

S1 EFDC-ACT User's Guide

S1.1 Preface

Environmental Fluid Dynamics Code–Automatic Calibration Toolkit (EFDC-ACT) was developed for automating the calibration of a 3D hydrodynamic and water quality model, EFDC, with more than two hundred parameters. EFDC-ACT is a multivariate and multiobjective automatic calibration toolkit based on **R 4.1.0**. In order to use EFDC-ACT, the user needs to install

1. **R 4.1.0**, available at <https://cloud.r-project.org/>.
2. **RStudio**, available at <https://www.rstudio.com/>.

To implement EFDC-ACT to your model, the user should go through three main steps: initialization, autocalibration, and post-analysis.

S1.2 Initialization: Preparation of all the files

In the initialization step, the user needs to prepare the following input files:

1. **EFDC input files**. The input files necessary to run the EFDC model. Details of which files are included can be found in the EFDC User Manual (Hamrick, 1996). These files should be placed in the **EFDC** folder.
2. **EFDC-ACT-par.yaml** This file provides EFDC-ACT with parameters for function selection, model information, control of the autocalibration, etc. These files should be placed in the root directory.
3. **Par_list.csv** This file contains information on the parameters that can be automatically calibrated by EFDC-ACT. Each parameter includes **fixed_value**, **minimum_value**, **maximum_value**, **is_auto_calibrate**, and **digits** five columns. When **is_auto_calibrate** is 0, the corresponding parameter value will use **fixed_value**, while when **is_auto_calibrate** is 1, the corresponding parameter value will be taken randomly from the range defined by **minimum_value** and **maximum_value**. All parameters are truncated according to the number of decimal places specified by **digits**. The user can also add parameters not listed in the file but present in the EFDC model directly, with only minor modifications in the code of the EFDC model. This file should be placed in the **data/raw** folder.
4. **Var_list.csv** This file contains information on the state variables corresponding to the autocalibration objective function. Each state variable containing seven columns **I**, **J**, **K**, **variable_code**, **statistic_code**, **statistic_minmax**, and **statistic_prec**. In these seven columns, **I**, **J**, and **K** represent the spatial location of the state variables, **variable_code** represents the code corresponding to the type of state variable, **statistic_code**, **statistic_minmax**, and **statistic_prec** represent the code corresponding to the type of statistic, whether the statistic is to be maximized or minimized, and the desired accuracy of the statistic. This file should be placed in the **data/raw** folder.

5. **Observation files.** In addition to the above information, the observations corresponding to the state variables should be placed in the **data/raw** folder as CSV files with the filename "**type-I-J-K.csv**" (e.g., TP-35-21-4.csv).

All of the above files have templates in their corresponding locations. Although EFDC-ACT already provides the available **EFDC.exe**, for those who want to compile EFDC source code themselves, we also give the FORTRAN subroutine files that need to be compiled together with the EFDC source code to create the interface for EFDC-ACT, including **CDALPHA_INPUT.f**, **ZBRADJ_INPUT.f**, and **WQ_INPUT.f** placed in the **src** folder. Users can also refer to these files to add more parameter interfaces themselves.

S1.3 Autocalibration: Automatic parameter calibration with the EFDC-ACT

After all the preparation work has been completed, users can open the **EFDC-ACT-main.R** with **RStudio** and source this R script. Then EFDC-ACT will automatically carry out the model calibration process or other functions selected by the users.

For coping with the multivariate, multi-objective optimization problem of EFDC model calibration, EFDC-ACT uses the *caRamel* package. It is a genetic algorithm-based multi-objective optimizer designed for complex model calibration. To cope with the calibration of time-consuming models, the *caRamel* package also applies parallel computing capabilities to accelerate the optimization process (Monteil et al., 2020). However, the parallel function of EFDC-ACT is still under development. Due to the highly complex and time-consuming characteristics of EFDC, the use of the *caRamel* package for automatic calibration can substantially reduce the difficulty for modelers in tuning parameters and comparing model results. In addition to the *caRamel* package, other R packages for optimization or model calibration can be incorporated into EFDC-ACT with a minor modification to the relevant functions.

