



HAL
open science

Particle filter for high frequency oxygen data assimilation in river systems

Shuaitao Wang, Nicolas Flipo, Thomas Romary, Masihullah Hasanyar

► **To cite this version:**

Shuaitao Wang, Nicolas Flipo, Thomas Romary, Masihullah Hasanyar. Particle filter for high frequency oxygen data assimilation in river systems. *Environmental Modelling and Software*, 2022, 151, pp.105382. 10.1016/j.envsoft.2022.105382 . hal-03635705

HAL Id: hal-03635705

<https://hal.science/hal-03635705v1>

Submitted on 22 Jul 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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 Particle filter for high frequency oxygen data assimilation in river systems

2 Shuaitao Wang^{a,1,*}, Nicolas Flipo^a, Thomas Romary^a, Masihullah Hasanyar^a

3 ^a*Geosciences and Geoengineering Department, MINES ParisTech, PSL University, 35 Rue Saint-Honoré 77300 Fontainebleau, France*

4 Abstract

5 High frequency water quality data measured by *in-situ* sensors allowed the development of auto-calibration methods
6 for water quality simulation programs. However, these methods consider static parameter values, which may be
7 unrealistic for microorganism activities. An alternative technique is to use data assimilation methods. This paper
8 presents a first application of a particle filter that assimilates dissolved oxygen (DO) data into the hydro-ecological
9 model ProSe. It demonstrates the capability of the approach for simulating DO concentrations and characterizing time-
10 varying physiological properties of living communities in contrasted trophic contexts. DO concentrations are better
11 estimated, especially during algal bloom when phytoplankton physiological parameters match the ones reported in
12 the literature. Despite the simulation capabilities related to phytoplankton, further improvements related to low flow
13 periods can still be achieved, especially concerning the heterotrophic bacteria properties as well as a finer description
14 of the biodegradable component of the organic matter flux at the system's boundaries.

15 *Keywords:* Data assimilation, Dissolved oxygen, Particle filter, Parameter estimation, Water quality modeling

16 Highlights

- 17 • A particle filtering method is applied for the first time in the Seine River system
- 18 • The filter improves significantly the simulation of oxygen and algae dynamics
- 19 • The filter can capture time-varying phytoplanktonic parameters during algal blooms
- 20 • The description of BDOM data need to be further investigated for low flow periods

21 Software availability

22 Name of software : ProSE-PA

*Corresponding author

Email address: shuaitaowang@outlook.com (Shuaitao Wang)

¹Current address: Sorbonne Université, CNRS, EPHE, UMR 7619 Metis, 4 place Jussieu 75005 Paris, France (shuaitao.wang@sorbonne-universite.fr)

Preprint submitted to Elsevier

March 18, 2022

23 Contact address: nicolas.flipo@mines-paristech.fr
24 Year first available: 2019
25 Program language: ANSI C
26 Operating system: Linux
27 Software access: Gitlab deposit under discussion
28 Availability and cost: Open source
29 Licence: Open licence under discussion

30 1. Introduction

31 Since the first water quality model on dissolved oxygen (DO) developed by [Streeter and Phelps \(1925\)](#), water
32 quality modeling takes into account more and more complex biogeochemical cycles in river systems. Numerous river
33 water quality simulation softwares have been developed over the past decades, such as the QUAL series ([Park and](#)
34 [Lee, 2002](#); [Pelletier et al., 2006](#)), AQUATOX ([Park et al., 1974, 1982](#)), MIKE11/MIKE HYDRO River ([DHI, 2007,](#)
35 [2017b](#)), QUASAR ([Whitehead et al., 1997](#)), RWQM ([Shanahan et al., 2001](#); [Reichert et al., 2001](#); [Vanrolleghem](#)
36 [et al., 2001](#)), RIVERSTRAHLER ([Billen et al., 1994](#); [Garnier et al., 1995](#); [Ruelland et al., 2007](#); [Thieu et al., 2009](#);
37 [Raimonet et al., 2018](#)) and ProSE ([Even et al., 1998, 2007b](#); [Flipo et al., 2004](#); [Vilmin et al., 2015b](#)).

38 Complex biogeochemical processes are represented by physico-chemical equations using a high number of param-
39 eters related to microorganisms' activities (growth, respiration, mortality, photosynthesis etc.) and physical processes
40 such as oxygen reaeration. All those parameter values need to be determined as much as possible based on laboratory
41 experiments. A numerical adjustment is required for the remaining parameters. This step is called calibration and is
42 based on the comparison between the output of the forward model and measured data. Finally the calibrated model
43 is validated over another period of time, which ends the two-step fitting procedure. The calibration procedure was
44 formerly achieved following a trial-error method, like the first applications of the RIVE model ([Billen et al., 1994](#);
45 [Garnier et al., 1995](#); [Even et al., 1998](#)) in the Seine basin, or with more elaborate approaches derived from learn-
46 ing strategies ([Vilmin et al., 2015b](#)). To reduce the bias due to user subjectivity, auto-calibration procedures have
47 been implemented into a few river water quality simulation programs, such as in SIMCAT ([Warn, 1987](#)), QUASAR
48 ([Whitehead et al., 1997](#)), QUAL2kw ([Pelletier et al., 2006](#)) and MIKE HYDRO River ([DHI, 2017a](#)). However, the
49 concept of an optimal parameter set having a biophysical meaning is questionable as illustrated by the equifinality,
50 where different parameter sets can produce equally good results ([Beven, 1989](#); [Polus et al., 2011](#)). Moreover, a static
51 optimal parameter over time is not realistic for the description of microorganism activities, such as phytoplankton
52 dynamics ([Mao et al., 2009](#); [Huang et al., 2013](#)) as well as phosphorus dynamics ([Huang and Gao, 2017](#)). The recent

53 study of Wang et al. (2019) confirmed this statement and demonstrated the need for a dynamic representation of the
54 most influential parameters of water quality models.

55 *In situ* sensors allow for a continuous monitoring of river water quality like DO concentrations measured every
56 15 minutes by the MeSeine network of SIAAP (the public sewage company of the Greater Paris area) and by the
57 CARBOSEINE project in the Seine River system (Groleau et al., 2013). These measurements have been used to
58 calibrate and validate the water quality program, ProSe, in the Seine river system (Vilmin et al., 2016, 2018). However,
59 the simulation of DO concentrations during low water periods remains difficult, particularly for dry years and algal
60 bloom periods (Vilmin et al., 2018). An alternative approach to combine the high frequency water quality data
61 with the water quality model is to use a data assimilation technique. Data assimilation methods enable to transform
62 deterministic models into stochastic models (with dynamic model parameters) as recommended by Reichert and
63 Mieleitner (2009) and Kattwinkel and Reichert (2017) for ecological and environmental research.

64 Data assimilation methods have been widely applied in meteorology (Courtier et al., 1994; Kalnay et al., 1996;
65 Courtier et al., 1998; Gauthier et al., 2007; Kleist and Ide, 2015; Yucel et al., 2015) and hydrology modeling (Ottlé
66 and Vidal-Madjar, 1994; Rodell et al., 2004; Moradkhani et al., 2005a,b; Weerts and El Serafy, 2006; Andreadis
67 et al., 2007; Salamon and Feyen, 2009; Moradkhani et al., 2012; Plaza et al., 2012; Paiva et al., 2013; Vrugt et al.,
68 2013; Abbaszadeh et al., 2018). Although the use of data assimilation based on an extended Kalman filter dates back
69 from the late 1970's, early 1980's with improved versions of Streeter Phelps-like, *i.e.* DO-BOD, models (Beck and
70 Young, 1976; Beck, 1978; Whitehead, 1978; Whitehead et al., 1981; Whitehead and Hornberger, 1984; Cosby and
71 Hornberger, 1984; Guo, 2003), its appropriation by the freshwater community remains very limited.

72 With the advances in computational power and high frequency measurements, the application of data assimila-
73 tion methods recently expended in biogeochemical modeling. The ensemble-based Kalman filter has been tested in
74 biogeochemical oceanography modeling (Soetaert and Gregoire, 2011; Simon and Bertino, 2012; Simon et al., 2012;
75 Gharamti et al., 2017; Yu et al., 2018). For water quality modeling, after having used mostly extended kalman filter
76 (Bowles and Grenney, 1978; Cosby et al., 1984; Ennola et al., 1998; Pastres et al., 2003; Mao et al., 2009), recent
77 studies focus on the forecast of algal bloom dynamics in river systems or lakes using ensemble-based Kalman filter
78 (Beck and Halfon, 1991; Huang et al., 2013; Kim et al., 2014; Page et al., 2018; Chen et al., 2019; Loos et al., 2020),
79 which became the most popular assimilation technique (Carrassi et al., 2018; Cho et al., 2020). However this method
80 assumes a Gaussian distribution of the observational and simulation errors. A Bayesian method, as the particle filter,
81 is required to evaluate forward model parameters' distributions (Carrassi et al., 2018). Such an application for fresh-
82 water modelling was also recommended by Huang et al. (2013) and, for the first time in the freshwater community,
83 implemented by Wang et al. (2019) on a synthetic case study. ^{c1}The main interest of a synthetic case study is that the

^{c1}SW: Text added.

84 [parameter values are prescribed so that the efficiency of the particle filter can be evaluated properly](#). This first study on
85 the PF applied to water quality issue demonstrated the interest of such an approach to identify community switch in
86 river systems as well as the most important bio-physical processes over time. ^{c2}[Moreover, the study lead to technical](#) ^{c2SW: Text}
87 [conclusions on the number of particles required, and how to handle parameters which are useful for the real oxygen](#) ^{added.}
88 [data assimilation](#).

89 In the current paper, we present a first application of the particle filter implemented into the PROSE program (Even
90 et al., 1998, 2007b; Flipo et al., 2004; Vilmin et al., 2015b), called hereafter PROSE-PA (Wang et al., 2019), assimilating
91 high frequency ^{c1}[measured](#) DO data. The study aims at proving the capacity of particle filter to simulate DO con- ^{c1SW: Text}
92 centrations and to characterize the physiological properties of phytoplankton and heterotrophic bacteria (represented ^{added.}
93 by model parameters) in contrasted hydrological and trophic contexts (low flow/high flow, autotrophic/heterotrophic)
94 within the Seine River system, ^{c2}[a real and more complex river system compared with the synthetic case study](#) (Wang ^{c2SW: Text}
95 et al., 2019). ^{added.}

96 The manuscript is organized as follows. The sections 2.1 and 2.2 present briefly the PROSE-PA program and the
97 particle filter for easy reading. One can consult the previous publications for the detailed description of PROSE-PA
98 program and the particle filtering algorithm (Wang et al., 2018, 2019). The study area, simulation period, hydrological
99 conditions, model input data and available DO measurements are presented in section 2.3. The simulations, statistical
100 criteria to evaluate the model performance and numerical settings are described in section 2.4. Then, the performances
101 of the PROSE model on the simulated DO concentrations , with and without data assimilation framework, are evaluated
102 separately (section 3.1). Following the simulated DO concentrations, the phytoplanktonic and bacterial parameters'
103 identifiability is successively illustrated in section 3.2. The uncertainties of input BDOM data (biodegradable dis-
104 solved organic matters) and the capacity of particle filter to identify phytoplanktonic parameters' values are discussed
105 in section 4. To finish, the last section gives briefly the conclusions (section 5).

106 2. Material and methods

107 Our goal being the improvement of the DO concentrations in water as well as the identification of dynamic micro-
108 organisms' parameters, we present hereafter the tools, the numerical set up and the details of the simulations.

109 2.1. The PROSE-PA software

110 The PROSE-PA software (Wang et al., 2019) is the most recent version of the PROSE software (Even et al., 1998,
111 2007b; Flipo et al., 2004; Vilmin et al., 2015b), which was used extensively to study the biogeochemical functioning
112 of the Seine River system (Even et al., 1998, 2004, 2007a; Flipo et al., 2004, 2007; Polus et al., 2011; Raimonet et al.,

113 2015; Vilmin et al., 2015b,a, 2016, 2018). It couples PROSE with a particle filtering algorithm in order to assimilate
 114 high frequency DO concentrations in a river system and to characterize the most influential model parameters (Fig.
 115 1).

116 PROSE-PA is composed of three independent C-libraries (a hydrodynamic library, a transport library and a biogeo-
 117 chemical library) and simulates in parallel the hydro-biogeochemical functioning of a river system with a Bayesian
 118 data assimilation algorithm (Particle Filter, Fig. 1). The hydrodynamic library calculates water heights and discharges
 119 by solving the 1D shallow water equations (Fig. 1a). The transport library uses the calculated hydraulic data to sim-
 120 ulate the advection and dispersion of both particulate and dissolved matters (Fig. 1a). The biogeochemical library,
 121 C-RIVE, is based on the community-centered RIVE model (Billen et al., 1994; Garnier et al., 1995). The RIVE model
 122 simulates the cycles of nutrients, carbon and dissolved oxygen in both the water column and an unconsolidated sedi-
 123 ment layer (Fig. 1b). Biogeochemical processes related to microorganisms, such as growth, mortality, photosynthesis
 124 and respiration, are explicitly formulated. The governing equations of C-RIVE are given in Wang et al. (2018).

125 2.2. Data assimilation framework

126 2.2.1. Particle Filter

127 Sequential data assimilation aims at integrating observations \mathbf{y}_t^* at each time step t into the forward model by
 128 conditioning the values of the state variable \mathbf{z}_t (namely DO concentrations and model parameters x_t) to the observed
 129 quantities ($f(\mathbf{z}_t|\mathbf{y}_{1:t}^*)$). Based on the Bayes' theorem (Bayes, 1763) and the Markov property (Markov, 1906), it can be
 130 summed up by the Bayesian filtering equation :

$$131 \quad f(\mathbf{z}_t|\mathbf{y}_{1:t}^*) = \frac{f(\mathbf{y}_t^*|\mathbf{z}_t)f(\mathbf{z}_t|\mathbf{y}_{1:t-1}^*)}{f(\mathbf{y}_t^*|\mathbf{y}_{1:t-1}^*)} = \frac{f(\mathbf{y}_t^*|\mathbf{z}_t)f(\mathbf{z}_t|\mathbf{z}_{t-1}, \mathbf{y}_{1:t-1}^*)f(\mathbf{z}_{t-1}|\mathbf{y}_{1:t-1}^*)}{f(\mathbf{y}_t^*|\mathbf{y}_{1:t-1}^*)} \quad (1)$$

132 where $\mathbf{y}_{1:t}^*$ represents the observation vector up to time t , $f(\mathbf{z}_t|\mathbf{y}_{1:t}^*)$ is the conditional probability of the state vari-
 133 ables knowing the observations up to time t , $f(\mathbf{y}_t^*|\mathbf{z}_t)$ is the probability to observe \mathbf{y}_t^* knowing \mathbf{z}_t (namely the likeli-
 134 hood), $f(\mathbf{z}_t|\mathbf{y}_{1:t-1}^*)$ is the conditional probability of the state variables knowing the observations up to time $t-1$ (prior
 135 knowledge), while $f(\mathbf{y}_t^*|\mathbf{y}_{1:t-1}^*)$ is a normalization constant that does not depend on \mathbf{z}_t . Therefore, to condition the state
 136 variables \mathbf{z}_t to the observations up to time t , we need to know how to compute the likelihood of the observations given
 137 the current values of the parameters ($f(\mathbf{y}_t^*|\mathbf{z}_t)$). The term $f(\mathbf{z}_t|\mathbf{y}_{1:t-1}^*)$ is obtained by propagating $f(\mathbf{z}_{t-1}|\mathbf{y}_{1:t-1}^*)$ through
 138 the forward model.

