



HAL
open science

Simulation du métabolisme de la Seine par assimilation de données en continu

Shuaitao Wang

► **To cite this version:**

Shuaitao Wang. Simulation du métabolisme de la Seine par assimilation de données en continu. Hydrologie. Université Paris sciences et lettres, 2019. Français. NNT : 2019PSLEM029 . tel-02388690

HAL Id: tel-02388690

<https://pastel.hal.science/tel-02388690v1>

Submitted on 2 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à MINES ParisTech

Simulation du métabolisme de la Seine par assimilation de données en continu

Soutenue par

Shuaitao WANG

Le 6 septembre 2019

École doctorale n°398

**Géosciences, Ressources
Naturelles et Environnement**

Spécialité

**Géosciences et géoingé-
nierie**

Composition du jury :

Mme Valérie MONBET Professeur, Université de Rennes 1	<i>Rapporteuse</i>
M. Peter VANROLLEGHEM Professeur, Université Laval	<i>Rapporteur</i>
Mme Josette GARNIER Directeur de Recherche, CNRS	<i>Présidente</i>
M. Jean-Marie MOUCHEL Professeur, Sorbonne Université	<i>Examineur</i>
M. Vincent ROCHER Expert eau et assainissement, SIAAP	<i>Examineur</i>
Mme Sabine SAUVAGE Ingénieur de Recherche, CNRS	<i>Examinatrice</i>
M. Nicolas FLIPO Maître de Recherche, MINES ParisTech	<i>Directeur de thèse</i>
M. Thomas ROMARY Assistant Professeur, MINES ParisTech	<i>Co-directeur de thèse</i>

Résumé

Les teneurs en oxygène dans le système des rivières résultent de nombreux processus physiques et biologiques. Ces teneurs en oxygène traduisent le métabolisme (la production primaire et la respiration) des cours d'eau, et sont utilisées comme un indicateur de qualité de l'eau. Le développement des sondes *in-situ* permet de suivre les concentrations en oxygène dissous en continu dans le système Seine depuis une vingtaine d'années. Ces mesures haute fréquence fournissent l'opportunité de mieux comprendre le fonctionnement écologique du système Seine en appliquant la méthode d'assimilation de données.

Cette thèse a pour but d'implémenter un schéma d'assimilation de données dans le modèle hydro-biogéochimique, PROSE, afin de comprendre le fonctionnement biogéochimique du système Seine soumis à une forte pression anthropique et déterminer l'évolution temporelle des propriétés physiologiques des communautés vivantes. Pour cela, ce travail se découpe en quatre étapes. Une nouvelle version parallélisée de PROSE, PROSE-P, a été développée dans un premier temps en couplant trois librairies (hydraulique, transport et biogéochimique). Dans un deuxième temps, une analyse de sensibilité des paramètres contrôlant l'oxygène du module biogéochimique, C-RIVE, a été réalisée. Cette analyse de sensibilité permet d'identifier les paramètres physiques et physiologiques auxquels la variation de concentration en O_2 est plus sensible. Un filtrage particulière a été ensuite implémenté dans PROSE-P en se basant sur les résultats de l'analyse de sensibilité. Le nouveau modèle PROSE disposant du schéma d'assimilation de données est appelé PROSE-PA. L'efficacité de PROSE-PA est d'abord testé sur un cas synthétique. Enfin, les mesures réelles de l'oxygène de l'année 2011 en Seine sont assimilées par PROSE-PA, afin de déterminer les propriétés physiologiques des micro-organismes.

Dans la première phase de ce travail, les formulations mathématiques

et les schémas numériques de la nouvelle version parallélisée de PROSE, PROSE-P, sont décrits. PROSE-P est ensuite validé par une comparaison des résultats d'une simulation en Seine avec la dernière version de PROSE, PROSE3.6.9. PROSE-P réduit le temps de calcul et montre une résolution plus stable du problème hydraulique ainsi qu'une meilleure représentation des échanges eau-air intégrant l'influence de la température. Ce travail est publié comme un rapport de technique dans la cadre du programme PIREN-SEINE (Wang et al., 2018a).

Dans la deuxième phase de ce travail, une analyse de sensibilité des paramètres du modèle biogéochimique C-RIVE en utilisant la méthode de Morris et les indices de Sobol est réalisée. La méthode la décomposition en fonctions empiriques orthogonales (EOF) est appliquée pour réduire la dimension des sorties du modèle (Fig. 4.1). Les paramètres les plus influents en fonction des contextes hydrologiques et trophiques du système Seine sont identifiés. Les résultats indiquent que la respiration de maintenance du phytoplancton et les paramètres photosynthétiques (i.e. la capacité photosynthétique, le taux de photosynthèse maximal, et les termes d'extinction lumineuse) sont les plus influents pendant les blooms algaux. Quant le système de la rivière devient hétérotrophe, l'oxygène dissous est plus sensible aux paramètres bactériens (moyenne et haute température de l'eau) et aux coefficients de réaération (basse température de l'eau). L'effet de la navigation sur la variation de concentration en O_2 est souligné. L'impact de la température de l'eau sur les résultats d'analyse de sensibilité est également étudié. Finalement, ce travail préliminaire à l'assimilation de données a permis de sélectionner les paramètres les plus influents du modèle. Cette partie est publiée dans le journal "Water Research" (Wang et al., 2018b). Le schéma d'assimilation de données par la suite permet de modifier la valeur de ces paramètres au cours du temps en fonction des teneurs en oxygène mesurées dans le milieu.

Dans la troisième phase de ce travail, un filtre particulaire est implémenté pour la première fois dans un modèle hydro-biogéochimique en se basant sur les résultats d'analyse de sensibilité pour assimiler les mesures en oxygène haute fréquence et estimer les paramètres du modèle. Afin d'évaluer les performances de PROSE-PA, un cas synthétique représentant 66 km de la Seine depuis Suresnes jusqu'à Andrésy est construit. Une observation virtuelle des concentrations en O_2 toutes les 15 minutes est générée à 5 stations

de mesures par PROSE-PA en utilisant les paramètres prédéfinis. L'objectif de cette partie est d'assimiler ces concentrations en O_2 toutes les 15 min, afin de retrouver les paramètres prédéfinis et les vraies concentrations en O_2 . Avec PROSE-PA, les vraies concentrations en O_2 sont parfaitement simulées. L'identifiabilité des paramètres dépend de l'état trophique du système. Les paramètres physiques, bactériens et phytoplanctoniques sont bien identifiés aux périodes où ils dominent la concentration en O_2 , PROSE-PA permet ainsi d'identifier les deux blooms algaux ainsi que le cycle nycthéral des paramètres liés à la photosynthèse. Cela démontre l'efficacité du filtre particulaire pour l'estimation des concentrations en O_2 et des paramètres en biogéochimie. Cette partie est publiée au journal "Water Research".

En fin, PROSE-PA est appliqué sur un cas réel qui décrit la Seine depuis l'amont de Paris jusqu'à l'estuaire pour l'année 2011. Avant d'assimiler les mesures réelles en O_2 , un diagnostic de la simulation des concentrations en O_2 par le modèle déterministe a été réalisé. Ce diagnostic permet d'atténuer les surestimations de l'oxygène dissous soulignées par [Vilmin \(2014\)](#) en spécialisant la matière organique dissoute (MOD) dans la rivière puis en calibrant les activités de la MOD biodégradable. Dès que cette étape est réalisée, PROSE-PA assimile les vraies mesures en O_2 de l'année 2011 à huit stations gérées par le SIAAP. Les résultats montrent une amélioration significative des concentrations en O_2 simulées et une bonne identification des propriétés physiologiques des phytoplanctons pendant les blooms algaux. Pour les paramètres bactériens, une bonne identifiabilité est aussi observée pour les périodes de fortes désoxygénations. Les résultats montrent que PROSE-PA devrait permettre la ré-analyse des données historiques d'oxygène mesurées depuis plus de vingt ans par le réseau MeSeine. De plus, il est possible d'intégrer PROSE-PA dans la gestion du système couplé réseau d'assainissement-station d'épuration (STEP)-milieu récepteur après avoir quantifié les incertitudes liées aux conditions aux limites (spéciation de la matière organique dissoute, déversoirs d'orage) et à la paramétrisation de la matière organique (ajustement des ratios de transformations de la MO détritique).

Abstract

Oxygen concentrations in river systems result from many physical and biological processes. The oxygen concentration reflects the metabolism of river systems and is used as an indicator of water quality. The development of in-situ sensors allows high-frequency monitoring of dissolved oxygen in the Seine River system in recent years. These continuous in-situ measurements offer a new opportunity to improve our understanding of the Seine River metabolism by a data assimilation approach.

The aim of this thesis is to couple a data assimilation scheme with the forward hydro-biogeochemical model, PROSE, in order to understand the biogeochemical functioning of the highly anthropogenized Seine River system and determine the temporal evolution of physiological properties related to micro-organism communities. This work is carried out in four steps. A new parallel version of the PROSE model, PROSE-P, is first developed, coupling three independent libraries (hydrodynamic, transport and biogeochemical). In a second step, the parameter sensitivity of the biogeochemical module, C-RIVE, of PROSE-P is achieved. The sensitivity analysis allows the identification of both physical and physiological parameters governing the variation of the dissolved oxygen in river systems. A particle filtering algorithm is then implemented into PROSE-P based on the results of the sensitivity analysis, the full structure being called PROSE-PA. The performances of the approach are assessed on a synthetic case study. Finally, the real time data assimilation with the PROSE-PA approach on the full 220 km Seine model is performed (from Paris to the entrance of the Seine River estuary).

In the first part of this work, the new parallel version of PROSE, PROSE-P, is described including the mathematical formulations, numerical schemes and parallel strategy. PROSE-P is then validated by comparing the results of a simulation of the full Seine model with those obtained by PROSE3.6.9.

PROSE-P reduces the computational cost and results in a more stable hydraulic resolution and an improvement of exchanges at the air-water interface. This section has been published as a technical report of the PIREN-SEINE research program (Wang et al., 2018a).

In the second part of this work, a time-dependent global sensitivity analysis for dissolved oxygen is performed using Morris and Sobol methods combined with a functional principal component analysis for dimension reduction. The aim of this part is to identify the twelve most important parameters of C-RIVE in different trophic contexts and to understand the biogeochemical functioning of river systems. The results indicate that the maintenance respiration of phytoplankton and the photosynthetic parameters (i.e. photosynthetic capacity, the maximal photosynthesis rate and light extinction coefficients) are the most influential parameters during algal blooms. When the river system becomes heterotrophic, the bacterial activities and the reaeration coefficients affect the most the dissolved oxygen concentration in the water column. The share between those two processes depends on the thermal state of the river system. An anthropogenic effect (ship navigation) on the variation of dissolved oxygen concentration has also been highlighted. The role of this effect evolves with thermal and trophic conditions. This section has been published in “Water Research” (Wang et al., 2018b).

In the third part of this work, the implementation of a particle filtering algorithm into PROSE-P is presented in order to assimilate high frequency dissolved oxygen data and estimate metabolism parameters in the Seine River system. The procedure is designed based on the results of the former sensitivity analysis of the model. The performances of the PROSE-PA approach are assessed on a synthetic case study that mimics 66 km of the Seine River. Synthetic dissolved oxygen data are generated using time varying parameters. This part aims at retrieving the predefined parameters by assimilating the synthetic observation data. The simulated dissolved oxygen concentrations coincide perfectly with the true values. The identification of the parameters depends on the trophic context and more surprisingly on the thermal state of the river system. The physical, bacterial and phytoplanktonic parameters can be retrieved properly, leading to the differentiation of two successive algal blooms by comparing the estimated posterior distribution of the optimal temperature for growth of phytoplankton. The photosynthetic parameters’ distributions following circadian cycles during algal blooms are

also discussed. This section has been published in “Water Research” ([Wang et al., 2019](#)).

Finally, the real time data assimilation with PROSE-PA approach on the full 220 km Seine model is in progress for the year 2011. The application of the assimilation scheme on the real dataset leads to a refinement of organic matter speciation between biodegradable and refractory carbon in the upstream inputs by the Seine, Marne and Oise Rivers. Results of this work reveal that data assimilation with a particle filter opens new opportunities in the development of tools for real-time urban water management. To this aim, it remains necessary to quantify input uncertainties (forcing data, outflows of WWTPs (Waste Water Treatment plants) and CSOs (Combined Sewer Overflows) etc.).

Remerciements

Je tiens à remercier en premier Nicolas Flipo, mon directeur de thèse, qui m’a encadré tout au long de ma thèse et qui a placé sa confiance en moi depuis le début de cette thèse. Je le remercie également pour ses conseils toujours pertinents et éclairés, ses encouragements. Je remercie Thomas Romary, mon co-directeur de thèse, qui m’a aidé à comprendre les connaissances en statistiques qui sont indispensables pour réussir cette thèse. Merci pour ses conseils pertinents, ses réponses efficaces et sa bonne humeur. Grâce à eux, la thèse s’est très bien déroulée. Je suis très content d’avoir passé trois ans avec eux.

Je remercie à tous les membres de jury pour les relectures, les commentaires ainsi que les discussions riches, les conseils pertinents lors de la soutenance qui améliorent fortement mon mémoire de thèse. Merci à Valérie Monbet et Peter Vanrolleghem, les deux rapporteurs, Josette Garnier, la présidente du jury, ainsi que Sabine Sauvage, Jean-Marie Mouchel et Vincent Rocher, les examinateurs.

Je tiens également à remercier les membres du comité de suivi de la thèse des deux premières années, Vincent Thieu, Jean-Marie Mouchel, Lauriane Vilmin et Vincent Rocher qui ont donné les conseils pour bien avancer la thèse.

Au niveau de la modélisation, ce travail n’aurait pas été possible sans la grande quantité et qualité des données d’entrée du modèle disponibles. Je remercie donc Lauriane Vilmin qui les a construits dans sa thèse. Merci aussi pour ses réponses toujours efficaces.

Cette thèse a été réalisée au sein de l’équipe Systèmes Hydrologiques et Réservoirs en collaborant avec l’équipe Géostatistique du Centre de Géosciences. Je remercie Patrick Goblet qui m’a envoyé la proposition de thèse et m’a accueilli le premier jour au laboratoire en me disant “nihao”. Merci aux membres de l’équipe que je croise quotidiennement : Elisabeth Cordier,

Agnès Rivière, Sophie Guillon, Pascal Viennot, Nicolas Gallois, Dominique Bruel, Heena Shabir, Alexandra Mattei, Asma Berrhouma, Fulvia Baratelli, Aurélien Bordet, Deniz Kilic, Jonathan Schuite, Mathias Maillot, Jacques Denizer. Merci pour “PROSE *for life TM*” que j’ai beaucoup aimé.

Je remercie mes collègues, amis français et chinois que j’ai rencontrés à Fontainebleau pour les week-ends agréables, les soirées conviviales, les balades dans la forêt.

Enfin, je tiens à remercier ma famille. Merci à mes parents, mes frères et ma soeur qui m’ont beaucoup soutenu. Merci à ma tante et mon oncle, grâce à qui j’ai pu venir étudier en France.

Merci à toutes et à tous!!!

Table des matières

Résumé	1
Abstract	5
Remerciements	9
1 Introduction	25
1.1 Modélisation de la qualité de l'eau en rivière	25
1.2 Les mesures haute fréquence	26
1.3 Assimilation de données	27
1.4 La Seine depuis l'amont de Paris jusqu'à l'estuaire	27
1.4.1 Évolution de qualité de l'eau de la Seine : l'Oxygène Dissous	30
1.4.2 Mesures haute fréquence de l'oxygène dissous de l'an- née 2011	32
1.5 Objectifs de la thèse	33
1.6 Cadre de la thèse : Le programme PIREN-SEINE	34
1.7 Structure du mémoire	35
I Modélisation de la qualité de l'eau en rivière	39
2 Assimilation de données et modèles de qualité de l'eau	41
2.1 Méthodes d'assimilation de données	42
2.1.1 Notations	43
2.1.2 Méthodes variationnelles	43
2.1.3 Filtre de Kalman	44
2.1.4 Le filtre de Kalman d'ensemble (EnKF)	45
2.1.5 Méthodes de Monte Carlo séquentielles	47

2.1.6	Applications	47
2.2	Revue des modèles de la qualité de l'eau en rivière	48
2.3	Autre logiciels	53
2.4	Conclusion	54
3	Mise en place d'une nouvelle version de ProSe : ProSe-P	57
3.1	Structure du code	57
3.2	Librairie hydraulique <i>libhyd</i>	58
3.2.1	Discrétisation de l'équation de conservation de la masse	59
3.2.2	Discrétisation de l'équation de conservation de la quantité de mouvement	60
3.2.3	Loi de fonctionnement d'un ouvrage hydraulique	62
3.3	Librairie de transport <i>libttc</i>	64
3.3.1	Discrétisation numérique	65
3.3.2	La réoxygénation aux barrages	65
3.4	Librairie biogéochimique <i>librive</i> (C-RIVE)	67
3.4.1	Reformulation à base physique du coefficient de réaération	67
3.5	Parallélisation des calculs de transport et de biogéochimie	69
3.6	Performances de PROSE-P	72
3.6.1	Cas test de barrage	72
3.6.2	Une simulation du modèle Seine	72
II	Etude de sensibilité des paramètres contrôlant l'oxygène	81
4	Etude de sensibilité du modèle biogéochimique C-RIVE	83
4.1	Introduction	85
4.2	Material and methods	88
4.2.1	Biogeochemical model C-RIVE	88
4.2.2	Mathematical modeling of C-RIVE	90
4.2.3	Case study	93
4.2.4	Parameters considered in the sensitivity analysis	95
4.2.5	Sensitivity analysis methodology	104
4.2.6	Simulation setting and computational cost	108
4.3	Results	109

4.3.1	Results of the Morris method over the low flow periods	109
4.3.2	Results of the Sobol' sensitivity analysis over the low flow period	111
4.3.3	Sensitivity analysis results during high flow period . . .	113
4.3.4	Selection of the most influential parameters of the C-RIVE model	114
4.4	Discussion	115
4.4.1	Parameters' behavior during an algal bloom	115
4.4.2	Parameters' behavior out of a algal bloom	116
4.4.3	Impact of the temperature on SA	117
4.4.4	Recommendations for water quality modeling	117
4.5	Conclusions	118
4.6	Informations Supplémentaires	121
4.6.1	Convergence de la méthode de Morris	121
4.6.2	L'effet de la température de l'eau sur le scénario Bloom	121
4.6.3	Sensibilité du modèle à la limitation de la croissance bactérienne	124
4.6.4	L'effet de la température sur la sensibilité à K_s	126
 III Assimilation de données d'oxygène par filtrage particulaire : ProSe-PA		127
 5 Preuve de concept sur un cas synthétique		129
5.1	Introduction	131
5.2	Material and methods	133
5.2.1	Synthetic case study	133
5.2.2	Data assimilation framework using particle filtering, the PROSE-PA approach	134
5.2.3	Description of the synthetic case study	140
5.2.4	Numerical setting and computational cost	145
5.3	Results : Oxygen simulation and parameter identification . . .	145
5.3.1	Effective sample size for resampling procedure	145
5.3.2	Almost perfect DO simulation with PROSE-PA	145
5.3.3	Parameter identification in the frame of the state classification	147
5.4	Discussion	152

5.4.1	Specifications for real time parameter identification . . .	152
5.4.2	Possibility to identify different phytoplanktonic communities	155
5.4.3	Circadian rhythm	156
5.4.4	Performances' sensitivity to the observation error . . .	157
5.5	Conclusions	157
5.6	Supplementary information	159
5.6.1	Importance sampling and definition of the importance weights in particle filtering	159
5.6.2	PROSE-PA <i>vs</i> Filtre particulaire standard	162
5.6.3	Test du nombre de particules	163
5.6.4	Concentrations en oxygène dissous simulées aux stations de validations	166
6	La Seine depuis l'amont de Paris jusqu'à l'estuaire	169
6.1	Diagnostic de la simulation des concentrations en O ₂ pour l'année 2011	170
6.1.1	Reproduction des concentrations en O ₂ par une simulation directe	170
6.1.2	Profil en long des matières organiques biodégradables (MOD1 et MOD2)	173
6.2	Spéciation de la matière organique dissoute à l'amont du domaine	173
6.2.1	Déficit en MOD biodégradable	175
6.2.2	Réactivités des MODs biodégradables	176
6.3	Performances de PROSE-PA pour l'année 2011	177
6.3.1	Stratégie de modélisation	178
6.3.2	La taille effective pour le ré-échantillonnage lors de la simulation	179
6.3.3	Concentrations en O ₂ simulées par PROSE-PA	180
6.3.4	Identifiabilité des propriétés physiologiques	184
6.3.5	Difficultés d'identification des paramètres	197
6.3.6	Recommandations pour la perturbation après le ré-échantillonnage	199
6.4	Conclusions	199

Conclusion et perspectives	203
A Résultats de la simulation d'une espèce de bactérie hétérotrophe avec la stratégie non bouclée	207
B Résultats de la simulation de deux espèces de bactérie hétérotrophe avec la stratégie bouclée	213
C Guide d'utilisation : ProSe-PA 0.30	219
C.1 Get Started	220
C.2 Définition des dossiers contenant les fichiers d'entrée du modèle	222
C.3 Paramétrisation générale	225
C.4 Simulation hydraulique	226
C.4.1 Paramétrisation pour hydraulique	226
C.4.2 Fichier d'initialisation de l'hydraulique	227
C.4.3 Description du réseau hydrographique	228
C.4.4 Trajectoire du centre des faces	230
C.4.5 Apports	231
C.4.6 Prise en compte de la courbure	232
C.5 Simulation de transport et biogéochimie	232
C.5.1 Les fichiers de paramètres des espèces	232
C.5.2 Paramétrisation pour transport	241
C.5.3 Paramétrisation pour biogéochimie	242
C.5.4 Les fichiers de description des compartiments	242
C.5.5 Les paramètres liés à la sédimentation-érosion	244
C.5.6 Les forçages météorologiques	244
C.5.7 Définition des intervalle des paramètres à assimiler	245
C.6 Définition des fichier de sorties	246
C.6.1 Sorties PKs, final states	249
C.6.2 Profils ponctuels	250
C.6.3 Profils en long hydraulique	251
C.6.4 Bilans hydraulique	251
C.6.5 Sorties graphiques hydraulique	251
C.6.6 Bilans biogéochimiques	252
C.7 Fichier de commande pour cas test barrage	253

Liste des tableaux

1.1	Station de mesure en oxygène dissous de l'année 2011 du SIAAP	33
2.1	Description des modèles de qualité de l'eau	55
3.1	Performance de calcul en parallèle	73
3.2	Performance de calcul des différentes versions	74
3.3	Répartition des temps de calcul des différentes parties de PROSE-P	74
3.4	Critères statistiques des débits simulés par PROSE-P et PROSE-PA3.6.9	76
3.5	Critères statistiques des concentrations en O ₂ simulées par les deux codes	78
4.1	Sensitivity analysis studies in hydrological and water quality modeling	87
4.2	Initial conditions of species simulated for the case study	95
4.3	The initial conditions of DOM, algal and bacterial biomass in different trophic and hydrological contexts	96
4.4	Parameter list and variation ranges	97
4.5	First order and total sensitivity index with 95% confidence interval (Pre-bloom)	119
4.6	First order and total sensitivity index with 95% confidence interval (Bloom)	120
4.7	First order and total sensitivity index with 95% confidence interval (Post-bloom)	120
5.1	True parameters considered in data assimilation	143
5.2	Statistic errors of the weighted average and the true state	147

5.3	Kling-Gupta Efficiency (KGE) at five monitoring sites for different observation errors	158
5.4	Thresholds of the relative observation error for the parameter identifiability	158
6.1	Répartition des différents types de matière organique dissoute	174
6.2	Déficit en MOD biodégradable pour la période 250-300 jours	175
6.3	Répartition de MOD3 en MOD1 et MOD2	177
6.4	Les valeurs des paramètres liés aux réactivités de la MOD biodégradable pour le balayage	177
6.5	Simulations correspondant à la RMSE minimale de chaque station entre les jours 250 et 300 en fonction des différentes paramétrisations pour calibrer les réactivités de la MOD biodégradable	178
6.6	Répartition des différents types de matière organique dissoute après la calibration	178
6.7	Paramètres numériques du filtre particulaire pour le cas réel .	178
6.8	Critères statistiques des concentrations en O_2 simulées aux stations de mesure du SIAAP par PROSE3.6.9 et PROSE-PA	183
6.9	Comparaison des moyennes temporelles des modes des paramètres identifiées par PROSE-PA avec celles utilisées dans la littérature	191
6.10	Critère statistique des concentrations en O_2 simulées aux stations de mesures du SIAAP pour deux stratégies de perturbations	200
C.1	Éléments à définir dans le fichier de commande	226
C.2	Éléments à définir dans le fichier de commande	227
C.3	Codes des espèces biogéochimique simulées	233
C.4	Éléments à définir dans le fichier de commande	241
C.5	Éléments à définir dans le fichier de commande	242

Table des figures

1.1	Domaine simulé	28
1.2	Débits moyens journaliers observés à la station de jaugeage d'Austerlitz entre le 1 janvier	29
1.3	Évolution de qualité de l'eau de la Seine à la station Poses entre 1971 et 2017	31
1.4	Mesures en oxygène dissous de l'année 2011 à Bougival	32
2.1	Schéma simplifié d'assimilation de données	42
2.2	Représentation schématique des différents étapes du filtre de Kalman d'ensemble	46
3.1	Schéma simplifié du logiciel PROSE	58
3.2	Les indices du schéma numérique	59
3.3	Représentation d'un seuil	62
3.4	Représentation d'une vanne	64
3.5	Schéma synthétique de C-RIVE	68
3.6	Schéma synthétique de calcul en parallèle dans PROSE-P	70
3.7	Performance théorique de la parallélisation	71
3.8	Comparaison d'un cas test de barrage en régime permanent	73
3.9	Les débits simulés par PROSE-P et PROSE3.6.9 pour l'année 2007-1	76
3.10	Les débits simulés par PROSE-P et PROSE3.6.9 pour l'année 2007-2	77
3.11	Le débit simulé des premiers 30 jours à Suresnes et Sartrou- ville pour l'année 2007	77
3.12	Les concentrations de l'oxygène dissous simulées par PROSE- P et PROSE pour l'année 2007	79

3.13	Les concentrations de l'oxygène dissous simulées par PROSE-P et PROSE pour l'année 2007	80
3.14	Les concentrations de l'oxygène dissous simulées pendant l'étiage à Chatou et Méricourt	80
4.1	Méthodologie de l'analyse de sensibilité du module biogéochimique C-RIVE	85
4.2	Schematic description of C-RIVE model	89
4.3	Synthetic schema of case study	93
4.4	Profile of algal and bacteria biomass simulated for the period 2007-2012	94
4.5	Sensitivity analysis methodology	104
4.6	Results of Morris method for low flow conditions	110
4.7	Results of Sobol method for low flow periods	112
4.8	Results of Morris and Sobol methods during high flow period	114
4.9	The evolution of the total sensitivity indices with temperature for high flow condition	118
4.10	Satabilité de la méthode de Morris pour le scénario Bloom	121
4.11	L'évolution des indices de Morris en fonction de la température de l'eau pour le scénario Bloom	122
4.12	L'évolution des indices totaux en fonction de la température de l'eau pour le scénario Bloom	123
4.13	Analyse de sensibilité basée sur la méthode de Morris intégrant la limitation de croissance bactérienne (K_s) pour le cas Post-bloom	124
4.14	Les indices de Sobol pour le scénario Post-bloom avec K_s	125
4.15	L'évolution des indices totaux de Sobol en fonction de la température pour le scénario Post-bloom avec K_s	126
5.1	The description of study area and the monitoring sites	135
5.2	The flowchart of PROSE-PA approach	141
5.3	State variable classification of synthetic case study over time	144
5.4	The evolution of effective sample size during assimilation	146
5.5	The evolution of effective sample size for days 183-185	146
5.6	Simulated DO concentrations at Bougival, Sartrouville and Andresy stations	148
5.7	Daily normalised weights of physical parameters	149

5.8	The normalised importance weights of bacterial parameters	150
5.9	The normalised importance weights of phytoplanktonic parameters	151
5.10	The posterior pdf of K_{navig} during low temperature period	153
5.11	The identification of parameters in different trophic and thermal contexts of river system	154
5.12	The posterior distributions of $T_{opt,pp}$ during the first and the second algal bloom	155
5.13	The posterior distributions of $P_{max,pp}$ during 184-186 days	156
5.14	Simulated DO concentrations at Bougival, Sartrouville and Andresy stations with 100 particles	163
5.15	Simulated DO concentrations at Bougival, Sartrouville and Andresy stations with 300 particles	164
5.16	Simulated DO concentrations at Bougival, Sartrouville and Andresy stations with 400 particles	165
5.17	Simulated DO concentrations at Colombes and Conflans-Sainte-Honorine stations with 100 particles	166
5.18	Simulated DO concentrations at Colombes and Conflans-Sainte-Honorine stations with 300 particles	167
5.19	Simulated DO concentrations at Colombes and Conflans-Sainte-Honorine stations with 500 particles	167
6.1	Chroniques de concentrations en O_2 observées et simulées avec le modèle direct aux stations SIAAP-1	171
6.2	Chroniques de concentrations en O_2 observées et simulées avec le modèle direct aux stations SIAAP-2	172
6.3	Profil longitudinal de MOD1 et MOD2 entre Suresnes et Méricourt pour le jour 264	174
6.4	L'évolution de la taille effective pour le ré-échantillonnage	179
6.5	L'évolution de la taille effective pour le ré-échantillonnage pour les jours 196-200	180
6.6	Concentrations en O_2 observées et simulées par PROSE-PA-1 aux stations de SIAAP en 2011	181
6.7	Concentrations en O_2 observées et simulées par PROSE-PA-1 aux stations de SIAAP	182
6.8	Classification de l'état du système en fonction de la biomasse algale et de la température à la station Bougival	185

6.9	L'évolution des poids normalisés pour les paramètres phyto- planctoniques	186
6.10	L'évolution de la distribution a posteriori de la température optimale pour le bloom de mars	187
6.11	L'évolution de la distribution a posteriori de la respiration de maintenance pour le bloom de mars	188
6.12	L'évolution de la distribution a posteriori du taux de photo- synthèse maximal le bloom de mars	189
6.13	L'évolution de la distribution a posteriori de la capacité pho- tosynthétique pour le bloom de mars	190
6.14	L'évolution des poids normalisés pour les paramètres physiques	192
6.15	L'évolution de la distribution a posteriori du coefficient de ré- aération liée à la navigation	193
6.16	L'évolution des poids normalisés pour les paramètres bactériens	194
6.17	L'évolution des distributions a posteriori des paramètres bac- tériens avant le bloom de mars	195
6.18	L'évolution des distributions a posteriori des paramètres bac- tériens entre les blooms	196
6.19	Corrélations entre les paramètres, les concentrations d'O ₂ à Meulan et les conditions aux limites pour les jours 240 à 290	198
A.1	Concentrations en O ₂ observées et simulées par PROSE-PA aux stations du SIAAP avec la stratégie du domaine des para- mètres non bouclé	208
A.2	Concentrations en O ₂ observées et simulées par PROSE-PA aux stations du SIAAP avec la stratégie du domaine des para- mètres non bouclé	209
A.3	L'évolution des poids normalisés pour les paramètres phyto- planctoniques avec la stratégie du domaine non bouclé	210
A.4	L'évolution des poids normalisés pour les paramètres bacté- riens avec la stratégie du domaine non bouclé	210
A.5	L'évolution des poids normalisés pour les paramètres phy- siques avec la stratégie du domaine non bouclé	211
B.1	Concentrations en O ₂ observées et simulées par PROSE-PA aux stations du SIAAP avec la stratégie du domaine des para- mètres bouclé et deux espèces bactériennes	214

B.2	Concentrations en O_2 observées et simulées par PROSE-PA aux stations du SIAAP avec la stratégie du domaine des paramètres bouclé et deux espèces bactériennes	215
B.3	L'évolution des poids normalisés pour les paramètres phytoplanktoniques avec la stratégie du domaine bouclé et deux espèces bactériennes	216
B.4	L'évolution des poids normalisés pour les paramètres bactériens avec la stratégie du domaine bouclé et deux espèces bactériennes	216
B.5	L'évolution des poids normalisés pour les paramètres physiques avec la stratégie du domaine bouclé et deux espèces bactériennes	217

Chapitre 1

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., 2018). Le métabolisme des fleuves dépend de la production d'oxygène par les producteurs primaires autotrophes, de la respiration bactérienne hétérotrophe, ainsi que des processus physiques de réaération à l'interface air-eau et des interactions à l'interface eau-sédiment (Odum, 1956; Vilmin et al., 2016).

1.1 Modélisation de la qualité de l'eau en rivière

La modélisation de la qualité de l'eau a commencé initialement par [Streeter and Phelps \(1925\)](#) qui ont étudié l'effet d'un rejet de matière organique sur les milieux récepteurs avec le modèle conceptuel de Streeter-Phelps via le terme de DBO (Demande biochimique en oxygène). Dans la seconde partie du XXe siècle, les processus biogéochimiques détaillés (cycle du carbone, des nutriments et diffusion etc.) ont été intégrés dans les modèles actuels grâce au River Continuum concept.

Le River Continuum concept développé par [Vannote et al. \(1980\)](#) correspond à un changement de paradigme pour l'hydro-écologie en introduisant le concept d'invariance de physiologie des communautés d'organismes d'amont en aval des cours d'eau ; les structures observées ne dépendant que du champ de contraintes hydro-morphologiques et anthropiques auquel le cours d'eau est soumis. Ce paradigme a permis de développer les modèles biogéochimiques actuels basés sur la notion de communautés d'espèces, qui

deviennent des variables des modèles dont le devenir est simulé au fil de l'eau. Ainsi les producteurs primaires et les bactéries hétérotrophes, sont-ils aujourd'hui explicitement simulés.

Cependant, de nombreux processus physiques nécessitent un grand nombre de paramètres caractéristiques (Flipo, 2013). Certains paramètres peuvent être estimés *in situ* ou en laboratoire, puis utilisées pour la paramétrisation des modèles (Garnier et al., 1995; Flipo et al., 2007a; Raimonet et al., 2015). 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 and Billen, 1993; Garnier et al., 1995, 1998). Ces paramètres physiologiques sont aussi importants que les mesures 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., 2007b; Flipo et al., 2007a; Escoffier et al., 2018; Vilmin et al., 2016). Du fait des complexités hydrologiques et biogéochimiques, les modèles de qualité de l'eau sont difficiles à calibrer et à valider (Arhonditsis and Brett, 2004). De plus, il est possible que différentes paramétrisations du modèle produisent des résultats similaires, ceci est référencé comme un problème d'équifinalité (Duan et al., 1992; Beven, 1989; Polus et al., 2011).

1.2 Les mesures haute fréquence

Depuis les années 1990, l'apparition de capteurs *in situ* en continu et à pas de temps fin (toutes les 15 minutes) permet de suivre très précisément l'évolution des teneurs en oxygène dans la Seine comme le réseau MeSeine du SIAAP (*Syndicat Intercommunal de l'Assainissement de l'Agglomération Parisienne*) et le projet CARBOSEINE (Groleau et al., 2012, 2013). Ces mesures haute fréquence montrent un cycle saisonnier avec de basses concentrations de l'oxygène dissous (sous-saturation) en été. Cette variation saisonnière est liée dans un premier temps à la solubilité de l'oxygène et dans un deuxième temps à l'activité des micro-organismes en fonction de la tem-

pérature (Wang et al., 2018b). Les suivis historiques des concentrations en oxygène dissous montrent aussi une forte amélioration de la qualité de l'eau depuis une dizaine d'années où le niveau de l'oxygène dissous ne descend pas au dessous de 4 mgO₂/L (Garnier et al., 2019). Cette amélioration est fortement due à l'implémentation de nouvelles technologies de traitement dans les stations d'épuration et à la mise en place de la DCE (Directive Cadre Européenne, (Parliament Council of the European Union, 2000)).

Grâce aux observations haute fréquence à l'échelle pluri-annuelle, le métabolisme de l'écosystème (la production primaire et la respiration) peut être estimé comme dans le cas de la Seine (Escoffier et al., 2018), le Rhône (Guillon et al., 2019) ou d'autres systèmes (Odum, 1956; Needoba et al., 2012). De plus, il devient possible de calibrer et simuler le fonctionnement biogéochimique de la Seine à l'échelle pluri-annuelle avec les mesures haute fréquence (Vilmin, 2014; Vilmin et al., 2016).

1.3 Assimilation de données

Cependant les modèles, aussi sophistiqués soient-ils, peinent à reproduire les teneurs en oxygène notamment en étiage (Vilmin, 2014), car leur paramétrisation est fortement incertaine dans le cas d'un transfert des propriétés physiologiques de communautés vivantes déterminées dans les conditions contrôlées du laboratoire au milieu naturel. De plus, ces propriétés physiologiques moyennes sont actuellement considérées constantes dans le temps par les modélisations déterministes. 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 et de corriger les variables d'état des modèles (concentrations) et leurs paramètres afin de mieux déterminer l'évolution des teneurs en oxygène du milieu et d'estimer les propriétés physiologiques temporelles des micro-organismes.

1.4 La Seine depuis l'amont de Paris jusqu'à l'estuaire

La Seine prend sa source au plateau de Langres (Champagne-Ardenne) à 471 mètre d'altitude et se jette dans la Manche au Havre, après s'être

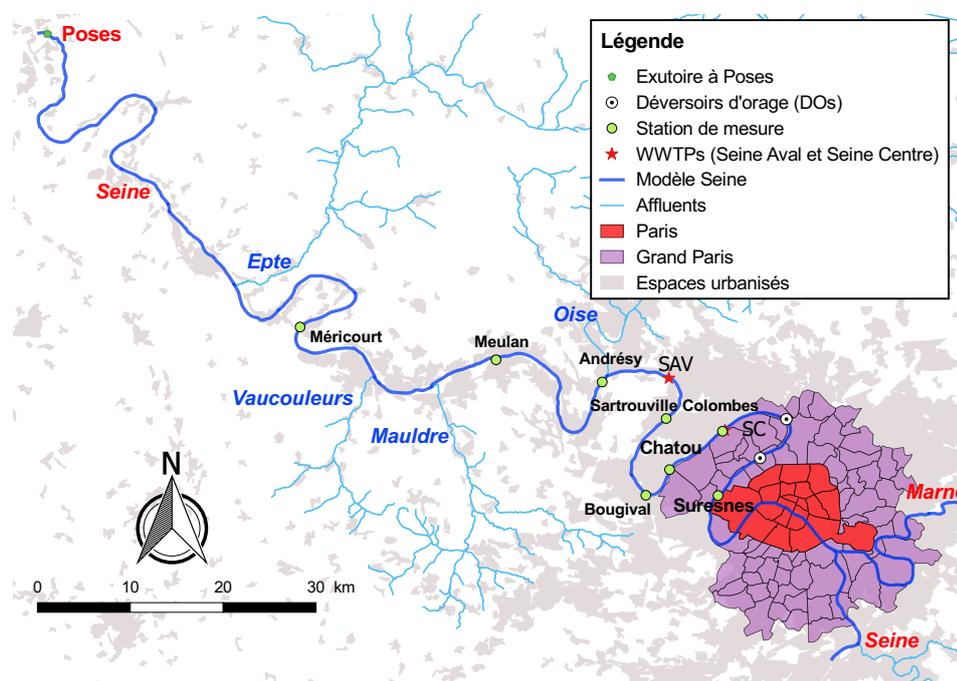


FIGURE 1.1 – Domaine simulé

écoulée sur 776 km. Son exutoire se trouve à Poses, à 166 km en amont du Havre. A partir de Poses, la marée commence à influencer le fonctionnement hydrodynamique. Dans cette thèse, on s'intéresse à une zone soumise à de fortes pressions anthropiques : la seine depuis l'amont de l'agglomération parisienne jusqu'à Poses (Fig. 1.1). Dans cette zone, les affluents Oise, Mauldre, Vaucouleurs, Epte ainsi qu'un tronçon de 25 km de la Marne sont inclus dans le modèle Seine. De plus, deux stations d'épuration (STEP), de Seine Aval et Seine Centre, se situent à l'aval de l'agglomération parisienne. Le modèle Seine mesure une longueur d'environ 220 km de long.

Les principaux affluents ont un régime hydrologique similaire à celui de la Seine : le régime pluvial océanique. Les années hydrologiques démarrent en septembre avec les débits minimums en été ($66.2 \text{ m}^3/\text{s}$ en 2011 à Austerlitz) et les débits maximums en hiver ($1020 \text{ m}^3/\text{s}$ en 2011 à Austerlitz, Fig. 1.2). L'année 2011 est caractérisée par une année quinquennale sèche avec un VCN30 (débit moyen minimal sur 30 jours consécutifs) de $84 \text{ m}^3/\text{s}$ pour la période entre le 16 juin 2011 et le 15 juillet 2011. Pendant l'étiage, le faible débit et la haute température de l'eau peuvent causer des nuisances

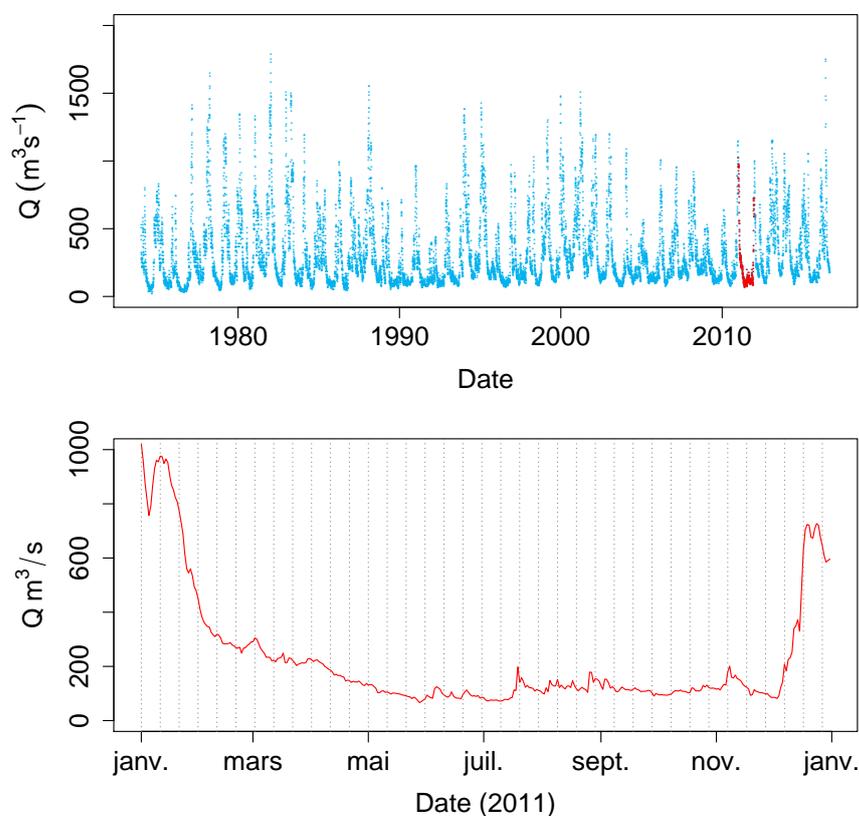


FIGURE 1.2 – Débits moyens journaliers observés à la station de jaugeage d'Austerlitz entre le 1 janvier 1974 et le 4 septembre 2016. Les points rouges correspondent à l'année 2011

de qualité de l'eau. Il y a eu trois blooms algaux dans la Seine en 2011. Cette année a été ainsi choisie pour ce travail.

L'oxygène dissous est fortement impacté par les rejets anthropiques : les rejets des STEP de l'agglomération parisienne gérées par le SIAAP, les déversements de temps de pluie (Déversoirs d'Orage, DOs) et les prélèvements et rejets de temps sec (Vilmin, 2014). Les DOs pendant les grands événements de pluies peuvent durer de une heure à vingt heures et engendrer une chute rapide de la concentration en oxygène. Ces chutes rapides d' O_2 peuvent dépasser plus de $2 \text{ mgO}_2/\text{L}$ (Vilmin, 2014). Une description de l'évolution des teneurs en oxygène en Seine est présentée dans la suite.

1.4.1 Évolution de qualité de l'eau de la Seine : l'Oxygène Dissous

Cette partie s'inspire d'un chapitre d'ouvrage sur le fonctionnement écologique de la Seine ([Garnier et al., 2019](#)), ainsi que de l'ouvrage de synthèse de l'évolution de la qualité au regard des phases d'aménagements urbains ([Rocher and Azimi, 2017](#)). L'oxygène dissous est utilisé depuis longtemps comme un indicateur de la qualité de l'eau des rivières. Depuis le XIX^e siècle, la population du bassin de la Seine a beaucoup augmenté, jusqu'à 17 millions en 2018. Le système de collecte des eaux usées dans l'agglomération parisienne commence au début du XX^e siècle. A la fin des années 1960, environ 60% des eaux usées collectées ont été rejetées directement dans la Seine. La qualité des eaux de la Seine étaient ainsi fortement dégradée à l'époque. Les teneurs en oxygène diminuent rapidement, d'amont en aval du fleuve, qui est en anoxie à l'aval des points de rejets ([Garnier et al., 2019](#)). La mise en place des stations d'épuration dans le bassin de la Seine commence en 1940 : la première station d'épuration à Achères (STEP, Seine Aval aujourd'hui). Ensuite, de nouvelles usines de traitement des eaux usées sont construites pour augmenter la capacité de traitement. Cette nouvelle phase de déploiement du système d'assainissement francilien a duré jusqu'aux années 1990 ([Rocher and Azimi, 2017](#)). Depuis une dizaine d'années, de nouvelles technologies de traitement des eaux usées sont mises en œuvre pour éliminer non seulement les rejets de matière organique mais aussi de nutriments ([Garnier et al., 2019](#)). L'évolution du système d'assainissement de la Seine a permis la reconquête du milieu en maintenant la concentration en oxygène dissous à des niveaux acceptables ($> 4 \text{ mgO}_2/\text{L}$) depuis une dizaine d'années (Fig. 1.3), ce qui permet le maintien des espèces piscicoles ([Warren et al., 1973](#)).

Le métabolisme de la Seine a évolué depuis les années 1990. L'évolution du système d'assainissement dans le bassin de la Seine a eu peu d'influence sur la production primaire, en revanche les implémentations des stations d'épuration diminuent la respiration de l'écosystème, notamment la respiration hétérotrophe ([Garnier et al., 2019](#)). Dans les années 1970, le système Seine est caractérisé comme étant hétérotrophe (respiration $>$ production) du fait des rejets de matière organique non traitée dans la Seine.

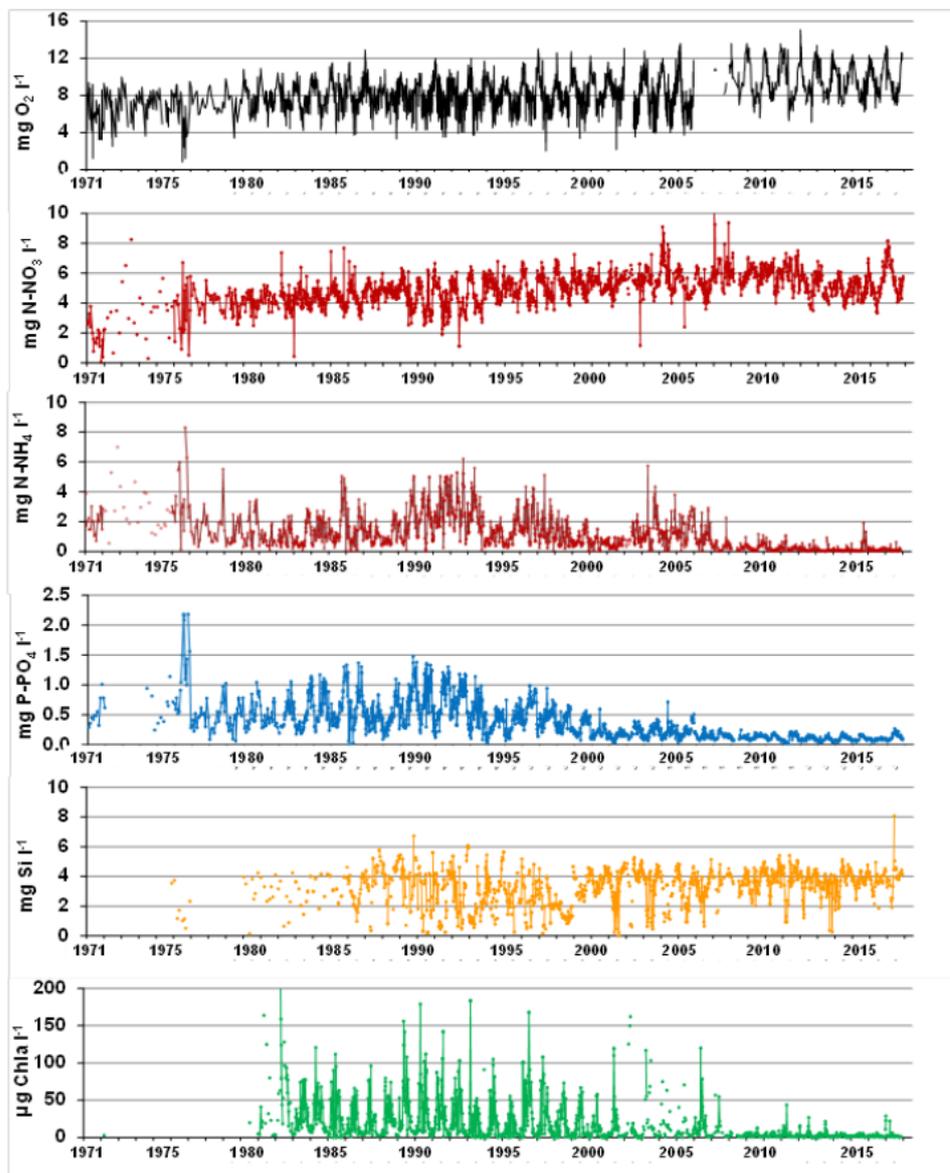


FIGURE 1.3 – Évolution de qualité de l'eau de la Seine à la station Poses entre 1971 et 2017 : l'oxygène dissous, ammonium, phosphates, silice dissous, biomasse phytoplanktonique en *chl a* (Garnier et al., 2019).

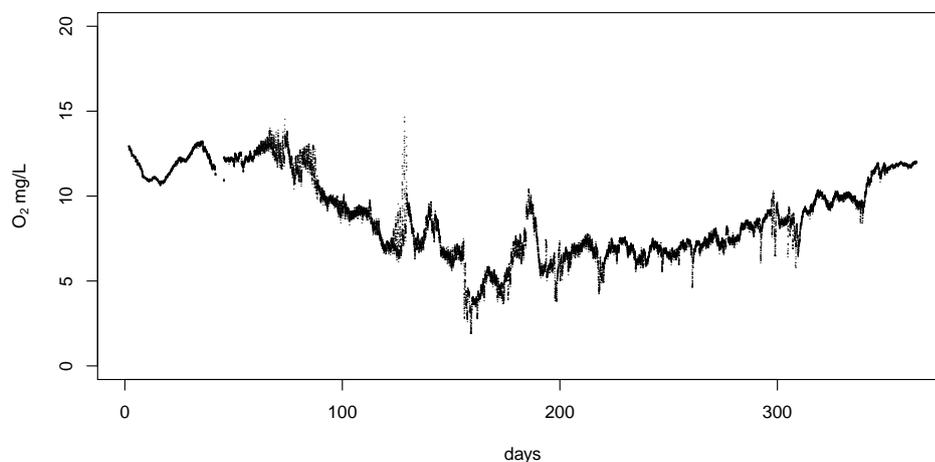


FIGURE 1.4 – Mesures en oxygène dissous de l'année 2011 à Bougival

1.4.2 Mesures haute fréquence de l'oxygène dissous de l'année 2011

L'ancien réseau d'auto-surveillance du SIAAP, socle du réseau actuel MeSeine, permet de mesurer les teneurs en oxygène en continu (toutes les 15 min). Les mesures haute fréquence peuvent être utilisées pour quantifier le métabolisme (la production et la respiration) du système (Needoba et al., 2012; Escoffier et al., 2018) et pour calibrer les modèles de qualité de l'eau (Billen et al., 2001; Vilmin et al., 2016). Cette quantification du métabolisme du système permet ensuite de comprendre le fonctionnement écologique du milieu.

Dans ce travail, les données en oxygène dissous de l'année 2011 à 8 stations de mesures gérées par le SIAAP sont utilisées (Tab. 1.1, Vilmin (2014)). Les 8 stations de mesure se situent à l'aval de l'agglomération parisienne et en aval des principaux rejets (Fig. 1.1). Ces données en continu sont assimilées dans le chapitre 6 (p.169) pour évaluer les performances du schéma d'assimilation présenté dans le chapitre 5.

L'objectif de cette partie n'était pas de décrire précisément le système Seine, mais de donner les éléments utiles à la lecture des chapitres suivants, à savoir que l'année 2011 est relativement sèche, présentant trois blooms alguaux. Pour une présentation plus détaillée, le lecteur curieux pourra se référer au chapitre 1 du travail de doctorat de L. Vilmin (Vilmin, 2014), ou

TABLE 1.1 – Station de mesure en oxygène dissous de l’année 2011 du SIAAP (Vilmin, 2014)

Cours d’eau	Station	Nombre de mesure	Fréquence de mesure
Seine	Suresnes	30948	quart-horaire
Seine	Colombes	15994	quart-horaire
Seine	Chatou (RD)	29626	quart-horaire
Seine	Bougival (RG)	33971	quart-horaire
Seine	Sartrouville	14912	quart-horaire
Seine	Andrésey (RG)	30586	quart-horaire
Seine	Meulan-en-Yvelines	16738	quart-horaire
Seine	Méricourt	32680	quart-horaire

La différence du nombre de mesures est liée au dysfonctionnement des sondes

encore aux ouvrages anniversaire du PIREN Seine (Meybeck et al., 1998; Billen et al., 2007; Garnier et al., 2019).

1.5 Objectifs de la thèse

L’objet de la thèse est de construire une version parallélisée du logiciel PROSE pour assimiler en continu les mesures d’oxygène dissous haute fréquence recueillies par le SIAAP afin de notamment déterminer l’évolution temporelle des propriétés physiologiques des communautés d’espèces autotrophes et hétérotrophes. Le modèle est mis en oeuvre sur les 220 km de Seine et Marne depuis l’amont de l’agglomération parisienne jusqu’à l’estuaire de la Seine soumis à de forts rejets anthropiques. Il s’agit du domaine simulé à l’échelle pluri-annuelle par Vilmin (2014) sur la période 2007-2012. Le travail est découpé selon les étapes suivantes :

- Programmation et débogage des trois librairies : hydrodynamique (*libhyd*), transport (*libttc*) et biogéochimie (*librive*)
- Couplage des trois librairies et parallélisation des calculs de transport et biogéochimie pour construire une nouvelle version de PROSE : PROSE-P
- Validation du logiciel couplé, PROSE-P, sur une simulation du modèle Seine
- Implémentation d’un schéma d’assimilation de données dans PROSE-

P pour développer une fonctionnalité dans PROSE-P qui permet d’assimiler les concentrations en oxygène dissous et d’estimer les propriétés physiologiques : PROSE-PA

- Validation de PROSE-PA sur un cas synthétique
- Application de PROSE-PA sur le modèle Seine pour l’année 2011

1.6 Cadre de la thèse : Le programme PIREN- Seine

Cette recherche s’inscrit dans le cadre du programme de recherche PIREN-SEINE. Le PIREN-SEINE est un Programme scientifique Interdisciplinaire de Recherche sur l’ENVIRONNEMENT créé en 1989, qui vise à mieux comprendre le fonctionnement du système formé par le réseau hydrographique de la Seine, son bassin versant et la société humaine qui l’habite à partir de mesures de terrain et de modélisations (<https://www.piren-seine.fr>). Ce programme réunit 20 équipes de recherche de toute la France (METIS, IPGP, Centre de Géosciences MINES ParisTech, LEESU et IMPMC etc.), 6 instituts nationaux (CNRS, CEA, INRA, IRSTEA, INERIS et IFSTTAR) ainsi que 10 acteurs publics et privés en charge de la gestion de l’eau dans la bassin de la Seine (AESN, SIAAP, SDEIF, EPTB Seine Grands lacs, etc.). Les instituts de recherche permettent de collaborer sur un grand nombre de thématiques comme l’hydrologie, l’hydrochimie, l’hydrogéologie, la biogéochimie, l’agronomie, l’écotoxicologie et l’histoire etc. (<https://www.piren-seine.fr>).

Le fonctionnement écologique de l’ensemble du système fluvial et sa modélisation, depuis les bactéries jusqu’aux poissons, sont basés sur l’étude fine des processus physiques, chimiques et biologiques des milieux. Les modèles développés par le PIREN-Seine simulent les variations écologiques et biogéochimiques de l’hydrosystème, depuis les ruisseaux jusqu’à l’entrée de l’estuaire.

Il est actuellement en fin de la phase 7 qui a débuté en septembre 2015 avec trois axes de recherche : “Axe 1 Territoire & Usages” qui mène des recherches sur les sources de contaminants des eaux et les interactions entre les territoires urbains et agricoles ; “Axe 2 Fonctionnement hydroécologique” qui étudie les processus physico-chimiques et écologiques à l’œuvre dans le réseau hydrographique, “Axe 3 Connaissance et devenir des territoires” qui inscrit les recherches du PIREN-SEINE dans le temps et au sein des problé-

matiques sociétales actuelles. Depuis 2016, le PIREN-SEINE dispose d'une cellule "Transfert de connaissances", qui permet de transférer des savoirs produits par les chercheurs aux partenaires et de renforcer la relation entre les professionnels de l'environnement et le grand public.

Cette thèse contribue aux travaux de l'Axe 2. Elle a pour but de mettre en synergie données et modèles afin de mieux comprendre et conceptualiser les fonctionnements physiques, chimiques et écosystémiques des corridors hydroécologiques. À terme, le but est de proposer des scénarios crédibles de l'évolution de l'aménagement des corridors pour les futurs du bassin, afin d'anticiper au mieux les éventuelles modalités de gestion des milieux à mettre en place.

1.7 Structure du mémoire

Le cœur de cette thèse est de développer un schéma d'assimilation de données dans le modèle hydro-biogéochimique PROSE, afin de mieux simuler les teneurs en oxygène en Seine et de déterminer l'évolution temporelle des propriétés physiologiques des micro-organismes. Ceci permet d'approfondir ensuite la compréhension du fonctionnement biogéochimique de la Seine. L'assimilation de données nécessite souvent un grand nombre de simulations qui est très coûteux pour analyser les variables d'état (les concentrations) et les paramètres du modèle. Afin de répondre à cette demande, une nouvelle version de PROSE devrait être développée. Le travail est ainsi réalisé en plusieurs étapes comme indiqué précédemment dans les objectifs de la thèse (p. 33).

Le manuscrit s'organise de la manière suivante :

- Dans la première partie, une revue des modèles de la qualité de l'eau en rivière (Section 2.2, p.48) et des méthodes d'assimilation de données classiques est réalisée (Section 2.1, p.42). Cette partie permet de synthétiser les avantages et les inconvénients des modèles de la qualité de l'eau et des méthodes d'assimilation de données basées sur la littérature.
- La deuxième partie est consacrée à la mise en place de la nouvelle version parallélisée de PROSE, PROSE-P, et sa validation par la comparaison des résultats d'une simulation du modèle Seine de PROSE3.6.9 et PROSE-P (Wang et al., 2018a). PROSE-P couple trois modules in-

dépendants : le module hydraulique (*libhyd*), le module de transport (*libttc*) et le module biogéochimique (*librive*). Les schémas numériques utilisés pour le module hydraulique et le module de transport sont détaillés (Section 3.2, p. 58 et Section 3.3, p.64). Les performances de PROSE-P (le gain de la parallélisation, la simulation hydraulique et les concentrations en oxygène dissous) sont évaluées. L'amélioration des performances comprend une réduction des temps de calcul grâce à la parallélisation, une plus grande robustesse du code hydraulique et une meilleure représentation des échanges eau-air intégrant l'influence de la température (Section 3.6, p.72).

- La troisième partie est composée d'une publication dans le journal *Water Research* (Wang et al., 2018b). Cette partie étudie essentiellement l'influence des paramètres du module biogéochimique sur la variation des concentrations en l'oxygène dissous (Chapitre 4, p.83). Les résultats montrent que l'oxygène dissous est plus sensible aux paramètres phytoplanctoniques (la photosynthèse, la respiration, et la température optimale) pour les périodes de bloom algal tandis qu'il est sensible aux paramètres bactériens (le rendement de croissance, la croissance, la température optimale) hors des périodes de bloom algal. Un effet thermique a été constaté. A basse température ($< 6^{\circ}\text{C}$), le coefficient de réaération liée à la navigation est le plus influent. Cette analyse de sensibilité permet de sélectionner un nombre réduit de paramètres pour l'assimilation de données.
- La dernière partie est consacrée au déploiement de l'assimilation de données dans PROSE-PA. Pour cela, un algorithme de filtrage particulaire est implémenté dans PROSE-P en se basant sur l'étude de sensibilité de la partie 4. Les performances de PROSE-PA sont dans un premier temps testées sur un cas synthétique qui simule 66 km de la Seine depuis Suresnes jusqu'à Andrésy. Il s'agit d'un domaine fortement influencé par les activités anthropiques (les rejets de STEP, les rejets de déversoirs d'orage). Cette section a été publiée dans le journal *Water Research* (Chapitre 5, p.129, (Wang et al., 2019)). Dans un deuxième temps, une assimilation des observations réelles de l'année 2011 est réalisée dans le modèle Seine depuis l'agglomération parisienne jusqu'à l'estuaire de la Seine (Chapitre 6, p.169). Pour cela, la répartition de la matière organique dissoute pour les condi-

tions aux limites en amont est d'abord corrigée. Une procédure de calibration des paramètres liés à l'utilisation de la matière organique dissoute est réalisée. La simulation de l'oxygène dissous est fortement améliorée par l'assimilation de données. La variabilité temporelle des paramètres physiologiques est déterminée.

Une notice d'utilisation de PROSE-PA est décrite dans l'annexe (Annexe C, p.219). Les résultats montrent que le filtrage particulaire est efficace pour assimiler les concentrations en oxygène dissous et estimer les paramètres du modèle. Ce travail ouvre la porte stochastique à la modélisation de la qualité de l'eau de la Seine. Il pose une base solide sur laquelle s'appuyer par le développement d'un système de gestion des eaux urbaines en temps réel.

Première partie

Modélisation de la qualité de
l'eau en rivière

Chapitre 2

L'assimilation de données et la modélisation de qualité de l'eau

Résumé

Dans ce chapitre, une revue bibliographique est réalisée. Elle permet au lecteur de rapidement acquérir les concepts sous-jacents aux modèles de qualité de l'eau, à l'assimilation de données et à l'implémentation de cette dernière dans l'exercice délicat de la modélisation des systèmes réels. Plusieurs méthodes d'assimilation de données existent dans la littérature. Les méthodes sont développées pour les modèles linéaires ou les modèles non linéaires.

Les modèles de simulation de la qualité de l'eau ont des spécificités (non linéaires, de nombreux paramètres, agent-centré, conceptuel, mécaniste etc.) qui permettent de sélectionner une méthode appropriée qui va permettre de caractériser les incertitudes.

Le filtrage particulaire est adapté aux modèles non linéaires de dimension réduite (par exemple, les modèles basés sur des communautés d'espèces). le choix est ainsi fait d'implémenter un filtrage particulaire avec le modèle de qualité des eaux PROSE, pour assimiler les données d'oxygène en continu disponibles en Seine.

2.1 Méthodes d'assimilation de données

L'assimilation de données (Bennett, 2002; Kalnay, 2003) est une procédure qui permet de relier les observations aux simulations, afin de mieux décrire un processus (Wikle and Berliner, 2007). On s'intéresse à estimer la variable d'état du système (ici l'oxygène dissous dans la colonne d'eau) et les paramètres du modèle PROSE (ici les propriétés physiologiques). Le principe de l'assimilation de données pour le cas de l'oxygène dissous est décrit en figure 2.1. L'oxygène dissous dans la colonne d'eau est initialement simulé avec le jeu de paramètres **JP1**. Les mesures haute fréquence nous permettent de calculer la différence entre simulation et observation. En fonction des différences, un nouveau jeu de paramètres est estimé (**JP2**). Ce nouveau jeu de paramètre est utilisé jusqu'à atteindre un nouveau point d'observation qui sert à estimer un nouveau jeu de paramètre **JP3**. Le processus se poursuit ainsi jusqu'à la fin de la simulation.

Il existe de nombreuses méthodes d'assimilation de données dans la littérature. Un tutoriel bayésien pour l'assimilation de données est présenté par Wikle and Berliner (2007). Carrassi et al. (2018) examinent les méthodes d'assimilation de données en géoscience. Quelques méthodes parmi les plus connues et populaires sont ensuite présentées de façon descriptive.

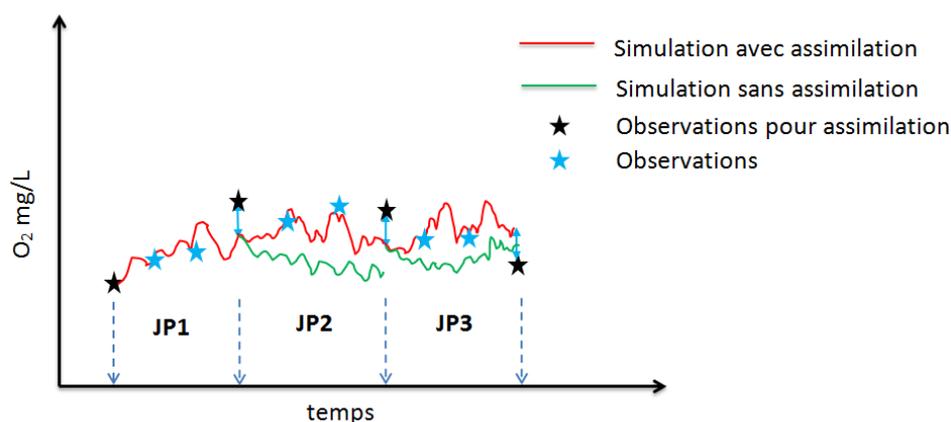


FIGURE 2.1 – Schéma simplifié d'assimilation de données séquentielle et continue, JP : Jeux de paramètres

2.1.1 Notations

- \mathbf{X} : La variable d'état (ici l'oxygène dissous)
- \mathbf{Y} : Les observations
- \mathbf{H} : L'opérateur d'observation
- \mathbf{M} : L'opérateur de modèle
- \mathbf{P} : La matrice de covariance d'erreur de prévision ou d'analyse de \mathbf{X}
- \mathbf{Q} : La matrice de covariance d'erreur du modèle ($\boldsymbol{\eta}$)
- \mathbf{R} : La matrice de covariance d'erreur d'observation ($\boldsymbol{\epsilon}$)
- $\boldsymbol{\epsilon}$: Les erreurs des observations $\mathbf{y} - \mathbf{H}\mathbf{x}$
- $\boldsymbol{\eta}$: Les erreurs du modèle $\mathbf{x} - \mathbf{M}\mathbf{x}$

Les lettres majuscules \mathbf{X} et \mathbf{Y} correspondent à des variables aléatoires. Les lettres minuscules \mathbf{x} et \mathbf{y} représentent les réalisations de \mathbf{X} et \mathbf{Y} . Les lettres grecques $\boldsymbol{\epsilon}$ et $\boldsymbol{\eta}$ correspondent aussi à des variables aléatoires. L'opérateur de modèle \mathbf{M} correspond à l'évolution temporelle des processus objectifs. L'opérateur d'observation \mathbf{H} relie l'observation \mathbf{y} avec la variable d'état \mathbf{x} simulée. On peut décrire le système en question par les équations suivantes.

$$\mathbf{x}_t = \mathbf{M}(\mathbf{x}_{t-1}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \quad (2.1)$$

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{R}_t) \quad (2.2)$$

2.1.2 Méthodes variationnelles

Les méthodes variationnelles (3D-Var, Méthode d'assimilation variationnelle tri-dimensionnelle, 4D-Var, Méthode d'assimilation variationnelle quadri-dimensionnelle) sont introduites par [Sasaki \(1955, 1958\)](#) dans les années 1950. L'approche variationnelle consiste à minimiser une fonction coût (fonction objectif, $\mathbf{J}(\mathbf{x})$) mesurant par exemple l'écart entre simulations et observations.

2.1.2.1 3D-Var

Le 3D-Var cherche à minimiser la différence au sens des moindres carrés entre l'état estimé et les observations. Le nouvel état analysé est utilisé comme point de départ de la prévision suivante. La fonction coût pour cette méthode est définie comme :

$$\mathbf{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^{true})^T \mathbf{Q}^{-1}(\mathbf{x} - \mathbf{x}^{true}) \quad (2.3)$$

La valeur minimale de \mathbf{J} peut être obtenue en annulant la dérivée $\nabla \mathbf{J}(\mathbf{x}) = 0$.

$$\nabla \mathbf{J}(\mathbf{x}) = \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}) + \mathbf{Q}^{-1}(\mathbf{x} - \mathbf{x}^{true}) \quad (2.4)$$

où \mathbf{H} désigne l'opérateur tangent-linéaire de \mathbf{H} .

L'équation 2.4 peut être résolue par des méthodes numériques comme la méthode de Newton, le gradient conjugué etc. Il existe plusieurs implémentations de cette méthode : 3D-Var classique (analyser à chaque pas de temps d'observation ou moyenner les observations sur une fenêtre temporelle définie), 3D-Var FGAT (analyser au pas de temps le plus proche sur une fenêtre temporelle définie) et 3D-Var incrémental (Thompson, 1969; Daget, 2007).

2.1.2.2 4D-Var

La méthode d'assimilation variationnelle quadri-dimensionnelle, 4D-Var, est considérée comme une extension temporelle du 3D-Var. Au lieu d'analyser à l'instant donné sur la fenêtre temporelle définie (3D-Var), 4D-Var cherche la trajectoire optimale sur la fenêtre de temps définie (Daget, 2007). 4D-Var prend ainsi en compte les observations, mais aussi leurs distributions spatiales et temporelles (dans la fenêtre de temps définie). 4D-Var permet donc d'obtenir la trajectoire la plus proche possible de l'ensemble des observations utilisées.

2.1.3 Filtre de Kalman

Le filtre de Kalman (Kalman, 1960) est une solution récursive pour des modèles linéaires, dont l'erreur d'observation ϵ est gaussienne. Il s'agit d'une approche séquentielle. Le filtre de Kalman introduit l'évolution de l'état du système entre deux instants t_i et t_{i+1} .

Notons la matrice de covariance d'erreur d'analyse, \mathbf{P}^a , la matrice de covariance de l'erreur de prévision, \mathbf{P}^f . Les opérateurs de modèle et d'observation sont tous linéaires. On décrit le système par les équations suivantes :

$$\mathbf{x}_t = \mathbf{M}_{t-1}(\mathbf{x}_{t-1}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \quad (2.5)$$

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{R}_t) \quad (2.6)$$

La matrice de covariance d'erreur de prévision est calculée (Eq. (2.7)).

$$\mathbf{P}_t^f = \mathbf{M}_{t-1} \mathbf{P}_{t-1}^a \mathbf{M}_{t-1}^T + \mathbf{Q}_t \quad (2.7)$$

Le filtre de Kalman est décrit de la manière suivante (Wikle and Berliner, 2007; Daget, 2007)

— Calcul de la matrice de gain \mathbf{K} à l'instant i :

$$\mathbf{K}_t = \mathbf{P}_t^f \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_t^f \mathbf{H}_t^T + \mathbf{R}_t)^{-1} \quad (2.8)$$

— Analyse \mathbf{x}_t^a et \mathbf{P}_t^a à l'instant t avec \mathbf{y}_t :

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_t^f) \quad (2.9)$$

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_t^f \quad (2.10)$$

— Prévision pour instant $t + 1$:

$$\mathbf{x}_{t+1}^f = \mathbf{M}_t(\mathbf{x}_t^a) \quad (2.11)$$

$$\mathbf{P}_{t+1}^f = \mathbf{M}_t \mathbf{P}_t^a \mathbf{M}_t^T + \mathbf{Q}_{t+1} \quad (2.12)$$

Il suffit ensuite de répéter cette procédure au cours du temps.

2.1.4 Le filtre de Kalman d'ensemble (EnKF)

Evensen (1994) a développé initialement la méthode filtre du Kalman d'ensemble (EnKF), puis l'a corrigée en 1998 (Burgers et al., 1998). Cette méthode utilise la méthode de Monte-Carlo pour estimer la matrice de covariance d'erreur de prévision en faisant l'hypothèse gaussienne à partir d'un ensemble de simulations (la moyenne et la matrice de covariance, Fig.2.2 et Eq. (2.13)).

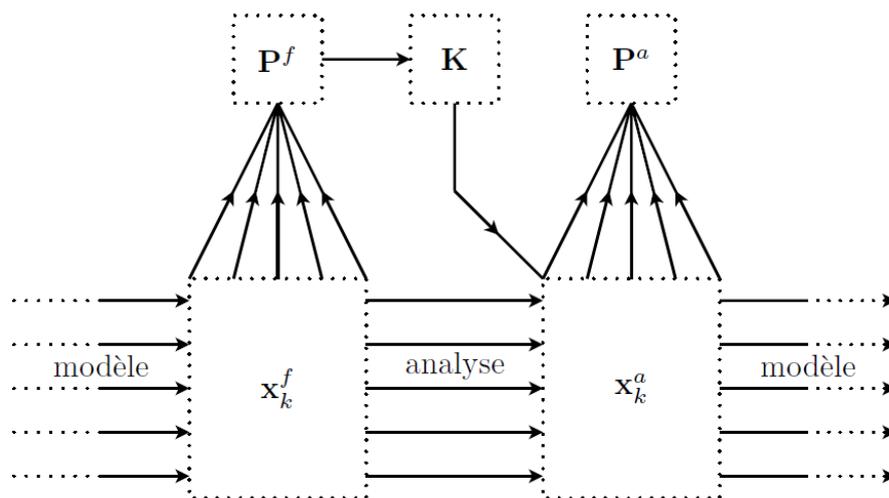


FIGURE 2.2 – Représentation schématique des différents étapes du filtre de Kalman d'ensemble (Daget, 2007)

$$\mathbf{P}_t^f = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_t^{f,i} - \hat{\mathbf{x}}_t^{f,i})(\mathbf{x}_t^{f,i} - \hat{\mathbf{x}}_t^{f,i})^T \quad (2.13)$$

$$\hat{\mathbf{x}}_t^{f,i} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_t^{f,i} \quad (2.14)$$

Avec, N le nombre de simulations. Le calcul de la matrice de gain \mathbf{K} et l'étape d'analyse \mathbf{x}_t^a s'effectuent de la même manière qu'avec le filtre de Kalman (Eq.(2.8) et (2.9)). EnKF est devenu maintenant l'approche la plus populaire pour l'assimilation de données à haute dimension dans les domaines hydrologique, atmosphérique et océanique. Il existe aussi des variations du filtre de Kalman d'ensemble qui ne sont pas présentées ici (Hendricks Franssen and Kinzelbach, 2008; Li et al., 2012; Sakov et al., 2012; Gharamti et al., 2013).

En plus des méthodes citées précédemment, il existe des approches qui combinent les méthodes variationnelles et les méthodes d'ensemble, des extensions du filtre de Kalman dans la littérature (Hamill and Snyder, 2000; Beck, 1987; Simon J. Julier, 1997; Wan and Van Der Merwe, 2000). Ces méthodes ne sont pas présentées ici.

2.1.5 Méthodes de Monte Carlo séquentielles

Une autre technique alternative est référencée comme les méthodes de Monte Carlo séquentielles (Doucet et al., 2001). Ces méthodes sont basées sur le théorème de Bayes (Bayes, 1763) et utilisent la méthode de Monte Carlo pour estimer la distribution a posteriori d'une variable d'état. Il est souvent difficile de tirer des réalisations à partir d'une distribution inconnue. Pour adresser ce problème, on utilise souvent la méthode de l'échantillonnage d'importance (*Importance Sampling (IS)*, (Liu, 2001)). Cette méthode est utilisée récursivement dans les approches séquentielles : *Sequential Importance Sampling (SIS)*, *Sequential Importance sampling Resampling (SIR)* (Doucet et al., 2001). SIS et SIR sont similaires. La différence entre les deux est la fréquence du ré-échantillonnage (à chaque pas de temps ou selon des critères) pour diminuer l'effet de dégénérescence (Moradkhani et al., 2005). Ces méthodes utilisent un ensemble de simulations comme l'EnKF. Les simulations sont appelées aussi particules. La technique est aussi référencée comme le filtrage particulaire dans la littérature. L'EnKF peut être vu comme une approximation du filtrage particulaire (Monbet, 2009). Cependant le filtrage particulaire nécessite plus de simulations que l'EnKF. Un filtrage particulaire est mis en œuvre et détaillé dans le chapitre 5.

2.1.6 Applications

De nombreuses applications des méthodes citées existent dans la littérature. Les méthodes variationnelles sont populaires dans le domaine atmosphérique (météorologique) (Kalnay et al., 1996; Barker et al., 2004; Dee et al., 2011). Hostache et al. (2010) ont assimilé les niveaux d'eau issus des images satellitales avec la méthode variationnelle 4D-Var. Le filtre de Kalman d'ensemble (EnKF) a été utilisé largement dans la littérature grâce à la nécessité d'un petit nombre de simulations par rapport au filtrage particulaire, soit en atmosphère (Lorenc, 2003; Annan et al., 2005), soit en hydrologie (Durand et al., 2008; Camporese et al., 2009; Hendricks Franssen et al., 2011; DeChant and Moradkhani, 2012; Paiva et al., 2013; Samuel et al., 2014; Shi et al., 2014; Massari et al., 2015). Le filtrage particulaire attire de plus en plus d'attention dans le domaine hydrologique (Moradkhani et al., 2005; Weerts and El Serafy, 2006; Salamon and Feyen, 2009; Plaza et al., 2012; Yan and Moradkhani, 2016; Abbaszadeh et al., 2018). La faisabilité du filtrage particulaire pour un modèle écologique est testée par

(Kattwinkel and Reichert, 2017).

En océanographie, la faisabilité de l'EnKF et des extensions de l'EnKF (Mao et al., 2009; Simon and Bertino, 2012; Simon et al., 2012; Gharamti et al., 2017; Yu et al., 2018) et 3D-Var (Cossarini et al., 2019) est testée sur les modèles biogéochimiques et écologiques. Pour l'assimilation de la qualité de l'eau de surface, les études récentes se concentrent sur les prévisions des blooms des algues soit en rivière (Kim et al., 2014) soit en lac (Huang et al., 2013; Page et al., 2018). L'assimilation des données de l'oxygène dissous haute fréquence est rarement publiée. Pastres et al. (2003) a utilisé l'EKF (extended Kalman filter) pour estimer trois paramètres d'un modèle simple de DO-chorophyll dans la lagune de Venise. Xue et al. (2012) a appliqué l'EnKF afin de définir la localisation optimale pour les mesures en oxygène dissous dans la baie du Massachusetts.

2.2 Revue des modèles de la qualité de l'eau en rivière

La modélisation de l'oxygène dissous fait intervenir de nombreux processus physiques et biogéochimiques. Pour les processus physiques, la réaération à l'interface eau-air et la réoxygénation aux ouvrages hydrauliques sont des sources d'oxygène importantes. La réaération est influencée par la vitesse du vent, la vitesse de l'écoulement, la diffusion moléculaire et la navigation tandis que l'oxygénation est liée à une chute d'eau en aval des barrages. Pour la part biogéochimique, la production primaire (photosynthèse) et la respiration des micro-organismes sont les processus les plus importants. La modélisation des micro-organismes est basée sur les communautés d'espèces ou les équations cinétiques chimiques. A l'exception des processus cités, la demande benthique en O_2 (le flux d'oxygène demandé par le sédiment) est soulignée par (Vilmin et al., 2016). La demande benthique en O_2 est modélisée explicitement ou simulée par un terme constant.

De nombreux modèles de simulation de la qualité des eaux en rivières sont disponibles (AQUATOX, modèles QUAL, modèles MIKE, QUASAR, WASP, SIMCAT, TOMCAT RWQM, et PROSE). Ces modèles sont agent-centrés ou basés sur les équations cinétiques. Une revue des modèles, notamment la simulation de l'oxygène dissous dans ces modèles est réalisée dans la suite. Cette synthèse des modèles s'inspire dans la littérature (Rauch et al.,

1998; Reichert et al., 2001a; Cox, 2003a; Flipo, 2005; Kannel et al., 2011; Sharma and Kansal, 2013; Wang et al., 2013; Vilmin, 2014).

AQUATOX

Il s'agit d'un modèle mécaniste d'écosystème qui simule l'impact des nutriments, matières organiques, produits toxiques sur les espèces dans le milieu aquatique. Il est basé sur un modèle biologique en milieu aquatique CLEAN (Park et al., 1974) et un modèle d'écotoxicologie PEST (Park et al., 1982). Dans les modèles cités, il est le seul modèle qui permet de simuler l'effet des polluants sur les poissons, et les invertébrés aquatiques (Park et al., 2008). Les processus liés à l'oxygène dissous comme la photosynthèse, la réaération, la respiration des vivants, la nitrification et la dégradation des matières organiques sous forme DBO sont simulés. La résolution du système se fait avec un schéma de Runge-Kutta d'ordre 4 ou 5. Il est capable de simuler l'impact des sédiments mono ou multicouche. De plus, le pH est aussi simulé dans ce modèle. Un outil d'analyse des incertitudes par Hypercube Latin intégré renforce l'intérêt de son utilisation. Cependant, le processus hydraulique n'est que simulé en régime permanent.

QUAL2kw

Il existe une série de modèles QUAL : QUAL2E/QUAL2EU, QUAL2k, QUAL2kw. Park and Lee (2002) a développé QUAL2k après l'identification des limites des modèles QUAL2E/QUAL2EU en 2002. L'amélioration principale apportée dans QUAL2k est d'ajouter la transformation du phytoplancton détritique en DBO (Demande Biologique en Oxygène) et la quantification de la production d'oxygène par les plantes.

QUAL2kw (Pelletier et al., 2006) est un modèle issu du modèle QUAL2k. Dans ce modèle, le processus hydraulique est considéré en régime permanent. Néanmoins, la qualité de l'eau est simulée en régime transitoire. De plus, il n'est pas capable de simuler les branches d'une rivière. Un bilan énergétique est intégré dans QUAL2kw. Il est ainsi possible de simuler les matières organiques, les nutriments, oxygène dissous, les phytoplancton, la température, pH, conductivité, alcalinité. La plupart des processus sont décrits par une cinétique du 1er ordre excepté l'oxygène dissous, le nitrate et phosphate. L'oxygène dissous est simulé par la photosynthèse, la respiration, la réaération, l'oxydation de la DBO, et la nitrification. Le phytoplancton est

explicitement simulé via la photosynthèse (Beer-Lambert), la respiration, la mortalité et la sédimentation. Les limitations en nutriments sont exprimées par les équations de Monod. Les échanges de nutriments, et d'oxygène dissous entre l'eau et le sédiment sont explicitement simulés. Le logiciel dispose en plus une procédure d'auto-calibration utilisant un algorithme génétique.

MIKE 11/MIKE HYDRO River

MIKE 11 est développé par le Danish Hydraulic Institute (DHI) depuis 1972. Il s'agit d'un modèle commercial couplé avec un module hydraulique qui résout les équations de Saint-Venant à une dimension en régime transitoire. Le transport des matières se fait avec un module de transport qui simule l'advection et la dispersion des substances dans l'eau. Le module de qualité de l'eau peut être lancé en choisissant le niveau de complexité du système. Par exemple pour la simulation de l'oxygène, les utilisateurs peuvent activer ou désactiver l'effet des sédiments (échanges eau-sédiment, demande benthique en oxygène), la demande en oxygène par la nitrification. Pour simuler l'oxygène dissous, la photosynthèse, la respiration, la réaération et la dégradation des matières organiques sont systématiquement pris en compte. Les biomasses bactériennes ne sont jamais explicitement représentées comme dans d'autres modèles cités. La consommation d'oxygène via l'hydrolyse et la minéralisation des matières organiques par les bactéries hétérotrophes est calculée avec une cinétique du 1er ordre en utilisant un terme de DBO. Comme MIKE 11 permet de simuler dynamiquement le débit et la qualité de l'eau (le régime transitoire, les interactions eau-sédiment, l'advection-dispersion et la température), il nécessite de renseigner un important jeu de données d'entrée (les données géométriques de la rivière, les conditions aux limites temporelles).

La nouvelle génération des modèles MIKE pour la modélisation de la qualité de l'eau en rivière s'appelle MIKE HYDRO River qui se base sur MIKE 11. Il est capable d'assimiler des données avec un filtre de Kalman d'ensemble (EnKF). Une étude récente qui utilise MIKE HYDRO River pour assimiler les hauteurs d'eau d'une rivière issues de satellite a été réalisée par [Schneider et al. \(2018\)](#). Une procédure d'auto-calibration est aussi intégrée dans MIKE HYDRO River.

QUASAR

QUASAR a été développé pour évaluer l'impact environnemental des polluants sur la qualité des eaux en rivière (Whitehead et al., 1997). Il est initialement développé à partir d'une étude de la rivière Bedford Ouse (Whitehead et al., 1981). C'est un modèle dynamique qui simule l'hydraulique, le transport et le devenir du carbone, des nutriments et l'oxygène en une dimension. Il est capable de simuler le système rivière en mode stochastique avec la méthode de Monte Carlo comme dans SIMCAT et TOMCAT (cf. p.2.2). Le modèle de simulation de l'oxygène dissous ressemble à celui de MIKE 11 et prend en compte la photosynthèse, la respiration, la réaération, l'oxydation de DBO, la nitrification et la demande benthique en oxygène. Un terme source lié à la reoxygénation aux barrages est en plus considéré. Cependant les biomasses des micro-organismes, phytoplancton, bactéries, ne sont pas explicitement simulées. Il n'existe pas d'échanges de nutriments entre l'eau et le sédiment. La demande benthique en oxygène est fixée comme dans la plupart des modèles cités. Ce modèle s'appuie comme MIKE sur un jeu de données d'entrée conséquent.

WASP 8

Le développement du modèle WASP a commencé dans les années 70s par USEPA (Wool et al., 2001). Il a été ensuite amélioré en WASP 7 (Ambrose and Wool, 2009). Il s'agit d'un modèle qui simule le transport (advection-dispersion, sédiment) et les cycles du carbone, des nutriments et de l'oxygène dans le milieu aquatique. Il est capable de simuler le système rivière à une, deux ou trois dimensions. Dans ce modèle, plusieurs sous-modèles existent : EUTRO pour la qualité de l'eau, TOXI pour les espèces toxiques, MERCURY et HEAT. Il est ainsi possible de simuler les produits toxiques, les métaux et la température. WASP permet d'attribuer des stoechiométries C :N :P spécifiques aux différentes fractions de matière organique. L'oxygène dissous est influencé par la réaération, la photosynthèse, la respiration, la dégradation de DBO et la demande benthique en O₂.

WASP 7 ne permet pas de simuler les macrophytes, ce qui est intégré dans WASP 8. WASP 8 a notamment apporté les implémentations sur le transport et la transformation des nanomatériaux (Ambrose et al., 2017).

SIMCAT et TOMCAT

SIMCAT est un modèle stochastique et déterministe développé par Angilan Water ([Warn, 1987](#)). Il utilise la méthode de Monte Carlo afin de prédire le comportement de la qualité de l'eau sous forme de moyennes et de quantiles. Il s'agit d'un modèle simplifié à l'extrême. Dans SIMCAT, on ne résout pas l'équation de transport (Advection-Dispersion). Pour simuler l'oxygène dissous, le modèle ne considère que la dégradation de la DBO et la réaération, tandis que la photosynthèse, la respiration, la nitrification et la demande benthique en oxygène ne sont pas pris en compte. La réaération est estimée avec un coefficient de réaération constant. De plus, il ne simule que le système en régime permanent à une dimension. Dans ce logiciel, une procédure d'auto-calibration est intégrée.

TOMCAT est développé par UK water utility company, Thames Water, dans les années 80s ([Bowden and Brown, 1984](#)). La conceptualisation de TOMCAT ressemble à celle de SIMCAT. Comme SIMCAT, il dispose aussi d'une approche stochastique de Monte Carlo, mais il permet des corrélations temporelles plus complexes. La transformation des solutés est identique à celle de SIMCAT à l'exception de la température et de l'oxygène dissous. Dans TOMCAT, l'effet de la nitrification est implémenté. De plus, le coefficient de réaération est calculé en fonction de la bathymétrie de la rivière et de la température.

Les deux modèles sont stochastiques et déterministes, mais simplifiés à l'extrême.

RWQM

Le RWQM est réalisé dans le cadre d'*IWA Task Group on River Water Quality Modelling* ([Shanahan et al., 2001](#); [Reichert et al., 2001b](#); [Vanrolleghem et al., 2001](#); [Reichert, 2001](#); [Reichert et al., 2001a](#); [Reichert and Vanrolleghem, 2001](#)) pour développer une base scientifique et technique des formalismes sur la modélisation de la qualité de l'eau en rivière. Il simule les cycles de C, O, N et P dans une rivière en utilisant une stoechiométrie constante. Il est le seul modèle qui est possible d'intégrer le fonctionnement du réseau d'assainissement pour étudier son impact sur le milieu récepteur (le logiciel WEST ([Vanhooren et al., 2003](#))).

RIVE et ProSe

Le logiciel PROSE a été initialement développé par [Even \(1995\)](#) à MINES ParisTech. Ce logiciel est constitué de trois modules : hydraulique, transport et biogéochimique. Le module hydraulique résout les équations de Saint Venant à une dimension en régime permanent ou transitoire. La résolution numérique se fait par un schéma semi-implicite aux différences finies sur grilles décalées. Le module de transport se base sur l'équation d'advection-dispersion. Le modèle RIVE, cœur biogéochimique du modèle PROSE, développé depuis 30 ans (<https://www.fire.upmc.fr/rive>), est basé sur des communautés d'espèces qui permet de réduire l'espace des paramètres. RIVE permet de simuler explicitement les processus biogéochimiques comme la photosynthèse, la respiration, la croissance ainsi que les cycles des nutriments dans plusieurs compartiments (EAU, VASE et PERIPHYTON). Les dynamiques des bactéries hétérotrophes et nitrifiantes sont explicitement simulées pour la dégradation des matières organiques, la nitrification et la dénitrification. La réaération du dioxygène est calculée en prenant compte la vitesse du vent, la vitesse de l'écoulement, la diffusion moléculaire et plus particulièrement l'effet de la navigation. La sédimentation et l'érosion des sédiments sont simulées simultanément. Les échanges des substances dissoutes entre l'eau et le sédiment se font par la diffusion. L'intégration du terme de puissance (remise en suspension des matières particulaires) transmise par la navigation pour l'érosion a significativement amélioré la simulation des blooms algaux en 2011 et en 2012 ([Vilmin, 2014](#); [Vilmin et al., 2015b](#)). Le modèle RIVE a été implémenté dans plusieurs codes numériques, notamment RIVERSTRAHLER ([Billen et al., 1994](#); [Garnier et al., 1995](#)), pyNuts-RIVERSTRAHLER ([Thieu et al., 2017](#)), PROSE ([Even, 1995](#); [Even et al., 1998](#); [Flipo et al., 2004, 2007b](#); [Vilmin et al., 2015b, 2016](#)) et QUAL-NET ([Minaudo et al., 2018](#)).

2.3 Autre logiciels

A l'exception des logiciels cités précédemment, il existe aussi des logiciels de simulation de la qualité de l'eau à l'échelle des bassins versant : SWAT ([Arnold et al., 1993](#)), HSPF ([Barnwell et al., 1987](#)), RIVERSTRAHLER ([Billen et al., 1994](#)). Ces logiciels reproduisent le cycle de l'eau et permettent de calculer des flux de matières au sein du réseau hydrographique.

2.4 Conclusion

Les processus liés à la simulation de l'oxygène dissous pour les modèles étudiés sont synthétisés dans le table 2.1. D'après la revue menée dans la section 2.2, p. 48, peu de logiciels de qualité de l'eau en rivière disposent d'un schéma d'assimilation de données. L'assimilation de données est devenue nécessaire pour mettre en synergie les données haute fréquence et les modèles, notamment pour la gestion des eaux urbaines en temps réel.

Parmi les méthodologies d'assimilation de données, le filtrage particulaire présente de nombreux arguments en sa faveur. En effet, il est bien adapté aux problèmes non linéaires de dimension réduites, ce qui est le cas du modèle centré sur les communautés d'espèces comme C-RIVE. Il permet également de quantifier les incertitudes pour des structures d'erreurs non gaussiennes, et de conserver le bilan de masse au moment du ré-échantillonnage (analyse).

Le choix est ainsi fait d'implémenter un filtrage particulaire avec le modèle de qualité des eaux ProSe, pour assimiler les données d'oxygène en continu disponibles en Seine (Chapitre 5, p.129). Aucun travail de ce type n'a été identifié dans la revue bibliographique.

TABLE 2.1: Description des modèles de qualité de l'eau

Modèles	Processus lié à la simulation de l'oxygène dissous					
	Hydraulique	Réaération	Réoxygénation aux bar-rages	Photosynthèse-Respiration	Bactéries hétérotrophes	Demande biologique en O2
AQUATOX	Permanent	Oui (vitesse du vent, hauteur d'eau)	Non	Oui(espèces-centré)	Non (DBO)	Oui (cinétique)
QUAL2kw	Permanent	Oui (vitesse de l'écoulement, hauteur d'eau)	Non	Oui	Non (DBO)	Oui (cinétique)
MIKE	Transitoire	Oui (vitesse de l'écoulement, hauteur d'eau, pente)	Non	Oui	Non (DBO)	Oui (cinétique)
QUASAR	Transitoire	Oui	Oui	Oui	Non (DBO)	Oui (cinétique)
WASP	Transitoire	Oui	Non	Oui	Non (DBO)	Oui (cinétique)
SIMCAT	Permanent	Oui	Non	Non	Non (DBO)	Non

TOMCAT	Permanent	Oui (largeur de la rivière, section mouillée)	Non	Non (DBO)	Non
RWQM	Transitoire	Oui	Non	Oui	Oui
PROSE	Transitoire	Oui (vitesse du vent, vitesse de l'écoulement, diffusion moléculaire, navigation)	Oui	Oui (espèces-centré)	Oui (benthos, diffusion)
RIVERSTRAHLER (basin-scale)	Oui	oui	Oui (espèces-centré)	Oui (espèces-centré)	Oui (benthos, diffusion)
SWAT (basin-scale)	Conceptuel	Oui (vitesse de l'écoulement, diffusion moléculaire, hauteur d'eau)	Oui	Oui (espèces-centré)	Oui (cinétique)
HSPF (basin-scale)	Conceptuel	Oui (vitesse du vent, vitesse de l'écoulement, hauteur d'eau)	Oui (espèces-centré)	non (DBO)	Oui (cinétique)

DBO : Demande Biologique en Oxygène (cinétique du 1er ordre)

Cinétique : Cinétique du 1er ordre

Chapitre 3

Mise en place d'une nouvelle version de ProSe : ProSe-P

Résumé

Ce chapitre présente la stratégie de construction de la nouvelle version parallélisée de PROSE, PROSE-P, et ses performances. Trois bibliothèques programmées en ANSI C (le module hydrodynamique, le module de transport et le module biogéochimique) sont validées séparément et couplées afin de développer PROSE-P, après avoir reformulé le coefficient de diffusion de dioxygène à base physique et fiabilisé les codes. La nouvelle version parallélisée de PROSE, PROSE-P, est vérifiée par une comparaison avec PROSE3.6.9 (Wang et al., 2018a). Les simulations de PROSE-P indiquent une réduction de temps de calcul se basant sur la parallélisation, une plus grande robustesse du code hydraulique et une meilleure représentation des échanges eau-air intégrant l'influence de la température. Il s'agit d'une étape importante pour développer un schéma d'assimilation de données dans le logiciel PROSE.

3.1 Structure du code

Il y a 3 modules pour un calcul séquentiel hydraulique, puis transport, et puis biogéochimique (Fig. 3.1). Chaque calcul s'appuie sur une bibliothèque spécifique programmée en C dont le codage a été initié dans les travaux antérieurs : le module hydraulique (*libhyd*) et le module biogéochimique (C-RIVE) par le travail de Lauriane Vilmin, et la bibliothèque de transport (*libttc*)

par Agnès Rivière dans le cadre du développement des outils de l'équipe SHR (Systèmes Hydrologiques et Réservoirs) du Centre de Géosciences suivant les recommandations des outils formulés (Flipo, 2013).

Dans cette thèse, ces travaux préliminaires ont été fiabilisés par un travail important de débogage, d'implémentation, de reformulation de concepts physiques, et la vérification de chaque librairie séparément. Une fois ces tâches effectuées, une nouvelle version parallélisée de PROSE a été développée en couplant les librairies fonctionnelles. Ce travail et sa vérification sont présentées dans la suite de ce chapitre.

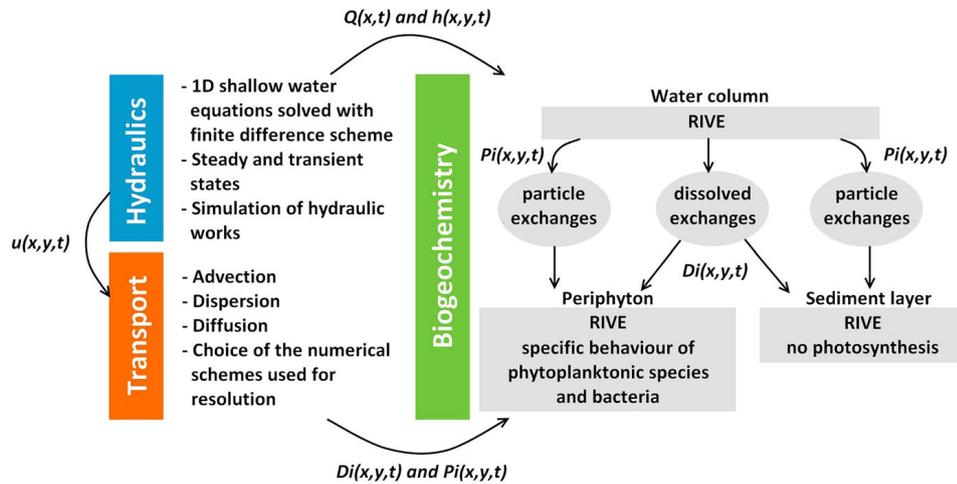


FIGURE 3.1 – Schéma simplifié du logiciel PROSE (d'après (Vilmin et al., 2015a))

3.2 Librairie hydraulique *libhyd*

Cette librairie hydraulique (Vilmin, 2012) résout les équations de Saint-Venant à une dimension avec un schéma numérique aux différences finies sur grille décalée. Le système d'équations comprend la conservation de la masse et de la quantité de mouvement (Eq (3.1) et (3.2)). Ce schéma donne finalement les débits aux faces de chaque maille et la cote au centre de chaque maille.

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q \quad (3.1)$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) = -gA \left(\frac{\partial Z}{\partial x} - i \right) \quad (3.2)$$

avec,

A : la surface mouillée, [m²]

Q : le débit, [m³.s⁻¹]

q : l'apport ou le prélèvement, [m³.s⁻¹]

Z : la cote de la surface libre [m]

i : la pente hydraulique, calculée avec la formule de Manning-Stricker,

[-], $Q = A \cdot K_s \cdot R_h^{3/2} \cdot \sqrt{i}$

x : la distance longitudinale

En faisant l'hypothèse d'une section de rivière dont la largeur (L) est grande devant la hauteur d'eau (h), on peut estimer $A=h \cdot L$.

3.2.1 Discrétisation de l'équation de conservation de la masse

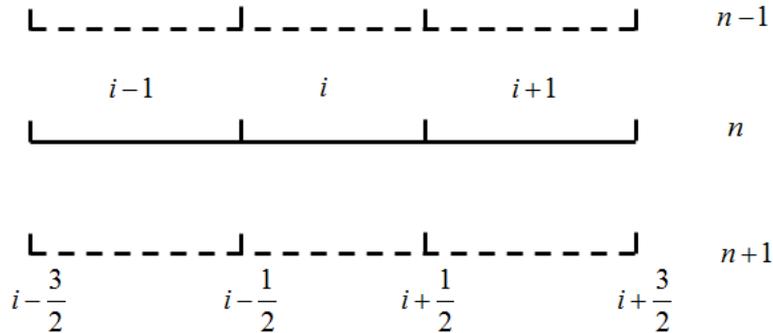


FIGURE 3.2 – Les indices du schéma numérique

Le schéma numérique utilisé pour la discrétisation des équations est un schéma semi-implicite aux différences finies sur grilles décalées, très proche d'un schéma aux volumes finis. Les charges hydrauliques sont calculées au centre des cellules, et les débits aux faces (Fig. 3.2).

$$\frac{\partial(h \cdot L)}{\partial t} + \frac{\partial Q}{\partial s} = q \quad (3.3)$$

$$CM1 + CM2 = CM3 \quad (3.4)$$

Chaque terme de l'équation 3 est noté CMi (Eq. 3.4) et discrétisé pas à pas :

$$\begin{aligned} CM1 &= \left(h \cdot \frac{\partial L}{\partial t} + L \cdot \frac{\partial h}{\partial t} \right)_i \\ CM1 &= \left(h \cdot \frac{\partial L}{\partial Z} \cdot \frac{\partial Z}{\partial t} + L \cdot \left(\frac{\partial Z}{\partial t} - \frac{\partial Z_f}{\partial t} \right) \right)_i \\ CM1 &= \left(\frac{\partial Z}{\partial t} \right)_i \left(h \cdot \frac{\partial L}{\partial Z} + L \right) \end{aligned} \quad (3.5)$$

$$CM2 = \frac{Q_{i+\frac{1}{2}} - Q_{i-\frac{1}{2}}}{\Delta x_i} \quad (3.6)$$

En substituant l'équation (3.5) et l'équation (3.6) dans l'équation (3.3), on obtient :

$$\left(\frac{\partial Z}{\partial t} \right)_i = \frac{1}{\left(h \cdot \frac{\partial L}{\partial Z} + L \right)_i} \left(-\frac{Q_{i+\frac{1}{2}} - Q_{i-\frac{1}{2}}}{\Delta x_i} + q_i \right) \quad (3.7)$$

Dont la discrétisation temporelle est la suivante :

$$\begin{aligned} LHS &= RHS \quad (3.8) \\ LHS &= -\frac{\theta}{\Delta t} \cdot Q_{i-\frac{1}{2}}^{n+1} + \frac{\alpha_i^n}{\Delta t} \cdot Z_i^{n+1} + \frac{\theta}{\Delta t} \cdot Q_{i+\frac{1}{2}}^{n+1} \\ RHS &= \alpha_i^n \cdot \frac{Z_i^n}{\Delta t} + \frac{(1-\theta) \cdot Q_{i-\frac{1}{2}}^n - (1-\theta) \cdot Q_{i+\frac{1}{2}}^n}{\Delta x_i} + \theta \cdot q_i^{n+1} + (1-\theta) \cdot q_i^n \end{aligned}$$

avec $\alpha_i^n = \left(h \cdot \frac{\partial L}{\partial Z} + L \right)_i^n$

3.2.2 Discrétisation de l'équation de conservation de la quantité de mouvement

En discrétisant spatialement l'équation (3.2), on obtient :

$$\begin{aligned} \left(\frac{\partial Q}{\partial t} \right)_{i+\frac{1}{2}} + 2 \cdot \frac{Q_{i+\frac{1}{2}}}{A_{i+\frac{1}{2}}} \cdot \left(\frac{\partial Q}{\partial x} \right)_{i+\frac{1}{2}} - \frac{Q_{i+\frac{1}{2}}^2}{A_{i+\frac{1}{2}}^2} \cdot \frac{A_{i+1} - A_i}{\Delta x_{i+\frac{1}{2}}} \\ = -g \cdot A_{i+\frac{1}{2}} \cdot \frac{Z_{i+1} - Z_i}{\Delta x_{i+\frac{1}{2}}} - \frac{C_f \cdot Q_{i+\frac{1}{2}}^2}{2 \cdot h_{i+\frac{1}{2}} \cdot A_{i+\frac{1}{2}}} \end{aligned} \quad (3.9)$$

avec $C_f = \frac{2gh}{K_s^2 R_h^{4/3}}$

Chacun des termes de l'équation (3.9) est explicité par la suite suivant les notations de l'équation (3.10).

$$QM1 + QM2 + QM3 = QM4 + QM5 \quad (3.10)$$

$$QM1 = \frac{Q_{i+\frac{1}{2}}^{n+1} - Q_{i+\frac{1}{2}}^n}{\Delta t} \quad (3.11)$$

Comme les débits sont calculés aux faces de chaque maille, le terme $\left(\frac{\partial Q}{\partial t}\right)_{i+\frac{1}{2}}$ est calculé par rapport à ses faces amont et aval, soit $i - \frac{1}{2}$, et $i + \frac{3}{2}$. QM2 s'écrit ainsi :

$$QM2 = 2 \cdot \frac{\theta \cdot Q_{i+\frac{1}{2}}^{n+1} + (1-\theta)Q_{i+\frac{1}{2}}^n}{A_{i+\frac{1}{2}}^n} \cdot \frac{Q_{i+\frac{3}{2}}^n - Q_{i-\frac{1}{2}}^n}{2 \cdot \Delta x_{i+\frac{1}{2}}}$$

$$QM2 = a_{i+\frac{1}{2}}^n \cdot \left(\theta \cdot Q_{i+\frac{1}{2}}^{n+1} + (1-\theta)Q_{i+\frac{1}{2}}^n \right) \quad (3.12)$$

$$QM3 = -\frac{Q_{i+\frac{1}{2}}^n \cdot \left(\theta \cdot Q_{i+\frac{1}{2}}^{n+1} + (1-\theta)Q_{i+\frac{1}{2}}^n \right)}{\frac{1}{4} \cdot (A_{i+1}^n + A_i^n)^2} \cdot \frac{A_{i+1}^n - A_i^n}{\Delta x_{i+\frac{1}{2}}}$$

$$QM3 = b_{i+\frac{1}{2}}^n \cdot \left(\theta \cdot Q_{i+\frac{1}{2}}^{n+1} + (1-\theta)Q_{i+\frac{1}{2}}^n \right) \quad (3.13)$$

$$QM4 = -\frac{g \cdot A_{i+\frac{1}{2}}^n}{\Delta x_{i+\frac{1}{2}}} \cdot \left(\theta \cdot (Z_{i+1}^{n+1} - Z_i^{n+1}) + (1-\theta) \cdot (Z_{i+1}^n - Z_i^n) \right)$$

$$QM4 = c_{i+\frac{1}{2}}^n \cdot \left(\theta \cdot (Z_{i+1}^{n+1} - Z_i^{n+1}) + (1-\theta) \cdot (Z_{i+1}^n - Z_i^n) \right) \quad (3.14)$$

$$QM5 = -\frac{C_f \cdot Q_{i+\frac{1}{2}}^n}{2 \cdot h_{i+\frac{1}{2}}^n \cdot A_{i+\frac{1}{2}}^n} \cdot \left(\theta \cdot Q_{i+\frac{1}{2}}^{n+1} + (1-\theta)Q_{i+\frac{1}{2}}^n \right)$$

$$QM5 = d_{i+\frac{1}{2}}^n \cdot \left(\theta \cdot Q_{i+\frac{1}{2}}^{n+1} + (1-\theta)Q_{i+\frac{1}{2}}^n \right) \quad (3.15)$$

Avec

$$a_{i+\frac{1}{2}}^n = \frac{Q_{i+\frac{3}{2}}^n - Q_{i-\frac{1}{2}}^n}{A_{i+\frac{1}{2}}^n \cdot \Delta x_{i+\frac{1}{2}}}, b_{i+\frac{1}{2}}^n = -\frac{4 \cdot Q_{i+\frac{1}{2}}^n}{\Delta x_{i+\frac{1}{2}}} \cdot \frac{A_{i+1}^n - A_i^n}{(A_{i+1}^n + A_i^n)^2}, c_{i+\frac{1}{2}}^n = -\frac{g \cdot A_{i+\frac{1}{2}}^n}{\Delta x_{i+\frac{1}{2}}}, d_{i+\frac{1}{2}}^n = -\frac{C_f \cdot Q_{i+\frac{1}{2}}^n}{2 \cdot h_{i+\frac{1}{2}}^n \cdot A_{i+\frac{1}{2}}^n}$$

On obtient la discrétisation finale :

$$Q_{i+\frac{1}{2}}^{n+1} - \theta g_{i+\frac{1}{2}}^n Z_{i+1}^{n+1} + \theta g_{i+\frac{1}{2}}^n Z_i^{n+1} = f_{i+\frac{1}{2}}^n Q_{i+\frac{1}{2}}^n + (1 - \theta) g_{i+\frac{1}{2}}^n (Z_{i+1}^n - Z_i^n) \quad (3.16)$$

Avec

$$e_{i+\frac{1}{2}}^N = a_{i+\frac{1}{2}}^N + b_{i+\frac{1}{2}}^N - d_{i+\frac{1}{2}}^N, \quad f_{i+\frac{1}{2}}^N = \frac{1 - \Delta t \cdot (1 - \theta) \cdot e_{i+\frac{1}{2}}^N}{1 + \Delta t \cdot \theta \cdot e_{i+\frac{1}{2}}^N}, \quad g_{i+\frac{1}{2}}^N = \frac{\Delta t \cdot c_{i+\frac{1}{2}}^N}{1 + \Delta t \cdot \theta \cdot e_{i+\frac{1}{2}}^N}$$

Les discrétisations (3.8) et (3.16) sont implémentées dans le code de *libhyd*. Les deux ensembles de variables cote et débit sont résolus simultanément dans un système qui comprend en premier lieu les cotes aux centres, puis les débits aux faces. Par conséquent pour un système de N éléments de calcul, les N premières lignes concernent les cotes au centre des éléments, et les N+p lignes suivantes concernent les débits aux faces. Le système comporte donc 2N+p lignes. p vaut 1 dans un système mono-axial. Par contre à chaque occurrence d'une branche p augmente de 1.

3.2.3 Loi de fonctionnement d'un ouvrage hydraulique

Le débit au droit d'un ouvrage hydraulique est calculé comme la somme des débits vérifiant les lois de fonctionnement des différents constituants de l'ouvrage. Il s'agit d'une relation de face qui est utilisée pour représenter l'équation de débit dans le système linéaire à 2N+P équations mentionnées ci-dessus. Les formules décrivant le fonctionnement des seuils et des sousverse au niveau d'une vanne noyée sont issues de [Even \(1995\)](#), et [Degoutte \(2002\)](#). Elles sont rappelées ci-après.

3.2.3.1 Loi de déversement d'un seuil

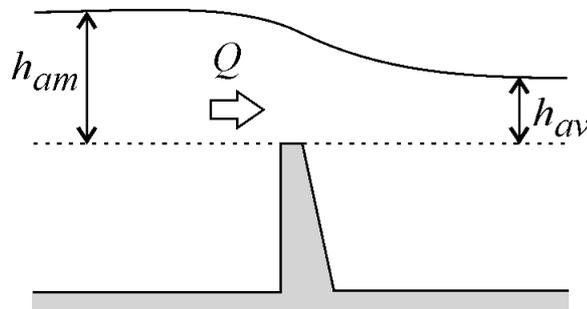


FIGURE 3.3 – Représentation d'un seuil

Si $h_{am} \geq 1.5h_{av}$, alors le seuil est dénoyé. L'écoulement aval n'a alors pas d'influence sur l'écoulement au droit du seuil. La loi de seuil s'écrit :

$$Q = \frac{2}{3} \cdot \mu_d \cdot l \cdot \sqrt{2g} \cdot h_{am}^{1.5} \quad (3.17)$$

La discrétisation suit une formulation explicite :

$$Q^n - \frac{2}{3} \mu_d \cdot l \cdot \sqrt{2 \cdot g \cdot h_{am}^{n-1}} \cdot Z_{am}^n = -\frac{2}{3} \mu_d \cdot l \cdot \sqrt{2 \cdot g \cdot h_{am}^{n-1}} \cdot h_s^n \quad (3.18)$$

Avec,

Q : le débit, [m³.s⁻¹]

μ_d : le coefficient du débit du seuil, 0.6, [-]

l : la largeur du seuil, [m]

h_{am} : la hauteur d'eau par rapport à la hauteur du seuil en amont du seuil, $h_{am} = Z_{am} - h_s$, [m]

Z_{am} : la cote d'eau en amont du seuil, [m]

Si $h_{am} < 1.5h_{av}$, le seuil est noyé. Le tirant d'eau aval influence l'écoulement au droit du seuil. La loi devient :

$$Q = \sqrt{3} \cdot \mu_d \cdot l \cdot h_{av} \cdot \sqrt{2 \cdot g \cdot (h_{am} - h_{av})} \quad (3.19)$$

Afin de discrétiser linéairement l'équation (3.19), on la réécrit comme,

$$Q = \sqrt{3} \cdot \mu_d \cdot l \cdot h_{av} \cdot \sqrt{2 \cdot g} \frac{Z_{am} - Z_{av}}{\sqrt{(h_{am} - h_{av})}} \quad (3.20)$$

Dont la forme de discrétisation explicite est la suivante :

$$Q^n - \sqrt{3} \cdot \mu_d \cdot l \cdot h_{av}^{n-1} \cdot \sqrt{2g} \left(\frac{Z_{am}^n}{(h_{am}^{n-1} - h_{av}^{n-1})} - \frac{Z_{av}^n}{(h_{am}^{n-1} - h_{av}^{n-1})} \right) = 0 \quad (3.21)$$

3.2.3.2 Loi de sousverses d'une vanne

Toutes les vannes sont considérées comme noyées. Le débit au droit de la vanne s'exprime comme suit :

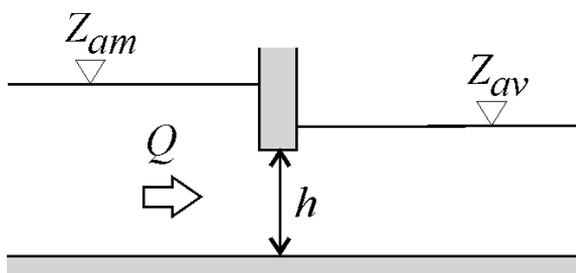


FIGURE 3.4 – Représentation d'une vanne

$$Q = \mu_d \cdot l \cdot h_v \cdot \sqrt{2 \cdot g \cdot (Z_{am} - Z_{av})} \quad (3.22)$$

Avec,

μ_d : le coefficient du débit, 0.6, [-]

h_v : la hauteur de la vanne, [m]

La discrétisation est suivante :

$$Q^n - \mu_d \cdot l \cdot h_{av}^n \cdot \sqrt{2g} \left(\frac{Z_{am}^n}{\sqrt{Z_{am}^{n-1} - Z_{av}^{n-1}}} - \frac{Z_{av}^n}{\sqrt{Z_{am}^{n-1} - Z_{av}^{n-1}}} \right) = 0 \quad (3.23)$$

3.3 Librairie de transport *libttc*

Cette librairie est utilisée après le calcul hydraulique afin de simuler l'advection et la dispersion des substances. L'équation de transport utilisée est issue de [de Marsily \(1981\)](#).

$$\omega \frac{\partial C}{\partial t} = \text{div}(D \overrightarrow{\text{grad}} C - C \vec{U}) \quad (3.24)$$

Avec,

ω : la porosité du milieu, ici $\omega = 1$ pour la rivière, [-]

C : la concentration volumique de chaque substance, [mmol.m³]

D : le coefficient de dispersion, [m.s⁻¹]

\vec{U} : la vitesse de Darcy, ici la vitesse de l'eau, [m.s⁻¹]

On peut exprimer l'équation (3.24) en débit pour une dimension,

$$\frac{\partial(AC)}{\partial t} + \frac{\partial(QC)}{\partial x} = AD \frac{\partial^2 C}{\partial x^2} + F_x \quad (3.25)$$

Avec,

A : la surface mouillée au centre de la maille, [m²]

Q : le débit aux faces de la maille, [m³.s⁻¹]

F_x : le flux d'apport ou le flux de prélèvement longitudinal, [mmol.m⁻¹.s⁻¹]

3.3.1 Discrétisation numérique

La discrétisation se fait en temps et en espace.

$$\begin{aligned} & \frac{A_i^n C_i^n - A_i^{n-1} C_i^{n-1}}{\Delta t} + \theta \frac{Q_i^n C_i^n - Q_{i-1}^n C_{i-1}^n}{\Delta x} + (1-\theta) \frac{Q_i^{n-1} C_i^{n-1} - Q_{i-1}^{n-1} C_{i-1}^{n-1}}{\Delta x} \\ = & D(A_i^n \theta \frac{C_{i+1}^n - 2C_i^n + C_{i-1}^n}{\Delta x^2} + A_i^{n-1} (1-\theta) \frac{C_{i+1}^{n-1} - 2C_i^{n-1} + C_{i-1}^{n-1}}{\Delta x^2}) + \theta F_x^n + (1-\theta) F_x^{n-1} \end{aligned} \quad (3.26)$$

3.3.2 La réoxygénation aux barrages

Cette partie s'inspire de la thèse de [Even \(1995\)](#) et ne présente que la discrétisation numérique. L'équation (3.26) sous forme implicite pour l'advection pure est la suivante pour un barrage situé à la face $i - \frac{1}{2}$:

$$\frac{A_i^n C_i^n - A_i^{n-1} C_i^{n-1}}{\Delta t} + \frac{Q_i^n C_i^n - Q_{i-1}^{n-1} C_{i-1}^{n-1}}{\Delta x} = F_x^n \quad (3.27)$$

Avec,

i : l'indice de la maille aval du barrage

$i - 1$: l'indice de la maille amont du barrage

Il est possible d'exprimer le terme $Q_{i-1}^n C_{i-1}^n$ en prenant en compte la réaération par $Q_{i-1}^n (C_{i-1}^n + \delta^n)$. Le terme source δ^n correspond à la partie de réoxygénation. L'oxygénation est généralement exprimée à l'aide d'un paramètre exprimant le rapport des déficits en oxygène amont et aval,

$$r = \frac{[O_2]^{sat}(T) - [O_2]_{am}}{[O_2]^{sat}(T) - [O_2]_{av}} \quad (3.28)$$

En exprimant $[O_2]_{av}$ par $C_{i-1}^n + \delta^n$, on trouve :

$$\delta^n = \frac{r-1}{r} ([O_2]^{sat}(T) - C_{i-1}^n) \quad (3.29)$$

Une distinction est faite en écoulement noyé et dénoyé. La formule de Holler donne la valeur théorique en écoulement dénoyé (Even, 1995) :

$$r_d = 1 + 0.21H \quad (3.30)$$

Pour un écoulement noyé, on utilise la formule d'Avery et Novack (Even, 1995).

$$r_n = 1 + A \cdot Fr_{av}^{0.21} \cdot Re^{0.75} \quad (3.31)$$

Avec,

A : 1.0043E-06

Fr : le nombre de Froude, $Fr = \frac{u^2}{gL_c}$, [-]

u : la vitesse de l'eau, [m.s⁻¹]

L_c : la longueur caractéristique, [m]

Re : le nombre de Reynolds, $Re = \frac{uL}{\nu}$

ν : la viscosité cinématique de l'eau, [m².s⁻¹]

Il distinguer les parts du débit noyées (Q_{i-1}^{nns}) et dénoyées (Q_{i-1}^{nds}) en surverse :

$$Q_{i-1}^n = Q_{i-1}^{nds} + Q_{i-1}^{nns} + Q_{i-1}^{nsous} \quad (3.32)$$

Avec Q_{i-1}^{nsous} le débit en sousverse, [m³.s⁻¹]

Le terme $Q_{i-1}^n C_{i-1}^n$ en tenant compte de la réoxygénation devient :

$$Q_{i-1}^{ns}(C_{i-1}^n + \delta^n) + Q_{i-1}^{nsous} C_{i-1}^n = Q_{i-1}^n C_{i-1}^n + Q_{i-1}^{nds} \delta^{nds} + Q_{i-1}^{nns} \delta^{nns} \quad (3.33)$$

Avec,

Q_{i-1}^{ns} : le débit en surverse estimé à partir des lois de déversement aux barrages, [m³.s⁻¹]

δ^{nds} : la partie de réoxygénation liée au débit dénoyé en susverse, [mmol.m⁻³]

δ^{nns} : la partie de réoxygénation liée au débit noyé en susverse, [mmol.m⁻³]

En intégrant l'équation (3.33) dans l'équation (3.27), on trouve :

$$\begin{aligned} \frac{A_i^n C_i^n - A_i^n C_i^{n-1}}{\Delta t} + \frac{Q_i^n C_i^n - Q_{i-1}^n C_{i-1}^n + Q_{i-1}^{ns} \cdot coef \cdot C_{i-1}^n}{\Delta x} \\ = F_x^n + \frac{Q_{i-1}^{ns} \cdot coef \cdot [O_2]^{sat}(T)}{\Delta x} \end{aligned} \quad (3.34)$$

Avec,

$$coef = frac_{debit} \cdot \frac{r_d - 1}{r_d} + (1 - frac_{debit}) \cdot \frac{r_n - 1}{r_n}, [-]$$

$frac_{debit}$: la fraction du débit dénoyé en surverse, [-]

3.4 Librairie biogéochimique *librive* (C-RIVE)

Cette librairie est appelée après le calcul de transport. La librairie *librive* est la version en librairie de C-RIVE. (Vilmin et al., 2012). C-RIVE est programmé en langage C, basée sur le modèle RIVE (<https://www.fire.upmc.fr/rive>). La résolution du problème biogéochimique se fait par un schéma explicite ou un schéma de Runge-Kutta d'ordre 4 (par défaut). Les formulations mathématiques liées à la simulation de l'oxygène dissous est détaillées dans la section 4.2.2 (p.4.2.2). Une formulation à base physique du coefficient de réaération intégrant l'influence de la température de l'eau est apportée dans ce travail.

3.4.1 Reformulation à base physique du coefficient de réaération

Le coefficient de réaération K_{rea} est calculé à partir d'une formule empirique développé par Thibodeaux et al. (1994).

$$K_{rea} = \sqrt{\frac{D_m \times V_{wa}}{h}} + (K_{wind} \times V_{wind}^{2.23} \times (D_m \times 10000)^{\frac{2}{3}} + K_{navig}) \quad (3.35)$$

Avec,

D_m : coefficient de diffusion moléculaire à l'interface eau-sédiment [m²/s]

V_{wa} : vitesse de l'écoulement [m/s]

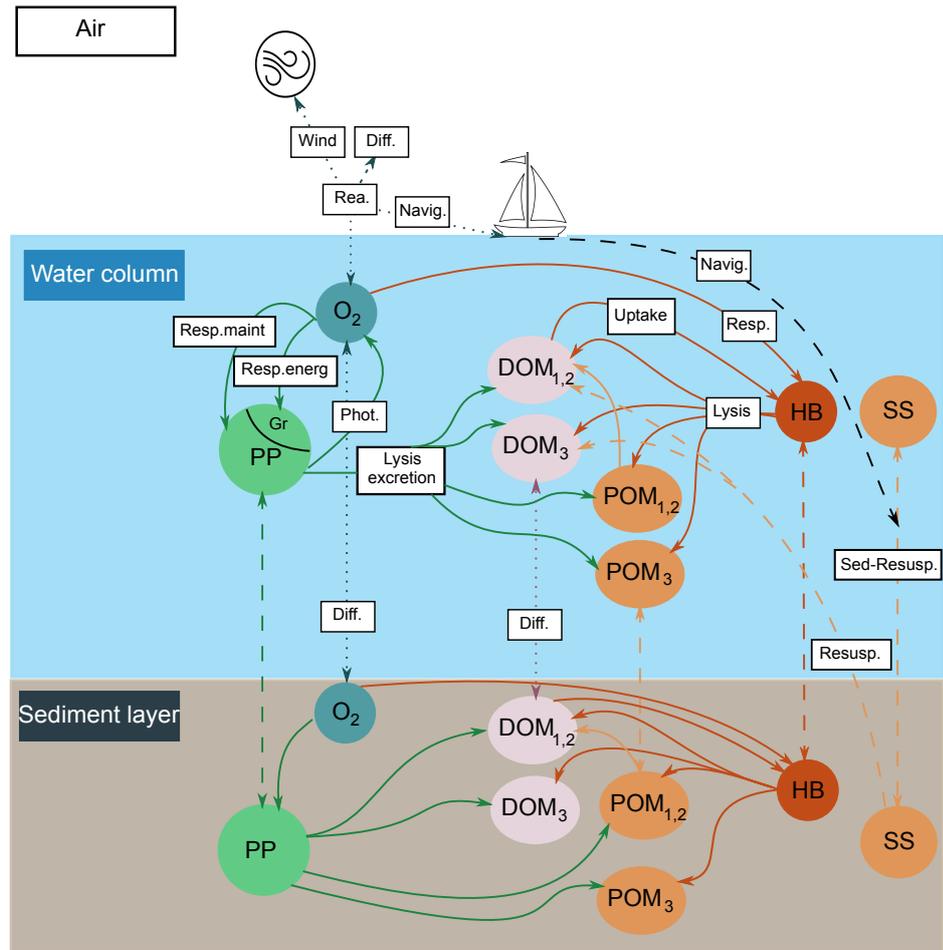


FIGURE 3.5 – Schéma synthétique de C-RIVE (Wang et al., 2018b), modifié à partir de Vilmin et al. (2018)

K_{wind} : coefficient de réaération liée à la vitesse du vent [m/s]

V_{wind} : vitesse du vent [m/s]

K_{navig} : coefficient de réaération lié à la navigation [m/s]

Dans PROSE3.6.9, D_m est prédéfini par l'utilisateur. En revanche, il est calculé avec la formule proposée par [Wilke and Chang \(1955\)](#) en fonction de la température dans PROSE-P.

$$D_m = 7.4 \times 10^{-8} \frac{T(\psi_{H_2O} M_{H_2O})^{\frac{1}{2}}}{\mu V_{O_2}^{0.6}} \quad (3.36)$$

Avec,

T : température en Kelvin [K]

ψ_{H_2O} : paramètre d'association pour l'eau 2.26, [-], ([Reid et al., 1977](#))

M_{H_2O} : masse molaire de l'eau, 18 [g/mol]

μ : viscosité de l'eau [cP]

V_{O_2} : volume molaire de l'oxygène, 25.6 cm³/mol

3.5 Parallélisation des calculs de transport et de biogéochimie

Dans PROSE-P, les résolutions de l'équation de transport de chaque substance sont indépendantes les unes des autres. Supposons que l'on modélise le devenir de n espèces. Il est possible de distribuer ces résolutions sur N processeurs, chaque processeur traite alors $\frac{n}{N}$ espèces, soit un temps de calcul potentiellement réduit d'un facteur N pour la résolution du problème de transport (Fig. 3.6). De même, le calcul biogéochimique se fait au sein d'une maille de discrétisation, il est possible de programmer un calcul en parallèle par maille (Fig. 3.6). Chaque processeur traite $\frac{m}{N}$ mailles (m : le nombre de mailles, N : le nombre de processeurs utilisé).

Nous nous intéressons maintenant aux performances du code parallélisé, PROSE-P. Si on renomme T_{hyd} (Fig. 3.6) en T_{np} pour désigner le temps de calcul de la partie non parallélisée (la lecture des données, le calcul hydraulique, le calcul des coefficients matriciels pour le transport, le passages des données entres les bibliothèques, l'impression des sorties etc.), le gain théorique se calcule par :

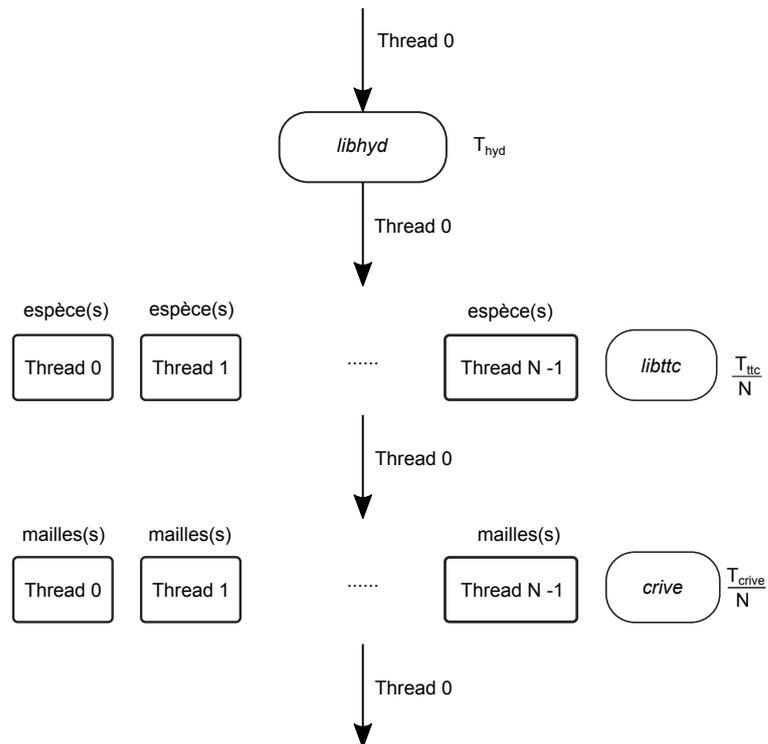


FIGURE 3.6 – Schéma synthétique de calcul en parallèle dans PROSE-P (T_{hyd} : le temps de calcul pour hydraulique; T_{ttc} : le temps de calcul pour le transport sans parallélisation; T_{crive} : le temps de calcul pour la biogéochimie sans parallélisation. N : le nombre de processeurs; Thread 0 : le processeur principal.)

$$\begin{aligned}
gain &= \frac{T_{tot,np}}{T_{tot,p}} \\
&= \frac{T_{tot,np}}{\frac{T_p}{N_{proc}} + T_{np}} \\
&= \frac{N_{proc} \cdot T_{tot,np}}{T_p + N_{proc} \cdot T_{np}} \\
&= \frac{N_{proc} \cdot T_{tot,np}}{T_{tot,np} + (N_{proc} - 1) \cdot T_{np}} \\
&= \frac{N_{proc}}{1 + (N_{proc} - 1)p} \tag{3.37}
\end{aligned}$$

avec,

$T_{tot,np}$: le temps de calcul avec un processeur, $T_{tot,np} = T_p + T_{np}$

$T_{tot,p}$: le temps de calcul avec N_{proc} processeurs

N_{proc} : le nombre de processeurs utilisé

p : la proportion du temps de calcul pour la partie non parallélisée (T_{np}),

$$p = \frac{T_{np}}{T_{tot,np}}$$

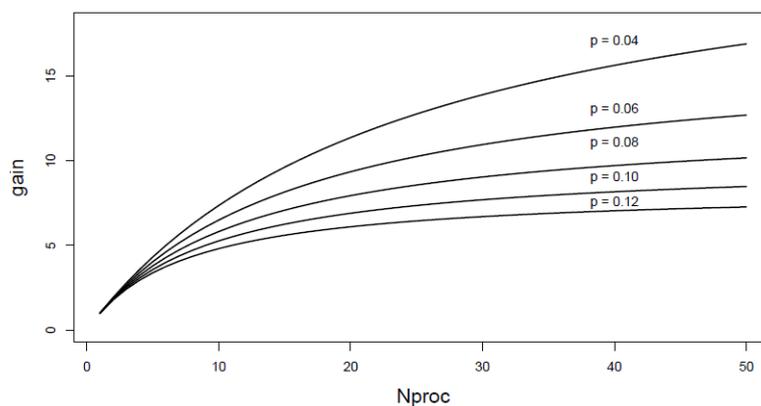


FIGURE 3.7 – Performance théorique de la parallélisation, p : la proportion du temps de calcul pour la partie non parallélisée

La performance théorique de la parallélisation en fonction du nombre de processeurs est tracé avec de différentes de valeurs p (Fig. 3.7).

3.6 Performances de ProSe-P

Les performances de PROSE-P sont évaluées d'abord sur un cas test de barrage, puisque les débogages du code *libhyd* ont été menés. Ensuite, une comparaison des résultats d'une simulation de modèle Seine entre PROSE-P et la dernière version de PROSE, PROSE3.6.9, est réalisée.

3.6.1 Cas test de barrage

Un cas test de barrage avec deux biefs a été construit pour comparer les résultats de PROSE-P en régime permanent avec ceux de HEC-RAS5.0.3 (Brunner, 2016) et de PROSE3.6.9 (Vilmin, 2014).

Les deux biefs mesurent 40 km avec une section rectangulaire de largeur de 10 m. La pente du fond est 0.001 au sein des biefs. Un seuil de largeur de 10 m avec un coefficient de débit (Eq. (3.17)) de 0.6 se positionne au milieu du domaine. Le coefficient de Strickler est 20 pour les deux biefs. Le débit imposé en amont du domaine est $28.92 \text{ m}^3 \cdot \text{s}^{-1}$ qui donne une hauteur d'eau théorique de 3 m. La charge imposée en aval est la hauteur théorique (3 m), ce qui implique que la surface libre sans barrage est parallèle à la cote du fond. Selon les résultats obtenus (Fig. 3.8), PROSE-P et PROSE3.6.9 aboutissent à des résultats très proches, qui correspondent également à ceux de HEC-RAS5.0.3. Néanmoins, il existe une différence ténue entre PROSE-P et PROSE3.6.9 à l'aval des biefs, où PROSE-P affiche des performances plus proches de celles d'HEC-RAS. Le travail de parallélisation du code a ainsi permis de légèrement améliorer les performances globales du code.

3.6.2 Une simulation du modèle Seine

Le modèle représente les 220 km de Seine et Marne depuis l'amont de l'agglomération parisienne (Choisy et Neuilly) jusqu'à l'estuaire de la Seine (Poses) soumis à de forts rejets anthropiques. Il s'agit du domaine simulé à l'échelle pluri-annuelle par Vilmin (2014). Le domaine est découpé en 959 mailles pour PROSE-P et 958 mailles pour PROSE3.6.9. La maille de différence est liée à la stratégie de remaillage qui diffère dans les deux codes. La période simulée est l'année de 2007 (365 jours). On compare les résultats (les débits et les concentrations en oxygène dissous) simulées par les deux codes. Les performances de PROSE-P seront évaluées par des critères statistiques.

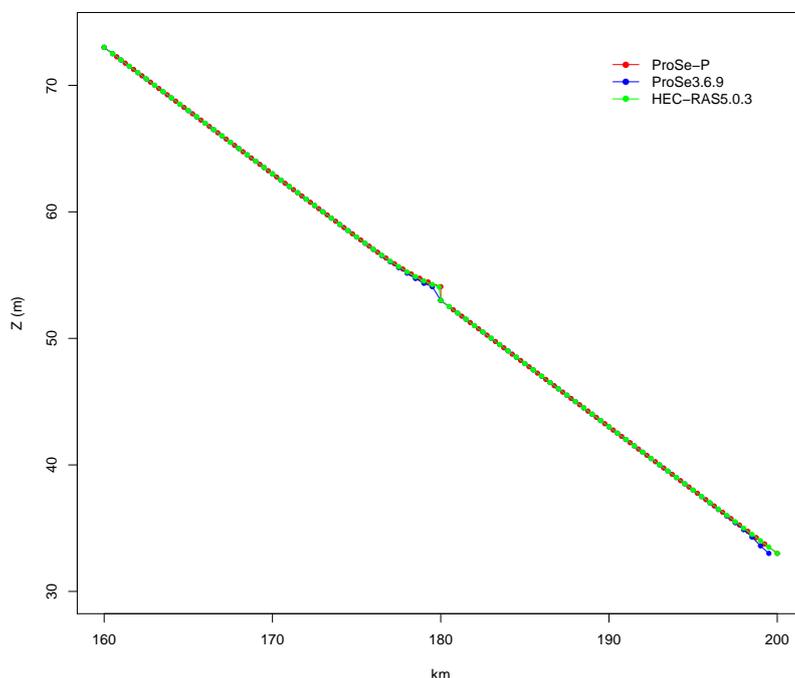


FIGURE 3.8 – Comparaison d’un cas test de barrage en régime permanent

3.6.2.1 Performance de calcul en parallèle pour ProSe-P

On analyse tout d’abord la performance du calcul en parallèle. Pour cela, on réalise deux simulations par PROSE-P, une avec un processeur, et l’autre avec 10 processeurs (Tab. 3.1).

TABLE 3.1 – Performance de calcul en parallèle

Code	N_{maille}	N_{proc}	$dt(min)$	$T(j)$	$t(min)$	gain
PROSE-P	959	10	10	365	23	4.7
PROSE-P	959	1	10	365	108	-

N_{maille} : le nombre de mailles.

dt : le pas de temps de simulation.

T : la période de simulation.

t : le temps de calcul.

D’après la simulation réalisée avec un processeur, p vaut 0.08 (Eq. (3.37)), soit 8% du $T_{tot,np}$. En théorie, on aurait un gain de 5.8 (Fig. 3.7). Pourtant, le gain réel est 4.7, ce qui correspond au gain de 7 processeurs. L’accès au

disque dur (physique) par d'autres utilisateurs pourrait être une explication de cet écart entre gain réel et gain théorique.

Il est alors intéressant de comparer les performances de PROSE-P avec celles de la version séquentielle de PROSE3.6.9, validée en fin de phase VI du PIREN-SEINE, soit la version 3.6.9 (Vilmin, 2014; Laborie et al., 2016).

TABLE 3.2 – Performance de calcul des différentes versions

Code	N_{maille}	N_{proc}	$dt(min)$	$T(j)$	$t(min)$	gain
PROSE-P	959	10	10	365	23	2.1
PROSE3.6.9	958	1	10	365	49	-

Le gain de PROSE-P avec 10 processeurs est de l'ordre 2 (Tab. 3.2) par rapport au PROSE. Il est un peu loin de nos attentes théoriques (4.7). Puisque PROSE-P avec un processeur est deux fois plus lent que PROSE3.6.9. (Tab. 3.1 p.73 et Tab. 3.2 p.74). Ceci est lié au calcul biogéochimique (*librive*). Son optimisation fera l'objet de travaux futurs.

TABLE 3.3 – Répartition des temps de calcul des différentes parties de PROSE-P

$T_{tot}(s)$	$T_{lecture}(s)$	$T_{hyd}(s)$	$T_{ttc}(s)$	$T_{rive}(s)$	$T_{passage}(s)$	$T_{output}(s)$
1387	1.89	240.14	371.84	500.35	245.51	27.24
100%	0.14%	17.32%	26.81%	36.07%	17.70%	1.96%

En regardant les différentes parties de temps de calcul de PROSE-P, la partie biologique et le temps de passages des données (transfert de concentrations, débit, vitesse de l'eau, hauteur d'eau etc. au *librive*, *libttc*) sont coûteux (> 50%).

3.6.2.2 Critères statistiques utilisés pour évaluer les performances de ProSe-P

Afin d'évaluer les performances de PROSE-P, deux critères statistiques sont calculés : RMSE (*Root Mean Square Error*, Eq.(3.38)) et KGE (*Kling-Gupta efficiency*, Eq.(3.39)). KGE est un critère proposé par Gupta et al. (2009), puis modifié en 2012 (Kling et al., 2012) pour évaluer les perfor-

mances des modèles pluie-débit en hydrologie. Il se base sur la décomposition du coefficient NSE (*Nash-Sutcliffe efficiency*) et MSE (*Mean Square Error*). Ce coefficient comprend finalement trois composantes : le coefficient de corrélation (r), le biais du modèle (α) et la variabilité (γ) (Kling et al., 2012). Les simulations sont d'autant meilleures que le critère KGE et ses composantes se rapprochent de 1.

$$RMSE = \sqrt{\frac{\sum(sim - obs)^2}{N}} \quad (3.38)$$

$$KGE = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (3.39)$$

r : coefficient de corrélation

$$\beta = \frac{\mu_{sim}}{\mu_{obs}}$$

$$\gamma = \frac{CV_{sim}}{CV_{obs}} = \frac{\sigma_{sim}/\mu_{sim}}{\sigma_{obs}/\mu_{obs}}$$

Avec,

μ : moyenne

σ : écart-type

CV : coefficient de variation

3.6.2.3 Validation de ProSe-P sur le calcul hydraulique (Q)

Les débits sont bien simulés par PROSE-P (Fig. 3.9 et Fig. 3.10) en comparant avec ceux simulés par PROSE3.6.9. Les simulations de PROSE3.6.9 sont utilisées comme observations dans les équations de RMSE (Eq.(3.38)) et KGE (Eq. (3.39)). La RMSE (Root Mean Square Error) maximale peut être obtenue à la station Méricourt avec une valeur de 7.61 m³/s (Tab. 3.4). Les critères KGEs sont tous supérieurs à 0.98, ce qui indique une cohérence quasiment parfaite entre les simulations de PROSE-P et celles de PROSE3.6.9. Le bilan de masse est bien conservé sur l'ensemble du domaine simulé. Mais les débits simulés par PROSE3.6.9 présentent des instabilités à chaque démarrage d'un nouveau pas de temps (pics régulièrement simulés, Fig. 3.11), ce qui conduit une différence absolue maximale de 118.06 m³/s à la station Suresnes. PROSE-P est ainsi plus stable, ce qui indique de

TABLE 3.4 – Critères statistiques des débits simulés par PROSE-P et PROSE-PA3.6.9

Stations	RMSE	KGE	r	β	γ
Austerlitz	1.06	0.9998	1.0000	1.0000	0.9998
Suresnes	4.73	0.9928	0.9997	1.0045	0.9944
Chatou	4.87	0.9915	0.9993	1.0070	0.9952
Bougival	2.61	0.9892	0.9993	0.9993	0.9892
Sartrouville	6.38	0.9919	0.9995	1.0045	0.9933
Andrésy	3.90	0.9975	0.9997	1.0074	0.9990
Triel	6.16	0.9943	0.9997	1.0031	0.9953
Méricourt	7.61	0.9942	0.9996	1.0030	0.9950

meilleures performances pour la résolution du problème hydraulique.

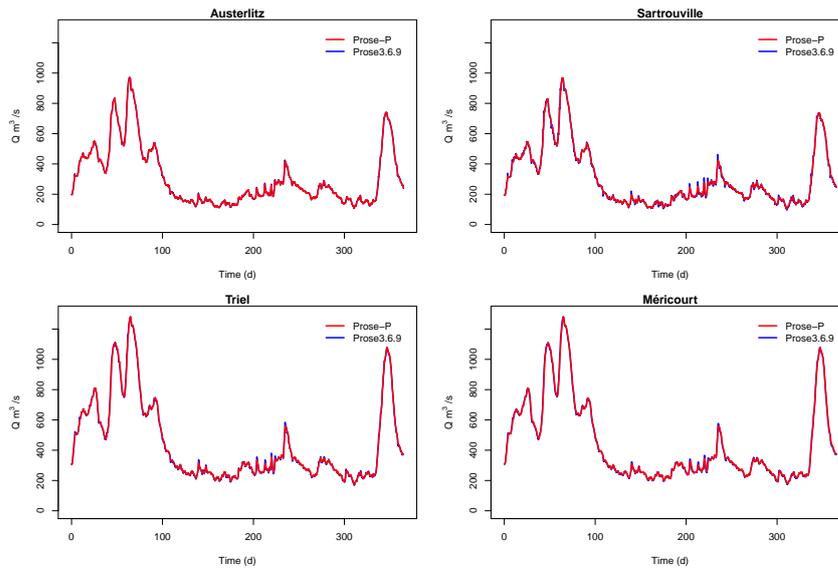


FIGURE 3.9 – Les débits simulés par PROSE-P et PROSE3.6.9 pour l'année 2007

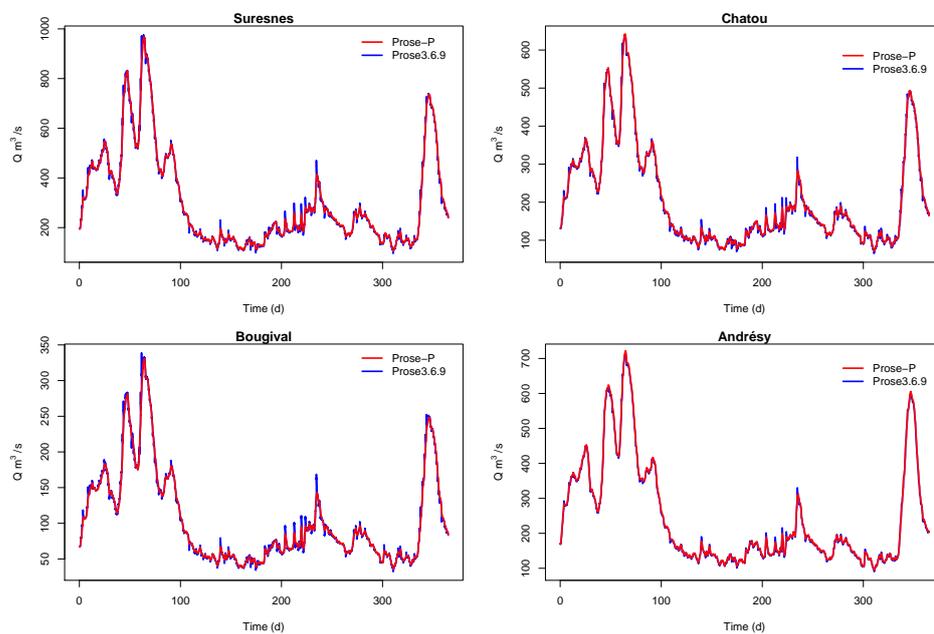


FIGURE 3.10 – Les débits simulés sur les branches par PROSE-P et PROSE3.6.9 pour l'année 2007

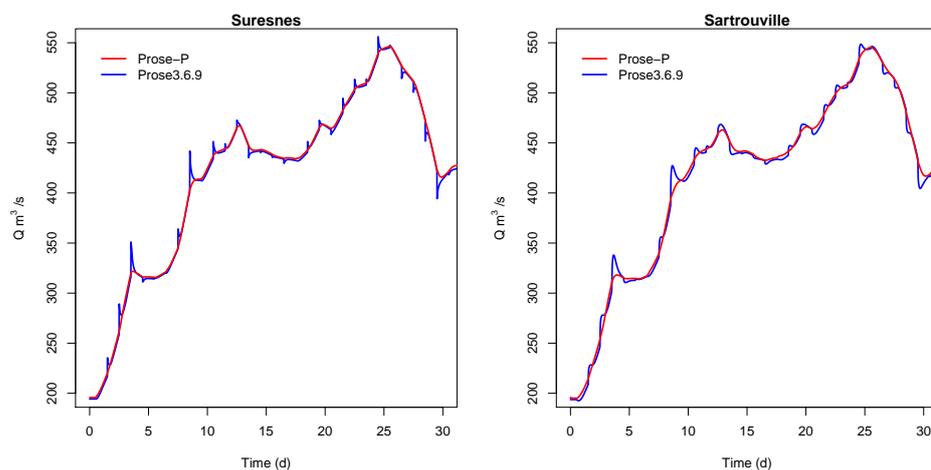


FIGURE 3.11 – Le débit simulé des premiers 30 jours à Suresnes et Sartrouville pour l'année 2007

3.6.2.4 Validation de ProSe-P sur le calcul biogéochimique (O₂)

TABLE 3.5 – Critères statistiques des concentrations en O₂ simulées par les deux codes

2007 Stations	PROSE-P					PROSE3.6.9				
	RMSE mgO ₂ /L	KGE	r	β	γ	RMSE mgO ₂ /L	KGE	r	β	γ
Suresnes	0.55	0.95	0.97	0.98	0.97	0.52	0.91	0.97	0.99	0.92
Chatou	0.92	0.72	0.96	1.02	0.72	1.02	0.67	0.96	1.04	0.68
Bougival	0.73	0.78	0.96	1.01	0.79	0.77	0.74	0.96	1.02	0.74
Sartrouville	0.97	0.48	0.65	1.06	0.62	1.02	0.52	0.68	1.08	0.65
Andrésy	0.69	0.78	0.96	1.03	0.78	0.78	0.74	0.96	1.05	0.75
Méricourt	0.90	0.68	0.95	1.03	0.69	0.90	0.70	0.94	1.04	0.71

Les RMSEs sont améliorées pour les simulations de PROSE-P, c’est notamment le cas aux stations Chatou, Sartrouville et Andrésy (Tab. 3.5). Globalement, les tendances des concentrations de l’oxygène dissous simulées par les deux codes sont bien cohérentes avec les observations. Les coefficients de corrélations sont très proches de 1 (r), sauf pour la station Sartrouville. Cependant, les simulations de PROSE-P comportent moins de biais que celles de PROSE3.6.9 (Tab. 3.5). Le biais maximal des simulation de PROSE-P est environ 6% à la station Sartrouville tandis qu’il est de 8% pour celles de PROSE3.6.9. Les deux codes ont tendance à surestimer les concentrations en oxygène dissous ($\beta > 1$ globalement), c’est notamment le cas en période de basses eaux. Cette surestimation est déjà soulignée par [Vilmin \(2014\)](#) pour PROSE3.6.9. Néanmoins, ces écarts sont légèrement atténués avec PROSE-P (Fig. 3.14). Les améliorations sont liées à la reformulation de la réaération aux barrages qui considère un rôle passif des vannes, et à la prise en compte de la température dans le calcul des échanges air-eau ([Wilke and Chang, 1955](#)), ce qui n’était pas le cas dans PROSE3.6.9. Il nécessite cependant de re-calibrer PROSE-P.

Conclusion

En conclusion, une nouvelle version parallélisée du code PROSE a été construite en se basant sur trois librairies C : *libhyd*, *libttc*, et *librive*.

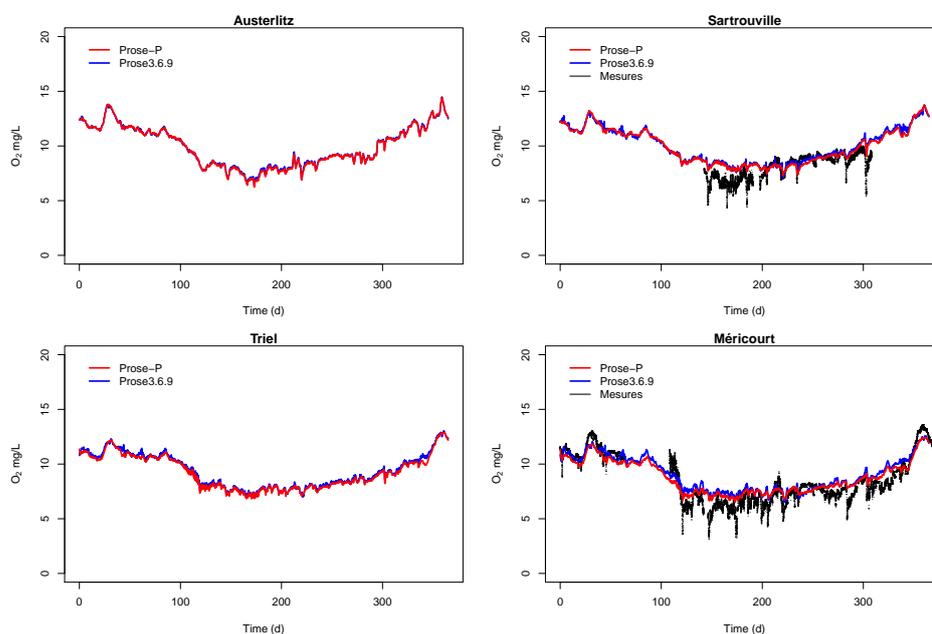


FIGURE 3.12 – Les concentrations de l’oxygène dissous simulées par PROSE-P et PROSE3.6.9 pour l’année 2007

Cette nouvelle version, appelée PROSE-P, est maintenant opérationnelle et affiche des performances améliorées par rapport à l’ancienne version séquentielle du code, PROSE3.6.9. L’amélioration des performances comprend, une réduction des temps de calcul se basant sur la parallélisation qui permet d’exploiter les nouvelles capacités des machines multi-processeurs, une plus grande robustesse du code hydraulique, et une meilleure représentation des échanges eau-air intégrant l’influence de la température. Il s’agit d’une étape importante dans le développement du code ouvrant la porte à un pan nouveau de recherche autour de l’intégration des données de suivi haute fréquence dans la caractérisation de l’état de la Seine.

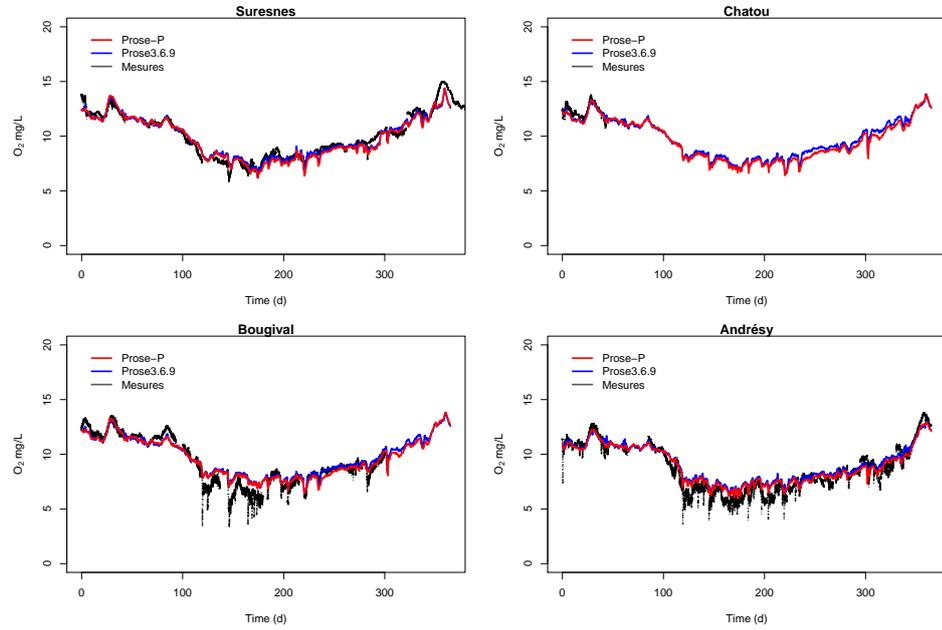


FIGURE 3.13 – Les concentrations de l'oxygène dissous simulées par PROSE-P et PROSEpour l'année 2007

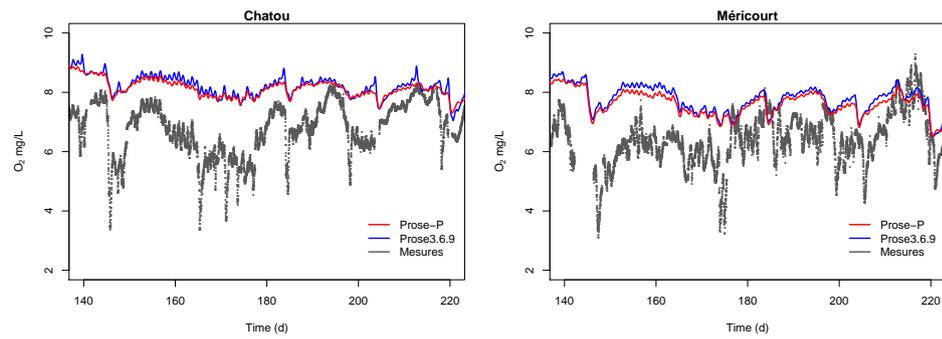


FIGURE 3.14 – Les concentrations de l'oxygène dissous simulées pendant l'été à Chatou et Méricourt

Deuxième partie

Etude de sensibilité des
paramètres contrôlant
l'oxygène

Chapitre 4

Etude de sensibilité du modèle biogéochimique C-RIVE

Résumé

Les modèles de qualité de l'eau simulant de nombreux processus biogéochimiques, sont souvent décrits par des équations mécanistes à base physique. Ces équations physiques nécessitent de définir la valeur de paramètres comme la croissance et la mortalité des micro-organismes (phytoplancton, bactéries hétérotrophes). Avant d'implémenter le schéma d'assimilation de données dans PROSE-P, il faut identifier les paramètres auxquels la variation de concentration en O_2 est sensible.

Ce chapitre présente une analyse de sensibilité des paramètres du modèle biogéochimique C-RIVE en utilisant la méthode de Morris et les indices de Sobol ([Wang et al., 2018b](#)). La méthode de la décomposition en fonctions empiriques orthogonales (EOF) est appliquée pour réduire la dimension des sorties du modèle (Fig. 4.1). L'objectif de cette analyse est d'identifier les paramètres les plus influents en fonction des contextes hydrologiques et trophiques du système Seine. Les résultats indiquent que la respiration de maintenance du phytoplancton et les paramètres photosynthétiques (i.e. la capacité photosynthétique, le taux de photosynthèse maximal, et les termes d'extinction lumineuse) sont les plus influents pendant les blooms algaux. Quand le système de la rivière devient hétérotrophe, l'oxygène dissous est

plus sensible aux paramètres bactériens (moyenne et haute température de l'eau) et aux coefficients de réaération (basse température de l'eau). L'effet de la navigation sur la variation de concentration en O_2 est souligné. L'impact de la température de l'eau sur les résultats d'analyse de sensibilité est également étudié.

Finalement, ce travail préliminaire à l'assimilation de données a permis de sélectionner les paramètres les plus influents du modèle. Le schéma d'assimilation de données implémenté par la suite permet de modifier la valeur de ces paramètres au cours du temps en fonction des teneurs en oxygène mesurées dans le milieu.

Abstract

Dissolved oxygen within water column is a key variable to characterize the water quality. Water quality modeling has been extensively developed for decades. However, complex biogeochemical cycles are described using a high number of parameters. Hence, parameters' uncertainty constitutes a major problem in the application of these models. Sensitivity analysis allows the identification of the most influential parameters in a model and a better understanding of the governing processes. This paper presents a time-dependent sensitivity analysis for dissolved oxygen using Morris and Sobol methods combined with an empirical orthogonal function (EOF) analysis for dimension reduction. The aim of this study is to identify the most important parameters of C-RIVE model in different trophic contexts and to understand the biogeochemical functioning of river systems. The results indicate that the maintenance respiration of phytoplankton and the photosynthetic parameters (i.e. photosynthetic capacity, the maximal photosynthesis rate and light extinction coefficients) are the most influential parameters during algal blooms. When the river system becomes heterotrophic, the bacterial activities (moderate and high temperature) and the reaeration coefficients (low temperature) affect the most the dissolved oxygen concentration in the water column. An anthropogenic effect (ship navigation) on variation of dissolved oxygen concentration has been identified and the role of this anthropogenic effect evolves with hydrological and trophic conditions.

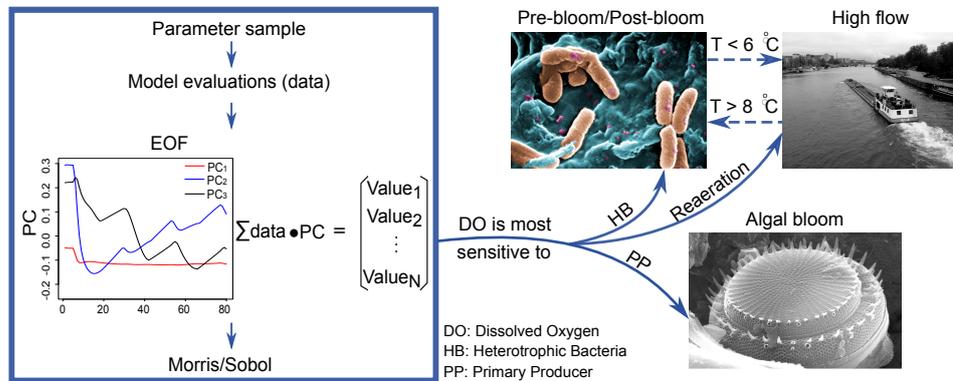


FIGURE 4.1 – Méthodologie de l'analyse de sensibilité du module biogéochimique C-RIVE (Wang et al., 2018b)

4.1 Introduction

Dissolved oxygen (DO) within water column is one of the most important variable to characterize the metabolism of a river system (Odum, 1956; Needoba et al., 2012; Escoffier et al., 2018) and its concentration in a river system is frequently considered as a good indicator of the water quality (Sánchez et al., 2007; Zahraeifard and Deng, 2012; Bailey and Ahmadi, 2014). In addition to massive fish mortality (Boët et al., 1999), low DO concentrations caused by heterotrophic respiration (dramatic degradation of organic matters, e.g., downstream wastewater treatment plant) engender smelling and taste problems (Bailey and Ahmadi, 2014).

The first water quality model was developed by Streeter and Phelps (1925). It is based on DO for understanding the impact of organic matters rejection on water quality of a stream. Then, more complex biogeochemical cycles, such as carbon cycle and nutrients' cycles were considered in the water quality modeling over the past decades (Billen et al., 1994; Garnier et al., 1995; Reichert et al., 2001b; Cox, 2003a; Flipo, 2005; Vilmin, 2014). DO concentration within water column results from multiple processes such as respiration, photosynthesis, sediment oxygen demand and reaeration at air-water interface (Cox, 2003b; Escoffier et al., 2018; Vilmin et al., 2016; Thibodeaux et al., 1994). These processes have been implemented in current water quality modeling. However, the complex biogeochemical cycles are described using a high number of parameters (Flipo, 2013). The high

number of parameters leads to non uniqueness issue of parameter identification, namely equifinality concept (Beven, 1989), which means that different processes can produce a same or similar effect (Ebel and Loague, 2006). Therefore, parameter estimation and model validation remain very difficult (Arhonditsis and Brett, 2004). Hence, parameters' uncertainty constitutes a major problem in the validation and extrapolation capabilities of models (Beven, 1989; Polus et al., 2011).

Sensitivity analysis (SA) is identified to be a necessary step in the model building process (Jakeman et al., 2006; Campolongo et al., 2007). SA allows the identification of the most influential parameters in a model and a better understanding of the governing processes, which permits the reduction of parameters' dimensionality in the optimization (Saltelli et al., 2000). The non-influential parameters are set to a fixed value (factor fixing) and the other parameters should be better estimated for the reduction of model output uncertainty (factor prioritisation) (Saltelli et al., 2004; Gamerith et al., 2013).

Sensitivity analysis applications can be found for hydrological modeling (Hall et al., 2005; Reusser et al., 2011; Zhai et al., 2014; Garambois et al., 2013), water quality modeling (Reichert and Vanrolleghem, 2001; Nossent et al., 2011; Sun et al., 2012; Gamerith et al., 2013; Cho et al., 2017) and particularly for DO (Bierman et al., 1994; Deflandre et al., 2006; Bailey and Ahmadi, 2014; Cox and Whitehead, 2005). Some SA methods used by these authors are shown in table 4.1. Reusser et al. (2011) summarized also SA methods used in hydrology and water quality modeling in literature. These studies related to DO identified the importance of many processes : algal processes, sediment oxygen demand and oxygen reaeration at air-water interface. However, the sensitivity analysis of water quality model in contrasted hydrological and trophic contexts has not been investigated yet.

In this paper, a synthetic river system representing the Seine is simulated by a biogeochemical model, C-RIVE, for 80 hours with a time step of 15 min. Two global sensitivity analysis (GSA) methods are applied to C-RIVE model, in different flow and biogeochemical conditions. The screening method proposed by Morris (1991) and improved by Campolongo et al. (2007) is widely applied in literature due to its low computational cost. However, since the Morris method is qualitative, it is usually used to reduce the number of analyzed parameters in order to apply another more subtle

TABLE 4.1 – Sensitivity analysis studies in hydrological and water quality modeling

Study	Model	Method	Evaluated model output
Hall and Tank (2005)	Lisflood-FP	Sobol method	Inundation performance measure
Reusser et al. (2011)	TOPMODEL, WaSiM-ETH	FAST, EFAST, Sobol method	Fast Fourier transformation
Zhai et al. (2014)	SWAT, QUAL2E	Morris, OAT	Nash-Sutcliffe coefficient
Nossent et al. (2011)	SWAT	Sobol method	Nash-Sutcliffe coefficient
Sun et al. (2012)	IHACRES	OAT local SA, Morris method, RSA	Nash-Sutcliffe coefficient
Gamerith et al. (2013)	SMUSI	Standardised regression coefficients (SRCs), Morris method	Total annual over-flow, peak discharge
Cho et al. (2017)	SWAT, QUAL2E	Relative sensitivity (Sr)	Water quality concentrations
Deflandre et al. (2006)	QUESTOR	EFAST	Nash-Sutcliffe efficiency
Bailey and Ahmadi (2014)	QUAL2E	Morris method	Mean concentrations, concentrations at each time step
Sincock et al. (2003)	QUASAR	RSA	Nash-Sutcliffe efficiency
Cox and Whitehead (2005)	O ²	Generalized Sensitivity Analysis, Monte Carlo techniques	Cumulative distribution functions
Tang et al. (2007b)	SAC-SMA-SNOW-17	Sobol, RSA, ANOVA, PEST	RMSE of discharge, RMSE of Box-Cox transformed discharge
Garambois et al. (2013)	MARINE	Sobol method	Discharge at each time step

(E)FAST : (extended) Fourier amplitude sensitivity test

RSA : Regional sensitivity analysis

PEST : Parameter estimation software

but more costly GSA method, such as the Sobol method (Sobol, 1993). The efficiency of Sobol method has been investigated by Tang et al. (2007b,a), who concluded that the Sobol method yielded the most robust sensitivity rankings. The objectives of this study are to :

- identify the most influential parameters of C-RIVE model on DO

concentrations within the water column

- improve the knowledge on the biogeochemical processes at work in different hydrological and trophic contexts

In the section 4.2.1 and 4.2.2, the C-RIVE biogeochemical model is presented, followed by the description of the case study (Section 4.2.3). The parameters considered in the SA are shown in the section 4.2.4. The sensitivity analysis methodology is described subsequently (Section 4.2.5). For the results, the Morris and Sobol sensitivity indices during low flow in different trophic contexts are firstly given (Section 4.3.1 and 4.3.2), followed by those during high flow (Section 4.3.3). The selection of influential parameters and the discussion of the parameters' behavior in different contexts are presented in section 4.4, followed by a brief conclusion (Section 4.5).

4.2 Material and methods

4.2.1 Biogeochemical model C-RIVE

The C-RIVE model is a C library of the RIVE model which has been implemented in a few numerical programs, mainly RIVERSTRAHLER (Billen et al., 1994; Garnier et al., 1995) and PROSE (Even, 1995; Even et al., 1998; Flipo et al., 2004, 2007b; Vilmin et al., 2018, 2016). C-RIVE simulates the cycling of carbon, nitrogen, phosphorus and oxygen in river systems. Biogeochemical cycles are simulated with a community-centered or agent-based model. Biogeochemical processes are simulated in the water column and the sediment layer. The exchanges of dissolved compounds at sediment-water interface are also simulated by diffusion (Flipo et al., 2004). Particulate matter sedimentation and resuspension by the energy transferred from water flow and river navigation have been developed by Even et al. (2004) and calibrated by Vilmin et al. (2015b). Raimonet et al. (2015) have used this model to investigate the nitrites production. Phytoplankton, heterotrophic bacteria and nitrifying bacteria are explicitly modeled by the means of growth and mortality rates weighted by organic matter and nutrients limitation. The performances of the hydro-ecological model, PROSE, are well characterized in Vilmin et al. (2018, 2016, 2015b), concerning errors on carbon, nitrogen, phosphorus and oxygen. A simplified schema of biogeochemical processes implemented in the C-RIVE model is shown in figure 4.2.

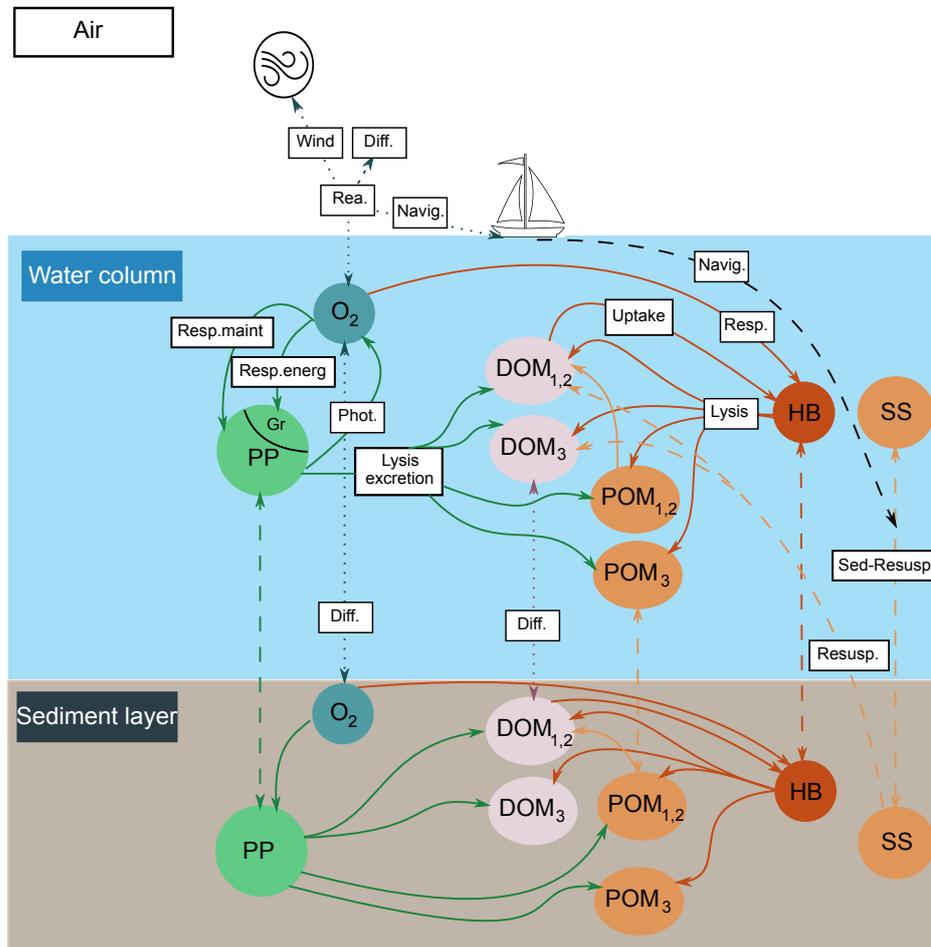


FIGURE 4.2 – Schematic description of C-RIVE model (modified from Vilmin et al. (2016)). Dotted lines : diffusive exchanges (Diff.); dashed lines : sedimentation-resuspension (Sed-Resusp.); solid lines : biogeochemical processes; PP : primary producer; Gr : growth; HB : heterotrophic bacteria; SS : suspended solids; DOM : dissolved organic matter; POM : particulate organic matter; subscripts 1, 2, and 3 refer to the highly biodegradable, moderately biodegradable, and nonbiodegradable fractions of organic matter. Rea. : reaeration; Navig. ; navigation; Resp. : respiration; Phot. : photosynthesis; Nitrifying bacteria are not represented.

4.2.2 Mathematical modeling of C-RIVE

The dissolved oxygen concentration within water column results from multiple processes :

$$\frac{d[O_2]}{dt} = \frac{d[O_2]}{dt}_{physical} + \frac{d[O_2]}{dt}_{bacterial} + \frac{d[O_2]}{dt}_{algal} \quad (4.1)$$

Physical processes are related to the reaeration and the diffusion at sediment-water interface :

$$\frac{d[O_2]}{dt}_{physical} = \frac{K_{rea}}{h}([O_2]_{sat}(T) - [O_2]) - \frac{D_s}{h}([O_2]_{water} - [O_2]_{sed}) \quad (4.2)$$

where

K_{rea} : reaeration coefficient (m/s)

h : water height (m)

$[O_2]_{sat}(T)$: maximum dissolved oxygen concentration saturation at temperature T (mg/L)

D_s : diffusion coefficient at sediment-water interface (m/s)

The reaeration coefficient K_{rea} is calculated from the empirical equation developed by [Thibodeaux et al. \(1994\)](#) :

$$K_{rea} = \sqrt{\frac{D_m \times V_{wa}}{h}} + (K_{wind} \times V_{wind}^{2.23} \times (D_m \times 10000)^{\frac{2}{3}} + K_{navig}) \quad (4.3)$$

where D_m is the molecular diffusivity of dissolved oxygen (m^2/s) at air-water interface depending on temperature ([Wilke and Chang, 1955](#)), V_{wa} is the water velocity (m/s), K_{wind} is the reaeration coefficient related to wind and V_{wind} is the wind speed at 10 m elevation (m/s) ([Escoffier et al., 2018](#); [Thibodeaux et al., 1994](#)). K_{navig} is the reaeration coefficient due to navigation introduced by [Vilmin \(2014\)](#).

The bacterial respiration is expressed as :

$$\frac{d[O_2]}{dt}_{bacterial} = -\tau_{HB}(1 - Y_{HB})uptake \quad (4.4)$$

$$uptake = \frac{\mu_{max,HB} e^{-\frac{(T-T_{opt,HB})^2}{\sigma_{HB}^2}} \frac{S}{S+K_S} [HB]}{Y_{HB}} \quad (4.5)$$

where

$[HB]$: heterotrophic bacteria biomass (mgC/L)

Y_{HB} : bacterial growth yield (-)

$uptake$: the highly biodegradable dissolved organic matter uptake for bacterial growth ($mgC/L/s$)

τ_{HB} : 1.0 ($molO_2/molC$) when considering the full oxidation of organic matter by the respiration process

$\mu_{max,HB}$: maximum growth rate of heterotrophic bacteria (/s)

$T_{opt,HB}$: optimal temperature for bacteria growth ($^{\circ}C$)

σ_{HB} : standard deviation of temperature function for bacteria ($^{\circ}C$)

S : the highly biodegradable dissolved organic matter MOD_1 (substrate) (mgC/L)

K_S : half-saturation constant for substrate uptake directly by bacterial growth (mgC/L)

The behavior of primary producers is represented using the AQUAPHY model (Lancelot et al., 1991). A phytoplankton specie is divided into three components :

$$[PP] = [F] + [R] + [M] \quad (4.6)$$

where

$[PP]$: primary producer biomass (mgC/L)

$[F]$: structural and functional macromolecules, which composes about 85% of $[PP]$ (mgC/L)

$[R]$: reserve products consisting of polysaccharides and/or lipids (mgC/L)

$[M]$: small metabolites for macromolecular synthesis (mgC/L)

The variation of DO by phytoplankton is mainly due to respiration and photosynthesis (Platt et al., 1977).

$$\frac{d[O_2]}{dt}_{algal} = p_{O_2} - r_{O_2,pp} \quad (4.7)$$

$$p_{O_2} = \tau_{PP} \frac{1}{h} \int_0^h P(z) dz [F] \quad (4.8)$$

$$P(z) = P_{\max}(1 - e^{-\frac{\alpha I(z)}{P_{\max}}})e^{-\frac{\beta I(z)}{P_{\max}}} \quad (4.9)$$

The irradiance is computed by Beer-Lambert law : $I(z) = I_0 e^{-\eta z}$. $\eta = \eta_{\text{water}} + \eta_{\text{chla}}[\text{chla}] + \eta_{\text{mes}}[SS]$.

where

z : depth (m)

h : water height (m)

τ_{PP} : 1.0 ($molO_2/molC$) when considering the full oxidation of organic matter by the respiration process

P_{\max} : maximum photosynthesis rate ($/s$)

α : photosynthetic capacity ($m^2s/\mu E/s$)

β : capacity of photoinhibition ($m^2s/\mu E/s$), set to zero in this study

$[\text{chla}]$: chlorophyll a (chl a) concentration = $\frac{[PP]}{C/\text{chla}}$ ($\mu gchla/L$)

$[SS]$: suspended solid concentration (mg/L)

η_{water} : light extinction coefficient for pure water (m^{-1})

η_{chla} : linear algal self-shading light extinction coefficient ($L/\mu gchla/m$)

η_{mes} : light extinction coefficient related to suspended solid ($L/mg/m$)

The phytoplankton respiration is expressed by the sum of two terms, associated with maintenance process and synthesis of new cellular material respectively.

$$r_{O_2,pp} = R_{m,pp}[F] + R_{\mu,pp}[F]\mu_{pp} \quad (4.10)$$

where

$R_{m,pp}$: maintenance respiration rate ($/s$)

$R_{\mu,pp}$: energetic cost of growth ($-$)

μ_{pp} : rate of phytoplankton growth ($/s$)

The rate of phytoplankton growth μ_{pp} is assumed to be limited by small metabolites $[M]$ and nutrients using a Monod kinetic :

$$\mu_{pp} = \mu_{\max,pp} e^{-\frac{(T-T_{opt,pp})^2}{\sigma^2}} \left(\frac{[M]}{[PP]} \right) \left(\frac{[N]}{[N] + K_N} \right) \left(\frac{[P]}{[P] + K_P} \right) \quad (4.11)$$

where

$[N]$ and $[P]$: nitrogen and phosphorus concentrations ($[N] = [NO_3^-] + [NH_4^+]$ and $[P] = [PO_4^{3-}]$) (mgN/L and mgP/L)

$\mu_{max,phy}$: maximum growth rate of phytoplankton ($/s$)

$T_{opt,phy}$: optimal temperature for phytoplankton growth ($^{\circ}C$)

σ_{phy} : standard deviation of temperature function for phytoplankton ($^{\circ}C$)

K_M , K_N and K_P : Half-saturation constant for phytoplankton, for N, for P ($-$, mgN/L and mgP/L)

4.2.3 Case study

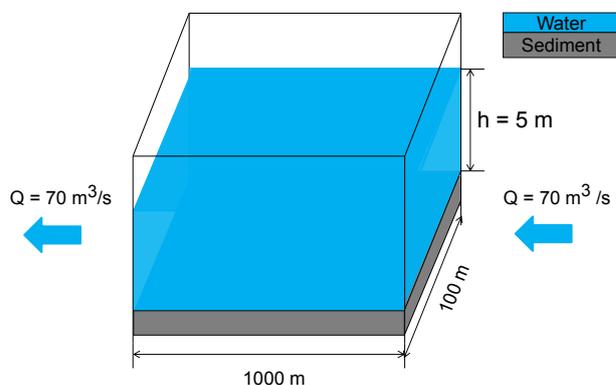


FIGURE 4.3 – Synthetic schema of case study (Q : discharge)

A case study, which mimics the Seine River, has been developed (Fig.4.3) and used to identify the most influential parameters of C-RIVE model. It is composed of a single 100m-wide cross-section, representative of a 1000m-long reach. Two compartments, namely the water column and the sediment layer, are simulated. The upstream and downstream concentrations are set to zero which means that the case study is a lagrangian approach allowing to follow the evolution of the water quality of a water body during its motion within the Seine River system.

4.2.3.1 Hydrological and biogeochemical conditions

The Seine River system is characterized by a temperate climate with oceanic influence. Maximum and minimum water flows occur in January

and August, respectively (Vilmin et al., 2016). Low flow periods are characterized by a discharge of $70 \text{ m}^3/\text{s}$. The maximum discharge of $300 \text{ m}^3/\text{s}$ is considered for high flow periods. Both values are representative of the observed discharge at Bougival (Escoffier et al., 2018).

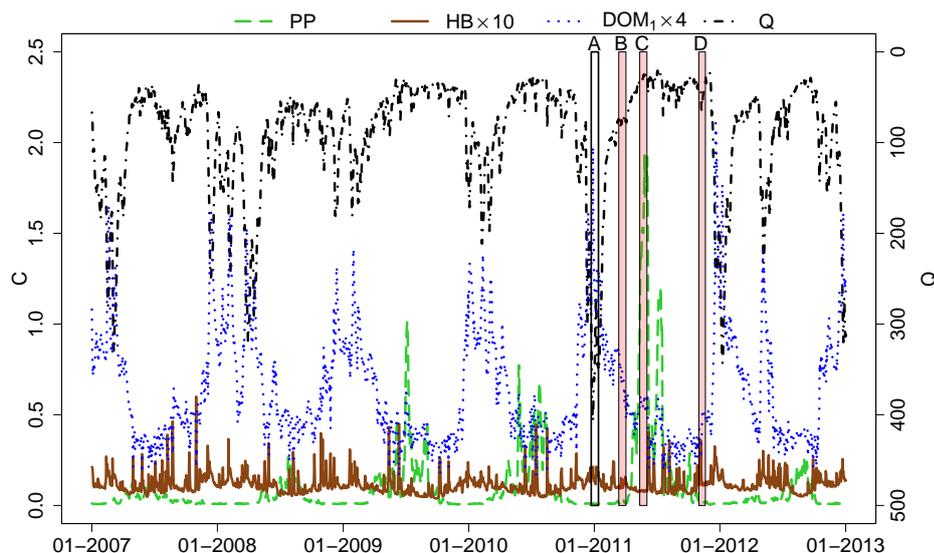


FIGURE 4.4 – Profile of algal and bacteria biomass simulated for the period 2007-2012. (PP (mgC/L) : Primary producer biomass; $HB \times 10$ (mgC/L) : Heterotrophic bacteria biomass multiply by 10; $DOM_1 \times 4$ (mgC/L) : Concentration of the highly biodegradable dissolved organic matter (substrate) multiply by 4; Q (m^3/s) : Discharge; A : High flow scenario; B : Pre-bloom scenario; C : Bloom scenario; D : Post-bloom scenario). Simulations from Vilmin et al. (2016).

According to algal blooms at Bougival in 2011 (Fig. 4.4), three trophic states can be distinguished during low flow periods : **Pre-bloom**, **Bloom** and **Post-bloom**.

- **Pre-bloom** with few bacteria and primary producers in the water column.
- **Bloom** where the phytoplankton is dominant.
- **Post-bloom** where a large amount of heterotrophic bacteria is growing in the river system following the mortality of the algal bloom.

For high flow periods, only one scenario is considered. Sediments are eroded, leading to an increase of DOM in the water column (Fig. 4.4). All the scenarii are built based on Vilmin et al. (2016) simulations.

4.2.3.2 Initial conditions

To determine the initial conditions describing the general situation in the Seine River system, we extract the biogeochemical concentrations simulated for the period 2007-2012 by Vilmin et al. (2018, 2016) at Bougival in the Seine River (Fig. 4.4). The initial biogeochemical concentrations for the case study are set to the mean concentrations over the period 2007-2012 (Tab. 4.2) except for the highly and moderately biodegradable dissolved organic matters (DOM_1 and DOM_2), temperature, phytoplankton and bacterial biomass which depend on the hydrological and trophic contexts and which are extracted from the simulated values (Fig. 4.4) and summarized in Tab. 4.3.

TABLE 4.2 – Initial conditions of species simulated for the case study

Species	$C_{ini,Water}$	$C_{ini,Sediment}$	Unit
NH_4^+	0.182	0.325	mgN/L
NO_2^-	0.041	0.043	mgN/L
NO_3^-	4.769	4.514	mgN/L
SS	16.820	95010	mg/L
DOM_3	1.948	1.944	mgC/L
POM_1	0.125	44	mgC/L
POM_2	0.232	696	mgC/L
POM_3	0.403	2555	mgC/L
PO_4^{3+}	0.100	0.266	mgP/L
O_2	9.931	6.654	mgO ₂ /L

DOM_3 : non biodegradable dissolved organic matter.

POM : particulate organic matter, index 1,2,3 corresponding to highly, moderately and non biodegradable fraction.

4.2.4 Parameters considered in the sensitivity analysis

The sensitivity to 31 parameters was analyzed (Tab. 4.4), including 3 physical parameters related to the reaeration at air-water interface and the sediment-water diffusion, 7 bacterial parameters associated with respiration, mortality and growth rate and 21 phytoplanktonic parameters mainly related to photosynthesis, respiration, mortality and growth rate. These 31

TABLE 4.3 – The initial conditions of DOM, algal and bacterial biomass in different trophic and hydrological contexts

Scenarios	Layers	HB	PP	DOM_1	DOM_2	Unite	mean T ($^{\circ}C$)
High flow	Water	0.021	0.010	0.305	0.678	mgC/L	3.9
	Sediment	0.032	0.003	0.182	0.655	mgC/L	
Pre-bloom	Water	0.007	0.010	0.120	0.506	mgC/L	13.9
	Sediment	0.016	0.002	0.117	0.669	mgC/L	
Bloom	Water	0.007	1.459	0.142	0.503	mgC/L	20.0
	Sediment	0.005	0.330	0.074	1.341	mgC/L	
Post-bloom	Water	0.023	0.010	0.100	0.535	mgC/L	13.4
	Sediment	0.016	0.003	0.115	1.281	mgC/L	

DOM_1 : highly biodegradable dissolved organic matter

DOM_2 : moderate biodegradable dissolved organic matter

parameters govern the variation of DO within the water column in the C-RIVE model. We account for the reaeration process through the ship navigation (K_{navig}) and the wind effect (K_{wind}) only (eq. (4.3)). Analyzing effect of weirs will require to set up a full hydrological model which is not the purpose of this paper and could be tackled with the PROSE model in future works.

Given previous researches carried out in the framework of the PIREN-SEINE program (<https://www.piren-seine.fr/>) (Garnier et al., 1995, 1998, 1992; Even et al., 1998; Flipo et al., 2004, 2007b; Vilmin, 2014), optimal parameter values for the Seine River system are available. 11 parameter variation ranges are therefore obtained applying a variation of 25% around these optimum values (Tab. 4.4). For the remaining 20 parameters, variation ranges are assessed based on literature review (Tab. 4.4).

In addition, the chemical composition of organic matter is the Redfield ratio $C : N : P = 106 : 16 : 1$ (Redfield et al., 1963) for all simulations. The sedimentary processes have been calibrated and validated by Vilmin et al. (2015b). Therefore, the calibrated sedimentary parameters are considered for all simulations. The release of DOM into water column by erosion process has been taken into account for the selected high flow period (Fig. 4.4).

TABLE 4.4: Parameter list and variation ranges

Parameters	Description	Min_Val	Ref_val	Max_val	Unit	References
η_{water}	Light extinction coefficient for pure water	0.2	0.32	0.8	[/m]	Smith and Baker (1981); Wetzel (1983); Morel et al. (2007)
D_s^*	Diffusion coefficient at sediment-water interface	0.00125	0.0015	0.001875	[m/s]	Garnier et al. (1995); Even et al. (2004); Flipo et al. (2004)
K_{navig}	Reaeration coefficient related to the navigation	0	0	0.05	[m/h]	Vilmin (2014)
K_{wind}^*	Reaeration coefficient related to wind effect	0.885	1.18	1.475	[m/h]	Thibodeaux et al. (1994); Escoffier et al. (2018)
Bacterial parameters						

$\mu_{max, hb}$	Maximum growth rate of bacteria	0.01	0.04	0.13	[/h]	Garnier et al. (1992); Servais and Garnier (1993); Zweifel et al. (1995); Sterner et al. (1995); Even (1995)
σ_{hb}^*	Standard deviation of temperature function for bacteria	12.75	17	21.25	[°C]	Garnier et al. (1992, 1998)
$K_{O_2, hb}^*$	Half-saturation constant for dissolved oxygen	0.375	0.5	0.625	[mgO ₂ /L]	Even et al. (1998); Garnier et al. (2000); Flipo et al. (2004)
$mort_{hb}$	Maximal mortality rate bacteria	0.01	0.02	0.08	[/h]	Servais et al. (1985); Garnier et al. (1995); Sterner et al. (1995)
$T_{opt, hb}$	Optimal temperature for bacterial growth	10	25	35	[°C]	Billen and Servais (1991); Garnier et al. (1998)

$V_{set, hb}$	Settling velocity of bacteria	0.0	0.005	0.1	$[m/h]$	Even et al. (1998); Garnier et al. (1992)
Y_{hb}	Bacterial growth yield	0.03	0.25	0.5	[-]	Kroer (1993); Benner et al. (1995); Giorgio and Cole (1998); Servais and Garnier (1993); Billen and Servais (1989)
Phytoplanktonic parameters						
α_{pp}	Photosynthetic capacity	0.0003	0.0012	0.0018	$[m^2 s / \mu E / h]$	Hanset (1989); Garnier et al. (1995); Flipo et al. (2004)
$\eta_{chla, pp}$	Linear algal self-shading light extinction coefficient	0.006	0.02	0.054	$[L / \mu g chla / m]$	Bailey and Ahmadi, 2014; Garnier et al., 1995; Morel et al., 2007; Wang et al., 2009)

$\eta_{ss,pp}^*$	Light extinction coefficient related to suspended solids	0.03	0.04	0.05	[L/mg/m]	Even (1995); Flipo et al. (2004); Vilmin (2014)
$\mu_{max,pp}$	Maximum growth rate of phytoplankton	0.035	0.12	0.125	[/h]	Hanset (1989); Cox (2003a); Cho et al. (2017)
σ_{pp}^*	Standard deviation of temperature function for phytoplankton	12.75	17	21.25	[°C]	Garnier et al. (1995, 1998); Descy et al. (2012)
$C/chla_{pp}$	Ratio of carbon to chlorophyll <i>a</i>	0.02	0.035	0.13	[mgC/ μ gchl <i>a</i>]	Even (1995); Hanset (1989); Sathyendranath et al. (2009); Jakobsen and Markager (2016)
$C_{R,pp}$	Catabolise of reserve products	0.1	0.2	0.25	[/h]	(Lancelot and Mathot, 1987; Hanset, 1989; Garnier et al., 1995)

$E_{C^{te},pp}^*$	Constant rate of excretion	0.00075	0.001	0.00125	[/h]	Billen et al. (1994); Garnier et al. (1995); Even et al. (2004)
$E_{phot,pp}^*$	Fraction of photosynthesis excreted	0.00045	0.0006	0.00075	[/h]	Billen et al. (1994); Garnier et al. (1995)
K_M	Half-saturation constant for small metabolites M	0.045	0.06	0.075	[-]	Billen et al. (1994); Garnier et al. (1995)
$K_{N,pp}$	Half-saturation constant for N	10	70	300	$[\mu gN/L]$	Cox (2003a); Billen et al. (1994)
$K_{NH_4^+,pp}^*$	Half-saturation constant for use of NH_4 by phytoplankton	67.5	90	112.5	$[\mu gN/L]$	Brion (1993); Even (1995)
$K_{P,pp}$	Half-saturation constant for P	1	46	50	$[\mu gP/L]$	Cox (2003a); Garnier et al. (1995); Cho et al. (2017)
$mort_{pp}$	Maximal mortality rate of phytoplankton	0.0	0.004	0.02	[/h]	Bruning et al. (1992); Even et al. (1998); Billen et al. (1994)

$P_{max,pp}$	Photosynthesis rate	0.09	0.4	0.546	[/h]	(Capblancq and Dauta, 1978; Hanset, 1989; Lancelot et al., 1991; Hanset, 1989)
$R_{\mu,pp}^*$	Energetic cost of growth	0.375	0.5	0.625	[-]	Lancelot et al. (1991); Garnier et al. (1995); Even et al. (1998)
$R_{m,pp}$	Respiration of maintenance	0.001	0.002	0.021	[/h]	Cox (2003a); Lancelot et al. (1991); Billen et al. (1994)
$S_{R,pp}$	Synthesis of reserve products	0.08	0.26	0.49	[/h]	Hanset (1989); Lancelot et al. (1991); Even (1995); Billen et al. (1994)
$T_{opt,pp}$	Optimum temperature for growth of phytoplankton	10	33	37	[°C]	Dauta et al. (1990); Garnier et al. (1995)

$V_{set,pp}$	Settling velocity for phytoplankton	0.003	0.05	0.075	$[m/h]$	Bailey and Ah-madi (2014); Cho et al. (2017); Garnier et al. (1995); Even et al. (1998)
	tob					

* *range = value calibrated* $\times (1 \pm 25\%)$

4.2.5 Sensitivity analysis methodology

The influence of input factors (X) on the model output (Y , DO concentration) which is a set of curves describing the variation of DO over time, is analyzed based on an innovative methodology (Fig.4.5). The methodology consists of three steps. First, the model output is a DO time series, which is reduced to perform a sensitivity analysis. Indeed, sensitivity indices are usually calculated for scalar outputs obtained from objective functions, such as the Nash-Sutcliffe efficiency (NSE) or the Root Mean Square Error (RMSE) (Deflandre et al., 2006; Sincock et al., 2003; Nossent et al., 2011). To emphasize the global sensitivity of input factors in terms of the entire model output in different contexts, we propose to transform the time-dependent model output of DO to one or several scalar values for each simulation by applying an empirical orthogonal function (EOF) analysis. This dimensionality reduction permits to compute a composite output value for each simulation. The variations of these composite values are then analyzed using the Morris screening technique to identify the most influential input parameters. Finally, the non-influential parameters are set to a fixed value and a variance decomposition-based method, the Sobol' sensitivity analysis (Sobol, 1993), is performed on the identified parameters.

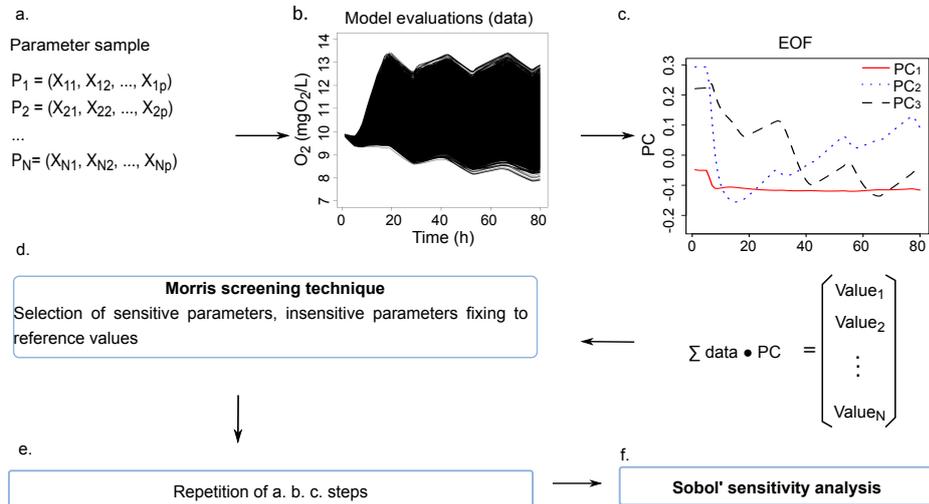


FIGURE 4.5 – Sensitivity analysis methodology

4.2.5.1 Dimensionality reduction

Temporal sensitivity is usually calculated for the overall model performance over a period of time (objective functions, e.g., the root mean square error RMSE of model outputs or the Nash-Sutcliffe efficiency NSE) (Deflandre et al., 2006; Nossent et al., 2011; Sincock et al., 2003). Other authors evaluate the mean/sum/max of the model output directly, or the model output at each time step (Reusser et al., 2011; Bailey and Ahmadi, 2014). By contrast, the calibration is not the objective of this paper. Our objective is rather to identify the parameters that are the most influential on the variation of DO during the time of the experiment. Therefore, we use an empirical orthogonal function (EOF) analysis to transform the model output. EOF is a factorial method that reduces the data dimensionality with a controlled loss of information (Wold et al., 1987). In our case, the data are the multiple model outputs obtained from the differing model parameters for each scenario. The EOF leads to a decomposition of the model outputs in terms of uncorrelated vectors called principal components (PC). The model output time series are represented by a linear combination of PC. The coefficients of the linear combination are then considered for the SA.

We use the function `prcomp` in R (R Core Team, 2014) to compute the PC and select m (Fig.4.5c) components which summarize at least 99% of the total model output variance. In the EOF framework, each principal component summarizes a temporal feature. Each simulation can be transformed into m values corresponding to the orthogonal projection of each model output times series onto the m principal directions (components). As these values summarize most of the total variability of the simulated data, they can be employed for SA. Since the principal components are uncorrelated, an indicator for SA is defined as the sum of the projected data of each simulation. This indicator is used in the subsequent sensitivity analysis where a scalar is required as an output (Lamboni et al., 2011).

4.2.5.2 Morris method

The Morris method (Morris, 1991; Campolongo et al., 2007) is a screening technique based on a discretization of the input parameters in n levels. The most used and simplest screening design is called “**One-At-a-Time**” (**OAT**) experiments, where each parameter varies one after another.

The Morris method performs r OAT designs and measures the relative

variation of the model output (E, Elementary Effect). In this paper, the model output is summarized by the indicator built from the EOF. Each parameter $X_i, i = 1, \dots, p$, is randomly selected in the input space and the variation direction is also random. The repetition (r times OAT) allows the scanning of the full input space. In practice, the space of input parameters is normalized on $[0, 1]$. Further, we assume that the input parameters are uniformly distributed in the input space. Since each OAT design provides p elementary effects (one per parameter) via $p + 1$ points in the input space, the cost of the Morris method is $r(p + 1)$ for r OAT designs.

To assess the influence of the input factors, two indices are computed : μ^* and σ . Let us denote the model function f with parameter set X and E_j^i the elementary effect of parameter X_j for the i -th repetition, $j = 1, \dots, p$ and $i = 1, \dots, r$. E_j^i is defined as :

$$E_j^i = \frac{f(X + \Delta e_j) - f(X)}{\Delta}, \quad (4.12)$$

where Δ is a value chosen in $\{1/(n - 1), \dots, 1 - 1/(n - 1)\}$ by the user, n is the number of discretization levels, e_j is a vector of the canonical basis. For any X_j selected in the input space, $X + \Delta e_j$ is always in the input space. [Campolongo et al. \(2007\)](#) recommend that Δ equals $n/2(n - 1)$. The indices are calculated as follows :

- $\mu_j^* = \frac{1}{r} \sum_{i=1}^r |E_j^i|$ (mean of the absolute values of the elementary effects)
- $\sigma_j = \sqrt{\frac{1}{r} \sum_{i=1}^r (E_j^i - \frac{1}{r} \sum_{i=1}^r E_j^i)^2}$ (standard deviation of the elementary effects)

μ_j^* estimates the overall impact of the parameter X_j on the model output and σ measures the non-linear and/or interaction effects with other parameters.

To interpret the sensitivity indices μ_j^* and σ_j , we draw the scatter plot $\mu_j^* - \sigma$. The Morris method allows the classification of the parameters in three types :

- Parameters having negligible effects with small μ^* and σ .
- Parameters having non-linear effects and/or interactions with other parameters with large μ^* and σ .
- Parameters having linear and additive effects with large μ^* and small σ .

The Morris method is adapted to models with large number of para-

eters. It allows the reduction of the number of parameters (factor fixing), discarding those with negligible effects, in order to use subsequently other more subtle but more costly SA methods, like Sobol' sensitivity analysis.

4.2.5.3 The Sobol' sensitivity analysis

The Sobol' sensitivity analysis (Sobol, 1993; Saltelli, 2002; Saltelli et al., 2010) is based on the decomposition of the output model variance. It can be used for non-linear and non monotonic models and functions. Consider a model represented by the function :

$$Y = f(X) = f(X_1, \dots, X_p), \quad (4.13)$$

where Y is the model output (here, dissolved oxygen concentration) and X (X_1, \dots, X_p) is the parameter set of the model. It is possible to present this model by a sum of elementary functions :

$$f(X) = f_0 + \sum_{i=1}^p f_i(X_i) + \sum_{i=1}^{p-1} \sum_{j=i+1}^p f_{ij}(X_i, X_j) + \dots + f_{1,\dots,p}(X_1, \dots, X_p) \quad (4.14)$$

If the parameters of the model (X_1, \dots, X_p) are mutually independent and each elementary function is square integrable with a zero mean (Eq. (4.15)), then f_0 is the expectation of the model output (Y), and this decomposition is unique (Sobol, 1993).

$$\int_0^1 f_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) dx_{i_k} = 0, \quad \forall k = 1, \dots, s, \quad \forall \{i_1, \dots, i_s\} \subseteq \{1, \dots, p\} \quad (4.15)$$

The total unconditional variance of the model output Y is defined as :

$$V(Y) = \int_{\Omega^p} f^2(X) dX - f_0^2 \quad (4.16)$$

where Ω^p is the space of input parameters with p dimensions.

The variance of each elementary function, called partial variance, is calculated as :

$$V_{i_1, \dots, i_s} = \int_0^1 \dots \int_0^1 f_{i_1, \dots, i_s}^2(X_{i_1}, \dots, X_{i_s}) dX_{i_1}, \dots, dX_{i_s} \quad (4.17)$$

where $1 \leq i_1 \leq \dots \leq i_s \leq p$ and $s = 1, \dots, p$. Under the hypothesis that the input parameters are mutually independent, the total unconditional variance is equal to the sum of the partial variances :

$$V(Y) = \sum_{i=1}^p V_i(X_i) + \sum_{i=1}^{p-1} \sum_{j=i+1}^p V_{ij}(X_i, X_j) + \dots + V_{1,\dots,p}(X_1, \dots, X_p) \quad (4.18)$$

The Sobol's indices are determined by the ratio of the partial variances to the total unconditional variance.

- First order index : $S_i = \frac{V_i}{V}$
- Second order index : $S_{ij} = \frac{V_{ij}}{V}$
- Total index : $S_{T_i} = S_i + \sum_{j \neq i} S_{ij} + \dots + S_{1,\dots,p}$

The first order index S_i indicates the individual variance contribution of the parameter X_i to the total unconditional variance, called also the “main effect”. The second order index S_{ij} explains the interaction effect between parameters X_i and X_j . The global impact of parameter X_i including the main effect and all its interactions with other parameters is given by the total index S_{T_i} . If the sum of all first order indices is less than 1, the model is non-additive.

4.2.6 Simulation setting and computational cost

To avoid the nutrient limitation of biogeochemical processes, the case study is simulated for 80 hours with a time step of 15 min. The mass balance is calculated every hour.

For the Morris method, each parameter range is discretized in 25 (n) levels, and a grid jump of 13 ($= n \times \frac{n}{2(n-1)}$) is used. According to a test of the number of **OAT** repetitions (not shown here), we repeat 1000 (r) times **OAT** designs which is sufficient to obtain a stable result. This repetition results in 32000 ($= r \cdot (p + 1) = 1000 \cdot (31 + 1)$) model evaluations. The 17 parameters identified by the Morris method are then analyzed by applying the Sobol' sensitivity analysis.

For the Sobol' sensitivity analysis, a sample size (n) of 20000 is selected, demanding 380000 ($= n \cdot (p + 2) = 20000 \cdot (17 + 2)$) simulations to calculate the first order and total sensitivity indices. This sample size is also suitable to ensure the convergence of sensitivity index values.

The sensitivity indices are computed by a PYTHON package SALIB (Herman and Usher, 2017). The output of SALIB gives also the 95% confidence intervals estimated by bootstrap analysis with a resample size of 100 (Efron and Tibshirani, 1993).

To decrease the computational cost, a parallel computing code in PYTHON has been developed and coupled to C-RIVE. The duration of the computations is 20 minutes for the Morris method and 4 hours for the Sobol' sensitivity analysis with 12 processors (Intel(R) Xeon(R) E5-2430 with a frequency of 2.50 GHz).

4.3 Results

In the following paragraphs, the Morris method and the Sobol's sensitivity analysis results are presented. First, the results of Morris method in three trophic contexts over the low flow periods are shown (Section 4.3.1). The subsequent section (Section 4.3.2) describes the outcome of Sobol' sensitivity analysis over the low flow periods. Finally, the results of sensitivity analysis during the high flow period are presented (Section 4.3.3).

4.3.1 Results of the Morris method over the low flow periods

4.3.1.1 Pre-bloom

During low flow, in a **Pre-bloom** configuration, the maximum growth rate of bacteria ($\mu_{max,hb}$), the bacterial growth yield (Y_{hb}), the rate of bacteria mortality ($mort_{hb}$), the optimal temperature for bacteria growth ($T_{opt,hb}$) and the reaeration coefficient related to navigation (K_{navig}) are the most influential parameters (Fig. 4.6A). Concerning the phytoplanktonic parameters, the most influential is the photosynthetic capacity (α_{pp}), but its effect remains negligible.

For **Pre-bloom** scenario, 8 influential parameters are identified using μ^* : $\mu_{max,hb}$, Y_{hb} , $mort_{hb}$, $T_{opt,hb}$, K_{navig} , K_{wind} , σ_{hb} and $V_{sed,hb}$ (see Table 4.4 and Section 4.2.2 for the definition for these parameters). The most influential parameters are related to bacterial activity. 6 parameters are related to bacteria and two are physical.

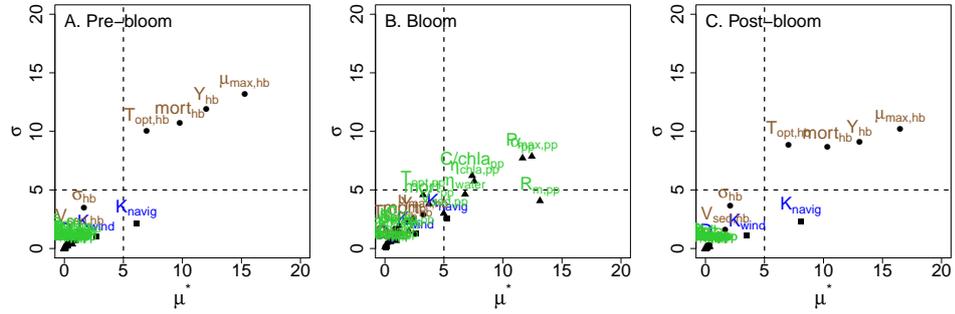


FIGURE 4.6 – Results of Morris method for low flow conditions (square (text blue) : physical parameters ; circle (text brown) : bacterial parameters ; triangle (text green) : phytoplanktonic parameters ; dashed lines : σ and μ^* thresholds determined by the results of Sobol method)

4.3.1.2 Bloom

During a bloom, parameters related to maintenance respiration ($R_{m,pp}$), and photosynthesis (the photosynthetic capacity, α_{pp} ; the maximum algal photosynthesis rate, $P_{max,pp}$; the light extinction coefficients, $\eta_{chla,pp}$ and η_{water}) are the most significant parameters (Fig. 4.6B). The ratio of carbon to chl *a* ($C/chla_{pp}$) is as important as the light extinction coefficients. The maximal growth rate ($\mu_{max,pp}$) and mortality rate ($mort_{pp}$) are insensitive. In contrast to the **Pre-bloom** scenario, the bacterial parameters and reaeration coefficients are less important during algal blooms. The most influential bacterial parameter is the maximal growth rate ($\mu_{max,hb}$).

For **Bloom** scenario, 11 parameters are assumed to be sensitive : $R_{m,pp}$, $P_{max,pp}$, α_{pp} , $\eta_{chla,pp}$, $C/chla_{pp}$, η_{water} , K_{navig} , $V_{sed,pp}$, $mort_{pp}$, $\mu_{max,hb}$ and $T_{opt,pp}$ (see Table 4.4 and Section 4.2.2 for the definition for these parameters). 8 parameters are related to the phytoplankton, two are physical parameters, and only one is bacterial.

4.3.1.3 Post-bloom

Interestingly, under **Post-bloom** conditions (Fig. 4.6C), the same parameter distribution is observed as under **Pre-bloom** conditions. The most influential parameters are $\mu_{max,hb}$, Y_{hb} , $mort_{hb}$, $T_{opt,hb}$ and K_{navig} . No phytoplanktonic parameter has a significant effect on the variation of DO in

water column.

For **Post-bloom** scenario, parameters identified are exactly the same as in **Pre-bloom** context : $\mu_{max,hb}$, Y_{hb} , $mort_{hb}$, K_{navig} , $T_{opt,hb}$, K_{wind} , σ_{hb} and $V_{sed,hb}$ (see Table 4.4 for the definition for these parameters). 6 parameters are related to bacteria and 2 to physical processes.

4.3.1.4 The σ and μ^* thresholds

The Morris method allows the classification of the parameters in three types according to the values of σ and μ^* . However, it is usually difficult to determine the σ and μ^* thresholds. The next section illustrates the consistency between the Morris method and Sobol sensitivity analysis (Fig. 4.6, 4.7 and 4.8).

According to the parameters selected, the σ and μ^* thresholds can be determined ($\sigma = 5$, $\mu^* = 5$) for the definition of the three areas, which defines the importance and independency of each parameter (see definition in section 4.2.5.2). Those values are reported on each Morris Figure to ease their readability (Fig. 4.6 and 4.8A). Further studies on river water quality models should be carried out to assess the universality of those values.

4.3.2 Results of the Sobol' sensitivity analysis over the low flow period

17 parameters are selected using the Morris method, among which 9 are related to phytoplankton, 6 to bacteria and 2 to oxygen reaeration at the air-water interface. The other parameters are set to their reference value (Tab. 4.4).

Figure 4.7 presents the first order (left) and the total sensitivity indices (right) of Sobol method. The parameters are ranked by total sensitivity index (S_{T_i}). Only the most influential parameters for each trophic state are shown.

4.3.2.1 Pre-bloom

The result of Sobol' sensitivity analysis for the **Pre-bloom** (Fig. 4.7A) indicates that the maximal growth rate for bacteria ($\mu_{max,hb}$) is the most influential parameter, which represents 46% of the variance of the model output (S_{T_i}), among which 26% is caused by the variation of $\mu_{max,hb}$ itself

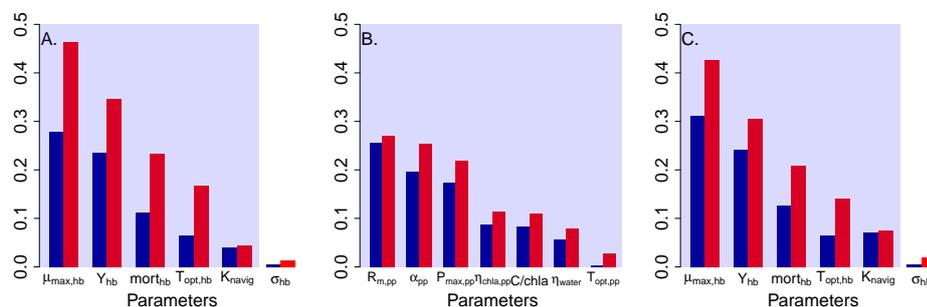


FIGURE 4.7 – Results of Sobol method for low flow periods in various trophic contexts : A) Pre-bloom, B) Bloom, C) Post-bloom. S_i in blue (left) and S_{T_i} in red (right); Parameters ranked by total index S_{T_i} ; The blue area highlights the parameters representing 95% of the total variance of the model output, V_T)

(S_i). In other words, the maximal growth rate for bacteria has an important interaction with other parameters (18%, respectively). The parameter ranked second is the bacterial growth yield (Y_{hb}) with a global impact of 35% (Fig. 4.7A). All bacterial parameters ($\mu_{max,hb}$, Y_{hb} , $mort_{hb}$ and $T_{opt,hb}$) show an interaction effect ($S_i < S_{T_i}$), while the reaeration coefficient related to the navigation (K_{navig}) has an independent effect shown also by the Morris method (Fig. 4.6A). The sum of all first order indices (S_i) is less than 1 (0.751, Table 4.5), which indicates that the C-RIVE module is not additive, as can be expected for a complex environmental model.

The phytoplanktonic parameters have no significant influence on the variation of DO concentration for **Pre-bloom** conditions. The photosynthetic capacity (α_{pp}) has a total sensitivity index less than 0.005.

4.3.2.2 Bloom

The maintenance respiration ($R_{m,pp}$) represents 27% of the total unconditional variance of the model output followed by the photosynthetic capacity (α_{pp}) with a contribution of 25% (Fig. 4.7B). However, the energetic cost of growth ($R_{\mu,pp}$) has no significant influence on the variation of DO. The consumption of DO by phytoplankton is mainly used to support and maintain life processes within cells of organisms.

The third ranked parameter is the maximal photosynthesis rate with a

contribution of 22% ($P_{max,pp}$). 19% of the variance of the simulated DO result from the variation of the light extinction coefficients (11% for $\eta_{chla,pp}$ and 8% for η_{water} respectively). The ratio of carbon to chl *a* ($C/chla_{pp}$) is also influential (11%).

The photosynthetic parameters (α_{pp} , $P_{max,pp}$, η and $C/chla_{pp}$) account for more than 70% of the variance of the simulated DO during a bloom. The photosynthesis and respiration are the most important processes governing the DO concentration during a bloom, as can be expected. The bacterial processes have no significant influence on the model output. Albeit the interactions between parameters are lesser during a bloom than during a pre-bloom, they are not negligible for α_{pp} and $P_{max,pp}$.

4.3.2.3 Post-bloom

After a bloom, the parameter ranking is the same as the one for **Pre-bloom** conditions. The maximal growth rate ($\mu_{max,hb}$) is the most influential parameter with a 43% contribution of the variance of the model output (Fig. 4.6C). The second ranked parameter is bacterial growth yield (Y_{hb}) with a 24% individual contribution and 7% of interactions with other parameters. Like **Pre-bloom** configuration, the river system is heterotrophic, the growth, mortality and respiration processes of bacteria mainly control the DO concentration within water column.

The values of the first order indices for bacterial parameters are less than the values of the total sensitivity indices ($S_i < S_{T,i}$), which proves the existence of parameter interactions. However, the reaeration coefficient related to the navigation (K_{navig}) has an independent effect ($S_i \approx S_{T,i}$, Fig. 4.7C), which is consistent with the result of the Morris method (Fig. 4.6C).

4.3.3 Sensitivity analysis results during high flow period

During high flow period, the reaeration coefficient related to the navigation (K_{navig}) becomes the most significant parameter with a contribution of 51% to the variance of simulated DO (Fig. 4.8B, S_T). Furthermore, the equality between the first index and the total index of K_{navig} indicates a linear effect of the reaeration coefficient. K_{wind} behaves the same way with a 7% contribution to the variance of simulated DO. The second ranked parameter is the bacterial growth yield (Y_{hb} , $S_T = 27\%$). The optimal temperature for

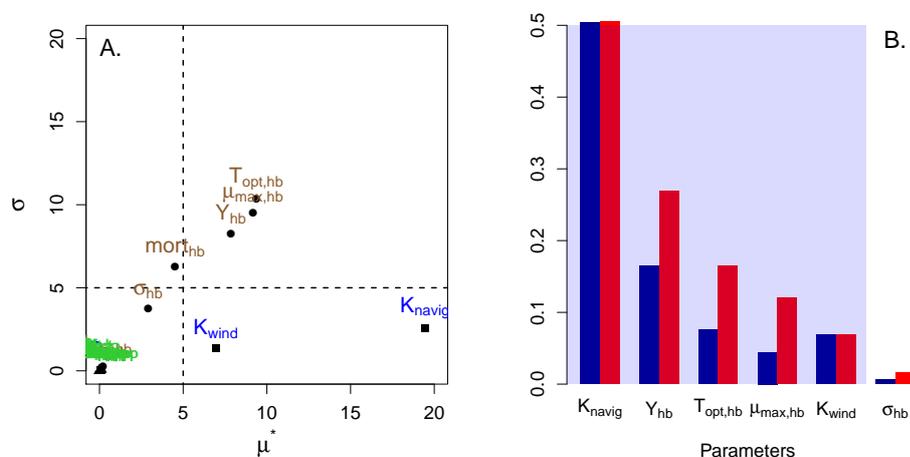


FIGURE 4.8 – Results of Morris(A) and Sobol (B) methods during high flow period (Dashed lines : σ and μ^* thresholds determined by the results of Sobol method) S_i in blue and S_{T_i} in red ; Parameters ranked by total index, S_{T_i} . The blue area highlights the parameters representing 95% of the total variance of the model output, V_T .

bacteria growth ($T_{opt,hb}$) and the maximum growth rate of bacteria ($\mu_{max,hb}$) are also influential (16% and 12%, respectively).

Like the **Pre-bloom** and **Post-bloom** configurations, the phytoplanktonic parameters have no significant influence on the variation of simulated DO. The influential parameters identified here are similar to those selected in **Pre-bloom** and **Post-bloom** conditions. However, the parameter ranking is different, especially the two reaeration coefficients (K_{navig} and K_{wind}) under high flow period. This result highlights the importance of oxygen reaeration, mostly controlled by human activities (navigation) during winter.

4.3.4 Selection of the most influential parameters of the C-RIVE model

We select the parameters summarizing at least 95% of the total unconditional variance of the model output for various trophic and hydrological conditions. The results are presented by the polygon (blue area) in the figures 4.7 and 4.8. To summarize the results, we distinguish two contrasted parameter selections : bloom (Fig. 4.7B) and no bloom (Fig. 4.7A, 4.7C and

4.8B).

During a bloom, 7 parameters are selected, among which 6 parameters are related to phytoplankton and one to the optical properties of pure water ($R_{m,pp}$, α_{pp} , $P_{max,pp}$, $\eta_{chla,pp}$, $C/chla_{pp}$, η_{water} and $T_{opt,pp}$ and see Table 4.4 for the definition for these parameters).

Out of algal blooms, 6 parameters can summarize more than 95% of the total variance of the model output for all conditions ($\mu_{max,hb}$, Y_{hb} , $mort_{hb}$, $T_{opt,hb}$, K_{navig} and K_{wind}). In contrast to a bloom, 2 parameters are related to oxygen reaeration and 4 to bacterial activities. No phytoplanktonic parameter is sensitive.

In conclusion, 13 parameters are identified as most influential for C-RIVE model.

4.4 Discussion

4.4.1 Parameters' behavior during an algal bloom

During an algal bloom, the maintenance respiration ($R_{m,pp}$) and the photosynthetic parameters (α_{pp} , $P_{max,pp}$, $\eta_{chla,pp}$, η_{water}) are the most sensitive. The bacterial and oxygen reaeration parameters have no significant effects on the variation of DO concentration. This outcome can be interpreted by the fact that phytoplankton metabolism is dominant and the river system is autotrophic during an algal bloom.

A recent study on the Seine River metabolism points out that the autotrophic respiration (AR) contributes significantly to the ecosystem respiration (ER) during phytoplankton bloom (Escoffier et al., 2018). The algal respiration has been identified as the most important process during summer months by Bailey and Ahmadi (2014). Bierman et al. (1994) and Cox and Whitehead (2005), who mentioned the sensitivity of oxygen depletion to algal respiration. We confirm this statement in our study. Furthermore, our study highlights that the consumption of DO by phytoplankton is mainly used to support and maintain life processes within cells of organisms ($R_{m,pp}$), but not by the energetic cost of the growth ($R_{\mu,pp}$, eq.(4.10)).

The influence of photosynthetic parameters indicates that the primary productivity is the main control on DO concentration during algal bloom, as already mentioned by Bierman et al. (1994) for the Mississippi River (United States). Furthermore, the light limitation on primary production in the Seine

River system has been reported in the literature (Garnier et al., 2001). Bierman et al. (1994) also stated that “dissolved oxygen concentrations appear very sensitive to changes in underwater light attenuation”. These studies are consistent with the influence of the light extinction coefficients ($\eta_{chla,pp}$ and η_{water}) in this paper. The light extinction coefficient related to suspended solid ($\eta_{ss,pp}$) is completely uninfluential in the case of the Seine River, which carries few sediments compared to large rivers (Delmas et al., 2012). During algal blooms, light availability is therefore mainly controlled by chl *a* concentration, and not by suspended solid concentration.

The influence of the ratio of carbon to chl *a* ($C/chla_{pp}$) is due to a strong coupling between light attenuation and chl *a* concentration (eq. (4.9)). As the ratio of carbon to chl *a* widely varies and depends on species and physiological state (Capblancq, 1982), it constitutes a factor of uncertainty in biogeochemical modeling.

4.4.2 Parameters’ behavior out of a algal bloom

During low flow periods without algal bloom, the depletion of DO is mainly due to bacterial respiration, as the river system is heterotrophic. According to the DO mass balance derived from our simulations with reference parameters, the sediment oxygen demand accounts for 23% of DO depletion within the water column for **Post-bloom** conditions and 37% for the **Pre-bloom** configuration. These proportions are consistent with the recent study on Seine River benthic dynamics carried out by Vilmin et al. (2016), who concluded that benthic respiration accounts for one third of the total river respiration during low flow period. However, no significant effect of the diffusion coefficient at the sediment-water interface (D_s) has been found in this paper. Indeed, the benthic flux is not only controlled by D_s , but also by the difference of DO concentrations between the water column and the sediment layer, in other words controlled by the bacterial parameters ($\mu_{max,hb}$, Y_{hb} and $mort_{hb}$ etc.). To summarize, the effect of the sediment demand on the DO in the water column (Cox and Whitehead, 2005; Deflandre et al., 2006) is related to the bacterial dynamics and growth in the sediment layer.

During high flow period, the oxygen reaeration becomes the most influential process due to low water temperatures and high water velocity. The variation of reaeration coefficients contributes to more than 60% of the

simulated DO variance (Fig. 4.8B). Although the oxygen reaeration has been identified as one of the most influential processes for several studies (Cox and Whitehead, 2005; Deflandre et al., 2006; Bailey and Ahmadi, 2014), we point out that the main control on water reaeration is the navigation (K_{navig}) and not the mixing due to the wind (K_{wind}). Therefore, water reaeration is mostly controlled by human activities whatever the hydrological and trophic conditions (Fig. 4.7 and 4.8B).

4.4.3 Impact of the temperature on SA

As stated previously **Post-bloom** and **High-flow** simulations are sensitive to the same parameters with a different ranking, the reaeration being the most influential for **High-flow**, while the growth rate of heterotrophic bacteria is the most important for **Post-bloom**. The difference between those two simulations relies on the discharge and the water temperature. In order to discriminate between those forcings, we simulate the **Post-bloom** and the **High-flow** over a wide range of ambient temperatures (Fig. 4.9). The total sensitivity indices for each temperature are calculated and normalized to 100%. The cumulation of the total normalized sensitivity indices (S'_T) versus temperature are placed in figure 4.9 for 6 parameters. The sensitivity of DO to parameters exhibits the same dependency on water temperature for the **Post-bloom** and **High-flow** scenarios. At low temperature ($T < 6\text{ }^\circ\text{C}$), the most influential parameter is the navigation reaeration (K_{navig}), while it is the maximal bacterial growth rate ($\mu_{max,hb}$) for temperature above $15\text{ }^\circ\text{C}$. Between $8\text{ }^\circ\text{C}$ and $15\text{ }^\circ\text{C}$, the bacterial parameters are the most influential without a clear dominance of one over the others.

4.4.4 Recommendations for water quality modeling

Although the effect of the oxygen reaeration rate is already recognized (Cox and Whitehead, 2005; Deflandre et al., 2006; Bailey and Ahmadi, 2014), the reaeration process requires further investigation. Especially, the anthropogenic effect, such as navigation, should be considered in water quality models used for low temperature simulations ($T < 6\text{ }^\circ$).

During algal bloom, this paper shows that the maintenance respiration ($R_{m,pp}$) is far more important than the energetic cost for growth ($R_{\mu,pp}$) in terms of autotrophic respiration (AR). Therefore, we recommend to describe

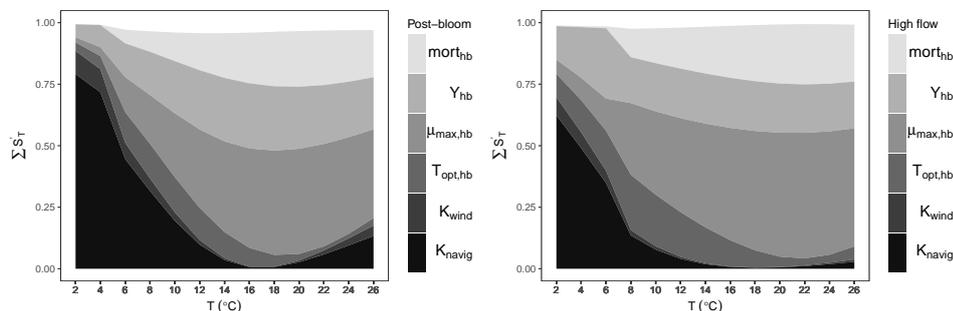


FIGURE 4.9 – The evolution of the total sensitivity indices (normalized to 100%, S_T^i) with temperature for Post-bloom (left) and High flow (right) conditions (From bottom to top : K_{navig} , K_{wind} , $T_{opt,hb}$, $\mu_{max,hb}$, Y_{hb} and $mort_{hb}$).

the algal respiration at least with two end-members ($R_{m,pp}$ and $R_{\mu,pp}$) for water quality modeling.

The sensitivity of the C-RIVE model to the ratio of carbon to chl *a* ($C/chla_{pp}$) shows the necessity of representing this ratio in water quality modeling explicitly. This should improve the capabilities of the models to simulate changing conditions, such as CSO (Combined Sewer Overflows) and their impact on the river system.

4.5 Conclusions

In this investigation, the aim was to study the parameters' influence of the C-RIVE biogeochemical model in different trophic and hydrological contexts. The time-dependent sensitivity analysis is realized using Morris method and Sobol method combined with an EOF.

- In general, it can be concluded that the Morris method and Sobol' sensitivity analysis combined with EOF are efficient methods to understand the behavior of the model parameters and improve the knowledge on the model processes.
- This paper points out that the sensitivity analysis results depend on the trophic and hydrological states. It also emphasizes the usefulness of performing sensitivity analysis on long temporal averages of model outputs.

TABLE 4.5 – First order and total sensitivity index with 95% confidence interval (Pre-bloom)

Results of the Pre-bloom scenario							
Rank	Parameters	S_i	95% CIs		$S_{T,i}$	95% CIs	
1	$\mu_{max,hb}$	0.278	0.265	0.291	0.463	0.449	0.477
2	Y_{hb}	0.236	0.224	0.248	0.347	0.336	0.358
3	$mort_{hb}$	0.111	0.101	0.121	0.233	0.224	0.242
4	$T_{opt,hb}$	0.066	0.058	0.074	0.167	0.159	0.176
5	K_{navig}	0.041	0.037	0.045	0.044	0.043	0.045
6	σ_{hb}	0.004	0.002	0.006	0.014	0.013	0.015
...
	Sum	0.751	0.695	0.806	1.288	1.244	1.333

- 6 phytoplanktonic parameters and one physical parameter represents 95% of the DO variance during a bloom. 4 bacterial parameters and 2 reaeration coefficients have been identified as most influential out of algal bloom.
- During a algal bloom, the maintenance respiration of phytoplankton ($R_{m,pp}$) and the photosynthetic parameters (α_{pp} , $P_{max,pp}$, $\eta_{chla,pp}$ and η_{water}) are the most important parameters. The results of **Pre-bloom** scenario are similar to those obtained under **Post-bloom** scenario. The 4 bacterial parameters ($\mu_{max,hb}$, Y_{hb} , $mort_{hb}$, $T_{opt,hb}$) and one reaeration coefficient (K_{navig}) are the most sensitive.
- Out of algal bloom, the water temperature plays an important role on the parameter ranking. At low temperatures ($T < 6$ ° C), the reaeration coefficients are the most sensitive, while the bacterial activities are dominant at moderate and high water temperatures.
- The difference between bloom and no bloom is mainly due to algal biomass. Hence, it is necessary to perform sensitivity analysis including boundary conditions (upstream concentrations, rejection of wastewater treatment plants etc.).
- An anthropogenic effect (ship navigation) on variation of DO has been identified and the role of this anthropogenic effect evolves with hydrological and trophic conditions.

TABLE 4.6 – First order and total sensitivity index with 95% confidence interval (Bloom)

Results of the Bloom scenario							
Rank	Parameters	S_i	95% CIs		$S_{T,i}$	95% CIs	
1	$R_{m,pp}$	0.257	0.247	0.267	0.270	0.264	0.276
2	α_{pp}	0.197	0.187	0.207	0.254	0.246	0.262
3	$P_{max,pp}$	0.173	0.165	0.181	0.220	0.214	0.226
4	$\eta_{chla,pp}$	0.088	0.082	0.094	0.115	0.112	0.118
5	$C/chla_{pp}$	0.084	0.076	0.090	0.109	0.105	0.113
6	η_{water}	0.057	0.052	0.062	0.079	0.077	0.081
7	$T_{opt,pp}$	0.003	0.000	0.006	0.027	0.026	0.028
...
	Sum	0.888	0.829	0.947	1.115	1.083	1.146

TABLE 4.7 – First order and total sensitivity index with 95% confidence interval (Post-bloom)

Results of the Post-bloom scenario							
Rank	Parameters	S_i	95% CIs		$S_{T,i}$	95% CIs	
1	$\mu_{max,hb}$	0.312	0.302	0.322	0.426	0.414	0.438
2	Y_{hb}	0.241	0.230	0.252	0.306	0.292	0.314
3	$mort_{hb}$	0.127	0.120	0.134	0.208	0.201	0.215
4	$T_{opt,hb}$	0.065	0.058	0.072	0.141	0.135	0.147
5	K_{navig}	0.071	0.066	0.076	0.075	0.073	0.077
6	σ_{hb}	0.005	0.002	0.008	0.019	0.018	0.020
7	K_{wind}	0.013	0.011	0.015	0.015	0.014	0.016
...
	Sum	0.839	0.791	0.887	1.197	1.161	1.234

4.6 Informations Supplémentaires

4.6.1 Convergence de la méthode de Morris

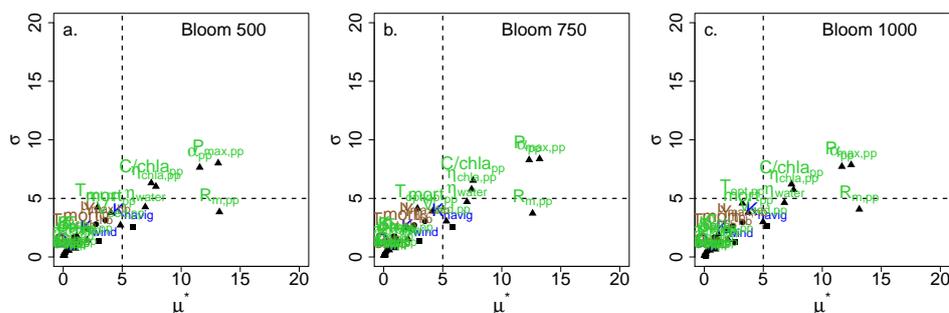


FIGURE 4.10 – Stabilité de la méthode de Morris pour le scénario Bloom.

La convergence de la méthode de Morris pour le scénario Bloom a été vérifiée. La figure 4.10 montre que le résultat de la méthode de Morris avec 500 répétitions est similaire que celle obtenue avec 1000 répétitions.

4.6.2 L'effet de la température de l'eau sur le scénario Bloom

Dans l'analyse de sensibilité précédente (cf. section 4.4.3, p.117), l'effet de la température de l'eau pour le scénario Bloom n'a pas été étudié. Afin de compléter cette analyse, une analyse de sensibilité par la méthode de Morris et la méthode de Sobol en fonction de la température de l'eau est réalisée pour le scénario Bloom (Fig. 4.11 et 4.12).

Les paramètres identifiés par la méthode de Morris (Fig. 4.11) sont globalement indépendants de la température du milieu, avec les valeurs μ^* et σ stables. Cependant, deux paramètres ont de légères variations sans pour autant remettre en cause leur caractère influent : $T_{opt,pp}$ et $R_{m,pp}$.

Afin de préciser l'importance de ces paramètres en fonction de T , une analyse de Sobol similaire à celle menée pour les bactéries hétérotrophes (cf. section 4.4.3, p.117) est ici menée pour les périodes de bloom algal. Les indices de Sobol montrent un effet similaire, mais avec un décalage de température (Fig. 4.12). Quand la température de l'eau est inférieure à 10 °C, $T_{opt,pp}$ est le plus influent. $R_{m,pp}$ ne domine plus la variation de la concentration en oxygène quand la température de l'eau dépasse 22 °C.

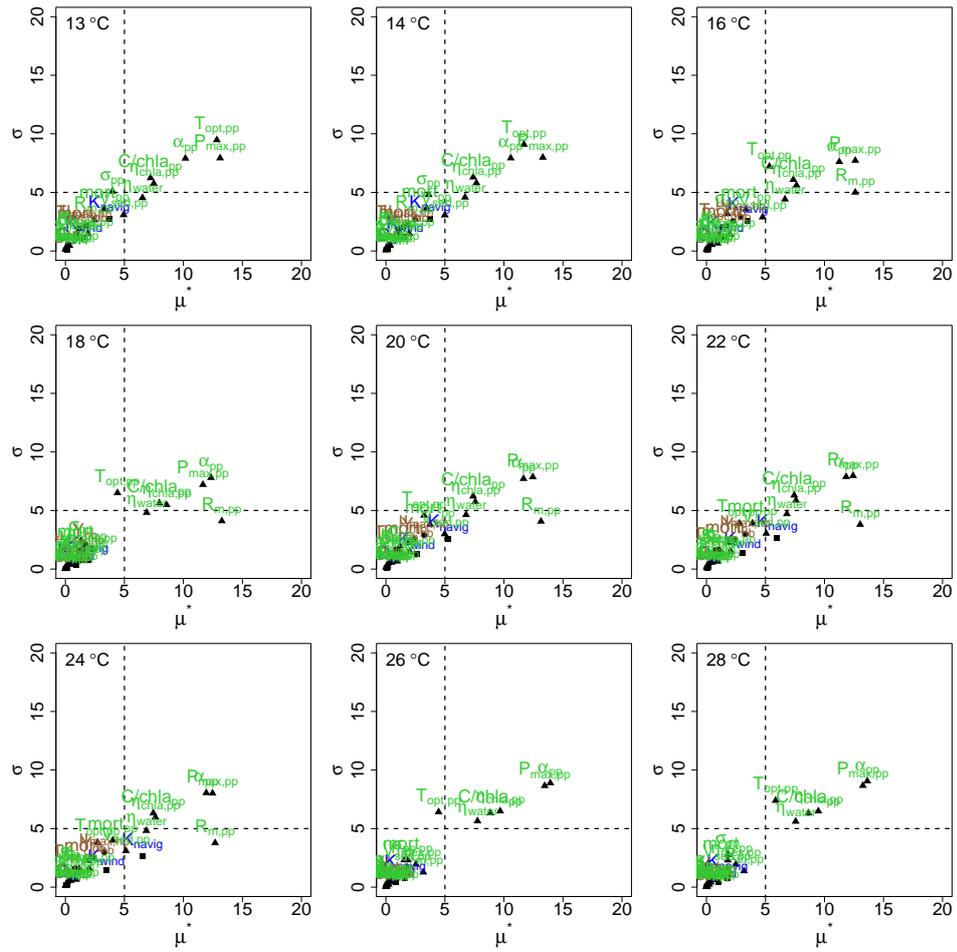


FIGURE 4.11 – L'évolution des indices de Morris en fonction de la température de l'eau pour le scénario Bloom.

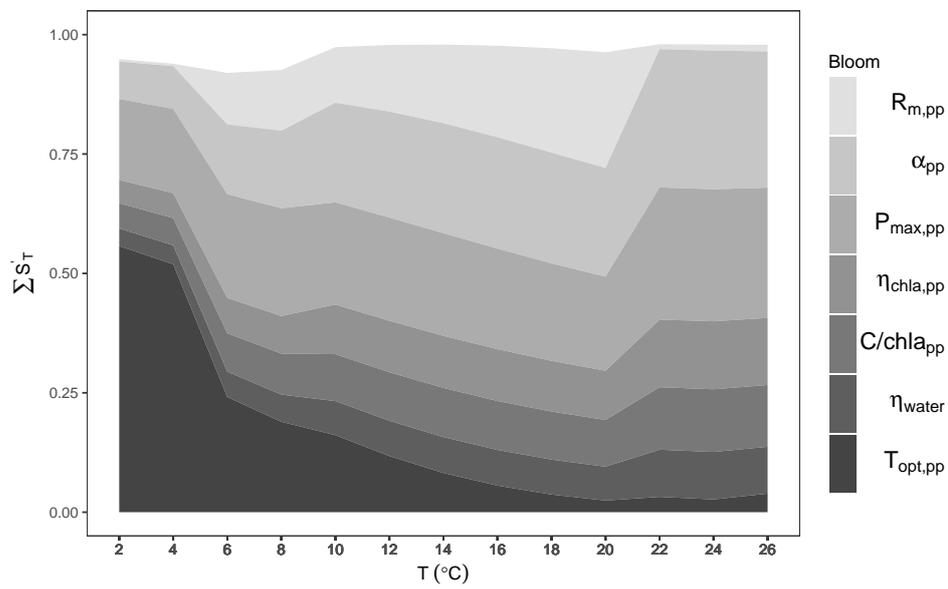


FIGURE 4.12 – L'évolution des indices totaux (normalisés en 100%) en fonction de la température de l'eau pour le scénario Bloom. Voir la table 4.4 pour la définition des paramètres.

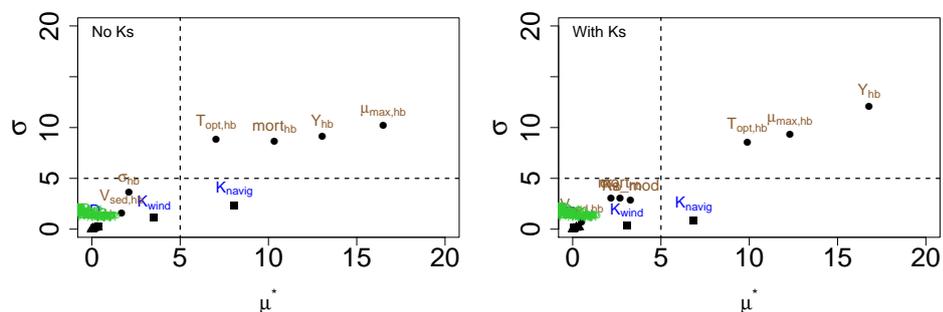


FIGURE 4.13 – Analyse de sensibilité basée sur la méthode de Morris intégrant la limitation de croissance bactérienne (K_s) pour le cas Post-bloom.

En conclusion, $T_{opt,pp}$ est le paramètre le plus influent à basse température et l'importance de $R_{m,pp}$ diminue à haute température. La sensibilité aux autres paramètres n'indique pas de dépendance à la température de l'eau.

4.6.3 Sensibilité du modèle à la limitation de la croissance bactérienne

La croissance des bactéries hétérotrophes dans RIVE est décrite par une équation de Monod limitée par la matière organique labile (MOD1) : $\mu_{hb} = \mu_{max,hb,T} \frac{MOD1}{MOD1+K_s}$. La constante de demi-saturation pour la croissance des bactéries hétérotrophes (K_s) a été supposée connue avec une valeur de 0.1 mgC/L (Garnier et al., 1992) dans le travail publié dans *Water Research* (Wang et al., 2018b).

Afin de compléter cette connaissance, une analyse de sensibilité pour le scénario Post-bloom est faite en intégrant ce paramètre (K_s). La gamme de K_s est obtenue en perturbant la valeur de référence de +/- 50% ([0.05-0.15] mgC/L).

L'analyse avec K_s ne change pas l'identification des paramètres sensibles (Fig. 4.13 et 4.14). Cependant le classement des paramètres change. Le rendement de croissance des bactéries (Y_{hb}) devient le plus influent. L'importance du taux de mortalité ($mort_{hb}$) diminue. Ces résultats montrent que K_s est peu influent. Il reste cependant à calibrer pour l'assimilation de données.

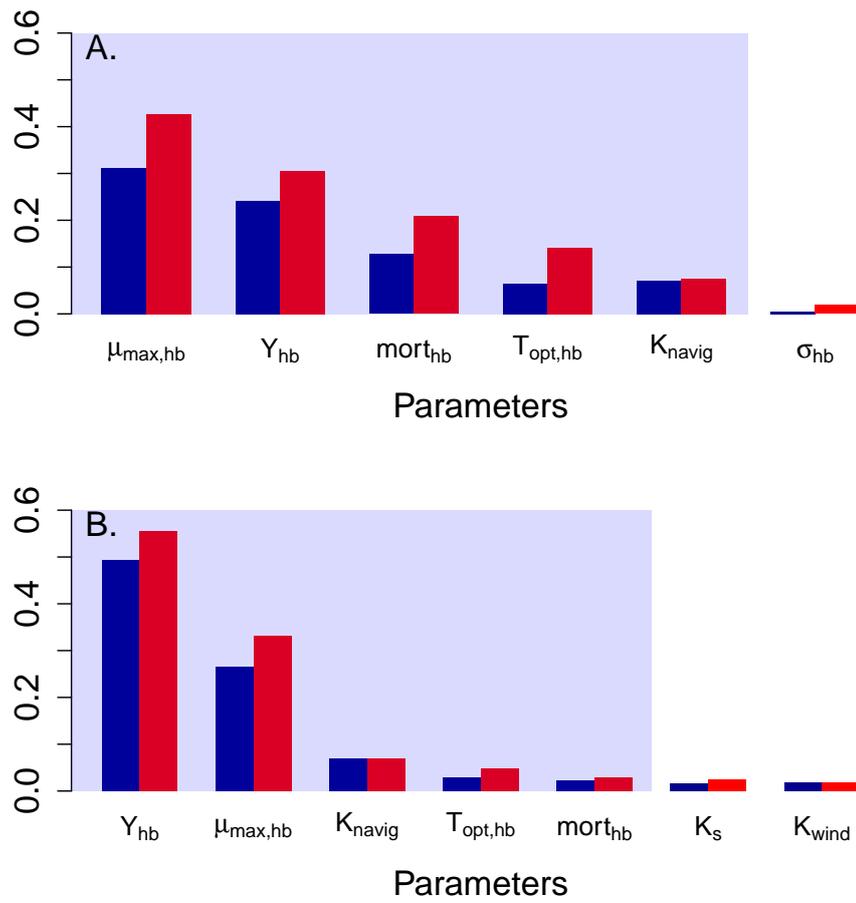


FIGURE 4.14 – Les indices de Sobol pour le scénario Post-bloom avec K_s . A. Pas de K_s B. Avec K_s . Polygone résume 95% de la variabilité des simulations.

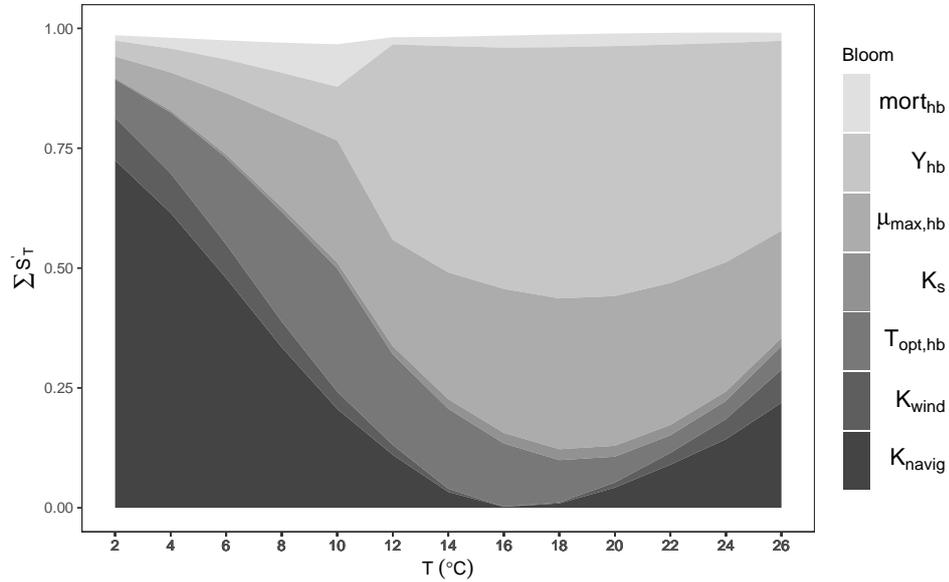


FIGURE 4.15 – L'évolution des indices totaux de Sobol (normalisés en 100%) en fonction de la température pour le scénario Post-bloom avec K_s

4.6.4 L'effet de la température sur la sensibilité à K_s

L'effet de la température de l'eau sur l'analyse de sensibilité à K_s est étudié. L'intégration de K_s dans le jeu des paramètres influents ne modifie pas les conclusions de Wang et al. (2018b) (A cet effet, comparer les figures 4.15 et 4.9). K_{navig} est toujours le plus sensible à basse température ($< 6^\circ\text{C}$). Le rendement de croissance (Y_{hb}) et le taux de croissance maximal ($mort_{hb}$) sont les deux paramètres les plus influents quand la température de l'eau dépasse 12°C . Une diminution de K_s favorise la croissance des bactéries hétérotrophes, ce qui rend le taux de mortalité ($mort_{hb}$) moins influent (Fig. 4.15). K_s lui même ne fait pas partie des paramètres les plus influents. Il n'est donc pas intégré au schéma d'assimilation de données.

Troisième partie

Assimilation de données d'oxygène par filtrage particulaire : ProSe-PA

Chapitre 5

Preuve de concept sur un cas synthétique

Résumé

Dans ce chapitre, un filtre particulaire est implémenté pour la première fois dans un modèle hydro-biogéochimique. Le nouveau modèle est appelé PROSE-PA. Les 12 paramètres sélectionnés dans le chapitre 4 sont inclus dans le filtre particulaire. Afin d'évaluer les performances de PROSE-PA, un cas synthétique représentant 66 km de la Seine depuis Suresnes jusqu'à Andrésy est construit. Ce domaine se trouve à l'aval de l'agglomération parisienne soumise à une forte pression anthropique. Les activités humaines comprennent notamment les rejets de deux stations d'épuration (Seine Aval et Seine centre) et les rejets de déversoirs d'orage en temps de pluie.

Une observation virtuelle des concentrations en O_2 toutes les 15 minutes est générée à 5 stations de mesures par PROSE-PA en utilisant les paramètres prédéfinis pour une année complète. Cette année comprend deux blooms algaux successifs caractérisés par deux propriétés physiologiques différentes (les valeurs des paramètres différentes). L'objectif de ce chapitre est d'assimiler ces concentrations en O_2 au pas de 15 min, afin de retrouver les paramètres prédéfinis et les vraies concentrations en O_2 . Cela teste l'efficacité du filtre particulaire couplé avec un modèle de qualité de l'eau pour estimer les concentrations en O_2 et les paramètres du modèle. Dans ce cas synthétique, on n'étudie que les incertitudes relatives aux paramètres, tandis que les conditions aux limites sont considérées parfaites.

Avec PROSE-PA, les vraies concentrations en O_2 sont parfaitement reconstituées. L'identifiabilité des paramètres dépend de l'état trophique du système. Les paramètres physiques, bactériens et phytoplanctoniques sont bien identifiés aux périodes où ils dominent la concentration en O_2 , PROSE-PA permet ainsi d'identifier les deux blooms algaux ainsi que le cycle nyctéméral des paramètres liés à la photosynthèse. Ce chapitre démontre l'efficacité du filtre particulière pour l'estimation des concentrations en O_2 et des paramètres en biogéochimie. PROSE-PA est maintenant opérationnel pour assimiler les vraies mesures dans le système Seine.

Abstract

The coupling of high frequency data of water quality with physically based models of river systems is of great interest for the management of urban socio-ecosystems. One approach to exploit high frequency data is data assimilation which has received an increasing attention in the field of hydrology, but not for water quality modeling so far. We present here a first implementation of a particle filtering algorithm into a community-centered hydro-biogeochemical model to assimilate high frequency dissolved oxygen data and to estimate metabolism parameters in the Seine River system. The procedure is designed based on the results of a former sensitivity analysis of the model (Wang et al., 2018b) that allows for the identification of the twelve most influential parameters all over the year. Those parameters are both physical and related to micro-organisms (reaeration coefficient, photosynthetic parameters, growth rates, respiration rates and optimal temperature). The performances of the approach are assessed on a synthetic case study that mimics 66 km of the Seine River. Virtual dissolved oxygen data are generated using time varying parameters. This paper aims at retrieving the predefined parameters by assimilating those data. The simulated dissolved oxygen concentrations match the reference concentrations. The identification of the parameters depends on the hydrological and trophic contexts and more surprisingly on the thermal state of the river. The physical, bacterial and phytoplanktonic parameters can be retrieved properly, leading to the differentiation of two successive algal blooms by comparing the estimated posterior distribution of the optimal temperature for phytoplankton growth. Finally, photosynthetic parameters' distributions following circadian cycles

during algal blooms are discussed.

5.1 Introduction

The coupling of high frequency data of water quality with physically based model of river systems is of great interest for the management of urban socio-ecosystems. One approach to exploit high frequency data is data assimilation which combines observations and short-range forecasts to estimate the distribution of the true state of a process (Wikle and Berliner, 2007). Assimilating high frequency data allows for the identification of the multiple sources of model uncertainty relative to parameters, model structure, forcing data (e.g. temperature, wind speed and solar radiation) and observations (Evensen, 2003), which limit the validation and application of water quality models (Beven, 1989; Polus et al., 2011).

A lot of data assimilation techniques exist in literature such as the variational methods (Sasaki, 1955, 1958), the Kalman filter (KF, (Kalman, 1960)), the extended Kalman filter (EKF, (Beck, 1987)), the ensemble Kalman filter (EnKF, (Evensen, 1994)) and the particle filter (PF, (Doucet et al., 2001)). A review of these methods can be found in the literature (Wikle and Berliner, 2007; Cappe et al., 2007; Särkkä, 2013). The variational methods minimizing a cost function have been widely used in numerical weather prediction (Courtier et al., 1994, 1998; Gauthier et al., 2007; Kleist and Ide, 2015; Yucel et al., 2015). The major drawbacks of variational methods are the numerical implementation complexity and the computational cost. The KF handles only linear models with Gaussian errors, which are not consistent with the hydrologic and water quality models.

To extend to nonlinear models, the EKF uses the linearized formula tangent to the forward model, which is an approximation and would be very costly to implement for high-dimensional systems (Evensen, 2003). The EnKF uses Monte Carlo samples to approximate the forecast distribution and then applies linear update formulas to obtain the posterior distribution (Evensen, 2003), while the particle filter based methods estimate the forecast and posterior distributions using discrete probability densities (Arulampalam et al., 2002) obtained via Bayes' theorem (Bayes, 1763). Since the EnKF and particle filter resolve the two major drawbacks related to the use of EKF and handle highly nonlinear models, the EnKF and particle

filter based methods have become the most commonly used sequential data assimilation techniques in hydrologic modeling for state-parameter estimation (Moradkhani et al., 2005; Weerts and El Serafy, 2006; Andreadis et al., 2007; Salamon and Feyen, 2009; Plaza et al., 2012; DeChant and Moradkhani, 2012; Vrugt et al., 2013; Shi et al., 2014; Abbaszadeh et al., 2018).

In biogeochemical oceanography, the efficiency of the ensemble-based Kalman filters has been assessed on ocean models for ocean biogeochemical state and parameter estimation (Simon and Bertino, 2012; Simon et al., 2012; Gharamti et al., 2017; Yu et al., 2018). Due to the lack of data, the difficulties in representing complex hydro-ecosystems and the computational cost, there are still relatively few applications about the state-parameter estimation by assimilating high-frequency dissolved oxygen (DO) concentrations in river systems. Pastres et al. (2003) have applied the EKF to update three parameters of a simple DO-chlorophyll model in the lagoon of Venice. Mao et al. (2009) used EKF to forecast algal blooms and dissolved oxygen dynamics in a marine ecosystem. The recent studies focus particularly on forecasting algal bloom dynamics using EnKF in river system (Kim et al., 2014) or in lakes (Huang et al., 2013; Page et al., 2018). Xue et al. (2012) applied also EnKF in the Massachusetts Bay to design the optimal monitoring sites for DO measurements.

However, as stated by Wikle and Berliner (2007, p. 10-11), "We assume that the forecast distribution can be characterized by its first two moments (or, equivalently, that it is Gaussian with mean and (estimated) variance/covariance matrix). ... However, in nonlinear cases, since Gaussianity cannot hold for all time, the EnKF must yield biased samples and estimates, even for unlimited sample sizes". The recent investigations concluded also that the assumption of a Gaussian error structure for the forecast distribution may not be realistic for hydrologic systems and phytoplankton dynamics and the authors suggested testing the particle filter based methods to overcome this problem (Plaza et al., 2012; Pasetto et al., 2012; DeChant and Moradkhani, 2012; Huang et al., 2013). The feasibility of Bayesian inference with a Particle Markov Chain Monte Carlo algorithm is tested on a simple predator-prey model in ecological research (Kattwinkel and Reichert, 2017).

In this paper, we present the PROSE-PA model, which consists in a first implementation of a particle filtering algorithm into the community-centered hydro-biogeochemical model PROSE (Even et al., 1998, 2007b; Flipo et al.,

2004; Vilmin et al., 2015b) in order to assimilate 15 min-DO data and estimate metabolisms' parameters in the Seine River system. The procedure is designed from the results of a sensitivity analysis of the biogeochemical module, C-RIVE, of the PROSE model that identified the twelve most influential parameters of the model (Wang et al., 2018b). Those parameters are both physical (water re-aeration by fluvial navigation) and physiological (for instance growth rate of heterotrophic bacteria and photosynthetic parameters of phytoplankton). The performances of PROSE-PA are assessed on a synthetic case study that mimics 66 km of the Seine River and generates virtual "observation" data of DO. The objectives of this study are to retrieve the predefined parameters used to generate the "observation" data and to distinguish two successive algal blooms by identifying different physiological properties.

The manuscript is organized as follows. The section 4.2 presents the PROSE model and the study area, followed by the PROSE-PA approach including the mathematical formulations of the particle filter, the resampling algorithm as well as the full numerical algorithm. The considered parameters the "observation" data and the input data for PROSE model are described in section 5.2.3. The numerical settings and the computational cost are given in section 5.3.1. Then, we evaluate the simulated DO concentrations by four statistical criteria (section 5.3.2) and show the identification of metabolisms' parameters in frame of state classification (section 5.3.3). The real time parameter identification is discussed in section 5.4.1, followed by the identifiability of two successive algal blooms by comparing the different physiological properties. The photosynthetic parameters' distributions following circadian cycles are described in section 5.4.3. The sensitivity of PROSE-PA performances to the observation error is discussed in section 5.4.4. To finish, a brief conclusion is given in section 5.5.

5.2 Material and methods

5.2.1 Synthetic case study

5.2.1.1 The ProSe model

The PROSE model (Even et al., 1998, 2007b; Flipo et al., 2004; Vilmin et al., 2015b), couples three libraries (a hydrodynamic library, a transport

library and a biogeochemical library), that simulate together the hydro-biogeochemical functioning of a river system. The hydraulic module solves the 1D shallow water equations with a finite volume scheme. The transport module simulates the advection and dispersion of both particulate and dissolved substances, including also water re-oxygenation by overflowing over hydraulic works. The biogeochemical library is based on the RIVE conceptual model which is a community-centered model (Billen et al., 1994; Garnier et al., 1995) (<https://www.fire.upmc.fr/rive>). The cycles of carbon, nutrients and dissolved oxygen are simulated in both water column and sediment layer. The compound exchanges between water column and benthic layers have been successively developed (Even et al., 2004, 2007a; Flipo et al., 2004, 2007b; Vilmin et al., 2015b). The PROSE model has been well validated and largely applied in the Seine River system (Even et al., 1998, 2004, 2007a; Polus et al., 2011; Raimonet et al., 2015; Vilmin et al., 2015a,b, 2016, 2018).

5.2.1.2 Study area

The study area is located downstream Paris city and consists of 66 km of the Seine River (Fig. 5.1). The bathymetry data of the channel is extracted from the full 220 km Seine model used in recent studies (Vilmin et al., 2016, 2018). This area is highly impacted by human activities including two Waste Water Treatment Plants (WWTPs, Seine Aval and Seine Centre) and two major Combined Sewer Overflows (CSOs). The Seine Aval (SAV) is the largest WWTP of Europe, which treats the effluents of over 6.5 million equivalent inhabitants (Rocher et al., 2011). During rain events, CSOs discharge large amount of suspended solids, organic matters and nutrients into the Seine River (Even et al., 2007b). Assimilating DO is therefore of great interest for the decision makers in this area.

5.2.2 Data assimilation framework using particle filtering, the ProSe-PA approach

5.2.2.1 State-space model

To begin with, we represent the PROSE model as a state-space model (Kalman, 1960). A state-space model uses inputs, state variables, outputs to describe the evolution of a system over time. In our case, it uses three equations (Eq. (5.1), (5.2), (5.3)). Let \mathbf{y} represents the state variable vector

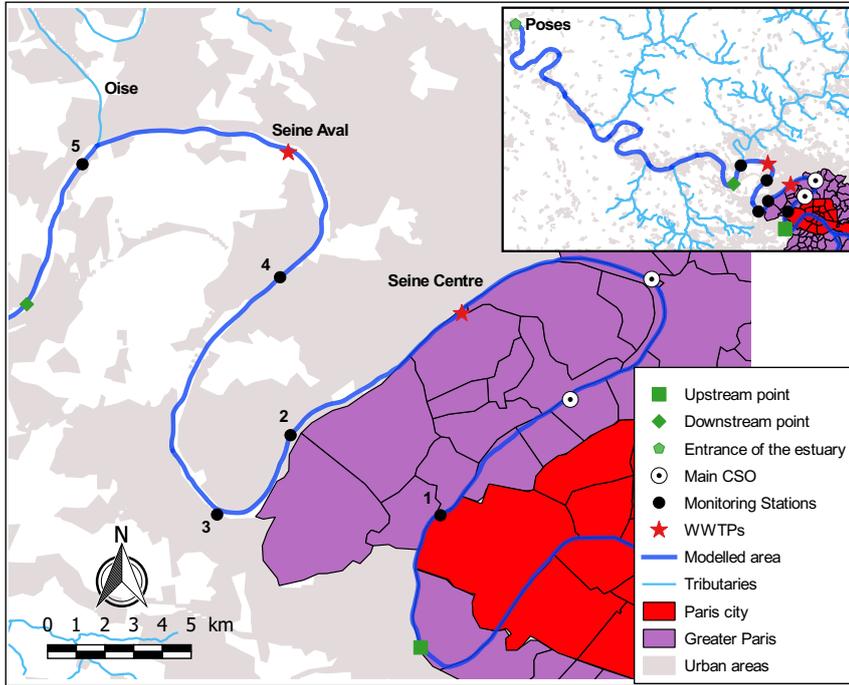


FIGURE 5.1 – The description of study area and monitoring sites. Full Seine model is shown at upper-right corner ; Monitoring stations from upstream to downstream : 1. Suresnes, 2. Chatou, 3. Bougival, 4. Sartrouville, 5. Andresy.

of the system in terms of physics (DO concentrations in this study) and \mathbf{x} be the vector of model parameters. The state variable and the model parameters are both modelled as random variables \mathbf{Y} , \mathbf{X} characterized by their probability distribution functions (pdf). \mathbf{x} and \mathbf{y} denote the realizations of random variables \mathbf{X} , \mathbf{Y} . The observation vector \mathbf{y}^* is a realization of the random variable \mathbf{Y}^* . The state-space model explains the temporal evolution of the system by assuming that it behaves as a Markov process (Markov, 1906). This means that the values at t depend on the values at $t - 1$ only :

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \boldsymbol{\eta}_t \quad (5.1)$$

$$\mathbf{y}_t = M(\mathbf{y}_{t-1}, \mu_t, \mathbf{x}_t) + \boldsymbol{\nu}_t \quad (5.2)$$

$$\mathbf{y}_t^* = h(\mathbf{y}_t) + \boldsymbol{\epsilon}_t \quad (5.3)$$

The transition of model parameter \mathbf{x} from $t - 1$ to t is described by a

Gaussian perturbation ($\boldsymbol{\eta}_t$, Eq. (5.1)). In the above equations, M is the forward model (in our case PROSE) and h is the observation operator relating the forecasted state variable (\mathbf{y}_t) to the observation (\mathbf{y}_t^*). In other words, h denotes the selection of model cells where observations are available. \mathbf{y}_{t-1} stands for the posterior state variable at previous time step ($t-1$) or equivalently for the prior state variable at current time step (t). The symbols $\boldsymbol{\mu}_t$ and \mathbf{x}_t represent respectively the prior forcing data (e.g. temperature, wind speed and solar radiation) and the prior model parameters at time step t . The variables $\boldsymbol{\nu}_t$ and $\boldsymbol{\epsilon}_t$ characterize the unknown model and measurement errors respectively. The PROSE model aims at conserving the mass balance. The model does it with accuracy, the relative errors being in the order 10^{-5} or 10^{-6} . This is far below observation errors. Therefore, no model errors are considered ($\boldsymbol{\nu}_t = 0$) in our case.

Note that the nomenclature we use here is not canonical. In dynamical system literature, the parameters would have been called the "states" because they are the hidden dynamic variables that govern the system's behaviour. In forward modelling and especially hydrology, the state variables designate the variables that are computed by the model, while the parameters represents the time varying functional inputs of the numerical model. We consider the latter as the reader is more likely to be accustomed to this definition. Please note that, with this definition, the input fluxes at the boundary of the model are not called parameters, but boundary conditions.

5.2.2.2 Bayesian inference

We seek to estimate the posterior pdf of both the random variable \mathbf{Y} and the parameter \mathbf{X} conditionally to the observations \mathbf{Y}^* . To simplify the writing, we define the state vector $\mathbf{Z} = [\mathbf{Y}^T, \mathbf{X}^T]^T$ and its realizations $\mathbf{z} = [\mathbf{y}^T, \mathbf{x}^T]^T$. Therefore, our goal can be reformulated in finding the conditional pdf of \mathbf{Z} knowing \mathbf{y}^* noted $f_{\mathbf{Z}}(\mathbf{z}|\mathbf{y}^*)$. Through Bayes' theorem (Bayes, 1763), the posterior pdf $f_{\mathbf{Z}}(\mathbf{z}|\mathbf{y}^*)$ can be deduced from the product of the prior pdf ($f_{\mathbf{Z}}(\mathbf{z})$) and the likelihood distribution ($f_{\mathbf{Y}^*}(\mathbf{y}^*|\mathbf{z})$), up to a normalizing constant independent from \mathbf{Z} :

$$f_{\mathbf{Z}}(\mathbf{z}|\mathbf{y}^*) \propto f_{\mathbf{Y}^*}(\mathbf{y}^*|\mathbf{z})f_{\mathbf{Z}}(\mathbf{z}) \quad (5.4)$$

The prior pdf $f_{\mathbf{z}}(\mathbf{z})$ gives the prior knowledge on \mathbf{z} before the measurements are taken. The likelihood $f_{\mathbf{Y}^*}(\mathbf{y}^*|\mathbf{z})$ describes the pdf of the observation data given the model predictions.

5.2.2.3 Sequential form

The formulation (5.4) does not relate to time. To introduce a sequential form for data assimilation, we define the temporal trajectories of the random variables \mathbf{Z} and \mathbf{Y}^* , $\mathbf{Z}_{1:t} \equiv \{\mathbf{Z}_1, \dots, \mathbf{Z}_t\}$ and $\mathbf{Y}_{1:t}^* \equiv \{\mathbf{Y}_1^*, \dots, \mathbf{Y}_t^*\}$, discretized over time by sequences of state $(\mathbf{Z}_i, \mathbf{Y}_i^*, i = 1, \dots, t)$, as well as their realizations $\mathbf{z}_{1:t}$, $\mathbf{y}_{1:t}^*$. The posterior pdf of the trajectory $\mathbf{Z}_{1:t}$ given observation $\mathbf{y}_{1:t}^*$ writes $f_{\mathbf{z}_{1:t}}(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*)$.

Next, we can write the posterior pdf $f_{\mathbf{z}_{1:t}}(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*)$ under a recursive form (Arulampalam et al., 2002; Doucet et al., 2001; Wikle and Berliner, 2007; Särkkä, 2013) owing to the Markovian nature of the process. We simplify the notations rewriting a pdf $f_{\mathbf{z}}(\mathbf{z})$ into $f(\mathbf{z})$.

$$f(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*) \propto f(\mathbf{y}_t^*|\mathbf{z}_{1:t}, \mathbf{y}_{1:t-1}^*)f(\mathbf{z}_{1:t}|\mathbf{y}_{1:t-1}^*) \quad (5.5)$$

$$= f(\mathbf{y}_t^*|\mathbf{z}_t)f(\mathbf{z}_t|\mathbf{z}_{1:t-1}, \mathbf{y}_{1:t-1}^*)f(\mathbf{z}_{1:t-1}|\mathbf{y}_{1:t-1}^*) \quad (5.6)$$

$$= f(\mathbf{y}_t^*|\mathbf{z}_t)f(\mathbf{z}_t|\mathbf{z}_{t-1})f(\mathbf{z}_{1:t-1}|\mathbf{y}_{1:t-1}^*), \quad (5.7)$$

where we use first the Bayes' theorem to get equation (5.5), then the Bayes' theorem and the Markov property for equation (5.6) and the Markov property to obtain the sequential relation (5.7). We propagate \mathbf{z}_t using the evolution equations (5.1) and (5.2) in order to model the distribution $f(\mathbf{z}_t|\mathbf{z}_{t-1})$. Note in case of a deterministic formulation ($(\boldsymbol{\nu}_t, \boldsymbol{\eta}_t) = (0, 0)$ in equations (5.1) and (5.2)), $f(\mathbf{z}_t|\mathbf{z}_{t-1})$ is a Dirac measure. Consequently, we can access the posterior pdf $f(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*)$ of the trajectory $\mathbf{Z}_{1:t}$ by updating it at each time step. This will be done through the use of sampling techniques.

5.2.2.4 The particle filter

The particle filter aims at approximating the posterior pdf $f(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*)$ of the trajectory $\mathbf{Z}_{1:t}$ knowing $\mathbf{y}_{1:t}^*$ by a set of particles (simulations) associated with weights $(\boldsymbol{\omega}_t)$. It is generally impossible to sample from $f(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*)$ directly. To address this problem, sequential importance sampling has been suggested (Doucet et al., 2000; Liu, 2001). For the ease of reading, the

principle of the importance sampling and the definition of the importance weights are detailed in 5.6.1. We give here the weight update formula for each particle directly, which is a consequence of the sequential decomposition of the posterior of the trajectory obtained in (Eq. (5.7)) .

$$\omega_t^i = f(\mathbf{y}_t^* | \mathbf{z}_t^i) \omega_{t-1}^i \quad (5.8)$$

$$\hat{\omega}_t^i = \frac{\omega_t^i}{\sum \omega_t^i} \quad (5.9)$$

where $f(\mathbf{y}_t^* | \mathbf{z}_t^i)$ denotes the likelihood function, which quantifies how \mathbf{y}_t^* is likely to be observed given \mathbf{z}_t^i at time t . ω_t^i and ω_{t-1}^i stand for the posterior and prior weights at time t . In other words, the posterior weight at time $t-1$ serves as a prior weight at time t . We also define the normalized importance weight, $\hat{\omega}_t^i$, associated with the particle i .

In practice, it is usually not necessary to estimate the posterior pdf $f(\mathbf{z}_{1:t} | \mathbf{y}_{1:t}^*)$ of the trajectory $\mathbf{Z}_{1:t}$ knowing $\mathbf{y}_{1:t}^*$. We are interested here in the marginal distribution of $f(\mathbf{z}_{1:t} | \mathbf{y}_{1:t}^*)$, called filtering distribution $f(\mathbf{z}_t | \mathbf{y}_{1:t}^*)$. It represents the distribution of the state variables knowing all the past observations as well as the current one. In this paper, we note the filtering posterior pdf at time t as $f(\mathbf{z}_t | \mathbf{y}_{1:t}^*)$ which can be approximated as,

$$f(\mathbf{z}_t | \mathbf{y}_{1:t}^*) \approx \sum_{i=1}^N \hat{\omega}_t^i \delta_{\mathbf{z}_t^i} \quad (5.10)$$

where $\delta(\cdot)$ is a Dirac delta function and N denotes the number of particles. That is the filtering distribution is approximated by a discrete distribution, whose probability mass function is defined by the normalized importance weights and charges the particles.

If we assume that the observation error is Gaussian, then the likelihood probability of each particle can be calculated using the pdf of the multivariate normal distribution :

$$\begin{aligned} \ln L(\mathbf{y}_t^* | \mathbf{z}_t^i) &= -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Sigma|) - \frac{1}{2} (\mathbf{y}_t^* - h(\mathbf{y}_t^i))^T \Sigma^{-1} (\mathbf{y}_t^* - h(\mathbf{y}_t^i)) \\ f(\mathbf{y}_t^* | \mathbf{z}_t^i) &= \frac{L(\mathbf{y}_t^* | \mathbf{z}_t^i)}{\sum_{i=1}^N L(\mathbf{y}_t^* | \mathbf{z}_t^i)} \end{aligned}$$

where m is the number of observation sites and Σ is the error covariance matrix of the observations.

5.2.2.5 Resampling algorithm

A common problem when using particle filtering is the degeneracy phenomena. This occurs when almost all the particles have zero or very small importance weights. Only a few particles have significant importance weights. Therefore, the discrete probability densities cannot represent the posterior pdf of a given state adequately. A resampling procedure reduces the degeneracy effect. The basic idea of the resampling is to discard particles that have a small weight and to duplicate particles with a large weight. Generally, it is not necessary to perform resampling at every time step, but only when necessary. A way to monitor the need of resampling is to estimate the “effective” number of particles by the effective sample size (N_{eff}). The N_{eff} cannot be evaluated exactly but it can be estimated by (Kong et al., 1994; Doucet et al., 2000) :

$$\widehat{N}_{eff} = \frac{1}{\sum_{i=1}^N (\hat{\omega}_t^i)^2} \quad (5.12)$$

When N_{eff} is below a threshold ($N_{thres} = \alpha \cdot N$) predefined by the user, resampling is performed. After the resampling step, all weights are reset to $1/N$. As the particles having an important weight may be copied many times, this results in sample impoverishment problem. To maintain the diversity of the ensemble, a random perturbation is added to the parameters’ value after the resampling step (eq. (5.13)).

$$\mathbf{x}_{t+1}^i = \mathbf{x}_{t,resampling}^i + \eta_t^i \quad \eta_t^i \sim N(0, (s \cdot \Phi)^2) \quad (5.13)$$

where s is a percent perturbation predefined by the user (0.1 in this study) and Φ denotes the parameter range.

The resampling technique used in this work is referred to as systematic resampling. The procedure of systematic resampling is thoroughly described in the literature (Kitagawa, 1996; Moradkhani et al., 2005; Li et al., 2015).

5.2.2.6 Numerical algorithm

The coupling of the particle filtering algorithm with the PROSE model is called PROSE-PA. PA stands for Parallel computing and data Assimilation. A full description of the PROSE-PA approach is given below (Fig. 5.2).

5.2.3 Description of the synthetic case study

5.2.3.1 Parameters considered and virtual “observation” data

Twelve parameters of the PROSE model were identified in different hydrological and trophic contexts in a previous work (Wang et al., 2018b). Those parameters are both physical and physiological (Table 5.1).

Five monitoring stations are located in the study area (Suresnes, Chatou, Bougival, Sartrouville, and Andresy, from upstream to downstream respectively. Fig. 5.1). The assimilation period is the year 2011 when algal blooms occurred in March and July in the Seine River (Vilmin et al., 2016). The two successive algal blooms are characterized by different physiological properties, especially the optimal temperature for growth of phytoplankton (Vilmin, 2014). A high-frequency DO (dissolved oxygen) dataset (\mathbf{y}^{ref} , every 15 minutes), which corresponds to the frequency of the real measurement, is generated using predefined parameters (Table 5.1). Three phytoplanktonic parameters are modified manually at day 139 in order to represent the two different algal blooms ($P_{max,pp}$, $R_{m,pp}$ and $T_{opt,pp}$, see Tab. 5.1 for parameter definition). In this case study, the “observation” data are obtained adding a Gaussian error to the reference dataset with a mean of zero and a standard deviation of $0.01 \times \mathbf{y}^{ref}$ ($\mathbf{y}_t^* = \mathbf{y}_t^{ref} + \boldsymbol{\epsilon}_t$, $\boldsymbol{\epsilon}_t \sim N(0, (0.01 \times \mathbf{y}_t^{ref})^2)$). A standard deviation of $0.01 \times \mathbf{y}^{ref}$ ensures that the 95% observation errors are less than $0.02 \times \mathbf{y}^{ref}$, which is coherent with the sensors used in the Seine River system (Garnier et al., 2019). The reference dataset and predefined parameters are assumed to be true and are assessed via data assimilation. In this study, no model errors are considered ($\boldsymbol{\nu}_t = 0$, see Eq. (5.2)). The major aims of this work are to retrieve the reference DO concentrations and the dominant predefined parameters identified by Wang et al. (2018b) over time.

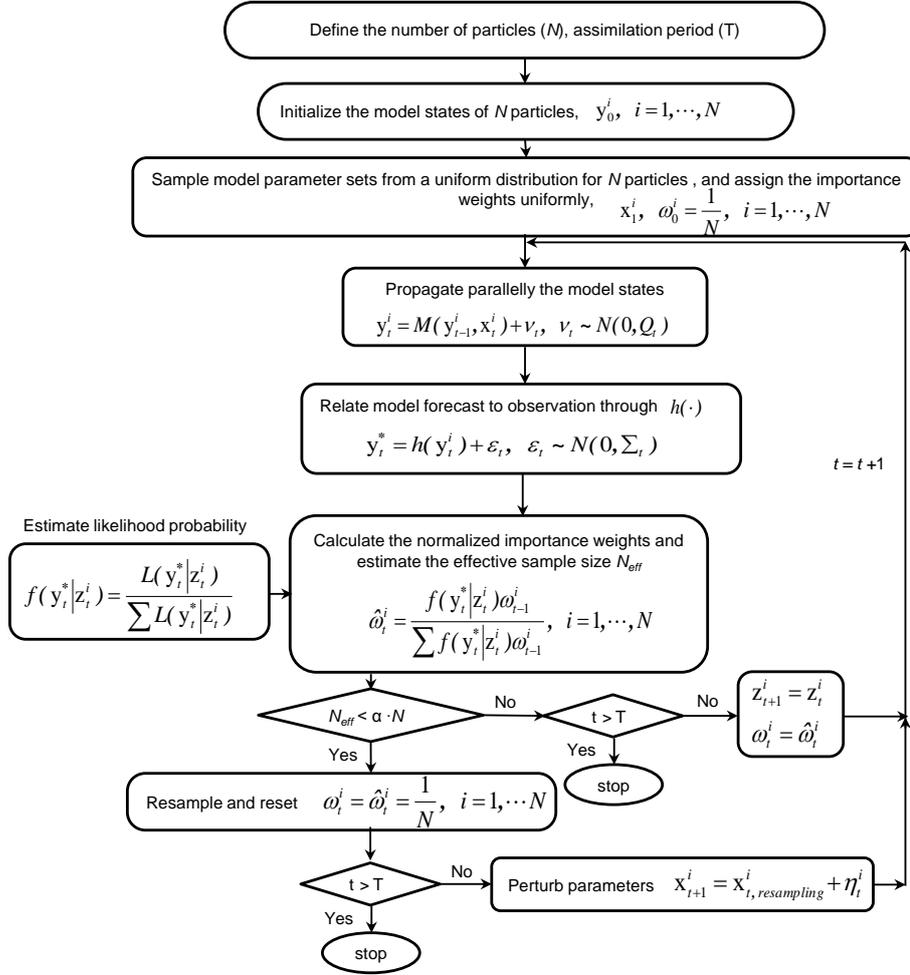


FIGURE 5.2 – The flowchart of PROSE-PA approach. The state \mathbf{z}^i denotes i th particle or a realization of random variable $\mathbf{Z} = [\mathbf{Y}^T, \mathbf{X}^T]^T$. \mathbf{y}^i and \mathbf{x}^i signify model state and parameter set of particle i . Q_t and Σ_t correspond to the model error covariance matrix and the observation error covariance matrix respectively. In this work, no model errors are considered, $\boldsymbol{\nu}_t = 0$.

5.2.3.2 Input data for ProSe model

Apart from hydraulic (time varying river discharge and CSOs data), geometric data (river channels), the PROSE model also requires time varying concentrations of the biogeochemical variables (micro-organism biomass, dissolved oxygen, nutrients, organic matters, suspended solids) entering the system and the meteorological data (solar irradiance, water temperature and wind speed). These data are used as boundary conditions of the PROSE model. The quantification of uncertainties on boundary conditions is an ambitious topic that extends far beyond the scope of this paper and will require the study of real systems. Here we only investigate uncertainties on parameters.

TABLE 5.1 – True parameters considered in data assimilation (Parameter range from Wang et al. (2018b))

Parameters	Description	Range	True March	True July	Unit
Physical parameters					
η_{water}	Light extinction coefficient for pure water	[0.2, 0.8]	0.32	0.32	[m^{-1}]
K_{navig}	Reaeration coefficient related to the navigation	[0, 0.05]	0.025	0.025	[$m.h^{-1}$]
Bacterial parameters					
$\mu_{max,hb}$	Maximum growth rate of bacteria	[0.01, 0.13]	0.04	0.04	[h^{-1}]
$mort_{hb}$	Maximal mortality rate of bacteria	[0.01, 0.08]	0.02	0.02	[h^{-1}]
$T_{opt,hb}$	Optimal temperature for bacterial growth	[10, 35]	25	25	[° C]
Y_{hb}	Bacterial growth yield	[0.03, 0.5]	0.25	0.25	[-]
Phytoplanktonic parameters					
α_{pp}	Photosynthetic capacity	[0.0003, 0.0018]	0.0012	0.0012	[$m^2.s.\mu E^{-1}.h^{-1}$]
$\eta_{chla,pp}$	Light extinction coefficient by algal self-shading	[0.006, 0.054]	0.02	0.02	[$L.\mu gchla^{-1}.m^{-1}$]
$Chla/C_{pp}$	Ratio of chlorophyll <i>a</i> to carbon	[50, 7.69]	28.57	28.57	[$\mu gchla.mgC^{-1}$]
$P_{max,pp}$	Maximum photosynthesis rate	[0.09, 0.546]	0.2	0.4	[h^{-1}]
$R_{m,pp}$	Respiration of maintenance	[0.001, 0.021]	0.002	0.01	[h^{-1}]
$T_{opt,pp}$	Optimal temperature for growth of phytoplankton	[10, 37]	13	23	[° C]

5.2.3.3 Qualitative description of the simulation period : state classification

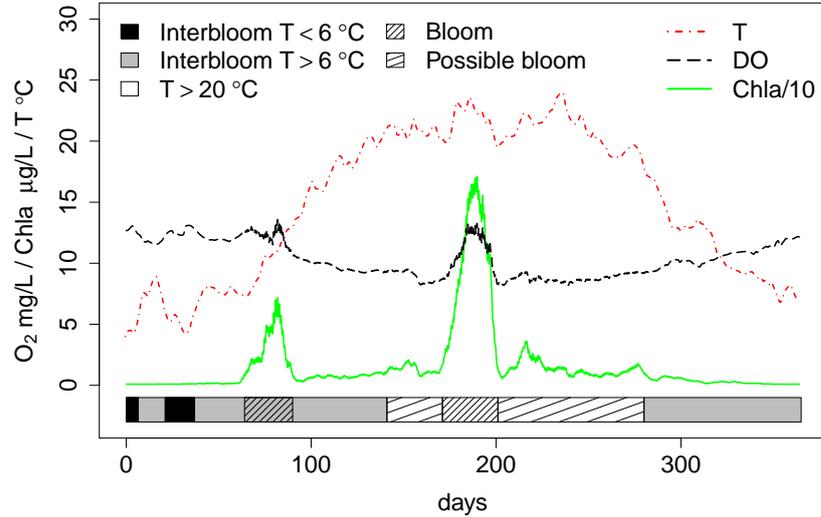


FIGURE 5.3 – State classification of synthetic case study over time, concluded by Wang et al. (2018b). The red line (dash-dotted) represents water temperature; The black line (dashed) denotes true DO concentrations at the Bougival station; The green line (solid) shows the evolution of *chl a* concentrations. The polygons at the bottom correspond to the different periods. The water temperature increases from black to white.

According to the former sensitivity analysis of the biogeochemical module, C-RIVE, of the PROSE model (Wang et al., 2018b), we can classify the year 2011 into different periods (Fig. 5.3). The polygons at the bottom show this classification (Fig. 5.3). The water temperature increases from black to white (color gradient) and the line shaded polygons represent algal bloom periods ($C_{chl a} > 10 \mu\text{g/L}$). Wang et al. (2018b) concluded that at low temperatures (Interbloom $T < 6 \text{ }^\circ\text{C}$, black polygons), the river system is controlled by the reaeration process (K_{navig}). At moderate and high water temperatures, the maximum growth rate of bacteria ($\mu_{max,hb}$) is most influential to variation of DO concentrations when no algal bloom occurs (Interbloom $T > 6 \text{ }^\circ\text{C}$, gray polygons). Once algal bloom occurs, phytoplanktonic parameters related to respiration ($R_{m,pp}$) and to photosynthesis (α_{pp} , $P_{max,pp}$ etc.) are dominant. A supplementary sensitivity analysis by

Sobol method (Sobol, 1993) during algal bloom shows that the optimal temperature ($T_{opt,pp}$) plays an important role on the growth of phytoplankton and that the influence of $R_{m,pp}$ (respiration of maintenance) depends on the water temperature (Fig. 4.12). When the water temperature exceeds 20 °C, the total sensitivity index of $R_{m,pp}$ decreases dramatically (Fig. 4.12).

5.2.4 Numerical setting and computational cost

According to a test of the number of particles (not shown here), we select 500 particles in this work which is sufficient to i) match the observed oxygen concentrations and ii) identify the posterior pdfs of sensible parameters. The OpenMP Application Programming Interface (API) is implemented in the code to simulate the 500 particles in parallel. For a 1 year-simulation period (365 days) at a 15-min time step, the computation takes 1.13 days with 20 processors (Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40 GHz). A resampling threshold of $\alpha = 0.3$ is chosen, corresponding to a minimum effective sample size of 150.

5.3 Results : Oxygen simulation and parameter identification

5.3.1 Effective sample size for resampling procedure

Although the effective sample size (N_{eff}) reduces fast after two months, synchronously with the start of the first algal bloom (day 64), the parameter perturbation allows the restoration of particle diversity (Fig. 5.4, $N_{eff} > 450$ after resampling), which indicates a proper exploration of the parameter space by the algorithm. This is achieved by selecting a configuration of the Gaussian random walk (Pearson, 1905) through its standard deviation. After several trials (not shown here), the standard deviation of the random walk has been set to 0.1 times the parameters' range (eq. (5.13)).

5.3.2 Almost perfect DO simulation with ProSe-PA

Figure 5.6 shows simulated DO concentrations at three downstream monitoring stations. Those stations are selected because they are less sensitive to input boundary conditions and therefore permit to evaluate the conceptual structure of the code and the conceptualization of processes. The model

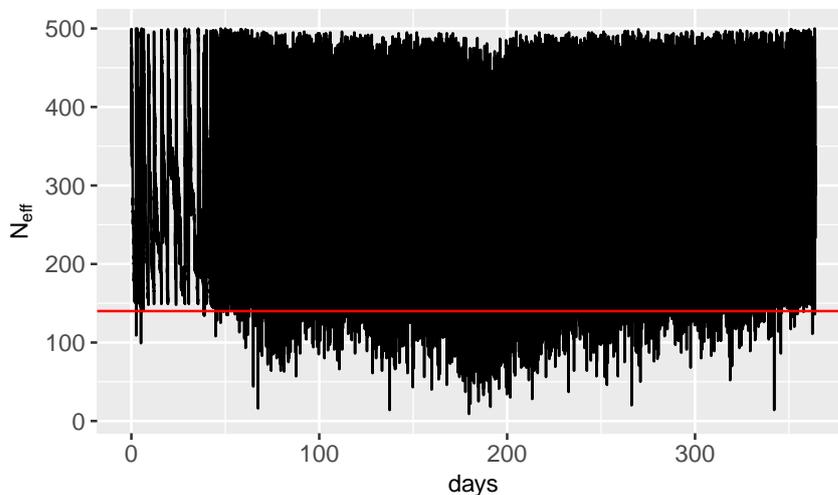


FIGURE 5.4 – The evolution of effective sample size during assimilation. Red line corresponds to the minimum effective sample size 150. Frequent resamplings ensure a proper searching of the parameter space.

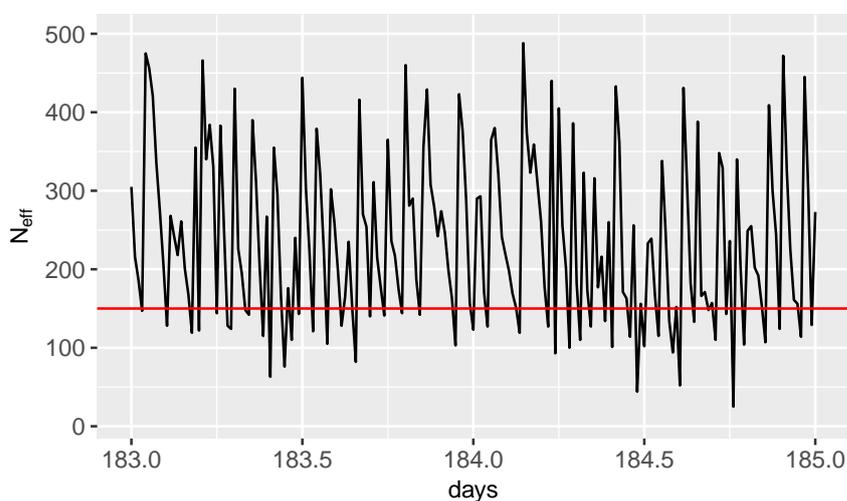


FIGURE 5.5 – The evolution of effective sample size for days 183-185. Red line corresponds to the minimum effective sample size 150. Frequent resamplings ensure a proper searching of the parameter space.

performances are evaluated by four statistical criteria (RMSE, Root-Mean-Square-Error ; MAE, Maximum Absolute Error ; NSE, Nash-Sutcliffe Efficiency ; R, Correlation coefficient).

The ensemble weighted average DO concentrations at all stations match the reference DO concentrations. Two algal blooms are well retrieved. The first starts at 64th day and the second around 171st day (Fig. 5.6). The maximum RMSE and MAE values between ensemble weighted mean and reference DO data are obtained at the Suresnes station with an error of 0.035 mgO₂/L and an error of 0.321 mgO₂/L respectively (Tab. 5.2). Furthermore, all NSEs and correlation coefficients (R) are close to 1, which signifies a perfect match of simulated DO concentration to the reference DO data. In addition, the 95 percentile confidence intervals are very narrow, which means a perfect simulation also.

TABLE 5.2 – Statistic errors of the weighted average of 500 particles/simulations and the true state at five monitoring sites

Stations	RMSE	NSE	MAE	R
Suresnes	0.0345	0.9995	0.3209	0.9998
Chatou	0.0173	0.9999	0.1347	0.9999
Bougival	0.0192	0.9998	0.1627	0.9999
Sartrouville	0.0201	0.9998	0.1765	0.9999
Andresy	0.0163	0.9998	0.1350	0.9999

RMSE : Root-Mean-Square-Error

NSE : Nash-Sutcliffe Efficiency

MAE : Maximum Absolute Error

R : Correlation coefficient

5.3.3 Parameter identification in the frame of the state classification

In this section, the identification of physical, bacterial and phytoplanktonic parameters is shown in plots displaying normalised weights over time (Fig. 5.7,5.8,5.9) for each parameter of interest. The daily normalised weights over time permits to represent the evolution of the posterior pdf of each parameter. Each parameter range is divided in 20 intervals. The sum of daily normalised weights in each interval which approximates the posterior pdf are shown by a image plot. The dashed black line represents the predefined value of parameter used to generate “observation” data.

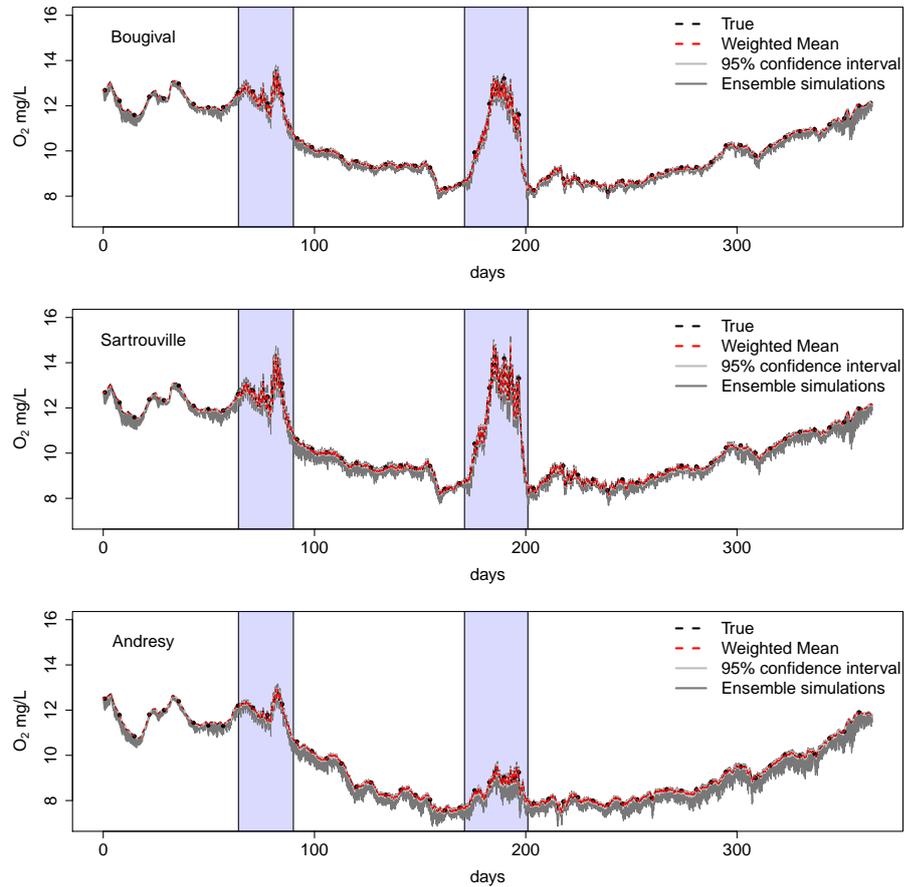


FIGURE 5.6 – Simulated DO concentrations at Bougival, Sartrouville and Andresy stations. The dashed red lines denote ensemble weighted means. The true states are represented by dashed black lines. To clarify the illustration, weekly true DO concentrations are shown by black points. The gray areas correspond to the ensemble simulations (dark gray areas) and the 95 percentile confidence intervals (light gray areas). The two algal bloom periods are shown by blue polygons.

5.3.3.1 Physical parameters

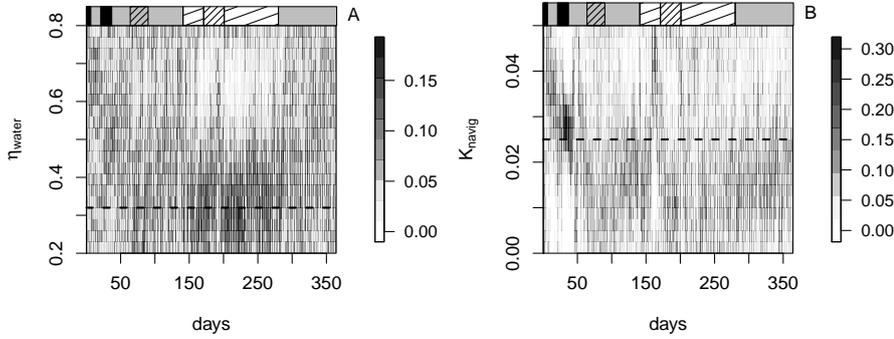


FIGURE 5.7 – Daily normalised weights of physical parameters : light extinction coefficient for pure water (η_{water}) and reaeration coefficient due to navigation (K_{navig}). The dashed line represents the predefined true value. Line shaded polygons correspond to algal bloom periods and water temperature increases from black to white (color gradient). See Fig. 5.3 for detailed state classification.

Two physical parameters are assimilated in this paper, the light extinction coefficient for pure water (η_{water}) and the reaeration coefficient related to the navigation (K_{navig}). Albeit light extinction by clear water (η_{water}) is a physical parameter, it is an important control for phytoplankton growth (Wang et al., 2018b). It is not very well determined most of the time, except during algal blooms when the posterior pdfs focus around the reference value of 0.32 m^{-1} (see the line shaded polygons in Fig. 5.7A). K_{navig} controls the reaeration process and is the most influential parameter to variation of DO in winter when the water temperature is below $6 \text{ }^\circ\text{C}$ (Wang et al., 2018b). In this period (see the second black polygon in Fig. 5.7B), the particle filter achieves a very fine identification of its pdf. Seeing other periods of the year, its effect remains negligible and logically it remains unidentified.

5.3.3.2 Bacterial parameters

In the previous sensitivity analysis (Wang et al., 2018b), the maximum growth rate of bacteria ($\mu_{max,hb}$) is the first ranked parameter out of algal bloom with a moderate water temperature ($T > 6 \text{ }^\circ\text{C}$) and the bacterial growth yield (Y_{hb}) is identified as the second influential parameter for inter algal bloom periods (both low and moderate water temperatures). $\mu_{max,hb}$

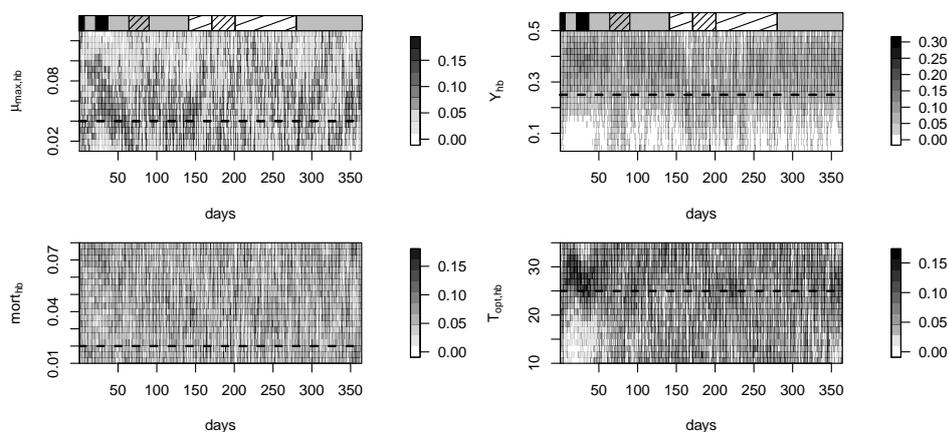


FIGURE 5.8 – The normalised importance weights of bacterial parameters. See Tab. 5.1 for parameter definition. The dashed line represents the pre-defined true value. Line shaded polygons correspond to algal bloom periods and water temperature increases from black to white (color gradient). See Fig. 5.3 for detailed classification.

can be retrieved well for several periods. These periods match globally the gray polygons which correspond to inter algal bloom periods with a water temperature $T > 6$ °C (Fig. 5.8). Y_{hb} is slightly overestimated during the simulation, except for the algal blooms periods (see the line shaded polygons in Fig. 5.8). The optimal temperature for bacterial growth ($T_{opt,hb}$) can be determined at the beginning of simulation. However, the maximal mortality rate of bacteria ($mort_{hb}$) spreads uniformly over the parameter range during the data assimilation period.

5.3.3.3 Phytoplanktonic parameters

It can be observed that the respiration of maintenance ($R_{m,pp}$), the photosynthetic capacity (α_{pp}), the light extinction coefficient by algal self-shading ($\eta_{chla,pp}$), and the optimal temperature for growth of phytoplankton ($T_{opt,pp}$) are well estimated during the first algal bloom (see the gray line shaded polygon in Fig 5.9), while the maximum photosynthesis rate ($P_{max,pp}$) is overestimated (Fig. 5.9). Similarly, almost all the phytoplanktonic parameters can be retrieved during the second algal bloom (see the clear line shaded polygons in Fig. 5.9), but not for $R_{m,pp}$ and $Chla/C_{pp}$. The ratio of chlorophyll *a* to carbon ($Chla/C_{pp}$) is unidentified during the simulation,

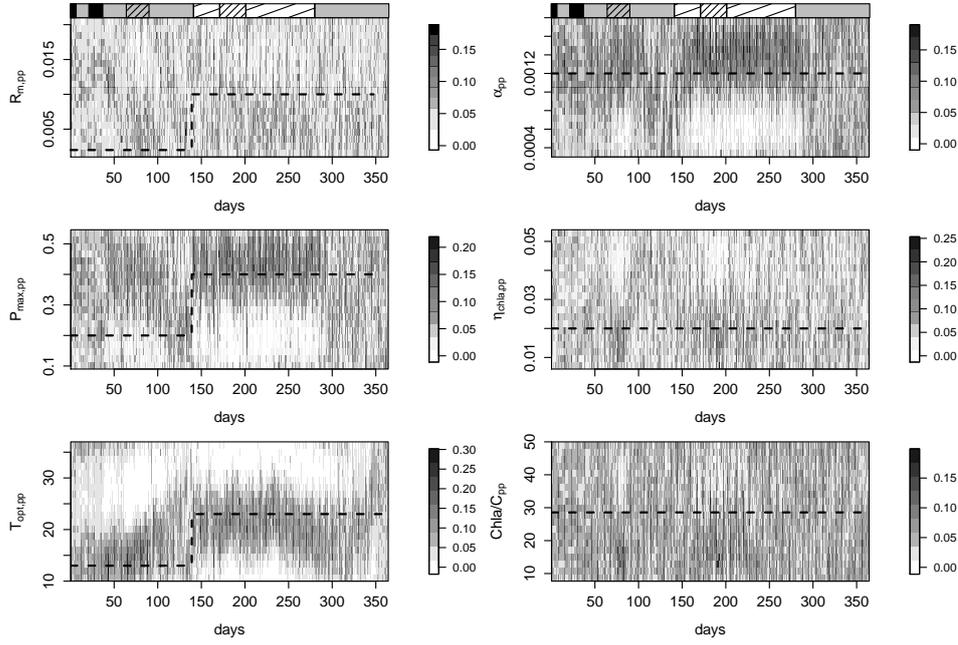


FIGURE 5.9 – The normalised importance weights of phytoplanktonic parameters. See Tab. 5.1 for parameter definition. The dashed line represents the predefined true value. Line shaded polygons correspond to algal bloom periods and water temperature increases from black to white (color gradient). See Fig. 5.3 for detailed classification.

because $Chla/C_{pp}$ has little effect on the variation of DO concentrations (Wang et al., 2018b). The particle filter captures the change of phytoplanktonic properties after 139 day for $P_{max,pp}$ and $T_{opt,pp}$. During algal blooms, five among seven parameters including η_{water} can be retrieved. Even though the two other parameters are not well estimated, we obtain satisfying simulated DO concentrations during algal blooms (Fig. 5.6). Although α_{pp} , η_{water} , $\eta_{chla,pp}$ and $T_{opt,pp}$, are always identified by the algorithm during blooms, this is not the case for $R_{m,pp}$ and $P_{max,pp}$ which are supposed to be important during algal bloom (Wang et al., 2018b). This is further discussed in the next section.

5.4 Discussion

5.4.1 Specifications for real time parameter identification

As shown by the results, the identification of the different parameters is not always ensured during the data assimilation period. This result can be explained by the parameters' influence in contrasted hydrological and trophic contexts. One parameter that is dominant on DO concentration can be well estimated, while the other uninfluential parameters have a relatively large uncertainty. In this section, we discuss the real time parameter identification. The posterior pdfs are estimated using weighted Kernel density estimation with the normalised importance weights.

5.4.1.1 Parameters identified between algal blooms

Out of algal blooms, the identified parameters are $\mu_{max,hb}$ and K_{navig} , which is coherent with the previous sensitivity analysis (Wang et al., 2018b). At low temperatures ($T < 6$ °C), bacterial and phytoplanktonic activities are limited, the reaeration (K_{navig}) controls DO concentration within water column. The river system is governed by physical processes. The daily posterior pdfs of K_{navig} at low temperature period (day 21-38) are shown (Fig. 5.10A). Although the pdfs of K_{navig} from day 21 to day 28 have more dispersion than those for day 29-38, their modes are centred on the reference value (Fig. 5.10A). However, the water temperature is below 6 °C at the beginning of the simulation (days 0-6), K_{navig} is not very well estimated (see the first black polygon in Fig. 5.7B). This can be explained by the fact that the prior pdf of K_{navig} is uniform at the start of the simulation and this period is too short.

When the water temperature increases ($T > 6$ °C), the heterotrophic bacteria develop and degrade organic matter. $\mu_{max,hb}$ becomes the most important parameter governing the DO concentration in river system (Wang et al., 2018b). The identification of $\mu_{max,hb}$ in those periods can be expected (see the gray polygons in Fig. 5.8). The pdfs of $\mu_{max,hb}$ are displayed for day 45-62 when the water temperature increases (Fig. 5.10B). It can be clearly seen that the modes of the pdfs of $\mu_{max,hb}$ move towards the reference value (Fig. 5.10B). The particle filter captures the change from physical control (K_{navig}) to bacterial control ($\mu_{max,hb}$).

Next, the mortality of phytoplankton at the end of algal blooms supplies

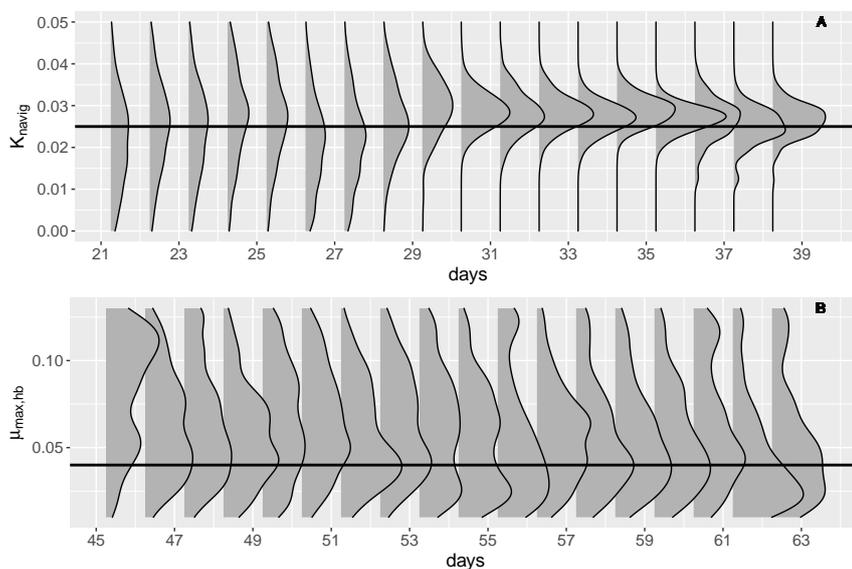


FIGURE 5.10 – The posterior pdf of K_{navig} during low temperature period (A) and the posterior pdf of $\mu_{max,hb}$ with a moderate temperature (B); Black lines represent the predefined values.

organic matter for bacterial growth. $\mu_{max,hb}$ is thus well estimated after algal blooms (see the gray polygons day 91-140 and day 280-300 in Fig. 5.8). This indicates that the algorithm adapts the evolution of the regime of river system from autotroph to heterotroph. However, $\mu_{max,hb}$ is not well estimated for day 300-325, even if this period corresponds to inter algal blooms. The correlation between the water temperature and the DO concentration in this period (Fig. 5.3) signifies that the DO concentration is controlled by water temperature and then is near the saturation concentration. The reaeration and micro-organism activities are not important in this period. Therefore, no parameter can be identified in this period.

5.4.1.2 Parameters identified during algal blooms

The maximum photosynthesis rate ($P_{max,pp}$) is third ranked parameter during algal blooms in the sensitivity analysis (Wang et al., 2018b). However, $P_{max,pp}$ is overestimated with a stable posterior pdf during the first algal bloom (see line shaded polygon day 64-90 in Fig. 5.9). The particles with $P_{max,pp}$ close to 0.2 h^{-1} (predefined value) are under weighted. That means the compensation of $P_{max,pp}$ with the other parameters exist. The parameter

interaction of $P_{max,pp}$ has been shown by the difference between the total sensitivity index and the first order sensitivity index (Wang et al., 2018b). Nonetheless, the DO concentrations have been well estimated.

The former sensitivity analysis which used almost constant water temperatures didn't highlight the influence of $T_{opt,pp}$. The supplementary result shows that the influence of $T_{opt,pp}$ depends on the water temperature (Fig. 4.12). Therefore, the water temperature is a crucial factor for algal blooms when $T_{opt,pp}$ is defined. This is the reason why $T_{opt,pp}$ can be identified before algal blooms, but not after algal blooms (Fig. 5.9).

Contrarily to the first algal bloom, the identification of $R_{m,pp}$ fails for the second algal bloom (Fig. 5.9). The supplementary sensitivity analysis shows that when the water temperature exceeds 20 °C, the total sensitivity index of $R_{m,pp}$ reduces dramatically (Fig. 4.12). The water temperatures are over 21 °C during the second bloom (Fig. 5.3). Therefore, $R_{m,pp}$ is not influential during the second algal bloom and remains unidentified. The other identified parameters match the sensitivity analysis ($\alpha_{pp}, \eta_{water}, \eta_{chla,pp}$).

In conclusion, the identification of the parameters depends on not only the hydrological and trophic contexts but also on the thermal state of the river system. The detailed identification of the parameters is resumed in figure 5.11.

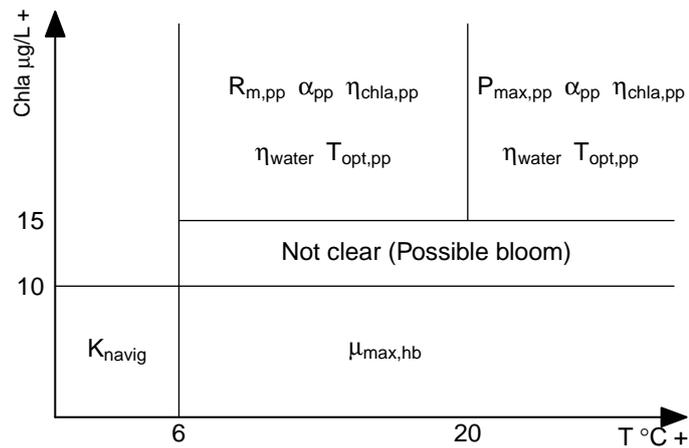


FIGURE 5.11 – The identification of parameters in different trophic (C_{chla}) and thermal (T °C) contexts of river system. See Tab. 5.1 for parameter definition.

5.4.2 Possibility to identify different phytoplanktonic communities

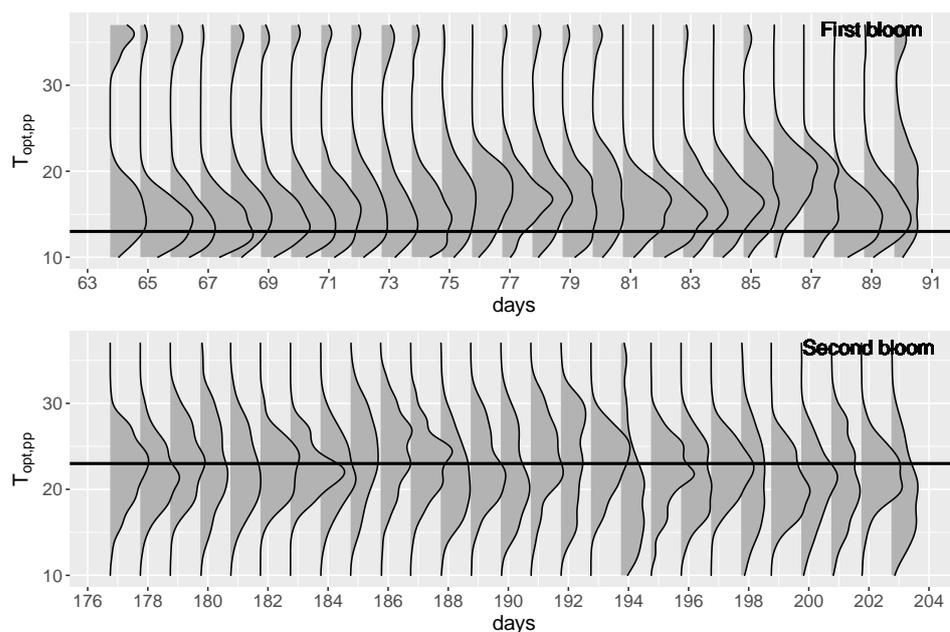


FIGURE 5.12 – The posterior distributions of $T_{opt,pp}$ during the first and the second algal bloom at 6 pm. Black line represents the predefined value.

In this paper, we define two different phytoplanktonic properties to distinguish blooms in March and in July ($R_{m,pp}$, $P_{max,pp}$ and $T_{opt,pp}$). The particle filter detects well the change of phytoplanktonic properties (Fig. 5.9), especially the optimal temperature for the growth of phytoplankton ($T_{opt,pp}$). The posterior pdfs of $T_{opt,pp}$ are stable during the two algal blooms (Fig. 5.12). It is therefore possible to differentiate phytoplanktonic communities in real time by comparing the posterior distributions of $T_{opt,pp}$. The different optimal temperatures that have been used for modeling the phytoplanktonic communities during spring and summer are the ones determined in the Loire and Seine rivers (Descy et al., 2012; Garnier et al., 1995; Vilmin, 2014).

5.4.3 Circadian rhythm

At night, the photosynthetic parameters are absolutely uninfluential and thus should remain unidentified. Therefore, a circadian rhythm (day and night) can be expected for the photosynthetic parameters. To study the circadian rhythm of photosynthetic parameters, the posterior pdfs of $P_{max,pp}$, α_{pp} and $T_{opt,pp}$ are displayed every 3 hours during days 184-186 (Fig. 5.13). It can be clearly noticed that the posterior pdfs of the three parameters have more dispersion at night (00 :00-6 :00) than those during the day (9 :00-18 :00). Their modes don't match the reference values at night, while the three parameters are well estimated during the day (Fig. 5.13). The PROSE-PA approach retrieves well the circadian rhythm of the photosynthetic parameters' influence.

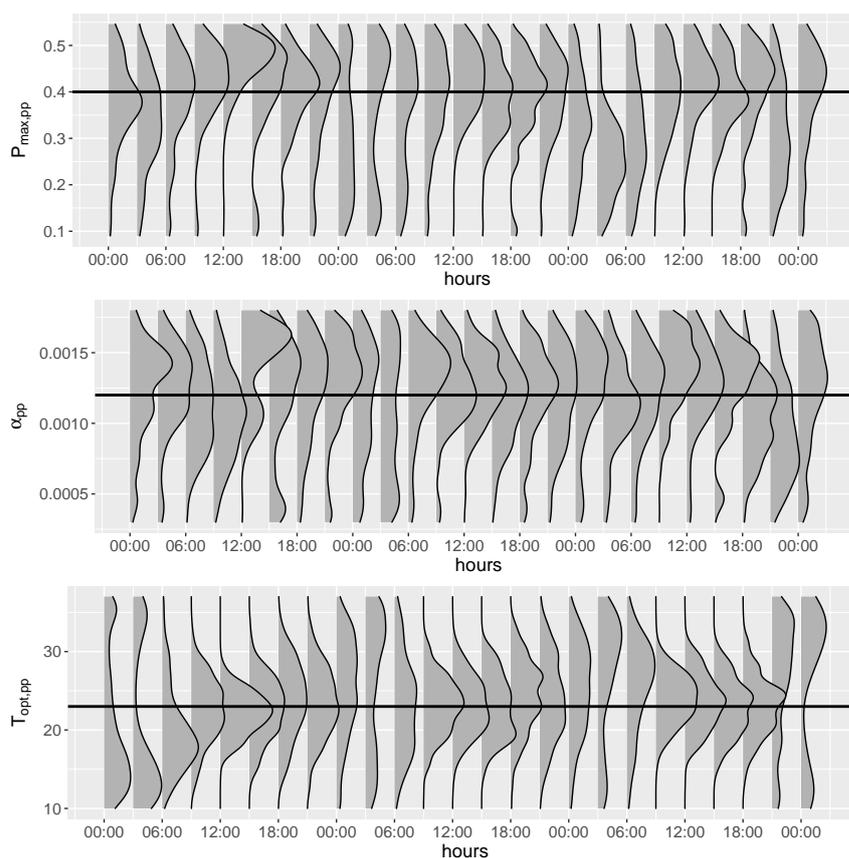


FIGURE 5.13 – The posterior distributions of $P_{max,pp}$, α_{pp} and $T_{opt,pp}$ during 184-186 days

5.4.4 Performances' sensitivity to the observation error

The proof of concept of the particle filter has been achieved assuming a relative error on observations of 1 %. In this section, we assess the impact of the relative error on both ProSe-PA ability to simulate oxygen concentrations and to identify parameter posterior pdfs. To this aim, various relative observation errors are tested from 1% to 10%. The Kling-Gupta Efficiencies (KGE) (Gupta et al., 2009; Kling et al., 2012) are calculated to evaluate the model performances. KGE is based on the decomposition of the mean squared error and NSE performance criteria (Gupta et al., 2009). KGEs range from -Inf to 1. Essentially, the closer to 1, the more accurate the model is.

The results show that the model performance decreases slightly with the increase of the observation relative error. Nevertheless, PROSE-PA retrieves in average the reference DO concentrations for all observation relative errors tested. KGEs are always larger than 0.96 (Tab. 5.3) whatever the relative error. The identifiability of the bacterial parameters ($\mu_{max,hb}$, Y_{hb} and $T_{opt,pp}$) and the reaeration coefficient related to navigation (K_{navig}) is not ensured when the observation relative error exceeds 5% while PROSE-PA is able to capture phytoplanktonic properties for all tested relative errors (1%-10%, Table 5.4), except for the ratio of chlorophyll *a* to carbon ($Chla/C_{pp}$) which has little influence on the variation of DO concentrations (Wang et al., 2018b).

5.5 Conclusions

In this work, we present a first implementation of particle filter into a hydro-biogeochemical model for metabolism's parameter estimation. The assimilation of a 15-min "observation" DO data is realized in the Seine River system on a synthetic case study.

- It can be concluded that the particle filter is an efficient method for the biogeochemical data assimilation and for the metabolism's parameter estimation in urban river systems.
- The PROSE-PA approach is capable to retrieve accurately the true DO concentrations at all stations.
- The identification of the metabolism parameters depends on the hydrological and trophic contexts and more surprisingly on the thermal state of the river system.

TABLE 5.3 – Kling-Gupta Efficiency (KGE) between the weighted mean and the true state at five monitoring sites for different observation relative errors

Stations	Observation relative errors [-]				
	0.010	0.025	0.050	0.075	0.100
Suresnes	0.9939	0.9917	0.9867	0.9776	0.9753
Chatou	0.9961	0.9964	0.9948	0.9764	0.9806
Bougival	0.9996	0.9980	0.9932	0.9625	0.9771
Sartrouville	0.9982	0.9975	0.9926	0.9824	0.9772
Andresy	0.9989	0.9987	0.9956	0.9915	0.9819

KGE (Kling-Gupta Efficiency) (Kling et al., 2012) : $KGE = 1 - \frac{\sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}}{\sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}}$

r : Correlation coefficient

β : Model bias. $\beta = \frac{\mu_{sim}}{\mu_{obs}}$, with μ the mean of the DO concentrations

γ : Variability. $\gamma = \frac{\sigma_{sim}/\mu_{sim}}{\sigma_{obs}/\mu_{obs}}$, with σ the standard deviation of the DO concentrations

TABLE 5.4 – Thresholds of the relative observation error for the parameter identifiability (1%-10%)

Parameters	Thresholds	Parameters	Thresholds	Parameters	Thresholds
$R_{m,pp}$	10% (T < 20 °C)	η_{water}	10%	Y_{hb}	< 7.5%
$P_{max,pp}$	10%	Chla/ C_{pp}	Not stable	$mort_{hb}$	Not stable
α_{pp}	10%	$T_{opt,pp}$	10%	$T_{opt,hb}$	< 5%
$\eta_{chla,pp}$	10%	$\mu_{max,hb}$	< 5%	K_{navig}	< 5%

10% : The parameters are always identified for the relative observation errors tested.

See the table 5.1 for the parameters' definition and units.

- The data assimilation method adapts to the trophic state's change of the Seine River system and the circadian cycle of photosynthetic parameters is well captured by the PROSE-PA approach.
- It is possible to distinguish phytoplanktonic species by identifying the different physiological properties via data assimilation. However, further research is needed to prove it.
- The PROSE-PA model can be transferable to waste water manager for assessing the impact of their practices now. However, it is necessary to quantify input uncertainties (forcing data, reject water of WWTPs and CSOs etc.), which are not considered in this paper.

5.6 Supplementary information

5.6.1 Importance sampling and definition of the importance weights in particle filtering

We suppose here that we want to simulate a random variable \mathbf{U} which may have a complex distribution. The importance sampling method uses an instrumental distribution called the importance distribution $\pi(\cdot)$ from which we know how to draw samples (Särkkä, 2013). The importance sampling relies on the following formula :

$$\int f_{\mathbf{U}}(\mathbf{u})d\mathbf{u} = \int \left[\frac{f_{\mathbf{U}}(\mathbf{u})}{\pi_{\mathbf{U}}(\mathbf{u})} \right] \pi_{\mathbf{U}}(\mathbf{u})d\mathbf{u} \quad (5.14)$$

where $\pi_{\mathbf{U}}(\mathbf{u})$ is a proposal pdf for the random variable \mathbf{U} from which we can draw samples and $f_{\mathbf{U}}(\mathbf{u})$ is the posterior pdf of given state \mathbf{U} from which we cannot directly or efficiently draw samples. The equation (5.14) transforms the expectation of \mathbf{U} into the expectation of the term $\left[\mathbf{u} \frac{f_{\mathbf{U}}(\mathbf{u})}{\pi_{\mathbf{U}}(\mathbf{u})} \right]$. We can perform Monte Carlo approximation to estimate the expectation of $g(\mathbf{U})$, denoted $\mathbf{E}(g(\mathbf{U}))$, for any function g :

$$\begin{aligned}
\mathbf{E}(g(\mathbf{U})) &= \int \left[g(\mathbf{u}) \frac{f_{\mathbf{U}}(\mathbf{u})}{\pi_{\mathbf{U}}(\mathbf{u})} \right] \pi_{\mathbf{U}}(\mathbf{u}) d\mathbf{u} & (5.15) \\
&\approx \frac{1}{N} \sum_{i=1}^N g(\mathbf{u}^i) \frac{f_{\mathbf{U}}(\mathbf{u}^i)}{\pi_{\mathbf{U}}(\mathbf{u}^i)} \\
\omega^i &= \frac{1}{N} \frac{f_{\mathbf{U}}(\mathbf{u}^i)}{\pi_{\mathbf{U}}(\mathbf{u}^i)}
\end{aligned}$$

where \mathbf{u}^i ($i = 1, \dots, N$) is i th sample (particle) drawn from the importance distribution $\pi_{\mathbf{U}}(\mathbf{u})$ and ω^i denotes the weight associated with i th sample (particle). With particular choices of g , we can hence retrieve the expectation of \mathbf{U} , its variance, quantiles, etc.

We can use importance sampling to address the problem of sampling directly from the posterior pdf $f(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*)$ of the trajectory $\mathbf{Z}_{1:t}$. Let the importance distribution, $\pi(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*)$, from equation (5.7) we can compute importance weights,

$$\omega_t = \frac{f(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*)}{\pi(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*)} \propto \frac{f(\mathbf{y}_t^*|\mathbf{z}_t)f(\mathbf{z}_t|\mathbf{z}_{t-1})f(\mathbf{z}_{1:t-1}|y_{1:t-1}^*)}{\pi(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*)} \quad (5.16)$$

Then, assuming the importance distribution $\pi(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*)$ to be Markovian, that is

$$\begin{aligned}
\pi(\mathbf{z}_{1:t}|\mathbf{y}_{1:t}^*) &= \pi(\mathbf{z}_t|\mathbf{z}_{1:t-1}, \mathbf{y}_{1:t}^*)\pi(\mathbf{z}_{1:t-1}|\mathbf{y}_{1:t-1}^*) & (5.17) \\
&= \pi(\mathbf{z}_t|\mathbf{z}_{t-1}, \mathbf{y}_t^*)\pi(\mathbf{z}_{1:t-1}|\mathbf{y}_{1:t-1}^*),
\end{aligned}$$

then equation (5.16) becomes,

$$\begin{aligned}
\omega_t &\propto \frac{f(\mathbf{y}_t^*|\mathbf{z}_t)f(\mathbf{z}_t|\mathbf{z}_{t-1})f(\mathbf{z}_{1:t-1}|y_{1:t-1}^*)}{\pi(\mathbf{z}_t|\mathbf{z}_{t-1}, \mathbf{y}_t^*)\pi(\mathbf{z}_{1:t-1}|\mathbf{y}_{1:t-1}^*)} & (5.18) \\
&\propto \frac{f(\mathbf{y}_t^*|\mathbf{z}_t)f(\mathbf{z}_t|\mathbf{z}_{t-1})}{\pi(\mathbf{z}_t|\mathbf{z}_{t-1}, \mathbf{y}_t^*)} \omega_{t-1}
\end{aligned}$$

In other words, we derive an update formula for the importance weights in the same form as the update formula for the posterior pdf ((5.7)). If we can draw N samples (particles) from the importance distribution $\pi(\mathbf{z}_t|\mathbf{y}_{1:t}^*)$,

$$\mathbf{z}_t^i \sim \pi(\mathbf{z}_t | \mathbf{y}_{1:t}^*) \quad i = 1, \dots, N \quad (5.19)$$

a weight update formula for each particle can be written as follows,

$$\begin{aligned} \omega_t^i &\propto \frac{f(\mathbf{y}_t^* | \mathbf{z}_t^i) f(\mathbf{z}_t^i | \mathbf{z}_{t-1}^i)}{\pi(\mathbf{z}_t^i | \mathbf{z}_{t-1}^i, \mathbf{y}_t^*)} \omega_{t-1}^i \\ \hat{\omega}_t^i &= \frac{\omega_t^i}{\sum \omega_t^i} \end{aligned} \quad (5.20)$$

where we define the normalized importance weight ($\hat{\omega}_t^i$) associated with the particle i . Normalizing the weights allows to evacuate the problem of the normalizing constant in the Bayesian representation of the posterior ((5.5)). Equation (5.20) shows that we only need to store the state \mathbf{z}_t at time t . In addition, it is usually not necessary to estimate the full pdf $f(\mathbf{z}_{1:t} | \mathbf{y}_{1:t}^*)$ in practice. In this work, we are interested in the marginal distribution $f(\mathbf{z}_t | \mathbf{y}_{1:t}^*)$ at time t , called filtering distribution. The filtering posterior pdf $f(\mathbf{z}_t | \mathbf{y}_{1:t}^*)$ at time t can be approximated as,

$$f(\mathbf{z}_t | \mathbf{y}_{1:t}^*) \approx \sum_{i=1}^N \hat{\omega}_t^i \delta(\mathbf{z}_t - \mathbf{z}_t^i) \quad (5.21)$$

where $\delta(\cdot)$ is a Dirac delta function.

The performance of the above algorithm depends on the quality of the importance distribution $\pi(\cdot)$. Typically, the optimal importance distribution is (Doucet et al., 2001; Särkkä, 2013),

$$\pi(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{y}_t^*) = f(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{y}_t^*) \quad (5.22)$$

In practice, however we generally do not know the distribution $f(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{y}_t^*)$ and we rather propagate \mathbf{z}_t using the evolution equations (5.1) and (5.2), which terms to draw samples from $f(\mathbf{z}_t | \mathbf{z}_{t-1})$. In that case, the equations (5.18) and (5.20) simplifies into

$$\omega_t \propto f(\mathbf{y}_t^* | \mathbf{z}_t) \omega_{t-1} \quad (5.23)$$

5.6.2 ProSe-PA vs Filtre particulaire standard

Le filtre particulaire proposé dans cette étude n'est pas complètement standard. En effet, les paramètres seraient perturbés à chaque pas de temps dans un filtre particulaire usuel. Le filtre standard a été comparé avec le filtre qu'on propose par essai erreur. Les résultats du filtre qu'on propose identifient mieux les paramètres bactériens (Pre-bloom et Post-bloom) et le coefficient de réaération (K_{navig}). Ces résultats sont en synergie avec les résultats de l'analyse de sensibilité. On justifie également ce choix du fait que cela a plus de sens d'un point de vu phénoménologique et biogéochimique en raison de la lenteur de l'évolution des paramètres par rapport à la fréquence d'acquisition des données d'oxygène dissous (15 min).

5.6.3 Test du nombre de particules

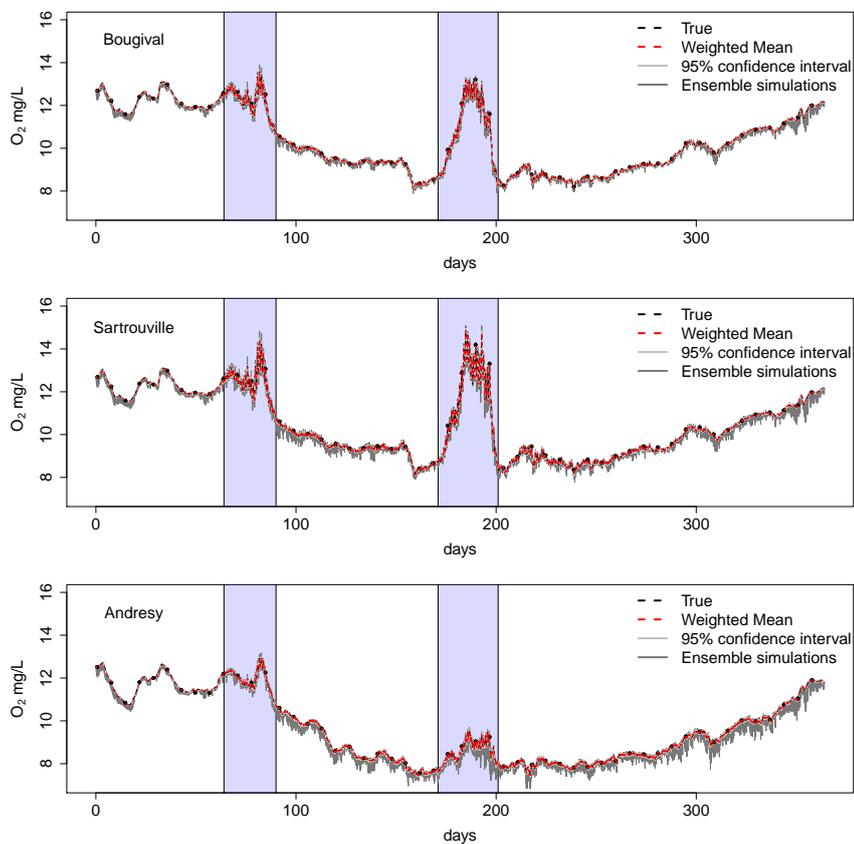


FIGURE 5.14 – Simulated DO concentrations at Bougival, Sartrouville and Andresy stations with 100 particles.

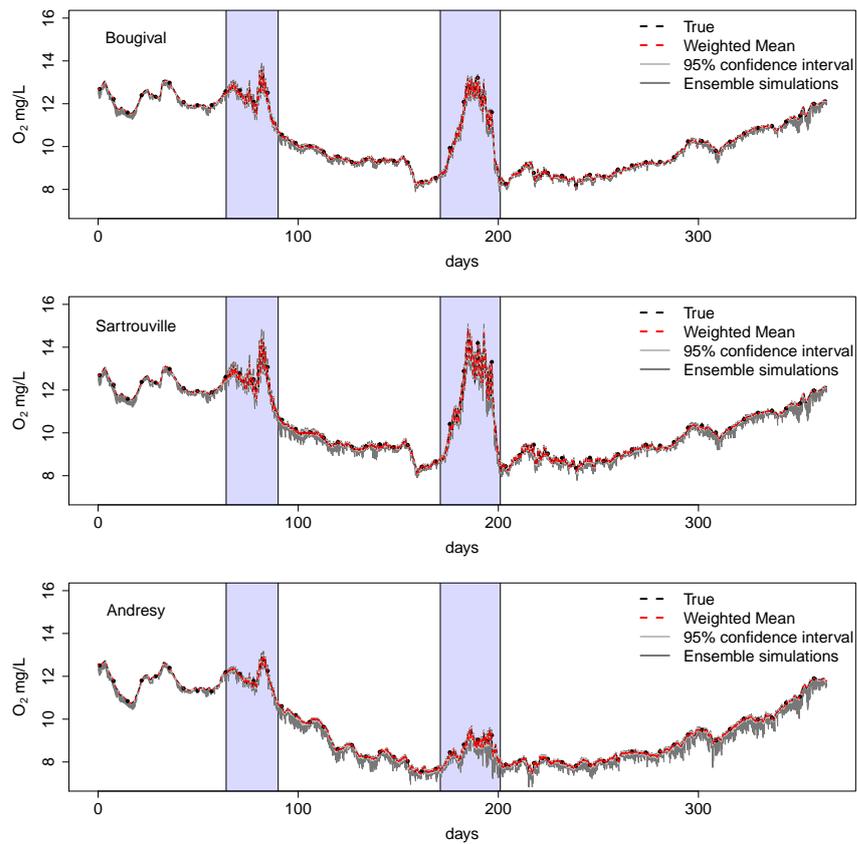


FIGURE 5.15 – Simulated DO concentrations at Bougival, Sartrouville and Andresy stations with 300 particles.

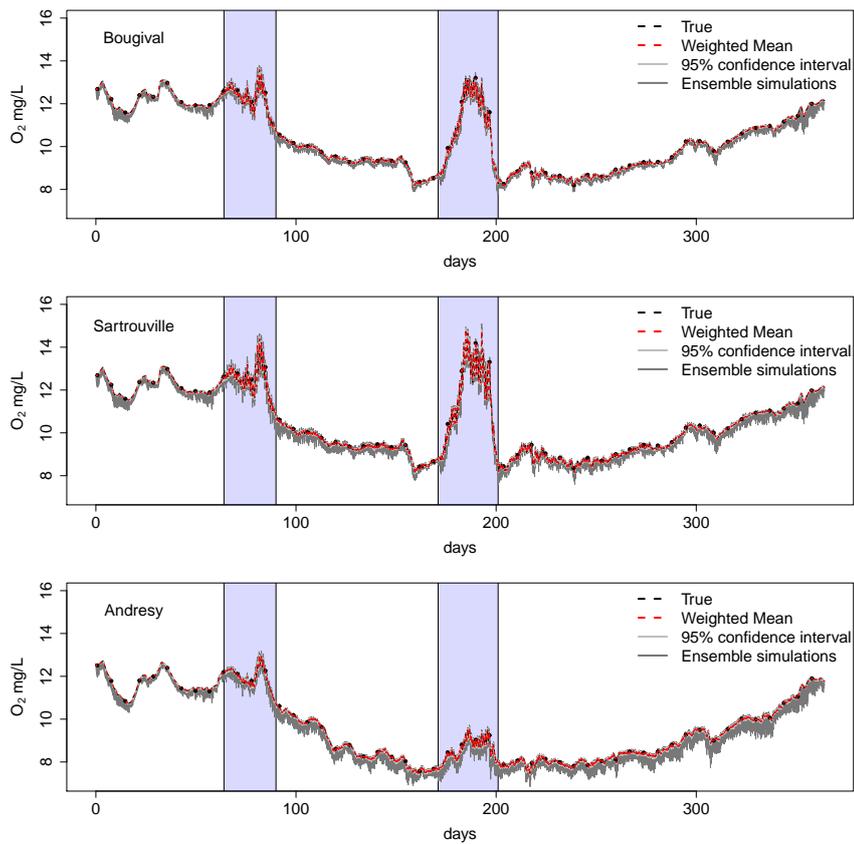


FIGURE 5.16 – Simulated DO concentrations at Bougival, Sartrouville and Andresy stations with 400 particles.

5.6.4 Concentrations en oxygène dissous simulées aux stations de validations

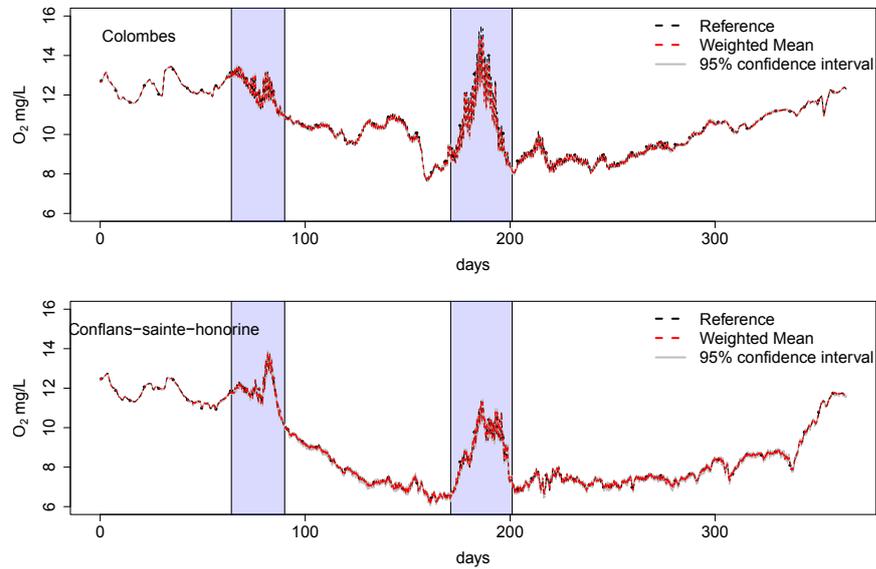


FIGURE 5.17 – Simulated DO concentrations at Colombes and Conflans-Sainte-Honorine stations (validation) with 100 particles.

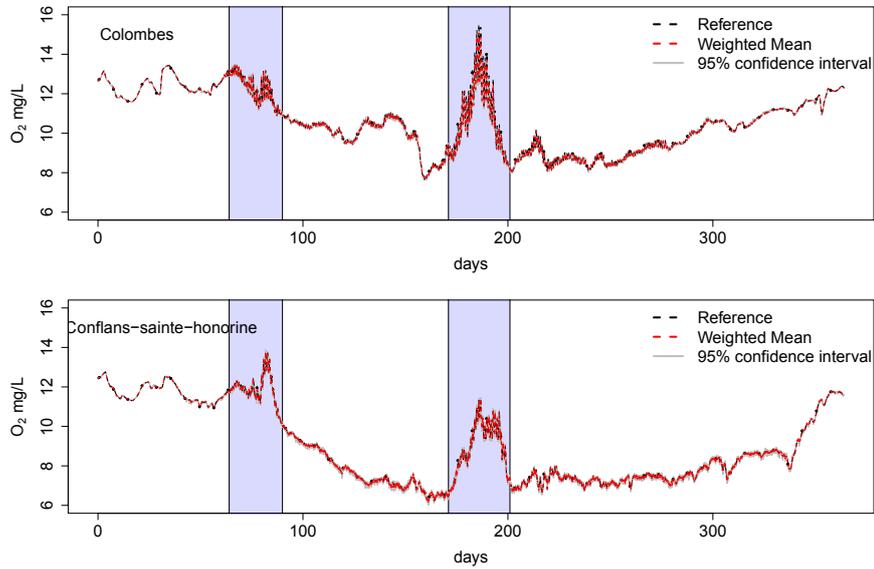


FIGURE 5.18 – Simulated DO concentrations at Colombes and Conflans-Sainte-Honorine stations (validation) with 300 particles.

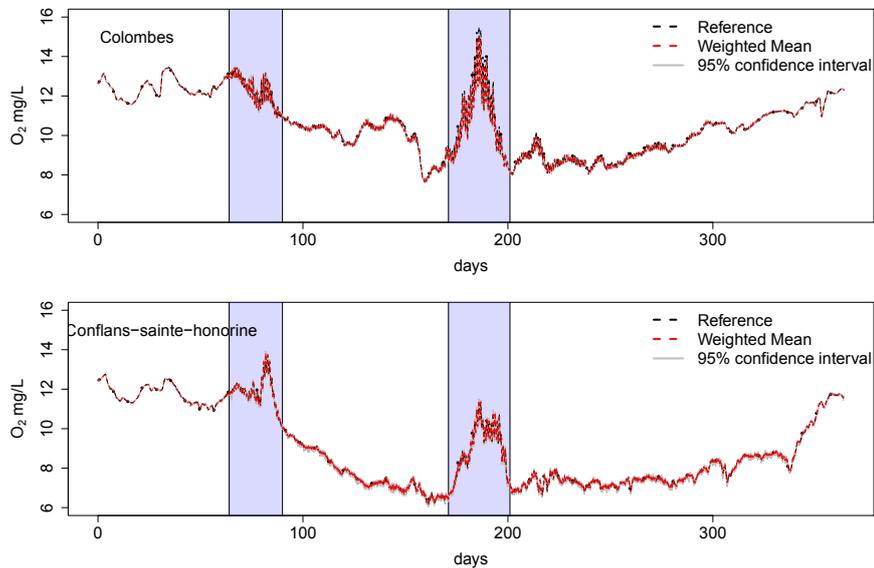


FIGURE 5.19 – Simulated DO concentrations at Colombes and Conflans-Sainte-Honorine stations (validation) with 500 particles.

Chapitre 6

La Seine depuis l'amont de Paris jusqu'à l'estuaire

Résumé

L'efficacité du filtre particulaire est démontrée dans le chapitre précédent sur un cas synthétique. Dans ce chapitre, PROSE-PA est appliqué sur un cas réel qui décrit la Seine depuis l'amont de Paris jusqu'à l'estuaire. Avant d'assimiler les mesures réelles en O_2 , un diagnostic de la simulation des concentrations en O_2 par le modèle direct a été réalisé. Ce diagnostic permet d'atténuer les surestimations de l'oxygène dissous soulignées par [Vilmin \(2014\)](#) en travaillant les clés de répartition de la matière organique dissoute (MOD) dans la rivière ainsi qu'une calibration des activités de la MOD biodégradable. Une fois ce réajustement réalisé, PROSE-PA assimile les mesures en O_2 de l'année 2011 à 8 stations gérées par le SIAAP. Les résultats montrent une amélioration significative des concentrations en O_2 simulées et une bonne identification des propriétés physiologiques des phytoplanctons pendant les blooms algaux. Pour les paramètres bactériens, une bonne identifiabilité est observée pour des périodes de fortes désoxygénations. En revanche, des oscillations de paramètres apparaissent, certainement dues à des effets de compensation entre paramètres et/ou de variations des conditions aux limites.

6.1 Diagnostic de la simulation des concentrations en O_2 pour l'année 2011

En 2011, trois blooms algaux ont été observés dans la Seine à la station Bougival (Groleau et al., 2014). Le bloom de mars est principalement constitué de diatomées de type *Synedra ulna*, *Cyclotella spp.* et *Nitzschia spp.*, tandis que le bloom de mai est dominé par les diatomées de type diatomeflagellées enkystées (Groleau et al., 2014; Escoffier et al., 2018). Enfin, le bloom de juillet est constitué notamment d'algues de l'espèce *Cyclotella spp.* (Escoffier et al., 2018).

Vilmin (2014) a calibré les températures optimales et les taux de croissance maximaux des cortèges d'espèces de mars et de juillet. Le bloom de mars est caractérisé par une température de 10 °C et un écart-type de 8 °C, tandis que le bloom de juillet a une température optimale plus élevée (23 °C) et un écart-type de 8 °C. Cette nouvelle paramétrisation a permis d'améliorer la reproduction des concentrations en O_2 .

Cependant, Vilmin (2014) a souligné que le modèle PROSE a tendance à surestimer les concentrations en O_2 en certaines stations. Dans la suite, la relation entre cette surestimation et la disponibilité de la MOD est mise en évidence.

6.1.1 Reproduction des concentrations en O_2 par une simulation directe

Une simulation pour l'année 2011 par PROSE-P est réalisée. Une seule espèce de phytoplancton et de bactérie hétérotrophe est considérée pour cette simulation. On peut observer des surestimations d' O_2 en temps sec soulignées par Vilmin (2014). C'est notamment le cas aux stations de Chatou, Sartrouville et Méricourt entre les jours 250 et 300 pour l'année 2011 (Fig. 6.1 et 6.2). Les écarts entre observations et simulations peuvent dépasser plus de 2 mg O_2 /L, notamment autour du jour 160. L'hypothèse est posée d'un déficit en MOD biodégradable dans les conditions aux limites ainsi qu'une mauvaise paramétrisation des cinétiques de la MOD.

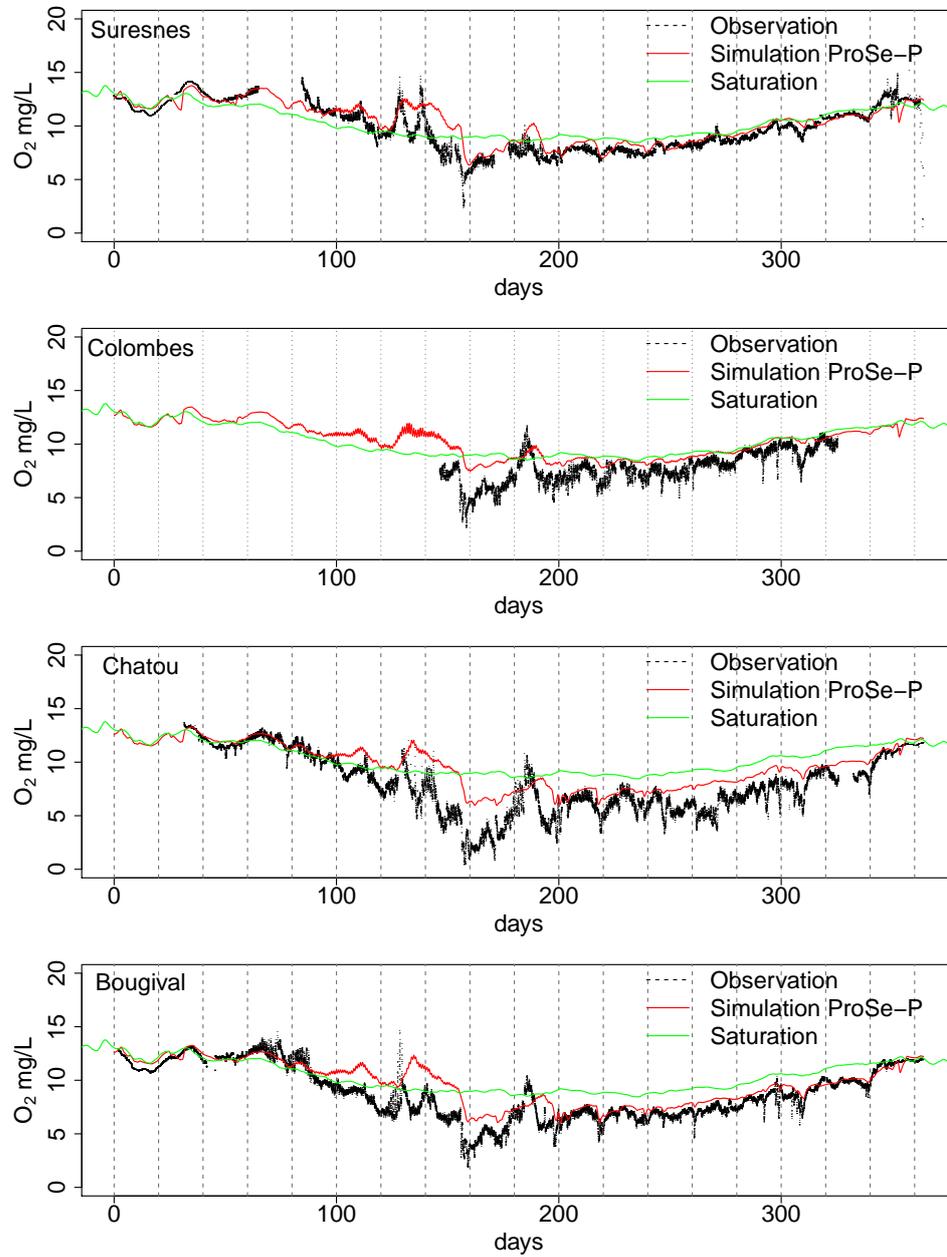


FIGURE 6.1 – Chroniques de concentrations en O₂ observées (ligne noire) et simulées (ligne rouge) avec le modèle direct PROSE-P aux stations SIAAP

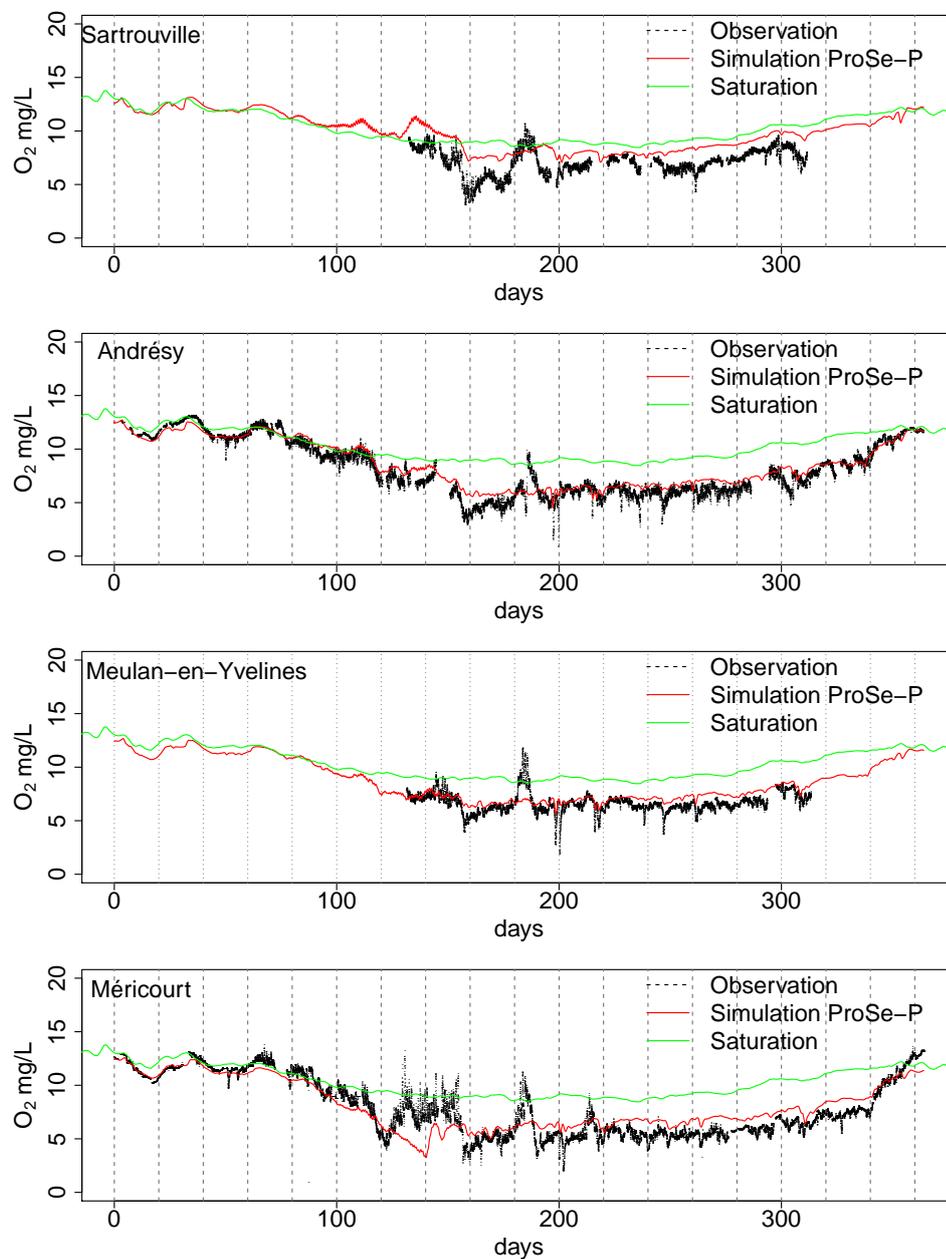


FIGURE 6.2 – Chroniques de concentrations en O₂ observées (ligne noire) et simulées (ligne rouge) avec le modèle direct PROSE-P aux stations SIAAP

6.1.2 Profil en long des matières organiques biodégradables (MOD1 et MOD2)

Afin de vérifier la disponibilité des matières organiques dissoutes biodégradables (MOD1 et MOD2), les profils en long des MOD1 et MOD2 simulés par PROSE-P sont tracés entre Suresnes (pk 674 km) et Méricourt (pk 779 km) (Fig. 6.3). Le PK signifie Point Kilométrique. Dans le modèle, le point aval du modèle (Poses) est fixé à 862 km. Le pk à un point donné se fait par un calcul algébrique.

Le profil en long indique que le prélèvement de MOD1 est limité avant Seine Aval (pk 720 km). Le niveau de MOD1 reste quasiment constant (0.25 mgC/L). Les concentrations en O₂ sont donc surestimées aux stations Châtou et Sartrouville. Ceci est lié à la constante de demi-saturation K_s fixée à 0.25 mgC/L. Cette valeur est deux fois plus grande que la valeur déterminée (0.1 mg/L) par Garnier et al. (1992). En revanche, les rejets de la station d'épuration Seine Aval permettent d'enrichir la concentration en MOD1 (0.53 mgC/L). Cet enrichissement en MOD1 assure une bonne reproduction des concentrations en O₂ à la station Andrésy (pk 729 km) (Fig. 6.2). Cependant la concentration en MOD1 est proche de 0.1 mgC/L à la station Méricourt (pk 779km), la croissance des bactéries est donc limitée. C'est pourquoi on observe une surestimation de l'oxygène dissous à la station Méricourt (Fig. 6.2).

De même, la transformation de MOD2 en MOD1 est aussi limitée (Fig. 6.3). La constante de demi-saturation K_{MOD2} est fixée à 2.5 mgC/L, ce qui est très élevé en comparaison des concentrations en MOD2 (< 0.7 mgC/L).

L'analyse des chroniques temporelles à plusieurs stations, couplée à celle des profils en long permet de relier la surestimation de l'oxygène à un déficit en MOD biodégradable et aux cinétiques que les bactéries contrôlent. La spéciation de la MOD (biodégradable et réfractaire) et la calibration des paramètres liés aux réactivités de la MOD biodégradable (K_s , K_{mod2}) sont donc nécessaires.

6.2 Spéciation de la matière organique dissoute à l'amont du domaine

Dans C-RIVE, la matière organique dissoute est décomposée en trois fractions : MOD1 (Matière Organique Dissoute rapidement biodégradable,

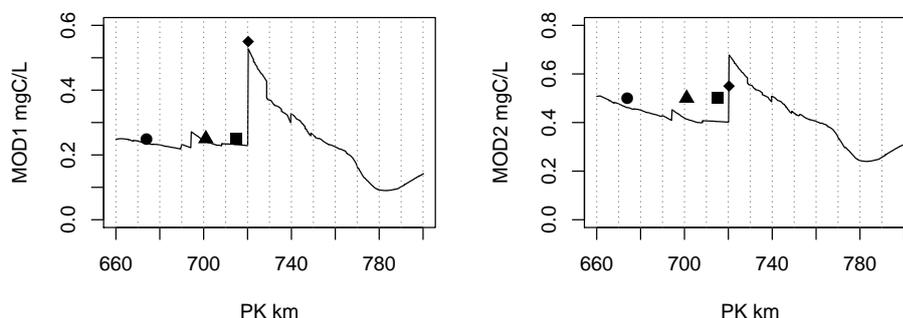


FIGURE 6.3 – Profil longitudinal de MOD1 et MOD2 entre Suresnes et Méricourt pour le jour 264. Les symboles de gauche à droite représentent les positions de Suresnes, Chatou, Sartrouville, Seine Aval. PK : point kilométrique, 862 km à Poses

en moins de 5 jours), MOD2 (Matière Organique Dissoute lentement biodégradable en moins de 45 jours) et MOD3 (Matière Organique Dissoute réfractaire). La répartition des trois types de MOD dans les rivières est estimée à l'aide de relations empiriques basées sur la concentration en carbone organique total (COT) (Vilmin, 2014, p.235). Ces relations sont présentées dans la table 6.1. Au total, 30% des MOD sont considérées comme biodégradables (MOD1 et MOD2).

D'après la simulation par le modèle déterministe (Fig. 6.1 et 6.2), il semble que 30% de la MOD ne soient pas suffisants pour assurer une bonne simulation de l'oxygène dissous en temps sec (250-300 jours). Cette relation (MOD biodégradable et réfractaire) devrait être corrigée. Pour corriger cette relation, les déficits en MOD biodégradable sont estimés à chaque station de mesure dans la suite.

TABLE 6.1 – Répartition des différents types de matière organique dissoute (Vilmin, 2014, p.235)

Variable PROSE	[MOD1]	[MOD2]	[MOD3]
Seine, Marne, Oise	0.072[COT] =0.09[MOD]	0.168[COT] =0.21[MOD]	0.56[COT] =0.7[MOD]

6.2.1 Déficit en MOD biodégradable

Les surestimations en O_2 traduisent un déficit de MOD biodégradable. On considère que ce déficit est lié notamment aux apports des grands fleuves (Seine, Marne et Oise). Les déficits en MOD labile peuvent être calculés entre 250 et 300 jours à chaque station de mesure (Eq. (6.1)). Puisque le débit est relativement stable pour cette période, l'influence des rejets des DOs est peu importante. De plus, les blooms des algues n'ont pas été observés dans la Seine. Le système de la rivière est ainsi moins compliqué que celui en bloom. La surestimation en O_2 est principalement liée aux réactivités de la MOD. Dans la suite ces déficits en matière biodégradable sont compensés en diminuant la part de MOD réfractaire dans la répartition de la MOD.

$$prel_{mod} = \frac{([O_2]_{simul} - [O_2]_{obs}) \cdot Q}{\tau_{hb} \cdot (1 - Y_{hb})} \quad (6.1)$$

$$ratio_{station} = \frac{prel_{mod}}{MOD3_{apport}} \quad (6.2)$$

Avec,

$[O_2]_{simul}$ [g/m³] : Concentrations en oxygène dissous simulées

$[O_2]_{obs}$ [g/m³] : Concentrations en oxygène dissous observées

Q : le débit [m³/s]

τ_{hb} [gO₂/gC] : Oxydation des matières organiques dissoutes par les bactéries hétérotrophes 2.67 [gO₂/gC]

$prel_{mod}$: Prélèvement des MOD1

MOD3_{apport} : MOD3 apportée par Seine, Marne et Oise

ratio_{station} : Déficit en ratio par rapport à la MOD3 apportée par Seine, Marne et Oise

TABLE 6.2 – Déficit en MOD biodégradable pour la période 250-300 jours par rapport à MOD3

Stations	Suresnes	Colombes	Chatou	Bougival
Déficit (%)	2.1	8.1	6.5	2.2
Stations	Sartrouville	Andrésy	Meulan-en-Yvelines	Méricourt
Déficit (%)	14.4	3.1	5.4	11.6

Les deux stations (Sartrouville et Méricourt) représentent les déficits les

plus grands (Tab. 6.2) avec des pourcentages de 14.4% et 11.6%. L'éloignement de ces stations des amonts et points de rejets expliquent ces forts déficits en O₂ du fait d'un déficit de MOD. A l'opposé, les stations proches des discontinuités sont bien simulées comme Suresnes et Andrésy (déficit de MOD de 2.1% et 3.1%).

Afin de corriger ces biais, la fraction de MOD biodégradable apportée par les grandes rivières (Seine, Marne et Oise) est augmentée en diminuant la fraction de MOD3. Selon le calcul (Tab. 6.2), environ 25% de la MOD3 apportée (14.4 + 11.6) par les rivières devraient être répartis en MOD1 et MOD2. Après la correction, 47.5% des MOD sont considérés comme biodégradables. Dans le système aquatique, la fraction de MOD biodégradable par rapport à MOD est entre 11% et 59% (Servais et al., 1987). Cette fraction corrigée (47.5 %) reste ainsi dans une gamme acceptable.

6.2.2 Réactivités des MODs biodégradables

Dans C-RIVE, la MOD1 est considérée comme substrat et est assimilée directement pour la croissance des bactéries hétérotrophes. L'hydrolyse exoenzymatique de MOD2 vers MOD1 est liée à la biomasse bactérienne. Ces deux processus sont exprimés à l'aide des équations de Monod. La répartition de MOD3 en MOD1 et MOD2 est modélisée suivant les équations (6.3) et (6.4) paramétrées à l'aide des valeurs de la table 6.3.

$$hyd_{mod2} = k_{hyd,max} \frac{MOD2}{MOD2 + K_{mod2}} BH_{tot} \quad (6.3)$$

$$\mu_{BH} = \mu_{BH,max} \frac{MOD1}{MOD1 + K_s} \quad (6.4)$$

Avec,

$k_{hyd,max}$ [/h] : Coefficient d'hydrolyse de MOD2 en MOD1

K_{mod2} [mgC/L] : Constante de demi-saturation pour l'hydrolyse de MOD2

BH_{tot} [mgC/L] : Biomasse totale des bactéries hétérotrophes

$\mu_{BH,max}$ [/h] : Taux de croissance maximal

K_s [mgC/l] : Constante de demi-saturation pour la croissance des bactéries hétérotrophes

Les trois paramètres (K_s , K_{mod2} et $K_{hyd,max}$) ne sont actuellement pas intégrés dans le schéma d'assimilation de données. Comme décrit dans la section précédente (Section 6.1.2, p.173), une calibration des paramètres est

TABLE 6.3 – Répartition (ratio) de 25% de MOD3 en MOD1 et MOD2

MOD biodégradable	Scénario 1	Scénario 2	Scénario 3
MOD1	0.00	0.25	0.50
MOD2	1.00	0.75	0.50

TABLE 6.4 – Les valeurs des paramètres liés aux réactivités de la MOD biodégradable pour le balayage

Ks [mgC/L]	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.15	
K_{mod2} [mgC/L]	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.10	1.20
$k_{hyd,max}$ [/h]	0.25	0.35	0.45	0.55	0.65	0.75						

nécessaire. Afin de calibrer ces trois paramètres en fonction de la répartition de MOD3 (Tab. 6.3), un balayage des trois paramètres est réalisé (Tab. 6.4). Au total, 2970 simulations ont été lancées, les RMSEs (*Root Mean Square Error*) sont calculées entre les jours 250 et 300 à chaque station. On sélectionne les simulations correspondant à la RMSE minimale de chaque station (Tab. 6.5). Les stations sont numérotées depuis l'amont vers aval : 1. Suresnes, 2. Colombes, 3. Chatou, 4. Bougival, 5. Sartrouville, 6. Andrésy, 7. Meulan-en-Yvelines, 8. Méricourt (Tab. 6.5).

A partir des résultats obtenus, on sélectionne la simulation nommée par 0.00_1.00_0.05_0.20_0.55. Cette simulation permet en effet de diminuer la RMSE à la fois pour les stations Chatou (3), Sartrouville (5) et Méricourt (8) où le modèle direct surestime beaucoup les concentrations en O₂ (Fig. 6.1 et 6.2). 25% des MOD3 sont totalement réparties en MOD2 (100%). Les constantes de demi-saturation faibles ($K_s = 0.05$ mgC/L et $K_{mod2} = 0.20$ mg/L) favorisent également la réactivité des MODs biodégradables. Les nouvelles clés de répartition de la MOD sont indiquées Table 6.6.

6.3 Performances de ProSe-PA pour l'année 2011

Suite à la redéfinition des clés de répartition de la MOD, le schéma d'assimilation de données est maintenant évalué en Seine pour l'année 2011. L'objectif de cette partie est d'identifier les propriétés physiologiques des communautés vivantes et de bien reproduire les concentrations en O₂. Les mesures en O₂ à chaque quart d'heure à 8 stations du SIAAP (Suresnes, Co-

TABLE 6.5 – Simulations correspondant à la RMSE minimale de chaque station entre les jours 250 et 300 en fonction des différentes paramétrisations pour calibrer les réactivités de la MOD biodégradable. La colonne Simulations est nommée avec le règle : ratio-MOD1_ratio-MOD2_K_s_K_{mod2}_K_{hyd,max}.

Simulations	1	2	3	4	5	6	7	8
0.00_1.00_0.15_1.50_0.25	0.40	0.78	1.57	0.41	1.20	0.59	0.64	0.67
0.25_0.75_0.05_0.60_0.75	0.96	0.53	0.88	0.79	0.78	0.48	0.37	0.69
0.00_1.00_0.05_0.20_0.55	0.94	0.55	0.83	0.87	0.75	0.49	0.37	0.67
0.00_1.00_0.13_1.50_0.35	0.46	0.72	1.46	0.39	1.11	0.55	0.56	0.77
0.00_1.00_0.05_0.40_0.55	0.92	0.53	0.92	0.92	0.72	0.55	0.51	1.02
0.50_0.50_0.05_0.70_0.75	0.97	0.54	0.87	0.79	0.79	0.47	0.37	0.69
0.50_0.50_0.06_0.20_0.65	1.02	0.56	0.88	0.77	0.81	0.48	0.36	0.65
Avant la calibration (Fig. 6.1 et 6.2)	0.46	1.05	2.12	0.86	1.60	0.86	0.97	1.30

TABLE 6.6 – Répartition des différents types de matière organique dissoute après la calibration

Variable PROSE	[MOD1]	[MOD2]	[MOD3]
Seine, Marne, Oise	0.09[MOD]	0.385[MOD]	0.525[MOD]

lombes, Chatou, Bougival, Sartrouville, Andrésy, Meulan-en-Yvelines, Méricourt) sont assimilées par PROSE-PA suivant le schéma du chapitre 5.

6.3.1 Stratégie de modélisation

Dans ce cas réel, une espèce phytoplanctonique et une espèce de bactérie hétérotrophe sont introduites dans PROSE-PA. Les biomasses des deux espèces sont obtenues par la somme des espèces identifiées dans le milieu (Vilmin, 2014). On suppose que l'erreur de mesure est proportionnelle aux valeurs mesurées (10%). La perturbation des paramètres après le ré-échantillonnage est identique à celle utilisée dans le cas synthétique soit 10% de la gamme de chaque paramètre (cf. chapitre 5. p.129). La table 6.7 récapitule la configuration du filtre.

TABLE 6.7 – Paramètres numériques du filtre particulaire pour le cas réel

Nombre de particules	Seuil de ré-échantillonnage	Erreur _{obs}	Perturbation
500	150	10%	10%

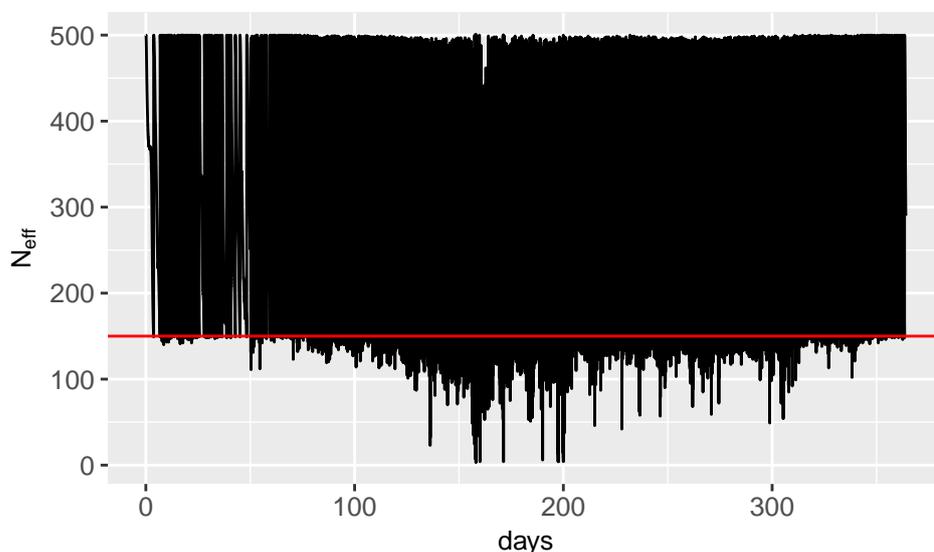


FIGURE 6.4 – L'évolution de la taille effective (N_{eff}) pour le ré-échantillonnage.

6.3.2 La taille effective pour le ré-échantillonnage lors de la simulation

La taille effective (N_{eff}) est un critère permettant d'évaluer la diversité de l'ensemble des particules simulées. Lorsque N_{eff} passe sous 150, un ré-échantillonnage qui consiste à parcourir l'espace des paramètres autour des valeurs les plus probables est nécessaire. La taille effective est globalement supérieure à 150 pour les 100 premiers jours de simulation (Fig. 6.4). Le filtre peut bien reproduire les concentrations en O_2 . Cependant, la taille effective a tendance à descendre plus fréquemment au dessous du seuil de 150 à partir du jour 100. Pour certains moments entre les jours 150 et 200, de fortes dégénérescences sont observées. Ce qui indique que quasiment toutes les particules ont un poids nul ou très faible. Le filtre a alors des difficultés pour corriger les concentrations en O_2 simulées. Des écarts importants entre observation et simulation peuvent être observés pour cette période (Fig. 6.6 et 6.7). Ces difficultés pourraient être liées aux incertitudes des conditions aux limites ou aux incertitudes de la transformation du phytoplancton détritique en MOD biodégradable pour cette période. Pour rappel, les conditions

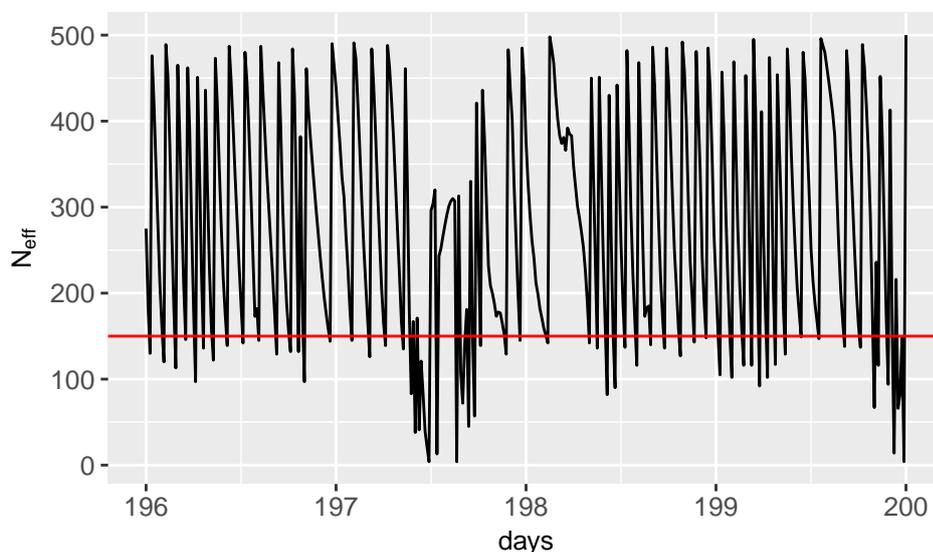


FIGURE 6.5 – L'évolution de la taille effective (N_{eff}) pour le ré-échantillonnage pour les jours 196-200.

aux limites sont considérées parfaites. Néanmoins, la diversité des particules est rétablie par la perturbation des paramètres après le ré-échantillonnage. Ce maintien de la diversité des particules par la perturbation assure ensuite la simulation de l'oxygène dissous après le jour 200 où la taille effective est globalement supérieure à 50 (Fig. 6.4).

6.3.3 Concentrations en O_2 simulées par ProSe-PA

Afin d'évaluer les performances de PROSE-PA sur le cas réel, les critères statistiques (RMSE et $KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}$, (Kling et al., 2012)) détaillés dans la section 3.6.2.2 sont utilisées (Eq. 3.39). Le filtrage particulaire (PROSE-PA) permet de bien reproduire les dynamiques de l' O_2 (Fig. 6.6 et 6.7). Les concentrations en O_2 simulées sont significativement améliorées par PROSE-PA (Tab. 6.8). En effet, la RMSE diminue de moitié par rapport aux performances de PROSE3.6.9 (Vilmin, 2014). Ces améliorations sont particulièrement impressionnantes aux stations Chatou (0.98 mg O_2 contre 2.42 mg O_2), Sartrouville (1.00 mg O_2 contre 2.50 mg O_2), Meulan-en-Yvelines (0.64 mg O_2 contre 1.58 mg O_2) et Méricourt (1.00 mg O_2 contre 1.91 mg O_2). De même, les coefficients KGEs pour les simulations de

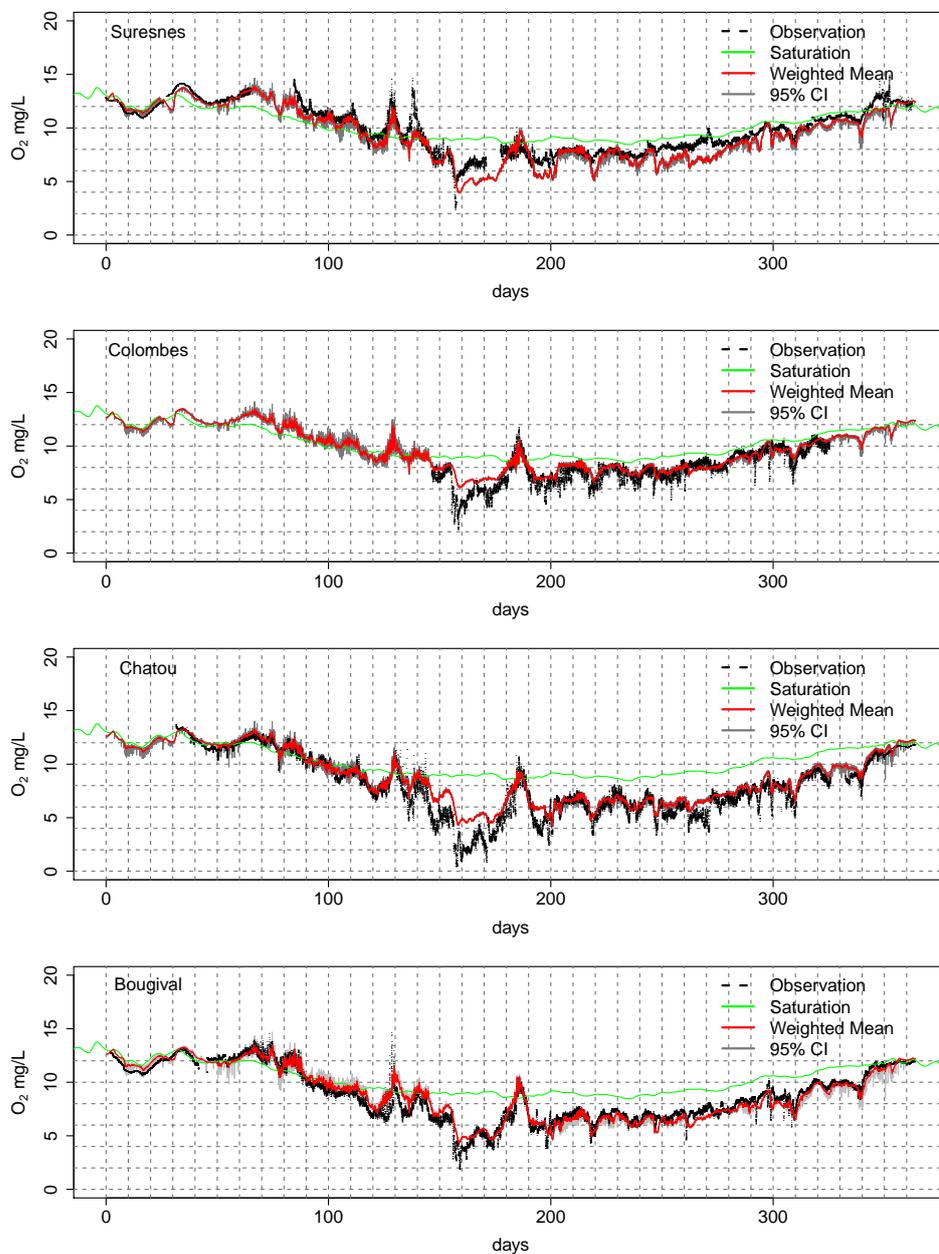


FIGURE 6.6 – Concentrations en O_2 observées (ligne noire) et simulées (ligne rouge) par PROSE-PA aux stations SIAAP en 2011 (le 1 janvier 2011 = jour 0). Les zones grises présentent les intervalles de confiance à 95%.

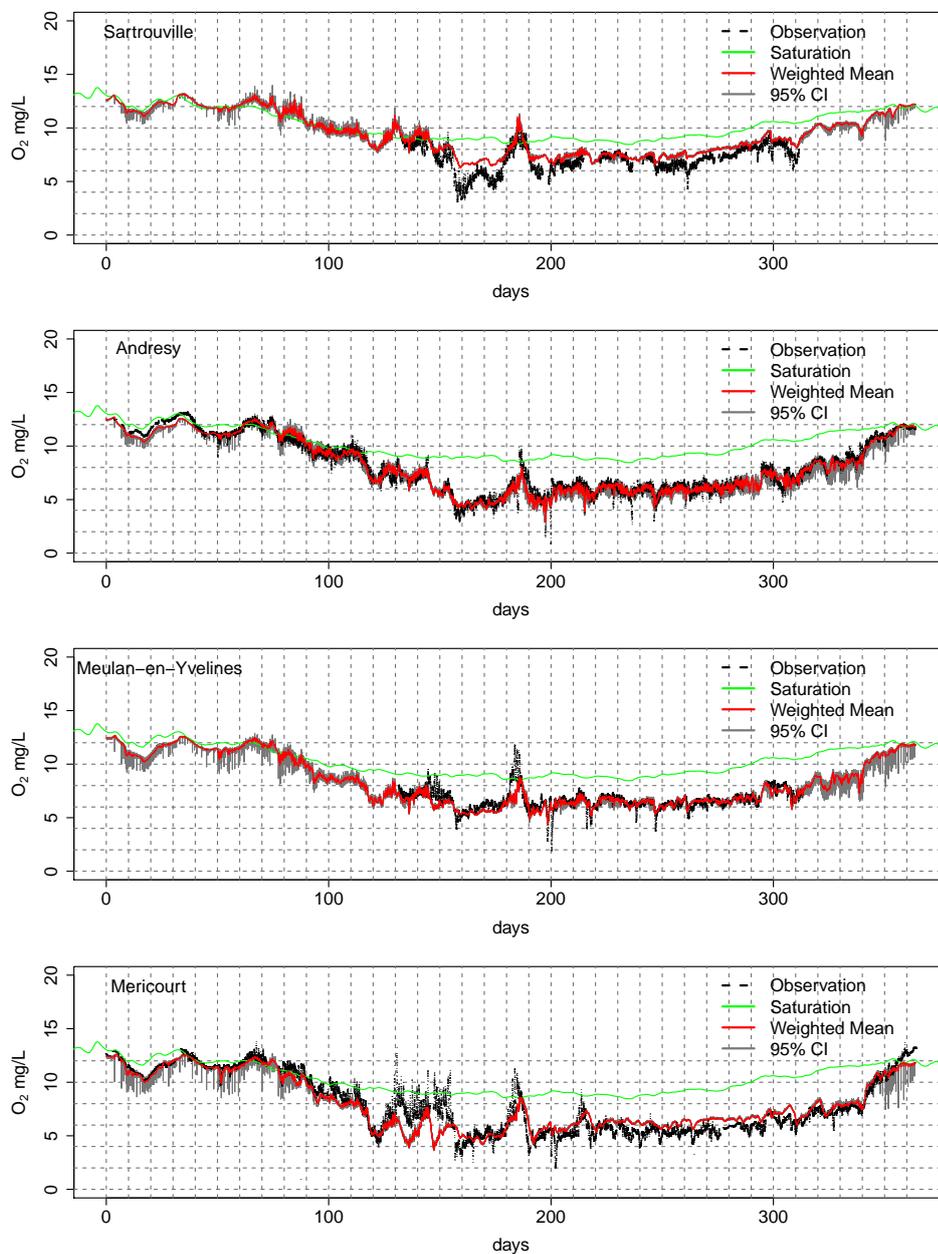


FIGURE 6.7 – Concentrations en O_2 observées (ligne noire) et simulées (ligne rouge) par PROSE-PA aux stations SIAAP en 2011 (le 1 janvier 2011 = jour 0). Les zones grises présentent les intervalles de confiance à 95%.

PROSE-PA sont tous supérieurs à 0.62, tandis que seulement trois coefficients KGEs (Suresnes, Bougival et Andrézy) étaient supérieurs à 0.62 pour les simulations de PROSE3.6.9 (Tab. 6.8). La corrélation entre concentrations simulées par PROSE-PA et concentrations observées est supérieure à 0.74 (Meulan-en-Yvelines). Le biais du modèle ($\beta = \frac{\mu_{sim}}{\mu_{obs}}$) est globalement inférieur à 7%, sauf pour la station Sartrouville. Le biais à la station Sartrouville passe en effet de 29% (PROSE3.6.9) à 11% (PROSE-PA). En revanche, la variabilité ($\gamma = \frac{\mu_{sim}\sigma_{obs}}{\mu_{obs}\sigma_{sim}}$) des simulations est légèrement sous-estimée par rapport aux observations à l'exception de la station Suresnes qui représente une surestimation de la variabilité de 15 %. Il faudrait noter que les concentrations en O_2 entre les jours 250 et 300 sont sous-estimées pour la station Suresnes certainement du fait de l'enrichissement de la MOD biodégradable par la modification des clés de répartition de la MOD. La meilleure simulation est obtenue à la station Andrézy où le coefficient KGE et ses composantes sont excellents entre 0.98 et 1.00.

Ces améliorations contribuent à une meilleure représentation des dynamiques des blooms algaux (Fig. 6.6 et 6.7). De plus, les creux d' O_2 engendrés par les rejets anthropiques sont bien reproduits par PROSE-PA. C'est notamment le cas aux stations Andrézy et Meulan-en-Yvelines qui se trouvent juste en aval de la station d'épuration Seine Aval (Fig. 6.7). Ces chutes d' O_2 étaient difficiles à simuler avec le modèle déterministe direct (Fig. 6.2).

TABLE 6.8 – Critère statistique des concentrations en O_2 simulées aux stations de mesure du SIAAP par PROSE3.6.9 et PROSE-PA

2011 Stations	PROSE-PA					PROSE3.6.9				
	RMSE mgO ₂ /L	KGE	r	α	γ	RMSE mgO ₂ /L	KGE	r	α	γ
Suresnes	0.92	0.83	0.95	0.94	1.15	1.11	0.70	0.88	1.04	0.73
Colombes	0.91	0.63	0.86	1.05	0.66	1.72	0.42	0.72	1.18	0.53
Chatou	0.98	0.79	0.96	1.07	0.80	2.42	0.49	0.75	1.19	0.64
Bougival	0.62	0.97	0.97	1.00	0.99	1.99	0.62	0.71	1.11	0.78
Sartrouville	1.00	0.60	0.87	1.11	0.63	2.50	0.45	0.54	1.29	1.07
Andrézy	0.48	0.98	0.98	0.99	1.00	0.84	0.85	0.95	1.02	0.86
Meulan	0.64	0.62	0.74	0.97	0.73	1.58	0.38	0.42	1.16	1.15
Méricourt	1.00	0.87	0.93	1.00	0.89	1.91	0.58	0.74	1.07	0.68

6.3.4 Identifiabilité des propriétés physiologiques

Afin d'analyser l'identification des propriétés physiologiques, la méthodologie de classification de l'état du système définie dans le cas synthétique (Section 5.2.3.3 pp.144) est utilisée. Trois blooms algaux ($C_{chla} > 12 \mu\text{gchla/L}$) sont identifiés (Fig. 6.8). Le premier bloom débute en mars (62ème jours) et dure environ deux mois. Ce premier bloom a une concentration maximale en *chla* de $75 \mu\text{gchla/L}$. La température de l'eau varie entre 8 et 20 °C. Le bloom de mai dure moins d'un mois avec la concentration maximale en *chla* de $85 \mu\text{gchla/L}$. La température de l'eau pour ce bloom est autour de 20 °C. Le bloom de juillet est beaucoup plus court (14 jours). La concentration maximale en *chla* atteint à $75 \mu\text{gchla/L}$.

De même, les poids normalisés journaliers des paramètres qui représentent l'évolution de la distribution a posteriori de chaque paramètre sont présentés par des image plots (Fig. 6.9, 6.14 et 6.16). Pour cela, la gamme de chaque paramètre est divisée en 20 intervalles. La somme des poids normalisés journaliers dans chacune intervalle approxime la distribution a posteriori de chaque paramètre. De plus, les distributions a posteriori estimées par une méthode à noyaux pondérés des paramètres sont également tracées pour les périodes intéressantes (Fig. 6.10, 6.11, 6.12, 6.15 et 6.17).

6.3.4.1 Paramètres phytoplanctoniques

Bloom de mars

Pour le bloom de mars, PROSE-PA identifie globalement deux propriétés physiologiques. La première est caractérisée par une température optimale ($T_{opt,pp}$) entre 10 et 20 °C (Fig. 6.9 et 6.10A.), tandis que la deuxième a une température optimale plus élevée entre 20 et 30 °C (Fig. 6.9 et 6.10B.). De même, la respiration de maintenance ($R_{m,pp}$) est plus forte au début du bloom de mars avec une moyenne des modes de 0.012 h^{-1} (Fig. 6.11A.) que celle de la fin du bloom de mars (0.006 h^{-1} , Fig. 6.11B.). Le changement du taux de photosynthèse maximal ($P_{max,pp}$) se passe plus tôt vers le 73ème jour (Fig. 6.12A.). La moyenne des modes passe en fait de 0.28 à 0.44 h^{-1} . La distribution a posteriori de la capacité photosynthétique (α_{pp}) est relativement stable pour le bloom de mars (Fig. 6.13). La moyenne des modes est $0.0015 \text{ m}^2 \cdot \text{s} \cdot \mu\text{E} \cdot \text{h}^{-1}$.

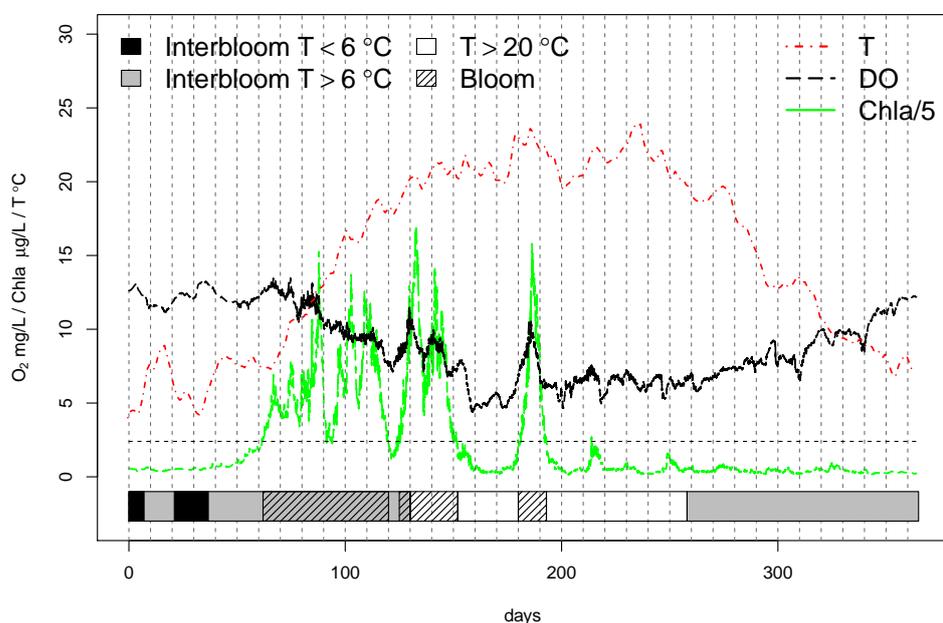


FIGURE 6.8 – Classification de l'état du système en fonction de la biomasse algale et de la température à la station Bougival

Bloom de mai

Pendant le bloom de mai, la respiration de maintenance (R_m) et le taux de photosynthèse maximal ($P_{max,pp}$) ne sont pas stables. Il n'est pas possible d'en identifier un mode principal. En revanche, les autres paramètres sont identifiables. La température optimale ($T_{opt,pp}$) identifiée est élevée ($30.9\text{ }^\circ\text{C}$) et les valeurs de la capacité photosynthétique (α_{pp}) et du terme d'extinction lumineuse liée à la concentration en chla ($\eta_{chla,pp}$) sont cohérentes avec celles utilisées par Garnier et al. (1995) et Vilmin (2014) ($0.0012\text{ m}^2\text{s}\cdot\mu\text{E}^{-1}\cdot\text{h}^{-1}$ et $0.022\text{ L}\cdot\mu\text{gchla}^{-1}\cdot\text{m}^{-1}$ respectivement).

Bloom de juillet

Selon les résultats obtenus (Fig. 6.9), l'identification des paramètres phytoplanctoniques pour le bloom de juillet est cohérente avec celle identifiée pendant le bloom de fin mars et début avril (la deuxième propriété physiologique du premier bloom, Tab. 6.9), à l'exception de la respiration de maintenance (R_m , 0.006 contre 0.014 h^{-1}), qui a été identifié peu influent

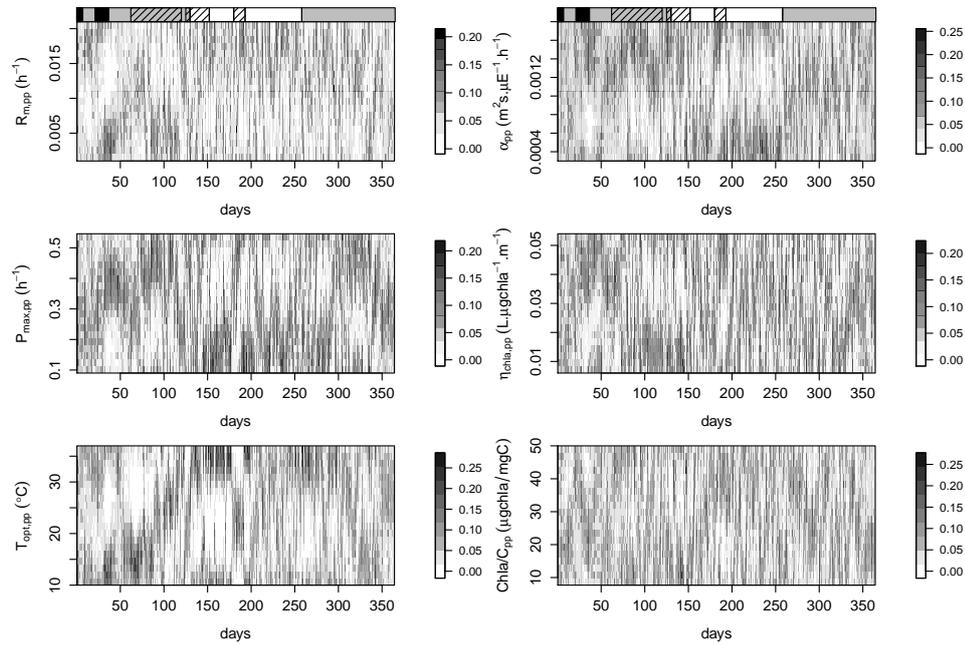


FIGURE 6.9 – L'évolution des poids normalisés pour les paramètres phytoplanctoniques. Les polygones barrés indiquent les périodes des trois blooms (Fig. 6.8) et le gradient de couleur signifie un gradient de température de froid à chaud (couleur de foncée à claire)

quand la température de l'eau dépasse 20 °C.

Par ailleurs, un changement de propriétés lié aux paramètres de photosynthèse est constaté à la fin du bloom de juillet (α_{pp} , P_{max} , $\eta_{chla,pp}$, Fig. 6.9). Le filtre particulaire représente ainsi la mortalité du phytoplancton liée à sa sénescence par une diminution du taux de photosynthèse ($P_{max,pp}$), de la capacité photosynthétique (α_{pp}) et du terme d'extinction lumineuse ($\eta_{chla,pp}$, η_{water}) (Fig. 6.14, jours 193-200).

6.3.4.2 Synthèse des résultats concernant des producteurs primaires

Selon [Escoffier et al. \(2014\)](#), les espèces de diatomées identifiées pour le bloom de mars sont *Synedra ulna*, *Cyclotella spp.*, *Nitzschia spp.*. [Garnier et al. \(1995\)](#) ont souligné que le bloom de printemps dans la Seine est dominé notamment par une espèce de diatomée centrique *Stephanodiscus hantzschii*

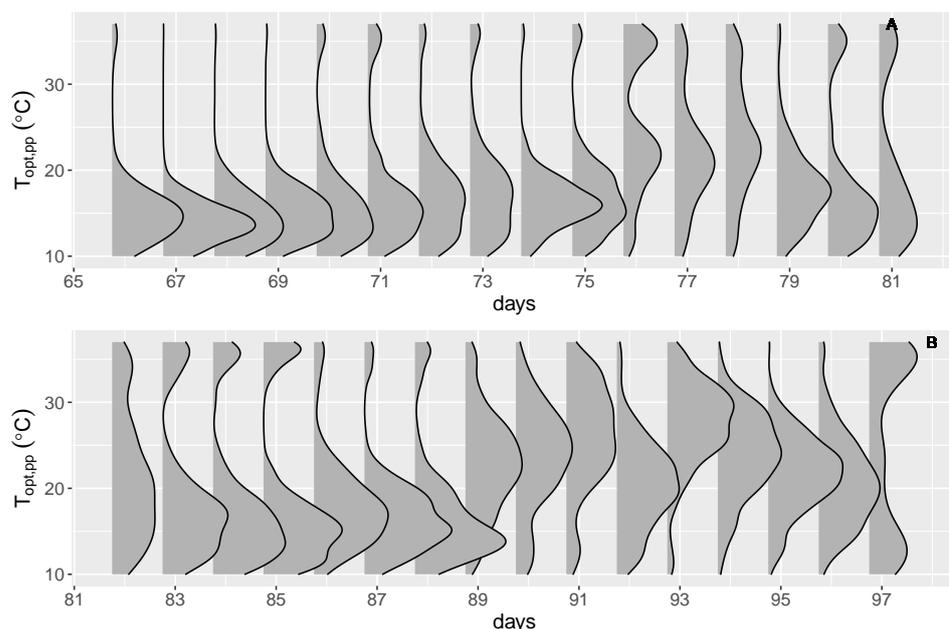


FIGURE 6.10 – L'évolution de la distribution a posteriori de la température optimale ($T_{opt,pp}$) pour le bloom de mars

qui est dans la même famille (*Stephanodiscaceae*) que celle de *Cyclotella spp.*. Descy et al. (2012) indiquent que l'espèce *Stephanodiscus spp.* croit en début de printemps, tandis que *Cyclotella spp.* qui est plus petite, croit en fin de printemps et en été. Descy et al. (2012) ont modélisé le phytoplancton dans la Loire en utilisant deux propriétés physiologiques différentes pour *Stephanodiscus spp.* et *Cyclotella spp.*. Les températures optimales utilisées pour les deux espèces sont 11.0 et 17.0 °C. Vilmin (2014) a utilisé 10.0 et 23.0 °C. Ces valeurs sont globalement identifiées par PROSE-PA pour le bloom de mars (deux) et le bloom de juillet (Tab. 6.9).

Pour le bloom de juillet, *Cyclotella spp.* est observée comme la seule espèce algale (Ecoffier et al., 2014). Les paramètres identifiés sont globalement cohérents pour le bloom de juillet et le bloom de fin mars (Tab. 6.9), à l'exception de la respiration de maintenance ($R_{m,pp}$). Ceci indique que l'espèce *Cyclotella spp.* pourrait également être aussi dominante en fin du bloom de mars avec une température optimale de 24.3 °C.

Pour le bloom de mai, même si l'identification des paramètres $R_{m,pp}$ et

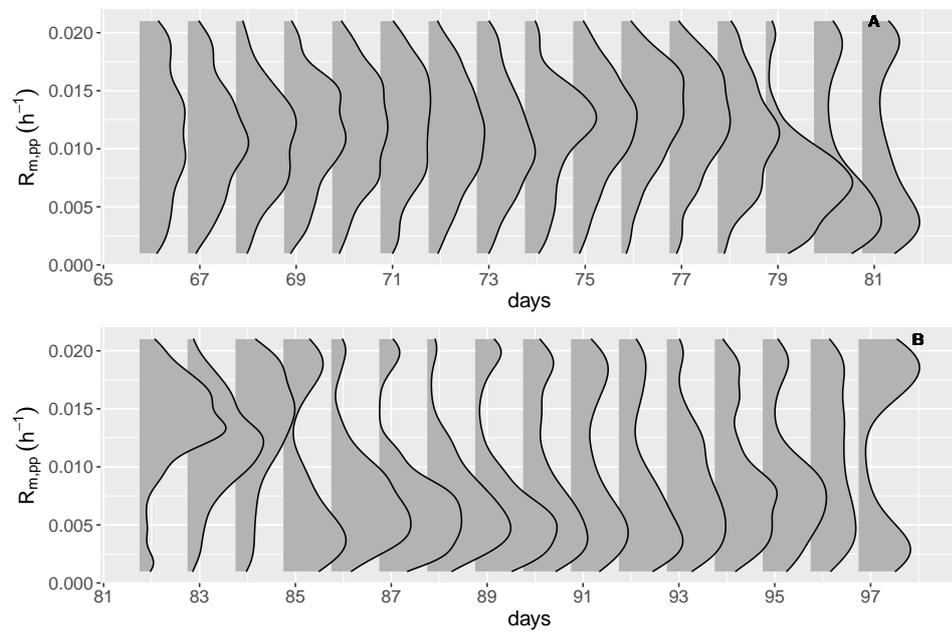


FIGURE 6.11 – L'évolution de la distribution a posteriori de la respiration de maintenance ($R_{m,pp}$) pour le bloom de mars, le 1 janvier 2011 = jour 0

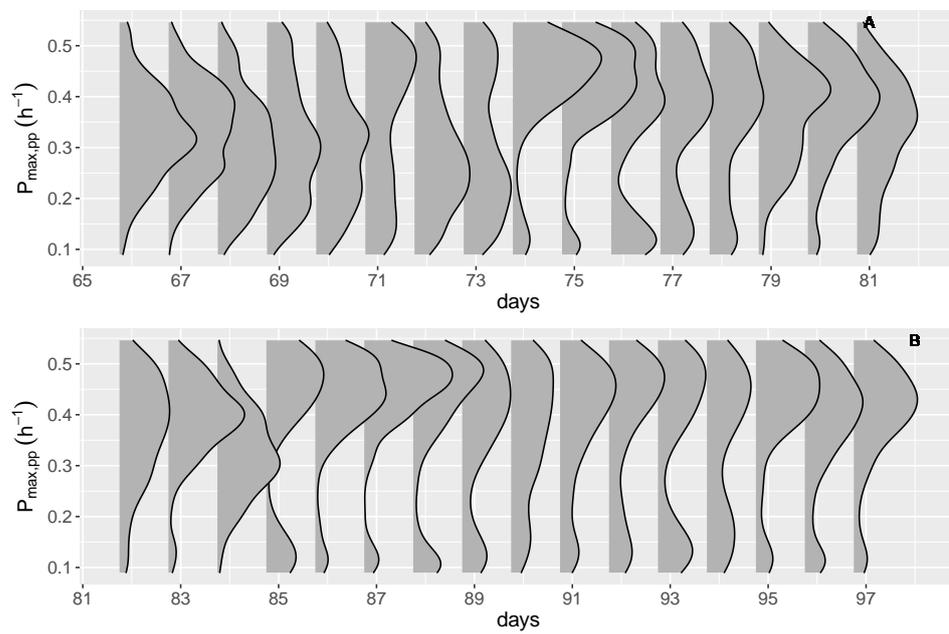


FIGURE 6.12 – L'évolution de la distribution a posteriori du taux de photosynthèse maximal ($P_{max,pp}$) pour le bloom de mars, le 1 janvier 2011 = jour 0

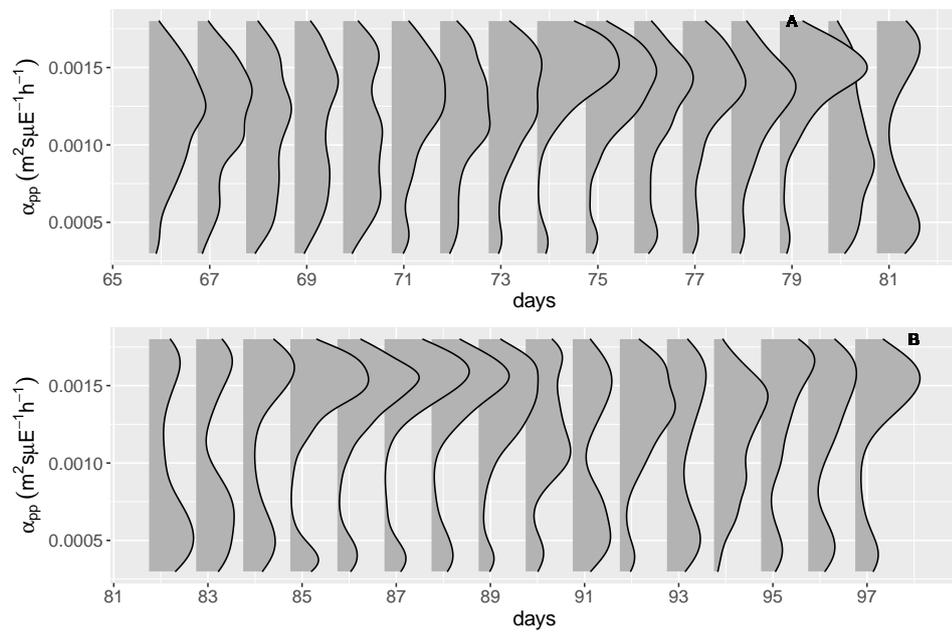


FIGURE 6.13 – L'évolution de la distribution a posteriori de la capacité photosynthétique (α_{pp}) pour le bloom de mars, le 1 janvier 2011 = jour 0

$P_{max,pp}$ n'est pas stable, les deux pics de concentrations en O_2 sont bien reproduits (120-150 jours). Le bloom de mai est principalement constitué par l'espèce dinoflagellées enkystées (Groleau et al., 2014). Un taux de croissance nul a été utilisé par Vilmin (2014) pour cette espèce. Cependant le taux de croissance de 0.025 h^{-1} correspondant à celui du bloom de juillet utilisé par Vilmin (2014) est appliqué dans cette étude. Ceci pourrait expliquer la surestimation des concentrations en chl_a à la station Bougival durant les mois d'avril et mai (Fig. 6.8). Pour réduire cette surestimation en chl_a , il est possible de simuler deux espèces de phytoplancton : une espèce à assimiler et une espèce dinoflagellées enkystées. Les paramètres de dinoflagellées enkystées seraient fixés. Une comparaison des valeurs des paramètres identifiées (les moyennes temporelles des modes sur la période de temps considérée) par PROSE-PA avec celles utilisées par Garnier et al. (1995), Descy et al. (2012) et Vilmin (2014) est synthétisées dans la table 6.9.

TABLE 6.9 – Comparaison des moyennes temporelles des modes des paramètres identifiées par PROSE-PA avec celles utilisées dans la littérature

Paramètres	PROSE-PA				Vilmin (2014)		Garnier et al. (1995)
	mars1	mars2	mai	juillet	mars	juillet	diatomée
$T_{opt,pp}$ [°C]	14.8	24.3	30.9	22.8	10.0	23.0	21.0
$R_{m,pp}$ [h^{-1}]	0.012	0.006	-	0.014	0.002	0.002	0.002
$P_{max,pp}$ [h^{-1}]	0.28	0.44	-	0.43	0.20	0.20	0.20
α_{pp} [$\text{m}^2\text{s}.\mu\text{E}^{-1}.\text{h}^{-1}$]	0.0013	0.0016	0.0012	0.0012	0.0012	0.0012	0.0012
$\eta_{chl_a,pp}$ [$\text{L}.\mu\text{gchl}_a^{-1}.\text{m}^{-1}$]	0.036	0.026	0.022	0.036	0.020	0.020	0.030
η_{water} [m^{-1}]	0.42	0.35	0.39	0.44	1.00	1.00	0.32

Descy et al. (2012) T_{opt} : *Stephanodiscus spp.* 11 °C ; *Cyclotella spp.* 17 °C ;

- : non stable

Voir la table 5.1, p.143 pour la signification des paramètres

6.3.4.3 Paramètres physiques

Extinction lumineuse de l'eau pure

La photosynthèse est réduite par les effets d'extinction lumineuse de l'eau

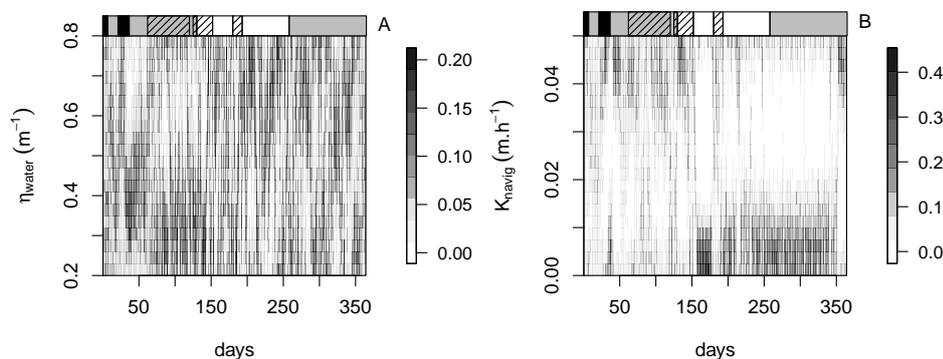


FIGURE 6.14 – L'évolution des poids normalisés pour les paramètres physiques

pure (η_{water}). η_{water} est relativement stable pour les périodes de blooms. La moyenne des modes pour les trois blooms est entre 0.35 et 0.44 m^{-1} (Tab. 6.9). Cette gamme est cohérente avec la valeur 0.32 m^{-1} identifiée par (Garnier et al., 1995).

Réaération

Le coefficient de réaération liée à la navigation est identifié comme le plus influent quand la température est inférieure à 6 °C (période symbolisée par un polygone noir, Fig. 6.14). Pour cette période (jours 21-36), les concentrations en O₂ observées se situent sous la saturation (Fig. 6.6 et Fig. 6.7). Elles sont légèrement surestimées par PROSE-PA pour certaines stations se trouvant en amont tandis que les concentrations en O₂ sont sous-estimées en hiver à l'aval. Une valeur proche de 0.05 $m.h^{-1}$ est ainsi identifié en hiver pour le coefficient de réaération liée à la navigation (Fig. 6.15). Pour les jours 33-38, un changement de l'effet de la navigation est observé (Fig. 6.15). Cet effet peut être également observé pour les jours 200-350 et la période entre le bloom de mai et le bloom de juillet (les 155-180, Fig. 6.14). Pendant ces périodes, les concentrations en O₂ observées sont sous saturées et le modèle estime des concentrations plus élevées que les observations. Le modèle cherche ainsi à réduire les écarts entre simulation et observation en annulant le coefficient de réaération ($K_{navig} = 0 m.h^{-1}$). En revanche, la variabilité des concentrations en O₂ est sous-estimée pour les périodes de blooms algaux. Une valeur de K_{navig} proche de 0.05 $m.h^{-1}$ est alors identi-

fiée.

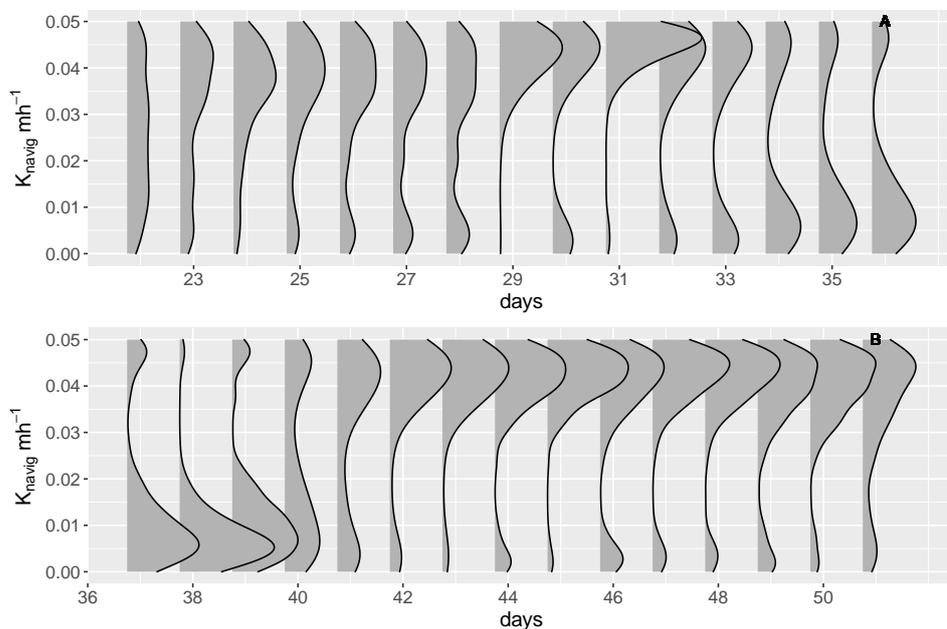


FIGURE 6.15 – L'évolution de la distribution a posteriori du coefficient de réaération liée à la navigation (K_{navig}) pour l'hiver, le 1 janvier 2011 = jour 0

6.3.4.4 Paramètres bactériens

Avant le bloom algal

Le taux de croissance maximal ($\mu_{max,hb}$), le rendement de croissance (Y_{hb}) et la température optimale pour la croissance des bactéries ($T_{opt,hb}$) peuvent être identifiés avant le bloom de mars (Fig. 6.16). $\mu_{max,hb}$ a une valeur relativement stable de 0.03 h^{-1} . De même, les modes des distributions a posteriori de Y_{hb} et $T_{opt,hb}$ sont près de 0.41 et de $31.1 \text{ }^\circ\text{C}$ (Fig. 6.17).

Entre deux blooms algaux

De même, une bonne identification des paramètres bactériens est observée pour la période entre le bloom de mai et le bloom de juillet (jours 156-170, Fig. 6.16). Le taux de croissance maximal ($\mu_{max,hb}$) est proche de la valeur maximale (0.13 h^{-1} , Fig. 6.18). Ce qui indique une forte croissance des

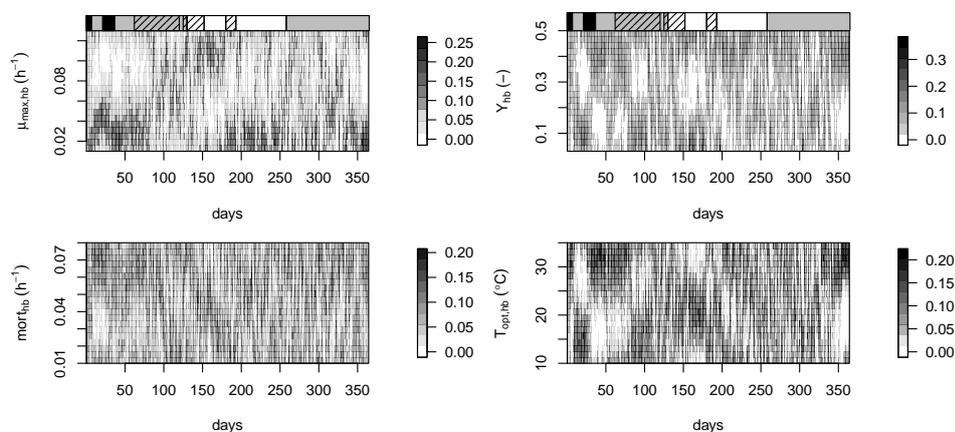


FIGURE 6.16 – L'évolution des poids normalisés pour les paramètres bactériens

bactéries hétérotrophes. En revanche, le rendement de croissance (Y_{hb}) est concentré sur une valeur faible (0.08, Fig. 6.18). Un rendement de croissance de 0.08 signifie que 92% du prélèvement de MOD labile est respirée. Cet effet est cohérent avec les concentrations en O_2 simulées pour cette période où elle sont globalement surestimées à l'exception à la station Suresnes (Fig. 6.6 et 6.7). Rappelons que le filtre a du mal à corriger les concentrations en O_2 pour cette période (cf. Section 6.4, p.179). Le schéma d'assimilation a tendance à favoriser les dynamiques des bactéries hétérotrophes pour mieux corriger les concentrations en O_2 . Les valeurs identifiées pour la température optimale (T_{opt} , 20.9 °C) et le taux de mortalité (0.048 h^{-1}) sont cohérentes avec celles utilisées par Garnier et al. (1995).

Après le bloom algal

Les mêmes valeurs de paramètres que celles trouvées entre le bloom de mai et le bloom de juillet peuvent être également identifiées pour quelques jours après le bloom de juillet (jours 195-200). Cependant, des oscillations des valeurs des paramètres sont observées après environ 200 jours. C'est notamment le cas pour le rendement de croissance (Y_{hb}), le taux de mortalité ($mort_{hb}$) et la température optimale ($T_{opt,pp}$). Ces oscillations révèlent un problème de compensation entre eux.

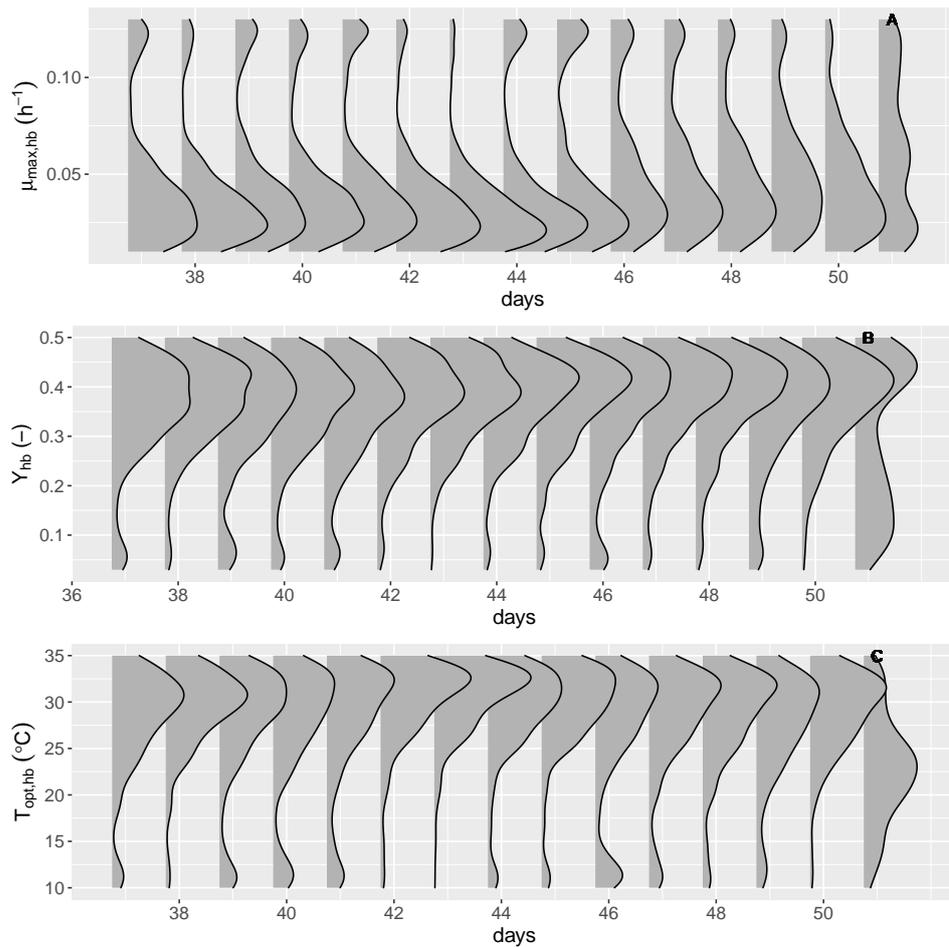


FIGURE 6.17 – L'évolution des distributions a posteriori des paramètres bactériens : le taux de croissance maximal ($\mu_{max,hb}$), le rendement de croissance (Y_{hb}) et la température maximale ($T_{opt,hb}$) avant le bloom de mars, le 1 janvier 2011 = jour 0

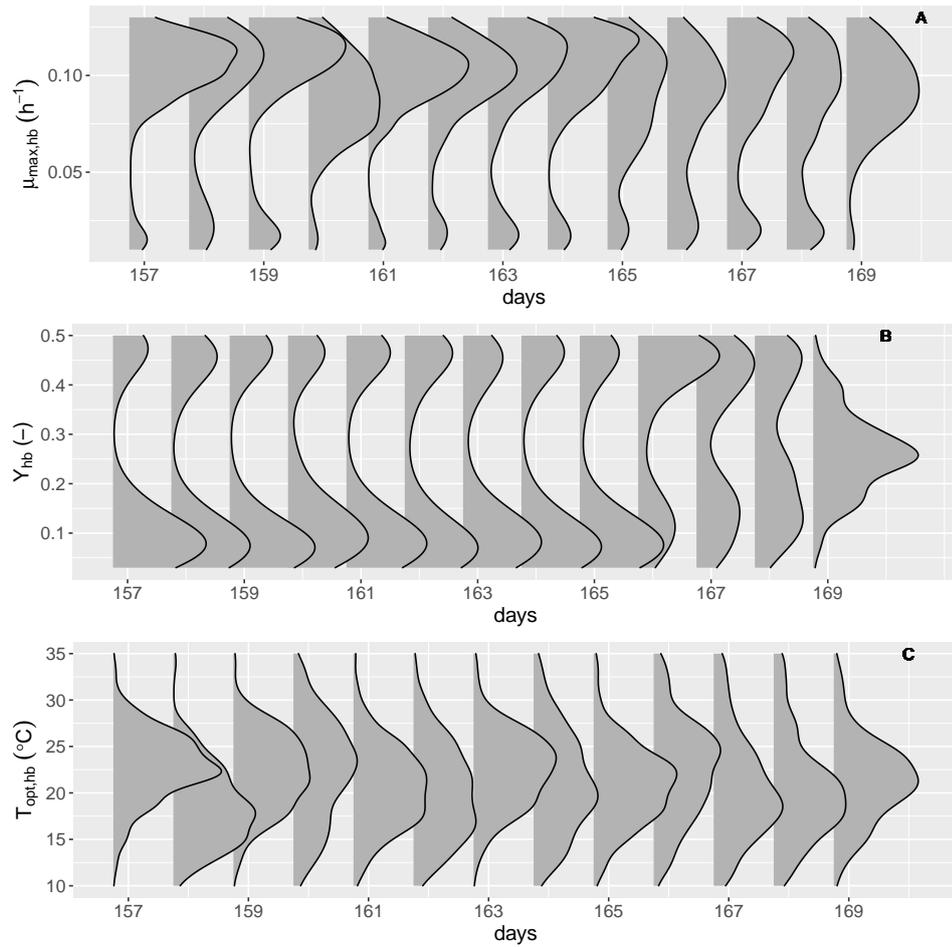


FIGURE 6.18 – L'évolution des distributions a posteriori des paramètres bactériens : le taux de croissance maximal ($\mu_{max,hb}$), le rendement de croissance (Y_{hb}) et la température maximale ($T_{opt,hb}$) entre le bloom de mai et le bloom de juillet, le 1 janvier 2011 = jour 0

6.3.5 Difficultés d'identification des paramètres

Les concentrations en O_2 sont significativement améliorées par l'assimilation de données. Cependant, il est difficile de les simuler pour la période entre le bloom de mai et le bloom de juillet où elles sont surestimées par PROSE-PA quand le filtre favorise les dynamiques des bactéries hétérotrophes. Ceci est lié au déficit en MOD à ce moment. Afin d'améliorer ces simulations, une piste de réflexion serait d'enrichir le milieu en MOD biodégradable à cette période. Il convient cependant de faire un test de sensibilité aux clés de répartition de la MOD détritique. On pourrait aussi imaginer inclure les clés de répartition de la MOD parmi les paramètres dynamiques.

Pour les jours 200-350, certains paramètres présentent un caractère oscillatoire dont on souhaite tester l'éventuelle synchronicité. Pour cela, les corrélations entre les concentrations en O_2 journalières et les valeurs correspondant aux maxima journaliers a posteriori des paramètres sont analysées conjointement aux concentrations d' O_2 aval à Meulan, au débit à Meulan et aux conditions aux limites décalées du temps de transfert de l'eau depuis l'amont jusqu'à Meulan. En analysant le débit à Austerlitz plus précisément, il apparaît relativement stable entre les jours 240 et 290. On s'intéresse donc en particulier à cette période pour étudier les relations entre ces différentes variables (Fig. 6.19).

Mises à part les relations connues entre les variables MOD et débit, il est difficile d'identifier des relations nettes entre :

- les variations des paramètres entre eux ;
- les variations des paramètres et les variables MOD, débit et oxygène ;
- l'oxygène et le débit.

Il semble toutefois que des relations de dépendance entre paramètres existent. Par exemple, lorsque le paramètre $R_{m,pp}$ est fort, $T_{opt,pp}$ n'est pas contraint, et réciproquement. En effet, si aucune relation n'existait, l'ensemble de l'espace des paramètres devrait être uniformément couvert.

Ces observations indiquent qu'un travail important est à mener pour intégrer les erreurs portées par les conditions aux limites au schéma d'assimilation de données de PROSE-PA.



FIGURE 6.19 – Corrélations entre les valeurs correspondant aux maxima journaliers a posteriori, les concentrations d' O_2 à Meulan et les conditions aux limites pour les jours 240 à 290. Voir la table 5.1 pour la signification des paramètres.

6.3.6 Recommandations pour la perturbation après le ré-échantillonnage

La perturbation des paramètres est effectuée après le ré-échantillonnage. Pour perturber les paramètres, deux stratégies sont envisagées dans PROSE-PA : le domaine des paramètres bouclé et le domaine des paramètres non bouclé. Les résultats présentés précédemment sont issus de la stratégie du domaine des paramètres bouclé. Une comparaison des deux stratégies est réalisée.

- Le domaine des paramètres bouclé : Les deux extrêmes (`val_min`, `val_max`) se connectent. Cette stratégie est utilisée pour le cas synthétique.
- Le domaine des paramètres non bouclé : Les deux extrêmes (`val_min`, `val_max`) sont des bornes infranchissables. Quand le paramètre se trouve en dehors du domaine (`val_max` ou `val_min`) après la perturbation, il reste inchangé.

Les concentrations en O₂ simulées par la stratégie du domaine des paramètres bouclés sont généralement meilleures que celles simulées par la stratégie du domaine des paramètres non bouclés, notamment à la station Méricourt (Tab. 6.10, Fig. A.1 et A.2). Une meilleure simulation de l'oxygène dissous assure une meilleure estimation des propriétés physiologiques. La stratégie du domaine bouclé permet de mieux explorer l'espace des paramètres au sens statistique.

En revanche, en utilisant la stratégie du domaine bouclé, la distribution a posteriori d'un paramètre est parfois bimodale (Fig. 6.11B et 6.12B). Avec la stratégie du domaine non bouclé, on peut éviter ce phénomène. Des tests devraient être réalisés pour mieux évaluer les avantages et les inconvénients des deux stratégies.

6.4 Conclusions

En conclusion, les données de l'oxygène dissous de l'année 2011 en Seine ont été assimilées par PROSE-PA après avoir redéfini les clés de répartition de la MOD et calibré les réactivités de la MOD. D'après les résultats obtenus, quelques conclusions positives peuvent être tirées.

- PROSE-PA améliore significativement les simulations des concentrations en O₂ en Seine. Les chutes rapides d'oxygène pour les périodes

TABLE 6.10 – Critère statistique des concentrations en O_2 simulées aux stations de mesures du SIAAP pour deux stratégies de perturbations

2011 Stations	PROSE-PA bouclé					PROSE-PA non bouclé				
	RMSE mgO ₂ /L	KGE	r	α	γ	RMSE mgO ₂ /L	KGE	r	α	γ
Suresnes	0.92	0.83	0.95	0.94	1.15	0.93	0.86	0.93	0.95	1.11
Colombes	0.91	0.63	0.86	1.05	0.66	1.00	0.61	0.85	1.07	0.65
Chatou	0.98	0.79	0.96	1.07	0.80	1.09	0.75	0.96	1.09	0.77
Bougival	0.62	0.97	0.97	1.00	0.99	0.70	0.93	0.96	1.02	0.95
Sartrouville	1.00	0.60	0.87	1.11	0.63	1.10	0.59	0.88	1.13	0.63
Andrésey	0.48	0.98	0.98	0.99	1.00	0.47	0.98	0.98	0.99	1.00
Meulan	0.64	0.62	0.74	0.97	0.73	0.86	0.63	0.64	0.93	0.97
Méricourt	1.00	0.87	0.93	1.00	0.89	1.51	0.76	0.88	0.94	1.19

de crise sont reconstituées.

- Les paramètres phytoplanctoniques sont bien identifiés pendant les blooms algaux. Les valeurs des paramètres (les moyennes temporelles des modes des paramètres) correspondent bien aux valeurs utilisées dans la littérature.
- Le schéma d'assimilation de données dans PROSE-PA permet d'identifier les propriétés bactériennes pour les périodes avant le bloom de mars et entre deux blooms algaux en 2011.

A l'exception des points positives, quelques questions restent à étudier.

- Une surestimation des concentrations en O_2 entre le bloom de mai et le bloom de juillet est souligné.
- La variabilité des concentrations en O_2 est globalement sous-estimée (la variation jour-nuit et la chute d' O_2 liée aux déversoirs d'orage).
- Des oscillations des valeurs correspondant aux maxima a posteriori des paramètres sont observées pour les jours 200-350.

Afin d'améliorer les performances de PROSE-PA, plusieurs pistes sont proposées.

- La spéciation de la MOD en Seine permet d'améliorer certainement la simulation des concentrations en O_2 .
- Il est possible d'intégrer les clés de répartition de la MOD dans les paramètres dynamiques.

-
- L'analyse sur la transformation du phytoplancton détritique en MOD et MOP atténuerait la surestimation des concentrations en O_2 pour la période entre le blomm de mai et le bloom de juillet.
 - Il convient de tester les simulations multi-espèces de phytoplancton et de bactérie.

Conclusion et perspectives

Conclusion

En conclusion, dans cette thèse, un filtre particulaire a été développé afin d'assimiler les concentrations en O_2 toutes les 15 minutes et de déterminer les propriétés physiologiques des communautés vivantes. Pour cela, diverses approches ont été mises en œuvre :

- Une nouvelle version de PROSE, PROSE-P, a été construite qui permet de fonctionner en mode parallèle. L'amélioration des performances comprend, une réduction des temps de calcul se basant sur la parallélisation qui permet d'exploiter les nouvelles capacités des machines multi-processeurs, une plus grande robustesse d'une simulation hydraulique, et une meilleure représentation des échanges eau-air intégrant les effets thermiques.
- L'analyse de sensibilité du module biogéochimique C-RIVE a été réalisée pour plusieurs conditions hydrologiques et trophiques. Cette analyse de sensibilité a permis d'identifier les paramètres sensibles (physiques, bactériens et phytoplanctoniques) qui contrôlent la variation d' O_2 dans la colonne d'eau en fonction des états trophique et thermique du système.
- L'implémentation pour la première fois d'un algorithme de filtrage particulaire couplé à un modèle de simulation de la qualité de l'eau : PROSE-PA.
- Les performances de PROSE-PA ont été évaluées dans un premier temps sur un cas synthétique en étudiant la capacité du filtre à capturer la variabilité temporelle continue des paramètres physiologiques des communautés de micro-organismes. Cette étude sur le cas synthétique a démontré l'efficacité du filtrage particulaire pour l'assimilation de données en biogéochimie.

- Les clés de répartition des MODs réfractaires apportées par les grands fleuves en MOD biodégradable ont été revues, et une calibration des réactivités de la MOD biodégradable a été menée afin d'atténuer les surestimations de l'oxygène dissous en temps sec.
- L'assimilation des vraies concentrations en O_2 en 2011 aux stations du SIAAP a été réalisée. Une amélioration significative de la simulation des concentrations en O_2 a été observée, notamment les chutes rapides d' O_2 des périodes de crise et les dynamiques des blooms algaux. Les successions de blooms algaux ont également été identifiées par le biais de propriétés physiologiques variables dans le temps mais relativement stable à l'échelle de l'évènement.
- Les paramètres bactériens ont été bien identifiés pour les périodes avant un bloom algal et entre deux blooms algaux. Cependant, l'identifiabilité des paramètres bactériens en temps sec reste à étudier. Pour cette période, un effet probable de compensation des paramètres est observé.

Les résultats de l'assimilation des vraies concentrations en O_2 montrent que le modèle PROSE-PA est opérationnel et prêt pour compléter notre connaissance sur le fonctionnement biogéochimique de la Seine et estimer l'évolution temporelle des propriétés physiologiques des communautés vivantes.

Perspectives

Un outil d'assimilation des concentrations en O_2 , PROSE-PA, a été développé et appliqué sur le modèle Seine depuis l'amont de l'agglomération parisienne jusqu'à l'estuaire de la Seine. Bien que les dynamiques des blooms algaux et les chutes rapides d' O_2 soient bien reproduites, des améliorations sont encore à apporter au logiciel.

- Les conditions aux limites des flux portent souvent une forte incertitude. La quantification des incertitudes liées aux conditions aux limites est nécessaire pour une simulation plus robuste, notamment l'estimation des propriétés physiologiques des communautés vivantes. Cette sensibilité du modèle liée aux conditions aux limites n'a pas été prise en compte dans ce travail. Cependant, la répartition des MODs réfractaires apportées par les grandes fleuves en MODs bio-

dégradables a permis d'améliorer la simulation d' O_2 pour les périodes d'étiage sévère. Une étude récente de la spéciation de MOD de l'amont à l'aval de l'agglomération parisienne par spectrométrie de fluorescence 3D a été réalisée par [Goffin et al. \(2017\)](#). La spectrofluorimétrie 3D présente la possibilité d'être utilisée en temps réel. Cette spéciation de MOD en ligne et en temps réel prévue dans la phase VIII du programme PIREN-SEINE va sûrement améliorer notre connaissance sur les flux de MODs biodégradables apportés par non seulement les grands fleuves mais aussi les rejets de station d'épuration et les déversoirs d'orage en temps de pluie. L'intégration de cette connaissance dans PROSE-PA ouvre de belles perspectives de progrès significatifs.

- PROSE-PA surestime les concentrations en O_2 pour les périodes entre le bloom de mai et le bloom de juillet. Ceci est certainement dû à un déficit de MOD biodégradable. Il convient donc d'étudier les sources de MOD biodégradable qui enrichissent le système. Pour cela, les clés de répartition de la MOD détritique post-bloom sont à reprendre. Pour le moment, 45% des détritiques du phytoplancton sont transformés en MOD biodégradable (30% pour MOD1 et 15% pour MOD2). Il serait également probable de compléter cette étude par celle des cinétiques d'hydrolyse de la MOP en MOD.
- Les bactéries hétérotrophes peuvent être divisées en deux catégories dans la Seine : petites et grosses ([Garnier et al., 1992](#)). Les grosses bactéries sont issues des rejets de station d'épuration. Leurs propriétés physiologiques peuvent ainsi être bien déterminées par des expériences de laboratoire. Il serait donc pertinent de centrer le schéma d'assimilation sur les bactériennes présentes dans le milieu, soit les petits bactériens. Il conviendrait alors de distinguer deux espèces dans le modèle en supposant une parfaitement connue. Un premier test simulant deux espèces de bactéries hétérotrophes a été réalisé ([Annexe B](#)). Les résultats montrent qu'on a besoin de bien différencier la répartition des bactéries hétérotrophes dans le système pour renseigner les conditions aux limites du modèle.

Le schéma d'assimilation de PROSE-PA est possible d'inclure une fonction multi-objective afin d'assimiler les autres variables comme nitrate, orthophosphates, ammonium. Au delà du métabolisme classique (C/O/N/P),

les processus liés à l'adsorption des contaminants, aux polluants organiques persistants et aux métaux restent à intégrer dans PROSE-PA. Application de cet outil avec ou sans l'assimilation de données à des jeux de données événementiel comme qu'il a été fait pour la crue de juin 2016 (Flipo et al., 2018) permettra de quantifier des flux de matières en plus de la caractérisation de l'état du système en crue.

Ces améliorations techniques, meilleure description de la spéciation de la MOD, ajustement des ratios de transformations de la MO détritique, et boucle microbienne multi-espèces devraient amener le code PROSE-P à un niveau de performances inégalées. Suite à ces modifications, PROSE-PA devrait permettre la ré-analyse des données historiques d'oxygène mesuré depuis plus de vingt ans par le réseau MeSeine. Une autre perspective alléchante est l'insertion de cet outil dans un chaîne intégrée d'aide à la décision pour le pilotage en temps réel d'une partie du cycle de l'eau. C'est à dire la gestion du système couplé réseau d'assainissement-STEP-milieu récepteur.

Annexe A

Résultats de la simulation d'une espèce de bactérie hétérotrophe avec la stratégie non bouclée

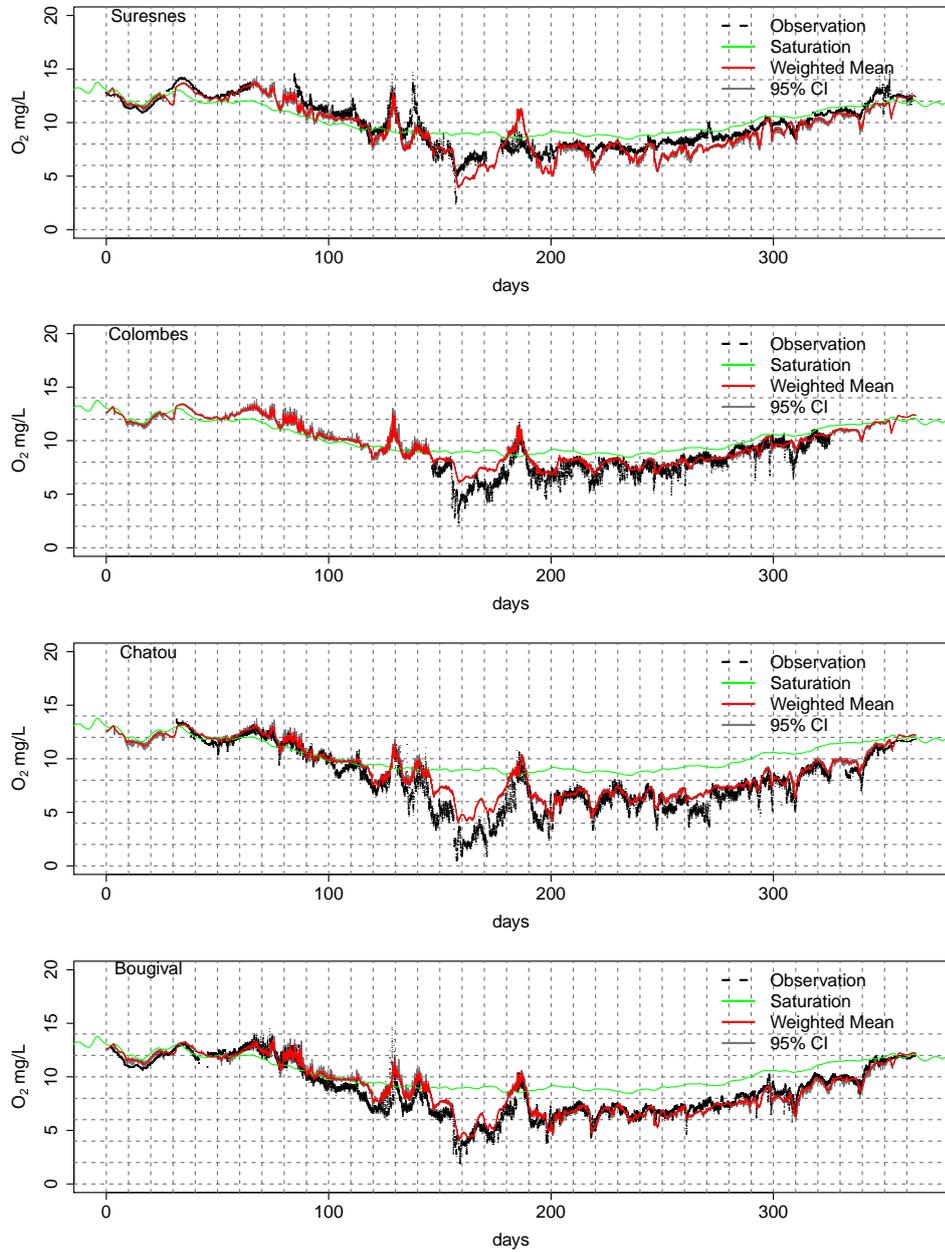


FIGURE A.1 – Concentrations en O_2 observées et simulées par PROSE-PA aux stations du SIAAP avec la stratégie du domaine des paramètres non bouclé

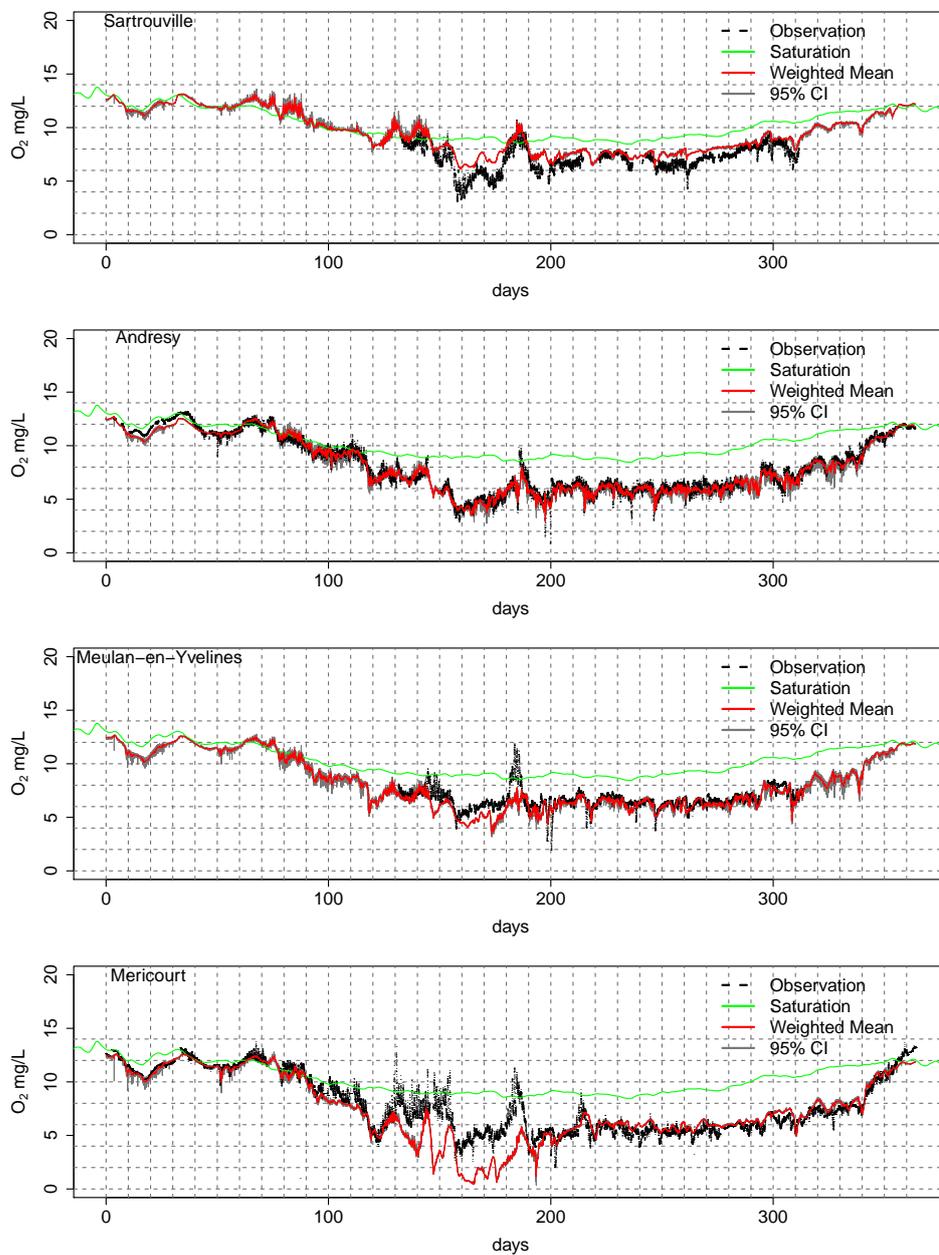


FIGURE A.2 – Concentrations en O_2 observées et simulées par PROSE-PA aux stations du SIAAP avec la stratégie du domaine des paramètres non bouclé

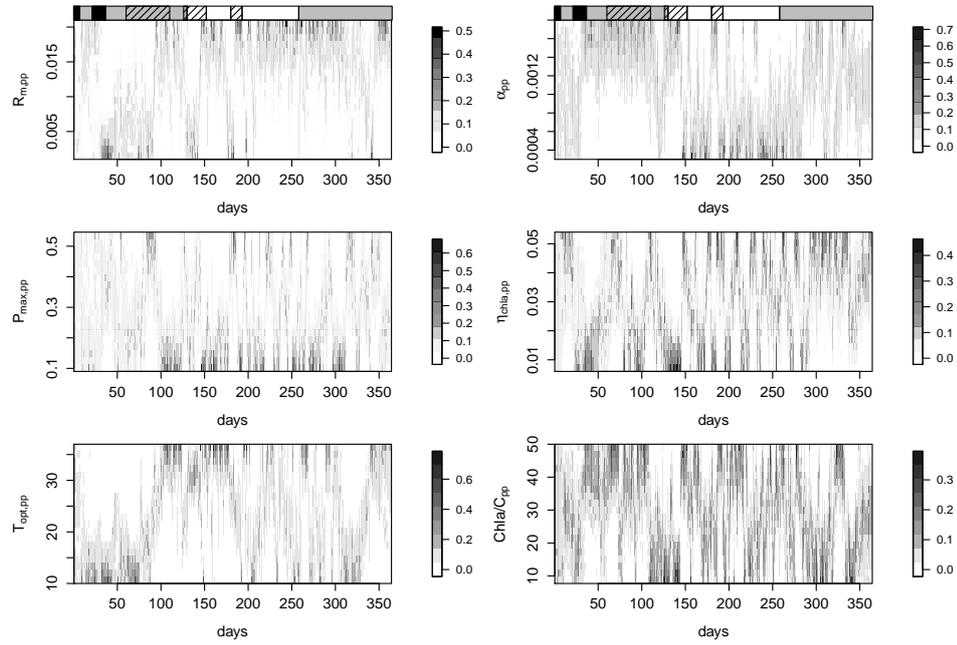


FIGURE A.3 – L'évolution des poids normalisés pour les paramètres phyto-
planctoniques avec la stratégie du domaine non bouclé. Voir la figure 6.8
pour la classification des polygones en haut et la table 5.1 pour les unités.

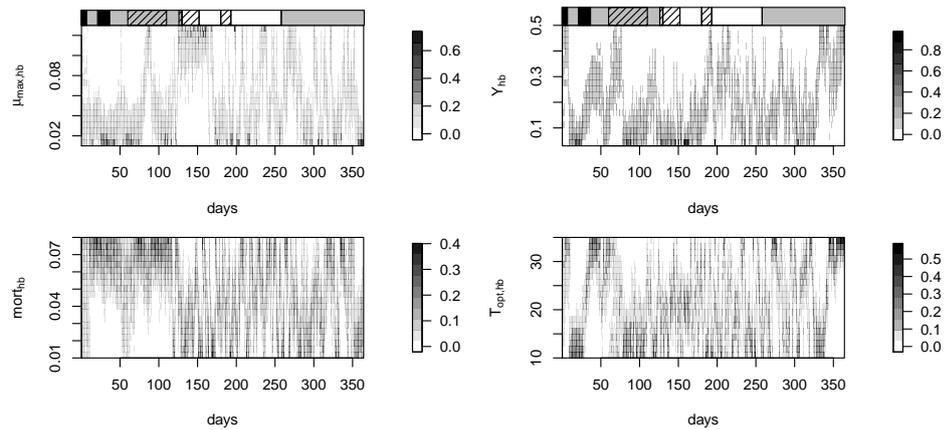


FIGURE A.4 – L'évolution des poids normalisés pour les paramètres bac-
tériens avec la stratégie du domaine non bouclé. Voir la figure 6.8 pour la
classification des polygones en haut et la table 5.1 pour les unités.

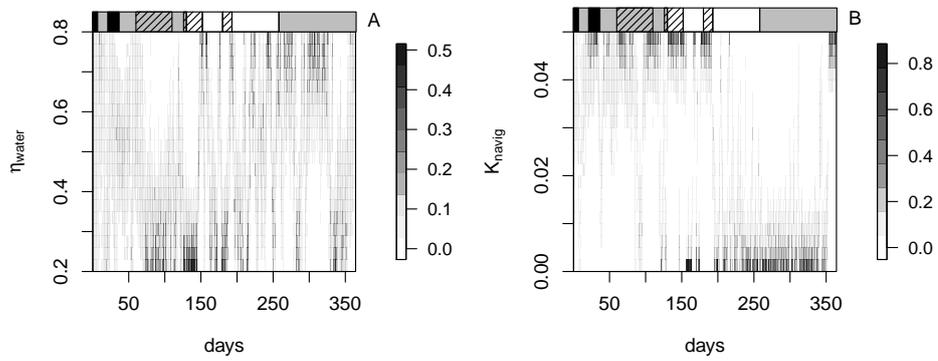


FIGURE A.5 – L'évolution des poids normalisés pour les paramètres physiques avec la stratégie du domaine non bouclé. Voir la figure 6.8 pour la classification des polygones en haut et la table 5.1 pour les unités.

Annexe B

Résultats de la simulation de deux espèces de bactérie hétérotrophe avec la stratégie bouclée

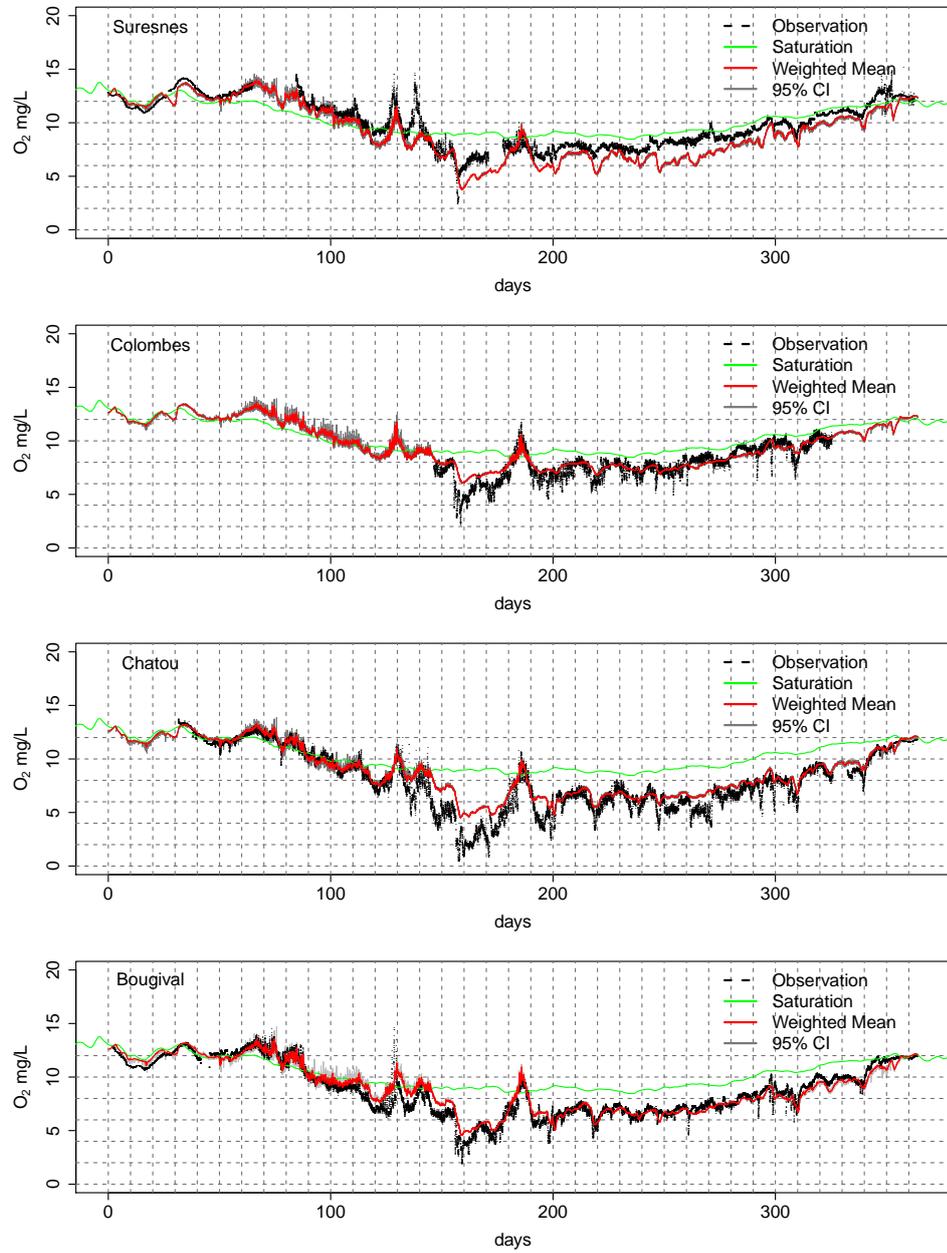


FIGURE B.1 – Concentrations en O_2 observées et simulées par PROSE-PA aux stations du SIAAP avec la stratégie du domaine des paramètres bouclé et deux espèces bactériennes-1

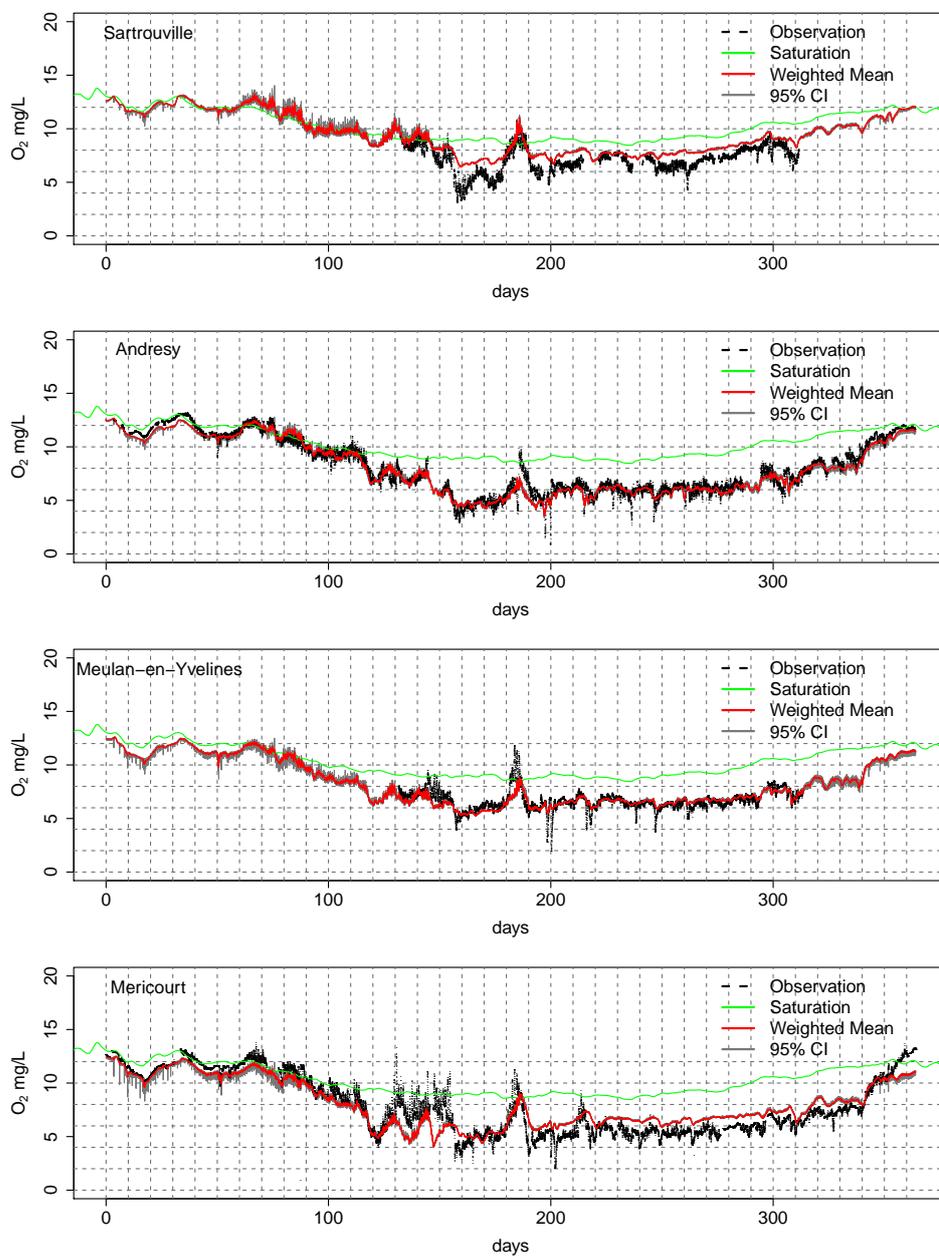


FIGURE B.2 – Concentrations en O_2 observées et simulées par PROSE-PA aux stations du SIAAP avec la stratégie du domaine des paramètres bouclé et deux espèces bactériennes-2

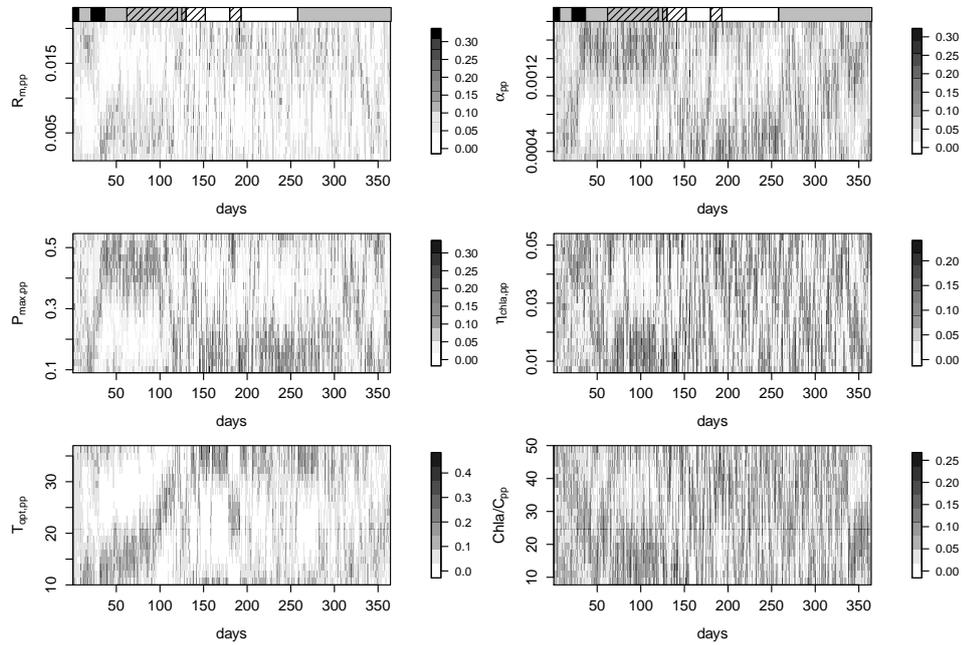


FIGURE B.3 – L'évolution des poids normalisés pour les paramètres phyto-planktoniques avec la stratégie du domaine bouclé et deux espèces bactériennes. Voir la figure 6.8 pour la classification des polygones en haut et la table 5.1 pour les unités.

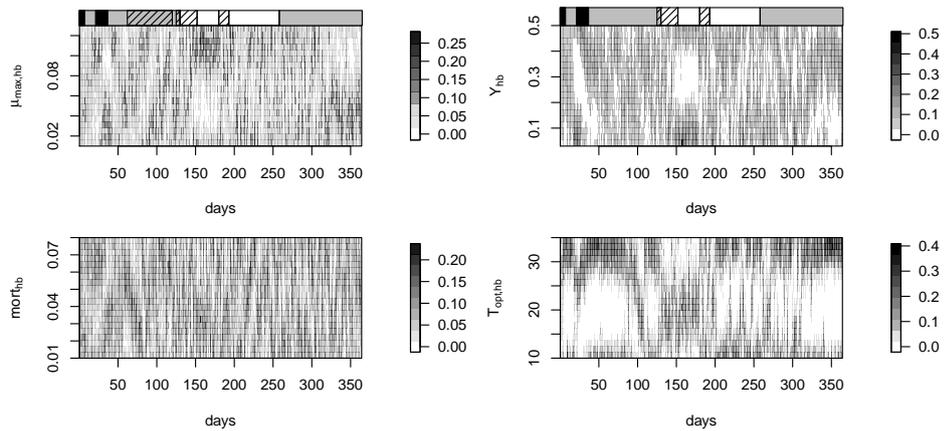


FIGURE B.4 – L'évolution des poids normalisés pour les paramètres bactériens avec la stratégie du domaine bouclé et deux espèces bactériennes. Voir la figure 6.8 pour la classification des polygones en haut et la table 5.1 pour les unités.

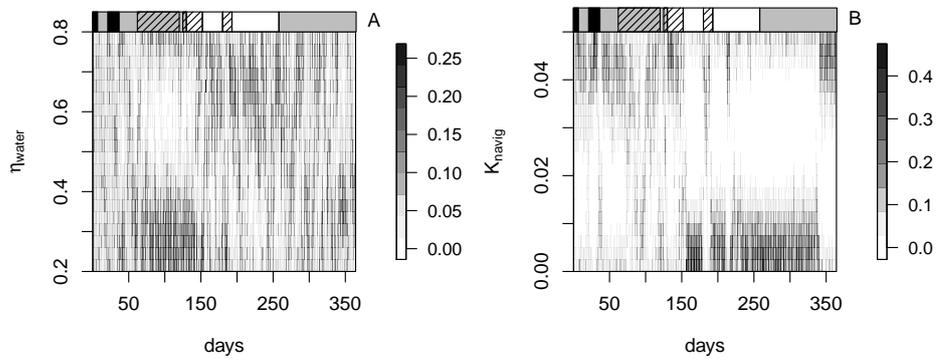


FIGURE B.5 – L'évolution des poids normalisés pour les paramètres physiques avec la stratégie du domaine bouclé et deux espèces bactériennes. Voir la figure 6.8 pour la classification des polygones en haut et la table 5.1 pour les unités.

Annexe C

Guide d'utilisation : ProSe-PA 0.30

Résumé

Cette notice présente la structure d'un fichier de commande permettant de réaliser une simulation PROSE0.30. Elle présente de plus la structure des fichiers d'entrée permettant de définir les géométries et les paramètres du modèle. Les aspects géomatiques permettant la construction des différents maillages ne sont pas détaillés.

C.1 Get Started

Cette partie décrit les étapes à suivre afin d'installer correctement le programme.

1. Installation GSL - GNU Scientific Library en mode développeur

Passer directement à l'étape 2 lors de la première installation. Revenir à cette section en cas de problème en 3.

Sous Ubuntu, l'utilisateur doit bénéficier des droits de superuser

```
> sudo apt-get install libgsl-dev
```

Le Makefile est configuré par défaut pour un système UBUNTU. Dans le cas d'un système REDHAT/FEDORA, voir les commentaires directement dans le Makefile concernant les variables d'environnement PREFIX_GSL et SUFFIX_GSL. Ultimately if nothing works, try to locate the static gsl library and its .h. Use the command `> locate libgsl` and `> locate gsl_matrix`

2. Création de "LIB_HYDROSYSTEM_PATH"

Dans un premier temps, à la racine de votre compte (tapez la commande "cd" dans un terminal), ouvrez le fichier .bashrc avec un éditeur de texte (par exemple `emacs .bashrc`) et ajoutez la ligne suivante : `export LIB_HYDROSYSTEM_PATH=$HOME/Programmes/LIBS`

Rendez ensuite les variables d'environnement effectives en tapant dans un terminal

```
source $HOME/.bashrc
```

Créez le dossiers suivants en tapant :

```
mkdir $HOME/Programmes
```

```
mkdir $HOME/Programmes/LIBS
```

3. Téléchargement et compilation de PROSE0.30 ainsi que de toutes les librairies :

Depuis \$HOME/Programmes, téléchargez prose avec la commande suivante :

```
>svn checkout http://svn.geosciences.fontainebleau.ensmp.fr/repos/prose
```

Se rendre ensuite dans `prose/trunk/src/` et entrez la commande suivante :

```
> ./make_pprose.sh all
```

En cas de problème de compilation relatif à `gsl` (`gsl/gsl_matrix.h not found`) retourner au point 1 de la procédure de démarrage.

Pour paralléliser plusieurs simulations en cas d'assimilation de données, il faut activer `-DCDA` dans le `Makefile` :

```
> OPTD = -DOMP $(OPTOMP) -DCOUPLED_RIVE -DCDA
```

Par défaut, la parallélisation de transport et biogéochimie est activé, mais pas pour l'assimilation de données :

```
> OPTD = -DOMP $(OPTOMP) -DCOUPLED_RIVE
```

PROSE0.30 est prêt à fonctionner par la commande :

```
> ./PROSE0.30 fichier_de_commande.COMM log_file.log
```

C.2 Définition des dossiers contenant les fichiers d'entrée du modèle

Un exemple de syntaxe du bloc de définition des dossiers d'entrée du modèle est présenté ci dessous. Cet exemple est tiré du fichier de commande de simulation du modèle Suresnes-Andrésy.

```
Input_folders = $HOME/Programmes/prose/trunk/test/DATA_BARRAGE/

Output_folder = $HOME/OUTPUT_PROSE_COUPLED/test_barrage_deux_biefs_steady

simulation =
{
  test_simulation #nom de la simulation

  time = {
    t_ini = [d] 0.
    t_end = [d] 10.
    dt = [min] 10.
    #beginning = 30 december 2011 00 00 #date when t = 0.
  }

  settings = {
    regime = steady #transient
    hyd = yes
    ttc = no
    rive = no
    solver = sp_prose #gc_prose
    #da = no
    #num_thread = 1
    #num_particules = 1
    #num_thread_par = 1
    #alpha_da = 0.3
    #seuil_chla = [ugchla/l] 0.
    #error_obs = 0.01
    #s_percent = 0.10
    #nb_comp_phy = 3
```

```
#dbo_oxy = [molO2/molC] 1.0

#species = { ... }
hydraulic = {
    ndim = 1

    calculate_curvature = no
    dx = [m] 500

    eps_Q = [m^3/s] 0.000001
    eps_Z = [m] 0.000001

    global_strickler = 20.
    dz = [m] 0.1
    upstream_Hmin = [m] 0.
    downstream_Hmax = [m] 20.
    theta = 0.9

    schem_type = ST_VENANT

    initialization = {
        init_Z_file = include seinetot_ini_Z
        init_Q_file = include seinetot_ini_Q
    }

network_description =
{

    singularities = { include singularities_desc }
    reaches = { include reaches_desc }
    profiles = { include profiles_desc }
    inflows = { include inflows_desc_hyd }
}
}
```

```
#transport = { theta_ttc = 1. }
#biology = { ... }
#param_range_da = { ... }
}
}

outputs =
{
  print_pk = { yes
              file_name = pks
            }

  final_state = { yes
                 file_name = seinetot_ini
                }

  #iterations = { yes }

  longitudinal_profiles = { ... }

  time_series = { ... }

  mass_balances = { ... }
  #mass_balances_bio = { ... }
}
```

Tout ce qui est précédé d'un caractère dièse correspond à des commentaires. Les chaînes de caractères précédées de `include` correspondent à des noms de fichiers, présents dans les répertoires définis comme `Input_folders`. Ces chemins définissent les dossiers contenant les fichiers d'entrée du modèle. "NPILE" chemins peuvent être définis. "NPILE" est fixé à 25. Il est préférable de définir le chemin complet vers le dossier contenant les fichiers d'entrée du modèle. Par la suite pour chaque fichier d'entrée défini dans le fichier de commande, l'existence du fichier d'entrée sera testé dans chaque dossier défini ici jusqu'à ce qu'il soit trouvé. En cas de non existence du fichier d'entrée dans un des dossiers définis dans ce bloc, un message d'er-

reur est écrit et l'exécution du programme est stoppée. La structure de ces fichiers est décrite plus loin. La syntaxe concernant les sorties de modèle est également développée dans la suite.

C.3 Paramétrisation générale

Cette partie définit les éléments nécessaires dans le fichier de commande pour simulation hydraulique ou complète ou assimilation de données.

Eléments à définir dans le fichier de commande				
nom	description	*	**	valeur par défaut
<i>Paramètres de temps</i>				
t_ini	instant initial		x	
t_end	instant final		x	
dt	pas de temps de calcul	x	x	
beginning	date correspondant à l'instant initial			
<i>Type de simulation</i>				
regime	type de simulation : permanent ou transitoire			STEADY
<i>Modules de simulation</i>				
hyd	Module hydraulique			YES
ttc	Module de transport			NO
rive	Module biogéochimique			NO
da	Assimilation de données (O_2)			NO
num_thread	Nombre de threads pour une simulation parallélisée de ttc et de rive			1
num_particules	Nombre de particules pour l'assimilation de données, N_{par}			1
num_thread_par	Nombre de threads pour paralléliser les particules en cas d'assimilation de données			1
<i>Biogéochimie</i>				

epsilon	Valeur seuil en-dessous de laquelle les quantités de matière en mmol et les masses en g sont considérées comme nulles	10^{-10}
nb_comp_phy	Nombre de compartiments constituant les espèces phytoplanctoniques (1 ou 3)	1
dbo_oxy	demande biologique en oxygène, taux de consommation lors de l'hydrolyse de la matière organique par les bactéries hétérotrophes et la production primaire	$1.2 \text{ molO}_2\text{molC}^{-1}$
<i>Assimilation de données</i>		
alpha_da	La taille effective minimum $\alpha \times N_{par}$	0.3
seuil_chla	Le seuil de la concentration en chl <i>a</i> pour déterminer si le système rivière est en cas de bloom algal	0. (pas de seuil)
error_obs	L'erreur associée aux observations	0.01
s_percent	La perturbation des paramètres (random walk)	0.1
solver	Solveur matricielle, libgc ou sparser	gc_prose

* Paramètre indispensable dans le cas d'une simulation en régime permanent

** Paramètre indispensable dans le cas d'une simulation transitoire

TABLE C.1: Éléments à définir dans le fichier de commande

C.4 Simulation hydraulique

C.4.1 Paramétrisation pour hydraulique

Éléments à définir dans le fichier de commande				
nom	description	*	**	valeur par défaut
<i>Tabulations verticales</i>				
dz	Pas des tabulations verticales			0.1 m
upstream_Hmin	Hauteur d'eau minimale sur le domaine			0.

<code>ndim</code>	nombre de dimensions	1
<code>downstream_Hmax</code>	Hauteur d'eau maximale sur le domaine	10. m
<i>Paramètres globaux</i>		
<code>curvature</code>	Valeur globale de la courbure	0.
<code>global_strickler</code>	Valeur globale du coefficient de Strickler	40.
<i>Options de calcul</i>		
<code>calculate_curvature</code>	calcul de la courbure par libhyd si les profils en travers sont géoréférencés	NO
<code>dx</code>	Pas de redécoupage longitudinal	aucun
<code>theta</code>	Degré d'implicite pour hydraulique	0.9
<code>eps_Q</code>	Précision de calcul des débits aux faces	$10^{-6} \text{ m}^3\text{s}^{-1}$
<code>eps_Z</code>	Précision de calcul des cotes aux centres	10^{-6} m
<code>schem_type</code>	Résolution des équations de Saint-venant	ST_VENANT

* Paramètre indispensable dans le cas d'une simulation en régime permanent

** Paramètre indispensable dans le cas d'une simulation transitoire

TABLE C.2: Eléments à définir dans le fichier de commande

C.4.2 Fichier d'initialisation de l'hydraulique

La partie `initialization` est facultative. Les fichiers `seinetot_init_Q.txt` et `seinetot_init_Z.txt` correspondent respectivement aux valeurs initiales de débit aux faces et de cotes aux centres des éléments. Si ces fichiers ne sont pas fournis par l'utilisateur. Des valeurs initiales sont calculées en fonction des conditions limites imposées. Pour initialiser l'hydraulique, l'utilisateur peut fournir deux fichier :

- un fichier d'initialisation des cotes au centre des éléments (`seinetot_init_Z.txt`), avec une première colonne contenant les numéros

- des éléments et une seconde les cotes ;
- un fichier d'initialisation des débits aux faces (`seinetot_init_Q.txt`), avec une première colonne contenant les numéros des faces et une seconde les débits ;

C.4.3 Description du réseau hydrographique

C.4.3.1 Définition des singularités

```
upstream_point;
```

```
middle_point;
```

```
downstream_dam
{
    pk = [km] 100.
    type = WATER_LEVEL
    Z(t) = { [d] [m]
            0.  0.599
            }
};
```

```
weir_and_gate_dam
{
    type = HYDWORK
    holler = 0.21
    hyd_structures = {
        weir = {
            mu = 0.6
            width = [m] 5.
            Zw(t) = { [d] [m]
                    0.  50.6
                    }
        }

        gate = {
            mu = 0.6
            width = [m] 5.
        }
    }
};
```

```

                                H(t) = { [d] [m]
                                           0.  0.2
                                           }
                                }
                                }
};

```

Les amonts, singularités «continues» et confluences ou diffluences sans ouvrages hydrauliques sont simplement nommés. Les autres singularités sont également décrites par leur type et une ou plusieurs loi de fonctionnement. Un pk doit être attribué à l'une des singularités avales. `downstream_dam` représente un barrage décrit par une loi de cote de la surface libre en fonction du temps. `weir_and_gate_dam` est un barrage constitué d'un seuil, représenté par sa cote de seuil en fonction du temps, et d'une vanne. `holler` est le coefficient de Holler pour la réoxygénation aux barrages.

C.4.3.2 Définition des biefs

Les biefs sont décrits par une singularité amont, une singularité aval et un numéro de branche. En effet, plusieurs biefs peuvent avoir le même amont et le même aval, dans le cas de confluences ou de diffluences. Le coefficient de Strickler peut être donné en fonction du débit en m^3s^{-1} . Cette fonction sera la même dans tout le bief. Si elle n'est pas définie ici, la valeur globale du coefficient de Strickler, définie dans le fichier de commande, ou, à défaut, égale à $40 \text{ m}^{\frac{1}{3}}\text{s}^{-1}$ sera adoptée pour le bief considéré.

```

upstream_point -> middle_point 1
{
    strickler = { 0. 40. }
};

middle_point -> downstream_dam 1;

```

C.4.3.3 Profils en travers

Les profils en travers sont décrits par leur nom (P1 et P2), la singularité amont (`upstream_point`) et le numéro de la branche à laquelle le profil appartient (1). `dx` correspond à la distance longitudinale jusqu'au prochain

profil en travers. La géométrie peut ensuite être définie de deux manières : soit avec le type `Absc Z` où `Absc` est l'abscisse transversale du point, soit avec le type `X Y Z` où les points sont données par leurs coordonnées géoréférencées. Les profils sont toujours décrits de la berge gauche vers la berge droite.

Si la courbure est connue par l'utilisateur en un point, celui-ci peut l'indiquer, comme fait pour `P2`.

```
P1 <- upstream_point 1
{
  dx = [km] 10.
  type = Absc Z
      0. 105.
      0. 100.
      10. 100.
      10. 105.
};
```

```
P2 <- upstream_point 1
{
  dx = [km] 10.
  curvature = [/m] 0.001
  type = X Y Z
      0. 0. 95.
      0. 0. 90.
      10. 0. 90.
      10. 0. 95.
};
```

C.4.4 Trajectoire du centre des faces

Si la bathymétrie du cours d'eau n'est pas connue, l'utilisateur peut fournir la trajectoire décrite par les milieux des profils en travers du cours d'eau. Le programme crée alors automatiquement les faces et les éléments nécessaires. Ceux-ci ont une géométrie rectangulaire. Dans le fichier de commande, la description des profils en travers, appelée comme suit :

```
profiles = { include profiles_desc }
```

peut être remplacée par une suite de points définis, dans l'ordre, par leurs coordonnées, leur cote de fond, la largeur du profil souhaité en ce point et la courbure :

```
trajectory = {
    [m] [km] [m] [m] [1/m]
    0.  0.  100. 10.  0.
    0.  10.  90.  10.  0.
    0.  20.  80.  10.  0.
    0.  30.  70.  10.  0.
}
```

La commande précédent va par exemple provoquer la création d'un bief rectiligne de 30 kilomètres, constitué de 4 faces de largeur 10 mètres. Pour que la courbure donnée soit prise en compte, l'utilisateur doit fixer la valeur de `calculate_curvature` dans le fichier de commande à `GIVEN`. La valeur `YES` induirait le calcul de la courbure en chaque face à partir de ses coordonnées.

C.4.5 Apports

```
CANAL_TEST : upstream_point 1 {
type = UPSTREAM_INFLOW
q = { [d] [m^3/s] 0. 5 }
o2 1 = { [d] [mgO2/l] 0. 12. }
...
};
```

```
AFFLUENT : middle_point 1 {
type = INFLUENT
x = [km] 1.
y = 1.
q = { [d] [m^3/s] 0. 0.5 }
o2 1 = { [d] [mgO2/l] 0. 12. }
phy 1 = { [d] [mgC/l] 0. 1.5 }
phy 2 = { [d] [mgC/l] 0. 1.5 }
...
};
```

Les apports doivent être décrits par leur nom (ici, `CANAL_TEST` et `AFFLUENT`), la singularité amont (`upstream_point` et `middle_point`) et le numéro de branche. Ils peuvent être de type `UPSTREAM_INFLOW`, `INFLUENT` ou `EFFLUENT`. `x` et `y` correspondent respectivement à la distance depuis la singularité amont et la position transversal de l'apport (comprise entre 0 et 1, 0 correspondant à la rive gauche et 1 la rive droite). Pour les espèces biogéochimiques, il faut indiquer le nom de la variable (`O2` ou `PHY` par exemple) et le numéro sous-espèce (1, 2).

C.4.6 Prise en compte de la courbure

L'utilisateur a trois options de prise en compte de la courbure, définies par la valeur de `calculate_curvature` dans le fichier de commande :

- `NO` : la courbure est considérée nulle en tout point du réseau hydrographique, on est dans le cas des équations de Saint-Venant rectilignes à une dimension ;
- `YES` : la courbure est calculée par le programme au niveau des points géoréférencés ; la valeur globale, définie dans le fichier de commande, ou la valeur ponctuelle, définie dans le fichier de description des profils en travers est employée aux autres points ;
- `GIVEN` : aucune courbure n'est calculée par le programme ; les valeurs ponctuelles données par l'utilisateur sont employées, ou, à défaut la valeur globale définie dans le fichier de commande.

C.5 Simulation de transport et biogéochimique

C.5.1 Les fichiers de paramètres des espèces

Les fichiers de description des espèces particulières doivent être lus avant ceux décrivant les espèces dissoutes. Le mot-clé pour cette partie est `species = { ... }`.

Les espèces peuvent avoir des comportements différents dans les différentes sortes de compartiments (eau, vase et périphyton). Si les paramètres biogéochimiques ne sont pas définis dans certains types de compartiments, les valeurs utilisées dans ces types de compartiments seront les premières

code	nom de l'espèce
PHY	phytoplancton
ZOO	zooplancton
BACT	bactérie hétérotrophe
BACTN	bactérie nitrifiante
BIF	bactérie indicatrice de contamination fécale
MES	matière en suspension minérale
MOP	matière organique particulaire
MOD	matière organique dissoute
SI_D	silice dissoute
NH4	ammonium
NO2	nitrite
NO3	nitrate
PO4	orthophosphate
O2	dioxygène
N2O	oxyde nitreux

TABLE C.3 – Codes des espèces biogéochimique simulées

définies par l'utilisateur. Par exemple, si l'utilisateur donne d'abord les paramètres dans l'eau puis dans les vases mais pas dans le périphyton, si une couche de périphyton est simulée, alors les espèces auront le même comportement que dans l'eau.

Ci-dessous, un exemple de fichier de description des espèces phytoplanctoniques. Un fichier de description d'une espèce doit commencer par le code de l'espèce (cf. Table C.3), puis sont données les descriptions de une plusieurs sous-espèces (`nom_sous_espece = { ... }`).

Pour chaque sous-espèce, différents paramètres peuvent être définis selon leur type : paramètres des espèces particulières (masse volumique et porosité), des espèces vivantes (température optimale), composition biochimique (ratios C/...), constantes de demi-saturation relatives à certaines substances (`kmich ...`). Les paramètres relatifs aux différents processus associés à la sous espèce peuvent ensuite être fixés. Dans le cas suivant, les processus associés à la sous-espèce phytoplanctonique `chloro` sont la photosynthèse, la croissance, la respiration, l'excrétion et la mortalité. L'advection-dispersion pour transport est également définie aussi pour chaque espèce.

PHY

`chloro =`

```
{
  transport_type = SOLUTE # pour transport
  media_type = FREE_WATER # pour transport
  advection = yes # pour transport
  diffusion = no # pour transport

  rho = [g/cm^3] 1.2
  phi = 0.9

  Topt = [°C] 30.
  sigma = [°C] 15.

  kmich phys = 0.06
  kmich N = [ugN/l] 70.
  kmich P = [ugP/l] 46.
  kmich Si = [mgSiO2/l] 0.
  kmich NH4 = [umol/l] 5.
  C/ P = [mC/mgP] 40.
  C/ N = [mgC/mgN] 7.
  C/ SI = [mC/mgSI] 0.
  C/ chla = [mC/ugchla] 0.035

  photosynthesis = { eta = [/m] 0.32
                    eta_mes = [l/mg/m] 0.04
                    eta_chla = [l/ugchla/m] 0.02
                    alpha = [m^2s/uE/h] 0.0012
                    beta = [m^2s/uE/h] 0.
                    pmax = [/h] 0.5
                  }

  growth = { sr = [/h] 0.375
            cr = [/h] 1.
            mumax = [/h] 0.125
          }

  excretion = { excr_cst = [/h] 0.001
```

```
        excr_phot = 0.0006
    }

    respiration = { maint = [/h] 0.024
                   energ = 0.5
    }

    mortality = { mort = [/h] 0.004
                  phi_lim_mort = [mgC/l] 2.1
                  delta = 1.
    }
}
```

Plusieurs sous-espèces peuvent être définies pour une même espèce, comme pour la matière organique particulaire ci-dessous.

MOP

```
mop1 =
{
    transport_type = SOLUTE # pour transport
    media_type = FREE_WATER # pour transport
    advection = yes # pour transport
    diffusion = no # pour transport

    rho = [g/cm^3] 1.2
    phi = 0.9

    compose_mo = 0.3

    kmich mod = [mgC/l] 0.25
    C/ P = [mgC/mgP] 40.
    C/ N = [mgC/mgN] 7.

    hydrolysis = { kc_hydr = [/h] 0.05
    }
}
```

```
mop2 =
{
  rho = [g/cm^3] 1.2
  phi = 0.9

  transport_type = SOLUTE
  media_type = FREE_WATER
  advection = yes
  diffusion = no

  compose_mo = 0.15

  kmich mod = [mgC/l] 0.25
  C/ P = [mgC/mgP] 40.
  C/ N = [mgC/mgN] 7.

  hydrolysis = { kc_hydr = [/h] 0.000225
                }
}

mop3 =
{
  rho = [g/cm^3] 1.2
  phi = 0.9

  transport_type = SOLUTE # pour transport
  media_type = FREE_WATER # pour transport
  advection = yes # pour transport
  diffusion = no # pour transport

  compose_mo = 0.05

  kmich mod = [mgC/l] 0.25
  C/ P = [mgC/mgP] 40.
  C/ N = [mgC/mgN] 7.
```

```

    hydrolysis = { kc_hydr = [/h] 0.
                  }
}

```

NH4

```

ammonium =
{
  ds = [m/s] 0.0015

  transport_type = SOLUTE # pour transport
  media_type = FREE_WATER # pour transport
  advection = yes # pour transport
  diffusion = no # pour transport
  diff_mol = [m^2/s] 0.1 # pour transport
}

```

Le processus d'adsorption associé à une espèce biogéochimique dissoute doit être décrit comme dans l'exemple ci-dessous. En plus des paramètres d'équilibre, une liste d'espèces particulières sur lesquelles l'espèce dissoute peut s'adsorber doit être fournie (`code_espece num_espece`).

PO4

```

phosphate =
{
  ads_desorption = { Kps = [mgP/l] 0.7
                    Pac = [mg/mg] 0.178
                    damping = [s] 600
                    adsorbs_on = { mop 1 mop 2 mop 3 }
                    }
}

```

Des réactions chimiques peuvent également être décrites. Une sous-espèce peut intervenir dans plusieurs réactions (`reaction = {...}`). Une réaction est décrite par un coefficient de dégradation d'ordre 1 ainsi que les autres réacteurs et produits avec les coefficients de stoechiométrie correspondants.

NO2 =

```
nitrite =
{
  reactions =
  {
    reaction = { decay = [/h] 0.005
                 other_reactors = { oxygene 0.5 }
                 products = { nitrate 1. }
               }
  }
}
```

L'exemple précédent traduit la réaction chimique : $NO_2 + \frac{1}{2}O_2 \rightarrow NO_3$. Les concentrations en nitrite, dioxygène et nitrate sont alors calculées de la manière suivante :

$$\begin{aligned}\frac{dC_{NO_2}}{dt} &= -\text{decay} \cdot C_{NO_2} \\ \frac{dC_{O_2}}{dt} &= 0.5 \cdot \text{decay} \cdot C_{NO_2} \\ \frac{dC_{NO_3}}{dt} &= \text{decay} \cdot C_{NO_2}\end{aligned}$$

où C_{NO_2} , C_{O_2} et C_{NO_3} sont en $\text{mmol}\cdot\text{m}^{-3}$.

La liste des paramètres pouvant être définis pour chaque processus est donnée dans la Table C.4.

Paramètres des espèces biogéochimiques			
nom	description	unité	valeur par défaut
<i>Paramètres des espèces particulières</i>			
rho	masse volumique	[ML ⁻³]	1000 gm ⁻³
phi	porosité		0.
Vsed	vitesse de sédimentation	[LT ⁻¹]	
scouring	paramètre d'arrachage phytoplanctonique		
<i>Paramètres des espèces dissoutes</i>			
Ds	coefficient de diffusion turbulente eau-sédiments	[L ² T ⁻¹]	

Paramètres des espèces vivantes

Topt	température optimale	°C	20 °C
sigma	écart-type de la fonction de température	°C	1 °C
prod_N2O	taux de production d'oxyde nitreux lors de la dénitrification	[T ⁻¹]	0.

Paramètres des espèces minérales

compose_mo	part de matière organique en laquelle sont dégradées les espèces vivantes lors de la mortalité		
------------	--	--	--

*Paramètres des espèces gazeuses**Paramètres de photosynthèse (photosynthesis)*

eta	terme de base de l'extinction lumineuse	[L ⁻¹]	
eta_chla	terme de l'extinction lumineuse fonction de la biomasse phytoplanctonique	[L ² M ⁻¹]	
eta_mes	terme de l'extinction lumineuse fonction de la matière en suspension	[L ² M ⁻¹]	
alpha	capacité photosynthétique		
beta	capacité de photoinhibition		
pmax	taux de production photosynthétique maximal		

Paramètres de croissance (growth)

mumax	taux de croissance maximale		
yield	rendement de croissance		
phy_lim_gr	concentration critique au-dessus de laquelle la croissance est freinée		
sr	taux de synthèse des produits de réserve phytoplanctoniques		
cr	taux de catabolyse des produits de réserve		
yield_nit	rendement de nitrification		

<code>kmich_no2</code>	constante de demi-saturation vis-à-vis des nitrites lors de la production d'oxyde nitreux
<code>cmin_o2</code>	concentration en oxygène dissous en-dessous de laquelle peut être produit de l'oxyde nitreux lors de la nitrification
<code>stoechio_nit</code>	stoechiométrie de la consommation en dioxygène lors de la nitrification
<code>fbis</code>	paramètre de correction de la croissance dans le cas d'une production d'oxyde nitreux

*Paramètres de respiration (**respiration**)*

<code>maint</code>	taux de respiration lié à la survie de la cellule
<code>energ</code>	taux de respiration liée à la dépense énergétique de croissance

*Paramètres de broutage (**grazing**)*

<code>graz</code>	taux de broutage
-------------------	------------------

*Paramètres d'excrétion (**excretion**)*

<code>excr_cst</code>	taux d'excrétion de base
<code>excr_phot</code>	taux d'excrétion liée à la photosynthèse

*Paramètres de mortalité (**mortality**)*

<code>mort</code>	taux de mortalité
<code>phi_lim_mort</code>	concentration critique au-dessus de laquelle la mortalité est accélérée
<code>delta</code>	facteur multiplicatif de la mortalité pour une concentration supérieure à <code>phi_lim_mort</code>

*Paramètres de réaction chimique (**reaction**)*

<code>decay</code>	taux de décroissance radioactive (ordre 1)
--------------------	--

	<i>Paramètres d'hydrolyse (hydrolysis)</i>
kc_hydr	constante de demi-saturation pour l'hydrolyse de la matière organique
	<i>Paramètres d'adsorption (ads_desorption)</i>
Kps	
Pac	Capacité d'adsorption maximale
damping	Coefficient d'amortissement de l'adsorption
a_Fr	paramètre dans le formalisme d'adsorption de Freundlich
b_Fr	paramètre dans le formalisme d'adsorption de Freundlich
Tc_Fr	temps de contact dans le formalisme d'adsorption de Freundlich
	<i>Paramètres de réaération (reaeration)</i>
rea_navig	coefficient de réaération due à la navigation
rea_wind	coefficient de réaération due au vent
D_Richey	coefficient de diffusion moléculaire à l'interface eau-air
	<i>Paramètres de transport (Advection-dispersion)</i>
transport_type	Le type de substance chaleur ou chimique (HEAT ou SOLUTE), pour l'instant pas de température
media_type	FREE_WATER pour la rivière
advection	le processus d'advection (YES ou NO)
diffusion	la diffusion moléculaire (YES ou NO)
diff_mol	le coefficient de diffusion moléculaire (YES [L ² T ⁻¹] ou NO)

TABLE C.4: Eléments à définir dans le fichier de commande

C.5.2 Paramétrisation pour transport

La paramétrisation de transport commence par `transport = { theta_ttc = 1. }` Un seul paramètre à définir est le `theta_ttc` pour le schéma numé-

rique. Pour l'instant $\theta_{ttc} = 1$ est accepté (implicite).

C.5.3 Paramétrisation pour biogéochimie

La paramétrisation de simulation biogéochimique commence par `biology = { ... }`. Trois paramètres généraux sont définie d'abord (`numerical_method`, `max_div_dt` et `dz_rive`).

Eléments à définir dans le fichier de commande			
nom	description	*	valeur par défaut
<i>Options de calcul</i>			
<code>numerical_method</code>	Méthode numérique de résolution des processus biogéochimiques (RUNGE_KUTTA ou EXPLICIT)		RUNGE_KUTTA
<code>max_div_dt</code>	Facteur de redécoupage maximal du pas de temps de calcul		1.0
<code>dz_rive</code>	Pas de discrétisation pour le calcul de la pénétration des radiations dans la colonne d'eau		0.1 m

* Paramètre indispensable

TABLE C.5: Eléments à définir dans le fichier de commande

C.5.4 Les fichiers de description des compartiments

Un fichier de description de compartiments doit débiter par le code du type de compartiments qui sera décrit (`WATER`, `VASE` ou `PERIPHYTON`). Les compartiments de ce type sont ensuite décrits du haut vers le bas de la section.

La description d'un compartiment (`nom_compartiment = { ... }`) comprend son état initial (masse, porosité, masse volumique), les concentrations initiales des espèces biogéochimiques (une concentration non initialisée sera considérée comme nulle, sauf pour l'oxygène dans l'eau qui sera initialisée à la saturation). Pour définir les concentrations initiales des espèces biogéochimiques (y compris les masses, les volumes), il est possible :

- Soit un fichier d'initialisation des concentrations des éléments, avec une première colonne contenant les numéros des éléments (commence

C.5.5 Les paramètres liés à la sédimentation-érosion

Les processus d'érosion et de sédimentation sont simulés simultanément. Cette partie commence par le mot clé `exchanges = { ... }`.

```
exchanges = {
    calc_sedim_eros = SIMULTANEOUS_PROSE
    eta_hyd = [g/cm^3] 0.003
    pnavig = [g/m^2/s] 0.003
}
```

C.5.6 Les forçages météorologiques

Les forçages météorologiques nécessaires au calcul de la biogéochimie sont : la température, le rayonnement solaire et le vent. Si aucune chronique n'est disponible, l'utilisateur peut donner les paramètres `mean`, `amplitude`, `delay` et `attenuation` correspondant à la température (T), au rayonnement solaire (I) et à la photopériode (λ). La température et le rayonnement solaire sont alors calculés de la manière suivante :

$$T(t_j) = \text{mean}_T - \text{amplitude}_T \cdot \cos\left(\frac{2\pi \cdot (t_j - \text{delay}_T)}{365.}\right) \quad (\text{C.1})$$

$$I(t_j) = \frac{\pi}{2} \text{attenuation}_I \left(\text{mean}_I - \text{amplitude}_I \cdot \cos\left(\frac{2\pi \cdot t_j}{365.}\right) \right) \cdot \sin\left(\frac{\pi * (t_h - 1)}{\lambda(t_j)}\right) \quad (\text{C.2})$$

$$\lambda(t_j) = \text{mean}_\lambda - \text{amplitude}_\lambda \cdot \cos\left(\frac{2\pi \cdot t_j}{365.}\right) \quad (\text{C.3})$$

t_j et t_h correspondent respectivement à la date en jours et en heures.

La syntaxe des forçages météorologiques dans les fichiers d'entrée est la suivante. En l'absence de chronique de vent, celui-ci est considéré comme nul.

```
temperature = { mean = [°C] 20.
                amplitude = [°C] 1.
                delay = [h] 5.
            }
```

```
wind = { [d] [m/s]
         0.5 1.
         1. 5.
         2. 6.
       }

radiation = { [d] [J/cm^2/h]
              0. 100.
            }

photoperiod = { mean = [h] 12.
                amplitude = [h] 3.
              }
```

C.5.7 Définition des intervalle des paramètres à assimiler

Les intervalles des paramètres à assimiler sont définies dans la partie `biology = { ... }`. Le syntaxe des paramètres dans le fichier de commande est la suivante. Il est indispensable pour l'assimilation de données.

```
param_range_da = { include param_range }
```

Le fichier `param_range` contient 4 colonnes : `nom du paramètre` `unité` `min_val` `max_val`. Un exemple est donné ci-dessous. Les intervalles ci-dessous sont issues des travaux de l'analyse de sensibilité de module `librive` ([Wang et al., 2018b](#)).

```
Rm_phy [/h] 0.001 0.021
a_phy [m^2s/uE/h] 0.0003 0.0018
pmax_phy [/h] 0.09 0.546
eta_chla_phy [l/ugchla/m] 0.006 0.054
c_chla_phy [mgC/ugchla] 0.02 0.13
eta_water [/m] 0.2 0.8
Topt_phy [°C] 10. 37.
mumax_bact [/h] 0.01 0.13
yield_bact 0.03 0.5
mort_bact [/h] 0.01 0.08
```

```
Topt_bact [°C] 10 35
Krea_navig [m/h] 0. 0.05
```

C.6 Définition des fichier de sorties

Le syntaxe des sorties dans le fichier de commande est montré ci-dessous. Plusieurs types de sorties sont décrits plus loin.

```
outputs =
{
  print_pk = { yes
              file_name = pks }
  final_state = { yes
                 file_name = seinetot_ini }
  #iterations = { yes }
  #mb_elements = { yes }
  time_series =
  {
    time_unit = [min]

    var = { Zfs    [m]
           hh     [m]
           Surf   [m^2]
           Peri   [m]
           Rh     [m]
           Vel    [m/s]
           Disch  [m^3/s]
           phy    [mgC/1]
           o2     [mgO2/1]
           ...
          }

  points = { SEINE
            [km] 715. 1 # Sartrouville
            [km] 706.3 1 # Aval Bougival
            [km] 702.4 2 #Aval Chatou
            [km] 730.2 1 # Aval Andrésey
```

```
        [km] 674.6  1 # Aval Suresnes 1
    }
points_obs = { SEINE
    [km] 715 1 o2 1 [s] [mgO2/l] include conc_init/o2_sartrouville_pprose_dia_0.01_t
    [km] 730.2 1 o2 1 [s] [mgO2/l] include conc_init/o2_andresy_pprose_dia_0.01_t
    [km] 706.3 1 o2 1 [s] [mgO2/l] include conc_init/o2_bougival_pprose_dia_0.01_t
    [km] 702.4 2 o2 1 [s] [mgO2/l] include conc_init/o2_chatou_pprose_dia_0.01_t
    [km] 674.6  1 o2 1 [s] [mgO2/l] include conc_init/o2_suresnes_pprose_dia_0.01_t
}

graphics = GNUPLOT
}

longitudinal_profiles =
{
    time = { t_ini = [s] 0.0
            t_end = [d] 0.1
            dt = [min] 1.
        }

    time_unit = [min]

    var = { Zfs    [m]
            H      [m]
            Surf   [m^2]
            Peri   [m]
            Rh     [m]
            Vel    [m/s]
            Disch  [m^3/s]
        }

    extent = { canal_test
                [km] 0.  1
                [km] 75. 1
            }
    extent = { canal_test
```

```
        [km] 75. 1
        [km] 100. 1
    }

    graphics = GNUPLOT
}

mass_balances =
{
    time = { t_ini = [s] 0.0
            t_end = [d] 0.1
            dt = [min] 1.
            }

    time_unit = [min]

    extent = { canal_test
              [km] 0. 1
              [km] 75. 1
              }
    extent = { canal_test
              [km] 75. 1
              [km] 100. 1
              }

    graphics = GNUPLOT
}

mass_balances_bio = {

    mb1 = { t_ini = [d] 1461.
           t_end = [d] 1491.
           n_steps = 1. #dt_mb = n_steps * dt_Simul

    species = { phy [mol]
               o2 [mol]
               #nh4 [mol]
               #no3 [mol]
```

```

        #no2 [mol]
        #bact [mol]
        #bactn [mol]
        #po4 [mol]
        #mod [mol]
        #mop [mol]
        #mes [kg]
    }
    #annex_var = { #carbone [mgC]
        #ntot [mgN]
        #ptot [mgP]
        #phyf [mol]
        #phyr [mol]
        #phys [mol]
        #chla [ugchla]
    #}
    }
}

```

Un fichier `mass_balance.txt` est créé dans le répertoire de résultats. Il donne le bilan de masse hydraulique sur l'ensemble du domaine simulé à chaque itération.

C.6.1 Sorties PKs, final states

L'utilisateur peut demander l'impression des pks des singularités et apports (sortie `print_pk`).

Dans le cas d'une simulation permanente, il peut demander un fichier d'itérations (sortie `iterations`) ainsi que deux fichiers contenant l'état hydraulique (cote de la surface libre aux centres des éléments et débit aux faces) après convergence (sortie `final_state`), pouvant par la suite être employé pour initialiser une autre simulation. Le fichier `iterations` donne les valeurs calculées des variables hydrauliques en toutes les faces et en tous les centres d'éléments à chaque itération. Attention, ce fichier peut devenir très volumineux dans le cas de simulations avec beaucoup de mailles et/ou si la convergence demande beaucoup d'itérations.

Pour une simulation permanente prenant compte le transport (biogéochimie), un fichier `conc_init` est créé dans le répertoire de résultats. Ce fichier sert à créer les fichiers d'initialisation des concentrations des éléments pour initialiser une autre simulation transitoire.

Pour une simulation transitoire prenant compte le calcul biogéochimique, `final_state` donne l'état hydraulique finale (`seinetot_ini_Q` et `seinetot_ini_Z`) et les concentrations finales dans l'eau et le sédiment (`conc_final_water` et `conc_final_vase`), les volumes finaux du sédiment (`volume_final_vase`) dans le répertoire de résultats. Ces fichiers finals peuvent être utilisés pour initialiser une autre simulation transitoire.

Le bilan de masse hydraulique par élément sur l'ensemble du domaine simulé est possible de créer par `mb_elements = { yes }`. Attention, ce fichier peut devenir très volumineux dans le cas de simulations avec beaucoup de mailles en régime transitoire.

Pour ces quatre derniers types de sorties, un nom de fichier peut être défini par l'utilisateur, mais il n'est pas possible de le définir pour les concentrations finales et les volumes finaux.

C.6.2 Profils ponctuels

Un ensemble de profils ponctuels (`time_series`) est défini par :

- son instant initial (`t_ini`), son instant final (`t_end`) , son pas de temps (`dt`), par défaut égaux à ceux de la simulation,
- une unité de temps (`time_unit`) souhaitée pour les sorties (par défaut la seconde),
- les variables souhaitées (`var = { ... }`) ainsi que leur unité,
- les PKs auxquels on souhaite la chronique (`points = { nom_rivière unité_pk PK numéro_de_branche ...}`).
- éventuellement les PKs auxquels on a les observations temporelles pour l'assimilation de données (`points_obs = { nom_rivière unité_pk PK numéro_de_branche la_variable_à_assimiler numéro_sous-espèce unité_temps unité_mesure temps val_mesure ...}`)

Plusieurs ensembles peuvent être définis dans le même fichier de commande.

C.6.3 Profils en long hydraulique

Un ensemble de profils en long pour hydraulique (`longitudinal_profiles`) est défini par :

- son instant initial (`t_ini`), son instant final (`t_end`) , son pas de temps (`dt`), par défaut égaux à ceux de la simulation,
- une unité de temps (`time_unit`) souhaitée pour les sorties (par défaut la seconde),
- les variables souhaitées (`var = { ... }`) ainsi que leur unité,
- les linéaires sur lesquels on souhaite le profil (`extent = { ...}`) — plusieurs linéaires peuvent être demandés pour un même ensemble de profils.

Plusieurs ensembles peuvent être définis dans le même fichier de commande.

C.6.4 Bilans hydraulique

Un ensemble de bilans (`mass_balances`) est défini par :

- son instant initial (`t_ini`), son instant final (`t_end`) , son pas de temps (`dt`), par défaut égaux à ceux de la simulation, (peut-être un bug pour un pas de temps différent de ceux de la simulation)
- une unité de temps (`time_unit`) souhaitée pour les sorties (par défaut la seconde),
- les tronçons sur lesquels on souhaite le bilan (`extent = { ...}`) — plusieurs linéaires peuvent être demandés pour un même ensemble de bilans.

Plusieurs ensembles peuvent être définis dans le même fichier de commande.

C.6.5 Sorties graphiques hydraulique

Des sorties graphiques hydraulique peuvent être demandées en ajoutant la ligne `graphics = GNUPLOT`. Un répertoire `gnuplot` est alors créé dans le répertoire de sorties. Ce nouveau répertoire contient les fichiers `gnuplot` permettant de tracer les profils en long et profils ponctuels demandés par l'utilisateur. Ils peuvent être utilisés avec la commande `gnu`.

C.6.6 Bilans biogéochimiques

Un ensemble de bilans biogéochimiques est défini par :

— son instant initial (`t_ini`), son instant final (`t_end`) , le pas de temps (`dt`). Le pas de temps `dt` est défini par une facteur de multiplication `n_steps = 1` (`dt = n_steps * dt_simul`).

— les variables souhaitées (`species = { ... }`) ainsi que leur unité

Pour l'instant, un bilan sur l'ensemble du domaine simulé est possible.

C.7 Fichier de commande pour cas test barrage

Exemple de fichier de commande en régime transitoire pour le cas test barrage.

```
Input_folders = $HOME/Programmes/prose/trunk/test/DATA_BARRAGE
               = $HOME/OUTPUT_PROSE_COUPLED/test_barrage_deux_biefs_steady
Output_folder = $HOME/OUTPUT_PROSE_COUPLED/test_barrage_deux_biefs_transient

simulation =
{
test_simulation

time = {
  t_ini = [d] 0.
  t_end = [d] 5.
  dt = [min] 10.
# beginning = 30 december 2011 00 00 #date when t = 0.
  }

settings = {

  regime = transient #steady
  hyd = no
  ttc = no
  rive = no
  solver = sp_prose #gc_prose #

hydraulic = {
  ndim = 1
  calculate_curvature = no
  dx = [m] 500
  eps_Q = [m^3/s] 0.000001
  eps_Z = [m] 0.000001
  global_strickler = 20.
  dz = [m] 0.1
  upstream_Hmin = [m] 0.
```

```
downstream_Hmax = [m] 20.
theta = 0.9
schem_type = ST_VENANT

initialization = {
    init_Z_file = include seinetot_ini_Z
    init_Q_file = include seinetot_ini_Q
}

network_description =
{
    singularities = { include singularities_desc }
    reaches = { include reaches_desc }
    profiles = { include profiles_desc }
    inflows = { include inflows_desc_hyd }
}
}
}

outputs =
{
print_pk = { yes
    file_name = pks }

final_state = { yes
    file_name = seinetot_ini }

#mb_elements = { yes }

longitudinal_profiles =
{
time = { t_ini = [d] 0.0
    t_end = [d] 5
    dt = [h] 12.
```

```
}
time_unit = [s]

var = { Zfs [m]
        hh [m]
        Surf [m^2]
        Peri [m]
        Rh [m]
        Vel [m/s]
        Disch [m^3/s]
}

extent = { canal_test1
          [km] 160.0 1
          [km] 200 1
}

graphics = GNUPLOT
}

time_series =
{
time = { dt = [min] 10. }
        time_unit = [s]
        var = { hh [m]
                zfs [m]
                surf [m^2]
                Disch [m^3/s]
                Vel [m/s]
        }
}

points = { CANAL_TEST1
          [km] 170. 1
          [km] 160. 1
          [km] 180. 1
          [km] 190. 1
```

```
    }  
}  
  
mass_balances = { time = { dt = [min] 10. }  
                  time_unit = [min]  
  
                  extent = { CANAL_TEST1  
                             [km] 180 1  
                             [km] 200 1  
                          }  
                  extent = { CANAL_TEST1  
                             [km] 160 1  
                             [km] 180 1  
                          }  
  
                  #graphics = GNUPLOT  
                  }  
  
}
```

Bibliographie

- Abbaszadeh, P., Moradkhani, H., Yan, H., 2018. Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo. *Advances in Water Resources* 111, 192–204.
- Ambrose, B., Avant, B., Han, Y., Knightes, C., Wool, T. A., 2017. Water quality assessment simulation program (WASP8) : Upgrades to the advanced toxicant module for simulating dissolved chemicals, nanomaterials, and solids. Tech. rep., USEPA, Washington, DC.
- Ambrose, B., Wool, T. A., 2009. WASP7 stream transport - model theory and user's guide. Tech. rep., USEPA, Washington, DC.
- Andreadis, K., Clark, E., Lettermaier, D., Alsdorf, D., 2007. Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters* 34, L10403.
- Annan, J., Hargreaves, J., Edwards, N., Marsh, R., 2005. Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Modelling* 8 (1), 135 – 154.
- Arhonditsis, G., Brett, M., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. *Marine Ecology Progress Series* 271, 13–26.
- Arnold, J. G., Allen, P. M., Bernhardt, G., 1993. A comprehensive surface-groundwater flow model. *Journal of Hydrology* 142 (1), 47 – 69.
- Arulampalam, M. S., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* 50 (2), 174–188.

- Bailey, R. T., Ahmadi, M., 2014. Spatial and temporal variability of in-stream water quality parameter influence on dissolved oxygen and nitrate within a regional stream network. *Ecological Modelling* 277, 87 – 96.
- Barker, D. M., Huang, W., Guo, Y.-R., Bourgeois, A. J., Xiao, Q. N., 2004. A three-dimensional variational data assimilation system for mm5 : Implementation and initial results. *Monthly Weather Review* 132 (4), 897–914.
- Barnwell, T., Brown, L., Whittemore, R., 1987. Systems Analysis in Water Quality Management. Pergamon Press New York, Ch. QUAL2E - A Case Study in Water Quality Modelling Software, pp. 377–387.
- Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* 53, 370–418.
- Beck, M., 1987. Water quality modelling : a review of the analysis of uncertainty. *Water Resources Research* 23 (8), 1393–1442.
- Benner, R., Opsahl, S., Chin-Leo, G., Richey, J. E., Forsberg, B. R., 1995. Bacterial carbon metabolism in the amazon river system. *Limnology and Oceanography* 40 (7), 1262–1270.
- Bennett, A. F., 2002. Inverse modeling of the ocean and atmosphere. Cambridge Univeristy Press, Cambridge, UK.
- Beven, K., 1989. Changing ideas in hydrology. The case of physically-based model. *Journal of Hydrology* 105, 157–172.
- Bierman, V. J., Hinz, S. C., Zhu, D.-W., Wiseman, W. J., Rabalais, N. N., Turner, R. E., 1994. A preliminary mass balance model of primary productivity and dissolved oxygen in the mississippi river plume/inner gulf shelf region. *Estuaries* 17 (4), 886–899.
- Billen, G., Garnier, J., Ficht, A., Cun, C., 2001. Modeling the response of water quality in the Seine River estuary to human activity in its watershed over the last 50 years. *Estuaries* 24 (6B), 977–993.
- 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, 119–137.

- Billen, G., Garnier, J., Mouchel, J.-M., Silvestre, M., 2007. The Seine system : Introduction to a multidisciplinary approach of the functioning of a regional river system. *Sciences of Total Environment* 375, 1–12.
- Billen, G., Servais, P., 1989. Modélisation des processus de dégradation bactérienne de la matière organique en milieu aquatique. In : et al., B. (Ed.), *Micro-organismes dans les écosystèmes océaniques*. Masson Paris, pp. 219–245.
- Billen, G., Servais, P., 1991. Modélisation du transport de polluants par l'estuaire de l'escault. cas du phosphore. Tech. rep., Groupe de Microbiologie des Milieux Aquatiques. Université Libre de Bruxelles, Belgique.
- Bowden, K., Brown, S. R., 1984. Relating effluent control parameters to river quality objectives using a generalised catchment simulation model. *Water Science and Technology* 16 (5–7), 197–206.
- Boët, P., Belliard, J., dit Thomas, R. B., Tales, E., 1999. Multiple human impacts by city of paris on fish communities in the Seine river basin, France. *Hydrobiologia* 410, 59–68.
- Brion, N., 1993. Etude de la nitrification dans l'estuaire de la Seine. Master, Groupe de Microbiologie des Milieux Aquatiques. Université Libre de Bruxelles, Belgique.
- Bruning, K., Lingeman, R., Ringelberg, J., 1992. Estimating the impact of fungal parasites on phytoplankton populations. *Limnol. Oceanogr.* 37 (2), 252–260.
- Brunner, G. W., 2016. HEC-RAS, River analysis system hydraulic reference manual. U.S. Army Corps of Engineers, Davis, CA.
- Burgers, G., Jan van Leeuwen, P., Evensen, G., 1998. Analysis scheme in the ensemble kalman filter. *Monthly Weather Review* 126 (6), 1719–1724.
- Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software* 22 (10), 1509–1518.
- Camporese, M., Paniconi, C., Putti, M., Salandin, P., 2009. Ensemble kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow. *Water Resources Research* 45, W10421.

- Capblancq, J., 1982. Phytoplankton et production primaire. Masson, pp. 1–48.
- Capblancq, J., Dauta, A., 1978. Phytoplankton et production primaire de la rivière lot. *Annls Limnol.* 14, 85–112.
- Cappe, O., Godsill, S. J., Moulines, E., 2007. An overview of existing methods and recent advances in sequential monte carlo. *Proceedings of the IEEE* 95 (5), 899–924.
- Carrassi, A., Bocquet, M., Bertino, L., Evensen, G., 2018. Data assimilation in the geosciences : An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews : Climate Change* 9 (5), e535.
- Cho, J., Her, Y., Bosch, D., 2017. Sensitivity of simulated conservation practice effectiveness to representation of field and in-stream processes in the little river watershed. *Environmental Modeling & Assessment* 22 (2), 159–173.
- Cossarini, G., Mariotti, L., Feudale, L., Mignot, A., Salon, S., Taillandier, V., Teruzzi, A., D’Ortenzio, F., 2019. Towards operational 3d-var assimilation of chlorophyll biogeochemical-argo float data into a biogeochemical model of the mediterranean sea. *Ocean Modelling* 133, 112 – 128.
- Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., Fisher, M., Pailleux, J., 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I : Formulation. *Quarterly Journal of the Royal Meteorological Society* 124 (550), 1783–1807.
- Courtier, P., Thépaut, J. N., Hollingsworth, A., 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society* 120 (519), 1367–1387.
- Cox, B. A., 2003a. A review of currently available in-stream water-quality models and their applicability for simulating dissolved oxygen in lowland rivers. *Sciences of Total Environment* 314–316, 335–377.
- Cox, B. A., 2003b. A review of dissolved oxygen modelling techniques for lowland rivers. *Sciences of Total Environment* 314–316, 303–334.

- Cox, B. A., Whitehead, P. G., 2005. Parameter sensitivity and predictive uncertainty in a new water quality model, Q^2 . *Journal of Environmental Engineering* 131 (1), 147–157.
- Daget, N., 2007. Revue des méthodes d'assimilation. Tech. rep., CERFACS, Toulouse, France.
- Dauta, A., Devaux, J., Piquemal, F., Boumnich, L., 1990. Growth rate of four freshwater algae in relation to light and temperature. In : Bonin, D., Golterman, H. (Eds.), *Fluxes Between Trophic Levels and Through the Water-Sediment Interface*. pp. 221–226.
- de Marsily, G., 1981. *Hydrogéologie quantitative*. Collection sciences de la Terre, Paris.
- DeChant, C. M., Moradkhani, H., 2012. Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research* 48 (4).
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kallberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The era-interim reanalysis : configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 137 (656), 553–597.
- Deflandre, A., Williams, R., Elorza, F., Mira, J., Boorman, D., 2006. Analysis of the QUESTOR water quality model using a fourier amplitude sensitivity test (FAST) for two UK rivers. *Science of The Total Environment* 360 (1), 290–304.
- Degoutte, G., 2002. *HYDRAULIQUE ET DYNAMIQUE FLUVIALE appliquées au diagnostic, à l'aménagement et la gestion des rivières*. ENGREF, Paris, France.

- Delmas, M., Cerdan, O., Cheviron, B., Mouchel, J.-M., Eyrolle, F., 2012. Sediment export from French rivers to the sea. *Earth Surface Processes and Landforms* 37, 754–762.
- Descy, J.-P., Leitao, M., Everbecq, E., Smitz, J.-S., Delière, J.-F., 2012. Phytoplankton of the River Loire, France : a biodiversity and modelling study. *Journal of Plankton Research* 34 (2), 120–135.
- Doucet, A., de Freitas, N., Gordon, N., 2001. *Sequential Monte Carlo Methods in Practice*. Springer, New York.
- Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing* 10 (3), 197–208.
- Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* 28 (4), 1015–1031.
- Durand, M., Andreadis, K., Alsdorf, D., Lettenmaier, D., Moller, D., Wilson, M., 2008. Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model. *Geophysical Research Letters* 35, L20401.
- Ebel, B., Loague, K., 2006. Physics-based hydrologic-response simulation : Seeing through the fog of equifinality. *Hydrological Processes* 20, 2887–2900.
- Efron, B., Tibshirani, R., 1993. *An Introduction to the Bootstrap*. Macmillan Publishers Limited., Boca Raton.
- Escoffier, N., Bensoussan, N., Vilmin, L., Flipo, N., Rocher, V., David, A., Métivier, F., Groleau, A., 2018. Estimating ecosystem metabolism from continuous multi-sensor measurements in the seine river. *Environmental Science and Pollution Research* 25 (24), 23451–23467.
- Escoffier, N., Bernard, C., Hamlaoui, S., Groleau, A., Catherine, A., 10 2014. Quantifying phytoplankton communities using spectral fluorescence : the effects of species composition and physiological state. *Journal of Plankton Research* 37 (1), 233–247.

- Even, S., 1995. Modélisation d'un écosystème fluvial : la Seine - le modèle PROSE. Ph.D. thesis, Centre d'Informatique Géologique, Ecole Nationale Supérieure des Mines de Paris, France.
- Even, S., Bacq, N., Ruelland, D., Billen, G., Garnier, J., Poulin, M., Théry, S., Blanc, S., 2007a. New tools for modelling water quality of hydrosystems : An application in the Seine River basin in the frame of the Water Framework Directive. *Sciences of Total Environment* 375 (1-3), 274–291.
- Even, S., Mouchel, J. M., Servais, P., Flipo, N., Poulin, M., Blanc, S., Chabanel, M., Paffoni, C., 2007b. Modeling the impacts of Combined Sewer Overflows on the river Seine water quality. *Sciences of Total Environment* 375 (1-3), 140–151.
- Even, S., Poulin, M., Garnier, J., Billen, G., Servais, P., Chesterikoff, A., Coste, M., 1998. River ecosystem modelling : Application of the PROSE model to the Seine river (France). *Hydrobiologia* 373, 27–37.
- Even, S., Poulin, M., Mouchel, J. M., Seidl, M., Servais, P., 2004. Modelling oxygen deficits in the Seine river downstream of combined sewer overflows. *Ecol. Model.* 173, 177–196.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research : Oceans* 99 (C5), 10143–10162.
- Evensen, G., 2003. The Ensemble Kalman Filter : theoretical formulation and practical implementation. *Ocean Dynamics* 53 (4), 343–367.
- Flipo, N., 2005. Modélisation intégrée des transferts d'azote dans les aquifères et les rivières : Application au bassin du Grand Morin. Ph.D. thesis, Centre d'Informatique Géologique, Ecole Nationale Supérieure des Mines de Paris.
- Flipo, N., 2013. Modélisation des hydrosystèmes continentaux pour une gestion durable de la ressource en eau. Habilitation à diriger des recherches, Université Pierre et Marie Curie - Paris VI, France, Habilitation thesis.
- Flipo, N., Even, S., Poulin, M., Tusseau-Vuillemin, M. H., Améziiane, T., Dauta, A., 2004. Biogeochemical modelling at the river scale : Plankton

- and periphyton dynamics - Grand Morin case study, France. *Ecological Modelling* 176, 333–347.
- Flipo, N., Jeannée, N., Poulin, M., Even, S., Ledoux, E., 2007a. Assessment of nitrate pollution in the Grand Morin aquifers (France) : combined use of geostatistics and physically-based modeling. *Environmental Pollution* 146 (1), 241–256.
- Flipo, N., Mouchel, J. M., Fisson, C., Wang, S., Le Gall, M., Ayrault, S., Labadie, P., Gasperi, J., Guillon, S., Budzinski, H., Evrard, O., Romary, T., Chautru, E., Abhervé, D., Chevillotte, G., Narcy, J. B., Catten, A., Meybeck, M., 2018. Les effets de la crue de juin 2016 sur la qualité de l'eau du bassin de la Seine. Fascicule #17 du PIREN-Seine, ISBN : 978-2-490463-05-3.
- Flipo, N., Rabouille, C., Poulin, M., Even, S., Tusseau-Vuillemin, M., Lalande, M., 2007b. Primary production in headwater streams of the Seine basin : the Grand Morin case study. *Sciences of Total Environment* 375, 98–109.
- Gamerith, V., Neumann, M. B., Muschalla, D., 2013. Applying global sensitivity analysis to the modelling of flow and water quality in sewers. *Water Research* 47 (13), 4600–4611.
- Garambois, P. A., Roux, H., Larnier, K., Castaings, W., Dartus, D., 2013. Characterization of process-oriented hydrologic model behavior with temporal sensitivity analysis for flash floods in mediterranean catchments. *Hydrology and Earth System Sciences* 17 (6), 2305–2322.
- Garnier, J., Billen, G., 1993. Ecological interactions in a shallow sand-pit (Créteil Lake, France). *Hydrobiologia* 275/276, 97–114.
- Garnier, J., Billen, G., Coste, M., 1995. Seasonal succession of diatoms and chlorophyceae in the drainage network of the river Seine : Observations and modelling. *Limnology and Oceanography* 40 (4), 750–765.
- Garnier, J., Billen, G., Hanset, P., Testard, P., Coste, M., 1998. Développement algual et eutrophisation dans le réseau hydrographique de la Seine. In : Meybeck, M., de Marsily, G., Fustec, E. (Eds.), *La Seine en son bassin-Fonctionnement écologique d'un système fluvial anthropisé*. Elsevier, Amsterdam, The Netherlands, pp. 593–626.

- Garnier, J., Billen, G., Sanchez, N., Leporcq, B., 2000. Ecological functioning of the Marne reservoir (upper Seine basin, France). *Regulated Rivers : Research & Management* 16, 51–71.
- 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.* 37, 83–94.
- Garnier, J., Marescaux, A., Guillon, S., Vilmin, L., Rocher, V., Billen, G., Thieu, V., Silvestre, M., Passy, P., Raimonet, M., Groleau, A., Théry, S., Tallec, G., Flipo, N., 2019. The Seine River Basin. *Handbook of Environmental Chemistry*. Springer, New York, USA, Ch. Ecological functioning of the Seine River : from long-term modelling approaches to high-frequency data analysis, p. In press.
- Garnier, J., Servais, P., Billen, G., Akopian, M., Brion, N., 2001. Lower seine river and estuary (france) carbon and oxygen budgets during low flow. *Estuaries* 24 (6), 964–976.
- Gauthier, P., Tanguay, M., Laroche, S., Pellerin, S., Morneau, J., 2007. Extension of 3DVAR to 4DVAR : Implementation of 4DVAR at the Meteorological Service of Canada. *Monthly Weather Review* 135 (6), 2339–2354.
- Gharamti, M., Tjiputra, J., Bethke, I., Samuelson, A., Skjelvan, I., Bentsen, M., Bertino, L., 2017. Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites. *Ocean Modelling* 112, 65 – 89.
- Gharamti, M. E., Hoteit, I., Valstar, J., 2013. Dual states estimation of a subsurface flow-transport coupled model using ensemble kalman filtering. *Advances in Water Resources* 60, 75 – 88.
- Giorgio, P. A. D., Cole, J. J., 1998. Bacterial growth efficiency in natural aquatic systems. *Annual Review of Ecology and Systematics* 29, 503–41.
- Goffin, A., Guérin, S., Rocher, V., Varrault, G., 2017. Caractérisation de l'évolution de la matière organique dissoute de l'amont à l'aval de l'agglomération parisienne pendant une année hydrologique par spectrométrie de fluorescence 3d. Tech. rep., PIREN Seine.

- Groleau, A., Escoffier, N., Bensoussan, N., Flipo, N., Poulin, M., Vilmin, L., Bernard, C., Rocher, V., Métivier, F., 2012. Biogéochimie haute résolution : Bilan méthodologique et présentation des observations pour l'année 2011. Tech. rep., PIREN Seine.
- Groleau, A., Escoffier, N., Bensoussan, N., Flipo, N., Poulin, M., Vilmin, L., Bernard, C., Rocher, V., Métivier, F., 2013. Réseau de mesures haute fréquence CarboSeine : validation métrologique et indicateurs du fonctionnement trophique. Tech. rep., PIREN-Seine.
- Groleau, A., Escoffier, N., Vilmin, L., Auge, A., Bensoussan, N., Poulin, M., Rocher, V., Métivier, F., Flipo, N., 2014. Apports croisés des mesures haute fréquence pour le calcul d'indicateurs et pour la modélisation du fonctionnement trophique de la Seine à l'aval de Paris. Tech. rep., PIREN-Seine.
- Guillon, S., Thorel, M., Flipo, N., Oursel, B., Claret, C., Fayolle, S., Bertrand, C., Rappelle, B., Piegay, H., Olivier, J.-M., Vienney, A., Marmonier, P., Franquet, E., 2019. Functional classification of artificial alluvial ponds driven by connectivity with the river : Consequences for restoration. *Ecological Engineering* 127, 394–403.
- Gupta, H. V., Kling, H., Yilmaz, K. K., Martinez, G. F., 2009. Decomposition of the mean squared error and nse performance criteria : Implications for improving hydrological modelling. *Journal of Hydrology* 377 (1), 80 – 91.
- Hall, J. W., Tarantola, S., Bates, P. D., Horritt, M. S., 2005. Distributed sensitivity analysis of flood inundation model calibration. *Journal of Hydraulic Engineering* 131, 117–126.
- Hall, R. O., Tank, J. L., 2005. Correcting whole-stream estimates of metabolism for groundwater input. *Limnology and Oceanography : Methods* 3, 222–229.
- Hamill, T. M., Snyder, C., 2000. A hybrid ensemble Kalman filter–3D variational analysis scheme. *Monthly Weather Review* 128 (8), 2905–2919.
- Hanset, P., 1989. Etude du développement du phytoplancton dans le réseau hydrographique de la Seine. élaboration d'un modèle idéalisé. Mas-

- ter, Groupe de Microbiologie des Milieux Aquatiques. Université Libre de Bruxelles, Belgique.
- Hendricks Franssen, H. J., Kaiser, H. P., Kuhlmann, U., Bauser, G., Stauffer, F., Müller, R., Kinzelbach, W., 2011. Operational real-time modeling with ensemble kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating. *Water Resources Research* 47 (2).
- Hendricks Franssen, H. J., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the ensemble kalman filter : Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research* 44 (9).
- Herman, J., Usher, W., 2017. SALib : An open-source python library for sensitivity analysis. *The Journal of Open Source Software* 2 (9).
- Hostache, R., Lai, X., Monnier, J., Puech, C., 2010. Assimilation of spatial distributed water levels into a shallow-water flood model. part ii : using a remote sensing image of mosel river. *Journal of Hydrology* 390 (3-4), 257–268.
- Huang, J., Gao, J., Liu, J., Zhang, Y., 2013. State and parameter update of a hydrodynamic-phytoplankton model using ensemble Kalman filter. *Ecological Modelling* 263, 81–91.
- Jakeman, A., Letcher, R., Norton, J., 2006. Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software* 21 (5), 602–614.
- Jakobsen, H. H., Markager, S., 2016. Carbon-to-chlorophyll ratio for phytoplankton in temperate coastal waters : Seasonal patterns and relationship to nutrients. *Limnology and Oceanography* 61 (5), 1853–1868.
- Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82 (1), 35–45.
- Kalnay, E., 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge, UK.

- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., Joseph, D., 1996. The ncep/n-car 40-year reanalysis project. *Bulletin of the American Meteorological Society* 77 (3), 437–472.
- Kannel, P. R., Kanel, S. R., Lee, S., Lee, Y.-S., Gan, T. Y., 2011. A review of Public Domain Water Quality Models for Simulating Dissolved Oxygen in Rivers and Streams. *Environ Model Assess* 16, 183–204.
- Kattwinkel, M., Reichert, P., 2017. Bayesian parameter inference for individual-based models using a Particle Markov Chain Monte Carlo method. *Environmental Modelling & Software* 87, 110–119.
- Kim, K., Park, M., Min, J., Ryu, I., Kang, M., Park, L. J., 2014. Simulation of algal bloom dynamics in a river with the ensemble Kalman filter. *Journal of Hydrology* 519, 2810–2821.
- Kitagawa, G., 1996. Monte Carlo filter and smoother for non-gaussian non-linear state space models. *Journal of Computational and Graphical Statistics* 5 (1), 1–25.
- Kleist, D. T., Ide, K., 2015. An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II : 4DEnVar and Hybrid Variants. *Monthly Weather Review* 143 (2), 452–470.
- Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper danube basin under an ensemble of climate change scenarios. *Journal of Hydrology* 424-425, 264 – 277.
- Kong, A., Liu, J. S., Wong, W. H., 1994. Sequential imputations and Bayesian missing data problems. *Journal of the American statistical association* 89 (425), 278–288.
- Kroer, N., 1993. Bacterial growth efficiency on natural dissolved organic matter. *Limnology and Oceanography* 38 (6), 1282–1290.
- Laborie, B., Rocher, V., Vilmin, L., Poulin, M., Raimonet, M., Benard, L., Bernier, J., Guérin, S., Cussonneau, A., Pouillaude, J., Escoffier, N., Groleau, A., Mouchel, J.-M., Flipo, N., 2016. Le modèle de prédiction

- de la qualité de la seine prose - améliorations apportées par les récents travaux de recherche (piren-phase vi). *TSM* (11), 43–66.
- 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), 450–459.
- Lancelot, C., Mathot, S., 1987. Dynamics of a phaeocystis-dominated spring bloom in belgian coastal waters. i. phytoplanktonic activities and related parameters. *Marine Ecology Progress Series* 37, 239–248.
- Lancelot, C., Veth, C., Mathot, S., 1991. Modelling ice-edge phytoplankton bloom in the Scotia-Weddell sea sector of the southern ocean during spring 1998. *J. Mar. System* 2, 333–346.
- Li, L., Zhou, H., Hendricks Franssen, H. J., Gómez-Hernández, J. J., 2012. Groundwater flow inverse modeling in non-multigaussian media : performance assessment of the normal-score ensemble kalman filter. *Hydrology and Earth System Sciences* 16 (2), 573–590.
- Li, T., Villarrubia, G., Sun, S., Corchado, J. M., Bajo, J., 2015. Resampling methods for particle filtering : identical distribution, a new method, and comparable study. *Frontiers of Information Technology & Electronic Engineering* 16 (11), 969–984.
- Liu, J. S., 2001. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, USA.
- Lorenc, A. C., 2003. The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society* 129 (595), 3183–3203.
- Mao, J., Lee, J. H., Choi, K., 2009. The extended Kalman filter for forecast of algal bloom dynamics. *Water Research* 43 (17), 4214 – 4224.
- Markov, A. A., 1906. Extension of the law of large numbers to dependent quantities (in russian). *Izvestiia Fiz.-Matem. Obsch. Kazan Univ.*,(2nd Ser.) 15, 135–156.
- Massari, C., Brocca, L., Tarpanelli, A., Moramarco, T., 2015. Data assimilation of satellite soil moisture into rainfall-runoff modelling : A complex recipe ? *Remote Sensing* 7 (9), 11403–11433.

- Meybeck, M., de Marsily, G., Fustec, É., 1998. La Seine en son bassin, Fonctionnement écologique d'un système fluvial anthropisé. Elsevier, Amsterdam, The Netherlands.
- Minaudo, C., Curie, F., Jullian, Y., Gassama, N., Moatar, F., 2018. Qual-net, a high temporal-resolution eutrophication model for large hydrographic networks. *Biogeosciences* 15 (7), 2251–2269.
- Monbet, V., 2009. Quelques apports à la modélisation stochastique en océanographie et météorologie. HDR.
- Moradkhani, H., Hsu, K., Gupta, H., Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters : Sequential data assimilation using the particle filter. *Water Resources Research* 41 (5).
- Morel, A., Gentili, B., Claustre, H., Babin, M., Bricaud, A., Ras, J., Tièche, F., 2007. Optical properties of the “clearest” natural waters. *Limnology and Oceanography* 52 (1), 217–229.
- Morris, M., 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics* 33, 161–174.
- Needoba, J. A., Peterson, T. D., Johnson, K. S., 2012. Molecular biological technologies for ocean sensing. Springer, New York, USA, Ch. Method for the Quantification of Aquatic Primary Production and Net Ecosystem Metabolism Using In Situ Dissolved Oxygen Sensors, pp. 73–101.
- Nossent, J., Elsen, P., Bauwens, W., 2011. Sobol sensitivity analysis of a complex environmental model. *Environmental Modelling & Software* 26 (12), 1515–1525.
- Odum, H. T., 1956. Primary production in flowing waters. *Limnology and Oceanography* 1, 795–801.
- Page, T., Smith, P. J., Beven, K. J., Jones, I. D., Elliott, J. A., Maberly, S. C., Mackay, E. B., Ville, M. D., Feuchtmayr, H., 2018. Adaptive forecasting of phytoplankton communities. *Water Research* 134, 74 – 85.
- Paiva, R. C. D., Collischonn, W., Bonnet, M.-P., de Gonçalves, L. G. G., Calmant, S., Getirana, A., Santos da Silva, J., 2013. Assimilating in situ

- and radar altimetry data into a large-scale hydrologic-hydrodynamic model for streamflow forecast in the amazon. *Hydrology and Earth System Sciences* 17 (7), 2929–2946.
- Park, R., I. Connolly, C., R. Albanese, J., S. Clesceri, L., W. Heitzman, G., H. Herbrandson, H., H. Indyke, B., R. Loehe, J., Ross, S., D. Sharma, D., W. Shuster, W., 1982. Modeling the fate of toxic organic materials in aquatic environments. Rept. EPA-600/S3-82-028. USEPA, Athens, GA.
- Park, R., O'Neill, R., Bloomfield, J., Shugart, H., Booth, R., Koonce, J., Adams, M., Clesceri, L., Colon, E., Dettman, E., Goldstein, R., Hoopes, J., Huff, D., Katz, S., Kitchell, J., Kohberger, R., LaRow, E., McNaught, D., Peterson, J., Zahorcak, C., 01 1974. A generalized model for simulating lake ecosystems. *Simulation* 23, 30–50.
- Park, R. A., Clough, J. S., Coombs Wellman, M., 2008. Aquatox : Modeling environmental fate and ecological effects in aquatic ecosystems. *Ecological Modelling* 213, 1–15.
- Park, S. S., Lee, Y. S., 2002. A water quality modeling study of the nakdong river, korea. *Ecological Modeling* 152, 65–75.
- Parliament Council of the European Union, 2000. Directive 2000/60/CE du 23 octobre 2000 établissant un cadre pour une politique communautaire dans le domaine de l'eau.
- Pasetto, D., Camporese, M., Putti, M., 2012. Ensemble Kalman filter versus particle filter for a physically-based coupled surface–subsurface model. *Advances in Water Resources* 47, 1–13.
- Pastres, R., Ciavatta, S., Solidoro, C., 2003. The extended Kalman Filter (EKF) as a tool for the assimilation of high frequency water quality data. *Ecological Modelling* 170 (2), 227–235.
- Pearson, K., 1905. The problem of the random walk. *Nature* 72 (1865), 294.
- Pelletier, G., Chapra, S., Tao, H., 03 2006. QUAL2Kw — a framework for modeling water quality in streams and rivers using a genetic algorithm for calibration. *Environmental Modelling and Software*, 419–425.

- Platt, T., Denman, K., Jassby, A., 1977. The sea - Ideas and Observations on Progress in the Study of the Seas - Marine Modelling. Vol. 6. John Wiley and Sons, New York, Ch. Modelling the productivity of phytoplankton, pp. 807–856.
- Plaza, D. A., De Keyser, R., De Lannoy, G. J. M., Giustarini, L., Matgen, P., Pauwels, V. R. N., 2012. The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter. *Hydrology and Earth System Sciences* 16 (2), 375–390.
- Polus, E., Flipo, N., de Fouquet, C., Poulin, M., 2011. Geostatistics for assessing the efficiency of distributed physically-based water quality model. Application to nitrates in the Seine River. *Hydrological Processes* 25 (2), 217–233.
- R Core Team, 2014. R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL <https://www.R-project.org>
- Raimonet, M., Vilmin, L., Flipo, N., Rocher, V., Laverman, A., 2015. Modeling the fate of nitrite in an urbanized river using experimentally obtained nitrifier growth parameters. *Water Research* 73, 373–387.
- Rauch, W., Henze, M., Koncsocs, L., Reichert, P., Shanahan, P., Somlyódy, L., Vanrolleghem, P., 1998. River water quality modelling : I. state of the art. *Water Science and Technology* 38 (11), 237–244.
- Redfield, A., Ketchum, B., Richards, F., 1963. The Sea. Ideas and Observations on Progress in the Study of the Seas. The Composition of the Sea-Water Comparative and Descriptive Oceanography. Vol. 2. Interscience Publishers, Geneva, Swiss, Ch. The influence of organisms on the composition of sea-water, pp. 26–77.
- Reichert, P., 2001. River Water Quality Model no.1 (RWQM1) : Case study II oxygen and nitrogen conversion processes in the river glatt (Switzerland). *Water Science and Technology* 43 (5), 51–60.
- Reichert, P., Borchardt, D., Henze, M., Rauch, W., Shanahan, P., Somlyódy, L., Vanrolleghem, P., 2001a. River Water Quality Model no.1. Tech. rep., IWA, London, UK.

- Reichert, P., Borchart, D., Henze, M., Rauch, W., Shanahan, P., Somlyódy, L., Vanrolleghem, P., 2001b. River Water Quality Model no. 1 : II. biochemical process equations. *Water Science and Technology* 43 (5), 11–30.
- Reichert, P., Vanrolleghem, P., 2001. Identifiability and uncertainty analysis of the river water quality model no. 1(RWQM1). *Water Science & Technology* 43 (7), 329–338.
- Reusser, D. E., Buytaert, W., Zehe, E., 2011. Temporal dynamics of model parameter sensitivity for computationally expensive models with the fourier amplitude sensitivity test. *Water Resources Research* 47 (7).
- Rocher, V., Azimi, S., 2017. Evolution de la qualité de la Seine en lien avec les progrès de l’assainissement. Editions Johannet, Paris, France.
- Rocher, V., Garcia-Gonzalez, E., Paffoni, C., Thomas, W., 2011. La production de nitrites lors de la dénitrification des eaux usées : un sujet sensible et complexe ! *L’Eau, l’Industrie, les Nuisances* 344, 80–83.
- Sakov, P., Oliver, D. S., Bertino, L., 2012. An iterative enkf for strongly nonlinear systems. *Monthly Weather Review* 140 (6), 1988–2004.
- Salamon, P., Feyen, L., 2009. Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *Journal of Hydrology* 376, 428–442.
- Saltelli, A., 2002. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications* 145, 280–297.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications* 181 (2), 259–270.
- Saltelli, A., Chan, K., Scott, E., 2000. *Sensitivity Analysis (Vol 1.)*. New York : Wiley.
- Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. *Sensitivity Analysis in Practice : A Guide to Assessing Scientific Models*. Halsted Press, New York, USA.

- Samuel, J., Coulibaly, P., Dumedah, G., Moradkhani, H., 2014. Assessing model state and forecasts variation in hydrologic data assimilation. *Journal of Hydrology* 513, 127 – 141.
- Sánchez, E., Colmenarejo, M. F., Vicente, J., Rubio, A., García, M. G., Travieso, L., Borja, R., 2007. Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. *Ecological Indicators* 7 (2), 315–328.
- Särkkä, S., 2013. *Bayesian Filtering and Smoothing*. Cambridge University Press, UK.
- Sasaki, Y., 1955. A fundamental study of the numerical prediction based on the variational principle. *Journal of the Meteorological Society of Japan*. Ser. II 33 (6), 262–275.
- Sasaki, Y., 1958. An objective analysis based on the variational method. *Journal of the Meteorological Society of Japan*. Ser. II 36 (3), 77–88.
- Sathyendranath, S., Stuart, V., Nair, A., Oka, K., Nakane, T., Bouman, H., Forget, M., Maass, H., Platt, T., 2009. Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. *Marine Ecology-progress Series* 383, 73–84.
- Schneider, R., Ridler, M.-E., Godiksen, P. N., Madsen, H., Bauer-Gottwein, P., 2018. A data assimilation system combining cryosat-2 data and hydrodynamic river models. *Journal of Hydrology* 557", 197 – 210.
- Servais, P., Billen, G., Hascoët, M.-C., 1987. Determination of the biodegradable fraction of dissolved organic matter in waters. *Water Research* 21 (4), 445–450.
- Servais, P., Billen, G., Vives-Rego, J., 1985. Rate of bacterial mortality in aquatic environments. *Applied and Environmental Microbiology* 49 (6), 1448–1454.
- Servais, P., Garnier, J., 1993. Contribution of Heterotrophic Bacterial Production to the Carbon Budget of the River Seine (France). *Microbial Ecology* 25, 19–33.

- Shanahan, P., Borchardt, D., Henze, M., Rauch, W., Reichert, P., Somlyódy, L., Vanrolleghem, P., 2001. River Water Quality Model no. 1 (RWQM1) : I. Modelling approach. *Water Science and Technology* 43 (5), 1–9.
- Sharma, D., Kansal, A., 2013. Assessment of river quality models : a review. *Reviews in Environmental Science and Bio/Technology* 12, 285–311.
- Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., Yu, X., 2014. Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter : A synthetic experiment. *Water Resources Research* 50 (1), 706–724.
- Simon, E., Bertino, L., 2012. Gaussian anamorphosis extension of the denkf for combined state parameter estimation : Application to a 1D ocean ecosystem model. *Journal of Marine Systems* 89 (1), 1 – 18.
- Simon, E., Samuelsen, A., Bertino, L., Dumont, D., 2012. Estimation of positive sum-to-one constrained zooplankton grazing preferences with the denkf : a twin experiment. *Ocean Science* 8 (4), 587–602.
- Simon J. Julier, J. K. U., 1997. New extension of the Kalman filter to non-linear systems.
- Sincock, A., Wheeler, H., Whitehead, P., 2003. Calibration and sensitivity analysis of a river water quality model under unsteady flow conditions. *Journal of Hydrology* 277, 214–229.
- Smith, R. C., Baker, K. S., Jan 1981. Optical properties of the clearest natural waters (200–800 nm). *Applied Optics* 20 (2), 177–184.
- Sobol, I., 1993. Sensitivity estimates for on linear mathematical models. *Mathematical Modelling and Computational Experiments* 4 (1).
- Sterner, R. W., Chrzanowski, T. H., Elser, J. J., George, N. B., 1995. Sources of nitrogen and phosphorus supporting the growth of bacterio- and phytoplankton in an oligotrophic canadian shield lake. *Limnology and Oceanography* 40 (2), 242–249.
- Streeter, H., Phelps, E. B., 1925. A study of the pollution and natural purification of the Ohio River. Tech. Rep. 146, U.S. Public Health Service, Treasury Department, Washington DC.

- Sun, X., Newham, L., Croke, B., Norton, J., 2012. Three complementary methods for sensitivity analysis of a water quality model. *Environmental Modelling & Software* 37 (Supplement C), 19–29.
- Tang, Y., Reed, P., van Werkhoven, K., Wagener, T., 2007a. Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis. *Water Resources Research* 43 (6).
- Tang, Y., Reed, P., Wagener, T., van Werkhoven, K., 2007b. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and Earth System Sciences* 11 (2), 793–817.
- Thibodeaux, L., Poulin, M., Even, S., 1994. A model for enhanced aeration of streams by motor vessels with application to the Seine river. *Journal of Hazardous Materials* 37 (3), 459–473.
- Thieu, V., Silvestre, M., Garnier, J., Billen, G., 2017. Introducing the biogeochemical pynuts-riverstrahler model to assess prospective scenario impact along the aquatic continuum in western eu-rivers. In : *Proceedings of ASLO Aquatic Sciences Meeting : "Moutains to the Sea"*. Honolulu, USA.
- Thompson, P. D., 1969. Reduction of analysis error through constraints of dynamical consistency. *Journal of Applied Meteorology* 8 (5), 738–742.
- Vanhooren, H., Meirlaen, J., Amerlinck, Y., Claeys, F., Vangheluwe, H., Vanrolleghem, P. A., 2003. WEST : Modelling biological wastewater treatment. *Journal of Hydroinformatics* 5, 27–50.
- Vannote, R., Minshall, G., Cummins, K., Sedell, J., Cushing, C., 1980. The river continuum concept. *Journal canadien des sciences halieutiques et aquatiques* 37 (1), 130–137.
- Vanrolleghem, P., Borchardt, D., Henze, M., Rauch, W., Reichert, P., Shanahan, P., Somlyódy, L., 2001. River Water Quality Model no.1 (RWQM1) : III biochemical submodel selection. *Water Science and Technology* 43 (5), 31–40.
- Vilmin, L., 2012. Module hydrodynamique HyCu. Fontainebleau, France.

- 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, France.
- Vilmin, L., Aissa-Grouz, N., Garnier, J., Billen, G., Mouchel, J. M., Poulin, M., Flipo, N., 2015a. Impact of hydro-sedimentary processes on the dynamics of soluble reactive phosphorus in the Seine River. *Biogeochemistry* 122, 229–251.
- Vilmin, L., Flipo, N., de Fouquet, C., Poulin, M., 2015b. Pluri-annual sediment budget in a navigated river system : The Seine River (France). *Sciences of Total Environment* 502, 48–59.
- Vilmin, L., Flipo, N., Escoffier, N., Groleau, A., 2018. 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* 25 (24), 23485–23501.
- Vilmin, L., Flipo, N., Escoffier, N., Rocher, V., Groleau, A., 2016. Carbon fate in a large temperate human-impacted river system : Focus on benthic dynamics. *Global Biogeochem. Cycles* 30 (7), 1086–1104.
- Vilmin, L., Flipo, N., Poulin, M., 2012. Le modèle de simulation biogéochimique C-RIVE. Tech. rep., PIREN-Seine.
- Vrugt, J. A., ter Braak, C. J., Diks, C. G., Schoups, G., 2013. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation : Theory, concepts and applications. *Advances in Water Resources* 51, 457–478.
- Wan, E. A., Van Der Merwe, R., Oct 2000. The unscented Kalman filter for nonlinear estimation. In : *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium* (Cat. No.00EX373). pp. 153–158.
- Wang, Q., Li, S., Jia, P., Qi, C., Ding, F., 2013. A review of surface water quality models. *The Scientific World Journal* 2013, 7.
- Wang, S., Flipo, N., Rivière, A., Romary, T., 2018a. Mise en place d'une nouvelle version de ProSe : ProSe-P. Tech. rep., PIREN-Seine.

- Wang, S., Flipo, N., Romary, T., 2018b. Time-dependent global sensitivity analysis of the C-RIVE biogeochemical model in contrasted hydrological and trophic contexts. *Water Research* 144, 341–355.
- Wang, S., Flipo, N., Romary, T., 2019. Oxygen data assimilation for estimating micro-organism communities' parameters in river systems. *Water Research* 165, 115021.
- Wang, X. J., Behrenfeld, M., Le Borgne, R., Murtugudde, R., Boss, E., 2009. Regulation of phytoplankton carbon to chlorophyll ratio by light, nutrients and temperature in the Equatorial Pacific ocean : a basin-scale model. *Biogeosciences* 6, 391–404.
- Warn, A. E., 1987. SIMCAT-a catchment simulation model for planning investment for river quality. In : Beck, M. B. (Ed.), *Systems Analysis in Water Quality Management*. Oxford : IAWPRC, Pergamon, pp. 211–218.
- Warren, C. E., Doudoroff, P., Shumway, D. L., 1973. Development of dissolved oxygen criteria for freshwater fish. Tech. rep., Washington, D.C.
- Weerts, A. H., El Serafy, G. Y. H., 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resources Research* 42 (9).
- Wetzel, R. G., 1983. *Limnology*, second edition. Saunders College Publishing, USA.
- Whitehead, P., Beck, B., O'Connell, E., 1981. A systems model of streamflow and water quality in the bedford ouse river system — ii. water quality modelling. *Water Research* 15, 1157–1171.
- Whitehead, P. G., Williams, R. J., Lewis, D. R., 1997. Quality simulation along river systems (quasar) : model theory and development. *Sciences of Total Environment* 194/195, 447–456.
- Wikle, C. K., Berliner, L. M., 2007. A Bayesian tutorial for data assimilation. *Physica D : Nonlinear Phenomena* 230 (1), 1–16.
- Wilke, C. R., Chang, P., 1955. Correlation of diffusion coefficients in dilute solutions. *AIChE Journal* 1 (2), 264–270.

- Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 2 (1), 37 – 52.
- Wool, T., Ambrose, R., Martin, J., and, E. C., 2001. Water Quality Analysis Simulation Program (WASP) version 6.0 - User's manual. Tech. rep., USEPA.
- Xue, P., Chen, C., Beardsley, R. C., 2012. Observing system simulation experiments of dissolved oxygen monitoring in Massachusetts Bay. *Journal of Geophysical Research : Oceans* 117 (C5).
- Yan, H., Moradkhani, H., 2016. Combined assimilation of streamflow and satellite soil moisture with the particle filter and geostatistical modeling. *Advances in Water Resources* 94, 364 – 378.
- Yu, L., Fennel, K., Bertino, L., Gharamti, M. E., Thompson, K. R., 2018. Insights on multivariate updates of physical and biogeochemical ocean variables using an ensemble kalman filter and an idealized model of upwelling. *Ocean Modelling* 126, 13 – 28.
- Yucel, I., Onen, A., Yilmaz, K., Gochis, D., 2015. Calibration and evaluation of a flood forecasting system : Utility of numerical weather prediction model, data assimilation and satellite-based rainfall. *Journal of Hydrology* 523, 49–66.
- Zahraeifard, V., Deng, Z., 2012. Modeling sediment resuspension-induced do variation in fine-grained streams. *Sciences of Total Environment* 441, 176–181.
- Zhai, X., Zhang, Y., Wang, X., Xia, J., Liang, T., 2014. Non-point source pollution modelling using soil and water assessment tool and its parameter sensitivity analysis in xin'anjiang catchment, china. *Hydrological Processes* 28 (4), 1627–1640.
- Zweifel, U. L., Wikner, J., Hagström, A., Lundberg, E., Norrman, B., 1995. Dynamics of dissolved organic carbon in a coastal ecosystem. *Limnol. Oceanogr.* 40 (2), 299–305.

RÉSUMÉ

Cette thèse a pour objectif d'implémenter un schéma d'assimilation de données dans le modèle hydro-biogéochimique PROSE, afin d'assimiler les mesures en continu d'oxygène dissous de la colonne d'eau et de déterminer l'évolution temporelle des propriétés physiologiques des communautés vivantes. Dans un premier temps, une nouvelle version parallélisée de PROSE, PROSE-P, est développée en couplant les bibliothèques hydraulique, de transport et biogéochimique (C-RIVE). Dans un deuxième temps, une analyse de sensibilité du module C-RIVE permet d'identifier un nombre restreint de paramètres influençant fortement les concentrations en oxygène dissous. Basé sur cette sélection, un algorithme de filtrage particulaire est implémenté afin d'assimiler séquentiellement les données haute fréquence d'O₂. Le couple PROSE-P-filtre particulaire, PROSE-PA, est ensuite appliqué sur un cas synthétique afin d'identifier les paramètres numériques pertinents et de valider l'efficacité du filtre particulaire pour les modèles de qualité de l'eau en rivière. Enfin, les mesures en continu d'O₂ de l'année 2011 en Seine sont assimilées par PROSE-PA. Les résultats montrent que PROSE-PA améliore significativement la simulation des concentrations en oxygène dissous, notamment les dynamiques algales et les chutes d'oxygène pendant les périodes de crise. L'application aux données réelles révèle cependant les limites de l'approche développée, notamment la sensibilité aux conditions aux limites. Plusieurs pistes sont proposées afin d'améliorer les performances de PROSE-PA.

MOTS CLÉS

Métabolisme, Assimilation de données, Oxygène dissous, Filtre particulaire PROSE-PA, Modélisation de la qualité de l'eau

ABSTRACT

The aim of the thesis is to implement a data assimilation scheme in the hydro-biogeochemical model PROSE, in order to assimilate continuous measurements of dissolved oxygen in the water column and to determine the temporal evolution of the physiological properties of the communities of living species. First, a new parallel version of PROSE, PROSE-P, is developed coupling the three packages: hydrodynamic, transport and biogeochemical (C-RIVE). Second, a sensitivity analysis of the C-RIVE model allows the identification of a limited number of influential parameters controlling the dissolved oxygen concentrations. Based on the selection, a particle filtering algorithm is implemented in order to assimilate sequentially the high frequency oxygen data. The coupling PROSE-P-particle filter, PROSE-PA is then applied on a synthetic case to tune the numerical settings for the data assimilation and to test the efficiency of the particle filter in river water quality models. Finally, the continuous measurements of dissolved oxygen of the year 2011 in the Seine River are assimilated by PROSE-PA. The results show that PROSE-PA improves significantly the simulation of the dissolved oxygen concentrations, especially the dynamics of algal blooms periods and the fast chute of O₂ for the critical periods. This application to the real oxygen data reveals however some limits of the developed approach, especially the sensitivity to the boundary conditions. Some ideas are proposed to improve the performances of PROSE-PA.

KEYWORDS

Metabolism, Data assimilation, Dissolved oxygen, Particle filter, PROSE-PA, Water quality modelling