In EFDC-ACT, the number of parameters and the number of state variables read from **Par_list.csv** and **Var_list.csv** are used as parameters **nobj** and **nvar** in the function **caRamel()**. The columns **minimum_value** and **maximum_value** in the file **Par_list.csv** are used as **bounds**, and the column **statistic_minmax** in the file **Var_list.csv** is used as **minmax**. Other necessary parameters such as **popsiz**, **archsiz**, **maxrun**, etc. are input by the user from the configuration file **EFDC-ACT-pars.yaml**.

S1.4 Post-analysis: Checking automatic calibration results

Once the entire autocalibration process of EFDC-ACT has been completed, the user can check the final optimization results in **caRamel_result.csv** placed in **reports** folder. For more flexibility in the analysis of model results, EFDC-ACT also offers the model results extraction and model results evaluation module as a post-analysis step. These functions can also be used independently for the results of EFDC model runs without EFDC-ACT. This module includes not only statistical evaluation techniques but also graphical evaluation techniques. The results of the statistics specified as the objective function will be stored in **Var_result.csv** in the **reports** folder at the end of each iteration, and the corresponding parameter values will be stored in **Par_result.csv** in the same folder. To give a visual comparison that statistics cannot provide, the model results and observed data for the specified state variables are used to plot the time series and are placed in the **figure** folder.