139 However, there is no analytical formulation for these probability distributions. Particle filtering and in particular
 140 the sampling importance resampling algorithm (SIR, (Doucet et al., 2000; Liu, 2001)) approximates the conditional

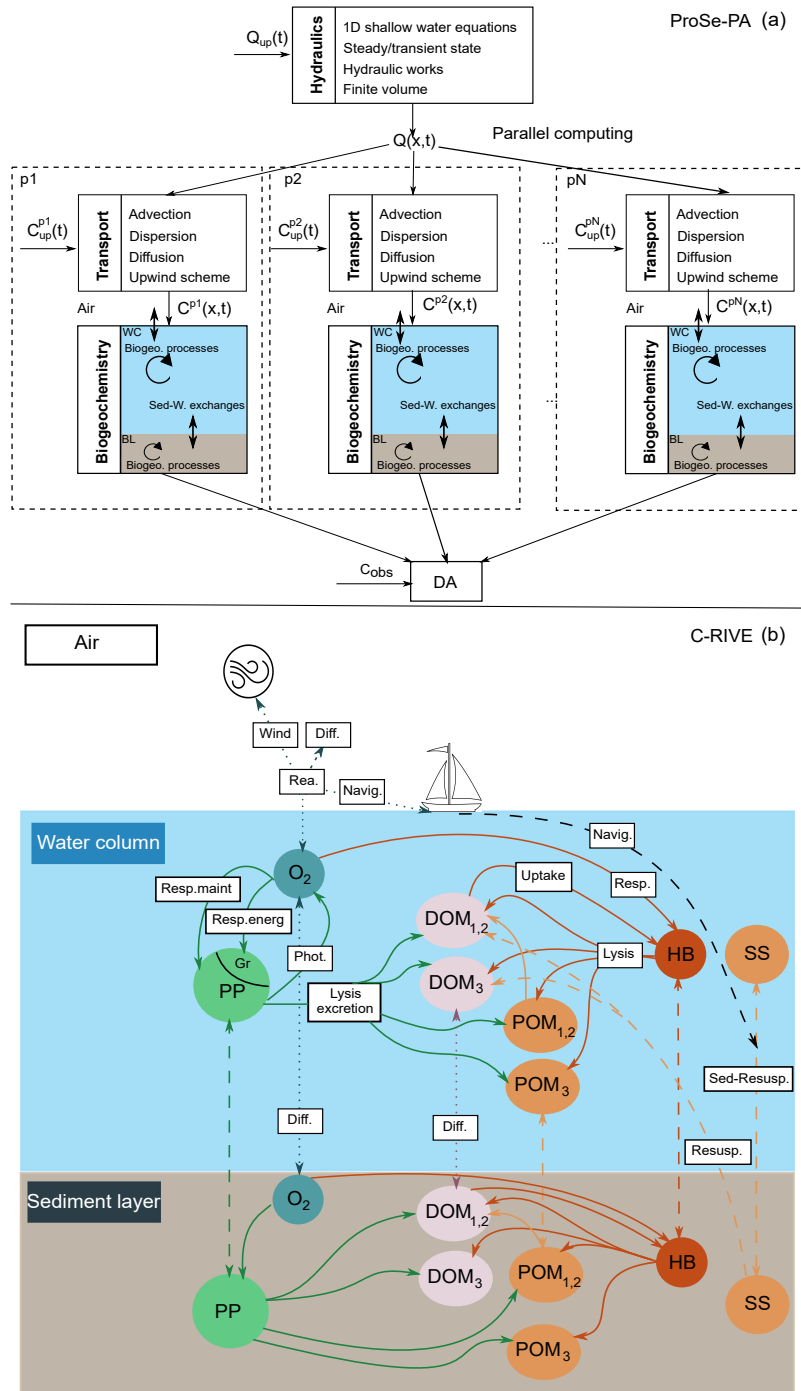


Fig. 1: (a) Schematic description of the ProSe-PA program with the hydraulic, transport, biogeochemical modules and the flowchart of particle filter (particles $p1, p2, \dots, pN$). (b) Simplified flowchart of the biogeochemical model (Wang et al., 2018). Dotted lines: diffusive exchanges (Diff.); dashed lines: sedimentation-resuspension (Sed-Resusp.); solid lines: biogeochemical processes; PP: primary producer; Gr: growth; HB heterotrophic bacteria; SS: suspended solids; DOM: dissolved organic matter; POM: particulate organic matter; subscripts 1, 2 and 3 refer to the fast biodegradable, slowly biodegradable and refractory fractions of organic matter. Rea.: reaeration; Navig.: navigation; Resp.: respiration; Phot.: photosynthesis. Nitrifying bacteria are not represented.

141 distributions ($f(\mathbf{z}_t|\mathbf{y}_{1:t}^*)$) by N weighted realizations of the state variables (the particles). Using an importance distri-
 142 bution (proposal distribution) $\pi(\mathbf{z}_t|\mathbf{y}_{1:t}^*)$, the weight update formula for each particle (Eq. (2)) is

$$143 \quad \omega_t^i = \frac{f(\mathbf{z}_t^i|\mathbf{y}_{1:t}^*)}{\pi(\mathbf{z}_t^i|\mathbf{y}_{1:t}^*)} \propto \frac{f(\mathbf{y}_t^*|\mathbf{z}_t^i)f(\mathbf{z}_t^i|\mathbf{z}_{t-1}^i, \mathbf{y}_{1:t-1}^*)f(\mathbf{z}_{t-1}^i|\mathbf{y}_{1:t-1}^*)}{\pi(\mathbf{z}_t^i|\mathbf{z}_{t-1}^i, \mathbf{y}_t^*)\pi(\mathbf{z}_{t-1}^i|\mathbf{y}_{1:t-1}^*)} \quad (2)$$

144 where $f(\mathbf{y}_t^*|\mathbf{z}_t^i)$ is the likelihood function, which indicates how \mathbf{y}_t^* is likely to be observed given \mathbf{z}_t^i at time t .

145 Typically, the standard importance distribution ($\pi(\mathbf{z}_t^i|\mathbf{z}_{t-1}^i, \mathbf{y}_t^*) = f(\mathbf{z}_t^i|\mathbf{z}_{t-1}^i, \mathbf{y}_{1:t-1}^*)$) is used (Doucet et al., 2001;
 146 Särkkä, 2013). In this case, the equation (2) can be transformed to,

$$147 \quad \omega_t^i \propto f(\mathbf{y}_t^*|\mathbf{z}_t^i)\omega_{t-1}^i \quad (3)$$

$$148 \quad \hat{\omega}_t^i = \frac{\omega_t^i}{\sum \omega_t^i}$$

149 where ω_t^i and ω_{t-1}^i represent the posterior and prior importance weights at time t . $\hat{\omega}_t^i$ is the normalised importance
 150 weight.

151 The filtering distribution $f(\mathbf{z}_t|\mathbf{y}_{1:t}^*)$ is then approximated by the weighted particles (eq. (4), Doucet et al. (2001)):

$$152 \quad f(\mathbf{z}_t|\mathbf{y}_{1:t}^*) \approx \sum_{i=1}^N \hat{\omega}_t^i \delta_{\mathbf{z}_t^i} \quad (4)$$

153 where $\delta(\cdot)$ is the Dirac function and N is the number of particles.

154 Assuming that the observation error is Gaussian, the likelihood of each particle i is given by the pdf of the multi-
 155 variate normal distribution.

$$156 \quad \ln L(\mathbf{y}_t^*|\mathbf{z}_t^i) = -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Sigma|) - \frac{1}{2} (\mathbf{y}_t^* - h(\mathbf{z}_t^i))^T \Sigma^{-1} (\mathbf{y}_t^* - h(\mathbf{z}_t^i)) \quad (5)$$

$$157 \quad f(\mathbf{y}_t^*|\mathbf{z}_t^i) = \frac{L(\mathbf{y}_t^*|\mathbf{z}_t^i)}{\sum_{i=1}^N L(\mathbf{y}_t^*|\mathbf{z}_t^i)}$$

158 where m is the number of monitoring stations and h denotes the observation operator which extracts the simulated
 159 dissolved oxygen concentrations at the monitoring sites from \mathbf{z}_t^i to the observation \mathbf{y}_t^* . Σ is the error covariance matrix
 160 of the observations. Considering that the observation errors at the monitoring stations are mutually independent, Σ
 161 is a diagonal matrix. Since we assumed that the observation error follows a normal distribution, the diagonal terms
 162 correspond to the variance of the measurement errors.

163 2.2.2. Resampling procedure

164 A well known and important problem when applying particle filtering is the degeneracy of the weights which
165 means that almost all the particles have a near zero importance weight after a simulation time and only a few particles
166 have a high importance weight. The discrete probability densities then cannot approximate adequately the filtering
167 posterior pdf $f(\mathbf{z}_t | \mathbf{y}_{1:t}^*)$ (eq. (4)). To reduce the degeneracy effect, a series of resampling methods were proposed
168 in the literature and reviewed by Li et al. (2015). Resampling aims at eliminating particles with small importance
169 weights and duplicating particles with high importance weights. The systematic resampling technique (Kitagawa,
170 1996; Moradkhani et al., 2005a; Li et al., 2015) has been implemented in ProSE-PA. The criterion for performing
171 resampling is based on the variation of the importance weights which indicates the degree of particle degeneracy. The
172 effective sample size (N_{eff}) defined by Kong et al. (1994) allows for monitoring the degree of particle degeneracy and
173 is approximated by \widehat{N}_{eff} (eq. (6)):

$$174 \quad \widehat{N}_{eff} = \frac{1}{\sum_{i=1}^N (\hat{\omega}_t^i)^2} \quad (6)$$

175 \widehat{N}_{eff} ranges from 1 to N (the number of particles). In practice, the resampling procedure is performed only when
176 \widehat{N}_{eff} falls below a user-defined threshold $N_{thres} = \alpha \cdot N$, with $\alpha = 0.3$ (Wang et al., 2019). After resampling, all
177 the importance weights are reset to $1/N$. As mentioned above, the particles with high importance weights may be
178 duplicated many times, which results in sample impoverishment. To restore the diversity of the particles, a random
179 perturbation is added to the parameter values after the resampling procedure (eq. (7)).

$$180 \quad \mathbf{x}_{t+1}^i = \mathbf{x}_{t,resampling}^i + \eta_t^i \quad \eta_t^i \sim N(0, (s \cdot \Phi_i)^2) \quad (7)$$

181 where s is a user-defined perturbation proportional to the parameter ranges, Φ_i .

182 More details about the mathematical formulations of ProSE-PA approach and a flowchart can be found in Wang
183 et al. (2019).

184 2.3. Study area and available data

185 2.3.1. Study area and simulation period

186 The study area starts from the upstream of Paris city up to the entrance of the Seine estuary at Poses (Fig. 2). This
187 area consists of 220 km stretch of the Seine River and 25 km of the Marne River. Four tributaries of the Seine River

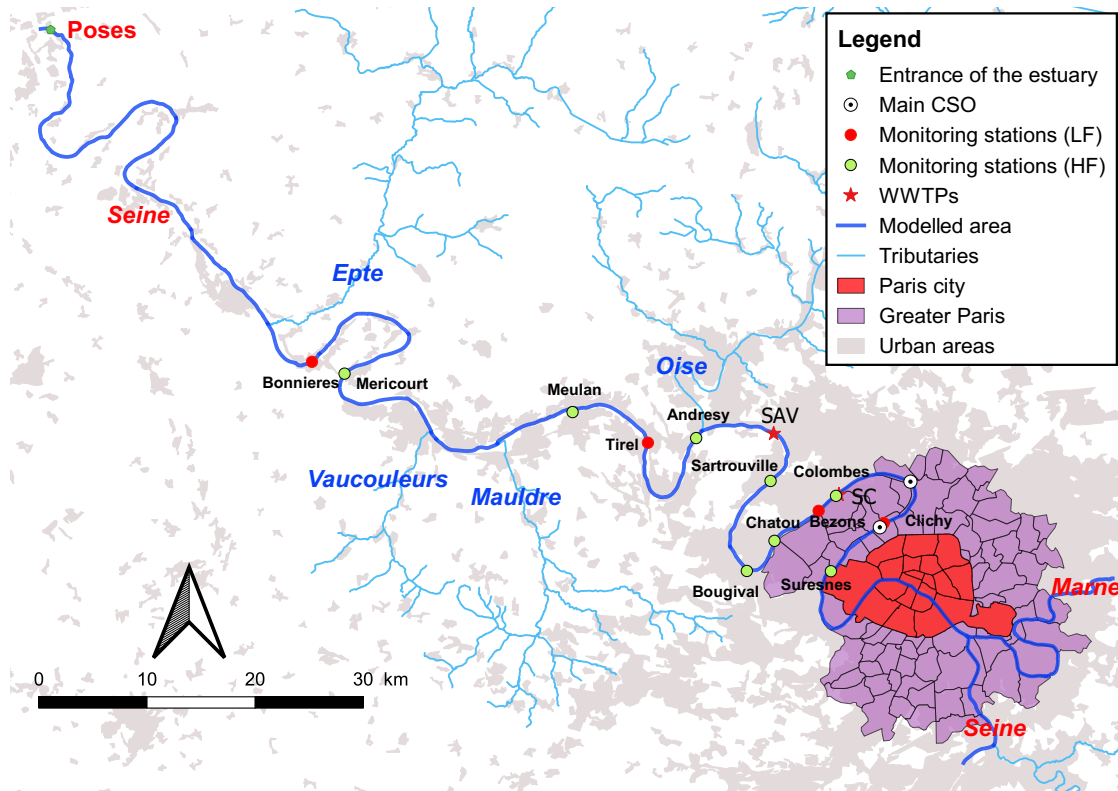


Fig. 2: The description of study area and monitoring sites. Monitoring stations (High frequency, in green) from upstream to downstream: 1. Suresnes, 2. Colombes, 3. Chatou, 4. Bougival, 5. Sartrouville, 6. Andresy, 7. Meulan, 8. Mericourt. Monitoring stations (Low frequency, in red) from upstream to downstream: 1. Clichy, 2. Bezons, 3. Triel, 4. Bonnières. Two major WWTPs, Seine Centre (SC) and Seine Aval (SAV), are represented by red stars.

