RESPONSES FOR GMD-2022-303

Title: *Recalibration of a three-dimensional water quality model with a newly developed autocalibration toolkit (EFDC-ACT v1.0.0): how much improvement will be achieved with a wider hydrological variability?*

June 3, 2023

Dr. Jeffrey Neal
Topic Editor
Geoscientific Model Development

Dear Jeff,

We highly appreciate you for giving us the opportunity to submit a revised draft of our manuscript for publication in the *GMD*. The comments of the editor and referees for the insightful and constructive recommendations are good for improving the quality of our manuscript. In the following pages are our main corrections to the manuscript and responses to the referee comments as points to points.

We have read referee comments carefully and have made the revision to improve the scientific reproducibility and presentation quality of our manuscript. Firstly, we rephrased our title based on suggestions and the title has now been changed to *"Recalibration of a three-dimensional water quality model with a newly developed autocalibration toolkit (EFDC-ACT v1.0.0): how much improvement will be achieved with a wider hydrological variability?"*. We emphasized the importance of understanding the model system and the parameter implications and added the description of the EFDC and the unique challenges in the introduction part. We restructured the materials and methods part, complemented the calibration with original strategy as the benchmarking point, and simplified the description about study area and data sources. We also appended to our results the time savings and efficiency of EFDC-ACT, describing the calibration results in more details and explaining the validation strategy and primary parameters selection process. We have also included the model calibrated with original strategy in Figure 3 to clearly screen the improvement. In addition, we also answered other questions from the referees.

Please find attached a revised version of our manuscript originally entitled "*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?*", which we would like to submit for your kind consideration. We hope that you will find the revised submission much improved. If not, we are willing to revise the manuscript until the reviewers are satisfied with the revision. Thank you very much for considering our paper for publication to **Geoscientific Model Development**.

Sincerely Yours,

Chen Zhang, Professor
Tianjin University

Referee comments:

Referee#1 (Dr. Hu, Fenjuan):

General Comments:

This paper presents the newly developed auto-calibration tool, EFDC-ACT, a R-package for auto calibrating the 3D hydrodynamic and water quality model, EFDC. The content and topic of this paper fits GMD's scope. It presents a modelling tool that advances the 3D model hydrodynamic and water quality calibration process. The tool is original and novel, the data is reliable and transparent. The results of the model calibration with the new tool are shown to be satisfactory. The manuscript has clear structure, background, methods, and results were clearly presented, and discussion is constructive, with fluent and precise language.

However, the paper lacks little consistence, starting introduction and methods with heavy description of 3D Yuqiao Reservoir model, and its chronic development over the years, and equally heavy description of the data sources regarding hydrological years (dry, wet and normal years), which diverted the focus of the paper. While the result section and discussion section are focused on the EFDC-ACT performance, which is very good.

To fit GMD's Development and technical paper scope, this study should focus mainly on the statistical and auto-calibration technics applied in 3D-hydrodyanmic model calibration practices and the novelty of this tool regarding this practices. The YRWQM (Yuqiao Reservior Water Quality Model) is a practice on EFDC model (a 3D hydrodynamic and water quality model), and the scientific questions of different hydrological years, and why it's causing bias in calibrating WQ parameters such as chla, are good argument and background for a better results of re-calibration, and therefore support the evidence the newly developed tool is an advance, but should not be main focus.

Response: We are grateful for thoughtful suggestions for improving our paper. Based on these comments, we carefully revised the manuscript. We hope the revised manuscript will meet the standards of GMD. If not, we are willing to revise the MS until the reviewers are satisfied with the revision. Below you will find our point-by-point responses to the referee comments/questions.

Sectional comments:

Comment 1:

Introduction*: the introduction part contained the main points of motivations of this study: difficulties in calibrating complex 3D hydrodynamic models (starting from line 35), and short data calibration creating equifinality problems in models. It also covered available auto calibration practices (PEST-EFDC, and PEST-Delft3D). The introduction should continue the path of analysis the current auto calibration methods and what's lacking in current practices, fx. Not so fit for EFDC calibration. A closer look at EFDC's parameters and calibration challenges will bring the focus near the initiation of EFDC-ACT development. I miss the overview of the EDDC model structure and its parameter sets, and specific challenges of calibrating such 3D complex model.*

Answer 1:

Thanks for your suggestion. we have added the description of the EFDC model structure, parameter sets, and the unique challenges of calibrating such a complex three-dimensional model in the introduction part. These corrections will further help the readers to recognize the innovative and important nature of the work in this paper.

