Particle filter for high frequency oxygen data assimilation in river systems

Shuaitao Wang, Nicolas Flipo, Thomas Romary, Masihullah Hasanyar

Particle filter for high frequency oxygen data assimilation in river systems

Shuaitao Wang[a,1,*], Nicolas Flipo[a], Thomas Romary[a], Masihullah Hasanyar[a]

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

Abstract

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

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

Highlights

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

Software availability

Name of software : ProSe-PA

*Corresponding author
Email address: shuaitaowang@outlook.com (Shuaitao Wang)
[1]Current address: Sorbonne Université, CNRS, EPHE, UMR 7619 Metis, 4 place Jussieu 75005 Paris, France (shuaitao.wang@sorbonne-universite.fr)

Contact address: nicolas.flipo@mines-paristech.fr

Year first available: 2019

Program language: ANSI C

Operating system: Linux

Software access: Gitlab deposit under discussion

Availability and cost: Open source

Licence: Open licence under discussion

1. Introduction

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

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

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

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

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

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

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

[c2] *SW: Text added.*

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

[c1] *SW: Text added.*

[c2] *SW: Text added.*

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

2. Material and methods

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

2.1. The ProSe-PA *software*

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

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

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

2.2. *Data assimilation framework*

2.2.1. *Particle Filter*

Sequential data assimilation aims at integrating observations $\mathbf{y}_t^*$ at each time step t into the forward model by conditioning the values of the state variable $\mathbf{z}_t$ (namely DO concentrations and model parameters x_t) to the observed 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 summed up by the Bayesian filtering equation :

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

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 variables 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 likelihood), $f(\mathbf{z}_t|\mathbf{y}_{1:t-1}^*)$ is the conditional probability of the state variables knowing the observations up to time $t-1$ (prior 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 variables $\mathbf{z}_t$ to the observations up to time t, we need to know how to compute the likelihood of the observations given 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 the forward model.

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

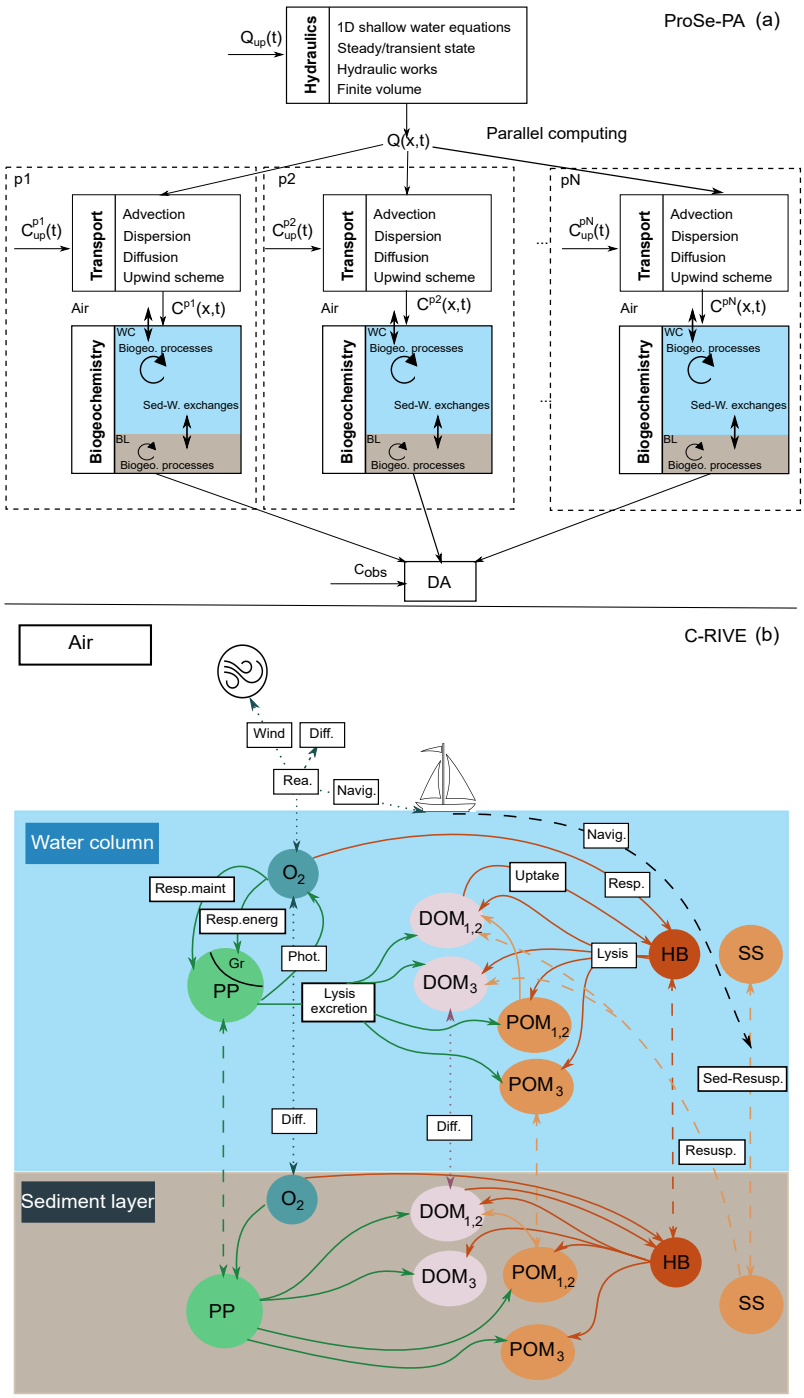

Fig. 1: (a) Schematic description of the ProSe-PA program with the hydraulic, transport, biogeochemical modules and the flowchart of particle filter (particles p1, p2, ... , pN). (b) Simplified flowchart of the biogeochemical model (Wang et al., 2018). Dotted lines: diffusive exchanges (Diff.); dashed lines: sedimentation-resuspension (Sed-Resusp.); solid lines: biogeochemical processes; PP: primary producer; Gr: growth; HB heterotrophic bacteria; SS: suspended solids; DOM: disolved 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.

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

$$\omega^i_t = \frac{f(\mathbf{z}^i_t|\mathbf{y}^*_{1:t})}{\pi(\mathbf{z}^i_t|\mathbf{y}^*_{1:t})} \propto \frac{f(\mathbf{y}^*_t|\mathbf{z}^i_t) f(\mathbf{z}^i_t|\mathbf{z}^i_{t-1}, \mathbf{y}^*_{1:t-1}) f(\mathbf{z}^i_{t-1}|\mathbf{y}^*_{1:t-1})}{\pi(\mathbf{z}^i_t|\mathbf{z}^i_{t-1}, \mathbf{y}^*_t)\pi(\mathbf{z}^i_{t-1}|\mathbf{y}^*_{1:t-1})} \tag{2}$$

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

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

$$\begin{aligned} \omega^i_t &\propto f(\mathbf{y}^*_t|\mathbf{z}^i_t)\omega^i_{t-1} \\ \hat{\omega}^i_t &= \frac{\omega^i_t}{\sum \omega^i_t} \end{aligned} \tag{3}$$

where ω^i_t and ω^i_{t-1} represent the posterior and prior importance weights at time t. $\hat{\omega}^i_t$ is the normalised importance weight.

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

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

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

Assuming that the observation error is Gaussian, the likelihood of each particle i is given by the pdf of the multivariate normal distribution.

$$\begin{aligned} \ln L(\mathbf{y}^*_t|\mathbf{z}^i_t) &= -\frac{m}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Sigma|) - \frac{1}{2}(\mathbf{y}^*_t - h(\mathbf{z}^i_t))^{\mathrm{T}}\Sigma^{-1}(\mathbf{y}^*_t - h(\mathbf{z}^i_t)) \\ f(\mathbf{y}^*_t|\mathbf{z}^i_t) &= \frac{L(\mathbf{y}^*_t|\mathbf{z}^i_t)}{\sum_{i=1}^{N} L(\mathbf{y}^*_t|\mathbf{z}^i_t)} \end{aligned} \tag{5}$$

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

2.2.2. *Resampling procedure*

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

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

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

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

where s is a user-defined perturbation proportional to the parameter ranges, $\mathbf{\Phi}_i$.

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

2.3. *Study area and available data*

2.3.1. *Study area and simulation period*

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

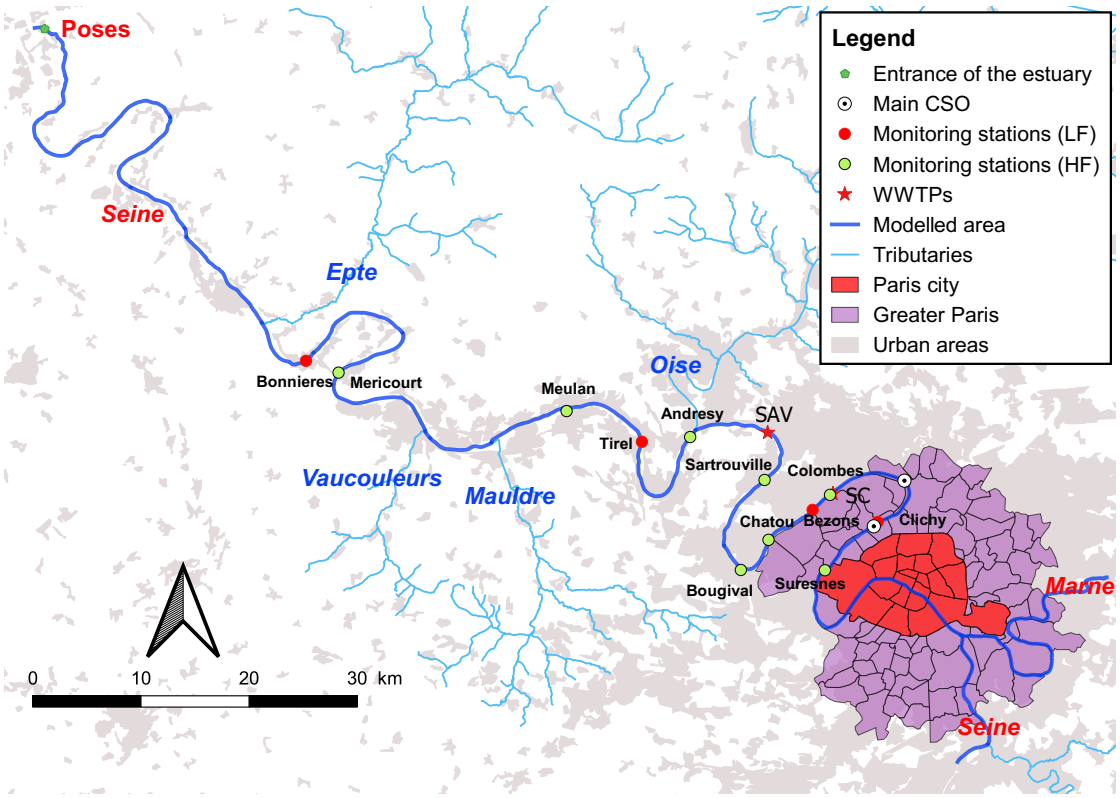

Fig. 2: The description of study area and monitoring sites. Monitoring stations (High frequency, in green) from upstream to downstream: 1. Suresnes, 2. Colombes, 3. Chatou, 4.Bougival, 5. Sartrouville, 6. Andresy, 7. Meulan, 8. Mericourt. Monitoring stations (Low frequency, in red) from upstream to downstream: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

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

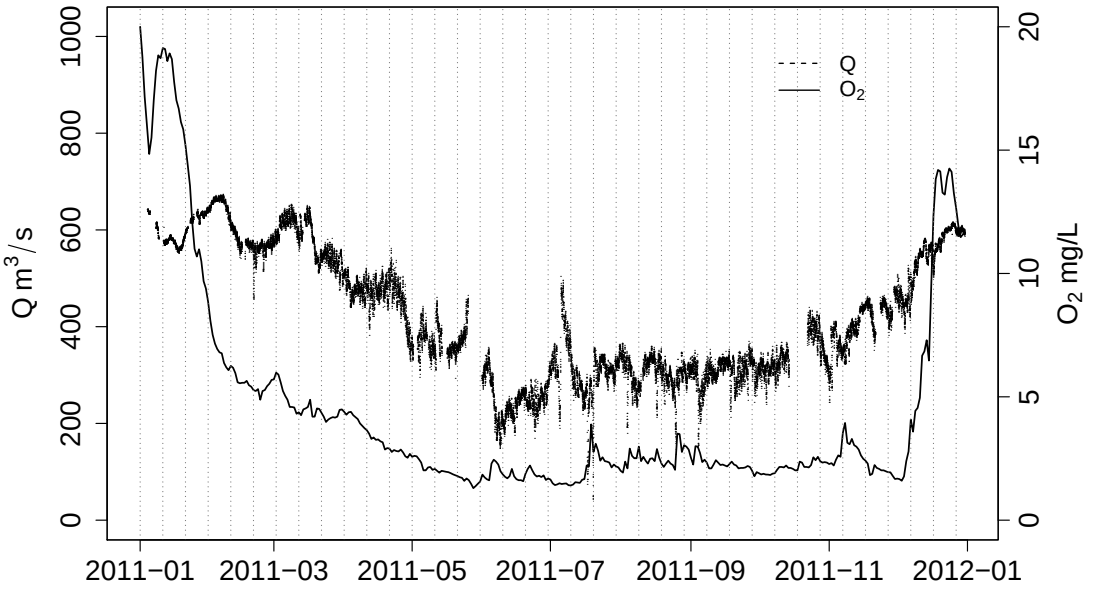

Fig. 3: Daily discharge (Q m^3/s) at the gauging station and 15-min measured DO concentrations (O_2 mg/L) at Andresy station in 2011.

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

2.3.2. Input data and available oxygen data

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

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

Eight monitoring stations managed by SIAAP provide 15-min time step DO concentrations in 2011 (Fig. 2 and Tab. 1). The differences in the number of measurements are related to sensor dysfunctioning. Four monitoring stations from RCS monitoring network provide low frequency DO measurements (Fig. 2 and Tab. 1). These high frequency measurements are assimilated by ProSe-PA (Wang et al., 2019) to improve the simulation of the biogeochemical functioning of the Seine River system and to characterize the most influential model parameters that were identified 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
Bonnieres	7	bimonthly	SEDIF

2.4. Simulations and numerical setting

2.4.1. Simulation types and statistical criteria

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

To evaluate the potential of ProSe-PA for the simulation of DO concentrations at the four low frequency monitoring stations (RCS), all the high frequency observed DO data are assimilated.

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

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

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

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

where

N_{obs}: Number of observations

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

r: Correlation coefficient

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

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

2.4.2. Parameters considered for data assimilation and numerical settings

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

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

The observation error is assumed to follow a Gaussian distribution with a mean zero and a standard deviation of 10% of the measured DO concentrations ($0.1y_t^*$), 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]	[$m^2.s.\mu E^{-1}.h^{-1}$]
$\eta_{chla,pp}$	Light extinction coefficient by algal self-shading	[0.006, 0.054]	[$L.\mu gchla^{-1}.m^{-1}$]
$Chla/C_{pp}$	Ratio of chlorophyll *a* to carbon	[7.69,50]	[$\mu gchla.mgC^{-1}$]
$P_{max,pp}$	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]

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

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

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

3. Results: DO simulation and model parameter characterization

3.1. Improved simulation of DO concentrations with Particle Filter

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

The particle filter improves the simulation of DO concentrations significantly (Tab. 3). Indeed, RMSEs obtained with ProSe-PA range from 0.71 mgO_2/L to 1.32 mgO_2/L while those obtained with forward model ProSe vary from 0.75 mgO_2/L to 1.61 mgO_2/L for the whole year of 2011. RMSEs for ProSe-PA decrease at all monitoring stations, especially at Bougival (47%, 0.74 mgO_2/L vs 1.40 mgO_2/L, Tab. 3), Sartrouville (33%, 1.14 mgO_2/L vs 1.69 mgO_2/L,

Tab. 3) and Meulan (29%, 0.70 mgO_2/L vs 0.99 mgO_2/L, Tab. 3). Moreover, KGEs obtained with ProSe-PA are closer to 1, except for the Sartrouville station (0.46 vs 0.53). At all stations, the simulated DO concentrations by ProSe-PA are more correlated with the measured DO concentrations (see r in Tab. 3). In terms of model bias for the full year 2011 (β), ProSe-PA and ProSe produce similar results. Both ProSe-PA and ProSe tend to overestimate DO concentrations ($\beta > 1$, Tab. 3). The maximum bias is estimated at the Sartrouville station with an overestimation of 12% for ProSe-PA and 19% for ProSe respectively (Tab. 3). The coefficient of variation (γ) explains the relative dispersion of the DO concentrations around the average of DO concentration. The results show that ProSe-PA evaluates more correctly the dispersion of observed DO concentrations where the coefficients of variation (γ) are much closer to 1 (Tab. 3), except for the Sartrouville (0.51 vs 0.78) and Meulan (0.60 vs 0.77) stations. However, both ProSe-PA and ProSe underestimate the amplitude of DO concentrations ($\gamma < 1$). Despite a lower KGE at Sartrouville station for the simulations of ProSe-PA, the results show the capacity of particle filter for predicting (cross-validation) the DO concentrations in river systems (higher RMSE at all stations).

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

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

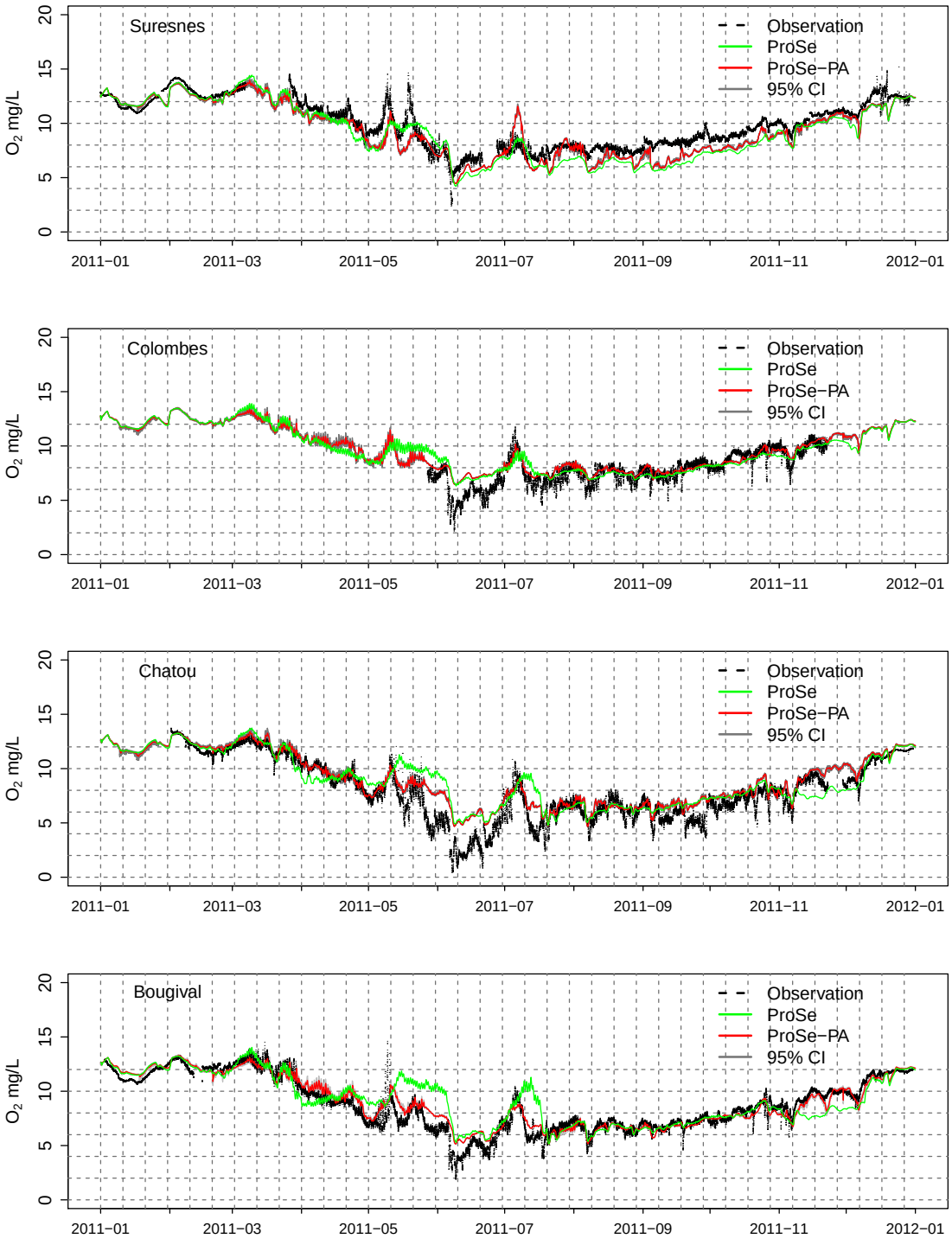

Fig. 4: Simulated DO concentrations at Suresnes, Colombes, Chatou and Bougival stations (cross-validation). The red lines denote ensemble weighted averages simulated by ProSe-PA. The green lines show the simulated DO concentrations by ProSe with static parameters. The observation data are represented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

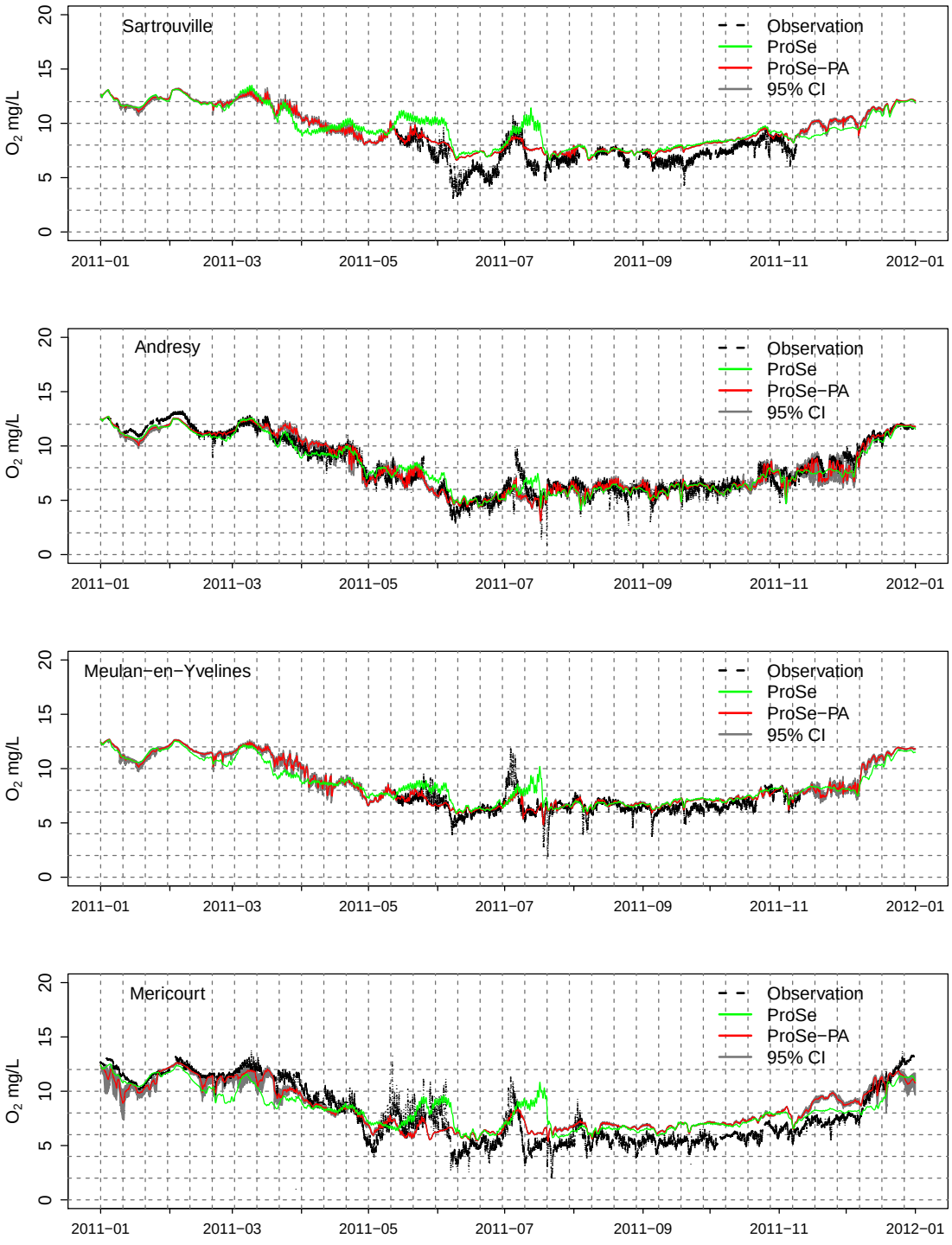

Fig. 5: Simulated DO concentrations at Sartrouville, Andresy, Meulan and Mericourt stations (cross-validation). The red lines denote ensemble weighted averages simulated by ProSe-PA. The green lines show the simulated DO concentrations by ProSe with static parameters. The observation data are represented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

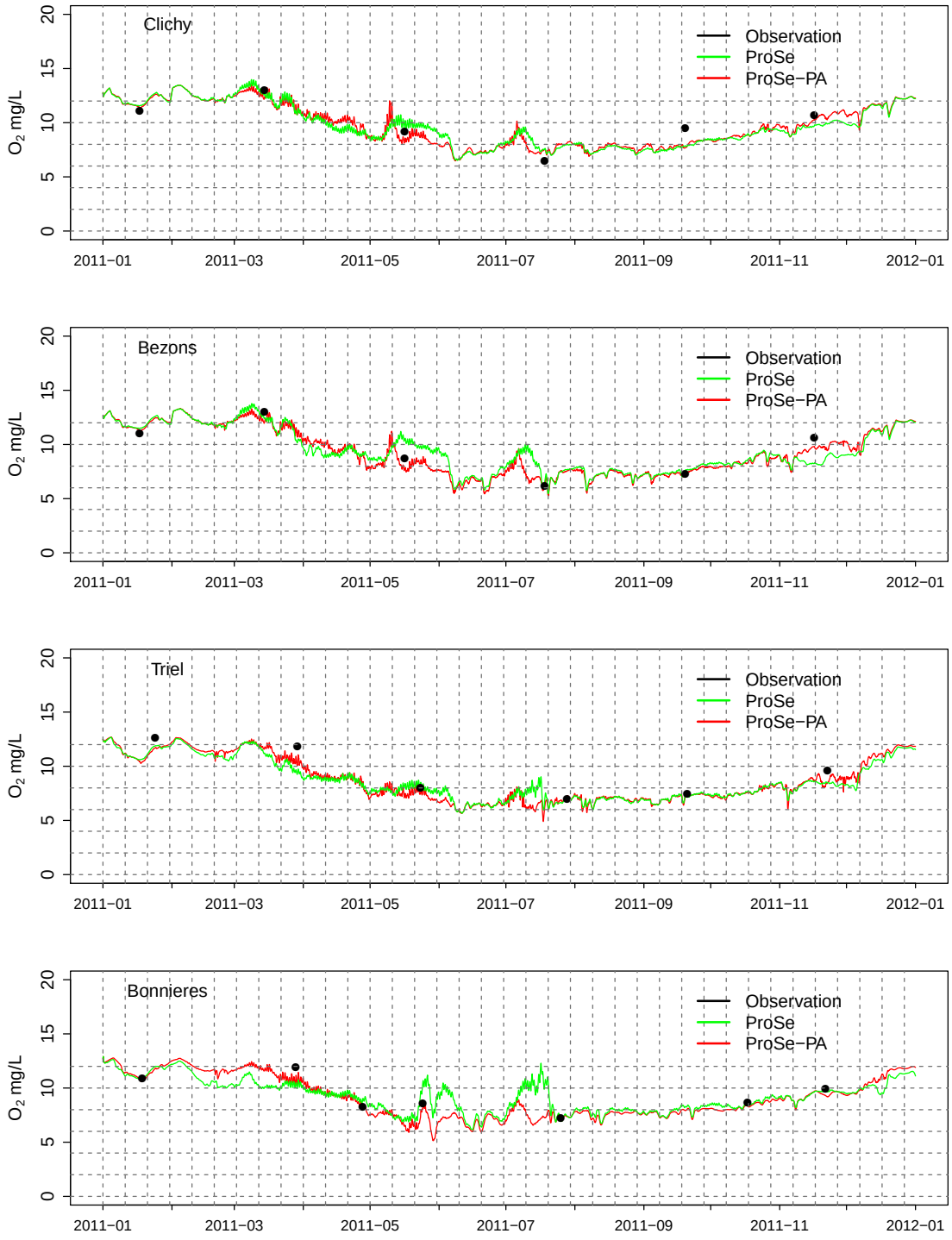

Fig. 6: Simulated DO concentrations at Clichy, Bezons, Triel and Bonnieres stations (cross-validation). The red lines denote ensemble weighted averages simulated by ProSe-PA. The green lines show the simulated DO concentrations by ProSe with static parameters.

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		P
	RMSE	KGE	r	β	γ	RMSE	KGE	r	β	γ	RMSE	KGE	RMSE	KGE	RMSE	KGE	
Suresnes	1.09	0.81	0.93	0.93	1.17	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	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	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	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	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	0.75	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	0.99	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	1.52	0.60	0.87	1.04	0.62	1.28	0.78	1.62	0.47	1.52	0.65	

3.2. Dynamic parameters' characterization

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

3.2.1. Phytoplanktonic parameters

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

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

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

Relative high values of the respiration of maintenance ($R_{m,pp}$), ranging from 0.012 h^{-1} to 0.018 h^{-1} (Fig. 8), 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 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 vary within the parameter range (Fig. 7). This behaviour can be observed for the light extinction coefficient by algal self-shading ($\eta_{chla,pp}$) and the ratio of chlorophyll a to carbon (Chla/C_{pp}) also, while no clear switch of the 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 this period is estimated. After 2011-04-15, the maximum photosynthesis rate of phytoplankton ($P_{max,pp}$) decreases with an average of 0.27 h^{-1}. The posterior distributions of the photosynthetic capacity (α_{pp}) from 2011-03-16 to

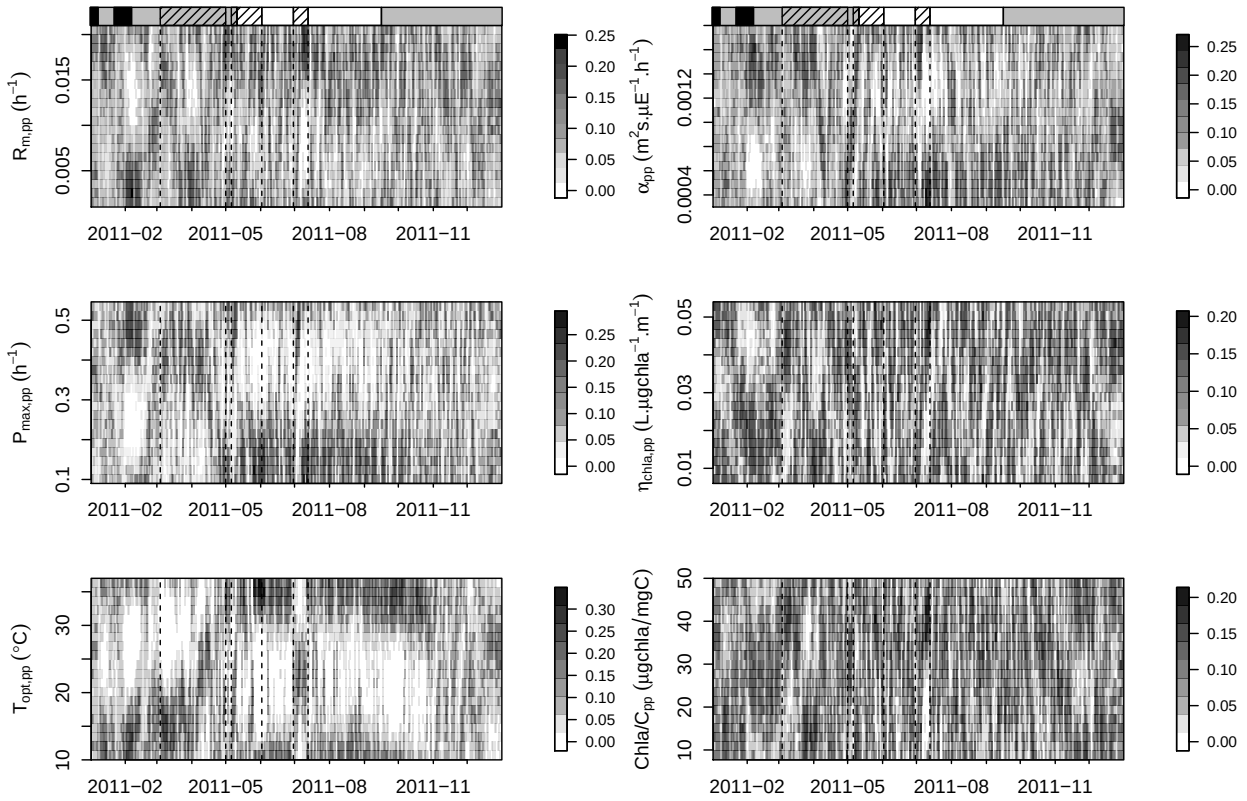

Fig. 7: Normalised importance weights of phytoplanktonic parameters. See Tab. 2 for parameter definition. Line shaded polygons correspond to algae bloom periods and water temperature increases from black to white. For the color gradient: black T < 6 °C, gray 6 °C < T < 20 °C, white T > 20 °C.

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

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

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

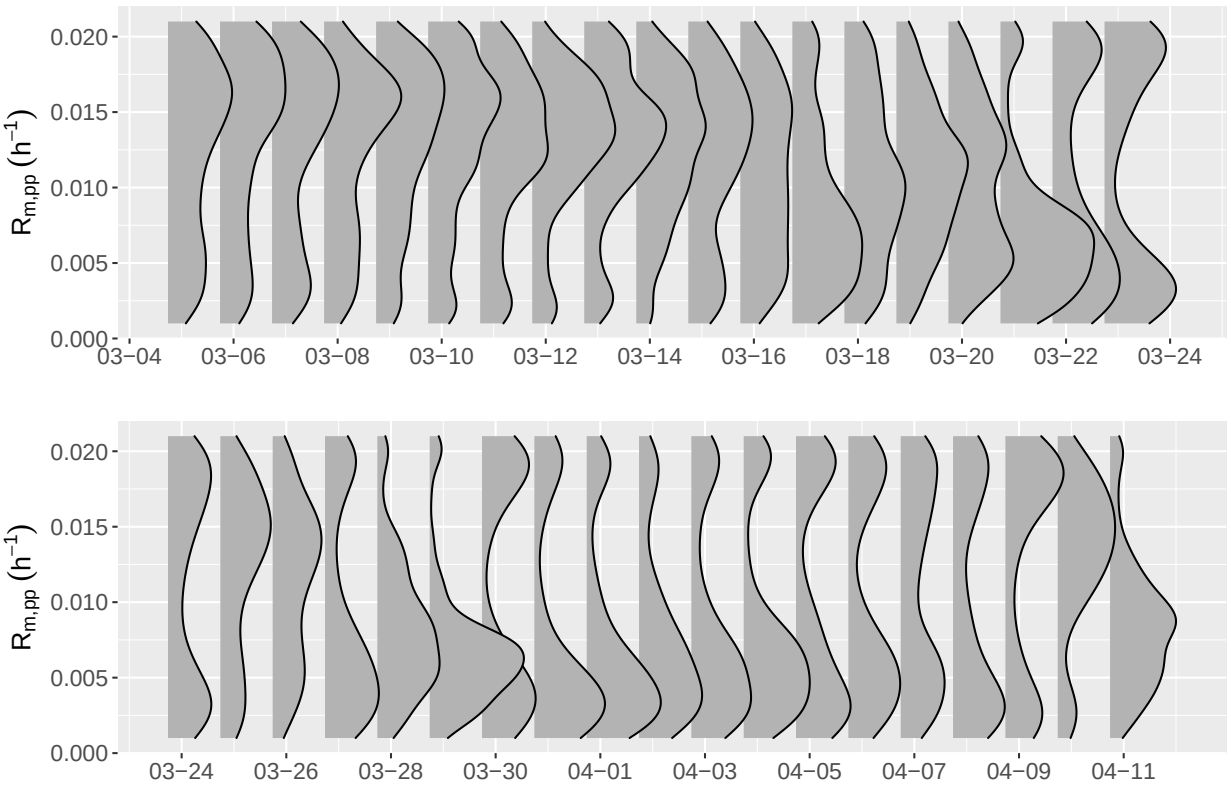

Fig. 8: Posterior pdfs of the respiration of maintenance during the bloom of March.

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

During the bloom of July, the optimal temperature for the growth of phytoplankton ($T_{opt,pp}$) is well determined 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 posterior densities are identified (0.015 h^{-1} and 0.034 L.μgchla^{-1}.m^{-1}, Fig. 7). A switch of the phytoplanktonic 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 represent the mortality of phytoplankton linked to its senescence by a decrease in the maximum photosynthesis rate ($P_{max,pp}$) and the photosynthetic capacity (α_{pp}).

3.2.2. Bacterial parameters

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

ProSe-PA identifies bacteria parameters before the bloom of March (Fig. 9 and Fig. 10). The averages of the 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 31.0 °C for the optimal temperature for growth of bacteria ($T_{opt,hb}$) (Fig. 10, 2011-02-06 to 2011-02-20).

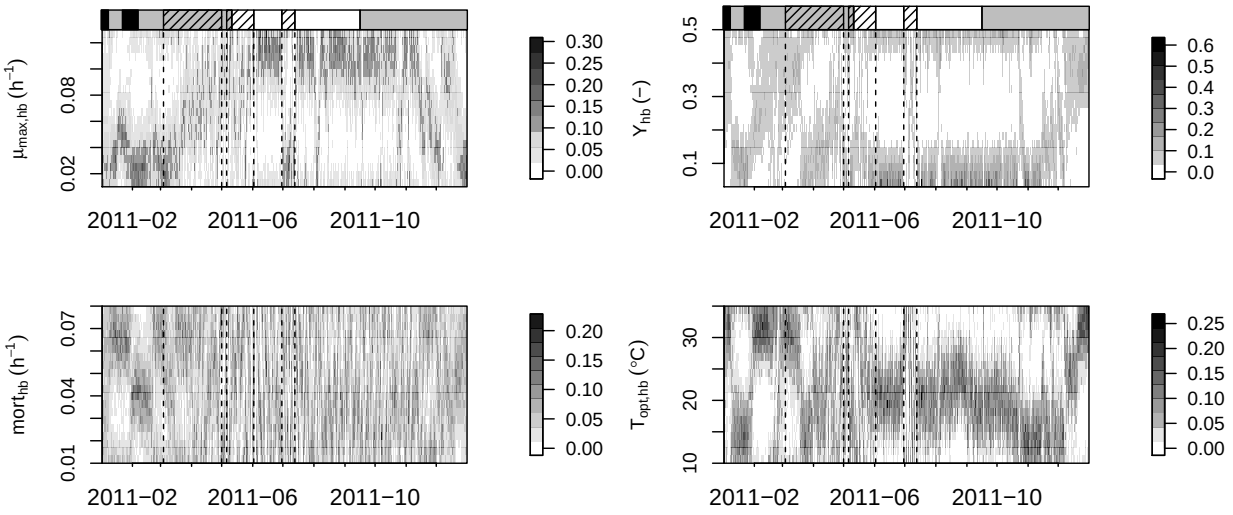

Fig. 9: Normalised importance weights of bacterial parameters. See Tab. 2 for parameter definition. Line shaded polygons correspond to algae bloom periods and water temperature increases from black to white. For the color gradient: black T < 6°C, gray 6 °C < T < 20°C, white T > 20 °C.

Between the bloom of May and the bloom of July, the bacterial activities are characterized by a high maximum 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

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

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

4. Discussion

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

The presented RMSEs of 0.71 to 1.32 mgO_2/L are relatively high (cross-validation) when comparing the measured and simulated DO concentrations (Tab. 3). As demonstrated in section 3.1, both ProSe-PA and ProSe overestimate 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 high frequency monitoring stations (Colombes, Chatou, Sartrouville and in a lesser extent Bougival and Mericourt). Slighter discrepancies of the same type are also observed in September and in October during dry periods, especially at Sartrouville and Mericourt stations. Those mismatches may have two origins, more or less independent from one another: uncertainties of BDOM concentrations in input data and particle filter degeneracy.

4.1.1. Particule Filter degeneracy

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

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

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

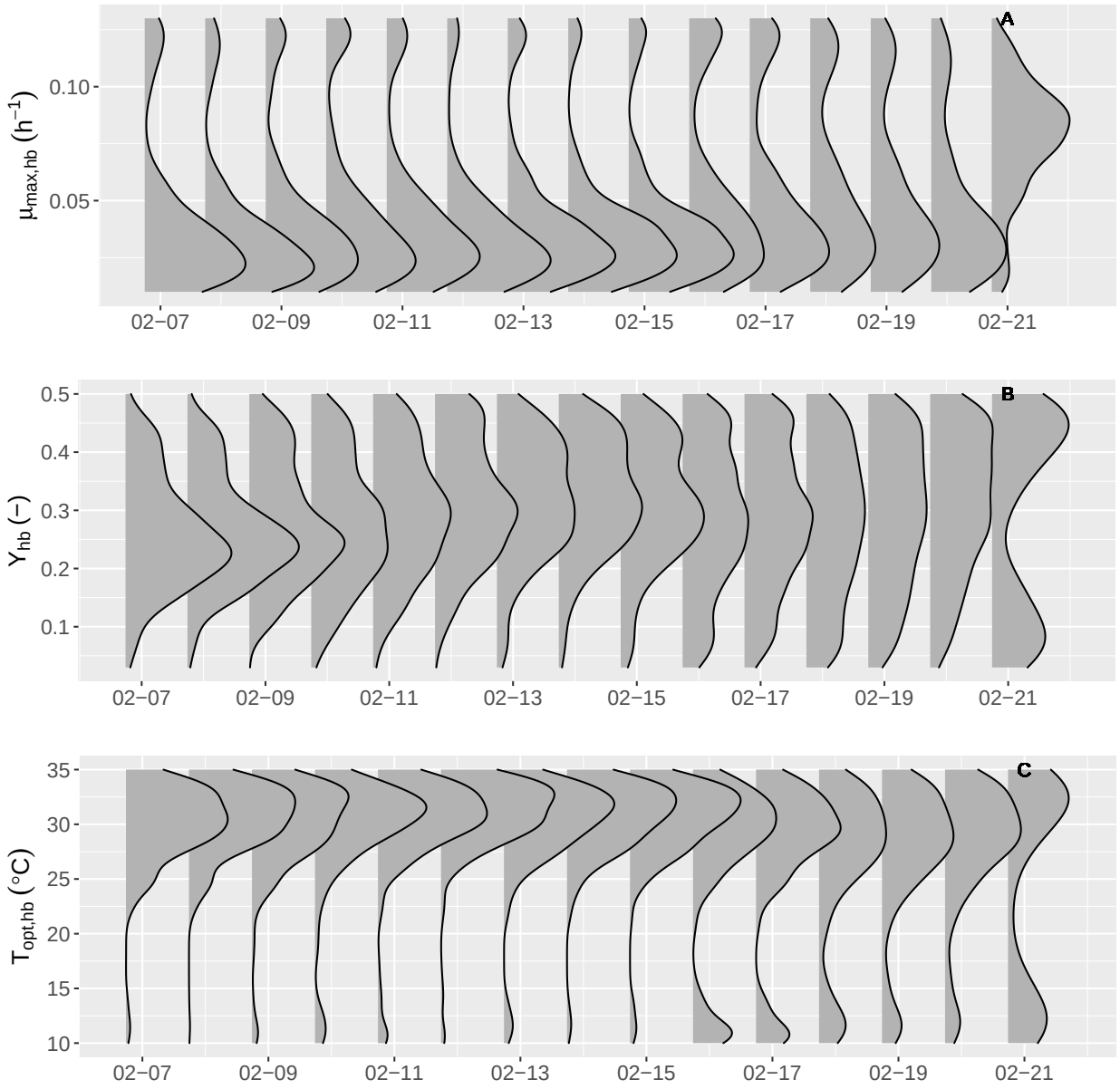

Fig. 10: Posterior pdfs of the bacterial parameters before the bloom of March.

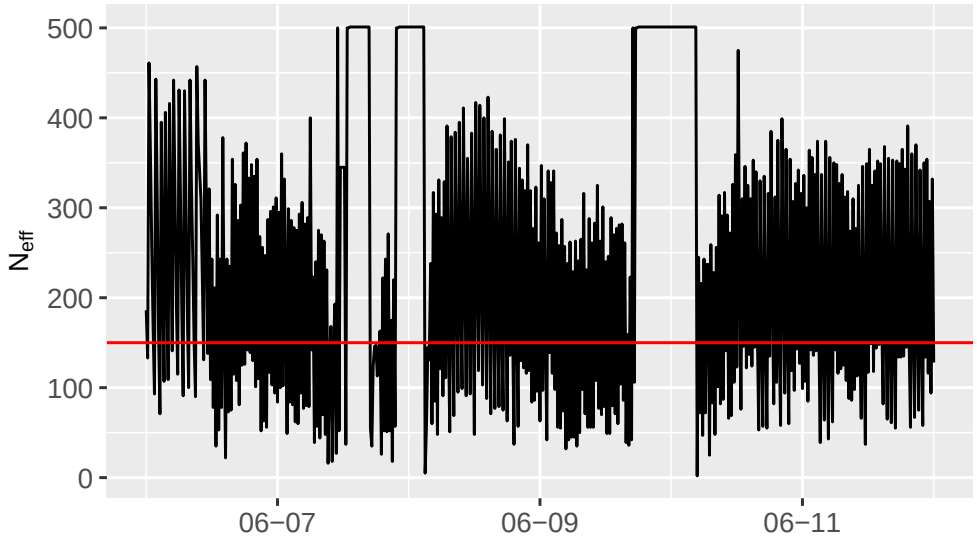

Fig. 11: Evolution of the effective sample size from 2011-06-05 to 2011-06-12. The red line corresponds to the minimum effective sample size of 150 below which a resampling procedure is carried out.

4.1.2. Description of BDOM in input data

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

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

It seems that the farther from an organic matter source the monitoring station is located, the larger the discrepancies between simulation and observation. This is true for the Suresnes to Sartrouville monitoring stations. ProSe-PA is able to simulate DO concentrations at Andresy station properly (a RMSE of 0.53 mgO_2/L when using eight stations for data assimilation). This station is located downstream of two important organic matter sources, the Oise River and the SAV WWTP (Fig. 2), which enrich the Seine River in BDOM (Fig. 12). Further downstream, when all the biodegradable organic matter has been consumed (see the mismatch at the Mericourt station), the particle filter degenerates again.

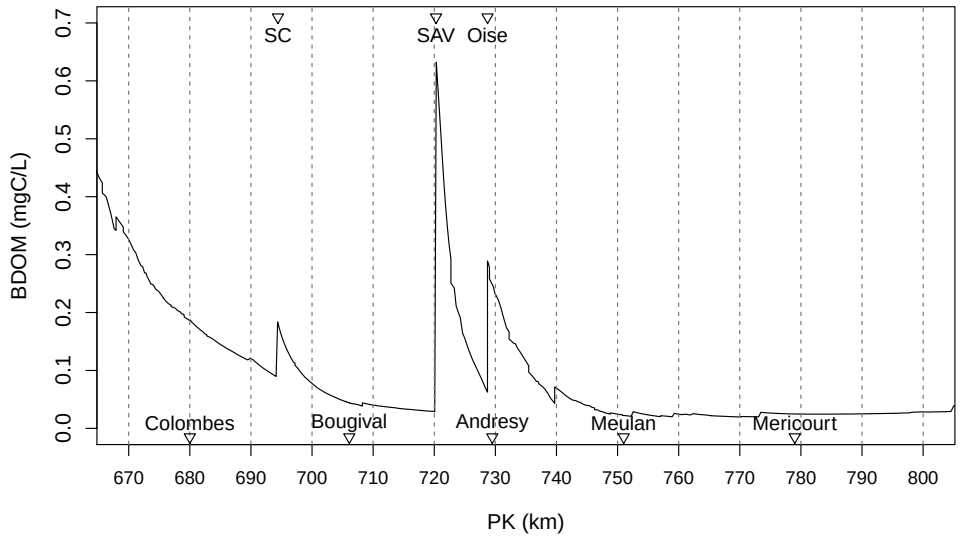

Fig. 12: Longitudinal profile of mean BDOM concentration from 2011-09-07 to 2011-09-27. Only left river reaches are presented here. PK: Kilometer Point, PK = 862 km at the entrance of the Seine river, Poses. SC: Seine Centre WWTP. SAV: Seine Aval WWTP. Oise: Oise river.

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

ProSe-PA simulates almost the same oxygen level at Meulan and Mericourt stations in September and October. However, the depletion of measured DO is observed at Mericourt station (Fig. 5). It means that supplementary BDOM is required to produce this depletion at Mericourt station. The same phenomenon can also be observed from Colombes station to Bougival station. The simulated DO concentrations varying from 7 mgO_2/L to 8 mgO_2/L are similar at Colombes, Chatou and Bougival stations (from 2011-09-07 to 2011-09-27). The DO concentrations are well simulated by ProSe-PA at Colombes and Bougival stations at this period. But the depletion of DO concentrations ($<$ 6 mgO_2/L, Fig. 4) during this period are not reproduced at Chatou station located at a right reach of the Seine River. This depletion of oxygen results from the lateral inflows containing BDOM (overflow of CSOs/WWTPs) before Chatou station. This indicates that some unknown overflow sources are not considered on the right reach of the Seine river before Chatou station.

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

of DOM have been reported in numerous studies (Meyer et al., 1987; Servais et al., 1987, 1995, 1999; Søndergaard and Middelboe, 1995; Wang et al., 2010; He et al., 2010; Kang and Mitchell, 2013). This portion depends on the hydrological and trophic contexts of a river, therefore varying temporally. Generally, the range of BDOM portion of 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, which seems a rather good estimate most of the time, but need to be refined when it leads to improper organic matter depletion.

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

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

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

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

5. Conclusions

A particle filter is applied for the first time to assimilate high frequency dissolved oxygen data in river system 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} [$m^2 s.\mu E^{-1}.h^{-1}$]	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/Chl$a$ [$mgC/\mu gchla$]	0.026	0.055	0.035	0.035	0.035	0.035

-: Oscillating modes

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

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

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

Declarations of interest

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

Acknowledgements

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

References

Abbaszadeh, P., Moradkhani, H., Yan, H., 2018. Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo. Advances in Water Resources 111, 192–204.

Andreadis, K., Clark, E., Lettermaier, D., Alsdorf, D., 2007. Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. Geophysical Research Letters 34, L10403.

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society of London 53, 370–418.

Beck, M., 1978. Random signal analysis in an environmental sciences problem. Applied Mathematical Modelling 2 (1), 23–29.

Beck, M., Halfon, E., 1991. Uncertainty, identifiability and the propagation of prediction errors: A case study of lake ontario. Journal of Forecasting 10 (1-2), 135–161.

Beck, M., Young, P., 10 1976. Systematic identification of do-bod model structure. Jnl. of Env. Eng. Division , American Society of Civil Engineers 102, 909–927.

Beven, K., 1989. Changing ideas in hydrology. The case of physically-based model. Journal of Hydrology 105, 157–172.

Billen, G., Garnier, J., Hanset, P., 1994. Modelling phytoplankton development in whole drainage networks: the RIVERSTRAHLER Model applied to the Seine river system. Hydrobiologia 289, 119–137.

Bowles, D., Grenney, W., 1978. Steady state river quality modeling by sequential extended kalman filters. Water Resources Research 14 (1), 84–96.

Carrassi, A., Bocquet, M., Bertino, L., Evensen, G., 2018. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. Wiley Interdisciplinary Reviews: Climate Change 9 (5), e535.

Chen, C., Huang, J., Chen, Q., Zhang, J., Li, Z., Lin, Y., 2019. Assimilating multi-source data into a three-dimensional hydro-ecological dynamics model using ensemble kalman filter. Environmental Modelling & Software 117, 188 – 199.

Cho, K. H., Pachepsky, Y., Ligaray, M., Kwon, Y., Kim, K. H., 2020. Data assimilation in surface water quality modeling: A review. Water Research 186, 116307.

Cosby, B., Hornberger, G., 1984. Identification of photosynthesis-light models for aquatic systems i. theory and simulations. Ecological Modelling 23 (1-2), 1–24.

Cosby, B., Hornberger, G., Kelly, M., 1984. Identification of photosynthesis-light models for aquatic systems ii. application to a macrophyte dominated stream. Ecological Modelling 23 (1), 25 – 51, modelling Primary Production.

Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., Fisher, M., Pailleux, J., 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. Quarterly Journal of the Royal Meteorological Society 124 (550), 1783–1807.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 PIREN-Seine, PIREN-Seine.

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 carbon and its major compound classes in the Pearl River estuary, China. Marine Chemistry 119 (1-4), 52–64.

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

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

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

Kang, P.-G., Mitchell, M. J., 2013. 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.

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

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

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

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

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

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

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

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

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

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

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

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

Moradkhani, H., DeChant, C. M., Sorooshian, S., 2012. Evolution of ensemble data assimilation for uncertainty quantification using the particle 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.

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

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

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

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

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

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

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

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

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

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

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

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

Thieu, V., Billen, G., Garnier, J., 2009. 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, 1740–1754.

Vanrolleghem, P., Borchardt, D., Henze, M., Rauch, W., Reichert, P., Shanahan, P., Somlyódy, L., 2001. River Water Quality Model no.1 (RWQM1): III biochemical submodel selection. Water Science and Technology 43 (5), 31–40.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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:
