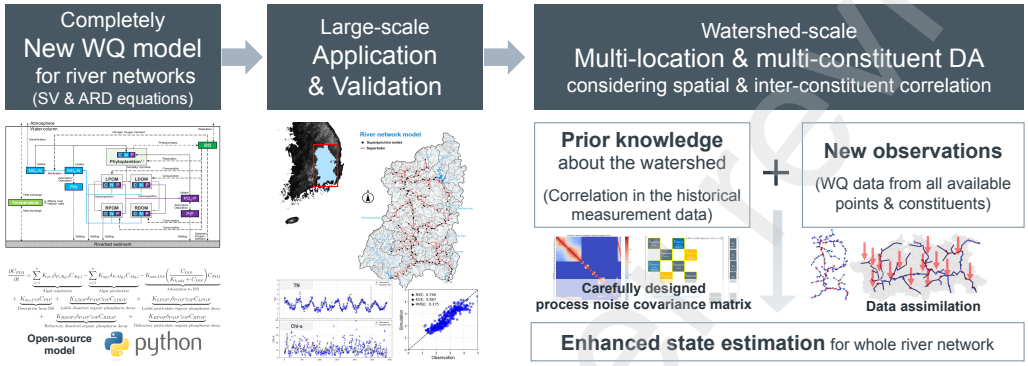
Graphical Abstract

EufoRiA: A new multi-constituent nutrient and harmful algal blooms model for river networks with online data assimilation

Min-Gyu Kim, Matthew Bartos



Effective tool for
Monitoring, prediction, and forecasting of WQ and HABs
in large river networks

Highlights

EufoRiA: A new multi-constituent nutrient and harmful algal blooms model for river networks with online data assimilation

Min-Gyu Kim, Matthew Bartos

- New physics-based nutrients and HABs model for online modeling of river basins is introduced.
- Support for unsteady hydraulics allows modeling of regulated river systems.
- Validation on large river basin shows competitive skill with existing HAB models.
- Online Kalman Filtering scheme improves predictions by fusing observations into model.
- Kalman Filter greatly improves estimation performance beyond model-only approaches.

EufoRiA: A new multi-constituent nutrient and harmful algal blooms model for river networks with online data assimilation

Min-Gyu Kim[a,b], Matthew Bartos[a]

[a]*Maseeh Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, Austin, 78712, TX, USA*
[b]*Korea Water Resources Corporation, Daejeon, Daedeok-gu, South Korea*

Abstract

Surface water quality impairment is an increasing challenge for water managers in the face of urbanization and climate change. While contaminant fate and transport models are essential for addressing water quality threats like harmful algal blooms, there is a lack of models designed for real-time simulation and decision support in large, regulated river basins. We propose EufoRiA, a new water quality model for river networks incorporating unsteady hydraulics, contaminant transport, reaction kinetics for 23 eutrophication-related constituents, and an online data assimilation scheme using Kalman Filtering that integrates real-time observations to improve model performance. Validating against long-term data from South Korea's Nakdong River, EufoRiA offers competitive performance with existing models in predicting constituents like nitrogen, phosphorus, and algae. Moreover, data assimilation significantly improves water quality constituent estimation compared to model-only approaches, particularly at ungaged locations. EufoRiA will enable enhanced decision-making for public safety and health against increasing water quality threats.

Keywords: Nutrient and HABs modeling, Watershed water quality management, River networks model, Kalman filtering, Data assimilation, Digital twins

1. Introduction

Effective management of water quality in surface water systems is crucial for providing clean and safe water resources to the public. In recent decades, urbanization and climate change have contributed to a rapid increase of point source and non-point source pollution from watersheds. These pollutants have led to high nutrient levels in surface water systems that result in ecosystem degradation and Harmful Algal Blooms (HABs) [1, 2]. Because HABs produce toxic substances like *Microcystin* that pose risks to human health (notably liver cancer [3]), HABs have been a global concern for decades. For example, Lake Erie in the US experienced an extreme HAB event in 2014 that disrupted the drinking water supply [4], and the Nakdong River in South Korea has suffered recurrent HABs during summer due to favorable environmental drivers, including excessive nutrient levels, high water temperatures, intensive solar radiation, and slow river velocity [5, 6]. Furthermore, climate change is expected to increase the frequency and severity of HABs because higher temperatures create favorable conditions for phytoplankton growth, and increased rainfall intensity leads to greater nutrient runoff from watersheds [7, 8, 9]. To address these challenges, measures to reduce HABs are essential for ensuring safe drinking water and protecting environmental health.

Researchers have implemented a variety of strategies to combat HABs that may be classified into either preventive or adaptive approaches [10]. Preventive measures focus on reducing nutrient loads from point- and nonpoint-pollution sources, as well as biochemical manipulations to limit HAB outbreaks. However, reducing pollution sources at the watershed scale typically requires effective regulations and stepwise reduction programs supported by long-term budgets [11]. Because reducing HABs through preventive methods is often limited by budgetary and political constraints, adaptive measures are also necessary to protect public health and ecosystems. These adaptive measures seek to respond to HABs in real-time, and may include early warning systems to limit public exposure as well as proactive control of hydraulic infrastructure like dams and weirs to minimize HAB proliferation [10, 12, 13]. These adaptative measures require reliable HAB prediction models that integrate broad-scale monitoring and accurate geographical and biochemical knowledge of the target river system.

In this study, we focus on the scientific modeling and forecasting of nutrient and algae dynamics at the river network scale as an adaptive measure

for reducing the adverse impacts of HABs. We begin by reviewing recent advances in physics-based modeling of nutrients and HABs in surface water systems, and identify several key knowledge gaps that limit the deployment of real-time models for watershed-scale nutrient management.

2. Background

Broadly, HABs modeling can be divided into two main approaches: empirical models that are derived from observed data, and physics-based models that incorporate prior information on contaminant fate and transport from known physical laws [13]. Although studies in recent decades have introduced novel empirical approaches based on machine learning algorithms [14, 15], limitations such as a lack of interpretability and transferability pose challenges for operational adoption [16]. Given these challenges, this study focuses on physics-based models, which are the most widely-used type of model in operational use today, and are suitable for scenario-based impact assessments of reduction measures and long-term planning requiring future projection capabilities [13].

A variety of physics-based models have been developed to predict HABs in surface water systems, with applications ranging from high-fidelity simulation of individual waterbodies to base-level assessment of large river systems [17, 18, 19, 20, 21]. For comprehensive water quality assessments, three-dimensional models like EFDC and CAEDYM are often chosen for their ability to simulate both unsteady hydraulics and complex water quality dynamics involving nutrients, algae and other constituents like epiphyton, macrophytes, metals, bacteria, and fish [22, 21]. Due to their high computational demands and input data requirements, these models are generally reserved for event-based simulation of individual estuaries and reservoirs with known water quality problems. For management of larger river systems, two-dimensional models offer a compromise between model fidelity and practical utility. Within this vein, CE-QUAL-W2 is a two-dimensional vertical hydrodynamics and water quality model that has been applied to rivers and reservoirs throughout the world, including in Korea where it serves as the primary water quality model used by the state water utility to simulate major rivers and their tributaries [23]. Finally, one-dimensional models offer a practical choice for water managers seeking to assess nutrient and HAB dynamics for large-scale river networks. Among these models, QUAL2E is one of the most widely used water quality models, owing to its long devel-

opment history and numerous application cases, with QUAL2K representing the latest installment in the QUAL series [18]. Likewise, HSPF is a network-scale one-dimensional watershed model that serves as a popular choice for nutrient fate and transport modeling in large river basins [24]. Existing one-dimensional models generally feature simplified hydraulics, with the QUAL models pertaining to steady-state conditions, and HSPF featuring a unidirectional kinematic wave routing scheme that is best suited for natural river systems where backwater effects are limited.

Apart from the development of physics-based models of nutrient and HAB dynamics, recent work has shown success in improving HAB forecasts through the use of data assimilation (DA) techniques that combine physically-based models together with real-time observational data [25]. In the context of nutrient and HABs modeling, DA seeks to incorporate water quality observations (e.g. chlorophyll measurements) into physically-based models of nutrient cycling and algal bloom dynamics to generate improved forecasts and estimate HAB potential in ungaged locations. Studies have explored a variety of different DA approaches including Extended Kalman Filtering (EKF) [26, 27, 28, 29], Ensemble Kalman Filtering (EnKF) [30, 31, 32, 33, 34, 35], Particle Filtering (PF) [36], Maximum Likelihood Estimation [**?** 37], and Variational Data Assimilation (3DVAR) [38]. These DA approaches have been integrated with a host of different water quality models, ranging from simplified '0-dimensional' tank models [39, 40, 41, 29, 28, 27, 26], to 1D models like HSPF [**?** 37], and high-resolution 3D models like EFDC [31, 38, 42]. While most studies focus on assimilation of a single constituent like temperature [31, 34, 38], sediments [43, 44], chlorophyll [32**?** , 40], or oxygen demand [36, 41, 28]; several recent studies have applied DA to multiple interdependent constituents such as algal biomass, phosphate, nitrate, dissolved oxygen, temperature, and other quantities relevant to HAB formation [45, 42, 37, 27].

While recent advances have shown promise in improving forecasts of HABs, challenges remain in applying these results to operational decision-making at the watershed scale. Existing models are generally either too computationally-intensive to apply for large river basins in real-time (EFDC, CAEDYM), or are too simplified to adequately capture the hydraulic behavior of regulated river networks (QUAL2K, HSPF), meaning that there is currently no suitable model to inform reservoir operations at the river basin scale. Moreover, a lack of built-in support for data assimilation means that most existing approaches largely rely on ensemble-based DA techniques like EnKF or PF, which use model outputs produced under iterated model runs to generate sample prob-

ability distributions of the underlying state variables [25]. While easy to implement, obtaining accurate results from these methods requires a large number of model runs, which is often more computationally intensive than direct approaches like EKF. Moreover, due to the lack of built-in DA support, existing studies mostly focus on assimilating a single constituent, with very few studies implementing assimilation of multiple reactive species. Given that algal growth is dependent on interactions between dozens of distinct reactive species, assimilation of a single constituent is unlikely to produce physically consistent estimates among water quality constituents, leading to degraded forecasting skill. In summary, there is a need for an online modeling system that combines a full-physics contaminant fate and transport model with built-in multi-constituent data assimilation functionality to better enable adaptive management of nutrients and algal blooms at the watershed scale.

To address these challenges, we introduce EufoRiA (Eutrophication model for River System Analysis), a new model built specifically for watershed-scale simulation and management of HABs that integrates unsteady hydraulics, multi-constituent contaminant transport and reaction kinetics, and online multi-constituent DA capabilities via an Extended Kalman Filtering (EKF) approach. The model is provided as open-source software, enabling extensions and integrations with existing models and interfaces. This study presents the detailed model development process and also presents comprehensive validation results in a large real-world river network—the Nakdong river basin in South Korea. Our key contributions are as follows:

- We introduce a new open-source physics-based model for online modeling of nutrient and HABs dynamics in large river networks that incorporates a physically-based hydraulic model and coupled reaction kinetics between 23 water quality constituents relevant to eutrophication and HAB formation.
- Validating the new model against long-term observed water quality data from a large, regulated river network, we find that EufoRiA achieves accuracy comparable to existing state-of-the-art models in predicting concentrations of water quality constituents.
- We derive and implement an efficient method for multi-constituent DA based on Extended Kalman Filtering (EKF) to assimilate water quality observations into the model. We show how the proposed DA technique

can be calibrated to integrate prior information on spatial and inter-constituent correlations via the Kalman Filter's process noise model.

- Through a holdout assessment, we show that our DA scheme significantly enhances estimation of water quality constituents within a large river network using real-world water quality observations. We find that DA performance is further improved through integration of prior information into the Kalman Filter's process noise model.

3. Methods

3.1. Development of new multi-constituent water quality model for large river networks

The EufoRiA model is a comprehensive physics-based nutrient and HABs model for river networks released as an open-source modeling toolkit in the Python programming language. This section describes the model development methods and procedure, including (i) model development history and description, (ii) governing equations and solution methods, (iii) biochemical reaction dynamics, (iv) simulation procedure at the software implementation level, (v) input data preparation and conversion, (vi) adaptive time-stepping method, and (vii) model application and validation.

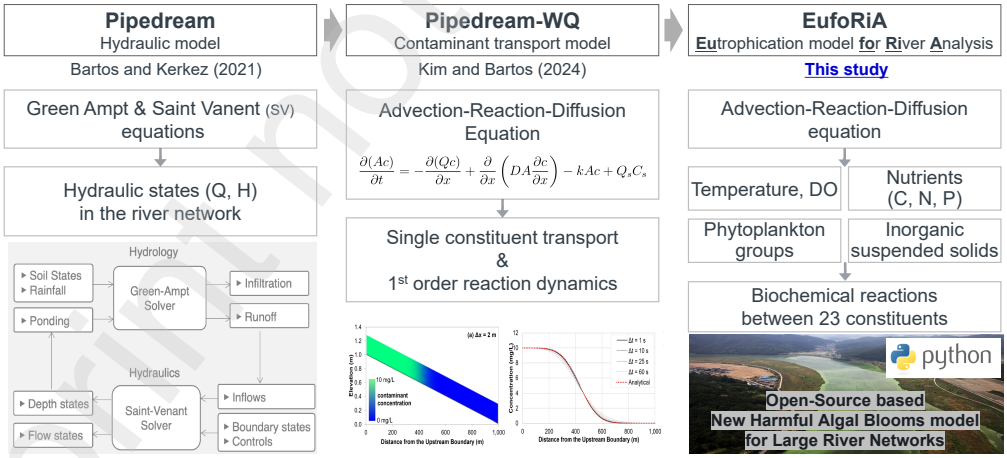


Figure 1: Brief model development history for EufoRiA

3.1.1. Model development history and brief description

The EufoRiA model simulates essential biochemical dynamics related to eutrophication in river network systems such as carbon, nitrogen, phosphorus, heat, dissolved oxygen, phytoplankton growth, and suspended sediments. In previous work, Kim and Bartos [46] introduced Pipedream-WQ, a new water quality model for natural and urban drainage networks based on the unsteady advection-reaction-diffusion (ARD) equation that incorporates real-time data assimilation using a Kalman Filtering scheme. Designed for real-time applications in networked drainage systems, this model features a novel implicit solver scheme based on the SUPERLINK algorithm [47], ensuring numerical stability and scalability to large networks. However, this earlier Pipedream-WQ model focused mainly on contaminant transport, and is thus limited to modeling a single water quality constituent with simple first-order reaction kinetics. In this study, we enhance the model's capabilities and develop an advanced version—EufoRiA—to represent water quality dynamics across 23 constituents related to eutrophication, enabling more effective predictions of nutrient fate and harmful algal blooms in river networks. The present study also evaluates this new nutrient model using long-term multi-constituent validation data from a large real-world river network. The distinct features of EufoRiA are summarized as follows:

- EufoRiA simulates nutrient dynamics in complex river networks by integrating an unsteady hydraulics model based on the Saint-Venant equations, unsteady transport using the advection-diffusion equation, and complex biochemical reactions for 23 water quality constituents into a unified modeling framework.
- Based on these capabilities, the EufoRiA model is suitable for simulating river networks under the influence of hydraulic infrastructure like dams and weirs, where contaminant fate and transport are affected by both unsteady hydraulics and backwater effects induced by impoundments.
- The EufoRiA model simultaneously represents both open channels and pressurized conduits, enabling simulation of urban stormwater systems in addition to natural river networks affected by built infrastructure.
- As an open-source model, EufoRiA enables extensions and integrations with other software, allowing for integration with data acquisition systems, visualization tools, and other process models.

3.1.2. *Governing equations and solution methods*

EufoRiA uses two sets of governing equations, corresponding to hydraulics and contaminant transport, respectively. For hydraulics, the physically-based Saint-Venant equations are used, consisting of the continuity equation (Equation 1) and momentum equation (Equation 2):

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = Q_s \tag{1}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}(Qu) + gA\left(\frac{\partial h}{\partial x} - S_0 + S_f + S_L\right) = 0 \tag{2}$$

Where A is the cross-sectional area of the river channel (m^2), Q is the flow rate at the given channel (m^3/s), u is the flow velocity (m/s), g is the gravitational force (m/s^2), h is the water depth (m), and S_0, S_f, and S_L represent the channel bottom slope, friction slope, and local head loss slope, respectively. For contaminant transport, the unsteady advection-reaction-diffusion equation is used (Equation 3) [48]:

$$\frac{\partial (Ac)}{\partial t} = -\frac{\partial (Qc)}{\partial x} + \frac{\partial}{\partial x}\left(DA\frac{\partial c}{\partial x}\right) - kAc + Q_s C_s \tag{3}$$

Where, c is the contaminant concentration (g/m^3), D is the diffusion coefficient (m^2/s), k is the first-order reaction coefficient ($1/s$), $Q_s c_s$ is the mass flow rate of contaminant into or out of the system boundary.

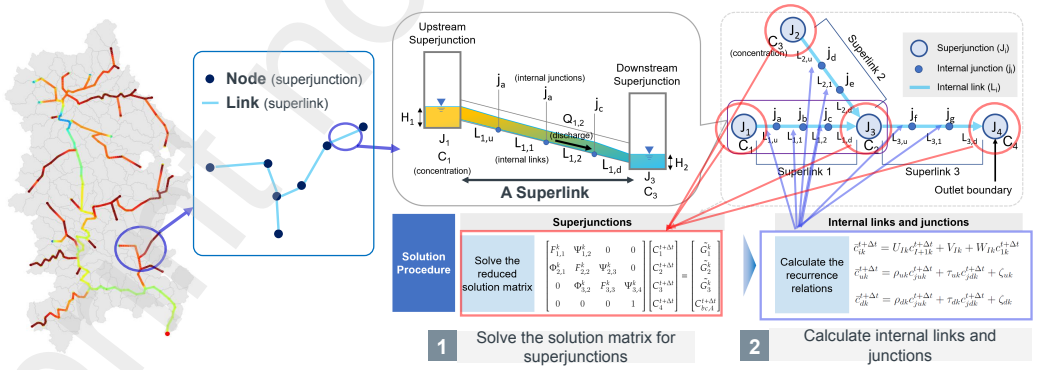


Figure 2: A schematic image of the Superlink algorithm applied for numerical solution of Saint-Venant and advection-reaction-diffusionequations

A staggered-grid implicit finite-difference scheme (SUPERLINK) is ap-

plied to obtain numerical solutions to these governing equations. Details of the numerical solutions for these equations are presented in previous research [49, 46]. Figure 2 illustrates the concept of the SUPERLINK algorithm. In this algorithm, computational elements are divided into four types—superjunctions, superlinks, junctions, and links. Superjunctions represent storage volumes (e.g. reservoirs or confluence points between tributaries) and may also include boundary-forcing inputs. Superlinks consist of chains of links and junctions connected in series and are used to represent river reaches and conduits. Within the SUPERLINK method, a sparse matrix equation is first solved to obtain the solutions to the three governing equations at superjunctions (in terms of hydraulic heads and contaminant concentrations at these elements); next, recurrence relations are used to calculate the solutions to governing equations at the internal links and junctions.

3.1.3. Biochemical reaction dynamics

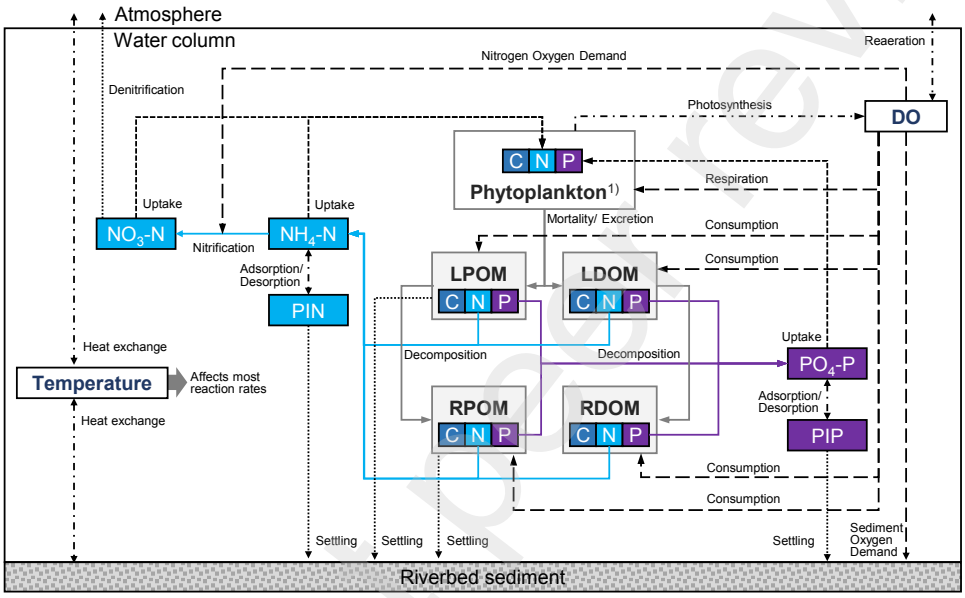
EufoRiA provides essential transport and biochemical dynamics models to simulate harmful algal blooms (HABs) in river networks, including key constituents such as water temperature, dissolved oxygen (DO), inorganic and organic nutrients (C, N, P), phytoplankton groups, and suspended sediments. Figure 3 shows a schematic diagram of the water quality reaction dynamics included in the EufoRiA model. In general, the dynamics for each constituent are described by coupled ordinary differential equations that are functions of both constituent concentrations, exogenous forcings (e.g. light and temperature), and species-specific parameters that inform reaction rates. For example, nitrate (NO_3) dynamics are described by the equation:

$$\frac{\partial C_{NO3}}{\partial t} = \underbrace{-\sum_{i=1}^{n}(1-P_{NH4})K_{ag,i}\delta_{n:alg,i}C_{alg,i}}_{\text{Algal production}} + \underbrace{K_{NH4}\gamma_{NH4}C_{NH4}}_{\text{Nitrification}} - \underbrace{K_{NO3}\gamma_{NO3}C_{NO3}}_{\text{Denitrification}} - \underbrace{\frac{K_{s,NO3}}{H}C_{NO3}}_{\text{Settling}} \tag{4}$$

Where C_{NO3} is the concentration of nitrate; $C_{alg,i}$ is the concentration of algal species i; C_{NH4} is the concentration of ammonium; $K_{ag,i}$, K_{NH4}, K_{NO3} and $K_{s,NO3}$ are growth/decay coefficients; γ_{NH4} and γ_{NO3} are temperature multiplier coefficients; P_{NH4} is the ammonium preference factor; $\delta_{n:alg,i}$ is

the stoichiometric parameter between C_{NO3} and $C_{alg,i}$; and H is the average hydraulic depth.

Sections S1–S10 in the Supplementary Information (SI) document provide a full list of all water quality constituents, a review and comparison of the biochemical dynamics in the existing models and the EufoRiA model, and the detailed reaction dynamics equations for all constituents.



1) **Phytoplakton**: Represented by biomasses of three phytoplakton groups = Diatom + Green algae + Cyanobacteria
2) **TN** (Total Nitrogen) = NO_3-N + NH_4-N + PIN + LPON + LDON + RPON + RDON + Algae-N
3) **TP** (Total Phosphorus) = PO_4-P + PIP + LPOP + LDOP + RPOP + RDOP+ Algae-P
4) **TOC** (Total Organic Carbon) = LPOC + LDOC + RPOC + RDOC+ Algae-C
5) **LPOM** (Labile Particulate Organic Matter) = LPO**C** + LPO**N** + LPO**P** : Three constituents are individually simulated in the model
6) **LDOM** (Labile Dissolved Organic Matter) = LDO**C** + LDO**N** + LDO**P**
7) **RPOM** (Refractory Particulate Organic Matter) = RPO**C** + RPO**N** + RPO**P**
8) **RDOM** (Refractory Dissolved Organic Matter) = RDO**C** + RDO**N** + RDO**P**

Figure 3: Schematic diagram of the water quality reaction kinetics in EufoRiA

Briefly describing the biochemical dynamics in EufoRiA, the *Temperature model* is based on the full heat balance equations between the atmosphere, water body, and riverbed (SI S4). The *DO model* includes depletion by organic and inorganic constituents, re-aeration from the atmosphere, and photosynthesis by phytoplankton (SI S5). The *Nutrients model* accounts for both inorganic and organic constituents of carbon, nitrogen, and phosphorus compounds. For inorganic constituents, the model simulates phosphate (PO_4-P), ammonium (NH_4-N), nitrate (NO_3-N), dissolved inorganic nitrogen (DIN), and phosphorus (DIP). Additionally, the organic nutrients (C, N,

P) are simulated as four different classes: labile dissolved, labile particulate, refractory dissolved, and refractory particulate (SI S6, S7, S8). The *Phytoplankton model* simulates three algal groups (i.e., Diatoms, Cyanobacteria, and Green algae) based on growth, respiration, excretion, mortality, and settling. The algal growth rate is calculated using nutrient-limiting factors, employing the Monod equation with temperature and light availability (SI S9). The *Suspended sediment model* is based on the transport and settling processes of inorganic suspended sediment (SI S10).

Table 1 compares the considered water quality constituents between existing models and EufoRiA. Compared to QUAL2K, an alternative river network-scale model for simulating nutrient dynamics, EufoRiA captures the major biochemical dynamics related to nutrient cycling and HAB formation while also accounting for unsteady hydraulics and contaminant transport. It should be noted that biological constituents not directly related to nutrient and HAB dynamics (e.g. Epiphyton, Macrophytes, Zooplankton, Bacteria, Fish) are omitted from the current implementation of EufoRiA. Some inorganic and chemical factors that may indirectly affect eutrophication dynamics (e.g. pH, alkalinity, riverbed sediment kinetics, silica, and metals) are also currently not included in the model, but are reserved as subjects for future work.

Oxygen demands used in traditional water quality analysis such as BOD (Biochemical Oxygen Demand), COD (Chemical Oxygen Demand), and CBOD (Carbonaceous Biochemical Oxygen Demand) are not explicitly modeled in EufoRiA for several reasons. First, these demands do not correspond to physical constituents with true masses and so the modeling of these constituents often relies on empirical formulations that are difficult to generalize or verify. In addition, the analytical techniques for BOD and COD do not ensure the thorough quantification of the existing organic compounds in the sample due to interfering factors [50]. In this regard, EufoRiA simulates only the TOC as a sum of the organic carbon constituents—LDOC, LPOC, RDOC, and RPOC—to represent the existence and reactions of organic carbons. Other DO demands like nitrification and bottom sediment in river channels are separately modeled using related equations (see SI Equation S.14). This approach is justified by the fact that some jurisdictions have recently moved away from oxygen demands and towards physical constituents as primary water quality indicators: for instance, in South Korea, the representative water quality criteria to evaluate the water quality grade of reservoirs and lakes was changed from COD to TOC after 2016 [51].

Table 1: Comparison of the considered water quality constituents between widely used existing models and EufoRiA

State variables	EFDC (3-D)	CAEDYM (3-D)	CE-QUAL-W2 (2-D)	QUAL2K (1-D)	EufoRiA (1-D)
Hydrodynamics and biochemical capability					
Unsteady hydraulics	✓	✓	✓	-	✓
Unsteady biochemical reactions	✓	✓	✓	-	✓
Basic water quality constituents					
Water temperature	✓	✓	✓	✓	✓
pH, alkalinity	-	✓	✓	✓	-
Dissolved oxygen and various oxygen demands					
DO	✓	✓	✓	✓	✓
BOD	-	✓	-	-	-
CBOD	-	-	✓	✓	-
COD	✓	-	-	-	-
Nutrients: Carbon, Nitrogen, and Phosphorus					
DIC (dissolved inorganic carbon)	-	✓	✓	✓	✓
LDOC[a]	✓	✓	✓	✓	✓
LPOC[a]	✓	✓	✓	✓	✓
RDOC[a]	✓	✓	✓	✓	✓
RPOC[a]	✓	✓	✓	✓	✓
Ammonium (NH4)	✓	✓	✓	✓	✓
Nitrate (NO3)	✓	✓	✓	✓	✓
LDON[a]	✓	✓	✓	✓	✓
LPON[a]	✓	✓	✓	✓	✓
RDON[a]	✓	✓	✓	✓	✓
RPON[a]	✓	✓	✓	✓	✓
Phosphate (PO4)	✓	✓	✓	✓	✓
LDOP[a]	✓	✓	✓	✓	✓
LPOP[a]	✓	✓	✓	✓	✓
RDOP[a]	✓	✓	✓	✓	✓
RPOP[a]	✓	✓	✓	✓	✓
Ecological constituents					
Algae (Phytoplankton)	✓	✓	✓	✓	✓
Epiphyton	✓	✓	✓	✓	-
Macrophyte and Zooplankton	-	✓	✓	-	-
Bacteria and Fish	-	✓	-	-	-
Suspended solids and others					
ISS (Inorganic suspended solids)	✓	✓	✓	✓	✓
Riverbed sediment kinetics	✓	✓	✓	✓	-
Silica	✓	-	-	✓	-
Metals	✓	-	-	✓	-

[a] LDO: Labile dissolved organic, LPO: Labile particulate organic, RDO: Refractory dissolved organic, RPO: Refractory particulate organic, for the last letter, C: Carbon, N: Nitrogen, P: Phosphorus

3.1.4. Simulation procedure at the software implementation level

The model structure of EufoRiA can be divided into three modules: hydraulics, constituent transport, and biochemical reactions between constituents. For hydraulics, the model uses the full dynamic unsteady Saint-Venant equation with an implicit upwind scheme for momentum transfer. The water quality model applies the unsteady advection-diffusion-reaction equation, also using an implicit upwind scheme for the advection component. For biochemical reactions, to facilitate the linkage between the transport and biochemical reaction modules, we use the 4th-order Runge-Kutta method, ensuring model accuracy with relatively large time steps while maintaining an explicit scheme. Like all explicit methods, the Runge-Kutta method may suffer from instability at large temporal step sizes, and thus careful selection of the time step is necessary. To address this issue, we propose an adaptive time stepping approach as described in SI S13.

Calculations for these three modules are performed sequentially at each time step, as described in Figure 4. First, the model simulates the hydraulics, including the flow rates and water depths across all elements in the river network. Based on the results of the hydraulic model, the model calculates the constituent transport using the advection-diffusion equation. Finally, the model simulates the biochemical reactions at each element. Once these steps are completed, the model proceeds to the next time step. To increase computational speed, the *numba* just-in-time compiler is used to accelerate numeric computations for the hydraulic, contaminant transport and biochemical reaction modules.

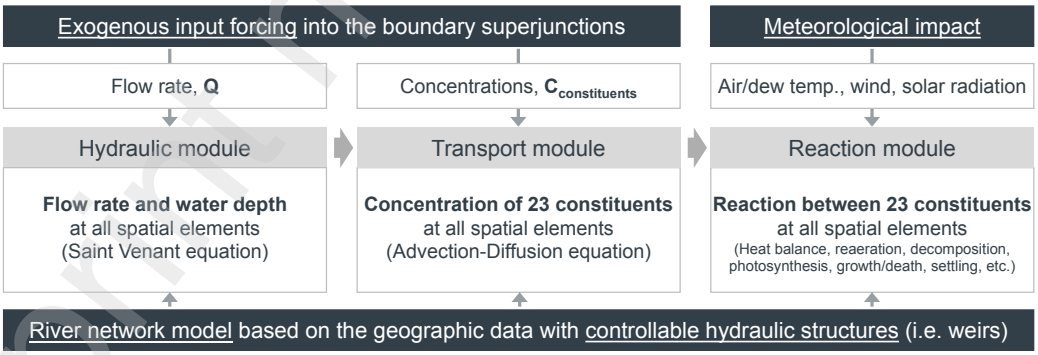


Figure 4: Brief description of the model simulation process at each time step

3.1.5. Model application and validation: A case study

As a case study, we apply EufoRiA to a complex real-world river network—the Nakdong River Basin in South Korea—to assess its effectiveness as a physics-based model for nutrients and HABs. This section provides details on the study area, data collection, river network model construction, and the calibration and validation processes.

Study area: Nakdong River basin in South Korea

The study area is the Nakdong River Basin, the second-largest watershed in South Korea, with a drainage area of 23,817 km^2 and a mainstem river length of 525 km [52]. The river basin is highly regulated, with 10 multi-purpose dams, 2 water supply dams, 8 operational large weirs along the mainstem, and 1 estuary bank. The river basin also contends with significant point and non-point pollution sources, including 21 large wastewater treatment plants (WWTPs) with capacities exceeding 20,000 m^3/day, several industrial complexes, and a population of approximately 13 million. Due to high pollutant loads, this river system experiences periodic HABs during the summer season. As is widely recognized, regulated river systems significantly impact algal bloom dynamics, typically in adverse ways [53]. The eight large weirs constructed along the mainstream between 2009 and 2011 have further exacerbated HAB severity and duration by increasing water retention time and reducing flow velocity. Moreover, as the Nakdong River serves as a drinking water source for 13 million people, ensuring water quality safety remains a critical concern. To address these challenges, water managers in South Korea have been working toward a comprehensive water quality and HABs prediction and forecasting system [54, 55]. In this context, we apply EufoRiA to this river basin to assess its skill in operational nutrient and HAB modeling. We focus on the river mainstem and its major tributaries, where the most validation data are available, and exclude the headwaters located upstream of the 10 multi-purpose dams (Figure 5).

Geographic and temporal dataset collection

Data required for model construction include geographical data, water quality measurements, flow rates, and meteorological data. Table 2 provides details on the datasets used in this research, including data sources and brief descriptions. In this study, we collect various geological datasets, including a 30×30 m digital elevation model (DEM), shapefiles of river channel networks

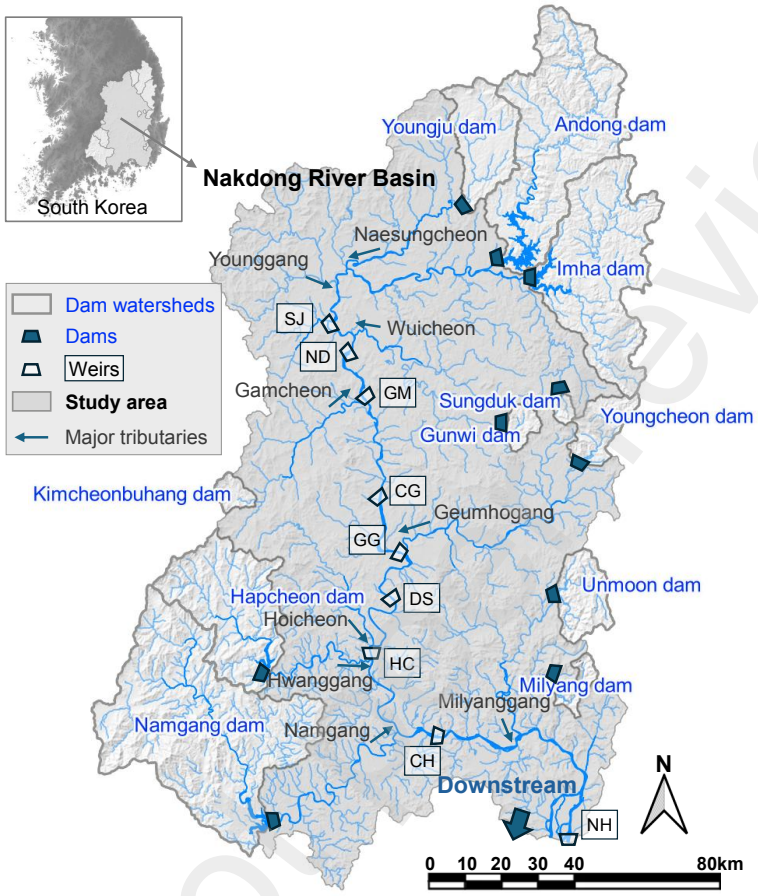


Figure 5: Study area: Nakdong River basin of South Korea (Where, SJ: Sang-Ju weir, ND: Nak-Dan weir, GM: Gu-Mi weir, CG: Chil-Gok weir, GG: Gangjeong-Goryung weir, DS: Dal-Sung weir, HC: Hapcheon-Changnyung weir, CH: Changnyung-Haman weir, NH: Nakdonggang-Hagoo weir)

and centerlines, and the official government-approved hydrologic unit catchment boundaries (141 catchments). In addition, streamflow data from major tributaries (8 stations) are obtained from the Water Management Information System (WAMIS) of South Korea, and representative meteorological data (4 stations) in the Nakdong River basin were collected from the Korea Meteorological Agency (KMA) website, including precipitation, air and dew point temperatures, wind velocity, and solar radiation.

Furthermore, river water quality data from 95 stations, covering 14 constituents such as temperature, DO, TN, TP, Chl-a, and TOC, are obtained

Table 2: Description of the spatial and temporal datasets used for the model application in the Nakdong River basin, South Korea

Data	Resolution	Description	Source
Digital elevation model	$30 \times 30m$	DEM of Nakdong River basin in Korea (TIFF)	National geographic information institute of Korea https://www.ngii.go.kr/
Catchment boundary	141 catchments	Official Hydrologic Unit Catchment of Korea (shape)	Water management information system of Korea http://www.wamis.go.kr
River channel	-	Channel boundaries of the mainstream and tributaries (shape)	Water management information system of Korea http://www.wamis.go.kr
Meteorological data	4 stations/hourly	Precipitation, air and dew point temperatures, wind velocity, and solar radiation data (time series from 2013 to 2022)	Korea Meteorological Administration https://data.kma.go.kr
Stream Flow	8 stations/daily	Observed flow rates from tributaries (time series from 2013 to 2022)	Water Management Information System of Korea http://www.wamis.go.kr
River water quality	95 stations/weekly, monthly	Observed water quality data from manual sampling stations (time series from 2013 to 2022)	National Institute of Environmental Research of Korea http://water.nier.go.kr
Effluent flow rate and water quality from WWTPs	21 stations/daily	Observed discharge and water quality data from the facilities (time series from 2013 to 2022)	National Institute of Environmental Research of Korea http://www.nier.go.kr
Discharging flow rate and water level of dams and weirs	12 dams and 8 weirs /daily	Observed discharge and water level(EL.m) data of the facilities (time series from 2013 to 2022)	Korea Water Resources Corporation http://www.water.or.kr

from the National Institute of Environmental Research (NIER) of Korea. Effluent water quality and flow rate data from 23 wastewater treatment plants (WWTPs) are also collected from NIER. Operational data for multipurpose dams and weirs, including discharge and water levels, are retrieved from the MyWater website of Korea Water Resources Corporation (K-water). All time-series datasets span 10 years, from 2013 to 2022. As mentioned in Section 3.1.5, the construction of large weirs has altered the environmental conditions of river channels, making it unreliable to apply water quality data from before 2012 for model calibration and validation. Figure 6 shows the locations of the measurement stations for different datasets.

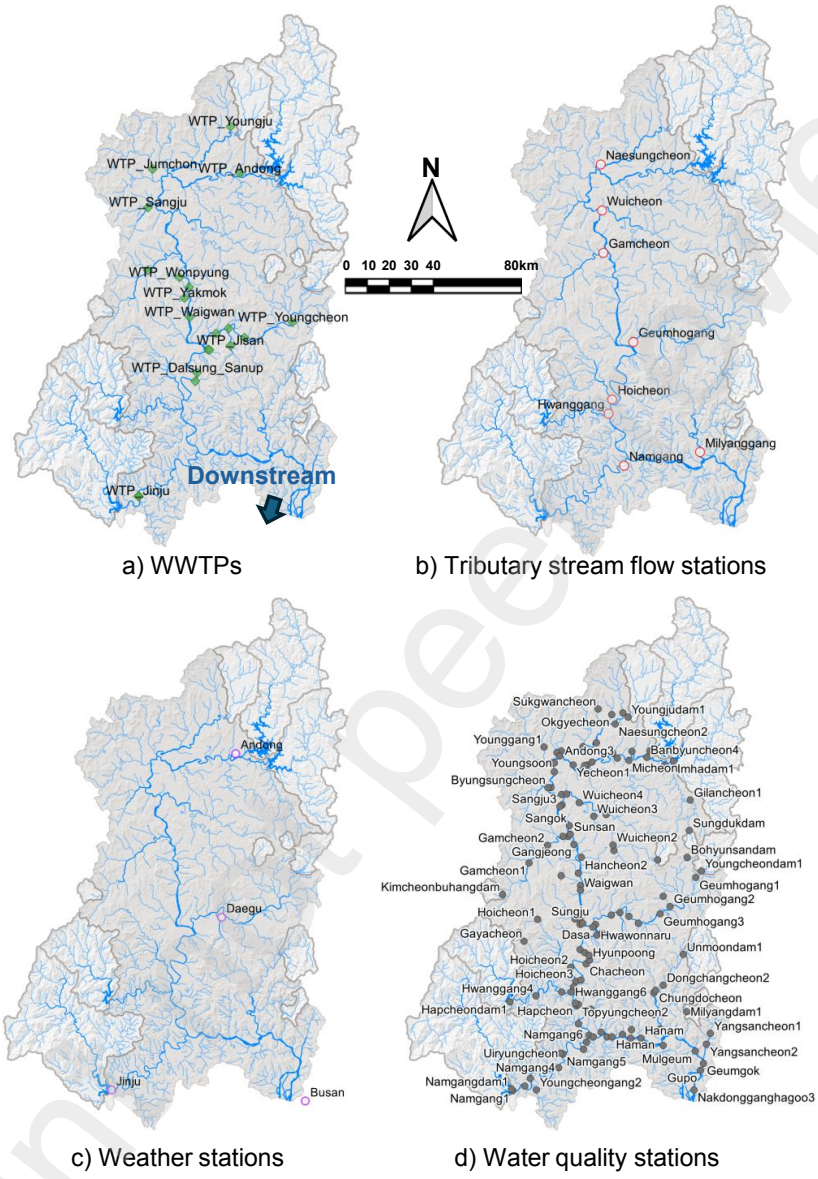


Figure 6: Locations of dataset: Nakdong River Basin in South Korea

Construction of the river network model

We construct a EufoRiA model for the Nakdong river basin using the prepared geographic data and previously constructed river channel section data, such as HEC-RAS geological data from K-water and other Korean agencies.