Please see the text in Section 1 as follows.

…The EFDC is a general-purpose model developed for simulating three-dimensional flow, transport, and biogeochemical processes in surface water systems, including rivers, lakes, estuaries, reservoirs, wetlands, and coastal regions (Hamrick, 1992). The hydrodynamic model consists of continuity, momentum, state, and transport equations for salinity and temperature. The water quality model consists of 22 state variables and associated kinetics (Ji, 2017). More than 200 parameters which govern the above process are spread over different cards in different input files, and the model results comprises lots of output files in different formats. Despite the emergence of numerous well established generic tools for automatic calibration, the process of linking these tools to EFDC input and output files is still cumbersome. For such a sprawling model system as the EFDC, a specific automatic calibration tool can eliminate much of the repetitive and unnecessary work. On the other hand, specific automatic calibration tools are needed to support multi-objective evaluation methods for three-dimensional model calibration, including evaluation at different locations on the horizontal plane, evaluation at different depths on the same grid, or a mixture of these objectives. **(Pages 2-3, lines 58-69)**

Comment 2:

Materials and Methods*: the most important part of this sections is 2.4, where the material of this study is the calibrated YRWQM from previous study, no need to present the study sites and data sources in detail. Methods also lacks part of benchmarking process, i.e. the former calibration of 2006,2007, which corresponding to the dry years in the re-calibrated period, should be a benchmarking point to compare the model calibrating performance. And the comparison of these two (the former YRWQM and the recalibrated YRWQM) model performances should not only include the calibration period, but also the validation period (the period where model parameters are not adjusted to data, to validate model's ability of presenting the reality).*

Answer 2:

Thanks for your helpful suggestions. We have restructured the Materials and Methods part and the revised Section 2 is structured as follows.

2.1 EFDC-Automatic Calibration Toolkit

The introduction to the automatic calibration toolkit has been moved to the first section to highlight its importance.

2.2 Chronicle of YRWQM

The original section 2.1, which describes the study area, have been simplified and merged with the original section 2.2. This will reduce unnecessary repetition of narrative. The description about the original YRWQM was simplified, and the details are demonstrated in the section 2.3.

Please see the text in Section 2.2 as follows.

The Yuqiao Reservoir (40°00'-40°04'N, 117°26'-117°37'E) is situated in Jixian County, Tianjin, China (Fig. 2). 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 Yuqiao Reservoir is the primary source of drinking, agricultural and industrial water for approximately 129 villages in the surrounding area (Zhang et al., 2019; Yu and Zhang, 2021). Previous studies have shown that Yuqiao Reservoir is a typical mesotrophic, phosphorus-limited environment (Chen et al., 2012; Zhang et al., 2020). (**Pages 5-6, lines 141-146**)

…The model was calibrated, validated, and employed to predict the variations of water quality resulting from agricultural pollution (Zhang et al., 2013). (**Page 6, lines 151-152**)

2.3 Benchmarking point: model calibration with the strategy of original YRWQM

This section is entirely new and focuses on describing the model calibration with the original strategy to establish a benchmarking point. This section will help the readers better understand the subsequent comparisons of model calibrating performance.

Please see the text in Section 2.3 as follows.

2.3 Benchmarking point: model calibration with the strategy of original YRWQM

As mentioned above, the original YRWQM was initially developed with the purpose of investigating the variations of water quality arising from agricultural pollution in the Yuqiao Reservoir. Both the hydrodynamic and water quality model were calibrated and validated with the observations collected in the six monitoring stations from 2006 to 2007 and performed satisfactorily. A more detailed description of the original YRWQM can be found in Zhang et al. (2013). It should be noted that both the calibration period (the year 2006) and validation period (six months of the year 2007) was under dry hydrological conditions.

