

High Frequency Data Assimilation in the ProSe-PA Water Quality Model : Focus on the drivers of river metabolism under low flow conditions

Masihullah Hasanyar

▶ **To cite this version:**

Masihullah Hasanyar. High Frequency Data Assimilation in the ProSe-PA Water Quality Model : Focus on the drivers of river metabolism under low flow conditions. Biodiversity and Ecology. Université Paris sciences et lettres, 2023. English. ⟨NNT : 2023UPSLM047⟩. ⟨tel-04496212⟩

HAL Id: tel-04496212

https://pastel.hal.science/tel-04496212v1

Submitted on 8 Mar 2024

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.



HAL Authorization

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT

DE L'UNIVERSITÉ PSL

Préparée à Mines Paris – PSL

Assimilation de données haute fréquence dans le modèle de qualité de l'eau ProSe-PA - Les facteurs du métabolisme fluvial dans des conditions de faible débit

Soutenue par

Masihullah HASANYAR

Le 28 juin 2023

École doctorale n°398

Géoscience, Ressources Naturelles et Environment

Spécialité

Géosciences et géo-ingénierie

Composition du jury :

M. Gilles VARRAULT Professeur, U. Paris-Est Créteil	*Président du jury* *Examinateur*
Mme. Florentina MOATAR Directrice de Recherche, INRAe Riverly	*Rapporteur*
Mme. Sabine SAUVAGE Ingénieure de Recherche, CNRS	*Rapporteur*
M. Vincent THIEU Maître de conférences, Sorbonne Université	*Examinateur*
M. Nicolas FLIPO Directeur de Recherche, Mines Paris – PSL	*Directeur de thèse*
M. Thomas ROMARY Maître assistant, Mines Paris – PSL	*Directeur de thèse*

MINES PARIS | PSL

Acknowledgement

I would like to express my deepest gratitude and appreciation first and foremost to my PhD supervisors, Nicolas Flipo and Thomas Romary, for their guidance, support, and unwavering belief in my abilities. They not only provided me with the opportunity to pursue my PhD in France but also helped me navigate the challenges of settling into a new country and successfully conducting my research.

I am also deeply grateful to my PhD advisor, Shuaitao Wang, whose insightful comments and expertise have been invaluable at every stage of my PhD. Without his guidance, this thesis would not have been possible.

I would like to express my heartfelt thanks to jury members, Florentina Moatar, Sabine Sauvage, Vincent Thieu, and Gilles Varrault, for their willingness to evaluate my work and presenting me with valuable questions that open the way for more deep understanding of the importance of my work and new approaches ahead of my thesis. I appreciate their expertise, insights, and constructive feedback. I would like to express my sincere gratitude to the members of my thesis follow-up committee during the last three years, Jean-Marie Mouchel, Edith Parlanti and Vincent Thieu, for their insightful and constructive feedback that greatly contributed to the advancement of my research. I am also deeply grateful to Josette Garnier for agreeing to be my co-author and for providing valuable feedback that significantly improved the quality of my work.

I am deeply grateful to the French government for financing my pre-doctorate studies through MOPGA master scholarship program in 2018. I would like to thank Salma Tafasca for sharing the PhD proposal with me back in 2019 and guiding me throughout the application process. This thesis was conducted in the context of two previous outstanding theses by Shuitao Wang and Lauriane Vilmin, whose exceptional work paved the way for my research. I conducted my research within the Hydrologic Systems and Reservoirs team, in close collaboration with the Geostatistics team. I would like to thank the team members, Agnès Rivière, Sophie Guillon, Nicolas Gallois, and Dominique Bruel, for their valuable insights and support throughout my research journey.

I also wish to acknowledge the friendship and support of my colleagues at Centre de Géosciences: Veronique Lachasse, Nathalie Dietrich, Antoine Collet, Deniz Kiliç, Nicolas Seigneur, Lucia Clarotto, Pierre Guillou, Loic Marlot, Carlos Assis, Léo de Souza, Christine El Khoury, Ferdinand Bhavsar, Charlie Garayt, Inès Ben Khaled, Razane Doucmak, Irina Sin, Firas Tayeb, Diana Camacho, Theo Rebert, Rose-Nelly Capito, Lise-Marie Girod, Fabien Bez and others with whom I have shared many memorable moments in work pauses, weekends, after work parties, and forest walks in Fontainebleau.

I owe a debt of gratitude to my wife, Sahar Hasanyar, whose patience and support during these three years have been the bedrock of my success. I am also immensely grateful to my daughter, Joséphine Hasanyar, whose birth in Oct 2022 was a source of inspiration to me and made my thesis journey even more meaningful.

Finally, I would like to dedicate this thesis to my late father, Haji Ghulam Hussain Hasanyar, and to my mother, Diba Hasanyar, for their role in bringing me up and instilling in me the values of

education, hard work and perseverance. Being a father myself, I understand now how tough has it been for them to bring us up in my home country Afghanistan. I thank my brothers, sisters and my uncles for their continous support.

Résumé: De l'oxygène au carbone organique dissous biodégradable à l'étiage par assimilation de données dans le modèle de simulation de la qualité des eaux ProSe-PA

1 Introduction

La qualité de l'eau est un aspect critique de la santé des rivières qui joue un rôle pivot dans le maintien de divers processus écologiques et le soutien des activités humaines. La qualité de l'eau des rivières revêt tout d'abord une importance écologique, car elle affecte la santé et la survie des organismes aquatiques. Ensuite, les humains dépendent des rivières propres pour l'eau potable, l'irrigation et les activités récréatives. Enfin, la qualité de l'eau des rivières a également des implications économiques, car des industries telles que le tourisme, la pêche, l'agriculture et les loisirs liés à l'eau dépendent fortement des écosystèmes fluviaux propres et sains. Cependant, les activités anthropiques telles que les pratiques agricoles et les rejets d'eaux usées provenant de sources industrielles et urbaines ont entraîné la dégradation de la qualité de l'eau.

Par conséquent, des programmes de surveillance ont été lancés pour suivre la qualité de l'eau au moyen de différents indicateurs tels que l'oxygène dissous (O_2), les nitrates ou alors le phosphore soluble réactif. Parmi ces indicateurs, l'oxygène dissous a été choisi comme principal indicateur de la qualité de l'eau, car il englobe les différents processus liés aux bactéries, aux producteurs primaires, aux micro-organismes nitrifiants et intègre également les processus physiques tels que la navigation et le vent. C'est pourquoi les problèmes environnementaux tels que l'eutrophisation et d'importants niveaux d'activité des bactéries hétérotrophes, qui surviennent en raison des activités anthropiques, entraînent une diminution de la concentration en O_2, ce qui entraîne ensuite une hausse de la mortalité des organismes aquatiques ainsi qu'un goût et une odeur indésirables dans la rivière.

Le bassin de la Seine, dont la qualité de l'eau s'est significativement améliorée depuis les années 1980 [Garnier et al., 2020] est pris comme site pilote de cette étude. Il représente 11,8 % (76 236 km^2) du territoire français et accueille 25 % (17 millions) de sa population [Flipo et al., 2020]. Il regroupe également une part importante des activités économiques nationales, à la fois industrielles, agricoles, de transport et de tourisme. Le bassin est soumis à des impacts anthropiques élevés, en particulier dans la région métropolitaine de Paris, où les stations d'épuration des eaux usées (STEP) et les apports pluvieux provenant des déversoirs d'orage imposent une forte charge de matière organique

dans la rivière. De plus, l'agriculture, le pâturage des animaux, l'industrie et le développement urbain sont des sources diffuses de pollution des cours d'eau.

Dans les années 1980, l'eutrophisation de la Seine a entraîné une diminution très forte de l'oxygène dissous [Garnier et al., 2020], aggravant la mortalité des organismes aquatiques et provoquant un goût et une odeur indésirables de la rivière. De plus, l'eutrophisation de l'eau due aux nutriments en excès augmente fortement les coûts de potabilisation de l'eau [Garnier et al., 2020]. Avec la construction de nouvelles stations d'épuration des eaux usées et l'amélioration de la technologie de traitement, des améliorations significatives de la qualité de l'eau de la rivière ont été obtenues. Cependant, plusieurs défis persistent, tels que le manque de dilution suffisante des rejets ponctuels pendant les périodes de faible débit estival, les rejets non traités des STEP pendant les périodes de maintenance, l'eutrophisation dans les zones côtières, les pollutions agricoles diffuses et la contamination des eaux souterraines par les nitrates [Garnier et al., 2020].

De plus, l'oxygène et la disponibilité de ces données à haute fréquence sont importants, car ils sont utilisés pour évaluer le métabolisme de la rivière en se basant sur les informations en oxygène d'une seule station, telles que développées par Odum [1956]. Des études récentes utilisent les données sur l'oxygène pour montrer la relation entre la superficie du bassin versant et l'activité métabolique en aval [Diamond et al., 2021] et pour étudier le passage des communautés phytoplanctoniques aux communautés de macrophytes [Diamond et al., 2022]. Ceci est en accord avec le concept de "conduit actif" proposé par Cole et al. [2007] et Battin et al. [2023], qui estiment qu'environ 40 % du carbone terrestre est échangé avec l'atmosphère.

Une autre utilisation des données sur l'oxygène réside dans le développement, l'étalonnage et la validation de modèles basés sur des principes physiques. Ces modèles sont au cœur de la gestion de la qualité de l'eau, ce qui permet d'évaluer les effets des rejets anthropiques existants et de prédire l'impact des futures infrastructures (STEP, déversoirs d'orage). Le premier modèle de qualité de l'eau pour simuler l'oxygène a été développé par Streeter and Phelps [1925], en incorporant l'oxygène et la demande biologique en oxygène. Ensuite, Vannote et al. [1980] ont élaboré le concept de continuum fluvial, ce qui a favorisé le développement de modèles centrés sur les communautés de microorganismes, tels que RIVE [Billen et al., 1994]. Finalement, le logiciel de modélisation de la qualité de l'eau ProSe a été créé en 1995 en couplant RIVE avec des équations d'hydrodynamique et de transport [Even et al., 1998, 2004, 2007, Flipo et al., 2004, Vilmin et al., 2016]. De nombreux autres modèles ont été créés, dont le principal problème est qu'ils fournissent une vue statique des communautés bactériennes et phytoplanctoniques des rivières. Cela explique les écarts importants observés dans leurs simulations, alors que des études montrent que ces communautés et d'autres paramètres du modèle ont tendance à évoluer avec le temps, ce qui signifie qu'une représentation dynamique de ces paramètres est nécessaire.

Ainsi, une solution consisterait à lancer des campagnes expérimentales pour déterminer la variation temporelle des paramètres du modèle, ce qui est une tâche difficile en raison du grand nombre de paramètres impliqués. L'étalonnage pourrait également être une autre méthode mais il peut entraîner une équifinalité[1]. Par conséquent, une stratégie efficace pour obtenir une représentation dynamique des processus hydro-biogéochimiques implique l'application de techniques d'assimilation de données. Ces méthodologies consistent à coupler des modèles basés sur des principes physiques avec des données d'observation à haute fréquence, telles que l'oxygène dissous, pour déduire ou estimer avec précision

[1] L'équifinalité décrit le fait qu'un système d'équations mal contraint présente une infinité de solutions, ce qui engendre une forte incertitude sur la valeur des paramètres contrôlant le fonctionnement du système modélisé par ce jeu d'équations

l'évolution temporelle de ces paramètres du modèle, ce qui permet d'obtenir de meilleurs résultats de simulation.

L'une des techniques d'assimilation les plus connues est la méthode de filtre de Kalman d'ensemble (EnKF, El Serafy et al. [2007]), qui a été d'abord mise en œuvre dans le modèle de qualité de l'eau Delft3D et reste aujourd'hui l'une des méthodes d'assimilation les plus populaires [Beck and Halfon, 1991, Carrassi et al., 2018, Page et al., 2018, Loos et al., 2020, Cho et al., 2020]. Ensuite, Babbar-Sebens et al. [2013] ont réalisé la première application d'une méthode variationnelle, 3DVAR, en assimilant des données de température par satellite dans un modèle de qualité de l'eau pour les réservoirs.

Enfin, la première mise en œuvre d'un filtre particulaire a été réalisée dans le modèle ProSe-PA pour la Seine, en assimilant des données d'oxygène dissous à haute fréquence [Wang, 2019, Wang et al., 2019, 2022a]. Le filtre particulaire est aussi performant que l'EnKF dans les systèmes légèrement non linéaires (par exemple pour modéliser le processus de réaération dans une rivière) ; il s'est même avéré plus efficace et précis que l'EnKF [Wang et al., 2023] dans les systèmes fortement non linéaires, par exemple dans la modélisation de rivières avec une forte activité des bactéries hétérotrophes. Le filtre particulaire de ProSe-PA a ainsi été en mesure de produire des simulations d'oxygène dissous acceptables et d'identifier les bonnes distributions *a posteriori* des paramètres du modèle. L'EnKF excelle uniquement dans les systèmes linéaires, où les paramètres et les variables d'état suivent une distribution gaussienne [Evensen et al., 2022]. L'avantage du filtre particulaire découle du fait qu'il pondère ses membres (particules) en fonction de leurs performances dans la prédiction de la vraie valeur, tandis que l'EnKF suppose une distribution gaussienne uniformément pondérée pour ses membres, ce qui est considéré comme une hypothèse biaisée [Wikle and Berliner, 2007]. De plus, il a été constaté que l'EnKF présentait un coût de calcul supérieur de 25 % par rapport au filtre particulaire [Wang et al., 2023].

ProSe-PA améliore significativement le modèle ProSe pour l'estimation du métabolisme de la rivière pendant les périodes de floraison et de débit élevé, et a pu estimer avec précision les paramètres de la production primaire et des processus physiques tels que la navigation. Cependant, il a échoué à représenter le métabolisme de la rivière pendant la période de faible débit en été, où les activités bactériennes hétérotrophes contrôlent la qualité de l'eau [Wang et al., 2022a].

La période de faible débit est importante car le manque de dilution rend la rivière plus sensible aux rejets urbains. Plusieurs autres modèles de qualité de l'eau rencontrent également des difficultés similaires en période de débit faible. Comme le profil longitudinal de carbone organique dissous biodégradable (CODB) simulé était bas le long de la plupart des stations pendant la période de faible débit, nous supposons que ces divergences d'oxygène proviennent de l'incertitude concernant le CODB des conditions aux limites du modèle. Cela signifie que si nous disposions d'un paramètre qui aurait pu augmenter la part de CODB des conditions aux limites, cela aurait pu entraîner une nouvelle diminution des niveaux d'oxygène et ainsi fournir une meilleure correspondance entre les observations et les simulations. De plus, en consultant la littérature, nous pouvons constater l'importance du carbone organique dissous (COD) et celle du CODB; par exemple, une augmentation du COD et du CODB est liée au changement des communautés bactériennes [Hullar et al., 2006, Crump et al., 2003] et à l'augmentation de la diversification des populations bactériennes [Landa et al., 2013].

L'hypothèse de recherche qui guide ce travail est que les écarts observés dans les niveaux d'oxygène en période d'étiage sont dus à une mauvaise estimation du CODB. Pour tester cette hypothèse, une stratégie scientifique est conçue. Tout d'abord, un modèle conceptuel est créé pour paramétrer la

répartition du carbone en ses constituants les plus petits (section 2, p.vi). Ensuite, une revue de la littérature est entreprise pour déterminer les gammes de variation de ces paramètres (section 3, p.vii). S'ensuit une analyse de sensibilité visant à distinguer les paramètres les plus influents sur l'évolution de l'oxygène dissous (section 4, p.ix). Enfin, les paramètres influents sont ajoutés au schéma d'assimilation de ProSe-PA dans le cadre d'une étude de cas synthétique (section 5, p.xi) et réelle (section 6, p.xv), afin d'observer si nous pouvons identifier les paramètres du modèle par l'assimilation des données d'oxygène et comment cela améliore la simulation du métabolisme de la rivière en période d'étiage.

2 Modèle conceptuel du carbone organique

Un modèle de répartition du carbone organique est proposé (Figure 1), composé de cinq paramètres. L'intégration du carbone organique total (COT) et des paramètres de la répartition du carbone constituent une nouveauté ajoutée au modèle C-RIVE de ProSe-PA. En effet, la version précédente de C-RIVE n'incluait aucun paramètre permettant de définir la répartition du carbone organique en leurs fractions correspondantes (voir Fig. 1 pour les définitions).

Le carbone organique total est divisé en formes dissoutes (COD) et particulaires (COP) par le paramètre t. Ensuite, le COD est divisé en fractions biodégradables (CODB) et réfractaires (CODR) par le paramètre b_1. Grâce au paramètre s_1, le CODB est à son tour divisé en (i) $CODB_1$ en tant que substrat limitant (COD rapidement biodégradable en 5 jours) et (ii) $CODB_2$ (COD lentement biodégradable en 45 jours). Le COP est divisé de manière similaire en ses fractions constitutives en utilisant les paramètres b_2 et s_2. Dans cette configuration, la seule information requise pour exécuter le logiciel est le COT provenant de données expérimentales. Par conséquent, les cinq paramètres de répartition de la matière organique (t, b_1, b_2, s_1, s_2) nous offrent la possibilité de mener une analyse de sensibilité pour quantifier leur influence sur la concentration en oxygène dissous.

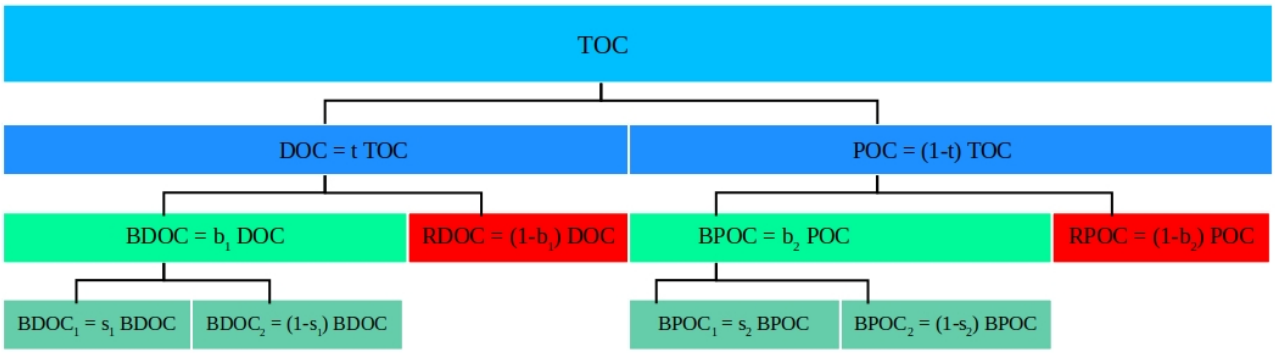


Figure 1: Répartition de la matière organique à partir du COT et des cinq paramètres t, b_1, s_1, b_2, s_2. COD : carbone organique dissous, COP : carbone organique particulaire, CODB : COD biodégradable, CODR : COD réfractaire, BDOC : COP biodégradable, COPR : COP réfractaire, $BDOC_1$ et $BDOC_2$: fractions de COD rapidement et lentement biodégradables, $BPOC_1$ et $BPOC_2$: fractions de COP rapidement et lentement biodégradables.

$$t = \frac{COD}{COT}$$

$$b_1 = \frac{CODB}{COD}$$
$$b_2 = \frac{COPB}{COP}$$
$$s_1 = \frac{CODB_1}{CODB}$$
$$s_2 = \frac{COPB_1}{COPB}$$

3 Revue bibliographique des paramètres de répartition du carbone organique

Les paramètres introduits pour la répartition du carbone sont soumis à un examen détaillé de la littérature, comprenant 150 publications.

Tout d'abord, les gammes de variation et les valeurs moyennes des cinq paramètres du modèle de carbone organique, ainsi que du COT et du COD, sont renseignées (Tab. 1). Ces valeurs peuvent être utilisées par les chercheurs en métabolisme des rivières et les modélisateurs de la qualité de l'eau pour avoir une vue d'ensemble de la variation de ces variables à l'échelle mondiale.

Table 1: Valeurs moyennes générales et gammes de variation des variables de matière organique

	COD [$mgCL^{-1}$]	COT [$mgCL^{-1}$]	t [-]	b_1 [-]	b_2 [-]	s_1 [-]	s_2 [-]
Min.	0.3	0.5	0.08	0.05	0.13	0.3	0.44
Moy.	5.0	9.0	0.57	0.23	0.35	0.59	0.62
Max.	19.3	39.0	0.96	0.54	0.52	0.95	0.95

Ensuite, l'influence des facteurs environnementaux (hydrologie, couverture terrestre et climat) sur la matière organique est étudiée via une analyse factorielle des données mixtes (AFDM) [Pagès, 2014]. Cette méthode permet de transformer l'ensemble de données en un nouvel espace, créant ainsi de nouvelles variables synthétiques appelées « dimensions ». Cela permet de mettre en lumière la relation entre les variables de l'ensemble des données (par exemple, les valeurs moyennes ou la couverture terrestre) et de déterminer les catégories importantes, appelées « groupes », qui influencent la variabilité globale de l'ensemble de données.

La couverture terrestre et le climat sont les principaux facteurs qualitatifs qui contribuent à la variance globale du COT (Fig. 2a). Les dimensions 1 et 2 représentent 45 % de la variance totale.

Les groupes[2], indiqués en lettres capitales, sont les zones humides, le climat continental, les terres arides en climat sec, et la forêt en climat tropical (Fig. 2b). Les autres catégories, en particulier celles liées à l'hydrologie, se trouvent très proches de l'origine et ne semblent pas jouer un rôle significatif dans la teneur en carbone.

Après avoir identifié les groupes pour le COT, le COD et les paramètres du modèle de répartition du carbone, le Table 2 énumère les valeurs moyennes et les gammes de variation des groupes identifiés pour chaque paramètre. À cette fin, les tableaux détaillés des variables de la matière organique sont

[2]Les groupes sont identifiés sur la base de l'analyse des dimensions 1, 2 et 3, qui représentent plus de 60 % de la variance totale. Les catégories sont identifiées dans l'espace des dimensions 1 et 2 (Fig. 2b)

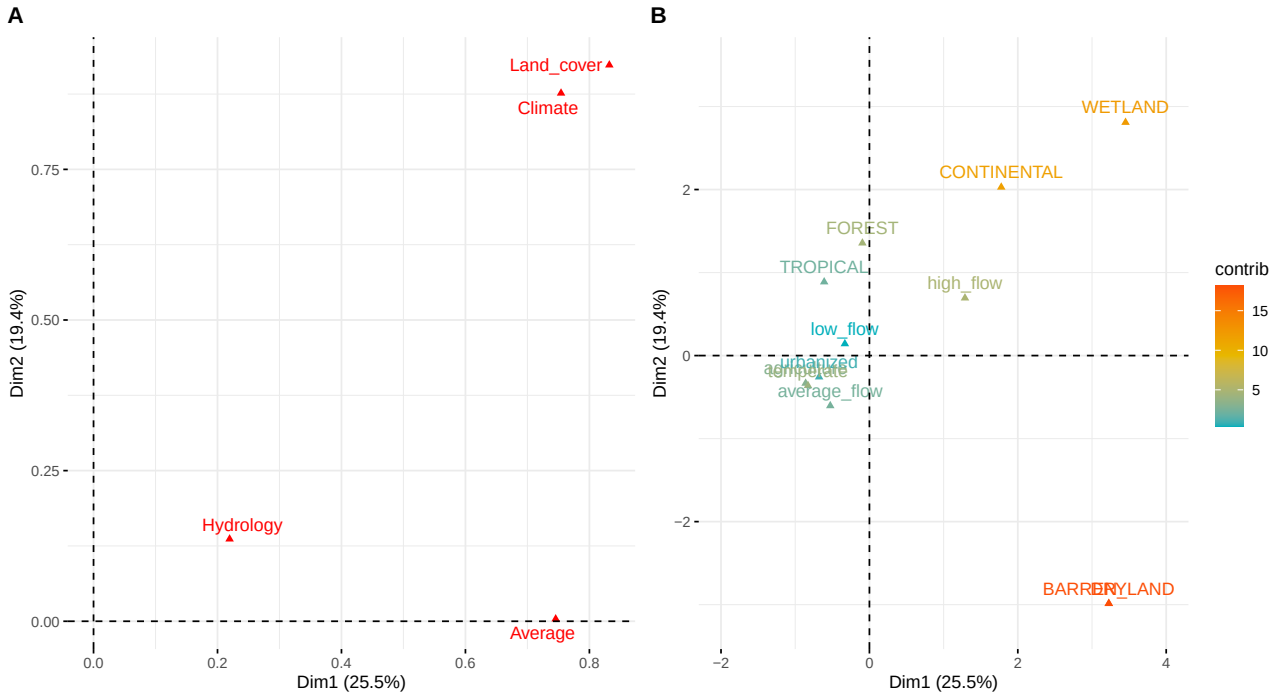


Figure 2: Résultats de l'AFDM pour COT. A) Coordonnées des facteurs dans le nouvel espace (dimensions 1 et 2). B) Représentation graphique des catégories de facteurs qualitatifs sur les dimensions 1 et 2 en fonction de leur contribution (en %) aux axes correspondants. Les catégories affichées en lettres capitales sont les groupes identifiés.

filtrés pour chaque groupe.

Table 2: Valeurs moyennes et gammes de variation des variables de la matière organique pour les groupes identifiés d'occupation du sol, de climat et d'hydrologie

Groupe	COD [mgC L^{-1}]			COT [mgC L^{-1}]			t [-]			b_1 [-]		
	min	moy	max	min	moy	max	min	moy	max	min	moy	max
zone humide	7.2	10.5	13.8	16.5	24.5	32.5				0.10	0.14	0.18
prairie	4.3	6.9	9.5				0.59	0.79	0.89			
zone urbanisé										0.23	0.35	0.48
agriculture										0.14	0.21	0.28
forêt										0.11	0.18	0.24
climat polaire	4.3	7.3	10.2				0.60	0.84	0.96			
climat continental				8.1	15.4	22.8				0.15	0.27	0.39
climat tropical										0.07	0.11	0.14
fort débit (hydrologie)										0.07	0.11	0.15
climate sec–terres arides	2.1	3.5	4.9	5.4	17.2	29.1	0.13	0.36	0.59			
climat tropical–forêt	2.4	4.6	6.7	3.0	5.1	8.3	0.31	0.43	0.53			

t = COD:COT, part de la partie dissoute du COT.
b_1 = CODB:COD, part de la partie biodégradable du COD.

Les résultats de cette étude contribuent aux recherches menées sur le métabolisme des rivières, car les paramètres du carbone organique présentés contrôlent les flux de carbone dans les systèmes fluviaux. Les chercheurs en métabolisme des rivières et les modélisateurs de la qualité de l'eau peuvent se référer au Table 2 afin d'obtenir des gammes de variation plus réduites pour leurs paramètres cibles.

En cas de chevauchement entre plusieurs groupes, les gammes peuvent être combinées pour obtenir une gamme plus large. Cependant, les gammes de variation globales présentées dans le Table 1 sont proposées pour être utilisées en cas de non-correspondance entre les caractéristiques de la rivière et les groupes identifiés. Si nous comparons ces deux tableaux pour b_1, la gamme de variation générale est de 0,05 à 0,54, tandis que dans une rivière urbanisée, cette gamme est réduite et varie entre 0,23 et 0,48.

4 Analyse de sensibilité pendant une période de faible débit estival

Dans cette partie, l'influence des paramètres de C-RIVE sur l'évolution de l'oxygène dissous est étudiée. Quatre ensembles de paramètres sont analysés : les paramètres bactériens (par exemple, le taux de croissance et la mortalité des bactéries), les paramètres physiques (par exemple, les constantes de navigation ou de vent), les paramètres de répartition du carbone organique (par exemple, le pourcentage de COD biodégradable) et les paramètres de dégradation du carbone organique (par exemple, la constante de prélèvement de la matière organique rapidement biodégradable).

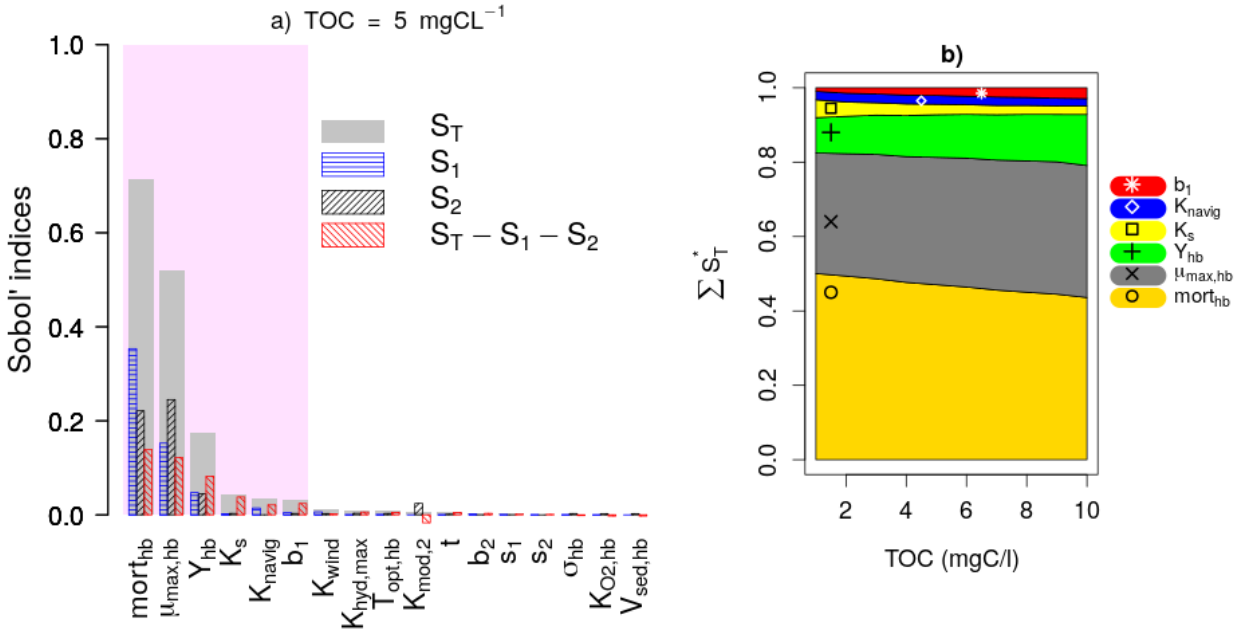


Figure 3: Analyse de sensibilité de Sobol pour tous les paramètres : (a) indices de sensibilité pour COT = 5 mgC.L^{-1}; (b) évolution des indices de sensibilité totale normalisés des paramètres influents avec le carbone organique total

Les résultats de l'analyse de sensibilité de Sobol pour COT = 5 mgC/l sont montrés sur la Figure 3a. Ils sont illustrés par un diagramme à barres des indices de sensibilité totale (S_T), de sensibilité de premier ordre (s_1) et de sensibilité de second ordre (s_2) des paramètres.

Les indices de sensibilité d'ordre supérieur sont également calculés en fonction de la différence entre l'indice total et les indices de premier et de deuxième ordre (S_T-S_1-S_2). Les paramètres sont classés en fonction de leur S_T et les paramètres les plus influents sont indiqués par la zone colorée en rose, qui comprend les paramètres constituant 95 % de la variance totale de la sortie du modèle.

La Figure 3b montre que quelle que soit la concentration de COT, la concentration en oxygène dissous est principalement contrôlée par le taux de mortalité bactérienne ($mort_{hb}$), le taux maximal

de croissance bactérienne ($\mu_{max,hb}$) et le rendement de croissance bactérien (Y_{hb}). En augmentant la concentration de COT, on observe une diminution progressive de l'influence de la mortalité, mais une augmentation de l'influence du rendement de croissance.

Le paramètre de la biodégradabilité du COD (b_1) fait également partie des paramètres influents. Étant donné que l'influence des paramètres de mortalité et de croissance contient une partie d'interactions complexes (sensibilité de second ordre s_2), ces dernières ont tendance à masquer l'influence des autres paramètres. Par conséquent, ils sont regroupés sous un nouveau terme appelé taux de croissance net (NG = croissance - mortalité), et une analyse affinée est réalisée afin d'examiner plus en détail le rôle des autres paramètres dans des conditions de forte et faible croissance nette.

4.1 Cas d'une forte croissance nette

En condition de croissance nette élevée, Y_{hb} et CODB contrôlent la qualité de l'eau en période de faible débit (Figure 4a). Cela est dû au fait que la communauté bactérienne consomme, en croissance nette élevée, une grande partie du CODB jusqu'à ce qu'il devienne un facteur limitant pour leur croissance. De plus, de très faibles interactions sont observées entre les paramètres, car la quasi-totalité de leur influence globale provient de leurs effets principaux ($S_T \approx s_1$ pour chaque paramètre), ce qui confirme que les interactions se produisent principalement en présence d'une condition de croissance nette élevée.

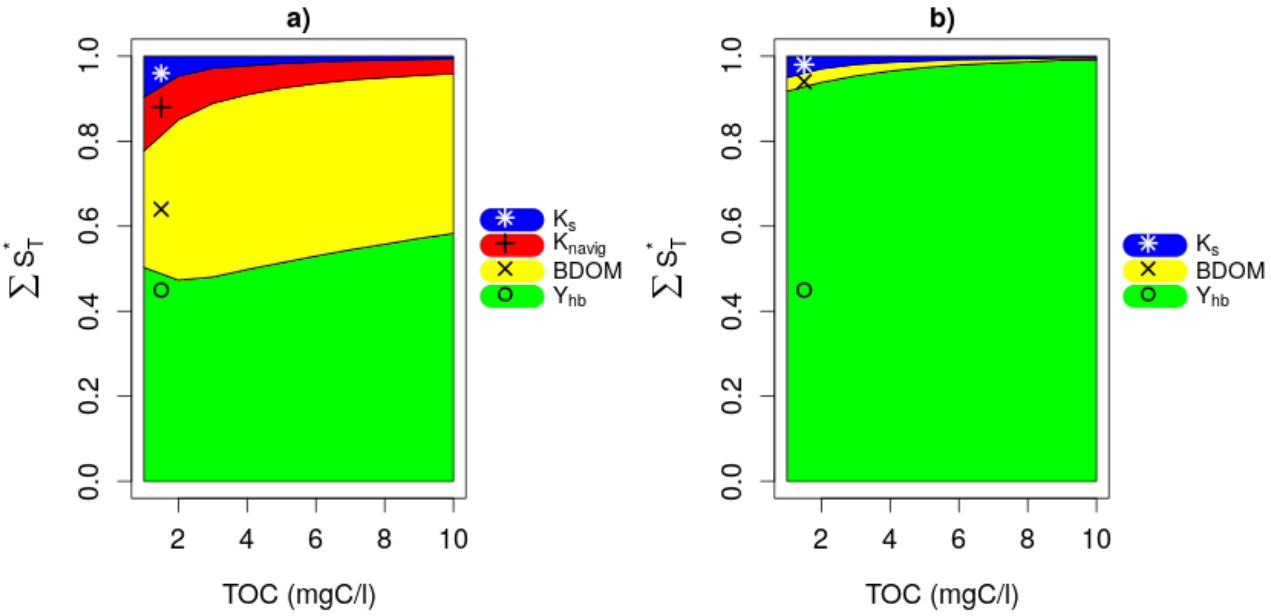


Figure 4: Résultats de l'analyse de sensibilité (a) Cas d'une forte croissance nette; (b) Cas d'une faible croissance nette

4.2 Cas d'une faible croissance nette

Y_{hb} est l'unique paramètre influent dans des conditions de faible croissance nette, et ce pour toutes les concentrations de COT (Figure 4b). En effet, le CODB n'est alors pas un facteur limitant de la croissance bactérienne qui est très faible par construction. Ainsi, seul le rendement de croissance influence la concentration en oxygène.

Le fait que le CODB soit influent pour une communauté bactérienne à forte croissance valide notre hypothèse de recherche selon laquelle les écarts dans les simulations de l'oxygène dissous sont dus à une sous estimation de la quantité de carbone biodégradable en amont du secteur modélisé, comme mis en lumière par Wang et al. [2022a], qui avaient déjà ajouté les paramètres bactériens influents tels que les taux de croissance, de mortalité, et le rendement au schéma d'assimilation de ProSe-PA, sans pour autant réduire significativement les écarts d'estimation de l'oxygène dissous en période d'étiage. En analysant les profils en long du CODB en Seine, les auteurs avaient alors constaté que le CODB présent dans le modèle était intégralement consommé par les bactéries, rendant impossible une augmentation de la respiration bactérienne.

Par conséquent, l'inclusion d'un paramètre b_1, qui représente la fraction de carbone organique dissous, dans le schéma d'assimilation de ProSe-PA est nécessaire.

5 Amélioration des simulations d'oxygène par intégration du CODB dans ProSe-PA sous un étude de cas synthétique

Le CODB est intégré dans le schéma d'assimilation de ProSe-PA parmi les paramètres assimilables du modèle (CODB = b_1 * t * COT). L'objectif de cette étude de cas synthétique est de vérifier si le CODB peut être identifié à partir des données de l'oxygène dissous dans différentes conditions environnementales.

5.1 Réflexions sur le positionnement des stations d'observation d'oxygène dissous pour l'assimilation de données

Tout d'abord, nous procédons à la détermination du positionnement idéal pour installer une station d'observation de l'oxygène dissous afin d'être en mesure d'inférer à la fois les paramètres de la physiologie des bactéries (taux de croissance maximal des bactéries et rendement) et la biodégradabilité du carbone organique dissous, par assimilation de ces données.

Un test est ainsi conduit visant à déplacer la station d'observation le long de la rivière, du point kilométrique 0 (kp0) au point kilométrique 100 (Figure 5). Les données sur l'oxygène dissous sont créées de manière synthétique dans cette étude en fixant les paramètres du modèle aux valeurs cibles, puis en bruitant les concentrations simulées avec un bruit blanc. Nous évaluons alors les performances de la procédure d'évaluation au regard de sa capacité à inférer les valeurs des paramètres physiques (distributions de probabilités centrées sur les valeurs ayant servi à générer les observations) et sur la racine carrée des écarts quadratiques moyens (RMSE) entre les valeurs d'oxygène simulées et les valeurs d'oxygène observées.

Les résultats sont analysés à la lumière des notions suivantes :

1. **Lieu d'observabilité**. Il s'agit du point en amont duquel aucun paramètre ne peut être détecté ou inféré par l'assimilation des données. En amont du lieu d'observabilité, le changement de la concentration en oxygène dissous, par rapport à la condition limite en amont est inférieur à l'erreur de mesure. L'emplacement de ce lieu (kp_{obs}) est déterminé à l'aide de la formule suivante :

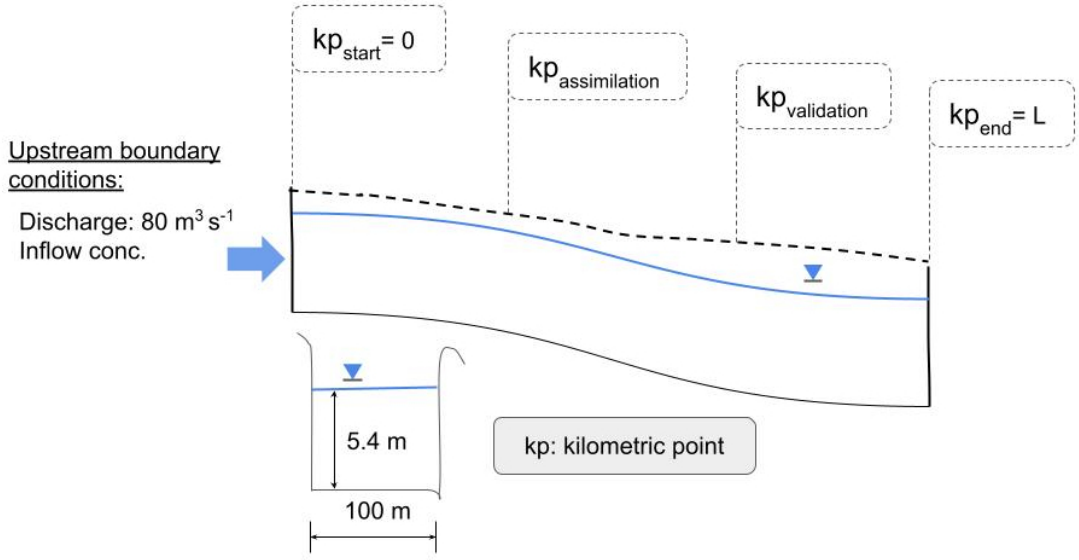


Figure 5: Schéma synthétique représentant un tronçon de L km de la Seine. kp_{start}: point kilométrique du point de départ du schéma synthétique. $kp_{assimilation}$: point kilométrique de la station de surveillance synthétique de l'oxygène dissous utilisée pour l'assimilation des données. $kp_{validation}$: point kilométrique de la station de surveillance de l'oxygène dissous utilisée pour l'évaluation des sorties du modèle. kp_{end} : point kilométrique de la fin du schéma synthétique.

$$kp_{obs} = vitesse \times t_{obs} \tag{1}$$

$$t_{obs} = \frac{\beta \times \sigma}{(1 - Y_{hb})\frac{\mu}{Y_{hb}} \times [HB] \times \frac{32}{12}} \tag{2}$$

Où μ est le taux de croissance effectif des bactéries, Y_{hb} le rendement de croissance des bactéries hétérotrophes, t_{obs} le temps nécessaire pour que le carbone organique atteigne le point kilométrique (pk) qui provoquerait une baisse de la concentration en oxygène dissous supérieure à l'erreur de mesure, β le score Z et est fixé à 1,96 pour un intervalle de confiance gaussien de 95%, et σ l'écart-type de l'erreur d'observation (1% de la concentration d'oxygène dissous saturée).

2. **Lieu de limitation**. Il se trouve à la distance minimale de la source de carbone à partir de laquelle la concentration du carbone organique dissous rapidement biodégradable (ou substrat, S) limite la croissance des bactéries.

3. **Lieu d'épuisement**. Il s'agit de la distance depuis la source de carbone organique où la concentration du substrat atteint une valeur critique telle que la croissance bactérienne ne peut plus compenser le taux de mortalité bactérienne ($\mu_{hb} = mort_{hb}$).

Les résultats de l'expérience confirment l'identification de ces trois lieux (Figure 6). Les valeurs de la RMSE sont initialement élevées avant d'atteindre le point d'observabilité, où aucun paramètre du modèle ne peut être inféré. Ensuite, la RMSE diminue dans la fenêtre d'observabilité-limitation, où

les paramètres de physiologie bactérienne sont récupérables. La RMSE reste également faible dans la fenêtre de limitation-épuisement, où les trois paramètres peuvent être inférés. Cependant, la valeur de RMSE augmente à nouveau en aval du point d'épuisement, où seul le paramètre de biodégradabilité du carbone organique peut être identifié.

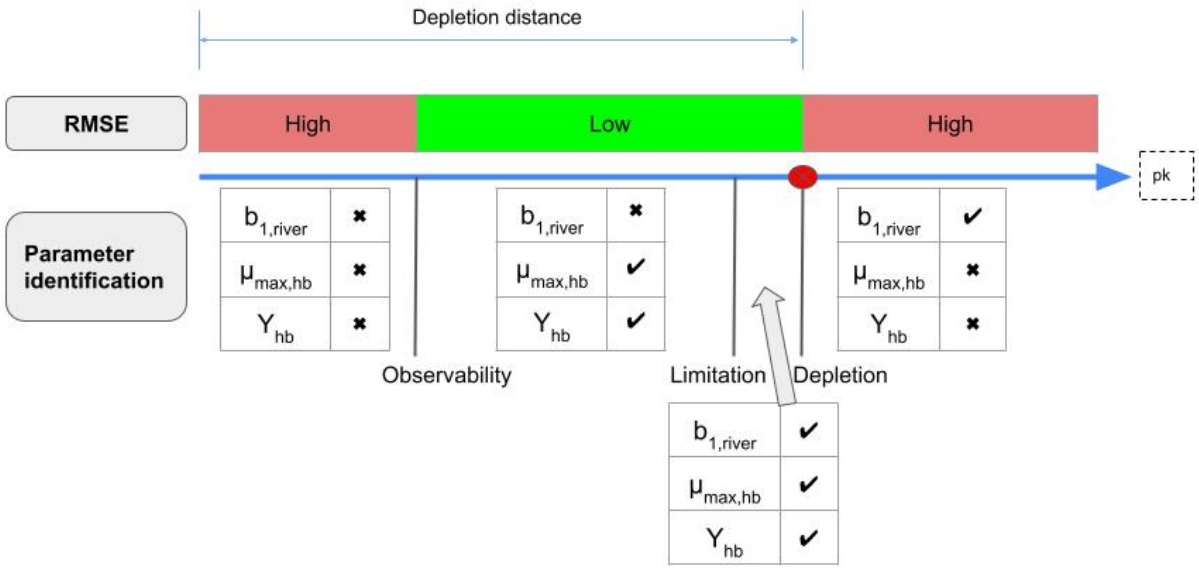


Figure 6: Résultats du déplacement de la station d'assimilation entre kp10 et kp90 : évolution de la RMSE à la station de validation (kp35) et de l'inférence des paramètres en fonction du pk de la station d'assimilation.

La position du point d'observabilité (kp_{obs}) est déterminée à une distance de 16 km de la source de carbone organique, en accord avec la valeur théorique calculée (17 km) (Eq. 2). L'absence de paramètres identifiables en amont de ce point est attribuée à une activité bactérienne insuffisante, qui ne parvient pas à induire une diminution significative des niveaux d'oxygène dissous. Par conséquent, ProSe-PA ne peut pas, dans ce secteur, inférer les paramètres bactériens à l'œuvre à partir de la série temporelle d'oxygène dissous. En revanche, en aval de ce point, $\mu_{max,hb}$ et Y_{hb} sont inférés car le changement d'oxygène dissous est plus important que l'erreur de mesure.

En ce qui concerne le point de limitation, les résultats montrent qu'il se situe entre les pk 38 et 39 pour une fraction de carbone organique dissous de 0,25, conduisant à une concentration limite du substrat ($S_{limiting}$) de 0,22 $mgCL^{-1}$.

De plus, la valeur critique marquant le point d'épuisement du substrat est déterminée à 0,01 mgC L^{-1}. Dans cette expérience, le point d'épuisement est détecté au pk 43. Cependant, la position du point d'épuisement augmente avec l'augmentation de b_1. Au-delà de ce point, seul b_1 peut être estimé avec précision, tandis que $\mu_{max,hb}$ et Y_{hb} présentent des densités de probabilité plutôt uniformes. Cela est dû au fait que, quelle que soit leur valeur réelle, le taux de croissance net des bactéries est négatif lorsque la concentration en substrat tombe en dessous de la valeur critique en raison de la prédominance de la mortalité sur le taux de croissance.

Afin de récupérer les trois paramètres, la station d'assimilation doit être située dans une fenêtre de limitation-épuisement, qui est généralement étroite. Cependant, la position de la fenêtre se déplace avec les changements de b_1 ou d'autres paramètres. Une stratégie à une seule station n'est donc pas

adaptée à la surveillance environnementale, car les valeurs des paramètres changent constamment dans des conditions naturelles.

5.2 Une stratégie à deux stations pour l'inférence du changement de biodégradabilité du carbone organique dissous (CODB)

Une approche optimale consisterait alors à utiliser une station d'assimilation dans la fenêtre d'observabilité-limitation pour détecter $\mu_{max,hb}$ et Y_{hb}, et une autre en aval du point d'épuisement pour détecter b_1, permettant ainsi la détection des trois paramètres avec un niveau acceptable de RMSE. Pour évaluer l'efficacité de cette stratégie, un test d'assimilation des données a été réalisé pour vérifier si nous pouvions identifier le changement de biodégradabilité du carbone organique dissous de 0,2 à 0,4.

La performance de ProSe-PA est évaluée sous des conditions de changement brusque et de changement progressif de la biodégradabilité du carbone organique dissous (Figure 7). ProSe-PA détecte avec succès les valeurs des paramètres du modèle dans les deux cas, comme en témoigne la concordance des valeurs cibles (ligne rouge) avec les densités de probabilités *a posteriori* de chaque paramètre. Cependant, on observe un retard de 5 à 6 jours dans la détection du changement lui-même, qui peut être attribué au temps nécessaire au carbone organique pour se déplacer depuis sa source jusqu'à la deuxième station d'observation située à 70 km dans la fenêtre d'épuisement. Avec une vitesse d'écoulement de 0,18 m/s, on estime que le carbone organique met environ 5 jours pour parcourir une distance de 70 km. Un retard similaire est également observé au début de la simulation (jours 1-5).

Dans le cas d'un changement abrupt de la biodégradabilité, la densité de probabilités *a posteriori* de b_1 présente également un changement brusque le jour 14 (Fig. 7(i)A), tandis que pour un changement progressif, la densité de probabilités *a posteriori* de b_1 change progressivement à partir du jour 14 (Fig. 7(ii)A).

Ainsi, pour la première fois, nous sommes en mesure d'estimer la biodégradabilité du carbone organique dissous dans les apports modélisés grâce aux données de suivi à haute fréquence de l'oxygène dissous. De plus, l'inclusion du paramètre b_1 améliore les simulations de l'oxygène dissous et aide à identifier les paramètres de croissance des bactéries.

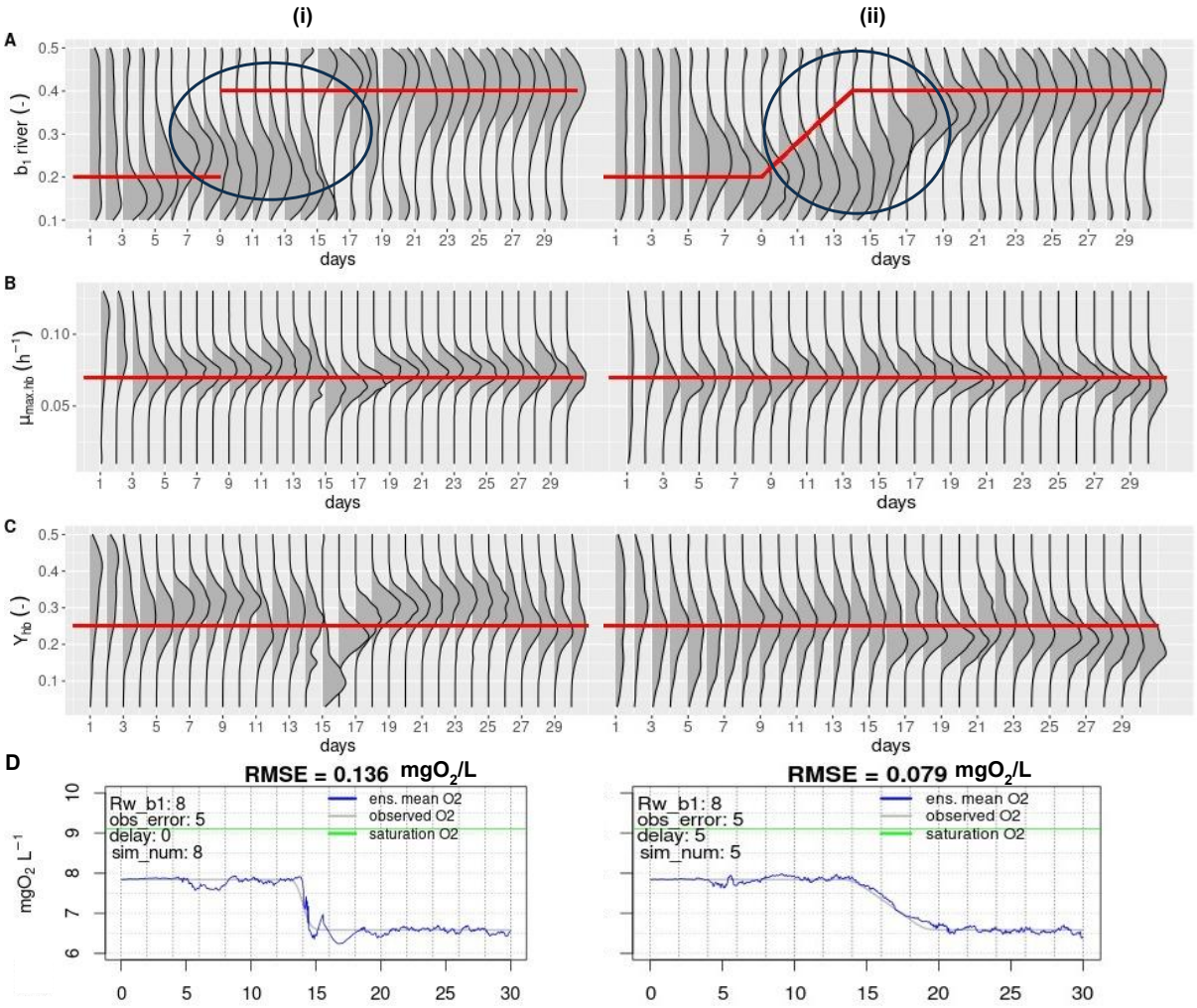


Figure 7: Inférence des paramètres cibles tout en détectant le changement de biodégradabilité du carbone organique dissous à partir des conditions aux limites avec un pas de temps d'une heure. (i) Changement brusque de b_1. (ii) Changement progressif de b_1. Les lignes rouges indiquent les valeurs cibles des paramètres à récupérer par l'assimilation des données. A, B et C) Distributions *a posteriori* de b_1, de taux de croissance et de rendement. D) Moyenne d'ensemble de l'oxygène dissous simulée par ProSe-PA par rapport à l'oxygène dissous synthétique observé à la station de validation.

6 Amélioration des simulations d'oxygène dissous dans la Seine grâce à l'inférence de la biodégradabilité du carbone organique dissous

Dans cette dernière section, les développements de ProSe-PA, notamment le nouveau modèle de répartition du carbone organique et l'ajout d'un nouveau paramètre assimilable représentant sa biodégradabilité (b_1), sont testés en situation réelle en Seine pour l'étiage 2011.

6.1 Zone d'étude et période d'étude

Le scénario réel examiné dans cette étude englobe une étendue de 220 km entre l'amont de la ville de Paris et l'estuaire de la Seine à Poses (Fig. 8), où les rivières Seine et Marne servent de conditions limites en amont, tandis que les affluents, tels que l'Oise, sont considérés comme des apports latéraux

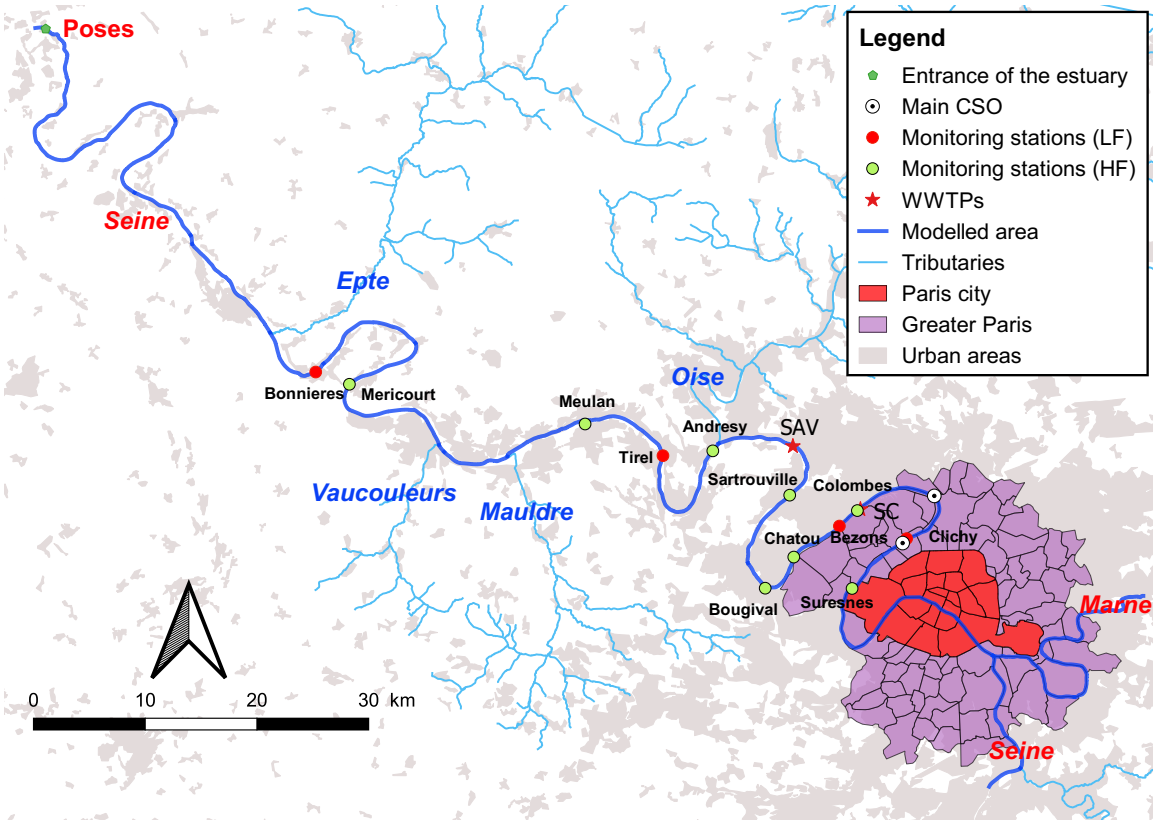


Figure 8: Tronçon de la Seine modélisé dans cette étude. Les huit stations de surveillance à haute fréquence (cercles verts) de l'amont vers l'aval : 1. Suresnes, 2. Colombes, 3. Chatou, 4. Bougival, 5. Sartrouville, 6. Andrésy, 7. Meulan, 8. Méricourt. Deux stations d'épuration (étoiles rouges) : Seine Aval (SAV) et Seine Centre (SC). Deux déversoirs d'orage principaux (cercles noirs) : Clichy et Garges-Epinay. (Tiré de Wang et al. [2022a]).

dans le modèle.

Le bassin de la Seine est à la fois industriel, agricole et fortement urbanisé, abritant 25 % de la population française [Flipo et al., 2020], et est donc soumis à d'importantes charges de carbone organique urbain provenant de cinq stations d'épuration des eaux usées (STEP). De plus, les eaux pluviales excédentaires pendant les événements de pluie sont évacuées dans la Seine via 156 déversoirs d'orage [Seidl, 1997, Seidl et al., 1998b, Even et al., 2007], dont 99 % se trouvent en amont de la station de surveillance de Bougival. Huit stations de surveillance fournissent des données sur l'oxygène dissous utilisées à des fins d'assimilation dans cette étude.

Pendant l'année 2011, une année sèche, la Seine a connu des conditions de débit faible pendant la majeure partie de l'année (Fig. 9A), ce qui a entraîné des concentrations d'oxygène dissous inférieures aux niveaux de saturation, indiquant un niveau élevé d'activité bactérienne hétérotrophe (Fig. 9B).

La période de faible débit en 2011 se situe entre les jours 100-340, coïncidant avec l'apparition de trois proliférations algales successives entre les jours 120-200 [Vilmin et al., 2016]. Ainsi, pour les besoins de cette étude, la période de débit faible est délimitée aux jours sans prolifération algale, c'est-à-dire entre les jours 200-300, au cours desquels des diminutions substantielles des niveaux d'oxygène sont discernables d'amont en aval. De plus, une période particulièrement sèche est également observée au cours de cette période de débit faible entre les jours 260 et 280.

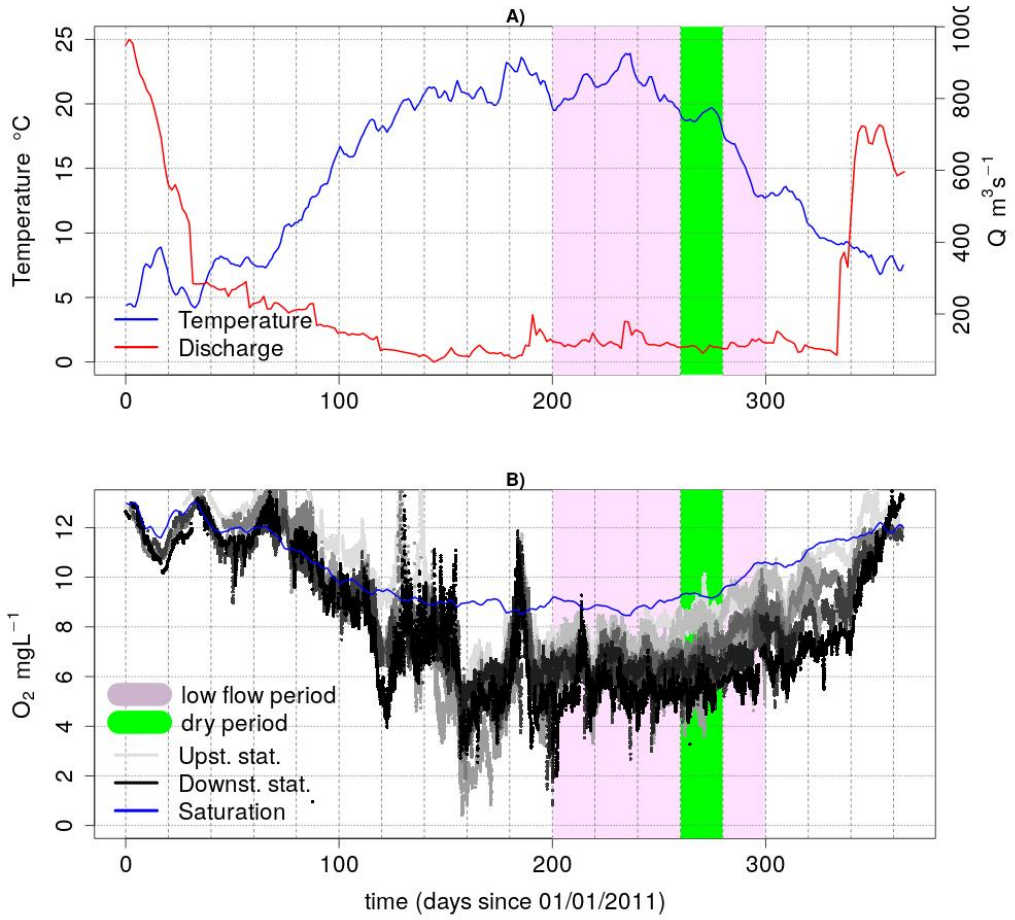


Figure 9: Débit, température et oxygénation de la Seine en 2011. Les polygones surlignés indiquent la période d'étiage (violet) et la période sèche (vert). A) Évolution de l'oxygène dissous aux huit stations de suivi (gradient de couleur : du gris au noir représentant les stations d'amont en aval). B) Débit journalier à la station de jaugeage et température à une station de surveillance.

6.2 Conception de scénarios d'assimilation des données pour tester le rôle du carbone organique dissous biodégradable

Le cas synthétique discuté dans les précédentes sections ne comprenait qu'une seule source de carbone, tandis que l'étude de cas réel présentée ici en fait intervenir plusieurs. Par conséquent, il est nécessaire de décider pour quelle source de carbone nous inférons b_1. Tout d'abord, nous avons identifié les principales sources en fonction de leur contribution au flux de carbone total, à savoir la station d'épuration des eaux usées de Seine-Aval, la Seine, l'Oise, la Marne, la station d'épuration des eaux usées de Seine-Amont, la station d'épuration des eaux usées de Seine-Centre et le déversement de temps sec et d'orage de Garche-Epinay (ancienne rivière). Compte tenu de la différence de biodégradabilité entre les stations d'épuration et les rivières en amont, b_1 est destiné à être inféré uniquement pour les rivières.

Pour une meilleure illustration du rôle du paramètre b_1, deux simulations d'assimilation des données sont lancées : une avec une part fixe de carbone dissous biodégradable (b_1 fixe) et l'autre avec une part variable de carbone dissous biodégradable (b_1 variable).

Les concentrations en oxygène dissous simulées par le scénario du carbone biodégradable fixe et le cas du carbone biodégradable variable sont comparées avec celles observées dans les huit stations de

surveillance en terme de RMSE.

6.3 Amélioration de la simulation de l'oxygène dissous par assimilation des données avec CODB variable

Les deux scénarios produisent des simulations en oxygène dissous acceptables au cours des 50 premiers jours. Cependant, au-delà du jour 250, le cas d'un ratio CODB:COD fixe présente des divergences importantes par rapport à l'oxygène observé, tandis que l'assimilation des données avec CODB:COD variable montre des performances remarquables, à l'exception des stations de Colombes et Bougival.

Les améliorations dues à l'assimilation des données avec inférence de la biodégradabilité sont confirmées en calculant la RMSE des deux scénarios pour la période de 250-300 (Tableau 3). Les résultats indiquent que l'assimilation des données avec une part variable de carbone biodégradable réduit significativement la RMSE à six des huit stations. La diminution la plus substantielle de la RMSE est observée à la station de Meulan, où la RMSE passe de 0,72 à 0,41 mgO_2L^{-1}, ce qui représente une réduction de 43 % des erreurs de simulation.

Table 3: RMSE des scénarios d'assimilation de données avec ratio de biodégradabilité du carbone organique dissous fixe et variable aux les stations de surveillance

Stations	CODB:COD fixe	CODB:COD variable	Variation de RMSE (%)
Suresnes	0.66	0.58	-12
Colombes	0.67	0.9	34
Chatou	1.62	1.31	-19
Bougival	0.45	0.48	7
Sartrouville	1.3	1.02	-22
Andresy	0.52	0.43	-17
Meulan	0.72	0.41	-43
Mericourt	1.19	0.71	-40

Les observations de l'oxygène dissous à la station de Meulan montrent ainsi une correspondance quasi parfaite avec la simulation du scénario du CODB variable (Figure 10 (i)).

En termes de réduction absolue de la RMSE, la diminution la plus significative est observée à la station de Méricourt, où elle diminue de 0,48 mgO_2L^{-1} (passant de 1,19 à 0,71 mgO_2L^{-1}, ce qui représente une réduction de 40 %). Ces résultats confirment l'hypothèse de Wang et al. [2022a], qui avait considéré le manque de CODB provenant des conditions aux limites comme une possible cause de la surestimation de l'oxygène dissous pendant les périodes de faible débit.

En revanche, des augmentations de la RMSE (34 % et 7 %) ne sont observées qu'à deux stations : Colombes et Bougival. Elle est presque négligeable à Bougival, d'environ 0,03 mgO_2L^{-1}. Cependant, à Colombes, l'augmentation de la RMSE pourrait être due à l'incertitude concernant le carbone organique provenant des nombreux déversoirs d'orage dans la partie en amont des secteurs modélisés.

En ce qui concerne l'inférence des paramètres du modèle, la croissance maximale des bactéries et le rendement des bactéries restent stables pendant cette période et sont bien identifiés dans les deux cas d'assimilation des données, avec une plage estimée de 0,04 à 0,06 h^{-1} dans le cas du carbone biodégradable fixe et de 0,02 à 0,04 h^{-1} dans le cas du carbone biodégradable variable.

En ce qui concerne b_1, sa valeur augmente à partir du jour 260 vers sa limite supérieure jusqu'à la fin de la période sèche (jour 280), pendant laquelle ProSe-PA infère les valeurs les plus élevées

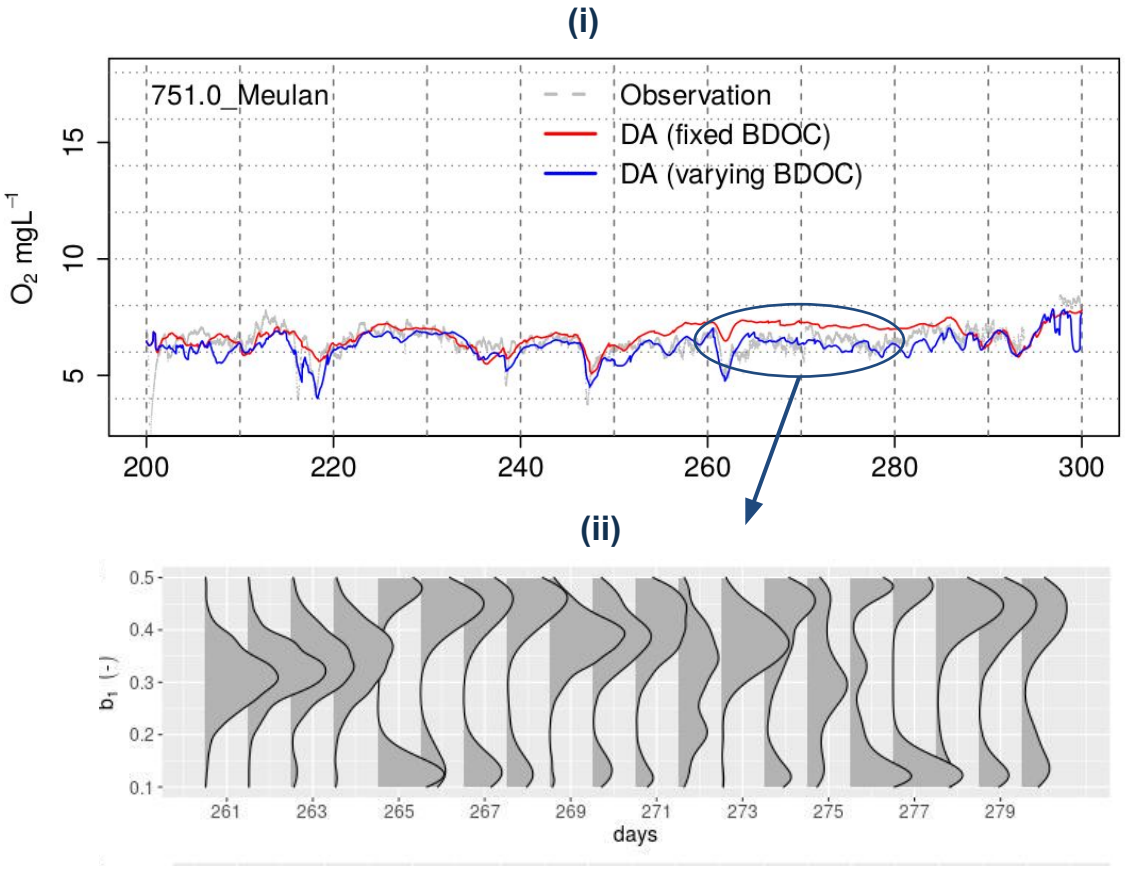


Figure 10: (i) Concentrations en oxygène dissous à la stations de Meulan simulées par assimilation des données avec CODB:COD fixe (ligne rouge) et avec CODB:COD variable (ligne bleue) par rapport à celles observées (ligne grise). (ii) Distribution a posteriori du paramètre b_1 pendant la période sèche inférée par l'assimilation des données avec CODB:COD variable. b_1 : ratio CODB:COD.

de b_1 valant 0,45. Ces valeurs sont cohérentes avec les conclusions de Varrault et al. [2016] pour la Seine pendant la période de faible débit de 2011. L'augmentation de b_1 signifie une augmentation de la proportion de carbone organique dissous biodégradable, ce qui se traduit par une amélioration significative de l'oxygène dissous simulé à plusieurs stations, notamment Chatou, Meulan et Méricourt. Cela confirme également l'hypothèse de Wang et al. [2022a] qui attribuent la surestimation des niveaux d'oxygène dissous pendant les périodes de faible débit dans la Seine à l'incertitude concernant le CODB.

De plus, l'assimilation des données avec CODB:COD variable fonctionne mieux pendant la période sèche (jours 260-280) (Figure 10 (ii)), ce qui peut être dû au fait que le métabolisme de la rivière est moins influencé par les déversements des déversoirs d'orage pendant de telles périodes. Cela permet alors au carbone provenant des sources principales d'avoir un impact plus important sur l'évolution de l'oxygène dissous, ce qui donne à ProSe-PA la possibilité d'estimer la biodégradabilité du COD de manière plus précise et d'améliorer les simulations. En dehors de la période sèche, les performances sont réduites, ce qui nécessite d'incorporer la biodégradabilité des déversoirs d'orage et des stations d'épuration des eaux usées, en plus de celle des affluents.

En ce qui concerne la représentation du CODB, les deux scénarios d'assimilation des données simulent des niveaux de carbone organique dissous similaires et se comportent de manière équivalente par

rapport au COD observé (Fig. 11). Cependant, l'assimilation des données avec un carbone biodégradable variable conduit à des concentrations plus élevées de carbone organique dissous biodégradable. Cette augmentation du CODB se traduit par des simulations d'oxygène améliorées tout en inférant le paramètre de biodégradabilité. Cependant, les valeurs du CODB devraient être comparées à celles obtenues dans l'étude de Varrault et al. [2016]. Si une sous-estimation est observée, cela signifierait que davantage de CODB doit entrer dans la rivière à partir d'autres sources telles que les déversoirs d'orage et les stations d'épuration des eaux usées, pour lesquelles un nouveau paramètre de biodégradabilité est nécessaire.

