/L
PO_4^{3+}	0.100	0.266	mgP/L
O_2	9.931	6.654	mgO ₂ /L

3.4 Paramètres à analyser

Afin de réaliser l'analyse de sensibilité, nous avons sélectionné 39 paramètres soit 5 paramètres physiques relatifs à la réaération air-eau et la diffusion turbulente eau-sédiment, 9 paramètres bactériens, et 25 paramètres liés au phytoplancton. Grâce à des études réalisées (Garnier et al., 1995; Garnier et al., 1998; Billen & Servais, 1991; Even et al., 1998; Flipo et al., 2004; Flipo et al., 2007b; Vilmin, 2014) dans le cadre du PIREN Seine, on peut trouver des références des paramètres pour la Seine (Tab. 2) La stoechiométrie pour la composition en carbone, azote et phosphore de la matière organique, est fixée au rapport de Redfield molaire C/N/P = 106/6/1 (Redfield et al., 1963).

Tableau 2. Liste des paramètres à analyser

Nom	Description	Val_min	Val_ref	Val_max	Unité	Réfs
Paramètres physiques						
D_m^*	Coefficient de diffusion moléculaire à l'interface eau-air	$1.35 \cdot 10^{-5}$	$1.8 \cdot 10^{-5}$	$2.25 \cdot 10^{-5}$	cm^2/s	(Vilmin, 2014)
$D_{m,Richey}^*$	Coefficient de diffusion moléculaire à l'interface eau-air lorsque l'eau est stagnante	$1.95 \cdot 10^{-5}$	$2.6 \cdot 10^{-5}$	$3.25 \cdot 10^{-5}$	cm^2/s	(Vilmin, 2014)
D_s^*	Coefficient de diffusion turbulente eau-sédiment	$1.25 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$1.875 \cdot 10^{-3}$	m/s	(Garnier et al., 1995; Even et al., 2004; Flipo et al., 2004)
K_{navig}^*	Coefficient de réaération due à la navigation	0	0	0.05	m/h	(Vilmin, 2014)
K_{wind}^*	Coefficient de réaération due au vent	88.5	118	147.5	cm/h	(Vilmin, 2014)
Paramètres bactériens						
$\mu_{max,bact}$	Taux de croissance maximal lié à la bactérie	0.015	0.04	0.04	$/h$	(Servais et al., 1985; Garnier et al., 1992)
ρ_{bact}^*	Masse volumique de la bactérie	1.1	1.2	1.5	g/cm^3	(Vilmin, 2014)
σ_{bact}^*	Ecart-type de la fonction de température	12.75	17	21.25	-	(Garnier et al., 1992; Garnier et al., 1998)
Φ_{bact}^*	Porosité de la bactérie	0.5	0.9	0.9	-	(Vilmin, 2014)
$K_{O_2,bact}^*$	Constante demi-saturation vis-à-vis de l'oxygène dissous	0.375	0.5	0.625	mgO_2/L	(Even et al., 1998; Flipo et al., 2004)
$mort_{bact}$	Taux de mortalité liée à la bactérie	0.01	0.02	0.03	$/h$	(Servais et al., 1985; Garnier et al., 1995; Even et al., 1998)
$T_{opt,bact}$	Température optimale pour la bactérie	10	25	35	$^{\circ}C$	(Billen & Servais, 1991; Garnier et al., 1998)
$V_{sed,bact}$	Vitesse de sédimentation de la bactérie	0.005	0.005	0.1	m/h	(Vilmin, 2014; Garnier et al., 1995)
Y_{bact}	Rendement de croissance lié à la bactérie	0.03	0.25	0.46	-	(Billen & Servais, 1991; Giorgio & Cole, 1998)
Paramètres phytoplanktoniques						
α_{phy}^*	Capacité photosynthétique	0.0009	0.0012	0.0015	$m^2s/\mu E/h$	(Garnier et al., 1995; Even et al., 2004; Flipo

$\eta_{chla,phy}$	Terme de l'extinction lumineuse fonction de biomasse phytoplanktonique	0.006	0.02	0.054	$L/\mu gchla/m$	et al., 2004) (Bailey & Ahmadi, 2014; Garnier et al., 1995)
$\eta_{mes,phy}^*$	Terme de l'extinction lumineuse fonction des matières en suspension	0.03	0.04	0.05	$L/mg/m$	(Flipo et al., 2004; Vilmin, 2014)
η_{phy}	Terme de base de l'extinction lumineuse	0.2	1.0	1.8	$/m$	(Flipo et al., 2004; Bailey & Ahmadi, 2014)
μ_{phy}	Taux de croissance maximal du phytoplancton	0.042	0.12	0.125	$/h$	(Even et al., 2007a; Cox, 2003; Cho et al., 2017)
ρ_{phy}^*	Masse volumique du phytoplancton	1.1	1.2	1.5	g/cm^3	(Vilmin, 2014)
σ_{phy}^*	Ecart-type de la fonction de température	12.75	17	21.25	-	(Garnier et al., 1995; Garnier et al., 1998; Descy et al., 2012)
Φ_{phy}^*	Porosité du phytoplancton	0.5	0.9	0.9	-	(Vilmin, 2014)
$C/chla_{phy}$	Ratio carbone chlorophylle a	0.0265	0.035	0.04375	$\mu g/L$	(Garnier et al., 1995; Flipo et al., 2007b; Vilmin, 2014)
$C_{R,phy}^*$	Taux de catabolyse des produits de réserve	0.15	0.2	0.25	$/h$	(Garnier et al., 1995)
$\delta_{lim,phy}^*$	Facteur multiplicatif de la mortalité pour une concentration supérieure à $lim_{mort,phy}^*$	1.0	1.0	1.25	-	(Vilmin, 2014)
$ener_{g,phy}$	Taux de respiration liée à la dépense énergétique de croissance	0.375	0.5	0.625	-	(Garnier et al., 1995)
$E_{C-te,phy}$	Taux d'excrétion de base	$7.5 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$	$1.25 \cdot 10^{-3}$	$/h$	(Garnier et al., 1995; Even et al., 2004)
$E_{phot,phy}^*$	Taux d'excrétion liée à la photosynthèse	$4.5 \cdot 10^{-4}$	$6.0 \cdot 10^{-4}$	$7.5 \cdot 10^{-4}$	$/h$	(Garnier et al., 1995)
K_{phy}^*	Constante demi-saturation vis-à-vis du quota cellulaire des petits métabolites	0.045	0.06	0.075	-	(Garnier et al., 1995)
$K_{N,phy}^*$	Constante demi-saturation vis-à-vis de l'azote	10	70	300	$\mu gN/L$	(Cox, 2003; Descy et al., 2012)
$K_{NH_4^+,phy}^*$	Constante demi-saturation vis-à-vis de l'ammonium	67.5	90	112.5	$\mu g/L$	(Vilmin, 2014)
$K_{P,phy}$	Constante demi-saturation vis-à-vis du phosphore	1	46	50	$\mu gP/L$	(Cox, 2003; Garnier et al., 1995; Cho et al.,

$lim^*_{mort,phy}$	Concentration critique au-dessus de la quelle la mortalité est accélérée	2.1	2.625	mgC/L	2017) (Vilmin, 2014)
$maint_{phy}$	Taux de respiration liée à la survie de la cellule	0.002	0.021	/h	(Cox, 2003; Garnier et al., 1995)
$mort^*_{phy}$	Taux de mortalité du phytoplancton	0.004	0.005	/h	(Garnier et al., 1995; Vilmin, 2014)
$P_{max,phy}$	Taux de production photosynthétique maximal	0.4	0.5	/h	(Capblancq & Dauta, 1978; Garnier et al., 1995)
$S^*_{R,phy}$	Taux de synthèse des produits de réserve	0.26	0.325	/h	(Garnier et al., 1995; Even et al., 2004)
$T_{opt,phy}$	Température optimale pour le phytoplancton	33	37	°C	(Garnier et al., 1995; Giorgio & Cole, 1998)
$V_{sed,phy}$	Vitesse de sédimentation pour le phytoplancton	0.05	0.075	m/h	(Bailey & Ahmadi, 2014; Cho et al., 2017; Garnier et al., 1995)

*: $range = Val_{ref} \times (1 \pm 25\%)$

4 Résultats et conclusions

4.1 L'analyse par la méthode de Morris

Pour la méthode de Morris, on a analysé les 3 scénarios (Partie 3.3) : **Pré-bloom**, **Bloom** et **Post-bloom**.

4.1.1 Pré-bloom

Les paramètres les plus sensibles sont ceux relatifs aux bactéries (Fig. 6). Le paramètre le plus sensible est le rendement de croissance bactérien (Y_{bp}), suivi par ordre décroissant du taux de croissance maximal ($\mu_{max,bp}$), de la mortalité ($mort_{bp}$), puis de l'optimum thermique de croissance ($T_{opt,bp}$). Le second groupe de paramètres sensibles concernent les paramètres physiques (D_m , coefficient de diffusion moléculaire air-eau ; K_{navig} , coefficient de réaération due à la navigation). Pour les paramètres de réaération, les σ sont beaucoup plus faibles que les μ^* . Ceci indique une influence plutôt linéaire sur la sortie du modèle (Partie 2.2). Par contre, les autres paramètres biologiques ont un effet d'interaction, ce qui est souvent attendu dans un modèle de simulation de la qualité des eaux. Pour finir les paramètres phytoplanctoniques sont peu sensibles. Néanmoins, parmi eux, la respiration de maintenance est la plus sensible ($maint_{phy}$).