Consistent with the original YRWQM calibration strategy, the dry years (the years 2006, 2007, 2010, and 2015) of the decade were selected as the calibration period to establish a benchmarking point for comparison with recalibrated YRWQM. The parameter set obtained from the calibration was implemented to simulate other years with different hydrological conditions to validate the model. The normal years (the years 2009, 2011, and 2014) and the wet years (the years 2008, 2012, and 2013) of the decade were also employed to calibrate and then validate the model with the same method. All of three different models and their parameter sets were eventually comprised with the recalibrated YRWQM to reveal improvements with a wider hydrological variability. (**Page 7, lines 166-179**)

We have also revised the Table 1 for consistency with the benchmarking point.

Table 1: Comparison of the recalibrated YRWQM, the original YRWQM (calibrated with the year 2006), and YRWQM calibrated with original strategy (the dry years 2006, 2007, 2010, and 2015). 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. (Revised Table)

State variables		Statistics	Recalibrated YRWQM	Original YRWQM	YRWQM calibrated with original strategy	
WSE		KGE	0.99	0.99	0.99	
		PBIAS(%)	0.03	-0.10	0.03	
TEM		KGE	0.91	0.91	0.91	
		PBIAS(%)	7.93	8.36	7.93	
TP		KGE	0.10	-1.60	0.11	
		PBIAS(%)	40	-78	20	
Chl *a*		KGE	0.30	-3.03	0.31	
		PBIAS(%)	36	-357	5	
DO		KGE	0.74	-0.19	0.24	
		PBIAS(%)	-2	-53	-22	
Parameters	Units	Descriptions	Recalibrated YRWQM	Original YRWQM	YRWQM calibrated with original strategy	Sensitivity
CChl	mg C / µg Chl	Carbon-to-Chlorophyll ratio for algae	0.080	0.083	0.060	Yes
PM	day^{-1}	The maximum growth rate for algae	2.77	2.00	5.10	Yes
Ke_b	m^{-1}	Background light extinction coefficient	0.410	0.475	0.475	No
TM1	°C	The lower optimal temperature for algal growth	21	9	22	No
TM2	°C	The upper optimal temperature for algal growth	28	15	26	No
KHP	mg/L	Phosphorus half-saturation for algae	0.0019	0.0010	0.0022	No
BMR	day^{-1}	Basal metabolism rate for algae	0.120	0.010	0.140	Yes
PRR	day^{-1}	Predation rate on algae	0.117	0.010	0.280	Yes
K_{RP}	day^{-1}	Minimum hydrolysis rate of refractory particulate organic phosphorus (RPOP)	0.005	0.001	0.067	No
K_{DP}	day^{-1}	Minimum mineralization rate of dissolved organic phosphorus (DOP)	0.060	0.005	0.070	Yes
REAC	/	Reaeration multiplier	1.4	0.9	1.1	Yes

2.4 YRWQM recalibration with EFDC-ACT

The original section 2.3, which describes data sources, have been simplified and integrated with the original section 2.4. The simplified part reduces unnecessary repetition of narrative.

Please see the text in Section 2.4 as follows.

The datasets required for YRWQM recalibration included meteorological data, discharge, precipitation, evaporation, water surface elevation, water temperature, and water quality data. Meteorological data 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. The data above were collected at a frequency of once a day from 2006 to 2015, except for 2012 when water surface elevation data were collected once a month (Zhang et al., 2019). While six monitoring stations were employed by the original YRWQM for calibration and validation (Zhang et al., 2013) to balance the cost with accuracy in calibration, water temperature and water quality data collected from monitoring station S2 were used in recalibrated model evaluation, which represented the water column at the center of the Yuqiao Reservoir. The water quality state variables included TP, Chl a, and DO concentrations (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. (**Page 8, lines 181-191**)

In addition to the above corrections, it should be noted that the model validation process was actually included in this study. YRWQM was recalibrated with a decade of observations, and then validated using data from three different time periods: dry, normal and wet years. Actually, the validation results had been represented in the Fig. 4. This approach to validation using different hydrological years demonstrates the robustness of the model and therefore we did not consider it necessary to include additional validation periods. We have elucidated the validation process more detailed in the caption of Figure 4.

Please see the text in the caption of Figure 4 as follows.

The YRWQM was recalibrated with the decade and calibrated with dry, normal, and wet years respectively. The parameter sets derived from four model calibration strategies were applied to the other hydrological years or the decade to validate the model under the different hydrological conditions. (**Page 13, lines 276-278**)

Comment 3:

***Results**: the current structure is very clear and sound, and I like esp. figure 4, a more clear way to present model performance. However, contrary to the conventional model application studies, the parameter values for calibrated model and performance of the calibration should not be the focus. This paper is stressing the advances of EFCD-ACT, and comparisons of calibration process (time consumption, models sets estimation etc.) and calibration performance (R2, RMSE, NSE, KGE, etc.), between the former YRWQM, and current YRWQM, should be addressed. The improvements and challenges regarding the new tool during application should be presented (section 3.2 and 3.3 should be presented in detail in my opinion).*

Answer 3:

We agree with referee and have made the corrections as suggested. The efficiency of the EFDC-ACT for YRWQM recalibration have been described in an additional paragraph in section 3.1, including time consumption, ease of manual operation, etc. These adjustments will help the readers to better understand the advantages of the automatic calibration toolkit. A comparison of the model performance of the original YRWQM with that of the recalibrated YRWQM has been shown (Table 1). The challenges regarding the new tool during application have been discussed in Section 4.3 and 4.4.

Please see the text in Section 3.1 as follows.

Before giving the complicated details of the analysis on model performance and parameters, we first give an overview of the implementation of the EFDC-ACT on the YRWQM recalibration. Both the manual recalibration and the automatic recalibration experiments with a modeling scale of one year were implemented under the same calculating workstation with Intel(R) Core(TM) i7-10700 CPU @ 2.90 GHz. In the manual recalibration, each iteration took an average of 8.25 h, with an average of 6.41 h spent on modeling and an average of 1.84 h spent on manual pre-processing (parameter adjustment, parameter set recording) and post-processing (result extraction, statistic calculation, time series plotting, model performance recording). The manual pre- and post-processing took 22.35% of the total time of each iteration. In the automatic recalibration, each iteration took an average of 6.43 h, with an average of 0.02 h (77 s) spent on automatic pre- and post-processing. The automatic pre- and post-processing took 0.36% of the total time of each iteration. In terms of time consumption, automatic recalibration with EFDC-ACT takes 21.99% less time than manual operation, thereby reducing the time consumed per iteration by 22.02% (1.82 h). From the perspective of labour saving, EFDC-ACT spared the modeler from tedious repetitive tasks such as extraction of results, calculation of statistical values, and plot of time series. These savings in time and labour provided us with abundant time to analyze and improve the recalibrated YRWQM. (**Pages 9, lines 212-224**)

Comment 4:

***Discussion**: excellent. This part is the best regarding this paper's purpose: presenting EFDC-ACT, and its technical and practices and performance. And its contribution to 3D model application practices.*

Answer 4:

Thank you for your positive comments on this section. We discussed some of the existing problems and tried to outline new challenges and directions for the future development of model calibration. We sincerely hope that this section will attract more new questions and ideas from our readers.

Comment 5:

***Data and source code availability**: EFDC-ACT src code is most available. However, observed data only contains the observation of the reservoir's inflow, outflow, temperature, WSE. However, no data on water quality and chla, which is presented in the calibration. In order to reproduce the study, meteo forcing and other input data for EFDC model are also needed. Data on data.cma.cn is a mess to dig out the data for this study. If there's concern of sharing these datasets or configure files of EFCD, should the authors address the reasons.*

Answer 5:

The Yuqiao Reservoir is an important source of drinking water and therefore the public may be sensitive to the relevant water quality conditions, especially Chl *a*. We have actually given observed values in the results (Fig. 3 and S4 in the Supplyment Material). On the other hand, the efficiency of the automatic calibration toolkit can be demonstrated without the YRWQM, and existing EFDC cases are available for testing. To demonstrate the scientific reproducibility of our study, we would like to share the datasets with editors and referees.

Referee#2:

General Comments:

The authors presented a recalibration effort to improve model performance and discussed the impacts of using wider hydrological variability on the robustness of model performance. The authors also provided a R based optimization tool to automate the calibration of the EFDC model. This tool will improve the efficiencies in calibration of sophisticated models and is valuable to the scientific society. At the same time, the write up of the manuscript needs some improvements. Therefore, my recommendation is revision.

Response: We sincerely appreciate your valuable suggestions to improve the quality of our manuscript. Based on these suggestions, we have carefully revised the manuscript. We hope the new coming manuscript will meet the standards of GMD. If not, we are willing to revise the MS until the reviewers are satisfied with the revision. Below you will find our point-by-point responses to the referee comments/questions.

Minor comments:

Comment 1:

More data and wider hydrological variability will improve the robustness of model performance is well-accepted. Logically, a model calibrated with wider hydrological variability is better. My recommendation is to focus on the quantitative improvement in this manuscript instead of asking the qualitative question. The title should be changed to something like "how much improvement will be achieved with a wider hydrological variability".

Answer 1:

This is quite a valuable idea and we have rephrased our title based on your suggestions after a full revision of the manuscript to make it more focused on the quantitative improvement of the model under a wider hydrological variability. The title has now been changed to *"Recalibration of a three-dimensional water quality model with a newly developed autocalibration toolkit (EFDC-ACT v1.0.0): how much improvement will be achieved with a wider hydrological variability?"* **(Page 1, lines 1-4)**. We have also revised the abstract and the conclusion parts for consistency with the title.

Please see the text in the abstract and the conclusion as follows.

…Furthermore, the KGEs improved by 43~202% in modeling TP-Chl a-DO process when compared to the models calibrated with only dry, normal and wet years. (**Page 1, lines17-18**)

…When compared to the models calibrated with only dry, normal and wet years, the KGEs improved by a maximum of 196%, 134%, 202% in modeling TP, Chl a, and DO respectively. (**Page 19, lines 407-409**)

Comment 2:

The correct way of conducting model calibration using different conditions (such as dry, normal, wet years) when long-term data are not available is to use one set of data to identify the parameter values, then using the identified parameter values to run other conditions. If the results in other conditions are good, we achieved a good model calibration. Otherwise, we go back to the first condition and identify another set of parameter value, and then again put them to other conditions. This process is *iterative until model results from all conditions achieve good results. This manuscript lists totally different parameter values for dry, normal, wet years/ This needs to be fixed.*

Answer 2:

The suggestions are valuable. Actually, we have recalibrated the model with decadal datasets in an iterative way and identified the first parameter set (Recalibrated YRWQM in Table 1 and red curve in Figure 5). We have also presented the model performance of the first set of parameters under different hydrological conditions through the model error statistics (Figure 4). We consider that such a long-term calibration process achieved the same goal as the proposed process, both of which essentially try to demonstrate the robustness of the model under different conditions.

As for the other three parameter sets, they were identified with different hydrological conditions respectively in the way suggested and their model performances in other conditions were also demonstrated in Figure 4. They are important arguments for revealing the impact of the hydrological variability on calibration results and form the basis of the relevant discussion in section 4.2.

For the above reasons, we think retaining all parameter sets would be a more appropriate approach. We have described the parameter sets and the calibration processes in more detail in the captain of Figure 4, which is the same as our response to the Comment 2 of Referee#1.

Please see the text in the caption of Figure 4 as follows.

The YRWQM was recalibrated with the decade and calibrated with dry, normal, and wet years respectively. The parameter sets derived from four model calibration strategies were applied to the other hydrological years or the decade to validate the model under the different hydrological conditions. (**Page 13, lines 276-278**)

Comment 3:

The manual calibration process is indeed a time-consuming process. However, it is a necessary process for the modeler to learn if the developed model is a good representation of the real waterbody. Furthermore, modelers inexperienced or unfamiliar with the model should not be the ones to conduct such manual calibration. It should be noted that modelers inexperienced or unfamiliar with the model could falsely use the auto-calibration tool so that the identified parameter values are not in reasonable range and perfectly fall into the problem of obtaining good model error statistics values while using wrong parameters. Auto-calibration should be viewed as an efficient way to refine calibration after using learning the system with the manual calibration.

Answer 3:

Thanks for your valuable insights. We have further emphasized in the introduction and the discussion parts that even with the aid of auto-calibration, we should spend time learning and understanding the model system and the parameter implications to avoid getting good model error statistics with the wrong parameters. It is necessary for the modelers to learn whether the model is a satisfactory representation of real waterbody.

Please see the text in the Section 1 and Section 4.3 as follows.

The first problem is that the conventional manual calibration method commonly used in process-based hydrodynamic and water quality models, whereas some steps, such as adjustment of inputs, tuning of parameters, evaluation of model performance and visualization of outputs is inefficient and does not guarantee optimal results. Firstly, these steps put modelers inexperienced or unfamiliar with the model through the wringer due to time-consuming and tedious tasks. … (**Page 2, lines 45-48**)

…Although with the aid of auto-calibration, modelers should spend time learning and understanding the model system and the parameter implications to avoid getting good model error statistics values with the wrong parameters. The auto-calibration should be viewed as an efficient way to refine calibration after learning the model system with the manual calibration. (**Page 18, lines 376-379**)

Comment 4:

The authors cited two modeling studies. One was used data from a decade observations and the other used data from two years. "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) recalibrated the Patuxent Estuary model with two years of data and adequately modeled Chl a and DO concentrations." The second example is not using long-term data.

Answer 4:

Thanks for the correction. We initially intended to cite the second model study mentioned as an example of a model recalibration using a wider hydrological variability (a wet year followed by a dry year), neglecting that this example did not use long-term data. We have deleted this literature here.

Comment 5:

The authors mentioned "The seasonal cycle of water temperatures was pronounced, with higher temperatures in summer and lower temperatures in winter.". This is not sufficient to describe a successful model development. Any temperature simulation will generate such seasonal trend given the right weather conditions and inflow temperatures. It is suggested that details should be added such as if highest temperature and lowest temperature are captured by the model.

Answer 5:

Thank you for your correction. In the Results section, we have added details that whether the model captured the highest and lowest temperatures.

Please see the text in the Section 3.1 as follows.

…The recalibrated YRWQM reproduced the seasonal cycle of water temperature with a KGE of 0.91. The highest and lowest water temperatures were grasped with a highest observed value of 31℃ and a corresponding simulated value of 28℃, and a lowest observed value of 0℃ and a same corresponding simulated value. (**Page 9, lines 228-230**)

Comment 6:

The authors mentioned "The 11 primary parameters in the water quality model were listed in Table 1, eight of which governed algal 220 kinetics (CChl, PM, Keb, TM1,TM2, KHP, BMR, PRR), two parameters influenced phosphorus cycling (KRP, KDP), and one affected reaeration (REAC). Among these parameters we also found six of them to be sensitive (Table 1)." It is suggested that the process of choosing the 11 primary parameters should be provided in the manuscript.

Answer 6:

We appreciate the suggestion. Since the EFDC has so many parameters, we first exclude those that are irrelevant or have little influence on the modeling state variables and kinetics. We then referred to the literature mentioned in Line 181 and the original YRWQM to further narrow down the parameter sets to obtain the parameters and their ranges for the model calibration, which were listed in Table S1. The primary parameters listed in Table 1 were further selected during our calibration process based on the biogeochemical characteristics of Yuqiao Reservoir and the model performance, and they significantly influenced the model results. Although some of the parameters are not sensitive to hydrological conditions, we have retained them in Table 1 in view of their significant influence on the model results.

Please see the text in the Section 3.1 as follows.

…These primary parameters were selected during our calibration process based on the biogeochemical characteristics of Yuqiao Reservoir and the model performance, and they significantly influenced the model results. (**Page 10, lines 253-254**)

Comment 7:

Figure 3 shows the model results after recalibration. The model results from the original calibration should be shown together so that readers can visually inspect the improvement of calibration.

Answer 7:

This suggestion is very valuable. We have added the model results from the original calibration strategy (with dry years) in Figure 3 as the orange curve.

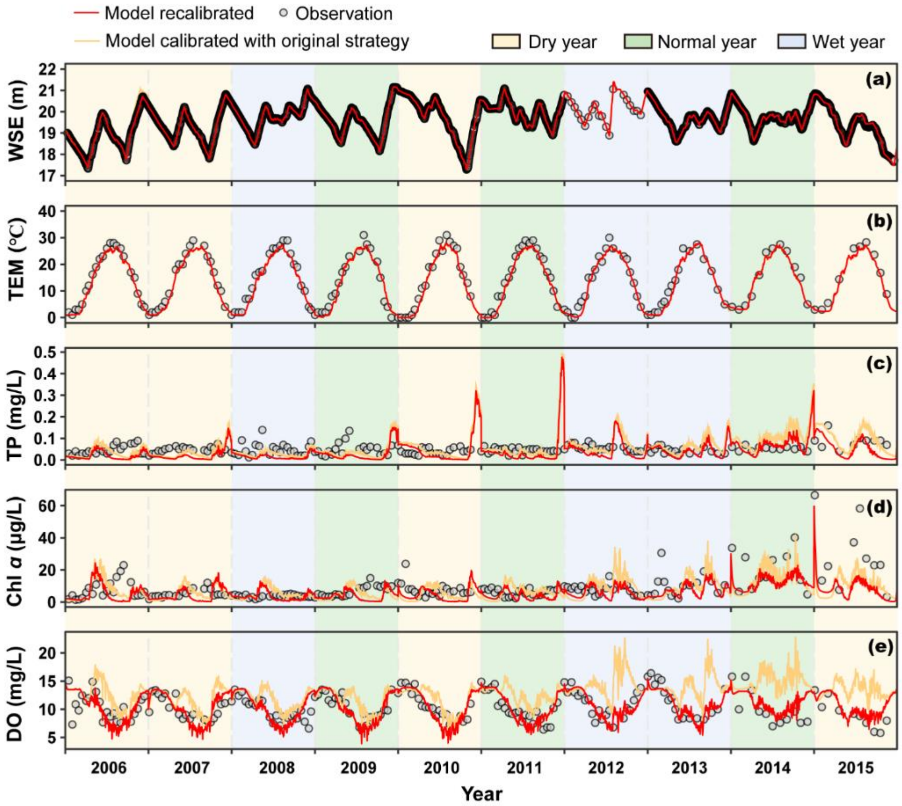

Figure 3: Performance of the model recalibrated and the model calibrated with original strategy (with dry years) at station S2 of the Yuqiao Reservoir (*n* = 3309 for WSE and 190 for other state variables). (Revised Figure)

Comment 8:

The values shown in Table 1 look like the recalibration replaced the algae from the spring growth algae to summer growth algae with much higher metabolism rates. Is there any algae study to support such change of algae type?

Answer 8:

Thank you for your questions, which are quite crucial. We did make lots of adjustments to the parameters associated with algal kinetics. As described in Section 4.1 (from line 282 to line 289), when we used the algal kinetic parameters of the original model directly into the recalibration, modeling Chl *a* concentrations that differed significantly from the observed values. Due to the narrow hydrological variability (only two dry years) of the observed data when constructing the original model (Zhang et al., 2013), the original parameter set was not reasonable and sufficient to describe the growth process of blue-green algae in summer with a low metabolism rate. Noting this problem, we adjusted the parameter ranges in recalibration, such as optimal temperature ranges for algal growth and metabolism rates, and ultimately the parameters identified in recalibration represented summer algal growth in a more accurate way. Some studies also indicated that algal biomass was significantly higher in summer (Ding et al., 2022, Single-variable method for predicting trends in chlorophyll a concentration based on the similarity of time series. Ecological Indicators, 140, 109027, DOI: 10.1016/j.ecolind.2022.109027), and that blue-green algae (cyanobacteria) associated with summer algal blooms did exist in Yuqiao Reservoir (Huo et al., 2018, Molecular detection of microbial communities associated with *Microcystis* vs *Synechococcus* dominated waters in Tianjin, China. Journal of Oceanology and Limnology, 36, 4, 1145-1156, DOI: 10.1007/s00343-018-7182-x).