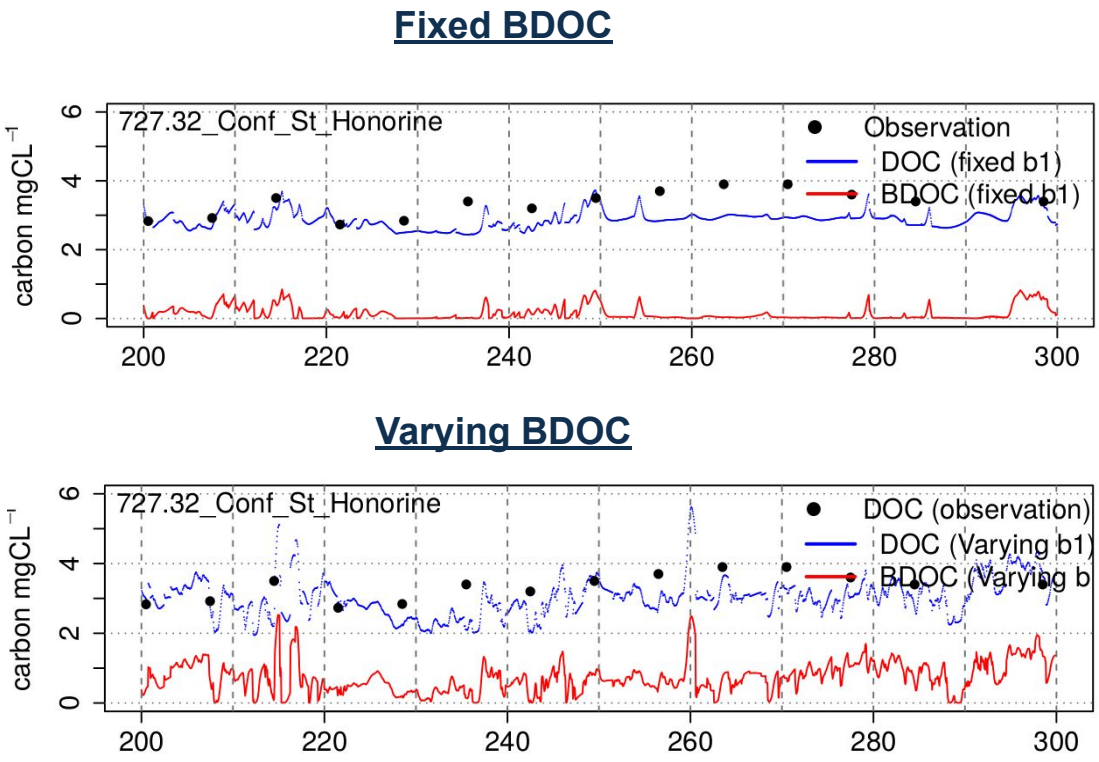


Figure 11: Simulations du COD et du CODB à Conflans-Saint-Honorine par l'assimilation des données avec CODB:COD fixe (en haut) et CODB:COD variable (en bas), et comparaison avec les concentrations en COD observées (cercles noirs).

L'incorporation du paramètre permettant d'estimer la fraction biodégradable du carbone organique dissous a ainsi non seulement amélioré la précision des simulations de l'oxygène dissous, mais a également facilité l'estimation de la biodégradabilité du carbone organique dissous en milieu naturel.

Conclusions

L'incertitude dans l'estimation du CODB génère des erreurs importantes sur la simulation de l'oxygène dissous à bas débit. L'inclusion d'une inférence dynamique de la fraction biodégradable du carbone organique dissous dans le schéma d'assimilation de données de ProSe-PA permet de significativement

améliorer la simulation de l'oxygène dissous et d'inférer les propriétés physiologiques des communautés bactériennes.

Dans cet objectif, une revue de la littérature a été réalisée, lors de laquelle nous avons précisé les gammes de variation des paramètres et montré que le climat et l'occupation du sol contrôlent la variabilité du carbone dans les milieux aquatiques.

Sur la base de l'analyse de sensibilité pour identifier les paramètres qui contrôlent le métabolisme de la rivière en période de faible débit, on a constaté que les paramètres bactériens contrôlent la qualité de l'eau à bas débit. Le CODB a également une influence sous des conditions d'activité bactérienne élevée. Il a donc été incorporé dans ProSe-PA et il a été prouvé que ProSe-PA est capable d'identifier la biodégradabilité du COD sans capteurs grâce aux données à haute fréquence de l'oxygène dissous, tant dans des études de cas synthétiques que réelles. Les emplacements optimaux pour la mise en place de stations de surveillance ont été identifiés. L'assimilation des données avec un CODB variable, facilitée par l'inclusion du paramètre de biodégradabilité du COD, a amélioré les simulations de l'oxygène et a contribué à l'inférence des paramètres du modèle.

En perspective, nous proposons la surveillance continue des communautés bactériennes et du CODB afin de valider les résultats de ProSe-PA, notamment pour étudier des situations extrêmes où le ratio de biodégradabilité du COD augmenterait dans des limites non encore documentées en milieux naturels.

Mots clés : modèle ProSe-PA, Assimilation de données, Qualité de l'eau, Métabolisme fluvial

Abstract

The main objective of this thesis is to improve the assessment of river metabolism by assimilating dissolved oxygen (DO) data in the ProSe-PA water quality model [Wang, 2019, Wang et al., 2022a] using synthetic and real case studies of the Seine river from upstream of Paris to its entry into the estuary. This study builds upon the pioneering work of Wang et al. [2019], who introduced the first particle filter in water quality modeling based on the previous calibration and validation of the hydro-biogeochemical processes occurring in the Seine river [Vilmin, 2014]. Despite improvements in oxygen simulation and the inference of physiological parameters governing micro-organism communities by Wang et al. [2019], discrepancies of about 1 mgO_2L^{-1} persisted during low-flow periods, where wastewater treatment plant outflows exert the highest pressure on the river system due to low dilution capacity.

The research hypothesis posits that uncertainties in organic carbon inflow from model boundary conditions cause the mismatch between observed and simulated DO. This research is carried out in four steps.

As a first step, a new partitioning scheme is introduced for organic carbon in ProSe-PA where 5 new parameters are proposed to distribute the total organic carbon into smaller fractions (model inputs). Then a detailed bibliography study is done to find out the variation range of the proposed parameters under different climate, land cover and hydrological conditions. This review gives access to information on variation of organic carbon variables such as dissolved organic carbon (DOC) and its biodegradable share, which is useful for the implementation of a Bayesian tool such as a particle filter.

In the second step, a Sobol' sensitivity analysis coupled with a dimension reduction method is conducted to determine the most influential parameters affecting DO evolution at low flow. Four sets of parameters are analyzed: bacterial parameters (e.g., maximum growth rate), physical parameters (e.g., navigation or wind constants), organic carbon partitioning parameters (e.g., biodegradable DOC percentage), and organic carbon degradation parameters (e.g., constant of substrate uptake). Results indicate that the maximum bacterial growth rate, growth yield, and biodegradability of DOC within 45 days dominate water quality under high bacterial net growth rate conditions (0.05 h^{-1}) at low flow.

The study's third step integrates the organic carbon partitioning model and its parameter for dissolved organic carbon biodegradability into the ProSe-PA model's data assimilation framework. A proof-of-concept test is performed to display the ProSe-PA model's capability to retrieve model parameters based on oxygen data assimilation. A synthetic case, which mimics the Seine river at summer low flow is devised whose results showcase the model's capacity to improve dissolved oxygen simulation and retrieves predefined parameters. Retrieval of parameters under varying environmental conditions necessitates the use of at least two assimilation stations. Additionally, the optimum perturbation percentages of the model parameters are configured to detect change in environmental conditions. Moreover, the model can identify both abrupt and gradual environmental shifts. This

study marks the first successful attempt to quantify organic carbon inflow in a river system, using a biogeochemical model and Bayesian framework with dissolved oxygen monitoring at two river stations.

As a final step, the performance of ProSe-PA model is evaluated in the real case of the Seine river, spanning a length of 220 km, during a low-flow period of the year 2011. The findings reveal improvements in the simulations of dissolved oxygen, concomitant with the quantification of the evolution of organic carbon biodegradability during summer low flow period.

Keywords : Prose-PA model, Water Quality, Data Assimilation, River metabolism

Contents

Acknowledgement **i**

Résumé: De l'oxygène au carbone organique dissous biodégradable à l'étiage par assimilation de données dans le modèle de simulation de la qualité des eaux ProSe-PA **iii**

1 Introduction . . . iii
2 Modèle conceptuel du carbone organique . . . vi
3 Revue bibliographique des paramètres de répartition du carbone organique . . . vii
4 Analyse de sensibilité pendant une période de faible débit estival . . . ix
4.1 Cas d'une forte croissance nette . . . x
4.2 Cas d'une faible croissance nette . . . x
5 Amélioration des simulations d'oxygène par intégration du CODB dans ProSe-PA sous un étude de cas synthétique . . . xi
5.1 Réflexions sur le positionnement des stations d'observation d'oxygène dissous pour l'assimilation de données . . . xi
5.2 Une stratégie à deux stations pour l'inférence du changement de biodégradabilité du carbone organique dissous (CODB) . . . xiv
6 Amélioration des simulations d'oxygène dissous dans la Seine grâce à l'inférence de la biodégradabilité du carbone organique dissous . . . xv
6.1 Zone d'étude et période d'étude . . . xv
6.2 Conception de scénarios d'assimilation des données pour tester le rôle du carbone organique dissous biodégradable . . . xvii
6.3 Amélioration de la simulation de l'oxygène dissous par assimilation des données avec CODB variable . . . xviii

Abstract **xxii**

List of Figures **xxix**

List of Tables **xxxiv**

1 Introduction **1**
1 General context . . . 1
2 A brief historical overview of water quality modeling: from static to dynamic representation . . . 2
3 Modeling and monitoring water Quality in the Seine River basin: advances in data assimilation and its challenges . . . 3
4 Thesis Objectives . . . 4
5 Thesis outline . . . 5

2 A review of variables representing the description and degradation of organic matter in world rivers under different climatic, hydrological, and land cover conditions 7
1 Introduction . . . 8
2 Material and methods . . . 11
2.1 Selection of OM variables for review . . . 11
2.1.1 OM descriptive variables in water quality models . . . 11
2.1.2 OM degradation variables in water quality models . . . 12
2.2 Data collection on OM variables . . . 13
2.2.1 Hydrological categories . . . 13
2.2.2 Land cover categories . . . 13
2.2.3 Climatic categories . . . 14
2.3 Factorial analysis of mixed data (FAMD) . . . 21
3 Results . . . 23
3.1 Overall average values and variation ranges of OM variables . . . 23
3.2 Identification of clusters among OM variables . . . 23
3.3 Average values and variation ranges per cluster . . . 25
4 Discussion . . . 26
4.1 Comparison of the FAMD results and the proposed variation ranges with the literature . . . 26
4.2 Uncertainties in characterizing OM variable values . . . 29
4.3 New direction accounting for photodegradation in water quality modeling . . . 31
4.4 The role of the qualitative approach in OM characterization: towards quantitative estimates . . . 33
5 Conclusions . . . 34

3 How much do bacterial growth properties and biodegradable dissolved organic matter control water quality at low flow? 36
1 Introduction . . . 38
2 Material and methods . . . 40
2.1 C-RIVE Biogeochemical model . . . 40
2.1.1 OM partitioning parameters: from total organic carbon to six OM fractions . . . 41
2.1.2 Parameters for SA and their variation ranges . . . 42
2.2 Case study . . . 42
2.3 Sensitivity analysis strategy . . . 44
2.3.1 Pre-analysis of the model with extreme parameter values . . . 44
2.3.2 Understanding river metabolism controls with multiple sensitivity analyses . . . 46
3 Results . . . 49
3.1 First SA: All parameters . . . 49
3.2 Second SA under high bacteria net growth rate . . . 51
3.3 Third SA under low bacteria net growth rate . . . 51
4 Discussion . . . 52
4.1 Hierarchy of the most influential parameters on DO during low flow period . . . 52
4.2 Limitations of the sensitivity analysis . . . 52
4.3 Consequences of the results on water quality monitoring in urban areas . . . 53
4.4 Consequences of the results on data assimilation (DA) in water quality modeling 54
5 Conclusions . . . 54

4 Quantifying heterotrophic bacteria parameters and dissolved organic carbon biodegradability through oxygen data assimilation in a river water quality model 56

1 Introduction . . . 58
2 Materials and Methods . . . 59
2.1 ProSe-PA . . . 60
2.1.1 Organic carbon repartition model . . . 61
2.1.2 Selection of model parameters for data assimilation . . . 62
2.2 Data assimilation in ProSe-PA using particle filtering . . . 63
2.3 Case study . . . 64
2.4 Design of data assimilation experiments . . . 66
2.4.1 Impact of DOC biodegradability on data assimilation . . . 66
2.4.2 Positioning of the assimilation station . . . 67
2.4.3 Detection of environmental changes . . . 68
2.5 Numerical settings and evaluation criteria . . . 71
3 Results and Discussion . . . 71
3.1 How does the information about the biodegradability of dissolved organic carbon impact oxygen data assimilation in a river water quality model? . . . 72
3.1.1 Inference of bacterial growth and yield only . . . 72
3.1.2 Inference of the biodegradability of dissolved organic carbon . . . 73
3.2 Optimal location of two assimilation stations in a river network . . . 76
3.2.1 Single station strategy . . . 76
3.2.2 A two station strategy for the detection of all parameters . . . 78
3.3 How to configure the data assimilation framework for detecting environmental changes? . . . 80
3.3.1 Detection of change in bacteria physiology . . . 80
3.3.2 Detection of change in DOC biodegradability . . . 81
4 Conclusions . . . 83

5 Improvement of high frequency oxygen data assimilation in river systems by incorporation of dissolved organic carbon biodegradability during low-flow periods 85
1 Introduction . . . 86
2 Study area and study period: Seine River during 2011 . . . 87
2.1 Situation of Seine River during low-flow periods of 2011 . . . 88
2.2 Model input data . . . 88
3 Developments in the source code of ProSe-PA to control organic matter inflow of model boundary conditions . . . 90
4 Data assimilation strategy for low-flow period . . . 92
4.1 Parameters considered for data assimilation . . . 93
4.2 Major carbon sources amongst boundary conditions . . . 93
4.3 Design of data assimilation experiments . . . 94
4.4 Numerical settings . . . 95
5 Results and Discussion . . . 96
5.1 Improved simulations of DO concentration thanks to the inference of the biodegradability of dissolved organic carbon . . . 96
5.2 Characterization of the evolution of model parameters . . . 97
5.3 BDOC concentration increases in DA with inference of biodegradability . . . 103
5.4 Data assimilation with two bacteria communities lead to stable characterization of model parameters . . . 104
5.5 Strategy for further enhancement of data assimilation simulations at low-flow . 109
6 Conclusion . . . 110

Conclusion and perspectives 112

List of publications 117

Appendices **119**

A Supplementary results of FAMD **120**
A1 Results of FADM for DOC, t and K_S variables . . . 121

B Overview of oxygen and organic matter related equations and parameters in C-RIVE **125**
B1 DO and OM processes in C-RIVE . . . 126
B1.1 Dissolved oxygen evolution equations . . . 126
B1.2 Organic matter degradation equations . . . 127
B2 Parameterization of organic matter partitioning and degradation . . . 128
B2.1 OM degradation parameters . . . 128
B2.2 OM partitioning parameters . . . 128

C Sensitivity analysis protocol **130**

D Detailed results of experiments for detecting change in model parameters **134**
D1 Monod curve . . . 135
D2 Detailed results of the different combinations of DA configuration parameters for detecting change in bacteria physiology and dissolved organic carbon biodegradability . . . 135

E ProSe-PA User Guide **138**
E1 Get Started . . . 140
E1.1 Installation GSL - GNU Scientific Library . . . 140
E1.2 Installation en mode utilisateur simple . . . 140
E1.3 Installation en mode développeur . . . 140
E1.4 Compilations spéciales pour parallélisation et assimilation de données . . . 141
E1.5 Lancement de ProSe . . . 142
E1.6 Workflow for launching simulations in ProSe . . . 142
E2 Définition des dossiers contenant les fichiers d'entrée du modèle . . . 143
E3 Paramétrisation générale . . . 147
E4 Simulation hydraulique . . . 149
E4.1 Paramétrisation pour hydraulique . . . 149
E4.2 Fichier d'initialisation de l'hydraulique . . . 150
E4.3 Description du réseau hydrographique . . . 150
E4.3.1 Définition des singularités . . . 150
E4.3.2 Définition des biefs . . . 152
E4.3.3 Profils en travers . . . 152
E4.4 Trajectoire du centre des faces . . . 153
E4.5 Apports . . . 154
E4.6 Prise en compte de la courbure . . . 155
E5 Simulation de transport et biogéochimique . . . 155
E5.1 Les fichiers de paramètres des espèces . . . 155
E5.2 Parametrization of the macrospecies . . . 165
E5.3 Paramétrisation pour transport des dissous . . . 167
E5.4 Paramétrisation pour transport de chaleur et l'échange atmosphérique . . . 167
E5.5 Paramétrisation pour biogéochimie . . . 169
E5.6 Les fichiers de description des compartiments . . . 169
E5.7 Les paramètres liés à la sédimentation-érosion . . . 171
E5.8 Les forçages météorologiques . . . 171
E5.9 Définition des intervalle des paramètres à assimiler . . . 172
E6 Définition des fichier de sorties . . . 173
E6.1 Sorties PKs, final states . . . 180

E6.2 Profils ponctuels 181
E6.3 Profils en long 181
E6.4 Sortie tube hydraulique 181
E6.5 Bilans hydraulique 182
E6.6 Sorties graphiques hydraulique 183
E6.7 Bilans biogéochimiques 183
E6.8 Bilan énergétique 184

Bibliography **185**

List of Figures

1 Répartition de la matière organique à partir du COT et des cinq paramètres t, b_1, s_1, b_2, s_2. COD : carbone organique dissous, COP : carbone organique particulaire, CODB : COD biodégradable, CODR : COD réfractaire, BDOC : COP biodégradable, COPR : COP réfractaire, $BDOC_1$ et $BDOC_2$: fractions de COD rapidement et lentement biodégradables, $BPOC_1$ et $BPOC_2$: fractions de COP rapidement et lentement biodégradables. vi

2 Résultats de l'AFDM pour COT. A) Coordonnées des facteurs dans le nouvel espace (dimensions 1 et 2). B) Représentation graphique des catégories de facteurs qualitatifs sur les dimensions 1 et 2 en fonction de leur contribution (en %) aux axes correspondants. Les catégories affichées en lettres capitales sont les groupes identifiés. viii

3 Analyse de sensibilité de Sobol pour tous les paramètres : (a) indices de sensibilité pour COT = 5 $mgC.L^{-1}$; (b) évolution des indices de sensibilité totale normalisés des paramètres influents avec le carbone organique total ix

4 Résultats de l'analyse de sensibilité (a) Cas d'une forte croissance nette; (b) Cas d'une faible croissance nette x

5 Schéma synthétique représentant un tronçon de L km de la Seine. kp_{start}: point kilométrique du point de départ du schéma synthétique. $kp_{assimilation}$: point kilométrique de la station de surveillance synthétique de l'oxygène dissous utilisée pour l'assimilation des données. $kp_{validation}$: point kilométrique de la station de surveillance de l'oxygène dissous utilisée pour l'évaluation des sorties du modèle. kp_{end} : point kilométrique de la fin du schéma synthétique. xii

6 Résultats du déplacement de la station d'assimilation entre kp10 et kp90 : évolution de la RMSE à la station de validation (kp35) et de l'inférence des paramètres en fonction du pk de la station d'assimilation. xiii

7 Inférence des paramètres cibles tout en détectant le changement de biodégradabilité du carbone organique dissous à partir des conditions aux limites avec un pas de temps d'une heure. (i) Changement brusque de b_1. (ii) Changement progressif de b_1. Les lignes rouges indiquent les valeurs cibles des paramètres à récupérer par l'assimilation des données. A, B et C) Distributions *a posteriori* de b_1, de taux de croissance et de rendement. D) Moyenne d'ensemble de l'oxygène dissous simulée par ProSe-PA par rapport à l'oxygène dissous synthétique observé à la station de validation. xv

8 Tronçon de la Seine modélisé dans cette étude. Les huit stations de surveillance à haute fréquence (cercles verts) de l'amont vers l'aval : 1. Suresnes, 2. Colombes, 3. Chatou, 4. Bougival, 5. Sartrouville, 6. Andrésy, 7. Meulan, 8. Méricourt. Deux stations d'épuration (étoiles rouges) : Seine Aval (SAV) et Seine Centre (SC). Deux déversoirs d'orage principaux (cercles noirs) : Clichy et Garges-Epinay. (Tiré de Wang et al. [2022a]). xvi

9 Débit, température et oxygénation de la Seine en 2011. Les polygones surlignés indiquent la période d'étiage (violet) et la période sèche (vert). A) Évolution de l'oxygène dissous aux huit stations de suivi (gradient de couleur : du gris au noir représentant les stations d'amont en aval). B) Débit journalier à la station de jaugeage et température à une station de surveillance. xvii
10 (i) Concentrations en oxygène dissous à la stations de Meulan simulées par assimilation des données avec CODB:COD fixe (ligne rouge) et avec CODB:COD variable (ligne bleue) par rapport à celles observées (ligne grise). (ii) Distribution a posteriori du paramètre b_1 pendant la période sèche inférée par l'assimilation des données avec CODB:COD variable. b_1 : ratio CODB:COD. xix
11 Simulations du COD et du CODB à Conflans-Saint-Honorine par l'assimilation des données avec CODB:COD fixe (en haut) et CODB:COD variable (en bas), et comparaison avec les concentrations en COD observées (cercles noirs). xx

2.1 OM partitioning into its constituent fractions in terms of the five partitioning ratios, namely, t, b_1, s_1, b_2, s_2. DOC: dissolved organic carbon, POC: particulate organic carbon, BDOC: biodegradable DOC, RDOC: refractory DOC, BPOC: biodegradable POC, RPOC: refractory POC, $BDOC_1$ & $BDOC_2$: rapidly and slowly biodegradable fractions of DOC, $BPOC_1$ & $BPOC_2$ rapidly and slowly biodegradable fractions of POC 11
2.2 FAMD results for TOC. A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters. 24
2.3 FAMD results for b_1. A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters. 25
2.4 Distribution of the TOC individuals on dimensions 1 and 2 per factor with the confidence ellipses. The point size represents the average TOC value and the colors represent the different categories of the three qualitative factors (climate, hydrology, and land cover) 27
2.5 Distribution of the b_1 individuals on dimensions 1 and 2 with the confidence ellipses. The point size represents the average b_1 value and the colors represent the different categories of the three qualitative factors (climate, hydrology, and land cover) 29
2.6 Photodegradation model (photodissolution and photoflocculation); DOC: dissolved organic carbon, POC: particulate organic carbon, BDOC: biodegradable DOC, RDOC: refractory DOC, BPOC: biodegradable POC, RPOC: refractory POC, DIC:dissolved inorganic carbon, k_{flocc}: photoflocculation rate, k_{diss}: photodissolution rate, $k_{DOC->mineral}$: transformation rate of DOC to minerals, $k_{POC->mineral}$: transformation rate of POC to minerals, b_1^p: b_1 value of the photodegraded DOC, b_2^p: b_2 value of the photodegraded POC 32

3.1 Schematic description of the OM-related processes accounted for by C-RIVE in the water column where OM is partitioned into six fractions of dissolved and particulate matter using the five partitioning parameters, namely, t, b_1, s_1, b_2, s_2; POM: particulate organic matter; DOM: dissolved organic matter; BDOM:biodegradable DOM, BPOM:biodegradable POM (subscripts 1, 2, and 3 refer to rapidly degradable, slowly degradable, and non-biodegradable fractions of OM, respectively); Blue dashed-double dotted lines: OM input from sources and partitioning between POM and DOM; Solid black lines: partitioning of DOM and POM into biodegradability pools ; Dotted black lines: Hydrolysis; Remaining solid lines: Biogeochemical processes. Resp.:Respiration; Photo,:Photosynthesis; PP: primary producers; HB: heterotrophic bacteria; WWTP: waste water treatment plant; CSO: combined sewer overflow 41

3.2 Synthetic scheme representing a reach of the Seine river (modified from Wang et al. [2018]). . . . 44
3.3 Eight plots of single simulations; (XX yy ZZ) Any parameter name written in capital or small letters means that its maximum or minimum value is used, respectively, in that specific single simulation. For example, plot #3 (BDOM NG y) that corresponds to simulation #3 in Table 3.3 is a simulation where the maximum values of BDOM and net growth and the minimum value of Y_{hb} are used . . . 47
3.4 Sobol SA results of **first SA: All parameters** (a) Sobol SA results for TOC = 5 $mgCL^{-1}$; (b) Evolution of the normalized total sensitivity indices of the influential parameters with TOC . . . 50
3.5 Results of (a) Second SA: Fixed high net growth; (b) Third SA: Fixed low net growth 51

4.1 The organic carbon partitioning model which distributes the total organic carbon (TOC) among its constituent fractions by using five ratios, namely, t, b_1, s_1, b_2, s_2. DOC: dissolved organic carbon, POC: particulate organic carbon, BDOC: biodegradable DOC, $DOC_{refractory}$: refractory DOC, BPOC: biodegradable POC, $POC_{refractory}$: refractory POC, $BDOC_{rapid}$ & $BDOC_{slow}$: rapidly and slowly biodegradable fractions of DOC, $BPOC_{rapid}$ & $BPOC_{slow}$ rapidly and slowly biodegradable fractions of POC (modified from Hasanyar et al. [2023a]) . . . 62
4.2 Synthetic scheme representing a L km reach of the Seine river. kp_{start}: kilometric point of the starting point of synthetic scheme. $kp_{assimilation}$: kilometric point of the synthetic DO monitoring station used for data assimilation. $kp_{validation}$: kilometric point of the DO monitoring station used for evaluation of model outputs. kp_{end}: kilometric point of the end of synthetic scheme. Please refer to Table 4.3 for reading case study properties' values. . . . 65
4.3 Result of DA without characterization of dissolved organic carbon biodegradability (based on exp. 1 setup in Table 4.3). The red line demonstrates the target parameter value which is supposed to be retrieved in DA. (i), (ii) and (iii) the different target b_1 cases of 0.25, 0.35 and 0.45, respectively. A and B) posterior pdfs of bacterial growth and yield. C) Ensemble mean of the simulated DO by ProSe-PA vs. observed DO at the validation station . . . 72
4.4 Results of DA with estimation of the biodegradability of dissolved organic carbon under high NG condition (based on exp. 2 setup in Table 4.3). The red line demonstrates the target parameter value which is supposed to be retrieved in DA. (i), (ii) and (iii) the different target b_1 cases of 0.25, 0.35 and 0.45, respectively. A, B and C) posterior pdfs of b_1, maximum growth rate of bacteria and bacteria yield. D) Ensemble mean of the simulated DO by ProSe-PA vs. synthetic observed DO. . . . 74
4.5 Results of DA with estimation of the biodegradability of dissolved organic carbon under low NG condition (exp. 2). The red line demonstrates the target parameter value which is supposed to be retrieved in DA. (i), (ii) and (iii) the different target b_1 cases of 0.25, 0.35 and 0.45, respectively. A, B and C) posterior pdfs of b_1, maximum growth rate of bacteria and bacteria yield. D) Ensemble mean of the simulated DO by ProSe-PA vs. synthetic observed DO. . . . 75
4.6 Results of moving assimilation station between kp10 to kp90 (exp. 3): evolution of RMSE at the validation station (kp35) with kp of the assimilation station, and the summary of parameter identification. Refer to section 2.4 for the definition of observability, limitation and depletion points. . . . 76
4.7 Longitudinal profiles (in kilometer) of the small monomeric substrate (rapidly biodegradable dissolved organic carbon, $BDOC_{rapid}$) for different b_1 values. The two vertical grey lines indicate the location of the two assimilation stations. The red dashed-line is the mean substrate limit. . . . 78

4.8 Results of the double station strategy. A, B, and C) the posterior pdfs of b_1, maximum bacterial growth rate and bacteria growth yield. Red lines show the target parameters values to be retrieved by DA. D) Ensemble mean of the simulated DO by ProSe-PA vs. observed DO . . . 79
4.9 Retrieval of target parameters while detecting change in bacteria physiology (exp. 4). (i) DA with 15 min time step. (ii) DA with 1 hr time step. Red lines show the target parameters values to be retrieved by data assimilation. Both cases have 4% perturbation for bacteria physiology parameters. . . . 81
4.10 Retrieval of target parameters while detecting change in biodegradability of dissolved organic carbon from boundary conditions with 1 hr time step (exp. 5). (i) Abrupt change in b_1. (ii) Gradual change in b_1. Red lines shows the target parameters values to be retrieved by data assimilation. . . . 82

5.1 Study Area . . . 87
5.2 Situation of Seine River in 2011. Highlighted polygons denote the low flow period (purple) and the dry period (green). A) Evolution of dissolved oxygen at the 8 monitoring stations. (color gradient: grey to black representing upstream to downstream stations). B) Daily discharge at the gauging station and temperature at a monitoring station. . . 89
5.3 The integration of *macrospecies* in the source code of ProSe-PA. `TOC` is the reading of the total organic carbon of any source. `threshold` is the reading of share of DOC in TOC (t parameter in the model). `share_mod` and `share_mop` reads the biodegradability shares of DOC and POC. `biodegradable` which is composed of `val` and `range` reads the average value and variation range of the biodegradability of DOC and POC (b_1 and b_2). `fast_biodegradable` is the reading of the share of rapidly biodegradable fractions of DOC and POC (s_1 and s_2). Only the parameter with a range is included in data assimilation framework. . . . 92
5.4 The fluxes of total organic carbon from seven major boundary conditions during the low-flow period of 2011 (in ton C day^{-1}). SAM: Seine Amont WWTP. SAV: Seine Aval WWTP. SC: Seine Centre WWTP. . . . 94
5.5 Ensemble weighted average DO concentrations at Suresnes, Colombes, Chatou and Bougival stations simulated by ProSe-PA vs. observed DO (grey line). The baseline simulation is denoted in red and the simulation incorporating the biodegradability of DOC in blue. . . . 98
5.6 Ensemble weighted average DO concentrations at Sartrouville, Andresy, Meulan and Mericourt stations simulated by ProSe-PA vs. observed DO (grey line). Baseline simulation is denoted in red and the simulation incorporating the biodegradability of DOC in blue. . . . 99
5.7 Normalised importance weights of model parameters. $\mu_{max,hb}$: maximum growth rate of bacteria. Y_{hb}: growth yield of bacteria. b_1: the parameter representing the biodegradability of dissolved organic matter in the proposed organic matter repartition model. . 100
5.8 Posterior pdfs of model parameters of the DA with inference of biodegradability during dry period. $\mu_{max,hb}$: maximum growth rate of bacteria. Y_{hb}: growth yield of bacteria. b_1: the parameter representing the biodegradability of dissolved organic matter in the proposed organic matter repartition model. . . . 101
5.9 Simulated BDOC and simulated DOC versus the observed DOC at St. Honorine station. 103
5.10 Ensemble weighted average DO concentrations at Suresnes, Colombes, Chatou and Bougival stations simulated by ProSe-PA vs. observed DO (grey line). The simulation with two bacteria communities is denoted in blue and the simulation with one bacteria is in blue. . . . 105

5.11 Ensemble weighted average DO concentrations at Sartrouville, Andresy, Meulan and Mericourt stations simulated by ProSe-PA vs. observed DO (grey line). The simulation with two bacteria communities is denoted in blue and the simulation with one bacteria is in blue. . . . 106
5.12 Normalised importance weights of model parameters in DA simulation with 2 bacteria communities. . . . 107
5.13 Longitudinal profile of mean BDOM concentration together with the mean concentrations of small and large bacteria from day 250 to 290. Only left river reaches are presented here. GE: Garge-Epinay, SCE: Seine Centre WWTP. SAV: Seine Aval WWTP. Oise: Oise river. SCE: Seine Gresilion WWTP. . . . 107
A1 FADM results for t (DOC:TOC). A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters. . . . 121
A2 Initial FADM results for DOC. A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters. . . . 122
A3 Refined FADM results for DOC (without dry-barren land clustering for better visualization of other quantitative variables. A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters. . . . 122
A4 FADM results of the Monod half-saturation constant for uptake of substrate (K_s in mgC L^{-1}). A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters. . . . 123
A5 Distribution of the individuals corresponding to the DOC:TOC ratio, t, on dimensions 1 & 2 with the confidence ellipses . . . 123
A6 Distribution of the individuals corresponding to DOC variables on dimensions 1 & 2 with the confidence ellipses . . . 124
A7 Distribution of the individuals corresponding to the Monod half-saturation constant for uptake of substrate (K_s in mgC L^{-1}) on dimensions 1 & 2 with the confidence ellipses 124
C1 Ensemble of the 260,000 DO simulations for TOC = 5 mgCL^{-1} in the second SA analysis 131
C2 a) Cumulative sum of EOF variances and b) time evolution of four (k) significant EOFs for TOC = 5 mgCL^{-1} in second SA . . . 132
D1 Monod curve and its slope plotted using a semi-saturation constant, K_S, of 0.025 mgC L^{-1} and a maximum growth rate of 0.07 h^{-1} (Eq. 4.13). The growth rate is normalized with the maximum rate of bacteria. . . . 135
D2 Detection of change in bacteria physiology (exp. 4). P_{yield}: perturbation % of bacteria growth yield parameter. P_{growth}: perturbation % of bacteria maximum growth parameter. The blue colored cell shows the best combination. . . . 136
D3 Detection of change in biodegradability of dissolved organic carbon of boundary conditions (exp. 5). P_{b1}: perturbation % of b_1 parameter. Changes types are 0 days (abrupt) and 5 days (gradual). The blue colored cell shows the best combination. . . . 137

List of Tables

1 Valeurs moyennes générales et gammes de variation des variables de matière organique vii
2 Valeurs moyennes et gammes de variation des variables de la matière organique pour les groupes identifiés d'occupation du sol, de climat et d'hydrologie viii
3 RMSE des scénarios d'assimilation de données avec ratio de biodégradabilité du carbone organique dissous fixe et variable aux les stations de surveillance xviii

2.1 Bibliography on TOC concentration (mgC L^{-1}) in experimental studies together with the hydrology, land cover, and climate of the sampled river 15
2.2 Bibliography on DOC concentration (mgC L^{-1}) in experimental studies together with the hydrology, land cover, and climate of the sampled river 16
2.3 Bibliography on t (DOC:TOC) variable in experimental studies together with the hydrology, land cover, and climate of the sampled river 18
2.4 Bibliography on the biodegradable fraction of DOC, b_1, in experimental studies together with the hydrology, land cover, and climate of the sampled river 19
2.5 Bibliography on the biodegradable fraction of POC, b_2, in experimental studies with their hydrology, land cover, and climate 20
2.6 Bibliography on the rapidly biodegradable fraction of BDOC and BPOC, s_1 and s_2, respectively 20
2.7 Bibliography on the Monod half-saturation constant for uptake of substrate,K_s, (mgC L^{-1}) in experimental studies together with the type of habitat or sampled environment and substrate type 21
2.8 Bibliography on the coefficient of organic matter hydrolysis ($k_{hyd,max}$ in h^{-1}) and the half-saturation constant for hydrolysis of $BDOC_2$ into $BDOC_1$ (K_{BDOC2} in mgC L^{-1}) 22
2.9 Overall average values and variation ranges of OM variables 23
2.10 Average values and variation ranges of OM variables for the identified clusters of land cover, climate, and hydrology 26

3.1 List of the parameters accounted for in the sensitivity analyses and their corresponding ranges of variation 43
3.2 Initial concentrations of the simulations 45
3.3 Definition of the eight single simulations achieved with extreme values of biodegradable dissolved organic matter, net growth of bacteria community and its associated yield . 46
3.4 The parameters considered in each of the four sensitivity analyses 49

4.1 List of model parameters considered for data assimilation and their corresponding variation ranges 63
4.2 Upstream boundary conditions 66

4.3 Properties of the case study experiments under different research questions. Columns 5–7 list the values of model parameters used for creation of synthetic DO monitoring data and for retrieval by ProSe-PA. L: length of case study reach. $\mu_{max,hb}$: maximum growth rate of bacteria. Y_{hb}: growth yield of bacteria. b_1: parameter representing the biodegradability of dissolved organic carbon. T: simulation period. NG: Net bacterial growth (values for high and low net growth are 0.05 and 0.01 h^{-1}, respectively). N_{assim}: number of assimilation stations. $kp_{assim,1,2}$: kilometric point of first and second assimilation stations. $kp_{valid.}$: kilometric point of validation station 70
4.4 Further details of the experiments for evaluating the capacity of ProSe-PA in detecting change in environmental conditions 70
4.5 Results of the optimum combination of data assimilation configuration parameters for detecting change in bacteria physiology and dissolved organic carbon biodegradability 80

5.1 DO concentrations in 2011 at the 8 monitoring stations to be used for data assimilation (retrieved from Wang et al. [2022a]) 90
5.2 Comparison of the organic carbon repartition model parameters used in different theses 91
5.3 List of model parameters considered for data assimilation and their corresponding variation ranges 93
5.4 Performances of the baseline DA scenario and DA with inference of biodegradability evaluated by RMSE. 97
5.5 Performances of the DA with two bacteria communities compared to DA with a single bacteria community evaluated by RMSE. 104
E1 Eléments à définir dans le fichier de commande 149
E2 Eléments à définir dans le fichier de commande 150
E3 Codes des espèces biogéochimique simulées 156
E4 Eléments à définir dans le fichier de commande 165
E5 Elements to be defined in the command file 167
E6 Eléments à définir dans le fichier de commande 169
E7 Description des termes d'un bilan pour une espèce 184

Chapter 1

Introduction

1 General context

The Seine river basin which constitutes 11.8% (76,236 km^2) of France territory and welcomes 25% (17 millions) of its population [Flipo et al., 2020], is a major industrial, agricultural, river navigation and tourism hub. The basin is subjected to high anthropogenic impacts particularly from the Paris metropolitan area where the wastewater treatment plants (WWTP) and short rainy spill inputs from combined sewage outflows (CSO) impose a large load of organic matter into the river. In addition, agricultural, animal grazing, industry and urban development are other sources of river pollution.

The environmental issues of the Seine river have been eutrophication and massive heterotrophic activity [Garnier et al., 2020] that has resulted in depletion of Dissolved Oxygen (DO) "the vital component for aquatic life and the overall health of the ecosystem" [Zahraeifard and Deng, 2012, Bailey and Ahmadi, 2014, Escoffier et al., 2016]. Lack of sufficient DO causes mortality of aquatic organisms, and undesirable taste and odor in the river which would eventually affect the tourism, water sports and entertainment industry as well. Furthermore, the eutrophication of drinking water due to excessive nutrients imposes an additional financial burden on water suppliers [Garnier et al., 2020]. The rapid growth of population in Paris city and its surrounding cities in the second half of the 19$^{\text{th}}$ century, coupled with the development of industrial fertilizers in 1900s made the circularity of human waste less attractive and led to direct release of sewer flows into the Seine River, degrading its water quality [Esculier and Barles, 2020]. At the end of the sixties, 60% of the collected wastewater were directly released to the river. However, since 1990s and after the construction of new treatment plants and the enhancement of treatment technology, there has been significant improvements in the river water quality. However, several challenges such as lack of enough dilution of point source releases during summer low flow periods, untreated releases from WWTPs during maintenance periods, eutrophication in coastal zones, diffuse agricultural loads and nitrate contamination of groundwater still exist [Garnier et al., 2020].

Given the challenges outlined above, water quality modeling has become an indispensable tool for managing water resources with regard to quality. It facilitates the examination of the effects of ongoing anthropization and pollution discharges from both point and non-point sources into water bodies. Additionally, water quality models are capable of predicting the potential impact of future river system developments, including the construction of new treatment facilities and combined sewer overflows.

As a result, they are valuable resources for devising remediation strategies aimed at improving water quality.

2 A brief historical overview of water quality modeling: from static to dynamic representation

The first water quality model for simulating DO was developed by Streeter and Phelps [1925] incorporating DO and the biological oxygen demand. Then, the River Continuum concept was proposed by Vannote et al. [1980] that made the basis for the current community based models by describing how the physical, chemical, and biological characteristics of a river change as it flows from its headwaters to its mouth. This helped ecologists understand the relationships between different parts of a river ecosystem, and how anthropogenic changes in one part of the ecosystem can affect the entire river.

Since then, several other water quality models have been developed that are reviewed in detail in [Reichert et al., 2001, Cox, 2003] such as: AQUATOX whose main capability is to simulate the effect of pollutants on fishes [Park et al., 1974]; the QUAL family of models that contains all the related process and has a auto-calibration capacity [Park and Lee, 2002]; RWQM who is the only model capable of studying the impact of sanitation networks on the receiving mediums [Vanrolleghem et al., 2001].

There are also models that are able to simulate water quality at the river basin scale, thereby analysing the biogeochemical functioning and metabolism of rivers within their hydrographic networks. One such model is the RIVERSTRAHLER model [Billen et al., 1994, Garnier et al., 1995, Thieu et al., 2009] that has been even coupled with a GIS [Ruelland et al., 2007, Thieu et al., 2011]. Another of such model is QUAL-NET that can simulate high temporal-resolution studies at a very low computational cost [Minaudo et al., 2018]. SWAT has also been used for modeling nitrate fluxes [Zettam et al., 2020] and even coupled with a lithological model to investigate the impact of ionic loadings [Lechuga-Crespo et al., 2021]

However, all these models provide a static view of river communities such as that of bacteria and phytoplanktonic communities. In other words, the model parameters representing the different hydro-biogeochemical processes do not change with time. Hence, a dynamic representation of the river system is necessary because studies show that the riverine communities undergo seasonal changes, respond to fluctuations in the organic matter loading, and are sensitive to modifications in the biodegradability of organic matter [Hullar et al., 2006, Crump et al., 2003]. Even an increase in BDOC is considered to increase the diversity of bacterial populations [Landa et al., 2013]. Consequently the physiological parameters of the river communities change accordingly. Nevertheless, the complexity of hydro-biogeochemical processes necessitates experimental campaigns to ascertain evolution of model parameters [Garnier et al., 1992b, 1995, Flipo et al., 2007], which is a challenging endeavor due to the large number of parameters involved. Furthermore, the calibration of water quality models is also an arduous task due to the intricate nature of hydro-biogeochemical processes [Arhonditsis et al., 2006], and may result in equifinality [Polus et al., 2011], whereby varying parameter combinations yield similar outcomes.

An effective strategy to attain a dynamic representation of hydro-biogeochemical processes entails the application of data assimilation techniques. These methodologies involve the coupling of models and observation data, such as Dissolved Oxygen (DO), to accurately determine the temporal evolu-

tion of critical model parameters that significantly impact the target observation data. Despite their early development in freshwater sciences, the usage of assimilation techniques remain limited in water quality modeling [Cho et al., 2020] among the scientific community. Beck and Young [1976] did the first implementation of a data assimilation technique, extended Kalman filter (EKF) [Cosby and Hornberger, 1984, Mao et al., 2009], in the water quality modeling in 1970s that used DO and BOD as state variables. Then, El Serafy et al. [2007] contributed to the first implementation of ensemble Kalman filter (EnKF) in Delft3D water quality model, which remains today the most popular assimilation method [Beck and Halfon, 1991, Carrassi et al., 2018, Page et al., 2018, Loos et al., 2020, Cho et al., 2020]. Next, Babbar-Sebens et al. [2013] did the first application of a variational method, 3DVAR, by assimilating satellite temperature data in a reservoir water quality model.

Finally, the first implementation of the particle filter (PF) method was achieved in the work of Wang [2019], Wang et al. [2019, 2022a] in ProSe-PA model for Seine River, assimilating high frequency DO data. PF is as good as EnKF in slightly non-linear systems (e.i. modeling of rearation process in a river), however it is proven to be more efficient and accurate than the EnKF [Wang et al., 2023] in strong non-linear systems for example in the modeling of rivers with dominant heterotrophic bacterial activity. It has been able to produce acceptable DO simulations and identify the correct posterior distributions of model parameters. EnKF only excels in linear systems where the parameters and state variables follow a Gaussian distribution [Evensen et al., 2022]. The advantage of PF stems from the fact that it weights its members (particles) based on their performance in predicting the true value, while EnKF assumes an equally weighted Gaussian distribution for its members, considered to be a biased assumption [Wikle and Berliner, 2007]. In addition, EnKF is found to have 25% more computational cost than PF [Wang et al., 2023].

3 Modeling and monitoring water Quality in the Seine River basin: advances in data assimilation and its challenges

In response to the aforementioned environmental challenges and to gain a better understanding of the ecological processes within the Seine watershed, the PIREN-Seine multidisciplinary research program (`https://www.piren-seine.fr`) was established in 1989 [Flipo et al., 2020]. Through this initiative, researchers embarked on projects that combined on-site surveys, data collection, and model development to create comprehensive tools for assessing the overall functioning of the system, which includes the river network, its watershed, and the social communities within it. In order to simulate the desired indicators, mathematical equations related to every aspect of the environment that influences the ecology of the river are incorporated into the models. Once these models are validated, they are used for management purposes.

The RIVE biogeochemical model [Billen et al., 1994, Garnier et al., 1995, Wang et al., 2018, Wang, 2021, Wang et al., 2022b], which is developed at the heart of PIREN-Seine research program, serves as an effective tool for simulating the complex cycles of carbon, oxygen, and other essential nutrients in the river. In order to provide a comprehensive understanding of water quality indicators and species along the Seine river under different trophic and hydrological conditions, the RIVE model was combined with transport and hydrodynamics modules to develop the ProSe software [Even et al., 1998, 2004, Flipo et al., 2004, 2007, Vilmin et al., 2015]. The calibration and validation of several hydro-biogeochemical processes in the ProSe software have been conducted in the PhD thesis of Vilmin

[2014] in addition to the assessment of the biogeochemical functioning of Seine River, especially with respect to the carbon cycle [Vilmin et al., 2016]. It also led to guidance on water quality monitoring [Vilmin et al., 2018].

Once again, as ProSe did not provide dynamic representation of hydro-biogeochemical processes, the need for development of data assimilation-based tool was felt which in turn required access to high frequency data in the Seine River basin. Within this framework and since 2000, the European Parliament [European Commission, 2000] introduced a directive in order to establish a framework for protection of European water resources based on which water managers must develop tools to enhance the water quality of rivers and monitor the quality indicators. Therefore, a high frequency data collection system was first tested under the CARBOSEINE project aiming at providing a tool to assess the resilience and vulnerability of the Seine over a sector extending from downstream of Paris to the confluence of Seine and Oise rivers [Escoffier et al., 2018, Groleau et al., 2016]. This network was taken over and further developed by the SIAAP (Service Interdépartemental pour l'Assainissement de l'Agglomération Parisienne) under the MeSeine label. Currently, dissolved oxygen (DO) data with a frequency of 15 min, and carbon, temperature and chlorophyll-a with lower frequencies are collected at several stations along the Seine river which not only allows continuous water quality monitoring, but also is utilized for improvement of water quality models in terms of better construction of the models and having data for comparison with simulation results.

As a result, thanks to the availability of high frequency DO data, ProSe-PA, a data assimilation software, was developed at Mines Paris – PSL, based on both the RIVE biogeochemical model and an innovative usage of high frequency oxygen data through the work of Wang [2019] that couples the ProSe model with a particle filter algorithm. ProSe-PA uses the high frequency DO data in order to correct over time the physiological parameters of species competing in the metabolism of river. An application using synthetic data [Wang et al., 2019] and data from the Seine river in 2011 [Wang et al., 2022a] was performed which not only provided a better characterization of physical, phytoplanktonic and bacterial parameters, but also provided significant improvements in simulated DO compared to the observed ones during both bloom and non-bloom periods. However, mismatches remain between the simulated and observed DO time series, in particular during non-bloom low flow periods. Wang et al. [2022a], who performed this study, hypothesized that they may be due to:

- Partitioning of Organic matter (OM) from boundary conditions (tributary rivers) and point sources such as WWTPs, CSOs into their dissolved, particulate and consequently to biodegradable fractions;
- Degradation kinetic of organic matter like the hydrolysis and uptake of organic matter by bacteria;
- Incorporation of distinct small and large bacteria species in the model
- Partitioning of phytoplankton lysis and excretion;

4 Thesis Objectives

In light of the two previous PhD theses [Vilmin, 2014, Wang et al., 2019] which leaded to the calibration and validation of the hydro-biogeochemical processes in ProSe as well as development of the data

assimilation tool (ProSe-PA), the overall objective of this thesis is to couple ProSe-PA with high frequency observation data (DO) in order to improve model accuracy by accounting for shortcomings and reducing uncertainties at low-flow period and further develop it into a management tool. It is intended to initially find the parameters governing DO at low flow by conducting a sensitivity analysis of ProSe-PA existing parameters together with new parameters proposed for addressing the uncertainties related to organic matter. Once the role of organic matter on DO evolution could be confirmed, its influential parameters will be integrated in the data assimilation scheme of ProSe-PA together with other influential parameters that represent bacteria and phytoplankton dynamics. It is indeed the first adaptation of a data assimilation-equipped water quality model to investigate the dynamics of low-flow condition. Hence, the thesis objectives are defined as:

1. Evaluating model uncertainties linked to parameterization of organic matter inflow and its degradation from different sources. This involves carrying out the following tasks:
   - Proposal of a parameterized organic matter partitioning model to account for the uncertainties related to organic matter from boundary condition
   - Identification of the parameters of organic matter degradation kinetics in ProSe-PA.
   - Bibliographical review to find the variation range of the above-mentioned parameters.
   - Sensitivity analysis of the model for finding the influential parameters among the mentioned parameters and the already existing parameters of ProSe-PA.
2. To reconstruct the river metabolism in terms of DO and organic matter for 2011 and to validate the improvements in simulations at low flow. This involves carrying out the following tasks:
   - Integration of the organic carbon partitioning model in the source code of ProSe-PA.
   - Incorporation of the influential parameters of the sensitivity analysis in the data assimilation framework of ProSe-PA.
   - Validation of these new developments by a synthetic case study to investigate if ProSe-PA can retrieve predefined target model parameters by oxygen data assimilation.
   - Application of ProSe-PA on a real case study for the low-flow period of 2011 together with testing the performance of ProSe-PAwith two distinct small and large bacteria communities in the model.

5 Thesis outline

To accomplish the thesis objectives and the corresponding tasks, this manuscript is organised into six chapters including the first chapter for introduction and the sixth chapter for conclusion and perspectives. The chapters are mostly written in article format as follows:

- The second chapter presents the results of a detailed literature review on the parameters of the organic matter partitioning model and degradation kinetics in world rivers under different climatic, hydrological and land cover conditions. This work is currently submitted for review in the journal of Earth Sciences Review. The purpose of this work is to not only find the variation range of the target parameters, but also to identify the main drivers of their variability using

the clustering method of "Factorial analysis of mixed data". The information collated in this literature review has been summarized into ten tables (section 2.2). This work is also partly published in the annual reports of PIREN-Seine [Hasanyar et al., 2020].

– The third chapter has been published in EGU journal of *Biogeosciences* [Hasanyar et al., 2023b]. It is devoted to the sensitivity analysis of bibliographical reviewed parameters and other parameters that already exist in ProSe-PA. In this paper, we applied Sobol's sensitivity analysis method along with the PCA dimension reduction technique to identify the significant parameters influencing water quality during summer low-flow periods. The findings indicated that the physiology of bacteria and the biodegradable fraction of dissolved organic matter (BDOC) from boundary conditions exert significant control over the evolution of dissolved oxygen during heterotrophic low-flow periods.

– Chapter 4: As the outcomes of the sensitivity analysis offer a promising avenue for integrating the influential parameters into data assimilation frameworks, the fourth chapter details the integration of the organic matter partitioning model in ProSe-PA in addition to incorporation of its influential parameter (biodegradability of DOC) in the assimilation framework of ProSe-PA. This paper is soon to be submitted to the journal of Water Resources Research. In this paper, we assessed the ability of ProSe-PA to retrieve the predefined values of model parameters including that of the biodegradability of dissolved organic carbon (DOC) by assimilating oxygen data, and validated the results using a synthetic test case. The study found that ProSe-PA was successful in quantifying biodegradability even under conditions of varying carbon input from boundary conditions. Incorporating this information improved the accuracy of dissolved oxygen simulations compared to those DA simulations where biodegradability of DOC was not inferred. This study is partly published in the annual reports of PIREN-Seine [Hasanyar et al., 2022].

– Chapter 5 assesses the performance of ProSe-PA in a real case scenario for the low-flow period of 2011. It demonstrates that assimilating oxygen data with the inference of biodegradability of dissolved organic carbon enhanced DO simulations in most monitoring stations. Furthermore, the evolution of the biodegradability parameter and bacterial growth was characterized over time. Although integrating data assimilation with both small and large bacteria communities provided greater stability in parameter characterization, it only improved simulations at a few stations, indicating the need for further exploration into the contribution of these two bacterial communities from boundary conditions. This manuscript is currently in a draft stage.

In addition, the appendix holds: the biogeochemical equations of dissolved oxygen and organic matter in ProSe-PA(appendix B1.1); a detailed description of the sensitivity analysis protocol used in Chapter 3 (appendix B2.2); and supplementary results of different analysis and simulations performed in this thesis.

Chapter 2

A review of variables representing the description and degradation of organic matter in world rivers under different climatic, hydrological, and land cover conditions

Manuscript status: resubmitted for peer review to the journal of Earth Sciences Review.

Résumé

La connaissance de la matière organique et de ses fractions constitutives dans les rivières est cruciale pour les études du métabolisme des rivières et de la modélisation de la qualité de l'eau. Comprendre comment les concentrations de carbone organique total (COT), de ses constituants tels que le carbone organique dissous et particulaire (COD, COP) et de leurs fractions biodégradables (CODB) varient en fonction des conditions hydrologiques, de couverture du sol et climatiques est essentiel pour étudier la dynamique de la matière organique dans les rivières. Ainsi, le but de cette étude est de fournir la gamme de variation de quatre variables descriptives de la matière organique, composées de deux indicateurs de qualité de l'eau (COT, COD) et de deux ratios de partitionnement de la matière organique (t = COD/COT et b_1 = CODB/COD).

Tout d'abord, une recherche bibliographique détaillée est réalisée pour collecter des données quantitatives expérimentales sur ces variables dans les conditions considérées (et leurs catégories correspondantes) : hydrologie (faible débit, débit moyen et fort débit), type de couverture du sol (forêt, agriculture, prairies, zones humides, zones urbanisées et terres stériles) et climat (tropical, sec, tempéré, continental et polaire). Les gammes de variation globales des variables sont : 0,3 à 19,3 mgC L^{-1} pour le COD, 0,5 à 39,0 mgC L^{-1} pour le COT, 0,08 à 0,96 pour t et 0,05 à 0,54 pour b_1. Ensuite, une méthode de réduction de la dimensionnalité (AFDM: analyse factorielle des données mixtes) est utilisée pour identifier les catégories les plus importantes (clusters) qui se distinguent dans le jeu de données et qui influencent la variabilité de chaque variable. Par conséquent, la gamme de variation et la valeur moyenne de chaque variable ne sont fournies que pour les clusters identifiés.

Les résultats de la AFDM démontrent le contrôle des variables de la matière organique par le

climat et la couverture du sol. Plus précisément, la biodégradabilité du COD augmente dans les zones urbanisées. Enfin, les incertitudes dans la mesure de la matière orgnanique, le rôle de sa mesure qualitative et les lacunes de recherche dans la modélisation de la qualité de l'eau sont discutées.

Abstract

Knowledge of organic matter (OM) and its constituent fractions in rivers is crucial for studies of river metabolism and water quality modeling. How the concentrations of total organic carbon (TOC), its constituents such as dissolved and particulate organic carbon (DOC, POC), and their biodegradable fractions (BDOC, BPOC) vary according to hydrological, land cover, and climatic conditions is essential in studying the dynamics of OM in rivers and streams. Thus, the purpose of this study is to provide the variation range of four OM descriptive variables consisting of two water quality indicators (TOC, DOC) and two OM partitioning ratios (t = DOC/TOC & b_1 = BDOC/DOC). Firstly, a detailed bibliographic search is performed to collect experimental quantitative data on the OM variables under the considered conditions (and their corresponding categories): hydrology (low flow, average flow, and high flow), land cover type (forest, agriculture, grasslands, wetlands, urbanized, and barren lands), and climate (tropical, dry, temperate, continental, and polar). The overall variation ranges of OM variables without discriminating the effect of climate, hydrology and land cover are: 0.3–19.3 mgC L^{-1} for DOC, 0.5–39.0 mgC L^{-1} for TOC, 0.08–0.96 for t, and 0.05–0.54 for b_1. Then, in order to discriminate their interconnection effects on OM variables, a dimensionality reduction method (FAMD: factorial analysis of mixed data) is used to identify the most important categories (clusters) that stand out in the dataset and drive the variability of each OM variable. Consequently, the variation range and the average value of each OM variable are provided only for the identified clusters. The results of FAMD demonstrate the control of OM variables by climate and land cover. More specifically, the biodegradability of DOC increases in urbanized land cover, while the largest OM content are reported in rivers connected to wetlands and the cluster of dry climate and barren lands. In addition, as an OM degradation parameter, the Monod half-saturation constant for uptake of substrate (K_S) is also classified according to its habitat and substrate types. Finally, the uncertainties in OM measurement, the role of qualitative OM measurement, and the research gaps in water quality modeling are discussed.

1 Introduction

The stream or river metabolism is an important aspect of the global aquatic carbon cycle [Battin et al., 2023, Cole et al., 2007]. Balance between gross primary production and ecosystem respiration in a

river ecosystem is a major issue for defining ecosystem metabolism and hence its auto-heterotrophic status [Garnier and Billen, 2007, Escoffier et al., 2018]. Research on river metabolism has shown the significant role of rivers in carbon sequestration and in CO_2/ CH_4 emissions, as well as in the transfer of carbon from land to oceans where rivers act like "active pipes" [Cole et al., 2007, Battin et al., 2009, Raymond et al., 2013, Hotchkiss et al., 2015, Stanley et al., 2016, Cole et al., 2021]. Conducting river metabolism research requires information on carbon worldwide, and its biodegradability under different conditions and flow regimes (hydrology). Nevertheless, such information is not readily available for all river ecosystems around the world since most carbon measurements have been made only in developed countries and, furthermore, rivers are subjected to global change including several other controlling factors. Climate change is known to alter river discharge and temperature [Blöschl et al., 2019, Yang et al., 2020, Seyedhashemi et al., 2022]. Land cover change through different uses of land, such as urbanization and increase in agricultural lands, is found to alter the composition, concentration, availability, and biodegradability of carbon [Quinton et al., 2010, Graeber et al., 2012, Hosen et al., 2014, Atkins et al., 2017, Liu et al., 2020]. These effects of global change have an impact on the carbon cycle and thereby on the estimation of river metabolism [Bernot et al., 2010a, Masese et al., 2017]. Therefore, a review is necessary to provide information on carbon levels and variations in rivers under different hydrological, land cover, and climatic conditions. Interestingly, this information could also be used to facilitate river metabolism research in rivers for which no carbon data exist, provided that information on their hydrology, climate, and basin land cover is documented.

Moreover, carbon also controls the dissolved oxygen in rivers due to microorganisms [Vannote et al., 1980]. Management tools (e.g., mechanistic or statistic water quality models) simulate river carbon fluxes and oxygenation, a good indicator of river quality, and can calculate mass balances (see for instance Vilmin et al. [2016, 2018], Segatto et al. [2020]). Furthermore, with the development of river monitoring networks with low-cost sensors enabling the integration of registered data, the use of such models has become more essential when evaluating the water quality and the ecological functioning of rivers [Wang et al., 2022a, 2023]. The purpose of this review is therefore to provide a quantitative analysis (average and variation ranges) of basic variables that represent carbon in rivers.

The information on organic matter (OM) variables first of all consists of total organic carbon (TOC), particulate organic carbon (POC), and dissolved organic carbon (DOC) concentrations that are needed for the determination of carbon flux and mass balance studies [Meybeck, 1982, Ran, 2013, Vilmin et al., 2016]. DOC is also the main substrate in the heterotrophic bacterial respiration in rivers and is useful for evaluating, for example, the production of drinking water from river [Yang et al.,

2019]. The biodegradable dissolved organic carbon (BDOC) and biodegradable particulate organic carbon (BPOC) concentrations are essential variables for modeling bacterial activity [Servais et al., 1995] and thus for understanding the functioning of the ecosystem and the biodegradable carbon flux of rivers into coastal zones [Wiegner et al., 2009, Etcheber et al., 2007]. BDOC as a substrate for bacterial respiration is of major interest in the estimation of CO_2 partial pressure (pCO_2) in rivers and CO_2 emission as a greenhouse gas (GHG) [Vilmin et al., 2016, Marescaux et al., 2018, 2020]. However, BDOC can be further split into rapidly and slowly biodegradable fractions ($BDOC_1$ & $BDOC_2$) that are important inputs of water quality models since bacteria consumption first occurs through the degradation of $BDOC_1$, $BDOC_2$ requiring the occurrence of an exoenzymatic activity [Billen, 1991]. Also formulated by Billen [1991], the degradation of OM in river systems is limited by the availability of small monomeric substrates, which is accounted for by a Monod half-saturation constant for uptake of substrate (K_S).

When reviewing the literature for similar work, we find that information provided by the other works does not satisfy the research needs regarding river metabolism and water quality modeling, especially because most of the river systems in the world today are impacted by global changes. Meybeck [1982] provided general ranges of values for DOC and TOC concentrations and fluxes as well as the DOC/TOC ratio, but these ranges do not discriminate rivers at low flow or high flow or do not differentiate between rivers with urbanized land cover and rivers with agricultural land cover. Dyer [1991] analyzed the average DOC and POC export rates for rivers on different continents while Ittekkot [1988] provided a review of POC levels and BPOC percentages for 13 world rivers. Ludwig et al. [1996] classified the DOC export and DOC/POC of world rivers according to climate and geographical location. Although the effect of climate on the variability of organic carbon has been confirmed [Olefeldt et al., 2013], no correlation between geographical location and organic carbon levels could be found. Indeed, hydrology, climate, land cover, and land use are considered the main controls of the variability of organic carbon [Rizinjirabake et al., 2018, Liu et al., 2020]. Huang et al. [2012] provided only mean annual DOC and POC concentrations in tropical rivers for some continents.

The OM information (average values and variation ranges) collected here from the literature for many rivers is intended to document the effect of drivers such as climate, hydrology, and land cover, including urbanization. A factorial analysis of mixed data (FAMD) [Escofier, 1979, Pagès, 2004] is used to study the relationship between OM variables and the different drivers through a clustering.

2 Material and methods

In this section, we first select the OM descriptive variables and their constituent fractions in world rivers in addition to OM degradation variables. Next, we present the data collected on these quantitative variables and the categories of different hydrological, climatic, and land cover conditions followed by the explanation of a dimensionality reduction method.

2.1 Selection of OM variables for review

2.1.1 OM descriptive variables in water quality models

In water quality models, the OM descriptive variables are mostly made up of two main water quality indicators: total organic carbon (TOC) and dissolved organic carbon (DOC). Then, these two variables are distributed into their constituent fractions based on a partitioning model illustrated in Figure 2.1. Here, TOC is initially divided into DOC and the particulate organic carbon (POC) parts. DOC and POC are subsequently partitioned into their corresponding biodegradable (BDOC, BPOC) and refractory fractions (RDOC, RPOC), respectively. In water quality models such as the RWQM [Reichert et al., 2001], OM was partitioned into a single pool for the biodegradable portion of DOC and POC. Other models such as RIVE further divide BDOC into $BDOC_1$ (rapidly biodegradable DOC in 5 days) and $BDOC_2$ (slowly biodegradable DOC in 45 days) [Billen et al., 1994, Garnier and Billen, 1994, Flipo et al., 2007]. Similarly POC is divided into $BPOC_1$ and $BPOC_2$ fractions.

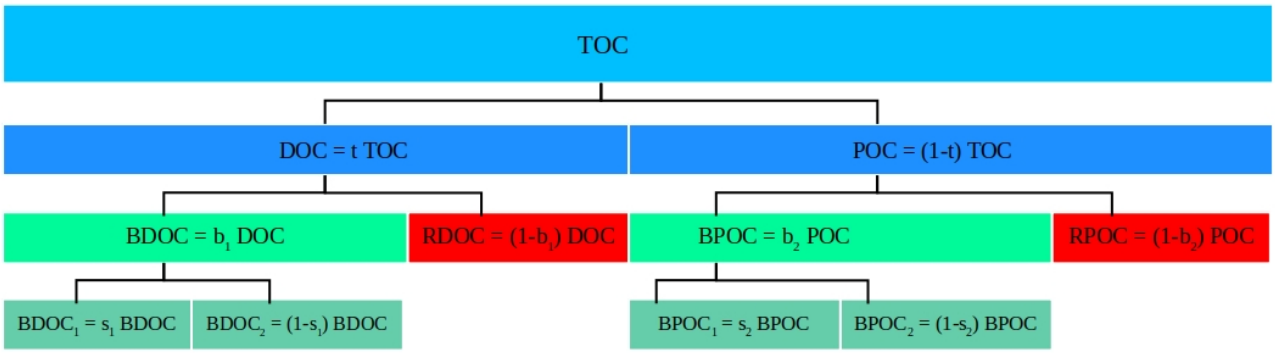


Figure 2.1: OM partitioning into its constituent fractions in terms of the five partitioning ratios, namely, t, b_1, s_1, b_2, s_2. DOC: dissolved organic carbon, POC: particulate organic carbon, BDOC: biodegradable DOC, RDOC: refractory DOC, BPOC: biodegradable POC, RPOC: refractory POC, $BDOC_1$ & $BDOC_2$: rapidly and slowly biodegradable fractions of DOC, $BPOC_1$ & $BPOC_2$ rapidly and slowly biodegradable fractions of POC

In order to establish a relationship between the different fractions of OM and to facilitate the

literature review, the following five ratios (variables), as depicted in Figure 2.1, are introduced to derive each of the OM fractions from top to bottom in terms of TOC and the ratios themselves:

$$t = \frac{DOC}{TOC}$$

$$b_1 = \frac{BDOC}{DOC}$$

$$b_2 = \frac{BPOC}{POC}$$

$$s_1 = \frac{BDOC_1}{BDOC}$$

$$s_2 = \frac{BPOC_1}{BPOC}$$

where,

TOC: total organic carbon (dissolved organic carbon (DOC) + particulate organic carbon (POC)) [mgC L^{-1}]

BDOC: biodegradable DOC (rapidly biodegradable DOC ($BDOC_1$) + slowly biodegradable DOC ($BDOC_2$)) [mgC L^{-1}]

BPOC: biodegradable POC (rapidly biodegradable POC ($BPOC_1$) + slowly biodegradable POC ($BPOC_2$)) [mgC L^{-1}]

RDOC: refractory DOC [mgC L^{-1}]

RPOC: refractory POC [mgC L^{-1}]

From this model, the two initial variables (TOC, DOC) and the five ratios are considered as descriptive variables for the literature review.

2.1.2 OM degradation variables in water quality models

Equations (2.1) and (2.2) represent the degradation of OM in the HSB model [Billen et al., 1988, Billen, 1991]. The degradation of organic matter is limited by the availability of small monomeric substrate ([S]), which is modelled by K_S (mgC L^{-1}), the Monod half-saturation constant for uptake of substrate (eq. (2.1)).

$$uptake_{HB} = \frac{\mu_{max,HB}\ e^{\frac{-(T-T_{opt,HB})^2}{\sigma^2_{HB}}} \frac{[S]}{[S]+K_S}[HB]}{Y_{HB}} \tag{2.1}$$

where, $\mu_{max,HB}$ (h^{-1}) is the maximum growth rate of bacteria, T is the temperature in °C, $T_{opt,HB}$

is the optimum temperature for the growth of bacteria in °C, σ_{HB} is the standard deviation of bacteria temperature function also expressed in °C. S (mgC L^{-1}) is the small monomeric substrate concentration (here, S = $BDOC_1$) and HB (mgC L^{-1}) is the concentration of heterotrophic bacteria.

$k_{hyd,max}$ and K_{BDOC_2} are two other variables (eq. (2.2)) that demonstrate the hydrolysis of $BDOC_2$ into $BDOC_1$.

$$hyd_{BDOC_2} = k_{hyd,max} \frac{[BDOC_2]}{[BDOC_2] + K_{BDOC_2}} [HB] \tag{2.2}$$

Here, hyd_{BDOC_2} (mgC L^{-1} h^{-1}) is the exoenzymatic hydrolysis of $BDOC_2$ into $BDOC_1$ where $k_{hyd,max}$ (h^{-1}) and K_{BDOC_2} (mgC L^{-1}) are the hydrolysis coefficient and the half-saturation constant for hydrolysis of $BDOC_2$ into $BDOC_1$, respectively. Similarly, POC is also hydrolyzed into $BDOC_2$ and $BDOC_1$ based on first-order kinetics; however, its hydrolysis parameters are ignored in this study since their rates are far slower than those of DOC [Garnier et al., 1999].

From this model, the three OM degradation variables (K_s, $k_{hyd,max}$, K_{BDOC2}) are selected for literature review.

2.2 Data collection on OM variables

A total of 150 publications were reviewed to extract information on the variation ranges and average values of the selected OM variables. In addition, the information related to hydrological, climatic, and land cover conditions of the environments where the OM variables had been sampled is also extracted. The following categories are considered for each of the conditions mentioned.

2.2.1 Hydrological categories

The categories considered for the hydrological condition are low flow, high flow, and average flow. Generally, specific discharge thresholds should be defined to differentiate between these three categories of flow. However, due to the lack of information on discharge and drainage area in the reviewed literature, the flow condition was accepted to be what the authors had described.

2.2.2 Land cover categories

The categories considered for land cover are forested, agricultural (vegetated), grasslands, wetlands, urbanized (built-up land), and barren areas (sedimentary plateaus or deserted areas).

2.2.3 Climatic categories

The climate categories are based on the Köppen–Geiger classification system [Köppen, 1884, Beck et al., 2018] where the world is divided into five main categories: tropical (equatorial), dry (arid and semi arid), temperate (warm), continental (snow), and polar (cold).

Table 2.1 - 2.6 detail the synthesis of the reviewed articles for extracting the information on the variation ranges and average values of the selected variables in rivers across the world under the different hydrological, climatic, and land cover conditions. It should be mentioned that most river basins have several land cover types; however, the land cover column in these tables is the dominant land cover type. This was done so in order to facilitate the implementation of the analytical method represented in the next section.

The OM degradation variable such as K_S (Table 2.7 - 2.8) is categorized according to sampling habitats and substrate types as it is a microorganism property.