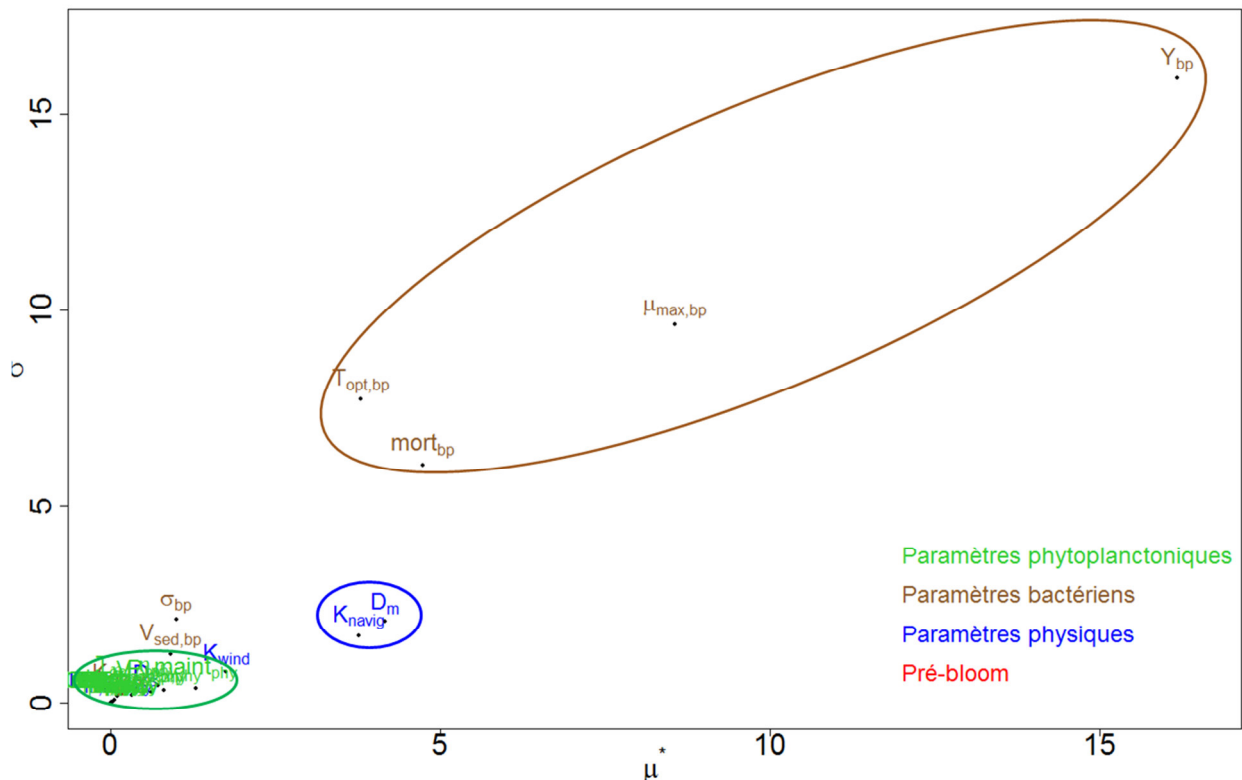


Figure 6. Analyse de sensibilité pré-bloom 80 heures (Morris) ($Bact_{EAU} = 0.007 \text{ mgC/L}$, $hy_{EAU} = 0.010 \text{ mgC/L}$, $MOD1_{EAU} = 0.2 \text{ mgC/L}$, $MOD2_{EAU} = 0.45 \text{ mgC/L}$, $Bact_{VASE} = 0.025 \text{ mgC/L}$, $Phy_{VASE} = 0.003 \text{ mgC/L}$, $MOD1_{VASE} = 0.13 \text{ mgC/L}$, $MOD2_{VASE} = 0.45 \text{ mgC/L}$), $r = 200$

4.1.2 Bloom

Pendant la simulation d'un **bloom**, le métabolisme phytoplanctonique est dominant et le système est qualifié d'autotrophe. Les paramètres phytoplanctoniques deviennent les plus sensibles (Fig. 7). Par ordre décroissant de sensibilité, on trouve la respiration de maintenance ($maint_{phy}$), à l'optimum thermique de croissance ($T_{opt,phy}$), et le taux maximum de photosynthèse ($P_{max,phy}$). Les paramètres physiques (D_m , K_{navig}) et bactériens (Y_{bp} , $\mu_{max,bp}$) sont quant à eux, aussi sensibles mais de manière moins importante que les paramètres phytoplanctoniques.