188 (Oise, Mauldre, Vaucouleurs and Epte) are taken into account in the model as lateral boundary conditions. Bearing
 189 almost 16 million inhabitants, the study area is highly impacted by human activities. Five Waste Water Treatment
 190 Plants (WWTPs) managed by the SIAAP are located in the modeled area. Only two major WWTPs (Seine Aval - SAV
 191 - and Seine Centre - SC) located downstream of Paris city are displayed in figure 2. SAV, the largest WWTP of Europe,
 192 treats the effluents of over 6.5 million equivalent inhabitants (Rocher et al., 2011). It has a mean water discharge of 17
 193 m³/s in summer, which contributes to almost one third of the discharge of the Seine River at the gauging station for
 194 2011. During large rain events, the water collected by the combined sewer system exceeds the treatment capacity of
 195 the WWTPs and overflows directly into the Seine River through many stormwater discharge pipes (more than 150).
 196 These water are called CSOs (Combined Sewer Overflows) that carry large amounts of suspended solids, organic
 197 matters and nutrients into the Seine River (Even et al., 2007b). The duration of the CSOs varies from 1 hour to 21
 198 hours and 90% of the CSOs last less than 10 hours. CSOs induce fast DO concentration drops by promoting bacterial
 199 activities (Fig. 3). Two major CSOs outlets are presented in figure 2.

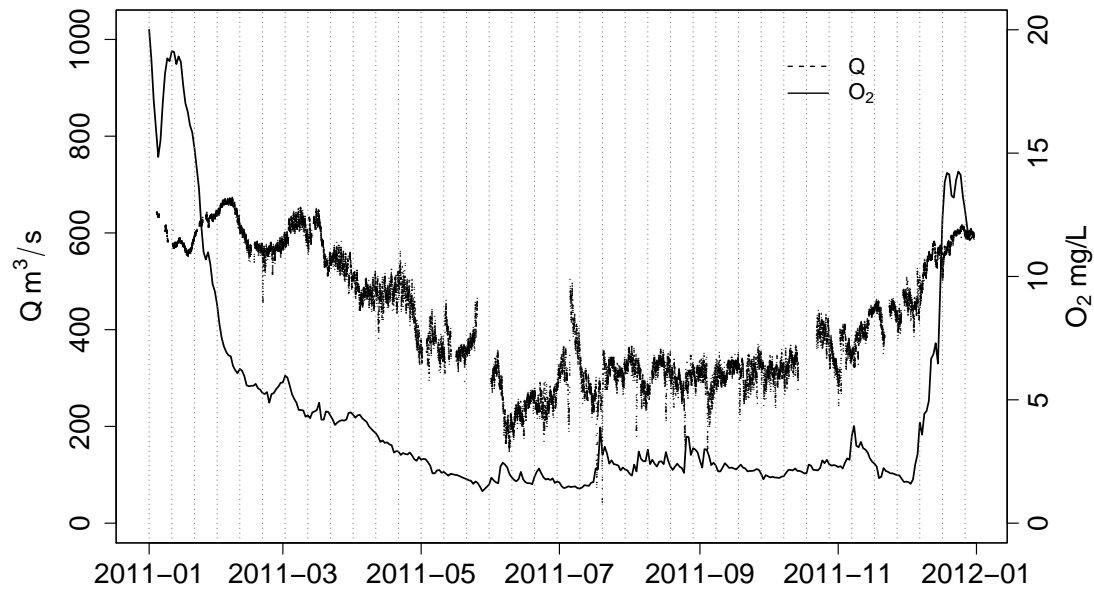


Fig. 3: Daily discharge (Q m³/s) at the gauging station and 15-min measured DO concentrations (O_2 mg/L) at Andresy station in 2011.

200 The Seine River is submitted to temperate climate conditions. It has a maximum discharge in winter and a
 201 minimum discharge in summer. The hydro-biogeochemical functioning of the Seine River is simulated for the year
 202 2011 (2011-01-01 to 2011-12-31). The year 2011 is a relatively dry year with a mean daily discharge of 226.6 m³/s, a
 203 maximum daily discharge of 1020 m³/s and a minimum daily discharge of 66.2 m³/s at the gauging station (obtained
 204 from Banque HYDRO database on 2020-06-23, www.hydro.eaufrance.fr, Fig. 3). Even though algal blooms
 205 almost disappeared in the Seine River (Garnier et al., 2020), three successive algal blooms occurred in the Seine River
 206 system in 2011 (Vilmin et al., 2016). The year 2011 therefore offers a large panel of hydrological and trophic contexts,
 207 which is challenging for the first test of the particle filter for mimicking a real river system.

208 2.3.2. Input data and available oxygen data

209 As for previous published studies (Vilmin et al., 2015b,a, 2016, 2018), daily water quality data (nutrients, organic
 210 matters, suspended solids, bacteria biomass) of the Seine, Marne and Oise rivers are provided by the public drinking
 211 water company of the Paris urban area (SEDIF). They are applied as lateral and upstream boundary conditions. For
 212 primary producers, weekly chl *a* measurements provided by SIAAP are used as the upstream boundary condition for
 213 Seine River while 15-min chl *a* concentrations provided by SEDIF define the upstream boundary condition of the
 214 Marne River. For smaller tributaries (Mauldre, Vaucouleurs and Epte), lower frequency water quality data from the

215 national river monitoring network (RCS) are available. River daily discharge are obtained from the national Banque
 216 HYDRO database (www.hydro.eaufrance.fr). The daily water flow and quality data of the five WWTPs provided
 217 by SIAAP, 151 CSOs and 26 small permanent effluents are also considered as lateral boundary conditions.

218 Eight monitoring stations managed by SIAAP provide 15-min time step DO concentrations in 2011 (Fig. 2 and
 219 Tab. 1). The differences in the number of measurements are related to sensor dysfunctioning. Four monitoring stations
 220 from RCS monitoring network provide low frequency DO measurements (Fig. 2 and Tab. 1). These high frequency
 221 measurements are assimilated by ProSE-PA (Wang et al., 2019) to improve the simulation of the biogeochemical
 222 functioning of the Seine River system and to characterize the most influential model parameters that were identified
 223 by Wang et al. (2018) beforehand.

Table 1: DO concentrations in 2011 provided by SIAAP and SEDIF

Station name	Number of measurements	Frequency	Data provider
Suresnes	30948	15 min	SIAAP
Colombes	15994	15 min	SIAAP
Chatou	29626	15 min	SIAAP
Bougival	33971	15 min	SIAAP
Sartrouville	14912	15 min	SIAAP
Andresy	30586	15 min	SIAAP
Meulan-en-Yvelines	16738	15 min	SIAAP
Mericourt	32680	15 min	SIAAP
Clichy	6	bimonthly	SEDIF
Bezons	6	bimonthly	SEDIF
Triel	6	bimonthly	SEDIF
Bonnières	7	bimonthly	SEDIF

224 2.4. Simulations and numerical setting

225 2.4.1. Simulation types and statistical criteria

226 To assess the performances of the particle filter, a cross-validation was realized. That means, we run 8 simulations
 227 with data assimilation. For each simulation, only the observed data at 7 high frequency monitoring stations are
 228 assimilated in order to predict the DO concentrations at the 8th monitoring station.

229 To evaluate the potential of ProSE-PA for the simulation of DO concentrations at the four low frequency monitor-
 230 ing stations (RCS), all the high frequency observed DO data are assimilated.

231 The results of a classic "static parameters" ProSE simulation without data assimilation is systematically displayed,
 232 for the sake of comparison with the "dynamic parameters" simulation provided by the particle filter. Calibrated
 233 parameters of the ProSE model (Garnier et al., 1995; Vilmin, 2014; Vilmin et al., 2015b) are used for the simulation
 234 without data assimilation.

235 The predicted DO concentrations are compared with the observations subsequently. Two statistical criteria, namely
 236 the Root Mean Square Error (RMSE) and the Kling Gupta Efficiency (KGE, [Kling et al. \(2012\)](#)), are calculated to
 237 evaluate the model performances. KGE is composed of three terms: the correlation coefficient, the model bias and the
 238 coefficient of variation (Eq. (9)). KGE ranges from -Inf to 1. Essentially, the closer to 1 the KGE, the more accurate
 239 the model.

$$240 \quad \text{RMSE} = \sqrt{\frac{\sum_{k=1}^{N_{obs}} (y^{sim,k} - y^{obs,k})^2}{N_{obs}}} \quad (8)$$

$$241 \quad \text{KGE} = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (9)$$

242 where

243 N_{obs} : Number of observations

244 $y^{sim,k}$ and $y^{obs,k}$: Simulated and observed DO concentrations

245 r : Correlation coefficient

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

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

248 2.4.2. Parameters considered for data assimilation and numerical settings

249 The twelve parameters considered for data assimilation in ProSE-PA were selected based on a previous sensitivity
 250 analysis of the C-RIVE model ([Wang et al., 2018](#)) and on the subsequent first proof of concept of a particle filter
 251 applied to water quality issues ([Wang et al., 2019](#)). Following the recommendations of [Wang et al. \(2019\)](#), the number
 252 of particles used in this study is set to 500.

253 The boundary conditions, namely the water quality data, the river discharge, the water flow and quality data of the
 254 WWTPs and the CSOs, are considered perfectly known. Although this may be a strong assumption, it allows for a
 255 first quantitative estimation of the particle filter performances for the simulation of real river systems. Especially the
 256 share of organic matter between its biodegradable and refractory fractions was calibrated using the forward model of
 257 [Vilmin et al. \(2018\)](#) configured with "static parameters", resulting in an overall better simulation of DO (Fig.4 and
 258 5). After calibration, the BDOM accounts for 47% of total DOM carried by Seine, Marne and Oise rivers, while this
 259 BDOM portion was 30% before calibration.

260 The observation error is assumed to follow a Gaussian distribution with a mean zero and a standard deviation of
 261 10% of the measured DO concentrations ($0.1y^*$), which is commonly acknowledged by the water quality community

Table 2: Parameters considered in data assimilation by ProSE-PA (Wang et al., 2019)

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

262 using EKF and EnKF (Mao et al., 2009; Kim et al., 2014). This relatively large observation error integrates both the
 263 measurement error itself and error due to uncertainties in sensor positioning.

264 The perturbation of the model parameters after the resampling procedure is randomly generated following a normal
 265 distribution with a mean zero and a standard deviation proportional to the parameter ranges. Based on the previous
 266 study (Wang et al., 2019), the value of s (proportion of the parameter range in equation (7)) is set to 0.1.

267 Eventually, 5-min time step simulations are run using 20 processors Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40
 268 GHz. The simulation with data assimilation takes 3.9 days for a 1-year simulation period (365 days).

269 3. Results: DO simulation and model parameter characterization

270 3.1. Improved simulation of DO concentrations with Particle Filter

271 The ensemble weighted average of the simulated DO concentrations and the 95% confidence intervals are cal-
 272 culated at the eight high frequency monitoring stations (Fig. 4 and Fig. 5). RMSEs and KGEs are calculated for
 273 the whole year of 2011, the algal bloom periods and the low flow periods without algal bloom (from 2011-07-20 to
 274 2011-12-06).

275 The particle filter improves the simulation of DO concentrations significantly (Tab. 3). Indeed, RMSEs obtained
 276 with ProSE-PA range from 0.71 mgO₂/L to 1.32 mgO₂/L while those obtained with forward model ProSE vary from
 277 0.75 mgO₂/L to 1.61 mgO₂/L for the whole year of 2011. RMSEs for ProSE-PA decrease at all monitoring stations,
 278 especially at Bougival (47%, 0.74 mgO₂/L vs 1.40 mgO₂/L, Tab. 3), Sartrouville (33%, 1.14 mgO₂/L vs 1.69 mgO₂/L,

279 Tab. 3) and Meulan (29%, 0.70 mgO₂/L vs 0.99 mgO₂/L, Tab. 3). Moreover, KGEs obtained with ProSE-PA are closer
280 to 1, except for the Sartrouville station (0.46 vs 0.53). At all stations, the simulated DO concentrations by ProSE-PA
281 are more correlated with the measured DO concentrations (see r in Tab. 3). In terms of model bias for the full year
282 2011 (β), ProSE-PA and ProSE produce similar results. Both ProSE-PA and ProSE tend to overestimate DO concen-
283 trations ($\beta > 1$, Tab. 3). The maximum bias is estimated at the Sartrouville station with an overestimation of 12% for
284 ProSE-PA and 19% for ProSE respectively (Tab. 3). The coefficient of variation (γ) explains the relative dispersion
285 of the DO concentrations around the average of DO concentration. The results show that ProSE-PA evaluates more
286 correctly the dispersion of observed DO concentrations where the coefficients of variation (γ) are much closer to 1
287 (Tab. 3), except for the Sartrouville (0.51 vs 0.78) and Meulan (0.60 vs 0.77) stations. However, both ProSE-PA and
288 ProSE underestimate the amplitude of DO concentrations ($\gamma < 1$). Despite a lower KGE at Sartrouville station for
289 the simulations of ProSE-PA, the results show the capacity of particle filter for predicting (cross-validation) the DO
290 concentrations in river systems (higher RMSE at all stations).

291 Besides improving the estimation of DO concentrations, the particle filter captures algae dynamics accurately
292 (Tab. 3, algal bloom period). The impact of algal blooms on DO concentrations is very difficult to simulate using the
293 forward model ProSE with static parameters. With ProSE, six values of KGEs are lower than 0.5, among which three
294 values are negative during algal bloom period (Tab. 3), while only two KGE values are below 0.5 for ProSE-PA. The
295 simulations at the four low frequency stations show also the improvement of algae dynamics simulation (Fig. 6).