S2 Location and discharge data of the Yuqiao Reservoir

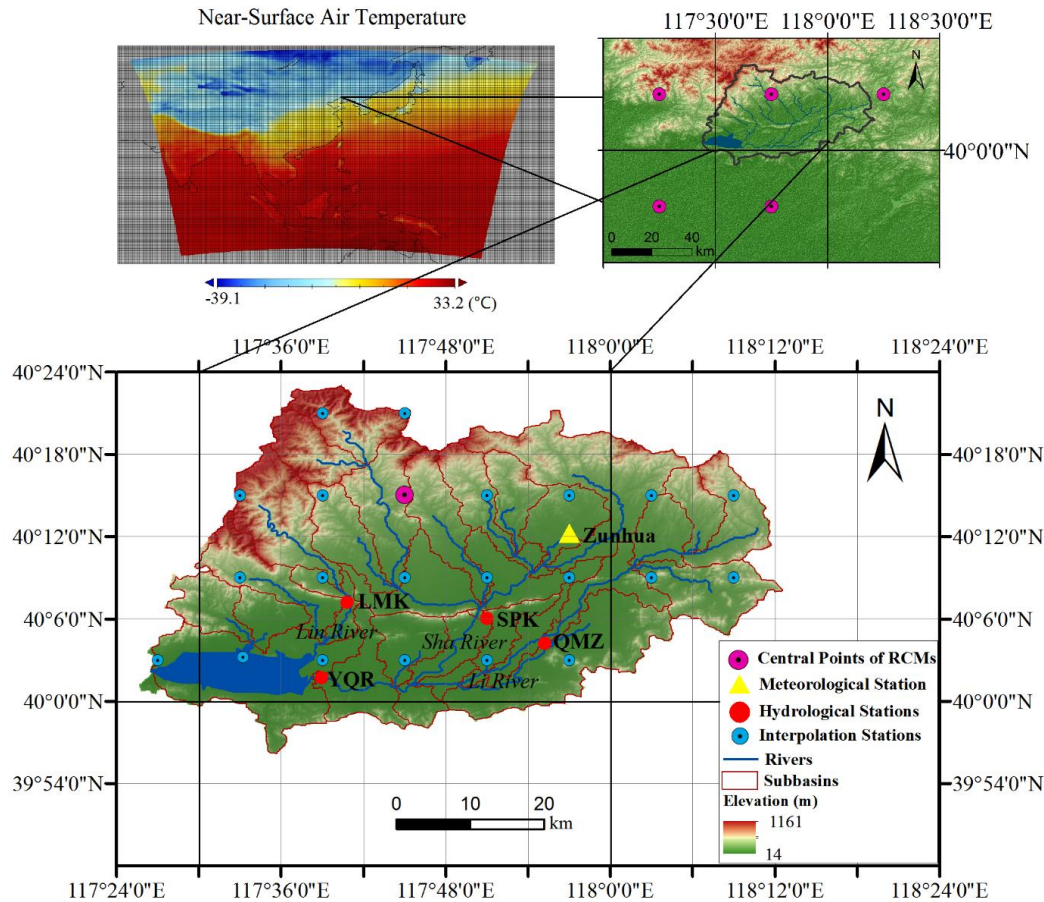


Figure S1: Location of the Yuqiao Reservoir watershed and the Yuqiao Reservoir. The spatial resolution of CORDEX-East Asia in the upper-left panel is $0.5^{\circ} \times 0.5^{\circ}$. Reproduced from Zhang et al. (2020).

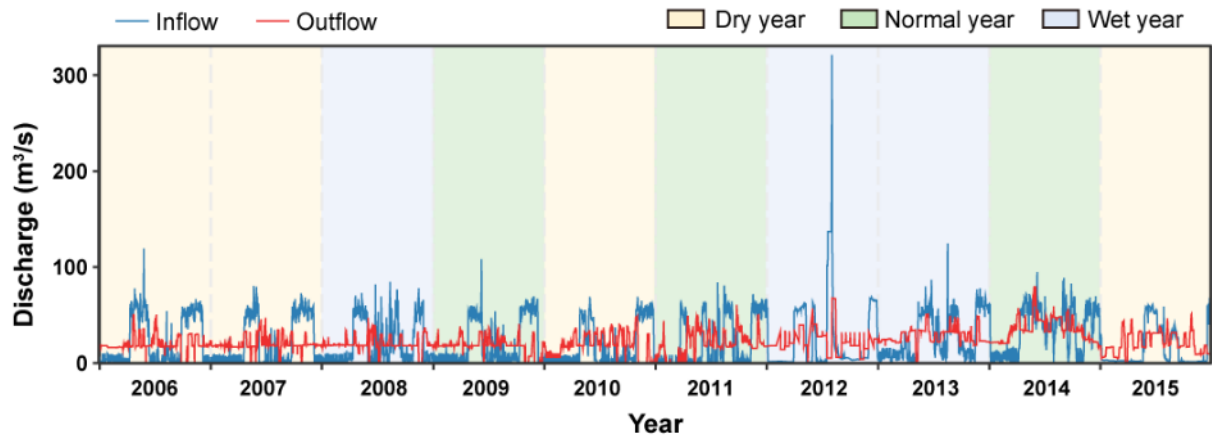


Figure S2: Inflow and outflow discharge data of the Yuqiao Reservoir from 2006 to 2015.

S3 Water quality model equations and parameter values

S3.1 Water quality model equations

- 70 The water quality state variables involved in this study include chlorophyll *a* (Chl *a*), total phosphorus (TP), and dissolved oxygen (DO). Their kinetic equations in Yuqiao Reservoir Water Quality Model (YRWQM) are described in full below. All equations and parameters in YRWQM are based on the EFDC framework. More detailed model principles can be found in Hamrick (1992) and Ji (2017).

S3.1.1 Model equations for Chl *a*

- 75 The measured Chl *a* concentrations and modeled algal biomass concentrations can be converted to each other using

$$\text{Chl } \alpha = \frac{B}{\text{CChl}} \quad (\text{S1})$$

where *B* is the algal biomass concentration as carbon (mg C/L) and CChl is the carbon-to-chlorophyll ratio for algae (mg C/μg Chl *a*).