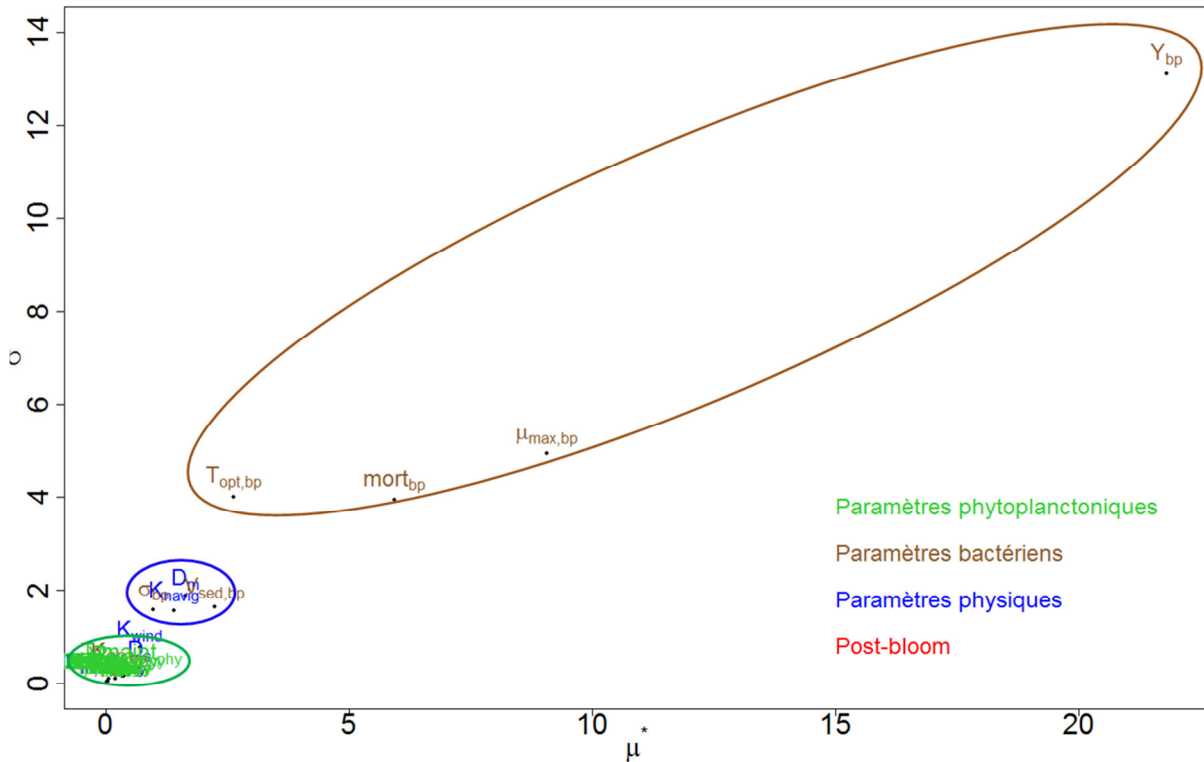


Figure 8. Analyse de sensibilité post-bloom 80 heures (Morris) ($Bact_{EAU} = 0.04 \text{ mgC/L}$, $Phy_{EAU} = 0.01 \text{ mgC/L}$, $r = 200$)

4.1.4 Bilan des analyse réalisées avec la méthode de Morris

Les trois analyses de Morris montrent que les situations de bloom présentent un contexte singulier, dans lequel les paramètres phytoplanctoniques - $maint_{phy}$ (taux de respiration liée à la survie de la cellule), $T_{opt,phy}$ (optimum thermique de croissance), $P_{max,phy}$ (taux maximum de photosynthèse) - sont les plus sensibles. Or de ces périodes, ce sont les paramètres bactériens - Y_{bp} (rendement de croissance), $\mu_{max,bp}$ (taux de croissance maximal), $T_{opt,bp}$ (température optimale), $mort_{bp}$ (taux de mortalité maximal) et σ_{bp} (écart-type de la fonction de température) - qui deviennent les plus sensibles, suivis des paramètres physiques (D_m , K_{wind} , K_{navig} , D_s). Ces paramètres restent relativement sensibles en condition de bloom.

Les scénarios **pré et post-bloom** sont très similaires, on distinguera donc seulement deux contextes dans la suite de ce travail : **Bloom**, et or bloom représenté par le scénario **Post-bloom**.

4.2 L'analyse par les indices de Sobol

Pour évaluer les indices de Sobol, un échantillonnage de dimension de 24000 (n) a été choisi. On sélectionne 19 (p) paramètres sensibles identifiés par la méthode de Morris (Fig. 9). On effectue ainsi 504000 ($= 24000(19 + 2)$)⁵ simulations au total. L'analyse de Morris a aussi permis d'identifier deux contextes, **Bloom** et or bloom (scenario **Post-bloom**), pour lesquels les indices de Sobol sont calculés.

⁵ $N = n(p + 2)$

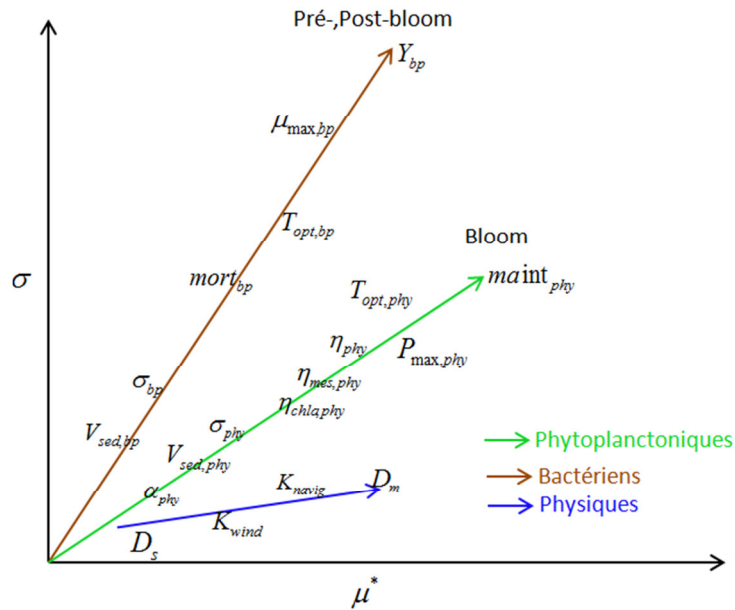


Figure 9. Les paramètres sensibles identifiés pour l'analyse des indices de Sobol par la méthode de Morris.