Table 2.1: Bibliography on TOC concentration (mgC L^{-1}) in experimental studies together with the hydrology, land cover, and climate of the sampled river

No.	Min.	Ref./Ave.	Max.	Hydrology	River location	References	Land cover	Climate
1		8.4		low flow	Clichy CSO, Seine River, France	Seidl et al. [1998b]	urbanized	temperate
2	1.6		7.6	low flow	Meuse River, Wallonia, Belgium	Lambert et al. [2017]	agriculture	temperate
3		2		low flow	Seine River, France	Flipo et al. [2018]	agriculture	temperate
4	2		3.2	low flow	Ups of Seine river, France	Varrault et al. [2016]	agriculture	temperate
5	0.51		1.82	low flow	Wailuku River, Hawaii, USA	Wiegner et al. [2009]	forest	tropical
6		6.55		low flow	Ups of Seine River, France	Servais et al. [1995]	agriculture	temperate
7	2		5	low flow	Rhone River, France	Sempéré et al. [2000]	agriculture	temperate
8	1.11		6.15	low flow	Han River, South Korea	Kim et al. [2013]	forest	continental
9	4		17	low flow	Yellow River, China	Ran [2013]	barren land	dry
10	3.5		8.7	low flow	Ups. of Pearl River estuary, China	He et al. [2010]	urbanized	temperate
11	18		27	low flow	Degero Stormyr, Sweden	Leach et al. [2016]	wetland	continental
12		4.96		low flow	Basse-Terre, Guadeloupe	Lloret et al. [2013]	forest	tropical
13	15		38	high flow	Degero Stormyr, Sweden	Leach et al. [2016]	wetland	continental
14		8		high flow	Seine River, France	Flipo et al. [2018]	agriculture	temperate
15	3.3		37	high flow	Han River, South Korea	Kim et al. [2013]	forest	continental
16	3.6	8	16.9	high flow	Wailuku River, Hawaii, USA	Wiegner et al. [2009]	forest	tropical
17	10		39	high flow	Yellow river, China	Ran [2013]	barren land	dry
18	10		25	high flow	Rhone River, France	Sempéré et al. [2000]	agriculture	temperate
19		6.16		high flow	Basse-Terre, Guadeloupe	Lloret et al. [2013]	forest	tropical
20	1.2		13.4	high flow	Meuse River, Wallonia, Belgium	Lambert et al. [2017]	agriculture	temperate
21	4		5	average flow	Ups of Seine River, France	Varrault et al. [2016]	agriculture	temperate
22	1.77		3.55	average flow	Danube River, Austria	Besemer et al. [2009]	grassland	temperate
23	2.3		8.2	average flow	Po River, Italy	Pettine et al. [1998]	agriculture	temperate
24	2.25		31.2	average flow	Yellow River, China	Zhang et al. [2013]	barren land	dry
25		8.9		average flow	Ups. of Schelde estuary, Belgium	Muylaert et al. [2005]	urbanized	temperate
26		4.9		average flow	Yangtze River, China	Qi [2014]	NA	temperate
27		1.7232		average flow	Columbia River, USA	Hill and Wheeler [2002]	forest	temperate
28		13.6		average flow	Downs. of WWTP, Seine River, France	Servais et al. [1995]	urbanized	temperate
29		9.2		average flow	Meuse River, Belgium	Servais et al. [1995]	urbanized	temperate
30		4.4		average flow	6 rivers, South Korea	Islam et al. [2019]	urbanized	continental
31		3.7		average flow	Saigon River, Vietnam	Nguyen et al. [2019]	agriculture	tropical
32		7.9		average flow	Saigon River, Vietnam	Nguyen et al. [2019]	urbanized	tropical
33	1.73		31.74	average flow	17 rivers, China	Liu et al. [2020]	agriculture	temperate

Table 2.2: Bibliography on DOC concentration (mgC L^{-1}) in experimental studies together with the hydrology, land cover, and climate of the sampled river

No.	Min.	Ref./Ave.	Max.	Hydrology	River location	References	Land cover	Climate
1	1.2		4.9	low flow	Meuse River, Wallonia, Belgium	Lambert et al. [2017]	agriculture	temperate
2		5.6		low flow	Epulu River, DR Congo	Spencer et al. [2010]	forest	tropical
3		4.8		low flow	Richmond River, Australia	Atkins et al. [2017]	forest	temperate
4		0.96		low flow	Capesterre River, Guadeloupe	Lloret et al. [2013]	forest	tropical
5		4.4		low flow	Ups of Seine River, France	Servais et al. [1995]	agriculture	temperate
6	2.4		5.7	low flow	Ups of Pearl estuary, China	He et al. [2010]	urbanized	temperate
7		3.6		low flow	Ob and Yenisei rivers, Russia	Köhler et al. [2003]	grassland	polar
8	5.3		7.3	low flow	Lena River, Russia	Martin et al. [1993]	grassland	continental
9	0.34	0.9	1.52	low flow	Wailuku River, HI, USA	Wiegner et al. [2009]	forest	tropical
10	1.34		4.1	low flow	Rhone River, France	Sempéré et al. [2000]	agriculture	temperate
11	0.87		3.23	low flow	Han River, South Korea	Kim et al. [2013]	forest	continental
12		1.7		low flow	White Clay Creek, PA, USA	Volk et al. [1997]	agriculture	temperate
13		4.00		low flow	Paris to Seine estuary, France	Garnier et al. [2001]	urbanized	temperate
14	6.6		9	high flow	Epulu River, DR Congo	Spencer et al. [2010]	forest	tropical
15		8.52		high flow	Richmond River, Australia	Atkins et al. [2017]	forest	temperate
16		2.16		high flow	Capesterre River, Guadeloupe	Lloret et al. [2013]	forest	tropical
17	2		15	high flow	Juruena River, Brazil	Johnson et al. [2011]	forest	tropical
18	0.9		6.6	high flow	Meuse River, Wallonia, Belgium	Lambert et al. [2017]	agriculture	temperate
19		10.8		high flow	Ob and Yenisei rivers, Russia	Köhler et al. [2003]	grassland	polar
20		10.5		high flow	Lena River, Russia	Martin et al. [1993]	grassland	continental
21	0.7		9.6	high flow	Wailuku River, HI, USA	Wiegner et al. [2009]	forest	tropical
22	0.9		8.25	high flow	Rhone River, France	Sempéré et al. [2000]	agriculture	temperate
23	1.4		5.5	high flow	Han River, South Korea	Kim et al. [2013]	forest	continental
24		9.2		high flow	White Clay Creek, PA, USA	Volk et al. [1997]	agriculture	temperate
25	1.50		3.30	average flow	Ups. of Seine river, France	Varrault et al. [2020]	agriculture	temperate
26	2.50		4.50	average flow	Ds. Of WWTP, Seine, France	Varrault et al. [2020]	urbanized	temperate
27	0.92		1.52	average flow	Minjiang River, China	Yang et al. [2019]	forest	temperate
28		1.98		average flow	Minjiang River, China	Yang et al. [2019]	forest	temperate
29		2.57		average flow	6 rivers, South Korea	Islam et al. [2019]	urbanized	continental
30		2.4		average flow	Saigon River, Vietnam	Nguyen et al. [2019]	agriculture	tropical

Table 2.2 continued .... (bibliography on DOC concentration (mgC L^{-1}))

No.	Min.	Ref./Ave.	Max.	Hydrology	River location	References	Land cover	Climate
31		3.6		average flow	Saigon river, Vietnam	Nguyen et al. [2019]	urbanized	tropical
32	1.25		10.41	average flow	17 rivers, China	Liu et al. [2020]	agriculture	NA
33		5.4		average flow	Rukarara River, Rwanda	Rizinjirabake et al. [2018]	forest	tropical
34		3.86		average flow	Rukarara River, Rwanda	Rizinjirabake et al. [2018]	agriculture	tropical
35	9		19.3	average flow	Stordalen, Sweden	Olefeldt et al. [2013]	wetland	continental
36	1.6		3.8	average flow	Yellow River, China	Zhang et al. [2013]	barren land	dry
37		7.6		average flow	Brandenburg, Germany	Graeber et al. [2012]	agriculture	continental
38		1.5		average flow	Brandenburg, Germany	Graeber et al. [2012]	forest	continental
39		12		average flow	Brandenburg, Germany	Graeber et al. [2012]	wetland	continental
40		5.8		average flow	Mullica River watershed, USA	Wiegner and Seitzinger [2004]	wetland	temperate
41		15.4		average flow	Mullica River watershed, USA	Wiegner and Seitzinger [2004]	wetland	temperate
42		8.45		average flow	Ds. of WWTP, Seine River, France	Servais et al. [1995]	urbanized	temperate
43		5.1		average flow	Meuse River, Belgium	Servais et al. [1995]	urbanized	temperate
44	2.5		6	average flow	Yellow River, China	Ran [2013]	barren land	dry
45	2.8		12	average flow	12 Russian rivers	Lobbes et al. [2000]	grassland	polar
46	3.6		12	average flow	Lena River, Russia	Lara et al. [1998]	grassland	continental
47	1.3		3.7	average flow	Po River, Italy	Pettine et al. [1998]	agriculture	temperate
48	1.2		2.8	average flow	Danube River, Austria	Besemer et al. [2009]	grassland	temperate
49		1.212		average flow	Columbia River, USA	Hill and Wheeler [2002]	forest	temperate
50	1.2		3.2	average flow	Yangtze River, China	Qi [2014]	NA	temperate
51	1.40		3.40	average flow	5 rivers, South Korea	Shin et al. [2016]	forest	continental
52		5.00		average flow	3 rivers in La Loire, France	Martin-Mousset et al. [1997]	agriculture	temperate
53		1.99		average flow	Arbatus Watershed, NY, USA	Kang and Mitchell [2013]	forest	temperate
54		5.04		average flow	Arbatus Watershed, NY, USA	Kang and Mitchell [2013]	wetland	temperate
55	8.86		13.12	average flow	Scheldt River, Belgium	Servais et al. [1987]	urbanized	temperate
56	3.5		4.94	average flow	Meuse River, Belgium	Servais et al. [1987]	agriculture	temperate
57		1.8		average flow	A forest river, Belgium	Servais et al. [1987]	forest	temperate
58	2.10		4.80	average flow	Ups. of estuary, Seine River	Servais and Garnier [2006]	agriculture	temperate

Table 2.3: Bibliography on t (DOC:TOC) variable in experimental studies together with the hydrology, land cover, and climate of the sampled river

No.	Min.	Ref./Ave.	Max.	Hydrology	River location	References	Land cover	Climate
1	0.545		0.911	low flow	Meuse River, Wallonia, Belgium	Lambert et al. [2017]	agriculture	temperate
2		0.82		low flow	Clichy CSO, Seine, France	Seidl et al. [1998b]	urbanized	temperate
3		0.67		low flow	Ups of Seine River, France	Servais et al. [1995]	agriculture	temperate
4		0.67		low flow	Po river, Italy	Pettine et al. [1998]	agriculture	temperate
5	0.62		0.8	low flow	Ups. of Pearl estuary, China	He et al. [2010]	urbanized	temperate
6	0.7		0.83	low flow	Ups. of Seine River, France	Varrault et al. [2016]	agriculture	temperate
7	0.6		0.8	low flow	Marne River, France	Varrault et al. [2016]	agriculture	temperate
8	0.2		0.72	low flow	Yellow River, China	Ran [2013]	barren land	dry
9	0.66	0.79	0.84	low flow	Wailuku River, HI, USA	Wiegner et al. [2009]	forest	tropical
10	0.67		0.82	low flow	Rhone River, France	Sempéré et al. [2000]	agriculture	temperate
11	0.52		0.78	low flow	Han River, South Korea	Kim et al. [2013]	forest	continental
12		0.19		low flow	Basse-Terre, Guadeloupe	Lloret et al. [2013]	forest	tropical
13		0.3		high flow	Po River, Italy	Pettine et al. [1998]	agriculture	temperate
14	0.333		0.884	high flow	Meuse River, Wallonia, Belgium	Lambert et al. [2017]	agriculture	temperate
15	0.09		0.33	high flow	Rhone River, France	Sempéré et al. [2000]	agriculture	temperate
16	0.19		0.57	high flow	Wailuku River, HI, USA	Wiegner et al. [2009]	forest	tropical
17	0.08		0.34	high flow	Yellow River, China	Ran [2013]	barren land	dry
18	0.15		0.42	high flow	Han River, South Korea	Kim et al. [2013]	forest	continental
19		0.35		high flow	Basse-Terre, Guadeloupe	Lloret et al. [2013]	forest	tropical
20	0.08		0.72	average flow	17 rivers, China	Liu et al. [2020]	agriculture	temperate
21		0.55		average flow	Meuse River, Belgium	Servais et al. [1995]	urbanized	temperate
22		0.62		average flow	6 rivers, South Korea	Islam et al. [2019]	urbanized	continental
23		0.46		average flow	Saigon River, Vietnam	Nguyen et al. [2019]	urbanized	tropical
24	0.4		0.83	average flow	Yangtze River, China	Qi [2014]	agriculture	temperate
25	0.12		0.71	average flow	Yellow River, China	Zhang et al. [2013]	barren land	dry
26	0.56		0.82	average flow	Danube River, Austria	Besemer et al. [2009]	grassland	temperate
27	0.56		0.96	average flow	12 Russian rivers	Lobbes et al. [2000]	grassland	polar
28		0.7		average flow	Columbia River, USA	Hill and Wheeler [2002]	forest	temperate
29		0.65		average flow	Saigon River, Vietnam	Nguyen et al. [2019]	agriculture	tropical
30		0.91		average flow	Ob and Yenisei rivers, Russia	Köhler et al. [2003]	grassland	polar

Table 2.4: Bibliography on the biodegradable fraction of DOC, b_1, in experimental studies together with the hydrology, land cover, and climate of the sampled river

No.	Min.	Ref./Ave.	Max.	Hydrology	River location	References	Land cover	Climate
1	0.1		0.38	low flow	Seine River, France	Varrault et al. [2016]	agriculture	temperate
2	0.18		0.4	low flow	Seine River, France	Varrault et al. [2020]	agriculture	temperate
3		0.4		low flow	Clichy CSO, Seine, France	Seidl et al. [1998b]	urbanized	temperate
4		0.27		low flow	Ups. of Seine river, France	Servais et al. [1995]	agriculture	temperate
5		0.4		low flow	Paris to Seine estuary, France	Garnier et al. [2001]	urbanized	temperate
6	0.18		0.24	low flow	8 streams in Alabama, USA	Shang et al. [2018]	agriculture	temperate
7		0.42		low flow	Ups. of Pearl estuary, China	He et al. [2010]	urbanized	temperate
8	0.11		0.36	low flow	Minjiang River, China	Yang et al. [2019]	urbanized	temperate
9		0.24		low flow	White Clay Creek, PA, USA	Volk et al. [1997]	agriculture	temperate
11		0.15		low flow	Wailuku River, Hawaii, USA:	Wiegner et al. [2009]	forest	tropical
12	0.33		0.54	low flow	Youngsan River, South Korea	Shin et al. [2016]	forest	continental
13		0.4		low flow	Llobregat River, Spain	Ribas et al. [1997]	forest	temperate
14	0.1		0.14	low flow	8 streams in Alabama, US	Shang et al. [2018]	forest	temperate
15		0.06		high flow	Wailuku River, Hawaii, USA:	Wiegner et al. [2009]	forest	tropical
16		0.06		high flow	Seine River, France	Varrault et al. [2020]	agriculture	temperate
17		0.26		high flow	White Clay Creek, PA, USA	Volk et al. [1997]	agriculture	temperate
18		0.05		high flow	Minjiang River, China	Yang et al. [2019]	forest	temperate
19	0.05		0.16	high flow	Youngsan River, South Korea	Shin et al. [2016]	forest	continental
20		0.14		high flow	Llobregat River, Spain	Ribas et al. [1997]	forest	temperate
21		0.12		average flow	Mullica River watershed, USA	Wiegner and Seitzinger [2004]	wetland	temperate
22	0.27		0.54	average flow	Scheldt River, Belgium	Servais et al. [1987]	urbanized	temperate
23		0.55		average flow	Downs. of wWTP, Seine, France	Servais and Garnier [1993]	urbanized	temperate
24	0.14		0.23	average flow	3 rivers in La Loire, France	Martin-Mousset et al. [1997]	agriculture	temperate
25	0.15		0.29	average flow	Ups. of Seine estuary, France	Servais and Garnier [2006]	agriculture	temperate
26		0.26		average flow	6 rivers, South Korea	Islam et al. [2019]	urbanized	continental
27		0.22		average flow	Mullica River watershed, USA	Wiegner and Seitzinger [2004]	wetland	temperate
28	0.19		0.35	average flow	Meuse River, Belgium	Servais et al. [1987]	agriculture	temperate
29	0.12		0.21	average flow	Seine River, France	Tusseau-Vuillemin et al. [2003]	agriculture	temperate
30		0.18		average flow	Arbatus Watershed, NY, USA	Kang and Mitchell [2013]	forest	temperate
31	0.05		0.54	average flow	5 rivers, South Korea	Shin et al. [2016]	forest	continental
32		0.11		average flow	A forest river, Belgium	Servais et al. [1987]	forest	temperate
33	0.05		0.22	average flow	Rhone River, France	Sempéré et al. [2000]	agriculture	temperate
34		0.07		average flow	Arbatus Watershed, NY, USA	Kang and Mitchell [2013]	wetland	temperate
35	0.05		0.27	average flow	Ups. of Schelde estuary, Belgium	Muylaert et al. [2005]	urbanized	temperate

Table 2.5: Bibliography on the biodegradable fraction of POC, b_2, in experimental studies with their hydrology, land cover, and climate

No.	Min.	Ref./Ave.	Max.	Hydrology	River location	References	Land cover	Climate
1		0.26		low flow	Ups. of Seine river, France	Servais et al. [1995]	agriculture	temperate
2	0.13		0.29	low flow	Danube River, Romania	Reschke et al. [2002]	urbanized	continental
3		0.6		low flow	Ups of Garonne estuary, France	Etcheber et al. [2007]	agriculture	temperate
4		0.61		low flow	Ups of Loire estuary, France	Etcheber et al. [2007]	agriculture	temperate
5	0.04		0.18	TSS> 150 mg/l	13 world rivers	Ittekkot [1988]	NA	NA
6	0.22	0.35	0.46	TSS< 150 mg/l	13 world rivers	Ittekkot [1988]	NA	NA
7		0.3		average flow	Meuse River, Belgium	Servais et al. [1995]	urbanized	temperate
8	0.14		0.23	average flow	Rhone River, France	Sempéré et al. [2000]	agriculture	temperate
9		0.35		average flow	world rivers	Spitzy and Ittekkot [1991]	NA	NA
10	0.13		0.52	average flow	Ups. of Seine estuary, France	Servais and Garnier [2006]	agriculture	temperate
11	0.2		0.47	average flow	Ups. of Garonne estuary, France	Etcheber et al. [2007]	agriculture	temperate
12		0.4		average flow	6 rivers, South Korea	Islam et al. [2019]	urbanized	continental
13		0.26		average flow	Seine River, France	Tusseau-Vuillemin et al. [2003]	agriculture	temperate

Table 2.6: Bibliography on the rapidly biodegradable fraction of BDOC and BPOC, s_1 and s_2, respectively

parameter	No	Min.	Ref./Ave.	Max.	Location	References
s_1	1	0.52		0.66	Fluvia River, Spain	Catalán et al. [2017]
	2		0.55		Ups. of estuary, Seine River	Servais and Garnier [2006]
	3	0.51		0.95	Raw influent to Acheres, France	Block et al. [1992]
	4	0.30		0.40	Mountainous stream, NC, USA	Qualls and Haines [1992]
	5	0.54		0.90	Red River, Vietnam	Nguyen et al. [2018]
	6		0.53		Urban effluent, Brussels, Belgium	Servais et al. [1995]
	7		0.62		Loch Creran coast, Scotland	Lønborg et al. [2009]
	8		0.90		Tana River, Kenya	Geeraert et al. [2016]
s_2	1		0.65		Ups. of the Seine estuary, France	Servais and Garnier [2006]
	2	0.44		0.95	Red River, Vietnam	Nguyen et al. [2018]
	3		0.52		Urban effluent, Brussels, Belgium	Servais et al. [1995]

Table 2.7: Bibliography on the Monod half-saturation constant for uptake of substrate,K_s, (mgC L^{-1}) in experimental studies together with the type of habitat or sampled environment and substrate type

No.	Min.	Ref./Ave.	Max.	Substrate	Habitat/sample source	Reference
1	0.014		0.144	glucose	mix habitat	Sepers [1977]; Billen [1984]
2		0.036		glucose	river	Servais [1986]
3		0.1		glucose	river	Servais et al. [1989]
4	0.001		0.02	glucose	estuary	Crawford et al. [1974]
5	0.0088		0.034	glucose	sea	Seki et al. [1972]
6		0.025		glucose	sea	Seki et al. [1974]
7	0.0088		0.027	glucose	sea	Seki et al. [1975]
8		0.0052		glucose	sea	Wright and Shah [1975]
9	0.006		0.22	glucose	mix habitat	Billen and Somville [1982]
10		0.012		glucose	sea	Vaccaro and Jannasch [1967]
11	0.015		0.275	glucose	sea	Hamilton and Preslan [1970]
12		0.008		glucose	sea	Takahashi and Ichimura [1971]
13	0.001		0.007	glucose	estuary	Hoppe [1978]
14	0.001		0.02	glucose	estuary	Gocke et al. [1981]
15		0.16		glucose	river	Brailsford et al. [2019]
16		0.26		glucose	river	Brailsford et al. [2019]
17	0.003		0.1	glucose	sea	Vaccaro and Jannasch [1966]
18		0.08		glucose	sea	Vaccaro and Jannasch [1966]
19	0.0128		0.0895	galactose	sea	Seki et al. [1972]
20		0.1459		galactose	sea	Seki et al. [1974]
21	0.0034		0.0136	galactose	sea	Seki et al. [1975]
22	0.0054		0.073	glycolate	sea	Wright and Shah [1975]
23		0.024		glycolate	river	Wright and Shah [1975]
24	0.0144		0.12	glycolate	mix habitat	Sepers [1977]; Billen [1984]
25		0.1		protein	estuary	Billen [1991]
26		0.1		protein	river	Garnier et al. [1992a]
27	0.0014		0.045	protein	mix habitat	Sepers [1977]; Billen [1984]
28	0.0024		0.12	acetate	mix habitat	Sepers [1977]; Billen [1984]
29		0.0024		acetate	sea	Wright and Shah [1975]
30	0.0024		0.024	acetate	sea	Billen and Somville [1982]
31	0.024		0.48	acetate	river	Billen and Somville [1982]

2.3 Factorial analysis of mixed data (FAMD)

On the one hand, it is not possible to find enough data in the literature to provide independent variation ranges of the OM variables for the different categories of hydrology, climate, and land cover. On the other hand, it is almost impossible to separate the effect of hydrology from land cover or climate on an OM value of, e.g., DOC concentration. Therefore, it is necessary to initially conduct a dimensionality reduction method that could identify the most important categories representing the overall variability for which variation ranges will be proposed in this work. Since the tabulated data of OM variables include both continuous data (quantitative, e.g., average values) and categorical data (qualitative, e.g., land cover or climate), the factorial analysis of mixed data (FAMD) [Pagès, 2014] is selected to transform the dataset into a new space where novel synthetic variables called 'dimensions'

Table 2.8: Bibliography on the coefficient of organic matter hydrolysis ($k_{hyd,max}$ in h^{-1}) and the half-saturation constant for hydrolysis of $BDOC_2$ into $BDOC_1$ (K_{BDOC2} in mgC L^{-1})

Parameter	No.	Min.	Ref./Ave.	Max.	References	Comments
K_{BDOC2} (mgC L^{-1})	1		2.5		Billen [1991]	model adjusted on exper. data
	2		2.5		Servais et al. [1989]	Meuse River OM experiment
	3	0.2		1.5	Wang et al. [2019]	ProSe-PA model calibration
$k_{hyd,max}$ (h^{-1})	1		0.25		Billen [1991]	model adjusted on exper. data
	2		0.25		Servais et al. [1989]	Meuse River OM experiment
	3	0.25		0.75	Wang et al. [2019]	ProSe-PA model calibration

are created. It uncovers the relationship between dataset variables (e.g., average values, land cover) and enables the identification of important categories that drive the overall variability in the dataset.

FAMD has been applied to find the main drivers of hydrological variability in different catchments [Adamovic et al., 2016], to study the effect of biochar on soil types [Fetjah et al., 2022], to determine the important features for crop yield efficiency [Basset-Mens et al., 2019], and it has also been applied in other studies related to agriculture, medicine, and road safety [Berger et al., 2019, González-Quintero et al., 2020, Abdelnour et al., 2022, Gonçalves et al., 2022].

In order to avoid confusion for readers between OM variables (e.g., DOC, b_1) and FAMD variables (e.g., average values, land cover), the following explanation is necessary. First the FAMD variables which are the columns of each dataset (Table 2.1 - 2.6) are called factors in this work. Each factor is divided into categories. For instance the categories of the hydrology factor are low flow, high flow and average flow. Finally, each category or group of categories, located in the same area of the FAMD space, that contribute significantly to the overall variance of each OM variable are called clusters.

In this study, the *Facto-MineR* package [Lê et al., 2008] in *R* is used to conduct FAMD and the *Factoextra* package [Kassambara and Mundt, 2020] in *R* is used to visualize the FAMD results. The individuals (table lines of literature data) for which a specific land cover or climate could not be associated, are removed from the analysis. The factors (table columns) used in this analysis are average (Ref./Ave. column in Table 2.1) as quantitative data, together with the categories that describe the three factors (hydrology, land cover, and climate). In this study 20 individuals are considered as the minimum for conducting FAMD. Thus, b_2, s_1, s_2, $k_{hyd,max}$, and K_{BDOC_2} are excluded from FAMD, however their overall variation ranges are proposed in section 3.1.

Once FAMD is conducted and the clusters of each OM variable are identified, the detailed tables of each OM variable are filtered according to those clusters for calculating their variation ranges and mean values.

3 Results

In this section, we look first at the overall average values and variation ranges of all OM variables without considering the hydrology, land cover and climate conditions. Next, clusters related to hydrology, land cover and climate conditions are identified in the dataset using FAMD for five OM variables (DOC, TOC, t, b_1, and K_S). It is followed by the presentation of the average values and variation ranges of the identified clusters.

3.1 Overall average values and variation ranges of OM variables

Table 2.9 presents the overall average values and variation ranges of all OM variables introduced in this study based on the collected dataset. This table could be used by river metabolism researchers and water quality modelers in order to have a general view on the variation of variables worldwide. However, if narrower ranges are needed under different categories of climate, hydrology and land cover, they can refer to section 3.3 where variation ranges are proposed only for the clusters found by FAMD.

Table 2.9: Overall average values and variation ranges of OM variables

	DOC [$mgCL^{-1}$]	TOC [$mgCL^{-1}$]	t [-]	b_1 [-]	b_2 [-]	s_1 [-]	s_2 [-]	K_S [$mgCL^{-1}$]	K_{BDOC2} [$mgCL^{-1}$]	$k_{hyd,max}$ h^{-1}
Min.	0.3	0.5	0.08	0.05	0.13	0.3	0.44	0.001	0.2	0.25
Avg.	5.0	9.0	0.57	0.23	0.35	0.59	0.62	0.073	2.0	0.33
Max.	19.3	39.0	0.96	0.54	0.52	0.95	0.95	0.6	2.50	0.75

t = DOC:TOC, share of the dissolved part of TOC.
b_1 = BDOC:DOC, share of the biodegradable part of DOC.
b_2 = BPOC:POC, share of the biodegradable part of POC.
s_1 = $BDOC_1$:BDOC, share of the rapidly biodegradable part of BDOC.
s_2 = $BPOC_1$:BPOC, share of the rapidly biodegradable part of BPOC.

3.2 Identification of clusters among OM variables

Figure 2.2A shows the result of FAMD for TOC where the FAMD factors are displayed on the new space. Land cover and climate are the main qualitative factors that stand out and contribute to the overall variance. In addition, dimensions 1 and 2 present 45% of the overall variance. Figure 2.2B displays the results of FAMD for the factor categories of TOC in the new space. They are highlighted based on their contributions to dimensions 1 and 2. The identified clusters of TOC are shown in capital letters upon investigation of additional dimensions (dim. 1, 2, and 3 consisted of more than 60 % of overall variance). The identified clusters are first of all wetland and continental climate. Even though they overlap in Figure 2.2B, barren land–dry climate also form a paired cluster. Another

paired cluster for tropical climate–forest could be identified thanks to information on dimension 3. Other categories, especially those of hydrology, are very close to the origin, and thus they cannot be considered as playing a significant role in the discrimination of carbon content.

The FAMD results of DOC and t (DOC:TOC) variables are presented in A1 where similar relationships among the factors and similar clusters were also identified for those two variables. The clusters for DOC are wetland, grassland, polar climate, dry climate–barren land, and tropical climate–forest categories. Except wetland, the same clusters are identified for the t ratio (DOC:TOC) as well.

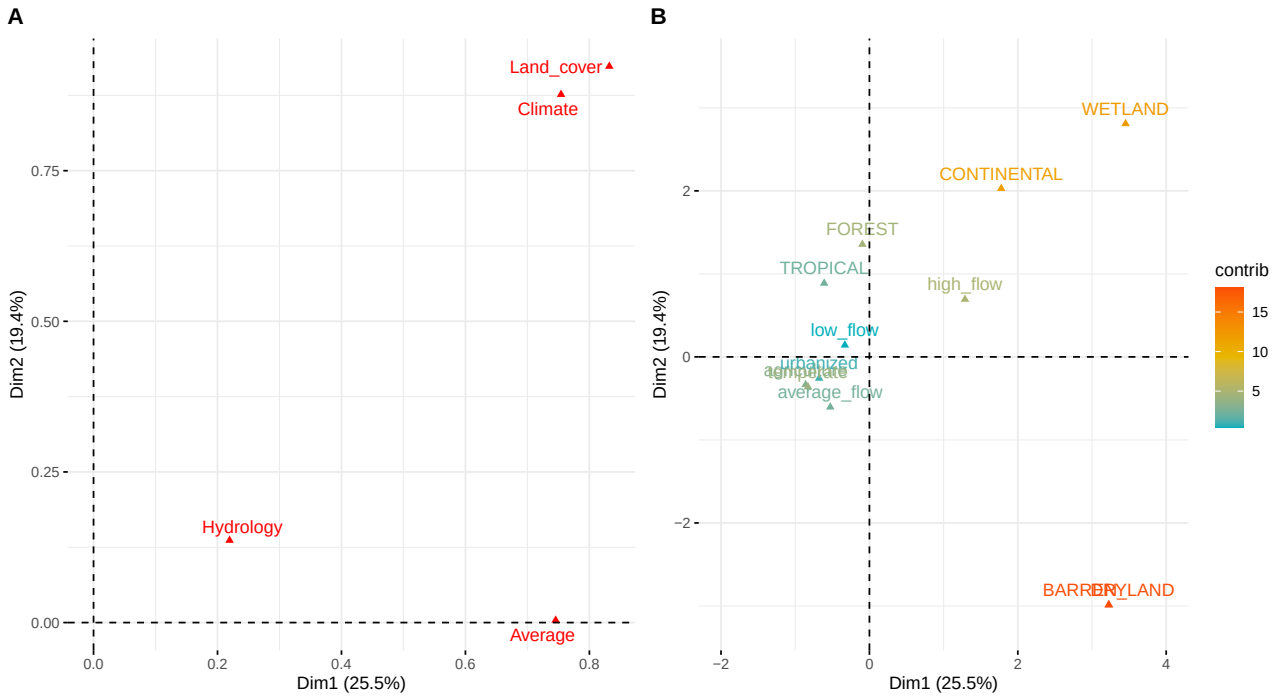


Figure 2.2: FAMD results for TOC. A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters.

Figure 2.3 displays the results of FAMD for the biodegradable fraction of DOC, b_1, on dimensions 1 and 2 that present 45% of the overall variance. Compared to TOC, DOC, and the proportion of DOC in TOC (t), hydrology explains the clustering of b_1 more than climate.

Upon investigation of dimensions 1, 2, and 3 (more than 60 % of overall variance), the identified clusters for b_1 are urbanized, wetland, agriculture, tropical climate, continental climate, and high-flow categories.

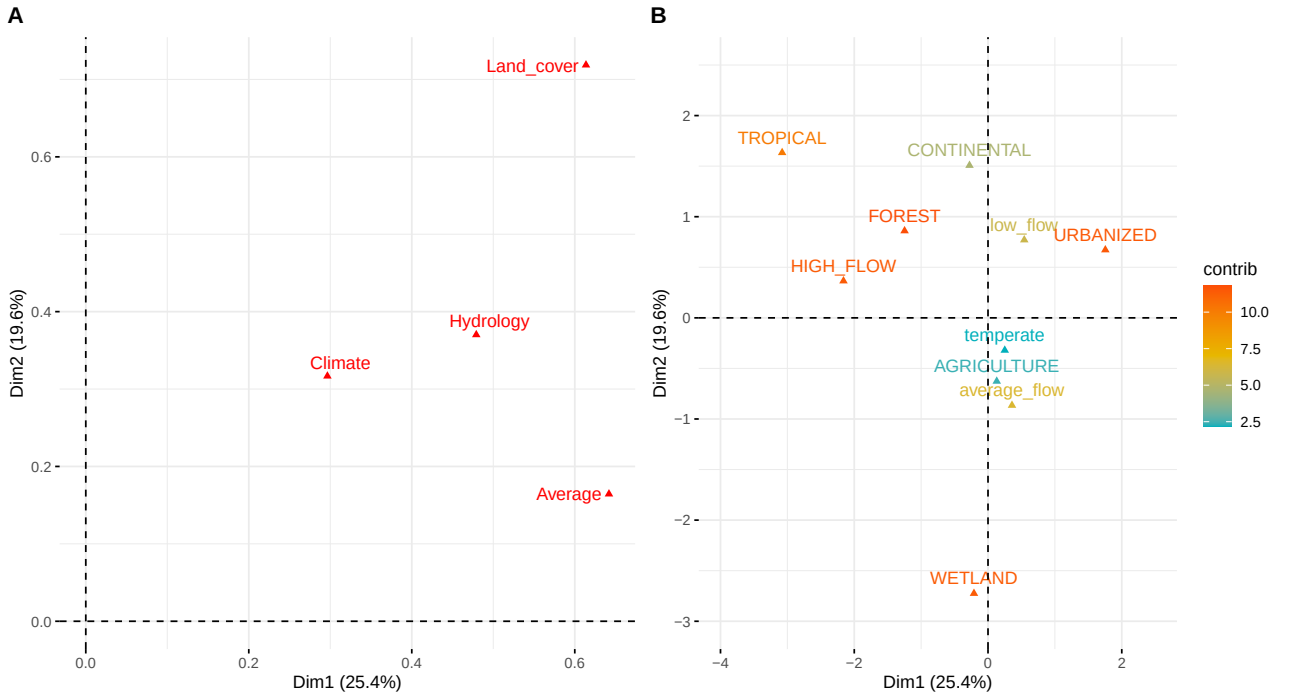


Figure 2.3: FAMD results for b_1. A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters.

Similarly, FAMD was also implemented on the K_S dataset (Figure A4) whose qualitative factors are habitat (sampling environment) and substrate (the material taken up by the heterotrophic bacteria). The identified clusters for K_S are river, estuary, and sea habitats together with galactose and protein substrates.

3.3 Average values and variation ranges per cluster

Table 2.10 lists the average values and variation ranges of the clusters that are identified for each OM variable in the previous section. For this purpose, the detailed tables of the OM variables are filtered for each cluster. For individuals that only had average values in the dataset, a 30 % coefficient (0.7 and 1.3) was used to estimate their minimum and maximum values, respectively. Then, the average of the minimum and maximum columns was calculated to estimate the variation ranges.

The average values and variation ranges for the clusters related to K_S are as follows: an average value of 0.133 mgC L^{-1} and a range of 0.071-0.195 mgC L^{-1} in rivers, an average value of 0.031 mgC L^{-1} and a range of 0.018-0.044 mgC L^{-1} in estuaries, and an average value of 0.042 mgC L^{-1} and a range of 0.017-0.067 mgC L^{-1} in marine coastal areas (sea). For substrate clusters, an average value of 0.069 mgC L^{-1} and a range of 0.039-0.098 mgC L^{-1} for galactose, and an average value of 0.074 mgC L^{-1} and a range of 0.047-0.102 mgC L^{-1} for proteins.

Table 2.10: Average values and variation ranges of OM variables for the identified clusters of land cover, climate, and hydrology

Cluster description	DOC [mgC L^{-1}]			TOC [mgC L^{-1}]			t [-] *			b_1 [-] **		
	min	avg	max	min	avg	max	min	avg	max	min	avg	max
wetland	7.2	10.5	13.8	16.5	24.5	32.5				0.10	0.14	0.18
grassland	4.3	6.9	9.5				0.59	0.79	0.89			
urbanized										0.23	0.35	0.48
agriculture										0.14	0.21	0.28
forest										0.11	0.18	0.24
polar climate	4.3	7.3	10.2				0.60	0.84	0.96			
continental climate				8.1	15.4	22.8				0.15	0.27	0.39
tropical climate										0.07	0.11	0.14
high flow hydrology										0.07	0.11	0.15
dry climate–barren land	2.1	3.5	4.9	5.4	17.2	29.1	0.13	0.36	0.59			
tropical climate–forest	2.4	4.6	6.7	3.0	5.1	8.3	0.31	0.43	0.53			

* t = DOC:TOC, share of the dissolved part of TOC.
** b_1 = BDOC:DOC, share of the biodegradable part of DOC.

4 Discussion

The results of this study contribute to the studies of river metabolism, as the presented OM variables control carbon fluxes in the river system. River metabolism researchers and water quality modelers can refer to Table 2.10 in order to obtain the variation ranges of their target OM variables if any of the climatic, hydrological, or land cover characteristics of the river under study matches with at least one or more of those identified clusters in the table. In the case of overlap between multiple clusters, the ranges could be crossed to get a wider range. However, overall variation ranges in Table 2.9 are proposed to be used in case of no match between the river characteristics and identified clusters.

In this section, we elaborate further on the identified clusters under FAMD and on the proposed variation ranges with related findings by the scientific community. As the results of this study demonstrated the quantitative variation of OM variables, we discuss the uncertainties in the measurement of these OM variables. Then we proceed to address a research gap in water quality modeling that is related to one of the main causes of uncertainty in OM measurement. Finally, we evaluate the role and value of qualitative analysis of OM compared to the quantitative approach presented in this paper.

4.1 Comparison of the FAMD results and the proposed variation ranges with the literature

As this is the first implementation of the FAMD method on OM variables in the literature, no similar study could be found for direct comparison with the results of this work. Nevertheless, the results of

the FAMD reveal the control of DOC, TOC, and t by climate and land cover conditions (Figure 2.2A, A1A and A3A). This is in accordance with the findings of Meybeck [1982] and Mulholland [2003], who consider that OM concentration is controlled by the aforementioned features. In addition, their first dimensions are mostly driven by the average factor as it is aligned with the dimension-1 axis. It means that along this axis, a gradient exists from high to low values of TOC, DOC and t variables. This could be verified in Figure 2.4 which shows the distribution of TOC individuals on dimensions 1 and 2 per FAMD factor together with their confidence ellipses. Points on the right side are bigger than the ones on the left side, meaning that the highest average TOC values are structured along with wetland, barren land, dry climate, and continental climate clusters. Similar results are found for t and DOC variables in Figure A5, and A6, respectively.

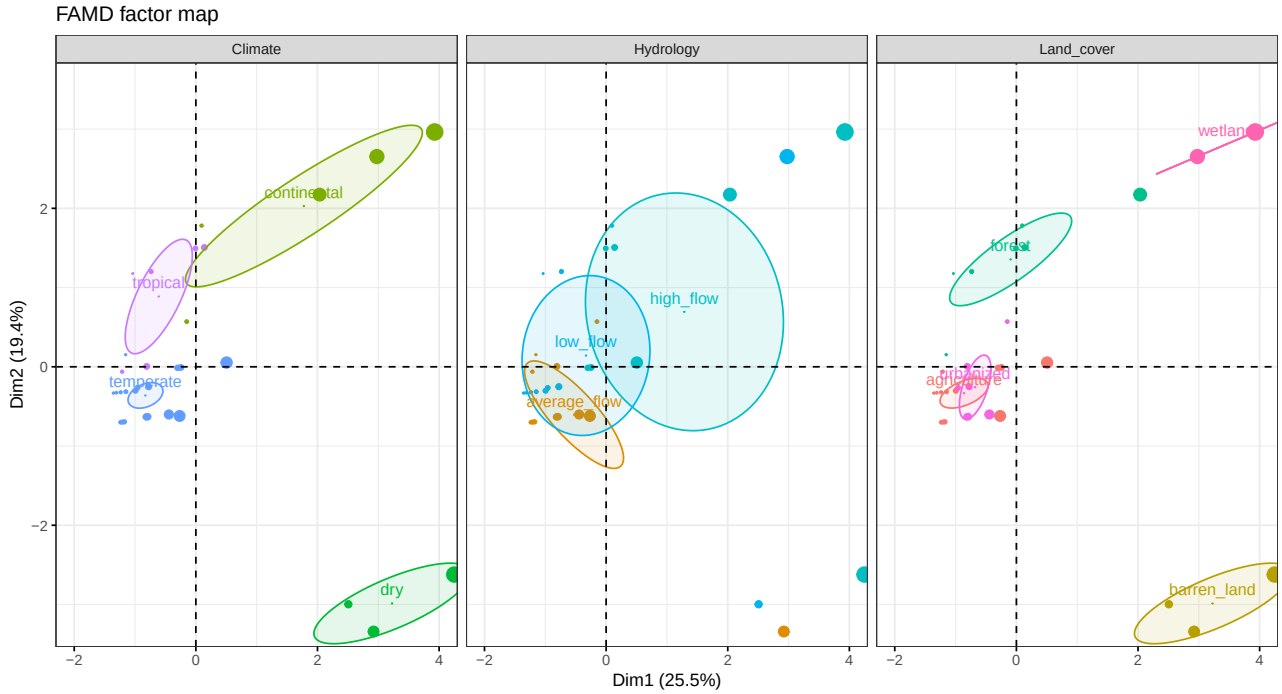


Figure 2.4: Distribution of the TOC individuals on dimensions 1 and 2 per factor with the confidence ellipses. The point size represents the average TOC value and the colors represent the different categories of the three qualitative factors (climate, hydrology, and land cover)

This tendency is explained by the highest average and range values of TOC (average of 24.5 mgC L^{-1} and a range of 16.5–32.5 mgC L^{-1}) and DOC (average of 10.5 mgC L^{-1} and a range of 7.2–13.8 mgC L^{-1}) in wetlands (Table 2.10). Indeed, wetlands are the most carbon-abundant land cover type [Lobbes et al., 2000, Shin et al., 2016]. In addition, wetlands have the highest concentration time (T_c) (travel time of the rainfall to the outlet of a watershed), and an increase in T_c is reported to decrease the runoff intensity [Rizinjirabake et al., 2018]. This may be another reason for the higher ranges of DOC and TOC in wetlands .

Dry-barren land has the second highest TOC average (17.2 mgC L^{-1}) and range values (5.4–29.1 mgC L^{-1}) and also contribute to dimension 1. However, dry-barren lands have lower DOC average (3.5 mgC L^{-1}) and range values (2.1–4.9 mgC L^{-1}) compared to other clusters (Table 2.10), which could be attributed to the higher POC content of barren lands and to their higher soil erosion rates [Ran, 2013]. This could be confirmed by the lower t average values of dry-barren lands (0.36). Continental climate is also important on dimension 1 of the TOC FAMD result. Its relatively high average (15.4 mgC L^{-1}) and range values (8.1–22.8mgC L^{-1}) could be due to the flushing of OM during the melting season [Holmes et al., 2008]. This is in accordance with the findings of Schmidt et al. [2011] that permafrost soils hold a huge amount of carbon that later on gets exposed to intense mineralization during the thawing season. The relatively higher DOC of continental regions may also be due to the high DOC content of the taiga regions [Lara et al., 1998]. However, the tropical–forest cluster has relatively lower average (5.1 mgC L^{-1}) and variation range values (3.0–8.3 mgC L^{-1}) and does not contribute on dimension 1.

Looking back at the FAMD results of b_1 (biodegradability of dissolved organic carbon) (Figure 2.3), once again, the average factor aligns with the first dimension. This is confirmed on Figure 2.5 where the urbanized land cover cluster is aligned on the right side of dimension 1 together with the average b_1 values. That is why high average (0.35) and range values (0.23–0.48) are found for b_1 in urbanized areas (Table 2.10). This is in accordance with the findings of Servais and Garnier [1993], Seidl et al. [1998b] and Servais et al. [1999], demonstrating that the releases from urbanized areas increase b_1 in the rivers. Hosen et al. [2014] state that there is a positive relationship between watershed impervious cover percentage (urbanized) and b_1.

Moreover, the high flow cluster is aligned along the left side of dimension 1 and thereby appears to have smaller average (0.11) and range values (0.07–0.15) compared to other clusters. This is explained by the dominance of low bioavailable [Wang et al., 2010], high-molecular-weight and hydrophobic OM [Kang and Mitchell, 2013, Varrault et al., 2016] at high flow periods during which interactions are noticeable between land cover and hydrology [Moatar et al., 2017]. High-molecular-weight OM has to undergo exo-enzymatic hydrolysis prior to the uptake by bacteria [Rogers, 1961]. Similarly, the tropical cluster also has small mean (0.11) and range values (0.07–0.14) and is placed on left side of dimension 1. This can be explained by an increased bacterial activity under tropical climate, which leads to the consumption of a significant amount of the biodegradable material [Gmach et al., 2020]. Therefore, lower b_1 values would be found in rivers of such regions.

The remaining four clusters for b_1 (wetland, agriculture, forest, and continental climate) do not

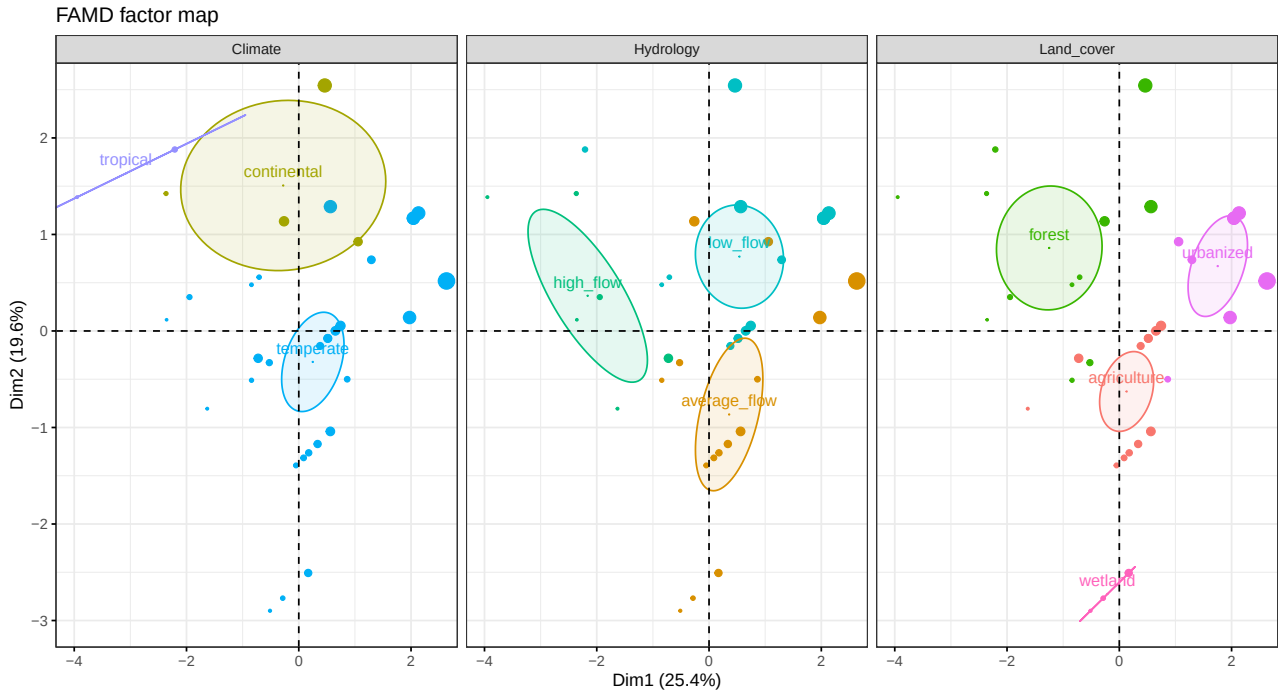


Figure 2.5: Distribution of the b_1 individuals on dimensions 1 and 2 with the confidence ellipses. The point size represents the average b_1 value and the colors represent the different categories of the three qualitative factors (climate, hydrology, and land cover)

contribute significantly to dimension 1 and thus have range values between those of urbanized and high flow clusters. For b_1 in agricultural areas, contradicting statements have been found in the literature. An increase in b_1 with the percentage of agricultural land has been reported by Graeber et al. [2015], Shang et al. [2018], while other studies demonstrate no impact of agricultural land on b_1 [Lu et al., 2013]. b_1 values in continental areas have the second highest mean (0.27) and variation range (0.15–0.39), as reported by Holmes et al. [2008], they argue that the high b_1 under continental and colder climates is due to flow mainly taking place during the spring snowmelt when both DOC and b_1 increase. Indeed, during the snowmelt period, the flow interacts with carbon-rich surface soil and transfers it to the river. In addition, bacterial activity is limited in cold temperatures, which prevents BDOC degradation. Therefore, b_1 is considered to increase in arctic rivers such as in Alaska, USA [Holmes et al., 2008]. Other procedures such as microbial degradation and photodegradation also have an impact on b_1 [Olefeldt et al., 2013], but they have not been sufficiently examined in the literature.

4.2 Uncertainties in characterizing OM variable values

The different OM variables contain various uncertainty and complexity sources that induce an uncertainty in the measured values. Water quality models and conventional tests therefore need to

accommodate these uncertainties.

These uncertainties stem from the OM origin. OM in aquatic environments originates from various sources that could be either autochthonous, i.e., in situ material such as phytoplankton and microorganisms, or allochthonous, i.e., from external sources such as OM carried by runoff from croplands or household and industrial wastes discharged to the environment. OM spatial heterogenity is another source of uncertainty. As OM moves downstream in a river, its heterogeneity increases due to the contribution of various OM sources, river hydraulics, effluent content, and biogeochemical processes [Zhang et al., 2021]. Mayer [1995] explains that the biodegradability or refractoriness of OM is not an intrinsic property, but rather a consequence of the interactions between the material and the surrounding environment [Arndt et al., 2013].

Other uncertainties are the methods to determine OM variables. If we consider BDOC, there are generally three types of methods for its determination in both water and soil samples: (i) incubation batch experiments that rely on the disappearance of DOC after 45 days and for which a natural inoculum is used [Servais et al., 1987, 1995, Khan et al., 1998]; (ii) bioreactors where bacteria are grown on glass beads through which the sample is passed [Kaplan and Newbold, 1995] requiring 3–5 months to produce colonization in the bioreactors [McDowell et al., 2006]; and (iii) kinetic methods where BDOC is estimated using an equation for the kinetics of DOC removal [Gregorich et al., 2003] or the kinetics of CO_2 evolution over a long period [Qualls, 2005].

Furthermore, OM characterization protocols have several limitations that induce uncertainties in the results and makes them incomparable since they are not conducted under similar conditions. The uncertainties are due to:

1. Incubation temperature, which usually varies between 20 and 25 °C and even up to 37 °C [Khan et al., 1999]. Higher temperature causes a faster degradation of BDOC [Khan et al., 1999].

2. Incubation duration, which is long and highly variable among studies (5–45 days up to several months) [Khan et al., 1999, Kadjeski et al., 2020, Kazmiruk et al., 2021] and requires effort to be conducted. Longer incubation tests will remove more DOC and thereby would result in higher BDOC concentrations [McDowell et al., 2006].

3. Inoculum type and volume, which differ from one protocol to another [Kadjeski et al., 2020]. An increase in inoculum is considered to decrease the incubation time. Khan et al. [1999].

4. Incubation type: suspended, sediment-attached, and glass-bead-attached inocula were compared and produced different biodegradabilities [Trulleyová and Rulík, 2004].

5. Sample pretreatment, sample filtration, and extraction are other sources of divergence among different methods [Zhang et al., 2021]. Gonsior et al. [2013] show that DOC is lower in surface samples compared to those just above the sediment layer. At the sediment/water interface, diffusion of DOC from the sediment and re-suspension of sediment occur, which increases the amount of DOC.

6. Light conditions during incubation is also important. Incubation tests are conducted under dark conditions whereas the *in situ* conditions include the effect of photodegradation, which is reported to increase the biodegradability of DOC [Mayer et al., 2011, Catalán et al., 2013, Schiebel et al., 2015] under the photodissolution process and that of POC under the photoflocculation process [Helms et al., 2013].

Considering these uncertainties and their role in the characterization of OM variables, there is a need for a universal method that could provide precise results within minimum time, cost, and effort and must be drawn from all the existing methods. Thus, efforts shall be made toward incorporation of these uncertainty sources in water quality models.

4.3 New direction accounting for photodegradation in water quality modeling

Having described the role of photodegradation as one of the sources of uncertainty in OM characterization, here it is considered as a research gap in OM modeling and a simplistic model is proposed for its integration in the water quality models.

None of the existing OM and water quality models integrate photodegradation, which is an environmental controlling factor whose contribution to OM dynamics is overwhelming according to the literature. For example, ProSe [Even et al., 1998, Flipo et al., 2004, Vilmin et al., 2015], a mechanistic hydro-biogeochemical model, considers the OM refractory fractions to be static whereas photodegradation converts refractory POC to DOC or to biodegradable dissolved fractions (BDOC) [Hu et al., 2022] thanks to ferric ions and nitrates acting as catalysts [Kieber et al., 2006, Koehler et al., 2016]. Indeed, the decomposing effect of solar radiation photons on photo-sensitive fractions of POC and DOC [Zhang et al., 2021] under *in situ* conditions modifies the composition and molecular size of OM, thereby transforming high-molecular-weight compounds to biodegradable low-molecular-weight compounds [Hansen et al., 2016]. Interestingly, photodegradation increases the bacterial activity up to 6-fold [Moran and Zepp, 1997]. This may be one of the reasons for the slight O_2 overestimation during the summer low-flow period by the river quality model ProSe-PA [Wang et al., 2022a].

Photodegradation could be included in the model as a function of influential factors such as solar irradiation (wavelength, time length), lightning events, presence of nitrates and ferric ions, light penetration [Hu et al., 2022], pH [Guo et al., 2022], and temperature [Porcal et al., 2015]. It is imperative that theoretical and experimental studies as well as calibration and validation efforts be carried out by the scientific community in order to determine the necessary equations for the role of each parameter in photodegradation.

On the basis of the reviewed literature, photodegradation (Figure 2.6) could be integrated in the water quality models using first-order kinetic equations for DOC and POC as proposed by Porcal et al. [2015] (eqs 2.3 and 2.4):

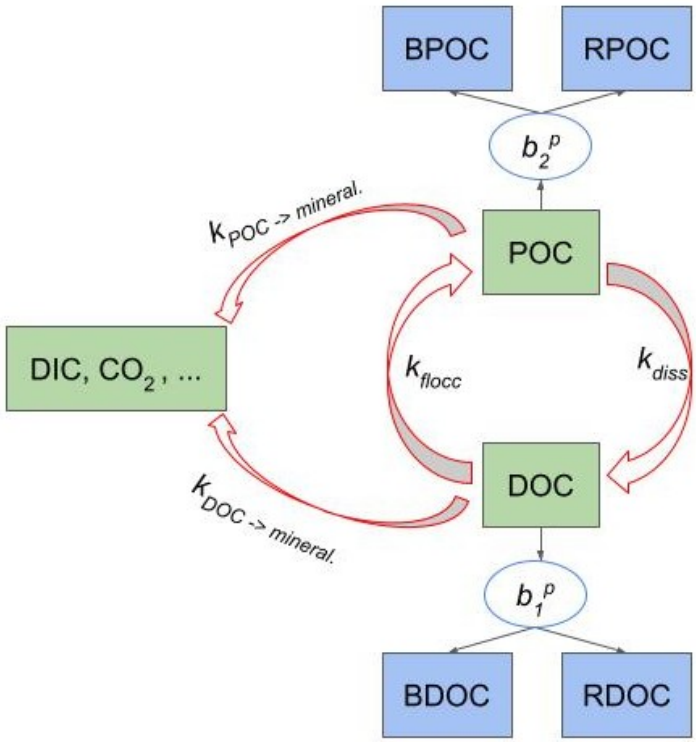


Figure 2.6: Photodegradation model (photodissolution and photoflocculation); DOC: dissolved organic carbon, POC: particulate organic carbon, BDOC: biodegradable DOC, RDOC: refractory DOC, BPOC: biodegradable POC, RPOC: refractory POC, DIC:dissolved inorganic carbon, k_{flocc}: photoflocculation rate, k_{diss}: photodissolution rate, $k_{DOC->mineral}$: transformation rate of DOC to minerals, $k_{POC->mineral}$: transformation rate of POC to minerals, b_1^p: b_1 value of the photodegraded DOC, b_2^p: b_2 value of the photodegraded POC

$$\frac{dDOC_{photo}}{dt} = k_{diss}\ [POC] - (k_{flocc} + k_{DOC->mineral})\ [DOC] \tag{2.3}$$

$$\frac{dPOC_{photo}}{dt} = k_{flocc}\ [DOC] - (k_{diss} + k_{POC->mineral})\ [POC] \tag{2.4}$$

Here, we consider that DOC it is transformed to POC at the photoflocculation rate of k_{flocc} and to other minerals at a rate of $k_{DOC->mineral}$ whereas it increases due to photodissolution at a rate of k_{diss}. The rate constants could be found through regression analysis and fitting on experimental data. Then DOC would be distributed between its biodegradable and refractory fractions based on b_1^p ratio, which would be found through a literature review. A similar concept exists for POC, which would be distributed between BPOC and RPOC based on b_2^p ratio.

4.4 The role of the qualitative approach in OM characterization: towards quantitative estimates

The quantitative approach discussed so far only measures the OM variables. It misses the OM quality that provides the tracing of the origin, the chemical composition, and humic content of OM. OM qualitative data tools such as specific ultraviolet absorbance at 254 nm ($SUVA_{254}$) and fluorescence spectroscopy are quite popular among the scientific community and river managers. If linked properly with OM quantitative data, they can be essential operational tools by providing the basis for a high-frequency OM data collection system.

These high-frequency data will lead to a significant reduction in the model uncertainties that are due to a lack of sufficient information on OM in sensitive periods such as low flows or in the aftermath of releases from combined sewage overflows. In addition, their major advantages are an easy set-up, being noninvasive, cheap, and high-sensitivity edge [Zhang et al., 2021] with the possibility of being deployed in real time [Kim et al., 2016].

$SUVA_{254}$ and fluorescence spectroscopy have been used for measurement of DOC in rivers and wastewater treatment plants (WWTPs) [Parlanti et al., 2000, 2002, Goffin et al., 2017]. Fluorescence spectrometry has been used (a) to provide information on the source of OM, whether it is allochthonous or autochthonous [Stedmon et al., 2003], (b) to classify the origins of DOC under low and high flow [Varrault et al., 2020], and (c) to determine the source of DOC in a wetland as being microbial [Palma et al., 2020]. Goffin et al. [2018] used 3D fluorescence spectroscopy to prepare a predictive model for BOD_5 in the dissolved fraction of WWTP raw water, while Hur and Cho [2012] correlated fluorescence spectroscopy indices with BOD and chemical oxygen demand (COD) in urban polluted rivers.

Thanks to an extensive deployment of *in situ* fluorescence monitoring, a common protocol is making future experiments comparable and credible proxies for DOC and BOD measurement as indicated in the review by Carstea et al. [2020]. These tools may be essential for measuring DOC concentrations, but in reality they only allow for a more or less correct estimation of DOC concentrations and are very

dependent on the type of water and on the characteristics of the DOC. When the DOC constituents absorbing at the wavelength used to measure the absorbance of the sample represent the majority of the DOC, the correlation can be good. In natural waters, however, this is rarely the case. And in any case, a calibration must be made between DOC and absorbance for each type of sample.

From the perspective of river water quality modeling, b_1 and specifically BDOC are important to be measured by qualitative and quantitative proxy methods because they may present the most influential fraction of OM for the biogeochemical functioning of water ecosystems [Hasanyar et al., 2020, 2023b]. This is not yet possible by spectroscopic techniques; nevertheless, Shin et al. [2016] reported a correlation between $SUVA_{254}$ and BDOC, and Hosen et al. [2014] have shown that the fluorescence method is capable of distinguishing between BDOC and RDOC. Moreover, Knapik et al. [2015] has found correlations between spectroscopic fluorescence peak intensities (labile and refractory peaks) and DOC. Thus, considering the ongoing research, it can be said that these methods are about to provide high-frequency BDOC and/or b_1 in real time in the very near future. Indeed, considering the capabilities of spectroscopic techniques, a complete new descriptor for the biodegradability of OM could be introduced based on the indices of these innovative tools.

5 Conclusions

The main objective of this study was to provide a quantitative review of variables representing the description and degradation of organic matter in world rivers. The literature was studied to extract quantitative information about the organic matter variables under different hydrological, land cover, and climatic conditions. Then the analytical method of factorial analysis of mixed data (FAMD) was implemented to identify clusters in the collected dataset based on different categories of hydrology, climate, and land cover. Finally, variation ranges and average values were proposed for the identified clusters under FAMD. The following conclusions are made from this study:

- Knowledge of the variation ranges and average values of OM variables provided in this study is essential in research related to river metabolism specifically for carbon transport from rivers to oceans, riverine carbon impact on global warming, and ecosystem functioning in addition to river water quality modeling and the design of drinking water and water treatment plants.
- The concentration of OM and its constituent fractions in a river depend on the river land cover and climate. The biodegradability of DOC depends on river hydrology in addition to climate and land cover.

- The FAMD method was able to identify clusters that stand out in the dataset. It also identified the climatic, land cover, and hydrological conditions that contribute to the variability of the OM variables.
- Urbanization and its effluents from water treatment plants and combined sewage overflows increase DOC biodegradability in rivers leading to increased bacterial activity. Such conditions result in the deterioration of water quality. Therefore, it is suggested to optimally design the location and distribution of urban effluents in the rivers in order to maintain a required level of water quality.
- Wetlands are carbon-abundant and provide more OM to a river than other land cover types. The cluster of dry climate and barren lands also contributes a high amount of OM to rivers.
- The average values and variation ranges of the OM variables provided in this study could be used by engineers interested in water management through the coupling of monitoring networks and river models. River models have uncertainties related to carbon inflow from tributary rivers and its diffuse and point sources. The variation ranges provided in this review can help reduce those uncertainties. Data assimilation, as one of the methods incorporated in river models to study the river metabolism, indeed requires information on average values and variation ranges of OM variables. In addition, sensitivity analysis of river models studying the role of OM also requires information on the variation ranges of parameters [Wang et al., 2018, Hasanyar et al., 2023b]. This review can help narrow the variation range of parameters representing OM inflow.
- Uncertainty sources exist in the experimental protocols for OM variables. There is a need for a universal method. Moreover, some uncertainties such as photodegradation could be incorporated in the water quality models.
- The link between qualitative and quantitative OM data can lead to an operational tool for a high-frequency OM data collection system.

Chapter 3

How much do bacterial growth properties and biodegradable dissolved organic matter control water quality at low flow?

Manuscript status: published at the EGU journal of Biogeosciences, https://doi.org/10.5194/bg-20-1621-2023

Résumé

Le développement d'outils de modélisation précis de la qualité de l'eau est nécessaire pour une gestion intégrée de la qualité de l'eau des systèmes fluviaux. Bien que certains modèles de qualité de l'eau puissent simuler avec précision les concentrations de l'oxygène dissous (OD) pendant les périodes de crue et les blooms de phytoplancton dans les rivières, des écarts persistent pendant les périodes de faible débit. Nous utilisons le modèle biogéochimique C-RIVE pour évaluer l'influence des paramètres du modèle sur les simulations d'OD à faible débit. Les paramètres étudiés sont liés à la communauté bactérienne (par exemple, taux de croissance bactérienne), à la matière organique (son partitionnement et sa dégradation) et à des facteurs physiques (par exemple, la réoxygénation de la rivière due à la navigation et au vent).

Les taux de croissance et de mortalité bactérienne sont de loin les deux paramètres les plus influents, suivis du rendement de croissance bactérienne. Les résultats de l'analyse de sensibilité affinée indiquent que la fraction biodégradable de la matière organique dissoute (CODB) et le rendement de croissance bactérienne sont les paramètres les plus influents dans des conditions de taux de croissance bactérienne

net élevé (= taux de croissance - taux de mortalité), tandis que le rendement de croissance bactérienne est indépendamment dominant dans des situations de faible croissance nette. Sur la base des résultats de cette étude, des propositions sont formulées pour la mesure in situ de la CODB dans un réseau de surveillance de la qualité de l'eau en milieu urbain fournissant des données à haute fréquence. Les résultats indiquent également la nécessité de surveiller la communauté bactérienne afin de détecter d'éventuels changements de communauté bactérienne après des événements transitoires tels que les déversements d'orage.

En outre, nous discutons de l'inclusion du paramètre de biodégradabilité du carbone dans les logiciels de modélisation statistique de la qualité de l'eau pour améliorer l'estimation de l'apport de matière organique à partir des conditions limites. Les résultats de cette étude soulignent l'importance de la communauté bactérienne et du CODB dans les modèles de qualité de l'eau pour une meilleure simulation de l'OD pendant les périodes de faible débit. Ils ont également des implications pratiques pour la surveillance et la gestion de la qualité de l'eau, notamment la nécessité de mesurer le CODB dans les réseaux de surveillance de la qualité de l'eau en milieu urbain, de surveiller la communauté bactérienne et d'inclure le CODB dans les logiciels de modélisation de la qualité de l'eau.

Abstract

The development of accurate water quality modeling tools is necessary for integrated water quality management of river systems. Even though some water quality models can simulate dissolved oxygen (DO) concentrations accurately during high flow periods and phytoplankton blooms in rivers, significant discrepancies remain during low flow periods, when the dilution capacity of the rivers is reduced. We use the C-RIVE biogeochemical model to evaluate the influence of controlling parameters on DO simulations at low flow. Based on a coarse model pre-analysis, three sensitivity analyses (SA) are carried out using the Sobol method. The parameters studied are related to bacterial community (e.g., bacterial growth rate), organic matter (OM; partitioning and degradation of OM into constituent fractions), and physical factors (e.g., reoxygenation of the river due to navigation and wind). Bacterial growth and mortality rates are found to be by far the two most influential parameters, followed by bacterial growth yield. More refined SA results indicate that the biodegradable fraction of dissolved organic matter (BDOM) and the bacterial growth yield are the most influential parameters under conditions of a high net bacterial growth rate (= growth rate – mortality rate), while bacterial growth yield is independently dominant in low net growth situations. Based on the results of this study, proposals are made for in situ measurement of BDOM under an urban area water quality monitoring

network that provides high-frequency data. The results also indicate the need for bacterial community monitoring in order to detect potential bacterial community shifts after transient events such as combined sewer overflows and modifications in internal processes of treatment plants. Furthermore, we discuss the inclusion of BDOM in statistical water quality modeling software for improvement in the estimation of organic matter inflow from boundary conditions.

1 Introduction

Dissolved oxygen (DO) has been considered the most important indicator of water quality in surface water resources [Streeter and Phelps, 1925, Odum, 1956, Escoffier et al., 2018], because it integrates the biological functioning of a system as well as the impact of anthropogenic forcing. It is the main variable used to evaluate river metabolism [Odum, 1956, Staehr et al., 2010, Demars et al., 2015] by comparing the gross primary production (GPP) with ecosystem respiration (ER) and defining whether an ecosystem is autotrophic or heterotrophic based on the net ecosystem production (NEP = GPP-ER) being positive or negative, respectively [Garnier et al., 2020]. Maintaining a sufficient level of DO is necessary for the overall health of rivers, not only because of the life dependency of water species [Garvey et al., 2007], but also for preventing smell and taste degradation [Bailey and Ahmadi, 2014].

The situation of rivers during low flow is of particular interest since studies have demonstrated that the river water quality during such flow periods is more vulnerable to degradation due to lower dilution rates. This is particularly the case if the river receives organic matter load from waste water treatment plants (WWTP) and combined sewer overflows (CSO). These OM loads lead to heterotrophic conditions in the river, where very low DO levels and high fish mortality can be observed [Seidl et al., 1998a, Even et al., 2004, Vilmin et al., 2016, Garnier et al., 2020]. Therefore, river water quality modeling has been one of the main research interests of water quality managers and researchers ever since the use of the very first water quality model [Streeter and Phelps, 1925] to more complex ones [Billen et al., 1994, Garnier et al., 1995, Even et al., 1998, Vanrolleghem et al., 2001, Flipo, 2005, Wang et al., 2013]. Its aim is to identify the main determinants of DO evolution and to forecast the response of aquatic systems to human-induced pressure, in particular due to waste water treatment plants (WWTPs) outflows.

Large discrepancies exist between DO simulations and observations during low-flow periods in water quality models. These mismatches were found in the QUESTOR model applied on the Thames (UK) [Hutchins et al., 2020], in the Riverstrahler model applied on the Mosel river (Germany) [Garnier et al., 1999], the Scheldt river (Belgium) [Thieu et al., 2009], and the Seine river (France) [Garnier

et al., 2020]. Yang et al. [2010] found the same results in the WASP model and noted that the uncertainty of the model lies in characterization of OM degradation and nitrification rates. Bailey and Ahmadi [2014] found similar results in the QUAL2E-OTIS water quality model. The ProSe model also has mismatches at low flow [Even et al., 2004, 2007, Vilmin et al., 2018, Garnier et al., 2020, Wang et al., 2022a]. Among the parameters that control DO concentration in water [Cox, 2003], Wang et al. [2022a] assume that the uncertainties related to the parameterization of OM degradation kinetics and OM biodegradability at system's boundaries (tributary rivers, WWTPs, and CSOs) play a major role in the discrepancies observed during low flow periods.

In order to objectively evaluate the controlling parameters of DO during such periods, a sensitivity analysis (SA) is conducted. Several applications of SA methods can be found for water quality modeling [Nossent et al., 2011, Bailey and Ahmadi, 2014, Cho et al., 2017, Wang et al., 2018]. Moreover, SA applications in hydrological and water quality modeling are summarized by [Reusser et al., 2011, Wang et al., 2018]. In this study, the Sobol method [Sobol, 1993] is chosen in order to understand the inter-parameter interactions.

For the first time, the influence of bacterial properties and that of the quantity and different fractions of OM sources are investigated on DO evolution at low flow using C-RIVE model [Vilmin et al., 2012, Wang et al., 2018]. It is conducted for better understanding of the short-term (5 days) and mid-term (45 days) effects of the rapidly and slowly biodegradable OM, respectively. To further understand the functioning of the Sobol' SA, the inter-parameter interactions are calculated to address how one parameter hides the influence of other parameters. On the basis of the SA results, suggestions are made for water quality monitoring in urban areas. Finally, proposals are made for a better integration of the influential parameters in data assimilation.

Based on the above discussion, we address three research questions:

1. What are the influential parameters controlling DO during a post-bloom summer low flow period where discrepancies are observed in different water quality models? Is a model that includes bacteria physiological parameters (growth and yield rates) alone sufficient to describe DO variation?

2. To what extent is the knowledge of the quantity of OM share, especially that of the biodegradable fraction of dissolved organic matter (BDOM), influential for water quality modeling?

3. What is the hierarchy (importance ranking) among the influential parameters ?

2 Material and methods

Here we represent the C-RIVE model (section 2.1) as the forward model of the study and identify the parameters that need to be included in the study. Then, two new sets of parameters are added to the study to account for the uncertainties related to the parameterization of OM degradation kinetics and its partitioning into different constituent fractions. This is followed by the determination of the variation ranges of the introduced parameters (section 2.1.2). Then, the strategy for conducting different SAs is detailed in section 2.3 to determine the influential parameters.

2.1 C-RIVE Biogeochemical model

C-RIVE is a C ANSI library that implements RIVE concepts [Billen et al., 1994, Garnier et al., 1995]. It simulates the cycles of carbon, oxygen, and other nutrients both in the water column and sediments of river systems (Fig. 3.1). The model is community centered and explicitly describes micro-organisms' communities, such as phytoplankton and heterotrophic bacteria. The physiological parameters of those communities were determined through multiple lab experiments. Both the RIVE model and its parameters were coupled in two river water quality models : RIVERSTRAHLER [Billen et al., 1994] and ProSe [Even et al., 1998]. These two models are calibrated and validated on real case applications in different river basins over the world such as in the Danube river (Romania and Bulgaria) [Garnier et al., 2002], in the Day-Nhue river (Vietnam) [Luu et al., 2021], in the Grand Morin river (France) [Flipo et al., 2004, 2007], in the Lule and Kalix rivers (Sweden) [Sferratore et al., 2008], in the Mosel river (Germany) [Garnier et al., 1999], in the Red river system (Vietnam and China) [Quynh et al., 2014], in the Scheldt river (Belgium and Netherlands) [Billen et al., 2005, Thieu et al., 2009], in the Seine river (France) [Even et al., 2004, 2007, Raimonet et al., 2015, Vilmin et al., 2015, 2016, Garnier et al., 2020], in the Somme river (France) [Thieu et al., 2009, 2010], and in the Zenne river (Belgium) [Garnier et al., 2013].

In the C-RIVE model, DO in the water column depends on physical, bacterial, and phytoplanktonic processes (Fig. 3.1). The physical and phytoplanktonic processes tend to provide DO while bacterial processes consume DO. The bacterial respiration, that is the main source of oxygen consumption, depends on the heterotrophic bacterial kinetics and the availability of substrate matter. These equations are accessible in previous publications [Billen et al., 1988, Servais et al., 1989, Billen, 1991, Wang et al., 2018]. For the readability of the paper, they are developed in the appendix sections B1.1 and B1.2.