The kinetic equation for algae is defined by

$$80 \quad \frac{\partial B}{\partial t} = (P - \text{BM} - \text{PR})B + \frac{\partial}{\partial z}(\text{WS} \cdot B) + \frac{\text{WB}}{V} \quad (\text{S2})$$

where *t* is time, *P* is the production rate of algae (day⁻¹), BM is the basal metabolism rate of algae (day⁻¹), PR is the predation rate of algae (day⁻¹), WS is the settling velocity of algae (m/day), WB is the external loads of algae (g C/day), and *V* is the volume (m³).

The production rate of algae is mathematically expressed as

$$85 \quad P = \text{PM} \cdot f_1(\text{PO}_4) \cdot f_2(I) \cdot f_3(T) \quad (\text{S3})$$

where PM is the maximum growth rate for algae (day⁻¹), *f*₁(*PO*₄) is the growth limiting function for orthophosphate concentration, *f*₂(*I*) is the growth limiting function for light intensity, and *f*₃(*T*) is the growth limiting function for temperature. These limiting function are defined by

$$f_1(\text{PO}_4) = \frac{\text{PO}_4}{\text{KHP} + \text{PO}_4} \quad (\text{S4})$$

$$90 \quad f_2(I) = \frac{2.718 \cdot \text{FD}}{\text{Kess} \cdot \Delta z} (e^{-\alpha_B} - e^{-\alpha_T}) \quad (\text{S5})$$

$$\text{Kess} = \text{Ke}_b + \text{Ke}_{\text{Chl}} \cdot \text{Chl } \alpha \quad (\text{S6})$$

$$\alpha_B = \frac{I_0}{\text{FD} \cdot I_s} \cdot \exp(-\text{Kess}[H_T + \Delta z]) \quad (\text{S7})$$

$$\alpha_T = \frac{I_0}{\text{FD} \cdot I_s} \cdot \exp(-\text{Kess} \cdot H_T) \quad (\text{S8})$$

$$I_s = \text{minimum}[(I_0)_{\text{avg}} \cdot e^{-\text{Kess} \cdot D_{\text{opt}}}, (I_s)_{\text{min}}] \quad (\text{S9})$$

$$95 \quad f_3(T) = \begin{cases} e^{-\text{KTG1}(T-\text{TM1})^2} & \text{if } T < \text{TM1} \\ 1.0 & \text{if } \text{TM1} \leq T \leq \text{TM2} \\ e^{-\text{KTG2}(T-\text{TM2})^2} & \text{if } T > \text{TM2} \end{cases} \quad (\text{S10})$$

where PO_4 is the orthophosphate concentration (mg/L), KHP is the half saturation constant for P (mg/L), FD is the fractional day length, K_{ess} is the total light extinction coefficient (m^{-1}), K_{eb} is the background light extinction coefficient (m^{-1}), K_{chl} is the light extinction coefficient for Chl *a* ($m^{-1}/(mg\ Chl/m^3)$), Δz is the layer thickness (m), I_0 is the daily total light intensity at water surface (langley/day), I_s is the optimal light intensity for algae (langley/day), H_T is the depth from the free surface to the top of the layer (m), $(I_0)_{avg}$ is the adjusted surface light intensity (langley/day), D_{opt} is the depth of maximum algal growth (m), $(I_s)_{min}$ is the specified minimum optimal light intensity for algae (langley/day), T is the temperature ($^{\circ}C$), TM1 is the lower optimal temperature for algal growth ($^{\circ}C$), TM2 is the upper optimal temperature for algal growth ($^{\circ}C$), KTG1 is the suboptimal temperature effect coefficient for algal growth ($^{\circ}C^{-1}$), and KTG2 is the superoptimal temperature effect coefficient for algal growth ($^{\circ}C^{-1}$).

The basal metabolism rate and predation rate of algae are expressed as

$$BM = BMR \cdot e^{KTB \cdot (T - TR)} \quad (S11)$$

$$PR = PRR \cdot e^{KTB \cdot (T - TR)} \quad (S12)$$

where BMR is the basal metabolism rate of algae at TR (day^{-1}), PRR is the predation rate of algae at TR (day^{-1}), KTB is the temperature effect coefficient ($^{\circ}C^{-1}$), and TR is the reference temperature ($^{\circ}C$).