4.2.1 Bloom

Le paramètre le plus sensible est $maint_{phy}$ (Fig. 10). L'effet total de $maint_{phy}$ est de l'ordre de 80% (Tab. 3). Son influence est 10 fois plus importante que celle du paramètre classé en deuxième place $T_{opt,phy}$ qui a une contribution de 7%. Les paramètres bactériens ne sont pas sensibles. Pour les paramètres sensibles, les écarts entre S_i et S_{T_i} indiquent des effets d'interaction avec d'autres paramètres (Tab. 3). Les 3 paramètres de réaération (D_m , K_{navig} , K_{wind}) sont peu sensibles et ont une influence linéaire sur la réponse du modèle. L'analyse par les indices de Sobol est cohérente avec l'analyse par la méthode de Morris pour le scénario **Bloom** (Partie 4.1.2, Fig. 7) et permet de restreindre le nombre de paramètres sensibles

Tableau 3. Les indices de Sobol estimés pour le scénario Bloom

Paramètres	S_i	Intervalle de confiance S_i	S_{T_i}	Intervalle de confiance S_{T_i}
$maint_{phy}$	0.776	± 0.015	0.802	± 0.013
$T_{opt,phy}$	0.044	± 0.004	0.068	± 0.002
$P_{max,phy}$	0.047	± 0.004	0.050	± 0.001
η_{phy}	0.024	± 0.003	0.027	± 0.001
$\eta_{chla,phy}$	0.021	± 0.003	0.025	± 0.001
$\eta_{mes,phy}$	0.015	± 0.002	0.017	± 0.000
D_m	0.011	± 0.002	0.012	± 0.000
α_{phy}	0.009	± 0.002	0.010	± 0.000
K_{navig}	0.008	± 0.002	0.009	± 0.000
σ_{phy}	0.003	± 0.002	0.008	± 0.000
Y_{bp}	0.003	± 0.001	0.003	± 0.000
K_{wind}	0.002	± 0.001	0.002	± 0.000
...
Somme	0.964		1.037	

L'intervalle de confiance est estimé par Bootstrap (Efron & Tibshirani, 1993). La taille de rééchantillonnage est égale à 100. Ce processus est intégré dans SALib. ... : Les paramètres ayant $S_i < 0.0005$ ne sont pas présentés.

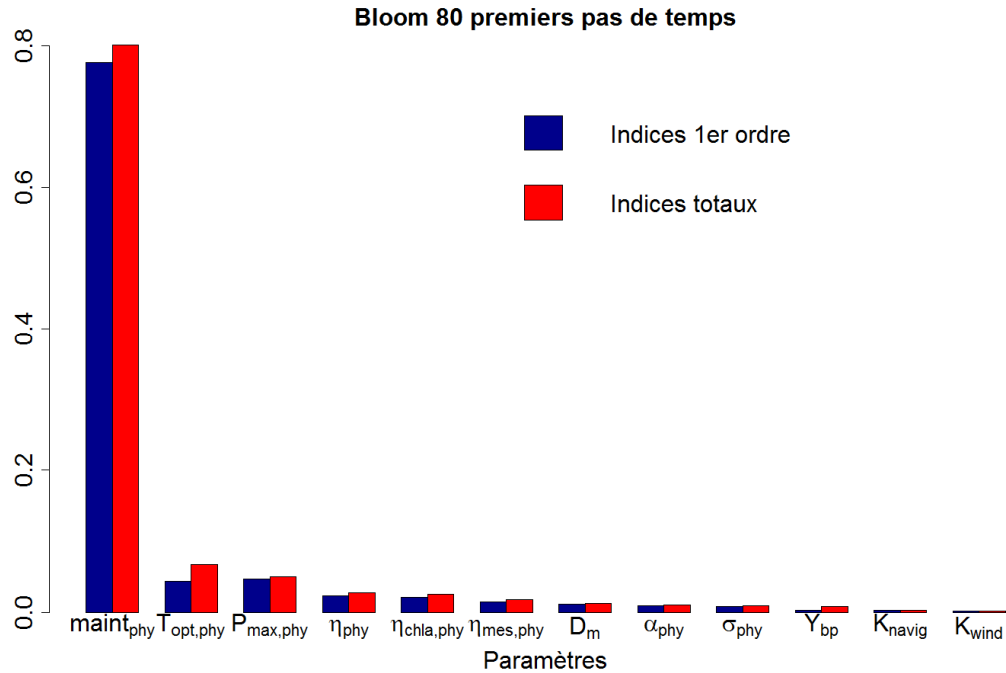


Figure 10. Représentation des indices de Sobol estimés pour le scénario Bloom

4.2.2 Post-bloom

Le paramètre le plus sensible est Y_{bp} (le rendement de croissance lié à la bactérie) qui a un effet total de 79% (Fig. 11&Tab. 4). Les 4 paramètres bactériens ($\mu_{max,bp}$, $mort_{bp}$, $T_{opt,bp}$ et $V_{sed,bp}$) sont sensibles (16%, 6%, 3% et 3%). Puisque le métabolisme phytoplanctonique est faible pour le scénario Post-bloom, Les effets totaux des paramètres phytoplanctoniques sont négligeables. Les paramètres physiques sont peu sensibles. L'analyse par les indices de Sobol est cohérente avec celle réalisée par la méthode de Morris (Fig. 8).

Tableau 4. Les indices de Sobol estimés pour le scénario Post-bloom

Paramètres	S_i	Intervalle de confiance S_i	S_{T_i}	Intervalle de confiance S_{T_i}
Y_{bp}	0.734	± 0.018	0.788	± 0.014
$\mu_{max,bp}$	0.120	± 0.007	0.155	± 0.004
$mort_{bp}$	0.041	± 0.004	0.057	± 0.002
$T_{opt,bp}$	0.017	± 0.003	0.028	± 0.001
$V_{sed,bp}$	0.013	± 0.003	0.025	± 0.001
D_s	0.002	± 0.001	0.006	± 0.000
D_m	0.003	± 0.001	0.006	± 0.000
K_{navig}	0.002	± 0.001	0.005	± 0.000
σ_{bp}	0.002	± 0.001	0.005	± 0.000
$maint_{phy}$	0.001	± 0.001	0.001	± 0.000
K_{wind}	0.001	± 0.001	0.001	± 0.000
...
Somme	0.936		1.078	

L'intervalle de confiance est estimé par Bootstrap (Efron & Tibshirani, 1993). La taille de rééchantillonnage est égale à 100. Ce processus est intégré dans SALib. ... : Les paramètres ayant $S_i < 0.0005$ ne sont pas présentés.

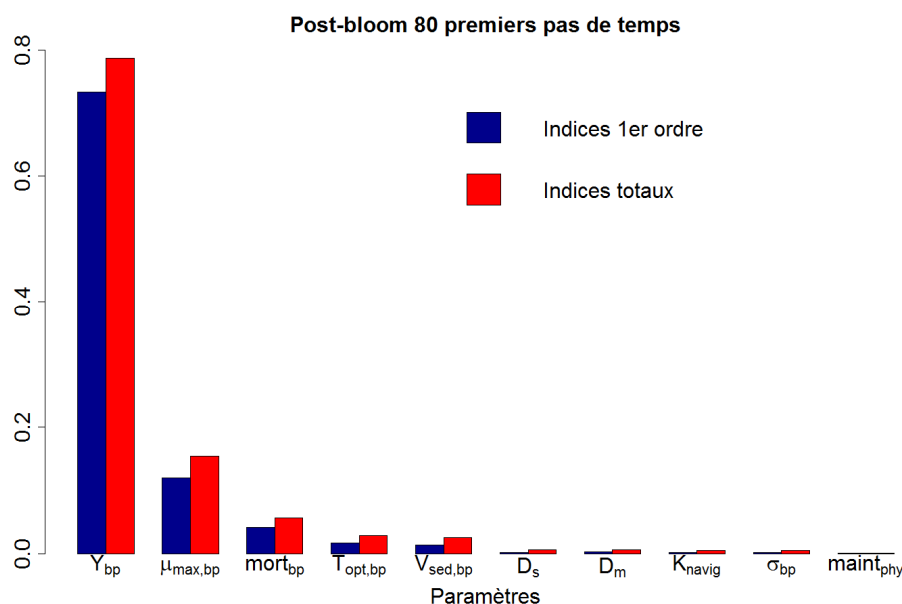


Figure 11. Représentation des indices de Sobol estimés pour le scénario Post-bloom

5 Sélection des paramètres à assimiler

Afin de sélectionner les paramètres à assimiler pour chaque scénario, on définit un critère qui permet d'identifier les paramètres contribuant à 99% de la variance des données simulées (Eq. 9) :

$$\frac{\sum_{i=1}^I S_{T_i}}{\sum_{i=1}^p S_{T_i}} \geq 0.99 \quad (9)$$

Avec I : le nombre de paramètres sélectionnés

p : le nombre de paramètres analysés

Ainsi, 16 paramètres sont sélectionnés pour l'assimilation de données (Fig. 12, Fig. 13, Tab. 5). Selon les résultats des scénarios simulés, on utilisera un traitement différencié des périodes **bloom/or bloom**. Huit paramètres seront assimilés en période or bloom, auxquels s'ajouteront huit paramètres supplémentaires (essentiellement relatifs aux communautés phytoplanctoniques) en période de bloom.

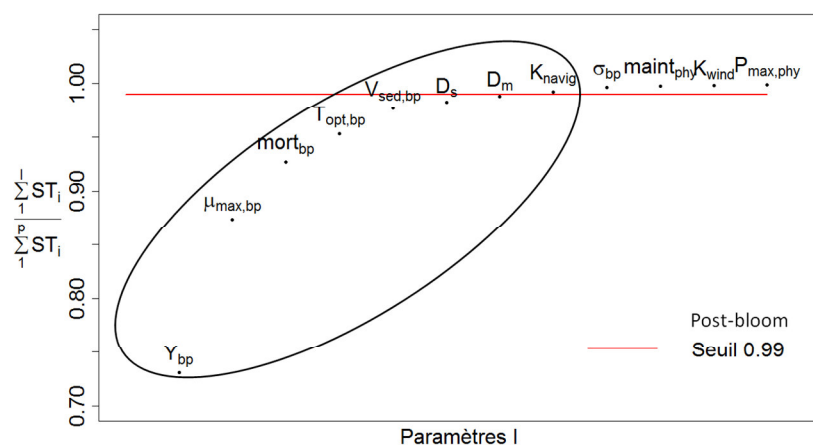


Figure 12. Le cumul des indices totaux normalisés pour le scénario Post-bloom

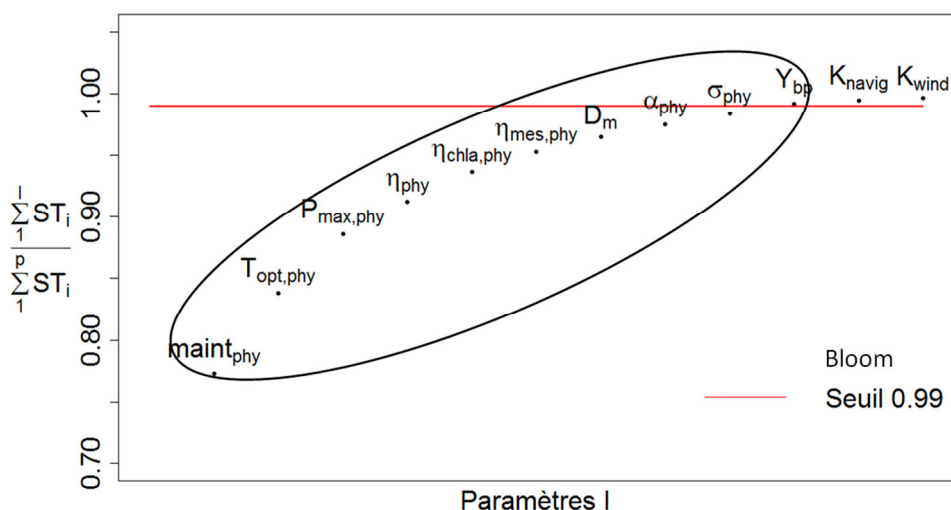


Figure 13. Le cumul des indices totaux normalisés pour le scénario Bloom

Tableau 5. Paramètres sélectionnés pour l'assimilation de données

Paramètres	Description
Paramètres physiques	
D_s	Coefficient de diffusion turbulente à l'interface eau-sédiment
D_m	Coefficient de diffusion moléculaire à l'interface eau-air
K_{navig}	Coefficient de réaération du à la navigation
Paramètres bactériens	
Y_{bp}	Rendement de croissance lié à la bactérie
$\mu_{max,bp}$	Taux de croissance maximal lié à la bactérie
$mort_{bp}$	Taux de mortalité maximal lié à la bactérie
$T_{opt,bp}$	Température optimale pour la bactérie
$V_{sed,bp}$	Vitesse de sédimentation de la bactérie
Paramètres phytoplanctoniques	
$maint_{phy}$	Taux de respiration lié à la survie de la cellule
$T_{opt,phy}$	Température optimale pour le phytoplancton
$P_{max,phy}$	Taux de production photosynthétique maximal
η_{phy}	Terme de base de l'extinction lumineuse
$\eta_{chla,phy}$	Terme de l'extinction lumineuse fonction de biomasse phytoplanctonique
$\eta_{mes,phy}$	Terme de l'extinction lumineuse fonction des matières en suspension
α_{phy}	Capacité photosynthétique
σ_{phy}	Écart-type de la fonction de température

6 Conclusions générales

L'analyse de sensibilité du module biogéochimique C-RIVE a été réalisée par la méthode de Morris et les indices de Sobol sur un cas test idéalisé qui caractérise le suivi d'une masse d'eau (Partie 3.1, p6). Selon les résultats obtenus, les paramètres sensibles dépendent de la biomasse relative des espèces phytoplanctoniques (Fig. 12, Fig. 13, Tab. 5). Pour les simulations **Pré-bloom** ou **Post-bloom**, la perturbation de l'oxygène dissous est principalement liée à l'activité bactérienne (système hétérotrophe), et notamment à leur rendement de croissance. En revanche, la variation de l'oxygène dissous est principalement engendrée par l'activité phytoplanctonique pour la simulation d'un **Bloom** (système autotrophe), notamment par leur respiration de maintenance.

Cette analyse de sensibilité permet de sélectionner 16 paramètres pour l'assimilation de données. Ces 16 paramètres seront à ajuster lors de l'assimilation de données. On souhaite réaliser les simulations les plus longues possibles afin de mieux cerner les évolutions physiologiques des espèces sur le temps long.

A cet effet, la sensibilité du modèle en période de crue devra être testée afin d'identifier les paramètres les plus sensibles à cette période. L'étape suivante de ce travail sera d'analyser la sensibilité du modèle à ses forçages. Pour cela nous utiliserons alors le logiciel prose appliqué à un tronçon de Seine. Vus les temps de calcul importants, il faudra au préalable reconstruire une version parallélisée du logiciel, qui sera validée sur la base des simulations réalisées par (Vilmin, 2014).