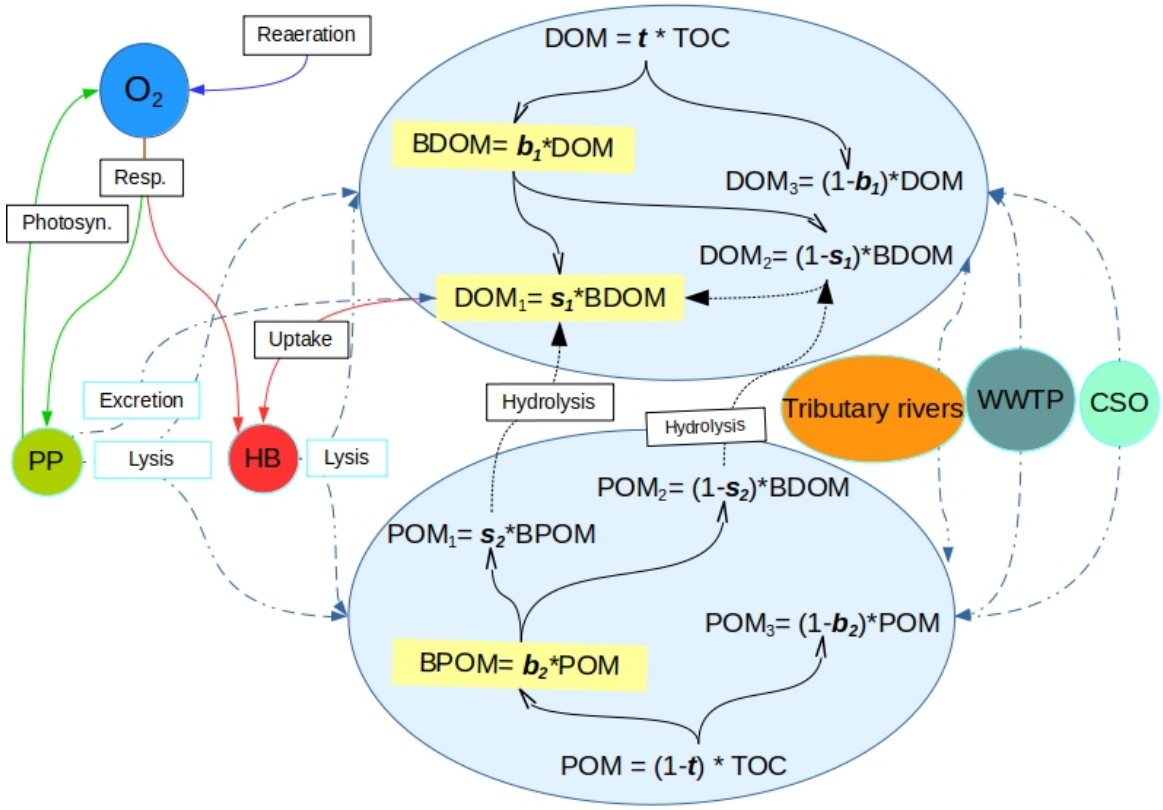


Figure 3.1: Schematic description of the OM-related processes accounted for by C-RIVE in the water column where OM is partitioned into six fractions of dissolved and particulate matter using the five partitioning parameters, namely, t, b_1, s_1, b_2, s_2; POM: particulate organic matter; DOM: dissolved organic matter; BDOM:biodegradable DOM, BPOM:biodegradable POM (subscripts 1, 2, and 3 refer to rapidly degradable, slowly degradable, and non-biodegradable fractions of OM, respectively); Blue dashed-double dotted lines: OM input from sources and partitioning between POM and DOM; Solid black lines: partitioning of DOM and POM into biodegradability pools ; Dotted black lines: Hydrolysis; Remaining solid lines: Biogeochemical processes. Resp.:Respiration; Photo,:Photosynthesis; PP: primary producers; HB: heterotrophic bacteria; WWTP: waste water treatment plant; CSO: combined sewer overflow

2.1.1 OM partitioning parameters: from total organic carbon to six OM fractions

The OM partitioning parameters are a novelty added in C-RIVE. Indeed, former version of C-RIVE did not include any parameter to define the partitioning of OM into DOM and POM and then further into their corresponding fractions $DOM_{1,2,3}$ and $POM_{1,2,3}$ (see Fig. 3.1 for the definitions). In the recent development of C-RIVE, the total organic carbon (TOC) is initially divided into dissolved (DOM) and particulate (POM) forms by t parameter (Fig. 3.1). Then, DOM is divided into biodegradable (BDOM) and refractory (DOM_3) fractions by b_1 parameter. Then, thanks to s_1 parameter, BDOM is further divided into (i) DOM_1 as the limiting substrate (rapidly biodegradable DOM in 5 days) and (ii) DOM_2 (slowly biodegradable DOM in 45 days). POM is similarly divided into its constituent fractions using b_2 and s_2 parameters. The equations concerning these five parameters and that of OM

degradation are accessible in appendix section B2. To further clarify, the only required forced information is TOC that comes from experimental data. Therefore, the five OM partitioning parameters (t, b_1, b_2, s_1, s_2) give us the possibility to conduct a sensitivity analysis to quantify their influence on DO concentration. Using this OM partitioning model and depending on these five parameters, we are able to convert time varying TOC of boundary conditions, per say river inflows, into time varying $DOM_{1,2,3}$ and $POM_{1,2,3}$ fractions.

2.1.2 Parameters for SA and their variation ranges

The influence of 17 parameters on oxygen concentrations are evaluated in this study (Tab. 3.1). Various types of parameters are identified: two physical parameters that account for O_2 re-aeration; seven bacterial parameters that account for bacteria growth, mortality, and respiration; three OM degradation parameters; and five OM partitioning parameters (Fig. 3.1).

Before proceeding to a SA, it is necessary to specify the range of variation of each parameter according to a *a prior* distribution based on former knowledge. Those distributions are assumed to be uniform within a range, whose definition relies on a literature review. The range of variation of the partitioning and degradation parameters (Tab. 3.1) is selected based on a detailed bibliographical review [Hasanyar et al., 2020, 2023a, Wang et al., 2018]. Table 3.1 also includes the range of variation of TOC, which represents the total organic matter input in the model due to the boundary conditions and varies from 1 to 10 mgC L^{-1} under low flow [Hasanyar et al., 2020].

2.2 Case study

The synthetic case developed by Wang et al. [2018] is adapted for the application of SAs on C-RIVE parameters during a low-flow period (Fig. 3.2). It is a river stretch with a width of 100 m and a length of 1000 m representing the Seine river. The low-flow period is characterized with a discharge of 80 m^3 s^{-1}. The simulation period is 45 day-long in order to be coherent with the experimental protocol of the BDOM measurement [Servais et al., 1995], that considers 45 days as a limit between refractory and slowly biodegradable organic matter. Moreover, a 45-day simulation period is also necessary for studying the long-term effect of TOC degradation on river metabolism.

Considering the discharge and the wet section, the numerical experiment can be viewed as a lagrangian one, where we follow a river body along a river network of the above mentioned dimension with a speed of 0.14 m s^{-1}.

Table 3.1: List of the parameters accounted for in the sensitivity analyses and their corresponding ranges of variation

Parameter	Description	Min. Val.	Max. Val.	Unit	References
TOC	Total organic carbon	1	10	[mgC L^{-1}]	
OM partitioning parameters					
t	ratio between dissolved and total organic matter (DOM/TOC)	0.4	0.9	[-]	Hasanyar et al. [2020, 2023a]
b_1	ratio between biodegradable DOM and DOM (BDOM/DOM)	0.1	0.5	[-]	
b_2	ratio between biodegradable POM and POM (BPOM/POM)	0.1	0.5	[-]	
s_1	ratio between rapidly biodegradable DOM and BDOM (DOM_1/BDOM)	0.4	0.95	[-]	
s_2	ratio between high biodegradable POM and BPOM (POM_1/BPOM)	0.4	0.95	[-]	
OM degradation parameters					
K_s	constant of semi saturation for bacterial substrate uptake	0.02	0.15	[mgC L^{-1}]	
K_{DOM2}	constant of semi saturation for the hydrolysis of DOM_2	0.2	1.5	[mgC L^{-1}]	
$k_{hyd,max}$	coefficient of the hydrolysis of DOM_2 to DOM_1	0.25	0.75	[h^{-1}]	
Bacterial parameters					
$T_{opt,hb}$	optimum temperature for bacterial growth	15	30	[°C]	Wang et al. [2018]
σ_{hb}	standard deviation of temperature function for bacterial growth	12.75	21.25	[°C]	
$V_{sed,hb}$	settling velocity of bacteria	0	0.1	[m h^{-1}]	
$K_{O_2,hb}$	Half-saturation constant for dissolved oxygen	0.375	0.625	[mgO_2 L^{-1}]	
$\mu_{max,hb}$	maximum growth rate of bacteria	0.01	0.07 *	[h^{-1}]	
Y_{hb}	bacterial growth yield	0.03	0.5	[-]	
$mort_{hb}$	bacterial mortality rate	0.01	0.08	[h^{-1}]	
Physical parameters					
K_{navig}	re-aeration coefficient due to navigation	0	0.05	[m h^{-1}]	
K_{wind}	re-aeration coefficient due to wind	0.885	1.475	[m h^{-1}]	

* The upper limit identified by Wang et al. [2018] is decreased from 0.13/h to 0.07/h in order to avoid complete DO depletion in simulations longer than 5 days

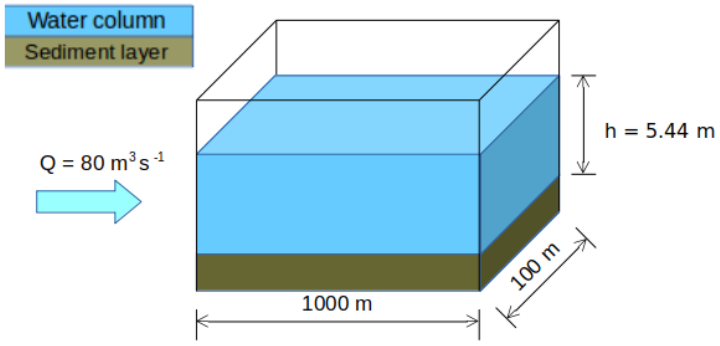


Figure 3.2: Synthetic scheme representing a reach of the Seine river (modified from Wang et al. [2018]).

Initial conditions

The initial concentrations for both water and sediment compartments (Tab. 3.2) are set based on the mean concentrations of the simulations at Bougival station during the 2007-2012 period [Vilmin et al., 2016] except for water temperature (depending on summer season), DO (depending on oxygen solubility), POM and DOM fractions (depending on the TOC concentration and partitioning parameters), and phytoplankton and bacterial biomass (depending on a post-bloom condition). As far as organic matter is concerned, TOC is first defined and then distributed into its *DOM* and *POM* fractions depending on the values of the five OM partitioning parameters (t, b_1, s_1, b_2, s_2). The hyporheic exchanges (between groundwater and river) are ignored in this work because the contribution of groundwater to a downstream river (Strahler order > 6), such as the Seine river at the crossing of the Paris urban area, is negligible with respect to the discharge of the river itself. For the Seine river the groundwater contribution along a 100 km is around 1 m^3 s^{-1} with respect to the Seine discharge amounting for 80 m^3 s^{-1} during severe low flow conditions [Pryet et al., 2015].

2.3 Sensitivity analysis strategy

Before defining in detail the SA, a coarse pre-analysis consisting in forward simulations of the C-RIVE model is conducted with extreme parameter values. Then, various SA are developed to assess the assumptions and conclusions put in place in the pre-analysis.

2.3.1 Pre-analysis of the model with extreme parameter values

First, we need to select certain parameters for the pre-analysis. We consider $\mu_{max,hb}$, $mort_{hb}$ and Y_{hb} as they were found to be influential in the study of Wang et al. [2018] under non-bloom situations.

Table 3.2: Initial concentrations of the simulations

No	Species	$C_{ini,water}$	$C_{ini,sediment}$	Unit
1	NH_4	0.12	0.33	[mgN L^{-1}]
2	NO_2	0.04	0.04	[mgN L^{-1}]
3	NO_3	7	4.54	[mgN L^{-1}]
4	TSS	16.82	95010	[mg L^{-1}]
5	PO_4	0.1	0.27	[mgP L^{-1}]
6	O_2	8.62	6.65	[mgO_2 L^{-1}]
7	HB	0.023	0.016	[mgC L^{-1}]
8	PP	0.010	0.003	[mgC L^{-1}]
9	DOM_1	f(TOC, partitioning parameters)	0.12	[mgC L^{-1}]
10	DOM_2	f(TOC, partitioning parameters)	1.28	[mgC L^{-1}]
11	DOM_3	f(TOC, partitioning parameters)	1.94	[mgC L^{-1}]
12	POM_1	f(TOC, partitioning parameters)	44	[mgC L^{-1}]
13	POM_2	f(TOC, partitioning parameters)	696	[mgC L^{-1}]
14	POM_3	f(TOC, partitioning parameters)	2555	[mgC L^{-1}]
15	T_{mean}	22.4 ± 3.0		°C

However, to decrease the number of parameters, $mort_{hb}$ and $\mu_{max,hb}$ are represented together as a single parameter called "net growth (NG)."

$$Net\ Growth\ (NG) = \mu_{max,hb} - mort_{hb}$$

Fixing $mort_{hb}$ = 0.02 h^{-1} at its reference value and $\mu_{max,hb}$ ranging between 0.022 h^{-1} and 0.07 h^{-1}, the net growth ranges from 0.002 h^{-1} to 0.05 h^{-1} while the range for Y_{hb} is taken from Table 3.1. As the OM partitioning parameters are not C-RIVE inputs, we consider BDOM to represent them in the model. Its range is given by Eq. (3.1)-(3.2) as follows:

$$BDOM_{min} = TOC_{ref} * t_{ref} * b_{1,min} \tag{3.1}$$

$$BDOM_{max} = TOC_{ref} * t_{ref} * b_{1,max} \tag{3.2}$$

Here, TOC_{ref} is a reference TOC value and fixed at 5 mgC L^{-1} (considered as the baseline concentration of TOC in the Seine river [Vilmin et al., 2016]), the reference t (t_{ref}) = 0.7 is the average value of t variation range and b_1 is taken from Table 3.1. This way BDOM varies following b_1 only and therefore remains statistically independent from the other parameters.

Eight simulations pertaining to eight different combinations of the minimum and maximum values of the net growth, its associated yield and BDOM are launched (Table 3.3) and accordingly for each combination, the evolution of DO, DOM_1, DOM_2, and BDOM is plotted (Fig. 3.3).

Table 3.3: Definition of the eight single simulations achieved with extreme values of biodegradable dissolved organic matter, net growth of bacteria community and its associated yield

Sim. No.	BDOM	Net growth	Y_{hb}
1	0.35	0.05	0.03
2	0.35	0.002	0.03
3	1.75	0.05	0.03
4	1.75	0.002	0.03
5	1.75	0.05	0.5
6	1.75	0.002	0.5
7	0.35	0.05	0.5
8	0.35	0.002	0.5

From the pre-analysis, we hypothesize that:

1. The net growth is one of the most influential parameters on DO because all simulations on the left side with a high net growth demonstrate significant depletion of DO than those on the right side having low net growth rates.

2. DO is sensitive to BDOM under high net growth conditions. This could be observed by comparing simulations 3 and 4 (under high net growth condition) with simulations 4 and 6 (under low net growth condition) (Fig. 3.3).

3. DO is not sensitive to BDOM under low net growth rates. Comparison of simulations 6 and 7 demonstrate that even a high BDOM coupled with low net growth (simulation 6) has less effect on DO than a low BDOM coupled with high net growth (simulation 7) .

2.3.2 Understanding river metabolism controls with multiple sensitivity analyses

Three sensitivity analyses (SAs) are derived to test the former three hypotheses. The detail of each SA parameterization is available in Table 3.4.

The first SA is conducted by assuming the general influence of net growth parameters in the pre-analysis, and in order to have a broader view of the model sensitivity with respect to all the model parameters. Based on the pre-analysis, we observed that the main effect due to BDOM is linked to high net growth rates, therefore, we can assume that the effect of parameters other than net growth parameters is demonstrated when they are coupled with a high net growth condition. In addition, since a significant interaction (the difference between the first and total sensitivity indices) is observed between net growth parameters in Wang et al. [2018], they are assumed to be hiding the influence of other parameters. We therefore implement a second SA where net growth parameters are fixed at their highest value. This SA removes the possibility of interactions among net growth and other model

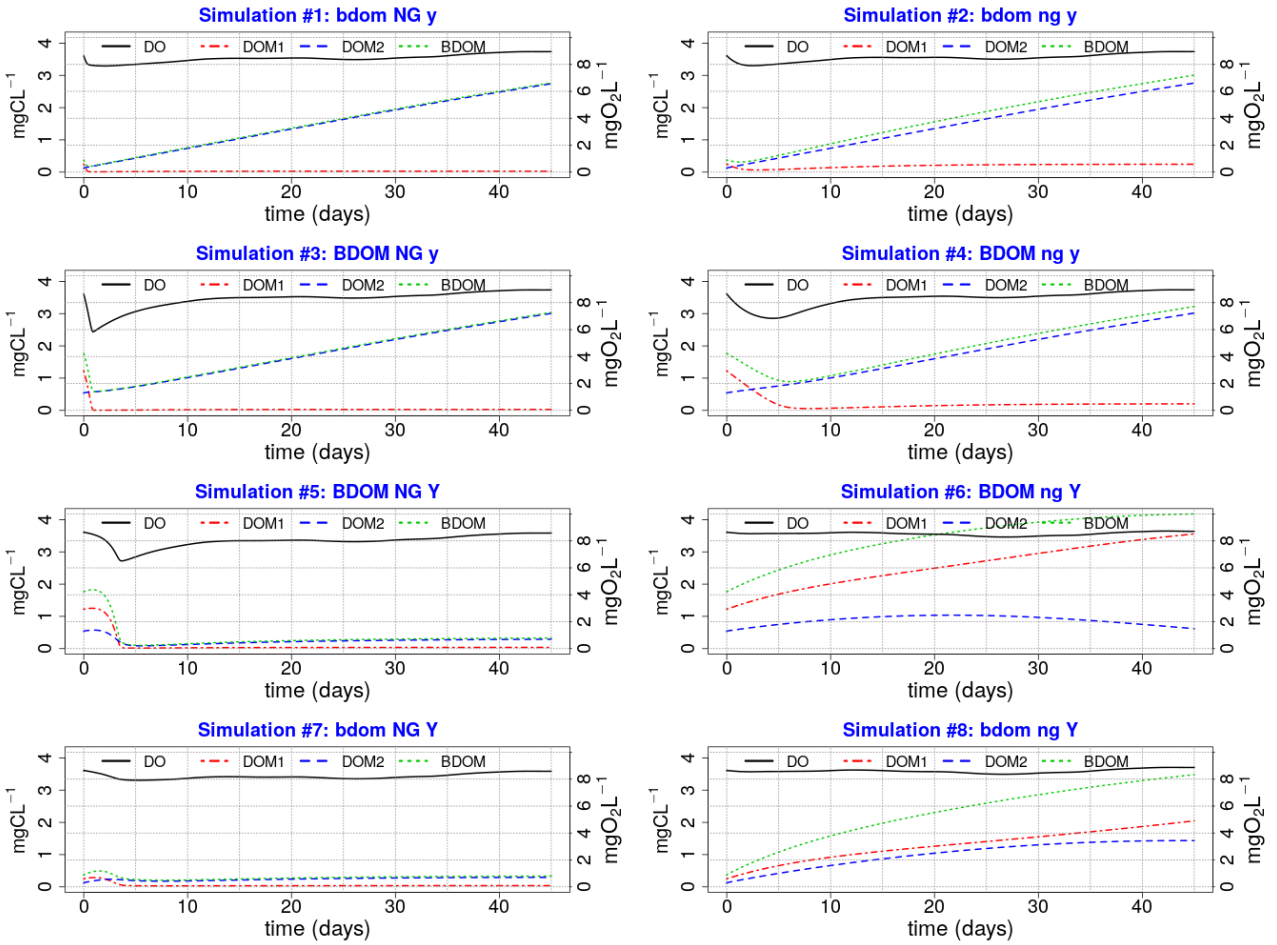


Figure 3.3: Eight plots of single simulations; (XX yy ZZ) Any parameter name written in capital or small letters means that its maximum or minimum value is used, respectively, in that specific single simulation. For example, plot #3 (BDOM NG y) that corresponds to simulation #3 in Table 3.3 is a simulation where the maximum values of BDOM and net growth and the minimum value of Y_{hb} are used

parameters. It results in a better evaluation of the model sensitivity with respect to the parameters whose influences might be hidden by the dominant and interacting parameters.

The third SA is performed to verify the second SA assumption that parameters other than net growth exert their influence only under a high net growth condition, and thus the same parameters could be deemed non-influential under a low net growth situation. The net growth parameter is fixed at its lowest value.

To summarize, the settings for the three SAs are as follows:

1. **First SA (All parameters included)**: There are 17 defined parameters (Table 3.1 & Table 3.4) in the model, the simulation period is **45 days**. It is repeated for each TOC concentration from 1 to 10 mgC L^{-1}, with a 1 mgC L^{-1} increase step.

2. **Second SA (Fixed high net growth)**: The influence of twelve parameters is evaluated

(Table 3.4). The bacteria net growth rate is fixed to its maximum value using the highest value of the bacteria growth rate ($\mu_{max,hb}$= 0.07 h^{-1}) and a mortality rate of 0.02 h^{-1}. Furthermore, to decrease the computational cost of the model, the three OM partitioning parameters (t, b_1 and b_2) from the first SA are narrowed to BDOM and BPOM whose variation ranges are calculated based on the following Eq. (3.3)-(3.6):

$$BDOM_{min} = TOC * t_{ref} * b_{1,min} \tag{3.3}$$

$$BDOM_{max} = TOC * t_{ref} * b_{1,max} \tag{3.4}$$

$$BPOM_{min} = TOC * (1 - t_{ref}) * b_{2,min} \tag{3.5}$$

$$BPOM_{max} = TOC * (1 - t_{ref}) * b_{2,max} \tag{3.6}$$

Where, t_{ref} is set to 0.7. TOC varies between 1 and 10 mgC L^{-1}, with a 1 mgC L^{-1} increase step. The time length is set to 5 days accordingly to the pre-analysis, which demonstrated that BDOM or precisely the substrate (DOM_1) was consumed in less than 5 days under the high net growth condition (simulations 1, 3, 5 & 7).

3. **Third SA (Fixed low net growth)**: The influence of twelve parameters is evaluated (Table 3.4). This SA is conducted in a similar way than the second SA except that this time $\mu_{max,hb}$ is fixed at a lower value of 0.022 h^{-1} in order to simulate a very low net growth rate condition of approximately 0.002 h^{-1}. This SA is also implemented for a **5-day** period of time and repeated 10 times to simulate TOC ranging from 1 to 10 mgC L^{-1}.

Each of the three aforementioned SAs is implemented based on an innovative SA methodology initially proposed in Wang et al. [2018] and adopted in this study, where the influence of input parameters (X) is evaluated on the C-RIVE model according to the variations of a large set of DO simulations (model output, Y). The necessary steps pursued for SA are summarized in appendix section B2.2.

Each SA is performed with a Python script that computes on an Intel(R) Xeon(R) E5-2640 (20 cores @ 2.4 GHz). The computational time is 12h per TOC value for the first SA, while it is reduced to only 3h per TOC value for the second and third ones.

Table 3.4: The parameters considered in each of the four sensitivity analyses

	1^{st} SA	2^{nd} SA	3^{rd} SA
OM partitioning parameters	t b_1 s_1 b_2 s_2	BDOM BPOM	BDOM BPOM
OM degradation parameters	K_s K_{DOM2} $k_{hyd,max}$	K_s K_{DOM2} $k_{hyd,max}$	K_s K_{DOM2} $k_{hyd,max}$
Bacterial parameters	$T_{opt,hb}$ σ_{hb} $V_{sed,hb}$ $K_{O_2,hb}$ Y_{hb} $\mu_{max,hb}$ $mort_{hb}$	$T_{opt,hb}$ σ_{hb} $V_{sed,hb}$ $K_{O_2,hb}$ Y_{hb}	$T_{opt,hb}$ σ_{hb} $V_{sed,hb}$ $K_{O_2,hb}$ Y_{hb}
Physical parameters	K_{navig} K_{wind}	K_{navig} K_{wind}	K_{navig} K_{wind}
total number of parameters	17	12	12

3 Results

This section presents the results of the three Sobol SAs during a summer low-flow period. The influential parameters of each analysis are discussed in the following paragraphs.

3.1 First SA: All parameters

Figure 3.4a presents the results of the Sobol SA method for TOC = 5 $mgCL^{-1}$. It is expressed by a bar plot of the total sensitivity (S_T), first-order sensitivity (S_1), and second-order sensitivity (S_2) indices of the parameters. The higher-order sensitivity indices are also calculated in terms of the difference between the total and the first- and second-order indices (S_T-S_1-S_2). The parameters are ranked based on their S_T and the most influential parameters are shown by the shaded area, which includes parameters constituting 95% of the total variance of the model output. The first SA is conducted for TOC values ranging from 1 to 10 $mgCL^{-1}$, which corresponds to ten runs. For each run, the evolution of the normalized total sensitivity indices (S_T^*) of the six most influential parameters is plotted (Fig. 3.4b).

DO is controlled by the bacterial mortality rate ($mort_{hb}$), maximum bacterial growth rate ($\mu_{max,hb}$), and bacterial yield (Y_{hb}), whatever the TOC concentration (Fig. 3.4b). By increasing TOC, we ob-

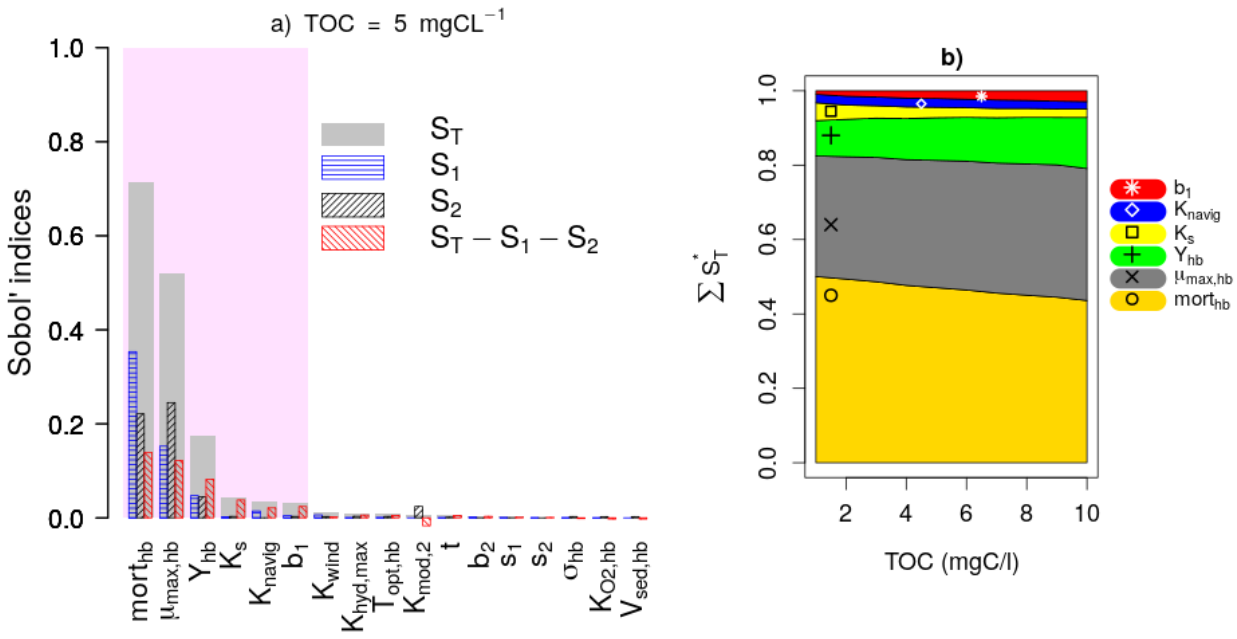


Figure 3.4: Sobol SA results of **first SA: All parameters** (a) Sobol SA results for TOC = 5 mgCL^{-1}; (b) Evolution of the normalized total sensitivity indices of the influential parameters with TOC

serve a gradual decrease in the influence of $mort_{hb}$, but an increase in the influence of Y_{hb}. This result obtained over the 45-day simulation period (Fig. 3.4) confirms the assumption made in the pre-analysis step (sec 2.3.1) regarding the overall dominance of bacterial parameters in long-term low-flow periods.

Apart from the constant of navigation (K_{navig}), which is a physical parameter, the other two influential parameters (K_s, b_1) are OM-related parameters that were introduced in this study. K_s seems to be more important in lower TOC concentrations compared to b_1 whose influence increases in higher TOC concentrations.

These results also confirm the assumptions made in the pre-analysis step that the dominant parameters tend to hide the influence of other parameters. Observing the second-order (S_2) and higher-order sensitivity (S_T-S_1-S_2) indices of $mort_{hb}$, $\mu_{max,hb}$ and Y_{hb} in Fig. 3.4a, very strong interactions can be highlighted between these parameters, i.e., a significant portion of their total sensitivity indices is due to their internal interactions. b_1 and K_s exert their influence only through higher-order interactions with these three parameters.

In the first SA, the three OM parameters (t, s_1 and s_2) are found to be non-influential, which means they can be excluded from SA by fixation in the second and third SA while calculating the variation ranges of BDOM and BPOM (Eq. (3.1)-(3.4)). Finally, the inclusion of b_1 among influential parameters out of the five OM parameters validates the selection of the BDOM concentration instead

of other OM components in the pre-analysis step.

3.2 Second SA under high bacteria net growth rate

The bacteria yield, Y_{hb}, and BDOM are the most influential parameters under high net growth rate conditions (Fig. 3.5a). This is due to the fact that the bacterial community manages to consume most of BDOM under a high net growth condition and then at some point, BDOM becomes a limiting factor for their growth. This result confirms the assumption made in the pre-analysis that the influence of parameters other than net growth parameters will be displayed if they are studied under a high net growth condition. The other important parameters are K_{navig} and K_s whose influence is reduced by increase in TOC. Moreover, very small interactions are observed between the parameters because almost all of their global influence stem from their main effects ($S_T \approx S_1$ for each parameter), which once again confirms the previous consideration that interactions are related to the effect of a varying net growth rate.

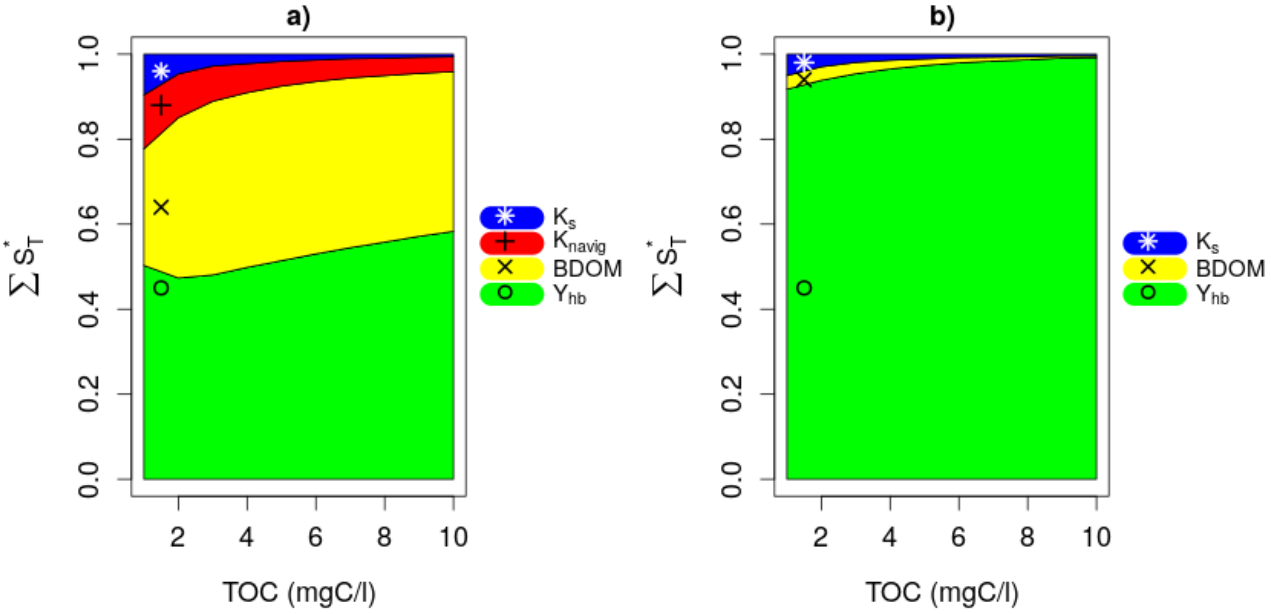


Figure 3.5: Results of (a) Second SA: Fixed high net growth; (b) Third SA: Fixed low net growth

3.3 Third SA under low bacteria net growth rate

The results of the third SA (Fig. 3.5b) reveal that Y_{hb} only is the influential parameter under a low net growth rate condition across whatever the TOC concentrations. This is due to the fact that, with such a low net growth rate, the bacterial community hardly grows at all and BDOM is subsequently not a limiting factor for such a small bacterial growth. This result confirms the assumption made in

the second SA by displaying all previous influential parameters except Y_{hb} as non-influential.

Finally, the second and third SA show that the influence of BDOM on DO increases with an increasing bacteria net growth rate.

4 Discussion

In this section, we propose first to synthesize our results, before analyzing the limitations in terms of physical processes accounted for. We then discuss what consequences do our results have on water quality monitoring in urban areas and on data assimilation.

4.1 Hierarchy of the most influential parameters on DO during low flow period

This study confirms that over a 45-day post-bloom summer low-flow period and whatever the TOC concentration, the bacterial net growth rate and Y_{hb} control the DO evolution. This is in agreement with the findings of the study by Wang et al. [2018], which was conducted over a 4-day simulation period. However, the mentioned bacteria physiological parameters are not sufficient to describe DO variation because the OM partitioning parameter, b_1, and the OM degradation parameter, K_s, are also influential at low flow under a high bacteria net growth rate.

Conducting the Sobol' SA for 5-days under a high bacterial net growth rate condition demonstrated the significant influence of BDOM and Y_{hb} at low flow. This is in connection with the findings of Hullar et al. [2006], Crump et al. [2003] that emphasized the importance of BDOM on bacteria population. This result is also in accordance with Bailey and Ahmadi [2014] who consider the model boundary condition as influential on DO but they do not specify which portion of the OM neither do they conduct their study at low flow. Nevertheless, this work is the first quantitative recognition of the role of BDOM on DO evolution that illustrates BDOM as the most influential fraction of the total organic matter entering a river system from its boundary conditions.

4.2 Limitations of the sensitivity analysis

This study is conducted under carbon source conditions where TOC varies between 1 and 10 mgC L^{-1}. It is also assumed that a carbon sink condition never happens. Indeed, carbon depletion would preclude the possibility to quantify the influence of the carbon on the DO concentration. Therefore the upper limit of $\mu_{max,hb}$ is reduced from 0.13 h^{-1} to 0.07 h^{-1} so as to avoid depletion of both carbon and oxygen during the simulation period.

Moreover, since C-RIVE currently lacks the impact of solar radiation on bacteria population, this process is not considered in this work. Nevertheless, it is recommended to be considered in future researches as solar radiation is found to damage the bacteria DNA impacting their growth and oxygen intake rates [Matallana-Surget and Wattiez, 2013]. This phenomenon could be included in the model in a simplistic approach either by increasing bacterial mortality or decreasing its growth rate by a given factor. However, it is necessary to find experimental data in order to quantify the effects of solar radiation that also depend on the type of bacterial community [Park et al., 2023].

4.3 Consequences of the results on water quality monitoring in urban areas

Rivers are highly sensitive to urban outflows at low flow [Seidl et al., 1998a, Huang et al., 2017] due to their low dilution capacity at this period of the year. Moreover, the construction or renewal of sanitary facilities, such as WWTP and CSO during transient events, induce changes in water quality due to changes in DOM, and BDOM concentrations [Servais and Garnier, 1993, Seidl et al., 1998b]. As a result, this induces potential shifts in bacterial communities, which have been found to be related to DOM source and its biodegradability [Hullar et al., 2006, Crump et al., 2003] such that an increase in BDOM is considered to increase the diversity of bacterial populations [Landa et al., 2013]. Even et al. [2004] therefore recommended a regular reassessment of the influential bacterial parameters. The sampling frequency in the monitoring stations should be set-up not only considering the temporal variability of the variable of interest, but also integrating possible successions of species.

Our study provides the list of important parameters for water quality modeling at low flow. For the bacterial parameters (mortality rate, maximum growth rate and growth yield), we propose the implementation of regular bacterial community sampling campaigns in the framework of the water quality monitoring program. Samples should thus be taken in river waters upstream from the major urban areas, and in the effluents of major WWTPs and CSOs.

Finally, considering the importance of BDOM at low flow periods, we propose the addition of BDOM measurement in the existing and new monitoring stations. Innovative methods such as specific ultraviolet absorbance at 254 nm ($SUVA_{254}$) and fluorescence spectroscopy [Parlanti et al., 2000, 2002, Goffin et al., 2017] could be proposed for high frequency measurement of OM and BDOM specifically. The monitoring stations shall be dense enough in space so that they could characterize the upstream tributary rivers and the outflow of anthropogenic sources such as WWTP and CSO. Such a spatial density and consideration of point and non-point pollution sources are necessary to cover all BDOM sources [Dixon et al., 1999, Polus et al., 2010, Do et al., 2012].

4.4 Consequences of the results on data assimilation (DA) in water quality modeling

Data assimilation is a method that combines observation data and physically based modeling in a statistical framework. It consists in sequentially updating the model parameters so that the output of the model will match the observation at each time step [Carrassi et al., 2018]. These techniques not only estimate the evolution of influential parameters thanks to observation data such as that of DO, but also provide enhanced simulation results of state variables.

The first assimilation tool that uses the particle filter technique (a statistical technique where numerous simulations are launched instead of one single simulation and are weighted based upon how well they reproduce the observed data) to couple with a water quality model is the ProSe-PA software [Wang et al., 2019, 2022a]. While studying a dry year using this software, mismatches were found between simulated and observed DO during low-flow periods [Wang et al., 2022a]. These mismatches were assumed to be due to insufficient biodegradable organic matter loading in the model caused by underestimated BDOM inputs to the Seine river. Therefore, based on our SA results, we can confirm that hypothesis and propose the incorporation of BDOM (the most influential OM-related parameter through the b_1 parameter) as a new component of ProSe-PA. It will facilitate not only the estimation of BDOM, but also improves the simulation results subsequently.

Another consequence of our results is that it would be appropriate to fix one or both of the $mort_{hb}$ and $\mu_{max,hb}$ parameters during data assimilation in order to facilitate parameter identification. In addition, the model of OM partitioning (Fig. 3.1) shall be explicitly included in the data assimilation software. The description of BDOM in each major organic matter source (tributary river, WWTP, and CSO) should also be independently represented in the DA scheme because each source brings its own specific contribution of organic matter and heterotrophic bacteria in urban rivers [Garnier et al., 1992a, Servais and Garnier, 1993, Seidl et al., 1998b].

5 Conclusions

The objective of this work was to investigate the role of organic matter loadings to river systems and the physiology of bacteria in river metabolism during a summer non-bloom low-flow period. New parameters were introduced to account for partitioning and degradation of OM. Then, the sensitivity of the C-RIVE model was analyzed against the newly introduced and the already existing model parameters. The following conclusions can be drawn from this study:

- The Sobol sensitivity analysis method proved very efficient in the identification of influential parameters on DO evolution in the C-RIVE model. Then, by fixing the interaction-inducing parameters, the influence of other parameters was assessed. This methodology may also be of interest for future sensitivity analysis where parameter interactions may hide the effect of other parameters;

- The river metabolism is dominated by bacterial activity at low flow during summer non-bloom periods. As a consequence, the net growth rate of bacteria, that combines their maximum growth rate ($\mu_{max,hb}$) and mortality rate ($mort_{hb}$), is the most important parameter under different TOC concentrations; therefore, it is essential to have a better estimation of the variation ranges of the growth and mortality rates of bacterial communities;

- Model response is very sensitive to the biodegradable share of DOM (BDOM) contributed by the boundary conditions. The effect of this parameter prevails at higher bacteria net growth rates occurring during summer low-flow periods when the organic matter brought by urban outflows is abundant in the river;

- Water quality monitoring networks shall continuously measure the influential parameters of this study in order to provide the water quality models with update values;

- More frequent sampling of autochtonous bacteria communities upstream and downstream of major urban areas and in major WWTP and CSO effluents will be of considerable interest to validate time varying parameter values estimated by data assimilation frameworks;

- The results of this study provide a list of influential and non-influential parameters. The latter can be fixed at their average or preferred value as per the literature, and the former can be introduced to the data assimilation tools in order to estimate their temporal evolution with assimilation of high-frequency DO data.

Chapter 4

Quantifying heterotrophic bacteria parameters and dissolved organic carbon biodegradability through oxygen data assimilation in a river water quality model

Manuscript status: Final draft to be submitted soon to Water Resources Research journal.

Résumé

La première mise en oeuvre d'un algorithme de filtre particulaire dans un modèle hydrobiogéochimique a permis de mieux représenter la simulation de l'oxygène dissous dans les systèmes fluviaux. Toutefois, des écarts significatifs persistent pendant les périodes estivales de faible débit. Lors d'une analyse de sensibilité, la partie biodégradable de la matière organique dissoute (CODB) provenant des conditions limites s'est avéré avoir une influence pendant les périodes de faible débit estival hors bloom où une activité de croissance nette bactérienne élevée est observée.

Par conséquent, le CODB est incorporé dans le logiciel d'assimilation de données ProSe-PA et comme preuve de concept, plusieurs études de cas synthétiques sont menées pour prouver que son ajout dans le schéma d'assimilation permet d'améliorer les simulations. Cette modification permet une meilleure simulation de l'oxygène dissous pendant les périodes de faible débit où la croissance nette est élevée, et aide également à mieux identifier la distribution postérieure des paramètres bactériens tels que la croissance bactérienne et le taux de rendement. Ainsi que elle permet de quantifier pour

la première fois la biodégradabilité du carbone organique dissous dans une rivière en fonction des données d'oxygène.

Ensuite, il est démontré qu'au moins deux stations de suivi d'oxygène dissous sont nécessaires pour identifier les paramètres du modèle. Enfin, ProSe-PA est ajusté pour détecter les changements dans les conditions environnementales dus aux changements dans la physiologie bactérienne ou la biodégradabilité du carbone organique dissous. L'étude vérifie également que ProSe-PA est capable d'identifier les changements instantanés et progressifs dans la biodégradabilité du carbone organique dissous.

Abstract

The first implementation of a particle filter algorithm in a hydro-biogeochemical model had improved dissolved oxygen (DO) simulation in river systems, yet mismatches between observed and simulated oxygen remain during summer low-flow periods. To address this issue, a sensitivity analysis was conducted, revealing the influence of the biodegradable fraction of dissolved organic carbon (BDOC) during non-bloom low-flow periods in summer when bacterial net growth activity is high. Therefore, in this study, BDOC is parameterized in ProSe-PA data assimilation software by the integration of an organic carbon partitioning model. As a proof of concept several case studies are developed which demonstrate that the incorporation of the parameter representing dissolved organic carbon biodegradability in the data assimilation scheme of ProSe-PA i) improves DO simulation during low-flow periods, ii) helps identify the posterior distribution of physiological parameters such as maximum bacterial growth rate and bacteria yield, and iii) for the first time quantifies the biodegradability of dissolved organic carbon in a river given oxygen data. Next, it is shown that at least two DO monitoring stations are necessary to identify model parameters whose locations are controlled by BDOC and bacterial activity. Finally, ProSe-PA is configured to detect changes in environmental conditions due to change in bacteria physiology or biodegradability of dissolved organic carbon from boundary conditions. It is achieved by evaluating the role of data assimilation configuration parameters such as the assimilation time step and the amount of perturbation of model parameters values. The study also verifies that ProSe-PA is capable of detecting abrupt and gradual changes in biodegradability of dissolved organic carbon from boundary conditions.

1 Introduction

Organic carbon is the main element determining the river metabolism, global carbon cycle, and human impact in aquatic environments [Battin et al., 2023] under the "active pipe" concept [Cole et al., 2007, Tranvik et al., 2018]. Since the river continuum concept, it is acknowledged that the state of a river system results from the interaction of micro-organisms with organic carbon loadings [Vannote et al., 1980]. River metabolism is evaluated using the dissolved oxygen (DO) concentration in the water column that results from the balance between gross primary production (e.g., photosynthesis by phytoplankton producing organic carbon) and ecosystem respiration (e.g., respiration of bacteria and phytoplankton) [Reichert et al., 2009, Diamond et al., 2021]. River metabolism has been studied since the 1950s [Odum, 1956] to evaluate the balance between organic carbon originating from allochthonous and autochtonous sources [Young et al., 2008] and to define the autotrophy or heterotrophy of an ecosystem [Garnier and Billen, 2007, Escoffier et al., 2018, Gurung et al., 2019, Rodríguez-Castillo et al., 2019, Garnier et al., 2020, Segatto et al., 2020].

Global change through human intervention disturbs river metabolism. River metabolism is influenced by organic carbon and nutrient loadings, water temperature, hydrology, light condition and land use [Uehlinger et al., 2003, Bernot et al., 2010b]. High inflows of organic carbon from different locations, either diffuse or point source, climate change-induced shifts in river discharge, temperature [Blöschl et al., 2019, Yang et al., 2020], and light availability have been responsible for the change in river and stream metabolism [Uehlinger et al., 2003, Battin et al., 2008, Beaulieu et al., 2013]. In particular, urbanization and outflows from treatment plants are found to increase heterotrophic bacterial activity and ecosystem respiration rates [Seidl et al., 1998a, Ruggiero et al., 2006, Izagirre et al., 2008, Tank et al., 2010, Bernot et al., 2010b, Landa et al., 2013, Garnier et al., 2020], being undesirable for aquatic life [Garvey et al., 2007] as they may lead to hypoxia [Diamond et al., 2023]. The organic carbon loadings from allochthonous sources may also induce a shift in the community and thereby a change in their physiological behaviors such as their growth and mortality rates [Garnier et al., 1992a]. To summarize, they shape the river metabolism and their impact is becoming more significant due to climate change, land use and flow regulation [Bernhardt et al., 2018].

River water quality models are used nowadays for management of water quality (e.g., DO levels) in river ecosystems, thereby, they are tools for assessment of river metabolism. An advanced software developed to assess the river metabolism, ProSe-PA [Wang et al., 2022a], couples the ProSe model [Even et al., 1998, 2004, Flipo et al., 2004, Vilmin et al., 2015] with a data assimilation framework using observed oxygen data in river system. Despite providing significant improvement in river metabolism

assessment at high flow and during phytoplankton blooms, there is still place for improvement during summer low-flow periods, which is a current weakness of most of the existing water quality models [Hasanyar et al., 2023b]. River metabolism at low-flow is of significant concern because anthropogenic pressures may overcome natural fluxes and therefore control water quality [Flipo et al., 2020]. This is usually due to low dilution capacity of rivers during such periods. The Seine River is found to be heterotrophic at the downstream of major sewage outflows during low-flow period [Garnier and Billen, 2007, Vilmin et al., 2016, Wang et al., 2022a].

Uncertainties related to the estimation of the fluxes of organic substrates circulating in the system and supporting the microbial loop [Azam et al., 1983, Allan and Castillo, 2009], in particular to that of organic carbon composition and its biodegradable dissolved organic carbon from model boundary conditions, is assumed to be one possible cause of flawed river metabolism estimates at low-flow [Wang et al., 2022a].

As the river metabolism at low-flow is partly controlled by the loadings of BDOC through different point and non-point sources, they need to be accounted for in water quality models. Unfortunately it is very rare to have access to this level of description of BDOC with sampling campaigns and even rarer with monitoring stations at high frequency, while such information exists for dissolved oxygen. This paper therefore proposes the first quantification of BDOC input of model boundary conditions based on *in situ* measurement of high frequency DO under a data assimilation framework and aimed at addressing the following research questions:

1. How does the information about the biodegradability of dissolved organic carbon (DOC) impact oxygen data assimilation in a river water quality model?

2. Where should the DO monitoring (assimilation) stations be located in a river network to improve river metabolism estimates using data assimilation?

3. How to configure the data assimilation framework for detecting environmental changes?

2 Materials and Methods

The ProSe-PA software is shortly presented, followed by first the integration of an organic carbon partitioning model and second the description of its data assimilation framework. Then, to retrieve the model parameters' values including that of dissolved organic carbon biodegradability, a synthetic case mimicking the Seine river at summer low-flow is created and several data assimilation experiments are set up for answering the research questions.

2.1 ProSe-PA

ProSe-PA is an hydro-biogeochemical software that is created by coupling ProSe hydro-biogeochemical model [Even et al., 1998, 2004, 2007, Flipo et al., 2004, 2007, Vilmin et al., 2015] with a particle filter [Wang, 2019, Wang et al., 2022a] to simulate water quality indicators in a river system. It performs data assimilation on high frequency data such as that of DO in order to provide DO simulation results and statistical distributions for the model parameters. ProSe-PA is composed of three libraries. First, the hydrodynamics library that calculates the hydraulics states of the river (discharge, velocity etc.), then the transport library uses these hydraulic outputs to account for advection, dispersion and diffusion of the dissolved and particulate material in the river. Finally, the C-RIVE library [Vilmin et al., 2012, Wang et al., 2018, Hasanyar et al., 2023b] utilizes the outputs of the two previous libraries to calculate the concentrations of water quality indicators along the river system by implementing the biogeochemical equations provided in the RIVE biogeochemical model [Billen et al., 1994]. RIVE holds the complete set of equations that include the kinetics of all the nutrients, organic carbon, O_2, micro-organisms (bacterial, phytoplanktonic and zooplanktonic) and that of the physical processes such as oxygen reaeration and sedimentation-erosion in water column and benthic layer. It has been validated in various studies [Garnier et al., 1995, Even et al., 1998, Thieu et al., 2009, Vilmin et al., 2016, Marescaux et al., 2020].

The DO concentration in a river depends on physical, bacterial, and phytoplanktonic processes as follows:

$$\frac{d[O_2]}{dt} = \frac{d[O_2]}{dt}_{physical} + \frac{d[O_2]}{dt}_{phytoplanktonic} + \frac{d[O_2]}{dt}_{bacterial} \tag{4.1}$$

However, as the bacterial process is found to dominate the metabolism during non-bloom low-flow period [Wang et al., 2018, Hasanyar et al., 2023b], we only look at the bacterial process and its subsequent equations. The effective growth rate (μ_{hb}) depends on the availability of small monomeric substrate (S) and water temperature (Eq. 4.2). The uptake of small monomeric substrate is then calculated by using the growth yield (Y_{hb}, Eq. 4.3). The absorbed small monomeric substrate not used for growth is respired (Eq. 4.4).

$$\mu_{hb} = \mu_{max,hb} \frac{[S]}{[S] + K_s} \, e^{\frac{-(T - T_{opt,hb})^2}{\sigma_{hb}^2}} \tag{4.2}$$

$$Upt_S = \frac{\mu_{hb}}{Y_{hb}} [HB] \tag{4.3}$$

$$\frac{d[O_2]}{dt}_{bacterial} = -\tau_{hb}(1 - Y_{hb})\, Upt_S \tag{4.4}$$

where, $\mu_{max,hb}$: Maximum growth rate of bacteria at $T_{opt,hb}$ [h^{-1}]

S: Small monomeric substrate to be uptaken by bacteria [mgC L^{-1}]

K_s: Monod half-saturation constant for bacterial growth (uptake constant) [mgC L^{-1}]

$T_{opt,hb}$: Optimum temperature for the growth of bacteria [° C]

σ_{hb}: Standard deviation of bacteria temperature function [° C]

Y_{hb}: Growth yield of heterotrophic bacteria [-]

$uptS$: Uptake of substrate for bacteria growth [mgC L^{-1} h^{-1}]

$[HB]$: Heterotrophic bacteria concentration (hereafter, called bacteria) [mgC L^{-1}]

τ_{hb}: 1.0 $[molO_2/molC]$ for full oxidation of organic carbon in the respiration process

2.1.1 Organic carbon repartition model

In the C-RIVE model, the total organic carbon (TOC) is firstly divided into dissolved (DOC) and particulate phases (POC). Then each phase is divided into three pools and each pool is characterized by its proper biodegradability. It is rather rapidly biodegradable ($BDOC_{rapid}$, $BPOC_{rapid}$) or slowly biodegradable ($BDOC_{slow}$, $BPOC_{slow}$) or even refractory ($DOC_{refractory}$, $POC_{refractory}$). It is to be noted that the small monomeric substrate (S) in equation 4.2 is indeed $BDOC_{rapid}$.

$$TOC = BDOC_{\text{rapid}} + BDOC_{\text{slow}} + DOC_{\text{refractory}} + BPOC_{\text{rapid}} + BPOC_{\text{slow}} + POC_{\text{refractory}} \tag{4.5}$$

Five new parameters (t, b_1, b_2, s_1, s_2,) have been introduced recently in C-RIVE to calculate the six pools of organic carbon directly from the total organic carbon (Fig. 4.1). The parameter t indicates the dissolved fraction of TOC and allows for splitting TOC into DOC and POC. The parameters b represent the biodegradable quantity of DOC (b_1) and POC (b_2). Finally, the slowly and rapidly biodegradable fractions are derived using s_1 for BDOC and s_2 for BPOC (Fig. 4.1). The average values and variation ranges of the five parameters were reviewed by Hasanyar et al. [2023a].

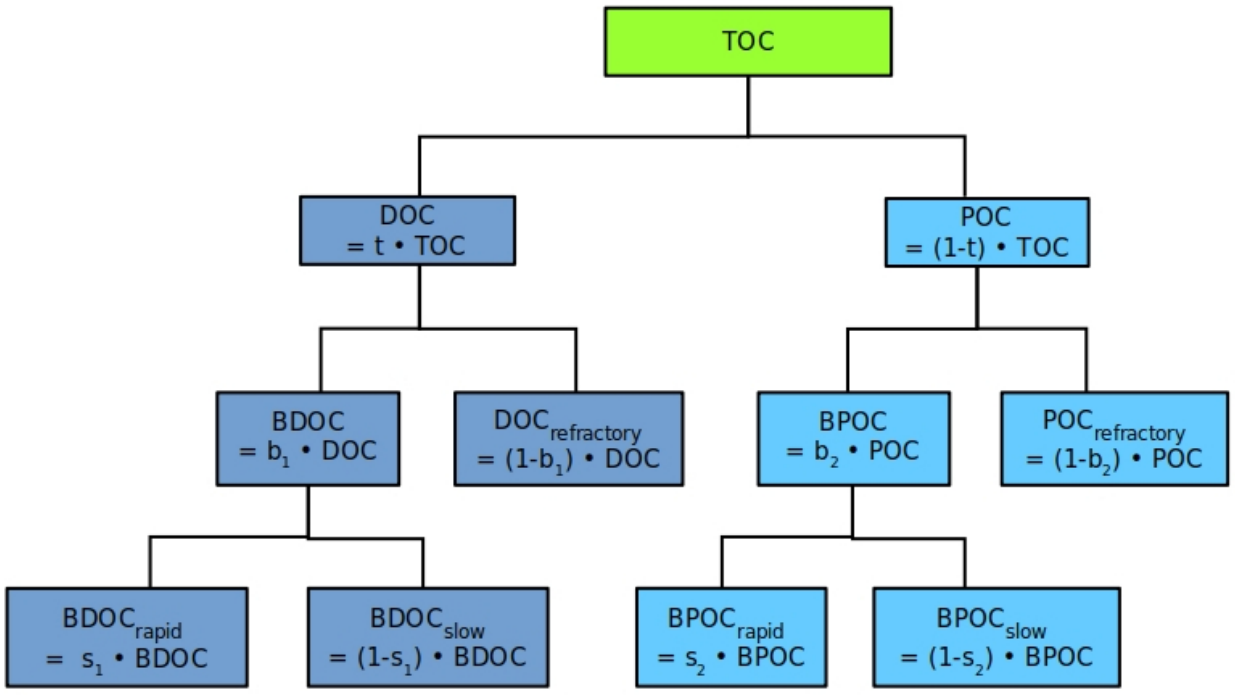


Figure 4.1: The organic carbon partitioning model which distributes the total organic carbon (TOC) among its constituent fractions by using five ratios, namely, t, b_1, s_1, b_2, s_2. DOC: dissolved organic carbon, POC: particulate organic carbon, BDOC: biodegradable DOC, $DOC_{refractory}$: refractory DOC, BPOC: biodegradable POC, $POC_{refractory}$: refractory POC, $BDOC_{rapid}$ & $BDOC_{slow}$: rapidly and slowly biodegradable fractions of DOC, $BPOC_{rapid}$ & $BPOC_{slow}$ rapidly and slowly biodegradable fractions of POC (modified from Hasanyar et al. [2023a])

2.1.2 Selection of model parameters for data assimilation

Among the parameters introduced in Eq. 4.2 - 4.4 , $\mu_{max,hb}$ and Y_{hb} are the influential bacteria physiology parameters based on a sensitivity analysis by Hasanyar et al. [2023b]. They also demonstrated that when looking at DO during heterotrophic low-flow periods, only the biodegradable dissolved organic carbon (BDOC) share of carbon matters. Therefore, the proportion of biodegradable carbon among the dissolved organic carbon (b_1 in eq. 4.6) is added as a new parameter in the particle filter of ProSe-PA:

$$b_1 = \frac{BDOC}{DOC} \tag{4.6}$$

As advocated by Hasanyar et al. [2023b], this study utilizes three key parameters for data assimilation, namely the maximum growth rate ($\mu_{max,hb}$) and growth yield (Y_{hb}) of bacteria, as well as the proportion of biodegradable dissolved organic carbon (b_1) present in the input boundary conditions. These parameters are presented in Table 4.1, along with their respective variation ranges.

Table 4.1: List of model parameters considered for data assimilation and their corresponding variation ranges

Parameter	Description	Min. val.	Max. val.	Unit
b_1	ratio between BDOC and DOC	0.1	0.5	[-]
$\mu_{max,hb}$	maximum growth rate of bacteria	0.01	0.13	[/h]
Y_{hb}	bacterial growth yield	0.03	0.5	[-]

2.2 Data assimilation in ProSe-PA using particle filtering

Wang [2019] first developed the ProSe-PA software, which coupled the hydro-biogeochemical model ProSe with a particle filter [Doucet et al., 2001]. The filter consists in launching N simulations (particles) where each particle represents a simulation with a vector of parameter values. Each parameter is sampled within its variation range (Table 4.1). As a result, N simulations pertaining to N different combination of the model parameters are launched which lead to N different outputs (DO concentrations) at the next time step. Then by comparing the simulated DO with observed DO, this method evaluates the effectiveness of each particle. Particles with parameter combinations that yielded better estimations of DO will receive higher weights, whereas those that produced poor estimations will receive lower ones. As a result, the distribution of each parameter can be approximated by a set of N weighted values that evolve through weight updating and resampling.

With $\mathbf{z}_t$ the state vector composed of the state variable $\mathbf{y}_t$ (DO concentration) and model parameter vector $\mathbf{x}_t$ of size n, and $\mathbf{y}_t^*$ the observations at time t, the data assimilation algorithm of ProSe-PA consists of the following steps (see Wang [2019] for more details):

1. **Model state and parameter initialization:** ProSe-PA samples uniformly and independently the values of the n model parameters and attaches them to each one of the N particles. Since the parameter values are uniformly sampled, ProSe-PA assigns them an initial uniform weight $\omega_0^i = 1/N, i = 1, .., N$.

2. **Forward model prediction:** N simulations are launched until the next time step.

3. **Particle weight calculation:** Once the simulation reaches the next time step where observed data $\mathbf{y}^*$ is available, ProSe-PA calculates a new normalized weight for each particle as follows:

$$\begin{aligned} \omega_t^i &\propto f(\mathbf{y}_t^*|\mathbf{z}_t^i)\omega_{t-1}^i \\ \hat{\omega}_t^i &= \frac{\omega_t^i}{\sum \omega_t^i} \end{aligned} \tag{4.7}$$

where ω_t^i and ω_{t-1}^i represent the posterior and prior importance weights at time t while $\hat{\omega}_t^i$ is the normalised importance weight. $f(\mathbf{y}_t^*|\mathbf{z}_t^i)$ is the likelihood of the particle i, which represents the probability to observe $\mathbf{y}_t^*$ given $\mathbf{z}_t^i$. Assuming the observation error to be Gaussian, it is calculated as

$$f(\mathbf{y}_t^*|\mathbf{z}_t^i) \propto \exp\left(-\frac{1}{2}(\mathbf{y}_t^* - h(\mathbf{z}_t^i))^T \Sigma^{-1} (\mathbf{y}_t^* - h(\mathbf{z}_t^i))\right) \tag{4.8}$$

where h is the observation operator which relates the observed DO ($\mathbf{y_t^*}$) and the simulated DO at time t. Σ is the observation error's covariance matrix.

4. **Particle resampling:** Another alluring aspect of the particle filter is the resampling step to avoid degeneracy of weights whenever the majority of the particles holds very small importance weights. The effective number of particles (N_{eff}) is estimated at every time step as

$$\widehat{N_{eff}} = \frac{1}{\sum_{i=1}^{N} (\hat{\omega}_t^i)^2}. \tag{4.9}$$

This quantity ranges from 1 to N and we resample as soon as $N_{eff} < \alpha \cdot N$, with α generally set around 0.3. The particles with the highest weights are duplicated and their parameter vectors are randomly perturbed in order to diversify the particle values. The amount of perturbation is controlled by a variable dependent parameter corresponding to a user-defined percentage of variation range of the considered variable. This parameter gives the value of the standard deviation of the centered Gaussian perturbation applied to each component of each resampled particle. The reader is referred to Wang et al. [2022a] for further details on particle filter method development in ProSe-PA software.

2.3 Case study

A synthetic case study having a length of L km is designed with a constant discharge of 80 m^3 s^{-1} in order to simulate the low-flow period (Fig. 4.2). The stretch has a width of 100 m and a flow depth of 5.4 m. The constant upstream boundary condition is obtained from Seine River and introduced in the case study. For DO, its saturation concentration at 20 °C is considered (9.1 mgO_2 L^{-1}, Table 4.2). The focus of the study is on the impact of heterotrophic bacteria activity on the water quality, thus no phytoplankton communities are considered.

A DO monitoring station is used for data assimilation ($kp_{assimilation}$) in the case study, and is called assimilation station in the remaining of the paper. A validation station is used to evaluate the model outputs ($kp_{validation}$). The reference DO data (y^{ref}) at those two stations are generated by running direct simulations with fixed values of the three model parameters and with a time step of 15-min. The DO observation data are created by perturbing the reference DO data with a centered Gaussian error (ϵ) with a standard deviation defined by the observation error (σ_{err}) (Eq. 4.10):

$$y^* = y^{ref} + \epsilon, \quad \epsilon \sim N(0, \sigma_{err}^2) \tag{4.10}$$

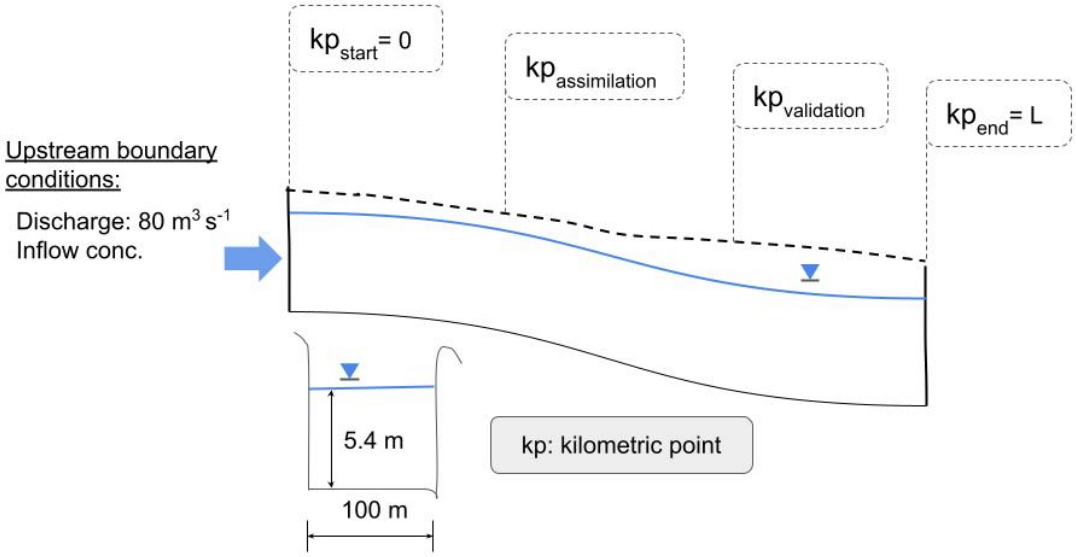


Figure 4.2: Synthetic scheme representing a L km reach of the Seine river. kp_{start}: kilometric point of the starting point of synthetic scheme. $kp_{assimilation}$: kilometric point of the synthetic DO monitoring station used for data assimilation. $kp_{validation}$: kilometric point of the DO monitoring station used for evaluation of model outputs. kp_{end}: kilometric point of the end of synthetic scheme. Please refer to Table 4.3 for reading case study properties' values.

For further information on creation of the synthetic DO time series and properties of the synthetic scheme such as L, $kp_{validation}$, etc., please refer to Table 4.3 and section 2.4.

Table 4.2: Upstream boundary conditions

No	Species	Concentrations (C_i)	Unit
1	TOC	6.0	[mgC/L]
2	NO_2	0.022	[mgN/L]
3	NO_3	3.70	[mgN/L]
4	TSS	3.73	[mg/L]
5	PO_4	0.072	[mgP/L]
6	O_2	9.1 (saturated)	[mgO_2/L]
7	HB	0.01	[mgC/L]
8	Phytoplankton	0.0	[mgC/L]
9	NH_4	0.11	[mgN/L]
10	T	20	°C

2.4 Design of data assimilation experiments

The following experiments are conducted to address the above mentioned research questions (Table 4.3 and 4.4).

2.4.1 Impact of DOC biodegradability on data assimilation

For observing the impact of biodegradability of DOC on data assimilation, experiments are designed to test how this information affects the characterization of model parameters and the simulation of DO concentrations at low-flow. A reference dataset is generated using a share of biodegradable dissolved organic carbon (b_1) of 0.35. A first DA experiment is then performed to retrieve the bacteria growth rate and its yield only using on purpose an inadequate value of 0.25 for b_1, in order to demonstrate the effect of an incorrect information about organic carbon composition. In a second experiment, b_1 is added as an assimilable model parameter. The objective is to evaluate if ProSe-PA is capable of identifying the parameter of dissolved organic carbon biodegradability and those of bacteria physiology thanks to oxygen data assimilation while providing acceptable DO simulation results. Three b_1 values of 0.25, 0.35 & 0.45 are targeted to be retrieved by ProSe-PA. Following the conclusions of Hasanyar et al. [2023b], this experiment is conducted under low and high maximum net growth rate conditions whereas the first experiment is conducted only under high maximum net growth rate conditions. The maximum net growth rate corresponds to the maximum growth rate minus the mortality rate. For the sake of brevity, it is simply referred to as the net growth rate (NG) in the remaining of the paper. The mortality rate of bacteria is fixed at 0.02 h^{-1}. High and low net growth conditions are achieved with maximum bacteria growth rates of 0.07 h^{-1} and 0.03 h^{-1}, respectively (exp. 2). The simulation period is 17 days.

2.4.2 Positioning of the assimilation station

To explore the optimal positioning of the assimilation station relative to an organic carbon source, the data assimilation procedure is replicated by shifting the assimilation station location along the river axis from kp10 to kp90 (exp. 3). The objective is to determine how the DO information location impacts the identification of model parameters and the oxygen simulations at the validation station. The results are analyzed in the light of the following notions:

1. **Observability point**. This is the point upstream of which no parameter could be detected. Upstream of the observability point, the change in DO concentration, compared to the upstream boundary condition at saturation, is lower than that of the measurement error. The location of this point ($kp_{observability}$) is found using the following formula:

$$kp_{observability} = flow\ velocity \times t_{observability} \tag{4.11}$$

$$t_{observability} = \frac{\beta \times \sigma}{(1 - Y_{hb})\frac{\mu}{Y_{hb}} \times [HB] \times \frac{32}{12}} \tag{4.12}$$

where, μ is the effective growth rate and calculated according to Eq. 4.2. $t_{observability}$ is the time required for the organic carbon to reach the pk that would cause a drop in the DO of more than the measurement error. β is the z score and set to 1.96 for a 95 % Gaussian confidence interval. σ is the standard deviation of the observation error (1% of the saturated oxygen concentration).

2. **Limitation point**. It is located at the minimum distance from carbon source for which the concentration of rapidly biodegradable dissolved organic carbon (or substrate, S) limits the bacteria growth. One challenge is to evaluate a mean limiting substrate concentration ($S_{mean,limiting}$) and a mean limiting growth rate ($\mu_{mean,limiting}$). For that the limitation is expressed in the model in the form of a Monod function (Eq. 4.13, Monod [1949]):

$$\mu = \mu_{max,hb}\frac{[S]}{[S] + K_s} \tag{4.13}$$

The limiting substrate concentration ($S_{mean,limiting}$), of 0.24 mgCL^{-1}, is visually detectable on the Monod curve considering a semi-saturation constant, K_S, of 0.025 mgC L^{-1} and a maximum growth rate of 0.07 h^{-1} (Eq. 4.13 and Fig. D1). This limit corresponds to the moment when the

slope of normalized growth rate ceases to decrease linearly and the curve has a slope of around 0.026 $h^{-1}mgC^{-1}L$ (**20°**) at this point. The corresponding limiting growth rate equals 0.063 h^{-1} beyond which growth is not limited anymore by the substrate. It corresponds to **90 %** of the maximal growth rate.

3. **Depletion point**. It is the distance from the organic carbon source where the substrate reaches a critical value such that the bacterial growth cannot compensate anymore the bacterial mortality rate ($\mu_{hb} = mort_{hb}$).

The simulation period is 17 days in these experiments as well. All these three points will be found experimentally for different b_1 values of 0.25, 0.35 and 0.45. A preliminary coarse-grained analysis is first conducted by moving the assimilation station with a 10 km intervals. Once the coarse targeted interval has been detected, the assimilation station is then moved with a smaller increment of 1 km until either the target parameter gets detected for the first time as the case of observability and limitation points (i.e., when the peak of b_1 pdf coincides with its target value) or some of them are no more detectable as in the case of depletion point.

2.4.3 Detection of environmental changes

To assess the performance of ProSe-PA in detecting changes in environmental conditions stemming from either changes in bacterial physiology (i.e., the maximal growth rate, $\mu_{max,hb}$) or biodegradability of dissolved organic carbon from the boundary condition (i.e., changes in b_1), two distinct experiments have been developed for customizing assimilation configuration parameters: assimilation time step and perturbation percentages.

The initial focus is on the changes in bacterial physiology (exp. 4) as bacterial parameters exert greater influence on water quality at low-flow compared to organic carbon partitioning parameters [Hasanyar et al., 2023b]. Moreover, during a bacteria community shift from small to large bacteria, the only physiological parameters that change are $\mu_{max,hb}$ and $mort_{hb}$ [Garnier et al., 1992b]. Here we investigate the change of $\mu_{max,hb}$ from 0.07 to 0.05 h^{-1} while $mort_{hb}$ is fixed in this study. The considered assimilation configuration parameters are assimilation time step (15 min and 1 hour) and perturbation percentages of $\mu_{max,hb}$ (4 % and 8 %) and Y_{hb} (4 % and 8 %). Perturbation percentage is found to play a role in detection of change [Wang et al., 2023] such that a high perturbation allows the particle filter to search a larger space and easily detect time-varying change. However, that is suspected to broader or widen the parameter pdf, to the detriment of accuracy. Thus, a good balance

between precision and detection capacity of parameters is necessary. Trying different assimilation time steps is important in the sense of decreasing the computational cost of DA.

To get reliable DA configuration parameters, each combination of these three parameters has been statistically estimated with 10 DA simulations. A total of 80 DA simulations have thus been launched. The RMSE and probability of success of each combination are evaluated based on the criteria explained in section 2.5. If the combinations demonstrate similar performances, a sample size of 50 is used to redetermine the best configuration. The simulation period is 17 days for detection of change in bacteria physiology with exact time of change being set to 9^{th} day.

Subsequently, the assimilation configuration parameters are fine-tuned for the detection of changes in the biodegradability of dissolved organic carbon from boundary conditions by employing the optimal configurations obtained in the previous experiment. Here we test the change of b_1 from 0.20 to 0.40. The configuration parameter comprises different b_1 perturbation percentages of b_1 (2 %, 4 %, 8 %) (exp. 5). The efficacy of these settings parameters is assessed in detecting both abrupt (0 days) and gradual (5 days) changes. As a result, a total of 60 DA simulations are conducted under this experimental framework. A longer simulation period of 30 days is adopted for the experiments of detection of change in biodegradability to allow for representation of gradual changes. The exact time of this environmental change is set to the 9^{th} day.

Table 4.3: Properties of the case study experiments under different research questions. Columns 5–7 list the values of model parameters used for creation of synthetic DO monitoring data and for retrieval by ProSe-PA. L: length of case study reach. $\mu_{max,hb}$: maximum growth rate of bacteria. Y_{hb}: growth yield of bacteria. b_1: parameter representing the biodegradability of dissolved organic carbon. T: simulation period. NG: Net bacterial growth (values for high and low net growth are 0.05 and 0.01 h^{-1}, respectively). N_{assim}: number of assimilation stations. $kp_{assim,1,2}$: kilometric point of first and second assimilation stations. $kp_{valid.}$: kilometric point of validation station

.