S3.1.2 Model equations for TP

TP is the sum of phosphorus in algae, refractory particulate organic phosphorus, dissolved organic phosphorus, and total phosphate. It can be expressed as

$$TP = APC \cdot B + RPOP + DOP + PO_4t \quad (S13)$$

where APC is the phosphorus-to-carbon ratio in algae (mg P/mg C), RPOP is the concentration of refractory particulate organic phosphorus (mg/L), DOP is the concentration of dissolved organic phosphorus (mg/L), and PO_4t is the concentration of total phosphate (mg/L).

The kinetic equation for RPOP is described as

$$\frac{\partial RPOP}{\partial t} = (FPR \cdot BM + FPRP \cdot PR) \cdot APC \cdot B - K_{RPOP} \cdot RPOP + \frac{\partial}{\partial z} (WS_{RP} \cdot RPOP) + \frac{WRPOP}{V} \quad (S14)$$

where FPR is the fraction of metabolized P produced as RPOP, FPRP is the fraction of predated P produced as RPOP, K_{RPOP} is the hydrolysis rate of RPOP (day^{-1}), WS_{RP} is the settling velocity of refractory particulate organic matter (m/day), and WRPOP is the external loads of RPOP (g P/day).

The kinetic equation for DOP is described as

$$\frac{\partial DOP}{\partial t} = (FPD \cdot BM + FPD P \cdot PR) \cdot APC \cdot B + K_{RPOP} \cdot RPOP - K_{DOP} \cdot DOP + \frac{WDOP}{V} \quad (S15)$$

where FPD is the fraction of metabolized P produced as DOP, FPD P is the fraction of predated P produced as DOP, K_{DOP} is the mineralization rate of DOP (day^{-1}), and WDOP is the external loads of DOP (g P/day).

The kinetic equation for PO_4t is described as

$$\frac{\partial \text{PO}_4\text{t}}{\partial t} = (\text{FPI} \cdot \text{BM} + \text{FPIP} \cdot \text{PR} - P) \cdot \text{APC} \cdot B + K_{\text{DOP}} \cdot \text{DOP} + \frac{\text{WPO}_4\text{t}}{V} \quad (\text{S16})$$

where FPI is the fraction of metabolized P produced as inorganic P, FPIP is the fraction of predated P produced as inorganic P, and WPO_4t is the external loads of inorganic P (g P/day).

130 The formulations for hydrolysis and mineralization rates are described as

$$K_{\text{RPOP}} = \left(K_{\text{RP}} + \frac{K_{\text{HP}}}{K_{\text{HP}} + \text{PO}_4\text{d}} \cdot K_{\text{RPalg}} \cdot B \right) \cdot e^{KT_{\text{MNL}}(T - TR_{\text{MNL}})} \quad (\text{S17})$$

$$K_{\text{DOP}} = \left(K_{\text{DP}} + \frac{K_{\text{HP}}}{K_{\text{HP}} + \text{PO}_4\text{d}} \cdot K_{\text{DPalg}} \cdot B \right) \cdot e^{KT_{\text{MNL}}(T - TR_{\text{MNL}})} \quad (\text{S18})$$

where K_{RP} is the minimum hydrolysis rate of RPOP (day^{-1}), K_{DP} is the minimum mineralization rate of DOP (day^{-1}), PO_4d is the concentration of dissolved phosphate, K_{RPalg} is the constant that relates the hydrolysis of RPOP to algal biomass
 135 ($\text{day}^{-1}/(\text{g C}/\text{m}^3)$), K_{DPalg} is the constant that relates mineralization of DOP to algal biomass ($\text{day}^{-1}/(\text{g C}/\text{m}^3)$), KT_{MNL} is the effect of temperature on hydrolysis and mineralization ($^{\circ}\text{C}^{-1}$), and TR_{MNL} is the reference temperature for hydrolysis and mineralization ($^{\circ}\text{C}$).