Bibliographie

- Billen, G., Garnier, J. & Hanset, P., 1994. Modelling phytoplankton development in whole drainage networks : The riverstrahler model applied to the Seine river system. *Hydrobiologia*, 289, pp. 119–137.
- Billen, G. & Servais, P., 1991. *Modélisation du transport de polluants par l'estuaire de l'escault. Cas du phosphore*. Technical report, Groupe de Microbiologie des Milieux Aquatiques. Université Libre de Bruxelles.
- Escoffier, N. et al., 2016. Estimating ecosystem metabolism from continuous multi-sensor measurements in the seine river. *Environmental Science and Pollution Research*.
- Even, S., 1995. *Modélisation d'un écosystème fluvial : la Seine - Le modèle ProSe*. PH. D. thesis. Ecole Nationale Supérieure des Mines de Paris.
- Even, S. et al., 2007a. Modeling the impacts of Combined Sewer Overflows on the river Seine water quality. *Sciences of Total Environment*, 373(1-3), pp.140–51.
- Even, S. et al., 1998. River ecosystem modelling : Application of the ProSe model to the Seine river (France). *Hydrobiologia*, 373, pp.27-37.
- Even, S. et al., 2004. Modelling oxygen deficits in the Seine river downstream of combined. *Ecol. Model.*, 173, pp.177-96.
- Flipo, N. et al., 2004. Biogeochemical modelling at the river scale : Plankton and periphyton dynamics - Grand Morin case study, France. *Ecol. Model.*, 176, pp.333-47.
- Flipo, N. et al., 2007b. Primary production in headwater streams of the Seine basin : the Grand Morin case study. *Sciences of Total Environment*, 375, pp.98–109.
- Garnier, J., Billen, J. & Coste, M., 1995. Seasonal succession of diatoms and chlorophycae in the drainage network of the river Seine : Observations and modelling. *Limnol. Oceanogr.*, 40(4), pp.750-65.
- Garnier, J. & Billen, G., 1993. Ecological interactions in a shallow sand-pit (Créteil Lake, France). *Hydrobiologia*, 275/276, pp.97-114.
- Garnier, J. et al., 1998. Développement algal et eutrophisation dans le réseau hydrographique de la Seine. In *La Seine en son bassin-Fonctionnement écologique d'un système fluvial anthropisé*. p.593-626.
- Garnier, J., Billen, G. & Servais, P., 1992. Physiological characteristics and ecological role of small and large sized bacteria in a polluted river (seine river, France). *Arch. Hydrobiol. Beih. Ergebn. Limnol.*, 37, pp.83-94.
- Giorgio, P.A.D. & Cole, J.J., 1998. Bacterial growth efficiency in natural aquatic systems. *Annual Review of Ecology and Systematics*, 29(1), pp.503-41.

- Groleau, A. et al., 2012. *Biogéochimie haute résolution : Bilan méthodologique et présentation des observations pour l'année 2011*. Technical report, PIREN Seine.
- Groleau, A. et al., 2013. *Réseau de mesures haute fréquence CarboSeine : validation métrologique et indicateurs du fonctionnement trophique*. Technical report, PIREN-Seine.
- Herman, J. & Usher, W., 2017. An open-source python library for sensitivity analysis. *The Journal of Open Source Software*, 2(9).
- Lamboni, M., Monod, H. & Makowski, D., 2011. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering and System Safety*, 96 (4), pp. 450 – 459.
- Morris, M., 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics* , 33, pp.161–74.
- Needoba, J.A., Peterson, T.D. & Johnson, K.S., 2012. Chapter Method for the Quantification of Aquatic Primary Production and Net Ecosystem Metabolism Using In Situ Dissolved Oxygen Sensors. In *Molecular Biological Technologies for Ocean Sensing*. pp.73-101.
- Nossent, J., Elsen, P. & Bauwens, W., 2011. Sobol sensitivity analysis of a complex environmental model. *Environmental Modelling&Software*, 26(12), pp.1515-25.
- Odum, H.T., 1956. Primary production in flowing waters. *Limnol. Oceanogr.*.
- Raimonet, M. et al., 2015. Modelling the fate of nitrite in an urbanized river using experimentally obtained nitrifier growth parameters. *Water Research*, 73, pp.373-87.
- Redfield, A.C., Ketchum, B.H. & Richards, F.A., 1963. The influence of organisms on the composition of sea-water. In *The Sea. Ideas and Observations on Progress in the Study of the Seas. The Composition of the Sea-Water Comparative and Descriptive Oceanography*. pp.26-77.
- Saltelli, A., 2002. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145, pp.280-97.
- Saltelli, A. et al., 2010. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2), pp.259-70.
- Servais, P., Billen, G. & Vives-Rego, J., 1985. Rate of bacterial mortality in aquatic environments. *Appl. envir. Microbiol.*, 49(6), pp.1448-54.
- Sobol, I., 1993. Sensitivity estimates for non linear mathematical models. *Mathematical Modelling and Computational Experiments*.
- Vilmin, L., 2014. *Modélisation du fonctionnement biogéochimique de la Seine de l'agglomération parisienne à l'estuaire à différentes échelles temporelles*. PH. D. thesis. Mines ParisTech.
- Vilmin, L. et al., 2015b. Impact of hydro-sedimentary processes on the dynamics of soluble reactive phosphorus in the Seine River. *Biogeochemistry*, 122, pp.229–51.
- Vilmin, L., Flipo, N., de Fouquet, C. & Poulin, M., 2015a. Pluri-annual sediment budget in a navigated river system : the Seine River (France). *Sciences of Total Environment*, 502, pp.48-59.
- Vilmin, L., Flipo, N., Escoffier, N. & Groleau, A., 2016b. Estimation of the water quality of a large urbanized river as defined by the european wfd : what is the optimal sampling frequency ? *Environmental Science and Pollution Research*.
- Vilmin, L. et al., 2016a. Carbon fate in a large temperate human-impacted river system : Focus on benthic dynamics. *in a large temperate human-impacted river system : Focus on benthic dynamics*, 30(7), pp.1086 - 1104.
- Wold, S., Esbensen, K. & Geladi, P., 1987. Principal component analysis. In *Chemometrics and Intelligent Laboratory Systems*. Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists. pp.37 – 52.