296 In a lesser extent ProSE-PA also improves the simulation of DO concentrations during low flow period without
297 algal bloom. But as for forward simulations with ProSE, the DO concentrations estimated with ProSE-PA are still
298 overestimated, which is related to large uncertainties on the organic matter composition as further discussed in the
299 discussion section.

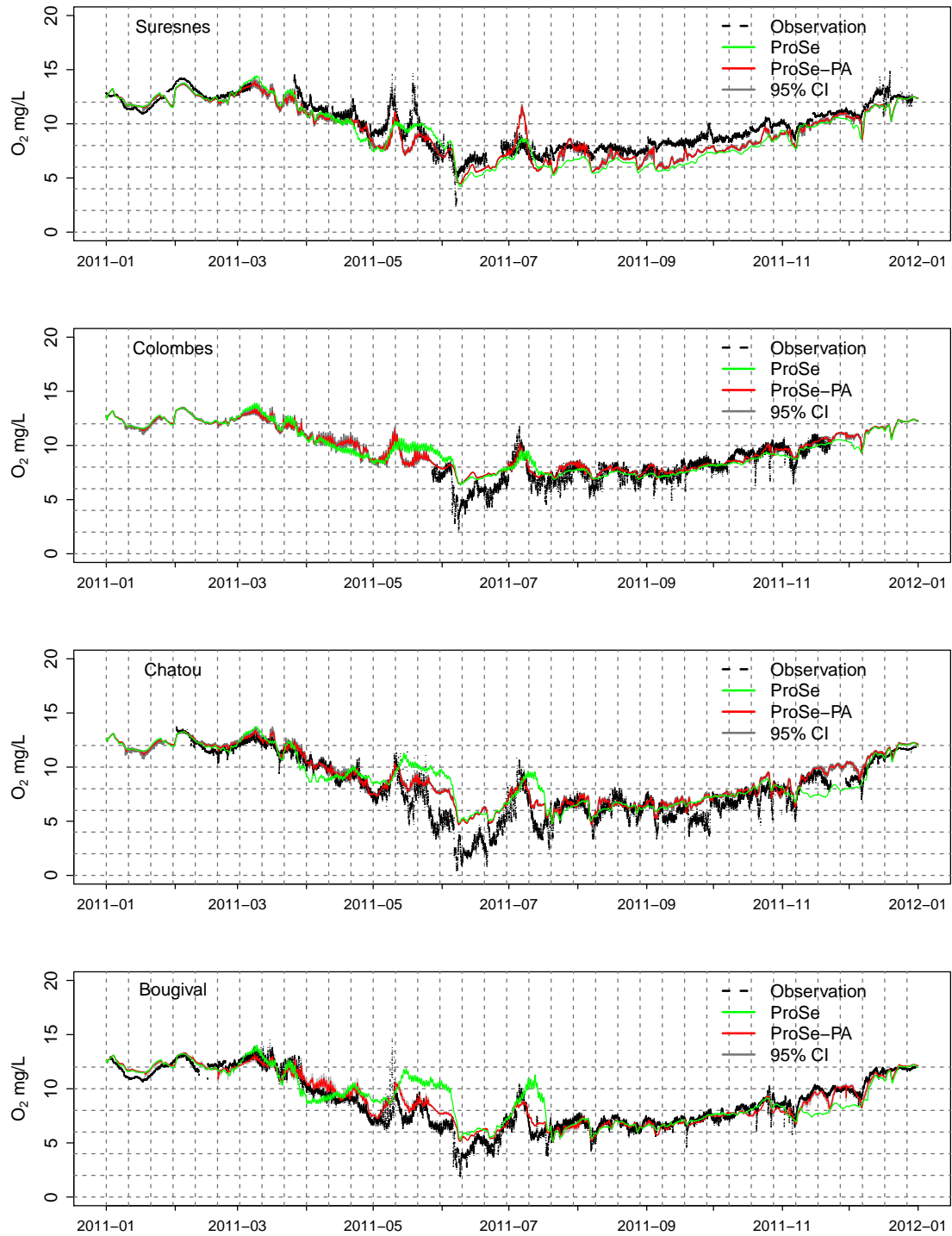


Fig. 4: Simulated DO concentrations at Suresnes, Colombes, Chatou and Bougival stations (cross-validation). The red lines denote ensemble weighted averages simulated by ProSe-PA. The green lines show the simulated DO concentrations by ProSe with static parameters. The observation data are represented by dashed black lines. The gray areas correspond to the 95 percentile confidence intervals (CI).

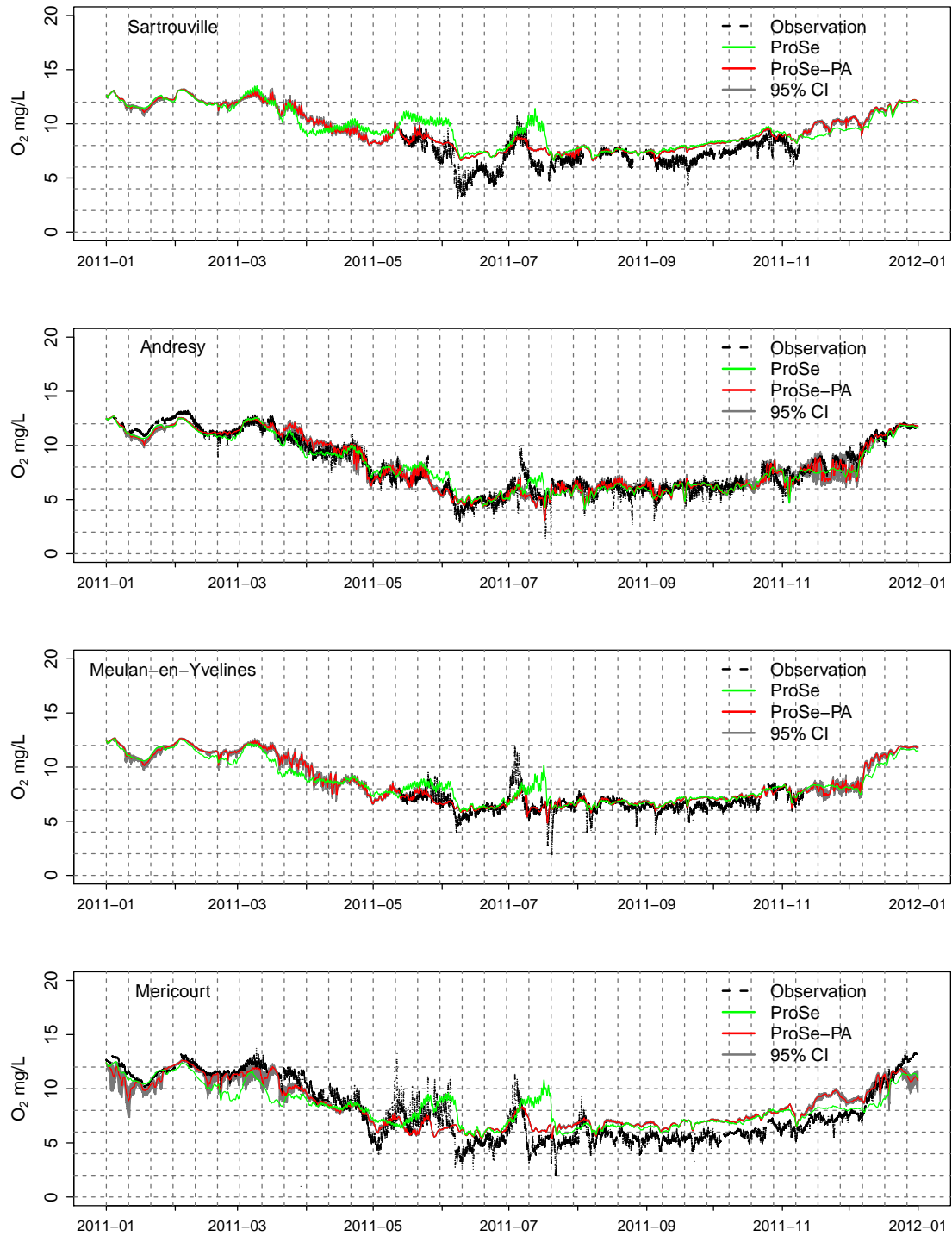


Fig. 5: Simulated DO concentrations at Sartrouville, Andresy, Meulan and Mericourt stations (cross-validation). The red lines denote ensemble weighted averages simulated by ProSe-PA. The green lines show the simulated DO concentrations by ProSe with static parameters. The observation data are represented by dashed black lines. The gray areas correspond to the 95 percentile confidence intervals (CI).

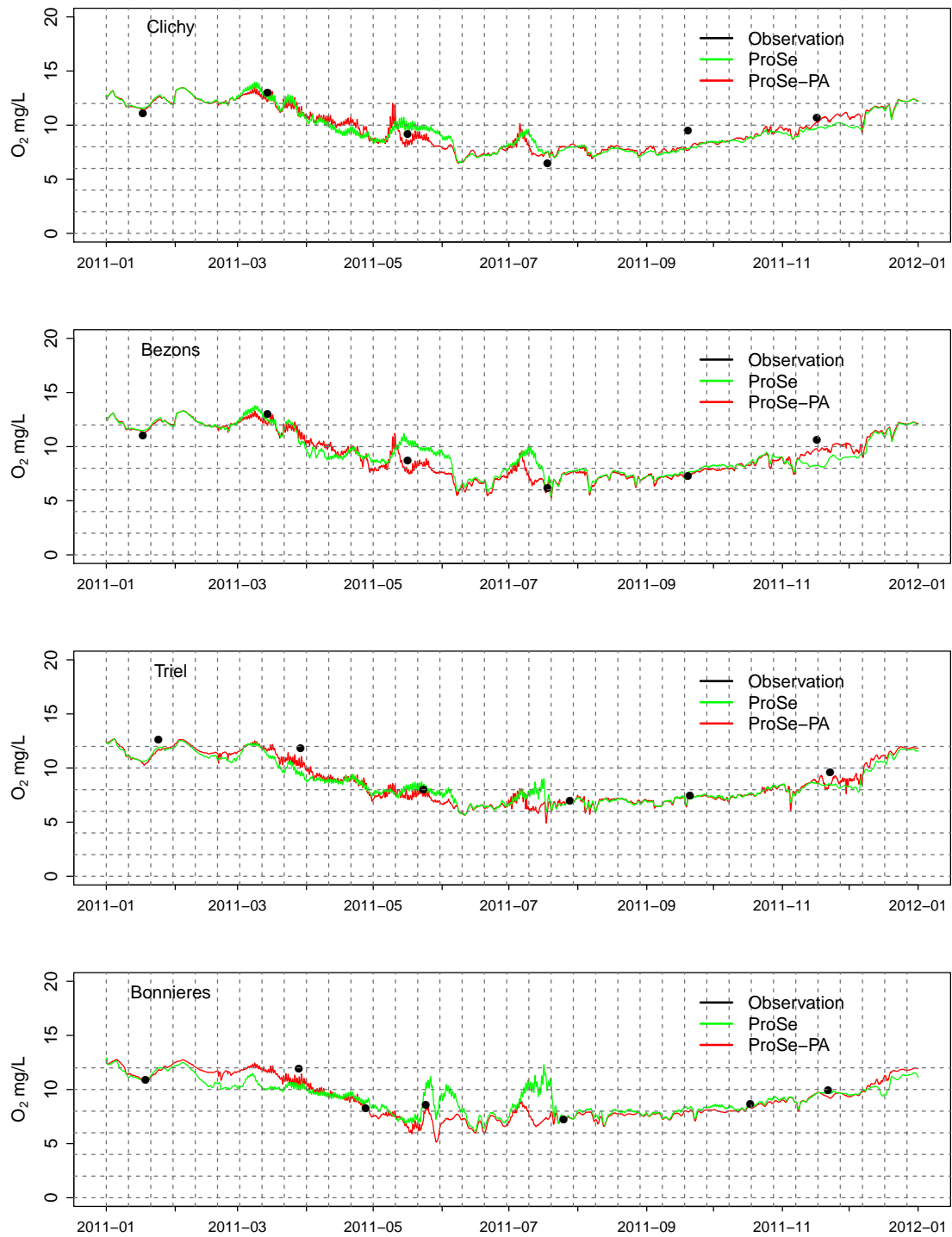


Fig. 6: Simulated DO concentrations at Clichy, Bezons, Triel and Bonnières stations (cross-validation). The red lines denote ensemble weighted averages simulated by ProSe-PA. The green lines show the simulated DO concentrations by ProSe with static parameters.

Table 3: Performances of the ProSE-PA and ProSE evaluated by RMSE and KGE

Stations	Average for 2011						Algal bloom period			Low flow period							
	ProSE-PA			ProSE			ProSE-PA	ProSE	ProSE-PA	ProSE	ProSE-PA	P					
	RMSE	KGE	τ	β	γ	γ	RMSE	KGE	τ	β	γ	RMSE	KGE				
Suresnes	1.09	0.81	0.93	0.93	1.17	1.28	1.26	0.70	0.93	0.91	1.28	1.40	0.74	1.19	0.76	1.07	0.64
Colombes	0.96	0.57	0.85	1.06	0.61	0.59	1.03	0.53	0.77	1.04	0.59	0.76	0.51	1.46	-0.05	0.65	0.75
Chatou	1.24	0.74	0.94	1.09	0.76	0.76	1.61	0.70	0.86	1.09	0.76	1.14	0.72	1.96	0.48	1.02	0.75
Bougival	0.74	0.90	0.96	1.02	0.92	0.88	1.40	0.80	0.83	1.04	0.88	0.92	0.72	2.00	0.44	0.51	0.93
Sartrouville	1.14	0.47	0.82	1.12	0.51	0.78	1.69	0.53	0.64	1.19	0.78	0.74	0.42	2.23	-0.05	0.89	0.68
Andresy	0.73	0.96	0.96	1.00	1.00	0.94	0.75	0.92	0.96	0.99	0.94	0.88	0.84	0.79	0.91	0.76	0.68
Meulan	0.71	0.48	0.67	1.03	0.60	0.77	0.99	0.45	0.51	1.08	0.77	1.07	0.24	1.43	-0.27	0.60	0.55
Mericourt	1.32	0.67	0.91	1.05	0.69	0.62	1.52	0.60	0.87	1.04	0.62	1.28	0.78	1.62	0.47	1.52	0.65

300 3.2. Dynamic parameters' characterization

301 To evaluate the performances of ProSE-PA on parameters' characterization, the temporal evolution of the posterior
302 pdf of the model parameters are displayed in a set of image plots (Fig. 7 and 9). High frequency oxygen data at eight
303 stations are assimilated to characterize the most influential parameters. The posterior pdfs (Figs. 8 and 10) are
304 estimated by weighted kernel density estimation using the normalised importance weights (Eq. (2)). The posterior
305 pdf can then give the parameter value at which the posterior pdf has a maximum density. This parameter value is
306 called the mode of the posterior pdf which is described as mode hereafter. To compare the parameter values estimated
307 by data assimilation with those reported in the literature, the average of the modes of a given parameter is calculated
308 for specific periods.

