Recalibration of a three-dimensional water quality model with a newly developed autocalibration toolkit (EFDC-ACT v1.0.0): does the model calibrated with a wider hydrological variability become more robust?

5 Chen Zhang, Tianyu Fu

State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300350, China

Correspondence to: Chen Zhang (emil@tju.edu.cn)

Abstract. Over recent years, the dramatic hydrological variability has challenged the robustness of process-based hydrodynamic and water quality models to long-term continuous biogeochemical processes. In this study, we developed a

10 new R-based autocalibration toolkit for the Environmental Fluid Dynamics Code (EFDC) and implemented it into the recalibration of the Yuqiao Reservoir Water Quality Model (YRWQM) with long-term observations from 2006 to 2015, including dry, normal, and wet years. The autocalibration toolkit facilitated recalibration and contributed to exploring how the model recalibrated with long-term observations performs more accurately and robustly. Previously, the original YRWQM was calibrated and validated with observations of dry years in 2006 and 2007, respectively. Compared to the

15 original YRWQM, the recalibrated YRWQM performed as well in water surface elevation with a Kling-Gupta Efficiency (KGE) of 0.99 and water temperature with a KGE of 0.91, while better in modeling total phosphorus (TP), chlorophyll *a* (Chl *a*), and dissolved oxygen (DO) with KGEs of 0.10, 0.30, and 0.74 respectively. Furthermore, to analyze the impact of different hydrological years as calibration periods on the model accuracy and robustness, we also calibrated the model in dry, normal, and wet years. The model calibrated in dry years overestimated DO concentrations, probably explained by the

20 parameter of algal growth rate increased by 84%. The model calibrated in wet years performed poorly for Chl *a* due to a 50% reduction in the carbon-to-chlorophyll ratio probably triggered by changes in the composition of the algal population. Our study suggests that calibrating process-based hydrodynamic and water quality models with long-term observations may be an important measure to improve the robustness of models under severe hydrological variability. The newly developed general automatic calibration toolkit and a possible hierarchical autocalibration strategy will also be a powerful tool for

25 future complex model calibration.

1 Introduction

Lakes and reservoirs fulfill the role of "sentinels" to climate change due to both their capacity to buffer synoptic-scale hydroclimatic extremes and their susceptibility to hydrological variability (Adrian et al., 2009; Williamson et al., 2009;

Mooij et al., 2019). In recent decades, dramatic hydrological variability has been widely detected and remarkably influenced

30   biogeochemical processes in lakes and reservoirs (Sinha et al., 2017; Grant et al., 2021; Kong et al., 2022; Salk et al., 2022). In a bid to delve into these variations, process-based hydrodynamic and water quality models have been increasingly popular tools since they can disentangle numerous intricate causal relations between exogenous drivers and environmental impacts within water bodies (Arhonditsis and Brett, 2004; Mooij et al., 2010; Fu et al., 2019). However, the accuracy and robustness of these models in the face of such intense hydrological variability have become a key issue.

35   Driven by the purpose of better understanding physical, chemical, and biological processes, the complexity of process-based hydrodynamic and water quality models continues unabated over recent years (Robson, 2014). However, the increased complexity of the model is a mixed blessing. It indeed helps us to examine biogeochemical processes in lakes and reservoirs, but when the complexity of the model exceeds a certain level, both the accuracy and the identifiability are diminished (McDonald and Urban, 2010). Higher dimensions, more state variables, and more specific details introduce more and more

40   parameters into models through drastic simplification of reality, which subsequently become a massive source of model uncertainty. Model calibration is one of the essential procedures in model setup to reduce the uncertainty from parameter estimations and to obtain a satisfactory parameter set to match simulated results with observed data (Jørgensen and Fath, 2011). Although calibration has been used extensively in process-based hydrodynamic and water quality models, there are still two notable problems.

45   The first problem is that the conventional manual calibration method commonly used in process-based hydrodynamic and water quality models is inefficient and does not guarantee optimal results. Firstly, it puts modelers inexperienced or unfamiliar with the model through the wringer due to time-consuming and tedious tasks. Secondly, the parameter set selected by this method may still suffer from dramatic uncertainty and interferences of subjective factors. With the development of computer technology and subsequent application in numerical simulation methods, the automatic calibration

50   method is burgeoning (Shimoda and Arhonditsis, 2016). Numerous modeling studies in recent decades have employed automatic calibration procedures in 2-D or lower-dimensional process-based hydrodynamic and water quality models in lakes or reservoirs (Rigosi et al., 2011; Huang, 2014; Luo et al., 2018). However, due to their high complexity and time-consuming calculation, there are few applications of automatic calibration procedures in 3-D hydrodynamic and water quality models. For example, the automated Parameter ESTimation software (PEST) was applied in the Environmental Fluid

55   Dynamic Code (EFDC) (Arifin et al., 2016), and optimization algorithms were applied in Delft3D (Xia et al., 2022; Xia and Shoemaker, 2021, 2022). The lack of automatic calibration procedures is a major hindrance to improving the efficiency of model calibration and indirectly causes the problem below.

The second problem is that many models were calibrated with only short-term observations with a narrow hydrological variability, then were employed in water ecosystem management decisions or future predictions. There are hidden perils in

60   the presumption that these calibrated models are adaptive to a wider hydrological variability. For example, in a previous study of the Spokane River and Lake Spokane Model, the use of a low flow period for calibration may result in an overestimation of inlake total phosphorus (TP) and chlorophyll a (Chl a), and an underestimation of minimal dissolved

oxygen (DO) (Zhang et al., 2018a). This issue has also arisen in studies of other models (Vaze et al., 2010; Nielsen et al., 2014; Basijokaite and Kelleher, 2021). A large number of parameters is almost impossible to be constrained by a narrow

65   hydrological variability (Janssen and Heuberger, 1995; Franks, 2009), thus triggering the equifinality problem, where several distinct parameter inputs produce the same model outputs called "good results for the wrong reasons" (Arhonditsis et al., 2007; Paudel, 2012). Even if the final parameter set chosen by the modeler satisfies the match between model results and observations under the current hydrological variability, there is no credibility that the model will be accurate as a robust prognostic tool under a wider hydrological variability (Arhonditsis et al., 2007). Therefore, the utilization of a longer period

70   of observations containing a wider range of hydrological years for calibration may be an important way to improve the identifiability of parameters in process-based hydrodynamic and water quality models. Benefiting from the continued accumulation of historical observations, numerous published models have been recalibrated using longer scales of data (James, 2016; Schnedler-Meyer et al., 2022). Cerco and Noel (2005) recalibrated the Chesapeake Bay model with a decade of observations resulting in a clear improvement in modeling primary production and light attenuation. Lung and Nice (2007)

75   recalibrated the Patuxent Estuary model with two years of data and adequately modeled Chl *a* and DO concentrations. Benefit from long-term data supports, exploring how the hydrological variability in the calibration period impact model calibration will help establish more accurate and robust models.

Long-term modeling is required to test the model's ability to reproduce ecosystems under different hydrological years. Based on the established Yuqiao Reservoir Water Quality Model (YRWQM) (Zhang et al., 2013, 2015, 2019), this study

80   aims to explore how long-term observations under a wider hydrological variability impact model calibration with the application of automatic calibration techniques. We hypothesized that models using observations with a longer period and wider hydrological variability for calibration perform more accurately and robustly. We first developed a new R-based autocalibration toolkit for the EFDC model. Second, we recalibrated the YRWQM by the toolkit with a decadal-scale observation under three hydrological situations: dry, normal, and wet periods. Finally, we compared the model performance,

85   parameters, and kinetic processes represented by parameters across the model calibration scenarios that used different split-sample approaches. These discrepancies will highlight the importance of the hydrological variability corresponding to the observed data for model calibration and deepen our understanding of biogeochemical processes in shallow lakes and reservoirs under the wide hydrological variability.

2 Materials and methods

90   ### 2.1 Study area

The Yuqiao Reservoir (40°00'-40°04'N, 117°26'-117°37'E) is situated in Jixian County, Tianjin, China (Fig. 1). The shallow reservoir has a length of 66 km from east to west and a width of 50 km from north to south, an average water depth of 4.74 m, a maximum water depth of 12.74 m, a total surface area of 86.6 km^2, and a storage capacity of 1.559 billion m^3. It

is situated within a basin area that covers 2060 km^2 (Fig. S1). The main tributaries of Yuqiao Reservoir include the Lin River,

95   Sha River, and Li River, and the outflow of the reservoir is through the Zhou culvert (Yu and Zhang, 2021).

Since the completion of the Luan River-Tianjin Water Diversion Project in 1983, the Yuqiao Reservoir has become the primary source of drinking, agricultural and industrial water for approximately 129 villages in the surrounding area. The water introduced into the reservoir from the Luan River through the Li River, the main channel for diverting water, has alleviated the water shortage problem in Tianjin City (Zhang et al., 2019). Previous studies have shown that Yuqiao

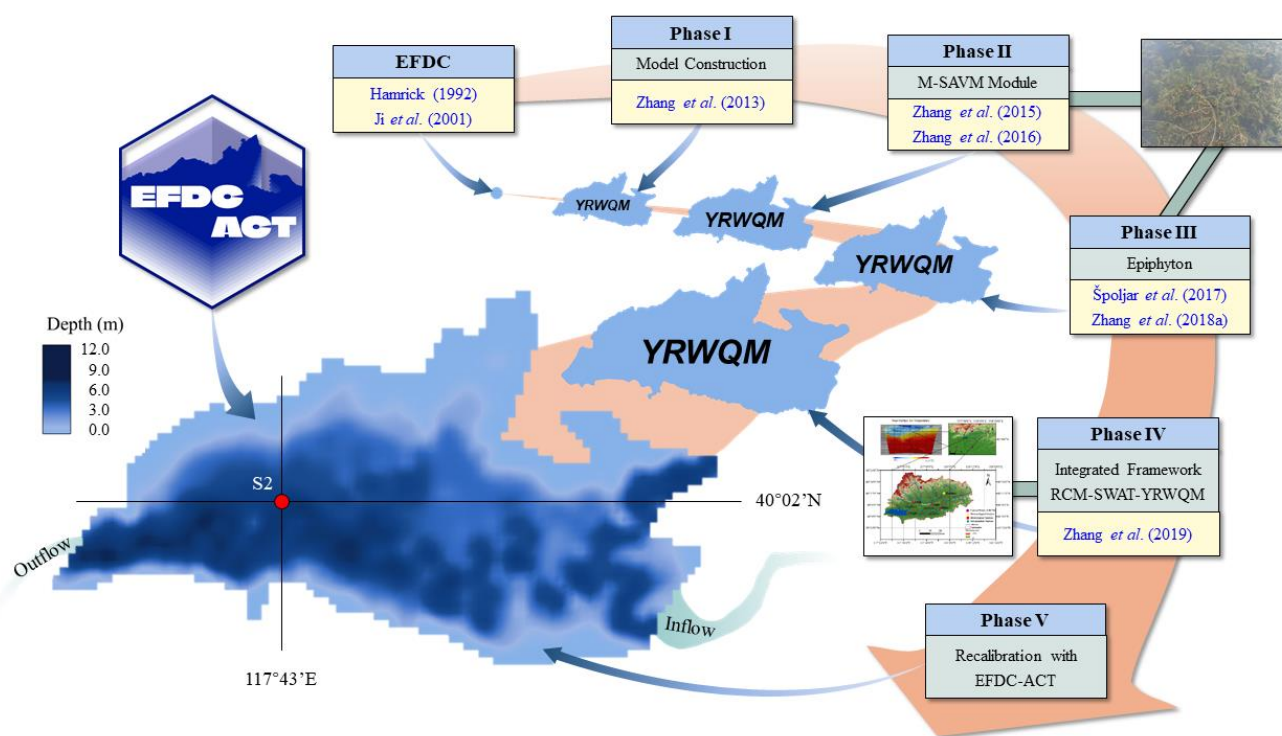
100   Reservoir is a typical mesotrophic, phosphorus-limited environment and that human-originated activities forcing (water diversions) obtained the maximum TP concentration contribution index (40%) (Chen et al., 2012; Zhang et al., 2020). If the phosphorus loading exceeds the threshold (150 μg P/L), eutrophication may cause a switch from the clear-water state to the turbid water state, and its restoration will be quite difficult (Zhang et al., 2016). Hence, the water quality of Yuqiao Reservoir is indeed worthy of attention.

105

Figure 1: Bathymetric map of the Yuqiao Reservoir and the chronicle of YRWQM. The YRWQM has undergone five phases of development and refinement: (I) Model construction; (II) M-SAVM module; (III) Epiphyton; (IV) Integrated framework RCM-SWAT-YRWQM; (V) Recalibration with EFDC-ACT.

2.2 Chronicle of YRWQM

110 The YRWQM (Zhang et al., 2013) is a regional hydrodynamic and water quality model developed under the framework of EFDC (Hamrick, 1992; Ji et al., 2001) to improve the understanding and management of the Yuqiao Reservoir. Since its inception, the model has undergone five phases of development and refinement (Fig. 1).

The original YRWQM was constructed to investigate how agricultural pollution by flood flows affects the water quality in the Yuqiao Reservoir. The model was calibrated with observations from 1 January to 31 December 2006 (one year) and

115 validated with observed data from 1 May to 31 October 2007 (six months) (Zhang et al., 2013). Subsequently, the YRWQM was coupled with a modified submerged aquatic vegetation model (M-SAVM) to study the development effect of submerged macrophytes (Zhang et al., 2015, 2016) and epiphyton (Špoljar et al., 2017; Zhang et al., 2018b) on the water quality indicators of the reservoir. An integrated climate-hydrological-water quality (RCM-SWAT-YRWQM) framework was also proposed to elucidate the effects of a changing climate on the trophic state (Zhang et al., 2019).

120 The YRWQM has become a powerful tool for research and management of the Yuqiao Reservoir water ecosystem through the above phases in the past decade. However, there may still be a risk of insufficient accuracy since the original YRWQM calibration does not take into account long-term hydrological variability. With the availability of the decadal-scale observations covering dry, normal, and wet periods and the design of new model calibration methods, it is time to examine whether a longer period of calibration can improve the accuracy and robustness of the model.

2.3 Data sources

Meteorological data, including atmospheric pressure, dry atmospheric temperature, relative humidity, wind speed, and wind direction, are available from China Meteorological Data Service Centre (http://data.cma.cn/). Discharge, precipitation, evaporation, and water surface elevation data used in the model were obtained from the Yuqiao Reservoir Administrative Bureau. All the data above were collected at a frequency of once a day from 2006 to 2015, except for 2012 when water

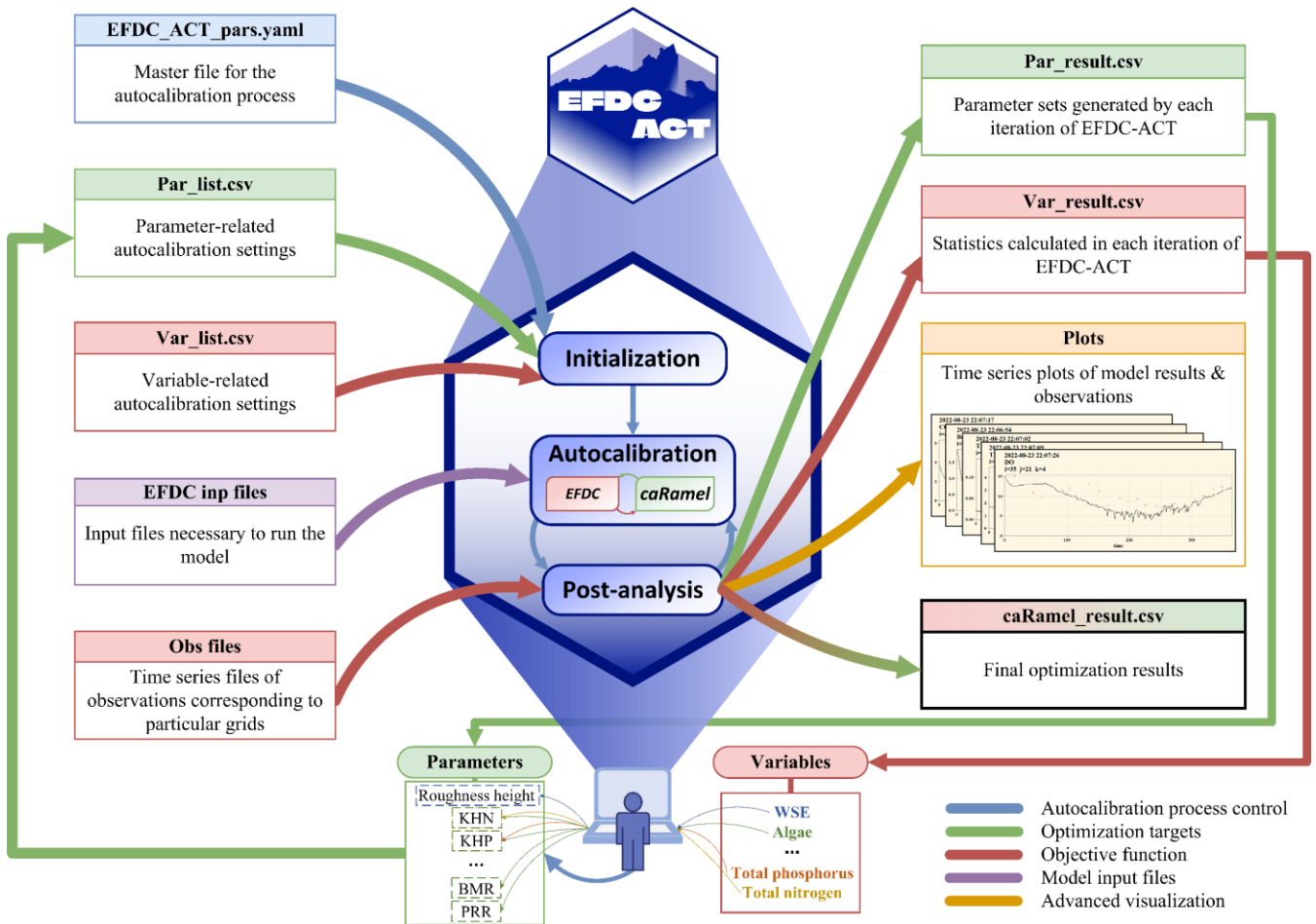
130 surface elevation data were collected once a month (Zhang et al., 2019).

There are seven monitoring stations located at the Yuqiao Reservoir, six of which were employed by the original YRWQM for calibration and validation (Zhang et al., 2013). In this study, to balance the cost with accuracy in calibration, water temperature and water quality data collected from monitoring station S2 were used in model evaluation, which represented the water column at the center of the Yuqiao Reservoir. Station S2 was employed to collect water temperature

135 and water quality data including TP, Chl *a*, and DO concentrations since 1989 (Zhang et al., 2015). All the water quality data was sampled, preserved, and analyzed monthly or semi-monthly from 2006 to 2015 according to the Standard Method for the Examination of Water and Wastewater Editorial Board (2002).

2.4 EFDC-Automatic Calibration Toolkit

EFDC-Automatic Calibration Toolkit (EFDC-ACT) was developed in this study for automating the calibration of a 3-D

140 hydrodynamic and water quality model, EFDC, with more than two hundred parameters. EFDC-ACT is a multi-parameter and multi-variable autocalibration toolkit based on R. Documentation and source code is shared and publicly available at https://doi.org/10.5281/zenodo.7438143. The conceptual overview of the EFDC-ACT is shown in Fig. 2. There are three main steps in EFDC-ACT: initialization, autocalibration, and post-analysis.

145 **Figure 2: Conceptual overview of EFDC-ACT. The main files consist of nine categories: (a) EFDC_ACT_pars.yaml: the EFDC-ACT master file; (b) Par_list.csv: parameter ranges and whether to calibrate; (c) Var_list.csv: objective state variable, spatial location, statistics used and accuracy expected; (d) EFDC inp files: input files for EFDC; (e) Obs files: CSV files containing observations; (f) Par_result.csv: autocalibration results of parameter values by each iteration; (g) Var_result.csv: autocalibration results of statistics for objective state variables by each iteration; (h) Plots: Time series plots for model results and observations; (i)**
150 **caRamel_result.csv: final autocalibration results containing optimal parameter sets and model evaluations.**

Before using EFDC-ACT, the user should prepare the necessary files including the EFDC-ACT master file, the Comma-Separated Values (CSV) file containing the parameter and variable information, the input file for the EFDC, and the

CSV file containing the observations. In the initialization step, EFDC-ACT checks and loads the master file, the parameter list, and the variable list. Then EFDC-ACT generates a matrix of parameter value ranges, sets the model evaluation statistics

155 as the objective functions, and launches the autocalibration process. More detail is in EFDC-ACT User's Guide (S1).

To maintain the diversity of the parameter sets while accelerating convergence, EFDC-ACT introduced the *caRamel* package (Monteil et al., 2020) in the autocalibration step. The *caRamel* package is a genetic algorithm-based multi-objective optimizer, incorporating the Multiobjective Evolutionary Annealing Simplex method (MEAS) (Efstratiadis and Koutsoyiannis, 2008) and Nondominated Sorting Genetic Algorithm II (ε-NSGA-II) (Reed and Devireddy, 2004). It is

160 suitable for highly complex, time-consuming hydrodynamic and water quality models like EFDC. EFDC-ACT controls the *caRamel* optimization according to the master file, feeding the parameter value range matrix into the *caRamel*. The *caRamel* generates the parameter set, then EFDC-ACT passes it into EFDC and starts the calculation. At the end of the model run, EFDC-ACT calculates the statistics based on the modeled and observed values and passes the result back to *caRamel* as the objective function value. As the parameter set is adjusted, the autocalibration process is repeated until the termination

165 criterion is reached, such as the maximum number of runs or the expected statistic results.

Automatic model evaluation is a potent tool for making models more transparent and credible (Alexandrov et al., 2011; Soares and Calijuri, 2021). To this end, EFDC-ACT provides model result extraction, statistical model evaluation, and graphical model evaluation in the post-analysis step. During the autocalibration process, the user can open the CSV files to view each iteration's parameter set and model evaluation results. After each iteration, EFDC-ACT plots the time series using

170 modeled and observed values, thus supplying the users with a visual comparison that statistics cannot afford. EFDC-ACT will also output the final optimization results in a CSV file when all iterations are complete.

According to the model evaluation guidelines proposed by Moriasi et al. (2007), the statistics used to evaluate the model performance include three categories: standard regression (R^2), dimensionless (NSE, KGE), and error index (MAE, RMSE, PBIAS, RSR). Kling-Gupta Efficiency (KGE) is included as an alternative to NSE in this study. KGE gives equal

175 weight to bias, linear correlation, and variability, avoiding the systematic underestimating of the variability (Gupta et al., 2009).

2.5 YRWQM recalibration with EFDC-ACT

Due to the lower stability of the hydrodynamic model compared to the water quality model, the recalibration of the YRWQM was divided into two parts: the hydrodynamic model recalibration and the water quality model recalibration. Both

180 parts of the YRWQM were recalibrated with EFDC-ACT. The parameter ranges listed in Table S1 were referenced from the original YRWQM and other literature (Wu and Xu, 2011; Zhang et al., 2013; Yi et al., 2016; Jiang et al., 2018; Zhao et al., 2020; Kim et al., 2021). KGE and PBIAS were used to evaluate the recalibrated model. Model performance is considered satisfactory when the KGE is greater than -0.41 in this study, meaning that the model improves upon the mean value benchmark (Knoben et al., 2019). The PBIAS describes the average tendency for simulated values to be greater or less than

185 observed values, with positive values indicating a model bias toward underestimation and negative values indicating a model
bias toward overestimation (Gupta et al., 1999).

The hydrodynamic model of YRWQM was recalibrated with the field data collected between 2006 and 2015 with a time step of 10 seconds. The objective function of EFDC-ACT included the KGE results of water surface elevation (WSE) and surface water temperature (TEM) at station S2 from 2006 to 2015. The decade included dry years (2006, 2007, 2010,
190 and 2015), normal years (2009, 2011, and 2014), and wet years (2008, 2012, and 2013). The parameters were automatically adjusted by EFDC-ACT until all the objective functions (KGEs) were greater than -0.41 or the number of iterations reached the maximum. In the water quality model recalibration, the objective function of EFDC-ACT included the KGE results of three water quality state variables (TP, Chl *a*, and DO) at stations S2 from 2006 to 2015. When all the objective functions (KGEs) were greater than -0.41 or the number of iterations exceeded the maximum, the autocalibration is considered
195 complete.

3 Results

3.1 Model recalibration

During the recalibration period (2006-2015), the bottom roughness height and the wind-drag multiplier were automatically calibrated. The hydrodynamics of the recalibrated YRWQM demonstrated good performance for WSE and TEM at station
200 S2 (Fig. 3). The recalibrated YRWQM remarkably reproduced the decadal variation of WSE in Yuqiao Reservoir with a KGE of 0.99. The seasonal cycle of water temperatures was pronounced, with higher temperatures in summer and lower temperatures in winter. The recalibrated YRWQM reproduced this characteristic with a KGE of 0.91. The modeled WSE and TEM both indicated that the hydrodynamic model in the recalibrated YRWQM is reliable and can be used for water quality modeling in the Yuqiao Reservoir during the recalibration period.
205 The water quality of the recalibrated YRWQM performed satisfactorily for the modeled TP concentration at station S2 with a KGE of 0.10. Most of the observations were evenly distributed with little variances on either side of the modeled values, with only a few observations away (Fig. 3). The modeled TP concentration peaked at the end of 2010 and 2011 and was beyond the range of observations. Nevertheless, the inter- and intra-annual variability in TP concentrations were still well captured and the model showed acceptable performance overall with a PBIAS of 40%. The model represented the
210 variation of Chl *a* concentration over the decade with a KGE of 0.30. Chl *a* concentration showed a clear double-peaked or multi-peaked pattern of intra-annual variation, with peaks occurring mostly in spring and autumn (Fig. 3). The modeled DO concentrations likewise showed good performance with a KGE of 0.74. DO concentration exhibited a pronounced seasonal cycle with lower concentrations in summer and higher concentrations in winter (Fig. 3).

The comparisons of the recalibrated model against the original model demonstrated the better accuracy and robustness
215 of the recalibrated YRWQM over the decade in the Yuqiao Reservoir (Table 1). The hydrodynamics of the recalibrated YRWQM performed as well as the original YRWQM, while the recalibrated YRWQM performed better than the original

YRWQM in modeling TP, Chl *a*, and DO concentrations. The water quality of the original YRWQM failed to reach a satisfactory result with KGEs of -1.60, -3.03, and -0.19 for TP, Chl *a*, and DO respectively, while the water quality of the recalibrated YRWQM performed well with KGEs of 0.10, 0.30, and 0.74 respectively as mentioned above. The 11 primary

220   parameters in the water quality model were listed in Table 1, eight of which governed algal kinetics (CChl, PM, Ke_b, TM1, TM2, KHP, BMR, PRR), two parameters influenced phosphorus cycling (K_{RP}, K_{DP}), and one affected reaeration (REAC). Among these parameters we also found six of them to be sensitive (Table 1). These parameters were   carbon-to-chlorophyll ratio for algae (CChl), the maximum growth rate for algae (PM), basal metabolism rate for algae (BMR), predation rate on algae (PRR), the minimum mineralization rate of dissolved organic phosphorus (K_{DP}), and reaeration

225   multiplier (REAC). More detailed model equations, parameter interpretations, and calibration results are listed in Sect. S3.

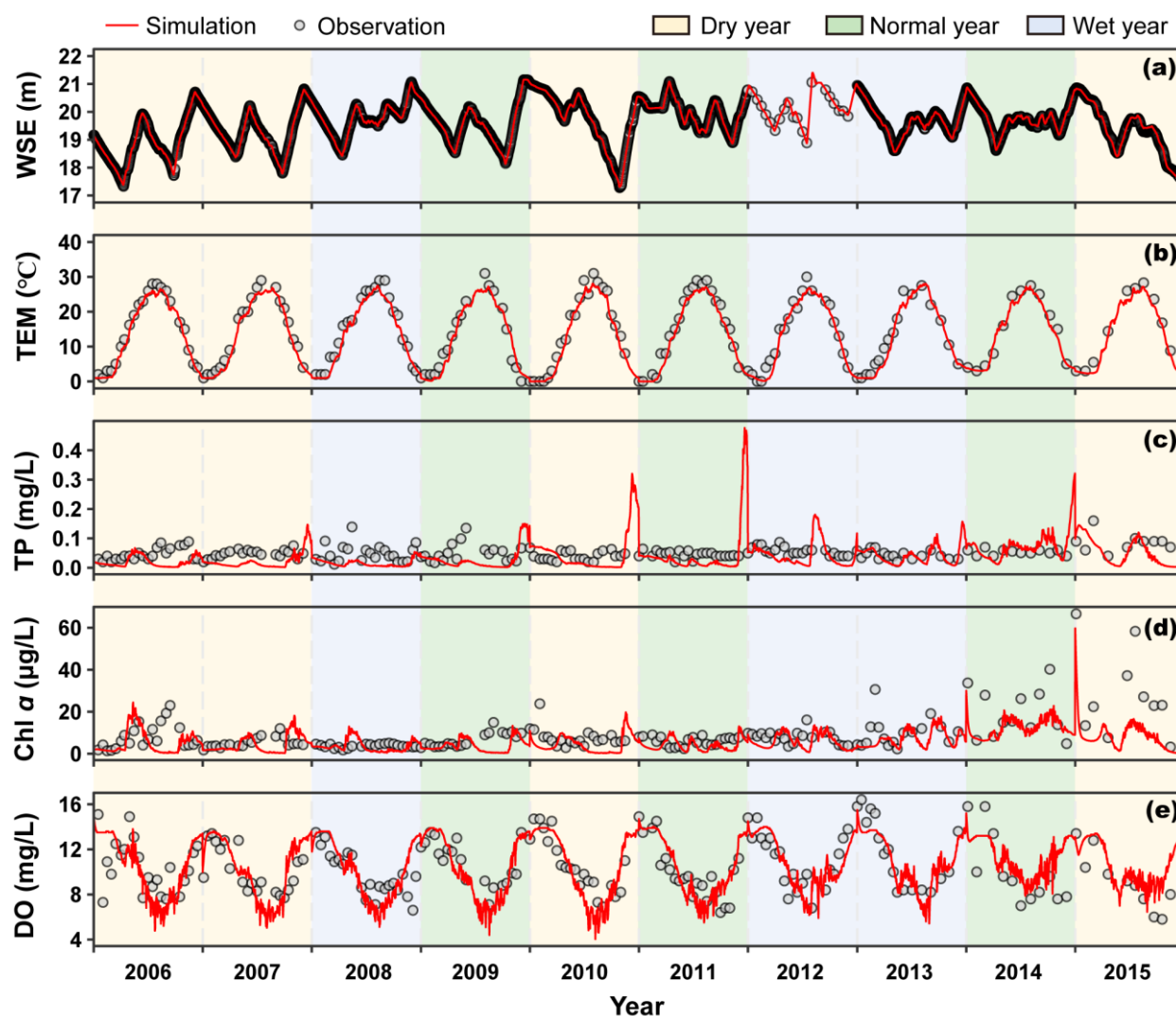



Figure 3: Performance of the recalibrated YRWQM at station S2 of the Yuqiao Reservoir (*n* = 3309 for WSE and 190 for other state variables).

Table 1: Comparison of the recalibrated YRWQM against the original YRWQM. The state variables being compared included WSE, TEM, TP, Chl *a*, and DO. The parameters being compared included eight governing algal kinetics (CChl, PM, Ke$_b$, TM1, TM2, KHP, BMR, PRR), two influencing phosphorus cycling (K$_{RP}$, K$_{DP}$), and one affecting the reaeration of DO (REAC). A more detailed explanation of the model equations and parameters is listed in Sect. S3.

State variables	Statistics	Recalibrated YRWQM	Original YRWQM	
WSE	KGE	0.99	0.99	
	PBIAS(%)	0.03	-0.10	
TEM	KGE	0.91	0.91	
	PBIAS(%)	7.93	8.36	
TP	KGE	0.10	-1.60	
	PBIAS(%)	40	-78	
Chl *a*	KGE	0.30	-3.03	
	PBIAS(%)	36	-357	
DO	KGE	0.74	-0.19	
	PBIAS(%)	-2	-53	

Parameters	Units	Descriptions	Recalibrated YRWQM	Original YRWQM	Sensitivity
CChl	mg C / µg Chl	Carbon-to-Chlorophyll ratio for algae	0.08	0.083	Yes
PM	day^{-1}	The maximum growth rate for algae	2.77	2	Yes
Ke$_b$	m^{-1}	Background light extinction coefficient	0.41	0.475	No
TM1	°C	The lower optimal temperature for algal growth	21	9	No
TM2	°C	The upper optimal temperature for algal growth	28	15	No
KHP	mg/L	Phosphorus half-saturation for algae	0.0019	0.0010	No
BMR	day^{-1}	Basal metabolism rate for algae	0.120	0.010	Yes
PRR	day^{-1}	Predation rate on algae	0.117	0.010	Yes
K$_{RP}$	day^{-1}	Minimum hydrolysis rate of refractory particulate organic phosphorus (RPOP)	0.005	0.001	No
K$_{DP}$	day^{-1}	Minimum mineralization rate of dissolved organic phosphorus (DOP)	0.060	0.005	Yes
REAC	/	Reaeration multiplier	1.4	0.9	Yes

3.2 Performance comparison between YRWQM recalibrated in decade and models calibrated in different hydrological years

There were obvious discrepancies in the performance of the recalibrated YRWQM in different hydrological years (Fig. 4). The ability of the recalibrated YRWQM to reproduce TP concentrations in the decade was the best with the highest KGE values. The recalibrated model evaluation for TP concentrations reflected satisfactory performance in dry and wet years with KGEs of -0.22 and 0.004 respectively, while in normal years the KGE value was less than -0.41. The recalibrated model showed reasonable performance for Chl *a* concentration and performed best in normal years with a KGE of 0.36. The recalibrated model succeeded in reproducing DO concentrations in different hydrological years, with all KGE values greater than 0.6 and the maximum KGE of 0.76 occurring in wet years.

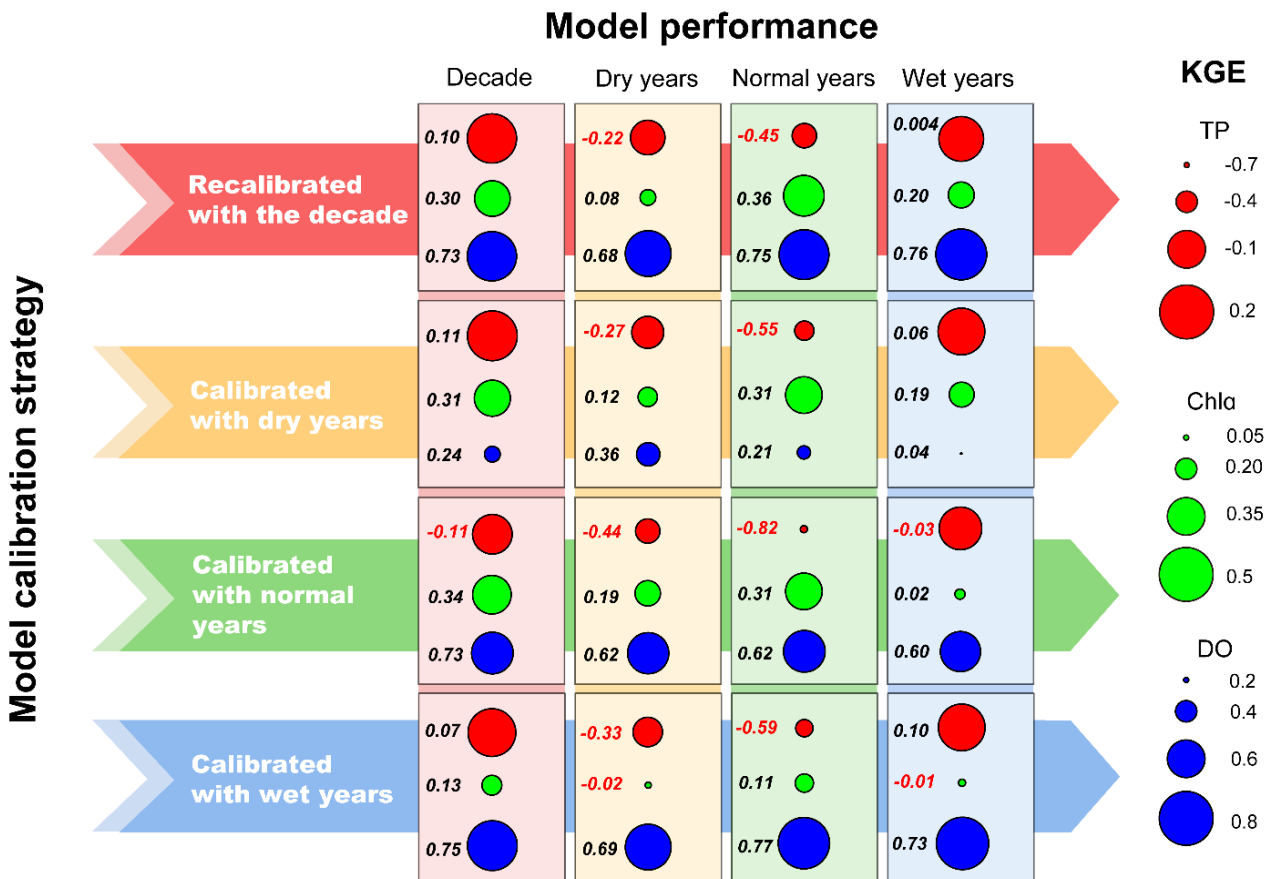



Figure 4: Performance of the models calibrated with different strategies.

In comparison to the YRWQM recalibrated with the decade, the other three models calibrated in different hydrological years showed distinct inferiority (Fig. 4). The model calibrated in dry years performed relatively poor in modeling DO with a KGE of 0.24 over the decade and a maximum KGE of 0.36 in dry years. The model calibrated in normal years failed to

obtain good evaluations in modeling TP with the lowest KGE values over the decade and in all three different hydrological

situations. The model calibrated in wet years showed relatively worse in modeling Chl *a* with all KGEs less than 0.2 and the

250   lowest KGE of -0.02 occurring in dry years. The results indicated that the YRWQM recalibrated with the decade

outperformed the other three models calibrated in a solely hydrological year, with the best robustness in modeling TP, Chl *a,*

and DO concentrations for a wide hydrological variability during the decade.

3.3 Parameters and kinetic processes comparison between recalibrated YRWQM and models calibrated within different hydrological years

255   Similar to the model performances, models employing different calibration strategies also had different parameter results

(Fig. 5). Most parameter values of the recalibrated YRWQM were within the parameter ranges of the other three calibrated

models, except for PRR, which had the lowest value of 0.12 among the four models. PRR represents the rate of predation on

algae by zooplankton or other aquatic organisms and algal predation is one of the main causes of algal reduction.

     Compared to models using other calibration strategies, the model calibrated in dry years had the highest PM of 5.1, the

260   highest PRR of 0.28, and the lowest REAC of 1.1. PM and PRR govern the growth and predation of algae respectively.

REAC represents the reaeration multiplier for the turbulence-induced and wind-induced surface reaeration coefficient, a

lower REAC value means less reaeration at the air-water interface. There were significantly lower BMR and higher K_{DP} in

the model calibrated in normal years, two parameters that represent the algal basal metabolism and the mineralization of

dissolved organic phosphorus into inorganic phosphorus, respectively. Most parameters of the model calibrated in wet years

265   were similar to the model recalibrated in the decade, except for one significantly lower value of CChl, which governs the

conversion between modeled and measured algal biomass.

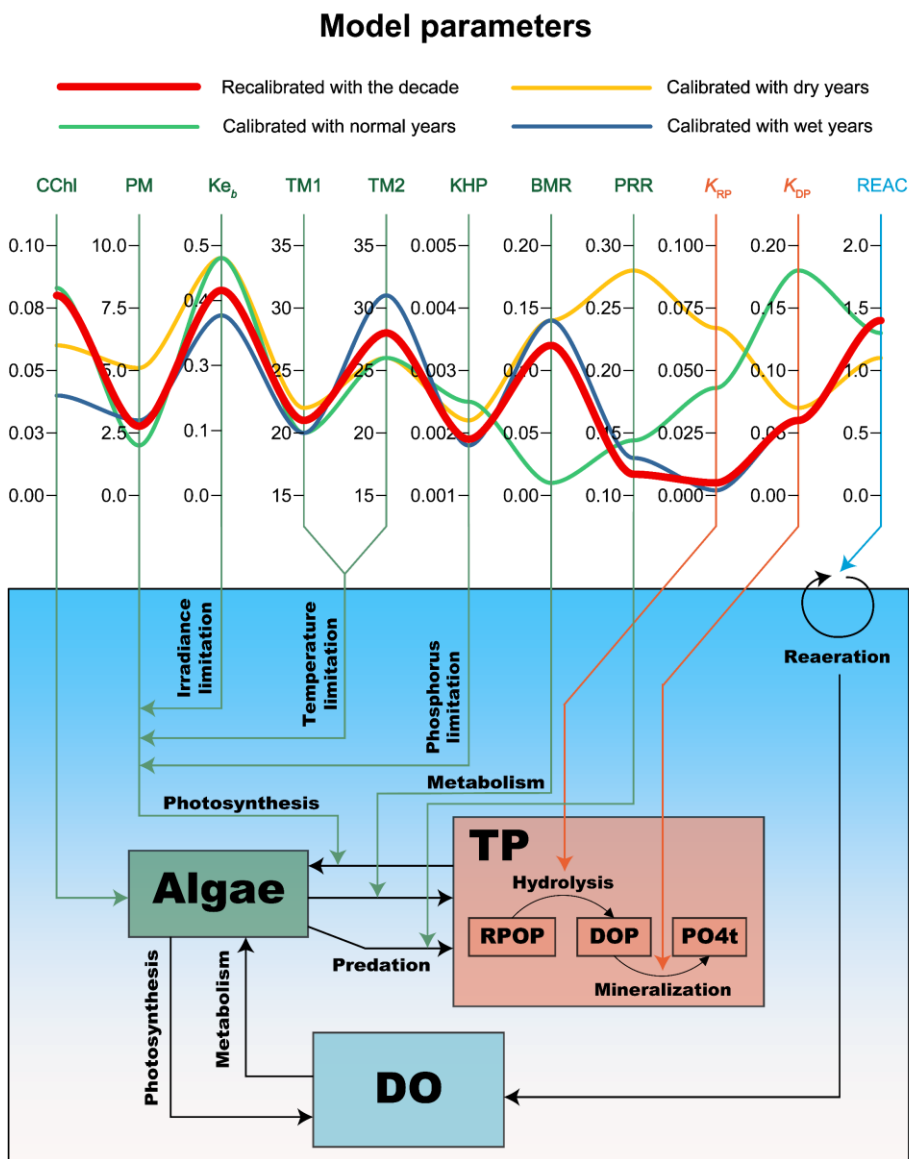



Figure 5: Parameters of the models calibrated with different hydrological years.

4 Discussion

4.1 Recalibrated YRWQM vs. Original YRWQM

Before embarking on the discussion of the discrepancies between the recalibrated YRWQM and original YRWQM, it is important to note that the model evaluations chose a single station (station S2, the center of the Yuqiao Reservoir) and three

state variables (TP, Chl *a*, DO), constrained by the complexity and computational cost of decadal-scale modeling. Nevertheless, the above indicators were considered capable of representing the main biogeochemical processes in Yuqiao

275 Reservoir, as previous statistical analyses and numerical models have indicated that Yuqiao Reservoir is a phosphorus-limited mesotrophic reservoir (Chen et al., 2012; Zhang et al., 2013; Xu et al., 2015). The model evaluations demonstrated that the recalibrated YRWQM performed equally well as the original YRWQM in terms of hydrodynamics, while the recalibrated YRWQM outperformed when it came to water quality. We supposed that the better performance probably stemmed from the recalibrated parameter values, especially sensitive parameters (Table 1). As described by Cerco and Cole

280 (1994) in the three-dimensional eutrophication model of Chesapeake Bay, the growth rate of algae was expressed as a multiplication of the maximum growth rate (PM) with a series of limiting factors in YRWQM, while the algal reduction was caused mainly by basal metabolism (BMR) and predation (PRR). These parameter values in the original YRWQM gave the algae too lenient growth conditions and motivated inaccurate algal outbreaks during the decade with a PBIAS of -357% (Table 1). These algal outbreaks may also be a potential reason for the overestimations of TP and DO concentrations with

285 PBIASs of -78% and -53%, respectively (Table 1). With the excessive algal outbreaks in the original YRWQM, the continuous enrichment of phosphorus in algae and the oxygen production process of excessive net photosynthesis prompted the final overestimations (Ji, 2017). As algal kinetics were accurately parameterized during the decade with a PBIAS of 36% in recalibrated YRWQM, satisfactory results were also obtained for TP and DO concentrations with PBIASs of 36% and -2% respectively (Table 1).

290 It should be noted that the modeled TP concentration peaks were not recorded in the observations in late 2010 and 2011 (Fig. 3). This may have been caused by the year-end water transfer, with inflow TP concentrations reaching 460 µg/L and 960 µg/L in December 2010 and 2011 respectively. It may also demonstrate that recalibrated YRWQM can provide a higher temporal resolution than observations and be potential as a hindcast model for reservoir management. Overall, the accuracy and robustness of the YRWQM have taken a solid step forward over a meticulous, long-term recalibration with EFDC-ACT.

295 **4.2 Why does the recalibrated YRWQM have better-performing parameters? Impact of the hydrological variability on calibration results**

Whereas it has been discussed above how updating parameter values improved the model accuracy and robustness of YRWQM, it is now more intriguing to see how this parameter updating was achieved by recalibration. We suppose that the observations with a wide hydrological variability may have contributed to the better-performing parameters, as the original

300 YRWQM was calibrated and validated in the only dry situation while the recalibrated YRWQM used decadal observations with a wide range of hydrological variability. Hydrological variability is one of the main causes of varying biogeochemical processes (Delpla et al., 2009; Li et al., 2020), and the changes in parameter values reflect the variability of these processes (Robson et al., 2018). James (2016) recalibrated the Lake Okeechobee Water Quality Model using 30 years of observations including a series of extreme hydro-meteorological events, thereby improving the quality of the parameters and the ability to

305 model nitrogen and phytoplankton. Many studies have also shown that the improvement in model parameters may be

triggered by calibration using long-term observations with greater hydrological variability (Cerco et al., 2004; Lung and Nice, 2007).

Among the four models, the parameters of the recalibrated YRWQM showed a proper trade-off with values almost falling within the range determined by the other models calibrated in specific hydrological years (Fig. 5). The model
310 calibrated in dry years performed as well as the recalibrated YRWQM for Chl *a* but failed to reproduce DO with a PBIAS of -21% (Fig. 4). This may be due to the highest algal growth rate (PM) causing excessive net photosynthesis (Fig.5). The drastic water level fluctuations of the Yuqiao Reservoir in dry years (Fig. 3) probably caused the decline of submerged macrophytes and the increase of phytoplankton, like other shallow water bodies (Furey et al., 2004; Krolová et al., 2013; Lu et al., 2018). However, it is necessary to analyze more observations of submerged macrophytes and couple the recalibrated
315 YRWQM with M-SAVM to gain a definite conclusion (Zhang et al., 2015). In the case of the model calibrated in wet years, the model performed poorly in modeling Chl *a* (Fig. 4) and the carbon-to-chlorophyll ratio (CChl) was the lowest (Fig. 5). Unlike a fixed value in the model, the value of CChl is more variable and depends on the makeup of the algae population, typically ranging from 0.015 to 0.1 (Bowie et al., 1985). Ren et al. (2019) also noted the differences in microbial composition between the dry and wet periods in Poyang Lake. A multi-species phytoplankton module that enables variable
320 CChl may contribute to more robust algal modeling. The above discussion pointed out the risks inherent in employing a model calibrated with a single hydrological year for climate change studies or management decisions. From the point of model accuracy and robustness, the use of long-term observations with sufficient hydrological variability to calibrate hydrodynamic and water quality models is probably the best option.

4.3 Highly efficient calibration with EFDC-ACT

325 The newly developed autocalibration toolkit, EFDC-ACT, eliminated a lot of hindrances to the recalibration of YRWQM. Compared to the conventional manual calibration method, it not only reduces a great deal of uncertainty from the subjective choice of parameters but also accelerates the convergence of the optimization process. As a generic autocalibration toolkit developed for models based on the EFDC framework, the EFDC-ACT supports the autocalibration of any combination of more than 200 parameters in the EFDC model. Meanwhile, the EFDC-ACT also incorporates automatic model evaluation
330 and advanced visualization of simulations and observations. Some process patterns can only be seen by time series plots and 2D plots. Statistics alone cannot reveal this kind of pattern (Bennett et al., 2013; Hipsey et al., 2020). The automated time series plots make the model results more visual and transparent at this point. The generated output files after each optimization iteration are overwritten and only the model parameters and evaluation results of each iteration are retained. This design ensures reproducibility while avoiding a large usage of hard disk space (Luo et al., 2018). The entire automatic
335 calibration framework proposed with EFDC-ACT can also be a reference to develop other automatic calibration tools for hydrodynamic and water quality models.

The *caRamel* algorithm adopted in EFDC-ACT has been demonstrated through case studies to obtain similar optimization results while speeding up convergence (Monteil et al., 2020). However, hundreds of parameters and the high

spatial and temporal complexity of EFDC bring about a time-consuming computation, making it difficult to reach the

340 recommended number of iterations of the *caRamel* algorithm. Furthermore, even with the support of optimization algorithms, how to obtain better calibration results faster is still a critical issue for the autocalibration of high-complexity models like EFDC.

4.4 Challenging high-complexity model autocalibration problems: a possible hierarchical autocalibration strategy introducing expert knowledge

345 To enable faster convergence of the model parameter optimization process, we propose a hierarchical autocalibration strategy based on EFDC-ACT. This strategy requires the modelers to orderly, automatically calibrate the model three times for different purposes. First, modelers formulate a large range of parameters based on literature or parameter implications, then run EFDC-ACT and perform a sensitivity analysis to find both the sensitive parameters and state variables. Although EFDC-ACT does not provide the functions for sensitivity analysis, there are a few R packages for sensitivity analysis, such

350 as the *sensitivity* package. A Bayesian framework integrating sensitivity, uncertainty, and identifiability analysis was also proposed for EFDC (Jia et al., 2018). The modelers will analyze the interactions between these sensitive variables according to expert knowledge, and variables with controlling effects will be the primary target for the second level of autocalibration. Next, the modelers target the sensitive variables and parameters identified in the first time and perform the autocalibration again until the model performs satisfactorily. Finally, the modelers hold the identified variables and parameters constant,

355 then auto-calibrate the model for the third time to determine the other insensitive state variables and parameters. With this hierarchical autocalibration strategy, EFDC-ACT can handle the parameter estimation of EFDC more competently. This strategy is a possible framework in the future, which is suitable not only for EFDC-ACT but also for other automatic calibration tools that do not produce sufficient iterations.

Even in the context of rapid advances in computer technology, expert knowledge is still indispensable to calibrating

360 highly complex models (Wood et al., 1990; Ostfeld and Salomons, 2005). With the emergence of automatic calibration tools, how to combine expert knowledge with them has become a new issue (Krueger et al., 2012; Xia and Shoemaker, 2022). The selection of key state variables in the hierarchical autocalibration strategy above is an example of the application of expert knowledge in an autocalibration tool. With the evolution of computer technology, the development of autocalibration tools, and the accumulation of observations, the hierarchical autocalibration strategy proposed above offers a possible workaround

365 to deal with enormous autocalibration problems in high-complexity models.

5 Conclusion

We developed a new automatic calibration toolkit, EFDC-ACT, and implemented it into the recalibration of the YRWQM with ten years (2006-2015) of observations in a wide range of hydrological variability. In comparison with the original YRWQM, the hydrodynamics of the recalibrated YRWQM performed as well over the decade, while the recalibrated model

370 performed significantly better in modeling TP, Chl *a*, and DO concentrations. We also calibrated the model in different hydrological years as calibration periods and compared their performance and parameters to the recalibrated YRWQM. Our analysis indicates that the recalibrated YRWQM accuracy and robustness improvement is derived from the constraining effect of observations with a wider hydrological variability. Such information will help to unravel how hydrological variability in the calibration periods affects the process-based hydrodynamic and water quality models, including their

375 parameters, kinetic processes, performance, and long-term robustness. Moreover, a general autocalibration toolkit developed in this study, EFDC-ACT, is substantially less time-consuming and more efficient for modelers than the conventional manual calibration method. The framework of EFDC-ACT and a possible hierarchical autocalibration strategy can also be a reference for future complex hydrodynamic and water quality model calibration. Finally, with our convenient autocalibration toolkit, it will be possible to explore the impact of the hydrological variability on more complex process-based

380 hydrodynamic and water quality models.

Code and data availability. The source code of the automatic calibration toolkit, EFDC-ACT, is freely available from https://doi.org/10.5281/zenodo.7438143 (Zhang and Fu, 2022) under the Creative Commons Attribution 4.0 International licence. Meteorological datasets are available from http://data.cma.cn/ (China Meteorological Data Service Centre). The dataset of other observed data is freely available from https://doi.org/10.5281/zenodo.7632109 (Zhang and Fu, 2023).

385 **Supplement.** The supplement related to this article is available online at: …

Author contributions. CZ designed the work, led the study, acquired the financial support, provided study resources, and conducted the research process; CZ and TF designed the methodology, developed the software, and wrote the initial manuscript draft; TF validated the reproducibility of results and prepared visualization; CZ reviewed and edited the manuscript.

390 **Competing interests.** The contact author has declared that none of the authors has any competing interests.

Acknowledgements. This research was supported by the National Natural Science Foundation of China (No.52079089). We also sincerely thank Prof. Zhengang Ji at the George Washington University for his constructive comments on the earlier draft.

Financial support. This research was supported by the National Natural Science Foundation of China (No.52079089).

395 **References**

Adrian, R., O'Reilly, C. M., Zagarese, H., Baines, S. B., Hessen, D. O., Keller, W., Livingstone, D. M., Sommaruga, R., Straile, D., Van Donk, E., Weyhenmeyer, G. A., and Winder, M.: Lakes as sentinels of climate change, Limnol. Oceanogr., 54, 2283–2297, https://doi.org/10.4319/lo.2009.54.6_part_2.2283, 2009.

Alexandrov, G. A., Ames, D., Bellocchi, G., Bruen, M., Crout, N., Erechtchoukova, M., Hildebrandt, A., Hoffman, F.,
400 Jackisch, C., Khaiter, P., Mannina, G., Matsunaga, T., Purucker, S. T., Rivington, M., and Samaniego, L.: Technical

assessment and evaluation of environmental models and software: Letter to the Editor, Environ. Model. Softw., 26, 328–336, https://doi.org/10.1016/j.envsoft.2010.08.004, 2011.

Arhonditsis, G. and Brett, M.: Evaluation of the current state of mechanistic aquatic biogeochemical modeling, Mar. Ecol. Prog. Ser., 271, 13–26, https://doi.org/10.3354/meps271013, 2004.

405 Arhonditsis, G. B., Qian, S. S., Stow, C. A., Lamon, E. C., and Reckhow, K. H.: Eutrophication risk assessment using Bayesian calibration of process-based models: Application to a mesotrophic lake, Ecol. Model., 208, 215–229, https://doi.org/10.1016/j.ecolmodel.2007.05.020, 2007.

Arifin, R. R., James, S. C., de Alwis Pitts, D. A., Hamlet, A. F., Sharma, A., and Fernando, H. J. S.: Simulating the thermal behavior in Lake Ontario using EFDC, J. Gt. Lakes Res., 42, 511–523, https://doi.org/10.1016/j.jglr.2016.03.011, 2016.

410 Basijokaite, R. and Kelleher, C.: Time-Varying Sensitivity Analysis Reveals Relationships Between Watershed Climate and Variations in Annual Parameter Importance in Regions With Strong Interannual Variability, Water Resour. Res., 57, https://doi.org/10.1029/2020WR028544, 2021.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., and 415 Andreassian, V.: Characterising performance of environmental models, Environ. Model. Softw., 40, 1–20, https://doi.org/10.1016/j.envsoft.2012.09.011, 2013.

Bowie, G. L., Mills, W. B., Porcella, D. B., Campbell, C. L., Pagenkopf, J. R., Rupp, G. L., Johnson, K. M., Chan, P. W. H., and Gherini, S. A.: Rates, Constants, and Kinetics Formulations in Surface Water Quality Modeling, U.S. Environmental Protection Agency, Washington, D.C., EPA/600/3-85/040., 1985.

420 Cerco, C.F. and Cole, T.: Three-Dimensional Eutrophication Model of Chesapeake Bay. Volume 1: Main Report., US Army Corps of Engineers Waterways Experiment Station., 1994.

Cerco, C. F. and Noel, M. R.: Incremental Improvements in Chesapeake Bay Environmental Model Package, J. Environ. Eng., 131, 745–754, https://doi.org/10.1061/(ASCE)0733-9372(2005)131:5(745), 2005.

Cerco, C. F., Noel, M. R., and Linker, L.: Managing for Water Clarity in Chesapeake Bay, J. Environ. Eng., 130, 631–642, 425 https://doi.org/10.1061/(ASCE)0733-9372(2004)130:6(631), 2004.

Chen, Y. Y., Zhang, C., Gao, X. P., and Wang, L. Y.: Long-term variations of water quality in a reservoir in China, Water Sci. Technol., 65, 1454–1460, https://doi.org/10.2166/wst.2012.034, 2012.

Delpla, I., Jung, A.-V., Baures, E., Clement, M., and Thomas, O.: Impacts of climate change on surface water quality in relation to drinking water production, Environ. Int., 35, 1225–1233, https://doi.org/10.1016/j.envint.2009.07.001, 2009.

430 Efstratiadis, A. and Koutsoyiannis, D.: Fitting Hydrological Models on Multiple Responses Using the Multiobjective Evolutionary Annealing-Simplex Approach, in: Practical Hydroinformatics, vol. 68, edited by: Abrahart, R. J., See, L. M., and Solomatine, D. P., Springer Berlin Heidelberg, Berlin, Heidelberg, 259–273, https://doi.org/10.1007/978-3-540-79881-1_19, 2008.

Franks, P. J. S.: Planktonic ecosystem models: perplexing parameterizations and a failure to fail, J. Plankton Res., 31, 1299–

435   1306, https://doi.org/10.1093/plankt/fbp069, 2009.

Fu, B., Merritt, W. S., Croke, B. F. W., Weber, T. R., and Jakeman, A. J.: A review of catchment-scale water quality and erosion models and a synthesis of future prospects, Environ. Model. Softw., 114, 75–97, https://doi.org/10.1016/j.envsoft.2018.12.008, 2019.

Furey, P. C., Nordin, R. N., and Mazumder, A.: Water Level Drawdown Affects Physical and Biogeochemical Properties of

440   Littoral Sediments of a Reservoir and a Natural Lake, Lake Reserv. Manag., 20, 280–295, https://doi.org/10.1080/07438140409354158, 2004.

Grant, L., Vanderkelen, I., Gudmundsson, L., Tan, Z., Perroud, M., Stepanenko, V. M., Debolskiy, A. V., Droppers, B., Janssen, A. B. G., Woolway, R. I., Choulga, M., Balsamo, G., Kirillin, G., Schewe, J., Zhao, F., del Valle, I. V., Golub, M., Pierson, D., Marcé, R., Seneviratne, S. I., and Thiery, W.: Attribution of global lake systems change to anthropogenic

445   forcing, Nat. Geosci., 14, 849–854, https://doi.org/10.1038/s41561-021-00833-x, 2021.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration, J. Hydrol. Eng., 4, 135–143, https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135), 1999.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80–91,

450   https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Hamrick, J. M.: A Three-Dimensional Environmental Fluid Dynamics Computer Code : Theoretical and computational aspects, Virginia Institute of Marine Science, College of William and Mary, https://doi.org/10.21220/V5TT6C, 1992.

Hipsey, M. R., Gal, G., Arhonditsis, G. B., Carey, C. C., Elliott, J. A., Frassl, M. A., Janse, J. H., de Mora, L., and Robson, B. J.: A system of metrics for the assessment and improvement of aquatic ecosystem models, Environ. Model. Softw., 128,

455   104697, https://doi.org/10.1016/j.envsoft.2020.104697, 2020.

Huang, Y.: Multi-objective calibration of a reservoir water quality model in aggregation and non-dominated sorting approaches, J. Hydrol., 510, 280–292, https://doi.org/10.1016/j.jhydrol.2013.12.036, 2014.

James, R. T.: Recalibration of the Lake Okeechobee Water Quality Model (LOWQM) to extreme hydro-meteorological events, Ecol. Model., 325, 71–83, https://doi.org/10.1016/j.ecolmodel.2016.01.007, 2016.

460   Janssen, P. H. M. and Heuberger, P. S. C.: Calibration of process-oriented models, Ecol. Model., 83, 55–66, https://doi.org/10.1016/0304-3800(95)00084-9, 1995.

Ji, Z.-G.: Hydrodynamics and water quality: modeling rivers, lakes, and estuaries, 2nd edition., John Wiley and Sons, Inc, Hoboken, NJ, https://doi.org/10.1002/9781119371946, 2017.

Ji, Z.-G., Morton, M. R., and Hamrick, J. M.: Wetting and Drying Simulation of Estuarine Processes, Estuar. Coast. Shelf

465   Sci., 53, 683–700, https://doi.org/10.1006/ecss.2001.0818, 2001.

Jia, H., Xu, T., Liang, S., Zhao, P., and Xu, C.: Bayesian framework of parameter sensitivity, uncertainty, and identifiability analysis in complex water quality models, Environ. Model. Softw., 104, 13–26, https://doi.org/10.1016/j.envsoft.2018.03.001, 2018.

Jiang, L., Li, Y., Zhao, X., Tillotson, M. R., Wang, W., Zhang, S., Sarpong, L., Asmaa, Q., and Pan, B.: Parameter uncertainty and sensitivity analysis of water quality model in Lake Taihu, China, Ecol. Model., 375, 1–12, https://doi.org/10.1016/j.ecolmodel.2018.02.014, 2018.

Jørgensen, S. E. and Fath, B. D.: Fundamentals of ecological modelling: applications in environmental management and research, 4th ed., Elsevier, Amsterdam ; Boston, 399 pp., 2011.

Kim, J., Seo, D., Jang, M., and Kim, J.: Augmentation of limited input data using an artificial neural network method to improve the accuracy of water quality modeling in a large lake, J. Hydrol., 602, 126817, https://doi.org/10.1016/j.jhydrol.2021.126817, 2021.

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrol. Earth Syst. Sci., 23, 4323–4331, https://doi.org/10.5194/hess-23-4323-2019, 2019.

Kong, X., Ghaffar, S., Determann, M., Friese, K., Jomaa, S., Mi, C., Shatwell, T., Rinke, K., and Rode, M.: Reservoir water quality deterioration due to deforestation emphasizes the indirect effects of global change, Water Res., 221, 118721, https://doi.org/10.1016/j.watres.2022.118721, 2022.

Krolová, M., Čížková, H., Hejzlar, J., and Poláková, S.: Response of littoral macrophytes to water level fluctuations in a storage reservoir, Knowl. Manag. Aquat. Ecosyst., https://doi.org/10.1051/kmae/2013042, 2013.

Krueger, T., Page, T., Hubacek, K., Smith, L., and Hiscock, K.: The role of expert opinion in environmental modelling, Environ. Model. Softw., 36, 4–18, https://doi.org/10.1016/j.envsoft.2012.01.011, 2012.

Li, X., Li, Y., and Li, G.: A scientometric review of the research on the impacts of climate change on water quality during 1998–2018, Environ. Sci. Pollut. Res., 27, 14322–14341, https://doi.org/10.1007/s11356-020-08176-7, 2020.

Lu, J., Bunn, S. E., and Burford, M. A.: Nutrient release and uptake by littoral macrophytes during water level fluctuations, Sci. Total Environ., 622–623, 29–40, https://doi.org/10.1016/j.scitotenv.2017.11.199, 2018.

Lung, W.-S. and Nice, A. J.: Eutrophication Model for the Patuxent Estuary: Advances in Predictive Capabilities, J. Environ. Eng., 133, 917–930, https://doi.org/10.1061/(ASCE)0733-9372(2007)133:9(917), 2007.

Luo, L., Hamilton, D., Lan, J., McBride, C., and Trolle, D.: Autocalibration of a one-dimensional hydrodynamic-ecological model (DYRESM 4.0-CAEDYM 3.1) using a Monte Carlo approach: simulations of hypoxic events in a polymictic lake, Geosci. Model Dev., 11, 903–913, https://doi.org/10.5194/gmd-11-903-2018, 2018.

McDonald, C. P. and Urban, N. R.: Using a model selection criterion to identify appropriate complexity in aquatic biogeochemical models, Ecol. Model., 221, 428-432, https://doi.org/10.1016/j.ecolmodel.2009.10.021, 2010.

Monteil, C., Zaoui, F., Le Moine, N., and Hendrickx, F.: Multi-objective calibration by combination of stochastic and gradient-like parameter generation rules – the caRamel algorithm, Hydrol. Earth Syst. Sci., 24, 3189–3209, https://doi.org/10.5194/hess-24-3189-2020, 2020.

Mooij, W. M., Trolle, D., Jeppesen, E., Arhonditsis, G., Belolipetsky, P. V., Chitamwebwa, D. B. R., Degermendzhy, A. G., DeAngelis, D. L., De Senerpont Domis, L. N., Downing, A. S., Elliott, J. A., Fragoso, C. R., Gaedke, U., Genova, S. N., Gulati, R. D., Håkanson, L., Hamilton, D. P., Hipsey, M. R., 't Hoen, J., Hülsmann, S., Los, F. H., Makler-Pick, V., Petzoldt, T., Prokopkin, I. G., Rinke, K., Schep, S. A., Tominaga, K., Van Dam, A. A., Van Nes, E. H., Wells, S. A., and Janse, J. H.: Challenges and opportunities for integrating lake ecosystem modelling approaches, Aquat. Ecol., 44, 633–667, https://doi.org/10.1007/s10452-010-9339-3, 2010.

Mooij, W. M., van Wijk, D., Beusen, A. H., Brederveld, R. J., Chang, M., Cobben, M. M., DeAngelis, D. L., Downing, A. S., Green, P., Gsell, A. S., Huttunen, I., Janse, J. H., Janssen, A. B., Hengeveld, G. M., Kong, X., Kramer, L., Kuiper, J. J., Langan, S. J., Nolet, B. A., Nuijten, R. J., Strokal, M., Troost, T. A., van Dam, A. A., and Teurlincx, S.: Modeling water quality in the Anthropocene: directions for the next-generation aquatic ecosystem models, Curr. Opin. Environ. Sustain., 36, 85–95, https://doi.org/10.1016/j.cosust.2018.10.012, 2019.

Moriasi, D. N., Arnold, J. G., Liew, M. W. V., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Transactions of the ASABE, 50, 885–900, https://doi.org/10.13031/2013.23153, 2007.

Nielsen, A., Trolle, D., Bjerring, R., Søndergaard, M., Olesen, J. E., Janse, J. H., Mooij, W. M., and Jeppesen, E.: Effects of climate and nutrient load on the water quality of shallow lakes assessed through ensemble runs by PCLake, Ecol. Appl., 24, 1926–1944, https://doi.org/10.1890/13-0790.1, 2014.

Ostfeld, A. and Salomons, S.: A hybrid genetic—instance based learning algorithm for CE-QUAL-W2 calibration, J. Hydrol., 310, 122–142, https://doi.org/10.1016/j.jhydrol.2004.12.004, 2005.

Paudel, R.: Does increased model complexity improve description of phosphorus dynamics in a large treatment wetland?, Ecol. Eng., 42, 283–294, https://doi.org/10.1016/j.ecoleng.2012.02.014, 2012.

Reed, P. and Devireddy, V.: Groundwater monitoring design: a case study combining epsilon dominance archiving and automatic parameterization for the NSGA-II, in: Advances in Natural Computation, vol. 1, World Scientific, 79–100, https://doi.org/10.1142/9789812567796_0004, 2004.

Ren, Z., Qu, X., Zhang, M., Yu, Y., and Peng, W.: Distinct Bacterial Communities in Wet and Dry Seasons During a Seasonal Water Level Fluctuation in the Largest Freshwater Lake (Poyang Lake) in China, Front. Microbiol., 10, 1167, https://doi.org/10.3389/fmicb.2019.01167, 2019.

Rigosi, A., Marcé, R., Escot, C., and Rueda, F. J.: A calibration strategy for dynamic succession models including several phytoplankton groups, Environ. Model. Softw., 26, 697–710, https://doi.org/10.1016/j.envsoft.2011.01.007, 2011.

Robson, B. J.: State of the art in modelling of phosphorus in aquatic systems: Review, criticisms and commentary, Environ. Model. Softw., 61, 339–359, https://doi.org/10.1016/j.envsoft.2014.01.012, 2014.

Robson, B. J., Arhonditsis, G. B., Baird, M. E., Brebion, J., Edwards, K. F., Geoffroy, L., Hébert, M.-P., van Dongen-Vogels, V., Jones, E. M., Kruk, C., Mongin, M., Shimoda, Y., Skerratt, J. H., Trevathan-Tackett, S. M., Wild-Allen, K., Kong, X., and Steven, A.: Towards evidence-based parameter values and priors for aquatic ecosystem modelling, Environ. Model. Softw., 100, 74–81, https://doi.org/10.1016/j.envsoft.2017.11.018, 2018.

Salk, K. R., Venkiteswaran, J. J., Couture, R., Higgins, S. N., Paterson, M. J., and Schiff, S. L.: Warming combined with experimental eutrophication intensifies lake phytoplankton blooms, Limnol. Oceanogr., 67, 147–158, https://doi.org/10.1002/lno.11982, 2022.

Schnedler-Meyer, N. A., Andersen, T. K., Hu, F. R. S., Bolding, K., Nielsen, A., and Trolle, D.: Water Ecosystems Tool (WET) 1.0 – a new generation of flexible aquatic ecosystem model, Geosci. Model Dev., 15, 3861–3878, https://doi.org/10.5194/gmd-15-3861-2022, 2022.

Shimoda, Y. and Arhonditsis, G. B.: Phytoplankton functional type modelling: Running before we can walk? A critical evaluation of the current state of knowledge, Ecol. Model., 320, 29–43, https://doi.org/10.1016/j.ecolmodel.2015.08.029, 2016.

Sinha, E., Michalak, A. M., and Balaji, V.: Eutrophication will increase during the 21st century as a result of precipitation changes, Science, 357, 405–408, https://doi.org/10.1126/science.aan2409, 2017.

Soares, L. M. V. and Calijuri, M. do C.: Deterministic modelling of freshwater lakes and reservoirs: Current trends and recent progress, Environ. Model. Softw., 144, 105143, https://doi.org/10.1016/j.envsoft.2021.105143, 2021.

Špoljar, M., Zhang, C., Dražina, T., Zhao, G., Lajtner, J., and Radonić, G.: Development of submerged macrophyte and epiphyton in a flow-through system: Assessment and modelling predictions in interconnected reservoirs, Ecol. Indic., 75, 145–154, https://doi.org/10.1016/j.ecolind.2016.12.038, 2017.

Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J.-M., Viney, N. R., and Teng, J.: Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies, J. Hydrol., 394, 447–457, https://doi.org/10.1016/j.jhydrol.2010.09.018, 2010.

Williamson, C. E., Saros, J. E., and Schindler, D. W.: Sentinels of Change, Science, 323, 887–888, https://doi.org/10.1126/science.1169443, 2009.

Wood, D. M., Houck, M. H., and Bell, J. M.: Automated Calibration and Use of Stream-Quality Simulation Model, J. Environ. Eng., 116, 236–249, https://doi.org/10.1061/(ASCE)0733-9372(1990)116:2(236), 1990.

Wu, G. and Xu, Z.: Prediction of algal blooming using EFDC model: Case study in the Daoxiang Lake, Ecol. Model., 222, 1245–1252, https://doi.org/10.1016/j.ecolmodel.2010.12.021, 2011.

Xia, W. and Shoemaker, C.: GOPS: efficient RBF surrogate global optimization algorithm with high dimensions and many parallel processors including application to multimodal water quality PDE model calibration, Optim. Eng., 22, 2741–2777, https://doi.org/10.1007/s11081-020-09556-1, 2021.

Xia, W. and Shoemaker, C. A.: A Repetitive Parameterization and Optimization Strategy for the Calibration of Complex and Computationally Expensive Process-Based Models With Application to a 3D Water Quality Model of a Tropical Reservoir, Water Resour. Res., 58, https://doi.org/10.1029/2021WR031054, 2022.

Xia, W., Akhtar, T., and Shoemaker, C. A.: A novel objective function DYNO for automatic multivariable calibration of 3D lake models, Hydrol. Earth Syst. Sci., 26, 3651–3671, https://doi.org/10.5194/hess-26-3651-2022, 2022.

Xu, Y., Xie, R., Wang, Y., and Sha, J.: Spatio-temporal variations of water quality in Yuqiao Reservoir Basin, North China, Front. Environ. Sci. Eng., 9, 649–664, https://doi.org/10.1007/s11783-014-0702-9, 2015.

Yi, X., Zou, R., and Guo, H.: Global sensitivity analysis of a three-dimensional nutrients-algae dynamic model for a large shallow lake, Ecol. Model., 327, 74–84, https://doi.org/10.1016/j.ecolmodel.2016.01.005, 2016.

Yu, R. and Zhang, C.: Early warning of water quality degradation: A copula-based Bayesian network model for highly efficient water quality risk assessment, J. Environ. Manage., 292, 112749, https://doi.org/10.1016/j.jenvman.2021.112749, 2021.

Zhang, C., Gao, X., Wang, L., and Chen, Y.: Analysis of agricultural pollution by flood flow impact on water quality in a reservoir using a three-dimensional water quality model, J. Hydroinformatics, 15, 1061–1072, https://doi.org/10.2166/hydro.2012.131, 2013.

Zhang, C., Gao, X., Wang, L., and Chen, X.: Modelling the role of epiphyton and water level for submerged macrophyte development with a modified submerged aquatic vegetation model in a shallow reservoir in China, Ecol. Eng., 81, 123–132, https://doi.org/10.1016/j.ecoleng.2015.04.048, 2015.

Zhang, C., Liu, H., Gao, X., and Zhang, H.: Modeling nutrients, oxygen and critical phosphorus loading in a shallow reservoir in China with a coupled water quality – Macrophytes model, Ecol. Indic., 66, 212–219, https://doi.org/10.1016/j.ecolind.2016.01.053, 2016.

Zhang, C., Huang, Y., Špoljar, M., Zhang, W., and Kuczyńska-Kippen, N.: Epiphyton dependency of macrophyte biomass in shallow reservoirs and implications for water transparency, Aquat. Bot., 150, 46–52, https://doi.org/10.1016/j.aquabot.2018.07.001, 2018a.

Zhang, C., Brett, M. T., Brattebo, S. K., and Welch, E. B.: How Well Does the Mechanistic Water Quality Model CE-QUAL-W2 Represent Biogeochemical Responses to Climatic and Hydrologic Forcing?, Water Resour. Res., 54, 6609–6624, https://doi.org/10.1029/2018WR022580, 2018b.

Zhang, C., Huang, Y., Javed, A., and Arhonditsis, G. B.: An ensemble modeling framework to study the effects of climate change on the trophic state of shallow reservoirs, Sci. Total Environ., 697, 134078, https://doi.org/10.1016/j.scitotenv.2019.134078, 2019.

Zhang, C., Yan, Q., Kuczyńska-Kippen, N., and Gao, X.: An Ensemble Kalman Filter approach to assess the effects of hydrological variability, water diversion, and meteorological forcing on the total phosphorus concentration in a shallow reservoir, Sci. Total Environ., 724, 138215, https://doi.org/10.1016/j.scitotenv.2020.138215, 2020.

Zhao, G., Gao, X., Zhang, C., and Sang, G.: The effects of turbulence on phytoplankton and implications for energy transfer with an integrated water quality-ecosystem model in a shallow lake, J. Environ. Manage., 256, 109954, https://doi.org/10.1016/j.jenvman.2019.109954, 2020.