In this process, we assume that all river channels are rectangular for simplicity, deriving the representative cross-sectional shape by considering the average width and depth of each river channel and section. Figure 7 shows the developed river network for this study.

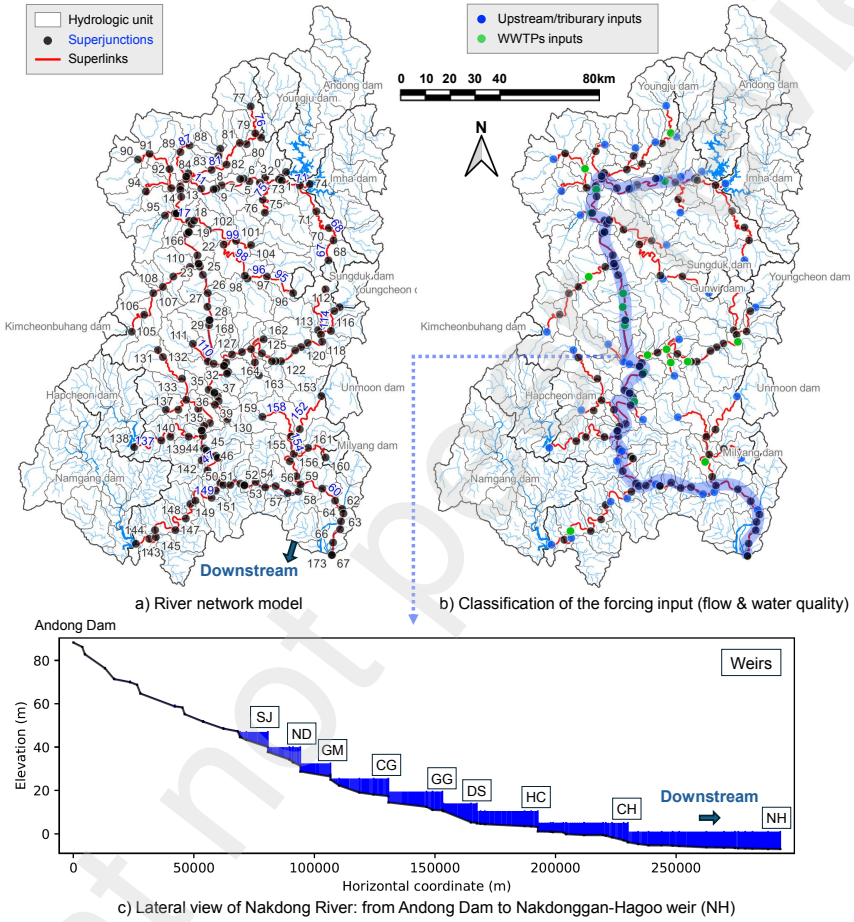


Figure 7: River network and classification of the forcing input (flow and water quality) categories

Within EufoRiA, the ultimate river network model consists of 164 superlinks (representing river reaches and reservoirs) and 174 superjunctions (representing confluences, control points, and boundary conditions). The superjunctions include 41 upstream input boundaries, including 10 multipurpose dam discharges and 18 WWTP input boundaries for 23 WWTPs facilities. Additionally, the superlinks contain 3,032 internal links and 3,196 internal

junctions for total 1,209 km of river network channels. We also represent eight large controllable weirs in the mainstream of the Nakdong River, which have a total storage capacity of 533.7 million m^3 at normal operation levels. For weir control, the control strategy described in Section S11 is applied such that the weir outflows produced by the model match the observed weir outflows over the period of record.

Calibration and validation

To assess the performance of EufoRiA, we calibrate the model to observed data over a calibration period (2013–2020) and then assess its performance against observed data over a validation period (2021–2022). The model's performance during these periods is evaluated based on the metrics described in Section 3.3 at eight weir locations along the Nakdong River mainstem, represented by white rectangular boxes (i.e., SJ, ND, GM, CG, GG, DS, HC, and CH) in Figure 5. Because EufoRiA is a one-dimensional model, we compare the model results with depth-averaged observed data at each weir location. We analyze model performance in terms of six key water quality constituents: water temperature, dissolved oxygen (DO), total nitrogen (TN), total phosphorus (TP), chlorophyll-a (Chl-a), and total organic carbon (TOC). Among these constituents, water temperature and DO are directly simulated in the model, while TN, TP, Chl-a, and TOC are derived from the modeled labile, refractory, particulate, and dissolved portions of each constituent based on the relationships described in Table 3.

Model calibration is achieved via a trial-and-error empirical calibration approach using visual comparisons and performance metrics, such as RMSE, NSE, KGE, and PBIAS. This model calibration approach is chosen due to the inherent challenge in applying optimization methods to calibrate a multi-constituent model containing over 100 parameters that govern water quality reaction kinetics. Because calibrating all parameters simultaneously is impractical, we prioritize the most influential parameters for more effective calibration, such as the nitrification/denitrification coefficients of nitrogen compounds, the maximum growth and mortality rates of algal groups, the settling rate for particulate constituents, and temperature multiplier-related coefficients for most constituents. The key parameter values associated with major nutrients and HABs applied in this study are presented in SI Tables S3–S9.

3.2. Data assimilation

We propose a novel data assimilation (DA) strategy based on Extended Kalman Filtering (EKF) to assimilate water quality sensor data into the process model while also incorporating prior knowledge about the system including observed correlations between water quality constituents. This section describes the development and evaluation of the DA procedure for EufoRiA, including (i) EKF scheme for the multi-constituent water quality model, (ii) observation matrix design, (iii) process noise covariance matrix design, and (iv) development of hold-out assessment scenarios to evaluate the performance of the proposed DA method. The distinct features of the EufoRiA DA scheme are summarized as follows:

- EufoRiA provides built-in data assimilation through an efficient Extended Kalman Filtering scheme that simultaneously assimilates water quality observations for up to 23 water quality constituents.
- We propose an observation model that effectively captures the biochemical relationships between common measurable water quality constituents (e.g. chlorophyll-a) and modeled constituents (e.g. algal biomass).
- We introduce a method for constructing the Extended Kalman Filter's process noise covariance matrix that integrates prior knowledge of spatial and inter-constituent correlations, with the ultimate effect of increasing DA accuracy.

3.2.1. Kalman Filtering scheme

In Kim and Bartos [46], a data assimilation scheme based on Kalman filtering [56] was developed to assimilate water quality observations into a single-constituent contaminant transport model based on the advection-reaction-diffusion equation. The present study extends this functionality to enable multi-constituent data assimilation within the new EufoRiA nutrient and HABs model. From the implicit numerical solution of the water quality transport equation (i.e. advection-reaction-diffusion equation), we may express the evolution of contaminant concentrations over time in terms of a state equation (Equation 5), along with the observed states in terms of an observation equation (Equation 6):

$$\mathbf{x}_{t+\Delta t} = A_t \mathbf{x}_t + B_t \mathbf{u}_{t+\Delta t} + \mathbf{y}_t + \mathbf{w}_t \tag{5}$$

$$\mathbf{z}_t = H_t \mathbf{x}_t + \mathbf{v}_t \tag{6}$$

Where, $\mathbf{x}_t$ is the n-dimensional state vector, $\mathbf{w}_t$ is a p-dimensional stochastic disturbance defined as zero-mean white noise, $\mathbf{v}_t$ is an m-dimensional vector of zero-mean white measurement noise, A_t is an $(n \times n)$ state transition matrix, B_t is an $(n \times \ell)$ input transition matrix, $\mathbf{u}_t$ is an $(n \times 1)$ vector of input forcing, $\mathbf{y}_t$ is an $(n \times 1)$ vector of constants, and H_t is the $(m \times n)$ observation matrix that expresses the observation function that maps the measurement data to the state variables in the model. Here, n represents number of states, p represents the number of disturbance inputs, and m represents the number of measurement points.

Based on the above state and observation equations, the Kalman recursion provides an algorithm to assimilate sensor observations into the model to yield the minimum mean-squared error estimator of system states [46, 56].

$$\hat{\mathbf{x}}_{t+\Delta t} = A_t \hat{\mathbf{x}}_t + B_t \mathbf{u}_{t+\Delta t} + \mathbf{y}_t + L_{t+\Delta t}[\mathbf{z}_{t+\Delta t} - \hat{\mathbf{z}}_t] \tag{7}$$

$$\hat{\mathbf{z}}_t = H_{t+\Delta t}(A_t \hat{\mathbf{x}}_t + B_t \mathbf{u}_{t+\Delta t} + \mathbf{y}_t) \tag{8}$$

$$L_t = P_t H_t^T (H_t P_t H_t^T + V_t)^{-1} \tag{9}$$

$$P_{t+\Delta t} = A_t (P_t - P_t H_t^T (H_t P_t H_t^T + V_t)^{-1} H_t P_t) A_t^T + W_t \tag{10}$$

Where, $\hat{\mathbf{x}}_t$ is the estimate of the state variable, $\hat{\mathbf{z}}_t$ is the estimate of the model output, L_t is the optimal Kalman gain, V_t is the measurement noise covariance matrix, P_t is the estimation error covariance matrix, W_t is the process noise covariance matrix. The Kalman gain L_t and the error covariance P_t are calculated at every time step and updated to correct the estimate of the system states $\hat{\mathbf{x}}_t$.

For the multi-constituent system, the matrices and vectors in Equations 7–10 may be augmented to represent all r constituents simultaneously. In this augmented form, the state transition matrix A_t will have a dimension of ((nr rows) $\times$ (nr columns)); the state vector $\mathbf{x}_t$ will have a dimension of ((nr rows) $\times$ (1 column)); and the observation matrix H_t will have a dimension of ((mr rows) $\times$ (nr columns)).

3.2.2. Observation matrix design

In the context of Kalman Filtering, the observation matrix represents a (linear) function that translates measured quantities (e.g. chlorophyll-a) into modeled quantities (e.g. algal concentrations). This observation model enables data assimilation for all modeling constituents using a smaller number

of observed constituents. In our model, the observation matrix is used to update all 23 modeled constituents based on 10 observed quantities that are commonly measured in surface water systems, including temperature, DO, TOC, TN, NH_4, NO_3, TP, PO_4, chlorophyll-a, and suspended sediments. Figure 8 shows the structure of the proposed observation matrix in this study that represents the following relationships among the water quality constituents in the model.

Table 3: Derived constituents in EufoRiA

Derived constituent		Simulated constituents in EufoRiA
Total organic carbon :	TOC =	LDOC + LPOC + RDOC + RPOC + c1×ALG1 + c2×ALG2 + c3×ALG3
Total nitrogen :	TN =	PIN + NH_4-N + NO_3-N + LDON + LPON + RDON + RPON + n1×ALG1 + n2×ALG2 + n3×ALG3
Total phosphorus :	TP =	PIP + PO_4-P + LDOP + LPOP + RDOP + RPOP + p1×ALG1 + p2×ALG2 + p3×ALG3
Chlorophyll-a :	Chl-a =	chl1×ALG1 + chl2×ALG1 + chl3×ALG3

* c1-c3, n1-n3, p1-p3, and chl 1-chl 3 are explained in Figure 8. ALG1-ALG3 are biomass of diatom, green algae, and cyanobacteria, repectively.

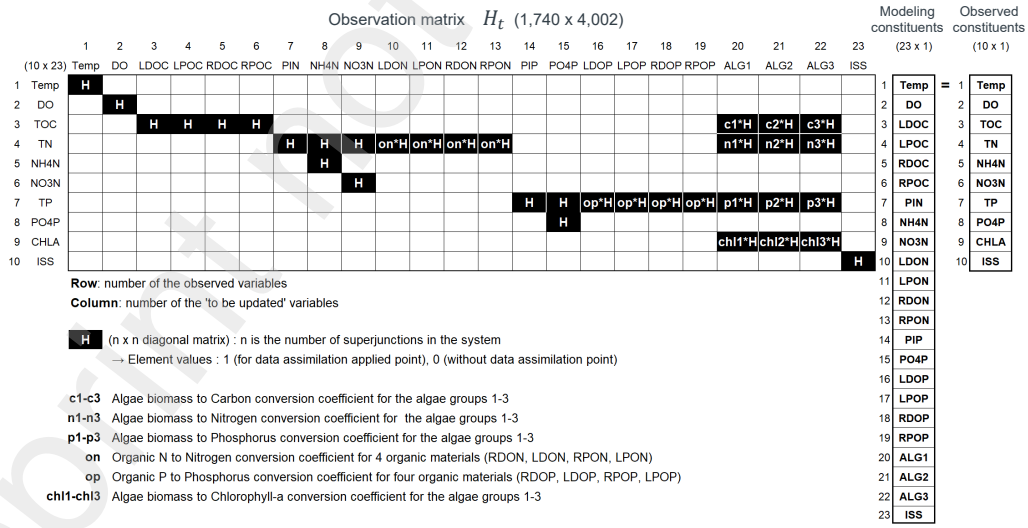


Figure 8: Building the observation matrix for the Kalman filtering

3.2.3. Process noise covariance matrix design

The process noise covariance matrix reflects the uncertainty in the process model, and thus specifying this matrix properly is crucial for obtaining accurate state estimation results within the Kalman Filter. In Kim and Bartos [46], it was shown that the process noise covariance matrix has a significant effect on the results of the Kalman filtering procedure, both in terms of the spatial distribution of the 'correction' applied by the Kalman filter, and on the accuracy of the ultimate state estimate. In the present study, we introduce and evaluate a new approach to calibrate the process noise covariance matrix based on observed correlations from the historical record. Specifically, we analyze and integrate the historical observed spatial and inter-species correlations among water quality constituents to construct the process noise covariance matrix, and then assess the improvement in performance afforded by incorporating this correlation structure into the Kalman filter.

The proposed method for specifying the process noise covariance accounts for both the spatial correlation in nutrient runoff inputs and the inter-constituent correlations between chemically-similar constituents. The spatial correlation component reflects the fact that nutrient loads are highly correlated in space [57] (e.g. fertilizer runoff from agricultural fields is spatially contiguous). Likewise, the interconstituent correlation reflects the fact that contaminants often exhibit significant correlations with one another due to similarities in chemical composition. For example, concentrations of nitrogen constituents such as total nitrogen, nitrate, and organic nitrogen groups are frequently correlated. To incorporate observed correlations into the DA framework, we first use historically observed water quality constituent concentrations from 2013–2020 (calibration period) to determine both reach-wise and constituent-wise covariances for each constituent. These covariances are then arranged into covariance matrices describing both the spatial and inter-constituent correlations. Additional details on this procedure are outlined in SI Sections S14 and S15. Based on the spatial and inter-constituent covariance matrices derived in the previous steps, a large block matrix representing the process noise covariance matrix for the entire system is constructed, as shown in Figure 9. In this figure, $\sum_{spatial\text{-}Ci}$ is the spatial correlation matrix between computational elements for the ith constituent, and $\sum_{const\text{-}CiCj}$ is the inter-constituent correlation matrix between the ith and jth water quality constituents at each spatial point.

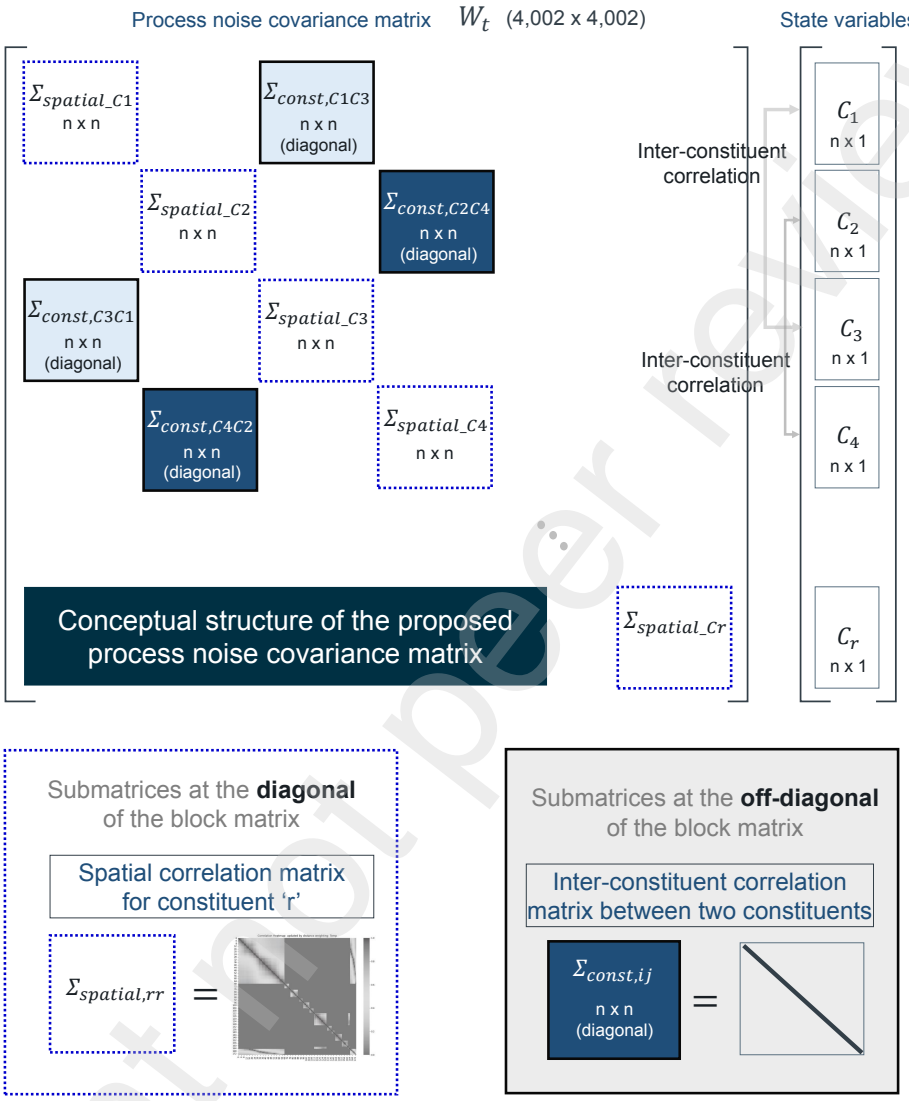


Figure 9: Building the system-wide process noise covariance matrix using the spatial and inter-constituent correlation matrices. Where, n: number of the spatial computational elements in the model = 174 in this study, r: number of modeling constituents = 23 from C_1 to C_{23}

3.2.4. Spatial holdout assessment

We evaluate the improvement in model skill attributable to the proposed DA scheme using a holdout assessment. Gage sites are first divided into

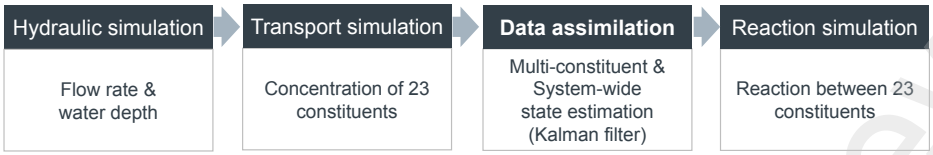


Figure 10: Data assimilation procedure in EufoRiA

assimilation points (where DA is applied) and holdout points (which are used only for evaluation). Sensor data are then assimilated into the model at the selected subset of assimilation points and the estimated contaminant concentrations are compared to measured contaminant concentrations at all sites. This procedure is repeated for a successively increasing number of assimilation points to determine how increasing the number of assimilated gages improves model accuracy. This holdout analysis is repeated using both the default (uncorrelated) process noise covariance matrix and the calibrated (correlated) process noise covariance matrix to assess how the inclusion of prior knowledge of correlations improves DA performance.

A total of 66 gage locations—37 in the mainstem of the Nakdong River and 29 in the tributaries—are used for DA holdout assessment. We apply a succession of 8 different holdout cases, ranging from 8–66 assimilation points (corresponding to 12–100% coverage), as shown in Figure 11. Case 0 represents the base scenario without DA application (model-only) and is not included in Figure 11. In case 1, DA is applied to 8 locations representing the weirs in the river mainstem. For cases 2 through 6, the number of assimilation points successively increases from 15 points to 56 points. Finally, in Case 7, all observed data from the 66 locations in the river network are utilized for DA. Modeling skill is assessed in terms of a suite of performance metrics as described in Section 3.3 (i.e., NSE, KGE, RMSE, and PBIAS). For each holdout trial, we compute the mean of each performance metric over all 66 gage locations.

3.3. Model performance evaluation metrics

We evaluate the performance of the multi-constituent water quality model in terms of the Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), Root Mean Squared Error (RMSE), and Percent Bias (PBIAS) performance metrics. NSE yields values between negative infinity and 1, and when it is greater than zero, the model can be considered a reasonable predictor (i.e. better than using the observed mean). The Nash-Sutcliffe Efficiency (NSE)

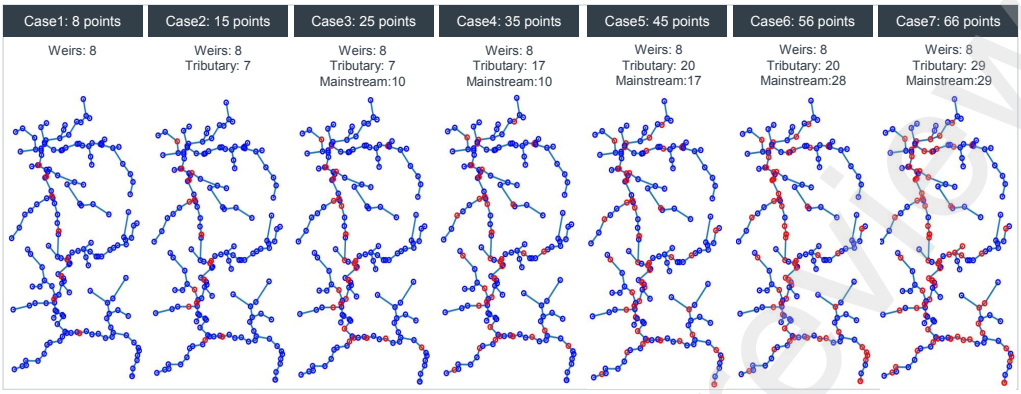


Figure 11: Scenarios for the spatial holdout assessment

is calculated using the following equation:

$$NSE = 1 - \frac{\sum_{i=1}^{N}(Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^{N}(Q_{obs,i} - \overline{Q_{obs}})^2} \tag{11}$$

Where, $Q_{obs,i}$ is i^{th} observed value, $Q_{sim,i}$ is i^{th} simulated value, $\overline{Q}_{obs}$ is a mean of the observed values, and N is the number of observed data

The Kling-Gupta Efficiency (KGE) is calculated using the following equation [58]:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \tag{12}$$

Where, r is the Pearson correlation coefficient between Q_{obs} and Q_{sim}, α is the ratio of the standard deviation of simulated values to the standard deviation of observed values $\left(\alpha = \frac{\sigma_{sim}}{\sigma_{obs}}\right)$, β is the ratio of the mean of simulated values to the mean of observed values $\left(\beta = \frac{\mu_{sim}}{\mu_{obs}}\right)$, σ_{sim} is the standard deviation of the simulated values, σ_{obs} is the standard deviation of the observed values, μ_{sim} is the mean of the simulated values, and μ_{obs} is the mean of the observed values.

The Root Mean Squared Error (RMSE) and the percentage bias (PBIAS) are computed as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(Q_{obs,i} - Q_{sim,i})^2}{N}} \tag{13}$$

$$PBIAS = \frac{\sum_{i=1}^{N}(Q_{obs,i} - Q_{sim,i})}{\sum_{i=1}^{N} Q_{obs,i}} \times 100 \quad (14)$$

When the RMSE of the model results is close to zero, the model can be considered highly accurate. Additionally, a PBIAS of zero indicates an unbiased model, while a positive PBIAS suggests that the model tends to underestimate the observed values, and a negative PBIAS indicates that the model tends to overestimate them.

4. Results

4.1. Model calibration and validation

To evaluate the model performance, simulation results without DA are first compared against established criteria for NSE, KGE, and PBIAS in watershed-scale water quality models. Ranges corresponding to satisfactory and unsatisfactory model performance for different constituents are shown in Table 4 [59]. More specifically, we apply the criteria for Nitrogen (Daily) and Phosphorus (Monthly) for NSE and KGE, and the criteria for Nutrients (Monthly) for PBIAS.

Table 4: Performance evaluation criteria for watershed-scale water quality models [59]

Measure	Component	Temporal scale	Very Good	Good	Satisfactory	Not satisfactory
NSE (KGE)	Nitrogen, Phosphorus	Monthly	$0.65 <$ NSE	$0.50 \leq$ NSE ≤ 0.65	$0.35 <$ NSE < 0.50	NSE ≤ 0.35
|PBIAS|(%)	Nitrogen, Phosphorus	Daily & Monthly	|PBIAS|(%) < 15	$15 <$ |PBIAS|(%) < 20	$20 \leq$ |PBIAS|(%) ≤ 30	$30 <$ |PBIAS|(%)

Figures 12 and 13 show the model calibration and validation results at the Changnyung-Haman (CH) weir, the most downstream weir in the Nakdong River. We select this location because the model output at this point reflects the accumulated uncertainties from numerous boundary input forcings as well as the complexity of the model itself. For reference, we provide results from other weirs in the Supplementary Information (SI) S18. According to the results, percent biases for all evaluated constituents are mostly within the range of $\pm 10\%$ with the exception of the calibration result for Chl-a, indicating that there are no significant biases in the model results. The validation results for water temperature, DO, TN, and TP show high accuracy with NSE of 0.402–0.986, KGE of 0.619–0.964, while the validation results of Chl-a and TOC show relatively lower accuracy with NSE of 0.272–0.305,

KGE of 0.376–0.513 (Table 5). Model performance across all constituents is comparable to results obtained from existing models in other studies (see SI Section S17 for a detailed comparison).

4.1.1. Water temperature

As seen in Figure 12, the temperature simulation results (black line) show excellent agreement with observed data (blue circle). Reproducing the spatial and temporal temperature pattern is crucial for ensuring the reliability of the river network water quality model because water temperature affects the reaction rates of most biochemical processes. According to the results, NSE values are 0.965 and 0.986, KGE are 0.926 and 0.964, RMSE are 1.604 and 1.011, and percent biases are +3.4% and -0.9% for the calibration and validation periods, respectively. According to the criteria in Table 4, these results are classified as 'very good' (Table 5). It should be noted that a longer simulation period during calibration causes higher variability and fluctuations, resulting in slightly lower performance across all metrics.

Water temperature in EufoRiA is affected by meteorological drivers, including air temperature, dew point temperature, wind velocity, and solar radiation, as well as hydraulic factors such as water flow rate and depth. Given the complexity of the river network and input data, as well as the significant depth and volume of the mainstream river channel (as shown in Figure 7) that strongly affect the river's heat balance, we conclude that EufoRiA effectively represents key heat balance processes, ensuring reliable water temperature modeling in impounded river networks like the Nakdong River.

4.1.2. Dissolved oxygen (DO)

After calibration, the model results for DO show good agreement with the observed data, with NSE values of 0.483 and 0.402, KGE values of 0.757 and 0.761, RMSE values of 1.690 and 1.640, and percent biases of +9.2% and +9.6% for the calibration and validation periods, respectively (Figures 12 and 13). These results can be assessed as 'satisfactory' to 'very good' performance in terms of NSE and KGE, and 'very good' performance in terms of percent bias by applying the criteria for 'Phosphorus' in Table 4.

Dissolved oxygen (DO) plays a pivotal role in maintaining the natural self-purification capacity of river ecosystems and acts as an electron acceptor in the decomposition of organic matter and oxidation processes, such as in the nitrification of ammonium (NH_4^+) to nitrate (NO_3^-). While reaeration at the water surface is the key mechanism for recovering DO concentration

in the water column, photosynthesis by phytoplankton also plays a critical role in nutrient-rich river systems. Large weirs in the river also limit oxygen exchange in the water column due to the increased water depth, resulting in hypoxia at the bottom of the river channel. Although these factors may affect the model results, in this study, we assume that the river channel is well-mixed vertically and horizontally because EufoRiA is a one-dimensional model. Even under this restriction, the model results show that EufoRiA appropriately represents the DO dynamics and seasonal patterns in the complicated river network of the Nakdong River.

4.1.3. Chlorophyll-a (Chl-a)

Most river water quality models such as EFDC, CAEDYM, QUAL2K, and CE-QUAL-W2 offer highly complicated algal growth models that incorporate various nutrient concentrations (e.g., nitrate, ammonium, and phosphate) and environmental drivers such as solar radiation, water temperature, and water depth. As a result, modeling HABs in the large-scale river network with high accuracy is typically extremely challenging work.

Considering these limitations and complexity, EufoRiA offers quite reasonable Chl-a modeling results (Figures 12 and 13), with NSE values of -0.001 and 0.305, KGE of 0.137 and 0.376, RMSE of 15.973 and 12.442, and percent biases of +13.1% and +7.2%, which are evaluated as 'not satisfactory' to 'satisfactory' in NSE and KGE, 'very good' in the percent bias. These results are comparable to those achieved in other studies (see SI Section S17).

In EufoRiA, Chl-a is calculated based on the three algal groups—diatom, green algae, and cyanobacteria. Even though these algal groups have different growth rates, optimal temperatures, and sensitivity to nutrient concentrations, we calibrate the model using only the observed value of Chl-a, resulting in limited accuracy. To improve model results, more detailed observation for algal groups should be required in the future research.

4.1.4. Total nitrogen (TN)

The calibrated model results for TN show good agreement with the observed data (Figures 12 and 13), with NSE values of 0.757 and 0.820, KGE values of 0.773 and 0.753, RMSE values of 0.383 and 0.292, and percent biases of +5.4% and +0.7% for the calibration and validation periods, respectively. These results can be classified as 'very good' performance (Table 5).

Accurate estimation of nutrient sources like nitrogen and phosphorus is crucial for reproducing the overall HAB dynamics at different spatial and

temporal scales. In EufoRiA, total nitrogen concentration is calculated as the sum of nitrate, ammonium, particulate inorganic nitrogen, four organic nitrogen groups, and internal nitrogen in three algal groups, rather than as an individual constituent. After nitrogen compounds enter the system (i.e., river network), the only pathways for total nitrogen removal in the model are settling and denitrification (SI S7). Excluding these two processes, total nitrogen in the system can be considered a conservative substance, while the relative composition among the above constituents continuously changes during the modeling process. In other words, the remaining total nitrogen in the system from the exogenous input loadings can be controlled by calibrating the settling and denitrification rates. In addition, total nitrogen concentration is higher during the low flow season than high flow season because of the impact of the nitrate in the base flow from groundwater discharge. We can confirm that EufoRiA model results reproduce this temporal pattern with reasonable accuracy.

4.1.5. Total phosphorus (TP)

The simulation results of TP show good agreement with observed data, with NSE of 0.485 and 0.398, KGE of 0.515 and 0.619, RMSE of 0.018 and 0.015, and percent biases of -1.3% and -7.7% for calibration and validation period, respectively (Figure 12 and 13). Applying the criteria in Table (Table 5), the results are evaluated as 'satisfactory' to 'good' in NSE and KGE, 'very good' in percent bias. Considering the model's complexity, uncertainties in input forcings and model parameters, we confirm that this level of performance metrics is reliable for the predicting the HABs in Nakdong River.

Phosphorus—especially phosphate as a reactive and soluble phosphorus—is the major limiting factor for algal growth in the Nakdong River due to the high background nitrogen concentration and the implementation of advanced phosphorus removal technologies in the wastewater treatment process. During the low flow period, phosphorus from point sources controls the total amount of available phosphorus for algal growth. However, during the high flow season, phosphorus originating from non-point sources in watersheds provides additional nutrients, resulting in summer HABs after rainfall events. The model output in Figure 12 and 13 show that EufoRiA reliably reproduces this seasonal pattern with satisfactory accuracy.

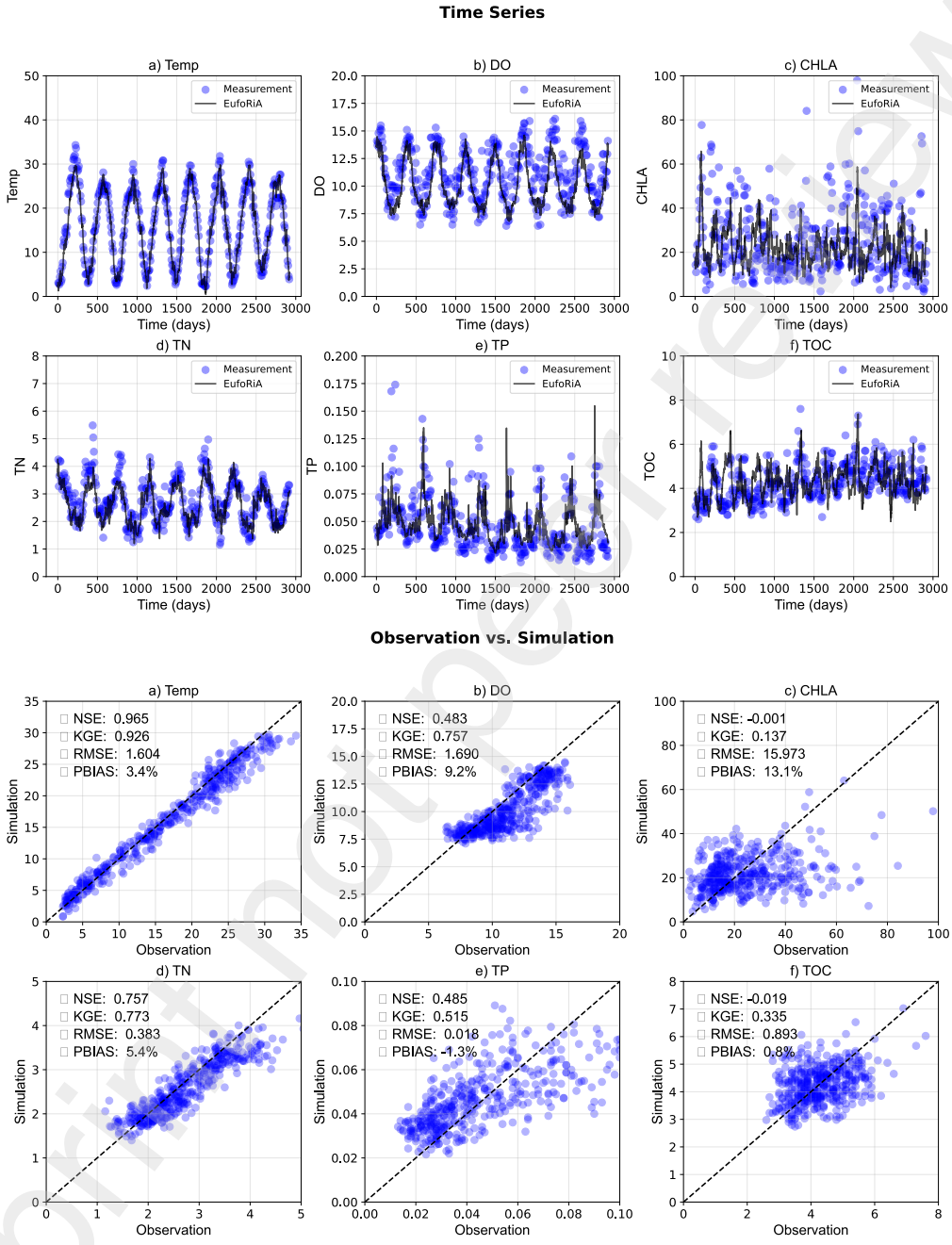


Figure 12: Calibration (2013-2020) results at Changnyung Haman (CH) weir

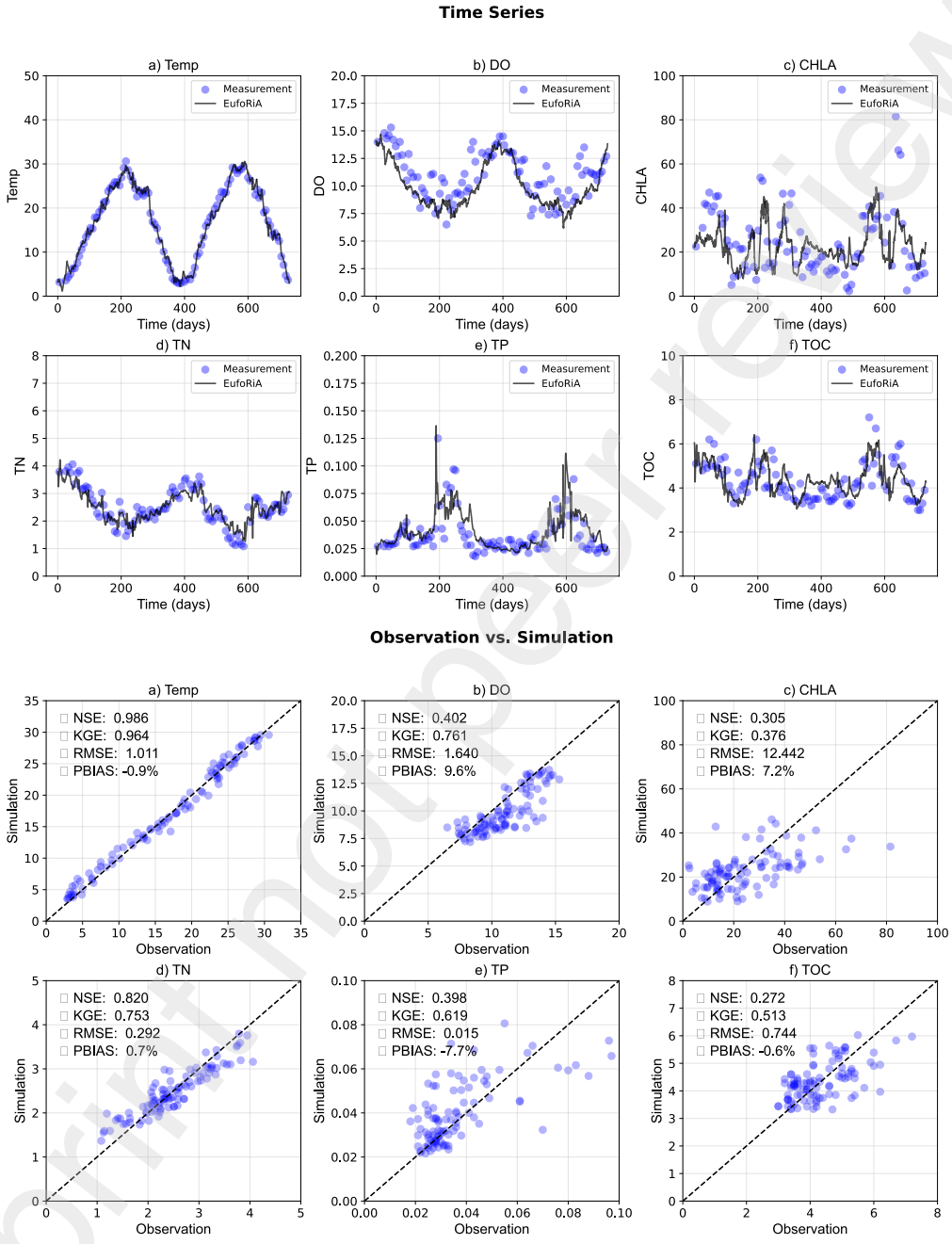


Figure 13: Validation results (2021-2022) at Changnyung Haman (CH) weir

4.1.6. Total organic carbon (TOC)

TOC model outputs show relatively less accurate results compared to the temperature, DO, and nutrients modeling results, with NSE of -0.019 and 0.272, KGE of 0.335 and 0.513, RMSE of 0.893 and 0.744, and percent biases of -0.8% and -0.6% for calibration and validation periods, respectively. According to Table 5, these results are classified as 'not satisfactory' to 'good' for NSE and KGE, 'very good' for the percent bias. As noted previously, these results are comparable to the performance of existing models (SI Section S17).

EufoRiA does not explicitly simulate the conventional oxygen demands such as BOD, CBOD, and COD because these quantities represent surrogate indicators to evaluate the oxygen demand reactions in the water, and do not exist as real physical quantities. Although many studies have sought to enhance modeling of these quantities, existing work has struggled to represent these quantities with satisfactory accuracy. Therefore, in this study, we simulate only four different organic carbon groups (i.e., LDOC, LPOC, RDOC, and RPOC) instead of simulating BOD, CBOD, or COD in the model. TOC model results also tend to correlate with Chl-a concentrations (i.e., algal biomass) given that the organic carbons are produced by photosynthesis during algal growth process. Therefore, to enhance the accuracy of TOC model results, the algal growth model accuracy must first be improved.

Table 5: Model calibration and validation results at the CH weir in the Nakdong-River: model performance metrics–NSE, KGE, RMSE, and PBIAS – for representative water quality constituents (We apply the criteria of daily nitrogen case in Table 4, and N.S. is Not Satisfactory.)

	NSE		KGE		RMSE		PBIAS	
	Calibration	Validation	Calibration	Validation	Calibration	Validation	Calibration	Validation
Temperature	0.965	0.986	0.926	0.964	1.604	1.011	+3.4%	-0.9%
Assessment result	Very good	Very good	Very good	Very good	-	-	Very good	Very good
Dissolved Oxygen	0.483	0.402	0.757	0.761	1.690	1.640	+9.2%	+9.6%
Assessment result	Satisfactory	Satisfactory	Very good	Very good	-	-	Very good	Very good
Chloropyll-a	-0.001	0.305	0.137	0.376	15.973	12.442	+13.1%	+7.2%
Assessment result	N.S.	N.S.	N.S.	Satisfactory	-	-	Very good	Very good
Total Nitrogen	0.757	0.820	0.773	0.753	0.383	0.292	+5.4%	+0.7%
Assessment result	Very good	Very good	Very good	Very good	-	-	Very good	Very good
Total Phosphorus	0.485	0.398	0.515	0.619	0.018	0.015	-1.3%	-7.7%
Assessment result	Satisfactory	Satisfactory	Good	Good	-	-	Very good	Very good
Total Organic Carbon	-0.019	0.272	0.335	0.513	0.893	0.744	-0.8%	-0.6%
Assessment result	N.S.	N.S.	N.S.	Good	-	-	Very good	Very good

4.2. Improvement in performance attributable to Data Assimilation

While EufoRiA provides results that are comparable to existing water quality models under the model-only simulation case (Section 4.1), data assimilation (DA) greatly improves the prediction performance by integrating observational data at in-stream locations into the numerical model. Analyzing the improvement in KGE, NSE, PBIAS, and RMSE attributable to DA for different sensor coverage densities, DA is found to significantly improve estimation performance in terms of all metrics considered, even under relatively sparse sensor coverage. Estimation performance is found to further improve when prior information on spatial and inter-constituent correlations is incorporated into the process noise covariance matrix. Sensors located along the river mainstem are found to be most effective at improving estimation performance relative to sensors located at smaller tributaries.

4.2.1. Model only vs. Data Assimilation (DA) applied simulation results

Based on the spatial holdout assessment described in Section 3.2.4, DA significantly improves model performance for all constituents considered, and a greater number of data assimilation locations leads to better estimation performance. Figure 14 shows results for six representative constituents in terms of improvement in KGE. Here, KGE is selected as a summary benchmark due to its ability to simultaneously measure correlation, bias, and variance of errors. In both the spatially correlated and uncorrelated cases, the mean KGEs of all observed points are consistently enhanced as the number of DA points increases from Case 1 (8 points, or 12% sensor coverage) to Case 7 (66 points, or 100% sensor coverage). Notably, even under Case 1—the scenario with the smallest number of DA points—we observe significant KGE improvements for all considered constituents, especially for total organic carbon and Chlorophyll-a. Under the model-only case, these two constituents showed the poorest performance, with Chlorophyll-a obtaining an mean KGE of 0.054 ('Not Satisfactory') over all gage locations, and TOC obtaining an mean KGE of 0.255 ('Not Satisfactory'). However, these results rapidly improve as data is assimilated. For Case 1 (12% sensor coverage), TOC achieves 'Satisfactory' performance, with a mean KGE of 0.389 for the uncorrelated case. By Case 3 (38% sensor coverage), TOC and Chlorophyll-a both achieve 'Good' performance, with mean KGEs of 0.492 and 0.419, respectively. Finally, by Case 6 (85% sensor coverage), TOC and Chlorophyll-a obtain 'Very Good' performance, with mean KGEs of 0.632 and 0.614, respectively. These

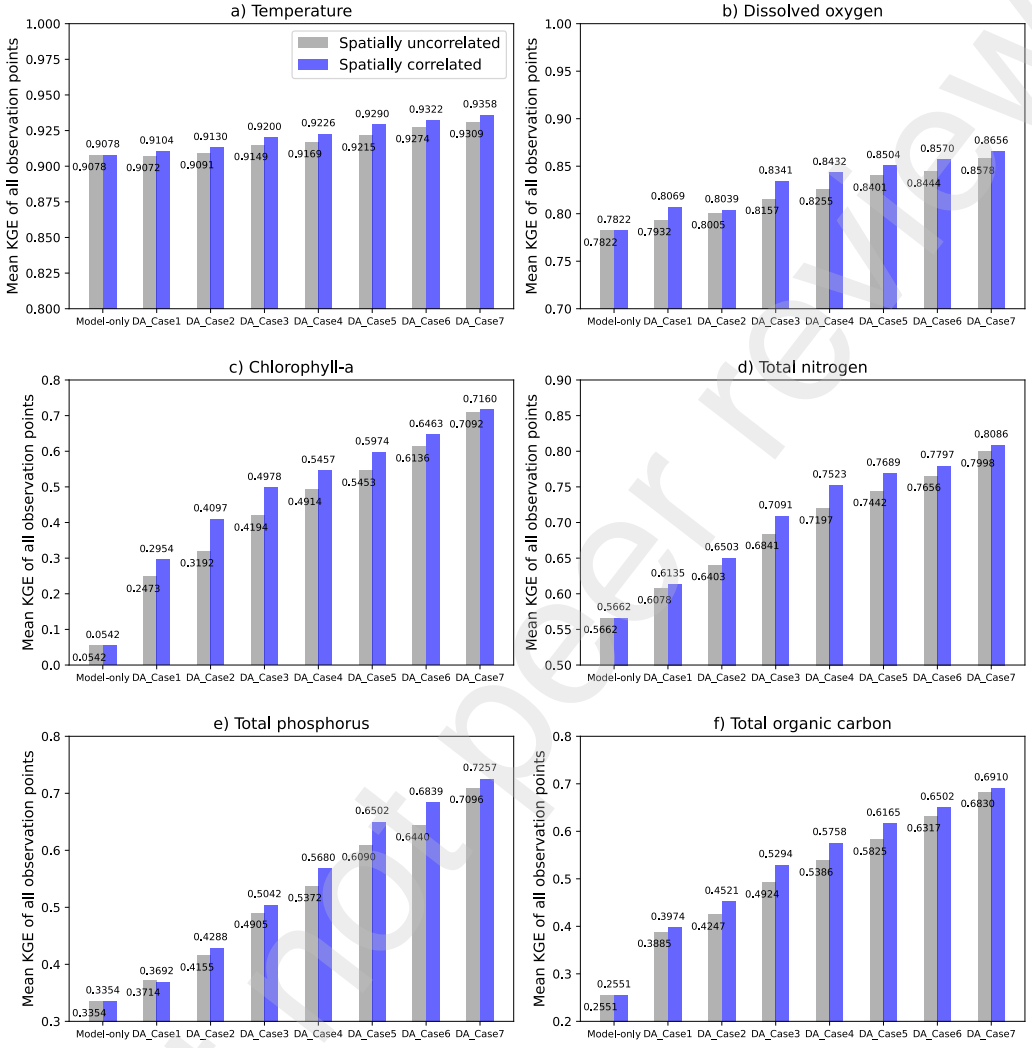


Figure 14: Result of the spatial holdout assessment based on the different numbers of data assimilation point scenarios

results indicate that DA enables accurate estimation of water quality constituents that are otherwise difficult to model even after extensive calibration. Moreover, even with a small number of DA locations, the mean state estimation results for the entire river network are significantly improved, especially when the DA locations are selected in the portions of the river mainstem that are impounded by weirs.

In addition to KGE, DA improves estimation accuracy with respect to all other performance metrics considered. Table 6 shows the mean NSE, KGE, RMSE, and PBIAS metrics for six water quality constituents at all gauge locations under the model-only case (Case 0) and DA-applied cases (Case 1, Case 4, and Case 7). We highlight here the results for Case 4 (representing 53% sensor coverage) and Case 7 (representing 100% sensor coverage). Under the model-only case, PBIAS ranges from 4.1–38% among constituents with a mean value of 15%. By Case 4, PBIAS is reduced to 3.8–25% with a mean of 10% (achieving all 'Good' and 'Very Good' scores, except for Chlorophyll-a which is 'Satisfactory'). Finally, by Case 7 PBIAS ranges from 3.1–16% with a mean of 7.3% (achieving all 'Very Good' scores except for Chlorophyll-a, which remains 'Satisfactory'). NSE scores are lower in general, but are still substantially improved by DA, with NSE improving from -1.05–0.92 (mean -0.01) under the model-only case, to -0.404–0.934 (mean 0.33) under Case 4, to 0.22–0.95 (mean 0.61) under Case 7 (achieving all 'Good' and 'Very Good' scores, except for TOC which remains 'Not Satisfactory'). In terms of the overall error, RMSE is reduced by 15–39% under Case 4, and by 28–58% under Case 7. In summary, DA improves modeling skill for all performance metrics considered, with the largest improvements again occurring for difficult-to-model constituents like TOC and Chlorophyll-a.

4.2.2. Impact of the process noise covariance matrix on DA performance

DA performance is further improved by incorporating prior knowledge on spatial and inter-constituent correlations into the Kalman Filter's process noise covariance matrix. Based on the scenarios described in Section 3.2.4, we compare the holdout results first assuming that process noise is uncorrelated (i.e. using a diagonal process noise covariance matrix), and second assuming that the process noise is correlated (i.e. using the process noise covariance matrix introduced in Section 3.2.3). Figure 14 shows that incorporating prior knowledge of spatial correlations consistently achieves better performance under all considered scenarios and constituents. Comparing the results with each spatially uncorrelated case, including spatial correlations in the process noise covariance matrix shows percent improvements in KGE of 0.003–0.007 for temperature, 0.003–0.018 for DO, 0.007–0.090 for Chlorophyll-a, 0.006–0.033 for total nitrogen, -0.002–0.041 for total phosphorus, and 0.009–0.037 for TOC under given DA scenarios. From these results, we confirm that the proposed calibrated process noise covariance matrix, which considers the correlations between the spatial locations based on the historical observed

Table 6: Comparison between the Model-only simulation results vs. DA applied (spatially uncorrelated) results: mean NSE, KGE, RMSE, and PBIAS for representative water quality constituents at all gauge locations during the validation period (2021-2022)

	Model-only	DA Case 1	DA Case 4	DA Case 7	Performance Improvement
NSE					(Max Absolute Change)
Temperature	0.918^{v}	0.920^{v}	0.934^{v}	0.950^{v}	+0.032
Dissolved Oxygen	0.620^{v}	0.644^{v}	0.733^{v}	0.810^{v}	+0.190
Chlorophyll-a	-0.447^{n}	-0.160^{n}	0.198^{n}	0.580^{v}	+1.027
Total Nitrogen	0.247^{n}	0.235^{n}	0.497^{g}	0.672^{v}	+0.425
Total Phosphorus	-0.368^{n}	-0.319^{n}	0.028^{n}	0.438^{g}	+0.806
Total Organic Carbon	-1.045^{n}	-0.822^{n}	-0.404^{n}	0.217^{n}	+1.262
KGE					(Max Absolute Change)
Temperature	0.908^{v}	0.907^{v}	0.917^{v}	0.931^{v}	+0.023
Dissolved Oxygen	0.782^{v}	0.793^{v}	0.826^{v}	0.858^{v}	+0.076
Chlorophyll-a	0.054^{n}	0.247^{n}	0.491^{g}	0.709^{v}	+0.655
Total Nitrogen	0.566^{v}	0.608^{v}	0.720^{v}	0.800^{v}	+0.234
Total Phosphorus	0.335^{s}	0.371^{s}	0.537^{g}	0.710^{v}	+0.375
Total Organic Carbon	0.255^{s}	0.389^{s}	0.539^{g}	0.683^{v}	+0.428
RMSE					(Max Percent Change)
Temperature	2.050	1.979	1.748	1.482	-27.7%
Dissolved Oxygen	1.280	1.180	1.045	0.871	-32.0%
Chlorophyll-a	12.26	10.04	7.483	5.110	-58.3%
Total Nitrogen	0.554	0.524	0.409	0.328	-40.8%
Total Phosphorus	0.022	0.020	0.015	0.011	-50.0%
Total Organic Carbon	0.990	0.869	0.701	0.538	-45.7%
PBIAS					(Max Percent Change)
Temperature	4.127^{v}	4.516^{v}	3.812^{v}	3.140^{v}	-23.9%
Dissolved Oxygen	5.743^{v}	4.956^{v}	4.388^{v}	3.585^{v}	-37.6%
Chlorophyll-a	38.34^{n}	31.99^{n}	25.18^{s}	15.92^{s}	-58.5%
Total Nitrogen	9.181^{g}	8.933^{v}	6.028^{v}	4.536^{v}	-50.6%
Total Phosphorus	16.73^{s}	16.62^{s}	11.47^{g}	8.378^{v}	-50.0%
Total Organic Carbon	15.07^{s}	14.42^{g}	11.65^{g}	8.173^{v}	-45.8%

v: very good; g: good; s: satisfactory; n: not satisfactory

water quality data, significantly enhances the Kalman Filter's overall state estimation performance.

4.2.3. Spatial analysis of DA performance improvements

We observe that the DA impacts tend to disappear more quickly for tributaries with small water volumes and short retention times. This result likely occurs because the water quality at a specific location along the river channel is immediately affected by advection both from the upstream reach and into the downstream reach. Collecting more detailed observational data at locations with large water volumes like reservoirs and river mainstems will thus generally result in greater temporal persistence of DA corrections and improved model skill. Figure 15 shows this tendency in the mainstem and

tributaries. While the mainstem river channel shows clearer performance enhancements between the model-only and DA applied cases, in the tributaries, the modifications of KGE are relatively small compared to the mainstem locations.

The spatial map of improvements in KGE for each constituent shown in Figure 15 clearly reflect the difference in DA effectiveness between main channel and tributary locations. More specifically, KGE enhancement in the mainstem and tributary locations are -0.04–0.121 and -0.007–0.056 for the temperature, 0.017–0.255 and 0.013–0.099 for DO, 0.018–1.209 and -0.005–1.347 for Chl-a, 0.066–0.699 and 0.014–0.523 for TN, 0.105–2.444 and 0.013–0.973 for TP, and 0.062–1.123 and -0.003–0.723 for TOC, respectively. Considering these results, it is a better strategy to choose locations for data assimilation in mainstem rivers or reservoirs within river networks that have substantial storage volumes, in order to enhance model performance under data-scarce conditions.

5. Discussion

5.1. Novel open-source-based river water quality model for online applications

For better management of water quality and HABs in river systems, water managers and researchers need effective online models with data assimilation capabilities. This study presents EufoRiA, which is to our knowledge the first comprehensive online water quality model for large river networks that combines full unsteady hydraulics, complex nutrient and HAB dynamics, and data assimilation capabilities. This framework enables online modeling and forecasting of HABs at the river network scale, with the ultimate goal of enabling water managers to proactively improve water quality and reduce HAB outbreaks via real-time control of hydraulic structures such as weirs, dams, and small-scale detention facilities [60].

Additionally, the open-source nature of our model facilitates active modification and extension, providing a general framework that may be tailored to specific real-world use cases. For instance, new reaction kinetics such as bottom nutrient exchange in the hyporheic zone or additional reaction terms for existing constituents in the model can be readily extended by adding new modules into the base model. Given its extensibility, we expect that EufoRiA provides a feasible and reliable software toolkit to build comprehensive online modeling systems for water quality management at the watershed scale.

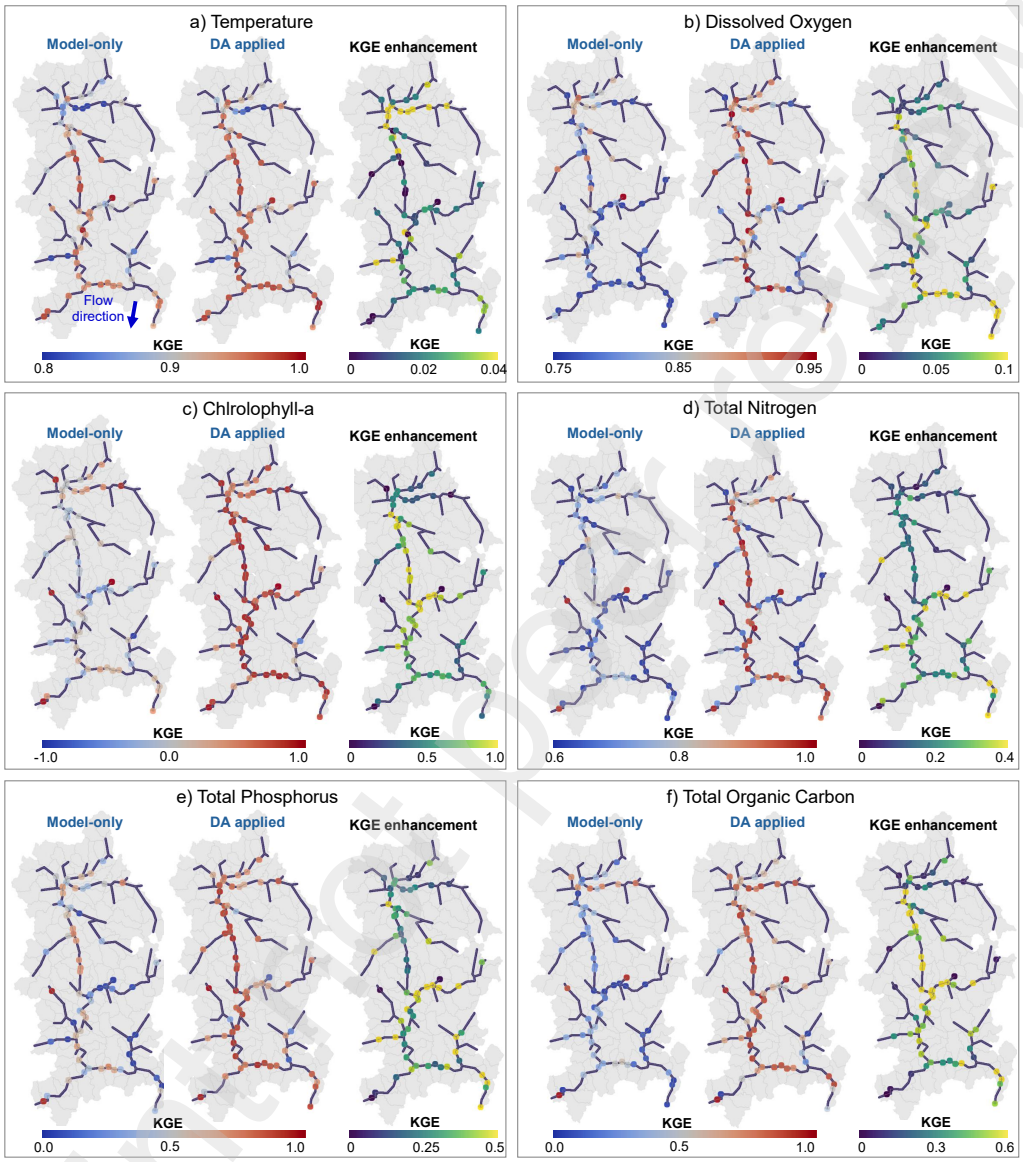


Figure 15: Change of the KGE at the all observation points based on the different numbers of data assimilation point scenarios: Physics-based model only (Base scenario) vs. 66 points data assimilation (Case 7 scenario), and KGE enhancement between two scenarios

5.2. Selection of the target modelling constituents and extensibility

The selection of water quality constituents in physics-based models depends on the specific objectives of water quality or ecological modeling. While many comprehensive models for rivers, lakes, and oceans incorporate highly complex water quality reaction kinetics to address various potential reactions in aquatic systems, highly complex models typically introduce greater uncertainty due to insufficient measurement data for input and verification, poorly understood biochemical processes, and the inherent heterogeneity of water systems [61]. Furthermore, increasing model complexity usually demands greater computational resources.

Therefore, a careful selection of model constituents is essential. For instance, the CAEDYM model represents one of the most sophisticated and advanced water quality and ecological models encompassing the eutrophication process. This model incorporates seven algal groups, five zooplankton groups, a jellyfish, five fish groups, a pathogen, a bacteria, four epiphyte groups, a seagrass, and three invertebrate groups, alongside fundamental nutrient kinetics [21]. However, given its complexity and the typical scarcity of measurement data, simultaneous modeling of these ecological processes without elaborately designed monitoring programs for specific purposes is typically infeasible in most practical implementations.

In this regard, EufoRiA omits relatively less influential constituents and interactions for HABs modelling. For instance, EufoRiA does not simulate constituents such as pH, alkalinity, conductivity, dissolved and particulate organic carbon, BOD, CBOD, COD, epiphyton, macrophyte, or zooplankton. In addition, while some researchers emphasize the importance of the hyporheic zone in nutrient cycling of river systems, EufoRiA does not yet include active interactions between river channels and bottom sediments. Including these additional components in the model would require further calibration and validation datasets, which are often unavailable in public periodic monitoring data. As discussed in Section 5.1, modifying the source code is always feasible and straightforward when calibration data becomes available. Therefore, we reserve these model extensions for future research.

5.3. Model performance and accuracy of EufoRiA: the necessity of DA methods

The new multi-constituent water quality model shows reasonable performance despite the fact that the chosen study area includes multiple natural and anthropogenic pollutant sources, complex interactions with hydraulic

infrastructure, and a large and heterogeneous river network model spanning nearly 24,000 km^2. While the simulation results for temperature, DO, nitrogen, and phosphorus show good agreement with the observed data, results for Chl-a and TOC are less accurate. The relatively low modeling accuracy for these quantities is caused by the inherent complexity in their reaction dynamics, which are subject to a large variety of environmental drivers such as meteorological conditions, nutrient levels and availability, and hydraulic conditions in the river channel. Researchers have applied various numerical models to assess and evaluate river water quality and algal bloom dynamics. However, accurate prediction of algal bloom dynamics is still highly challenging when using physics-based numerical models alone. Our results confirm the limitation of this model-only approach. To address this issue, we propose the integration of data assimilation methods to incorporate real-time sensor data into physics-based nutrient and HAB models. By incorporating real-time water quality observations into physics-based models, data assimilation enables more reliable prediction of highly uncertain constituents including Chl-a and TOC. Based on the results in Section 4.2.1, we conclude that model results after applying the DA method show much better performance compared to the model-only simulation results.

5.4. Integration of prior knowledge into DA process noise model

This study introduces a new method to improve DA performance by specifying the Kalman Filter's process noise model based on historical observations. Given that watershed-scale nutrient modeling is highly uncertain with a large number of degrees of freedom, it is crucial to ensure model accuracy based on all available knowledge about the target surface water system. The two major sources of uncertainty that are represented with the Kalman Filtering framework include measurement error (e.g. sensor noise) and modeling error (e.g. uncertainty in model parameters and exogenous inputs). While measurement error can be readily characterized from sensor manufacturer specifications, model uncertainty is much more difficult to quantify, given that model parameters and nutrient inputs to the river network are generally unknown. For this reason, the model uncertainty (process noise) is often taken to be uncorrelated by default. However, our findings show that incorporating observed correlations into the process noise covariance matrix significantly improves predictive performance of the model. The process noise covariance is thus a crucial parameter for obtaining accurate model results, and should be calibrated using all available data where possible.

5.5. Limitations and future research directions

Improvement of biochemical reaction kinetics and model structure

As discussed in the previous section, we ignore several biochemical and physical relationships such as nutrient exchange with underlying sediment layers, complicated sorption/desorption of nutrients, and additional constituents like zooplankton. Depending on the circumstances of the applied water body, these reactions may be critical factors for modeling of nutrients and harmful algal blooms. Model users must carefully consider the applicability of the EufoRiA model in light of the most relevant dynamics for their particular use case. In addition, this version of EufoRiA uses one-dimensional rectangular cross-sections for river channels. To represent more complex phenomena caused by thermal and density stratification in deep reservoirs, additional methods are needed to simulate stratification within computational elements. Furthermore, representation of more complicated channel cross-sectional geometries may provide more reliable model results.

Computational performance

The computational speed of EufoRiA is sufficient for real-time modeling of large river networks, but may pose a constraint for applications requiring large numbers of model runs such as model calibration, multi-scenario analysis, or real-time control. For the Nakdong river case study, the model requires roughly 1 hour for a 1-year simulation without data assimilation, and 5-6 times longer when DA is applied with hourly updates. For a 1-week forecast, this means the model requires about 1-2 minutes without DA, and 10-15 minutes with DA (using 13th Gen Intel Core i9-13900K CPU). This performance makes our model a practical choice for online modeling of large river basins. However, for applications like model predictive control, which require repeated calculations over an ensemble of scenarios, improvements to model speed may be required.

Application cases in real-world river networks

Because EufoRiA is a newly-developed model, there is currently a lack of case studies that demonstrate the applicability of the model across different climates, ecosystems, and nutrient loading scenarios. Additional case studies are needed to assess the model's performance under different use cases and accumulate results and insights from these experiences.

Coupled simulation with a rainfall-runoff model in watershed scale

Although EufoRiA offers highly sophisticated nutrient and HAB dynamics modeling within river networks, its capabilities remain restricted to the river channel. While point source pollution is represented through specific boundary inputs, most non-point source pollutants originate from the upstream watershed. Therefore, a separate watershed pollutant runoff and loading module is needed to fully integrate hydrology and pollutant dynamics at the watershed scale while incorporating relevant transport paths such as overland flow, interflow, groundwater baseflow. Most basin-scale rainfall-runoff and water quality models, such as SWAT, HSPF, and SWMM, follow a semi-distributed approach, incorporating both watershed runoff and river routing components. Similarly, integrating a hydrologic module into EufoRiA would enable more effective management of river basins that are impacted by both point and nonpoint-source pollutants.

6. Conclusion

This study introduces EufoRiA, an open-source physics-based model designed for online nutrient and HAB modeling in large regulated river basins. Featuring a hydraulic model based on the Saint-Venant equations, a contaminant transport model based on the unsteady advection-diffusion equation, a coupled reaction kinetics model incorporating 23 water quality constituents, and an efficient data assimilation scheme based on Extended Kalman Filtering, EufoRiA is designed to enable real-time simulation and management of nutrient cycling and eutrophication dynamics at the river basin scale. Validating EufoRiA against long-term water quality observations from the Nakdong river basin in South Korea, the model demonstrates good performance in reproducing observed water quality, with high accuracy for temperature, DO, TN, and TP, and relatively less accurate results for Chl-a and TOC, although performance is comparable to existing modeling studies. Applying data assimilation to fuse water quality measurements into the model drastically improves forecasting skill for all performance metrics considered. A spatial holdout assessment confirms that performance improvements attributable to DA persist even under relatively sparse sensor coverage. Moreover, we show that calibrating the process noise covariance matrix to incorporate historically observed spatial and inter-constituent correlations significantly improves state estimation performance.

EufoRiA provides a foundation for future digital twin models of river networks that will enable online simulation of nutrient and eutrophication dynamics. Towards this end, EufoRiA integrates a fully open-source software framework; advanced physics-based modeling of unsteady hydraulics, contaminant fate and transport, and multispecies reaction kinetics; and an efficient watershed-scale data assimilation scheme based on Extended Kalman Filtering that continuously updates the process model based on real-time observed data. However, further enhancements focusing on computational efficiency optimization, accuracy improvements for complex organic constituents like Chl-a and TOC, and integration with watershed runoff models are needed to enable a more holistic water management system. Through these efforts, we expect that EufoRiA will serve as an effective tool for water quality management in regulated river basins, helping water managers address the significant challenges of water quality deterioration and HABs caused by climate change.

CRediT authorship contribution statement

Min-Gyu Kim: Writing – original draft, Conceptualization, Methodology, Software, Investigation, Visualization. **Matthew Bartos**: Writing – review and editing, Supervision, Methodology, Software, Validation, Funding acquisition

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.

Acknowledgement

This study was supported in part by the National Science Foundation, United States under Grant No. 2340176, and the Korea Water Resources corporation.

Supplementary Information: A. Supplementary Information

Supplementary information for this article is attached as a supplementary information document. This document includes a list of state variables, a list

of reaction parameters and their calibrated values (S1), brief review of the biochemical dynamics in the existing models (S1), comprehensive theoretical backgrounds for all considered biochemical reaction kinetics (S4–S10), calibration and validation results for representative eight weir locations (S18), monthly boxplot for eight weir locations during the validation period (S19), and the normalized target diagrams [62] for representative six water quality constituents based on the different numbers of DA location scenarios (S20).

Software and data availability

Source code and data for the pre-review version of this work is available at: `https://github.com/minsky97/EufoRiA`.

- Software Name: EufoRiA
- Developer: Min-Gyu Kim and Matthew Bartos
- Contact: mgkim@utexas.edu and mdbartos@utexas.edu
- Date first available: June 26, 2025
- Programming Language: Python
- Documentation: Detailed documentation on theoretical backgrounds, parameters, and additional model outputs is available in Supplementary Information A. S1

References

[1] J. Huisman, G. A. Codd, H. W. Paerl, B. W. Ibelings, J. M. H. Verspagen, P. M. Visser, Cyanobacterial blooms, Nature Reviews Microbiology 16 (2018) 471–483. doi:10.1038/s41579-018-0040-1.

[2] H. W. Paerl, T. G. Otten, Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls, Microbial Ecology 65 (2013) 995–1010. doi:10.1007/s00248-012-0159-y.

[3] L. E. Fleming, C. Rivero, J. Burns, C. Williams, J. A. Bean, K. A. Shea, J. Stinn, Blue green algal (cyanobacterial) toxins, surface drinking water, and liver cancer in Florida, Harmful Algae 1 (2002) 157–168. doi:10.1016/S1568-9883(02)00026-4.

[4] S. B. Watson, C. Miller, G. Arhonditsis, G. L. Boyer, W. Carmichael, M. N. Charlton, R. Confesor, D. C. Depew, T. O. Höök, S. A. Ludsin, G. Matisoff, S. P. McElmurry, M. W. Murray, R. Peter Richards, Y. R. Rao, M. M. Steffen, S. W. Wilhelm, The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia, Harmful Algae 56 (2016) 44–66. doi:10.1016/j.hal.2016.04.010.

[5] K. B. Kim, M.-K. Jung, Y. F. Tsang, H.-H. Kwon, Stochastic modeling of chlorophyll-a for probabilistic assessment and monitoring of algae blooms in the Lower Nakdong River, South Korea, Journal of Hazardous Materials 400 (2020) 123066. doi:10.1016/j.jhazmat.2020.123066.

[6] S. Bae, D. Seo, Analysis and modeling of algal blooms in the Nakdong River, Korea, Ecological Modelling 372 (2018) 53–63. doi:10.1016/j.ecolmodel.2018.01.019.

[7] M. L. Wells, B. Karlson, A. Wulff, R. Kudela, C. Trick, V. Asnaghi, E. Berdalet, W. Cochlan, K. Davidson, M. De Rijcke, S. Dutkiewicz, G. Hallegraeff, K. J. Flynn, C. Legrand, H. Paerl, J. Silke, S. Suikkanen, P. Thompson, V. L. Trainer, Future HAB science: Directions and challenges in a changing climate, Harmful Algae 91 (2020) 101632. doi:10.1016/j.hal.2019.101632.

[8] H. W. Paerl, W. S. Gardner, K. E. Havens, A. R. Joyner, M. J. McCarthy, S. E. Newell, B. Qin, J. T. Scott, Mitigating cyanobacterial harmful algal blooms in aquatic ecosystems impacted by climate change and anthropogenic nutrients, Harmful Algae 54 (2016) 213–222. doi:10.1016/j.hal.2015.09.009.

[9] H. W. Paerl, J. Huisman, Blooms Like It Hot, Science 320 (2008) 57–58. doi:10.1126/science.1155398.

[10] C. R. Anderson, S. K. Moore, M. C. Tomlinson, J. Silke, C. K. Cusack, Living with Harmful Algal Blooms in a Changing World, in: Coastal and Marine Hazards, Risks, and Disasters, Elsevier, 2015, pp. 495–561. doi:10.1016/B978-0-12-396483-0.00017-0.

[11] C. Lancelot, V. Thieu, A. Polard, J. Garnier, G. Billen, W. Hecq, N. Gypens, Cost assessment and ecological effectiveness of nutrient reduction options for mitigating Phaeocystis colony blooms in the Southern North Sea: An integrated modeling approach, Science of The Total

Environment 409 (2011) 2179–2191. doi:10.1016/j.scitotenv.2011.02.023.

[12] W. Guan, M. Bao, X. Lou, Z. Zhou, K. Yin, Monitoring, modeling and projection of harmful algal blooms in China, Harmful Algae 111 (2022) 102164. doi:10.1016/j.hal.2021.102164.

[13] D. K. Ralston, S. K. Moore, Modeling harmful algal blooms in a changing climate, Harmful Algae 91 (2020) 101729. doi:10.1016/j.hal.2019.101729.

[14] R. C. Cruz, P. Reis Costa, S. Vinga, L. Krippahl, M. B. Lopes, A Review of Recent Machine Learning Advances for Forecasting Harmful Algal Blooms and Shellfish Contamination, Journal of Marine Science and Engineering 9 (2021) 283. doi:10.3390/jmse9030283.

[15] Q. V. Ly, X. C. Nguyen, N. C. Lê, T.-D. Truong, T.-H. T. Hoang, T. J. Park, T. Maqbool, J. Pyo, K. H. Cho, K.-S. Lee, J. Hur, Application of Machine Learning for eutrophication analysis and algal bloom prediction in an urban river: A 10-year study of the Han River, South Korea, Science of The Total Environment 797 (2021) 149040. doi:10.1016/j.scitotenv.2021.149040.

[16] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, L. Kagal, Explaining Explanations: An Overview of Interpretability of Machine Learning, in: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), IEEE, Turin, Italy, 2018, pp. 80–89. doi:10.1109/DSAA.2018.00018.

[17] W. Astuti, R. Govindaraju, Characterization and Modeling of Harmful Algal Blooms: A Review, Journal of Hydraulic Engineering 151 (2025) 03125001. doi:10.1061/JHEND8.HYENG-14108.

[18] S. C. Chapra, G. Pelletier, H. Tao, Qual2k: a modeling framework for simulating river and stream water quality, version 2.11: documentation and users manual, Civil and Environmental Engineering Dept., Tufts University, Medford, MA 109 (2008).

[19] A. Y. K. Kyeong Park, A Three-Dimensional Hydrodynamic-Eutrophication Model (HEM-3D) : Description of water quality and sediment process submodels (1995). doi:10.21220/V5ZH9N.

[20] T. Cole, S. Wells, CE-QUAL-W2: A Two-dimensional, Laterally Averaged, Hydrodynamic and Water Quality Model Version 3.5, 2006.

[21] M. Hipsey, J. Romero, J. Antenucci, D. Hamilton, Computational aquatic ecosystem dynamics model: Caedym v2, Contract Research Group, Centre for Water Research, University of Western Australia 90 (2006).

[22] T. Tetra, The environmental fluid dynamics code: Theory and computation, volume 3: Water quality module, Technical report (2007).

[23] T. M. Cole, S. A. Wells, Ce-qual-w2: A two-dimensional, laterally averaged, hydrodynamic and water quality model, version 3.5 (2006).

[24] B. Bicknell, J. Imhoff, J. Kittle, A. Donigian, R. Johanson, Hydrological simulation program–Fortran (HSPF): User's manual for release 12, US EPA Environmental Research Lab, 1997.

[25] K. H. Cho, Y. Pachepsky, M. Ligaray, Y. Kwon, K. H. Kim, Data assimilation in surface water quality modeling: A review, Water Research 186 (2020) 116307. doi:10.1016/j.watres.2020.116307.

[26] J. Mao, J. H. Lee, K. Choi, The extended Kalman filter for forecast of algal bloom dynamics, Water Research 43 (2009) 4214–4224. doi:10.1016/j.watres.2009.06.012.

[27] A. Voutilainen, T. Pyhälahti, K. Y. Kallio, J. Pulliainen, H. Haario, J. P. Kaipio, A filtering approach for estimating lake water quality from remote sensing data, International Journal of Applied Earth Observation and Geoinformation 9 (2007) 50–64. doi:10.1016/j.jag.2006.07.001.

[28] R. Pastres, S. Ciavatta, C. Solidoro, The Extended Kalman Filter (EKF) as a tool for the assimilation of high frequency water quality data, Ecological Modelling 170 (2003) 227–235. doi:10.1016/S0304-3800(03)00230-8.

[29] K. Ennola, J. Sarvala, G. Dévai, Modelling zooplankton population dynamics with the extended Kalman filtering technique, Ecological Modelling 110 (1998) 135–149. doi:10.1016/S0304-3800(98)00057-X.

[30] C. Chen, J. Huang, Q. Chen, J. Zhang, Z. Li, Y. Lin, Assimilating multi-source data into a three-dimensional hydro-ecological dynamics model using Ensemble Kalman Filter, Environmental Modelling & Software 117 (2019) 188–199. doi:10.1016/j.envsoft.2019.03.028.

[31] A. Javaheri, M. Babbar-Sebens, R. N. Miller, S. L. Hallett, J. L. Bartholomew, An adaptive ensemble Kalman filter for assimilation of multi-sensor, multi-modal water temperature observations into hydrodynamic model of shallow rivers, Journal of Hydrology 572 (2019) 682–691. doi:10.1016/j.jhydrol.2019.03.036.

[32] T. Page, P. J. Smith, K. J. Beven, I. D. Jones, J. A. Elliott, S. C. Maberly, E. B. Mackay, M. De Ville, H. Feuchtmayr, Adaptive forecasting of phytoplankton communities, Water Research 134 (2018) 74–85. doi:10.1016/j.watres.2018.01.046.

[33] J. Huang, J. Gao, An improved Ensemble Kalman Filter for optimizing parameters in a coupled phosphorus model for lowland polders in Lake Taihu Basin, China, Ecological Modelling 357 (2017) 14–22. doi:10.1016/j.ecolmodel.2017.04.019.

[34] A. Javaheri, M. Babbar-Sebens, R. N. Miller, From skin to bulk: An adjustment technique for assimilation of satellite-derived temperature observations in numerical models of small inland water bodies, Advances in Water Resources 92 (2016) 284–298. doi:10.1016/j.advwatres.2016.03.012.

[35] D. Shao, Z. Wang, B. Wang, W. Luo, A Water Quality Model with Three Dimensional Variational Data Assimilation for Contaminant Transport, Water Resources Management 30 (2016) 4501–4512. doi:10.1007/s11269-016-1432-5.

[36] S. Wang, N. Flipo, T. Romary, Oxygen data assimilation for estimating micro-organism communities' parameters in river systems, Water Research 165 (2019) 115021. doi:10.1016/j.watres.2019.115021.

[37] S. Kim, D.-J. Seo, H. Riazi, C. Shin, Improving water quality forecasting via data assimilation – Application of maximum likelihood ensemble filter to HSPF, Journal of Hydrology 519 (2014) 2797–2809. doi:10.1016/j.jhydrol.2014.09.051.

[38] M. Babbar-Sebens, L. Li, K. Song, S. Xie, On the Use of Landsat-5 TM Satellite for Assimilating Water Temperature Observations in 3D Hydrodynamic Model of Small Inland Reservoir in Midwestern US, Advances in Remote Sensing 02 (2013) 214–227. doi:10.4236/ars.2013.23024.

[39] B. Beck, P. Young, Systematic Identification of DO-BOD Model Structure, Journal of the Environmental Engineering Division 102 (1976) 909–927. doi:10.1061/JEEGAV.0000554.

[40] P. Whitehead, Modelling algal behaviour in the river thames, Water Research 18 (1984) 945–953. doi:10.1016/0043-1354(84)90244-6.

[41] B. Cosby, G. Hornberger, Identification of photosynthesis-light models for aquatic systems I. Theory and simulations, Ecological Modelling 23 (1984) 1–24. doi:10.1016/0304-3800(84)90116-9.

[42] K. Kim, M. Park, J.-H. Min, I. Ryu, M.-R. Kang, L. J. Park, Simulation of algal bloom dynamics in a river with the ensemble Kalman filter, Journal of Hydrology 519 (2014) 2810–2821. doi:10.1016/j.jhydrol.2014.09.073.

[43] N. Margvelashvili, J. Parslow, M. Herzfeld, K. Wild-Allen, J. Andrewartha, F. Rizwi, E. Jones, Development of operational data-assimilating water quality modelling system for South-East Tasmania, in: OCEANS'10 IEEE SYDNEY, IEEE, Sydney, Australia, 2010, pp. 1–5. doi:10.1109/OCEANSSYD.2010.5603601.

[44] A. Vodacek, Y. Li, A. J. Garrett, Remote sensing data assimilation in environmental models, in: 2008 37th IEEE Applied Imagery Pattern Recognition Workshop, IEEE, Washington, DC, USA, 2008, pp. 1–5. doi:10.1109/AIPR.2008.4906472.

[45] S. Loos, C. M. Shin, J. Sumihar, K. Kim, J. Cho, A. H. Weerts, Ensemble data assimilation methods for improving river water quality forecasting accuracy, Water Research 171 (2020) 115343. doi:10.1016/j.watres.2019.115343.

[46] M.-G. Kim, M. Bartos, A digital twin model for contaminant fate and transport in urban and natural drainage networks with online state estimation, Environmental Modelling & Software 171 (2024) 105868. doi:10.1016/j.envsoft.2023.105868.

[47] Z. Ji, General Hydrodynamic Model for Sewer/Channel Network Systems, Journal of Hydraulic Engineering 124 (1998) 307–315. doi:10.1061/(ASCE)0733-9429(1998)124:3(307).

[48] J. L. Martin, S. C. McCutcheon, Hydrodynamics and transport for water quality modeling, CRC press, 1998.

[49] M. Bartos, B. Kerkez, Pipedream: An interactive digital twin model for natural and urban drainage systems, Environmental Modelling & Software 144 (2021) 105120. doi:10.1016/j.envsoft.2021.105120.

[50] J. A. Aguilar-Torrejón, P. Balderas-Hernández, G. Roa-Morales, C. E. Barrera-Díaz, I. Rodríguez-Torres, T. Torres-Blancas, Relationship, importance, and development of analytical techniques: COD, BOD, and, TOC in water—An overview through time, SN Applied Sciences 5 (2023) 118. doi:10.1007/s42452-023-05318-7.

[51] MOE, The 2nd Master plan for water environmental management, Ministry of Environment of South Korea, Daejeon, South Korea, 2015.

[52] S. S. Park, Y. S. Lee, A water quality modeling study of the Nakdong River, Korea, Ecological Modelling 152 (2002) 65–75. doi:10.1016/S0304-3800(01)00489-6.

[53] J. Kim, J. R. Jones, D. Seo, Factors affecting harmful algal bloom occurrence in a river with regulated hydrology, Journal of Hydrology: Regional Studies 33 (2021) 100769. doi:10.1016/j.ejrh.2020.100769.

[54] J. Kim, T. Lee, D. Seo, Algal bloom prediction of the lower Han River, Korea using the EFDC hydrodynamic and water quality model, Ecological Modelling 366 (2017) 27–36. doi:10.1016/j.ecolmodel.2017.10.015.

[55] H. Y. Kim, C. Shin, Y. Park, J. Moon, Water Resources Management in the Republic of Korea: Korea's Challenge to Flood & Drought with Multi-purpose Dam and Multi-regional Water Supply System, IDB Publications (2018). doi:10.18235/0001532.

[56] R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering 82 (1960) 35–45. doi:10.1115/1.3662552.

[57] A. T. Cahill, F. Ungaro, M. B. Parlange, M. Mata, D. R. Nielsen, Combined spatial and kalman filter estimation of optimal soil hydraulic properties, Water Resources Research 35 (1999) 1079–1088. doi:10.1029/1998WR900121.

[58] H. V. Gupta, H. Kling, K. K. Yilmaz, G. F. Martinez, Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling, Journal of hydrology 377 (2009) 80–91. doi:10.1016/j.jhydrol.2009.08.003.

[59] D. N. Moriasi, M. W. Gitau, N. Pai, P. Daggupati, Hydrologic and water quality models: Performance measures and evaluation criteria, Transactions of the ASABE 58 (2015) 1763–1785. doi:10.13031/trans.58.10715.

[60] J. Oh, M. Bartos, Model predictive control of stormwater basins coupled with real-time data assimilation enhances flood and pollution control under uncertainty, Water Research 235 (2023) 119825. URL: https://www.sciencedirect.com/science/article/pii/S0043135423002609. doi:https://doi.org/10.1016/j.watres.2023.119825.

[61] K.-E. Lindenschmidt, The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling, Ecological Modelling 190 (2006) 72–86. doi:https://doi.org/10.1016/j.ecolmodel.2005.04.016.

[62] J. K. Jolliff, J. C. Kindle, I. Shulman, B. Penta, M. A. Friedrichs, R. Helber, R. A. Arnone, Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment, Journal of Marine Systems 76 (2009) 64–82. doi:10.1016/j.jmarsys.2008.05.014.

[63] R. V. Thomann, J. A. Mueller, Principles of surface water quality modeling and control (1987).

[64] D. K. Brady, W. L. Graves, J. C. Geyer, Surface heat exchange at power plant cooling lakes (1969).

[65] E. L. Thackston, J. W. Dawson III, Recalibration of a reaeration equation, Journal of environmental engineering 127 (2001) 317–321.

[66] K. W. Thornton, A. S. Lessem, A temperature algorithm for modifying biological rates, Transactions of the American Fisheries Society 107 (1978) 284–287. doi:10.1577/1548-8659(1978)107<284:ATAFMB>2.0.CO;2.

[67] R. V. Thomann, J. J. Fitzpatrick, Calibration and Verification of a Mathematical Model of the Eutrophication of the Potomac Estuary, DC Department of Environmental Sciences, 1982.

[68] H.-S. Kim, S.-Y. Kim, J. Park, M. Han, The Fractionation Characteristics of Organic Matter in Pollution Sources and River, Journal of Korean Society on Water Environment 33 (2017) 580–586. doi:10.15681/KSWE.2017.33.5.580.

[69] Y. Du, X. Liu, M. Hu, X. Liu, W. Peng, C. Liu, K. Rinke, B. Boehrer, Y. Wang, Resolving spatially complex interactions between hydrodynamics and biogeochemical processing in a large reservoir with metalimnetic oxygen deficits, Journal of Hydrology 644 (2024) 132060. doi:10.1016/j.jhydrol.2024.132060.

[70] D. Brito, T. B. Ramos, M. C. Gonçalves, M. Morais, R. Neves, Integrated modelling for water quality management in a eutrophic reservoir in south-eastern Portugal, Environmental Earth Sciences 77 (2018) 40. doi:10.1007/s12665-017-7221-5.

[71] S. M. Hong, A. Abbas, S. Kim, D. H. Kwon, N. Yoon, D. Yun, S. Lee, Y. Pachepsky, J. Pyo, K. H. Cho, Autonomous calibration of EFDC for predicting chlorophyll-a using reinforcement learning and a real-time monitoring system, Environmental Modelling & Software 168 (2023) 105805. doi:10.1016/j.envsoft.2023.105805.

[72] D. H. Lee, J. H. Kim, M.-H. Park, M. K. Stenstrom, J.-H. Kang, Automatic calibration and improvements on an instream chlorophyll a simulation in the HSPF model, Ecological Modelling 415 (2020) 108835. doi:10.1016/j.ecolmodel.2019.108835.

[73] H. H. Bui, N. H. Ha, T. N. D. Nguyen, A. T. Nguyen, T. T. H. Pham, J. Kandasamy, T. V. Nguyen, Integration of SWAT and QUAL2K for water quality modeling in a data scarce basin of Cau River basin in

Vietnam, Ecohydrology & Hydrobiology 19 (2019) 210–223. doi:10.1016/j.ecohyd.2019.03.005.

[74] J. Kim, A. Jonoski, D. P. Solomatine, P. L. M. Goethals, Water Quality Modelling for Nitrate Nitrogen Control Using HEC-RAS: Case Study of Nakdong River in South Korea, Water 15 (2023) 247. doi:10.3390/w15020247.

[75] J. Pyo, L. J. Park, Y. Pachepsky, S.-S. Baek, K. Kim, K. H. Cho, Using convolutional neural network for predicting cyanobacteria concentrations in river water, Water Research 186 (2020) 116349. doi:10.1016/j.watres.2020.116349.

[76] D. Kim, Y. Kim, B. Kim, Simulation of eutrophication in a reservoir by CE-QUAL-W2 for the evaluation of the importance of point sources and summer monsoon, Lake and Reservoir Management 35 (2019) 64–76. doi:10.1080/10402381.2018.1530318.

[77] M. T. Brett, S. K. Ahopelto, H. K. Brown, B. E. Brynestad, T. W. Butcher, E. E. Coba, C. A. Curtis, J. T. Dara, K. B. Doeden, K. R. Evans, L. Fan, J. D. Finley, N. J. Garguilo, S. M. Gebreeyesus, M. K. Goodman, K. W. Gray, C. Grinnell, K. L. Gross, B. R. E. Hite, A. J. Jones, P. T. Kenyon, A. M. Klock, R. E. Koshy, A. M. Lawler, M. Lu, L. Martinkosky, J. R. Miller-Schulze, Q. T. N. Nguyen, E. R. Runde, J. M. Stultz, S. Wang, F. P. White, C. H. Wilson, A. S. Wong, S. Y. Wu, P. G. Wurden, T. R. Young, G. B. Arhonditsis, The modeled and observed response of Lake Spokane hypolimnetic dissolved oxygen concentrations to phosphorus inputs, Lake and Reservoir Management 32 (2016) 246–258. doi:10.1080/10402381.2016.1170079.

[78] B. Tasnim, X. Fang, J. S. Hayworth, D. Tian, Simulating Nutrients and Phytoplankton Dynamics in Lakes: Model Development and Applications, Water 13 (2021) 2088. doi:10.3390/w13152088.

Supplementary Information

for

EufoRiA: A new multi-constituent nutrient and harmful algal blooms model for river networks with online data assimilation

Min-Gyu Kim & Matthew Bartos

mgkim@utexas.edu & mdbartos@utexas.edu

Civil, Architectural and Environmental Engineering,
University of Texas at Austin

S1. Definition of the state variables and parameters

Table S1: State variables including hydraulics and water quality constituents in EufoRiA (1)

Variable	Description	Unit
T_w	Water temperature	$°C$
$A_{surface}$	Surface area of the control volume	m^2
A_{bottom}	Bottom area of the control volume	m^2
V	Volume of the control volume	m^3
H	Average hydraulic depth	m
C_{DO}	Dissolved oxygen concentration	mg/L
$C_{DO,sat}$	Saturation DO concentration	mg/L
C_{DO}	DO concentration	mg/L
$C_{Alg,Diatom}$	Algal concentration of Diatom	mg/L
$C_{Alg,Green}$	Algal concentration of Green algae	mg/L
$C_{Alg,Cyano}$	Algal concentration of Cyanobacteria	mg/L
C_{LDOC}	Labile dissolved organic carbon concentration	mg/L
C_{LPOC}	Labile particulate organic carbon concentration	mg/L
C_{RDOC}	Refractory dissolved organic carbon concentration	mg/L
C_{RPOC}	Refractory particulate organic carbon concentration	mg/L
C_{NH4}	Ammonium concentration	mg/L
C_{NO3}	Nitrate (NO_3-N+NO_2-N) concentration	mg/L

Table S2: State variables including hydraulics and water quality constituents in EufoRiA (2)

Variable	Description	Unit
C_{PIN}	Particulate inorganic nitrogen concentration	mg/L
C_{LDON}	Labile dissolved organic nitrogen concentration	mg/L
C_{LPON}	Labile particulate organic nitrogen concentration	mg/L
C_{RDON}	Refractory dissolved organic nitrogen concentration	mg/L
C_{RPON}	Refractory particulate organic nitrogen concentration	mg/L
C_{PO4}	Phosphate concentration	mg/L
C_{PIP}	Particulate inorganic phosphorus concentration	mg/L
C_{LDOP}	Labile dissolved organic phosphorus concentration	mg/L
C_{LPOP}	Labile particulate organic phosphorus concentration	mg/L
C_{RDOP}	Refractory dissolved organic phosphorus concentration	mg/L
C_{RPOP}	Refractory particulate organic phosphorus concentration	mg/L
C_{ISS}	Inorganic suspended solids concentration	mg/L

Table S3: Parameters related to the biochemical reactions in EufoRiA (1)

Parameter	Description	Unit	Applied Value
K_{SOD}	Sediment oxygen demand (SOD)	$mg/m^2/day$	0.12–0.50
$K1_{SOD}$	K1 parameter for the temperature multiplier for SOD	-	0.01
$K2_{SOD}$	K2 parameter for the temperature multiplier for SOD	-	0.99
$T1_{SOD}$	T1 parameter for the temperature multiplier for SOD	-	0.0
$T2_{SOD}$	T2 parameter for the temperature multiplier for SOD	-	23.0
δ_{NH_4-N}	Oxygen stoichiometric coefficient for nitrification	-	4.57
δ_{OC}	Oxygen stoichiometric coefficient for organic carbon	-	0.3
$\delta_{ag,Diatom}$	Oxygen stoichiometric coefficient for algal growth for Diatom	-	1.7
$\delta_{ag,Green}$	Oxygen stoichiometric coefficient for algal growth for Green algae	-	1.7
$\delta_{ag,Cyano}$	Oxygen stoichiometric coefficient for algal growth for Cyanobacteria	-	1.7

Table S4: Parameters related to the biochemical reactions in EufoRiA (2)

Parameter	Description	Unit	Applied Value
$K_{Chl,Diatom}$	Algal biomass (mg/L) to Chlorophyll-a (mg/m^3) ratio of diatom	-	0.07
$K_{ag,max,Diatom}$	Maximum algal growth rate of diatom	$1/day$	4.75
$K_{ar,max,Diatom}$	Maximum algal respiration rate of diatom	$1/day$	0.019
$K_{am,max,Diatom}$	Maximum algal mortality rate of diatom	$1/day$	0.019
$K_{ae,max,Diatom}$	Maximum algal excretion rate of diatom	$1/day$	0.019
$K_{s,Diatom}$	Settling rate of diatom	$1/day$	0.01
$K1_{Diatom}$	K1 parameter for the temperature multiplier for diatom	-	0.50
$K2_{Diatom}$	K2 parameter for the temperature multiplier for diatom	-	0.99
$K3_{Diatom}$	K3 parameter for the temperature multiplier for diatom	-	0.99
$K4_{Diatom}$	K4 parameter for the temperature multiplier for diatom	-	0.01
$T1_{Diatom}$	T1 parameter for the temperature multiplier for diatom	-	0.01
$T2_{Diatom}$	T2 parameter for the temperature multiplier for diatom	-	6.0
$T3_{Diatom}$	T3 parameter for the temperature multiplier for diatom	-	10.0
$T4_{Diatom}$	T4 parameter for the temperature multiplier for diatom	-	11.0
$\delta_{C,Diatom}$	Stoichiometric equivalent between algal biomass and carbon for green diatom	-	0.6
$\delta_{N,Diatom}$	Stoichiometric equivalent between algal biomass and nitrogen for green diatom	-	0.08
$\delta_{P,Diatom}$	Stoichiometric equivalent between algal biomass and phosphorus for diatom	-	0.005

Table S5: Parameters related to the biochemical reactions in EufoRiA (3)

Parameter	Description	Unit	Applied Value
$K_{Chl,Green}$	Algal biomass (mg/L) to Chlorophyll-a (mg/m^3) ratio for green algaes	-	0.07
$K_{ag,max,Green}$	Maximum growth rate of green algae	$1/day$	2.55
$K_{ar,max,Green}$	Maximum respiration rate of green algae	$1/day$	0.019
$K_{am,max,Green}$	Maximum mortality rate of green algae	$1/day$	0.019
$K_{ae,max,Green}$	Maximum excretion rate of green algae	$1/day$	0.019
$K_{s,Green}$	Settling rate of green algae	$1/day$	0.005
$K1_{Green}$	K1 parameter for the temperature multiplier for green algae	-	0.01
$K2_{Green}$	K2 parameter for the temperature multiplier for green algae	-	0.99
$K3_{Green}$	K3 parameter for the temperature multiplier for green algae	-	0.99
$K4_{Green}$	K4 parameter for the temperature multiplier for green algae	-	0.01
$T1_{Green}$	T1 parameter for the temperature multiplier for green algae	-	14.0
$T2_{Green}$	T2 parameter for the temperature multiplier for green algae	-	16.0
$T3_{Green}$	T3 parameter for the temperature multiplier for green algae	-	18.0
$T4_{Green}$	T4 parameter for the temperature multiplier for green algae	-	22.0
$\delta_{C,Green}$	Stoichiometric equivalent between algal biomass and carbon for green algae	-	0.6
$\delta_{N,Green}$	Stoichiometric equivalent between algal biomass and nitrogen for green algae	-	0.08
$\delta_{P,Green}$	Stoichiometric equivalent between algal biomass and phosphorus for green algae	-	0.005

Table S6: Parameters related to the biochemical reactions in EufoRiA (4)

Parameter	Description	Unit	Applied Value
$K_{Chl,Cyano}$	Algal biomass (mg/L) to Chlorophyll-a (mg/m^3) ratio	-	0.07
$K_{ag,max,Cyano}$	Maximum growth rate of Cyanobacteria	$1/day$	3.45
$K_{ar,max,Cyano}$	Maximum respiration rate of Cyanobacteria	$1/day$	0.019
$K_{am,max,Cyano}$	Maximum mortality rate of Cyanobacteria	$1/day$	0.019
$K_{ae,max,Cyano}$	Maximum excretion rate of Cyanobacteria	$1/day$	0.019
$K1_{Cyano}$	K1 parameter for calculating the temperature multiplier for Cyanobacteria	-	0.01
$K2_{Cyano}$	K2 parameter for calculating the temperature multiplier for Cyanobacteria	-	0.99
$K3_{Cyano}$	K3 parameter for calculating the temperature multiplier for Cyanobacteria	-	0.99
$K4_{Cyano}$	K4 parameter for calculating the temperature multiplier for Cyanobacteria	-	0.05
$T1_{Cyano}$	T1 parameter for calculating the temperature multiplier for Cyanobacteria	-	20.0
$T2_{Cyano}$	T2 parameter for calculating the temperature multiplier for Cyanobacteria	-	25.0
$T3_{Cyano}$	T3 parameter for calculating the temperature multiplier for Cyanobacteria	-	28.0
$T4_{Cyano}$	T4 parameter for calculating the temperature multiplier for Cyanobacteria	-	32.0
$\delta_{C,cyano}$	Stoichiometric equivalent between algal biomass and carbon for cyanobacteria	-	0.6
$\delta_{N,cyano}$	Stoichiometric equivalent between algal biomass and nitrogen for cyanobacteria	-	0.08
$\delta_{P,cyano}$	Stoichiometric equivalent between algal biomass and phosphorus for cyanobacteria	-	0.005

Table S7: Parameters related to the biochemical reactions in EufoRiA (5)

Parameter	Description	Unit	Applied Value
K_{LDOC}	Labile dissolved organic carbon decay rate	$1/day$	0.002
K_{LPOC}	Labile particulate organic carbon decay rate	$1/day$	0.002
K_{RDOC}	Refractory dissolved organic carbon decay rate	$1/day$	0.002
K_{RPOC}	Refractory particulate organic carbon decay rate	$1/day$	0.002
$K_{LDOC \rightarrow RDOC}$	Decomposition rate from labile to refractory DOC	$1/day$	0.01
$K_{LPOC \rightarrow RPOC}$	Decomposition rate from labile to refractory POC	$1/day$	0.01
$K_{s,LPOC}$	Settling coefficient for LPOC	m/day	0.002
$K_{s,RPOC}$	Settling coefficient for RPOC	m/day	0.002
K_{NH4}	Nitrification rate	$1/day$	0.18
$K1_{NH4}$	K1 parameter for calculating the temperature multiplier for nitrification	-	0.05
$K2_{NH4}$	K2 parameter for calculating the temperature multiplier for nitrification	-	0.99
$T1_{NH4}$	T1 parameter for calculating the temperature multiplier for nitrification	-	0.01
$T2_{NH4}$	T2 parameter for calculating the temperature multiplier for nitrification	-	15.0
K_{NO3}	Denitrification rate	$1/day$	0.012
$K1_{NO3}$	K1 parameter for calculating the temperature multiplier for denitrification	-	0.0001
$K2_{NO3}$	K2 parameter for calculating the temperature multiplier for denitrification	-	0.99
$T1_{NO3}$	T1 parameter for calculating the temperature multiplier for denitrification	-	10.0
$T2_{NO3}$	T2 parameter for calculating the temperature multiplier for denitrification	-	20.0
$K_{s,NO3}$	Settling coefficient for nitrate	m/day	0.000

Table S8: Parameters related to the biochemical reactions in EufoRiA (6)

Parameter	Description	Unit	Applied Value
$K_{s,PIN}$	Settling coefficient for particulate inorganic nitrogen	m/day	0.005
P_{NH4}	Ammonium preference factor by NH4 and NO3 concentration	-	0.000
$P_{d,N}$	Portion factor when algae decay to LDON	-	0.5
K_{LDON}	Labile dissolved organic nitrogen decay rate	$1/day$	0.02
K_{LPON}	Labile particulate organic nitrogen decay rate	$1/day$	0.02
K_{RDON}	Refractory dissolved organic nitrogen decay rate	$1/day$	0.02
K_{RPON}	Refractory particulate organic nitrogen decay rate	$1/day$	0.02
$K_{LDON \rightarrow RDON}$	Decomposition rate from labile to refractory DON	$1/day$	0.02
$K_{LPON \rightarrow RPON}$	Decomposition rate from labile to refractory PON	$1/day$	0.02
$K_{s,LPON}$	Settling coefficient for LPON	m/day	0.002
$K_{s,RPON}$	Settling coefficient for RPON	m/day	0.002
$P_{d,P}$	Portion factor when algae decays to LDOP	-	0.6
$K_{ads,PIP}$	Adsorption coefficient to PIP	$1/day$	0.02
$K_{des,PIP}$	Desorption coefficient from PIP	$1/day$	0.005
$K_{s,PIP}$	Settling coefficient for PIP	m/day	0.05
K_{LDOP}	Labile dissolved organic phosphorus decay rate	$1/day$	0.015
K_{LPOP}	Labile particulate organic phosphorus decay rate	$1/day$	0.015
K_{RDOP}	Refractory dissolved organic phosphorus decay rate	$1/day$	0.015
K_{RPOP}	Refractory particulate organic phosphorus decay rate	$1/day$	0.015

Table S9: Parameters related to the biochemical reactions in EufoRiA (7)

Parameter	Description	Unit	Applied Value
$K_{LDOP \rightarrow RDOP}$	Decomposition rate from labile to refractory DOP	$1/day$	0.01
$K_{LPOP \rightarrow RPOP}$	Decomposition rate from labile to refractory POP	$1/day$	0.01
$K_{s,LPOP}$	Settling coefficient for LPOP	m/day	0.04
$K_{s,RPOP}$	Settling coefficient for RPOP	m/day	0.04
$K_{h,N,Diatom}$	Half saturation coefficient of nitrogen for Diatom	mg/L	0.1
$K_{h,N,Green}$	Half saturation coefficient of nitrogen for Green algae	mg/L	0.1
$K_{h,N,Cyano}$	Half saturation coefficient of nitrogen for cyanobacteria	mg/L	0.1
$K_{h,NH4}$	Half saturation coefficient for ammonium	mg/L	0.1
$K_{h,P,Diatom}$	Half saturation coefficient of nitrogen for Diatom	mg/L	0.001
$K_{h,P,Green}$	Half saturation coefficient of nitrogen for Green algae	mg/L	0.001
$K_{h,P,Cyano}$	Half saturation coefficient of nitrogen for Cyanobacteria	mg/L	0.001
$K_{h,PIP}$	Half saturation coefficient for adsorption to PIP	mg/L	0.005

S2. Review of the existing models focusing on the biochemical reactions

Based on a literature review of widely used existing models (QUAL2K, EFDC, CAEDYM, and CE-QUAL-W2), we include the most essential constituents, such as temperature, dissolved oxygen, organic and inorganic nutrients (N, P, C), and algal groups (Green, Diatom, and Cyanobacteria) in EufoRiA.

The Environmental Fluid Dynamics Code (EFDC) is a surface water modeling system that includes three-dimensional hydrodynamics and a water quality model. This model was first developed at the Virginia Institute of Marine Science (VIMS) and has been applied to hundreds of water bodies, including lakes, estuaries, oceans, and rivers worldwide. For water quality modeling, the initial approach involved coupling EFDC's hydrodynamic code with the external water quality model, WASP. Later, a fully integrated code was developed based on the nutrient and sediment kinetics equations of the CE-QUAL-ICM model [22].

The Computational Aquatic Ecosystem Dynamics Model (CAEDYM) is one of the most comprehensive and complex water quality and ecological models, incorporating various nutrients, algae, epiphyton, macrophytes, and higher organisms such as fish. CAEDYM requires an external three-dimensional hydrodynamics model, the Estuary, Lake, and Coastal Ocean Model (ELCOM) [21]. In Korea, the ELCOM-CAEDYM model is implemented as a reservoir model for K-water's integrated water quality modeling system, particularly due to its strong capability to reproduce thermal stratification in density-stratified lakes.

CE-QUAL-W2 is a two-dimensional vertical hydrodynamics and water quality model actively maintained by the USACE and the Water Quality Research Group at Portland State University. The latest version of this model was released on February 11, 2023, [23]. It has been applied to numerous rivers and reservoirs in various countries. In Korea, K-water utilizes this model as a river model for major large rivers and their tributaries.

QUAL2E is one of the most widely used water quality models. Its long development history and numerous application cases make it a reliable choice for solving water quality problems. QUAL2K is the latest version of the QUAL series model, providing fundamental biochemical reaction kinetics for one-dimensional river networks. Since our EufoRiA model also incorporates one-dimensional river networks, reviewing the reaction relationships

presented by QUAL2K is necessary and valuable [18].

There are some controversial differences in reaction kinetics among existing models, particularly in how each model handles organic matter. For example, in the EFDC model, algal mortality increases organic matter groups—including labile and refractory forms—regardless of whether they are dissolved or particulate [22]. However, in the CE-QUAL-W2 and CAEDYM models, algal mortality results in an increase only in labile dissolved and particulate organic matter.

Regarding the transformation of organic matter, the CAEDYM model allows labile particulate organic matter to decompose into labile dissolved organic matter, while refractory particulate organic matter decays into refractory dissolved organic matter. In contrast, the CE-QUAL-W2 model permits labile dissolved and particulate organic matter to decompose into refractory dissolved and particulate organic matter. Additionally, the EFDC model follows a slightly different reaction scheme for organic matter. QUAL2K, on the other hand, uses a relatively simple approach to simulate organic nutrients [18]. In EufoRiA, we adopt the organic matter kinetics of the CE-QUAL-W2 model (Figure S1).

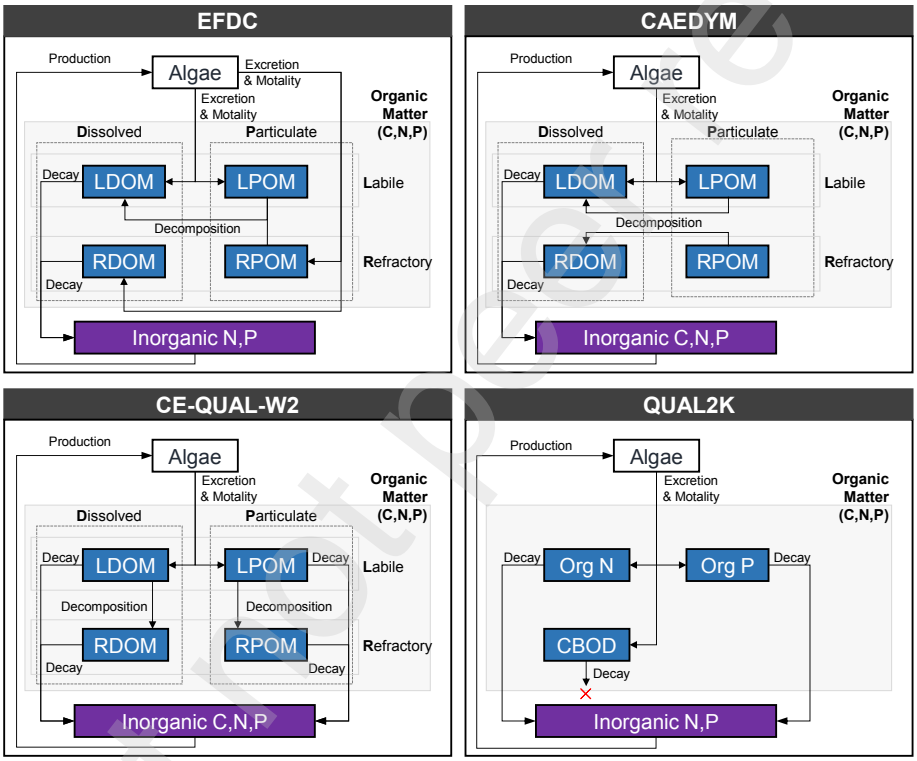


Figure S1: Comparison of the organic and nutrient cycle between models

S3. Limitations of existing models

Existing physics-based HAB models like CE-QUAL-W2, EFDC, and CAEDYM are primarily applied to individual river channels, estuaries, reservoirs, and oceans as opposed to river networks consisting of multiple interconnected reaches. These models focus on accurately capturing the stratification of heat, nutrient, and algae concentrations within these large waterbodies, which requires computationally-intensive 2D or 3D simulations. However, in complex river networks consisting of shallow and well-mixed streams, higher-dimensional models face challenges such as numerical instability due to rapidly changing dry and wet conditions, the need for small time steps, and difficulty in representing irregular horizontal and depth grids under complicated river bathymetry. Given these limitations, 1D models often present a more practical solution for simulation of water quality dynamics at the watershed scale. However, existing 1D models feature simplified hydraulics that limit their applicability in managed river basins that are affected by backwater effects and hydraulic controls. QUAL2K—the most widely applied 1D model—is a steady-state model and thus does not account for unsteady flow induced by storm events or controlled hydraulic structures like movable weirs [18]. While HSPF supports unsteady flow, its unidirectional kinematic wave routing scheme does not capture backwater effects created by impoundments, making it poorly-suited to studying the effects of hydraulic controls in regulated river basins. In summary, there is a strong need for a watershed-scale nutrient and HABs model that accounts for both unsteady hydraulics and multi-constituent contaminant fate and transport.

S4. Water Temperature

In EufoRiA, we adopt the full heat balance model in the existing models including QUAL2K and CE-QUAL-W2 ([18], [20]). Most theoretical backgrounds and equations are referenced and refined from these models.

S4.1. Transport model of the water temperature

We can express the advection-diffusion transport of water temperature using the following equation. In the modeling process, temperature is simulated as an arbitrary constituent.

$$\frac{\partial AT_w}{\partial t} = \frac{\partial}{\partial x}\left(DA\frac{\partial T_w}{\partial x}\right) - \frac{\partial QT_w}{\partial x} + \text{source/sink} \tag{S.1}$$

Where, D is the diffusion coefficient (m^2/sec), A is the cross-sectional area of links (m^2), Q is the flow rate (m^3/sec).

S4.2. Heat balance of the water column

The QUAL2K documentation [18] provides detailed heat balance relationships and equations for a one-dimensional river model. Since this study does not aim to develop a heat exchange model, we adopted a similar approach for water temperature modeling.

In open channels that interact with the atmosphere and solar radiation, the temperature change within a control volume V can be calculated using the following relationship. This calculation is performed separately from the transport equation:

$$A_{surface}J_{total} = \rho C_p V\frac{dT_w}{dt} \tag{S.2}$$

Where J_{total} is the total heat flux through the surface area of the control volume (W/m^2), ρ is the water density (g/cm^3), C_p is the specific heat capacity of water, and V is the control volume (m^3). In temperature modeling of river channels, the variations in water density and specific heat with temperature are negligible. Therefore, we assume that $\rho = 1\ g/cm^3$ and $C_p = 1$ (i.e. we ignore these terms in the model.).

The total heat flux consists of both radiative and non-radiative components (QUAL2K documentation, 2008). These components include net solar shortwave radiation (J_{sn}), net atmospheric longwave radiation (J_{an}), longwave back radiation from the water (J_{br}), conduction and convection (J_c),

evaporation and condensation (J_e), and sediment heat exchange (J_s). The total heat flux can be expressed as follows (Figure S2).

$$J_{total} = J_{sn} + J_{an} - J_{br} - J_c - J_e + J_s \tag{S.3}$$

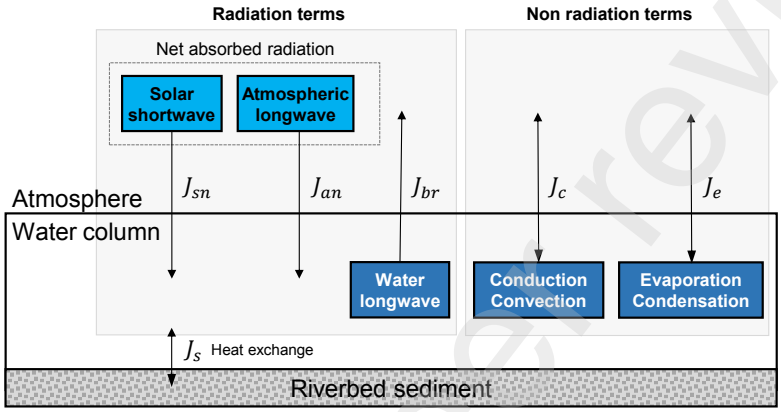


Figure S2: Total heat balance model for water temperature modeling

Table S10: Comparison of water temperature model between models

Reaction	EFDC (3-D)	CAEDYM[b] (3-D)	CE-QUAL-W2 (2-D)	QUAL2K (1-D)	EufoRiA (1-D)	Functions[a]	Method
Solar shortwave radiation(+)	✓	✓	✓	✓	✓	$f(I)$	-
Atmospheric longwave radiation(+)	✓	✓	✓	✓	✓	$f(T_{air}, e_{air})$	4th-order
Water longwave radiation(−)	✓	✓	✓	✓	✓	$f(T_w)$	4th-order
Conduction/convection(+,−)	✓	✓	✓	✓	✓	$f(T_w, T_{air}, u_{wind})$	1st-order
Evaporation/condensation(+,−)	✓	✓	✓	✓	✓	$f(u_{wind}, e_s, e_{air})$	zero-order
Sediment heat exchange(+,−)	✓	✓	✓	✓	✓	$f(T_w, T_{sed})$	1st-order
Ice formation/melt(+,−)	✓	-	✓	-	-	$f(T_w, T_{ice}, etc.)$	1st-order

[a] Where, I is solar radiation intensity (W/m^2), T_{air} is the air temperature, u_{wind} is the wind velocity (m/s), e_{air} is the vapor pressure of the atmosphere ($mmHg$), T_{sed} is the riverbed sediment temperature, and T_{ice} is the ice temperature.

S4.3. Calculation of heat flux terms

S4.3.1. Net solar shortwave radiation, J_{sn}

In EufoRiA, because we directly apply the solar radiation measurement data from KMA (Korea Meteorological Administration), our model does not

include the equations described in QUAL2K [18], to estimate net solar shortwave radiation from extraterrestrial radiation and cloud cover data.

S4.3.2. Atmospheric longwave radiation, J_{an}

Because the atmosphere has own temperature, it emits longwave radiation according to the Stefan-Boltzmann law, contributing to heat inflow into the water body. This term accounts for the attenuation effect of the atmosphere and the reflectance of the water surface [18], [63]:

$$J_{an} = \sigma(T_{air} + 273.15)^4(A + 0.031\sqrt{e_{air}})(1 - R_L) \tag{S.4}$$

where σ is the Stefan-Boltzmann constant ($= 11.7 \times 10^{-8}\ cal/(cm^2 dK^4)$), T_{air} is the air temperature ($^\circ C$), A is an empirical coefficient (ranging from 0.5 to 0.7), e_{air} is the atmospheric vapor pressure ($mmHg$), and R_L is the reflectance coefficient ($\simeq 0.03$).

S4.3.3. Water longwave back radiation, J_{br}

Similar to the atmosphere, the water body emits longwave radiation based on its temperature [18]:

$$J_{br} = \epsilon\sigma(T_w + 273.15)^4 \tag{S.5}$$

where ϵ is the emissivity of water ($\simeq 0.97$), and T_w is the water temperature.

S4.3.4. Conduction and convection, J_c

Heat exchange between air and water occurs through both conduction and convection. Wind-induced mixing enhances convective heat transfer, while conductive heat transfer occurs at the air-water interface. When the air temperature is higher than the water temperature, heat flows into the water body, and vice versa. This process is described as follows [18]:

$$J_c = c_1 f(U_w)(T_w - T_{air}) \tag{S.6}$$

where c_1 is the Bowen coefficient ($\simeq 0.47\ mmHg^\circ C^{-1}$), $f(U_w) = 19.0 + 0.95U_w^2$ proposed by Brady et al. [64], and U_w is the wind velocity (m/s).

S4.3.5. Evaporation and condensation, J_e

According to Dalton's law, evaporation results in heat loss from the water body, while condensation leads to heat gain when the atmospheric vapor

pressure exceeds the saturated vapor pressure. This process is expressed by the following equation:

$$J_e = f(U_w)(e_s - e_{air}) \tag{S.7}$$

where e_s is the saturated vapor pressure of water ($mmHg$), and e_{air} is the atmospheric vapor pressure ($mmHg$).

S4.3.6. Sediment heat exchange

The heat balance for the bottom sediment bed is expressed as [18]:

$$\frac{dT_s}{dt} = -\frac{J_s}{\rho_s C_{sed} H_{sed}} \tag{S.8}$$

where T_s is the temperature of the bottom sediment bed, J_s is the heat flux between the water and sediment bed, ρ_s is the sediment density, C_{sed} is the specific heat capacity of the sediment, and H_{sed} is the effective depth of the sediment bed.

The heat flux from the sediment bed to the water is given by:

$$J_s = \rho_s C_{sed} \frac{\alpha_s}{0.5 H_{sed}} (T_s - T_w) \tag{S.9}$$

Substituting this flux into the heat balance equation:

$$\frac{dT_s}{dt} = -\frac{\rho_s C_{sed} \frac{\alpha_s}{0.5 H_{sed}} (T_s - T_w)}{\rho_s C_{sed} H_{sed}} \tag{S.10}$$

$$\frac{dT_s}{dt} = -\frac{\alpha_s}{0.5 H_{sed}^2} (T_s - T_w) \tag{S.11}$$

S4.4. Summary of the water temperature model

The governing equations for water temperature transport and heat balance in EufoRiA are as follows:

For the transport of water temperature:

$$\frac{\partial A T_w}{\partial t} = \frac{\partial}{\partial x}\left(DA\frac{\partial T_w}{\partial x}\right) - \frac{\partial Q T_w}{\partial x} + \text{source/sink} \tag{S.12}$$

For the heat balance of the control volume of water:

$$\frac{dT_w}{dt} = \frac{1}{\rho C_p H}\Big\{ J_{sn} + \sigma(T_{air}+273.15)^4(A+0.031\sqrt{e_{air}})(1-R_L) - \epsilon\sigma(T_w+273.15)^4 \\ -c_1 f(U_w)(T_w - T_{air}) - f(U_w)(e_s - e_{air}) - \rho_s C_{sed}\frac{\alpha_s}{0.5H_{sed}}(T_w - T_s)\Big\} \tag{S.13}$$

where H is the average hydraulic depth of the control volume (m). In EufoRiA , this is further divided into the depth at junctions and the average depth of links, given by $H = A/B$, where A is the cross-sectional area of the link and B is the link width. To reduce the uncertainty of relevant parameters, we apply $(\rho_s C_{sed}\alpha_s)/(0.5H_{sed}) = 0.146$ in the model.

S5. Dissolved Oxygen

S5.1. Conceptual reaction kinetics

Dissolved oxygen dynamics are described by the schematic in Figure S2.

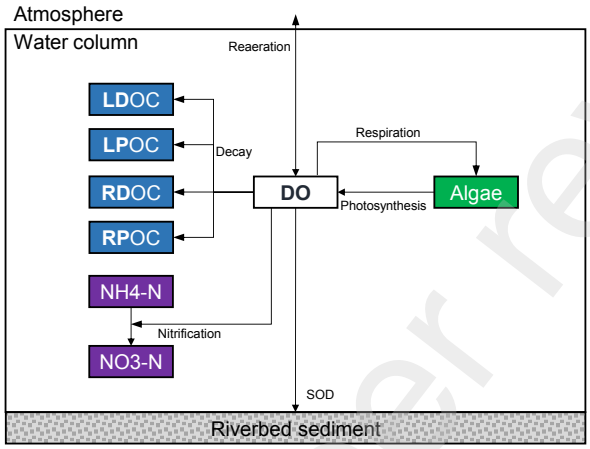


Figure S3: Dissolved oxygen model

Table S11: Comparison of dissolved oxygen kinetics between models

Reaction	EFDC (3-D)	CAEDYM (3-D)	CE-QUAL-W2 (2-D)	QUAL2K (1-D)	EufoRiA (1-D)	Functions [a]	Method
Reaeration(+)	✓	✓	✓	✓	✓	$f(u_{water}, u_{wind}, H)$	1st-order
Algal production(+)	✓	✓	✓	✓	✓	$f(C_{Alg}, N, P, I, T_w)$	Monod
Algal respiration(−)	✓	✓	✓	✓	✓	$f(C_{Alg}, T_w)$	1st-order
Nitrification(−)	✓	✓	✓	✓	✓	$f(C_{\mathrm{NH_4N}}, T_w)$	1st-order
SOD(−) [b]	✓	✓	✓	✓	✓	-	zero-order
Organic carbon decay(−)	✓	✓	✓	-	✓	$f(C_{OC}, T_w)$	1st-order
COD decay(−)	✓	-	-	-	-	$f(C_{COD}, T_w)$	1st-order
CBOD decay(−)	✓	-	✓	✓	-	$f(C_{CBOD}, T_w)$	1st-order
Suspended sediment decay(−)	-	-	-	✓	-	$f(C_{ISS}, T_w)$	1st-order
Epiphyton production(+)	✓	✓	✓	✓	-	$f(C_{Epi}, N, P, I, T_w)$	1st-order
Epiphyton respiration(−)	✓	✓	✓	✓	-	$f(C_{Epi}, T_w)$	1st-order
Macrophyte production(+)	-	✓	✓	-	-	$f(C_{Mac}, N, P, I, T_w)$	1st-order
Macrophyte respiration(−)	-	✓	✓	-	-	$f(C_{Mac}, T_w)$	1st-order
Zooplankton respiration(−)	-	✓	✓	-	-	$f(C_{Zoo}, T_w)$	1st-order

[a] Where, u_{water} is water velocity(m/s), u_{wind} is wind velocity(m/s), T_w is the water temperature, C_k are the concentrations of the constituents name of k, N and P mean the concentrations of the nitrogen and phosphorus groups, and I is the light intensity.
[b] Sediment oxygen demand of the riverbed is applied as zero-order reaction.

S5.2. Governing Equations

S5.2.1. Dissolved oxygen

The dissolved oxygen dynamics are described by:

$$\frac{\partial C_{DO}}{\partial t} = \underbrace{A_{surface} K_L (C_{DO,sat} - C_{DO})}_{reaeration} + \underbrace{\sum_{i=1}^{n} (K_{ag,i} - K_{ar,i}) \delta_{O:Alg,i} C_{Alg,i}}_{\text{Algal production and respiration}} - \underbrace{K_{LDOC} \delta_{O:OC} \gamma_{OC} C_{LDOC}}_{\text{Labile dissolved organic carbon decay}} - \underbrace{K_{LPOC} \delta_{O:OC} \gamma_{OC} C_{LPOC}}_{\text{Labile particulate organic carbon decay}} - \underbrace{K_{RDOC} \delta_{O:OC} \gamma_{OC} C_{RDOC}}_{\text{Refractory dissolved organic carbon decay}} - \underbrace{K_{RPOC} \delta_{O:OC} \gamma_{OC} C_{RPOC}}_{\text{Refractory particulate organic carbon decay}} - \underbrace{K_{NH4} \delta_{O:NH4} \gamma_{NH4} C_{NH4}}_{\text{Nitrification}} - \underbrace{K_{SOD} \gamma_{OC} \frac{A_{bottom}}{V}}_{\text{0-order SOD}} \tag{S.14}$$

S5.2.2. Reaeration coefficient from atmosphere, K_L

The reaeration coefficient can be expressed the sum of hydraulic and wind reaeration coefficient as follows.[18]

$$K_L = K_{L,h} + \frac{K_{L,w}}{H} \tag{S.15}$$

There hydraulic reaeration coefficient $K_{r,h}$([65]) and the wind reaeration coefficient are described by:

- If $H < 0.61m$, use the Owens-Gibbs formula: $k_{ah} = 5.32 \frac{U^{0.67}}{H^{1.85}}$
- If $H > 0.61m$ and $H > 3.45U^{2.5}$, use the O'Connor-Dobbins formula: $k_{ah} = 3.93 \frac{U^{0.5}}{H^{1.5}}$
- Otherwise, use the Churchill formula: $k_{ah} = 5.026 \frac{U}{H^{1.67}}$

In addition, $K_{L,w}$ is described by:

$$K_{L,w} = 0.0986 U_{w,10}^{1.64} \tag{S.16}$$

where H is the hydraulic depth(m) and $U_{w,10}$ is the wind velocity at 10m height.

The saturated oxygen concentration $C_{DO,sat}$ can be calculated by:

$$C_{DO,sat} = \exp\left(-139.34411 + \frac{1.575701 \times 10^5}{T_{abs}} - \frac{6.642308 \times 10^7}{T_{abs}^2} + \frac{1.2438 \times 10^{10}}{T_{abs}^3} - \frac{8.621949 \times 10^{11}}{T_{abs}^4}\right) \tag{S.17}$$

where, $T_{abs} = T_w + 273.15$ is the absolute temperature

S6. Carbon

S6.1. Conceptual reaction kinetics

The carbon cycle is illustrated in Figure S4. In EufoRiA, we ignore the dissolved inorganic carbon (DIC) term in this figure.

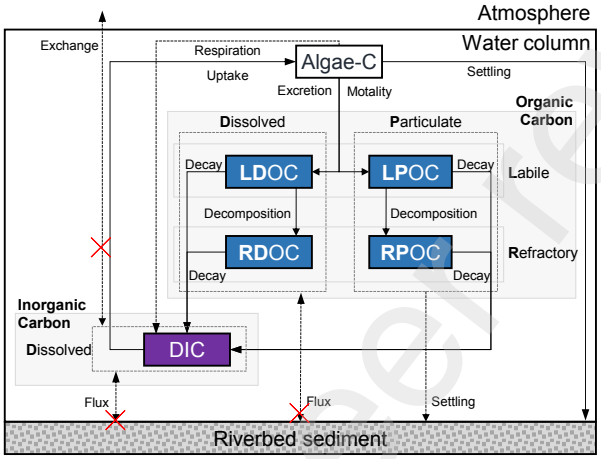


Figure S4: Carbon model of EufoRiA

Table S12: Comparison of carbon reaction kinetics between models - (1)

Reaction	EFDC (3-D)	CAEDYM (3-D)	CE-QUAL-W2 (2-D)	QUAL2K (1-D)	EufoRiA -WQ (1-D)	Functions	Method
LDOC(Labile Dissolved Organic Carbon)							
Algal mortality/excretion(+)	✓	✓	Not applied [a]	Not applied[b]	✓	$f(C_{Alg}, T_w)$	1st-order
Decomposition to RDOC(−)	-	-			✓	$f(C_{LDOC}, T_w)$	1st-order
Decomposition from LPOC(+)	✓	✓			-	$f(C_{LPOC}, T_w)$	1st-order
Decomposition from RPOC(+)	✓	-			-	$f(C_{RPOC}, T_w)$	1st-order
Decay to DIC(−)	✓	✓			✓	$f(C_{LDOC}, T_w)$	1st-order
Riverbed sediment flux(+,−)	-	✓			-	$f(C_{LDOC}, T_w)$	0,1st order
Epiphyton mortality/excretion(+)	✓	✓			-	$f(C_{Epi}, T_w)$	1st-order
Macrophyte mortality/excretion(+)	-	✓			-	$f(C_{Mac}, T_w)$	1st-order
Zooplankton mortality/excretion(+)	-	✓			-	$f(C_{Zoo}, T_w)$	1st-order
LPOC(Labile Particulate Organic Carbon)							
Algal mortality/excretion(+)	✓	✓	Not applied [a]	Not applied[b]	✓	$f(C_{Alg}, T_w)$	1st-order
Decomposition to RPOC(−)	-	-			✓	$f(C_{LPOC}, T_w)$	1st-order
Decomposition to LDOC(−)	✓	✓			-	$f(C_{LPOC}, T_w)$	1st-order
Decay to DIC(−)	-	-			✓	$f(C_{LPOC}, T_w)$	1st-order
Resuspension(+)	-	✓			-	$f(u_{water})$	-
Settling(−)	✓	✓			✓	$f(C_{LPOC})$	1st-order
Epiphyton mortality/excretion(+)	✓	✓			-	$f(C_{Epi}, T_w)$	1st-order
Macrophyte mortality/excretion(+)	-	✓			-	$f(C_{Mac}, T_w)$	1st-order
Zooplankton mortality/excretion(+)	-	✓			-	$f(C_{Zoo}, T_w)$	1st-order

[a] No information is provided. The manual does not describe the reaction kinetics for this constituent.

Table S13: Comparison of carbon reaction kinetics between models - (2)

Reaction	EFDC (3-D)	CAEDYM (3-D)	CE-QUAL-W2 (2-D)	QUAL2K (1-D)	EufoRiA (1-D)	Functions	Method
RDOC(Refractory Dissolved Organic Carbon)							
Algal mortality/excretion(+)	✓	-	Not applied [a]	Not applied[b]	-	$f(C_{Alg}, T_w)$	1st-order
Decomposition from LDOC(+)	-	-			✓	$f(C_{LDOC}, T_w)$	1st-order
Decomposition from RPOC(+)	-	✓			-	$f(C_{RPOC}, T_w)$	1st-order
Decay to DIC(−)	✓	✓			✓	$f(C_{RDOC}, T_w)$	1st-order
Riverbed sediment flux(+,−)	-	✓			-	$f(C_{RDOC}, T_w)$	0,1st order
Epiphyton mortality/excretion(+)	✓	-			-	$f(C_{Epi}, T_w)$	1st-order
RPOC(Refractory Particulate Organic Carbon)							
Algal mortality/excretion(+)	✓	-	Not applied [a]	Not applied[b]	-	$f(C_{Alg}, T_w)$	1st-order
Decomposition from LPOC(+)	-	-			✓	$f(C_{LPOC}, T_w)$	1st-order
Decomposition to RDOC(−)	-	✓			-	$f(C_{RPOC}, T_w)$	1st-order
Decomposition to LDOC(−)	✓	-			-	$f(C_{RPOC}, T_w)$	1st-order
Decay to DIC(−)	-	-			✓	$f(C_{RPOC}, T_w)$	1st-order
Resuspension(+)	-	✓			-	$f(u_{water})$	-
Settling(−)	✓	✓			✓	$f(C_{RPOC})$	1st-order
Epiphyton mortality/excretion(+)	✓	-			-	$f(C_{Epi}, T_w)$	1st-order

[a] Organic carbon in CE-QUAL-W2 is categorized as organic matters, specifically into LDOM, LPOM, RDOM, and RPOM.
[b] QUAL2K simulates total inorganic carbon and CBOD species without differentiating between the labile, refractory, dissolved, and particulate groups.

S6.2. Governing equations

S6.2.1. Organic carbons

Organic carbon (OC) is classified into four distinct groups.

Labile dissolved organic carbon(LDOC) dynamics are described by:

$$\frac{\partial C_{LDOC}}{\partial t} = \underbrace{\sum_{i=1}^{n} P_{d,C}(K_{am,i} + K_{ae,i})\delta_{C:Alg,i} C_{alg,i}}_{\text{algal mortality and excretion}} - \underbrace{K_{LDOC}\gamma_{OC} C_{LDOC}}_{\text{decay to DIC}} - \underbrace{K_{LDOC\rightarrow RDOC}\gamma_{OC} C_{LDOC}}_{\text{labile to refractory decomposition}} \tag{S.18}$$

Labile particulate organic carbon(LPOC) dynamics are described by:

$$\frac{\partial C_{LPOC}}{\partial t} = \underbrace{\sum_{i=1}^{n}(1-P_{d,C})(K_{am,i}+K_{ae,i})\delta_{C:Alg,i}C_{Alg,i}}_{\text{algal mortality and excretion}} - \underbrace{K_{LPOC}\gamma_{OC}C_{LPOC}}_{\text{decay to DIC}} - \underbrace{K_{LPOC\to RPOC}\gamma_{OC}C_{LPOC}}_{\text{labile to refractory decomposition}} - \underbrace{\frac{K_{s,LPOC}}{H}C_{LPOC}}_{\text{settling}} \tag{S.19}$$

Refractory dissolved organic carbon(RDOC) dynamics are described by:

$$\frac{\partial C_{RDOC}}{\partial t} = \underbrace{K_{LDOC\to RDOC}\gamma_{OC}C_{LDOC}}_{\text{labile to refractory decomposition}} - \underbrace{K_{RDOC}\gamma_{OC}C_{RDOC}}_{\text{decay to DIC}} \tag{S.20}$$

Refractory particulate organic carbon(RPOC) dynamics are described by:

$$\frac{\partial C_{RPOC}}{\partial t} = \underbrace{K_{LPOC\to RPOC}\gamma_{OC}C_{LPOC}}_{\text{labile to refractory decomposition}} - \underbrace{K_{RPOC}\gamma_{OC}C_{RPOC}}_{\text{decay to carbon}} - \underbrace{\frac{K_{s,RPOC}}{H}C_{RPOC}}_{\text{settling}} \tag{S.21}$$

Total organic carbon(C_{TOC}) can be calculated using the following equation.

$$C_{TOC} = C_{LDOC} + C_{LPOC} + C_{RDOC} + C_{RPOC} + \sum \delta_{C:Alg,i}C_{Alg,i} \tag{S.22}$$

The particulate and dissolved inorganic carbon are ignored in EufoRiA, as inorganic carbon is not a limiting factor for algal growth in most models.

S6.2.2. Temperature multipliers

The temperature multiplier γ_{OC} and γ_{NH4} can be calculated using the temperature coefficients K_1, K_2, T_1, T_2 for each constituent. The same expression

will be applied for other constituents including γ_{ON}, γ_{OP}, γ_{PO4}, and etc.

$$\lambda_T = \frac{K_1 e^{\gamma_1 (T_w - T_1)}}{1 + K_1 e^{\gamma_1 (T_w - T_1)} - K_1} \tag{S.23}$$

where, $\gamma_1 = \frac{1}{T_2 - T_1} ln \frac{K_2(1-K_1)}{K_1(1-K_2)}$

S7. Nitrogen

S7.1. Conceptual reaction kinetics

The nitrogen cycle is illustrated in Figure S5.

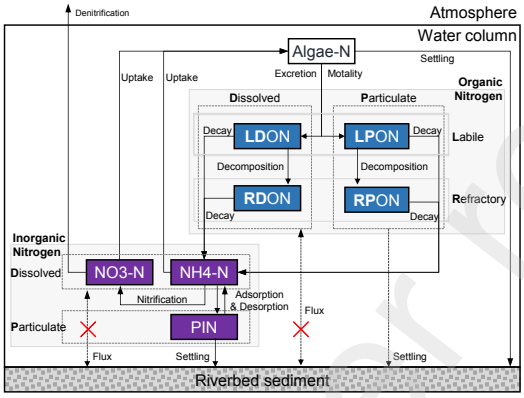


Figure S5: Nitrogen model of EufoRiA

Table S14: Comparison of nitrogen reaction kinetics between models - (1)

Reaction	EFDC (3-D)	CAEDYM (3-D)	CE-QUAL-W2 (2-D)	QUAL2K (1-D)	EufoRiA (1-D)	Functions	Method
NH4-N(Ammonium)							
Algal production(+)	✓	✓	✓	✓	✓	$f(C_{Alg}, N, P, I, T_w)$	Monod
Algal respiration(−)	✓	✓	✓	✓	✓	$f(C_{Alg}, T_w)$	1st-order
Nitrification(−)	✓	✓	✓	✓	✓	$f(C_{NH4}, T_w)$	1st-order
Decay from POM(+)	-	-	✓	-	✓	$f(C_{POM}, T_w)$	1st-order
Decay from DOM(+)	✓	✓	✓	✓	✓	$f(C_{DOM}, T_w)$	1st-order
Riverbed sediment release(+)	✓	✓	✓	✓	-	-	zero-order
Riverbed sediment absorb(−)	✓	✓	-	-	-	$f(C_{NH4})$	1st-order
CBOD decay(+)	-	-	✓	-	-	$f(C_{CBOD}, T_w)$	1st-order
Desorption from PIN	-	✓	✓	-	-	-	zero-order
Adsorption to PIN(−)	-	✓	-	-	-	$f(C_{NH4})$	1st-order
Epiphyton production(+)	✓	✓	✓	✓	-	$f(C_{Epi}, C_{NH4}, T_w)$	Monod
Epiphyton respiration(−)	✓	✓	✓	✓	-	$f(C_{Epi}, T_w)$	1st-order
Macrophyte production(+)	-	✓	✓	-	-	$f(C_{Mac}, C_{NH4}, T_w)$	Monod
Macrophyte respiration(−)	-	✓	-	-	-	$f(C_{Mac}, T_w)$	1st-order
Zooplankton respiration(−)	-	✓	✓	-	-	$f(C_{Zoo}, T_w)$	1st-order
PIN(Particulate Inorganic Nitrogen)							
Adsorption from NH4(+)	✓	✓	Not applied [a]	Not applied [a]	✓	$f(C_{NH4}, C_{ISS})$	Monod
Desorption to NH4(−)	✓	✓			✓	$f(C_{PIN})$	1st-order
Settling(−)	✓	✓			✓	$f(C_{PIN})$	1st-order
Resuspension(+)	✓	✓			-	$f(u_{water})$	-

[a] No information is provided. The manual does not describe the reaction kinetics for this constituent.

Table S15: Comparison of nitrogen reaction kinetics between models - (2)

Reaction	EFDC (3-D)	CAEDYM (3-D)	CE-QUAL-W2 (2-D)	QUAL2K (1-D)	EufoRiA (1-D)	Functions	Method
NO3+NO2(Nitrate)							
Nitrification(+)	✓	✓	✓	✓	✓	$f(C_{NH4}, T_w)$	1st-order
Algal production(−)	✓	✓	✓	✓	✓	$f(C_{Alg}, N, P, I, T_w)$	1st-order
Denitrification(-)	✓	✓	✓	✓	✓	$f(C_{NO3}, T_w)$	1st-order
Riverbed sediment release(+)	✓	✓	-	-	-	-	zero-order
Riverbed sediment absorb(−)	✓	✓	-	-	✓	$f(C_{NO3})$	1st-order
Epiphyton production(−)	✓	✓	✓	✓	-	$f(C_{Epi}, C_{NO3}, T_w)$	1st-order
LDON(Labile Dissolved Organic Nitrogen)							
Algal mortality/excretion(+)	✓	✓	✓	Not applied [a]	✓	$f(C_{Alg}, T_w)$	1st-order
Decomposition to RDON(−)	-	-	✓		✓	$f(C_{LDON}, T_w)$	1st-order
Decomposition from LPON(+)	✓	✓	-		-	$f(C_{LPON}, T_w)$	1st-order
Decomposition from RPON(+)	✓	-	-		-	$f(C_{RPON}, T_w)$	1st-order
Decay to NH4(−)	✓	✓	✓		✓	$f(C_{LDON}, T_w)$	1st-order
Riverbed sediment flux(+,−)	-	✓	-		-	$f(C_{LDON}, T_w)$	0,1st order
Epiphyton mortality/excretion(+)	✓	✓	✓		-	$f(C_{Epi}, T_w)$	1st-order
Macrophyte mortality/excretion(+)	-	✓	✓		-	$f(C_{Mac}, T_w)$	1st-order
Zooplankton mortality/excretion(+)	-	✓	-		-	$f(C_{Zoo}, T_w)$	1st-order
LPON(Labile Particulate Organic Nitrogen)							
Algal mortality/excretion(+)	✓	✓	✓	Not applied[a]	✓	$f(C_{Alg}, T_w)$	1st-order
Decomposition to RPON(−)	-	-	✓		✓	$f(C_{LPON}, T_w)$	1st-order
Decomposition to LDON(−)	✓	✓	-		-	$f(C_{LPON}, T_w)$	1st-order
Decay to NH4(−)	-	-	✓		✓	$f(C_{LPON}, T_w)$	1st-order
Resuspension(+)	-	✓	-		-	$f(u_{water})$	-
Settling(−)	✓	✓	✓		✓	$f(C_{LPON})$	1st-order
Epiphyton mortality/excretion(+)	✓	✓	✓		-	$f(C_{Epi}, T_w)$	1st-order
Macrophyte mortality/excretion(+)	-	✓	✓		-	$f(C_{Mac}, T_w)$	1st-order
Zooplankton mortality/excretion(+)	-	✓	✓		-	$f(C_{Zoo}, T_w)$	1st-order
RDON(Refractory Dissolved Organic Nitrogen)							
Algal mortality/excretion(+)	✓	-	-	Not applied[a]	-	$f(C_{Alg}, T_w)$	1st-order
Decomposition from LDON(+)	-	-	✓		✓	$f(C_{LDON}, T_w)$	1st-order
Decomposition from RPON(+)	-	✓	-		-	$f(C_{RPON}, T_w)$	1st-order
Decay to NH4(−)	✓	✓	✓		✓	$f(C_{RDON}, T_w)$	1st-order
Riverbed sediment flux(+,−)	-	✓	-		-	$f(C_{RDON}, T_w)$	0,1st order
Epiphyton mortality/excretion(+)	✓	-	-		-	$f(C_{Epi}, T_w)$	1st-order
RPON(Refractory Particulate Organic Nitrogen)							
Algal mortality/excretion(+)	✓	-	-	Not applied[a]	-	$f(C_{Alg}, T_w)$	1st-order
Decomposition from LPON(+)	-	-	✓		✓	$f(C_{LPON}, T_w)$	1st-order
Decomposition to RDON(−)	-	✓	-		-	$f(C_{RPON}, T_w)$	1st-order
Decomposition to LDON(−)	✓	-	-		-	$f(C_{RPON}, T_w)$	1st-order
Decay to NH4(−)	-	-	✓		✓	$f(C_{RPON}, T_w)$	1st-order
Resuspension(+)	-	✓	-		-	$f(u_{water})$	-
Settling(−)	✓	✓	✓		✓	$f(C_{RPON})$	1st-order
Epiphyton mortality/excretion(+)	✓	-	-		-	$f(C_{Epi}, T_w)$	1st-order
Macrophyte mortality/excretion(+)	-	-	✓		-	$f(C_{Mac}, T_w)$	1st-order

[a] QUAL2K simulates the organic nitrogen without distinguishing the labile, refractory, dissolved, and particulate groups.

S7.2. Governing equations

S7.2.1. Ammonium ($NH_4 - N$)

Ammonium reactions are described by:

$$\frac{\partial C_{NH4}}{\partial t} = \underbrace{\sum_{i=1}^{n} P_{NH4} K_{ag,i} \delta_{N:Alg,i} C_{alg,i}}_{\text{Algal respiration}} - \underbrace{\sum_{i=1}^{n} P_{NH4} K_{ag,i} \delta_{N:Alg,i} C_{alg,i}}_{\text{Algal production}} + \underbrace{K_{LDON} \delta_{n:on} \gamma_{ON} C_{LDON}}_{\text{Labile dissolved organic nitrogen decay}} + \underbrace{K_{LPON,i} \delta_{N:ON} \gamma_{ON} C_{LPON}}_{\text{Labile particulate organic nitrogen decay}} + \underbrace{K_{RDON,i} \delta_{N:ON} \gamma_{ON} C_{RDON}}_{\text{Refractory dissolved organic nitrogen decay}} + \underbrace{K_{RPON} \delta_{N:ON} \gamma_{ON} C_{RPON}}_{\text{Refractory particulate organic nitrogen decay}} - \underbrace{K_{NH4} \gamma_{NH4} C_{NH4}}_{\text{Nitrification}} \tag{S.24}$$

S7.2.2. Particulate inorganic nitrogen (PIN)

Particulate inorganic nitrogen dynamics are described by:

$$\frac{\partial C_{PIN}}{\partial t} = K_{ads,PIN} \left(\frac{C_{ISS}}{K_{h,PIN} + C_{ISS}} \right) C_{NH4} - K_{des,PIN} C_{PIN} - \frac{K_{s,PIN}}{H} C_{PIN} \tag{S.25}$$

S7.2.3. Nitrate ($NO_3 - N$)

Nitrate (NO2 + NO3) reactions dynamics are described by:

$$\frac{\partial C_{NO3}}{\partial t} = \underbrace{-\sum_{i=1}^{n} (1 - P_{NH4}) K_{ag,i} \delta_{n:alg,i} C_{alg,i}}_{\text{Algal production}} + \underbrace{K_{NH4} \gamma_{NH4} C_{NH4}}_{\text{Nitrification}} - \underbrace{K_{NO3} \gamma_{NO3} C_{NO3}}_{\text{Denitrification}} - \underbrace{\frac{K_{s,NO3}}{H} C_{NO3}}_{\text{Settling}} \tag{S.26}$$

S7.2.4. Organic nitrogen

Labile dissolved organic nitrogen (LDON) dynamics are described by:

$$\frac{\partial C_{LDON}}{\partial t} = \underbrace{\sum_{i=1}^{n} P_{d,N}(K_{am,i} + K_{ae,i})\delta_{N:Alg,i}C_{Alg,i}}_{\text{Algal mortality and excretion}} - \underbrace{K_{LDON}\gamma_{ON}C_{LDON}}_{\text{Decay to NH4}} - \underbrace{K_{LDON\rightarrow RDON}\gamma_{ON}C_{LDON}}_{\text{Labile to refractory decomposition}} \tag{S.27}$$

Labile particulate organic nitrogen (LPON) dynamics are described by:

$$\frac{\partial C_{LPON}}{\partial t} = \underbrace{\sum_{i=1}^{n}(1 - P_{d,N})(K_{am,i} + K_{ae,i})\delta_{N:Alg,i}C_{Alg,i}}_{\text{Algal mortality and excretion}} - \underbrace{K_{LPON}\gamma_{ON}C_{LPON}}_{\text{Decay to NH4}} - \underbrace{K_{LPON\rightarrow RPON}\gamma_{ON}C_{LPON}}_{\text{Labile to refractory decomposition}} - \underbrace{\frac{K_{s,LPON}}{H}C_{LPON}}_{\text{Settling}} \tag{S.28}$$

Refractory dissolved organic nitrogen (RDON) dynamics are described by:

$$\frac{\partial C_{RDON}}{\partial t} = \underbrace{K_{LDON\rightarrow RDON}\gamma_{ON}C_{LDON}}_{\text{Labile to refractory decomposition}} - \underbrace{K_{RDON}\gamma_{ON}C_{RDON}}_{\text{Decay to NH4}} \tag{S.29}$$

Refractory particulate organic nitrogen (RPON) dynamics are described by:

$$\frac{\partial C_{RPON}}{\partial t} = \underbrace{K_{LPON\rightarrow RPON}\gamma_{ON}C_{LPON}}_{\text{Labile to refractory decomposition}} - \underbrace{K_{RPON}\gamma_{ON}C_{RPON}}_{\text{Decay to NH4}} - \underbrace{\frac{K_{s,RPON}}{H}C_{RPON}}_{\text{Settling}} \tag{S.30}$$

Total nitrogen concentration (C_{TN}) is calculated by:

$$C_{TN} = C_{NH4} + C_{NO3} + C_{LDON} + C_{LPON} + C_{RDON} + C_{RPON} + \sum \delta_{N:Alg,i} C_{Alg,i} \tag{S.31}$$

S8. Phosphorus

S8.1. Conceptual reaction kinetics

The phosphorus cycle is illustrated in Figure S6.

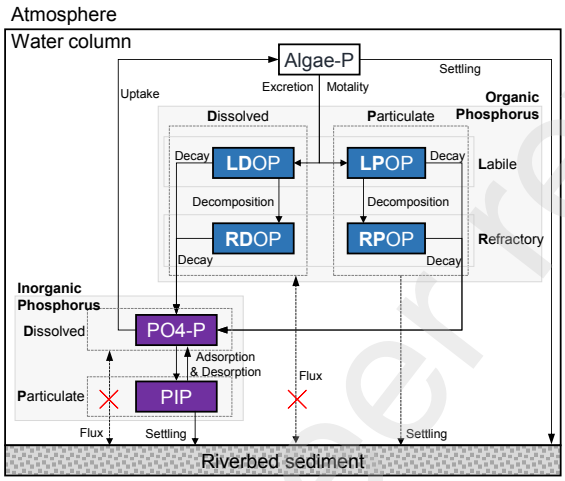


Figure S6: Phosphorus model of EufoRiA

Table S16: Comparison of phosphorus reaction kinetics between models - (1)

Reaction	EFDC (3-D)	CAEDYM (3-D)	CE-QUAL-W2 (2-D)	QUAL2K (1-D)	EufoRiA (1-D)	Functions	Method
PO4-P(Phosphate)							
Algal production(+)	✓	✓	✓	✓	✓	$f(C_{Alg}, N, P, I, T_w)$	Monod
Algal respiration(−)	✓	-	✓	✓	✓	$f(C_{Alg}, T_w)$	1st-order
Decay from POM(+)	-	-	✓	-	✓	$f(C_{POM}, T_w)$	1st-order
Decay from DOM(+)	✓	✓	✓	✓	✓	$f(C_{DOM}, T_w)$	1st-order
Riverbed sediment release(+)	✓	✓	✓	-	-	-	zero-order
Riverbed sediment absorb(−)	✓	✓	-	✓	-	$f(C_{NH4})$	1st-order
CBOD decay(+)	-	-	✓	-	-	$f(C_{CBOD}, T_w)$	1st-order
Suspended sediment release(+)	-	✓	✓	-	-	-	zero-order
Suspended sediment adsorption(−)	-	✓	✓	-	-	$f(C_{NH4})$	1st-order
Epiphyton production(+)	✓	✓	✓	✓	-	$f(C_{Epi}, C_{NH4}, T_w)$	Monod
Epiphyton respiration(−)	-	-	✓	✓	-	$f(C_{Epi}, T_w)$	1st-order
Macrophyte production(+)	-	✓	✓	-	-	$f(C_{Mac}, C_{NH4}, T_w)$	Monod
Macrophyte respiration(−)	-	-	✓	-	-	$f(C_{Mac}, T_w)$	1st-order
Zooplankton respiration(−)	-	-	✓	-	-	$f(C_{Zoo}, T_w)$	1st-order

[a] No information is provided. The manual does not describe the reaction kinetics for this constituent.

Table S17: Comparison of phosphorus reaction kinetics between models - (2)

Reaction	EFDC (3-D)	CAEDYM (3-D)	CE-QUAL-W2 (2-D)	QUAL2K (1-D)	EufoRiA (1-D)	Functions	Method
PIP(Particulate Inorganic Phosphorus)							
Adsorption from PO4(+)	✓	✓	✓	Not applied [a]	✓	$f(C_{PO4}, C_{ISS})$	Monod
Desorption to PO4(−)	✓	✓	✓		✓	$f(C_{PIP})$	1st-order
Settling(−)	✓	✓	✓		✓	$f(C_{PIP})$	1st-order
Resuspension(+)	✓	✓	-		-	$f(u_{water})$	-
LDOP(Labile Dissolved Organic Phosphorus)							
Algal mortality/excretion(+)	✓	✓	✓	Not applied [a]	✓	$f(C_{Alg}, T_w)$	1st-order
Decomposition to RDOP(−)	-	-	✓		✓	$f(C_{LDOP}, T_w)$	1st-order
Decomposition from LPOP(+)	✓	✓	-		-	$f(C_{LPOP}, T_w)$	1st-order
Decomposition from RPOP(+)	✓	-	-		-	$f(C_{RPOP}, T_w)$	1st-order
Decay to PO4(−)	✓	✓	✓		✓	$f(C_{LDOP}, T_w)$	1st-order
Riverbed sediment flux(+,−)	-	✓	-		-	$f(C_{LDOP}, T_w)$	0,1st order
Epiphyton mortality/excretion(+)	✓	✓	✓		-	$f(C_{Epi}, T_w)$	1st-order
Macrophyte mortality/excretion(+)	-	✓	✓		-	$f(C_{Mac}, T_w)$	1st-order
Zooplankton mortality/excretion(+)	-	✓	-		-	$f(C_{Zoo}, T_w)$	1st-order
LPOP(Labile Particulate Organic Phosphorus)							
Algal mortality/excretion(+)	✓	✓	✓	Not applied[a]	✓	$f(C_{Alg}, T_w)$	1st-order
Decomposition to RPOP(−)	-	-	✓		✓	$f(C_{LPOP}, T_w)$	1st-order
Decomposition to LDOP(−)	✓	✓	-		-	$f(C_{LPOP}, T_w)$	1st-order
Decay to PO4(−)	-	-	✓		✓	$f(C_{LPOP}, T_w)$	1st-order
Resuspension(+)	-	✓	-		-	$f(u_{water})$	-
Settling(−)	✓	✓	✓		✓	$f(C_{LPOP})$	1st-order
Zooplankton predation(−)	-	✓	✓		-	$f(C_{LPOP})$	1st-order
Epiphyton mortality/excretion(+)	✓	✓	✓		-	$f(C_{Epi}, T_w)$	1st-order
Macrophyte mortality/excretion(+)	-	✓	✓		-	$f(C_{Mac}, T_w)$	1st-order
Zooplankton mortality/excretion(+)	-	✓	✓		-	$f(C_{Zoo}, T_w)$	1st-order
RDOP(Refractory Dissolved Organic Phosphorus)							
Algal mortality/excretion(+)	✓	-	-	Not applied[a]	-	$f(C_{Alg}, T_w)$	1st-order
Decomposition from LDOP(+)	-	-	✓		✓	$f(C_{LDOP}, T_w)$	1st-order
Decomposition from RPOP(+)	-	✓	-		-	$f(C_{RPOP}, T_w)$	1st-order
Decay to PO4(−)	✓	✓	✓		✓	$f(C_{RDOP}, T_w)$	1st-order
Riverbed sediment flux(+,−)	-	✓	-		-	$f(C_{RDOP}, T_w)$	0,1st order
Epiphyton mortality/excretion(+)	✓	-	-		-	$f(C_{Epi}, T_w)$	1st-order
RPOP(Refractory Particulate Organic Phosphorus)							
Algal mortality/excretion(+)	✓	-	-	Not applied[a]	-	$f(C_{Alg}, T_w)$	1st-order
Decomposition from LPOP(+)	-	-	✓		✓	$f(C_{LPOP}, T_w)$	1st-order
Decomposition to RDOP(−)	-	✓	-		-	$f(C_{RPOP}, T_w)$	1st-order
Decomposition to LDOP(−)	✓	-	-		-	$f(C_{RPOP}, T_w)$	1st-order
Decay to PO4(−)	-	-	✓		✓	$f(C_{RPOP}, T_w)$	1st-order
Resuspension(+)	-	✓	-		-	$f(u_{water})$	-
Settling(−)	✓	✓	✓		✓	$f(C_{RPOP})$	1st-order
Epiphyton mortality/excretion(+)	✓	-	-		-	$f(C_{Epi}, T_w)$	1st-order
Macrophyte mortality/excretion(+)	-	-	✓		-	$f(C_{Mac}, T_w)$	1st-order

[a] QUAL2K simulates the organic phosphorus without distinguishing the labile, refractory, dissolved, and particulate groups.

S8.2. Governing equations

S8.2.1. Phosphate ($PO_4 - P$)

Phosphate reaction dynamics are described by:

$$\frac{\partial C_{PO4}}{\partial t} = \underbrace{\sum_{i=1}^{n} K_{ar,i}\delta_{P:Alg,i}C_{Alg,i}}_{\text{Algal respiration}} - \underbrace{\sum_{i=1}^{n} K_{ag,i}\delta_{P:Alg,i}C_{Alg,i}}_{\text{Algal production}} - \underbrace{K_{ads,PIP}\left(\frac{C_{ISS}}{K_{h,ads}+C_{ISS}}\right)C_{PO4}}_{\text{Adsorption to ISS}}$$
$$+ \underbrace{K_{des,PIP}C_{PIP}}_{\text{Desorption from ISS}} + \underbrace{K_{LDOP}\delta_{P:OP}\gamma_{OP}C_{LDOP}}_{\text{Labile dissolved organic phosphorus decay}} + \underbrace{K_{LPOP,i}\delta_{P:OP}\gamma_{OP}C_{LPOP}}_{\text{Labile particulate organic phosphorus decay}}$$
$$+ \underbrace{K_{RDOP,i}\delta_{P:OP}\gamma_{OP}C_{RDOP}}_{\text{Refractory dissolved organic phosphorus decay}} + \underbrace{K_{RPOP}\delta_{P:OP}\gamma_{OP}C_{RPOP}}_{\text{Refractory particulate organic phosphorus decay}} \tag{S.32}$$

S8.2.2. Particulate inorganic phosphorus (PIP)

Particulate inorganic phosphorus dynamics are described by:

$$\frac{\partial C_{PIP}}{\partial t} = K_{ads,PIP}\left(\frac{C_{ISS}}{K_{h,PIP}+C_{ISS}}\right)C_{PO4} - K_{des,PIP}C_{PIP} - \frac{K_{s,PIP}}{H}C_{PIP} \tag{S.33}$$

S8.2.3. organic phosphorus

Labile dissolved organic phosphorus(LDOP) dynamics are described by:

$$\frac{\partial C_{LDOP}}{\partial t} = \underbrace{\sum_{i=1}^{n} P_{d,P}(K_{am,i}+K_{ae,i})\delta_{P:Alg,i}C_{Alg,i}}_{\text{Algal mortality and excretion}} - \underbrace{K_{LDOP}\gamma_{OP}C_{LDOP}}_{\text{Decay to PO4}} - \underbrace{K_{LDOP\rightarrow RDOP}\gamma_{OP}C_{LDOP}}_{\text{Labile to refractory decomposition}} \tag{S.34}$$

Labile particulate organic phosphorus (LPOP) dynamics are described by:

$$\frac{\partial C_{LPOP}}{\partial t} = \underbrace{\sum_{i=1}^{n}(1-P_{d,P})(K_{am,i}+K_{ae,i})\delta_{OP:Alg,i}C_{Alg,i}}_{\text{Algal mortality and excretion}} - \underbrace{K_{LPOP}\gamma_{OP}C_{LPOP}}_{\text{Decay to PO4}}$$
$$- \underbrace{K_{LPOP\rightarrow RPOP}\gamma_{OP}C_{LPOP}}_{\text{Labile to refractory decomposition}} - \underbrace{\frac{K_{s,LPOP}}{H}C_{LPOP}}_{\text{Settling}} \quad \text{(S.35)}$$

Refractory dissolved organic phosphorus (RDOP) dynamics are described by:

$$\frac{\partial C_{RDOP}}{\partial t} = \underbrace{K_{LDOP\rightarrow RDOP}\gamma_{OP}C_{LDOP}}_{\text{Labile to refractory decomposition}} - \underbrace{K_{RDOP}\gamma_{OP}C_{RDOP}}_{\text{Decay to PO4}} \quad \text{(S.36)}$$

Refractory particulate organic phosphorus (RPOP) dynamics are described by:

$$\frac{\partial C_{RPOP}}{\partial t} = \underbrace{K_{LPOP\rightarrow RPOP}\gamma_{OP}C_{LPOP}}_{\text{Labile to refractory decomposition}} - \underbrace{K_{RPOP}\gamma_{OP}C_{RPOP}}_{\text{Decay to PO4}} - \underbrace{\frac{K_{s,RPOP}}{H}C_{RPOP}}_{\text{Settling}} \quad \text{(S.37)}$$

Total phosphorus concentration (C_{TP}) can be calculated using the following equation.

$$C_{TP} = C_{PO4} + C_{LDOP} + C_{LPOP} + C_{RDOP} + C_{RPOP} + \sum \delta_{P:Alg,i}C_{Alg,i} \quad \text{(S.38)}$$

S9. Algae (Phytoplankton)

S9.1. Conceptual reaction kinetics

Figure S7 shows the algal growth kinetics in EufoRiA .

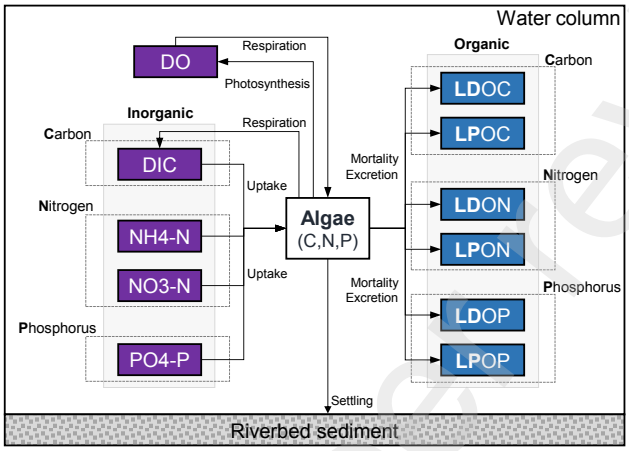


Figure S7: Algae model

Table S18: Comparison of algal reaction kinetics between models

Reaction	EFDC[a] (3-D)	CAEDYM[b] (3-D)	CE-QUAL-W2 (2-D)	QUAL2K[b] (1-D)	EufoRiA (1-D)	Functions	Method
Species	4 groups	7 groups	N groups	1 group	N groups	-	-
Production(+)	✓	✓	✓	✓	✓	$f(C_{Alg}, N, P, I, T_w)$	Monod
Respiration(−)	✓	✓	✓	✓	✓	$f(C_{Alg}, T_w)$	1st-order
Mortality(−)	✓	✓	✓	✓	✓	$f(C_{Alg}, T_w)$	1st-order
Excretion(−)	✓	✓	✓	✓	✓	$f(C_{Alg}, T_w)$	1st-order
Settling(−)	✓	✓	✓	✓	✓	$f(C_{Alg}, T_w)$	1st-order
Resuspension(+)	-	✓	-	-	-	$f(u_{water})$	1st-order
Predation(−)	✓	✓	✓	-	-	$f(C_{Zoo}, C_{Alg}, C_{POM})$	Monod

[a] EFDC considers respiration, mortality, and excretion as a basal metabolism
[b] CAEDYM and QUAL2K do not separate mortality and excretion

S9.2. Governing equations

In EufoRiA, algal growth, respiration, excretion, mortality, and settling are modeled by the following equation.

$$\frac{\partial C_{Alg,i}}{\partial t} = \underbrace{(K_{ag,i} - K_{ar,i} - K_{ae,i} - K_{am,i})C_{Alg,i}}_{\text{Algal growth, respiration, excretion, and mortality}} - \underbrace{\frac{K_{s,Alg,i}}{H}C_{Alg,i}}_{Settling} \quad \text{(S.39)}$$

S9.2.1. Algal growth rate

Algal growth rate is calculated using limitations of nutrients and light with temperature multipliers.

$$K_{ag,i} = \gamma_{ar,i}\gamma_{af,i}\lambda_{min}K_{ag,max,i} \tag{S.40}$$

where, $K_{ag,i}$ is algal growth rate for species i $(1/day)$, $K_{ag,max,i}$ is maximum algal growth rate for species i $(1/day)$, $\gamma_{ar,i}$ is temperature multiplier for rising curve for species i, $\gamma_{af,i}$ is temperature multiplier for falling curve for species i, and λ_{min} is minimum limitation factor.

S9.2.2. Temperature multiplier

We adopt the same temperature multiplier model as CE-QUAL-W2 [66]. This expression provides the capability to set the optimal temperature range for specific algal groups.

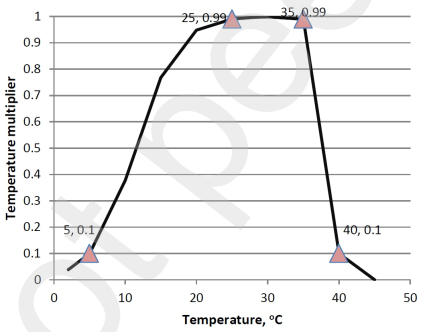


Figure S8: Example of the temperature multiplier function (source: [23])

$$\lambda_T = 0 \qquad \text{where} \quad T_w \leq T_1$$

$$\lambda_T = \underbrace{\frac{K_1 e^{\gamma_1(T_w - T_1)}}{1 + K_1 e^{\gamma_1(T_w - T_1)} - K_1}}_{\gamma_{ar}} \underbrace{\frac{K_4 e^{\gamma_2(T_4 - T_w)}}{1 + K_4 e^{\gamma_2(T_4 - T_w)} - K_4}}_{\gamma_{af}} \qquad \text{where} \quad T_1 < T_w < T_4$$

$$\lambda_{T_w} = 0 \qquad \text{where} \quad T_w \geq T_4 \tag{S.41}$$

where, $\gamma_1 = \frac{1}{T_2 - T_1} ln \frac{K_2(1-K_1)}{K_1(1-K_2)}$, $\gamma_2 = \frac{1}{T_4 - T_3} ln \frac{K_3(1-K_4)}{K_4(1-K_3)}$

S9.2.3. Nutrient and light limitation factor

The nutrient and light limitation factor (λ_{min}) can be calculated using two different expressions as follows.

$$\lambda_{min} = \min\left(\frac{C_{NH4} + C_{NO3}}{K_{h,N} + C_{NH4} + C_{NO3}}, \frac{C_{PO4}}{K_{h,P} + C_{PO4}}, \lambda_{light}\right) \tag{S.42}$$

Alternatively,

$$\lambda_{min} = \min\left(\frac{C_{NH4} + C_{NO3}}{K_{h,N} + C_{NH4} + C_{NO3}}, \frac{C_{PO4}}{K_{h,P} + C_{PO4}}\right)\lambda_{light} \tag{S.43}$$

where, $P_{h,n}$ is half-saturation coefficient for nitrogen, $P_{h,p}$ is half-saturation coefficient for phosphorus. The light attenuation factor λ_{light} is defined using Steele's equation (1962) as follows.

$$\lambda_{light} = \frac{2.7182}{k_e H}\left(e^{-\frac{PAR(0)}{K_{Lp}}} e^{k_e H} - e^{-\frac{PAR(0)}{K_{Lp}}}\right) \tag{S.44}$$

where, $PAR(0) = 0.47 I_0$, I_0 is solar radiation at the water surface W/m^2, K_{Lp} is the PAR at which algal growth is optimal.

The light extinction coefficient k_e is defined as follows.

$$k_e = k_{eb} + \alpha_{ISS} C_{ISS} + \alpha_{POM} C_{POM} + \alpha_{Alg} \sum_{i}^{n} C_{Alg,i} \tag{S.45}$$

where k_{eb} is the background light extinction coefficient ($1/m$), α_{ISS} is the light extinction due to inorganic suspended solids ($m^3/m/g$), α_{POM} is the light extinction due to particulate organic matters ($m^3/m/g$), and α_{Alg} is the light extinction due to algal group concentrations ($m^3/m/g$).

The ammonium preference factor P_{NH4} is calculated as follows [67, 23]:

$$\begin{aligned} P_{NH4} = C_{NH4}\frac{C_{NO3}}{(K_{h,NH4} + C_{NH4})(K_{h,NH4} + C_{NO3})} \\ + C_{NH4}\frac{K_{h,NH4}}{(C_{NH4} + C_{NO3})(K_{h,NH4} + C_{NO3})} \end{aligned} \tag{S.46}$$

S9.2.4. Other algal rates

Algal respiration, excretion, and mortality rates are calculated using the following expressions.

$$\begin{aligned} K_{ar,i} &= \lambda_T K_{ar,max,i} \\ K_{ae,i} &= (1-\lambda_{light})\lambda_T K_{ae,max,i} \\ K_{am,i} &= \lambda_T K_{am,max,i} \end{aligned} \tag{S.47}$$

S10. Inorganic suspended solids

Each model deals with the suspended solids or sediments differently. It is difficult to distinguish between suspended solids and sediments in many cases. In EufoRiA we consider that the inorganic suspended solids are equal to suspended sediments. We apply the settling of suspended solids as follows.

$$\frac{\partial C_{ISS}}{\partial t} = -\frac{K_{s,ISS}}{H} C_{ISS} \tag{S.48}$$

Additionally, we consider the adsorption/release of phosphate and ammonium in the water column using two nutrient variables: particulate inorganic nitrogen (PIN) and particulate inorganic phosphorus (PIP). The equations can be found in the previous chapters.

S11. Representing hydraulic structures and controls

EufoRiA provides the capability to model in-stream structures like weirs, which significantly impact the flow regime of rivers both spatially and temporally. Due to the retention effect of these weirs, accurately representing them is crucial for simulating water quality constituents and phenomena like harmful algal blooms in impounded and regulated river network models. In EufoRiA, we not only represent these in-stream hydraulic structures but also control them based on the opening ratio at each time step between the bottom and maximum operational elevations, to reproduce the observed water levels—used as the reference value for control—at each storage weir. To achieve reliable control of weirs in the model, we implement Proportional-Integral-Derivative (PID) control method as follows:

$$u(t) = K_p e(t) + K_i \int_0^t e(t)dt + K_d \frac{de(t)}{dt} \tag{S.49}$$

Where, $u(t)$ is the control value (e.g., weir opening ratio–from 0 (close) - 1 (fully open)– for each storage weir), K_p, K_i, and K_d are the proportional, integral, and derivative gain respectively, $e(t) = h_{model}(t) - h_{obs}(t)$ is the error between simulated water level at the weir location (h_{model}) and the observed water level at the weir location (h_{obs}).

To derive the incremental PID control, we differentiate both sides of Equation (S.49) by time:

$$\frac{du(t)}{dt} = K_p \frac{de(t)}{dt} + K_i e(t) + K_d \frac{d^2 e(t)}{dt^2} \tag{S.50}$$

Applying the Backward scheme:

$$\begin{aligned} \frac{u(t_k) - u(t_{k-1})}{\Delta t} = K_p \left(\frac{e(t_k) - e(t_{k-1})}{\Delta t} \right) + K_i e(t_k) \\ + K_d \left(\frac{\frac{e(t_k) - e(t_{k-1})}{\Delta t} - \frac{e(t_{k-1}) - e(t_{k-2})}{\Delta t^*}}{\Delta t} \right) \end{aligned} \tag{S.51}$$

Where, Δt is the time interval between the time k and $k-1$ (k is the current time step.), Δt^* is the time interval between the time $k-1$ and $k-2$, $e(t_k)$, $e(t_{k-1})$, and $e(t_{k-2})$ are the error between the observed water level and simulated water level at the time k, $k-1$, and $k-2$, respectively.

By rearranging Equation (S.51), we obtain the following incremental PID control expression:

$$u(t_k) = u(t_{k-1}) + K_p(e(t_k) - e(t_{k-1})) + K_i e(t_k)\Delta t + \frac{K_d}{\Delta t}(e(t_k) - 2e(t_{k-1}) + e(t_{k-2})) \tag{S.52}$$

Although we use the adaptive time-stepping method described in Section S13, for simplification, we assume that the time step, Δt, is constant—that is, $\Delta t = \Delta t^*$ in Equation (S.51). Based on Equation (S.52), the opening ratios of all weirs are adjusted at every time step to ensure that the simulated water levels at each weir closely follow the observed or desired control values.

S12. Input data preparation and conversion

Preparing the input time series data, including meteorological and water quality data for multi-constituent water quality models, is typically a tedious and time-consuming task for modelers. For instance, in CE-QUAL-W2, users must prepare numerous separate input files, such as flow rate, water quality, and meteorological time series, for each branch or tributary. In contrast, EufoRiA offers an efficient approach to streamline these processes. Figure S9 and Table S19 illustrate the conceptual input data conversion process and provide details.

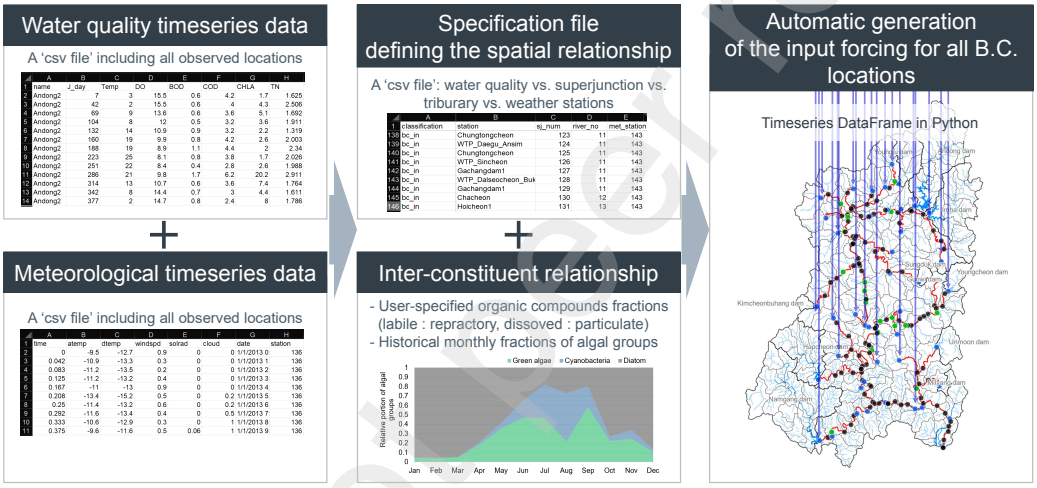


Figure S9: Input forcing timeseries data conversion process in EufoRiA

In EufoRiA, modelers prepare a single 'CSV' file containing all observed water quality datasets—at daily or longer frequencies (e.g., weekly, monthly), with no requirement for fixed intervals—classified by the observation station's name. More specifically, to build the input dataset, the model automatically generates time series of water quality data for specific input boundary locations, referencing a table that matches the observation station names with the corresponding boundary superjunctions using a single water quality dataset file. Similarly, meteorological data is prepared in a 'CSV' file, distinguished by its station code number. EufoRiA then converts these datasets into input time series for each relevant boundary input location, using a specification file that defines the spatial relationships between the water quality stations, superjunction indices, meteorological station numbers, and other relevant information. In addition, the initial values for water quality constituents at all

computational elements in the model are automatically assigned based on the user-defined simulation start date and input dataset. For flow rate boundary inputs, the model applies the area-ratio allocation method to determine runoff flow rates from associated input locations. Additionally, measured point source flow rates, such as dam releases and wastewater treatment discharges, are incorporated as boundary inputs.

Table S19: Conversion equations between observed and modeled water quality constituents

Nutrient group	Measured constituents	Modeling constituents	Conversion methods
Carbon	TOC	LDOC	LDOC = f_{LDOC}[a]× TOC
		LPOC	RDOC = f_{RDOC}[a]× TOC
		RDOC	LPOC = f_{LPOC}[a]× TOC
		RPOC	RPOC = f_{RPOC}[a]× TOC
Nitrogen	TN	NO_3-N	-
	DTN	NH_4-N	-
	NO_3-N	LDON	LDON = f_{LDON}[b]× (DTN − NO_3-N − NH_4-N)
	NH_4-N	LPON	RDON = f_{RDON}[b]× (DTN − NO_3-N − NH_4-N)
		RDON	LPON = f_{LPON}[b]× (TN − DTN − Algae-N)
		RPON	RPON = f_{RPON}[b]× (TN − DTN − Algae-N)
Phosphorus	TP	PO_4-P	-
	DTP	LDOP	LDOP = f_{LDOP}[c]× (DTP − PO_4-P)
	PO_4-P	RDOP	RDOP = f_{RDOP}[c]× (DTP − PO_4-P)
		LPOP	LPOP = f_{LPOP}[c]× (TP − DTP − Algae-P)
		RPOP	RPOP = f_{RPOP}[c]× (TP − DTP − Algae-P)

[a] The fractions of LDOC, RDOC, LPOC, and RPOC in TOC, respectively
[b] The fractions of labile and refractory in the dissolved and particulate organic nitrogen, respectively
[c] The fractions of labile and refractory in the dissolved and particulate organic phosphorus, respectively

Because not all 23 of the modeled constituents can be observed or specified directly, a set of conversion relationships are used to build the full input data for the model based on known relationships with observed constituents. For this task, the model uses user-specified information on the relative proportions of labile vs. refractory and dissolved vs. particulate organic compounds. For example, total nitrogen, nitrate, and ammonium data are used to calculate LDON, LPON, RDON, and RPON, relying on prior research that defines the relative proportions and relationships between these constituents and the observed data. A similar approach is applied to organic carbon and phosphorus. Specifically, for organic carbon, we refer to a previous study on the relative fractions between LDOC, LPOC, RDOC, and RPOC in the

Han River of South Korea ([68]). Table S19 summarizes the data conversion methods. Additionally, the model converts chlorophyll-a concentration into the biomass of three algal groups using their historical monthly patterns.

S13. Adaptive time-stepping method

EufoRiA employs an adaptive time-stepping method to reduce simulation times by using a large temporal step size during baseflow conditions, and using a fine temporal step size during storm events to avoid truncation error and instability. In coupled hydrodynamic and water quality modeling, ensuring both numerical stability and accuracy is critical for producing reliable simulation outcomes. While the fully implicit numerical scheme used for the hydraulics and water quality solvers in this study ensures unconditional stability, large time intervals may still compromise the accuracy of the numerical solution. This issue is especially critical in coupled modeling, where even minor volume balance errors between the computational elements in the hydraulic solver can significantly affect water quality results, leading to abrupt concentration peaks or even negative values due to the high sensitivity of concentration calculations in low-volume conditions. To maintain accuracy while ensuring reasonable computational speed, proper time interval selection are essential in the coupled simulation of the hydraulic and water quality model.

Given the complexity of our model structure—which integrates distinct solvers for hydraulics and water quality transport—and the absence of a theoretical framework for determining a universally optimal time step across different solvers, we adopt a practical adaptive time-stepping method. The proposed time interval (Δt) at the specific time is determined using Equation S.53.

$$\Delta t = k_{time} \frac{1}{N \times k_{percentile}} \sum_{ik=1}^{N \times k_{perceltile}} \frac{\Delta x_{ik}}{|Q_{ik}/A_{ik}|} \tag{S.53}$$

Where, N is the number of all links in the model, k_{time} is a user specified parameter to adjust the magnitude of Δt, $k_{percentile}$ is a user specified parameter to limit the averaging range, Δx_{ik} is the length of the ikth link in the model, A_{ik} and Q_{ik} are the water cross-sectional area and the flow rate at the ikth link in the model, respectively.

This expression resembles the widely known Courant-Friedrichs-Lewy (CFL) condition, $\Delta t \leqslant C_{max} \Delta x/u$, where C_{max} is a numerical method-dependent constant, u is the flow velocity, and Δx is the spatial discretization element in the numerical model. In our approach, users specify the minimum and maximum allowable time intervals. If the calculated Δt falls within this range, it is applied directly as the next time interval. Additionally, to mitigate the

impact of extremely large flow velocities ($|Q_{ik}/A_{ik}|$) in the model network, we introduce a coefficient, $k_{percentile}$, which restricts the averaging range to a subset of $\Delta x_{ik}/(|Q_{ik}/A_{ik}|)$ values. For instance, if $k_{percentile} = 0.7$, the model computes Δt for the next time interval by averaging the top 70% of these values, sorted in descending order. Through the case study described in 3.1.5, this approach is found to perform effectively under properly chosen parameters in Equation S.53.

S14. Spatial correlation between the locations in the river network

To incorporate observed spatial correlations in contaminant dynamics into the DA framework, we first use historically observed water quality constituent concentrations from 2013–2020 (calibration period) to determine reach-wise correlations for each constituent. We then incorporate these spatial correlations into the process noise covariance matrix used by the Kalman Filter. The underlying assumption is that exogenous inputs (such as fertilizer runoff loads) are spatially correlated, allowing integration of prior knowledge into the process noise covariance matrix to improve the DA results [57]. This assumption is particularly valid in river networks, where contaminant runoff loads are significantly correlated in space.

However, implementing spatial correlations requires careful consideration of distance weighting to localize their impact. Because several water quality constituents are heavily influenced by surrounding environmental factors (e.g., water temperature is primarily a function of meteorological phenomena rather than location), spurious spatial correlations may occur between distantly-separated points that lack causal physical connections. To avoid this issue, we limit the correlation impact based on the distance between spatial elements applying a simple distance weight function as follows, with considering the relationship of $Cov(X, Y) = \rho_{XY}\sigma_X\sigma_Y$:

$$\Sigma_{spatial,C_n,ij} = \rho_{ij}\sigma_i\sigma_j e^{-\frac{d_{ij}}{\phi}} \tag{S.54}$$

Where $\Sigma_{spatial,C_n,ij}$ is the covariance matrix that has elements consisting of ith and jth spatial covariance for nth water quality constituent, C_n means the nth constituent, ρ_{ij} is the Pearson's correlation for nth water quality constituent between ith and jth spatial elements, σ_i and σ_j is the standard deviation at i and jth spatial elements, d_{ij} is the Euclidean distance (m) between ith and jth spatial elements, ϕ is the user parameter (m) that defines the effective distance for correlation impact. In this study, $\phi = 20,000\ m$ is applied as the effective distance.

In addition, because different tributaries exhibit distinct water quality characteristics and are subject to nutrient loads originating from different upstream areas, the spatial correlation effects should be limited between locations in different tributaries. Therefore, we ignore correlations between tributaries that are separated by drainage divides to maintain physical consistency. Figure S10 shows the procedure to develop the spatial correlation matrices in EufoRiA.

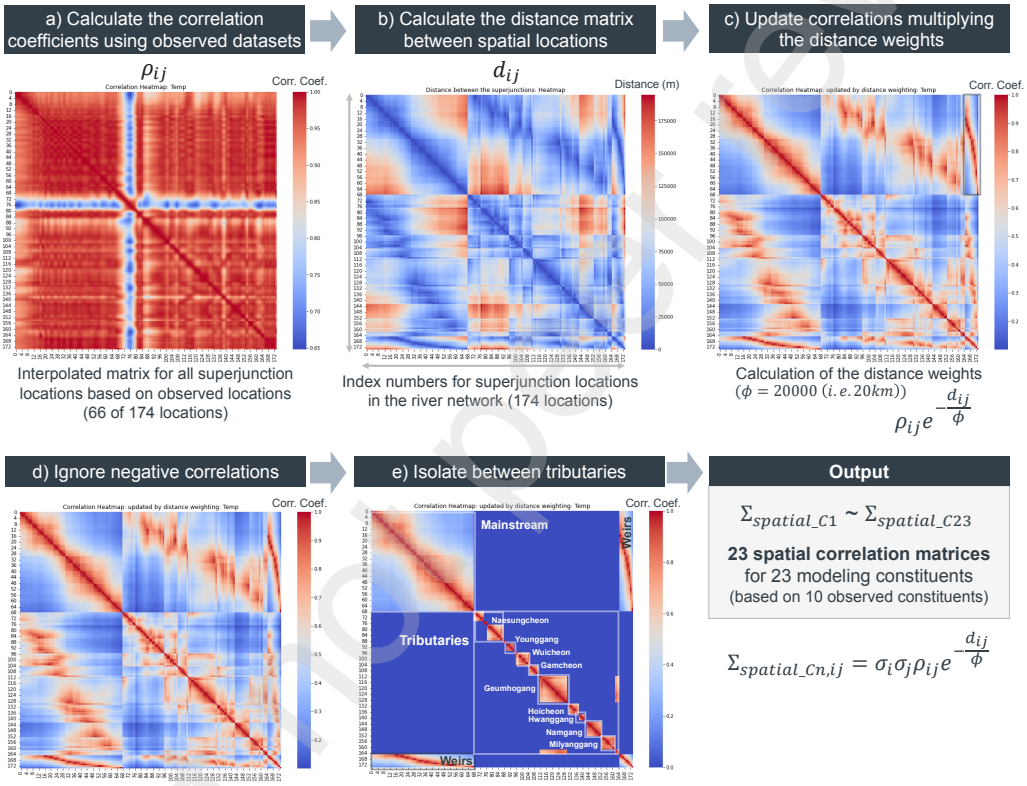


Figure S10: Procedure for developing the spatial correlation matrices between the different superjunction locations in the river network

S15. Inter-constituent correlations between water quality constituents

Water quality constituents often exhibit significant correlations due to similarities in chemical composition. For example, concentrations of nitrogen constituents such as total nitrogen, nitrate, and organic nitrogen groups are frequently correlated. By leveraging these correlations, we aim to enhance the estimation of unmeasured water quality constituents, such as organic nitrogen groups, using the measured constituents like total nitrogen or nitrate. As before, we capture these correlations within the process noise covariance matrix based on historically-observed correlations between constituents. First, inter-constituent correlations are analyzed for the calibration period (2013–2020), and the resulting correlations are used to construct the inter-constituent process noise covariance matrix. This inter-constituent process noise covariance matrix, together with the spatial correlation matrix, is then used inform the Kalman Filter during the validation period (2021–2022) and the resulting improvement in model skill is assessed.

S16. Details of the data assimilation procedure during numerical modeling

As discussed in Section 3.1.4, EufoRiA simulates coupled hydraulic and water quality dynamics in three steps, which include (i) the hydraulic simulation, (ii) the water quality transport simulation, and (iii) the simulation of reactions between the water quality constituents. Within this framework, data assimilation is applied between the water quality transport and reaction steps of the simulation. Thus, the state and input transition matrices for the Kalman filter represent the system dynamics as described by the hydraulics and advection-diffusion equation instead of the water quality reaction dynamics. This design choice is made given that the integration of the water quality reaction dynamics and transport equations is challenging given the complex reactions between the large number of constituents considered. However, considering the relatively small time steps of several minutes in this model, we can assume that system dynamics affecting the water quality concentrations during each time interval can be primarily represented by the spatial hydraulics and transport phenomenon rather than reaction kinetics.

S17. Comparison of model performances across different studies

To compare the performance of EufoRiA with existing models across various application scenarios, we collected performance metrics from different studies as presented below. Table S20 shows the performance of various models across different studies for predicting Chlorophyll-a and nutrients.

We also include the performance metrics presented in Table 5 to enable direct comparison with other assessment results from the literature. In comparison with literature data, the results of EufoRiA show relatively good performance for temperature, total nitrogen, and total phosphorus. Additionally, we can conclude that the DO and Chlorophyll-a results of EufoRiA also show comparable-level performances to existing research.

However, direct comparison would be inappropriate due to differences in target water bodies, simulation periods, target constituents, and performance metrics. Therefore, these values should be used for reference purposes only. Additionally, it should be noted that most existing applications have focused on modeling mainstem rivers or individual reservoirs, rather than complex river networks with multiple upstream boundary conditions and numerous WWTPs as point-source inputs, as implemented in this study.

Table S20: Model performance across different studies (Cal: Calibration, Val: Validation. When no distinction is made between calibration and validation periods, results are classified as calibration cases.)

Author	Model	Constituent	NSE		R^2		RMSE		PBIAS	
			Cal	Val	Cal	Val	Cal	Val	Cal	Val
This study (Table 5)	EufoRiA	Temperature	0.97	0.99	-	-	1.60	1.01	+3.4%	-0.9%
		DO	0.48	0.40	-	-	1.69	1.64	+9.2%	+9.6%
		Chl-a	-0.00	0.31	-	-	15.97	12.44	+13.1%	+7.2%
		TN	0.76	0.82	-	-	0.38	0.29	+5.4%	+0.7%
		TP	0.49	0.40	-	-	0.02	0.02	-1.3%	-7.7%
		TOC	-0.02	0.27	-	-	0.89	0.74	-0.8%	-0.6%
Du et al. [69]	CE-QUAL-W2	Temperature	0.95	-	0.95	-	0.46	-	-	-
		Chl-a	0.55	-	0.54	-	0.3	-	-	-
		TN	0.59	-	0.58	-	0.24	-	-	-
		TP	0.55	-	0.55	-	0.01	-	-	-
Brito et al. [70]	CE-QUAL-W2	Temperature	-	-	-	-	1.8	-	-	-
		DO	-	-	-	-	7.6	-	-	-
		Chl-a	-	-	-	-	62.9	-	-	-
		Nitrate	-	-	-	-	0.4	-	-	-
		Orthophosphoate	-	-	-	-	0.48	-	-	-
Hong et al. [71]	EFDC	Chl-a	-	-	0.213	-	8.434	-	-	-
Lee et al. [72]	HSPF	Temperature	0.96	0.96	0.96	0.96	-	-	+1.62%	+3.09%
		Nitrate	-1.05	0.61	0.17	0.63	-	-	-2.45%	-6.45%
		Phosphoate	-0.45	-0.35	0.24	0.06	-	-	+36.46%	+53.34%
		Chl-a	0.22	0.18	0.36	0.25	-	-	+6.60%	+33.44%
Bui et al. [73]	SWAT+QUAL2K	Temperature	0.53	-	0.72	-	-	-	-1.68%	-
		BOD	0.26	-	0.45	-	-	-	+12.56%	-
		COD	0.16	-	0.38	-	-	-	+6.49%	+53.34%
		TN	0.62	-	0.77	-	-	-	+13.73%	-
		TP	0.24	-	0.32	-	-	-	+9.92%	-
Kim et al. [74]	HEC-RAS	Nitrate	0.30–0.83	-0.16–0.76	0.44–0.90	0.37–0.79	-	-	-5.0–10.3%	0.4–10.0%
Pyo et al. [75]	EFDC-NIER	Temperature	-	-	-	-	0.92–1.15	-	-	-
		DO	-	-	-	-	1.35–3.27	-	-	-
		BOD	-	-	-	-	0.83–1.04	-	-	-
		TN	-	-	-	-	0.32–0.67	-	-	-
		TP	-	-	-	-	0.01–0.04	-	-	-
		Chl-a	-	-	-	-	9.52–19.93	-	-	-
Kim et al. [76]	CE-QUAL-W2	Temperature	-	-	0.89-0.95	-	1.64–2.32	-	-	-
		TP	-	-	0.14–0.77	-	0.01–0.07	-	-	-
		Chl-a	-	-	0.27–0.36	-	6.7–13.2	-	-	-
		TOC	-	-	0.11–0.43	-	1.2	-	-	-
Brett et al. [77]	CE-QUAL-W2	Temperature	0.82	-	-	-	1.3	-	6%	-
		TP	-0.1	-	-	-	0.01	-	58%	-
		Chl-a	0.14	-	-	-	4.6	-	66%	-
		DO	0.85	-	-	-	1.2	-	12%	-
Tasnim et al. [78]	MINLAKE2020	Temperature	0.50–0.98	-	0.95–0.99	-	0.98–1.88	-	-	-
		DO	-0.12–0.80	-	0.87–0.93	-	1.79–3.42	-	-	-

S18. Model Calibration and validation result graphs for SJ, ND, GM, CG, GG, DS, and HC weirs

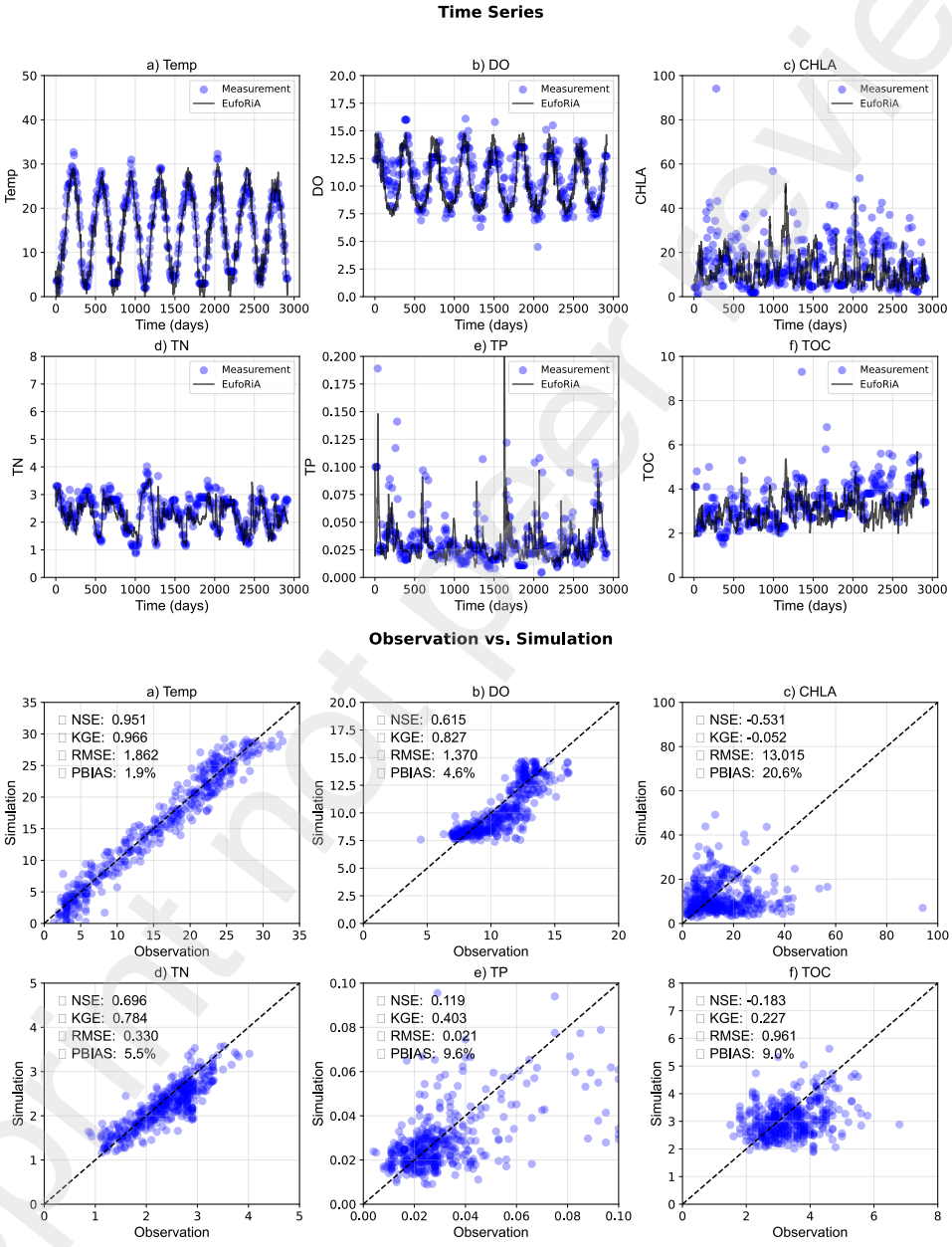


Figure S11: Calibration (2013-2020) results for Sangju (SJ) weir

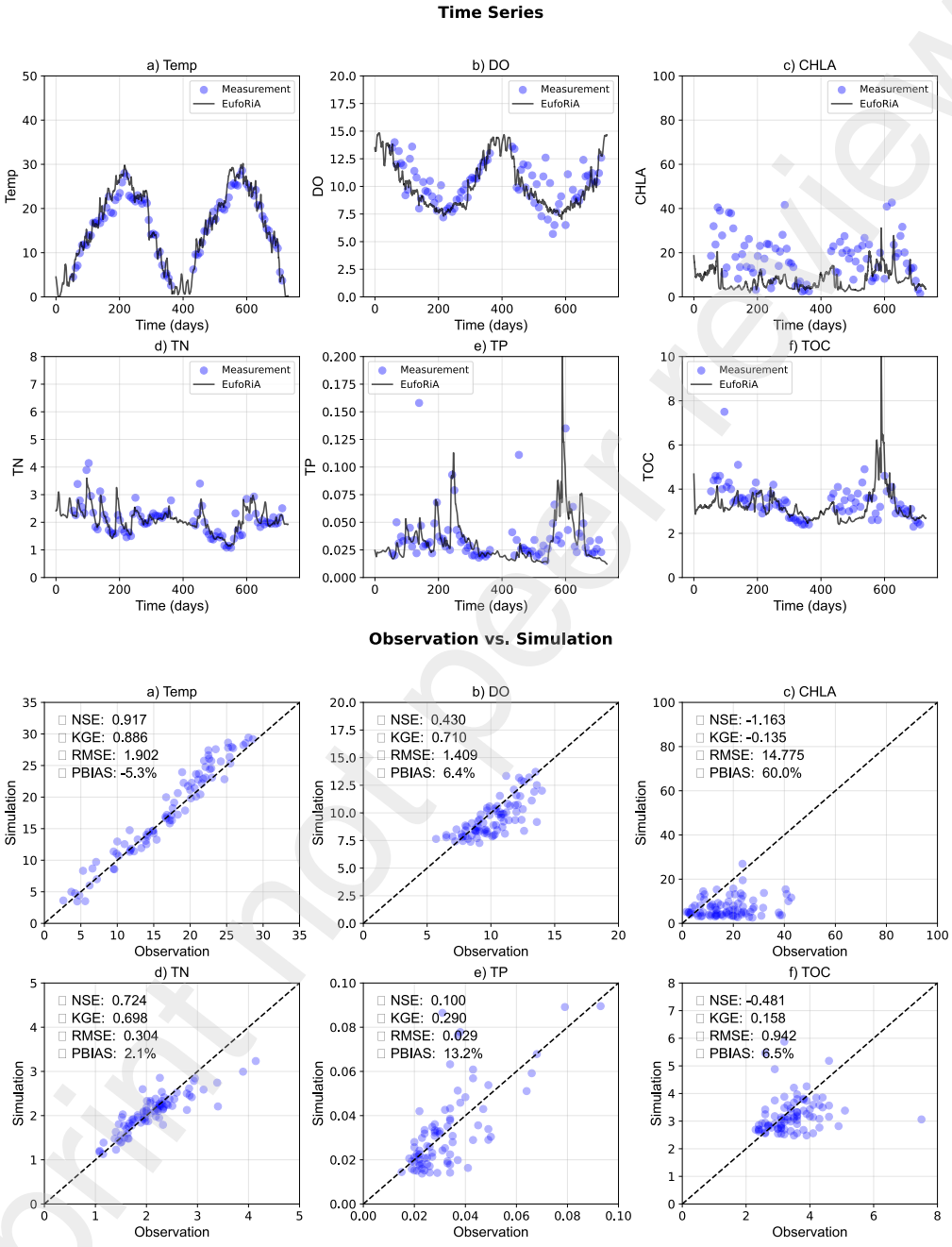


Figure S12: Validation (2021-2022) results for Sangju (SJ) weir

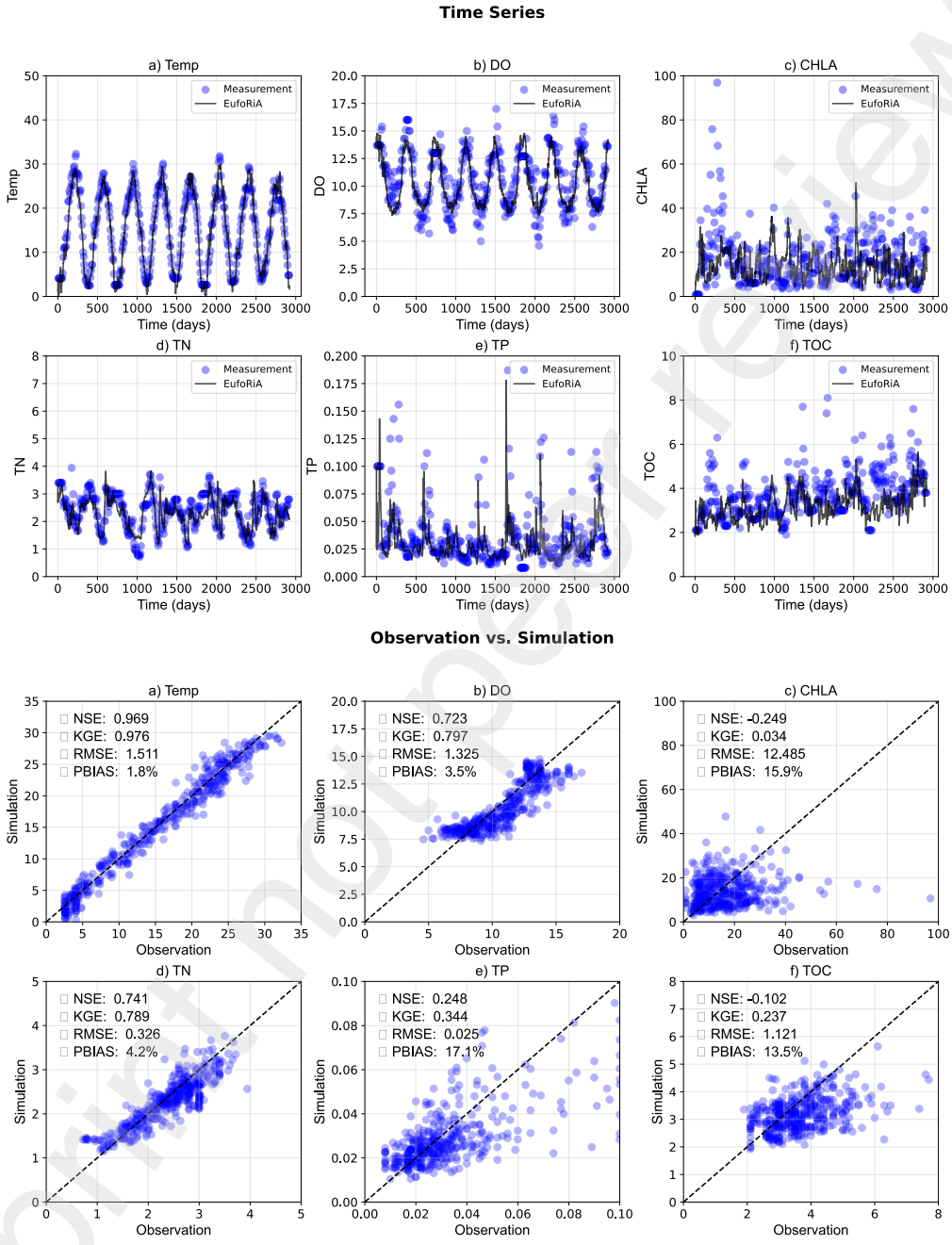


Figure S13: Calibration (2013-2020) results for Nakdan (ND) weir

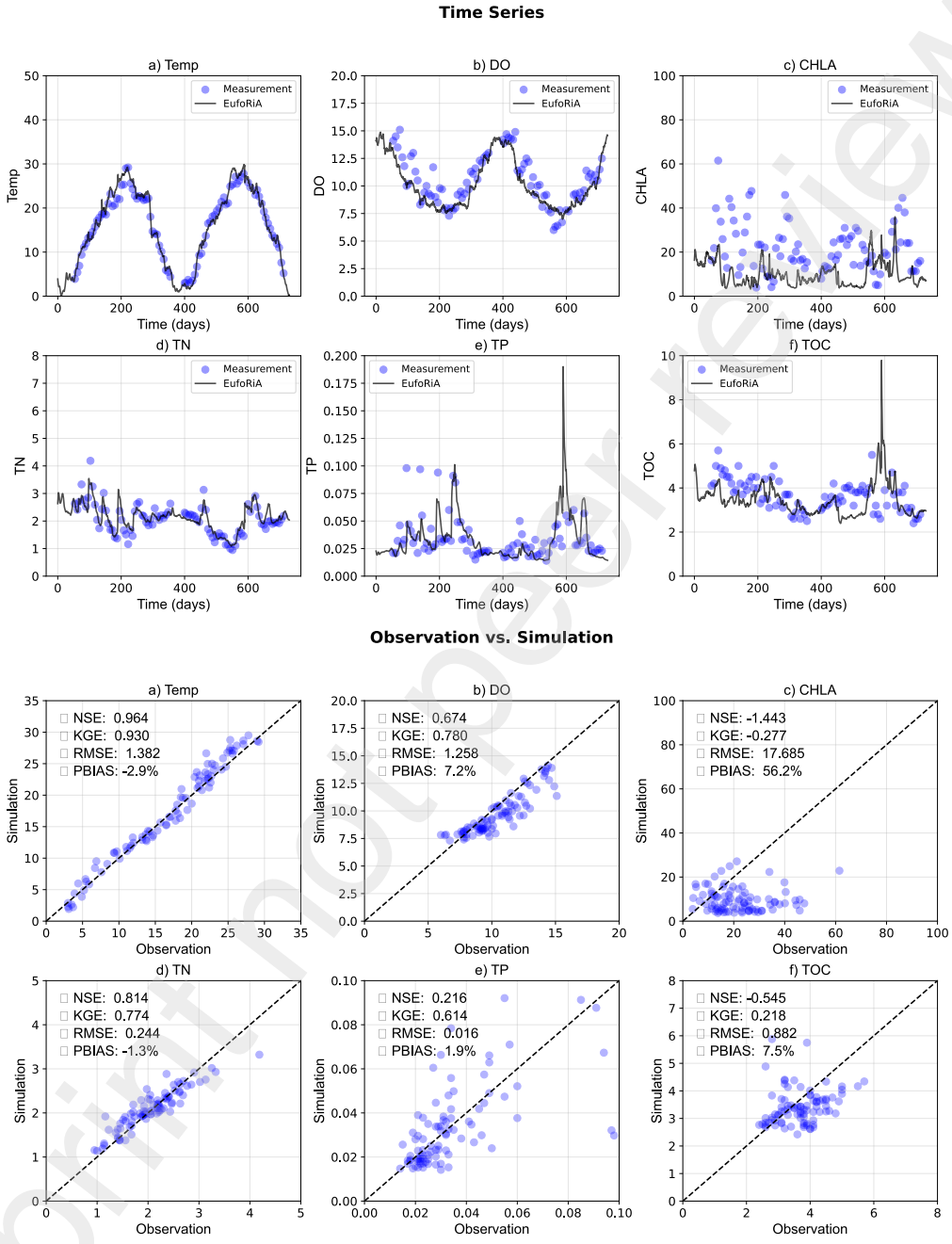


Figure S14: Validation (2021-2022) results for Nakdan (ND) weir

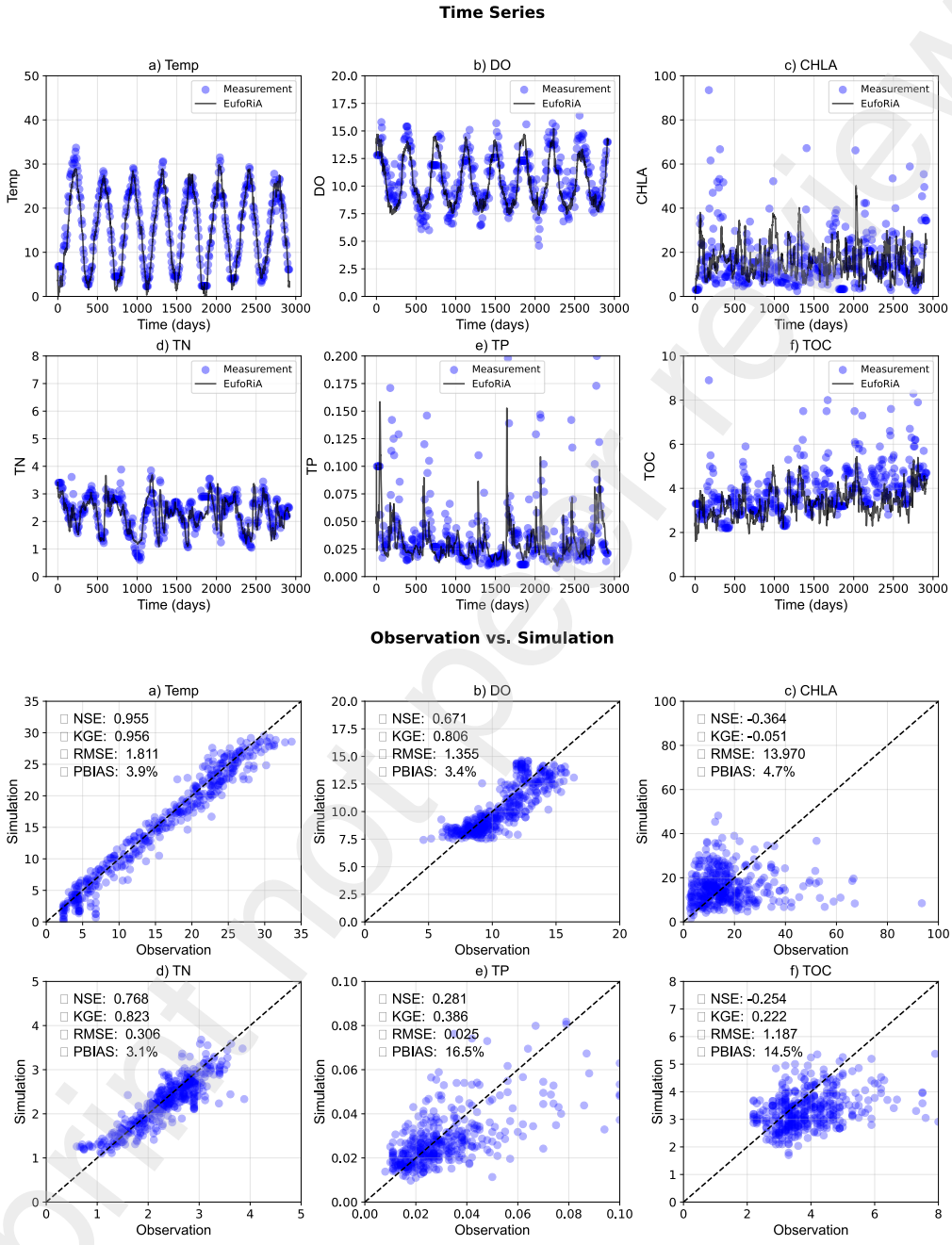


Figure S15: Calibration (2013-2020) results for Gumi (GM) weir

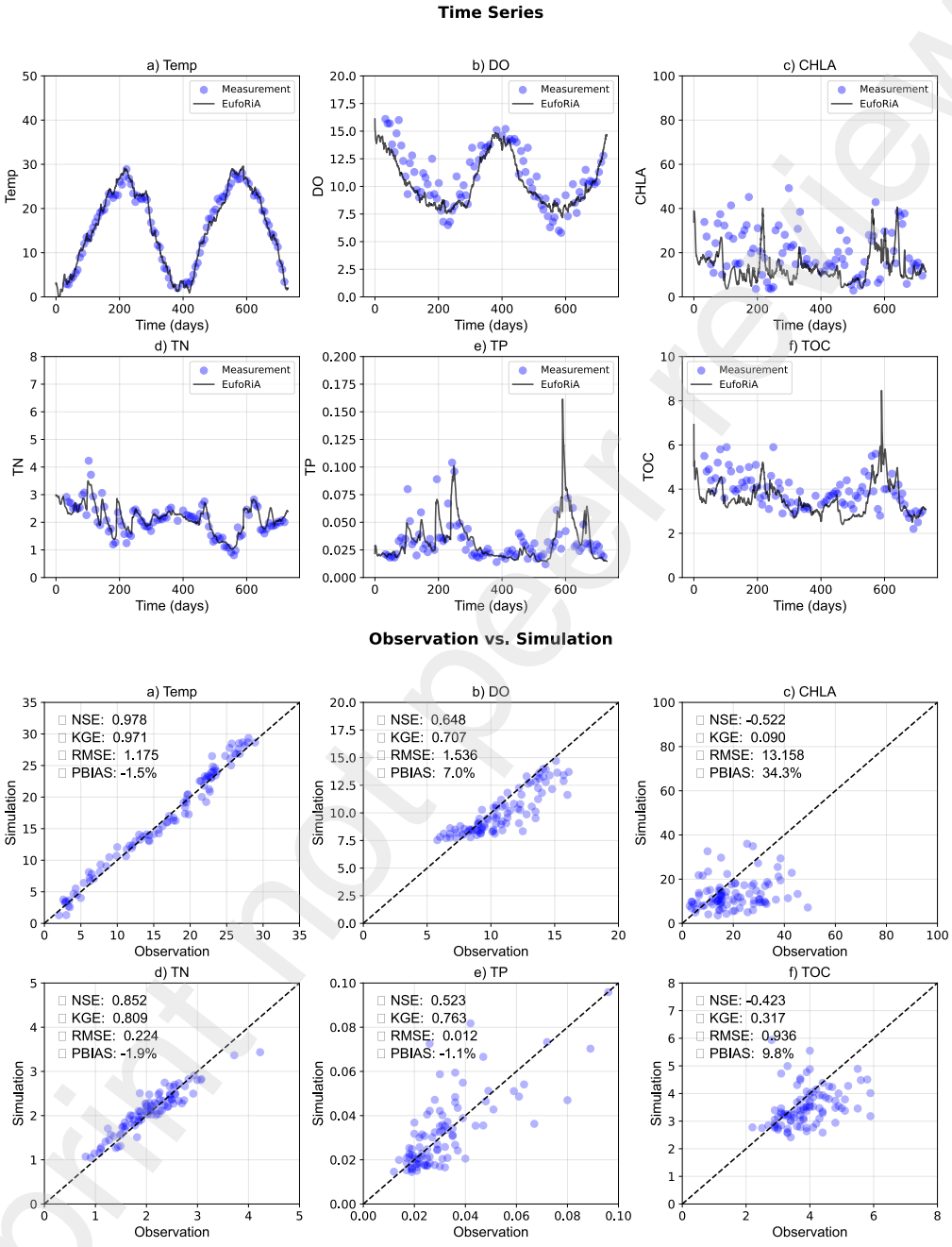


Figure S16: Validation (2021-2022) results for Gumi (GM) weir

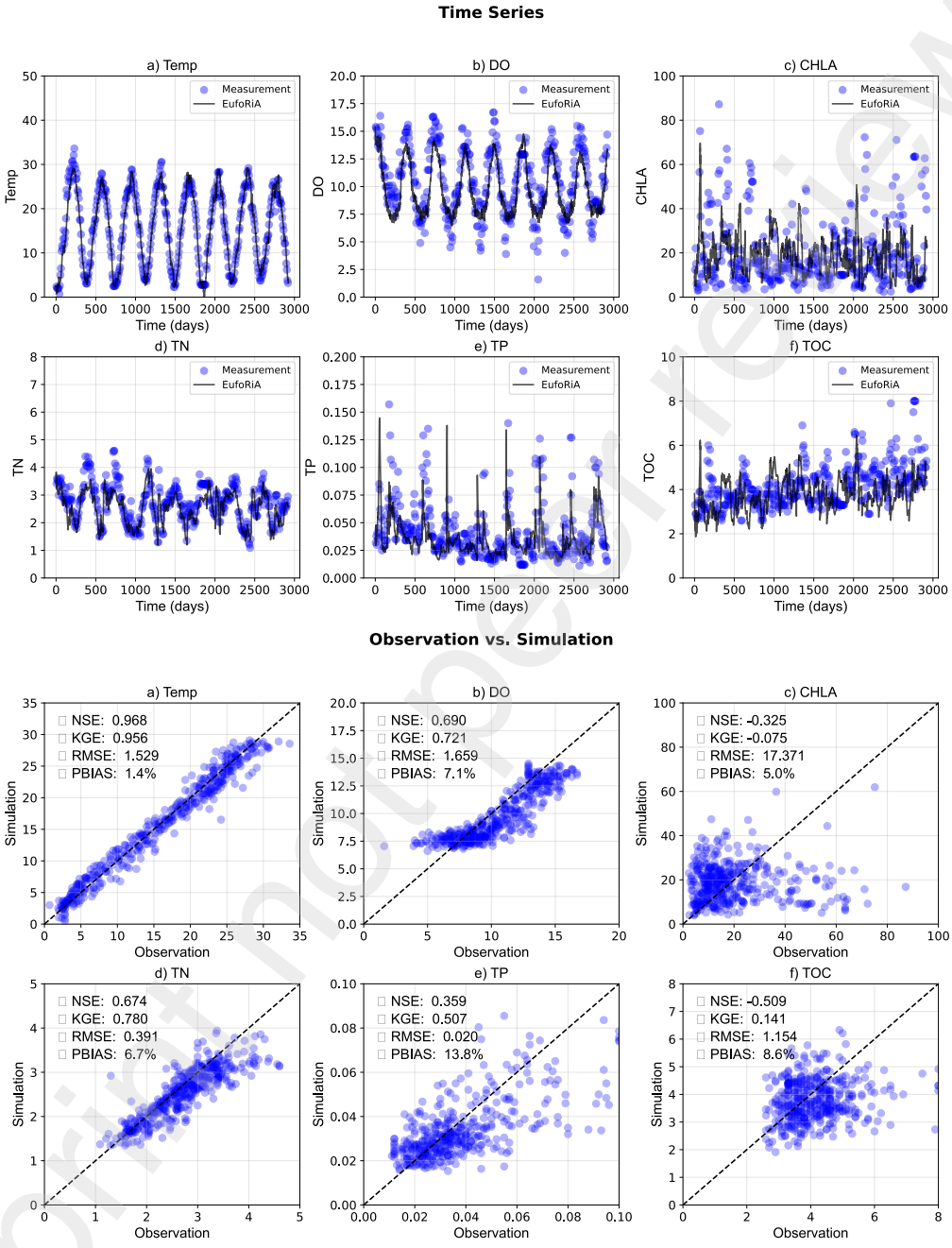


Figure S17: Calibration (2013-2020) results for Chilgok (CG) weir

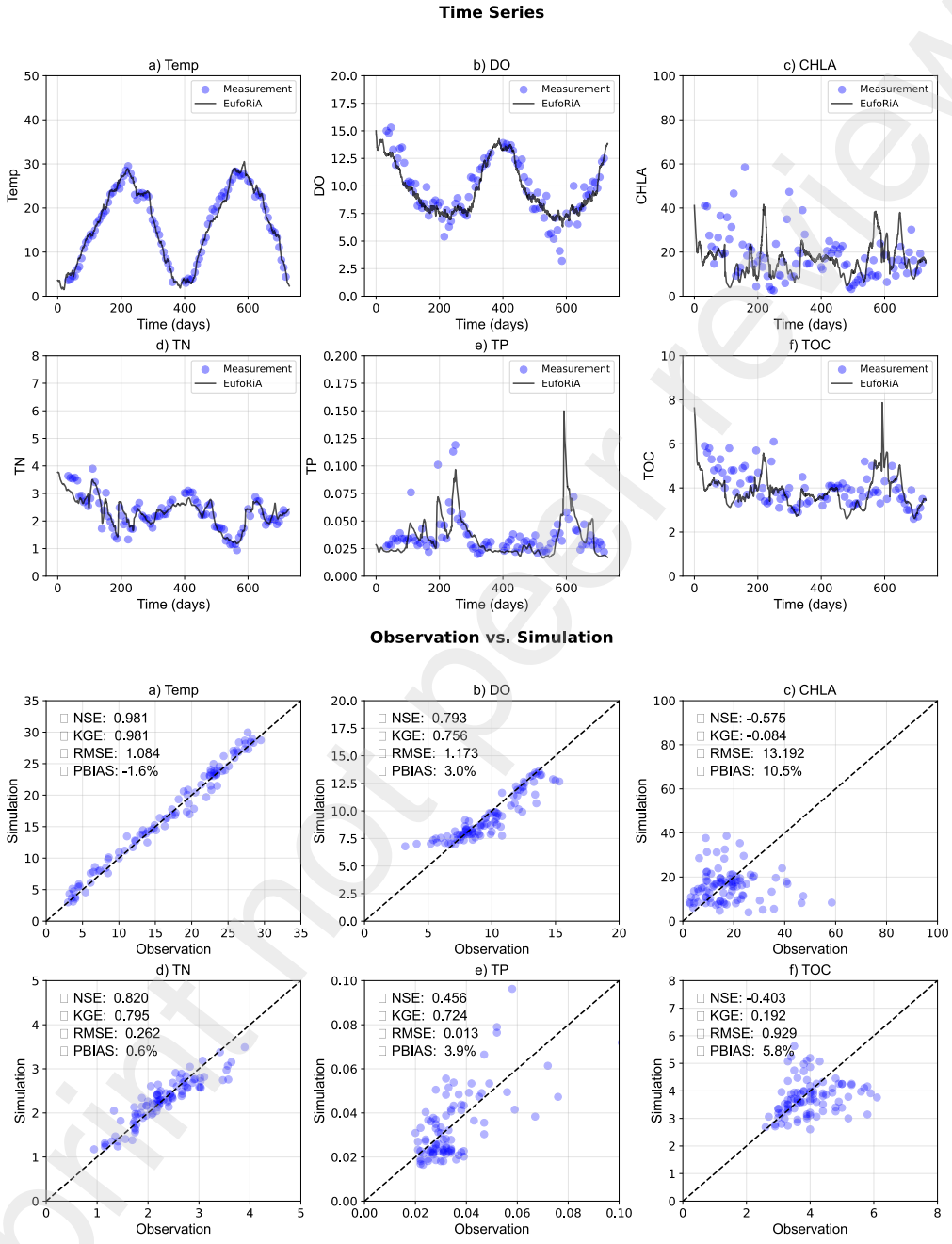


Figure S18: Validation (2021-2022) results for Chilgok (CG) weir

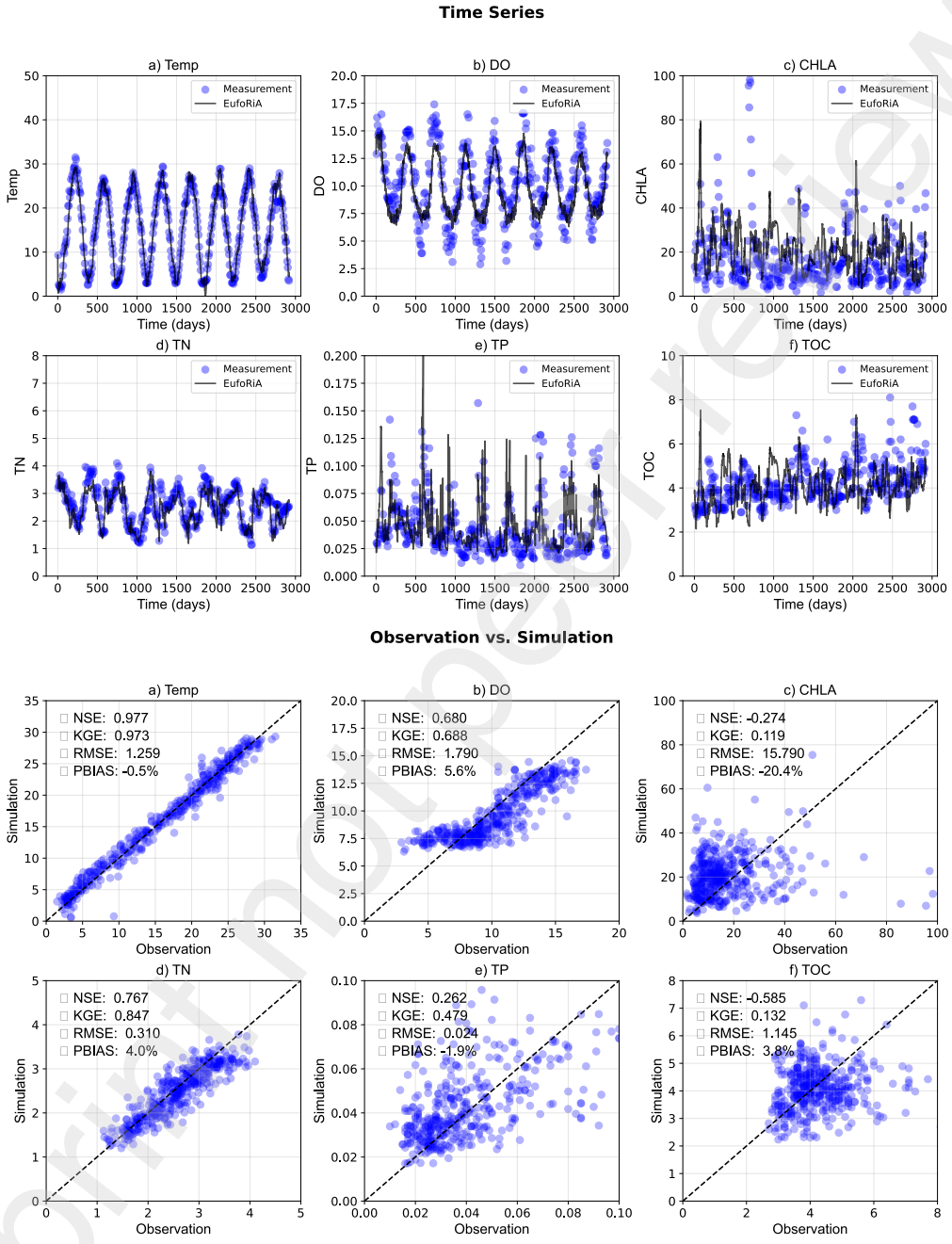


Figure S19: Calibration (2013-2020) results for Gangjeong-Goryung (GG) weir

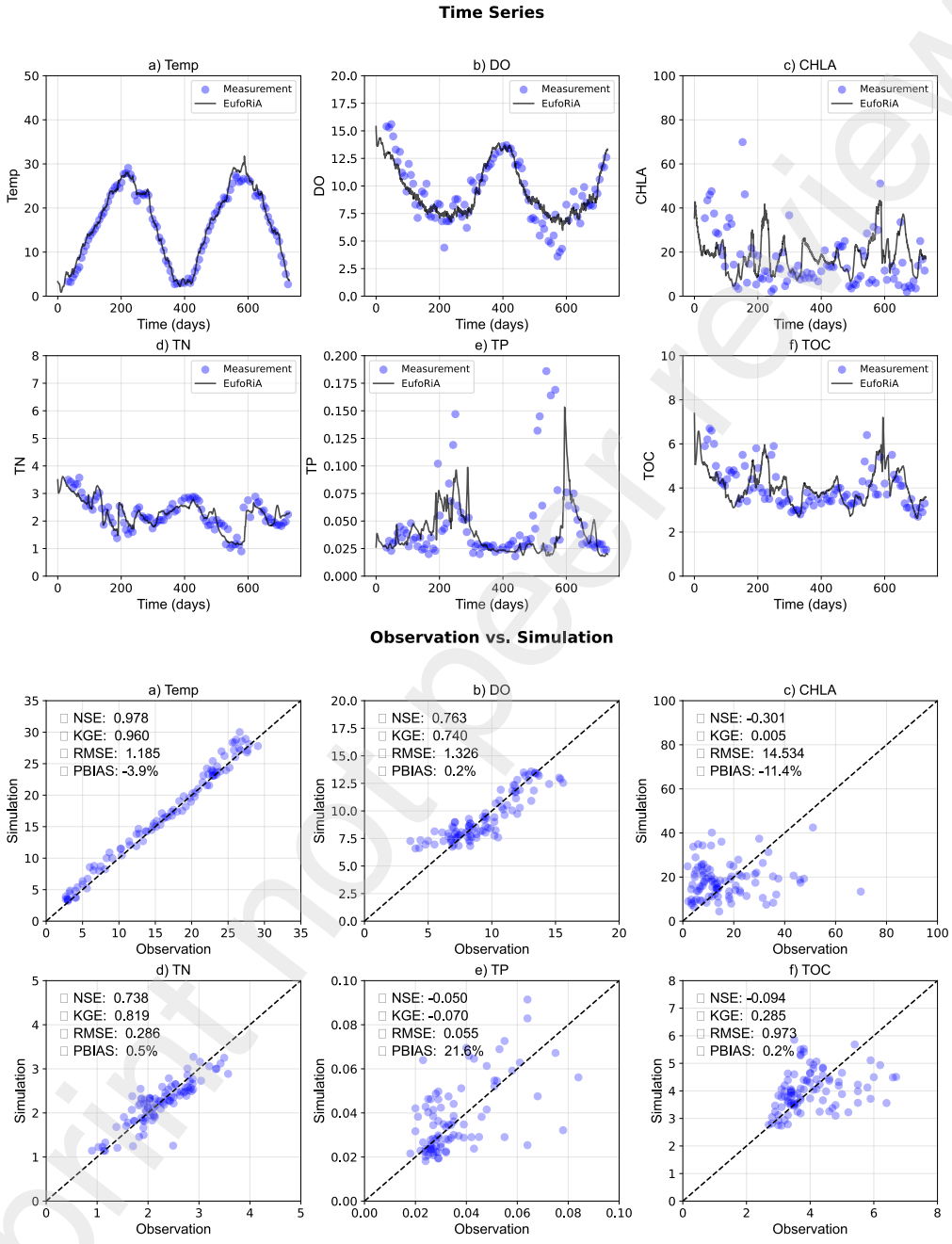


Figure S20: Validation (2021-2022) results for Gangjeong-Goryung (GG) weir

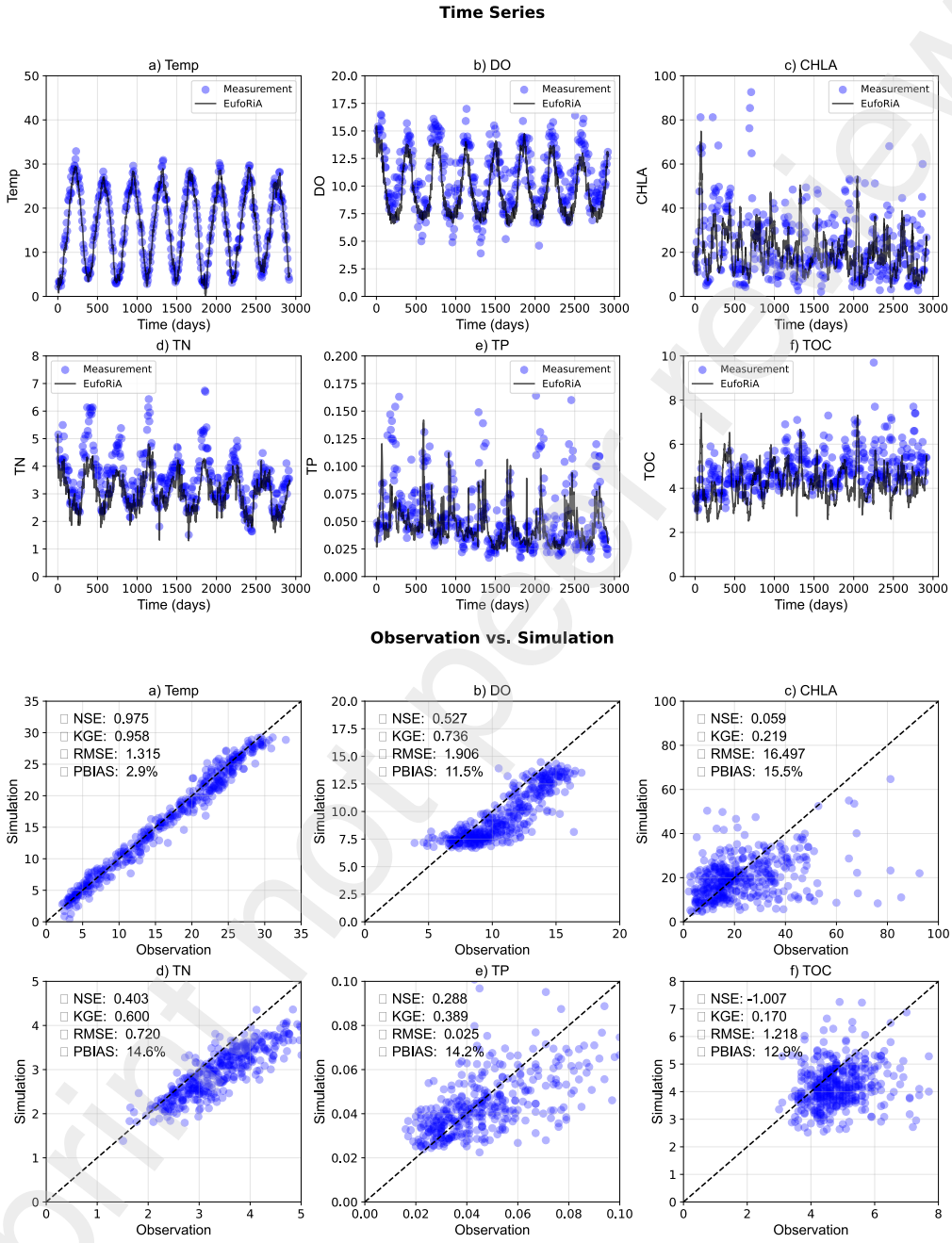


Figure S21: Calibration (2013-2020) results for Dalsung (DS) weir

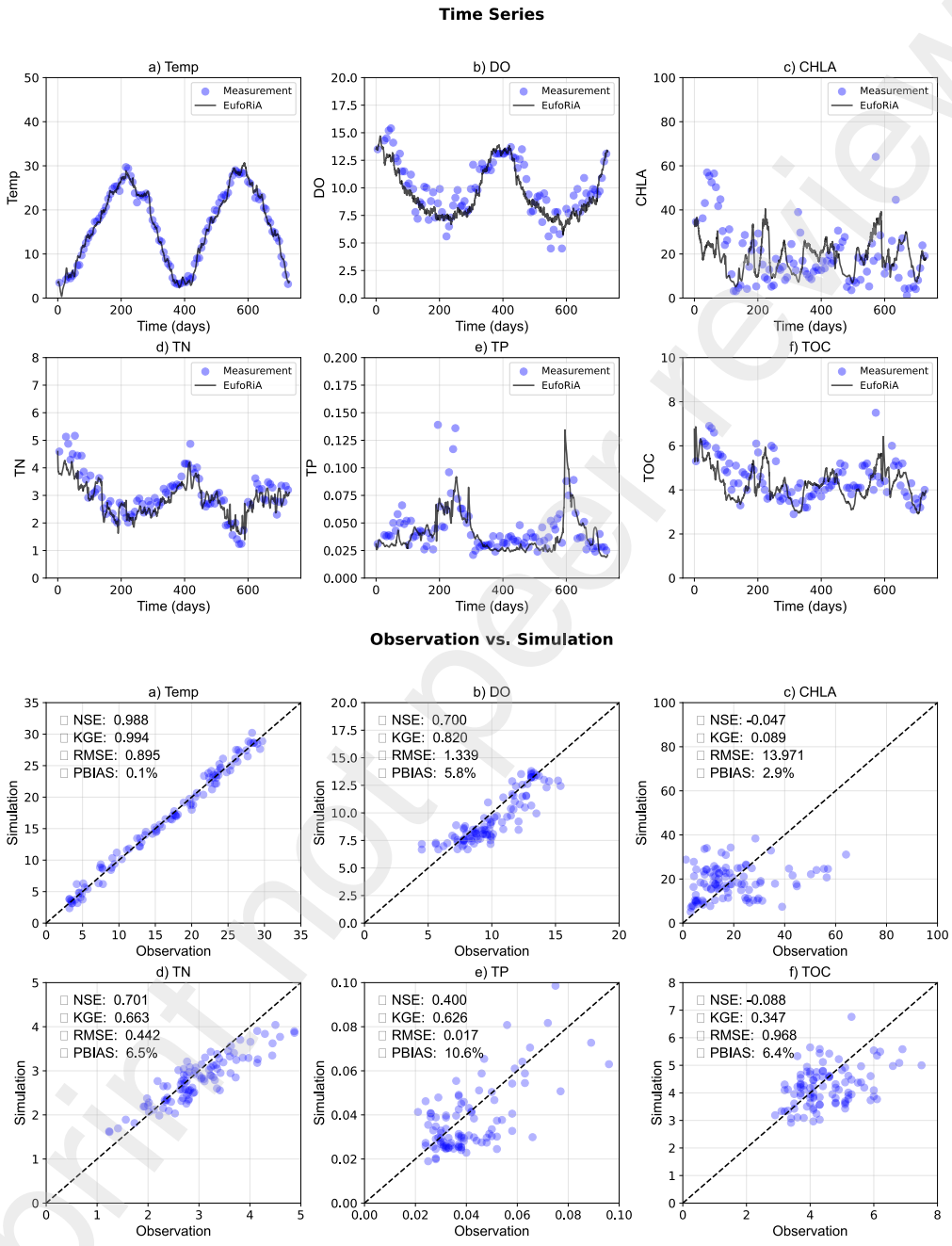


Figure S22: Validation (2021-2022) results for Dalsung (DS) weir

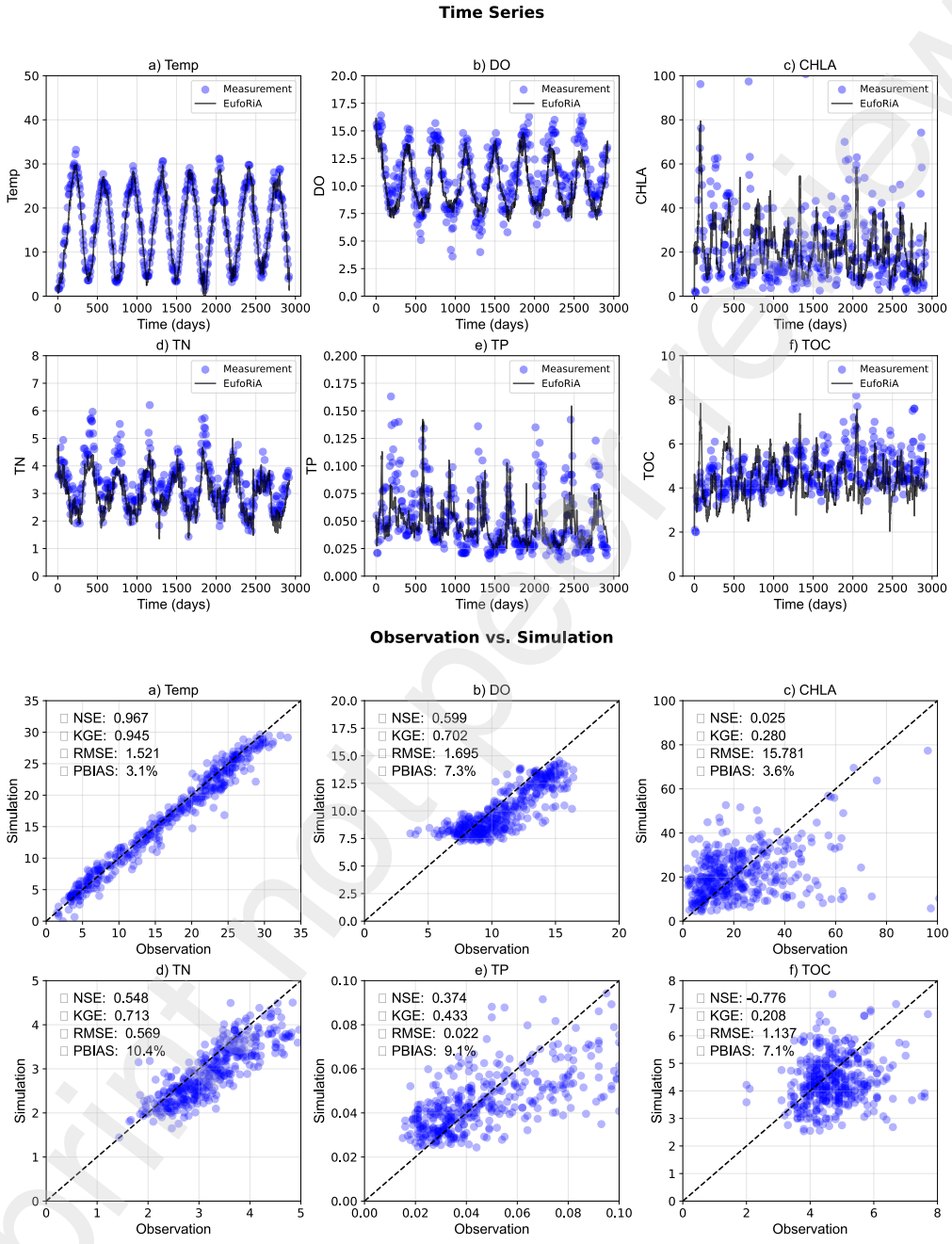


Figure S23: Calibration (2013-2020) results for Hapcheon-Changnyung (HC) weir

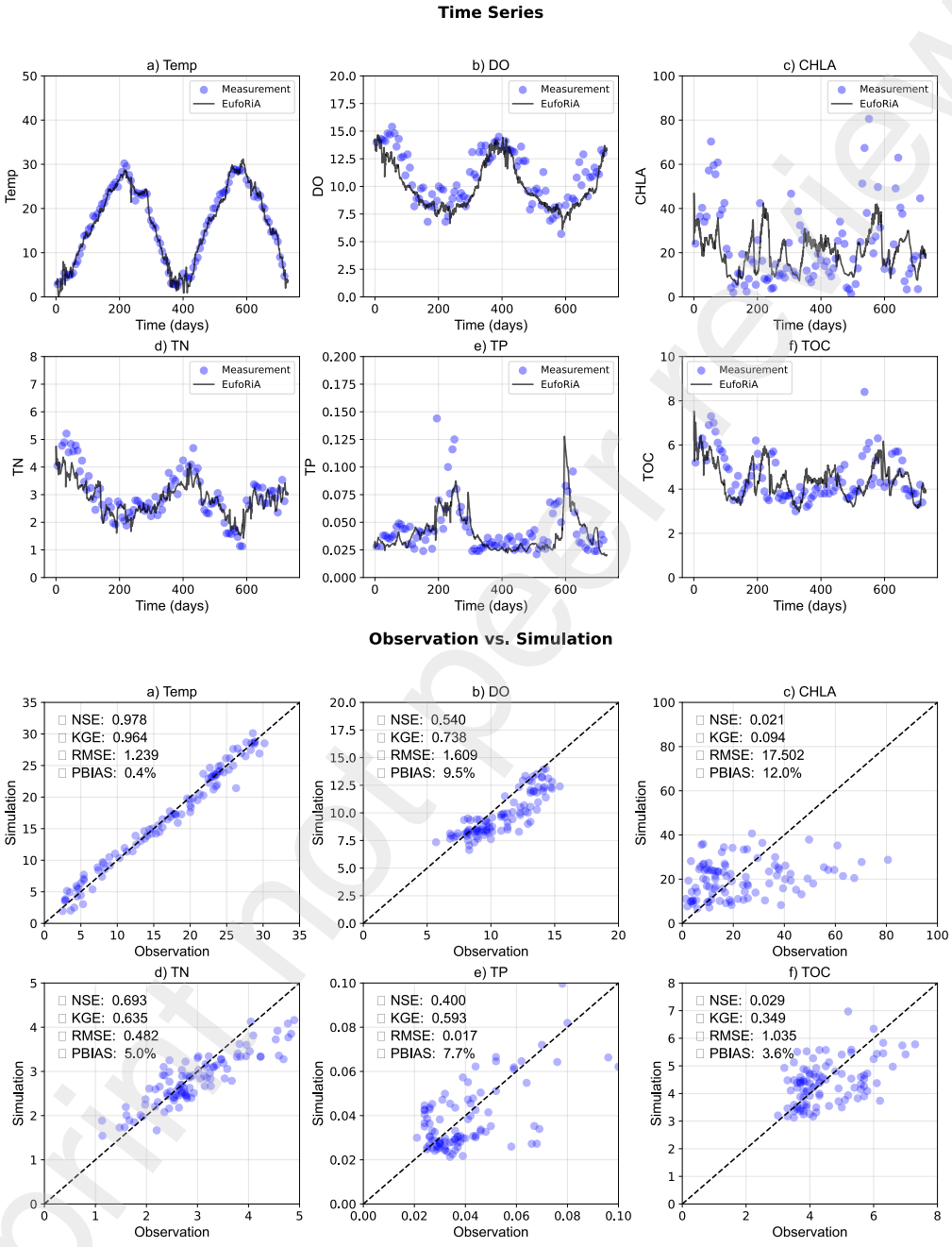


Figure S24: Validation (2021-2022) results for Hapcheon-Changnyung (HC) weir

S19. Monthly boxplots for validation periods (2021-2022)

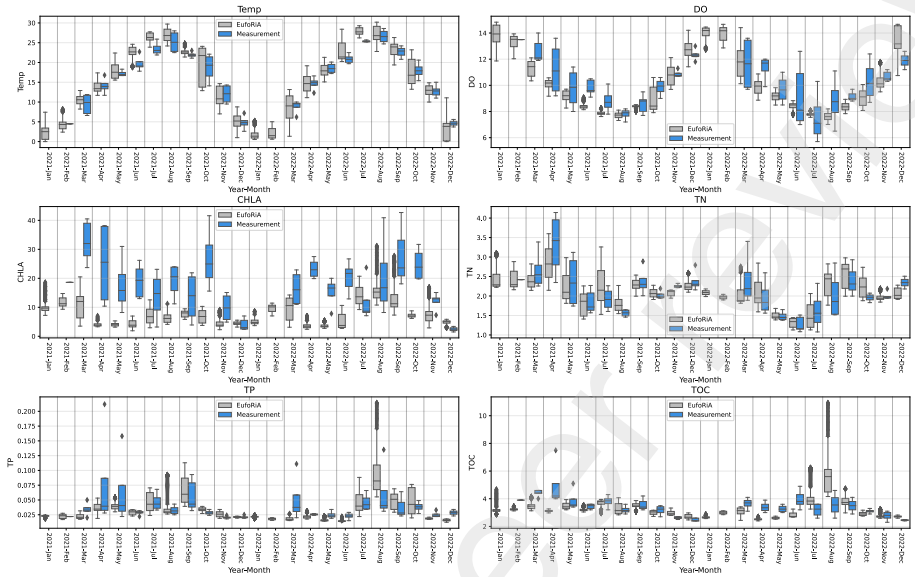


Figure S25: Monthly boxplot for Sangju (SJ) weir

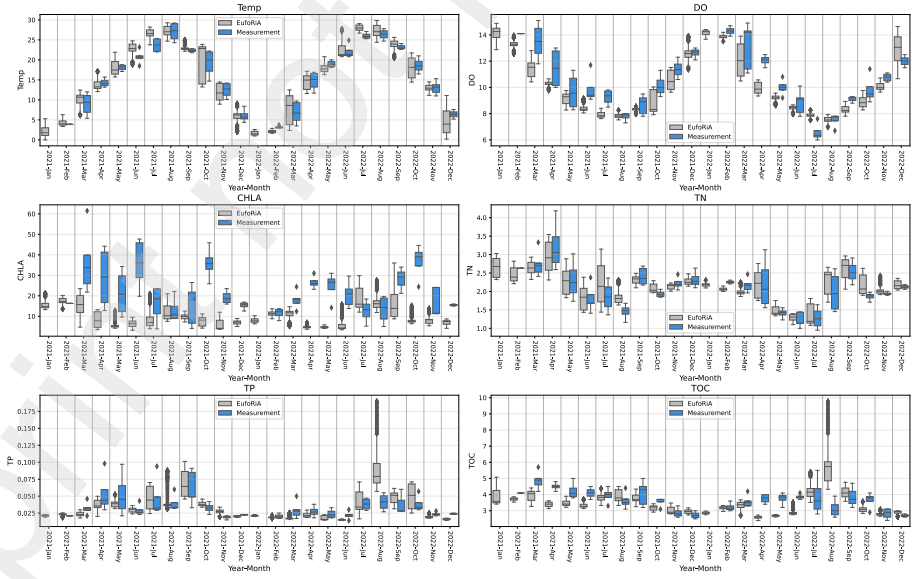


Figure S26: Monthly boxplot for Nakdan (ND) weir

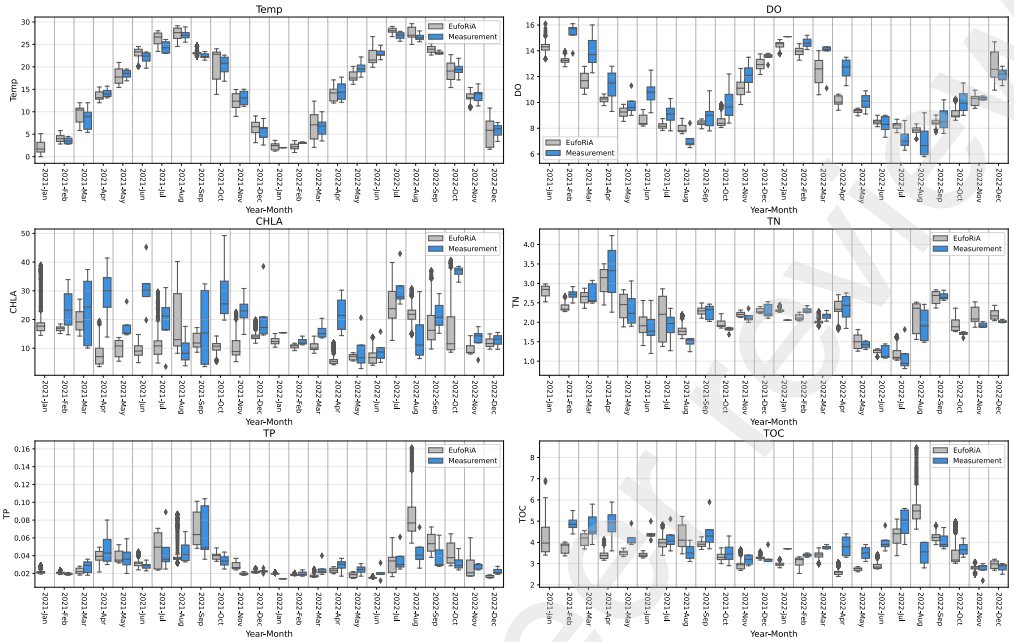


Figure S27: Monthly boxplot for Gumi (GM) weir

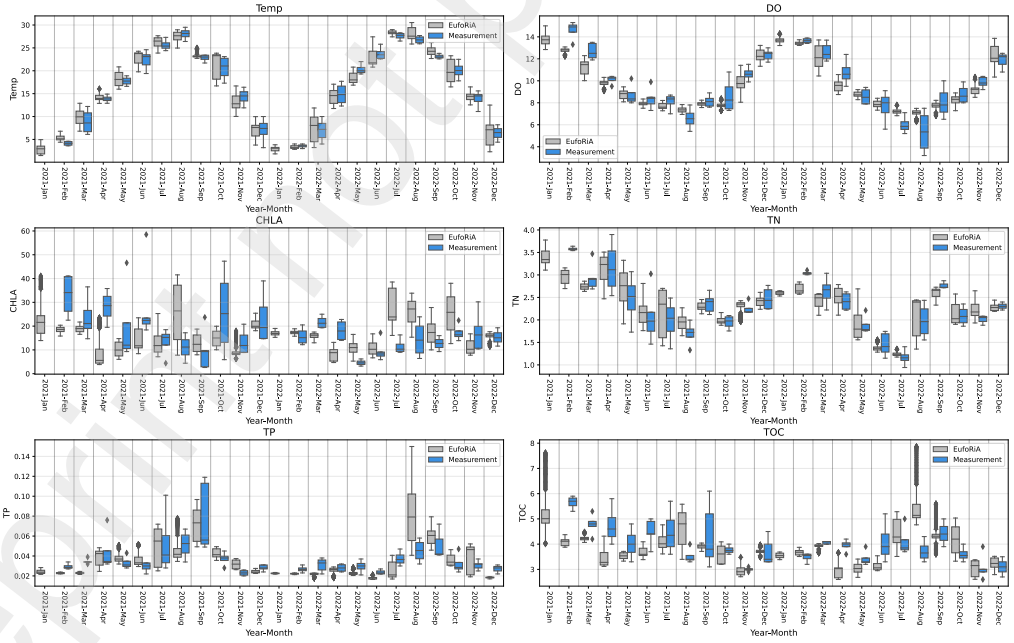


Figure S28: Monthly boxplot for Chilgok (CG) weir

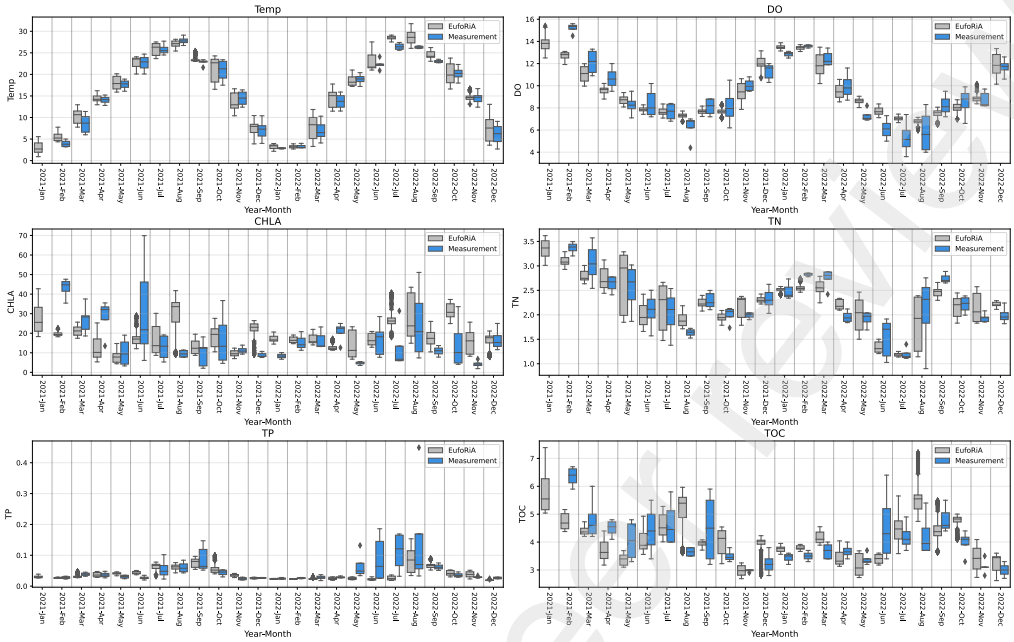


Figure S29: Monthly boxplot for Gangjeong-Goryung (GG) weir

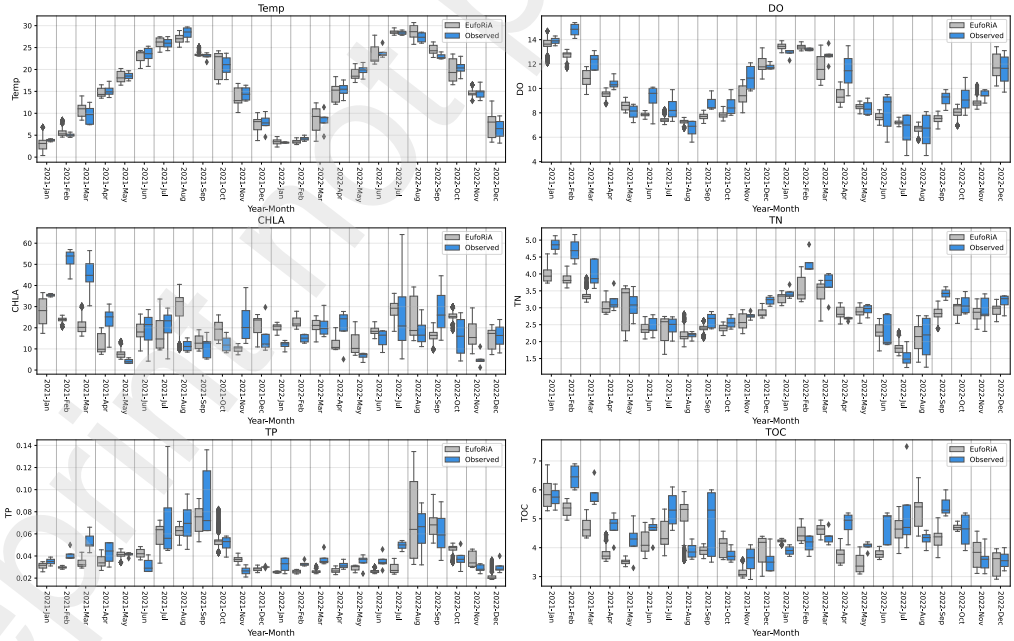


Figure S30: Monthly boxplot for Dalsung (DS) weir

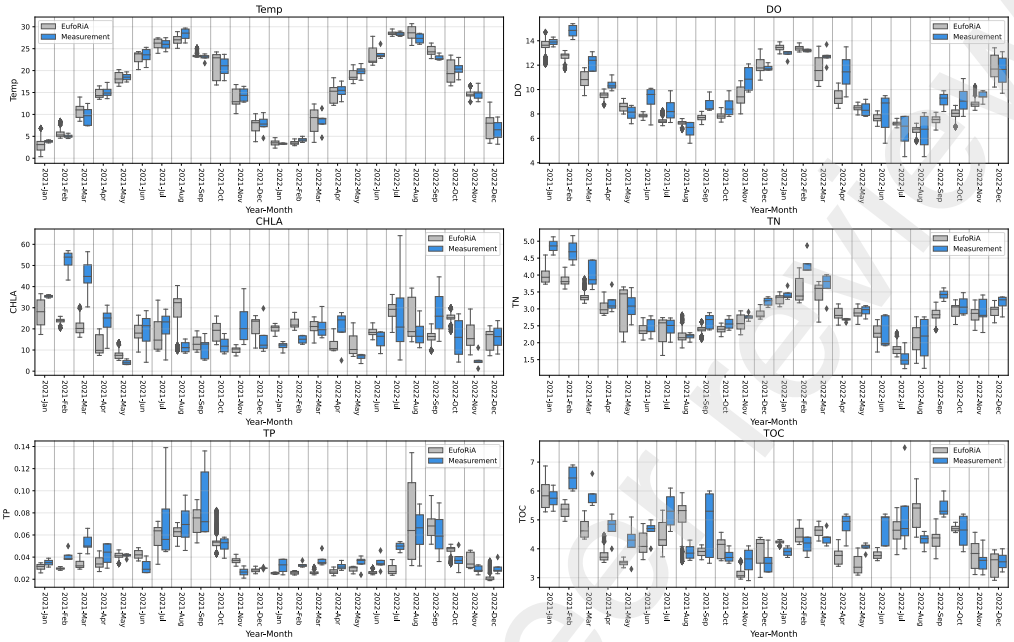


Figure S31: Monthly boxplot for Hapcheon-Changnyung (HC) weir

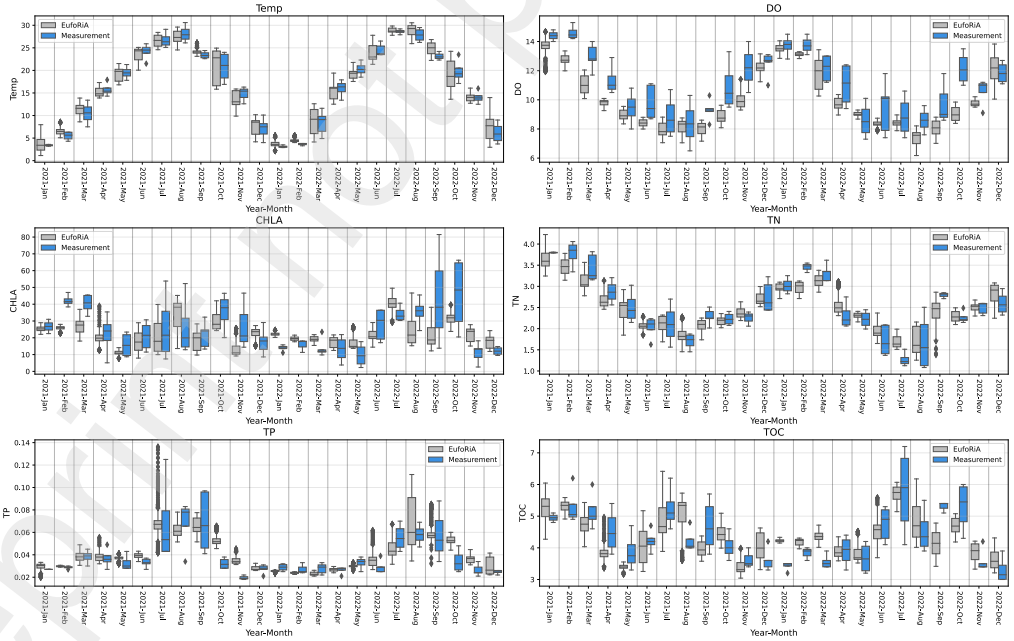


Figure S32: Monthly boxplot for Changnyung-Haman (CH) weir

S20. The normalized target diagrams

The target diagram was introduced by Jolliff et al. [62] to visually depict how both the pattern statistics and the bias (mean value differences) contribute to the magnitude of the total Root-Mean-Square Difference (RMSD).

In the normalized target diagram, y-axis represents the normalized bias as follows:

$$B^* = \frac{\overline{m} - \overline{r}}{\sigma_r} \tag{S.55}$$

Where, B^* is the normalized bias, $\overline{m}$ is the mean of model output values, $\overline{r}$ is the mean of observation values, σ_r is the standard deviation of observation values.

In addition, the x-axis represents the normalized unbiased RMSD calculated by:

$$RMSD' = \left(\frac{1}{N} \Sigma_{n=1}^{N} \left[(m_n - \overline{m}) - (r_n - \overline{r}) \right]^2 \right)^{0.5} \text{sign}(\sigma_m - \sigma_r)/\sigma_r \tag{S.56}$$

Where, $RMSD'$ is the normalized unbiased RMSD, m_n is the nth model output value, r_n is the nth observed value, σ_m is the standard deviation of the model output values.

Finally, RMSD can be calculated using the following expression.

$$RMSD = \sqrt{(B^*)^2 + (RMSD')^2} \tag{S.57}$$

In this section, we present normalized target diagrams comparing uncorrelated versus correlated DA cases for major water quality constituents across seven DA scenarios. Each point represents one of 66 observation points depicted in each graph. The color of each point indicates its RMSD value according to the legend provided.

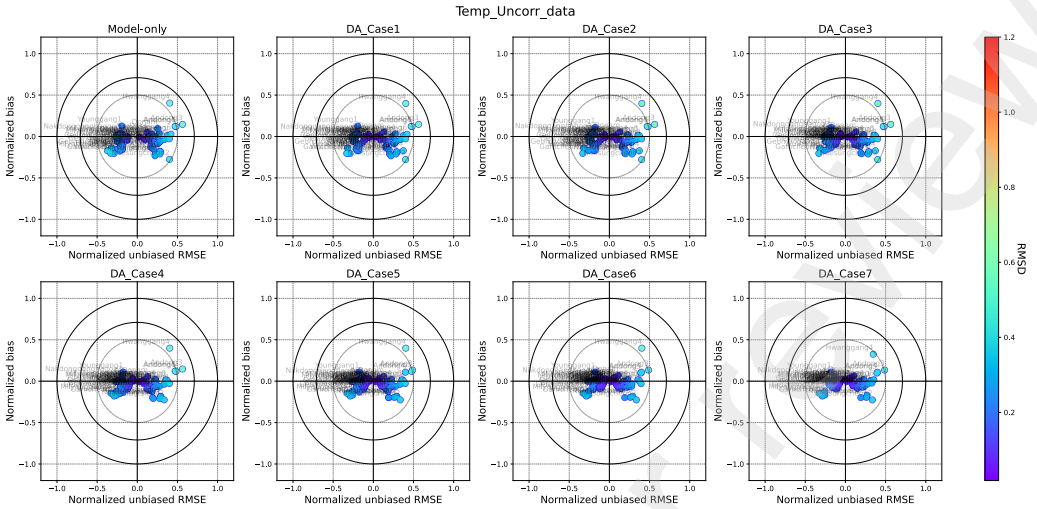


Figure S33: Normalized target diagrams for temperature model: spatially uncorrelated cases

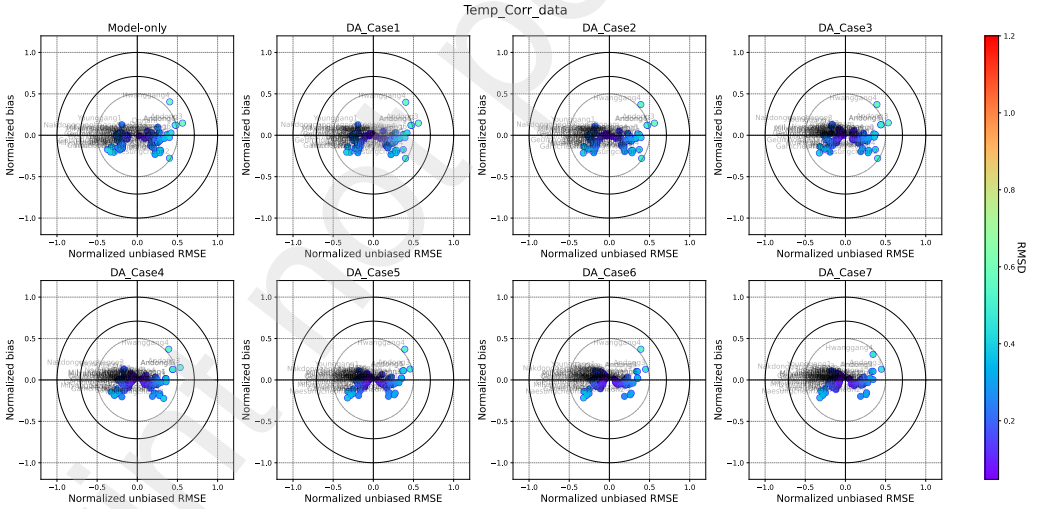


Figure S34: Normalized target diagrams for temperature model: spatially correlated cases

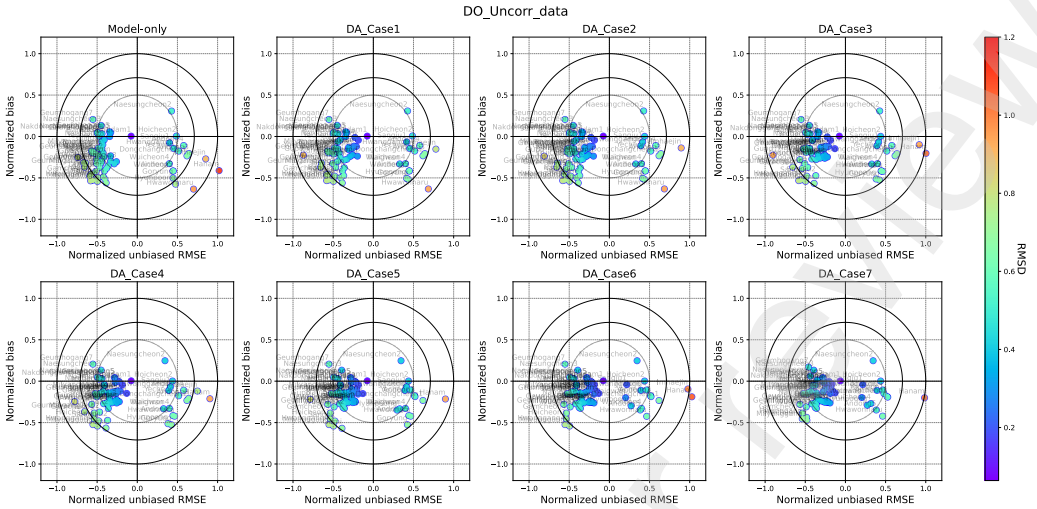


Figure S35: Normalized target diagrams for dissolved oxygen (DO) model: spatially uncorrelated cases

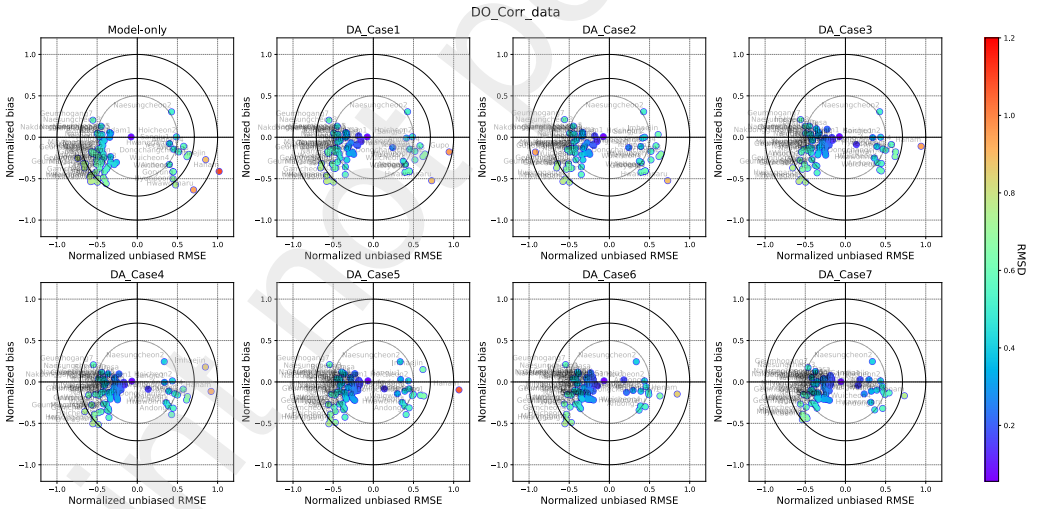


Figure S36: Normalized target diagrams for dissolved oxygen (DO) model: spatially correlated cases

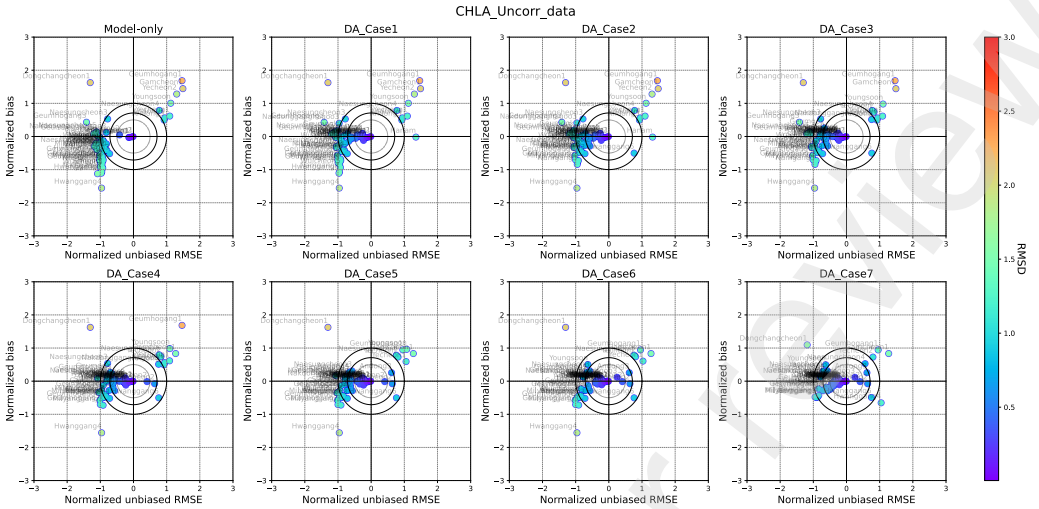


Figure S37: Normalized target diagrams for Chlorophyll-a (Chl-a) model: spatially uncorrelated cases

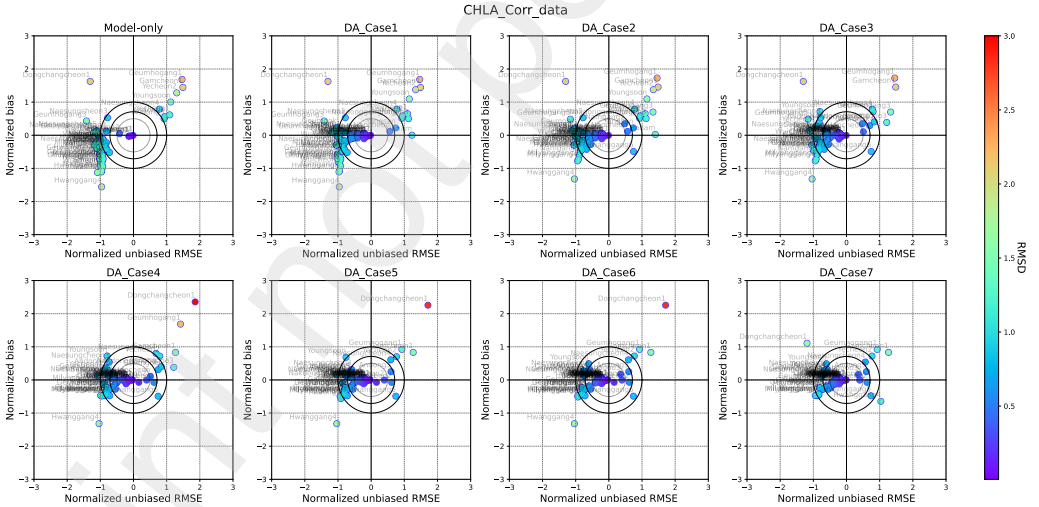


Figure S38: Normalized target diagrams for Chlorophyll-a (Chl-a) model: spatially correlated cases

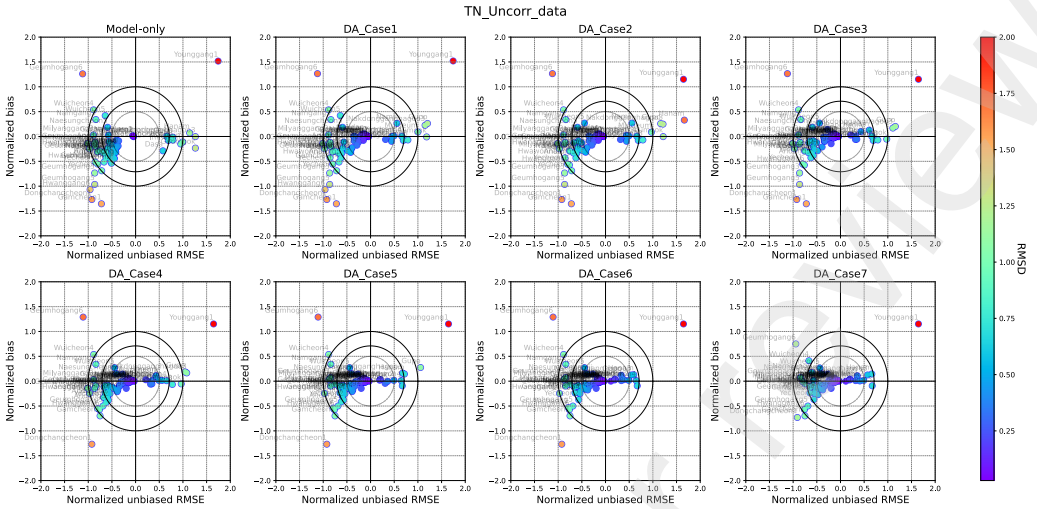


Figure S39: Normalized target diagrams for total nitrogen (TN) model: spatially uncorrelated cases

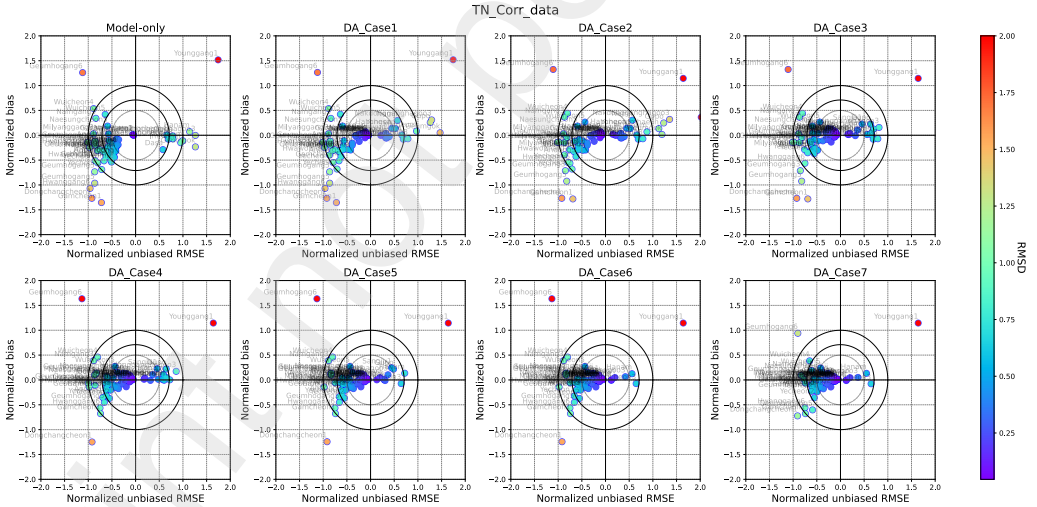


Figure S40: Normalized target diagrams for total nitrogen (TN) model: spatially correlated cases

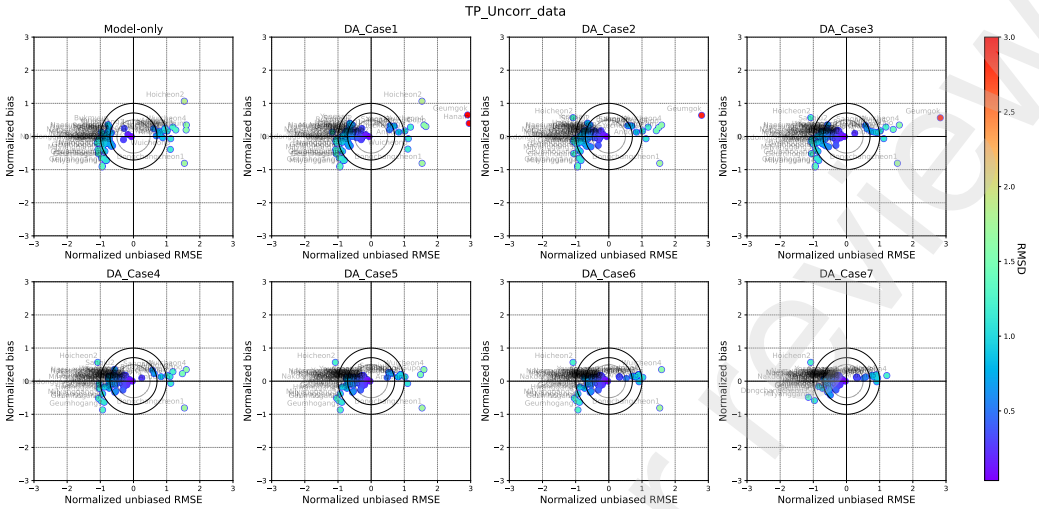


Figure S41: Normalized target diagrams for total phosphorus (TP) model: spatially uncorrelated cases

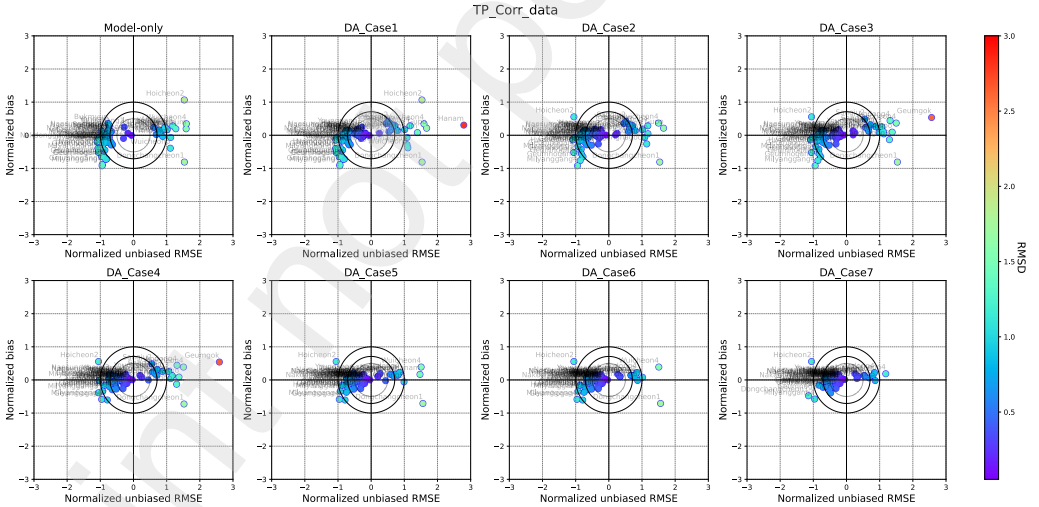


Figure S42: Normalized target diagrams for total phosphorus (TP) model: spatially correlated cases

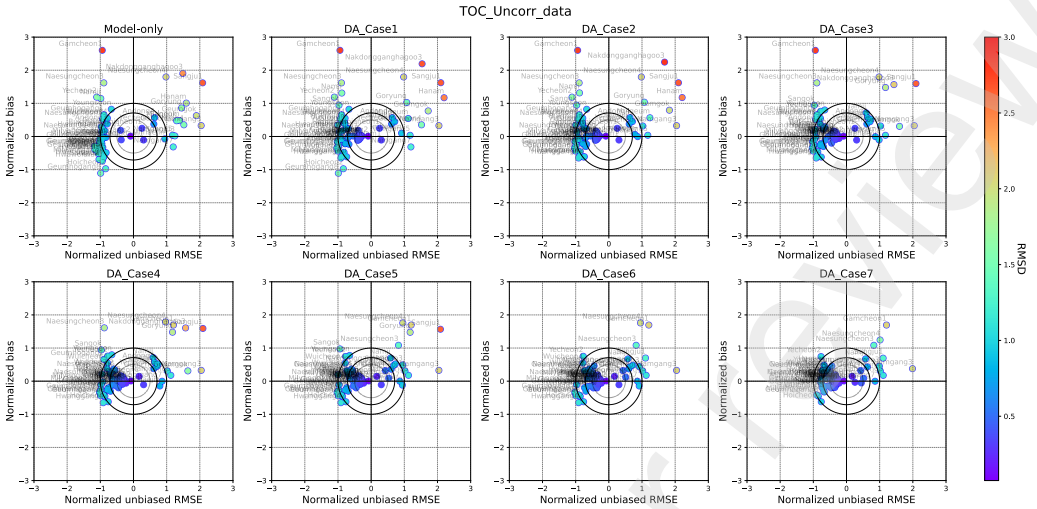


Figure S43: Normalized target diagrams for total organic carbon (TOC) model: spatially uncorrelated cases

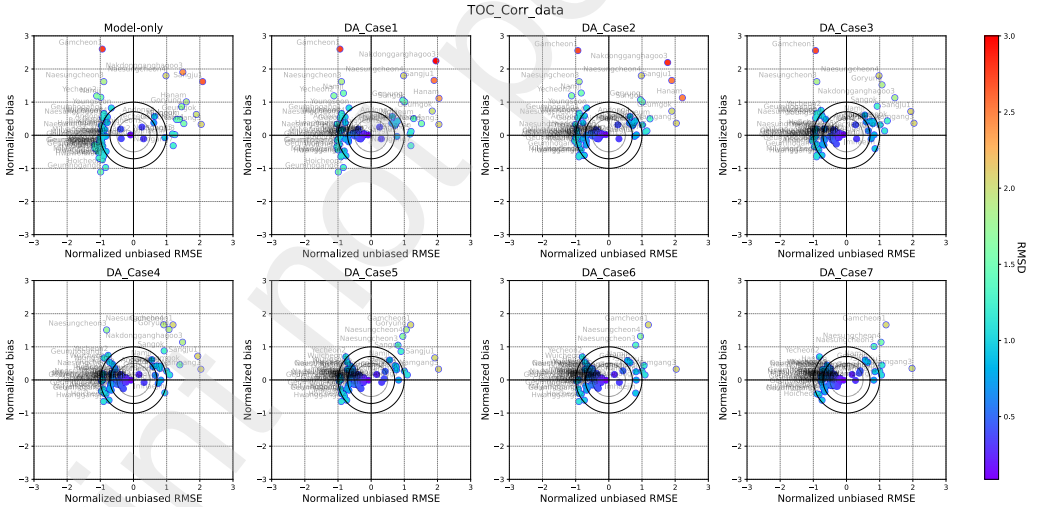


Figure S44: Normalized target diagrams for total organic carbon (TOC) model: spatially correlated cases

S21. RMSE maps for all superjunctions by DA scenarios

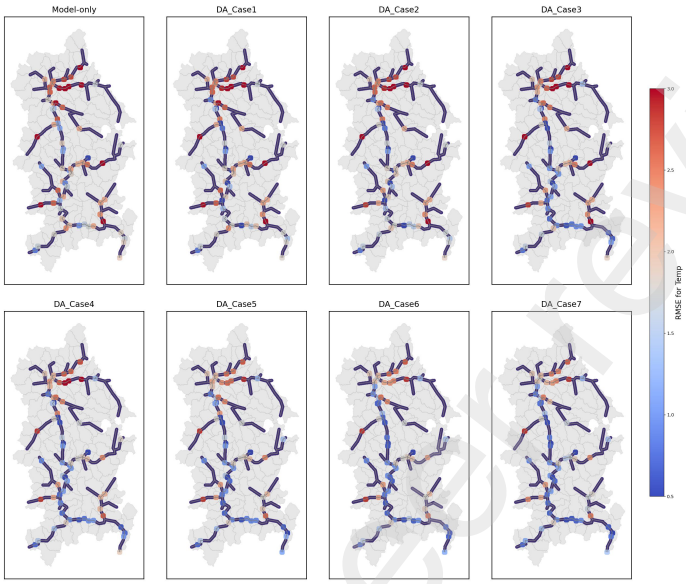


Figure S45: RMSE map for all superjunctions by DA scenarios: Temperature

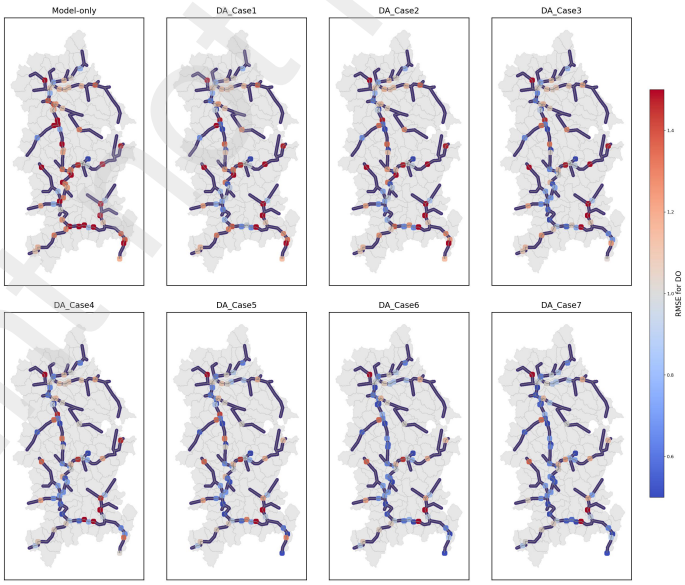


Figure S46: RMSE map for all superjunctions by the DA scenarios: Dissolved oxygen

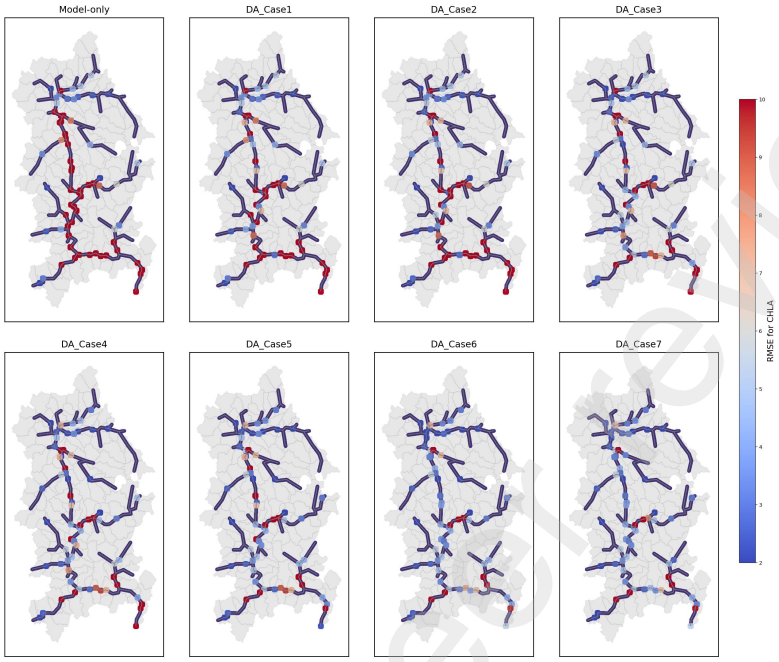


Figure S47: RMSE map for all superjunctions by the DA scenarios: Chlorophyll-a

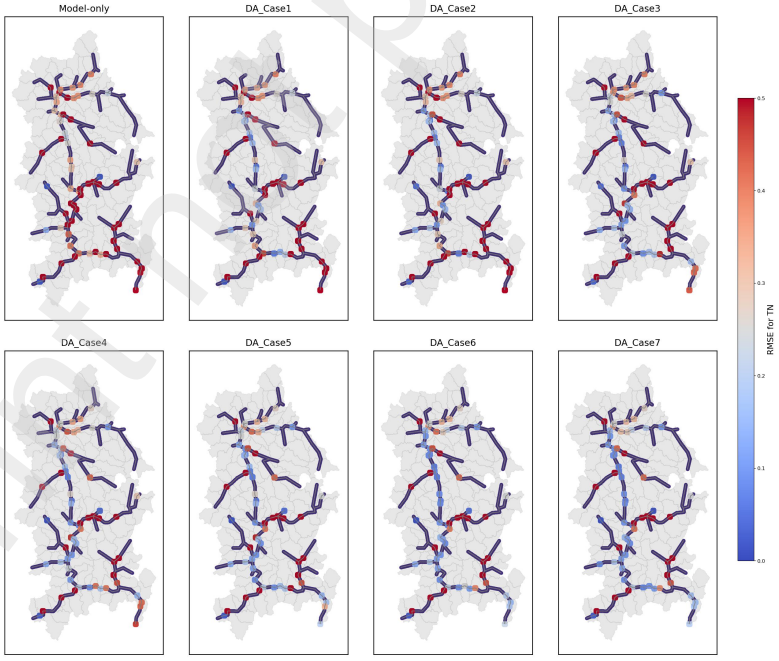


Figure S48: RMSE map for all superjunctions by the DA scenarios: Total nitrogen

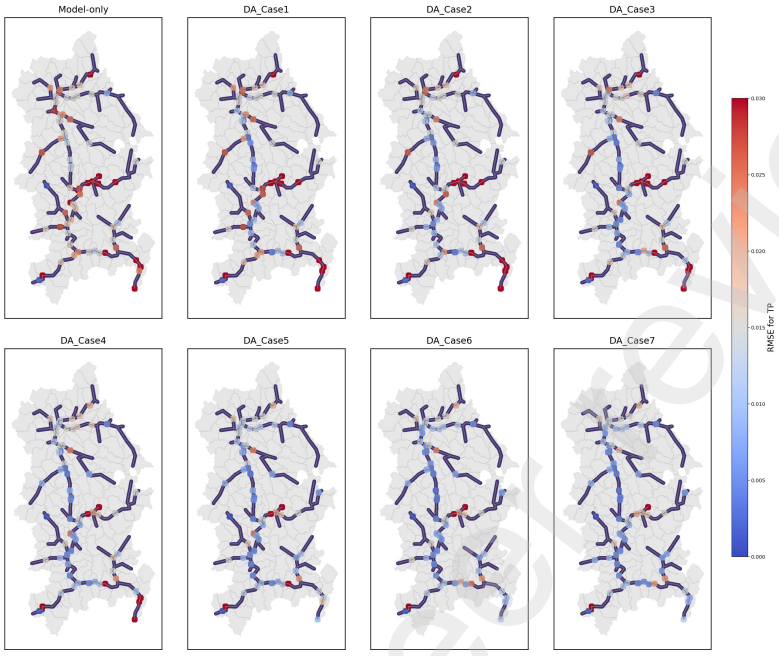


Figure S49: RMSE map for all superjunctions by the DA scenarios: Total phosphorus

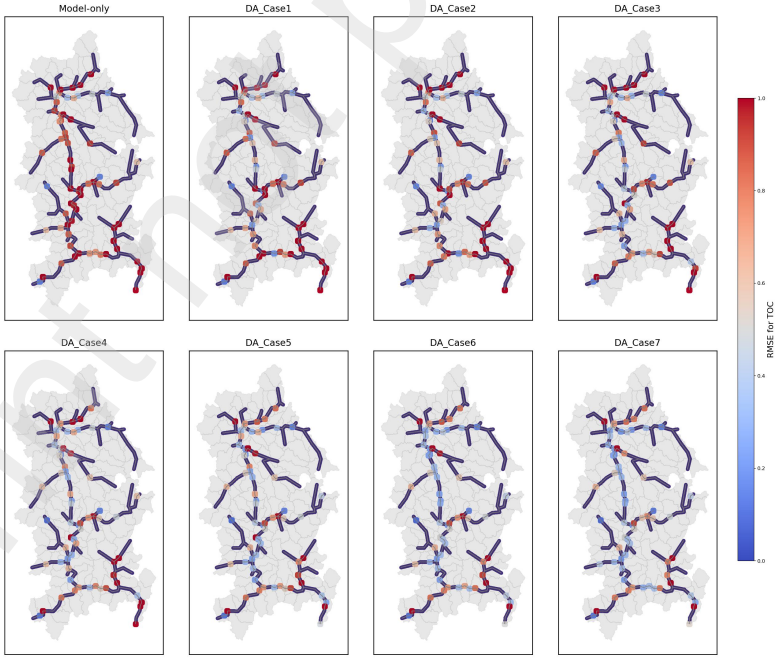


Figure S50: RMSE map for all superjunctions by the DA scenarios: Total organic carbon