309 3.2.1. Phytoplanktonic parameters

310 Three successive algal blooms occurred in the Seine River system in 2011 (Vilmin et al., 2016). These periods
311 are represented by the line shaded polygons (Fig. 7, chl $a > 12 \mu\text{g/L}$). Phytoplanktonic parameters affect the DO
312 concentrations only during algal blooms (Wang et al., 2019). Therefore, the results of each algal bloom are described
313 independently.

314 Bloom of March and April (first line shaded polygon, from 2011-03-04 to 2011-04-30).

315 The results illustrate that the optimal temperature for the growth of phytoplankton ($T_{opt,pp}$) increases after April
316 (Fig. 7). Before April (from 2011-03-04 to 2011-03-31), low optimal temperatures ($12 \text{ }^\circ\text{C} - 20 \text{ }^\circ\text{C}$) with an average
317 mode of $15.7 \text{ }^\circ\text{C}$ is estimated, while moderate optimal temperatures ($22 \text{ }^\circ\text{C} - 28 \text{ }^\circ\text{C}$) with an average of $24.7 \text{ }^\circ\text{C}$
318 are obtained from 2011-04-07 to 2011-04-18. ProSE-PA identifies then two optimal temperatures representing two
319 different physiological properties during the bloom of March and April (Fig. 7). The bloom seems therefore divided
320 into two phases.

321 Relative high values of the respiration of maintenance ($R_{m,pp}$), ranging from 0.012 h^{-1} to 0.018 h^{-1} (Fig. 8),
322 are estimated in the early stage of the bloom (2011-03-04 to 2011-03-15). An average of 0.015 h^{-1} is obtained
323 over this period. Then, the mode of $R_{m,pp}$ moves towards 0.006 h^{-1} (Fig. 8). At the end of the bloom, the modes
324 vary within the parameter range (Fig. 7). This behaviour can be observed for the light extinction coefficient by
325 algal self-shading ($\eta_{chla,pp}$) and the ratio of chlorophyll a to carbon (Chla/C_{pp}) also, while no clear switch of the
326 maximum photosynthesis rate ($P_{max,pp}$) can be observed before 2011-04-15. An average of 0.39 h^{-1} for $P_{max,pp}$ during
327 this period is estimated. After 2011-04-15, the maximum photosynthesis rate of phytoplankton ($P_{max,pp}$) decreases
328 with an average of 0.27 h^{-1} . The posterior distributions of the photosynthetic capacity (α_{pp}) from 2011-03-16 to

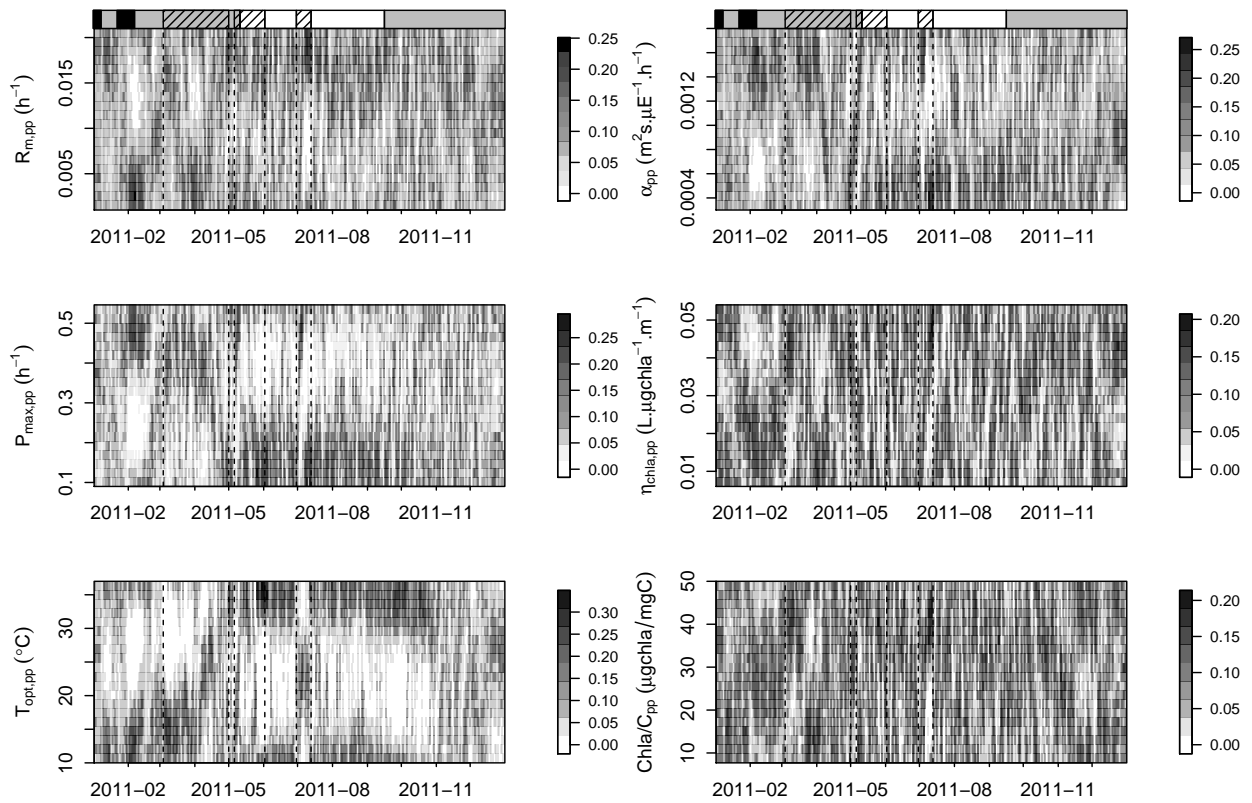


Fig. 7: Normalised importance weights of phytoplanktonic parameters. See Tab. 2 for parameter definition. Line shaded polygons correspond to algae bloom periods and water temperature increases from black to white. For the color gradient: black $T < 6\text{ }^{\circ}\text{C}$, gray $6\text{ }^{\circ}\text{C} < T < 20\text{ }^{\circ}\text{C}$, white $T > 20\text{ }^{\circ}\text{C}$.

329 2011-04-05 are stable with modes ranging from $0.013\text{ m}^2.\text{s}.\mu\text{E}^{-1}.\text{h}^{-1}$ to $0.016\text{ m}^2.\text{s}.\mu\text{E}^{-1}.\text{h}^{-1}$. An average of 0.0015
 330 $\text{m}^2.\text{s}.\mu\text{E}^{-1}.\text{h}^{-1}$ is calculated for this period.

331 *Bloom of May (second line shaded polygon, from 2011-05-06 to 2011-06-02).*

332

333 The posterior pdfs of the model parameters vary during the bloom in May (Fig. 7). The modes fluctuate widely.
 334 Even if the posterior pdfs of the phytoplanktonic parameters are unstable during this period, the simulated peaks of DO
 335 concentrations by ProSE-PA are well synchronized with the observed peaks which can not be correctly reproduced
 336 using ProSE (static parameter setting, Fig. 4 and 5). The results indicate that a time-varying parameter setting is
 337 necessary to reproduce the algae dynamics as mentioned by Mao et al. (2009), Huang et al. (2013) and Wang et al.
 338 (2019).

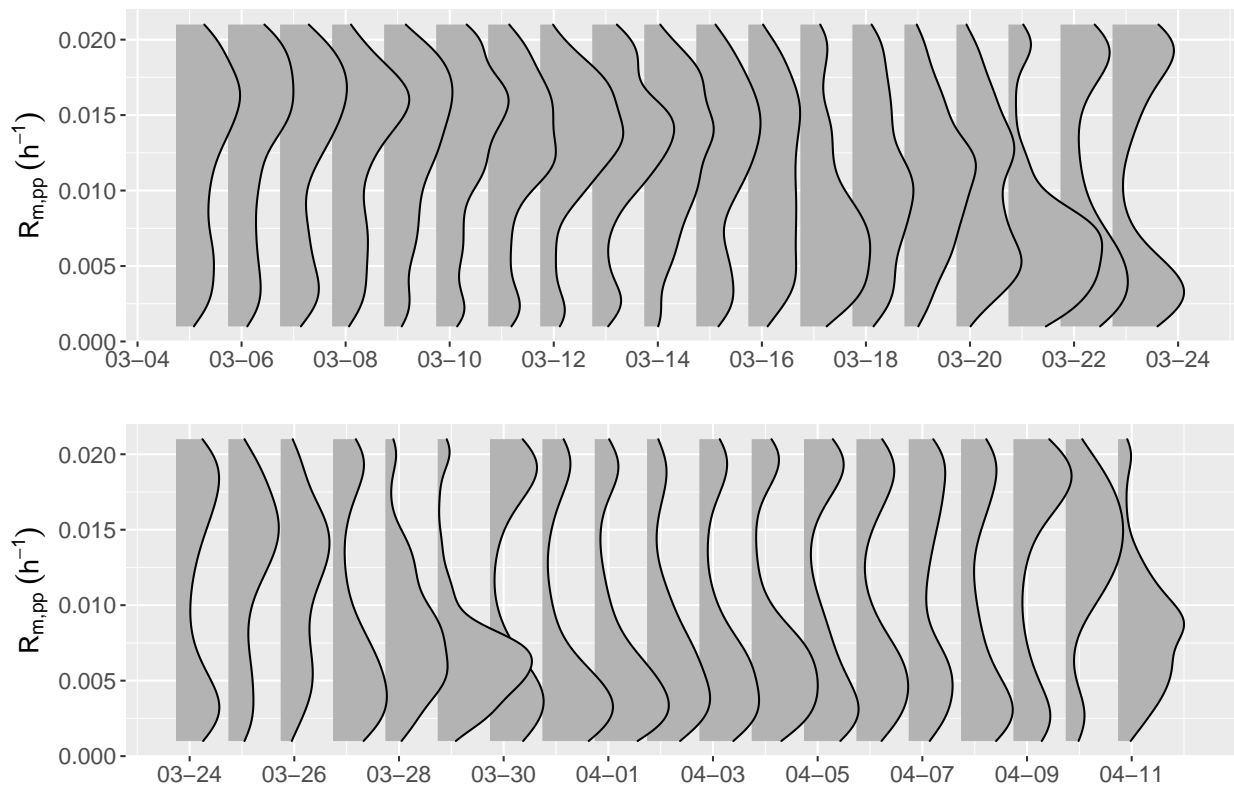


Fig. 8: Posterior pdfs of the respiration of maintenance during the bloom of March.

339 Bloom of July (third line shaded polygon, from 2011-06-30 to 2011-07-13).

340

341 During the bloom of July, the optimal temperature for the growth of phytoplankton ($T_{opt,pp}$) is well determined
 342 with a average mode of 22.4 °C (Fig. 7). High values of $R_{m,pp}$ and $\eta_{chla,pp}$ corresponding to the maximum of the
 343 posterior densities are identified (0.015 h⁻¹ and 0.034 L.μgchla⁻¹.m⁻¹, Fig. 7). A switch of the phytoplanktonic
 344 properties (α_{pp} and $P_{max,pp}$) can be observed at the end of the bloom, which indicates that the particle filter is able to
 345 represent the mortality of phytoplankton linked to its senescence by a decrease in the maximum photosynthesis rate
 346 ($P_{max,pp}$) and the photosynthetic capacity (α_{pp}).

347 3.2.2. Bacterial parameters

348 The bacterial parameters are mostly influential out of algal bloom periods and with moderate to high water tem-
 349 perature ($T > 6$ °C) (Wang et al., 2018).

350 PROSE-PA identifies bacteria parameters before the bloom of March (Fig. 9 and Fig. 10). The averages of the
 351 modes are estimated at 0.026 h⁻¹ for the maximum growth rate ($\mu_{max,hb}$), 0.29 for the bacterial growth yield (Y_{hb}), and
 352 31.0 °C for the optimal temperature for growth of bacteria ($T_{opt,hb}$) (Fig. 10, 2011-02-06 to 2011-02-20).

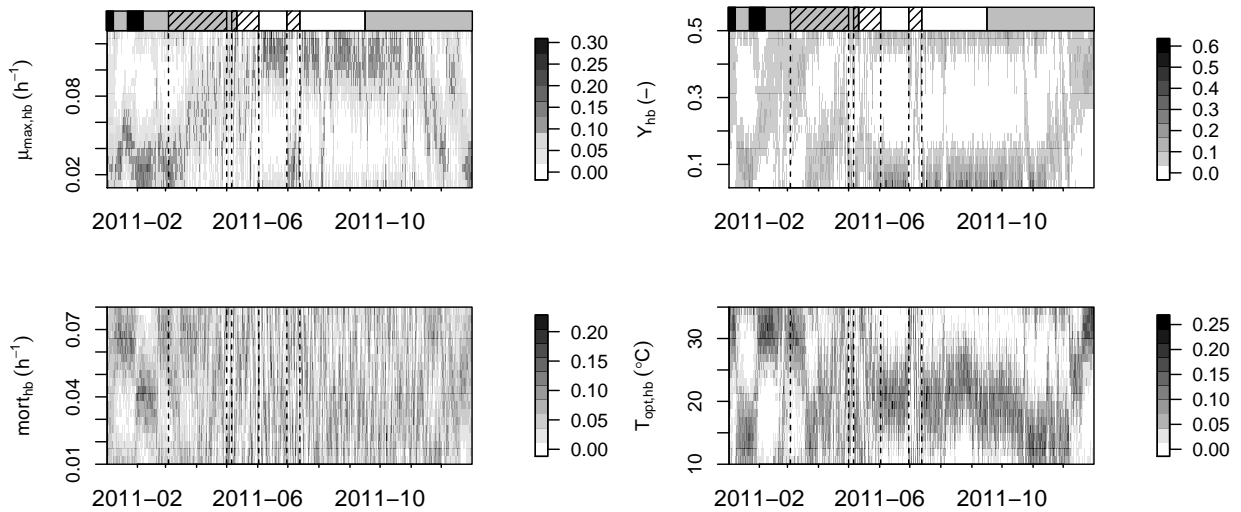


Fig. 9: Normalised importance weights of bacterial parameters. See Tab. 2 for parameter definition. Line shaded polygons correspond to algae bloom periods and water temperature increases from black to white. For the color gradient: black $T < 6$ °C, gray 6 °C $< T < 20$ °C, white $T > 20$ °C.

353 Between the bloom of May and the bloom of July, the bacterial activities are characterized by a high maximum
 354 growth rate ($\mu_{max,hb}$, 0.112 h⁻¹) and a low bacterial growth yield (Y_{hb} , 0.07, Fig. 9). The optimal temperature for the

355 growth of bacteria ($T_{opt,pp}$) is estimated at 20.7 °C.

356 After the bloom of July (before December), the physiological properties of the bacteria determined by particle
357 filter are similar to those identified between the bloom of May and the bloom of July (Fig. 9). At the end of the
358 simulation, the bacterial properties identified by particle filter are similar to those before the bloom of March (Fig. 9).

359 4. Discussion

360 4.1. Possible origin of mismatches between simulation and observation during critical periods

361 The presented RMSEs of 0.71 to 1.32 mgO₂/L are relatively high (cross-validation) when comparing the measured
362 and simulated DO concentrations (Tab. 3). As demonstrated in section 3.1, both ProSE-PA and ProSE overestimate
363 DO concentrations between the bloom of May and July (Fig. 4 and 5, 2011-06-03 to 2011-06-29) at five of the eight
364 high frequency monitoring stations (Colombes, Chatou, Sartrouville and in a lesser extent Bougival and Mericourt).
365 Slighter discrepancies of the same type are also observed in September and in October during dry periods, especially
366 at Sartrouville and Mericourt stations. Those mismatches may have two origins, more or less independent from one
367 another: uncertainties of BDOM concentrations in input data and particle filter degeneracy.

368 4.1.1. Particule Filter degeneracy

369 Periods when the mismatch between simulation and observation is large lead to the degeneracy of the particle
370 filter (section 2.2.2), meaning that during those periods of time almost all particles have a zero or small normalised
371 importance weight. The filter degeneracy is particularly important between the bloom of May and July (Fig. 11), and
372 in a lesser extent in September and in October during dry periods.

373 For any reason, the degeneracy effect indicates that the particle filter cannot correct the simulated DO concen-
374 trations by perturbing the model parameter values only. The constant N_{eff} for the three critical periods (Fig. 11)
375 illustrates that the differences between the simulated DO concentrations and the observations are so important that all
376 importance weights are close to zero (eq. (5)). To maintain the functioning of the particle filter, when all importance
377 weights goes to zero, the observation values at these moments are practically ignored. It technically explains why
378 high RMSEs occur during these critical periods.

379 As a matter of fact, the fact that the particle filter degenerates during critical periods of time invalidates one of
380 our main hypothesis, that is that model inputs, or boundary conditions, are perfectly known, and points out the need
381 for further research dedicated to the integration of uncertainties on these variables in the assimilation framework, and
382 especially the quantification of the biodegradable fraction of the organic matter.

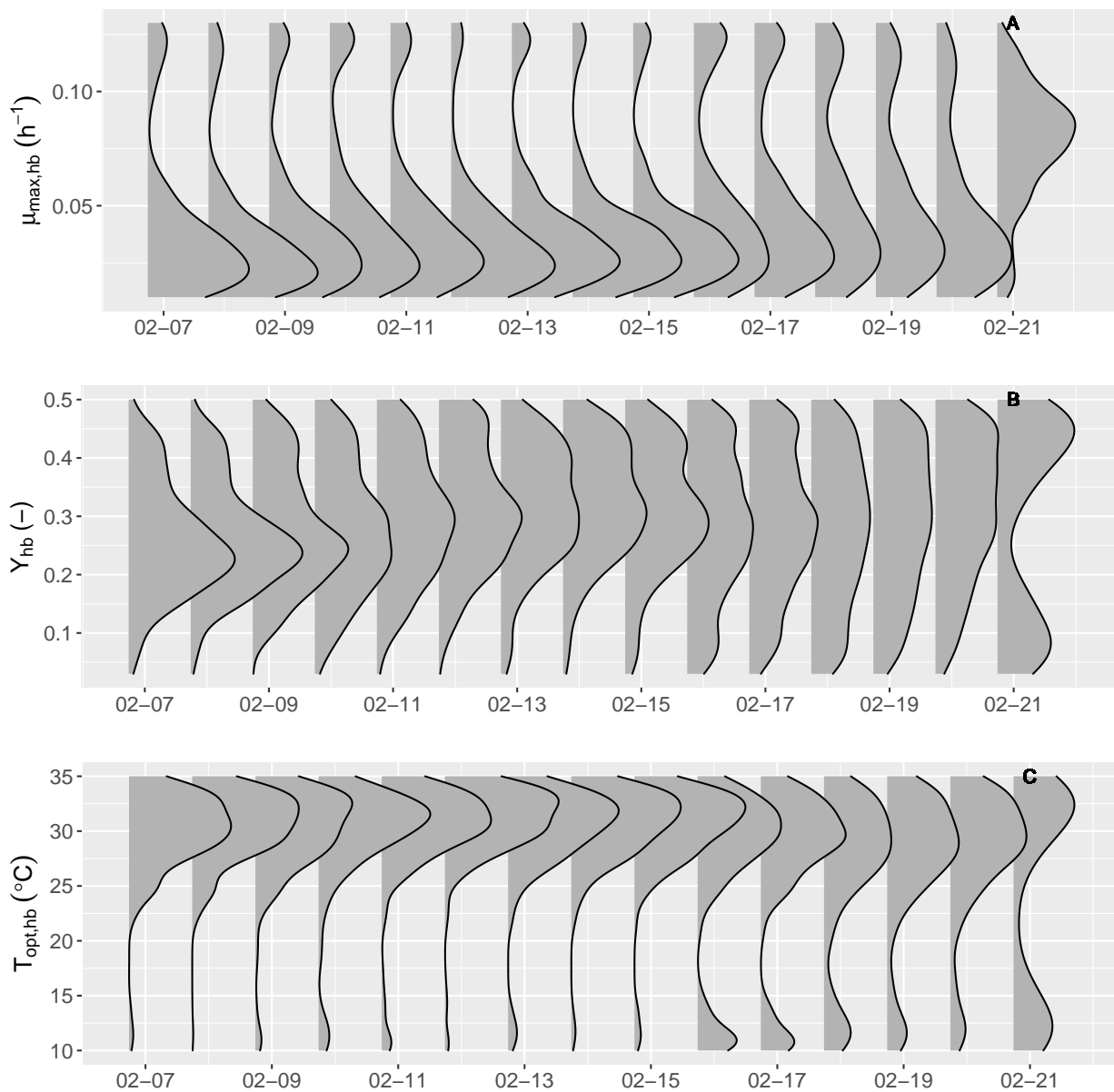


Fig. 10: Posterior pdfs of the bacterial parameters before the bloom of March.

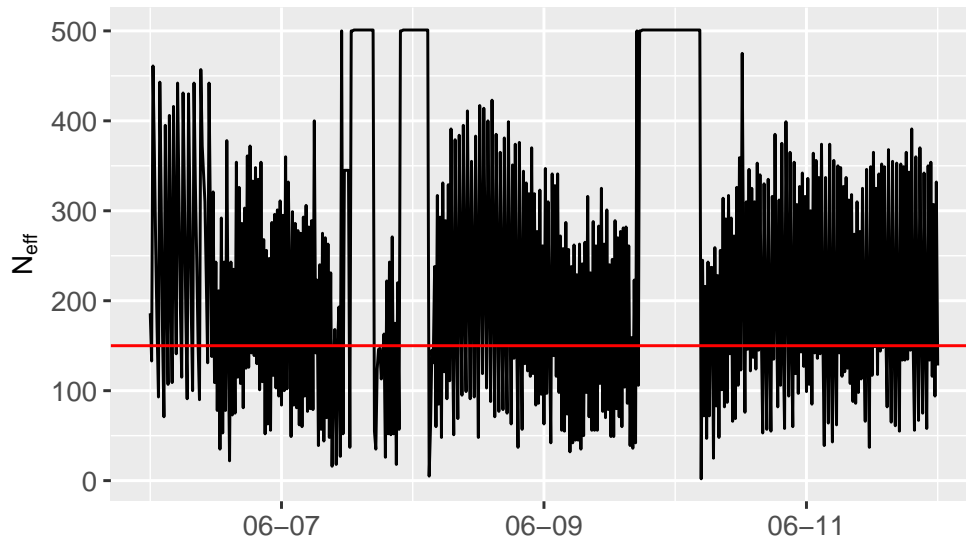


Fig. 11: Evolution of the effective sample size from 2011-06-05 to 2011-06-12. The red line corresponds to the minimum effective sample size of 150 below which a resampling procedure is carried out.

383 4.1.2. Description of BDOM in input data

384 During the critical periods mentioned above, although almost all the biodegradable organic matter is consumed by
 385 heterotrophic bacteria, the particle filter cannot mitigate the overestimation of oxygen. Before and after the bloom of
 386 July, it reaches its limit indeed by increasing $\mu_{max,hb}$ (maximal growth rate) value and decreasing Y_{hb} value (growth
 387 yield, Fig. 9). High maximum growth rate ($\mu_{max,hb}$, 0.112 h^{-1}) and low bacterial growth yield (Y_{hb} , 0.07) increase
 388 oxygen consumption by heterotrophic bacteria and decrease DO concentrations in water column but not sufficiently
 389 to reduce the gap between simulation and observation (Fig. 4 and 5).

390 Furthermore, the 95% confidence intervals of the DO concentrations simulated by ProSE-PA are very narrow
 391 during these periods (see for instance at Sartrouville station), which indicates that the model parameters have a low
 392 influence on DO concentrations.

393 It seems that the farther from an organic matter source the monitoring station is located, the larger the discrepancies
 394 between simulation and observation. This is true for the Suresnes to Sartrouville monitoring stations. ProSE-PA is
 395 able to simulate DO concentrations at Andresy station properly (a RMSE of $0.53 \text{ mgO}_2/\text{L}$ when using eight stations
 396 for data assimilation). This station is located downstream of two important organic matter sources, the Oise River
 397 and the SAV WWTP (Fig. 2), which enrich the Seine River in BDOM (Fig. 12). Further downstream, when all
 398 the biodegradable organic matter has been consumed (see the mismatch at the Mericourt station), the particle filter
 399 degenerates again.

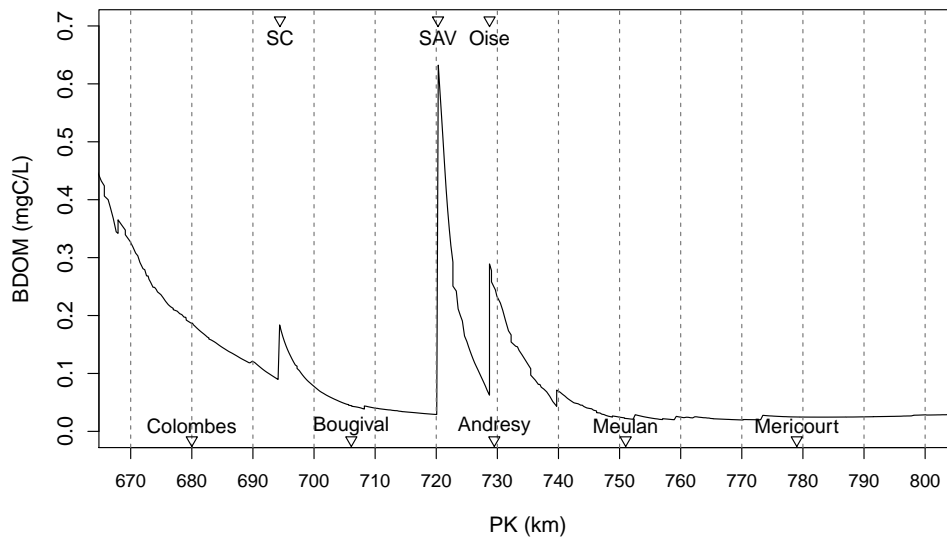


Fig. 12: Longitudinal profile of mean BDOM concentration from 2011-09-07 to 2011-09-27. Only left river reaches are presented here. PK: Kilometer Point, PK = 862 km at the entrance of the Seine river, Poses. SC: Seine Centre WWTP. SAV: Seine Aval WWTP. Oise: Oise river.

400 The overestimation of the DO concentrations during the critical periods can therefore be explained by a deficit of
 401 BDOM in the water column, as illustrated by the longitudinal profile of mean BDOM concentration from 2011-09-07
 402 to 2011-09-27 (Fig. 12). The BDOM are shortly degraded downstream of SAV WWTP and the Oise River. Simulated
 403 BDOM concentrations are less than 0.02 mgC/L at Meulan and Mericourt stations, which limits the bacterial activities
 404 (Fig. 12). This is the reason why the simulations of ProSE-PA and ProSE are similar during low flow periods without
 405 algal bloom.

406 ProSE-PA simulates almost the same oxygen level at Meulan and Mericourt stations in September and October.
 407 However, the depletion of measured DO is observed at Mericourt station (Fig. 5). It means that supplementary
 408 BDOM is required to produce this depletion at Mericourt station. The same phenomenon can also be observed from
 409 Colombes station to Bougival station. The simulated DO concentrations varying from 7 mgO₂/L to 8 mgO₂/L are
 410 similar at Colombes, Chatou and Bougival stations (from 2011-09-07 to 2011-09-27). The DO concentrations are well
 411 simulated by ProSE-PA at Colombes and Bougival stations at this period. But the depletion of DO concentrations (< 6
 412 mgO₂/L, Fig. 4) during this period are not reproduced at Chatou station located at a right reach of the Seine River. This
 413 depletion of oxygen results from the lateral inflows containing BDOM (overflow of CSOs/WWTPs) before Chatou
 414 station. This indicates that some unknown overflow sources are not considered on the right reach of the Seine river
 415 before Chatou station.

416 To conclude, the biodegradable organic matter carried by rivers, CSOs towards the Seine river is quite uncertain
 417 and dismiss the hypothesis of perfectly known boundary conditions. The uncertainties of the biodegradable portion

418 of DOM have been reported in numerous studies (Meyer et al., 1987; Servais et al., 1987, 1995, 1999; Søndergaard
419 and Middelboe, 1995; Wang et al., 2010; He et al., 2010; Kang and Mitchell, 2013). This portion depends on the
420 hydrological and trophic contexts of a river, therefore varying temporally. Generally, the range of BDOM portion of
421 DOM appears to be 0.05-0.54 (Hasanyar et al., 2020). The portion of BDOM used in this paper is 0.47 for rivers,
422 which seems a rather good estimate most of the time, but need to be refined when it leads to improper organic matter
423 depletion.

424 4.2. Synthesis of the phytoplanktonic parameters' values characterized by ProSE-PA (particle filter)

425 To interpret the identification of the phytoplanktonic parameters by ProSE-PA (particle filter), the averages of the
426 posterior pdfs' modes are calculated for the algal blooms of March and July (Tab. 4).

427 Two optimal temperatures for the growth of phytoplankton ($T_{opt,pp}$) are estimated for the algal bloom of March
428 (15.7 °C vs 24.7 °C respectively). A $T_{opt,pp}$ of 22.4 °C is identified for the bloom of July. The different physiological
429 properties identified by ProSE-PA reflect different phytoplanktonic communities in the Seine River. The algal species
430 of the Seine River have been investigated for the last 30 years by the PIREN-SEINE program. Groleau et al. (2014)
431 revealed that the phytoplanktonic species identified in March 2011 are *Synedra ulna*, *Cyclotella sp.*, *Nitzschia sp.* and
432 the specie identified in July is *Nitzschia sp.*. Garnier et al. (1995) highlighted that the spring algal bloom in Seine
433 River is mainly composed of a species of centric diatoms, *Stephanodiscus hantzschii*. Descy et al. (2012) mentioned
434 also that the *Stephanodiscus sp.* grows normally in early spring while the *Nitzschia sp.* grows at the end of spring or
435 in summer. This confirms the consistence of $T_{opt,pp}$ values identified for the blooms of March-April and July (24.7
436 °C vs 22.4 °C, probably *Nitzschia sp.*). Moreover, Descy et al. (2012) used 11 °C and 17 °C to model the different
437 phytoplanktonic communities growing in spring (*Stephanodiscus sp.*) and in summer (*Nitzschia sp.*) for the Loire
438 river while Vilmin (2014) obtained the calibrated values of 10 °C and 23 °C for the Seine River. These studies are in
439 agreement with the observed switch of $T_{opt,phy}$ of phytoplankton during the algal bloom of March-April.

440 To summarize, the parameter values identified by ProSE-PA are in agreement with those reported by Descy et al.
441 (2012); Vilmin (2014), and Garnier et al. (1995) for some periods and with the prior knowledge of phytoplanktonic
442 communities on Seine River Groleau et al. (2014). The different parameter values indicate the necessity of a time-
443 varying parameter setting for simulating algal bloom dynamics in river systems properly.

444 5. Conclusions

445 A particle filter is applied for the first time to assimilate high frequency dissolved oxygen data in river system
446 and to characterize the physiological properties of living communities (represented by model parameter values) under

Table 4: Average modes of the phytoplanktonic parameters compared to the values used in previous studies

Parameters	ProSE-PA			Vilmin (2014)		Garnier et al. (1995)
	March	April	July	March	July	Diatoms
$T_{opt,pp}$ [°C]	15.7	24.7	22.4	10.0	23.0	21.0
$R_{m,pp}$ [h^{-1}]	0.015	0.006	0.015	0.002	0.002	0.002
$P_{max,pp}$ [h^{-1}]	0.39	0.27	0.42	0.20	0.20	0.20
α_{pp} [$m^2s.\mu E^{-1}.h^{-1}$]	0.0015	-	0.0013	0.0012	0.0012	0.0012
$\eta_{chla,pp}$ [$L.\mu gchla^{-1}.m^{-1}$]	0.034	0.019	0.032	0.020	0.020	0.030
$C/Chla$ [$mgC/\mu gchla$]	0.026	0.055	0.035	0.035	0.035	0.035

-: Oscillating modes

447 contrasted hydrological and trophic contexts. In spite of the assumption that the boundary conditions (input water
 448 quality data) are perfectly known, this article shows that the particle filter improves significantly the simulation of the
 449 DO concentrations in the Seine River. Particularly, particle filter allows a more accurate simulation of the algal bloom
 450 dynamics which is very difficult to reproduce using forward model configured with "static parameter".

451 The different physiological properties of phytoplankton can be characterized using particle filter. The identified
 452 phytoplanktonic parameter values during the algal blooms are consistent with the values reported in the literature and
 453 with the prior knowledge on the phytoplanktonic communities in the Seine River system. This paper demonstrates the
 454 efficiency of the particle filter to predict the dissolved oxygen concentrations and to capture the variability of phyto-
 455 planktonic parameters during algal blooms which cannot be described by a forward model that uses static parameters
 456 only. This study demonstrates the necessity of a time-varying parameter setting for simulating microorganism activi-
 457 ties.

458 Nevertheless, further improvement of DO concentration simulated by the particle filter can be achieved by intro-
 459 ducing a time-varying BDOM content in the boundary conditions represented by rivers, CSOs, and WWTPs. Such an
 460 advance should also improve the capability of ProSE-PA to characterize heterotrophic bacteria properties (parameters)
 461 during low flow without algae bloom.

462 **Declarations of interest**

463 The authors declare that there are no known conflicts of interest associated with this work and there has been no
 464 significant financial supports for this paper that could have influenced its outcome. The authors alone are responsible
 465 for the content and writing of this article.

466 **Acknowledgements**

467 This work is a contribution to the PIREN-SEINE research program, part of the french Long Term Socio-Ecological
468 Research (LTSER) site "Zones Ateliers Seine". We would like to thank the data providers. The high frequency oxygen
469 observations from MeSeine network were provided by the public sewage company of the Greater Paris area (SIAAP)
470 and the low frequency oxygen data from the national river monitoring network (RCS) was supplied by the drinking
471 water company of the Paris urban area (SEDIF).

472 **References**

- 473 Abbaszadeh, P., Moradkhani, H., Yan, H., 2018. Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte
474 Carlo. *Advances in Water Resources* 111, 192–204.
- 475 Andreadis, K., Clark, E., Lettermaier, D., Alsdorf, D., 2007. Prospects for river discharge and depth estimation through assimilation of swath-
476 altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters* 34, L10403.
- 477 Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* 53,
478 370–418.
- 479 Beck, M., 1978. Random signal analysis in an environmental sciences problem. *Applied Mathematical Modelling* 2 (1), 23–29.
- 480 Beck, M., Halfon, E., 1991. Uncertainty, identifiability and the propagation of prediction errors: A case study of lake ontario. *Journal of Forecasting*
481 10 (1-2), 135–161.
- 482 Beck, M., Young, P., 10 1976. Systematic identification of do-bod model structure. *Jnl. of Env. Eng. Division , American Society of Civil Engineers*
483 102, 909–927.
- 484 Beven, K., 1989. Changing ideas in hydrology. The case of physically-based model. *Journal of Hydrology* 105, 157–172.
- 485 Billen, G., Garnier, J., Hanset, P., 1994. Modelling phytoplankton development in whole drainage networks: the RIVERSTRAHLER Model applied
486 to the Seine river system. *Hydrobiologia* 289, 119–137.
- 487 Bowles, D., Grenney, W., 1978. Steady state river quality modeling by sequential extended kalman filters. *Water Resources Research* 14 (1), 84–96.
- 488 Carrassi, A., Bocquet, M., Bertino, L., Evensen, G., 2018. Data assimilation in the geosciences: An overview of methods, issues, and perspectives.
489 *Wiley Interdisciplinary Reviews: Climate Change* 9 (5), e535.
- 490 Chen, C., Huang, J., Chen, Q., Zhang, J., Li, Z., Lin, Y., 2019. Assimilating multi-source data into a three-dimensional hydro-ecological dynamics
491 model using ensemble kalman filter. *Environmental Modelling & Software* 117, 188 – 199.
- 492 Cho, K. H., Pachevsky, Y., Ligaray, M., Kwon, Y., Kim, K. H., 2020. Data assimilation in surface water quality modeling: A review. *Water Research*
493 186, 116307.
- 494 Cosby, B., Hornberger, G., 1984. Identification of photosynthesis-light models for aquatic systems i. theory and simulations. *Ecological Modelling*
495 23 (1-2), 1–24.
- 496 Cosby, B., Hornberger, G., Kelly, M., 1984. Identification of photosynthesis-light models for aquatic systems ii. application to a macrophyte
497 dominated stream. *Ecological Modelling* 23 (1), 25 – 51, modelling Primary Production.
- 498 Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., Fisher, M., Pailleux, J., 1998. The ECMWF
499 implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quarterly Journal of the Royal Meteorological Society*
500 124 (550), 1783–1807.

501 Courtier, P., Thépaut, J. N., Hollingsworth, A., 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. Quarterly
502 Journal of the Royal Meteorological Society 120 (519), 1367–1387.

503 Descy, J.-P., Leitao, M., Everbecq, E., Smitz, J.-S., Delière, J.-F., 2012. Phytoplankton of the River Loire, France: a biodiversity and modelling
504 study. Journal of Plankton Research 34 (2), 120–135.

505 DHI, 2007. MIKE 11. A Modelling System for Rivers and Channels. Reference Manual. DHI.

506 DHI, 2017a. DHI Simulation Engine for 1D river and urban modelling. Reference Manual. DHI.

507 DHI, 2017b. MIKE HYDRO Basin Water Quality Model-MIKE ECO Lab Template Documentation. Scientific Documentation. DHI.

508 Doucet, A., de Freitas, N., Gordon, N., 2001. Sequential Monte Carlo Methods in Practice. Springer, New York.

509 Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential monte carlo sampling methods for bayesian filtering. Statistics and Computing 10 (3),
510 197–208.

511 Ennola, K., Sarvala, J., Dévai, G., 1998. Modelling zooplankton population dynamics with the extended kalman filtering technique. Ecological
512 Modelling 110 (2), 135 – 149.

513 Even, S., Bacq, N., Ruelland, D., Billen, G., Garnier, J., Poulin, M., Théry, S., Blanc, S., 2007a. New tools for modelling water quality of
514 hydrosystems: An application in the Seine River basin in the frame of the Water Framework Directive. Sciences of Total Environment 375 (1-
515 3), 274–291.

516 Even, S., Mouchel, J. M., Servais, P., Flipo, N., Poulin, M., Blanc, S., Chabanel, M., Paffoni, C., 2007b. Modeling the impacts of Combined Sewer
517 Overflows on the river Seine water quality. Sciences of Total Environment 375 (1-3), 140–151.

518 Even, S., Poulin, M., Garnier, J., Billen, G., Servais, P., Chesterikoff, A., Coste, M., 1998. River ecosystem modelling: Application of the ProSe
519 model to the Seine river (France). Hydrobiologia 373, 27–37.

520 Even, S., Poulin, M., Mouchel, J. M., Seidl, M., Servais, P., 2004. Modelling oxygen deficits in the Seine river downstream of combined sewer
521 overflows. Ecol. Model. 173, 177–196.

522 Flipo, N., Even, S., Poulin, M., Tusseau-Vuillemin, M. H., Améziane, T., Dauta, A., 2004. Biogeochemical modelling at the river scale: Plankton
523 and periphyton dynamics - Grand Morin case study, France. Ecological Modelling 176, 333–347.

524 Flipo, N., Rabouille, C., Poulin, M., Even, S., Tusseau-Vuillemin, M., Lalande, M., 2007. Primary production in headwater streams of the Seine
525 basin: the Grand Morin case study. Sciences of Total Environment 375, 98–109.

526 Garnier, J., Billen, G., Coste, M., 1995. Seasonal succession of diatoms and chlorophyceae in the drainage network of the river Seine: Observations
527 and modelling. Limnology and Oceanography 40 (4), 750–765.

528 Garnier, J., Marescaux, A., Guillon, S., Vilmin, L., Rocher, V., Billen, G., Thieu, V., Silvestre, M., Passy, P., Raimonet, M., Groleau, A., Théry, S.,
529 Tallec, G., Flipo, N., 2020. The Handbook of Environmental Chemistry. Handbook of Environmental Chemistry. Springer, Berlin, Heidelberg,
530 Ch. Ecological Functioning of the Seine River: From Long-Term Modelling Approaches to High-Frequency Data Analysis, pp. 1–28.

531 Gauthier, P., Tanguay, M., Laroche, S., Pellerin, S., Morneau, J., 2007. Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the
532 Meteorological Service of Canada. Monthly Weather Review 135 (6), 2339–2354.

533 Gharamti, M., Tjiputra, J., Bethke, I., Samuelsen, A., Skjelvan, I., Bentsen, M., Bertino, L., 2017. Ensemble data assimilation for ocean biogeo-
534 chemical state and parameter estimation at different sites. Ocean Modelling 112, 65 – 89.

535 Groleau, A., Escoffier, N., Bensoussan, N., Flipo, N., Poulin, M., Vilmin, L., Bernard, C., Rocher, V., Métivier, F., 2013. Réseau de mesures haute
536 fréquence CarboSeine : validation métrologique et indicateurs du fonctionnement trophique. Tech. rep., PIREN-Seine.

537 Groleau, A., Escoffier, N., Vilmin, L., Auge, A., Bensoussan, N., Poulin, M., Rocher, V., Métivier, F., Flipo, N., 2014. Apports croisés des mesures
538 haute fréquence pour le calcul d'indicateurs et pour la modélisation du fonctionnement trophique de la Seine à l'aval de Paris. Tech. rep.,
539 PIREN-Seine.

540 Guo, H., 06 2003. A stochastic water quality forecasting system for the yiluo river. *Journal of Environmental Informatics - J ENVIRON INFORM*
541 1, 18–32.

542 Hasanyar, M., Flipo, N., Romary, T. and Wang, S., Yari, A., 2020. Rôle de la matière organique dans le métabolisme des rivières à bas débit. *Rapport*
543 PIREN-Seine, PIREN-Seine.

544 He, B., Dai, M., Zhai, W., Wang, L., Wang, K., Chen, J., Lin, J., Han, A., Xu, Y., 2010. Distribution, degradation and dynamics of dissolved organic
545 carbon and its major compound classes in the Pearl River estuary, China. *Marine Chemistry* 119 (1-4), 52–64.

546 Huang, J., Gao, J., 2017. An improved ensemble kalman filter for optimizing parameters in a coupled phosphorus model for lowland polders in
547 lake taihu basin, china. *Ecological Modelling* 357, 14 – 22.