S3.1.3 Model equations for DO

The net change in DO concentration should be determined by the sum of the DO source and DO sinks, that is

$$140 \text{ Net change of DO} = \text{photosynthesis} - \text{respiration} - \text{nitrification} + \text{reaeration} - \text{SOD} + \text{external loads} \quad (\text{S19})$$

where reaeration is a process by which oxygen is transferred across the interface between the atmosphere and a waterbody, usually resulting in the net transfer of oxygen to the water. The change of DO concentration due to reaeration is expressed by
 reaeration = $\text{REAC} \cdot K_r (\text{DO}_s - \text{DO})$ (S20)

where REAC is the reaeration multiplier, K_r is the reaeration coefficient (day^{-1}), and DO_s is the saturation concentration of
 145 DO ($\text{g O}_2/\text{m}^3$).

The parameters affecting the DO source and sinks were not adjusted except for REAC. Detailed equations for these source and sink terms can be found in Hamrick (1992) and Ji (2017).

S3.2 Water quality model parameter calibration results

150 Table S1: Parameter ranges and calibrated values during model calibration. The parameter ranges 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).

Parameters	Units	Ranges			Model		Model		Model		Equations
			Recalibrated	Original	calibrated		calibrated		calibrated		
					with	dry	with	normal	with	wet	
			YRWQM	YRWQM	years		years		years		
CChl	mg C / µg Chl	0.04~0.085	0.08	0.083	0.06		0.083		0.04		Eq. (S1)
WS	m/day	0.05~0.1	0.08	0.1	0.06		0.08		0.08		Eq. (S2)
PM	day ⁻¹	1~5.5	2.77	2	5.1		2		3		Eq. (S3)
KHP	mg/L	0.001~0.0025	0.0019	0.001	0.0022		0.0025		0.0018		Eq. (S4)
Ke _b	m ⁻¹	0.3~0.55	0.41	0.475	0.475		0.475		0.36		Eq. (S6)
Ke _{Chl}	m ⁻¹ /(mg Chl/m ³)	0.01~0.07	0.04	0.041	0.04		0.07		0.04		Eq. (S6)
D _{opt}	m	1~2	1.9	1	1		1		1.3		Eq. (S9)
TM1	°C	9~27	21	9	22		20		20		Eq. (S10)
TM2	°C	25~31	28	15	26		26		31		Eq. (S10)
KTG1	°C ⁻¹	0.0001~0.01	0.005	0.0001	0.005		0.008		0.006		Eq. (S10)
KTG2	°C ⁻¹	0.0001~0.01	0.011	0.0001	0.011		0.009		0.012		Eq. (S10)
BMR	day ⁻¹	0.01~0.14	0.12	0.01	0.14		0.01		0.14		Eq. (S11)
PRR	day ⁻¹	0.01~0.3	0.117	0.01	0.28		0.144		0.13		Eq. (S12)
WS _{RP}	m/day	0.005~1	0.2	0.1	0.142		0.081		0.12		Eq. (S14)
K _{RP}	day ⁻¹	0.001~0.01	0.005	0.001	0.067		0.043		0.002		Eq. (S17)
K _{DP}	day ⁻¹	0.005~0.3	0.060	0.005	0.07		0.18		0.06		Eq. (S18)
REAC	/	0.5~1.5	1.4	0.9	1.1		1.3		1.4		Eq. (S20)

S4 Water quality model performance calibrated with different hydrological years during the decade at station S2 of the Yuqiao Reservoir

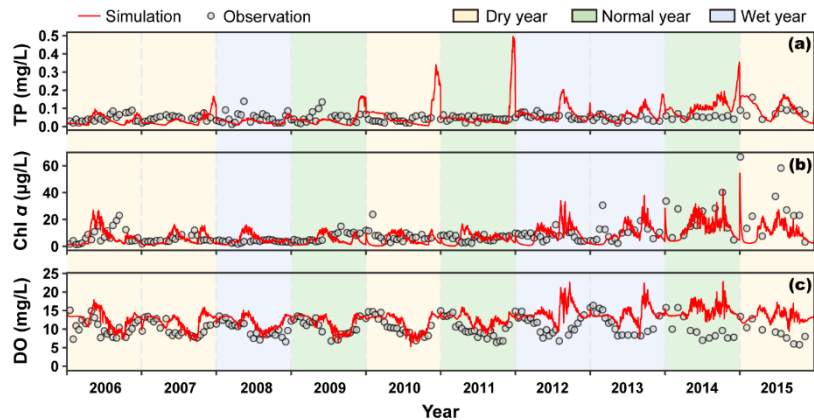


Figure S3: Performance of the water quality model calibrated with dry years at station S2 of the Yuqiao Reservoir ($n = 190$).

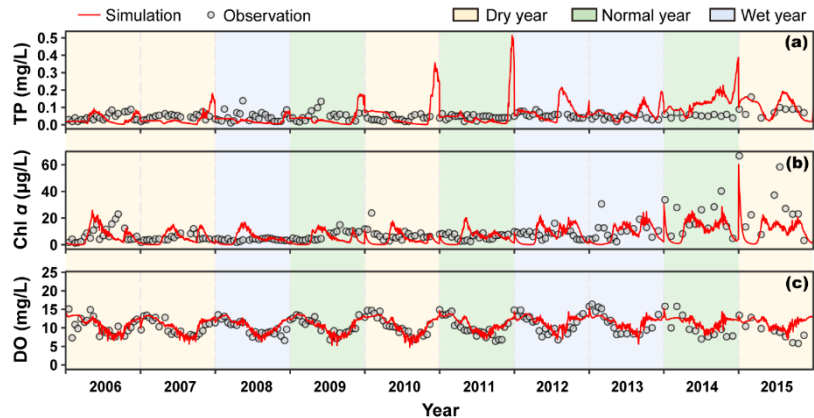


Figure S4: Performance of the water quality model calibrated with normal years at station S2 of the Yuqiao Reservoir ($n = 190$).

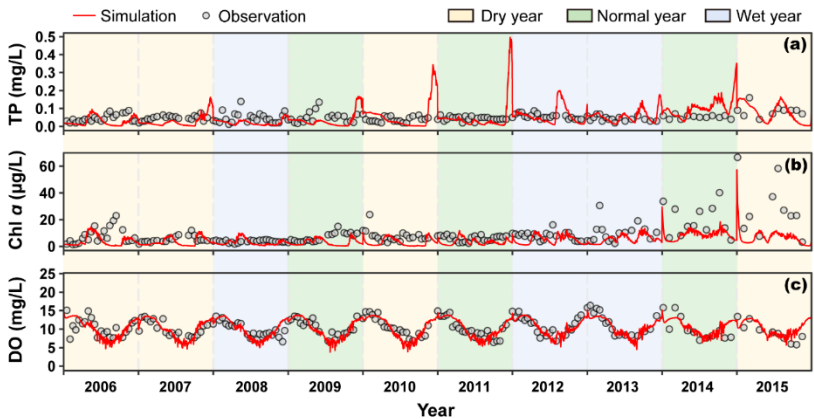


Figure S5: Performance of the water quality model calibrated with wet years at station S2 of the Yuqiao Reservoir ($n = 190$).

References:

- Hamrick, J. M.: A Three-Dimensional Environmental Fluid Dynamics Computer Code : Theoretical and computational aspects, 65, <https://doi.org/10.21220/V5TT6C>, 1992.
- Hamrick, J. M.: User's Manual for the Environmental Fluid Dynamics Computer Code, Special Reports in Applied Marine Science and Ocean Engineering (SRAMSOE) No. 331. Virginia Institute of Marine Science, College of William and Mary. <https://doi.org/10.21220/V5M74W224>, 1996.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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., 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.