Res. Ques.	Exper. #	Description	L (km)	$\mu_{max,hb}$ (h^{-1})	Y_{hb} (-)	b_1 (-)	T (days)	NG	N_{assim}	$kp_{assim,1}$ (km)	$kp_{assim,2}$ (km)	$kp_{valid.}$ (km)
1	1^{st}	fixed b_1 (DA without b_1)	50	0.07	0.25	0.35	17	high	2	20	40	30
	2^{nd}	b_1 considered for DA	50	0.07	0.25	0.25, 0.35, 0.45	17	high	2	20	40	30
		b_1 considered for DA	50	0.03	0.25	0.25, 0.35, 0.45	17	low	2	20	40	30
2	3^{rd}	DA with moving assimilation station	100	0.07	0.25	0.25, 0.35, 0.45	17	high	1	10-90	-	35
3	4^{th}	change in $\mu_{max,hb}$	100	0.07 → 0.05	0.25	0.25	17	high	2	35	70	40
	5^{th}	change in b_1	100	0.07	0.25	0.2 → 0.4	30	high	2	35	70	60

Table 4.4: Further details of the experiments for evaluating the capacity of ProSe-PA in detecting change in environmental conditions

Exp. #	DA configuration parameters	values	No. of samples	Total no. of simulations
4th	P_{Growth}: perturbation % of bacteria growth rate	4 %, 8 %	10	80
	P_{Yield}: perturbation % of bacteria growth yield	4 %, 8 %	10	
	assimilation time step	15min, 1 hr	10	
5th	P_{b_1}: perturbation % of dissolved organic carbon biodegradability parameter	2 %, 4 %, 8 %	10	60
	type of change	abrupt and gradual (0 and 5 days)	10	

2.5 Numerical settings and evaluation criteria

The number of particles is set to 4000, with a resampling threshold α of 0.3 corresponding to an effective sample size of 1200 (Eq. 4.9). For studying the impact of biodegradability of DOC and location of assimilation station on DA, the standard deviation of observation error (Eq. 4.10) is set to 1% of the observed values, the perturbation percentages are fixed to 1% for $\mu_{max,hb}$ and Y_{hb}, and 2% for b_1 because low perturbations are found to produce accurate results in experiments with stationary parameter values [Wang et al., 2023].

However, in the experiments aimed at detecting changes in model parameters, the observation error is set to 5% in order to facilitate detection of environmental changes, while the perturbation percentage is studied as a data assimilation configuration parameter. The computations are performed using 20 processors (Intel(R) Xeon(R) E5-2640 with a frequency of 2.40GHz) and take 3.5 hours to complete for each experiment with an assimilation period of 17 days.

To evaluate the model performance, the root mean square error (RMSE) is employed by comparing the simulated DO against the synthetic observed DO at the validation station. The RMSE is calculated for the full simulated period. Furthermore, the success in retrieving or identifying of target parameters values under constant environmental conditions is visually evaluated by verifying if the peak of the posterior distribution function (pdf) of each model parameter aligns with the target parameter value. As experiments under changing environmental conditions necessitate selecting the best DA setting configuration, requiring no fewer than ten simulations, a criterion is established to evaluate the probability of success in parameter retrieval. Firstly, the last 5–day window of each simulation is inspected, and one parameter pdf per day is extracted. Secondly, the 80 % confidence interval of each parameter pdf is calculated. Thirdly, a parameter is considered retrieved if its target value falls within the defined confidence interval on at least four out of five days. Lastly, a simulation is deemed successful if all three model parameters meet the parameter retrieval criterion.

3 Results and Discussion

Here we present the results for answering the different research questions. The purpose is to verify if ProSe-PA is capable of retrieving the target parameters under fixed and varying environmental conditions.

3.1 How does the information about the biodegradability of dissolved organic carbon impact oxygen data assimilation in a river water quality model?

3.1.1 Inference of bacterial growth and yield only

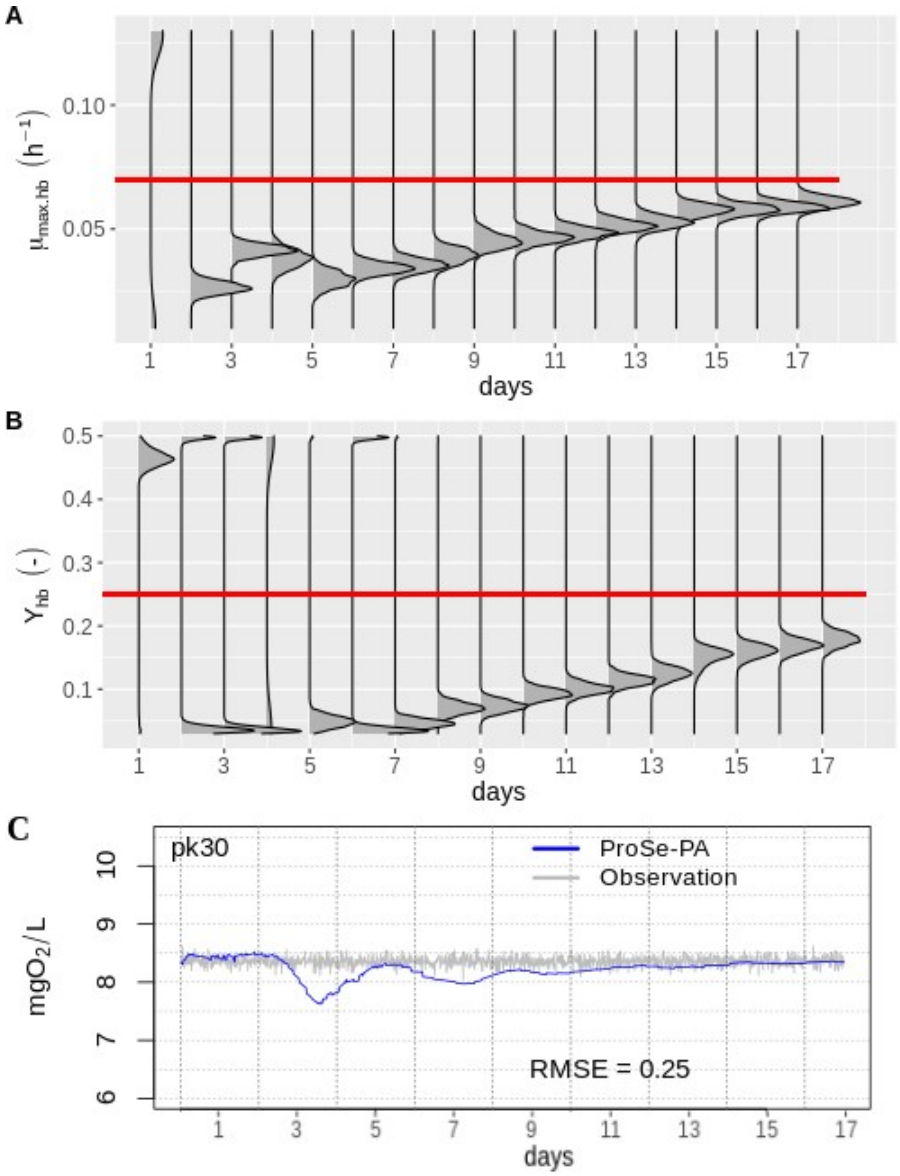


Figure 4.3: Result of DA without characterization of dissolved organic carbon biodegradability (based on exp. 1 setup in Table 4.3). The red line demonstrates the target parameter value which is supposed to be retrieved in DA. (i), (ii) and (iii) the different target b_1 cases of 0.25, 0.35 and 0.45, respectively. A and B) posterior pdfs of bacterial growth and yield. C) Ensemble mean of the simulated DO by ProSe-PA vs. observed DO at the validation station

If the BDOC load provided to ProSe-PA is inadequate, the model fails to retrieve the target parameter values of the physiological properties of bacterial communities shown in red (Figure 4.3A and B). Due to the 3–4 day delay in model boundary conditions' arrival at the first assimilation station, the parameter pdf peaks remain unchanged during the first days of the simulations. However,

upon arrival of the incorrect BDOC (fixed at 25% of DOC instead of the actual value of 35%) at the assimilation station, ProSe-PA converges towards the true value very slowly but ultimately fails to reach it over the 17-day simulation period. Moreover, ProSe-PA is unable to reproduce the observed dissolved oxygen concentrations at the validation station (Figure 4.3C), where consistent underestimation occurs during the simulation period. Thus, it is essential to include biodegradability information in the DA framework. The impact of lack of BDOC information on data assimilation during low-flow periods is also emphasized by Wang et al. [2022a].

3.1.2 Inference of the biodegradability of dissolved organic carbon

DA at high bacterial net growth condition (Fig. 4.4): Incorporating the parameter representing biodegradable dissolved organic carbon (b_1) in the assimilation scheme of ProSe-PA under a condition of high net bacteria growth rate leads to a remarkable agreement between the simulated and observed dissolved oxygen concentrations. It simultaneously provides accurate estimates for both the bacterial growth rate ($\mu_{max,hb}$) and growth yield (Y_{hb}) as well as for b_1 (Fig. 4.4(i) and Fig. 4.4(ii)). Hasanyar et al. [2023b] have brought to light the significant influence of BDOC on the evolution of dissolved oxygen at low-flow occuring under high net growth conditions. This finding elucidates the possibility of identifying b_1 parameter in these experiments.

In case of high biodegradability of DOC ($b_1 = 0.45$), even though wider posterior distributions of b_1 are obtained (Fig. 4.4(iii)A), the dissolved oxygen concentrations are simulated accurately as well as the bacterial physiological properties ($\mu_{max,hb}$ and Y_{hb}). This limitation stems from the fact that the rapidly biodegradable fraction of BDOC ($BDOC_{rapid}$), which is considered as the small monomeric "substrate" utilized for the growth of bacteria (Eq. 4.2 and 4.3), is available in large quantities at the assimilation station. It is assumed that a narrow peaked b_1 distribution is identifiable only when the amount of available substrate is less than a given threshold that limits bacterial growth. Thus, we assume that the oxygen consumption by the bacterial community is no longer restricted by the amount of organic carbon for a b_1 value greater than a given threshold, as discussed in detail in section 3.2 and 3.2.2 where an appropriate strategy is proposed to detect all DOC biodegradabilities.

DA at low bacterial net growth condition (Fig. 4.5): Upon incorporating b_1 into the assimilation scheme under low net growth rate conditions, the simulated DO accurately matches the observed DO for all three b_1 scenarios, while effectively estimating the target values for bacterial growth and yield parameters (Fig. 4.5). However, the target value for the amount of biodegradable dissolved

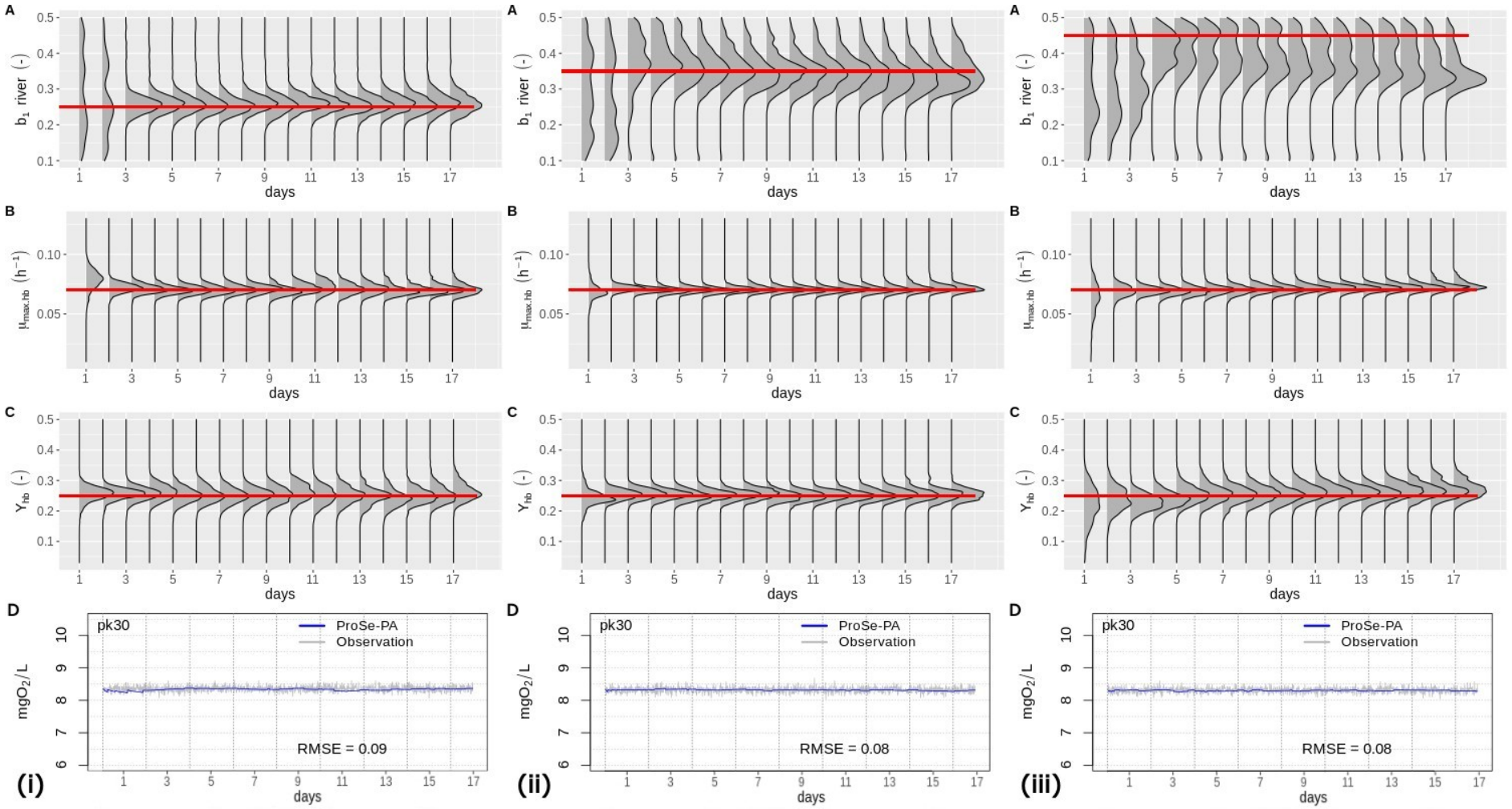


Figure 4.4: Results of DA with estimation of the biodegradability of dissolved organic carbon under high NG condition (based on exp. 2 setup in Table 4.3). The red line demonstrates the target parameter value which is supposed to be retrieved in DA. (i), (ii) and (iii) the different target b_1 cases of 0.25, 0.35 and 0.45, respectively. A, B and C) posterior pdfs of b_1, maximum growth rate of bacteria and bacteria yield. D) Ensemble mean of the simulated DO by ProSe-PA vs. synthetic observed DO.

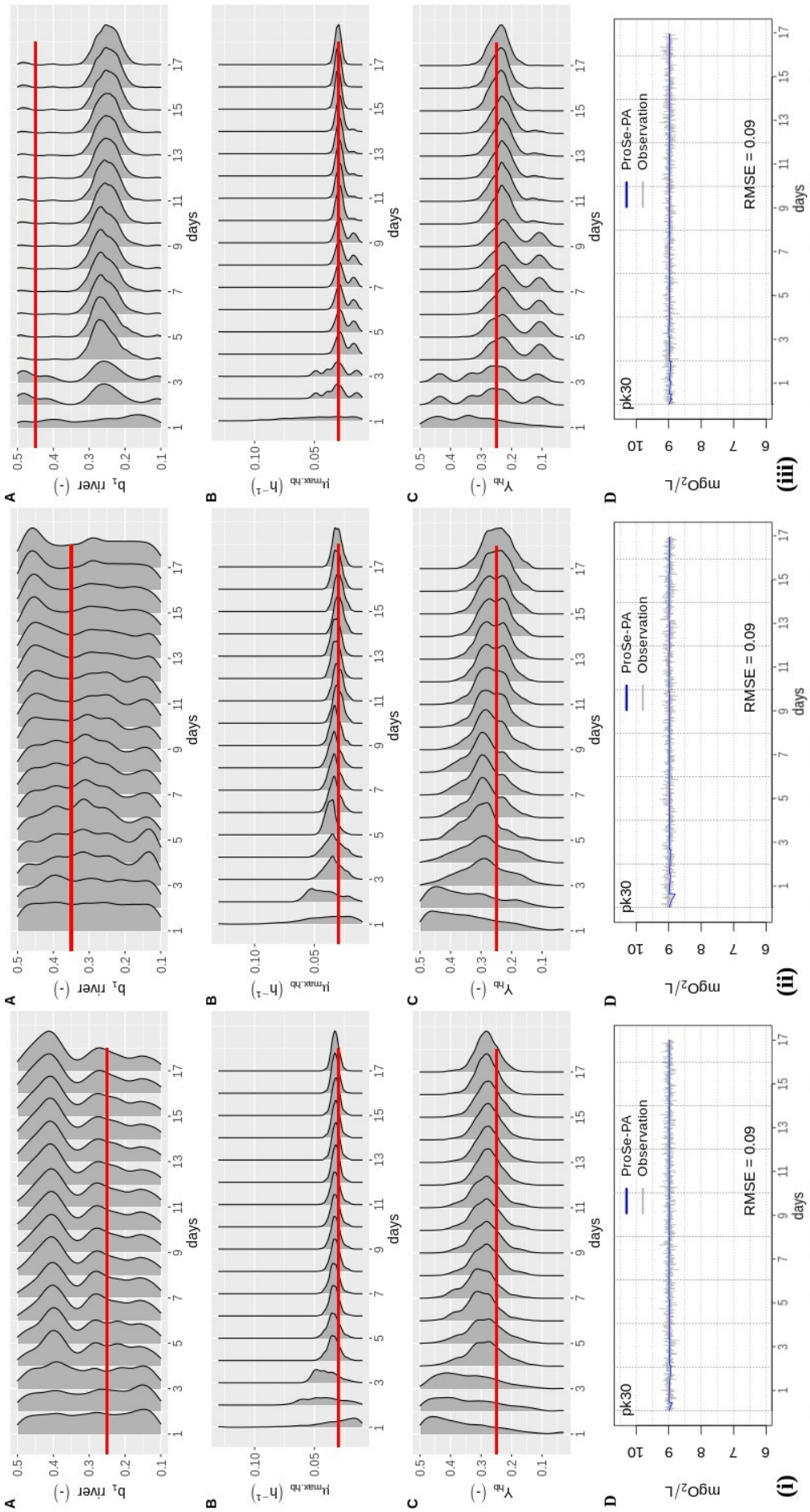


Figure 4.5: Results of DA with estimation of the biodegradability of dissolved organic carbon under low NG condition (exp. 2). The red line demonstrates the target parameter value which is supposed to be retrieved in DA. (i), (ii) and (iii) the different target b_1 cases of 0.25, 0.35 and 0.45, respectively. A, B and C) posterior pdfs of b_1, maximum growth rate of bacteria and bacteria yield. D) Ensemble mean of the simulated DO by ProSe-PA vs. synthetic observed DO.

organic carbon remains unretrieved, and instead, ProSe-PA identifies a quasi-uniform distribution for b_1. This outcome can be explained by the fact that BDOC exerts negligible influence on DO levels under low net growth conditions, as attested by a sensitivity analysis [Hasanyar et al., 2023b].

3.2 Optimal location of two assimilation stations in a river network

This section pertains to the optimal placement of assimilation stations and the minimum number of stations required to accurately estimate model parameters. The DA framework relies on a monitoring network that shall lead to satisfactory DO estimates over time, as well as accurate parameter estimates for DOC biodegradability and bacterial physiology.

3.2.1 Single station strategy

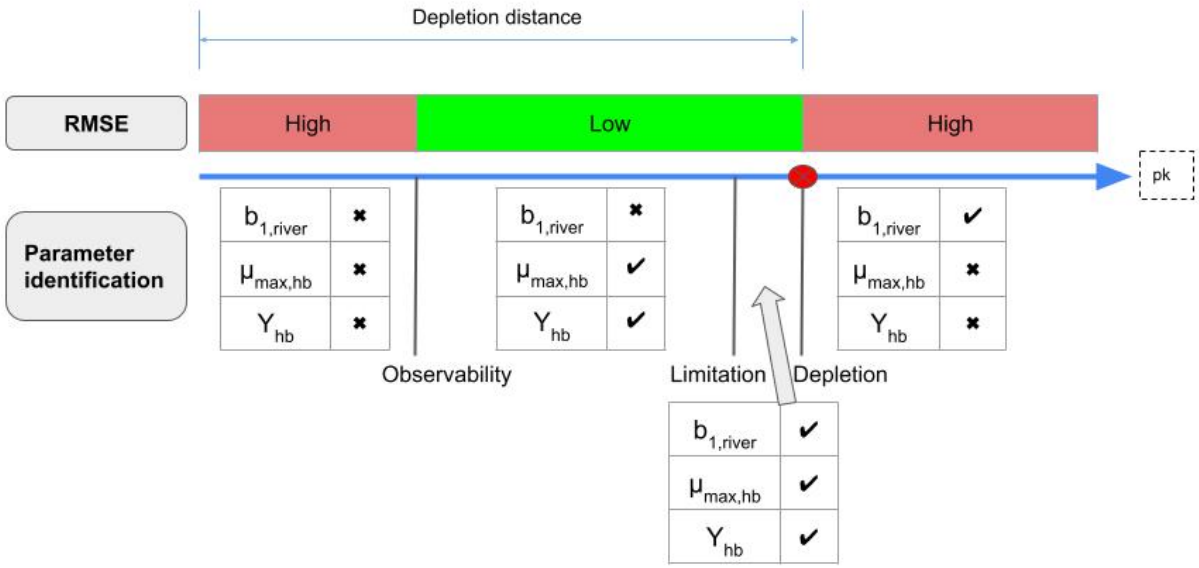


Figure 4.6: Results of moving assimilation station between kp10 to kp90 (exp. 3): evolution of RMSE at the validation station (kp35) with kp of the assimilation station, and the summary of parameter identification. Refer to section 2.4 for the definition of observability, limitation and depletion points.

The assimilation station is moved along the river axis, from kp10 to kp90 (exp. 3). The experiment results confirm the identification of the three points that were previously hypothesized in the DA experiments' setup (Fig. 4.6). The RMSE values are initially high before reaching the observability point, where no model parameter could be retrieved. Subsequently, the RMSE decreases within the observability-limitation window, where bacteria physiology parameters are detectable. The RMSE also remains low within the limitation-depletion window, where all three parameters can be inferred. However, the RMSE value rises again downstream of the depletion point, where only the organic

carbon biodegradability parameter (b_1) can be identified.

The location of the observability point ($kp_{observability}$) is determined to be at a distance of 16 km from the organic carbon source which is in agreement with the value (17 km) calculated theoretically (Eq. 4.12). The absence of identifiable parameters upstream of this point is attributed to inadequate bacterial activity, which fails to induce a significant drop in DO levels. Therefore, ProSe-PA is unable to draw any inferences from the DO time series. Conversely, downstream of this point, $\mu_{max,hb}$ and Y_{hb} are observable as the change in DO is more significant than the measurement error.

As for the limitation point, the experimental results demonstrate that the limitation point is located between kp38–39 for b_1=0.25 leading to a limiting substrate concentration ($S_{limiting}$) of 0.22 mgCL^{-1}. For b_1=0.35 and 0.45, the limitation points are found to be located between kp42–43 and kp45-46, and their limiting substrate concentrations are 0.25 and 0.27 mgCL^{-1}, respectively. Considering a semi-saturation constant, K_S, of 0.025 mgC L^{-1} and a maximum growth rate of 0.07 h^{-1}, these substrate values lead to limiting growth rates of 0.62–0.64 h^{-1} (Eq. 4.13) which is in agreement with the theoretical developments of section 2.4.2, that led to a mean limiting substrate concentration of 0.24 mgCL^{-1} and a mean limiting growth rate of 0.063 h^{-1}.

Furthermore, the critical value that marks the point of substrate depletion is determined to be 0.01 mgC L^{-1}. In this experiment, the depletion point is detected to be at kp43 for b_1 of 0.25. However, the location of depletion point increases with increase in b_1. Beyond this point, only b_1 can be accurately estimated, while $\mu_{max,hb}$ and Y_{hb} exhibit rather uniform pdfs. This is because, regardless of their actual values, the net bacterial growth rate is negative when the substrate concentration falls below the critical value due to the dominance of mortality over growth rate.

In order to retrieve the three parameters, the assimilation station station must be located in a limitation-depletion window, whose length varies with K_S and is generally narrow. Nevertheless the window location shifts with changes in b_1 or other parameters as depicted in Fig. 4.7. A one station strategy is therefore not suitable for environmental monitoring because parameters values are always changing in natural conditions. Back to the high NG condition, let's focus on the b_1 growth from 0.25 to 0.45 to illustrate changing environmental conditions. With an assimilation station at kp40, we showed earlier that the filter was not able to infer the highest content of biodegradable organic matter (b_1=0.45, see Figure 4.4). Plotting the mean limiting substrate concentration (0.24 mgCL^{-1}) on longitudinal profiles of the substrate ($BDOC_{rapid}$, see Fig. 4.7), it is apparent that the substrate concentration for b_1 values of 0.25 and 0.35 at the second assimilation station (kp40) is lower than the growth limiting substrate level, while for b_1 value of 0.45, it exceeds the limit. Thus, we assume that to

quantify b_1, at least one assimilation station must be located at a sufficient distance from the organic carbon source, where the rapidly biodegradable fraction of dissolved organic carbon (substrate) is lower than the growth limiting substrate value.

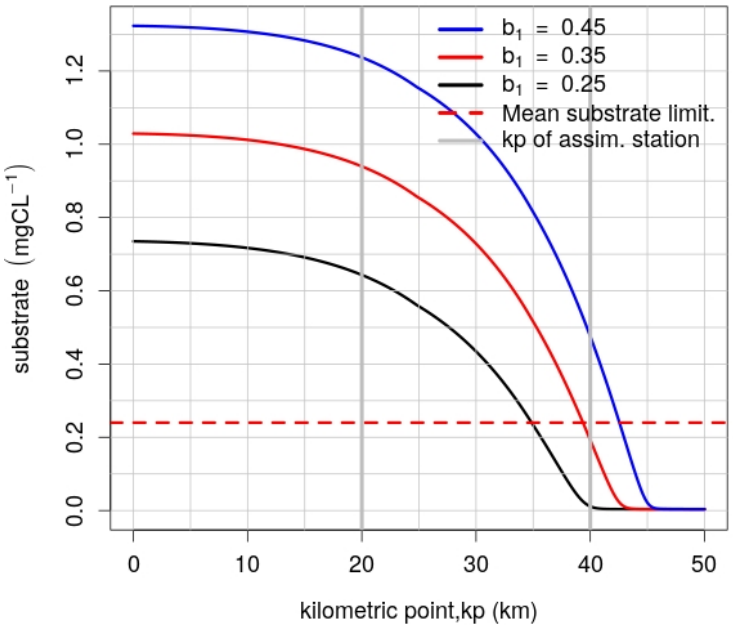


Figure 4.7: Longitudinal profiles (in kilometer) of the small monomeric substrate (rapidly biodegradable dissolved organic carbon, $BDOC_{rapid}$) for different b_1 values. The two vertical grey lines indicate the location of the two assimilation stations. The red dashed-line is the mean substrate limit.

3.2.2 A two station strategy for the detection of all parameters

An optimal approach would be to use one assimilation station in the observability-limitation window to detect $\mu_{max,hb}$ and Y_{hb}, and another downstream of the limitation point to detect b_1, enabling the detection of all three parameters with an acceptable level of RMSE. To assess the effectiveness of this strategy, a DA test was conducted using two assimilation stations at pk25 (in the observability-limitation window) and pk50 (in the depletion window, given the target b_1 value of 0.45 whose depletion distance is 46 km as shown in Fig. 4.7). The result of this test demonstrates not only a good match between the three parameter pdfs and their target values, but also an accurate oxygen simulation (Fig. 4.8).

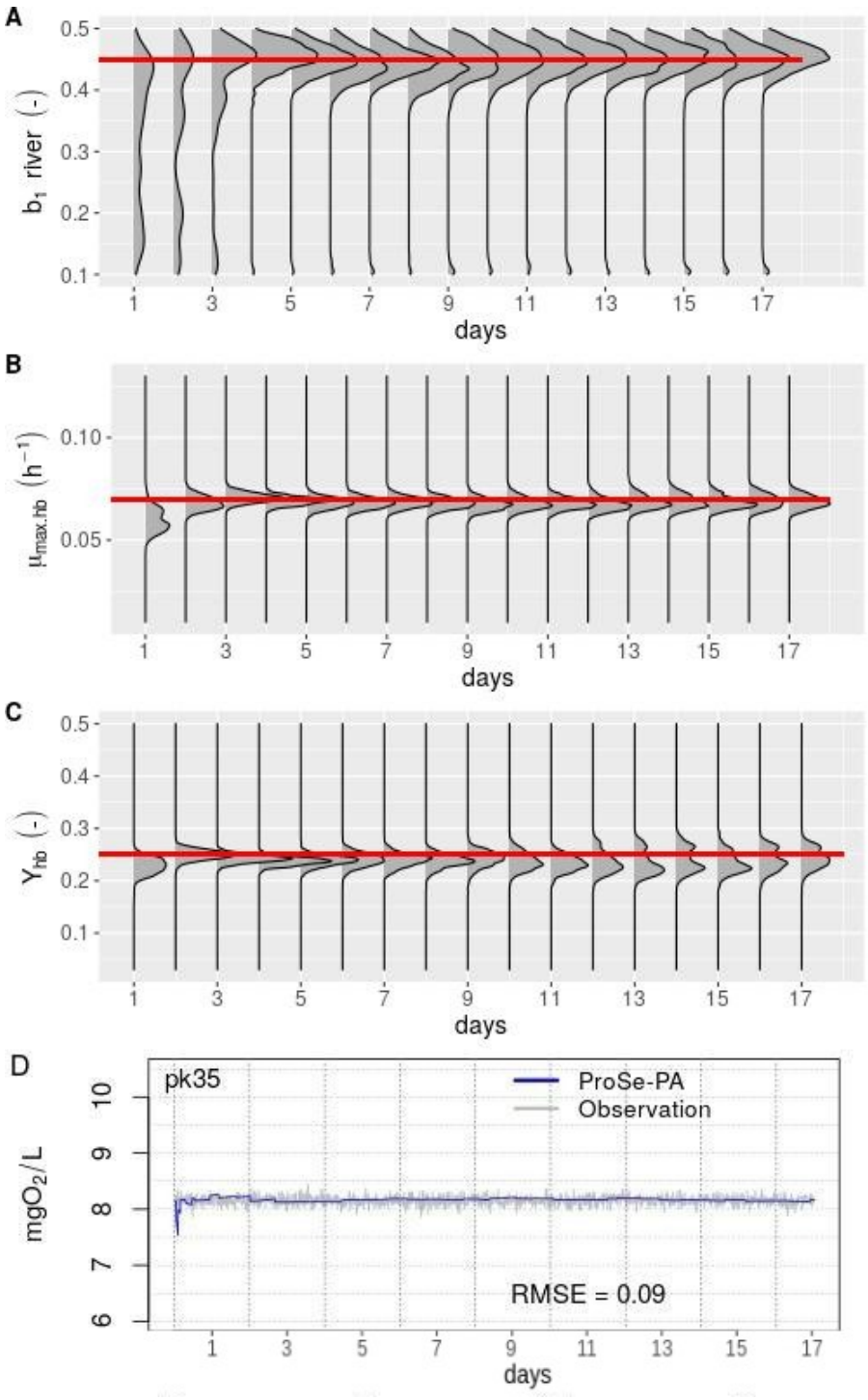


Figure 4.8: Results of the double station strategy. A, B, and C) the posterior pdfs of b_1, maximum bacterial growth rate and bacteria growth yield. Red lines show the target parameters values to be retrieved by DA. D) Ensemble mean of the simulated DO by ProSe-PA vs. observed DO

3.3 How to configure the data assimilation framework for detecting environmental changes?

In this section, the initial step involves placing two assimilation stations, one at kp35 (within the observability-limitation window) and the other one at kp70 (within the depletion window) as proposed in section 3.2.1. The capacity of ProSe-PA in detecting changes in the parameter of bacteria community ($\mu_{max,hb}$) and biodegradability of organic carbon from boundary conditions (b_1) is then assessed. For this purpose, the assimilation setting parameters, that consist of the assimilation time step and the perturbation percentages of model parameters, are configured.

3.3.1 Detection of change in bacteria physiology

Among the 60 different combinations of DA configuration parameters (Table 4.4), the optimal combination for successful detection of changes in bacteria physiology consists of an assimilation time step of 15 minutes and a perturbation of 4% for both the maximum growth rate and growth yield of bacteria (Table 4.5). Notably, this combination achieved a success rate of 100% and a mean RMSE of 0.10 $mgO_2.L^{-1}$, indicating its superior performance compared to other combinations.

It is worth mentioning that another combination with a similar assimilation time step and maximal growth rate perturbation % but a different growth yield perturbation % (8%) also demonstrated a 100% success rate and a mean RMSE of 0.11 $mgO_2.L^{-1}$ (Fig. D2). However, upon conducting 50 DA simulations for each combination, the former combination exhibited the most outstanding performance, with 49 successes among the 50 simulations (49/50) for detecting changes in bacteria physiology.

Table 4.5: Results of the optimum combination of data assimilation configuration parameters for detecting change in bacteria physiology and dissolved organic carbon biodegradability

Change in environmental condition	DA configuration parameters	Optimal comb.	Success (/10)	Mean RMSE
Change in bacteria physiology	Perturbation % of maximum growth rate	4%	10	0.10
	Perturbation % of growth yield	4%		
	assimilation time step	15 min		
Change in biodegradability of DOC	Perturbation % of DOC biodegradability parameter	8%	8	0.08
	type of change	gradual		

It is noteworthy that the same perturbation percentage of 4% for both parameters was found to provide the highest success rate (41/50) even for a 1-hour DA time step (Fig. D2). Therefore, this

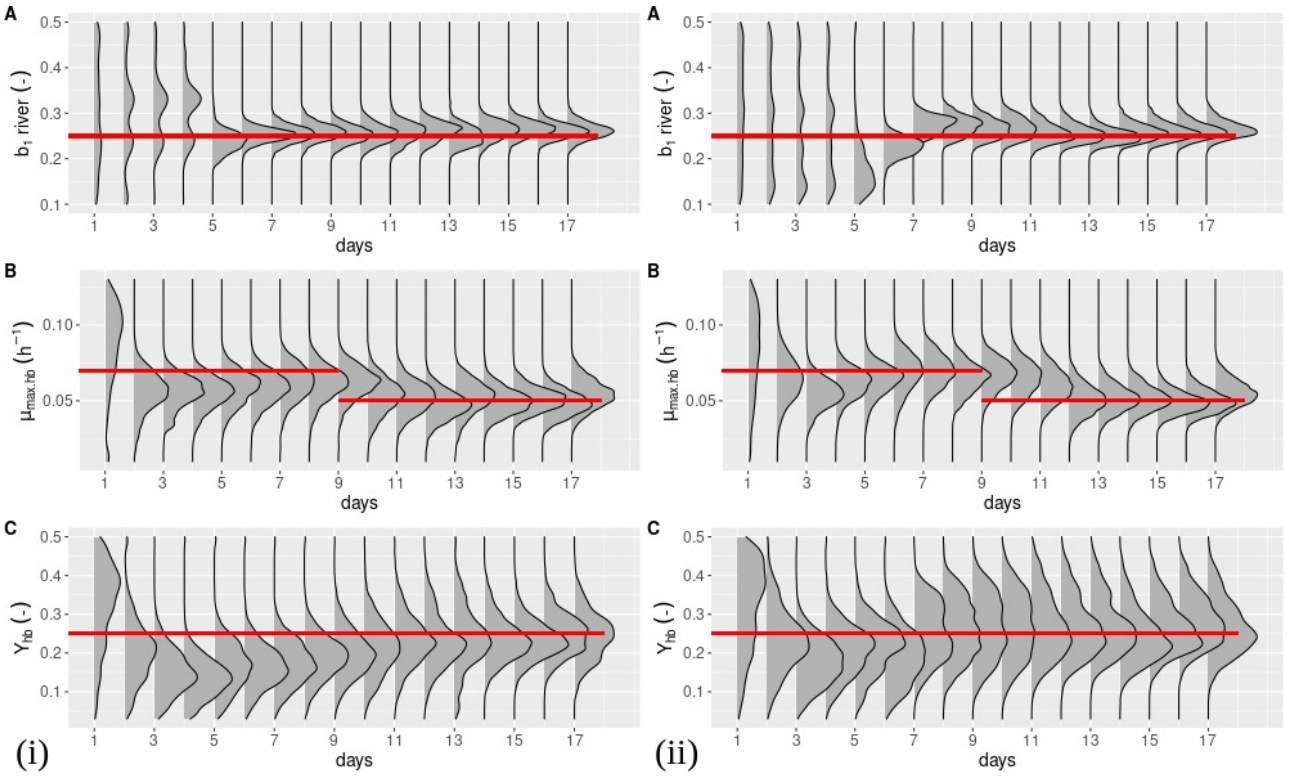


Figure 4.9: Retrieval of target parameters while detecting change in bacteria physiology (exp. 4). (i) DA with 15 min time step. (ii) DA with 1 hr time step. Red lines show the target parameters values to be retrieved by data assimilation. Both cases have 4% perturbation for bacteria physiology parameters.

perturbation percentage is deemed to be the most effective configuration for detecting changes in bacteria physiology, regardless of the assimilation time step.

Moreover, the results indicate that a 15-minute assimilation time step yields slightly lower mean RMSE values and faster detection speed compared to a 1-hour assimilation time step (Fig. 4.9). However, both configurations are successful in detecting model parameters and the 1-hour assimilation time step is selected as a reasonable configuration for detecting changes in bacteria physiology in this study. The choice of a higher assimilation time step is pragmatic and relies on the computational cost of a 1hr assimilation time step compared to a 15 min time step, *per se* four times lower for the former.

3.3.2 Detection of change in DOC biodegradability

Utilizing the optimal configuration parameters, it is found that the most effective b_1 perturbation percentage for detecting changes in dissolved organic carbon biodegradability is 8% (Table 4.5). This perturbation percentage corresponds to gradual changes in biodegradability. The success rate of this configuration is 80%, with a mean RMSE of 0.08 $mgO_2.L^{-1}$.

Moreover, it was observed that gradual changes in biodegradability, as opposed to abrupt changes,

led to better model simulations, with a lower RMSE of 0.08 $mgO_2.L^{-1}$ compared to 0.29 $mgO_2.L^{-1}$, and a higher likelihood of detecting model parameters, with 8 out of 10 compared to 5 out of 10, respectively (Fig. D3). ProSe-PA is therefore better equipped to detect changes in organic carbon biodegradability from sources with gradual changes, such as tributary rivers, rather than those with abrupt changes, such as combined sewage overflows.

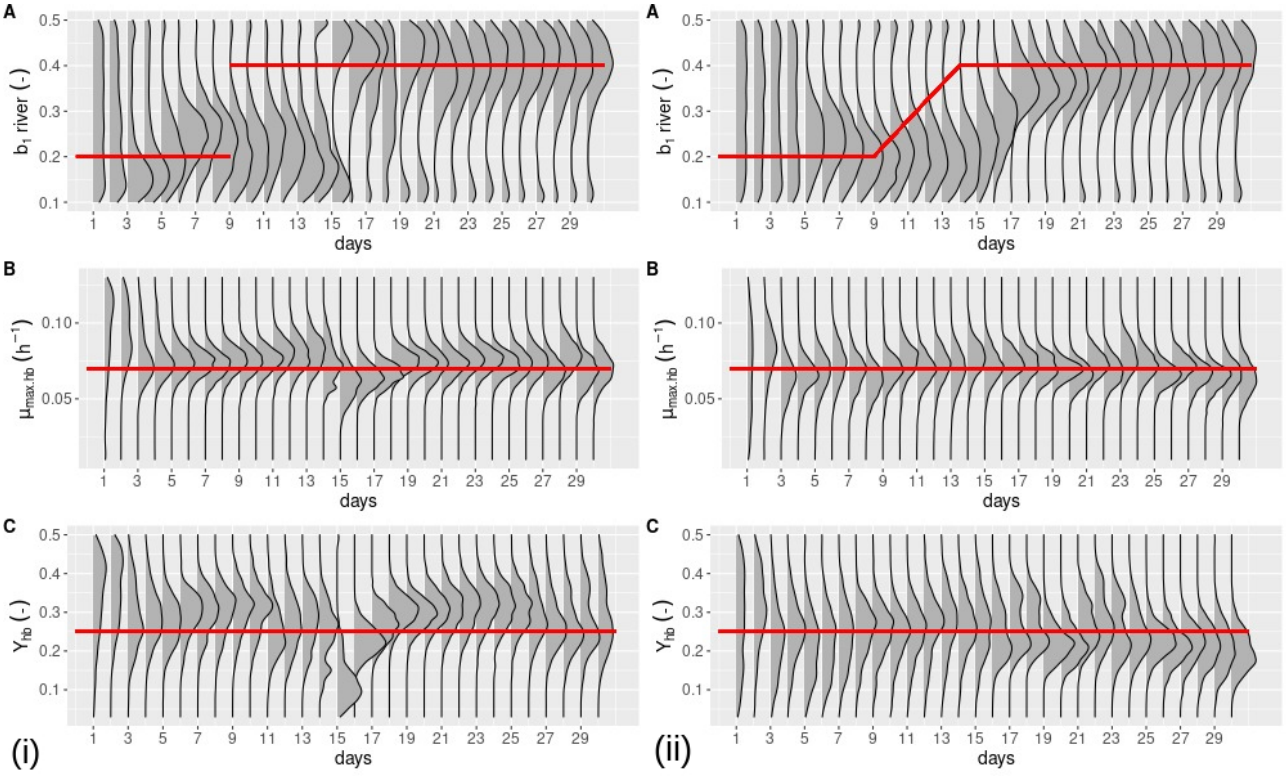


Figure 4.10: Retrieval of target parameters while detecting change in biodegradability of dissolved organic carbon from boundary conditions with 1 hr time step (exp. 5). (i) Abrupt change in b_1. (ii) Gradual change in b_1. Red lines shows the target parameters values to be retrieved by data assimilation.

ProSe-PA detects successfully the model parameter values in case of both abrupt and gradual changes in dissolved organic carbon biodegradability, using a perturbation percentage of 8% for b_1 (Fig. 4.10). However, there is a delay of 5 to 6 days in detecting the change itself, which can be attributed to the time taken by organic carbon to travel from its source at kp0 to the second assimilation station at kp70, located within the depletion window. With a flow velocity of 0.18 m/s, it is estimated that it takes approximately 5 days for the organic carbon to reach kp70. A similar delay is also observed at the beginning of the simulation (days 1-5).

In the case of an abrupt change in biodegradability, the pdf of b_1 also exhibits an abrupt change on day 14 (Fig. 4.10(i)A), whereas for a gradual change, the pdf of b_1 changes gradually starting from day 14 (Fig. 4.10(ii)A). Additionally, the computationally optimal configuration parameters for detection

of changes in bacteria physiology, consisting of a 1-hour assimilation time step and 4% perturbation for both bacteria growth rate and growth yield, also proved to be effective for detecting changes in organic carbon biodegradability and thereby confirms the efficiency of the protocol which was used in the design of experiments for detecting environmental changes.

Overall, the findings shed light on the importance of carefully selecting the DA configuration parameters to achieve optimal performance in detecting environmental changes, while also considering practical constraints and computational resources. They provide valuable insights for future investigations aiming to improve their applicability in various domains, including environmental monitoring.

4 Conclusions

The objective of this study is to ascertain whether ProSe-PA is capable of quantifying bacteria physiology and dissolved organic carbon biodegradability through the assimilation of oxygen data during a summer low-flow period. To achieve this goal, a mono channel like synthetic test case is developed, which reflects the conditions present in the Seine river that flows through the urban area of Paris city. To enable parameterization of organic carbon biodegradability, a novel organic carbon partitioning model is integrated into the ProSe-PA water quality model, in addition to the pre-existing parameters that represented bacteria physiology.

To begin with, this study demonstrated that the presence of uncertain information regarding the biodegradability of dissolved organic carbon from model boundary conditions can result in erroneous oxygen simulation outcomes, while simultaneously hindering the quantification of model parameters.

Next, the inclusion of the parameter of organic carbon biodegradability in the data assimilation framework under high net growth conditions led to a remarkable agreement between the simulated and observed oxygen. For the first time, we were able to quantify the biodegradability of organic carbon through oxygen data assimilation.

However, it has been validated that under low net growth conditions, the biodegradability of organic carbon cannot be effectively observed because DO is insensitive to BDOC in such conditions. Therefore, data assimilation characterizes a quasi uniform distribution for biodegradability of organic carbon during low growth conditions.

Furthermore, the study revealed that when investigating the identifiability of model parameters under varying levels of substrate availability, the biodegradability of organic carbon cannot be accurately identified if the organic carbon substrate is not a limiting factor for bacterial growth, The limit is found to be at around 90 % of maximum bacteria growth rate.

The usage of at least two assimilation stations is prescribed for retrieval of predefined parameters under varying environmental conditions (change in model parameter values). The significance of having multiple assimilation stations along a river stretch is highlighted by the fact that each parameter exhibits an optimal detection window within which the respective parameter values can be accurately estimated by the ProSe-PA model. To be specific, for identification of bacteria physiology parameters, an assimilation station placed within the observability–limitation window is deemed optimal, while for the parameter of organic carbon biodegradability, an assimilation station situated downstream of the limitation point is considered preferable.

To detect change in bacteria physiology and DOC biodegradability, appropriate assimilation settings such as assimilation time step and perturbation % of parameters are prescribed. With such a configuration, ProSe-PA is able to detect abrupt and gradual changes in biodegradability of dissolved organic carbon from boundary conditions.

Chapter 5

Improvement of high frequency oxygen data assimilation in river systems by incorporation of dissolved organic carbon biodegradability during low-flow periods

Manuscript status: It is currently in a draft stage, and the journal is to be decided.

Résumé

Les résultats et recommandations des chapitres précédents sont appliqués ici dans une étude de cas réel. Le but est d'améliorer la simulation du métabolisme de la Seine au cours d'une saison de faible débit en 2011.

Les résultats montrent que l'assimilation de données d'oxygène avec estimation de la biodégradabilité du carbone organique dissous donne non seulement une très bonne correspondance entre l'oxygène dissous observé et simulé dans la majorité des stations de surveillance d'oxygene dissous, et nous permet également d'estimer l'évolution temporelle des paramètres de biodégradabilité et de croissance bactérienne. Les valeurs trouvées pour la biodégradabilité sont cohérentes avec les données expérimentales de la littérature scientifique.

L'assimilation de données avec des petites et des grandes espèces bactériennes a également été incluse dans ProSe-PA avec un test qui a montré une plus grande stabilité dans l'estimation des paramètres du modèle par rapport au cas d'une seule espèce bactérienne. Cependant, cela a conduit

à l'amélioration des simulations dans seulement quelques stations, ce qui montre la nécessité d'une analyse plus approfondie sur la part des deux communautés bactériennes provenant des conditions aux limites.

Abstract

The results and recommendations from the previous chapters are applied here in a real case study to improve the simulation of Seine River metabolism during a low-flow season in 2011.

The results showed that the assimilation of oxygen data with estimation of the biodegradability of dissolved organic carbon not only improved DO simulations in majority of the DO monitoring stations, but also can characterize the temporal evolution of the parameter of biodegradability and that of bacterial growth. The values found for biodegradability is consistent with the experimental data in the literature.

Data assimilation with both small and large bacterial species has also been included in ProSe-PA, with a test showing greater stability in model parameter estimation compared to the case of a single bacterial species. However, this has only led to improvements in simulations in a few stations, highlighting the need for a more in-depth analysis of the proportion of the two bacterial communities coming from boundary conditions.

1 Introduction

Data assimilation including the estimation of carbon biodegradability is applied to the simulation of river metabolism in the Seine River during the low-flow period of 2011. The Seine River low-flow condition is a significant concern owing to the absence of dilution and the significant impact of outflows from urban point sources, such as water treatment plants and combined sewage overflows. Despite the remarkable improvements in the simulations of dissolved oxygen (DO) by ProSe-PA, the existence of mismatches between observed and simulated oxygen during the low-flow period of 2011 was still noticed [Wang et al., 2019, 2022a]. It was hypothesized that the uncertainties in the organic carbon inflow and its degradation kinetics from model boundary conditions contributed to these discrepancies. To address these uncertainties, a new organic carbon repartition model was developed, allowing for the parameterization of the organic carbon inflow and its subsequent integration into the ProSe-PA data assimilation framework. Therefore, a bibliography review was conducted to determine the proposed parameter's variation range (Chapter 2), followed by a sensitivity analysis to identify the influential

parameters on oxygen evolution at low-flow (Chapter 3). Among the newly proposed parameters, only the parameter representing the biodegradability share of dissolved organic carbon (b_1) was found to control metabolism during low flow, in addition to bacterial parameters such as maximum growth rate and growth yield. Thus, a new organic carbon model was incorporated into ProSe-PA, and its influential parameter was integrated among the existing model parameters considered for data assimilation. As a proof of concept, and using a synthetic case study, it was demonstrated that ProSe-PA is capable of identifying not only the biodegradability of dissolved organic carbon, but also the alterations in its biodegradability, through oxygen data assimilation (Chapter 4). This chapter aims to assess the efficiency of these developments in a real case scenario.

2 Study area and study period: Seine River during 2011

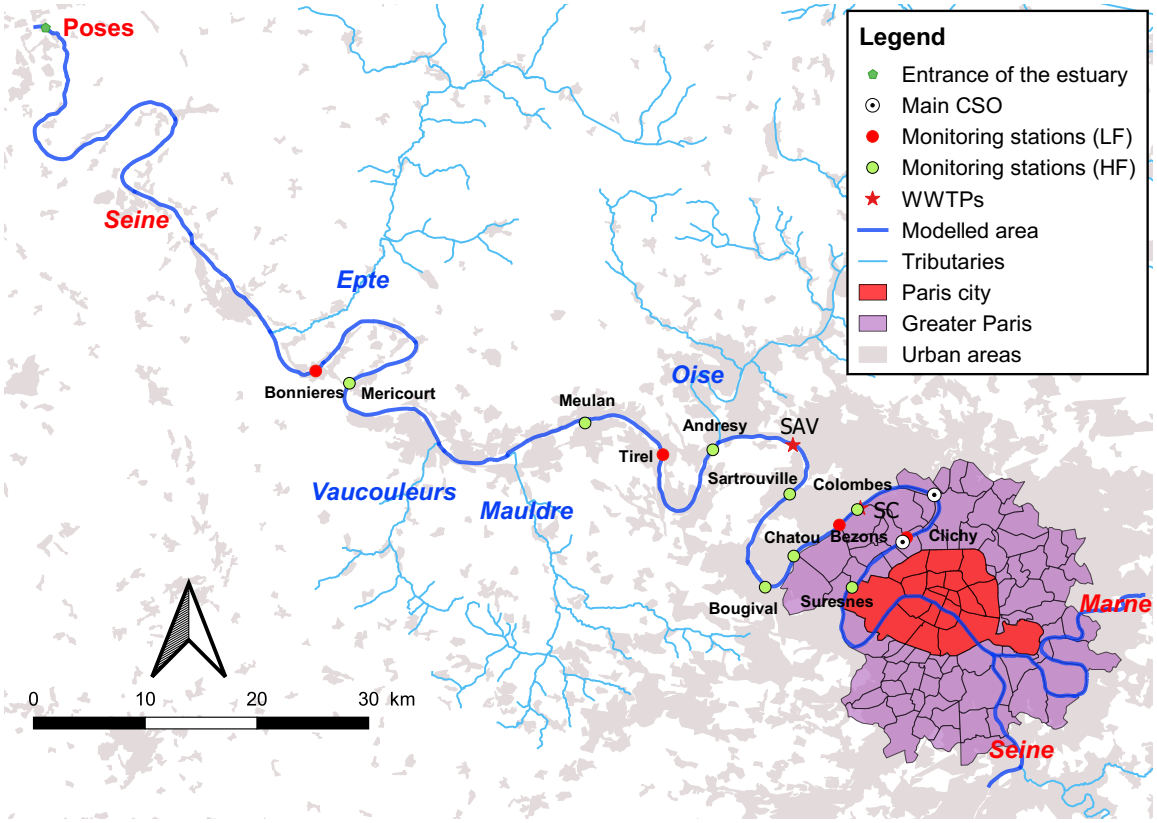


Figure 5.1: The Seine River stretch modeled in this study. The 8 high frequency monitoring stations (green circles) from upstream to downstream: 1. Suresnes, 2. Colombes, 3. Chatou, 4.Bougival, 5. Sartrouville, 6. Andresy, 7. Meulan, 8. Mericourt. 2 treatment plants (red stars): Seine Aval (SAV) and Seine Centre (SC). 2 main CSOs (black circles): Clichy and Garges-Epinay. (Retrieved from Wang et al. [2022a]).

The real case scenario examined in this study encompasses a 220 km stretch between the upstream of Paris city and the Seine estuary at Poses (Fig. 5.1), where the Seine and Marne rivers serve as

the upstream boundary conditions, while their tributaries, such as Oise, are considered as lateral inflows into the model. Furthermore, the Seine River basin is industrial, agricultural and heavily urbanized as it hosts 25 % of French population [Flipo et al., 2020] and therefore subjected to huge urban organic carbon loads from 5 wastewater treatment plants (WWTP) whose data is provided by SIAAP. Additionally, the excess stormwater during rain events is drained into the Seine River via the 156 combined sewage overflows (CSO) [Seidl, 1997, Seidl et al., 1998b, Even et al., 2007], with 99 % of them located upstream of the Bougival monitoring station.

2.1 Situation of Seine River during low-flow periods of 2011

Located in a temperate climate, the Seine River has a rainfall fed regime, that exhibits high flow periods during the winter season and low flow periods during the summer. During the dry 2011 year, the Seine River experienced low-flow conditions for most of the year (Fig. 5.2A), resulting in dissolved oxygen concentrations that are below saturation levels, indicating a high level of heterotrophic bacterial activity (Fig. 5.2B). The Austerlitz gauging station recorded 2011 as a relatively dry year having a mean daily discharge of 226.6 m^3s^{-1} (to be compared to the pluri-decadal average of 320 m^3s^{-1} mentioned by Flipo et al. [2020]), with the maximum daily discharge reaching 1020 m^3s^{-1} and the minimum daily discharge dropping to 66.2 m^3s^{-1}.

The low-flow period of 2011 occurs between days 100-340, which coincides with the occurrence of three successive blooms between days 120-200 [Vilmin et al., 2016]. Thus, for the scope of this study, the low-flow period is delimited to the non-bloom days of 200-300, during which substantial decreases in oxygen levels are discernible while flowing downstream from the upstream monitoring stations. Moreover, a particularly dry period is also observed during this low-flow period between days 260 and 280.

2.2 Model input data

This study uses daily water quality data from Seine, Marne, and Oise rivers provided by SEDIF, as well as weekly chl *a* measurements from SIAAP for Seine River and 15-min chl *a* concentrations from SEDIF for Marne River. For smaller tributaries, lower frequency water quality data from the national river monitoring network (RCS) are used. River daily discharge data are obtained from Hydroportail database, and the daily water flow and quality data of the five WWTPs are also provided by SIAAP.

In order to carry out oxygen data assimilation, we make use of the high-frequency dissolved oxygen (DO) concentration data furnished by SIAAP for the 8 monitoring stations depicted in Figure 5.2B

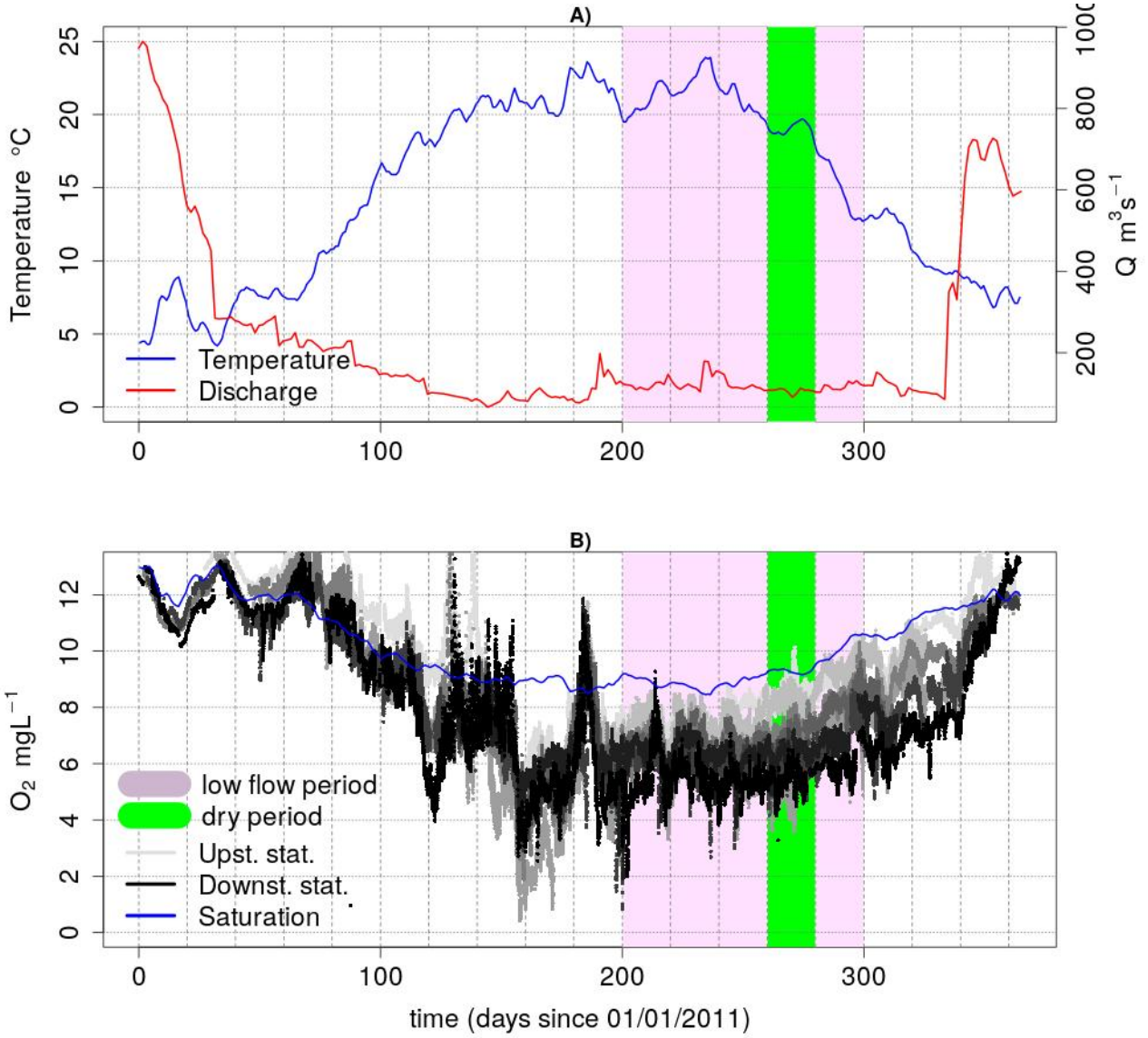


Figure 5.2: Situation of Seine River in 2011. Highlighted polygons denote the low flow period (purple) and the dry period (green). A) Evolution of dissolved oxygen at the 8 monitoring stations. (color gradient: grey to black representing upstream to downstream stations). B) Daily discharge at the gauging station and temperature at a monitoring station.

and listed in Table 5.1. The primary aim of this data assimilation exercise is to enhance the simulation of river metabolism during low-flow periods thanks to these 8 assimilation stations, and facilitate the identification of key model parameters that are known to exert a significant influence on the DO simulations at low-flow (Table 5.3 and Hasanyar et al. [2023b]).

Table 5.1: DO concentrations in 2011 at the 8 monitoring stations to be used for data assimilation (retrieved from Wang et al. [2022a])

Station no.	kilom. point (kp)	Station name	Number of measurements	Frequency
1	674.00	Suresnes	30948	15 min
2	694.53	Colombes	15994	15 min
3	701.40	Chatou	29626	15 min
4	706.40	Bougival	33971	15 min
5	715.00	Sartrouville	14912	15 min
6	729.50	Andresy	30586	15 min
7	751.00	Meulan	16738	15 min
8	779.00	Mericourt	32680	15 min

3 Developments in the source code of ProSe-PA to control organic matter inflow of model boundary conditions

As the fundamental premise of this thesis is centered around the integration of a novel organic carbon repartition model to parameterize the inflow of organic carbon from model boundary conditions, it is important to examine how organic carbon flow was treated formerly.

As elaborated in earlier chapters, ProSe-PA recognizes the organic carbon contribution of any boundary condition in terms of six small fractions that comprise $BDOC_{rapid}$, $BDOC_{slow}$, $DOC_{refractory}$, $BPOC_{rapid}$, $BPOC_{slow}$, and $POC_{refractory}$. Vilmin [2014] established these fractions based on coefficients derived from historical measurements that correlated BOD_5 with TOC, and its biodegradability [Servais et al., 1987, 1998, Tusseau-Vuillemin et al., 2002, Servais and Billen, 2009]. By converting these coefficients into the five distribution parameters of the organic carbon distribution model proposed in this study, it is determined that their values for b_1, t, and b_2 are well within the variation ranges, whereas the values used for s_1 and s_2 are beyond the variation ranges proposed in Chapter 2 (Table 5.2). Another difference is that the model permits the use of a range of values instead of fixed values.

Furthermore, Wang et al. [2019] conducted an extensive calibration exercise by performing 2970 direct simulations (without data assimilation) to calibrate these six fractions for the Seine, Marne,

Table 5.2: Comparison of the organic carbon repartition model parameters used in different theses

t	b_1	s_1	b_2	s_2	References
0.8	0.3	0.3	0.3	0.3	[Vilmin, 2014]
0.8	**0.475**	**0.189**	0.3	0.3	[Wang et al., 2019]
0.1–0.9	0.05–0.54	0.3–0.95	0.13–0.52	0.44–0.95	Chapter 2; [Hasanyar et al., 2023a]

t = DOC:TOC, share of the dissolved part of TOC.
b_1 = BDOC:DOC, share of the biodegradable part of DOC.
b_2 = BPOC:POC, share of the biodegradable part of POC.
s_1 = $BDOC_{rapid}$:BDOC, share of the rapidly biodegradable part of BDOC.
s_2 = $BPOC_{rapid}$:BPOC, share of the rapidly biodegradable part of BPOC.

and Oise rivers, which would improve DO simulations. This resulted in an increase in the proportion of BDOC (b1) and a decrease in that of $DOC_{refractory}$. Rather than following such a computationally intensive strategy for each simulations, we decided to include the biodegradable dissolved organic carbon as a parameter of ProSe-PA. This study therefore integrated the organic carbon repartition model (Fig. 4.1) into ProSe-PA, introducing the 5 parameters as a new species named *macrospecies* (Fig. 5.3):

The *macrospecies* presents a marked divergence in input processing methodology. Rather than requiring the six small fractions as user inputs, the model takes the total organic carbon (TOC) concentration of each carbon source as direct input and utilizes the defined values of the repartition model parameters to transform it into the six small fractions. Moreover, the inclusion of b_1 parameter among the model parameters considered for data assimilation enables the characterization of the evolution of carbon biodegradability, leading to enhanced simulation outcomes.

These advancements confer two distinct capacities for the integration of organic carbon. Firstly, the *macrospecies* may be employed directly to alter the fractions of various carbon sources and assess their influence on river metabolism. Secondly, the oxygen data assimilation technique can be utilized to estimate the biodegradability evolution of the major carbon sources.

```
macrospecies = {
TOC = {
   total_organic_carbon
   threshold = 0.7

    share_mod = {
   biodegradable = {  val = 0.30
                      range = { 0.1 , 0.5 }
                   }
   fast_biodegradable = { val = 0.7 }
   }
   share_mop = {
   biodegradable = { val = 0.3
                   }
   fast_biodegradable = { val = 0.7 }
   }
 }
}
```

Figure 5.3: The integration of *macrospecies* in the source code of ProSe-PA. `TOC` is the reading of the total organic carbon of any source. `threshold` is the reading of share of DOC in TOC (t parameter in the model). `share_mod` and `share_mop` reads the biodegradability shares of DOC and POC. `biodegradable` which is composed of `val` and `range` reads the average value and variation range of the biodegradability of DOC and POC (b_1 and b_2). `fast_biodegradable` is the reading of the share of rapidly biodegradable fractions of DOC and POC (s_1 and s_2). Only the parameter with a range is included in data assimilation framework.

4 Data assimilation strategy for low-flow period

Within this section, we initially select the model parameters to be considered for data assimilation. Then, we delineate the major sources of organic carbon that must be taken into account for data assimilation during the low-flow period of 2011. Subsequently, we design data assimilation experiments to demonstrate how the incorporation of the organic carbon repartition model enhances the estimation of river metabolism and characterizes dissolved organic carbon biodegradability during low-flow conditions.

4.1 Parameters considered for data assimilation

Similar to chapter 4, this study utilizes three key parameters for data assimilation, namely the maximum growth rate ($\mu_{max,hb}$) and growth yield (Y_{hb}) of bacteria, as well as the proportion of biodegradable dissolved organic carbon (b_1) present in the input boundary conditions. These parameters are presented in Table 5.3, along with their respective variation ranges. The reader is referred to Hasanyar et al. [2023a] for further information on the organic carbon partitioning parameters and their variation ranges.

Table 5.3: List of model parameters considered for data assimilation and their corresponding variation ranges

Parameter	Description	Min.	Max.	Unit
b_1	ratio between BDOC and DOC	0.1	0.5	[-]
$\mu_{max,hb}$	maximum growth rate of bacteria	0.01	0.13	[h^{-1}]
Y_{hb}	bacterial growth yield	0.03	0.5	[-]

4.2 Major carbon sources amongst boundary conditions

Upon investigation of the total organic carbon fluxes of all model boundary conditions, which include 7 tributary rivers, 5 WWTPs, and 156 CSOs, only seven major carbon sources are identified during the low-flow period of 2011 (Fig. 5.4). In agreement with Vilmin et al. [2016], these sources accounted for 92 % of the total carbon flux during this period. The Seine Aval WWTP and Seine River were found to be the most significant carbon contributors, accounting for 32 % and 27.4 %, respectively. The Oise River and Marne River were of moderate size contributors, contributing 12.9 % and 9.7 %, respectively. Meanwhile, Seine Amont WWTP (4.4 %), Seine Centre WWTP (3.6 %), and Garge-Epinay (2.9 %, old river) had smaller shares in comparison to other carbon sources.

In the context of data assimilation, it is crucial to determine the biodegradability (b_1) of the identified major carbon sources. In this regard, it is important to note that WWTPs have higher biodegradability and a greater proportion of dissolved and particulate carbon compared to tributary rivers [Servais and Garnier, 1993]. The value of b_1 is therefore different for the two types of carbon sources. To address this issue, the inference of b_1 with DA is achieved for permanent rivers, while for WWTPs, the first capability of the organic carbon model, namely the use of *macrospecies* is employed to fix the biodegradable DOC inflow from WWTPs at 50%. The value was selected in accordance with Servais and Garnier [1993] as the biodegradability of DOC from Seine Aval WWTP outflows was experimentally measured to be between 51–61%. Furthermore, other experimental studies have

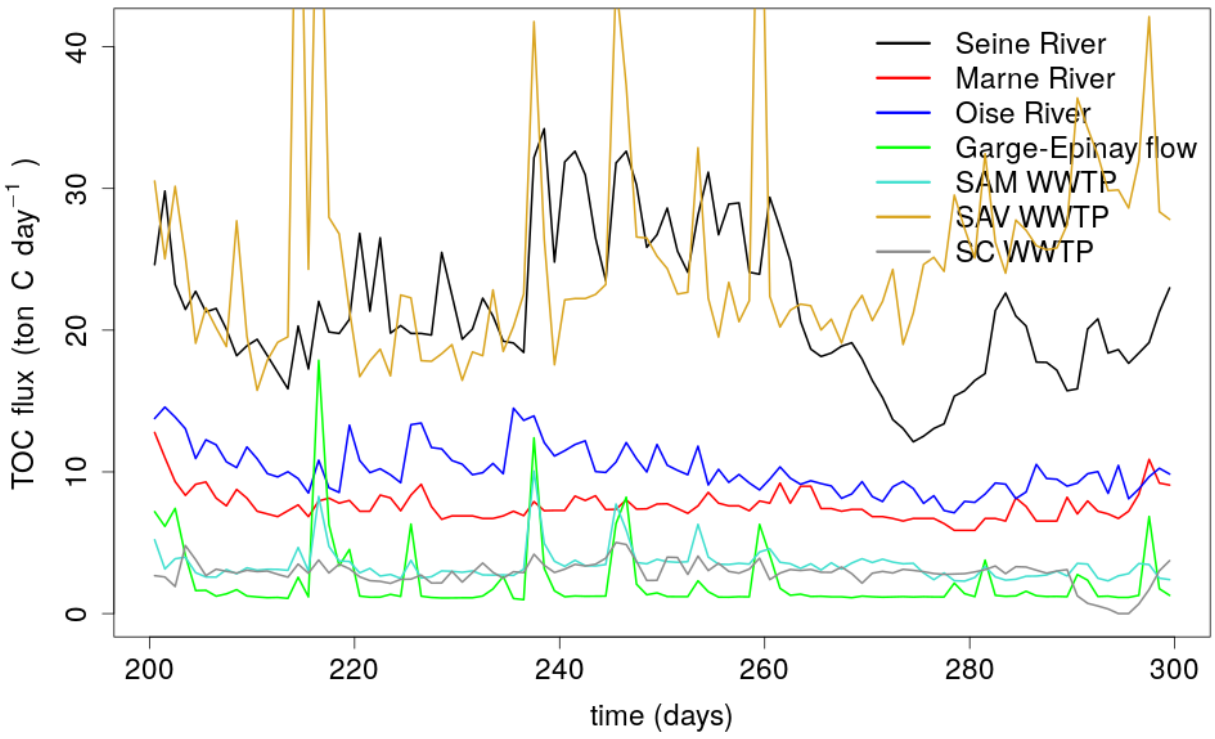


Figure 5.4: The fluxes of total organic carbon from seven major boundary conditions during the low-flow period of 2011 (in ton C day^{-1}). SAM: Seine Amont WWTP. SAV: Seine Aval WWTP. SC: Seine Centre WWTP.

shown a substantial increase in the biodegradability of DOC in the Seine River downstream of major treatment plants. Varrault et al. [2016] reported an average of 60% during the low-flow periods of 2011-2014, with a peak of 75% and 85% in the Marne and upstream stretches of the Seine River, respectively. Garnier et al. [2001] had found a range of 40-50% at the downstream of Seine Aval WWTP. Although these high biodegradabilities might be attributed to phytoplanktonic biomasses, given that our study period is sufficiently post-bloom, we can infer that these high values stem from outflows of the treatment plants.

4.3 Design of data assimilation experiments

In this study, we examine how the biodegradable fraction of dissolved organic carbon affects model parameter characterization and the simulation of dissolved oxygen concentrations during low-flow period of 2011. Additionally, we evaluate our ability to estimate the biodegradability of organic carbon sources introduced by model boundary conditions through oxygen data assimilation.

In addition, we aim to adapt the experimental setup to account for the impact of river metabolism upstream and downstream of Seine Aval WWTP. This is because the upstream and downstream stretches of the WWTP exhibit different metabolic characteristics during low-flow, with the down-

stream stretch experiencing higher rates of respiration than the upstream stretch [Vilmin et al., 2016]. Moreover, the stretch downstream of Seine Aval WWTP is dominated by large bacteria, as a result of urban outflows, while the upstream sections mainly contain small bacteria [Garnier et al., 1992b,a, Servais and Garnier, 1993]. Large bacteria have been found to possess higher growth and mortality rates, while small bacteria exhibit the opposite trend.

To assess the efficiency of the new developments in ProSe-PA, we will conduct three distinct DA experiments.

- The first experiment will serve as a baseline scenario where we do not characterize the biodegradability of any organic carbon source, but rely on the carbon input as in thesis of Vilmin [2014]. This scenario will have only two parameters for DA: maximum bacterial growth rate and growth yield.