548 Huang, J., Gao, J., Liu, J., Zhang, Y., 2013. State and parameter update of a hydrodynamic-phytoplankton model using ensemble Kalman filter.
549 *Ecological Modelling* 263, 81–91.

550 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M.,
551 Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., Joseph, D., 1996. The
552 ncep/ncar 40-year reanalysis project. *Bulletin of the American Meteorological Society* 77 (3), 437–472.

553 Kang, P.-G., Mitchell, M. J., 2013. Bioavailability and size-fraction of dissolved organic carbon, nitrogen, and sulfur at the Arbutus Lake watershed,
554 Adirondack Mountains, NY. *Biogeochemistry* 115 (1-3), 213–234.

555 Kattwinkel, M., Reichert, P., 2017. Bayesian parameter inference for individual-based models using a Particle Markov Chain Monte Carlo method.
556 *Environmental Modelling & Software* 87, 110–119.

557 Kim, K., Park, M., Min, J., Ryu, I., Kang, M., Park, L. J., 2014. Simulation of algal bloom dynamics in a river with the ensemble Kalman filter.
558 *Journal of Hydrology* 519, 2810–2821.

559 Kitagawa, G., 1996. Monte Carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical*
560 *Statistics* 5 (1), 1–25.

561 Kleist, D. T., Ide, K., 2015. An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II: 4DnEnVar and
562 Hybrid Variants. *Monthly Weather Review* 143 (2), 452–470.

563 Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper danube basin under an ensemble of climate change scenarios. *Journal of*
564 *Hydrology* 424-425, 264 – 277.

565 Kong, A., Liu, J. S., Wong, W. H., 1994. Sequential imputations and Bayesian missing data problems. *Journal of the American statistical association*
566 89 (425), 278–288.

567 Li, T., Villarrubia, G., Sun, S., Corchado, J. M., Bajo, J., 2015. Resampling methods for particle filtering: identical distribution, a new method, and
568 comparable study. *Frontiers of Information Technology & Electronic Engineering* 16 (11), 969–984.

569 Liu, J. S., 2001. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, USA.

570 Loos, S., Shin, C. M., Sumihar, J., Kim, K., Cho, J., Weerts, A. H., 2020. Ensemble data assimilation methods for improving river water quality
571 forecasting accuracy. *Water Research* 171, 115343.

572 Mao, J., Lee, J. H., Choi, K., 2009. The extended Kalman filter for forecast of algal bloom dynamics. *Water Research* 43 (17), 4214 – 4224.

573 Markov, A. A., 1906. Extension of the law of large numbers to dependent quantities (in russian). *Izvestiia Fiz.-Matem. Obsch. Kazan Univ.*, (2nd
574 Ser.) 15, 135–156.

575 Meyer, J. L., Edwards, R. T., Risley, R., 1987. Bacterial growth on dissolved organic carbon from a blackwater river. *Microbial Ecology* 13 (1),
576 13–29.

577 Moradkhani, H., DeChant, C. M., Sorooshian, S., 2012. Evolution of ensemble data assimilation for uncertainty quantification using the particle
578 filter-Markov chain Monte carlo method. *Water Resources Research* 48 (12).

579 Moradkhani, H., Hsu, K., Gupta, H., Sorooshian, S., 2005a. Uncertainty assessment of hydrologic model states and parameters: Sequential data
580 assimilation using the particle filter. *Water Resources Research* 41 (5).

581 Moradkhani, H., Sorooshian, S., Gupta, H. V., Houser, P. R., 2005b. Dual state-parameter estimation of hydrological models using ensemble
582 Kalman filter. *Advances in Water Resources* 28 (2), 135–147.

583 Ottlé, C., Vidal-Madjar, D., 1994. Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the hapex-
584 mobilhy region. *Journal of Hydrology* 158 (3), 241 – 264.

585 Page, T., Smith, P. J., Beven, K. J., Jones, I. D., Elliott, J. A., Maberly, S. C., Mackay, E. B., Ville, M. D., Feuchtmayr, H., 2018. Adaptive
586 forecasting of phytoplankton communities. *Water Research* 134, 74 – 85.

587 Paiva, R. C. D., Collischonn, W., Bonnet, M.-P., de Gonçalves, L. G. G., Calmant, S., Getirana, A., Santos da Silva, J., 2013. Assimilating in situ
588 and radar altimetry data into a large-scale hydrologic-hydrodynamic model for streamflow forecast in the amazon. *Hydrology and Earth System
589 Sciences* 17 (7), 2929–2946.

590 Park, R., I. Connolly, C., R. Albanese, J., S. Clesceri, L., W. Heitzman, G., H. Herbrandson, H., H. Indyke, B., R. Loehe, J., Ross, S., D. Sharma,
591 D., W. Shuster, W., 1982. Modeling the fate of toxic organic materials in aquatic environments. Rept. EPA-600/S3-82-028. USEPA, Athens,
592 GA.

593 Park, R., O’Neill, R., Bloomfield, J., Shugart, H., Booth, R., Koonce, J., Adams, M., Clesceri, L., Colon, E., Dettman, E., Goldstein, R., Hoopes,
594 J., Huff, D., Katz, S., Kitchell, J., Kohberger, R., LaRow, E., McNaught, D., Peterson, J., Zahorcak, C., 01 1974. A generalized model for
595 simulating lake ecosystems. *Simulation* 23, 30–50.

596 Park, S. S., Lee, Y. S., 2002. A water quality modeling study of the nakdong river, korea. *Ecological Modeling* 152, 65–75.

597 Pastres, R., Ciavatta, S., Solidoro, C., 2003. The extended Kalman Filter (EKF) as a tool for the assimilation of high frequency water quality data.
598 *Ecological Modelling* 170 (2), 227–235.

599 Pelletier, G., Chapra, S., Tao, H., 03 2006. QUAL2Kw — a framework for modeling water quality in streams and rivers using a genetic algorithm
600 for calibration. *Environmental Modelling and Software*, 419–425.

601 Plaza, D. A., De Keyser, R., De Lannoy, G. J. M., Giustarini, L., Matgen, P., Pauwels, V. R. N., 2012. The importance of parameter resampling for
602 soil moisture data assimilation into hydrologic models using the particle filter. *Hydrology and Earth System Sciences* 16 (2), 375–390.

603 Polus, E., Flipo, N., de Fouquet, C., Poulin, M., 2011. Geostatistics for assessing the efficiency of distributed physically-based water quality model.
604 Application to nitrates in the Seine River. *Hydrological Processes* 25 (2), 217–233.

605 Raimonet, M., Thieu, V., Silvestre, M., Oudin, L., Rabouille, C., Vautard, R., Garnier, J., 2018. Landward perspective of coastal eutrophication
606 potential under future climate change: The seine river case (france). *Frontiers in Marine Science* 5, 136.

607 Raimonet, M., Vilmin, L., Flipo, N., Rocher, V., Laverman, A., 2015. Modelling the fate of nitrite in an urbanized river using experimentally
608 obtained nitrifier growth parameters. *Water Research* 73, 373–387.

609 Reichert, P., Borchardt, D., Henze, M., Rauch, W., Shanahan, P., Somlyódy, L., Vanrolleghem, P., 2001. River Water Quality Model no.1. Tech.
610 rep., IWA, London, UK.

611 Reichert, P., Mieleitner, J., 2009. Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent param-
612 eters. *Water Resources Research* 45 (10).

613 Rocher, V., Garcia-Gonzalez, E., Paffoni, C., Thomas, W., 2011. La production de nitrites lors de la dénitrification des eaux usées: un sujet sensible
614 et complexe ! *L’Eau, l’Industrie, les Nuisances* 344, 80–83.

615 Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M.,
616 Entin, J. K., Walker, J. P., Lohmann, D., Toll, D., 2004. The global land data assimilation system. *Bulletin of the American Meteorological
617 Society* 85 (3), 381 – 394.

618 Ruelland, D., Billen, G., Brunstein, D., Garnier, J., 2007. Seneque: a multi-scaling gis interface to the riverstrahler model of the biogeochemical
619 functioning of river systems. *The Science of the total environment* 375 (1-3), 257–73.

620 Salamon, P., Feyen, L., 2009. Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential
621 data assimilation with the particle filter. *Journal of Hydrology* 376, 428–442.

622 Särkkä, S., 2013. *Bayesian Filtering and Smoothing*. Cambridge University Press, UK.

623 Servais, P., Billen, G., Hascoët, M.-C., 1987. Determination of the biodegradable fraction of dissolved organic matter in waters. *Water Research*
624 21 (4), 445–450.

625 Servais, P., Garnier, J., Demarteau, N., Brion, N., Billen, G., 1999. Supply of organic matter and bacteria to aquatic ecosystems through waste
626 water effluents. *Water Research* 33, 3521–3531.

627 Servais, P., Laurent, P., Billen, G., Gatel, D., 1995. Development of a model of bcod and bacterial biomass fluctuations in distribution systems.
628 *Revue des Sciences de l'Eau* 8, 427–462.

629 Shanahan, P., Borchardt, D., Henze, M., Rauch, W., Reichert, P., Somlyódy, L., Vanrolleghem, P., 2001. River Water Quality Model no. 1
630 (RWQM1): I. Modelling approach. *Water Science and Technology* 43 (5), 1–9.

631 Simon, E., Bertino, L., 2012. Gaussian anamorphosis extension of the denkf for combined state parameter estimation: Application to a 1D ocean
632 ecosystem model. *Journal of Marine Systems* 89 (1), 1 – 18.

633 Simon, E., Samuelsen, A., Bertino, L., Dumont, D., 2012. Estimation of positive sum-to-one constrained zooplankton grazing preferences with the
634 denkf: a twin experiment. *Ocean Science* 8 (4), 587–602.

635 Soetaert, K., Gregoire, M., 2011. Estimating marine biogeochemical rates of the carbonate ph system—a kalman filter tested. *Ecological Modelling*
636 222 (12), 1929 – 1942.

637 Søndergaard, M., Middelboe, M., 1995. A cross-system analysis of labile dissolved organic carbon. *Marine Ecology Progress Series* 118, 283–294.

638 Streeter, H., Phelps, E. B., 1925. A study of the pollution and natural purification of the Ohio River. Tech. Rep. 146, U.S. Public Health Service,
639 Treasury Department, Washington DC.

640 Thieu, V., Billen, G., Garnier, J., 2009. Nutrient transfer in three contrasting NW European watersheds: The Seine, Somme, and Scheldt Rivers. A
641 comparative application of the Seneque/Riverstrahler model. *Water Research* 43, 1740–1754.

642 Vanrolleghem, P., Borchardt, D., Henze, M., Rauch, W., Reichert, P., Shanahan, P., Somlyódy, L., 2001. River Water Quality Model no.1 (rwqm1):
643 III biochemical submodel selection. *Water Science and Technology* 43 (5), 31–40.

644 Vilmin, L., 2014. Modélisation du fonctionnement biogéochimique de la seine de l'agglomération parisienne à l'estuaire à différentes échelles
645 temporelles. Ph.D. thesis, MINES ParisTech, France.

646 Vilmin, L., Aissa-Grouz, N., Garnier, J., Billen, G., Mouchel, J. M., Poulin, M., Flipo, N., 2015a. Impact of hydro-sedimentary processes on the
647 dynamics of soluble reactive phosphorus in the Seine River. *Biogeochemistry* 122, 229–251.

648 Vilmin, L., Flipo, N., de Fouquet, C., Poulin, M., 2015b. Pluri-annual sediment budget in a navigated river system: The Seine River (France).
649 *Sciences of Total Environment* 502, 48–59.

650 Vilmin, L., Flipo, N., Escoffier, N., Groleau, A., 2018. Estimation of the water quality of a large urbanized river as defined by the european WFD:
651 what is the optimal sampling frequency? *Environmental Science and Pollution Research* 25 (24), 23485–23501.

652 Vilmin, L., Flipo, N., Escoffier, N., Rocher, V., Groleau, A., 2016. Carbon fate in a large temperate human-impacted river system: Focus on benthic
653 dynamics. *Global Biogeochem. Cycles* 30 (7), 1086–1104.

654 Vrugt, J. A., ter Braak, C. J., Diks, C. G., Schoups, G., 2013. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation:
655 Theory, concepts and applications. *Advances in Water Resources* 51, 457–478.

656 Wang, S., Flipo, N., Romary, T., 2018. Time-dependent global sensitivity analysis of the C-RIVE biogeochemical model in contrasted hydrological

657 and trophic contexts. *Water Research* 144, 341–355.

658 Wang, S., Flipo, N., Romary, T., 2019. Oxygen data assimilation for estimating micro-organism communities' parameters in river systems. *Water*
659 *Research* 165, 115021.

660 Wang, X., Chen, X., Liu, S., Ge, X., 2010. Effect of molecular weight of dissolved organic matter on toxicity and bioavailability of copper to
661 lettuce. *Journal of Environmental Sciences* 22 (12), 1960–1965.

662 Warn, A. E., 1987. SIMCAT-a catchment simulation model for planning investment for river quality. In: Beck, M. B. (Ed.), *Systems Analysis in*
663 *Water Quality Management*. Oxford: IAWPRC, Pergamon, pp. 211–218.

664 Weerts, A. H., El Serafy, G. Y. H., 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-
665 runoff models. *Water Resources Research* 42 (9).

666 Whitehead, P., 1978. Modelling and operational control of water quality in river systems. *Water Research* 12 (6), 377–384.

667 Whitehead, P., Beck, B., O'Connell, E., 1981. A systems model of streamflow and water quality in the bedford ouse river system-ii. water quality
668 modelling. *Water Research* 15 (10), 1157–1171.

669 Whitehead, P., Hornberger, G., 1984. Modelling algal behaviour in the river thames. *Water Research* 18 (8), 945–953.

670 Whitehead, P. G., Williams, R. J., Lewis, D. R., 1997. Quality simulation along river systems (quasar): model theory and development. *Sciences of*
671 *Total Environment* 194/195, 447–456.

672 Yu, L., Fennel, K., Bertino, L., Gharamti, M. E., Thompson, K. R., 2018. Insights on multivariate updates of physical and biogeochemical ocean
673 variables using an ensemble kalman filter and an idealized model of upwelling. *Ocean Modelling* 126, 13 – 28.

674 Yucel, I., Onen, A., Yilmaz, K., Gochis, D., 2015. Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction
675 model, data assimilation and satellite-based rainfall. *Journal of Hydrology* 523, 49–66.

2. The filter improves significantly the simulation of oxygen and algae dynamics
3. The filter can capture time-varying phytoplanktonic parameters during algal blooms
4. The description of BDOM data need to be further investigated for low flow periods

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: