Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Review

Data assimilation in surface water quality modeling: A review

Kyung Hwa Cho[a], Yakov Pachepsky[b,*], Mayzonee Ligaray[c], Yongsung Kwon[d], Kyung Hyun Kim[e]

[a] School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, 689-798, Republic of Korea
[b] Environmental Microbial and Food Safety Laboratory, USDA-ARS, Beltsville, MD 20705 USA
[c] Institute of Environmental Science and Meteorology, College of Science, University of the Philippines Diliman, Quezon City 1101, Philippines
[d] Division of Ecological Assessment Research, National Institute of Ecology, Seocheon 33657, Republic of Korea
[e] Watershed and Total Load Management Research Division, National Institute of Environmental Research, Ministry of Environment, Hwangyong-ro 42, Seogu, Incheon, Republic of Korea

ARTICLE INFO

Article history:
Received 17 October 2019
Revised 9 August 2020
Accepted 15 August 2020
Available online 16 August 2020

Keywords:
Water quality model
Data assimilation
Variational data assimilation
Extended Kalman filter
Ensemble Kalman filter

ABSTRACT

Data assimilation (DA) techniques are powerful means of dynamic natural system modeling that allow for the use of data as soon as it appears to improve model predictions and reduce prediction uncertainty by correcting state variables, model parameters, and boundary and initial conditions. The objectives of this review are to explore existing approaches and advances in DA applications for surface water quality modeling and to identify future research prospects. We first reviewed the DA methods used in water quality modeling as reported in literature. We then addressed observations and suggestions regarding various factors of DA performance, such as the mismatch between both lateral and vertical spatial detail of measurements and modeling, subgrid heterogeneity, presence of temporally stable spatial patterns in water quality parameters and related biases, evaluation of uncertainty in data and modeling results, mismatch between scales and schedules of data from multiple sources, selection of parameters to be updated along with state variables, update frequency and forecast skill. The review concludes with the outlook section that outlines current challenges and opportunities related to growing role of novel data sources, scale mismatch between model discretization and observation, structural uncertainty of models and conversion of measured to simulated vales, experimentation with DA prior to applications, using DA performance or model selection, the role of sensitivity analysis, and the expanding use of DA in water quality management.

1. Introduction

A water quality model is the mathematical representation of pollutant fate and transport within a water body that may be coupled with a mathematical representation of the movement of pollutants from land movement of pollutants from land -based sources to a water body (Kebede 2009). The water quality models include description of physical, chemical and biological mechanisms affecting fate and transport of pollutants. Surface water quality models are critically important tools for managing our nations' surface waters as they help local communities and environmental managers better understand how surface waters change in response to pollution and how to protect them. Water quality specialists use models for many purposes such as assessing water quality conditions and causes of impairment, predicting how surface waters will respond to changes in their watersheds and the environment (e.g., future growth, climate change), and forecasting quantitative benefits of new surface water protection policies (EPA 2018). Streeter and Phelps (1925) firstly introduced their model as a solution of the first-order differential equation to simulate DO decline by the decomposition of organic matters and DO increased by reaeration. Water quality modeling has since advanced from considering a simple first-order decomposition of organic matter to a complex multiple biochemically-mediated processes, from point-source models to both point and nonpoint sources as well as from 1-dimensional steady-state models to 3-dimensional dynamic models (Wang et al., 2013). In this regard, the number of model parameters considerably increased, thereby increasing the prediction uncertainty. It was eventually realized that the optimal estimates of the evolving water quality attributes should be obtained by jointly considering the outputs of the model with the data from ongoing observations. This can be achieved by using mathematical techniques of data assimilation (DA) and their computational implementations (Asch et al., 2016; Fletcher 2017).

[*] Corresponding author at: 10300 Baltimore Ave., Bldg. 173
E-mail address: Yakov.pachepsky@usda.gov (Y. Pachepsky).

https://doi.org/10.1016/j.watres.2020.116307
0043-1354/Published by Elsevier Ltd.

The data assimilation (DA) is the methodology whereby observational data are combined with output from a numerical model to produce an optimal estimate of the evolving state and/or parameters of the system (O'Neill 2003). As new observations of a water quality become available the DA may include updates of state variables (e.g., dissolved oxygen concentration or phytoplankton biomass), model parameters (e.g., decomposition or reaeration rate), and boundary conditions (e.g., tributary input to the modeled water body). The DA allows us to update modeling results and/or achieve model improvements each time new observations become available. The update is based on the consideration of the uncertainties in data and in model predictions. The more certain measurement data are and the less certain modeling results are, the closer the modeling results become to measurements after updates. The status of the pollution is simulated as the spatiotemporal variation of the pollutant concentrations and water quality attributes that affect the fate of pollutants. These concentrations and attributes are the state variables of the modeled aquatic systems. Coefficients in the surface water quality models describe the site-specific conditions; they are known as model parameters.

The DA has drawn much attention in surface hydrology wherein review works and monographs analyzed the DA potential to improve the hydrologic model performance (Montzka et al., 2012; Moradkhani and Sorooshian 2009; Park and Xu 2013). Its applications to water quality modeling have slower developments than its applications in hydrology (Robinson and Lermusiaux 2000; Sun et al., 2016), partly because it was quite challenging to carry out real-time and high-frequency monitoring of water quality at multiple locations. Beginning from 2005, the total annual number of peer-reviewed publications on water quality modeling steadily increased by about a hundred (data from the Scopus database). This growth occurred due to multiple reasons which includes the development of remote sensing algorithms, in situ sensors, advances in the modeling of water bodies, progress in understanding processes affecting water quality, regulatory actions, and development of pre- and post-processing capabilities. Around the same timeframe, the DA was applied for the first time to the three-dimensional water quality model for a large natural water body (Vodacek et al., 2008). Since then, various DA methods were utilized in modeling various types of natural water constituents, with the most recent focus on harmful algae blooms (Loos et al., 2020). Various DA applications in the surface water models may include updates of state variables (e.g., dissolved oxygen concentration or phytoplankton biomass), model parameters (e.g., dissolved organic matter decomposition rate), and boundary conditions (e.g., the tributary input to the main stream).

While the water quality modeling in general was extensively reviewed (Cho et al., 2016; Ji 2017; Wellen et al., 2015), reviews of DA in water quality modeling have not been published. There is a need for a systematic state-of-the-art presentation of this fast-developing field. The currently, published research allows one to evaluate and compare existing approaches and advances in DA applications to the surface water quality modeling as well as identify future research prospects. In this work, we first review the application of each DA methods, then discuss the factors of DA performance, and finally highlight research gaps and opportunities in DA application in surface water quality modelling.

2. Data assimilation applications in surface water quality models

Table 1 summarizes the reported DA applications for the water quality modeling. DA applications started with batch systems simulated with ordinary differential equations, where flow was not considered explicitly. Dissolved oxygen, biological oxygen demand, and chlorophyll-a concentrations were the primary (i.e., response) variables of interest. As the interest to managing water quality in large water bodies increased, DA applications have been developed for hydrodynamic models (e.g., EFDC – the environmental fluids dynamics code, Delft3D-WAQ, ALGE, and 3DHED) and watershed water quality models (e.g., HSPF – hydrologic simulation system FORTRAN (EPA, 2018b) and Coupled P model). EFDC and HSPF are most often found in literature. These two models typify the model classes that have been of primary interest to water quality DA. The EFDC can simulate water flow and water quality constituent transport in geometrically and dynamically complex water bodies, such as rivers, stratified estuaries, lakes, and coastal seas. The code is capable of simulating salinity, temperature, sediment, contaminant, and eutrophication variables. On the other hand, the HSPF is a continuous simulation, lumped parameter, watershed-scale model. Any time step (minimum of 1 min) can be used, although the typical time step used is 1 h. The subbasin is subdivided into hydrological response units, defined as relatively homogeneous areas based on land use and hydrologic properties of both pervious and impervious land, which HSPF can simulate separately. Most of the recently used models are scalable, and they have been applied to simulate streamflow and water quality at different spatial scales – from ponds to large complex lakes. Most modeling work using DA concentrated on biotic components of aquatic ecosystems In the last decade, DA applications in water quality modeling focused on simulations of eutrophication and harmful algal blooms.

2.1. Data assimilation methods in surface water quality research and management

To-date, three major methods have been utilized to assimilate water quality data into the water quality models (Table 1): (a) the variational data assimilation, (b) the extended Kalman Filter (EKF), and (c) Ensemble Kalman Filter (EnKF). The earliest application of data assimilation was published by Beck and Young (1976). They were the first to introduce the EKF method to explain the relationship between DO and BOD. No applications of EKF have been published after 2009. Historically, the variational data assimilation was applied in environmental science before EKF and EnKF, but its surface water quality applications only appeared in 2013. The EnKF has become the most popular DA method suitable in incorporating multiple source observations. Currently, the particle filter (PF) method was applied to assimilate DO concentration by Wang et al., (2019).

2.2. Variational data assimilation

The variational data assimilation approach was first introduced to minimize the discrepancy between model outputs and observations in meteorology and oceanography. The methods have been actively used to update the state variables from the late 1980s (Lawless 2013). The objective of this approach is to find the values of update variables (x_0) which minimizes the weighted least squares distance to the background (non-updated modeling result) update variables x_b plus the weighted least squares distance to the measurement in the assimilation window, as shown in Fig 1a. The cost function to minimize is:

$$\mathcal{J}\ (x_0) = \frac{1}{2}(x_0 - x_b)^T B^{-1}(x_0 - x_b) + (Hx_0 - y)^T\ R_i^{-1}(Hx_0 - y) \quad (1)$$

Here, x_0 is the vector of update variables which is sought as the result of data assimilation, $x_b$is the background values of update variables which is usually the modeling result for the update time, B is the covariance matrix of the background error, R is the covariance matrix of the observation error, H is the observation operator which maps the vector of update variables into the observation space, and y is the observation. This data assimilation scheme is

Table 1
DA applications in water quality modeling.

Authors	Year	Model, spatial dimension[a]	DA method[b]	Total number of model parameters	Selected parameters for DA[c]	State variables[d]	Observation (satellite)[e]	Aquatic system
Beck and Young	1976	CSTR,0D	EKF	5	2 or 4	DO, BOD	I and L	River
Whitehead and Hornberger	1984	NN, 0D	EKF	9	3 (SA)	*Chl-a*	L	River
Cosby and Hornberger	1984	NN, 0D	EKF	5	2	DO	Synthetic data	-
Cosby et al.	1984	NN, 0D	EKF	6	4	DO	I	River
Ennola et al.	1998	NN, 0D	EKF	4	4	Rotifera biomass	L	Sewage treatment pond
Pastres et al.	2003	NN, 0D	EKF	13	3	DO	I	Lagoon
Voutilainen et al.	2005	NN, 0D	EKF	-	-	TSS, *Chl-a*, CDOM	R (synthetic)	Lake
El Serafy et al.	2007	Delft3D-WAQ, 3D	EnKF	No information	IC	SPM	R (MERIS)	Coastal water
Vodacek et al.	2008	ALGE, 3D	EnKF	-	IC	TSS	R (MODIS)	Lake
Mao et al.	2009	NN, 0D	EKF	30	2	*Chl-a*, DO	I and L	Coastal waters
Margvelashvili et al.	2010	SHOCK-EMS, 3D	EnKF	Numerous	3	TSS	Not clear	Coastal waters
Babbar-Sebens et al.	2013	EFDC, 3D	3DVAR	6	IC	Water Temperature	R (Landsat-5TM)	Reservoir
Huang et al.	2013	NN, 2D	EnKF	-	1	*Chl-a*	L	Lake
Kim et al. [a]	2014	EFDC	EnKF	Numerous	BC	Streamflow, *Chl-a*, Phosphate-ion	L	River
Kim et al. [b]	2014	HSPF, 1D	MLEF	Numerous	BC	Streamflow, BOD, DO, *Chl-a*, NO3, phosphate-ion, water temperature	L	River basin
Shao et al.	2016	NN, 1D	EnKF	No information	BC	Sucrose	L	River
Javaheri et al.	2016	NN, 3D	EnKF	6	IC	Water Temperature	R (Landsat-5TM)	Reservoir
Riazi et al.	2016	HSPF, 1D	MLEF	Numerous	IC	*Chl-a*	L	River Basin
Huang and Gao	2017	Coupled P, 0D	EnKF	-	2	Phosphorus	L	Lake Basin
Page et al.	2018	PROTECH, 0D	EnKF	-	12	*Chl-a*	L	Lake
Javaheri et al.	2019	EFDC, 3D	EnKF	6	IC	Water Temperature	R (Landsat 7)	River
Chen et al.	2019	3DHED, 3D	EnKF	120	IC	Cyanobacteria biomass	R (MERIS)	Lake
Wang et al	2019	PROSE	PF	-	12	DO		

[a] NN – no name, 0D – batch system without spatial dimension, EFDC – Environmental Fluid Dynamics Code, HSPF – Hydrological Simulation Proram – Fortran, 3DHED - hydro-ecological dynamics model, PROTECH - Phytoplankton RespOnses To Environmental Change;

[b] EKF – extended Kalman filter, EnKF- ensemble Kalman filter, 3DVAR – three dimensional variational data assimilation, MLEF - maximum likelihood ensemble filter,

[c] SA – based on the Sensitivity Analysis ranking, IC – initial conditions, BC – boundary conditions,

[d] DO – dissolved oxygen, BOD biological oxygen demand, *Chl-a* – chlorophyl*l A*, TSS – total suspended solids, CDOM – colored dissolved organic matter, SPM – suspended particulate matter,

[e] I – in situ, L – laboratory, R remote sensing, MODIS - moderate resolution imaging spectroradiometer, MERIS - the medium resolution imaging spectrometer.

defined as the 3-dimensional variational data assimilation (3DVAR) as introduced by Sasaki (1958). It is illustrated in Fig 1a.

The 3DVAR data assimilation was only recently applied in water quality modeling, and was used for updating boundary or initial conditions. Shao et al. (2016) studied DA for a one-dimensional convective-dispersive transport model of the tracer in a river reach with four sampling stations. The authors demonstrated that the performance of a water quality model on the fate and transport of contaminants can be improved by introducing the 3DVAR data assimilation which resulted in a decrease of almost three-times the root-mean-squared error (RMSE). The DA in three-dimensional spatial setup was first applied by Babbar-Sebens et al. (2013) who used the 3DVAR approach for incorporating spatially continuous remote sensing temperature observations from the multi-spectral Landsat-5 TM band and spatially discrete *in situ* observations to change initial conditions of the EFDC model applied at a eutrophic Eagle Creek Reservoir in Central Indiana. The vector of initial conditions had 300 elements corresponding to different locations at the reservoir water surface, and the genetic algorithm was used for minimization of the cost function.

Some features of the 3DVAR method require caution in its applications to water quality problems. It remains challenging to apply 3DVAR in updating both the state variable and the parameters of the water quality models which consist of many different biochemical parameters. It has been shown that the 3DVAR data assimilation can be inherently unstable if the observation operator is unbounded (Marx and Potthast 2012). High computational cost may impose limitation of the variational assimilation applicability.

If the observation operator is strongly nonlinear, the 4D variational algorithms (Rabier and Liu 2003) that use more observation data within the observation window time, are more appropriate. Application of such algorithms require integration of the additional adjoint model equations. So far, no applications of these algorithms to water quality modeling were attempted.

2.3. The extended Kalman Filter (EKF)

The extended Kalman Filter (EKF), was obtained by modifying the original Kalman filter method using empirical assumptions (Beck and Young 1976; Young 1974). The original Kalman filter linearizes the estimation of mean and covariance while the EKF is the nonlinear version of Kalman filter. The EKF includes two steps, forecast and analysis, that are carried out for each time interval between two updates. As illustrated in Fig. 1b, the EKF creates the background vector of update variables x_j^f at a time j from the previous time step x_{j-1}^a :

$$x_j^f = f\left(x_{j-1}^a\right) \tag{2}$$

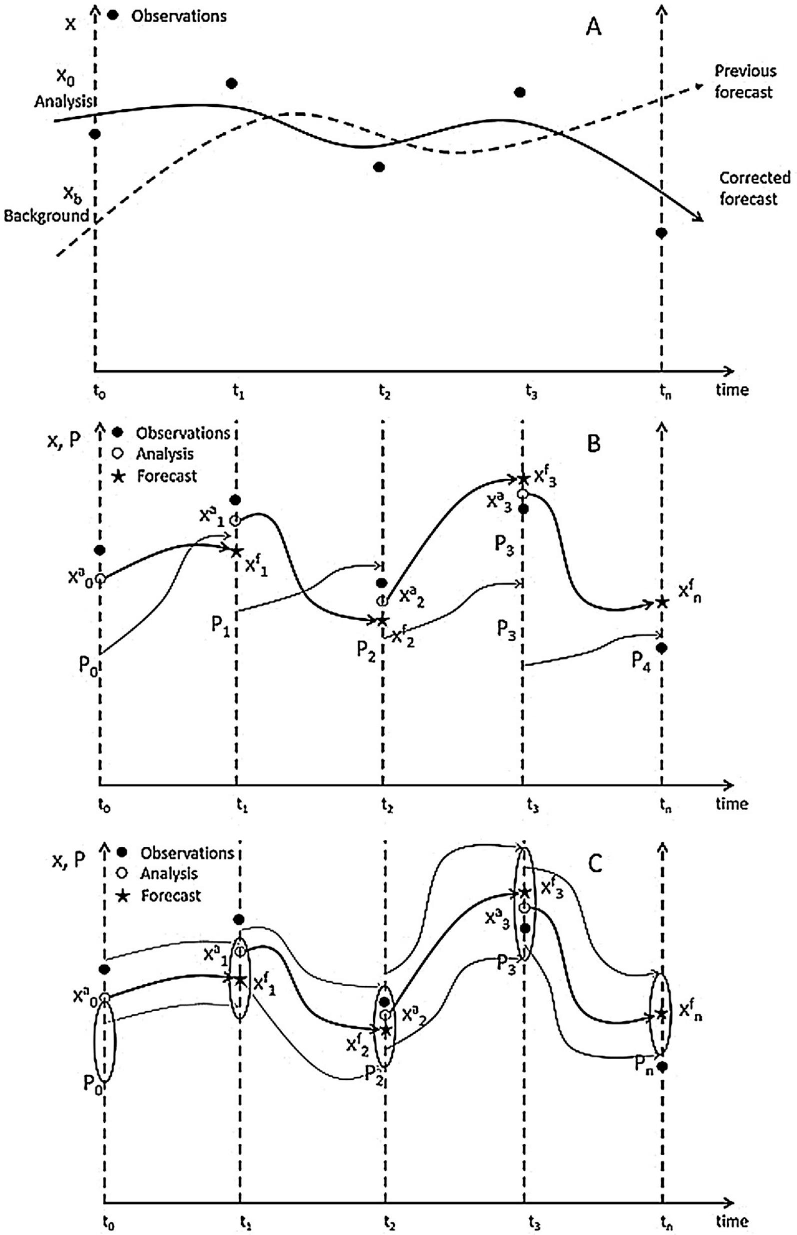

Fig. 1. Schematic diagrams for three different DA techniques; (A) 3DVAR, (B) EKF, and (C) EnKF (Modified from Reichle et al., 2002).

where f (.) is the nonlinear water quality model. It also quantifies the uncertainty of the estimate (i.e., the system error, or model error covariance, P_j) from the previous time step. During the analysis step, the EKF obtains the vector of update variables (x_j^a) from x_j^f using the observation y_j as follows:

$$x_j^a = x_j^f + K_j \left[y_j \; - - \; Hx_j^f \right] \tag{3}$$

Here H is the observation operator which maps the vector of update variables into the observation space, and K_j is the Kalman filter matrix defined as

$$K_j = P_j H_j^T \left[H_j P_j^b H_j^T + R_j \right]^{-1} \tag{4}$$

Here P_j is the error covariance matrix for update variables at time j, and $R_j$is the covariance matrix of the measurement error.

The EKF had been applied during the 1970s and 1980s with simple water quality models to estimate the model parameters and elucidate the performance of the models. In 1976, Beck and Young initiated the application of EKF to the interaction of dissolved oxygen and biochemical oxygen demand in a river. They compared the performance of three relevant models and their kinetic parameters, including reaeration and decay rate. This study was the first to demonstrate the ability of EKF to provide valuable insight on model structure identification. Also, it underscored the EKF limitations in terms of statistically ineffective estimation of parameters. The usefulness of the EKF for evaluating model adequacy was further investigated by Cosby and Hornberger (1984) who estimated DO-associated parameters in photosynthesis-light models for aquatic systems. Five different models generated synthetic data, the random error was applied to modeling results and to parameters. EKF application led to the absence of errors of Type I (i.e., failing to identify the correct model) and errors of Type II (i.e., preferring the incorrect model).

The effect of the inherent variability of water quality attribute (chlorophyll-a, *chl-a*) on the efficiency of EKF was investigated by Whitehead and Hornberger (1984) with data on algae populations in the river Thames. The EKF method was applied to estimate nine algae-related parameters resulting in either an incorrect estimation or a collinearity among parameters. The generalized sensitivity analysis selected three of the most significant parameters (e.g., algal growth rate, a power term on the self-shading factor, and the optimal solar radiation) before applying the EKF for the parameter estimation. This approach effectively resolved the issues and was recommended to make reliable forecasts on algae behavior. Further research on the use of EKF to update parameters in a DO-chlorophyll model from the high frequency field measurements was reported by Pastres et al. (2003) for the lagoon of Venice. These authors concluded that EKF has proved to be a useful tool for the updating of the estimates of the parameters of a simple DO-chlorophyll model, which can be used for linking the high frequency data to meteorological forcings, such as solar radiation and wind, and to other low frequency measurements of water quality parameters, such as the concentrations of *chl-a* and nutrients. In particular, by applying EKF, they found that the DO dynamic was not very sensitive to the tidal effect in the lagoon.

The reduced-order iterated EKF (ROIEKF), the dimension-reduced form of EKF proposed by Cane et al. (1996), was first introduced to include remote sensing observations in the simulation of water quality in a Finnish Lake (Voutilainen et al., 2007). The filtering approach combined prior information of water quality from an evolution model with TSS, Chl, and CDOM from remote sensing instruments, attempting better estimation of water quality. In the data processing, they simply assumed the Gaussian distribution of the system and the observation error of three water quality parameters in the evolution model, but did not include the model parameters as the source of uncertainty. The better performance of the filtering approach than the conventional least-square method was demonstrated in this work.

The first EKF application with simultaneous state and parameter estimation in water quality modeling was reported by Ennola et al. (1998) for modeling the zooplankton population (*Filinia longiseta*). The authors used bootstrap to resample the data from the original measurement dataset and made model runs with the resampled data in an attempt to test the reliability of the method. The EKF appeared to be effective in handling the measurement errors by being relatively insensitive to the level of these errors in the study. However, the authors found two unexpected limitations in applying the EKF to their problem: the time delay in the response to the change of parameters and the overcorrection of parameters after a time delay. These two limitations, which were also observed by Argentesi et al. (1987), were strongly dependent on the errors. Ennola et al. (1998) recommended that in their case the method needed relatively good data. The noise of single samples should be at most 25% to 40%, and the sampling interval should also be short enough for the changes in population dynamics to be detected. Another example of EKF application with joint update of state and model parameters is provided in the work of Mao et al. (2009) who used the advanced ecosystem model to explain the algal bloom events in a marine fish culture zone in Hong Kong. They found that the performance of EKF for chl-a was highly influenced by sampling intervals and prediction lead times. Measurement frequencies could be unequal for different water quality attributes; e. g., the more frequent DO data could compensate for smaller frequency of algal biomass measurements.

Overall, the EKF has been successfully applied to update the state vector and the model parameters of various water quality-related models, with aquatic systems spanning the complexity range from relatively simple DO dynamics to zooplankton behavior (Table 1). The primary advantages of the method are that it allows the estimation of the temporal trajectory of model parameters and allows one to directly see the influence of forcings on the model outputs. The Gaussianity assumption for model and measurement errors can be unrealistic. The linearization in the EKF can introduce uncontrollable errors when strong nonlinearities exist. In addition, if the initial estimate of the state is wrong, or if the process is modeled incorrectly, the filter may quickly diverge, owing this to its linearization. Another problem with the EKF is that the estimated covariance matrix tends to underestimate the true covariance matrix and therefore risks become inconsistent in the statistical sense without the addition of "stabilizing noise" (Huang et al., 2008). Maintaining the covariance matrix of the model errors may be computationally expensive in cases of many assimilation locations, and can be subject to errors related to the nonlinearity of the model.

2.4. Ensemble Kalman filter (EnKF)

The ensemble Kalman filter was introduced to alleviate conceptual and computational problems related to the determination of the model error covariance matrix P in Eq. (4). EnKF approximates the model error covariance matrix with the sample covariance matrix obtained from the ensemble of model runs (Evensen 1994; Houtekamer and Mitchell 1998). Fig. 1c shows the schematics of EnKF operation; each ensemble includes a set of possible model trajectories and their distribution serves as the source of information to find the covariance matrix of update the state variable P that is used in the Kalman filter Eq. (4). Fig. 2 illustrates the typical procedure of EnKF application with the water quality model (Chen et al., 2019). Here, the model was initialized and then run with each ensemble member to estimate the covariance matrix mentioned above. This procedure is applied until the end of the simulation period is reached.

The EnKF is currently the most popular technique of data assimilation in water quality studies, especially when the data sources are remote sensing platforms. In the first application of EnKF, MERIS sensor data on suspended particulate matter was assimilated in the sediment transport model Delft3D-WAQ (El Serafy et al., 2007). It was applied to assimilate the Landsat 7 TM surface water temperature data in the EFDC model (Javaheri et al., 2016). MODIS imagery data on TSS was assimilated in the sediment transport module of the ALGE hydrodynamic model (Vodacek et al., 2008). Assimilating data from multiple sources with EnKF was found beneficial. Page et al. (2018) used buoy and water quality observations to assimilate them into the phytoplankton community model (PROTECH). Chen et al. (2019) assimilated multiple-source data (in situ and remote sensing measurements) into the three-dimensional hydro-ecological dynamics (3DHED) model targeting the cyanobacterial blooms in Lake Taihu. Kim et al. (2014a) expanded application of the EnKF to the assimilation of observations of the multiple water quality parameters (PO_4-P and chl-a) in the EFDC model coupled with the HSPF model.

The number of ensemble members affects results of the EnKF applications. Too small ensembles can cause underestimation of error variances and overestimation of error cross-covariance for the variable update, i.e. analysis of states, in the model. To address these problems, various techniques are used to perform the variance inflation, i.e. increase of variance of the update variables obtained from the ensemble members, and variance localization, i.e. reduction of the Kalman gain matrix in (4). Javaheri et al. (2016) applied covariance inflation and covariance localization in EnKF to assimilate the multi-spectral Landsat-5 TM band temperature data in 300 locations. Experimentation is required to choose parameters of the covariance adjustment procedures.

The large size of the covariance matrices may make results of minimization of the cost function sensitive to noise. The maximum likelihood ensemble filter (MLEF) method addresses this problem by preconditioning, i.e. transformation that replaces the

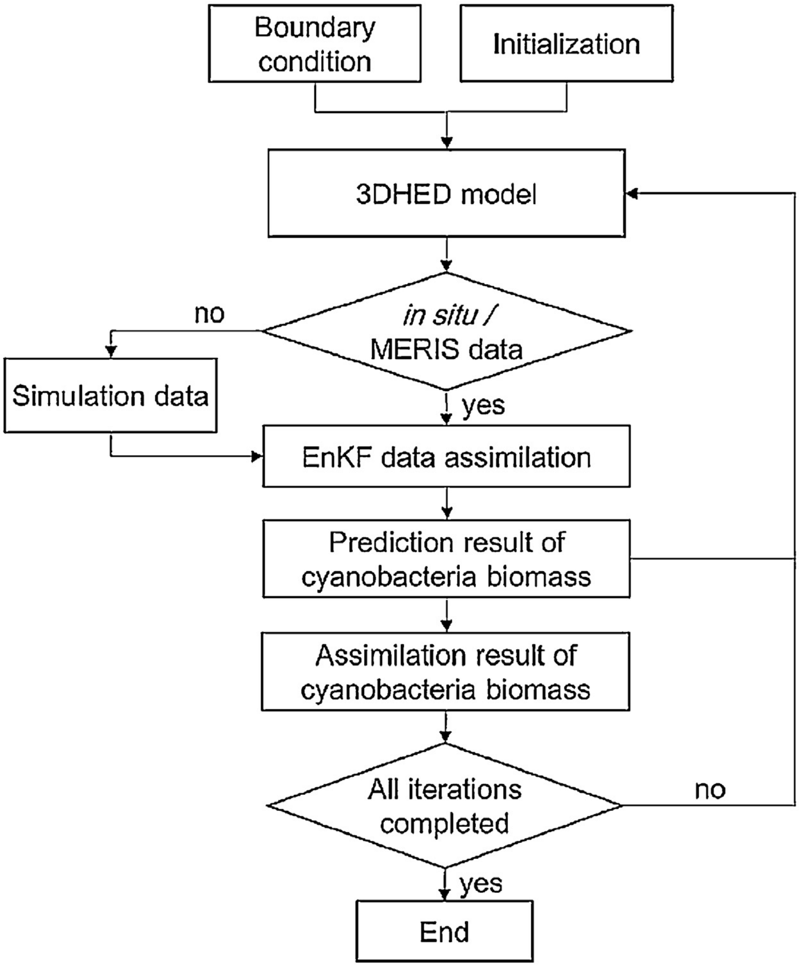

Fig. 2. Typical flowchart of the ensemble Kalman filter applications (adapted from Chen et al., 2019).

minimization of traditional cost function by the minimization procedure that is less sensitive to noise. Kim et al. (2014b) and Riazi et al. (2016) used the MLEF to forecast multiple water quality variables and flow using the HSPF and demonstrated that this DA method can effectively deal with highly nonlinear hydrological and biochemical observation equations.

2.5. Particle filter (PF)

The 3DVAR, EKF, and EnKF assume the Gaussianity of the error statistics for both a priori estimate i.e., background field x_b and observation. The Gaussianity assumption, however, is not necessarily correct because it is not adequate for inherently nonnegative variables in environmental fields. This assumption can be unrealistic for a highly nonlinear feature of water quality (e.g., nonlinear algal responses from various environments) in an aquatic system. One way to address this problem is to use the particle filter (PF) method, which is based on the Sequential Monte Carlo (SMC) simulations and is proposed as an alternative of the Kalman filter-based method. The PF uses the full prior probability function (PDF) without any assumptions on the form of prior PDF. The PF represents the posterior density function using several independent random samples. This method has been broadly used in hydrological modeling studies for a long time (Moradkhani et al., 2005; Weerts and El Serafy 2006), and discussion in water quality modeling about its applicability has started (Huang et al., 2013; Margvelashvili et al., 2010). Franssen and Neuweiler (2015) noted that the future hybrid approaches of EnKF and PF method will be increasingly applied and further developed. Currently, Wang et al. (2019) demonstrated that the PF method is an effective method for assimilation of a 15-min DO observation data in the Seine River system. In addition, it was able to identify the temporal variation of phytoplankton communities by estimating the optimal temperature for the growth of phytoplankton.

3. The timeline of DA applications in water quality modeling

Fig. 3 presents the timeline of changes in water quality model types and data assimilation methods. The milestones were the change in the dimensions of the flow model, introduction of the extended Kalman filter applications, start of using DA to support operational decisions in water quality management, use of remote sensing data, demonstration of the applicability of the 3DVAR DA method, introduction of the MLEF technique and beginning to use DA in watershed scale modeling, current application of the PF method. Evolution of the DA-supported modeling followed the availability of new types of data and quantifying their uncertainty.

4. Influence of data and model uncertainty on DA process

Water sampling, in situ measurements with sensors and remote sensing data have been employed in projects that included water quality data assimilation. Both analytical errors and environmental variability contribute to the uncertainty in observations *per se* and to the uncertainty of the conversion of measurements to the update variables.

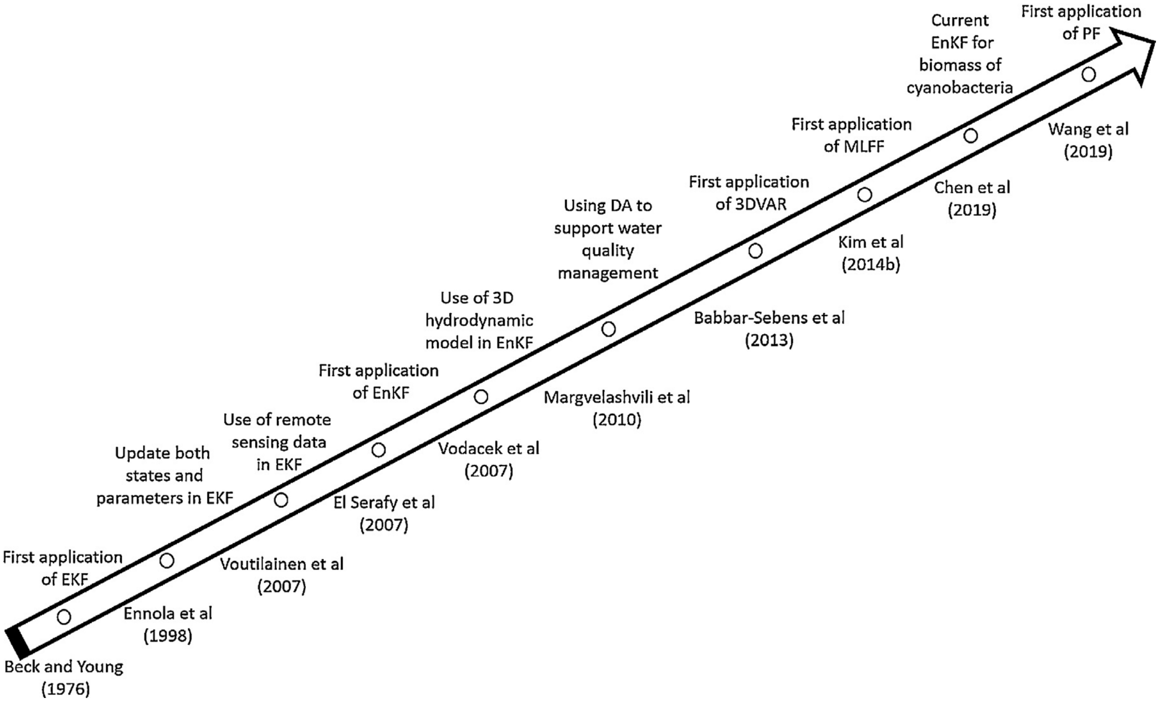
First application of PF
Current EnKF for biomass of cyanobacteria
Wang et al (2019)
First application of MLFF
Chen et al (2019)
First application of 3DVAR
Kim et al (2014b)
Using DA to support water quality management
Babbar-Sebens et al (2013)
Use of 3D hydrodynamic model in EnKF
Margvelashvili et al (2010)
First application of EnKF
Vodacek et al (2007)
Use of remote sensing data in EKF
El Serafy et al (2007)
Update both states and parameters in EKF
Voutilainen et al (2007)
First application of EKF
Ennola et al (1998)
Beck and Young (1976)

Fig. 3. Development of DA applications in the water quality modeling.

The mismatch between spatial and temporal domains of measurement and modeling presents the major difficulty in data assimilation for water quality modeling. Babbar-Sebens et al. (2013) assimilated satellite and sensor temperature data in a water quality model and commented that since the spatial resolution is smaller for TM images than the model grid system, the unit pixel for each data set would be mismatched in spatial scale. For example, when two data sets were overlapped with each other, there would be more than one pixel laid over the single model grid cell. When the random locations were picked, the corresponding value extracted from the TM images would not necessarily be the dominant pixel value located within its model grid size. Similarly, when identifying the in situ measurement locations, the corresponding values would not necessarily be dominant pixel values located within its model grid size. Also, there are substantially inaccurate representations of remote sensing observations on water bodies which are smaller than the pixel size of the satellite image. High uncertainty can also be encountered at the edge of the physical boundaries of water bodies since each cell in the observation includes the composite information from land and water.

The subgrid heterogeneity can complicate imagery data assimilation (Balsamo et al., 2018). An example of such heterogeneity is shown in Fig. 4 with drone-based imagery data for phycocyanin distribution in the Deachung reservoir, Korea (Kwon et al., 2020). Homogeneity of scenes in boxes of 1 and 2 is reflected in symmetrical distributions and the average is expected to give a good representation of the aggregated information for the whole grid cell (pixel). The heterogeneous scene 3 has the distribution of the subgrid reflectance that is far from symmetrical, and no univocal judgement can be made about the aggregation rule for further assimilation.

The mismatch of model and measurement scale in the vertical direction presents another issue that needs to be researched for assimilation of the remote sensing data in water quality models. Vertical gradients in top 20-25 cm layer in fresh water sources can be very high due to steep changes in water absorbance with depth (e. g., Maraccini et al. (2016)). The layer depth reflected in remote sensing-based estimates of water quality parameters differs depending on the type of parameter and properties of the water body. Giardino et al. (2015) used airborne imaging spectrometry and bio-optical algorithms to retrieve concentrations of suspended particulate matter, chlorophyll-a and colored dissolved organic matter. They validated retrievals with data acquired from the top 1 m layer in the study lake. On the other hand, Javaheri et al. (2019) found substantial biases in remotely sensed observations of temperature in surface water with Landsat 7, and removed these biases before the actual data assimilation steps were conducted. El Serafy et al. (2007) emphasized the need to account for the optical depth of the remote sensing products to remove the discrepancies between the remote sensing observations and the model output.

Temporal stability, i.e. presence of a persistent spatial pattern in deviations from average, was observed for various water quality parameters. Examples for concentrations of the fecal indicator organism (*Escherichia coli*) can be found in Pachepsky et al. (2017) and Stocker et al. (2019). This creates the spatially dependent bias that must be removed before the covariance matrix of errors are built. If this bias is not removed, the naïve computation of the covariance matrix for DA purposes will result in overly high values (Feng et al., 2011). The bias needs to be removed because all DA methods usually assume a zero-mean white noise error (Javaheri et al. 2016).

The bias correction may need to be applied to modeling results, too. The bias caused by structural deficiencies of the model can result in physically meaningless update variable values of a water quality model after data assimilation update (Eyre, 2016). Riazi et al. (2016) used the MLEF data assimilation algorithm with the HSPF water quality model, and employed the statistical bias correction procedure to account for systematic errors so that the DA solution may be found within the dynamic range of the model. The linear relationship between the truth and model prediction was used for the correction. The determination of the slope and the intercept of those equation was independent on the DA up-

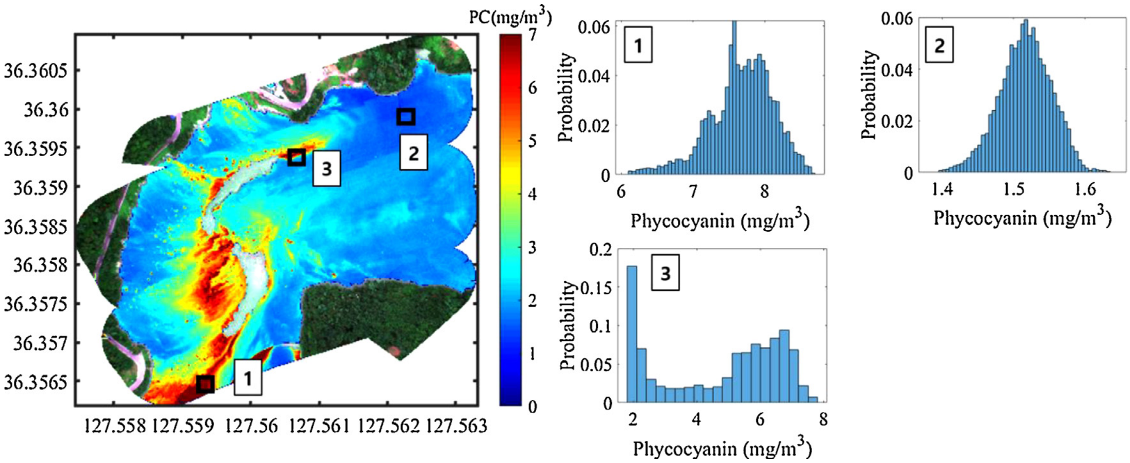

Fig. 4. Phycocyanin distribution retrieved from drone-borne hyperspectral image (Left panel, Kwon et al., 2020) and histograms of phycocyanin (PC) concentrations (Right panels) in three grid cells of a water quality model; 1) high concentration region, 2) low concentration region, and 3) heterogeneous region.

date, and DA was applied with and without bias correction to analyze the effect of this correction. Both spatial bias and model bias were removed in surface water temperature values in the work of Javaheri et al. (2019) before the remote sensing data assimilation update was applied.

It is notoriously difficult to evaluate the uncertainty of modeling results, if system uncertainty that arises from the structural errors in the model (Lawless 2013). In an application of the 3DVAR method, Shao et al. (2016) suggested the use of the applied National Meteorological Center (NMC) method to estimate the background error covariance, referring to the work of Parrish and Derber (1992). The method consists in approximation of the background variances by the variance of simulation results for consecutive update times. In EKF applications, Cosby and Hornberger (1984) assumed that the covariance matrix of system error can be obtained from either the variance of the innovations or the gain for a given R matrix. Pastres et al. (2003) estimated the constant Q using a preliminary two-step analysis which utilized the time series of the residuals and measurements.

Reports on assimilating data from multiple sources appear to be contradictory. Babbar-Sebens et al. (2013) assimilated temperature data from satellite observations and from *in situ* measurements, and found that improvements with respect to data from one data source occurred in parallel with worsening results with respect to another source. The authors explain it by spatial and temporal mismatch between the two sources. On other hand, Chen et al. (2019) assimilated in situ and satellite data relevant to modeling of harmful algal blooms and found that multi-source data was helpful in improving the model performance.

The EnKF method is designed to obtain the uncertainty of modeling results explicitly from the ensemble simulations. The number of ensemble members can be limited due to the heavy computational load of the 3-dimensional modeling. Javaheri et al. (2019) proposed to create small ensembles by perturbing the inputs and model initial conditions via the Latin hypercube sampling method. In such case the covariance matrix of modeling results can be inaccurate. Kim et al. (2014b) demonstrated that the information about the usefulness of DA can be obtained even with limited knowledge about the model result uncertainty. However, small ensembles in some cases have caused divergence and spurious correlations (Javaheri et al., 2016). Kim et al. (2014b) indicated that the application of the maximum likelihood ensemble filter includes the computation of the information matrix that allows the ensemble size to be judged. Overall, the effect of the ensemble size on the assimilation results must be researched for the task in hands.

5. Updating state variables and/or parameters and its influence on DA performance

The subsequent studies with EKF have generated the state-parameter vectors to simultaneously update the state variable and the associated parameters; Whitehead and Hornberger (1984) and Cosby and Hornberger (1984) explored the uncertainty of algae-associated parameters to determine the state variable (i.e., DO concentrations); they choose three significant parameters selected by the sensitivity analysis. Pastres et al. (2003) also applied a sensitivity analysis to select important parameters and included one state variable (DO) into the state-parameter vector. Mao et al. (2009) updated eight water quality variables and two most important parameters, algal growth rate and settling velocity, which were identified by a sensitivity analysis. Huang et al. (2013) used the, hydrodynamic-phytoplankton model that included a total of 15 parameters, with the most sensitive parameter being the maximum growth rate of phytoplankton. Considering the heavy computational burden due to the high spatial resolution (250 m × 250 m) and a relatively large ensemble size (100) of EnKF, only the most sensitive parameter was updated in this EnKF application. However, updating the model parameters did not always produce the improvement of DA performance. Fig. 5 illustrates the absolute error of *chl-a* generated from three different DA strategies, showing that updating the parameter does not improve the modeling result and the parameter uncertainty is relatively insignificant to other uncertainties (e.g., model structure and forcing input). Huang and Gao (2017) divided the year of 2014 into 24 sub-periods and then investigated the two sensitive parameters for each sub-period to be assimilated in the EnKF of the coupled phosphorus (P) model. They found much improved DA results with the parameter dynamic than DA results without the parameter dynamic.

Some researchers reported results of updating water quality variables that have not been measured along with other variables that were measured. Kim et al. (2014a) simulated river hydrodynamics and water quality. Although only *chl-a* data was involved in the assimilation, phosphate was selected among other water quality variables for update to evaluate the effect of *chl-a* assimilation on those variables. It turned out that the phosphate simulation was not improved by the *chl-a* data, which was due to weak correlation between the two variables in the model ensemble. Mao et al. (2009) applied data assimilation in modeling of algal bloom and observed that more frequent DO data can compensate for less frequent algal biomass measurements.

Estimating uncertainty of model parameters, as well as boundary conditions, is much more difficult than that of state variables.

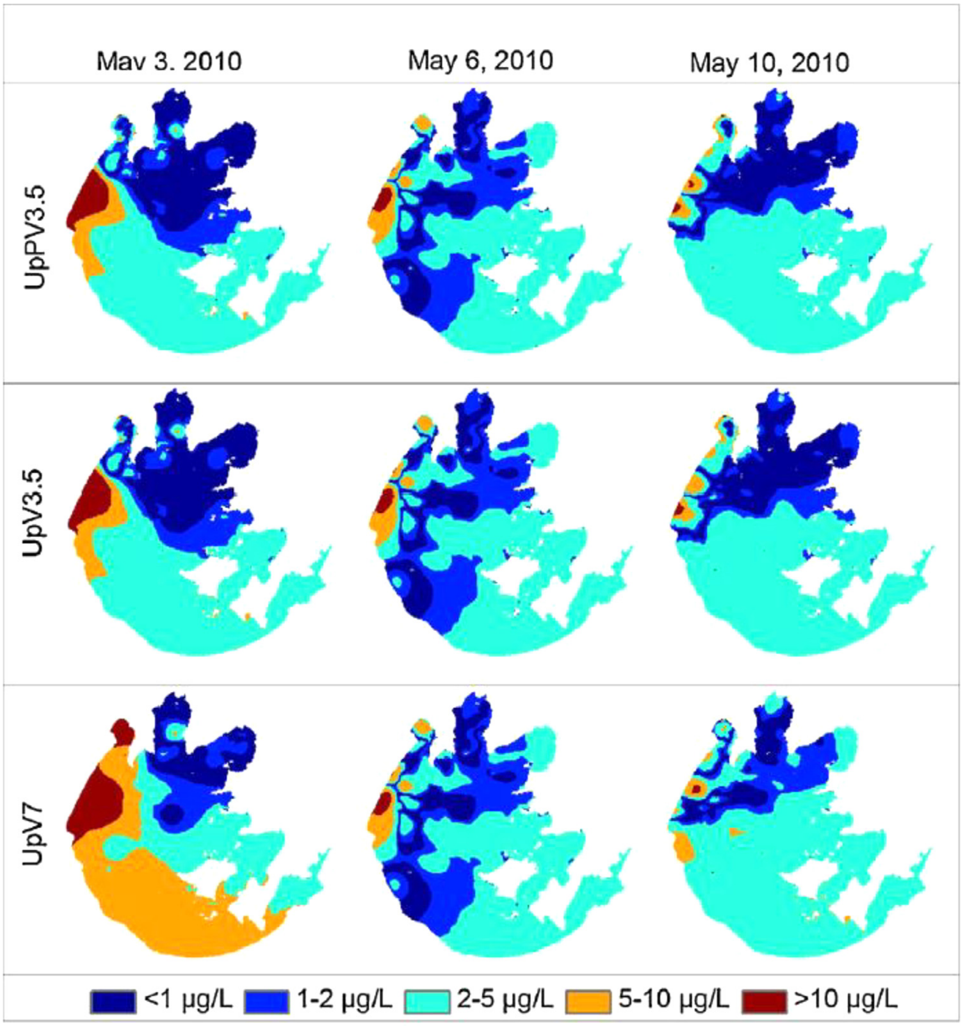

Fig. 5. Absolute errors of the water quality model estimating chlorophyll-a (*chl-a*)with EnKF application; UpPV3.5 updated both the initial *chl-a* and parameter updated twice a week. UpV3.5 and UpV7 updated only the initial *chl-a* at twice a week and once a week, respectively (Huang et al., 2013).

Huang et al. (2013) noted that such estimates are mostly determined empirically (e.g., with a certain percent of the initial value); 5–35% of the initial value was generally used as the relative standard deviation of the observational error. Considering that phytoplankton varies significantly even in the short term in Lake Taihu, the authors used 35% of the initial value as the standard deviation of the observational error. The standard deviation of maximum growth rate of phytoplankton was empirically set as 0.1 that is large enough to account for its uncertainty. The noise added to the forcing data was proportional to their magnitudes. The proportionality factor was set to 0.1, that was assumed to be reasonable to describe uncertainty from forcing data.

6. Influence of frequency of updates and forecast skill on DA process

The DA performance is very sensitive to the update time interval. Kim et al. (2014b) performed the sensitivity analysis to determine the optimal assimilation window and found 7 days to be sufficient for the watershed scale modeling. In simulations of the algal bloom dynamics, Mao et al. (2009) assimilated environmental data from multiple sources which had different observations frequency; chlorophyll (1-day interval), DO (2 hour), hydrometeorological data (1 hour), and nutrient data (bi-weekly). The EKF performance was evaluated for lower frequency of *chl-a* (1-, 2-, and 3-day) and frequency of DO (6-, 12-, and 24-hour). In general, longer update time interval resulted in lower accuracy of prediction on *chl-a*. The authors observed that the DO update time interval became very influential in model performance with the 3-day *chl-a* update time interval. Shorter DO sampling time still showed a high correlation, but longer DO sampling interval resulted in dramatic deterioration of the model performance.

The update time interval has not necessarily been constant. Javaheri et al. (2019) assimilated multiple-sensor water temperature data in the EFDC model, and applied an adaptive EnKF to determine the optimal time to assimilate *in situ* measurement into the model by introducing the threshold error. Whenever the error of the model for in situ measurements is greater than the threshold value, new *in situ* measurements were assimilated into the model.

The deterioration of forecast skills with the forecast time increase was noted by several authors. Javaheri et al. (2019) reported that error in the water temperature predicted by the updated model reverted in less than two days to the same level as that of an un-updated model. Page et al. (2018) determined that there was a general reduction in forecast skill with increasing forecasting period but forecasts for up to four or five days showed noticeably greater promise than those for longer periods. Mao et al. (2009) found that predictions with 1 to2 dsay lead-time were highly correlated with the observations (r = 0.7–

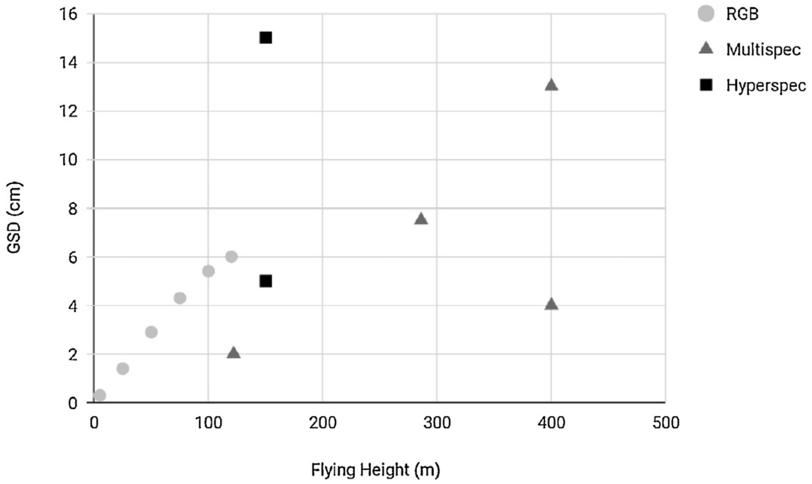

Fig. 6. Ground sampling distance and Flying height (m) for different camera/sensor type (adopted from Kislik et al., 2018).

0.9); the correlation stayed at a high level for a lead-time of 3 days (r = 0.6–0.7). Hydrodynamic modeling may affect the suitability of update intervals for the water quality assessment. Kim et al. (2014a) found that the influence of DA applied to the EFDC simulating river water quality did not last for a long time in the upper reaches of the river where the flow velocity is relatively high.

7. Outlook: challenges and opportunities

Data assimilation is becoming a staple in water quality management (Riazi et al., 2016; Romas et al., 2018). Table 1 shows that DA is applied to different target variables of water quality control. The availability of monitoring data defines the usability of the DA. For example, the DA can be especially efficient in modeling of the spread of waterborne pathogenic organisms because the extensive monitoring data become available in case of outbreaks. Successful applications of data assimilation in water quality modeling can stimulate research directed towards further improvements.

The number of possible data sources for data assimilation in water quality modeling is steadily growing. Currently, unmanned aerial vehicles (UAVs) with optical sensors became popular for environmental monitoring due to their high spatial resolutions (Kislik et al., 2018). The drone-borne hyperspectral imagery has sub-meter spatial resolution (Fig. 6) and optical resolutions on water quality that satellite and airborne imagery cannot provide. The UAV with optical sensors have been recently adopted for monitoring algal blooms in surface water bodies and demonstrated its ability to quantify algal species using various indexes including Normalized Difference Vegetation Index (NDVI), Green Leaf Index (GLI), and Algal Bloom Detection Index (AI) (Goldberg et al., 2016; Honkavaara et al., 2013; Jang et al., 2016; Kim et al., 2016; Su and Chou 2015; Xu et al., 2018). Images from UAVs promise to be excellent resources of data that can be assimilated into water quality models.

Data uncertainty topics are critical for the efficiency of the data assimilation, and research is required. Not only spatial but also temporal mismatch between observations and models may need to be resolved. Availability of high frequency measurements may result in time steps of model encompassing several measurement time steps. It is not obvious what statistic of the measurement dataset obtained during the model time step should be used in comparisons of modeling results obtained during this time step. Both temporal and spatial scales may be different for different types of data when multiple data sources are used in DA.

The data assimilation algorithms have many modifications that account for the specifics of the problem at hand. Off-the-shelf versions will not necessarily work in a satisfactory manner. The conversion of update variables to observations (reflected by the observation operator H in Eqs. (1) and (4)) may have an uncertainty strongly dependent on the range of the update variable values as controlled by the sensitivity of the sensor. The influence of missing data on data assimilation results is not known and needs to be clarified. The structural uncertainty of the model can be caused by its hydrology/hydrodynamics module, that limits the efficiency of the data assimilation applied to chemical and biological parameters (i.e. Riazi et al. (2016)). Another limitation may arise due to the effect of the update frequency on the improvement of the model performance after DA. Preliminary experimentation with DA before application presents a worthy research topic.

Most of DA applications were developed using the HSPF model for watersheds and the EFDC for water bodies (Table 1). Other models that are perfect candidates for DA applications are as follows: DRAINMOD (Skaggs et al., 2012), ProSe (Even et al., 1998; Flipo et al., 2004; Vilmin et al., 2015), AGNPS (Young et al., 1989), SWAT+ (SWAT+, 2020) in watershed water quality modeling and DYRESM (Hamilton and Schladow 1997), PROTECH (Reynolds et al., 2001) or HYPE (Lindström et al., 2010) for modeling water bodies. Besides, as the frameworks supporting modularity in water quality modeling are being developed (Whelan et al., 2014), the opportunities for DA applications will increase. The uncertainty in model predictions are in part dependent on model structure, and it is possible that the efficiency of the DA may be one of the criteria for the model selection.

The sensitivity analysis is expected play an influential role in controlling the DA efficiency if the joint state and parameter updates are undertaken. Selecting parameters to update may lead to more robust DA results. We did not find published examples of applying the sensitivity analysis in conjunction with DA. There appears to be a need in research on combining the DA and the sensitivity analysis.

8. Conclusion

The number of the papers on data assimilation in water quality modeling is relatively small. Each of these papers summarizes a large project specifically focused on DA application. However, the volume of the DA applications in the water quality arena can be much larger. Most of modeling works that involve calibration with monitoring data provide material for the data assimilation studies. Such studies can provide the valuable information about the possible time-dependence of model parameters as well as about reliability of calibration results.

The data assimilation in water quality modeling is steadily developing. Each of applications of data assimilation in water quality modeling has provided useful insights into the functioning of complex aquatic systems. Data assimilation is an efficient means of model accuracy improvement. It also offers the opportunity of tracking parameter changes in the varying environment. The possibility to improve the knowledge about the system with each new observation enriches monitoring information and may help to guide and correct the monitoring program. Explicit accounting for uncertainties in data and models will help to reach informed decisions on managing water quality. Lately, DA applications have begun to address regulatory and management needs with watershed-scale hydrological and 3D hydrodynamic models acquiring improved water quality components.

Data acquisition capabilities for the DA in water quality arena are steadily improving, especially in applications to harmful algal blooms where new remote sensing and proximal sensing platforms offer a treasure trove on useful information. The multisource DA has been tried. Research appears to be needed on unresolved issues concerning compatibility in space and time between model resolution, data resolution, and data from different sources. More insight must be gained on data uncertainty of key DA input. Also, an understanding needs to be developed on how the absence of monitoring influential compartments of aquatic systems, such as bottom sediments and periphyton, affect the performance of DA on water quality models.

Declaration of Competing Interest

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.

Acknowledgement

This research was supported by the Basic Core Technology Development Program for the Oceans and the Polar Regions of the National Research Foundation (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2016M1A5A1027457) and USDA-REE-ARS project 8042-12630-011-00D.

References

Argentesi, F., De Bernardi, R., Di Cola, G., Manca, M., 1987. Mathematical modeling of Daphnia populations. In: Peters, R.H., De Bernardi. Mem. Ist, Ital. Idrobiol 45, 389–412.

Asch, M., Bocquet, M., Nodet, M., 2016. Data assimilation: methods, algorithms, and applications. SIAM.

Babbar-Sebens, M., Li, L., Song, K., Xie, S., 2013. On the use of Landsat-5 TM satellite for assimilating water temperature observations in 3D hydrodynamic model of small inland reservoir in Midwestern US. Adv. Remote Sensing 2, 214.

Balsamo, G., Agusti-Panareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., Bousserez, N., Boussetta, S., Brown, A., Buizza, R., 2018. Satellite and in situ observations for advancing global Earth surface modelling: a review. Remote Sensing 10, 2038.

Beck, B., Young, P., 1976. Systematic identification of DO-BOD model structure. J. Env. Eng. Div. 102, 909–927.

Cane, M.A., Kaplan, A., Miller, R.N., Tang, B., Hackert, E.C., Busalacchi, A.J., 1996. Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter. J. Geophys. Res.: Oceans 101, 22599–22617.

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. Env. Modell. Software 117, 188–199.

Cho, K.H., Pachepsky, Y.A., Oliver, D.M., Muirhead, R.W., Park, Y., Quilliam, R.S., Shelton, D.R., 2016. Modeling fate and transport of fecally-derived microorganisms at the watershed scale: state of the science and future opportunities. Water Res. 100, 38–56.

Cosby, B.J., Hornberger, G.M., 1984. Identification of photosynthesis-light models for aquatic systems I. Theory and simulations. Ecol. Modell. 23, 1–24.

El Serafy, G.Y., Blaas, M., Eleveld, M.A., Van Der Woerd, H.J., 2007. Data assimilation of satellite data of suspended particulate matter in Delft3D-WAQ for the North Sea.. In: Proceedings of the Joint EUMETSAT/AMS Conference, Darmstadt, Germany, pp. 1–8 Citeseer.

Ennola, K., Sarvala, J., Dévai, G., 1998. Modelling zooplankton population dynamics with the extended Kalman filtering technique. Ecol. Modell. 110, 135–149.

EPA, U.S., 2018. Surface Water Quality Modeling https://www.epa.gov/waterdata/surface-water-quality-modeling.

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). In: Oceans, Rivers and Lakes: Energy and Substance Transfers at Interfaces. Springer, pp. 27–45.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res.: Oceans 99, 10143–10162.

Eyre, J.R., 2016. Observation bias correction schemes in data assimilation systems: A theoretical study of some of their properties. Quarterly Journal of the Royal Meteorological Society 142 (699), 2284–2291.

Feng, L., Palmer, P.I., Yang, Y., Yantosca, R.M., Kawa, S.R., Paris, J.D., Matsueda, H., Machida, T., 2011. Evaluating a 3-D transport model of atmospheric CO_2 using ground-based, aircraft, and space-borne data. Atmos. Chem. Phys. 11, 2789–2803.

Fletcher, S.J., 2017. Data assimilation for the geosciences: From theory to application. Elsevier.

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

Franssen, H.J.H., Neuweiler, I., 2015. Data assimilation for improved predictions of integrated terrestrial systems. AdWR 86, 257–259.

Giardino, C., Bresciani, M., Valentini, E., Gasperini, L., Bolpagni, R., Brando, V.E., 2015. Airborne hyperspectral data to assess suspended particulate matter and aquatic vegetation in a shallow and turbid lake. Remote Sens. Environ. 157, 48–57.

Goldberg, S.J., Kirby, J.T., & Licht, S.C. (2016). Applications of aerial multi-spectral imagery for algal bloom monitoring in Rhode Island. SURFO Technical Report No. 16-01, 28

Hamilton, D.P., Schladow, S.G., 1997. Prediction of water quality in lakes and reservoirs. Part I—Model description. Ecol. Modell. 96, 91–110.

Honkavaara, E., Hakala, T., Kirjasniemi, J., Lindfors, A., Mäkynen, J., Nurminen, K., Ruokokoski, P., Saari, H., Markelin, L., 2013. New light-weight stereosopic spectrometric airborne imaging technology for high-resolution environmental remote sensing case studies in water quality mapping. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 1, W1.

Houtekamer, P.L., Mitchell, H.L., 1998. Data assimilation using an ensemble Kalman filter technique. Monthly Weather Rev. 126, 796–811.

Huang, G.P., Mourikis, A.I., Roumeliotis, S.I., 2008. Analysis and improvement of the consistency of extended Kalman filter based SLAM.. In: 2008 IEEE International Conference on Robotics and Automation. IEEE, pp. 473–479.

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. Ecol. Modell. 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. Ecol. Modell. 263, 81–91.

Jang, S.W., Yoon, H.J., Kwak, S.N., Sohn, B.Y., Kim, S.G., Kim, D.H., 2016. Algal bloom monitoring using UAVs imagery. Adv. Sci. Technol. Lett. 138, 30–33.

Javaheri, A., Babbar-Sebens, M., Miller, R.N., 2016. From skin to bulk: An adjustment technique for assimilation of satellite-derived temperature observations in numerical models of small inland water bodies. Adv. Water Res. 92, 284–298.

Javaheri, A., Babbar-Sebens, M., Miller, R.N., Hallett, S.L., Bartholomew, J.L., 2019. An adaptive ensemble Kalman filter for assimilation of multi-sensor, multi-modal water temperature observations into hydrodynamic model of shallow rivers. J. Hydrol. 572, 682–691.

Ji, Z.-G., 2017. Hydrodynamics and water quality: modeling rivers, lakes, and estuaries. John Wiley & Sons.

Kebede, A. (2009). Water quality modeling: An overview, https://files.nc.gov/ncdeq/Water%20Quality/Planning/TMDL/Modeling/Modeling%20101%20for%20FON%20stakeholder%20May09.pdf

Kim, H.-M., Yoon, H.-J., Jang, S.W., Kwak, S.N., Sohn, B.Y., Kim, S.G., Kim, D.H., 2016. Application of unmanned aerial vehicle imagery for algal bloom monitoring in river basin. Int. J. Control and Auto. 9, 203–220.

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

Kim, S., Seo, D.-J., Riazi, H., Shin, C., 2014b. Improving water quality forecasting via data assimilation – Application of maximum likelihood ensemble filter to HSPF. J. Hydrol. 519, 2797–2809.

Kislik, C., Dronova, I., Kelly, M., 2018. UAVs in support of algal bloom research: a review of current applications and future opportunities. Drones 2, 35.

Kwon, Y.S., Pyo, J., Kwon, Y.-H., Duan, H., Cho, K.H., Park, Y., 2020. Drone-based hyperspectral remote sensing of cyanobacteria using vertical cumulative pigment concentration in a deep reservoir. Remote Sens. Environ. 236, 111517.

Lawless, A.S., 2013. Variational data assimilation for very large environmental problems. Large Scale Inverse Problems: Comput. Methods Applic. Earth Sci. 2, 55–90.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., Arheimer, B., 2010. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. Hydrol. Res. 41, 295–319.

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 Res. 171, 115343.

Mao, J.Q., Lee, J.H.W., Choi, K.W., 2009. The extended Kalman filter for forecast of algal bloom dynamics. Water Res. 43, 4214–4224.

Maraccini, P.A., Mattioli, M.C.M., Sassoubre, L.M., Cao, Y., Griffith, J.F., Ervin, J.S., Van De Werfhorst, L.C., Boehm, A.B., 2016. Solar inactivation of enterococci and Escherichia coli in natural waters: effects of water absorbance and depth. Environ. Sci. Technol. 50, 5068–5076.

Margvelashvili, N., Parslow, J.S., Herzfeld, M., Wild-Allen, K., Andrewartha, J., Rizwi, F., Jones, E., 2010. Development of operational data-assimilating water quality modelling system for South-East Tasmania.. OCEANS'10 IEEE SYDNEY 1–5.

Marx, B.A., Potthast, R.W., 2012. On instabilities in data assimilation algorithms. GEM-Int. J. on Geomathematics 3, 253–278.

Montzka, C., Pauwels, V., Franssen, H.-J.H., Han, X., Vereecken, H., 2012. Multivariate and multiscale data assimilation in terrestrial systems: A review. Sensors 12, 16291–16333.

Moradkhani, H., Hsu, K.L., Gupta, H., Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water Resour. Res. 41.

Moradkhani, H., Sorooshian, S., 2009. General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In: Hydrological modelling and the water cycle. Springer, pp. 1–24.

O'Neill, A. (2003). Introduction to data assimilation. ESA-ESRIN, Frascati, Rome, Italy, https://earth.esa.int/documents/973910/979015/oneill1-2.pdf.

Pachepsky, Y., Kierzewsky, R., Stocker, M., Mulbry, W., Millner, P., Shelton, D., 2017. Temporal stability of E. coli concentration patterns in two irrigation ponds in Maryland.. EGU General Assembly Conference Abstracts 3763.

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

Park, S.K., Xu, L., 2013. Data assimilation for atmospheric, oceanic and hydrologic applications. Springer Science & Business Media.

Parrish, D.F., Derber, J.C., 1992. The national meteorological center's spectral statistical-interpolation analysis system. Monthly Weather Rev. 120, 1747–1763.

Pastres, R., Ciavatta, S., Solidoro, C., 2003. The Extended Kalman Filter (EKF) as a tool for the assimilation of high frequency water quality data. Ecol. Modell. 170, 227–235.

Rabier, F., Liu, Z., 2003. Variational data assimilation: theory and overview.. ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean.

Reichle, R.H., McLaughlin, D.B., Entekhabi, D., 2002. Hydrologic data assimilation with the ensemble Kalman filter. Monthly Weather Rev. 130, 103–114.

Reynolds, C., Irish, A., Elliott, J., 2001. The ecological basis for simulating phytoplankton responses to environmental change (PROTECH). Ecol. Modell. 140, 271–291.

Riazi, H., Kim, S., Seo, D.-J., Shin, C., Kim, K., 2016. Improving Operational Water Quality Forecasting with Ensemble Data Assimilation. J. Water Manage. Modell..

Robinson, A.R., Lermusiaux, P.F., 2000. Overview of data assimilation. Harvard Rep. In Physical/Interdisciplinary Ocean sci. 62, 1–13.

Romas, E., Tzimas, A., Kandris, K., Pechlivanidis, I., Boultadakis, G., Giannakoulias, A., Schenk, K., Giardino, C., Bresciani, M., 2018. Operational short-term water quantity and quality forecasting in reservoirs intended for potable water production. EGUGA 7090.

Sasaki, Y., 1958. An objective analysis based on the variational method. J. Meteorol. Soc. Jpn. Ser. II 36, 77–88.

Shao, D., Wang, Z., Wang, B., Luo, W., 2016. A water quality model with three dimensional variational data assimilation for contaminant transport. Water Resour. Manage. 30, 4501–4512.

Skaggs, R.W., Youssef, M., Chescheir, G., 2012. DRAINMOD: Model use, calibration, and validation. Trans. ASABE 55, 1509–1522.

Stocker, M.D., Pachepsky, Y.A., Hill, R.L., Sellner, K.G., Macarisin, D., Staver, K.W., 2019. Intraseasonal variation of E. coli and environmental covariates in two irrigation ponds in Maryland, USA. Sci. Total Environ. 670, 732–740.

Streeter, H., Phelps, E.B., 1925. A study of the pollution and natural purification of the Ohio river, III, factors concerned in the phenomena of oxidation and reaeration. US public health service. Public Health Bul. 146, 75.

Su, T.-C., Chou, H.-T., 2015. Application of multispectral sensors carried on unmanned aerial vehicle (UAV) to trophic state mapping of small reservoirs: a case study of Tain-Pu reservoir in Kinmen, Taiwan. Remote Sens. 7, 10078–10097.

Sun, L., Seidou, O., Nistor, I., Liu, K., 2016. Review of the Kalman-type hydrological data assimilation. Hydrol. Sci. J. 61, 2348–2366.

SWAT+, 2020. https://swat.tamu.edu/software/plus/ (Accessed 19 August, 2020).

Vilmin, L., Flipo, N., De Fouquet, C., Poulin, M., 2015. Pluri-annual sediment budget in a navigated river system: the Seine River (France). Sci. Total Environ. 502, 48–59.

Vodacek, A., Li, Y., Garrett, A.J., 2008. Remote sensing data assimilation in environmental models.. 2008 37th IEEE Appl. Imagery Pattern Recog. Workshop 1–5.

Voutilainen, A., Pyhälahti, T., Kallio, K.Y., Pulliainen, J., Haario, H., Kaipio, J.P., 2007. A filtering approach for estimating lake water quality from remote sensing data. Int. J. Appl. Earth Obs. Geoinf. 9, 50–64.

Wang, Q., Li, S., Jia, P., Qi, C., Ding, F., 2013. A Review of Surface Water Quality Models. The Scientific World J. 2013, 7.

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

Weerts, A.H., El Serafy, G.Y., 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resour. Res. 42.

Wellen, C., Kamran-Disfani, A.-R., Arhonditsis, G.B., 2015. Evaluation of the current state of distributed watershed nutrient water quality modeling. Environ. Sci. Technol. 49, 3278–3290.

Whelan, G., Kim, K., Pelton, M.A., Soller, J.A., Castleton, K.J., Molina, M., Pachepsky, Y., Zepp, R., 2014. An integrated environmental modeling framework for performing quantitative microbial risk assessments. Environ. Modell. Software 55, 77–91.

Whitehead, P.G., Hornberger, G.M., 1984. Modelling algal behaviour in the river thames. Water Res. 18, 945–953.

Xu, F., Gao, Z., Jiang, X., Shang, W., Ning, J., Song, D., Ai, J., 2018. A UAV and S2A data-based estimation of the initial biomass of green algae in the South Yellow Sea. Mar. Pollut. Bull. 128, 408–414.

Young, P., 1974. Recursive approaches to time-series analysis. Bull. Inst. Maths. Appl. 10, 209–211.

Young, R., Onstad, C., Bosch, D., Anderson, W., 1989. AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds. J. Soil Water Conserv. 44, 168–173.