- In the second experiment, we will infer the biodegradability of dissolved organic carbon for the four major rivers by adding b_1 as the third parameter. This will help us investigate the role of biodegradability in DO simulation and model parameter characterization. At this stage, we presume that the four rivers possess similar biodegradabilities, owing to their shared climate and land cover, as well as exposure to similar anthropogenic influences; therefore, a solitary parameter will be inferred for them. Adding a separate parameter for each river will add to the non-linearity of the model.

- Lastly, we will examine the effect of differences in river metabolism at the upstream and downstream sides of Seine Aval WWTP using not only one but two bacteria communities. For this purpose, we will run a DA simulation with organic carbon biodegradability characterization similar to the previous experiment. The difference is that all tributary rivers will be considered to contain only small bacteria, while WWTPs and CSOs will be assumed to carry large bacteria. In this context, the physiological parameters of small bacteria are to be estimated by data assimilation, based on the ranges suggested in Table 5.3, while those of large bacteria are held constant, with their maximum growth rate fixed at 0.15 h^{-1}, mortality rate at 0.05 h^{-1}, and their growth yield at 0.25 [Servais and Garnier, 1993, Billen et al., 1988, 1990].

4.4 Numerical settings

The numerical settings of these experiments are selected according to the optimum data assimilation settings configured in the synthetic case study (Chapter 4). The assimilation time step and the

observation error are set to 1 hour and 5%, respectively. The perturbation percentages are 4% for bacteria maximum growth rate and growth yield, and 8% for b_1. The number of particles is set to 3000 with a resampling threshold α of 0.3. To evaluate the performance of each experiment, the root mean square error (RMSE) is employed by comparing the simulated DO against the observed DO at the the eight monitoring stations.

The computations are performed using 40 processors (Intel(R) Xeon(R) Gold 6248R with a frequency of 3.00GHz) which takes 11 hours for the 100 days DA simulation of the low-flow period.

5 Results and Discussion

5.1 Improved simulations of DO concentration thanks to the inference of the biodegradability of dissolved organic carbon

The ensemble weighted average dissolved oxygen (DO) concentrations of the baseline scenario and the case with inference of the biodegradability of dissolved organic carbon are calculated at the eight monitoring stations and subsequently compared with the corresponding DO observations (Fig. 5.5 and 5.6). Overall, significant improvements are observed at all stations except for Colombes. Both simulations produce acceptable results during the first 50 days with minimal discrepancies from the observed DO. However, beyond day 250, the baseline DA scenario exhibited significant mismatches with the observed DO, whereas DA with biodegradability inference showed remarkable performance, except for the Colombes station during days 250-300. The DO observations at the Meulan station showed a close-to-perfect match with the simulation of biodegradability inference scenario.

The improvements due to the DA with inference of biodegradability is confirmed by calculating the RMSE of both scenarios for the period of 250-300 (Table 5.4). The results indicate that DA with biodegradability inference significantly reduces the RMSE at six out of the eight stations. The most substantial decrease in RMSE is observed at the Meulan station, where the RMSE drops from 0.72 to 0.41 mgO_2L^{-1}, representing a 43% reduction. In terms of absolute reduction in RMSE, the most significant decrease is noticed at the Mericourt station, where the RMSE decreases by 0.48 mgO_2L^{-1} (from 1.19 to 0.71 mgO_2L^{-1}, representing a 40% reduction). These results confirm the hypothesis of Wang et al. [2022a] who had considered the lack of BDOC from boundary conditions as a possible cause of overestimation of DO during low flow periods.

Conversely, increases in RMSE (34% and 7%) are noticed at only two stations: Colombes and Bougival. The increase in RMSE at Bougival is almost negligible being about 0.03 mgO_2L^{-1}. However,

Table 5.4: Performances of the baseline DA scenario and DA with inference of biodegradability evaluated by RMSE.

Stations	Baseline DA	DA with biodeg.	Reduction in RMSE (%)
Suresnes	0.66	0.58	-12
Colombes	0.67	0.9	34
Chatou	1.62	1.31	-19
Bougival	0.45	0.48	7
Sartrouville	1.3	1.02	-22
Andresy	0.52	0.43	-17
Meulan	0.72	0.41	-43
Mericourt	1.19	0.71	-40

at Colombes, the increase in RMSE could be due to uncertainty in organic carbon from the numerous CSOs in the upstream part of the modelled stretches.

Moreover, DA with biodegradability inference performs best during the dry period (days 260–280), which may be due to the fact that river metabolism is less influenced by the outflows of CSOs during such periods. This allows carbon from major sources to have a greater impact on DO evolution, enabling ProSe-PA to estimate biodegradability more accurately and improve simulations. Out of the dry period, the performance is reduced which translates into the need for incorporating the biodegradability of CSOs and WWTPs as well in addition to that of tributary rivers.

In addition, compared to the baseline case, certain drops in DO due to peak fluxes of organic carbon from dominant sources (Fig. 5.4) are also relatively better simulated (see for example Chatou station around days 294 and 300, Meulan station around days 218, 248, and 262, and Bougival and Sartrouville stations around day 294 but with a lag, that may indicate that the estimation of the net growth rate of the bacteria community could be better estimated.

5.2 Characterization of the evolution of model parameters

The model performance is also evaluated by characterizing the temporal evolution of the posterior pdfs of model parameters, based on their importance weights (Fig. 5.7), for the two DA scenarios. In the baseline scenario, only the evolution of bacteria physiology parameters is estimated (Fig. 5.7A), whereas in the inference of biodegradability scenario, that of b_1 is also estimated by the particle filter (Fig. 5.7B).

Prior to day 235, the bacteria physiology parameters evolve similarly in both simulations, which accounts for the comparable DO simulations and small deviations from the observed DO. Thus, we focus on the period between days 250-290 where mismatches were noticed between the baseline DA

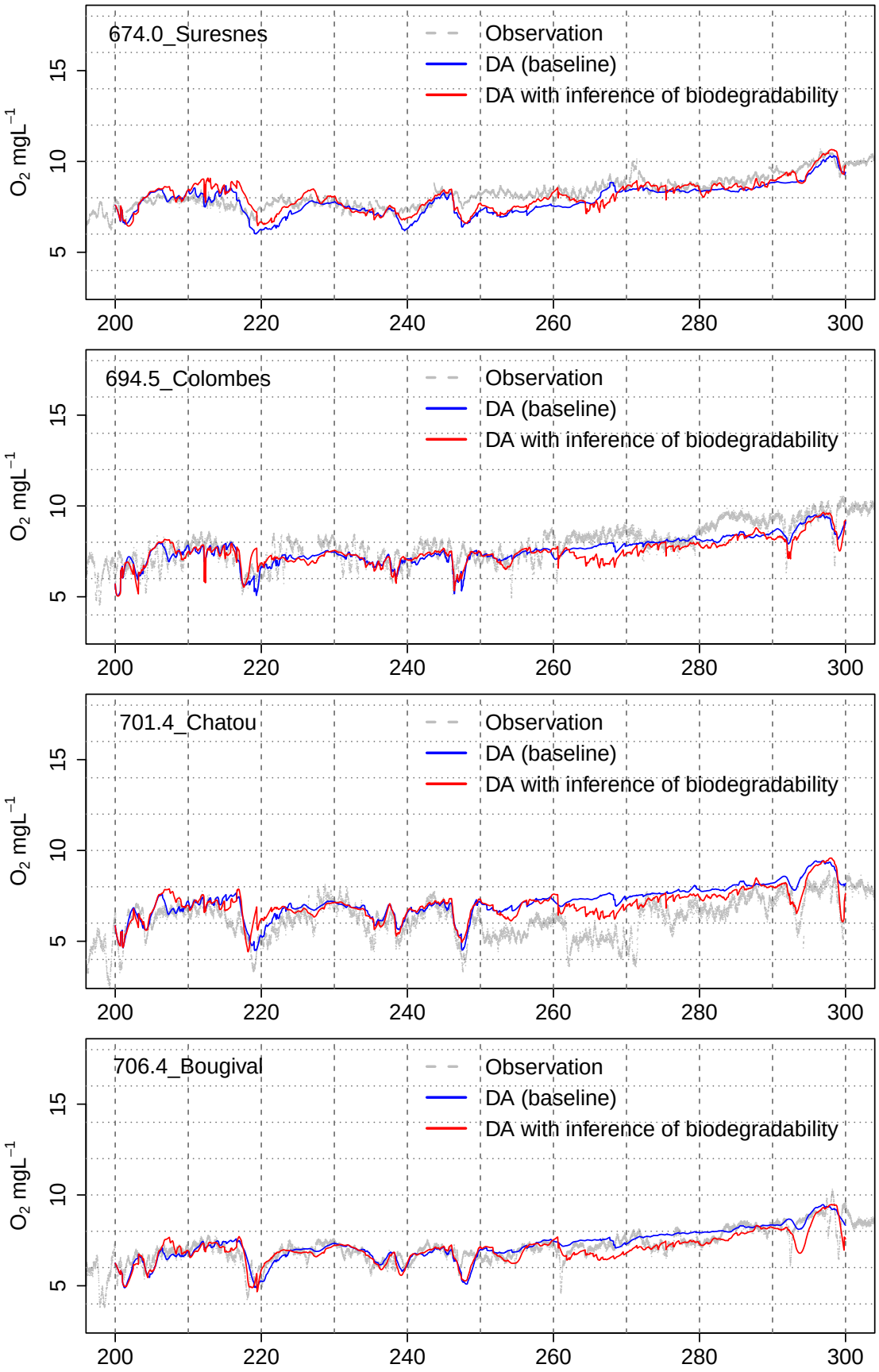


Figure 5.5: Ensemble weighted average DO concentrations at Suresnes, Colombes, Chatou and Bougival stations simulated by ProSe-PA vs. observed DO (grey line). The baseline simulation is denoted in red and the simulation incorporating the biodegradability of DOC in blue.

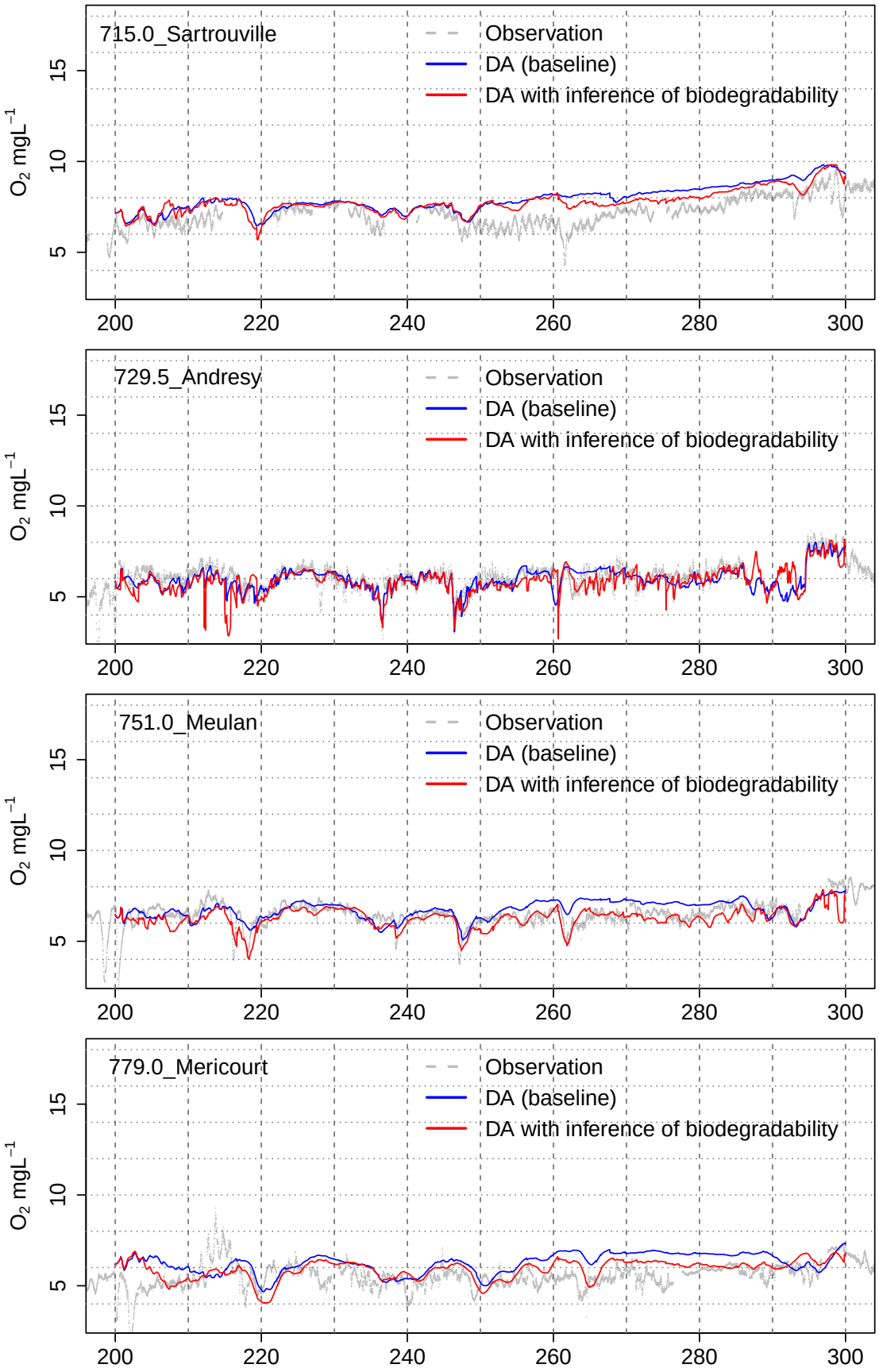
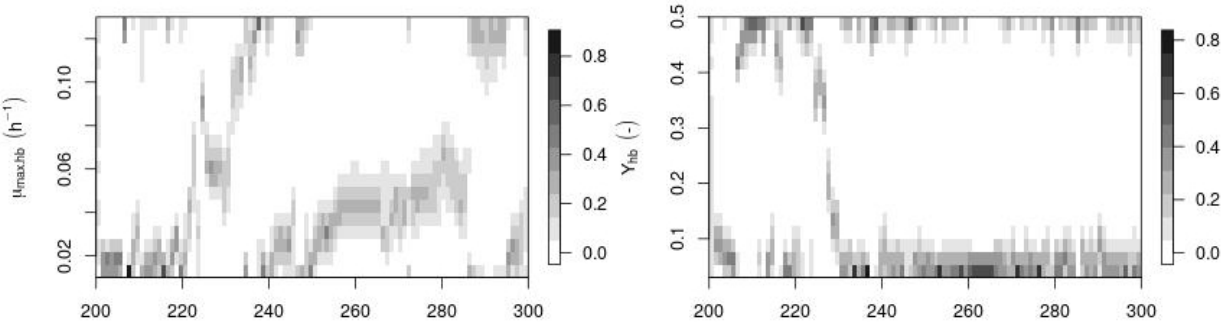
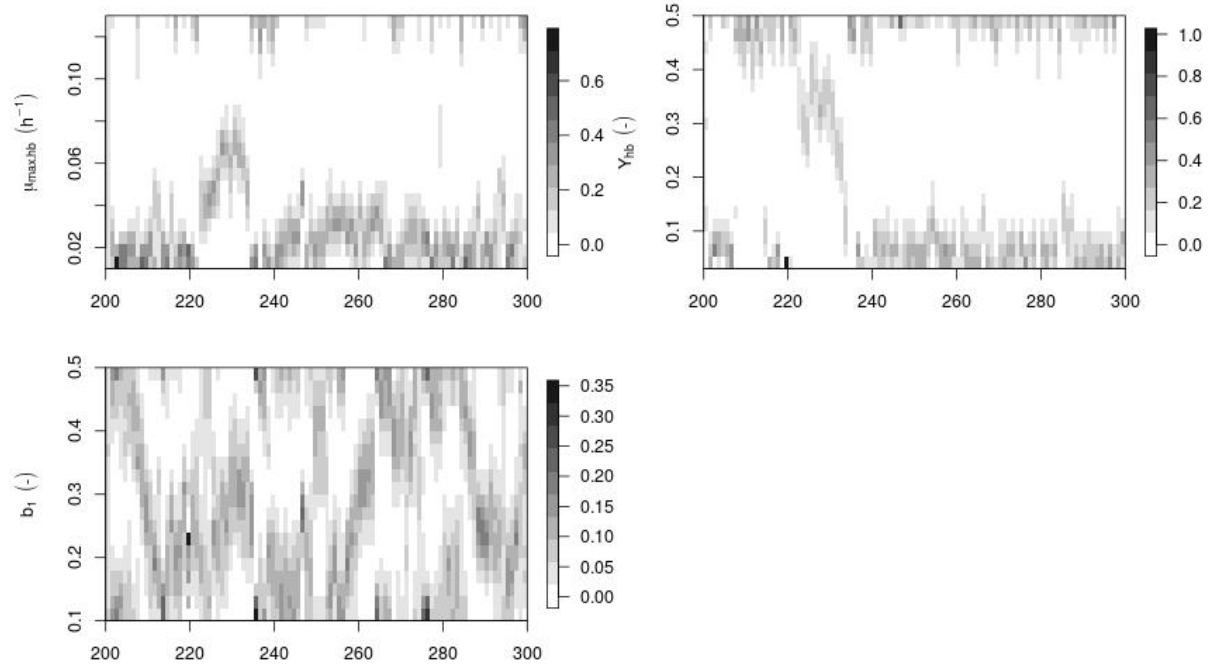


Figure 5.6: Ensemble weighted average DO concentrations at Sartrouville, Andresy, Meulan and Mericourt stations simulated by ProSe-PA vs. observed DO (grey line). Baseline simulation is denoted in red and the simulation incorporating the biodegradability of DOC in blue.



(a) Baseline DA



(b) DA with inference of biodegradability

Figure 5.7: Normalised importance weights of model parameters. $\mu_{max,hb}$: maximum growth rate of bacteria. Y_{hb}: growth yield of bacteria. b_1: the parameter representing the biodegradability of dissolved organic matter in the proposed organic matter repartition model.

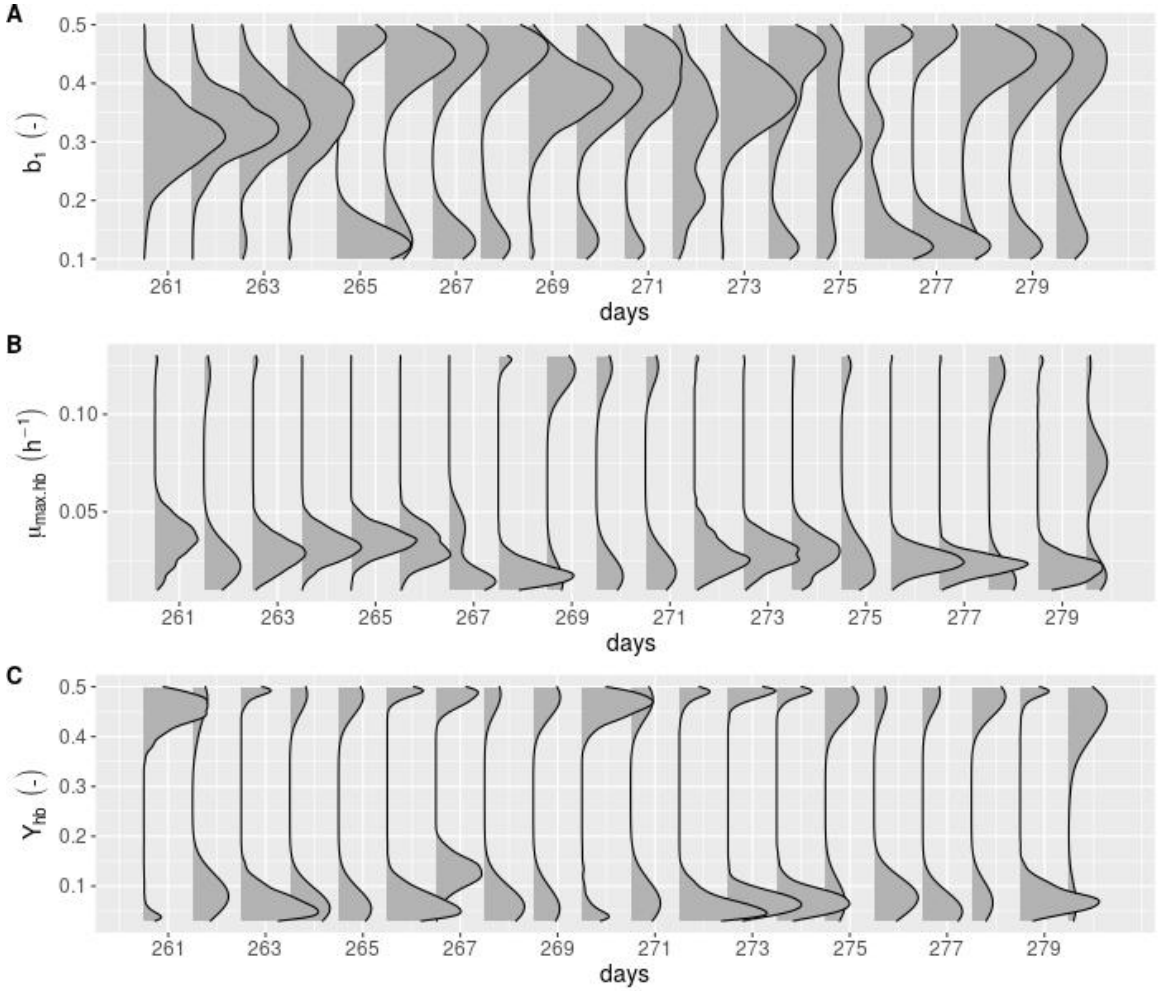


Figure 5.8: Posterior pdfs of model parameters of the DA with inference of biodegradability during dry period. $\mu_{max,hb}$: maximum growth rate of bacteria. Y_{hb}: growth yield of bacteria. b_1: the parameter representing the biodegradability of dissolved organic matter in the proposed organic matter repartition model.

simulation and observed DO (Fig. 5.5 and 5.6). The maximum bacteria growth is stable during this period and is well-identified in both DA cases, with an estimated range of 0.04–0.06 h^{-1} in the baseline case and 0.02–0.04 h^{-1} in the inference of biodegradability case. Furthermore, bacteria growth yield also remains stable in both cases and depict values less than 0.1.

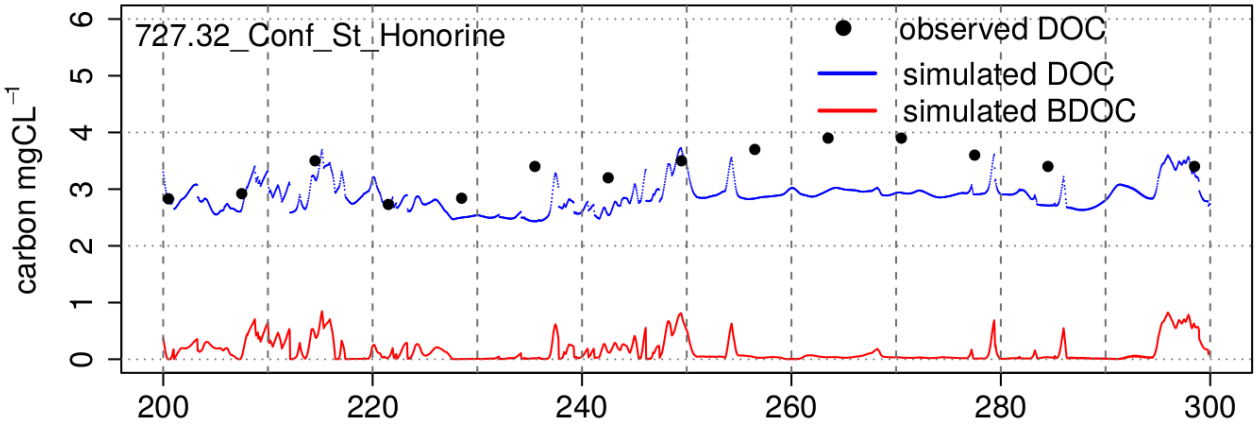
Regarding b_1, it increases from day 260 towards its upper limit until the end of the dry period (day 280) during which ProSe-PA infers the highest values of b_1=0.45. These values are consistent with the findings of Varrault et al. [2016] for the Seine river during low-flow of 2011. The increase in b_1 signifies a rise in the proportion of biodegradable dissolved organic carbon, resulting in a significant improvement in simulated DO at several stations, including Chatou, Meulan, and Mericourt (Fig. 5.5 and 5.6). It also confirms the hypothesis of Wang et al. [2022a] that attributes the overestimation of DO levels during low flow periods in Seine River to the uncertainty in BDOC. The incorporation of the parameter estimating the biodegradable fraction of dissolved organic carbon has not only facilitated the estimation of biodegradability but also improved the accuracy of DO simulations.

Focusing specifically on the dry period (days 260–280), during which the best performance of DA with the inference of biodegradability occurs, all three parameters in particular b_1, are well identified (Figure 5.8). The bi-modal nature of b_1 observed during days 240–260 and 280–290 can be attributed to two factors. First, DO is underestimated at the two upstream stations (Suresnes and Colombes) which require lower biodegradability values for Seine and Marne rivers to decrease oxygen consumption. Second, DO is overestimated at the remaining six stations, which necessitate a higher input of biodegradable DOC from Seine, Marne, Garge-Epinay and Oise rivers to reduce overestimations (this will be confirmed in the DA with 2 bacteria community). Another explanation for the bi-modal shape of b_1 posterior pdf is the perturbation percentage and the "looped" resampling method used in this work. We configured in section 3.3 a higher perturbation percentage for b_1 (8%) compared to $\mu_{max,hb}$ and Y_{hb} (4%), thus when the particle filter resamples parameters to diversify the particle distribution, the amount of perturbation for b_1 is larger as it searches within 8% of its variation range. Then, as we use looped system, the moment the resampled value falls out of the variation range, its corrected to a value on the opposite limit.

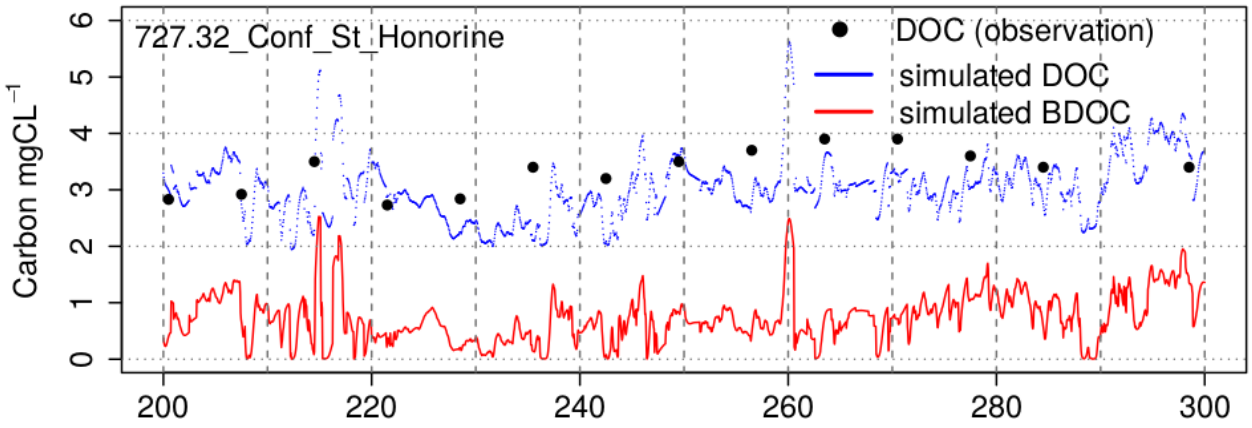
Furthermore, the analysis of TOC fluxes from the major carbon sources during 2011 (Fig.5.4) indicates a decrease in the contribution of Seine River during dry period (days 260-280), which in turn reduces the system's reliance on carbon from the Seine River. As a result, b_1 could be inferred only for the Garges-Epinay and Oise rivers, leading to a better identification of b_1 and improvement in simulated DO. Interestingly, b_1 is once again well estimated to be low (0.25) for days 290-300 when DO

levels increase in all stations, indicating lower demand for BDOC (Figure 5.7B). These results confirm the findings of our sensitivity analysis [Hasanyar et al., 2023b] that depicted b_1 as being influential on DO evolution at low flow and whose evolution could be therefore estimated by data assimilation.

5.3 BDOC concentration increases in DA with inference of biodegradability



(a) Baseline DA



(b) DA with inference of biodegradability

Figure 5.9: Simulated BDOC and simulated DOC versus the observed DOC at St. Honorine station.

Both DA scenarios simulate similar DOC levels and perform equally compared to the observed DOC as shown here at St. Honorine station (Fig. 5.9). However, the DA with inference of biodegradability leads to higher BDOC concentrations at most of the stations. This increase in BDOC results in improved oxygen simulations while inferring with the parameter of biodegradability. However, the BDOC values shall be compared with those obtained in Varrault et al. [2016]. If an underestimation is observed, that will mean that more BDOC needs to enter the river from other sources such as CSOs and WWTPs for which a new parameter of biodegradability is required.

Table 5.5: Performances of the DA with two bacteria communities compared to DA with a single bacteria community evaluated by RMSE.

Stations	DA with 2 bact.	DA with 1 bact.	Reduction in RMSE (%)
Suresnes	0.52	0.58	-12
Colombes	0.57	0.9	-58
Chatou	1.75	1.31	25
Bougival	0.54	0.48	11
Sartrouville	1.41	1.02	28
Andresy	1.2	0.43	64
Meulan	0.57	0.41	28
Mericourt	1.08	0.71	34

5.4 Data assimilation with two bacteria communities lead to stable characterization of model parameters

As previously discussed, the Seine River exhibits distinct metabolic behaviors between its upstream and downstream regions in relation to the overestimation of DO at 2 upstream stations and its underestimation at the remaining 6 stations [Vilmin et al., 2016]. Moreover, Njapou [2022] has calculated the indicators of river metabolism in the Seine River and has found that, during the low-flow period of 2011, both Suresnes and Andresy experience net heterotrophy. However, there is a notable increase in the rate of ecosystem respiration as one moves downstream towards Andresy, coinciding with a corresponding increase in the number of urban outflows.

To address the spatial difference in metabolism, one could contemplate manipulating the ratio of contributions from large and small bacteria for each model boundary condition. The variation in metabolism may indeed stem from differences in the respiration rates of small and large bacteria. The previous two simulations were conducted using a single bacteria community. In the present investigation, we consider two bacterial communities - large and small - with a concentration ratio of 0:1 in tributary rivers (Seine, Marne, and Oise), and 1:0 in CSOs and WWTPs. Although ProSe-PA is capable of estimating the evolution of physiological parameters of a single community, we have decided to infer the model parameters by DA for small bacteria and constrain those of large bacteria, as explicated in section 4.3.

In this study, DA is carried out using two bacteria communities and compared with the observed oxygen and the simulated DO from the DA utilizing only one bacterial community (i.e., the previous case of biodegradability inference) (Fig. 5.10 and 5.11).

The simulations for DA with two bacterial communities reduced RMSEs at two rather upstream stations, namely Suresnes (11%) and Colombes (57%) during the period of days 250-300, in comparison

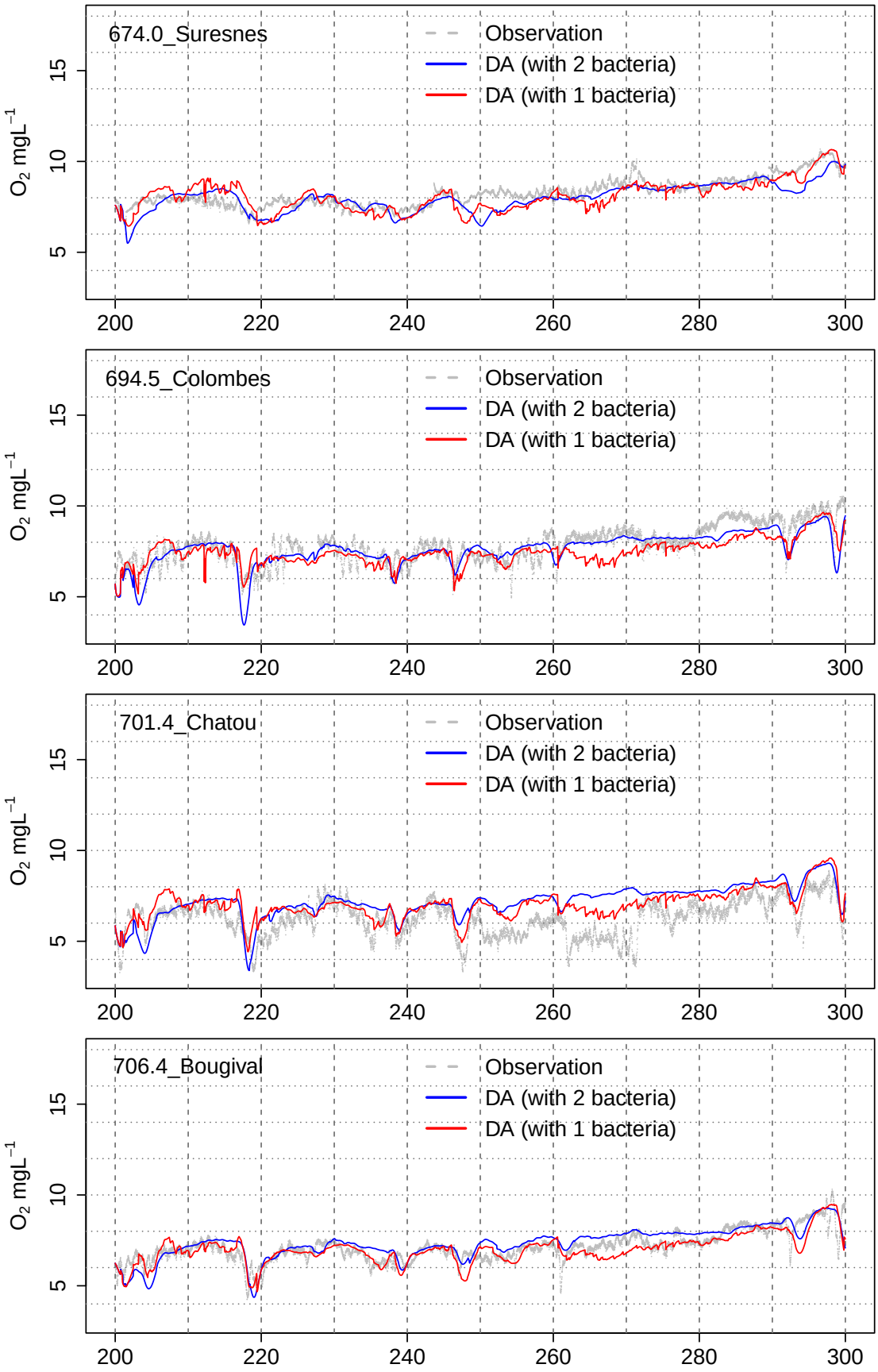


Figure 5.10: Ensemble weighted average DO concentrations at Suresnes, Colombes, Chatou and Bougival stations simulated by ProSe-PA vs. observed DO (grey line). The simulation with two bacteria communities is denoted in blue and the simulation with one bacteria is in blue.

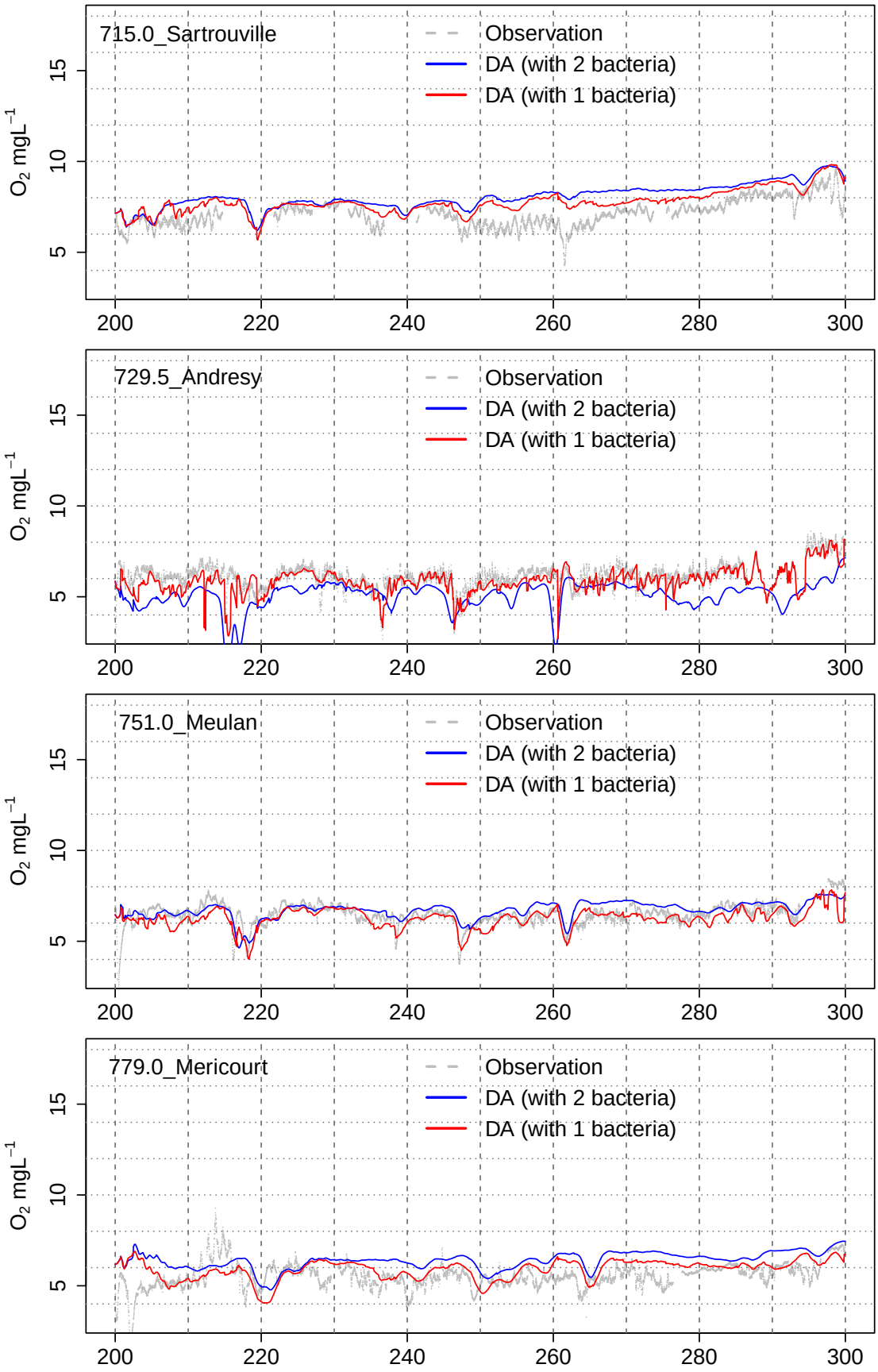


Figure 5.11: Ensemble weighted average DO concentrations at Sartrouville, Andresy, Meulan and Mericourt stations simulated by ProSe-PA vs. observed DO (grey line). The simulation with two bacteria communities is denoted in blue and the simulation with one bacteria is in blue.

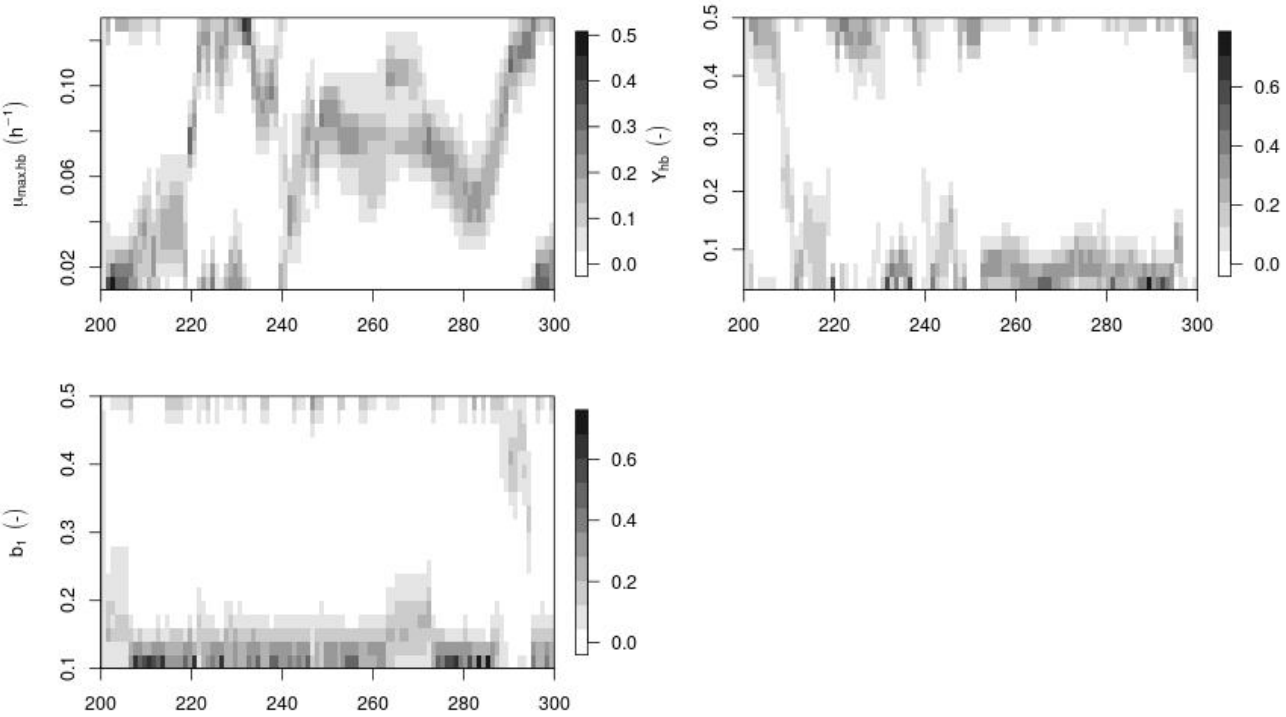


Figure 5.12: Normalised importance weights of model parameters in DA simulation with 2 bacteria communities.

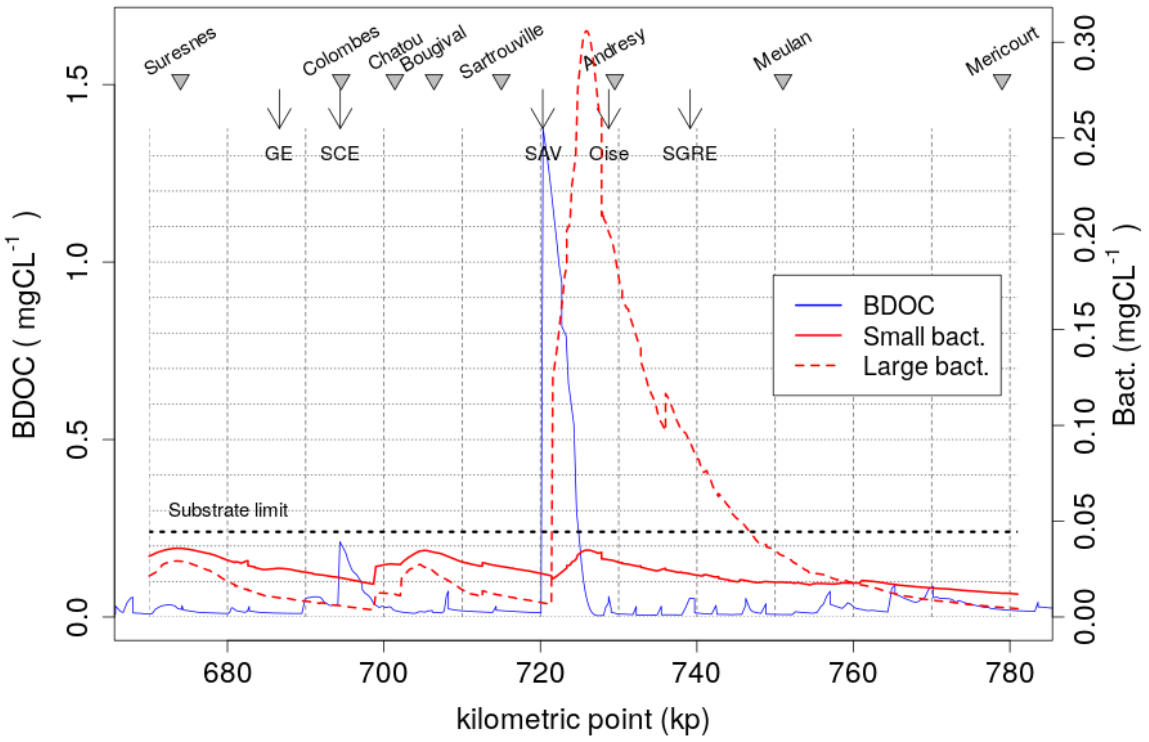


Figure 5.13: Longitudinal profile of mean BDOM concentration together with the mean concentrations of small and large bacteria from day 250 to 290. Only left river reaches are presented here. GE: Garge-Epinay, SCE: Seine Centre WWTP. SAV: Seine Aval WWTP. Oise: Oise river. SCE: Seine Gresilion WWTP.

to the case where only one bacterial community was employed (Table 5.5). However, higher RMSEs, indicating overestimations, are observed at the more downstream stations except for Andresy, where a significant underestimation has been noticed. This underestimation could potentially be attributed to the fact that the outflow from Seine Aval WWTP contains only large bacteria, which have higher growth rates leading to a massive consumption of BDOC and subsequent drop in oxygen levels. This situation may lead to lack of BDOC and active bacteria for the downstream stations like Meulan and Mericourt (Fig. 5.13), thereby leading to further overestimations compared to the case of a single bacterial community. Lack of BDOC and bacteria is evident in the longitudinal profile of Seine River.

However, the most significant outcome of using the two bacteria communities framework is the stability of bacteria and biodegradability parameters over time (Fig. 5.12). Specifically, the biodegradability of dissolved organic carbon has been well identified and estimated to be very low (ranging between 0.10 to 0.15) throughout the whole period, as opposed to the case of DA with only one bacterial community, which had identified higher values for b_1 (Fig. 5.7B). This confirms the hypothesis developed in chapter 4 based on which b_1 was supposed to be well detected if the amount of substrate is below a limit (0.24 $mgCL^{-1}$). The mean BDOC concentration for the period of 250-290 is observed to be well below the substrate limit (Fig. 5.13). However this range of 0.1-0.15 for biodegradability of DOC is not in accordance with the experimental data for Seine [Varrault et al., 2016]. Another explanation could be that, on one hand, we assimilate the bacteria parameters for the small bacteria, which are predominantly present in the upstream section coming from Seine, Marne, and Garge-Epinay rivers (upstream of Suresnes and Colombes station). On the other hand, DA with one bacterial community had underestimated the dissolved oxygen levels at these two stations. Therefore, as hypothesized in the previous section, ProSe-PA tries to bring down b_1 to reduce the underestimation gap at Suresnes and Colombes. Therefore, this explains why RMSEs are reduced at these two stations. This also explains the further overestimation at three stations downstream of Colombes (Chatou, Bougival, and Satrouville), as the estimated low values of b_1 lead to a lack of BDOC at these stations (Fig. 5.13) that are already overestimated by the DA with a single bacterial community. Furthermore, the underestimation of dissolved oxygen by the release of only large bacteria from Seine Aval WWTP, allows ProSe-PA to infer low b_1 values for Oise River.

In the case of DA with one bacterial community, ProSe-PA was bi-modal or shifting between the two extremes in certain periods, as the contradicting metabolic situations of the monitoring stations demanded both low and high values of b_1. This shows that different sections of the Seine River are controlled by different organic carbon sources.

5.5 Strategy for further enhancement of data assimilation simulations at low-flow

By combining the findings of DA with one and two bacteria communities and taking into account the role of biodegradability inference, it becomes apparent that although there were overall improvements in dissolved oxygen (DO) simulations, there are still mismatches at several stations. Therefore, two issues need to be addressed: firstly, the discrepancies themselves, and secondly, the metabolic nature of the discrepancies, which include underestimation at upstream stations and overestimation at downstream stations.

To solve these issues, we propose first of all to utilize the model with two bacteria communities, which can effectively account for the metabolic nature of the discrepancies. However, this requires a re-evaluation of the current concentration ratios between the large and small bacteria in each carbon source group (rivers, WWTPs, CSOs) based on an extensive review of the literature. Additionally, the fixed values utilized for the growth and yield of large bacteria must be scrutinized based on an analysis of the literature, as ProSe-PA is unable to assimilate the bacterial properties of both communities simultaneously. Therefore, one community must be fixed while the other is included in the assimilation scheme.

Subsequently, the discrepancies could be mitigated by inferring the biodegradabilities of the combined sewer overflows and wastewater treatment plants. The biodegradability of WWTPs is not presently inferred in this study as it is only rised to a fixed value of 0.5 thanks to the *macrospecies* compared to the value of 0.3 in the baseline scenario (Table 5.2), whereas that of CSOs is not considered at all. Moreover, they have higher biodegradabilities than the tributary rivers, therefore, necessitating the incorporation of a novel parameter to estimate their biodegradabilities. This is particularly crucial in accounting for the transitory aspect of the problem.

It is also important to consider the potential impact of other factors that may affect the DO concentrations in the Seine River, such as temperature and light conditions. Currently, for example two time series of these two variables are applied throughout the entire span of the model, disregarding the possible spatial variations that may exist. These factors can affect the growth and metabolism of bacteria, as well as the diffusion of oxygen in the water. Therefore, incorporating these factors into the model may help to further improve the accuracy of DO simulations.

Moreover, it is important to continue monitoring the Seine River and collecting data to update and validate the model. This can help to identify new sources of discrepancies and improve the accuracy of the model predictions. In particular, it is necessary to validate the parameter of biodegradability by comparing it with data from measurement campaigns in the Seine River.

6 Conclusion

In this study, dissolved oxygen data of the low period of 2011 is assimilated by ProSe-PA in order to evaluate the impact of the new organic carbon repartition model on DO simulations and parameter characterization in a real case scenario. For this purpose, a study area having a length of 220 km with 8 oxygen monitoring stations is selected. Then, a 100-day low-flow period is identified as a study period. Followingly, the major sources of carbon are identified whose biodegradabilities are supposed to be estimated by ProSe-PA. Three model parameters are considered for data assimilation, namely maximum bacteria growth rate, bacteria growth yield and the parameter representing the biodegradability of DOC in the proposed organic carbon repartition model. Then, three DA scenarios are considered that consist of: 1) a baseline case representing a simulation without inference of biodegradability, and two simulations with inference of biodegradability using 2) one and 3) two bacteria communities (large and small) frameworks. Based on the results, the following conclusions could be made:

- Data assimilation using a single bacteria community and in the presence of the parameter representing the biodegradability of DOC improves the DO simulation during low-flow periods compared to the baseline case (in six out of eight stations).

- ProSe-PA identifies the evolution of organic carbon biodegradability from the considered carbon sources with a posterior pdf. The best performance coincides with the dry period during which a high biodegradability in DOC of major sources is estimated. Therefore, those DO monitoring stations where DO was overestimated in the baseline scenario, experienced considerable improvements in their DO simulations with the inference of biodegradability. On the contrary, it deteriorated the simulations of those stations where DO was underestimated in the baseline scenario case.

- DA with two bacteria communities led to stable characterization of the model parameters compared to the case of one bacteria communities during which bi-modal pdfs were observed in certain periods. The evolution of bacterial parameters were inferred only for the small bacteria. The DO simulations improved only at the upstream monitoring stations where small bacteria dominated and that had experienced underestimation in the two previous assimilation scenarios.

- To further enhance DA simulations, we propose utilizing a two-bacteria communities model coupled with bibliographical re-evaluation of the ratios of small and large bacteria contribution from different groups of model boundary conditions. In addition, the macrospecies helped in

increasing the biodegradability of WWTPs from 0.3 to 0.4, but we propose independent inferring of the biodegradabilities of combined sewer overflows and wastewater treatment plants by data assimilation.

Conclusion and perspectives

Conclusions

In light of the previous theses [Vilmin, 2014, Wang et al., 2019], the purpose of this thesis was to improve the estimation of river metabolism at summer low-flow periods by addressing the main shortcomings of ProSe-PA that are uncertainties related to organic carbon from boundary conditions, role of degradation kinetics of organic carbon, and integration of distinct small and large bacteria communities in the model. For this purpose, several steps have been followed:

- A new organic repartition model consisted of five parameters was created to address the uncertainties related to organic carbon that enter the system from the model boundary conditions. This parameterization allows the user to manipulate the share of dissolved, particulate, biodegradable and refractory fractions of organic carbon for any boundary condition. In addition, the three parameters related to degradation kinetics of organic carbon, such as the constant for uptake of monomeric substrate, which already existed in the model, were also identified for subsequent sensitivity analysis together with the parameters of organic carbon repartition model.
- An extensive bibliographical review was conducted to collect information on variation ranges of the existing kinetic and the new repartition model parameters, under different conditions of hydrology, land cover and climate. This review also covered the main water quality variables of total organic carbon (TOC) and dissolved organic carbon (DOC). In addition to proposal of overall variation ranges for the considered parameters, a clustering method was also employed to identify the most important drivers of the variability of each parameter for whom narrowed variation ranges were proposed. Our analysis showed that the organic carbon concentration in a river is generally controlled by climate and land cover. In particular, urbanization is related to an increase in the biodegradability of DOC.

– A sensitivity analysis was conducted on model parameters to find the influential model parameters on dissolved oxygen evolution under low-flow conditions. The analysis showed that overall the bacterial parameters such as its growth and mortality rates control DO evolution. However, under the conditions of high bacterial net growth activity which is a property of summer low-flow periods, the biodegradable fraction of dissolved organic carbon is also influential. The model was not sensitive enough to the parameters of organic carbon degradation, therefore, the corresponding assumed uncertainties could be ignored.

– Observing the utility of the organic carbon repartition model and the role of its parameter of biodegradability on oxygen at low flow, it was integrated in ProSe-PA. In addition, the parameter of biodegradability was also in the data assimilation framework of ProSe-PA. Then, the performance of ProSe-PA was evaluated on a synthetic case and it was validated that ProSe-PA is capable of quantifying the biodegradability of DOC based on oxygen data assimilation under fixed and varying conditions of carbon input from boundary conditions. This inference of biodegradability contributed in accurate simulation of DO compared to the case where biodegradability was not inferred. This study also led to an optimum configuration of data assimilation settings for better identification of model parameters.

– Finally, the performance of ProSe-PA and its new repartition model was evaluated in a real case scenario during the low-flow period of 2011. It was shown that oxygen data assimilation with inference of the biodegradability of dissolved organic carbon improved DO simulations in majority of the DO monitoring stations. In addition, the temporal evolution of the parameter of biodegradability and that of bacterial growth was characterized.

– Data assimilation with small and large bacteria communities was also integrated in ProSe-PA with a test which illustrated more stability in the characterization of model parameters compared to the case with a single bacteria community. However, it led to enhancement of simulations only in a few stations which shows the need for further investigation on the share of the two bacterial communities from boundary conditions.

Perspectives and recommendations

Despite the remarkable improvements during a low-flow period, there is still room for more enhancements in ProSe-PA.

– First, for the DA with a single bacteria community, as the particle filter infers values towards the upper limit of b_1, we need to explore high ratio of b_1. It means that its upper limit shall be increased to let the particle filter search more BDOC leading to further decrease in oxygen concentration at stations where overestimations are observed.

– The utilization of two bacteria community-based oxygen data assimilation framework with inference of biodegradability is proposed. This is also in agreement with the opposite metabolisms observed at the two upstream DO monitoring stations compared to the other downstream stations. However, this necessitates two more efforts to reduce the overestimations of river metabolism: first the ratio of large to small bacteria contribution from boundary conditions need to be investigated either by literature review or experimental campaigns. A continuous monitoring of bacterial communities is also necessary based on the results of our sensitivity analysis which depicted their control on river metabolism at low-flow.

  Next, at least one new parameter is needed to be added in ProSe-PA for inferring the biodegradability of WWTP and transient sources such as CSOs as this study only allowed inference of this parameter for major tributary rivers (Seine, Marne, Oise and Garge-Epinay). As for WWTPs, the macrospecies was used to rise their biodegradabilities from the baseline scenario of 0.3 to 0.5, but the setback is that it is a fixed value. Moreover, WWTPs and CSOs have higher biodegradabilities than tributary rivers which means that ProSe-PA cannot provide stable estimates of biodegradability if we are going to infer all major carbon sources. The transitory aspect of CSOs is important as 90% of CSOs happen in less than 10 hours, inducing fast drop in DO levels while boosting bacterial activity [Wang et al., 2022a]. Another solution would be to condition the inference of bidegradability on the type of inflow, thereby, permitting inference of each source group at a time.

  Another strategy would be to switch the roles of small and large bacteria and utilize the stable results of the DA experiment where growth parameters were inferred for small bacteria. Indeed, another DA experiment could be conducted but this time with inference for big bacteria and fixation of small bacteria dynamics using the stable values from the previous experiment.

– Experimental campaigns of BDOC measurement are also necessary for validating the estimation of biodegradability of DOC by ProSe-PA. This is important in a sense that changes in DOC and BDOC are found to induce potential shifts in bacterial communities [Hullar et al., 2006, Crump et al., 2003] such that an increase in BDOC is considered to increase the diversity of

bacterial populations [Landa et al., 2013].

– A new type of weight called *assimilation station weights* could be implemented in the particle filter of ProSe-PA to limit the inference of each organic carbon source only at its zone of influence. As the carbon from upstream sources may not reach downstream stations, we need to infer the biodegradabilities of such sources only based on those DO monitoring stations that happen to be in their zone of influence. It could be achieved by calculating a dynamic length of influence for each carbon source similar to the equation for $pk_{observability}$ as a function of TOC input from the source, flow velocity, and bacteria physiology properties.

– The parameters of phytoplankton dynamics may also be included in the DA simulations to allow detection of daily fluctuations in DO.

– Photodegradation could be included in the model as it is found in this study to be a new source of uncertainty due to its role in spatial dynamicity of biodegradability of organic carbon and bacterial activity [Hansen et al., 2016, Hu et al., 2022]. A simplistic model to account for biodegradability is proposed in Hasanyar et al. [2023a].

– Finally, new sources of high frequency data could be incorporated in the model to achieve Multi-target Data Assimilation. The purpose is to have two or more sources for data assimilation. In particular, this could be done for organic carbon data because recent researches using methods like 3D fluorescence spectroscopy and $SUVA_{254}$ [Goffin et al., 2017, Varrault et al., 2020, Parlanti et al., 2000, 2002, Hosen et al., 2014, Knapik et al., 2015, Shin et al., 2016] have been promising and are about to provide us with high frequency evolution of DOC and even BDOC in real time in the very near future. Another source of multi-target DA could be the use of pCO_2 high frequency data to help improve the estimation of river metabolism by better balancing the biotic and abiotic processes. pCO_2 will help infer the parameters of abiotic processes such as temperature, pH, solar radition, total alkalinity [Marescaux et al., 2020] and whose main driver of variability is considered to be DOC [Marescaux et al., 2018].

Moreover, the developments in this thesis such as the *macrospecies* could be used to study the impact of different scenarios of organic carbon load from the new infrastructure developments like new WWTPs and CSOs along the Seine River by running ProSe-PA simulations both with or without data assimilation. The *macrospecies* allows exploiting the shares of biodegradable, dissolved, particulate and refractory fractions of organic carbon from these urban outflows.

This thesis also opens the door to another application which is to establish a "WWTP-sanitation network-river management system based on real time data assimilation in river". Indeed, the goal is to use ProSe-PA to specify the necessary level of treatment that would allow wastewater treatment plants to discharge an input load as high as possible while respecting the EU Water Directive, to maintain the water quality for 90% of the time above a good quality threshold.

List of publications

Peer-reviewed articles

Masihullah Hasanyar, Thomas Romary, Shuaitao Wang, Nicolas Flipo. "*How much do bacterial growth properties and biodegradable dissolved organic matter control water quality at low flow?* ". Biogeosciences.
DOI: https://doi.org/10.5194/bg-20-1621-2023

Co-authored articles

Shuaitao Wang, Nicolas Flipo, Thomas Romary, Masihullah Hasanyar. "*Particle filter for high frequency oxygen data assimilation in river systems*". Environmental Modelling & Software.
DOI: https://doi.org/10.1016/j.envsoft.2022.105382

Articles in preparation

Masihullah Hasanyar, Nicolas Flipo, Shuaitao Wang, Josette Garnier, Edith Parlanti, Thomas Romary. "*A review of variables representing the description and degradation of organic matter in world rivers under different climatic, hydrological, and land cover conditions*". Earth Sciences Review.

Masihullah Hasanyar, Nicolas Flipo, Thomas Romary, Shuaitao Wang. "*Quantifying heterotrophic bacteria parameters and dissolved organic carbon biodegradability through oxygen data assimilation in a river water quality model* ". Water Resources Research.

Masihullah Hasanyar, Nicolas Flipo, Thomas Romary, Shuaitao Wang. "*Improvement of high frequency oxygen data assimilation in river systems by incorporation of dissolved organic carbon biodegradability during low-flow periods*". Not decided yet.

International conferences

Masihullah Hasanyar, Nicolas Flipo, Thomas Romary, Shuaitao Wang. "*The role of organic matter and bacterial physiology on river metabolism at low flow*". EGU General Assembly, Gather Online, 2021.

Technical reports (in French)

Masihullah Hasanyar, Nicolas Flipo, Thomas Romary, Shuaitao Wang, Anice Yari. "*Rôle de la matière organique dans le métabolisme des rivières à bas débit*". Piren-Seine Phase 8 - Rapport 2020.
DOI: https://doi.org/10.26047/PIREN.rapp.ann.2020.vol19

Masihullah Hasanyar, Nicolas Flipo, Thomas Romary, Shuaitao Wang, Anice Yari. "*Dans quelle mesure les propriétés de croissance bactérienne et la matière organique biodégradable dissoute contrôlent-elles la qualité de l'eau à bas* ". Piren-Seine Phase 8 - Rapport 2021.

Masihullah Hasanyar, Nicolas Flipo, Thomas Romary, Shuaitao Wang, Anice Yari. "*Amélioration des simulations d'oxygène par intégration de la composition de la matière organique dans un modèle de qualité de l'eau*". Piren-Seine Phase 8 - Rapport 2022.
DOI: https://doi.org/10.26047/PIREN.rapp.ann.2022.vol19

Appendices

A Supplementary results of FAMD

A1 Results of FADM for DOC, t and K_S variables

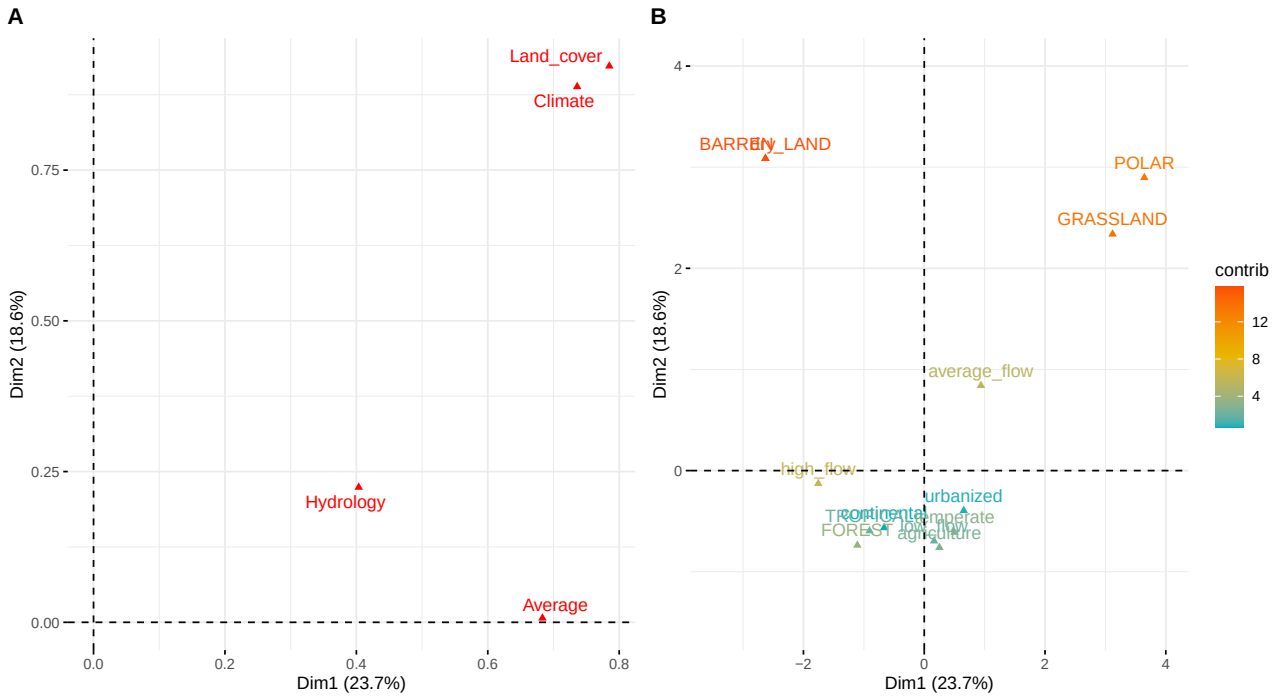


Figure A1: FADM results for t (DOC:TOC). A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters.

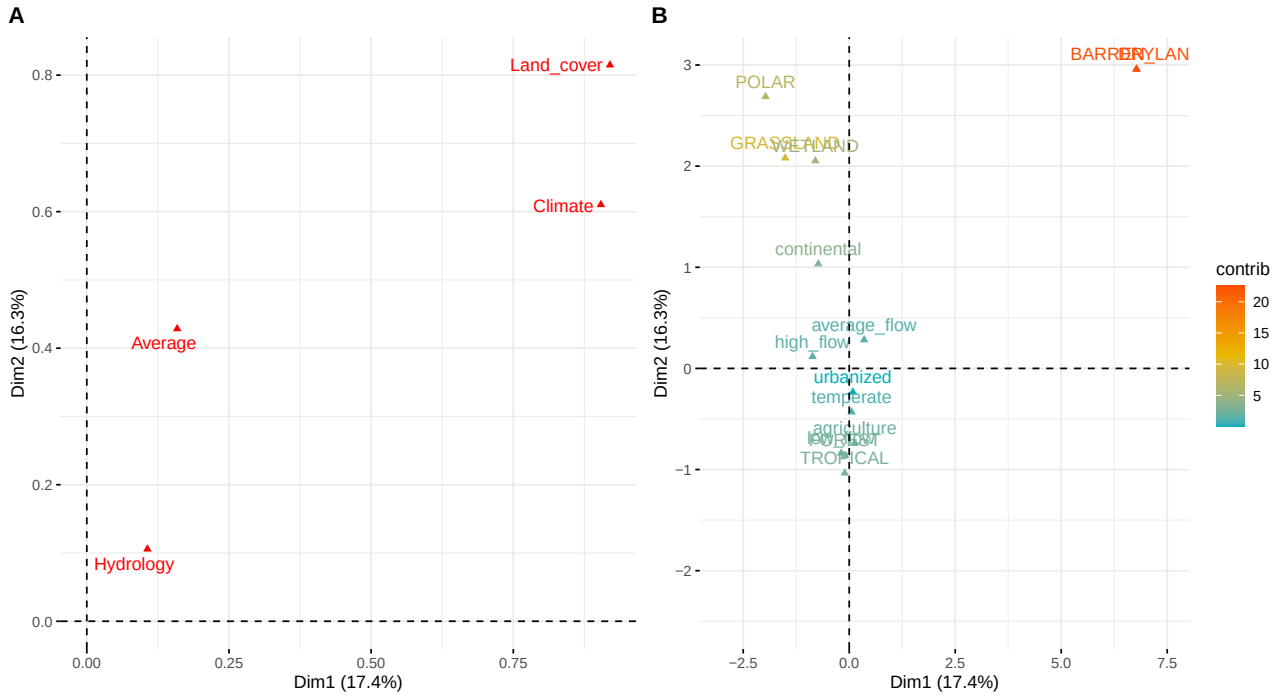


Figure A2: Initial FADM results for DOC. A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters.

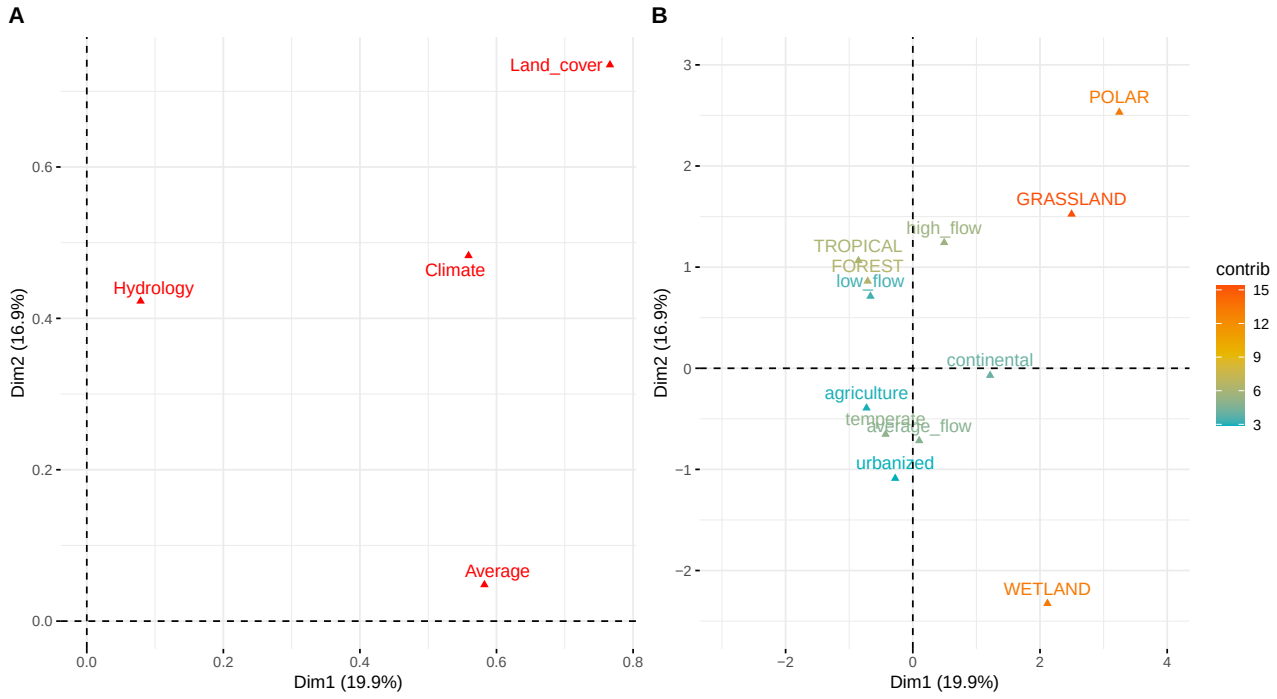


Figure A3: Refined FADM results for DOC (without dry-barren land clustering for better visualization of other quantitative variables. A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters.

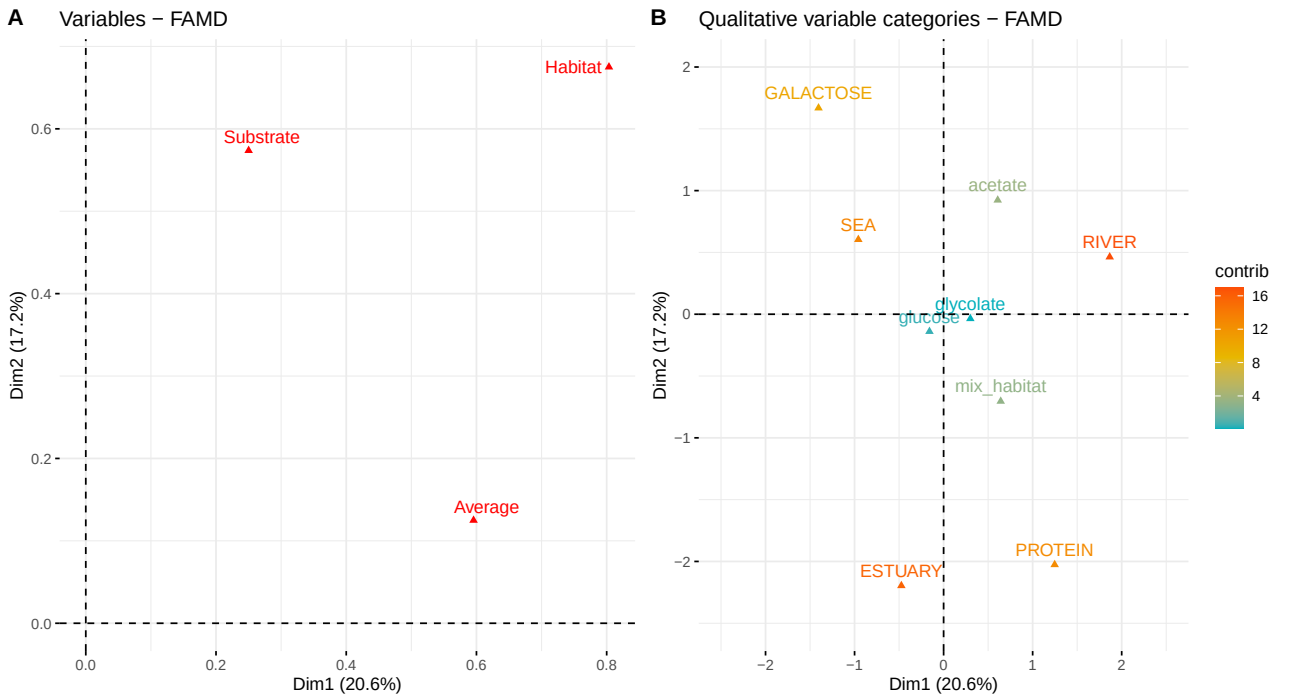


Figure A4: FADM results of the Monod half-saturation constant for uptake of substrate (K_s in mgC L^{-1}). A) Coordinates of the factors in the new space (dimensions 1 and 2). B) Plot of the categories of qualitative factors on dimensions 1 and 2 based on their contributions (in %) to the corresponding axes. Categories displayed in capital letters are the identified clusters.

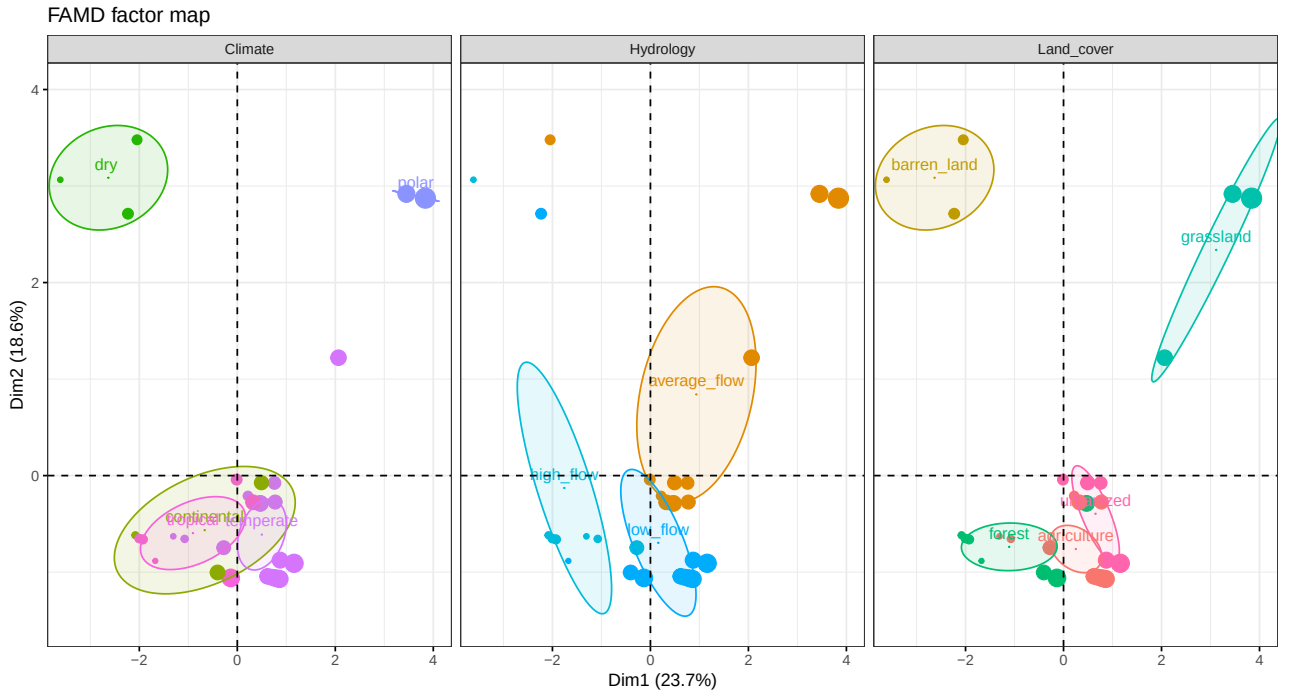


Figure A5: Distribution of the individuals corresponding to the DOC:TOC ratio, t, on dimensions 1 & 2 with the confidence ellipses

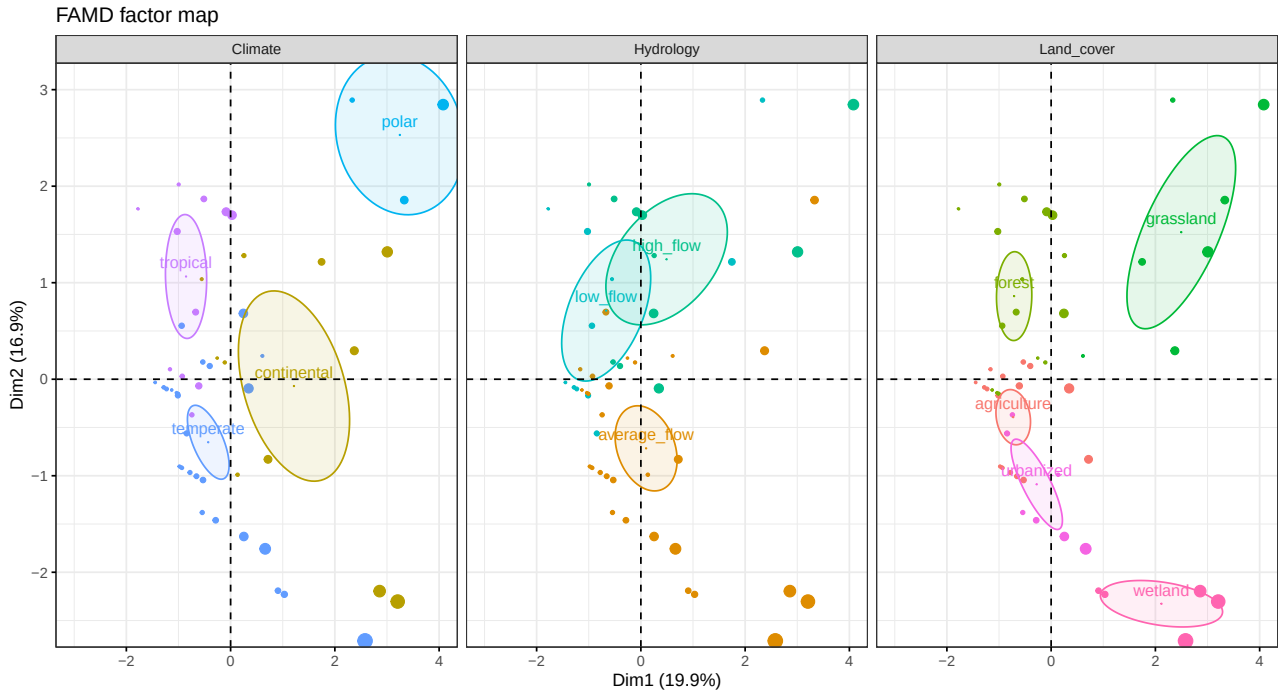


Figure A6: Distribution of the individuals corresponding to DOC variables on dimensions 1 & 2 with the confidence ellipses

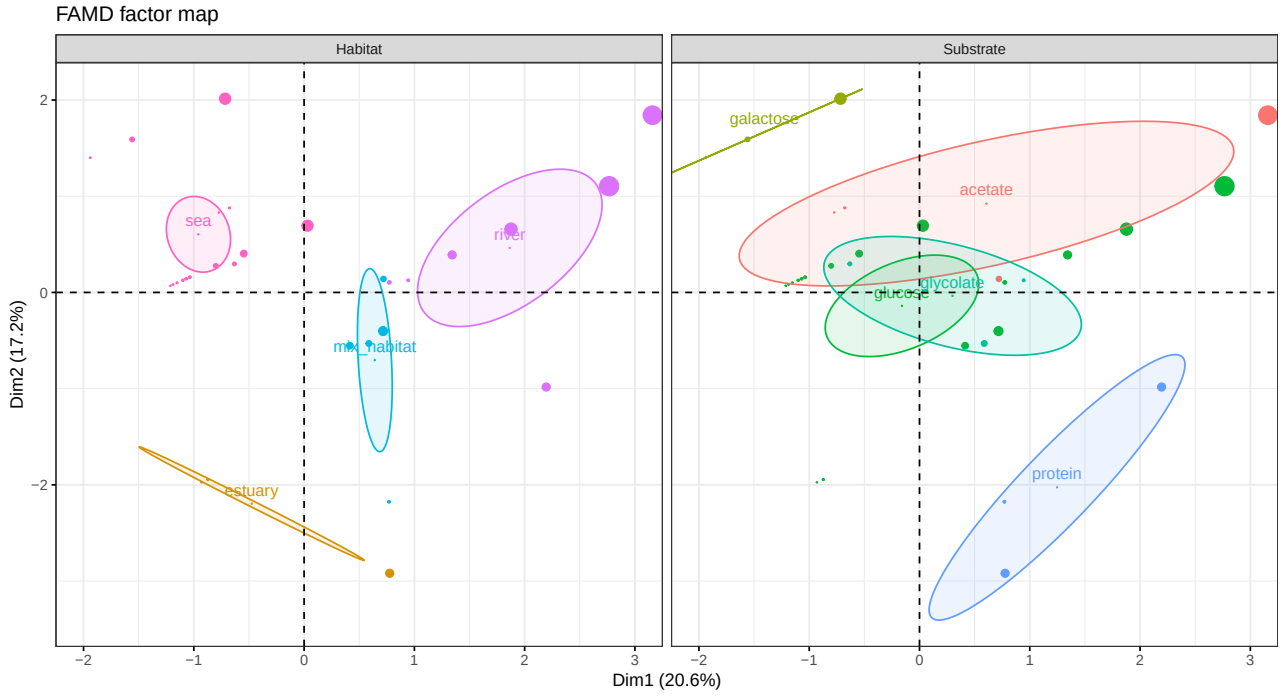


Figure A7: Distribution of the individuals corresponding to the Monod half-saturation constant for uptake of substrate (K_s in mgC L^{-1}) on dimensions 1 & 2 with the confidence ellipses

B Overview of oxygen and organic matter related equations and parameters in C-RIVE

B1 DO and OM processes in C-RIVE

B1.1 Dissolved oxygen evolution equations

DO in the water column depends on physical, bacterial, and phytoplanktonic processes:

$$\frac{d[O_2]}{dt} = \frac{d[O_2]}{dt}_{physical} + \frac{d[O_2]}{dt}_{phytoplanktonic} + \frac{d[O_2]}{dt}_{bacterial} \tag{B1}$$

The physical process depends on reaeration due to dams, wind, navigation, the oxygen holding capacity of water, and the diffusion of oxygen between the water-sediment interface as follows:

$$\frac{d[O_2]}{dt}_{physical} = \frac{K_{rea}}{h}([O_2]_{sat}(T) - [O_2]) - \frac{D_s}{h}([O_2]_{water} - [O_2]_{sed}) + \frac{d[O_2]}{dt}_{dams} \tag{B2}$$

where,

h: water depth [m]

$[O_2]_{sat}(T)$: the saturated oxygen concentration at temperature T [mgO_2L^{-1}]

D_s: the coefficient of diffusion between water and sediment layer [ms^{-1}]

K_{rea} : [ms^{-1}] the reoxygenation coefficient calculated from the empirical formula of Thibodeaux et al. [1994] as follows:

$$K_{rea} = \sqrt{\frac{D_m V_{wat}}{h}} + (K_{wind} V_{wind}^{2.23} (D_m * 10^4)^{2/3} + K_{navig}) \tag{B3}$$

where,

K_{wind}: reoxygenation coefficient due to wind [ms^{-1}]

V_{wind}: wind speed [ms^{-1}]

V_{wat}: river flow velocity [ms^{-1}]

K_{navig}: reoxygenation coefficient due to navigation [ms^{-1}] [Vilmin, 2014]

D_m: molecular diffusivity of DO [m^{2}s^{-1}]

The phytoplanktonic process depends on phytoplankton respiration ($R_{O_2,pp}$) and photosynthesis ($P_{O_2,pp}$) as follows:

$$\frac{d[O_2]}{dt}_{phytoplanktonic} = P_{O_2,pp} - R_{O_2,pp} \tag{B4}$$

And the bacterial process that is the main source of oxygen consumption depends on the heterotrophic bacterial kinetics and the availability of substrate matter (S, considered to be the rapidly biodegradable dissolved organic matter, DOM_1, in this model) as follows:

$$\frac{d[O_2]}{dt}_{bacterial} = -\tau_{HB}(1 - Y_{HB})\ uptake \tag{B5}$$

$$uptake = \frac{\mu_{max,HB} e^{\frac{-(T-T_{opt,HB})^2}{\sigma^2_{HB}}} \frac{[S]}{[S]+K_S}[HB]}{Y_{HB}} \tag{B6}$$

where,

$[HB]$: the concentration of heterotrophic bacteria (hereafter, called bacteria) $[mgCL^{-1}]$

τ_{HB}: 1.0 $[molO_2/molC]$ for full oxidation of OM in the respiration process

Y_{HB}: the growth yield of heterotrophic bacteria [-]

$uptake$: the uptake of substrate (here S = DOM_1) for bacteria growth $[mgCL^{-1}h^{-1}]$

$T_{opt,HB}$: optimum temperature for the growth of bacteria $[°C]$

$\mu_{max,HB}$: the maximum growth rate of bacteria at $T_{opt,HB}$ $[/h]$

σ_{HB}: standard deviation of bacteria temperature function $[°C]$

K_s: Monod half-saturation constant for bacterial growth (uptake constant) $[mgCL^{-1}]$

B1.2 Organic matter degradation equations

The degradation of OM happens through the uptake of small monomeric organic substrates (S, here $S = DOM_1$) by heterotrophic bacteria on the basis of the HSB model [Billen et al., 1988, Servais et al., 1989, Billen, 1991] and presented by Eq. (B7) and (B9). These substrates are either the direct input (P_S) of DOM_1 from OM sources or produced from the exoenzymatic hydrolysis of the macromolecular fractions of both dissolved (DOM_2) and particulate (POM_1, POM_2) organic matter [Billen, 1991] or they originate from the phytoplankton excretion, which produces more easily utilizable OM (DOM_1) and microorganism lysis products that are macromolecular matter [Larsson and Hagstrom, 1979, Garnier and Benest, 1990, Billen, 1991].

$$\frac{d[S = DOM_1]}{dt} = hyd_{DOM_2} + hyd_{POM_{1,2}} - uptake_{HB} + P_S + P_E + P_L \tag{B7}$$

where,

P_S, P_E, P_L: represent DOM_1 from the direct input of OM sources, phytoplankton excretion, and microorganism lysis, respectively $[mgCL^{-1}h^{-1}]$

hyd_{DOM_2}: hydrolysis of DOM_2 into DOM_1 based on the exoenzymatic hydrolysis equation of Michaelis-

Menten $[mgCL^{-1}h^{-1}]$

$hyd_{POM_{1,2}}$: hydrolysis of POM_1 and POM_2 into DOM_1 and DOM_2, respectively, by first-order kinetics $[mgCL^{-1}h^{-1}]$

$$hyd_{DOM_2} = k_{hyd,max}\frac{[DOM_2]}{[DOM_2] + K_{DOM2}}[HB] \tag{B8}$$

$$uptake_{HB} = \mu_{max,HB}\frac{[DOM_1]}{[DOM_1] + K_s}[HB] \tag{B9}$$

where,

$uptake_{HB}$: uptake or consumption of DOM_1 by heterotrophic bacteria $[mgCL^{-1}h^{-1}]$

$k_{hyd,max}$: coefficient for hydrolysis of DOM_2 into DOM_1 $[/h]$

K_{DOM2}: constant of semi-saturation for the hydrolysis of DOM_2 $[mgCL^{-1}]$

B2 Parameterization of organic matter partitioning and degradation

In order to account for the uncertainties related to the parameterization of OM degradation kinetics and its partitioning into different constituent fractions, the following two sets of parameters are introduced:

B2.1 OM degradation parameters

The parameters related to OM degradation are K_s (represents uptake of DOM_1 by bacteria), K_{DOM2} and $k_{hyd,max}$ (represent hydrolysis of DOM_2 to DOM_1), which have been defined in section B1.2 and that already exist in C-RIVE. Hydrolysis parameters of POM are not considered in this study because the rate of hydrolysis of $POM_{1,2}$ is slower than that of DOM_2 by an order of magnitude of 100 to 1000 [Billen et al., 1994].

B2.2 OM partitioning parameters

The following five parameters are introduced to represent the partitioning of OM:

$$t = \frac{DOM}{TOC}$$

$$b_1 = \frac{BDOM}{DOM}$$

$$s_1 = \frac{DOM_1}{BDOM}$$

$$b_2 = \frac{BPOM}{POM}$$

$$s_2 = \frac{POM_1}{BPOM}$$

where,

TOC: total organic matter or carbon (= DOM + POM) [$mgCL^{-1}$]

$BDOM$: biodegradable DOM (= DOM_1 + DOM_2) [$mgCL^{-1}$]

$BPOM$: biodegradable POM (= POM_1 + POM_2) [$mgCL^{-1}$]

t: ratio between dissolved and total organic matter [-]

b_1: ratio between biodegradable DOM and DOM [-]

s_1: ratio between rapidly biodegradable DOM and biodegradable DOM [-]

b_2: ratio between biodegradable POM and POM [-]

s_2: ratio between rapidly biodegradable POM and biodegradable POM [-]

$DOM_{1,2,3}$ and $POM_{1,2,3}$ were state variables in the former version of C-RIVE. They used to be forced information provided by user. They are now defined by the proposed partitioning model which has the above-mentioned five parameters.

C Sensitivity analysis protocol

SA methodology steps

1. **Input parameter identification**: Initially, a set of D input parameters (Table 4) are identified with their corresponding ranges of variation (Table 1).

2. **Parameter sampling and model input creation**: Saltelli's extension of the Sobol sequence [Saltelli, 2002] implemented in PYTHON SALIB package [Herman and Usher, 2017] is employed to create different combinations of the input parameters, which are designed to produce optimized simulations and efficient analysis results. Considering a sample size of 10,000 (N) (needed for stable results based on Nossent et al. [2011]), a matrix with a size of N(2D+2) × D is created for each SA analysis where every row represents one set of input parameters for the model.

3. **Model simulation**: In this step, the model inputs are launched into C-RIVE for the simulation period considered with a 1-min time step. As an output, a DO time series matrix with a size of $N(2D+2) \times M$, where M is the number of output time steps based on a 15 min output time step [1], is created corresponding to each input matrix created in the previous step. Figure C1 demonstrates the ensemble of 260,000 [=N(2D+2)] DO simulations of TOC = 5 $mgCL^{-1}$ in the second SA analysis (TOC = 5 $mgCL^{-1}$ is used in this study to represent the TOC range of 1-10 $mgCL^{-1}$ and to show the results in case they are similar across all TOC concentrations).

4. **Dimensionality reduction**: The empirical orthogonal function (EOF) method is an adaptation of principal component analysis (PCA) to study a phenomenon that changes with a continuous variable, such as time, and is applied to transform the output data from one coordinate system into another by introducing new uncorrelated (orthogonal) variables (principal components) [Jolliffe and Cadima, 2016]. EOF is adopted to transform the model output, which is a DO

[1] M = 45- or 5-day simulation period × 24 hrs × 3600 min / (1-min simulation time step × 15-min output time step) = 4230 or 480 depending on the simulation period, respectively

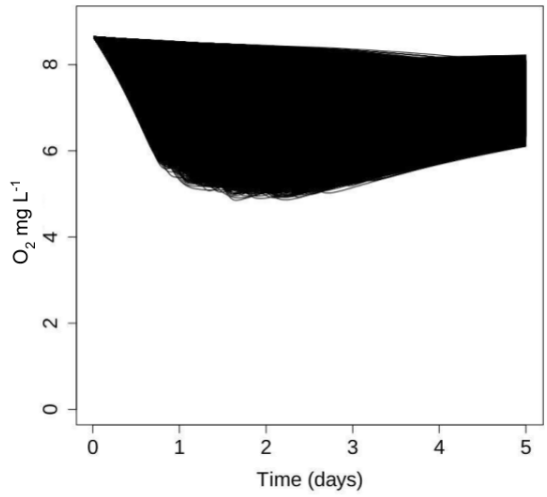


Figure C1: Ensemble of the 260,000 DO simulations for TOC = 5 mgCL^{-1} in the second SA analysis

times series matrix composed of M columns into a smaller matrix where each simulation can be represented by a linear combination of EOFs. The coefficients of this linear combination are indeed orthogonal projections that maximize the variance while transforming the data from a higher-dimensional space into a lower one. The way EOF decreases dimensionality is such that it ranks the components based on the maximized variance. In other words, most of the information is kept in the first few components, thereby making it possible to reduce the number of dimensions without losing a considerable amount of information [Wold et al., 1987]. In this study, the first k EOF elements that constitute at least 99% of the total model variance are considered to represent each single simulation of the DO time series as shown for TOC = 5 mgCL^{-1} in the second SA (Fig. C2a), where four (k) significant EOFs are found such that the first EOF (EOF_1) represents almost 55% of the total variance. Figure C2b illustrates the evolution of the eigenvalues of the four (k) EOFs with time, which are consequently used to represent each simulation in terms of the k new coordinates . Thereby, an $N(2D+2) \times M$ matrix is converted into a new matrix of $N(2D+2) \times k$, which will be subjected to the Sobol SA method. The R *prcomp* function is used to conduct the EOF analysis.

5. **Sobol sensitivity analysis**: The Sobol SA method [Sobol, 1993, Saltelli et al., 2010] is applied in this study to evaluate the sensitivity of the model output against the input parameters. It is a variance-based method that classifies the parameters based on their contribution to and/or influence on the total variance of the model output [Brookes et al., 2015]. It is a convenient method to be used for SA of complex models that involve interactions between parameters. In

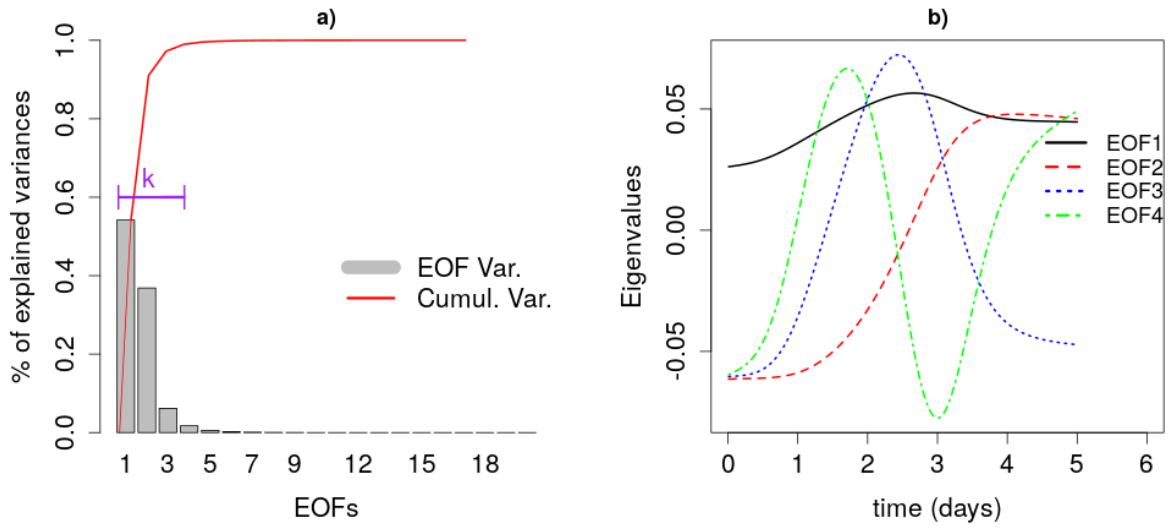


Figure C2: a) Cumulative sum of EOF variances and b) time evolution of four (k) significant EOFs for TOC = 5 mgCL^{-1} in second SA

this method, the model output (Y) is expressed as a function of D parameters:

$$Y = f(X) = f(X_1, ..., X_D), \tag{C1}$$

such that the model output could be decomposed by elementary functions:

$$f(X) = f_0 + \sum_{i=1}^{D} f_i(X_i) + \sum_{i=1}^{D-1} \sum_{j=i+1}^{D} f_{ij}(X_i, X_j) + ... + f_{1,...,D}(X_1, ..., X_D) \tag{C2}$$

Here f_0 is the expectation of the model output and each one of the elementary functions have a zero mean and can be computed by integration:

$$f_0 = \int_{[0,1]^D} f(X) d_X \tag{C3}$$

$$f_i(X_i) = -f_0 + \int_{[0,1]^{D-1}} f(X) d_{X \sim i} \tag{C4}$$

$$f_{ij}(X_i, X_j) = -f_0 - f_i(X_i) - f_j(X_j) + \int_{[0,1]^{D-2}} f(X) d_{X \sim (ij)} \tag{C5}$$

On the other hand, the total unconditional model variance could be defined as:

$$V(Y) = \int_{[0,1]^D} f^2(X) d_X - f_0^2 \tag{C6}$$

Thereby, the total unconditional variance of the model can be expressed as:

$$V(Y) = \sum_{i=1}^{D} V_i(X_i) + \sum_{i=1}^{D-1} \sum_{j=i+1}^{D} V_{ij}(X_i, X_j) + ... + V_{1,...,D}(X_1, ..., X_D) \tag{C7}$$

where, V_i is the partial variance of the i_{th} parameter and V_{ij} is the interaction effect of the i_{th} and j_{th} parameters. The partial variance is calculated as:

$$V_{i_1,...,i_s} = \int_0^1 ... \int_0^1 f^2_{i_1,...,i_s}(X_{i_1}, ..., X_{i_s}) dX_{i_1}, ...dX_{i_s} \tag{C8}$$

where $s = 1, ..., D$ and f_i is an elementary function. Therefore, the first-order Sobol SA indices can be computed as follows:

$$S_i = \frac{V_i}{V} \tag{C9}$$

S_i is also called as the "main effect" because it represents the contribution of a single input parameter i on the total variance. The total sensitivity index (S_{Ti}), also called "global effect," is another index that represents the sum of the first-order index (S_i) and the effect of the interaction between the parameters and is calculated as follows:

$$S_{Ti} = S_i + \sum_{j \neq i} S_{ij} + ... \tag{C10}$$

here, $S_{ij} = \frac{V_{ij}}{V}$ is called the "second-order index" and measures the interaction between a pair of parameters X_i and X_j. Therefore, the sum of second-order interactions of any parameter X_A with other parameters (X_B, ..., X_D) is considered to represent the second-order index of each parameter (S_2) as follows:

$$S_{2,A} = \sum_{j} S_{Aj} \tag{C11}$$

Since the output of previous step is a matrix of k vectors corresponding to the k EOFs, the Sobol indices of parameters are initially calculated k times for each EOF and then added while being weighted by the variance of the corresponding EOF.

D Detailed results of experiments for detecting change in model parameters

D1 Monod curve

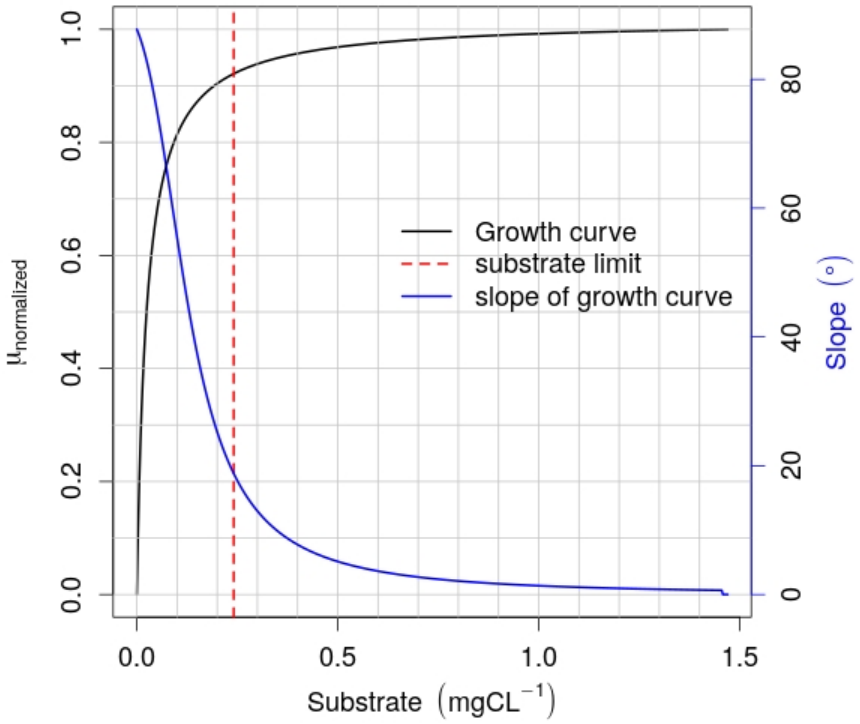


Figure D1: Monod curve and its slope plotted using a semi-saturation constant, K_S, of 0.025 mgC L^{-1} and a maximum growth rate of 0.07 h^{-1} (Eq. 4.13). The growth rate is normalized with the maximum rate of bacteria.

D2 Detailed results of the different combinations of DA configuration parameters for detecting change in bacteria physiology and dissolved organic carbon biodegradability

Figure D2: Detection of change in bacteria physiology (exp. 4). P_{yield}: perturbation % of bacteria growth yield parameter. P_{growth}: perturbation % of bacteria maximum growth parameter. The blue colored cell shows the best combination.

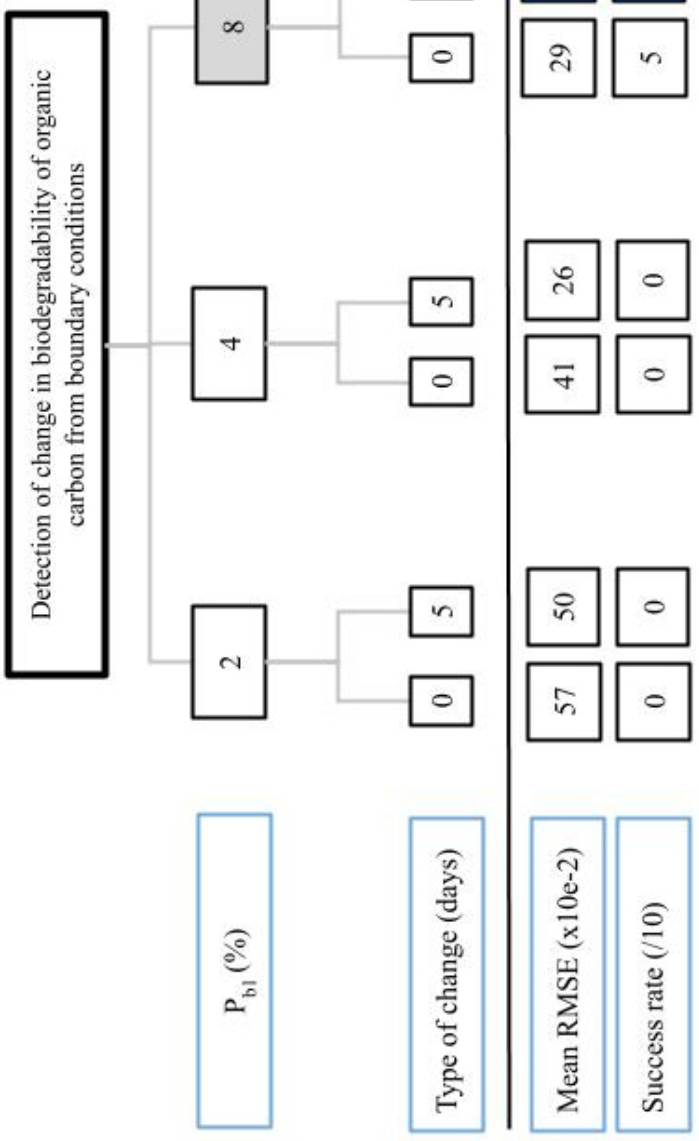


Figure D3: Detection of change in biodegradability of dissolved organic carbon of boundary conditions (exp. 5). P_{b1}: perturbation % of b_1 parameter. Changes types are 0 days (abrupt) and 5 days (gradual). The blue colored cell shows the best combination.

E ProSe-PA User Guide

ProSe-PA0.67

Guide d'utilisation - User guide

Nicolas Flipo, Shuaitao Wang, Aurélien Bordet, Lauriane Vilmin, Masihullah Hasanyar

Abstract

Cette notice présente la structure d'un fichier de commande permettant de réaliser une simulation ProSe0.67. 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.

E1 Get Started

Cette partie décrit les étapes à suivre afin d'installer correctement le programme en mode utilisateur ou développeur.

E1.1 Installation GSL - GNU Scientific Library

Passer directement à la section E1.2 lors de la première installation en mode utilisateur, ou à la section E1.3 en mode développeur. Revenir à cette section en cas de problème lors de l'installation.

Pour installer la version statique de libgsl 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 RED-HAT/FEDORA, voir les commentaires directement dans le Makefile concernant les variables d'environnement PREFIX_GSL et SUFFIX_GSL.

Si rien ne marche, essayer de localiser la librairie statique de GSL et ses .h. Utiliser la commande `> locate libgsl` and `> locate gsl_matrix`

E1.2 Installation en mode utilisateur simple

Cette section décrit la procédure d'installation du code pour un utilisateur simple de ProSe.

À l'endroit où vous souhaitez installer le programme télécharger les sources de ProSe : `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
```

En cas de problème de compilation relatif à gsl (gsl/gsl_matrix.h not found) retourner à la section E1.1.

E1.3 Installation en mode développeur

Pour pouvoir développer dans le code, il est conseillé de suivre l'ensemble des étapes suivantes pour un OS administré en **bash** :

1. 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
   ```

2. Téléchargement et compilation de ProSe0.67 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 à la section E1.1 de la procédure de démarrage.

E1.4 Compilations spéciales pour parallélisation et assimilation de données

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
```

E1.5 Lancement de ProSe

Que ce soit en mode utilisateur ou développeur, ProSe0.67est prêt à fonctionner par la commande :

```
> ./prose-pa0.67 fichier_de_commande.COMM log_file.log
```

E1.6 Workflow for launching simulations in ProSe

1) First of all, a steady state simulation shall be launched for the desired period in order to get initialization condition (initial height, discharge and species concentration along the river) for the transient simulations:

```
> ./prose-pa0.67 steady_state_command.COMM steady.log
```

2) The conc_init file (concentration unit is g m^{-3} for MES and mmol m^{-3} for other species) from the output folder of steady state simulation shall be copied to the layers/conc_init folder so that it could be used in the compartment files (watercomp). Then, the shell script creer_conc_init shall be launched so that the considered initial species concentrations be divided into separate files each of them (to be sure the column number is good for each species in cree_conc_init):

```
> ./creer_conc_init
```

3) The steady state output folder address shall be given in the command file of the transient state command file:

```
Input_folders = $HOME/Programmes/prose/trunk/test/SimRivStretch
                ...
          = $HOME/OUTPUT_PROSE_COUPLED/Steady_sim
```

4) The transient state command file (direct simulation or data assimiliation) could now be launched.

```
> ./prose-pa0.67 transient_state_command.COMM transient.log
```

E2 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 = {
           date_format = FR_TS #must before t_ini if we use date, FR_TS by default
           year0_julian = 2007 #must before t_ini if we use date, 1850 by default


           t_ini = [d] 0. #01/01/2011 00:00:00 # 2011-01-01 #we can use date
           t_end = [d] 10. #01/01/2012 00:00:00 # 2011-01-01 #we can use date
           dt = [min] 10.
           #beginning = 30 december 2011 00 00 #date when t = 0.
          }


settings = {
       regime = steady #transient
       hyd = yes
       ttc = no
       rive = no
       #dynamic_T=no #for heat transport
       #tube = no
       solver = sp_prose #gc_prose
       #da = no
       #num_thread = 1
```

```
        #num_particules = 1
        #num_thread_par = 1
        #num_tube_def = 5
        #alpha_da = 0.3
        #seuil_chla = [ugchla/l] 0.
        #error_obs = 0.01
        #s_percent = 0.10
        #default_t_inflows = [°C] 20. #default water temperature for inflows
        #nb_comp_phy = 3
        #dbo_oxy = [molO2/molC] 1.0

#species = { ... }

#macrospecies = { ... }


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 }
         }
}
### libseb, surface energy balance for temperature
#Meteo = { ... }
### heat transport
#Heat_transport = { ... }
### solute transport
#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 = { ... }
   #energy_balances_heat = { ... }
   #tube_output = {
   #    mesh_tube = { ... }
   #   hyd_tube = { ...  }
   #}

}
```

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'erreur est écrit et l'exécution du programme est stoppée. La structure de ces fichiers et décrite plus loin. La syntaxe concernant les sorties de modèle est également développée dans la suite.

E3 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
	Paramètres de temps			
`date_format`	Format de la date, quand on utilise les dates pour les données d'entrée ou de sortie (FR_TS or US_TS)			FR_TS
`year0_julian`	Année initiale pour calculer julian day			1850
`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			
	Type de simulation			
`regime`	type de simulation : permanent ou transitoire			STEADY
	Modules de simulation			
`hyd`	Module hydraulique			YES
`ttc`	Module de transport			NO
`rive`	Module biogéochimique			NO
`tube`	Module hydraulique pseudo-2D par tube			NO
`dynamic_T`	Transport de chaleur			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
`num_tube_def`	Nombre de tubes pour chaque bief	1
`default_t_inflows`	Température des apports par défaut	20 °C
	Biogéochimie	
`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 $\mathrm{molO_2molC^{-1}}$
	Assimilation de données	
`alpha_da`	La taille effective minimum $\alpha \times N_{par}$	0.3
`seuil_chla`	Le seuil de la concentration en chl a 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 après ré-échantillonnage (random walk, $\epsilon \sim N(0, 0.1(max - min))$)	0.1
`random_walk`	Type de random walk (loop ou not_loop)	loop
`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 E1: Eléments à définir dans le fichier de commande

E4 Simulation hydraulique

E4.1 Paramétrisation pour hydraulique

Eléments à définir dans le fichier de commande				
nom	description	*	**	valeur par défaut
	Tabulations verticales			
`dz`	Pas des tabulations verticales			0.1 m
`upstream_Hmin`	Hauteur d'eau minimale sur le domaine			0.
`ndim`	nombre de dimensions			1
`downstream_Hmax`	Hauteur d'eau maximale sur le domaine			10. m
	Paramètres globaux			
`curvature`	Valeur globale de la courbure			0.
`global_strickler`	Valeur globale du coefficient de Strickler			40.
	Options de calcul			
`calculate_curvature`	calcul de la courbure par libhyd si les profiles en travers sont géoréférencés			NO
`dx`	Pas de redécoupage longitudinal			aucun
`theta`	Degré d'implicite pour hydraulique			0.9

`eps_Q`	Précision de calcul des débits aux faces	10^{-6} m^3s^{-1}
`eps_Z`	Précision de calcul des cotes aux centres	10^{-6} m
`schem_type`	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 E2: Eléments à définir dans le fichier de commande

E4.2 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 limits 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 ;

E4.3 Description du réseau hydrographique

E4.3.1 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 égalemnt 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.

E4.3.2 Définition des biefs

Les biefs sont décrits par une singularité amont, une singularité avale 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 $m^{\frac{1}{3}}s^{-1}$ sera adoptée pour le bief considéré.

```
upstream_point -> middle_point 1
{
   strickler = { 0. 40. }
};

middle_point -> downstream_dam 1;
```

E4.3.3 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.
};
```

E4.4 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, le 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] [/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.

E4.5 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`.

`UPSTREAM_INFLOW` décrit une condition amont de débit. `INFLUENT` identifie un affluent non décrit explicitement. `EFFLUENT` identifie un rejet ou un prélèvement (débit négatif).

ATTENTION : Conceptuellement, dans le code, il n'y a pas de différences de traitement entre les deux types `INFLUENT` et `EFFLUENT`. Cette distinction n'est pas claire. Il faut légèrement reprendre l'algorithme et distinguer `EFFLUENT` en deux types : `EFFLUENT` et `UPTAKE` pour que l'utilisateur n'ait plus à définir de débits négatifs pour les prélève-

ments. En plus il faudra bien distinguer tous ces cas dans les colonnes du fichier de sortie mass_balance

`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).

E4.6 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.

E5 Simulation de transport et biogéochimique

E5.1 Les fichiers de paramètres des espèces

Les fichiers de description des espèces particulaires 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 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 E3), puis sont données les

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 E3: Codes des espèces biogéochimique simulées

descriptions de une plusieurs sous-espèces (`nom_sous_espece = { ... }`).

Pour chaque sous-espèce, différents paramètres peuvent être définis selon leur type : paramètres des espèces particulaires (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 processes 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.

Pour les bactéries hétérotrophes et nitrifiantes, il est possible de définir une limitation de croissance des bactéries dans les sédiments par syntaxe K_{BACT} `= [mgC/l] 0.0003`. Le taux de croissance sera ajusté en fonction de la biomasse bactérienne ($BACT$) dans les sédiments.

$$\mu = \mu_{max}\left(\frac{\frac{1}{BACT}}{\frac{1}{BACT} + \frac{1}{K_{BACT}}}\right) \tag{E1}$$

où,

μ_{max} : Taux de croissance maximal pour la bactérie (hétérotrophe ou nitrifiante), [h-1]

$BACT$: Biomasse bactérienne totale dans le sédiment, [mgC/L]

K_{BACT} : Facteur de limitation, [mgC/L]

μ : Taux de croissance ajusté en fonction de la biomasse bactérienne, [h-1]

`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 particulaires 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 : $NO_2 + \frac{1}{2}O_2 -> NO_3$. Les concentrations en

nitrite, dioxygène et nitrate sont alors calculées de la manière suivante :

$$\begin{aligned} \frac{dC_{NO_2}}{dt} &= -\texttt{decay} \cdot C_{NO_2} \\ \frac{dC_{O_2}}{dt} &= 0.5 \cdot \texttt{decay} \cdot C_{NO_2} \\ \frac{dC_{NO_3}}{dt} &= \texttt{decay} \cdot C_{NO_2} \end{aligned}$$

où C_{NO_2}, C_{O_2} et C_{NO_3} sont en mmol·m^{-3}.

La liste des paramètres pouvant être définis pour chaque processus est donnée dans la Table E4.

Paramètres des espèces biogéochmiques

nom	description	unité	valeur par défaut
	Paramètres des espèces particulaires		
`rho`	masse volumique	[ML^{-3}]	1000 gm^{-3}
`phi`	porosité		0.
`Vsed`	vitesse de sédimentation	[LT^{-1}]	
`scouring`	paramètre d'arrachage phytoplanctonique		
	Paramètres des espèces dissoutes		
`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	$[L^2M^{-1}]$
`eta_mes`	terme de l'extinction lumineuse fonction de la matière en suspension	$[L^2M^{-1}]$
`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
`kmich_no2`	constante de demi-saturation vis-à-vis des nitrites lors de la production d'oxyde nitreux
`cmin_o2`	concentration en oxygène dissous en-dessous de laquelle peut être produit de l'oxyde nitreux lors de la nitrification
`stoechio_nit`	stoechiométrie de la consommation en dioxygène lors de la nitrification

`fbis`	paramètre de correction de la croissance dans le cas d'une production d'oxyde nitreux

Paramètres de respiration (`respiration`)

`maint`	taux de respiration lié à la survie de la cellule
`energ`	taux de respiration liée à la dépense énergétique de croissance

Paramètres de broutage (`grazing`)

`graz`	taux de broutage

Paramètres d'excrétion (`excretion`)

`excr_cst`	taux d'excrétion de base
`excr_phot`	taux d'excrétion liée à la photosynthèse

Paramètres de mortalité (`mortality`)

`mort`	taux de mortalité
`phi_lim_mort`	concentration critique au-dessus de laquelle la mortalité est accélérée
`delta`	facteur multiplicatif de la mortalité pour une concentration supérieure à `phi_lim_mort`

Paramètres de réaction chimique (`reaction`)

`decay`	taux de décroissance radioactive (ordre 1)

Paramètres d'hydrolyse (`hydrolysis`)

`kc_hydr`	constante de demi-saturation pour l'hydrolyse de la matière organique

	Paramètres d'adsorption (`ads_desorption`)	
`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	
	Paramètres de réaeration (`reaeration`)	
`rea_navig`	coefficient de réaeration due à la navigation	
`rea_wind`	coefficient de réaeration due au vent	
`D_Richey`	coefficient de diffusion moléculaire à l'interface eau-air	
	Paramètres de transport (`Advection-dispersion`)	
`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 ou NO)	$[L^2T^{-1}]$

Table E4: Eléments à définir dans le fichier de commande

E5.2 Parametrization of the macrospecies

The macrospecies are indeed other species that exist in the river but are not considered under the species section. This could be parameters of the repartition of the total organic carbon (TOC) that

could be used to divide TOC first of all into MOD and MOP and then into their smaller fractions of MOD_1, MOD_2, MOD_3, MOP_1, MOP_2, MOP_3. For better understanding of these parameters, please visit here

```
macrospecies = {

TOC = {
   total_organic_carbon
   related_to MOD MOP , threshold = 0.7 # MOD = t*TOC, MOP = (1-t)*TOC
   share_mop = {
   biodegradable = { val = 0.3  #b2
                    # range = { 0.15, 0.5 }
                   }
   fast_biodegradable = { val = 0.7 } #s2
   }
   share_mod = {
   biodegradable = {  val = 0.30 #b1
                      # range = { 0.1 , 0.5 }
                    }
   fast_biodegradable = { val = 0.7 } #s1
   }
}
}
```

The list of parameters which could be defined under this item is in Table: E5.

Macrospecies parameters

name	description	unit	value
	Parameters of TOC macrospecie		
`threshold`	value of ratio between MOD and TOC (t)		
`share_mop`	definition of parameters for repartition of MOP		
`share_mod`	definition of parameters for repartition of MOD		

`biodegradable`	the ratio between biodegradable portion of MOP (BMOP) and MOP itself (b_2 = BMOP/MOP; similarly its the ratio between biodegradable portion of MOD and MOD itself (b_1 = BMOD/MOD)
`fast_biodegradable`	the ratio between fast biodegradable portion of MOP (MOP1) and BMOP (s_2 = MOP1/BMOP; similarly its the ratio between fast biodegradable portion of MOD (MOD1) and BMOD (s_1 = MOD1/BMOD
`val`	mean value of the parameter
`range`	variation range of the parameter

Table E5: Elements to be defined in the command file

E5.3 Paramétrisation pour transport des dissous

La paramétrisation de transport commence par `transport = { theta_ttc = 1. }` Un seul paramètre à définir est le theta_ttc pour le schéma numérique. Pour l'instant theta_ttc = 1 est accepté (implicite).

E5.4 Paramétrisation pour transport de chaleur et l'échange atmosphérique

Il est possible de modéliser le transport de chaleur (Température). Dans ce cas là, la température au sein de chaque maille serait différent l'un de l'autre. Ces températures seront utilisées pour la modélisation biogéochimique.

Le transport de chaleur est activé par la syntaxe `dynamic_T = yes` défini dans le bloc `setting`. La paramétrisation du transport de chaleur est assez simple. Comme le transport se fait par un schéma implicite, `theta_t` défini par utilisateur sera remplacé par 1.0. Pour la prochaine version de libttc ainsi ProSe-P, il est possible de programmer un schéma semi-implicite.

```
Heat_transport = {
theta_t = 0.8 # will be repalced by 1. in ProSe-P
```

```
init_T_file = [°C] include  seine_init_test_T


}
```

Le bilan énergétique à l'interface air-eau est possible calculé par la librairie libseb avec la syntaxe `METEO`. Dans ce bloc `METEO`, les données météorologiques de type SAFRAN comme la température de l'air, la pression atmosphérique, le rayonnement par onde longue et courte, l'humidité de l'air et la vitesse du vent sont définies.

```
Meteo = {
met_corresp = safran_forcing_to_reach_new_choisy_poses.txt
met_data = {
folder = $HOME/DATA_PROSE/DATA_BORDET/SAFRAN_1990_2019
                year0_meteo = 1840 #par defaut initial year for meteo data
forcings = {
T_air = TMPH_19902019.txt
P_atm = PRESSURE_GRID_SAFRAN.dat
H_sw = SWDH_19902019.txt
H_lw = LWDH_19902019.txt
Humid = HRLH_19902019.txt
U_wind = VNTH_19902019.txt
}
}
# #riverside_param = riv_param.txt # --> optional
}
```

La syntaxe `met_corresp` définit un table de correspondance entre les biefs et les mailles SAFRAN.

```
Amont Aval Voie Id_safran
conf_cygnes diff_saint_germain 1 1563
conf_grande_jatte diff_saint_denis 1 1454
diff_saint_germain conf_saint_germain 1 1563
suresnesg conf_puteaux 1 1454
suresnes suresnesg 1 1454
```

```
conf_seguin suresnes 1 1562
diff_saint_louis diff_cite 1 1564
diff_cite conf_saint_louis 2 1564
...
```

La syntaxe `folder` donne le chemin d'accès des données météorologiques et `year0_meteo` définit l'année initiale pour calculer julian day dans ces données.

E5.5 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
	Options de calcul		
`numerical_method`	Méthode numérique de résolution des processus biogéochimiques (RUNGE_KUTTA ou EXPLICIT)		`RUNGE_KUTTA`
`max_div_dt`	Facteur de redécoupage maximal du pas de temps de calcul		1.0
`dz_rive`	Pas de discrétisation pour le calcul de la pénétration des radiations dans la colonne d'eau		0.1 m

* Paramètre indispensable

Table E6: Eléments à définir dans le fichier de commande

E5.6 Les fichiers de description des compartiments

Un fichier de description de compartiments doit débuter 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 par 1) et une seconde les concentrations;
- Soit deux valeurs (un entier et une concentration) qui signifie que les concentrations sont les mêmes pour tous les éléments.

```
WATER


water1_s1 =
{
   porosity = 1.
   rho_layer = [kg/l] 1.

   initial_concentrations = { phy 1 = [mgC/l] include conc_init/phy_init # 1 10
                              po4 1 = [mgP/l] include conc_init/po4_init # 1 10
                              ...
                              mop 1 = [mgC/l] include conc_init/mop1_init # 1 5
                              mop 2 = [mgC/l] include conc_init/mop2_init # 1 5
                              mop 3 = [mgC/l] include conc_init/mop3_init # 1 5
                              ...
                            }
}
```

Dans la vase et le périphyton, la masse volumique et la porosité du milieu peuvent changer en fonction des espèces sédimentées, érodées, arrachées...

```
VASE


vase1_s1 =
{
   mass_layer = [kg] include mass_init #1 1000.
```

```
   porosity = 0.9
   rho_layer = [kg/l] 2

   initial_concentrations = { mes 1 = [mg/l] include conc_init/mes_init # 1 10
                              bactn 1 = [mgC/l] include conc_init/bactn1_init # 1 2.
                              nh4 1 = [mgN/l] include conc_init/nh4_init # 1 10.
                              ...
                            }
}
```

E5.7 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
              }
```

E5.8 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 :

$$T(t_j) = \texttt{mean}_T - \texttt{amplitude}_T \cdot \cos\left(\frac{2\pi \cdot (t_j - \texttt{delay}_T)}{365.}\right) \quad \text{(E2)}$$

$$I(t_j) = \frac{\pi}{2}\texttt{attenuation}_I \left(\texttt{mean}_I - \texttt{amplitude}_I \cdot \cos\left(\frac{2\pi \cdot t_j}{365.}\right)\right) \cdot \sin\left(\frac{\pi * (t_h - 1)}{\lambda(t_j)}\right) \quad \text{(E3)}$$

$$\lambda(t_j) = \texttt{mean}_\lambda - \texttt{amplitude}_\lambda \cdot \cos\left(\frac{2\pi \cdot t_j}{365.}\right) \quad \text{(E4)}$$

t_j et t_h correspondent respectivement à la date en jours et en heures.

La synthaxe 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.
              }
```

E5.9 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 colonne : `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`. L'utilisateur peut choisir les paramètres à ajuster en commentant (#) ceux qu'on ne veut pas ajuster.

```
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
b1_river 0.1 0.5
```

However, the param_range file can also contain the random walk (s_percent) of each parameter as its 5th column: `parameter name parameter unit min_val max_val s_percent`. This gives the capability of running simulations where each parameter has different random walk values.

```
mumax_bact [/h] 0.01 0.13 0.03
yield_bact 0.03 0.5 0.030
b1_river 0.1 0.5 0.030
```

E6 Définition des fichier de sorties

Le syntaxe des sorties dans le fichier de commande est montré ci-dessous. Plusieurs type de sortie 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/l]
          o2 [mgO2/l]
          TW [°C] #water temperature
          TA [°C] # air temperature

          ...
        }

  points = { SEINE
        [km] 715. 1 # Sartrouville
        [km] 706.3 1 # Aval Bougival
        [km] 702.4 2 #Aval Chatou
        [km] 730.2  1 # Aval Andrésy
        [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_t13
        [km] 730.2 1 o2 1 [s] [mgO2/l] include conc_init/o2_andresy_pprose_dia_0.01_t13
        [km] 706.3 1 o2 1 [s] [mgO2/l] include conc_init/o2_bougival_pprose_dia_0.01_t13
        [km] 702.4 2 o2 1 [s] [mgO2/l] include conc_init/o2_chatou_pprose_dia_0.01_t13
```

```
        [km] 674.6  1 o2 1 [s] [mgO2/l]  include conc_init/o2_suresnes_pprose_dia_0.01_t13
}

  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]
          o2 [mgO2/l]
          nh4 [mgN/l]
          no3 [mgN/l]
          TW [°C] #water temperature
          TA [°C] # air temperature
          #...
        }

  extent = { canal_test
             [km] 0.   1
```

```
                [km] 75.  1
              }
     extent = { canal_test
                [km] 75.  1
                [km] 100. 1
              }

     graphics = GNUPLOT
   }
tube_output = {
  mesh_tube = {

           t_ini = [d] 1551. # 2011-04-01
           t_end = [d] 1764. # 2011-10-31
           dt = [h] 12.
           time_unit = [d]
            extent = { canal_test
                [km] 0.   1
                [km] 75.  1
              }
                }

  hyd_tube = {
           t_ini = [d] 1551. # 2011-04-01
           t_end = [d] 1764. # 2011-10-31
           dt = [h] 12.

           time_unit = [d]
           extent = { canal_test
                [km] 0.   1
                [km] 75.  1
              }
```

```
            }

}
   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]
             #}


      extent = { canal_test
           [km] 75.  1
           [km] 100. 1
         }
         }

mb2 = { t_ini = [d] 1481.
        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]
                  #}


           extent = { canal_test
                [km] 25.  1
                [km] 100. 1
              }
              }
              }

energy_balances_heat = {

  eb1 = { t_ini = [d] 1461  #  one year
          t_end = [d] 1462
          n_steps = 1. #dt_mb = n_steps * dt_Simul
         time_unit = [d]
         heat_unit = [J]
         extent = { canal_test
```

```
            [km] 0. 1
            [km] 100. 1
  }
        }
}

}
```

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.

E6.1 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 en compte le transport (biogéochimie), un fichier `conc_init` est créé dans le répertoire de résultats. Ce fichier sert à générer 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 finales du sédiment (`volume_final_vase`) dans le répertoire de résultats. Ces fichiers finaux peuvent être utilisés pour initialiser une autre simulation transitoire.

Le bilan de masse hydraulique par élément sur l'ensemble du domaine simulé est créé 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 finals.

E6.2 Profils ponctuels

Un ensemble de profils ponctuels (`time_series`) est défini par :

- son instant initial (`t_ini`, jour julian ou format date), son instant final (`t_end` jour julian ou format date) , 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.

E6.3 Profils en long

Un ensemble de profils en long (`longitudinal_profiles`) est défini par :

- son instant initial (`t_ini` jour julian ou format date), son instant final (`t_end` jour julian ou format date) , 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.

E6.4 Sortie tube hydraulique

Un ensemble de sortie pour les tubes hydrauliques (pseudo-2D) est possible. L'utilisateur peut définir deux types de sortie dans ce bloc : `mesh_tube` (les coordonnées des tubes : points centraux et sommets des tubes (5 points pour un tube)) et `hyd_tube` (les données hydrauliques associées à chaque tube).

- 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 linéaires sur lesquels on souhaite le profil (`extent = { ...}`). Si l'on ne définit pas `extent`, les sorties sont pour l'ensemble du domaine simulé

Plusieurs ensembles peuvent être définis dans le même fichier de commande.
Il est possible de visualiser les tubes dans Qgis.

- Couche -> ajouter une couche -> ajouter une couche de texte délimité
- Choisir le fichier à visualiser
- Définition de la géométrie -> Well known text (WKT)
- Champ de géométrie : field_3 pour polygone et field_2 pour point
- Type de géométrie : Polygone ou point qu'on veut visualiser
- SCR de la géométrie : Lambert 93 (LAMB93) ou Lambert91 (LAMB91)
- Ajouter

E6.5 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.

E6.6 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`.

E6.7 Bilans biogéochimiques

Un ensemble de bilans biogéochimiques (mass_balances_bio = {}) 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
- les tronçons sur lesquels on souhaite le bilan (`extent = { ...}`)

Plusieurs ensembles peuvent être définis dans le même fichier de commande.

Description des termes d'un bilan pour une espèce

nom	description
`HYDROLYSIS`	Hydrolyse des matières organiques
`RAD_DECAY`	Réaction définie par utilisateur (équation E2)
`ADS_DESORPTION`	Adsorption-désorption
`MORTALITY`	Mortalité pour les espèces vivantes
`Grazing`	Broutage de Zooplankton
`GROWTH`	Croissance des espèces vivantes
`RECYCLING`	Reminéralisation pour les nutriments
`RESP`	Respiration des espèces vivantes
`EXCR`	Excrétion des PHY
`PHOT`	Photosynthèse
`CR_SR`	Cataobolyse et synthèse des produits de réserve **S**
`REA`	Réaeration

DEGAS	Dégazage
REOX	Réoxygénation due aux barrages
DIFFUSION	Échange dissous entre WATER et VASE
SEDIMENTATION_EROSION	Sédimentation et érosion
FIN	Masse (flux) entrant du domaine (PK amont)
FOUT	Masse (flux) sortant du domaine (PK aval)
FIN_LATERAL	Masse (flux) latéral (apport et prélèvement)
MINI	Masse initiale
MEND	Masse finale pour dt
EXCEPTION	Procédure d'exception
ERROR	Erreur du bilan

Table E7: Description des termes d'un bilan pour une espèce

E6.8 Bilan énergétique

Un ensemble de bilans énergétique (energy_balances_heat = {}) 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 unités associées pour le temps `time_unit =` et pour l'énergie `heat_unit =`
- les tronçons sur lesquels on souhaite le bilan (`extent = { ...}`)

Plusieurs ensembles peuvent être définis dans le même fichier de commande (conseillé).

Bibliography

C. Abdelnour, D. Ferreira, M. van de Beek, N. Cedres, K. Oppedal, L. Cavallin, F. Blanc, O. Bousiges, L.-O. Wahlund, A. Pilotto, A. Padovani, M. Boada, J. Pagonabarraga, J. Kulisevsky, D. Aarsland, A. W. Lemstra, and E. Westman. Parsing heterogeneity within dementia with Lewy bodies using clustering of biological, clinical, and demographic data. *Alzheimer's Research & Therapy*, 14(1):14, Jan. 2022. ISSN 1758-9193. doi: 10.1186/s13195-021-00946-w.

M. Adamovic, F. Branger, I. Braud, and S. Kralisch. Development of a data-driven semi-distributed hydrological model for regional scale catchments prone to Mediterranean flash floods. *Journal of Hydrology*, 541:173–189, Oct. 2016. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2016.03.032.

J. D. Allan and M. M. Castillo. *Stream Ecology: Structure and Function of Running Waters*. Springer, Dordrecht, 2. ed., reprinted edition, 2009. ISBN 978-1-4020-5582-9.

G. B. Arhonditsis, B. A. Adams-Vanharn, L. Nielsen, C. A. Stow, and K. H. Reckhow. Evaluation of the current state of mechanistic aquatic biogeochemical modeling: Citation analysis and future perspectives. *Environmental Science & Technology*, 40(21):6547–6554, Nov. 2006. ISSN 0013-936X. doi: 10.1021/es061030q.

S. Arndt, B. Jørgensen, D. LaRowe, J. Middelburg, R. Pancost, and P. Regnier. Quantifying the degradation of organic matter in marine sediments: A review and synthesis. *Earth-Science Reviews*, 123:53–86, Aug. 2013. ISSN 00128252. doi: 10.1016/j.earscirev.2013.02.008.

M. L. Atkins, I. R. Santos, and D. T. Maher. Seasonal exports and drivers of dissolved inorganic and organic carbon, carbon dioxide, methane and δ13C signatures in a subtropical river network. *Science of The Total Environment*, 575:545–563, Jan. 2017. ISSN 00489697. doi: 10.1016/j.scitotenv.2016.09.020.

F. Azam, T. Fenchel, J. Field, J. Gray, L. Meyer-Reil, and F. Thingstad. The Ecological Role of Water-Column Microbes in the Sea. *Marine Ecology Progress Series*, 10:257–263, 1983. ISSN 0171-8630, 1616-1599. doi: 10.3354/meps010257.

M. Babbar-Sebens, L. Li, K. Song, and S. Xie. On the Use of Landsat-5 TM Satellite for Assimilating Water Temperature Observations in 3D Hydrodynamic Model of Small Inland Reservoir in Mid-western US. *Advances in Remote Sensing*, 2(3):214–227, Aug. 2013. doi: 10.4236/ars.2013.23024.

R. T. Bailey and M. Ahmadi. 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, 2014. ISSN 0304-3800. doi: 10.1016/j.ecolmodel.2014.01.015.

C. Basset-Mens, B. Rhino, A. Ndereyimana, U. Kleih, and Y. Biard. Eco-efficiency of tomato from Rwamagana district in Rwanda: From field constraints to statistical significance. *Journal of Cleaner Production*, 229:420–430, Aug. 2019. ISSN 0959-6526. doi: 10.1016/j.jclepro.2019.04.256.

T. J. Battin, L. A. Kaplan, S. Findlay, C. S. Hopkinson, E. Marti, A. I. Packman, J. D. Newbold, and F. Sabater. Biophysical controls on organic carbon fluxes in fluvial networks. *Nature Geoscience*, 1 (2):95–100, Feb. 2008. ISSN 1752-0894, 1752-0908. doi: 10.1038/ngeo101.

T. J. Battin, S. Luyssaert, L. A. Kaplan, A. K. Aufdenkampe, A. Richter, and L. J. Tranvik. The boundless carbon cycle. *Nature Geoscience*, 2(9):598–600, Sept. 2009. ISSN 1752-0894, 1752-0908. doi: 10.1038/ngeo618.

T. J. Battin, R. Lauerwald, E. S. Bernhardt, E. Bertuzzo, L. G. Gener, R. O. Hall, E. R. Hotchkiss, T. Maavara, T. M. Pavelsky, L. Ran, P. Raymond, J. A. Rosentreter, and P. Regnier. River ecosystem metabolism and carbon biogeochemistry in a changing world. *Nature*, 613(7944):449–459, Jan. 2023. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-022-05500-8.

J. J. Beaulieu, C. P. Arango, D. A. Balz, and W. D. Shuster. Continuous monitoring reveals multiple controls on ecosystem metabolism in a suburban stream. *Freshwater Biology*, 58(5):918–937, 2013. ISSN 1365-2427. doi: 10.1111/fwb.12097.

H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, 5(1):180214, Dec. 2018. ISSN 2052-4463. doi: 10.1038/sdata.2018.214.

M. Beck and E. Halfon. Uncertainty, identifiability and the propagation of prediction errors: A case study of lake ontario. *Journal of Forecasting*, 10(1-2):135–161, 1991. doi: 10.1002/for.3980100109.

M. Beck and P. Young. Systematic identification of do-bod model structure. *Jnl. of Env. Eng. Division , American Society of Civil Engineers*, 102:909–927, 1976.

J. Berger, M. Vigan, B. Pereira, T. T. Nguyen, R. Froissart, N. Belmatoug, F. Dalbiès, A. Masseau, C. Rose, C. Serratrice, Y.-M. Pers, I. Bertchansky, F. Camou, M. Bengherbia, C. Bourgne, C. Caillaud, M. Pettazzoni, A. Berrahal, J. Stirnemann, F. Mentré, and M. G. Berger. Intra-monocyte Pharmacokinetics of Imiglucerase Supports a Possible Personalized Management of Gaucher Disease Type 1. *Clinical Pharmacokinetics*, 58(4):469–482, Apr. 2019. ISSN 0312-5963, 1179-1926. doi: 10.1007/s40262-018-0708-8.

E. S. Bernhardt, J. B. Heffernan, N. B. Grimm, E. H. Stanley, J. W. Harvey, M. Arroita, A. P. Appling, M. J. Cohen, W. H. McDowell, R. O. Hall, J. S. Read, B. J. Roberts, E. G. Stets, and C. B. Yackulic. The metabolic regimes of flowing waters. *Limnology and Oceanography*, 63(S1), Mar. 2018. ISSN 0024-3590, 1939-5590. doi: 10.1002/lno.10726.

M. J. Bernot, D. J. Sobota, R. O. Hall Jr, P. J. Mulholland, W. K. Dodds, J. R. Webster, J. L. Tank, L. R. Ashkenas, L. W. Cooper, C. N. Dahm, S. V. Gregory, N. B. Grimm, S. K. Hamilton, S. L. Johnson, W. H. Mcdowell, J. L. Meyer, B. Peterson, G. C. Poole, H. M. Valett, C. Arango, J. J. Beaulieu, A. J. Burgin, C. Crenshaw, A. M. Helton, L. Johnson, J. Merriam, B. R. Niederlehner, J. M. O'brien, J. D. Potter, R. W. Sheibley, S. M. Thomas, and K. Wilson. Inter-regional comparison of land-use effects on stream metabolism. *Freshwater Biology*, 55(9):1874–1890, 2010a. ISSN 1365-2427. doi: 10.1111/j.1365-2427.2010.02422.x.

M. J. Bernot, D. J. Sobota, R. O. Hall Jr, P. J. Mulholland, W. K. Dodds, J. R. Webster, J. L. Tank, L. R. Ashkenas, L. W. Cooper, C. N. Dahm, S. V. Gregory, N. B. Grimm, S. K. Hamilton, S. L. Johnson, W. H. Mcdowell, J. L. Meyer, B. Peterson, G. C. Poole, H. M. Valett, C. Arango, J. J. Beaulieu, A. J. Burgin, C. Crenshaw, A. M. Helton, L. Johnson, J. Merriam, B. R. Niederlehner, J. M. O'brien, J. D. Potter, R. W. Sheibley, S. M. Thomas, and K. Wilson. Inter-regional comparison of land-use effects on stream metabolism. *Freshwater Biology*, 55(9):1874–1890, 2010b. ISSN 1365-2427. doi: 10.1111/j.1365-2427.2010.02422.x.

K. Besemer, B. Luef, S. Preiner, B. Eichberger, M. Agis, and P. Peduzzi. Sources and composition of organic matter for bacterial growth in a large European river floodplain system (Danube, Austria).

Organic Geochemistry, 40(3):321–331, 2009. ISSN 01466380. doi: 10.1016/j.orggeochem.2008.12.005.

G. Billen. Heterotrophic Utilization and Regeneration of Nitrogen. In J. E. Hobbie and P. J. leB. Williams, editors, *Heterotrophic Activity in the Sea*, pages 313–355. Springer US, Boston, MA, 1984. ISBN 978-1-4684-9010-7. doi: 10.1007/978-1-4684-9010-7_15.

G. Billen. Protein Degradation in Aquatic Environments. In T. D. Brock and R. J. Chróst, editors, *Microbial Enzymes in Aquatic Environments*, pages 123–143. Springer New York, New York, NY, 1991. ISBN 978-1-4612-7793-4 978-1-4612-3090-8. doi: 10.1007/978-1-4612-3090-8_7.

G. Billen and M. Somville. Utilization of primary products by planktonic and benthic bacteria. Concerted Actions Oceanography. Final Report. Technical Report 3, Biological Oceanography Committee, 1982.

G. Billen, C. Lancelot, E. Becker, and P. Servais. Modelling microbial processes (phyto- and bacterioplankton) in the Schelde estuary. *Hydrobiological Bulletin*, 22(1):43–55, 1988. ISSN 0165-1404, 1573-5125. doi: 10.1007/BF02256781.

G. Billen, P. Servais, and S. Becquevort. Dynamics of bacterioplankton in oligotrophic and eutrophic aquatic environments: Bottom-up or top-down control? *Hydrobiologia*, 207(1):37–42, 1990. ISSN 0018-8158, 1573-5117. doi: 10.1007/BF00041438.

G. Billen, J. Garnier, and P. Hanset'. Modelling phytoplankton development in whole drainage networks: The RIVERSTRAHLER Model applied to the Seine river system. *Hydrobiologia*, 289: 119–137, 1994. doi: 10.1007/BF00007414.

G. Billen, J. Garnier, and V. Rousseau. Nutrient fluxes and water quality in the drainage network of the Scheldt basin over the last 50 years. *Hydrobiologia*, 540(1):47–67, May 2005. ISSN 1573-5117. doi: 10.1007/s10750-004-7103-1.

J. Block, L. Mathieu, P. Servais, D. Fontvieille, and P. Werner. Indigenous bacterial inocula for measuring the biodegradable dissolved organic carbon (BDOC) in waters. *Water Research*, 26(4): 481–486, 1992. ISSN 00431354. doi: 10.1016/0043-1354(92)90049-A.

G. Blöschl, J. Hall, A. Viglione, R. A. P. Perdigão, J. Parajka, B. Merz, D. Lun, B. Arheimer, G. T. Aronica, A. Bilibashi, M. Boháč, O. Bonacci, M. Borga, I. Čanjevac, A. Castellarin, G. B. Chirico, P. Claps, N. Frolova, D. Ganora, L. Gorbachova, A. Gül, J. Hannaford, S. Harrigan, M. Kireeva,

A. Kiss, T. R. Kjeldsen, S. Kohnová, J. J. Koskela, O. Ledvinka, N. Macdonald, M. Mavrova-Guirguinova, L. Mediero, R. Merz, P. Molnar, A. Montanari, C. Murphy, M. Osuch, V. Ovcharuk, I. Radevski, J. L. Salinas, E. Sauquet, M. Šraj, J. Szolgay, E. Volpi, D. Wilson, K. Zaimi, and N. Živković. Changing climate both increases and decreases European river floods. *Nature*, 573 (7772):108–111, Sept. 2019. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-019-1495-6.

F. L. Brailsford, H. C. Glanville, P. N. Golyshin, P. J. Johnes, C. A. Yates, and D. L. Jones. Microbial uptake kinetics of dissolved organic carbon (DOC) compound groups from river water and sediments. *Scientific Reports*, 9(1):11229, 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-47749-6.

V. J. Brookes, D. Jordan, S. Davis, M. P. Ward, and J. Heller. Saltelli Global Sensitivity Analysis and Simulation Modelling to Identify Intervention Strategies to Reduce the Prevalence of Escherichia coli O157 Contaminated Beef Carcasses. *PLoS ONE*, 10(12):1–22, 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0146016.

A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9 (5):e535, 2018. doi: 10.1002/wcc.535.

E. M. Carstea, C. L. Popa, A. Baker, and J. Bridgeman. In situ fluorescence measurements of dissolved organic matter: A review. *Science of The Total Environment*, 699:134361, Jan. 2020. ISSN 00489697. doi: 10.1016/j.scitotenv.2019.134361.

N. Catalán, B. Obrador, M. Felip, and J. L. Pretus. Higher reactivity of allochthonous vs. autochthonous DOC sources in a shallow lake. *Aquatic Sciences*, 75(4):581–593, Oct. 2013. ISSN 1015-1621, 1420-9055. doi: 10.1007/s00027-013-0302-y.

N. Catalán, J. P. Casas-Ruiz, D. von Schiller, L. Proia, B. Obrador, E. Zwirnmann, and R. Marcé. Biodegradation kinetics of dissolved organic matter chromatographic fractions in an intermittent river: Biodegradation Dynamics of DOM Fractions. *Journal of Geophysical Research: Biogeosciences*, 122(1):131–144, 2017. ISSN 21698953. doi: 10.1002/2016JG003512.

J. Cho, Y. Her, and D. Bosch. 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, 2017. doi: 10.1007/s10666-016-9530-6.

K. H. Cho, Y. Pachepsky, M. Ligaray, Y. Kwon, and K. H. Kim. Data assimilation in surface water

quality modeling: A review. *Water Research*, 186:116307, Nov. 2020. ISSN 0043-1354. doi: 10.1016/j.watres.2020.116307.

J. J. Cole, Y. T. Prairie, N. F. Caraco, W. H. McDowell, L. J. Tranvik, R. G. Striegl, C. M. Duarte, P. Kortelainen, J. A. Downing, J. J. Middelburg, and J. Melack. Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget. *Ecosystems*, 10(1):172–185, Feb. 2007. ISSN 1435-0629. doi: 10.1007/s10021-006-9013-8.

J. J. Cole, O. Hararuk, and C. T. Solomon. Chapter 7 - The Carbon Cycle: With a Brief Introduction to Global Biogeochemistry. In K. C. Weathers, D. L. Strayer, and G. E. Likens, editors, *Fundamentals of Ecosystem Science (Second Edition)*, pages 131–160. Academic Press, Jan. 2021. ISBN 978-0-12-812762-9. doi: 10.1016/B978-0-12-812762-9.00007-1.

B. Cosby and G. Hornberger. Identification of photosynthesis-light models for aquatic systems i. theory and simulations. *Ecological Modelling*, 23(1-2):1–24, 1984. doi: 10.1016/0304-3800(84)90116-9.

B. Cox. A review of currently available in-stream water-quality models and their applicability for simulating dissolved oxygen in lowland rivers. *The Science of The Total Environment*, 314–316: 335–377, 2003. ISSN 00489697. doi: 10.1016/S0048-9697(03)00063-9.

C. C. Crawford, J. E. Hobbie, and K. L. Webb. The Utilization of Dissolved Free Amino Acids by Estuarine Microorganisms. *Ecology*, 55(3):551–563, 1974. ISSN 0012-9658. doi: 10.2307/1935146.

B. C. Crump, G. W. Kling, M. Bahr, and J. E. Hobbie. Bacterioplankton Community Shifts in an Arctic Lake Correlate with Seasonal Changes in Organic Matter Source. *Applied and Environmental Microbiology*, 69(4):2253–2268, Apr. 2003. doi: 10.1128/AEM.69.4.2253-2268.2003.

B. Demars, J. Thompson, and J. Manson. Stream metabolism and the open diel oxygen method: Principles, practice, and perspectives. *Limnology and Oceanography: Methods*, 13(7):356–374, 2015. doi: 10.1002/lom3.10030.

J. S. Diamond, S. Bernal, A. Boukra, M. J. Cohen, D. Lewis, M. Masson, F. Moatar, and G. Pinay. Stream network variation in dissolved oxygen: Metabolism proxies and biogeochemical controls. *Ecological Indicators*, 131:108233, Nov. 2021. ISSN 1470-160X. doi: 10.1016/j.ecolind.2021.108233.

J. S. Diamond, F. Moatar, M. J. Cohen, A. Poirel, C. Martinet, A. Maire, and G. Pinay. Metabolic regime shifts and ecosystem state changes are decoupled in a large river. *Limnology and Oceanography*, 67(S1), Feb. 2022. ISSN 0024-3590, 1939-5590. doi: 10.1002/lno.11789.

J. S. Diamond, F. Moatar, R. Recoura-Massaquant, A. Chaumot, J. Zarnetske, L. Valette, and G. Pinay. Hypoxia is common in temperate headwaters and driven by hydrological extremes. *Ecological Indicators*, 147:109987, Mar. 2023. ISSN 1470-160X. doi: 10.1016/j.ecolind.2023.109987.

W. Dixon, G. K. Smyth, and B. Chiswell. Optimized selection of river sampling sites. *Water Research*, 33(4):971–978, Mar. 1999. ISSN 0043-1354. doi: 10.1016/S0043-1354(98)00289-9.

H. T. Do, S.-L. Lo, P.-T. Chiueh, and L. A. Phan Thi. Design of sampling locations for mountainous river monitoring. *Environmental Modelling & Software*, 27–28:62–70, Jan. 2012. ISSN 1364-8152. doi: 10.1016/j.envsoft.2011.09.007.

A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001. doi: 10.1007/978-1-4757-3437-9.

K. R. Dyer. Biogeochemistry of Major World Rivers, Scope 42 edited by E.T. Degens, S. Kempe, and J.E. Richey. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 1:189–190, 1991. doi: 10.1002/aqc.3270010209.

G. Y. El Serafy, M. Blaas, M. A. Eleveld, and H. J. van der Woerd. Data assimilation of satellite data of suspended particulate matter in delft3d-waq for the north sea. In *Proceedings of the Joint EUMETSAT/AMS Conference, Darmstadt, Germany*, pages 1–8. Citeseer, 2007.

N. Escoffier, N. Bensoussan, L. Vilmin, N. Flipo, V. Rocher, A. David, F. Métivier, and A. Groleau. Estimating ecosystem metabolism from continuous multi-sensor measurements in the Seine River. *Environmental Science and Pollution Research*, 25(24):23451–23467, 2016. ISSN 0944-1344, 1614-7499. doi: 10.1007/s11356-016-7096-0.

N. Escoffier, N. Bensoussan, L. Vilmin, N. Flipo, V. Rocher, A. David, F. Métivier, and A. Groleau. Estimating ecosystem metabolism from continuous multi-sensor measurements in the Seine River. *Environmental Science and Pollution Research*, 25(24):23451–23467, 2018. ISSN 0944-1344, 1614-7499. doi: 10.1007/s11356-016-7096-0.

B. Escofier. Traitement simultané de variables qualitatives et quantitatives en analyse factorielle. *Cahiers de l'analyse des données*, 4(2):137–146, 1979. URL http://eudml.org/doc/87944.

F. Esculier and S. Barles. Past and Future Trajectories of Human Excreta Management Systems: Paris in the Nineteenth to Twenty-First Centuries. In N. Flipo, P. Labadie, and L. Lestel, editors,

The Seine River Basin, volume 90, pages 117–140. Springer International Publishing, Cham, 2020. ISBN 978-3-030-54259-7 978-3-030-54260-3. doi: 10.1007/698_2019_407.

H. Etcheber, A. Taillez, G. Abril, J. Garnier, P. Servais, F. Moatar, and M.-V. Commarieu. Particulate organic carbon in the estuarine turbidity maxima of the Gironde, Loire and Seine estuaries: Origin and lability. *Hydrobiologia*, 588(1):245–259, 2007. ISSN 0018-8158, 1573-5117. doi: 10.1007/s10750-007-0667-9.

EU. European Commission. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy, 2000.

S. Even, M. Poulin, J. Garnier, G. Billen, P. Servais, A. Chesterikoff, and M. Coste. River ecosystem modelling: Application of the PROSE model to the Seine river (France). *Hydrobiologia*, 373(0): 27–45, 1998. ISSN 1573-5117. doi: 10.1023/A:1017045522336.

S. Even, M. Poulin, J.-M. Mouchel, M. Seidl, and P. Servais. Modelling oxygen deficits in the Seine River downstream of combined sewer overflows. *Ecological Modelling*, 173(2):177–196, Apr. 2004. ISSN 0304-3800. doi: 10.1016/j.ecolmodel.2003.08.019.

S. Even, J. M. Mouchel, P. Servais, N. Flipo, M. Poulin, S. Blanc, M. Chabanel, and C. Paffoni. Modeling the impacts of Combined Sewer Overflows on the river Seine water quality. *Science of The Total Environment*, 375(1-3):140–151, 2007. doi: 10.1016/j.scitotenv.2006.12.007.

G. Evensen, F. C. Vossepoel, and P. J. van Leeuwen. Fully Nonlinear Data Assimilation. In *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*, pages 95–110. Springer International Publishing, Cham, 2022. ISBN 978-3-030-96709-3. doi: 10.1007/978-3-030-96709-3_9.

D. Fetjah, Z. Idardare, B. Ihssane, L. F. Z. Ainlhout, and L. Bouqbis. Seasonal Paspalum vaginatum Physiological Characteristics Change with Agricultural Byproduct Biochar in Sandy Potting Soil. *Biology*, 11(4):560, Apr. 2022. ISSN 2079-7737. doi: 10.3390/biology11040560.

N. Flipo. *Modélisation intégrée des transferts d'azote dans les aquifères et les rivières : Application au bassin du Grand Morin.* PhD thesis, CIG, 2005.

N. Flipo, S. Even, M. Poulin, M.-H. Tusseau-Vuillemin, T. Ameziane, and A. Dauta. Biogeochemical modelling at the river scale: Plankton and periphyton dynamics: Grand Morin case study, France. *Ecological Modelling*, 176(3):333–347, 2004. ISSN 0304-3800. doi: 10.1016/j.ecolmodel.2004.01.012.

N. Flipo, C. Rabouille, M. Poulin, S. Even, M.-H. Tusseau-Vuillemin, and M. Lalande. Primary production in headwater streams of the Seine basin: The Grand Morin river case study. *Science of The Total Environment*, 375(1):98–109, 2007. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2006.12.015.

N. Flipo, J.-M. Mouchel, C. Fisson, S. Wang, and M. Le Gall. *Les Effets de La Crue de Juin 2016 Sur La Qualité de l'eau Du Bassin de La Seine*. Fascicule #17 Du PIREN-Seine. ARCEAU-IdF, 2018. ISBN 978-2-490-46305-3.

N. Flipo, L. Lestel, P. Labadie, M. Meybeck, and J. Garnier. Trajectories of the Seine River Basin. In *The Seine River Basin*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2020. doi: 10.1007/698_2019_437.

J. Garnier and D. Benest. Seasonal coupling between phyto- and bacterioplankton in a sand pit lake (Creteil lake, France). *Hydrobiologia*, 207:71–77, 1990.

J. Garnier and G. Billen. Ecological interactions in a shallow sand-pit lake (Lake Créteil, Parisian Basin, France): A modelling approach. *Hydrobiologia*, 275(1):97–114, 1994. ISSN 1573-5117. doi: 10.1007/BF00026703.

J. Garnier and G. Billen. Production vs. respiration in river systems: An indicator of an "ecological status". *Science of the Total Environment*, 375(1-3):110–124, 2007. doi: 10.1016/j.scitotenv.2006.12.006.

J. Garnier, G. Billen, P. Servais, and P. Servais. Physiological characteristics and ecological role of small and large sized bacteria in a polluted river (Seine River, France). *Arch. Hydrobiol. Beih.*, 37: 83–94, 1992a.

J. Garnier, P. Servais, and G. Billen. Bacterioplankton in the Seine River (France): Impact of the Parisian urban effluent. *Canadian Journal of Microbiology*, 38(1):56–64, 1992b. ISSN 0008-4166, 1480-3275. doi: 10.1139/m92-009.

J. Garnier, G. Billen, and M. Coste. Seasonal succession of diatoms and Chlorophyceae in the drainage network of the Seine River: Observation and modeling. *Limnology and Oceanography*, 40(4):750–765, 1995. ISSN 1939-5590. doi: 10.4319/lo.1995.40.4.0750.

J. Garnier, G. Billen, and L. Palfner. Understanding the oxygen budget and related ecological processes in the river Mosel: The RIVERSTRAHLER approach. In J. Garnier and J.-M. Mouchel, editors,

Man and River Systems, pages 151–166. Springer Netherlands, Dordrecht, 1999. ISBN 978-90-481-5393-0 978-94-017-2163-9. doi: 10.1007/978-94-017-2163-9_17.

J. Garnier, P. Servais, G. Billen, M. Akopian, and N. Brion. Lower Seine River and Estuary (France) Carbon and Oxygen Budgets during Low Flow. *Estuaries*, 24(6):964, 2001. ISSN 01608347. doi: 10.2307/1353010.

J. Garnier, G. Billen, E. Hannon, S. Fonbonne, Y. Videnina, and M. Soulie. Modeling transfer and retention of nutrients in the drainage network of the danube river. *Estuar. Coast. Shelf*, 54:285–308, 2002.

J. Garnier, N. Brion, J. Callens, P. Passy, C. Deligne, G. Billen, P. Servais, and C. Billen. Modeling historical changes in nutrient delivery and water quality of the Zenne River (1790s–2010): The role of land use, waterscape and urban wastewater management. *Journal of Marine Systems*, 128:62–76, Dec. 2013. ISSN 0924-7963. doi: 10.1016/j.jmarsys.2012.04.001.

J. Garnier, A. Marescaux, S. Guillon, L. Vilmin, V. Rocher, G. Billen, V. Thieu, M. Silvestre, P. Passy, M. Raimonet, A. Groleau, S. Théry, G. Tallec, and N. Flipo. Ecological Functioning of the Seine River: From Long-Term Modelling Approaches to High-Frequency Data Analysis. In *The Handbook of Environmental Chemistry*, pages 1–28. Springer Berlin Heidelberg, 2020.

J. E. Garvey, M. R. Whiles, and D. Streicher. A hierarchical model for oxygen dynamics in streams. *Canadian Journal of Fisheries and Aquatic Sciences*, 64(12):1816–1827, 2007. ISSN 0706-652X, 1205-7533. doi: 10.1139/f07-144.

N. Geeraert, F. O. Omengo, G. Govers, and S. Bouillon. Dissolved organic carbon lability and stable isotope shifts during microbial decomposition in a tropical river system. *Biogeosciences*, 13(2): 517–525, 2016. ISSN 1726-4189. doi: 10.5194/bg-13-517-2016.

M. R. Gmach, M. R. Cherubin, K. Kaiser, and C. E. P. Cerri. Processes that influence dissolved organic matter in the soil: A review. *Scientia Agricola*, 77(3):e20180164, 2020. ISSN 1678-992X. doi: 10.1590/1678-992x-2018-0164.

K. Gocke, R. Dawson, and G. Liebezeit. Availability of dissolved free glucose to heterotrophic microorganisms. *Marine Biology*, 62(2-3):209–216, 1981. ISSN 0025-3162, 1432-1793. doi: 10.1007/BF00388184.

A. Goffin, S. Guérin, V. Rocher, and G. Varrault. 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. Technical report, PIREN-Seine phase VII, 2017.

A. Goffin, S. Guérin, V. Rocher, and G. Varrault. Towards a better control of the wastewater treatment process: Excitation-emission matrix fluorescence spectroscopy of dissolved organic matter as a predictive tool of soluble BOD 5 in influents of six Parisian wastewater treatment plants. *Environmental Science and Pollution Research*, 25(9):8765–8776, Mar. 2018. ISSN 1614-7499. doi: 10.1007/s11356-018-1205-1.

M. G. M. Gonçalves, F. A. P. Avalos, J. V. dos Reis, M. V. Costa, S. H. G. Silva, G. C. Poggere, N. Curi, and M. D. de Menezes. Pedology-based management class establishment: A study case in Brazilian coffee crops. *Precision Agriculture*, 23(3):1027–1050, June 2022. ISSN 1573-1618. doi: 10.1007/s11119-021-09873-0.

M. Gonsior, P. Schmitt-Kopplin, and D. Bastviken. Depth-dependent molecular composition and photo-reactivity of dissolved organic matter in a boreal lake under winter and summer conditions. *Biogeosciences*, 10(11):6945–6956, Nov. 2013. ISSN 1726-4189. doi: 10.5194/bg-10-6945-2013.

R. González-Quintero, R. Barahona-Rosales, D. M. Bolívar-Vergara, N. Chirinda, J. Arango, H. A. Pantévez, G. Correa-Londoño, and M. S. Sánchez-Pinzón. Technical and environmental characterization of dual-purpose cattle farms and ways of improving production: A case study in Colombia. *Pastoralism*, 10(1):19, Sept. 2020. ISSN 2041-7136. doi: 10.1186/s13570-020-00170-5.

D. Graeber, J. Gelbrecht, M. T. Pusch, C. Anlanger, and D. von Schiller. Agriculture has changed the amount and composition of dissolved organic matter in Central European headwater streams. *Science of The Total Environment*, 438:435–446, 2012. ISSN 00489697. doi: 10.1016/j.scitotenv.2012.08.087.

D. Graeber, I. G. Boëchat, F. Encina-Montoya, C. Esse, J. Gelbrecht, G. Goyenola, B. Gücker, M. Heinz, B. Kronvang, M. Meerhoff, J. Nimptsch, M. T. Pusch, R. C. S. Silva, D. von Schiller, and E. Zwirnmann. Global effects of agriculture on fluvial dissolved organic matter. *Scientific Reports*, 5(1):16328, Dec. 2015. ISSN 2045-2322. doi: 10.1038/srep16328.

E. G. Gregorich, M. H. Beare, U. Stoklas, and P. St-Georges. Biodegradability of soluble organic matter in maize-cropped soils. *Geoderma*, 113(3):237–252, May 2003. ISSN 0016-7061. doi: 10.1016/S0016-7061(02)00363-4.

A. Groleau, N. Escoffier, L. Vilmin, A. Auge, M. Poulin, V. Rocher, F. Metivier, and N. Flipo. Apports croisés des mesures hautes fréquence pour le monitoring de la qualité de l'eau, le calcul d'indicateurs de métabolisme et la modélisation du fonctionnement trophique de la Seine à l'aval de Paris. Technical report, PIREN-Seine, 2016.

Y. Guo, Z. Guo, J. Wang, Z. Ye, L. Zhang, and J. Niu. Photodegradation of three antidepressants in natural waters: Important roles of dissolved organic matter and nitrate. *Science of The Total Environment*, 802:149825, Jan. 2022. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2021.149825.

A. Gurung, T. Iwata, D. Nakano, and J. Urabe. River Metabolism along a Latitudinal Gradient across Japan and in a global scale. *Scientific Reports*, 9(1):4932, Dec. 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-41427-3.

R. D. Hamilton and J. E. Preslan. OBSERVATIONS ON HETEROTROPHIC ACTIVITY IN THE EASTERN TROPICAL PACIFIC1: HETEROTROPHIC ACTIVITY IN THE PACIFIC. *Limnology and Oceanography*, 15(3):395–401, 1970. ISSN 00243590. doi: 10.4319/lo.1970.15.3.0395.

A. M. Hansen, T. E. C. Kraus, B. A. Pellerin, J. A. Fleck, B. D. Downing, and B. A. Bergamaschi. Optical properties of dissolved organic matter (DOM): Effects of biological and photolytic degradation. *Limnology and Oceanography*, 61(3):1015–1032, 2016. ISSN 1939-5590. doi: 10.1002/lno.10270.

M. Hasanyar, N. Flipo, T. Romary, S. Wang, and A. Yari. Rôle de la matière organique dans le métabolisme des rivières à bas débit. PIREN-Seine phase 8 - Rapport 2020, Mines Paris – PSL, 2020.

M. Hasanyar, N. Flipo, T. Romary, and S. Wang. Amélioration des simulations d'oxygène par intégration de la composition de la matière organique dans un modèle de qualité de l'eau. Technical report, Mines Paris – PSL, PIREN-Seine phase 8 - Technical report, 2022.

M. Hasanyar, N. Flipo, T. Romary, and S. Wang. A review of water quality variables and parameters characterizing organic matter in world rivers. manuscript submitted for publication. available upon request from masihullah.hasanyar@mines-paristech.fr. Mines Paris – PSL, 2023a.

M. Hasanyar, T. Romary, S. Wang, and N. Flipo. How much do bacterial growth properties and biodegradable dissolved organic matter control water quality at low flow? *Biogeosciences*, 20(8): 1621–1633, Apr. 2023b. ISSN 1726-4170. doi: 10.5194/bg-20-1621-2023.

B. He, M. Dai, W. Zhai, L. Wang, K. Wang, J. Chen, J. Lin, A. Han, and Y. Xu. Distribution, degradation and dynamics of dissolved organic carbon and its major compound classes in the Pearl River estuary, China. *Marine Chemistry*, 119(1-4):52–64, 2010. ISSN 03044203. doi: 10.1016/j.marchem.2009.12.006.

J. R. Helms, J. Mao, K. Schmidt-Rohr, H. Abdulla, and K. Mopper. Photochemical flocculation of terrestrial dissolved organic matter and iron. *Geochimica et Cosmochimica Acta*, 121:398–413, Nov. 2013. ISSN 00167037. doi: 10.1016/j.gca.2013.07.025.

J. Herman and W. Usher. SALib: An open-source python library for sensitivity analysis. *Journal of Open Source Software*, 2(9):97, 2017. doi: 10.21105/joss.00097.

J. Hill and P. Wheeler. Organic carbon and nitrogen in the northern California current system: Comparison of offshore, river plume, and coastally upwelled waters. *Progress in Oceanography*, 53 (2-4):369–387, 2002. ISSN 00796611. doi: 10.1016/S0079-6611(02)00037-X.

R. M. Holmes, J. W. McClelland, P. A. Raymond, B. B. Frazer, B. J. Peterson, and M. Stieglitz. Lability of DOC transported by Alaskan rivers to the Arctic Ocean. *Geophysical Research Letters*, 35(3):L03402, 2008. ISSN 0094-8276. doi: 10.1029/2007GL032837.

H.-G. Hoppe. Relations between active bacteria and heterotrophic potential in the sea. *Netherlands Journal of Sea Research*, 12(1):78–IN4, 1978. ISSN 00777579. doi: 10.1016/0077-7579(78)90026-1.

J. D. Hosen, O. T. McDonough, C. M. Febria, and M. A. Palmer. Dissolved Organic Matter Quality and Bioavailability Changes Across an Urbanization Gradient in Headwater Streams. *Environmental Science & Technology*, 48(14):7817–7824, July 2014. ISSN 0013-936X, 1520-5851. doi: 10.1021/es501422z.

E. R. Hotchkiss, R. O. Hall Jr, R. A. Sponseller, D. Butman, J. Klaminder, H. Laudon, M. Rosvall, and J. Karlsson. Sources of and processes controlling CO2 emissions change with the size of streams and rivers. *Nature Geoscience*, 8(9):696–699, Sept. 2015. ISSN 1752-0894, 1752-0908. doi: 10.1038/ngeo2507.

B. Hu, P. Wang, C. Wang, and T. Bao. Photogeochemistry of particulate organic matter in aquatic systems: A review. *Science of The Total Environment*, 806:150467, Feb. 2022. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2021.150467.

J. Huang, H. Yin, S. Chapra, and Q. Zhou. Modelling Dissolved Oxygen Depression in an Urban River in China. *Water*, 9(7):520, 2017. ISSN 2073-4441. doi: 10.3390/w9070520.

T.-H. Huang, Y.-H. Fu, P.-Y. Pan, and C.-T. A. Chen. Fluvial carbon fluxes in tropical rivers. *Current Opinion in Environmental Sustainability*, 4(2):162–169, May 2012. ISSN 18773435. doi: 10.1016/j.cosust.2012.02.004.

M. A. J. Hullar, L. A. Kaplan, and D. A. Stahl. Recurring Seasonal Dynamics of Microbial Communities in Stream Habitats. *Applied and Environmental Microbiology*, 72(1):713–722, Jan. 2006. ISSN 0099-2240, 1098-5336. doi: 10.1128/AEM.72.1.713-722.2006.

J. Hur and J. Cho. Prediction of BOD, COD, and Total Nitrogen Concentrations in a Typical Urban River Using a Fluorescence Excitation-Emission Matrix with PARAFAC and UV Absorption Indices. *Sensors (Basel, Switzerland)*, 12(1):972–986, Jan. 2012. ISSN 1424-8220. doi: 10.3390/s120100972.

M. Hutchins, G. Harding, H. Jarvie, T. Marsh, M. Bowes, and M. Loewenthal. Intense summer floods may induce prolonged increases in benthic respiration rates of more than one year leading to low river dissolved oxygen. *Journal of Hydrology X*, 8:100056, 2020. ISSN 25899155. doi: 10.1016/j.hydroa.2020.100056.

M. J. Islam, C. Jang, J. Eum, S.-m. Jung, M.-S. Shin, Y. Lee, Y. Choi, and B. Kim. C:N:P stoichiometry of particulate and dissolved organic matter in river waters and changes during decomposition. *Journal of Ecology and Environment*, 43(1):4, Dec. 2019. ISSN 2288-1220. doi: 10.1186/s41610-018-0101-4.

V. Ittekkot. Global trends in the nature of organic matter in river suspensions. *Nature*, 332(6163): 436–438, 1988. ISSN 0028-0836, 1476-4687. doi: 10.1038/332436a0.

O. Izagirre, U. Agirre, M. Bermejo, J. Pozo, and A. Elosegi. Environmental controls of whole-stream metabolism identified from continuous monitoring of Basque streams. *Journal of the North American Benthological Society*, 27(2):252–268, June 2008. ISSN 0887-3593, 1937-237X. doi: 10.1899/07-022.1.

M. S. Johnson, E. G. Couto, M. Abdo, and J. Lehmann. Fluorescence index as an indicator of dissolved organic carbon quality in hydrologic flowpaths of forested tropical watersheds. *Biogeochemistry*, 105 (1):149–157, Sept. 2011. ISSN 1573-515X. doi: 10.1007/s10533-011-9595-x.

I. T. Jolliffe and J. Cadima. Principal component analysis: A review and recent developments. *Philo-*

sophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374 (2065):20150202, Apr. 2016. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2015.0202.

M. Kadjeski, C. Fasching, and M. A. Xenopoulos. Synchronous Biodegradability and Production of Dissolved Organic Matter in Two Streams of Varying Land Use. *Frontiers in Microbiology*, 11, 2020. ISSN 1664-302X.

P.-G. Kang and M. J. Mitchell. Bioavailability and size-fraction of dissolved organic carbon, nitrogen, and sulfur at the Arbutus Lake watershed, Adirondack Mountains, NY. *Biogeochemistry*, 115(1-3): 213–234, 2013. ISSN 0168-2563, 1573-515X. doi: 10.1007/s10533-013-9829-1.

L. A. Kaplan and J. D. Newbold. Measurement of streamwater biodegradable dissolved organic carbon with a plug-flow bioreactor. *Water Research*, 29(12):2696–2706, Dec. 1995. ISSN 0043-1354. doi: 10.1016/0043-1354(95)00135-8.

A. Kassambara and F. Mundt. *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses*, 2020.

Z. V. Kazmiruk, D. W. Capelle, C. M. Kamula, S. Rysgaard, T. Papakyriakou, and Z. A. Kuzyk. High biodegradability of riverine dissolved organic carbon in late winter in Hudson Bay, Canada. *Elementa: Science of the Anthropocene*, 9(1):00123, May 2021. ISSN 2325-1026. doi: 10.1525/elementa.2020.00123.

E. Khan, R. W. Babcock, I. H. Suffet, and M. K. Stenstrom. Method development for measuring biodegradable organic carbon in reclaimed and treated wastewaters. *Water Environment Research*, 70(5):1025–1032, July 1998. ISSN 10614303. doi: 10.2175/106143098X123354.

E. Khan, S. King, R. W. B. Jr, and M. K. Stenstrom. Factors Influencing Biodegradable Dissolved Organic Carbon Measurement. *Journal of Environmental Engineering*, 125(6):514–521, 1999. ISSN 0733-9372, 1943-7870. doi: 10.1061/(ASCE)0733-9372(1999)125:6(514).

R. J. Kieber, R. F. Whitehead, and S. A. Skrabal. Photochemical production of dissolved organic carbon from resuspended sediments. *Limnology and Oceanography*, 51(5):2187–2195, 2006. ISSN 1939-5590. doi: 10.4319/lo.2006.51.5.2187.

C. Kim, J. B. Eom, S. Jung, and T. Ji. Detection of Organic Compounds in Water by an Optical Absorbance Method. *Sensors*, 16(1):61, Jan. 2016. ISSN 1424-8220. doi: 10.3390/s16010061.

J.-K. Kim, S. Jung, J.-s. Eom, C. Jang, Y. Lee, J. S. Owen, M.-S. Jung, and B. Kim. Dissolved and particulate organic carbon concentrations in stream water and relationships with land use in multiple-use watersheds of the Han River (Korea). *Water International*, 38(3):326–339, 2013. ISSN 0250-8060, 1941-1707. doi: 10.1080/02508060.2013.769411.

H. G. Knapik, C. V. S. Fernandes, J. C. R. de Azevedo, M. M. dos Santos, P. Dall'Agnol, and D. G. Fontane. Biodegradability of anthropogenic organic matter in polluted rivers using fluorescence, UV, and BDOC measurements. *Environmental Monitoring and Assessment*, 187(3):104, Mar. 2015. ISSN 0167-6369, 1573-2959. doi: 10.1007/s10661-015-4266-3.

B. Koehler, E. Broman, and L. J. Tranvik. Apparent quantum yield of photochemical dissolved organic carbon mineralization in lakes. *Limnology and Oceanography*, 61(6):2207–2221, 2016. ISSN 1939-5590. doi: 10.1002/lno.10366.

H. Köhler, B. Meon, V. V. Gordeev, A. Spitzy, and R. M. W. Amon. Dissolved organic matter (DOM) in the estuaries of Ob and Yenisei and the adjacent Kara Sea, Russia. In *Proceedings in Marine Science*, volume 6, pages 281–310. Elsevier (Amsterdam), 2003.

W. Köppen. Die Wärmezonen der Erde, nach der Dauer der heissen, gemässigten und kalten Zeit und nach der Wirkung der Wärme auf die organische Welt betrachtet (The thermal zones of the earth according to the duration of hot, moderate and cold periods and to the impact of heat on the organic world),(translated and edited by V OLKEN E. and S. B R ÖNNIMANN . – Meteorol. Z. 20 (2011), 351–360). *Meteorologische Zeitschrift*, 1:215–226, 1884. ISSN 0941-2948. doi: 10.1127/0941-2948/2011/105.

T. Lambert, S. Bouillon, F. Darchambeau, C. Morana, F. A. E. Roland, J.-P. Descy, and A. V. Borges. Effects of human land use on the terrestrial and aquatic sources of fluvial organic matter in a temperate river basin (The Meuse River, Belgium). *Biogeochemistry*, 136(2):191–211, 2017. ISSN 0168-2563, 1573-515X. doi: 10.1007/s10533-017-0387-9.

M. Landa, M. T. Cottrell, D. L. Kirchman, S. Blain, and I. Obernosterer. Changes in Bacterial Diversity in Response to Dissolved Organic Matter Supply in a Continuous Culture Experiment. *Aquatic Microbial Ecology*, 69(2):157–168, May 2013. ISSN 0948-3055, 1616-1564. doi: 10.3354/ame01632.

R. J. Lara, V. Rachold, G. Kattner, H. W. Hubberten, G. Guggenberger, A. Skoog, and D. N. Thomas. Dissolved organic matter and nutrients in the Lena River, Siberian Arctic: Characteristics

and distribution. *Marine Chemistry*, 59(3):301–309, Jan. 1998. ISSN 0304-4203. doi: 10.1016/S0304-4203(97)00076-5.

U. Larsson and A. Hagstrom. Phytoplankton exudate release as an energy source for the growth of pelagic bacteria. *Marine Biology*, 52:199–206, 1979. doi: 10.1007/BF00398133.

S. Lê, J. Josse, and F. Husson. FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, 25(1):1–18, 2008. doi: 10.18637/jss.v025.i01.

J. A. Leach, A. Larsson, M. B. Wallin, M. B. Nilsson, and H. Laudon. Twelve year interannual and seasonal variability of stream carbon export from a boreal peatland catchment: PEATLAND STREAM CARBON EXPORT. *Journal of Geophysical Research: Biogeosciences*, 121(7):1851–1866, July 2016. ISSN 21698953. doi: 10.1002/2016JG003357.

J. L. Lechuga-Crespo, S. Sauvage, E. Ruiz-Romera, C. George, and J. M. Sánchez-Pérez. SWATLitho: A hydrogeochemical model to estimate daily geochemical loads at the catchment scale. *Environmental Modelling & Software*, 135:104893, Jan. 2021. ISSN 1364-8152. doi: 10.1016/j.envsoft.2020.104893.

D. Liu, Y. Bai, X. He, C.-T. A. Chen, T.-H. Huang, D. Pan, X. Chen, D. Wang, and L. Zhang. Changes in riverine organic carbon input to the ocean from mainland China over the past 60 years. *Environment International*, 134:105258, Jan. 2020. ISSN 01604120. doi: 10.1016/j.envint.2019.105258.

E. Lloret, C. Dessert, L. Pastor, E. Lajeunesse, O. Crispi, J. Gaillardet, and M. Benedetti. Dynamic of particulate and dissolved organic carbon in small volcanic mountainous tropical watersheds. *Chemical Geology*, 351:229–244, Aug. 2013. ISSN 00092541. doi: 10.1016/j.chemgeo.2013.05.023.

J. M. Lobbes, H. P. Fitznar, and G. Kattner. Biogeochemical characteristics of dissolved and particulate organic matter in Russian rivers entering the Arctic Ocean. *Geochimica et Cosmochimica Acta*, 64(17):2973–2983, 2000. ISSN 00167037. doi: 10.1016/S0016-7037(00)00409-9.

C. Lønborg, K. Davidson, X. A. Álvarez–Salgado, and A. E. Miller. Bioavailability and bacterial degradation rates of dissolved organic matter in a temperate coastal area during an annual cycle. *Marine Chemistry*, 113(3-4):219–226, 2009. ISSN 03044203. doi: 10.1016/j.marchem.2009.02.003.

S. Loos, C. M. Shin, J. Sumihar, K. Kim, J. Cho, and A. H. Weerts. Ensemble data assimilation

methods for improving river water quality forecasting accuracy. *Water Research*, 171:115343, Mar. 2020. ISSN 0043-1354. doi: 10.1016/j.watres.2019.115343.

Y. Lu, J. E. Bauer, E. A. Canuel, Y. Yamashita, R. M. Chambers, and R. Jaffé. Photochemical and microbial alteration of dissolved organic matter in temperate headwater streams associated with different land use. *Journal of Geophysical Research: Biogeosciences*, 118(2):566–580, 2013. ISSN 2169-8961. doi: 10.1002/jgrg.20048.

W. Ludwig, J.-L. Probst, and S. Kempe. Predicting the oceanic input of organic carbon by continental erosion. *Global Biogeochemical Cycles*, 10(1):23–41, 1996. ISSN 1944-9224. doi: 10.1029/95GB02925.

M. T. Luu, T. D. Dinh, D. A. Trinh, and N. T. Doc. Water quality in an urbanized river basin impacted by multi-pollution sources: From comprehensive surveys to modelling. *ScienceAsia*, 47 (1):86, 2021. ISSN 1513-1874. doi: 10.2306/scienceasia1513-1874.2021.014.

J. Q. Mao, J. H. W. Lee, and K. W. Choi. The extended Kalman filter for forecast of algal bloom dynamics. *Water Research*, 43(17):4214–4224, Sept. 2009. ISSN 0043-1354. doi: 10.1016/j.watres. 2009.06.012.

A. Marescaux, V. Thieu, A. Borges, and J. Garnier. Seasonal and spatial variability of the partial pressure of carbon dioxide in the human-impacted seine river in france. *Scientific Reports*, 8(1): 13961, 2018. doi: 10.1038/s41598-018-32332-2.

A. Marescaux, V. Thieu, N. Gypens, M. Silvestre, and J. Garnier. Modeling inorganic carbon dynamics in the Seine River continuum in France. *Hydrology and Earth System Sciences*, 24(5):2379–2398, May 2020. ISSN 1607-7938. doi: 10.5194/hess-24-2379-2020.

J. M. Martin, D. M. Guan, F. Elbaz-Poulichet, A. J. Thomas, and V. V. Gordeev. Preliminary assessment of the distributions of some trace elements (As, Cd, Cu, Fe, Ni, Pb and Zn) in a pristine aquatic environment: The Lena River estuary (Russia). *Marine Chemistry*, 43(1):185–199, July 1993. ISSN 0304-4203. doi: 10.1016/0304-4203(93)90224-C.

B. Martin-Mousset, J. P. Croue, E. Lefebvre, and B. Legube. Distribution et caractérisation de la matière organique dissoute d'eaux naturelles de surface. *Water Research*, 31(3):541–553, 1997. ISSN 0043-1354. doi: 10.1016/S0043-1354(96)00259-X.

F. O. Masese, J. S. Salcedo-Borda, G. M. Gettel, K. Irvine, and M. E. McClain. Influence of catchment land use and seasonality on dissolved organic matter composition and ecosystem metabolism in

headwater streams of a Kenyan river. *Biogeochemistry*, 132(1):1–22, Jan. 2017. ISSN 1573-515X. doi: 10.1007/s10533-016-0269-6.

S. Matallana-Surget and R. Wattiez. Impact of Solar Radiation on Gene Expression in Bacteria. *Proteomes*, 1(2):70–86, July 2013. ISSN 2227-7382. doi: 10.3390/proteomes1020070.

L. Mayer, K. Thornton, and L. Schick. Bioavailability of organic matter photodissolved from coastal sediments. *Aquatic Microbial Ecology*, 64(3):275–284, Sept. 2011. ISSN 0948-3055, 1616-1564. doi: 10.3354/ame01530.

L. M. Mayer. Sedimentary organic matter preservation: An assessment and speculative synthesis—a comment. *Marine Chemistry*, 49(2):123–126, Apr. 1995. ISSN 0304-4203. doi: 10.1016/0304-4203(95)00011-F.

W. H. McDowell, A. Zsolnay, J. A. Aitkenhead-Peterson, E. G. Gregorich, D. L. Jones, D. Jödemann, K. Kalbitz, B. Marschner, and D. Schwesig. A comparison of methods to determine the biodegradable dissolved organic carbon from different terrestrial sources. *Soil Biology and Biochemistry*, 38 (7):1933–1942, July 2006. ISSN 0038-0717. doi: 10.1016/j.soilbio.2005.12.018.

M. Meybeck. Carbon, nitrogen, and phosphorus transport by world rivers. *Am. J. Sci.; (United States)*, 282:4:401–450, 1982. doi: 10.2475/ajs.282.4.401.

C. Minaudo, F. Curie, Y. Jullian, N. Gassama, and F. Moatar. QUAL-NET, a high temporal-resolution eutrophication model for large hydrographic networks. *Biogeosciences*, 15(7):2251–2269, Apr. 2018. ISSN 1726-4170. doi: 10.5194/bg-15-2251-2018.

F. Moatar, B. W. Abbott, C. Minaudo, F. Curie, and G. Pinay. Elemental properties, hydrology, and biology interact to shape concentration-discharge curves for carbon, nutrients, sediment, and major ions. *Water Resources Research*, 53(2):1270–1287, Feb. 2017. ISSN 0043-1397, 1944-7973. doi: 10.1002/2016WR019635.

J. Monod. The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1):371–394, 1949. doi: 10.1146/annurev.mi.03.100149.002103.

M. A. Moran and R. G. Zepp. Role of photoreactions in the formation of biologically labile compounds from dissolved organic matter. *Limnology and Oceanography*, 42(6):1307–1316, 1997. ISSN 1939-5590. doi: 10.4319/lo.1997.42.6.1307.

P. J. Mulholland. Large-Scale patterns in dissolved organic carbon concentration, flux and sources. In *Aquatic Ecosystems: Interactivity of Dissolved Organic Matter*, pages 139–159. Academic Press, 2003. ISBN 978-0-12-256371-3.

K. Muylaert, R. Dasseville, L. De Brabandere, F. Dehairs, and W. Vyverman. Dissolved organic carbon in the freshwater tidal reaches of the Schelde estuary. *Estuarine, Coastal and Shelf Science*, 64(4):591–600, 2005. ISSN 02727714. doi: 10.1016/j.ecss.2005.04.010.

H. T. M. Nguyen, G. Billen, J. Garnier, T. P. Q. Le, Q. L. Pham, S. Huon, and E. Rochelle-Newall. Organic carbon transfers in the subtropical Red River system (Viet Nam): Insights on CO2 sources and sinks. *Biogeochemistry*, 138(3):277–295, 2018. ISSN 0168-2563, 1573-515X. doi: 10.1007/s10533-018-0446-x.

T. T. Nguyen, J. Némery, N. Gratiot, E. Strady, V. Q. Tran, A. T. Nguyen, J. Aimé, and A. Peyne. Nutrient dynamics and eutrophication assessment in the tropical river system of Saigon – Dongnai (southern Vietnam). *Science of The Total Environment*, 653:370–383, Feb. 2019. ISSN 00489697. doi: 10.1016/j.scitotenv.2018.10.319.

P. Njapou. Indicateurs metaboliques operationnels pour le suivi de la qualité de la Seine par la mesure haute frequence. PFE report M2, Université Paris-Saclay, Paris, 2022.

J. Nossent, P. Elsen, and W. Bauwens. Sobol' sensitivity analysis of a complex environmental model. *Environmental Modelling & Software*, 26(12):1515–1525, 2011. ISSN 13648152. doi: 10.1016/j.envsoft.2011.08.010.

H. T. Odum. Primary Production in Flowing Waters1. *Limnology and Oceanography*, 1(2):102–117, 1956. ISSN 00243590. doi: 10.4319/lo.1956.1.2.0102.

D. Olefeldt, N. Roulet, R. Giesler, and A. Persson. Total waterborne carbon export and DOC composition from ten nested subarctic peatland catchments-importance of peatland cover, groundwater influence, and inter-annual variability of precipitation patterns: WATERBORNE CARBON EXPORT FROM SUBARCTIC CATCHMENTS. *Hydrological Processes*, 27(16):2280–2294, July 2013. ISSN 08856087. doi: 10.1002/hyp.9358.

T. Page, P. J. Smith, K. J. Beven, I. D. Jones, J. A. Elliott, S. C. Maberly, E. B. Mackay, M. De Ville, and H. Feuchtmayr. Adaptive forecasting of phytoplankton communities. *Water Research*, 134:74–85, May 2018. ISSN 0043-1354. doi: 10.1016/j.watres.2018.01.046.

J. Pagès. Analyse factorielle de données mixtes. *REVUE DE STATISTIQUE APPLIQUÉE*, 52(4): 93–111, 2004.

J. Pagès. *Multiple factor analysis by example using R*. CRC Press, 2014. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85053979425&partnerID=40&md5=65969f850c0da1e91aeb80113f449612. Cited by: 113.

D. Palma, M. Sleiman, O. Voldoire, A. Beauger, E. Parlanti, and C. Richard. Study of the dissolved organic matter (DOM) of the Auzon cut-off meander (Allier River, France) by spectral and photoreactivity approaches. *Environmental Science and Pollution Research*, 27(21):26385–26394, July 2020. ISSN 0944-1344, 1614-7499. doi: 10.1007/s11356-020-09005-7.

H. Park, S. S. A. Shah, G. Korshin, I. Angelidaki, and K.-H. Choo. The impact of sunlight on fouling behaviors and microbial communities in membrane bioreactors. *Journal of Membrane Science*, 672: 121443, Apr. 2023. ISSN 03767388. doi: 10.1016/j.memsci.2023.121443.

R. Park, R. O'Neill, J. Bloomfield, H. Shugart, R. Booth, J. Koonce, M. Adams, L. Clesceri, E. Colon, E. Dettman, R. Goldstein, J. Hoopes, D. Huff, S. Katz, J. Kitchell, R. Kohberger, E. LaRow, D. McNaught, J. Peterson, and C. Zahorcak. A generalized model for simulating lake ecosystems. *Simulation*, 23:30–50, 01 1974.

S. S. Park and Y. S. Lee. A water quality modeling study of the nakdong river, korea. *Ecol. Model.*, 152:65–75, 2002.

E. Parlanti, K. Wörz, L. Geoffroy, and M. Lamotte. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. *Organic Geochemistry*, 31(12):1765–1781, Dec. 2000. ISSN 0146-6380. doi: 10.1016/S0146-6380(00)00124-8.

E. Parlanti, B. Morin, and L. Vacher. Combined 3D-spectrofluorometry, high performance liquid chromatography and capillary electrophoresis for the characterization of dissolved organic matter in natural waters. *Organic Geochemistry*, 33(3):221–236, 2002. doi: 10.1016/S0146-6380(01)00154-1.

M. Pettine, L. Patrolecco, M. Camusso, and S. Crescenzio. Transport of Carbon and Nitrogen to the Northern Adriatic Sea by the Po River. *Estuarine, Coastal and Shelf Science*, 46(1):127–142, 1998. ISSN 02727714. doi: 10.1006/ecss.1997.0303.

E. Polus, C. de Fouquet, N. Flipo, and M. Poulin. Caractérisation spatiale et temporelle de la qualité

des « Masses d'Eau Cours d'Eau ». *Revue des sciences de l'eau*, 23(4):415–429, Dec. 2010. ISSN 1718-8598, 0992-7158. doi: 10.7202/045101ar.

E. Polus, N. Flipo, C. de Fouquet, and M. Poulin. Geostatistics for assessing the efficiency of a distributed physically-based water quality model: Application to nitrate in the Seine River. *Hydrological Processes*, 25(2):217–233, Jan. 2011. ISSN 08856087. doi: 10.1002/hyp.7838.

P. Porcal, P. J. Dillon, and L. A. Molot. Temperature Dependence of Photodegradation of Dissolved Organic Matter to Dissolved Inorganic Carbon and Particulate Organic Carbon. *PLOS ONE*, 10 (6):e0128884, June 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0128884.

A. Pryet, B. Labarthe, F. Saleh, M. Akopian, and N. Flipo. Reporting of stream-aquifer flow distribution at the regional scale with a distributed process-based model. *Water Resources Management*, 29:139–159, 2015. doi: 10.1007/s11269-014-0832-7.

W. Qi. Organic micropollutants in the Yangtze River: Seasonal occurrence and annual loads. *Science of the Total Environment*, 472:789–799, 2014.

R. G. Qualls. Biodegradability of Fractions of Dissolved Organic Carbon Leached from Decomposing Leaf Litter. *Environmental Science & Technology*, 39(6):1616–1622, Mar. 2005. ISSN 0013-936X, 1520-5851. doi: 10.1021/es049090o.

R. G. Qualls and B. L. Haines. s1_Biodegradability of Dissolved Organic Matter in Forest Throughfall, Soil Solution, and Stream Water. *Soil Science Society of America Journal*, 56(2):578–586, 1992. ISSN 03615995. doi: 10.2136/sssaj1992.03615995005600020038x.

J. N. Quinton, G. Govers, K. Van Oost, and R. D. Bardgett. The impact of agricultural soil erosion on biogeochemical cycling. *Nature Geoscience*, 3(5):311–314, May 2010. ISSN 1752-0908. doi: 10.1038/ngeo838.

L. T. P. Quynh, G. Garnier, G. Billen, L. N. Thiem, and C. V. Minh. Preminary results of riverstrahler model application to the red river system (vietnam). *Vietnam Journal of Chemistry*, 47(1):110, 2014. doi: 10.15625/4550.

M. Raimonet, L. Vilmin, N. Flipo, V. Rocher, and A. Laverman. Modelling the fate of nitrite in an urbanized river using experimentally obtained nitrifier growth parameters. *Water Research*, 73: 373–387, 2015. doi: 10.1016/j.watres.2015.01.026.

L. Ran. Spatial and seasonal variability of organic carbon transport in the Yellow River, China. *Journal of Hydrology*, 498:76–88, 2013.

P. A. Raymond, J. Hartmann, R. Lauerwald, S. Sobek, C. McDonald, M. Hoover, D. Butman, R. Striegl, E. Mayorga, C. Humborg, P. Kortelainen, H. Dürr, M. Meybeck, P. Ciais, and P. Guth. Global carbon dioxide emissions from inland waters. *Nature*, 503(7476):355–359, Nov. 2013. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature12760.

P. Reichert, D. Borchardt, M. Henze, W. Rauch, P. Shanahan, L. Somlyódy, and P. Vanrolleghem. River Water Quality Model no. 1 (RWQM1): II. Biochemical process equations. *Water Science and Technology*, 43(5):11–30, Mar. 2001. ISSN 0273-1223, 1996-9732. doi: 10.2166/wst.2001.0241.

P. Reichert, U. Uehlinger, and V. Acuña. Estimating stream metabolism from oxygen concentrations: Effect of spatial heterogeneity. *Journal of Geophysical Research*, 114(G3):G03016, Sept. 2009. ISSN 0148-0227. doi: 10.1029/2008JG000917.

S. Reschke, V. Ittekkot, and N. Panin. The Nature of Organic Matter in the Danube River Particles and North-western Black Sea Sediments. *Estuarine, Coastal and Shelf Science*, 54(3):563–574, 2002. ISSN 02727714. doi: 10.1006/ecss.2000.0665.

D. E. Reusser, W. Buytaert, and E. Zehe. Temporal dynamics of model parameter sensitivity for computationally expensive models with the fourier amplitude sensitivity test. *Water Resources Research*, 47(7), 2011. ISSN 1944-7973. doi: 10.1029/2010WR009947.

F. Ribas, J. Frías, J. Huguet, and F. Lucena. Efficiency of various water treatment processes in the removal of biodegradable and refractory organic matter. *Water Research*, 31(3):639–649, 1997. ISSN 00431354. doi: 10.1016/S0043-1354(96)00301-6.

F. Rizinjirabake, A. M. Abdi, D. E. Tenenbaum, and P. Pilesjö. Riverine dissolved organic carbon in Rukarara River Watershed, Rwanda. *Science of The Total Environment*, 643:793–806, Dec. 2018. ISSN 00489697. doi: 10.1016/j.scitotenv.2018.06.194.

T. Rodríguez-Castillo, E. Estévez, A. M. González-Ferreras, and J. Barquín. Estimating Ecosystem Metabolism to Entire River Networks. *Ecosystems*, 22(4):892–911, June 2019. ISSN 1432-9840, 1435-0629. doi: 10.1007/s10021-018-0311-8.

H. J. Rogers. The Dissimilation of High Molecular Weight Substances. In *The Bacteria*, volume 2, pages 261–318. Academic Press, i. c. gunsalus and r. y. stanier edition, 1961.

D. Ruelland, G. Billen, D. Brunstein, and J. Garnier. SENEQUE: A multi-scaling GIS interface to the Riverstrahler model of the biogeochemical functioning of river systems. *Science of The Total Environment*, 375(1):257–273, Apr. 2007. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2006.12.014.

A. Ruggiero, A. G. Solimini, and G. Carchini. Effects of a waste water treatment plant on organic matter dynamics and ecosystem functioning in a mediterranean stream. *Ann. Limnol. - Int. J. Lim.*, 42(2):97–107, 2006. doi: 10.1051/limn/2006014.

A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, 2002. ISSN 00104655. doi: 10.1016/S0010-4655(02)00280-1.

A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010. ISSN 00104655. doi: 10.1016/j.cpc.2009.09.018.

H. N. Schiebel, X. Wang, R. F. Chen, and F. Peri. Photochemical Release of Dissolved Organic Matter from Resuspended Salt Marsh Sediments. *Estuaries and Coasts*, 38(5):1692–1705, 2015. ISSN 1559-2723.

M. W. I. Schmidt, M. S. Torn, S. Abiven, T. Dittmar, G. Guggenberger, I. A. Janssens, M. Kleber, I. Kögel-Knabner, J. Lehmann, D. A. C. Manning, P. Nannipieri, D. P. Rasse, S. Weiner, and S. E. Trumbore. Persistence of soil organic matter as an ecosystem property. *Nature*, 478(7367):49–56, Oct. 2011. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature10386.

P. Segatto, T. Battin, and E. Bertuzzo. Modeling the coupled dynamics of stream metabolism and microbial biomass. *Limnology and Oceanography*, 65(7):1573–1593, 2020. doi: 10.1002/lno.11407.

M. Seidl. *Caractérisation des rejets urbains de temps de pluie et de leurs impacts sur l'oxygénation de la Seine*. Theses, Ecole Nationale des Ponts et Chaussées, Sept. 1997. URL https://pastel.archives-ouvertes.fr/tel-00523123.

M. Seidl, P. Servais, M. Martaud, C. Gandouin, and J. M. Mouchel. Organic carbon biodegradability and heterotrophic bacteria along a combined sewer catchment during rain events. *Water Science and Technology*, 37(1):25–33, 1998a. ISSN 0273-1223. doi: 10.1016/S0273-1223(97)00752-X.

M. Seidl, P. Servais, and J. Mouchel. Organic matter transport and degradation in the river Seine (France) after a combined sewer overflow. *Water Research*, 32(12):3569–3580, 1998b. ISSN 00431354. doi: 10.1016/S0043-1354(98)00169-9.

H. Seki, T. NAKAI, and H. OTOBE. REGIONAL DIFFERENCES ON TURNOVER RATE OF DISSOLVED MATERIALS IN THE PACIFIC OCEAN AT SUMMER OF 1971. *Arch. Hydrobiol*, 71:79–89, 1972.

H. Seki, T. Nakai, and H. Otobe. Turnover rate of dissolved materials in the Philippine Sea at winter of 1973. *Archiv für Hydrobiologie*, 73(2):238–244, 1974. ISSN 0003-9136. doi: 10.1127/archiv-hydrobiol/73/1974/238.

H. Seki, SEKI. H, YAMAGUCHI. Y, and ICHIMURA. S. TURNOVER RATE OF DISSOLVED ORGANIC MATERIALS IN A COASTAL REGION OF JAPAN AT SUMMER STAGNATION PERIOD OF 1974. *Arch. Hydrobiol. Beih.*, 75:297–305, 1975.

R. Sempéré, B. Charrière, F. Van Wambeke, and G. Cauwet. Carbon inputs of the Rhône River to the Mediterranean Sea: Biogeochemical implications. *Global Biogeochemical Cycles*, 14(2):669–681, 2000. ISSN 08866236. doi: 10.1029/1999GB900069.

A. B. J. Sepers. The utilization of dissolved organic compounds in aquatic environments. *Hydrobiologia*, 52(1):39–54, 1977. ISSN 0018-8158, 1573-5117. doi: 10.1007/BF02658081.

P. Servais. *Etude de la dégradation de la matière organique par les bactéries hétérotrophes en rivière. Développement d'une démarche méthodologique et application à la Meuse belge.* PhD thesis, Université Libre de Bruxelles, Faculté des Sciences, 1986.

P. Servais and G. Billen. Note sur le calcul des apports ponctuels à prendre en compte dans les modèles ProSe et Sénèque à partir des données disponibles sur les rejets de STEPs. Technical report, PIREN Seine, 2009.

P. Servais and J. Garnier. Contribution of heterotrophic bacterial production to the carbon budget of the river Seine (France). *Microbial Ecology*, 25(1):19–33, 1993. ISSN 0095-3628, 1432-184X. doi: 10.1007/BF00182127.

P. Servais and J. Garnier. Organic carbon and bacterial heterotrophic activity in the maximum turbidity zone of the Seine estuary (France). *Aquatic Sciences*, 68(1):78–85, 2006. ISSN 1015-1621, 1420-9055. doi: 10.1007/s00027-005-0809-y.

P. Servais, G. Billen, and M.-C. Hascoet. DETERMINATION OF THE BIODEGRADABLE FRACTION OF DISSOLVED ORGANIC MATTER IN WATERS. *Water Research*, 21(4):445–450, 1987. doi: 10.1016/0043-1354(87)90192-8.

P. Servais, G. Billen, J. Martinez, and J. Vives-Rego. Estimating bacterial mortality by the disappearance of ^{3} H-labeled intracellular DNA. *FEMS Microbiology Letters*, 62(2):119–125, 1989. ISSN 03781097, 15746968. doi: 10.1111/j.1574-6968.1989.tb03664.x.

P. Servais, A. Barillier, and J. Garnier. Determination of the biodegradable fraction of dissolved and particulate organic carbon in waters. *Annales de Limnologie - International Journal of Limnology*, 31(1):75–80, 1995. ISSN 0003-4088. doi: 10.1051/limn/1995005.

P. Servais, G. Billen, J. Garnier, Z. Idlafkih, J. Mouchel, M. Seidl, and M. Meybeck. Carbone organique : origine et biodégradabilité. In *La Seine en son bassin*, pages 483–525. Elsevier, 1998.

P. Servais, J. Garnier, N. Demarteau, N. Brion, and G. Billen. Supply of organic matter and bacteria to aquatic ecosystems through waste water effluents. *Water Research*, 33(16):3521–3531, 1999. ISSN 0043-1354. doi: 10.1016/S0043-1354(99)00056-1.

H. Seyedhashemi, J.-P. Vidal, J. S. Diamond, D. Thiéry, C. Monteil, F. Hendrickx, A. Maire, and F. Moatar. Regional, multi-decadal analysis on the Loire River basin reveals that stream temperature increases faster than air temperature. *Hydrology and Earth System Sciences*, 26(9):2583–2603, May 2022. ISSN 1607-7938. doi: 10.5194/hess-26-2583-2022.

A. Sferratore, G. Billen, J. Garnier, E. Smedberg, C. Humborg, and L. Rahm. Modelling nutrient fluxes from sub-arctic basins: Comparison of pristine vs. dammed rivers. *Journal of Marine Systems*, 73(3-4):236–249, 2008. doi: 10.1016/j.jmarsys.2007.10.012.

P. Shang, Y. Lu, Y. Du, R. Jaffé, R. H. Findlay, and A. Wynn. Climatic and watershed controls of dissolved organic matter variation in streams across a gradient of agricultural land use. *Science of The Total Environment*, 612:1442–1453, 2018. ISSN 00489697. doi: 10.1016/j.scitotenv.2017.08.322.

Y. Shin, E.-J. Lee, Y.-J. Jeon, J. Hur, and N.-H. Oh. Hydrological changes of DOM composition and biodegradability of rivers in temperate monsoon climates. *Journal of Hydrology*, 540:538–548, 2016. ISSN 00221694. doi: 10.1016/j.jhydrol.2016.06.004.

I. M. Sobol. Sensitivity Estimates for Nonlinear Mathematical Models. *Mathematical modelling and computational experiments*, 1(4):407–414, 1993.

R. G. M. Spencer, P. J. Hernes, R. Ruf, A. Baker, R. Y. Dyda, A. Stubbins, and J. Six. Temporal controls on dissolved organic matter and lignin biogeochemistry in a pristine tropical river, Democratic

Republic of Congo. *Journal of Geophysical Research*, 115(G3):G03013, Aug. 2010. ISSN 0148-0227. doi: 10.1029/2009JG001180.

A. Spitzy and V. Ittekkot. Dissolved and particulate organic matter in Rivers. In *Ocean Margin Processes in Global Change*, pages 5–17. John Wiley, Chichester, 1991.

P. Staehr, D. Bade, M. van de Bogert, G. Koch, C. Williamson, P. Hanson, J. Cole, and T. Kratz. Lake metabolism and the diel oxygen technique: State of the science. *Limnology and Oceanography: Methods*, 8(NOV):628–644, 2010. doi: 10.4319/lom.2010.8.0628.

E. H. Stanley, N. J. Casson, S. T. Christel, J. T. Crawford, L. C. Loken, and S. K. Oliver. The ecology of methane in streams and rivers: Patterns, controls, and global significance. *Ecological Monographs*, 86(2):146–171, 2016. ISSN 1557-7015. doi: 10.1890/15-1027.

C. A. Stedmon, S. Markager, and R. Bro. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Marine Chemistry*, 82(3):239–254, Aug. 2003. ISSN 0304-4203. doi: 10.1016/S0304-4203(03)00072-0.

H. Streeter and E. B. Phelps. A study of the pollution and natural purification of the Ohio River. Technical Report 146, U.S. Public Health Service, Treasury Department, Washington DC, 1925.

M. Takahashi and S. Ichimura. Glucose uptake in ocean profiles with special reference to temperature. *Marine Biology*, 11(3):206–213, 1971. ISSN 0025-3162, 1432-1793. doi: 10.1007/BF00401269.

J. L. Tank, E. J. Rosi-Marshall, N. A. Griffiths, S. A. Entrekin, and M. L. Stephen. A review of allochthonous organic matter dynamics and metabolism in streams. *Journal of the North American Benthological Society*, 29(1):118–146, Feb. 2010. ISSN 0887-3593. doi: 10.1899/08-170.1.

L. Thibodeaux, M. Poulin, and S. Even. A model for enhanced aeration of streams by motor vessels with application to the Seine river. *Journal of Hazardous Materials*, 37(3):459–473, June 1994. ISSN 0304-3894. doi: 10.1016/0304-3894(93)E0101-7.

V. Thieu, G. Billen, and J. Garnier. Nutrient transfer in three contrasting NW European watersheds: The Seine, Somme, and Scheldt Rivers. A comparative application of the Seneque/Riverstrahler model. *Water Research*, 43(6):1740–1754, 2009. ISSN 00431354. doi: 10.1016/j.watres.2009.01.014.

V. Thieu, J. Garnier, and G. Billen. Assessing the effect of nutrient mitigation measures in the watersheds of the Southern Bight of the North Sea. *Science of The Total Environment*, 408(6): 1245–1255, Feb. 2010. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2009.12.031.

V. Thieu, G. Billen, J. Garnier, and M. Benoît. Nitrogen cycling in a hypothetical scenario of generalised organic agriculture in the Seine, Somme and Scheldt watersheds. *Regional Environmental Change*, 11(2):359–370, June 2011. ISSN 1436-378X. doi: 10.1007/s10113-010-0142-4.

L. J. Tranvik, J. J. Cole, and Y. T. Prairie. The study of carbon in inland waters—from isolated ecosystems to players in the global carbon cycle. *Limnology and Oceanography Letters*, 3(3):41–48, June 2018. ISSN 2378-2242, 2378-2242. doi: 10.1002/lol2.10068.

Š. Trulleyová and M. Rulík. Determination of biodegradable dissolved organic carbon in waters: Comparison of batch methods. *Science of The Total Environment*, 332(1):253–260, Oct. 2004. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2004.04.018.

M.-H. Tusseau-Vuillemin, J. Garnier, P. Servais, and L. Laroche. Charges domestiques spécifiques et rejets de station d'épuration. Technical report, PIREN-Seine, 2002.

M.-H. Tusseau-Vuillemin, J. Dispan, J.-M. Mouchel, and P. Servais. Biodegradable fraction of organic carbon estimated under oxic and anoxic conditions. *Water Research*, 37(9):2242–2247, 2003. ISSN 00431354. doi: 10.1016/S0043-1354(02)00582-1.

U. Uehlinger, B. Kawecka, and C. T. Robinson. Effects of experimental floods on periphyton and stream metabolism below a high dam in the swiss alps river spol). *Aquatic Sciences - Research Across Boundaries*, 65(3):199–209, Sept. 2003. ISSN 1015-1621, 1420-9055. doi: 10.1007/s00027-003-0664-7.

R. F. Vaccaro and H. W. Jannasch. STUDIES ON HETEROTROPHIC ACTIVITY IN SEAWATER BASED ON GLUCOSE ASSIMILATION1: GLUCOSE ASSIMILATION IN SEAWATER. *Limnology and Oceanography*, 11(4):596–607, 1966. ISSN 00243590. doi: 10.4319/lo.1966.11.4.0596.

R. F. Vaccaro and H. W. Jannasch. Variations in Uptake Kinetics for Glucose by Natural Populations in Seawater. *Limnology and Oceanography*, 12(3):540–542, 1967.

R. Vannote, G. W. Minshall, K. Cummins, J. Sedell, and C. Cushing. The River Continuum Concept. *Can. J. Fish. Aquat. Sci.*, 37(1):130–137, 1980. doi: 10.1139/f80-017.

P. Vanrolleghem, D. Borchardt, M. Henze, W. Rauch, P. Reichert, P. Shanahan, and L. Somlyódy. River Water Quality Model no.1 (RWQM1): III biochemical submodel selection. *Water Science and Technology*, 43(5):31–40, 2001.

G. Varrault, P. T. Nguyen, Z. Matar, and C. Bonnot. La matière organique dans le bassin de la Seine : Variabilité, sources et influence sur la spéciation des micropolluants. In *Biogéochimie de l'axe Fluvial*, Rapport de Synthèse 2011 – 2015, pages 84–129. PIREN-Seine, rapport de synthèse 2011 – 2015 edition, 2016.

G. Varrault, E. Parlanti, Z. Matar, J. Garnier, P. T. Nguyen, S. Derenne, V. Rocher, B. Muresan, Y. Louis, C. Soares-Pereira, A. Goffin, M. F. Benedetti, A. Bressy, A. Gelabert, Y. Guo, and M.-A. Cordier. Aquatic Organic Matter in the Seine Basin: Sources, Spatio-Temporal Variability, Impact of Urban Discharges and Influence on Micro-pollutant Speciation. In N. Flipo, P. Labadie, and L. Lestel, editors, *The Seine River Basin*, The Handbook of Environmental Chemistry, pages 217–242. Springer International Publishing, Cham, 2020. ISBN 978-3-030-54260-3. doi: 10.1007/698_2019_383.

L. Vilmin. *Modélisation Du Fonctionnement Biogéochimique de La Seine de l'agglomération Parisienne à l'estuaire à Différentes Échelles Temporelles.* These de doctorat, Paris, ENMP, 2014.

L. Vilmin, N. Flipo, and M. Poulin. Le modèle de simulation biogéochimique C-RIVE. Technical report, PIREN-Seine, 2012.

L. Vilmin, N. Flipo, C. de Fouquet, and M. Poulin. Pluri-annual sediment budget in a navigated river system: The Seine River (France). *Science of The Total Environment*, 502:48–59, 2015. ISSN 00489697. doi: 10.1016/j.scitotenv.2014.08.110.

L. Vilmin, N. Flipo, N. Escoffier, V. Rocher, and A. Groleau. Carbon fate in a large temperate human-impacted river system: Focus on benthic dynamics. *Global Biogeochemical Cycles*, 30(7): 1086–1104, 2016. ISSN 08866236. doi: 10.1002/2015GB005271.

L. Vilmin, N. Flipo, N. Escoffier, and A. Groleau. 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, Aug. 2018. ISSN 0944-1344, 1614-7499. doi: 10.1007/s11356-016-7109-z.

C. J. Volk, C. B. Volk, and L. A. Kaplan. Chemical composition of biodegradable dissolved organic matter in streamwater. *Limnology and Oceanography*, 42(1):39–44, 1997. ISSN 00243590. doi: 10.4319/lo.1997.42.1.0039.

Q. Wang, S. Li, P. Jia, C. Qi, and F. Ding. A Review of Surface Water Quality Models. *The Scientific World Journal*, 2013:e231768, June 2013. ISSN 2356-6140. doi: 10.1155/2013/231768.

S. Wang. *Simulation Du Métabolisme de La Seine Par Assimilation de Données En Continu*. These de doctorat, Paris Sciences et Lettres, 2019.

S. Wang. Dynamique des producteurs primaires dans le modèle de biogéochimie aquatique RIVE unifié. Rapport 2021, Sorbonne Université, 2021.

S. Wang, N. Flipo, and T. Romary. Time-dependent global sensitivity analysis of the C-RIVE biogeochemical model in contrasted hydrological and trophic contexts. *Water Research*, 144:341–355, 2018. ISSN 00431354. doi: 10.1016/j.watres.2018.07.033.

S. Wang, N. Flipo, and T. Romary. Oxygen data assimilation for estimating micro-organism communities' parameters in river systems. *Water Research*, 165:115021, 2019. ISSN 00431354. doi: 10.1016/j.watres.2019.115021.

S. Wang, N. Flipo, T. Romary, and M. Hasanyar. Particle filter for high frequency oxygen data assimilation in river systems. *Environmental Modelling & Software*, 151:105382, May 2022a. ISSN 1364-8152. doi: 10.1016/j.envsoft.2022.105382.

S. Wang, V. Thieu, M. Silvestre, G. Billen, J. Garnier, A. Marescaux, L. Weidenfeld, and N. Flipo. Modèle biogéochimique RIVE unifié pour simuler la qualité des eaux de surface continentale. Rapport 2022, Sorbonne Université, Paris, 2022b.

S. Wang, N. Flipo, and T. Romary. Which filter for data assimilation in water quality models? Focus on oxygen reaeration and heterotrophic bacteria activity. *Journal of Hydrology*, page 129423, Mar. 2023. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2023.129423.

X. Wang, X. Chen, S. Liu, and X. Ge. Effect of molecular weight of dissolved organic matter on toxicity and bioavailability of copper to lettuce. *Journal of Environmental Sciences*, 22(12):1960–1965, 2010. ISSN 10010742. doi: 10.1016/S1001-0742(09)60346-6.

T. N. Wiegner and S. P. Seitzinger. Seasonal bioavailability of dissolved organic carbon and nitrogen from pristine and polluted freshwater wetlands. *Limnology and Oceanography*, 49(5):1703–1712, 2004. ISSN 1939-5590. doi: 10.4319/lo.2004.49.5.1703.

T. N. Wiegner, R. L. Tubal, and R. A. MacKenzie. Bioavailability and export of dissolved organic matter from a tropical river during base- and stormflow conditions. *Limnology and Oceanography*, 54(4):1233–1242, 2009. ISSN 00243590. doi: 10.4319/lo.2009.54.4.1233.

C. K. Wikle and L. M. Berliner. A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, 230(1):1–16, June 2007. ISSN 0167-2789. doi: 10.1016/j.physd.2006.09.017.

S. Wold, K. Esbensen, and P. Geladi. Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3):37–52, 1987.

R. T. Wright and N. M. Shah. The trophic role of glycolic acid in coastal seawater. I. Heterotrophic metabolism in seawater and bacterial cultures. *Marine Biology*, 33(2):175–183, 1975. ISSN 0025-3162, 1432-1793. doi: 10.1007/BF00390723.

C.-P. Yang, W.-S. Lung, J.-T. Kuo, and J.-H. Liu. Water Quality Modeling of a Hypoxic Stream. *Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management*, 14(2):115–123, Apr. 2010. ISSN 1090-025X, 1944-8376. doi: 10.1061/(ASCE)HZ.1944-8376.0000021.

L. Yang, Q. Cheng, W.-E. Zhuang, H. Wang, and W. Chen. Seasonal changes in the chemical composition and reactivity of dissolved organic matter at the land-ocean interface of a subtropical river. *Environmental Science and Pollution Research*, 26(24):24595–24608, Aug. 2019. ISSN 0944-1344, 1614-7499. doi: 10.1007/s11356-019-05700-2.

X. Yang, T. M. Pavelsky, and G. H. Allen. The past and future of global river ice. *Nature*, 577(7788): 69–73, Jan. 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-019-1848-1.

R. G. Young, C. D. Matthaei, and C. R. Townsend. Organic matter breakdown and ecosystem metabolism: functional indicators for assessing river ecosystem health. *Journal of the North American Benthological Society*, 27(3):605 – 625, 2008.

V. Zahraeifard and Z. Deng. VART Model–Based Method for Estimation of Instream Dissolved Oxygen and Reaeration Coefficient. *Journal of Environmental Engineering*, 138(4):518–524, Apr. 2012. ISSN 0733-9372, 1943-7870. doi: 10.1061/(ASCE)EE.1943-7870.0000494.

A. Zettam, A. Taleb, S. Sauvage, L. Boithias, N. Belaidi, and J. M. Sanchez-Perez. Applications of a SWAT model to evaluate the contribution of the Tafna catchment (north-west Africa) to the nitrate load entering the Mediterranean Sea. *Environmental Monitoring and Assessment*, 192(8):510, July 2020. ISSN 1573-2959. doi: 10.1007/s10661-020-08482-0.

H. Zhang, Y. Zheng, X. C. Wang, Y. Wang, and M. Dzakpasu. Characterization and biogeochemical implications of dissolved organic matter in aquatic environments. *Journal of Environmental Management*, 294:113041, Sept. 2021. ISSN 0301-4797. doi: 10.1016/j.jenvman.2021.113041.

L. J. Zhang, L. Wang, W.-J. Cai, D. M. Liu, and Z. G. Yu. Impact of human activities on organic carbon transport in the Yellow River. *Biogeosciences*, 10(4):2513–2524, 2013. ISSN 1726-4189. doi: 10.5194/bg-10-2513-2013.

RÉSUMÉ

Cette recherche vise à améliorer l'évaluation du métabolisme des rivières en assimilant les données de l'oxygène dissous dans le modèle de qualité de l'eau ProSe-PA, en utilisant des études de cas synthétiques et réelles de la Seine. L'étude repose sur un travail antérieur en introduisant un nouveau schéma de répartition pour le carbone organique et en utilisant un cadre bayésien avec surveillance de l'oxygène dissous. La recherche montre que les incertitudes dans l'apport en carbone organique entraînent des écarts entre les niveaux d'oxygène dissous observés et simulés. L'étude révèle que les paramètres bactériens et de partitionnement du carbone organique sont les plus influents pour déterminer la qualité de l'eau. La performance du modèle ProSe-PA est évaluée dans des cas réels et synthétiques de la Seine, montrant des améliorations dans les simulations d'oxygène dissous et la quantification de la biodégradabilité du carbone organique pendant les périodes de faible débit.

MOTS CLÉS

modèle ProSe-PA, Assimilation de données, Qualité de l'eau, Métabolisme fluvial

ABSTRACT

The goal of this research is to improve the assessment of river metabolism by assimilating dissolved oxygen data in the ProSe-PA water quality model, using synthetic and real case studies of the Seine river. The study builds on previous work by introducing a new partitioning scheme for organic carbon and using a Bayesian framework with dissolved oxygen monitoring. The research shows that uncertainties in organic carbon inflow cause discrepancies in observed and simulated dissolved oxygen levels. The study finds that bacterial parameters and organic carbon partitioning parameters are the most influential in determining water quality. The ProSe-PA model's performance is evaluated in both real and synthetic cases of the Seine river, showing improvements in dissolved oxygen simulations and quantification of organic carbon biodegradability during low flow periods.

KEYWORDS

Prose-PA model, Water Quality, Data Assimilation, River metabolism