

RESPONSE TO REVIEWERS

“Evaluating the performance of CE-QUAL-W2 version 4.5 sediment diagenesis model”
August 2025

Dear Editor-in-Chief

Thank you for the opportunity to submit a revised version of our manuscript, “Evaluating the Performance of the CE-QUAL-W2 Version 4.5 Sediment Diagenesis Model,” for consideration in Geoscientific Model Development (GMD). We sincerely appreciate the time and effort you, the Associate Editor, and the Reviewers have devoted to evaluating our work. We have carefully addressed all reviewer comments and incorporated the suggested revisions.

I look forward to hearing from you.

Sincerely,
Manuel Almeida

Reviewer #1

The authors have written an interesting, well-developed methods paper where they compare CE-QUAL-W2's zero-order sediment model against the full sediment diagenesis (SD) model introduced in V4 of the CE-QUAL-W2 water-quality model. I

believe many water-quality modellers using CE-QUAL-W2 are reluctant to try the new SD model due to the sheer number of coefficients in the compartment, so it is interesting that the authors were able to model their waterbody mostly using the default parameters of the diagenesis model. While the authors primarily discussed the results for DO, there does seem to be value in collecting a few sediment samples where possible, based on the better results for TP, TN and (potentially) Chl-a with the SD model.

The introduction was good and provided sufficient context for why the authors thought the work was of interest to the water-quality modelling community.

Author response: We sincerely thank the reviewer for providing thoughtful feedback. We have carefully considered each of the comments and have made corresponding revisions to enhance the manuscript accordingly. We appreciate the recognition of our efforts to compare the zero-order sediment model with the full sediment diagenesis (SD) model in CE-QUAL-W2. We believe that for long-term studies, it is especially relevant to implement a more comprehensive sediment model. As the reviewer correctly noted, a key motivation for this study was to demonstrate that the SD model can produce reasonable and improved results even when using mostly default parameter values—an important consideration for practitioners who may be hesitant to adopt the model due to its complexity.

The methods were sufficient although the information regarding the configuration and calibration of the main water quality model could be more in-depth (e.g., appendix table of most the important coefficients) rather than leaving the reader to have to search through the CE-QUAL-W2 user manual.

Author response: Thank you for your comment. We agree that providing more detailed information on the configuration and calibration of the water quality model enhances the clarity and usefulness of the manuscript. In response, we have added tables A2 to A8 summarizing the most important coefficients and parameters used in the CE-QUAL-W2 model setup. This addition allows readers to understand the model calibration without needing to refer to the user manual. We believe this change improves the transparency and reproducibility of our modeling approach. The following sentence was included in the manuscript:

PAGE 13 LINE 285

“Tables A2 through A8 display the most significant CE-QUAL-W2 coefficients obtained after the calibration process.”

I found one or two sections needed rereading several times to fully understand the objectives of the study and the model setup. Section 2.2 combines the model configuration (e.g., bathymetry, algal groups), a summary of the following method section, and a summary of the overall modelling approach, and I believe this could be better structured by separating the model set-up. Note that machine learning is not my area of expertise and so I am unable to comment on the derivation of the forcing datasets for water-quality.

Author response: Thank you for your comment. We agree with the reviewer's suggestion. Accordingly, two new sections have been added: Section 2.2.1 – Model Setup and Section 2.3 – Modeling Approach. The Methods section has been revised as follows:

PAGE 5 LINE 134-141

2.2.1 Model Setup

The bathymetry of the Torrão reservoir was initially defined using a Digital Elevation Model (DEM) provided by Energias de Portugal, S.A. (EDP) and structured according to the methodology outlined in Wells (2021). The reservoir comprises one main branch (the Tâmega River), three tributaries and one distributed tributary (Fig. 1). Tributaries 1 and 2 are depicted in Fig 1. Tributary 3 represents the inflow from the Douro River into the pump-back system of the Torrão Reservoir. The bathymetric map includes 27 segments, each measuring 1000 meters in length, and a maximum number of 58 layers, each with a depth of 1 meter. Following this preliminary step, the reservoir boundary conditions (including water quality, hydrology, meteorology, and sediment characterization) were defined according to the methods described in Section 2.4. Due to the lack of available information, the model structure only includes a single algae group (Diatoms).

PAGE 6 LINE 145 to PAGE 7 LINE 163

2.3 Modeling approach

To thoroughly evaluate the capability of CE-QUAL-W2 in modeling dissolved oxygen using the sediment diagenesis module, the four available SOD modeling approaches were considered: Zero-order model; First-order model; Zero/First-order model (Hybrid model) and the sediment diagenesis model (SG model). The models were calibrated for the 2016–2021 period (see Section 2.5). During the results analysis, the performance metrics obtained during each model's calibration process were compared, along with the SOD values across the bottom layers of each model. A sensitivity analysis was conducted following calibration to evaluate each model's response: a) to varying POC, PON, and POP values in the case of the SG model; b) to different SOD values in the Zero-order and Hybrid models and c) to varying the initial first order sediment concentration in the case of the First-order model. Section 2.6 details the methodological approach used for the sensitivity analysis. To assess the sensitivity of each model to reductions in external organic matter (OM) and phosphorus ($\text{PO}_4\text{-P}$) inputs, two separate scenario analyses were conducted. The first scenario involved an 80% reduction in OM inflow load, while the second applied an 80% reduction in both OM and $\text{PO}_4\text{-P}$ inflow loads. These reductions were implemented specifically in the main reservoir branch (Branch 1 – Tâmega River), where the majority of nutrient and organic inputs occur. Each sediment model—SD, Zero-order, First-order, and Hybrid—was run under baseline conditions and under both reduction scenarios. The impact on DO dynamics was evaluated using time series of depth- and segment-averaged DO concentrations. Each model—SD, Zero-order, First-order, and Hybrid—was run under baseline conditions and then under this reduced-loading scenario. The evaluation of model performance, along with the results of the sensitivity analysis, provided deeper insights into simulating SOD dynamics using the sediment diagenesis approach in comparison to the other SOD formulations.

The results were well presented visually and with plenty of discussion provided by the authors. However, I was unable to follow what was being discussed and shown regarding TOC and POC in Section 3.3 (lines 310 to 325). It was not clear to me if the black line and circles show TOC or POC as the legend (TOC) and y-axis/caption (POC) are for different variables, nor could I follow how it was concluded that the particulate fraction of organic carbon constituted 40% of the TOC. Lines 310 to 320 and Figure 4 should be clarified.

Author response: Thank you for your comment. We agree with the reviewer that this section was not sufficiently clear. During the sediment characterization, we assumed that the POC was equal to the observed TOC value, primarily because POC was not directly measured. The simulation that used the POC value derived from the observed TOC data is Run 5, which was calibrated and referred to as the W2_SD model (Run 5 – baseline). Run 2 produced the best performance based on the NSE and RMSE criteria. The mean sediment concentration used to characterize Run 5 was 17,712 mg/L, calculated as the average of the following TOC values: 24,000; 20,064; 21,408; 19,296; and 5,376 mg/L (see Table 3). For Run 2, the mean value was 7,085 mg/L, based on the average of 9,600; 8,026; 8,563; 7,718; and 2,150 mg/L (also from Table 3). The value used in Run 2 (7,085 mg/L) is approximately 40% of the value used in Run 5 (17,712 mg/L). Therefore, if Run 5 assumes that POC is equal to the full TOC value, and Run 2 provides the best fit to observed data, it is reasonable to infer that the initial POC value should be approximately 40% of the observed TOC. Additionally, Figure 4 was corrected—the legend now reads initial POC value instead of initial TOC. This section has been revised as follows to improve clarity. Please note that an issue was identified with the initial metric estimates, and all performance metrics were recalculated accordingly. The manuscript results have been updated to reflect this correction.

PAGE 20 LINE 351-371

“The SOD values strongly influence the water column DO; therefore, this parameter was considered to support this analysis. Figure 7 shows the SOD values from the reservoir bottom layer, predicted by the SD model for Runs 1 to 6, compared with the RMSE (Fig7A) and the NSE (Fig7B) values obtained between the predicted water column DO profiles and the mean initial POC values (across all sites values) for each run. These results suggest that Run 4 was the best modeling solution. Considering the results obtained for Run 5 (baseline), Run 4 reduced the RMSE from 2.015 mg/L (Run 5) to 2.011 mg/L (Run 4) and increased the NSE from 0.714 (Run 5) to 0.716 (Run 4). The average SOD value in the bottom layer of the reservoir (across all model segments) decreased from 1.162 g O₂/m²day (Run 5) to 1.071 g O₂/m²day (Run 4). Although the reduction is modest and had only a minor effect on the DO profile predictions (Fig. 9), it suggests that the initial POC values used in Run 5 were likely overestimated. This outcome aligns with the assumption made in Run 5, where all observed TOC was considered to exist entirely as POC. In contrast, Run 4 was characterized using a lower average sediment concentration. Specifically, the mean value used in Run 4 (14170 mg/L) represents approximately 80% of the TOC value used in Run 5 (17712 mg/L), which was derived from observed TOC measurements (see Table 3). This comparison suggests that a more realistic estimate is that about 80% of the total organic carbon exists in particulate form, with the remainder composed of dissolved organic carbon. Run 4 and Run 5 show negligible differences

in the predicted water temperature and DO profiles (Fig. 8 and 9). Table A10 presents the performance metrics for water temperature, DO, TN, TP, BOD₅, and Chl-a obtained for Run 4. While this run improved the DO simulation in the reservoir, results for the other constituents remained very similar to those of Run 5 (baseline). Overall, the water temperature profiles are very well captured by all models (Fig. 8), reflecting their robustness in simulating thermal dynamics. In contrast, DO profiles are more complex and challenging to model due to their sensitivity to multiple interacting processes. Nevertheless, the models were able to capture the main seasonal and vertical trends in DO concentrations, including stratification patterns and general oxygen depletion in bottom layers during warmer months (Fig.9)."

Furthermore, while this paper is of interest for those of us using the CE-QUAL-W2 model, and could be cross-transferred to other waterbodies using the CE-QUAL-W2 model, the authors did not attempt to place their findings in the context of the broader water-quality modelling science, and how this work may contribute. I think this should be added to the discussion to strengthen this submission.

Author response: Thank you for this comment. We appreciate your suggestion to broaden the context of our findings within the field of water-quality modeling science. In response, we have revised the discussion to clarify how our study contributes more broadly to sediment oxygen demand modeling in CE-QUAL-W2 and to the wider field of water-quality modeling.

We now emphasize that while the study's primary focus was to evaluate the performance of the sediment diagenesis (SD) model, the inclusion of alternative formulations (Zero-order, First-order, and Hybrid models) not only allowed for a direct performance comparison but also provided practical insights into model applicability under varying system conditions. We discuss the relative strengths and limitations of each approach, emphasizing how their performance relates to model structure, data availability, and application scale (e.g., short- vs long-term simulations).

Additionally, we highlight how the findings align with broader principles in ecological and environmental modeling, such as model parsimony (Burnham and Henderson, 2002) and user expertise (Piccolroaz et al., 2024). These insights are transferable to other water bodies and modeling frameworks, particularly where users face similar trade-offs between model complexity and data constraints. These revisions aim to better position the study within the broader water-quality modeling literature and demonstrate its relevance beyond the specific application to our study reservoir.

PAGE 31 LINE 538-561

It is important to emphasize that this study was primarily designed to evaluate the performance of the sediment diagenesis model. However, by incorporating alternative SOD modeling approaches, it inevitably allowed for a comparative ranking of model performance, highlighting the relative strengths and limitations of each formulation. The performance limitations of the Zero-order and First-order models can be attributed to their structural simplifications. Specifically, the Zero-order model's strong temperature dependence, coupled with its disregard for the dynamics of organic matter loading, reduces its ability to capture temporal variability driven by external inputs. Similarly, the lower accuracy of the First-order model likely stems from its exclusion of anaerobic decay processes and limited representation of sediment biogeochemistry, which becomes especially relevant under low-oxygen conditions.

The Hybrid model outperformed all other approaches. Considering the principle of parsimony (Occam's razor) (Burnham and Henderson, 2002), the simpler Hybrid model proved more effective than the complex SD model, making it the preferred choice for simulating SOD dynamics in the reservoir. These findings underscore the importance of selecting models that align with the specific characteristics of the system being studied. Simpler models, such as the Hybrid model, may be adequate for steady-state conditions, short- to medium-term forecasts, or scenarios with limited data. The zero-order SOD component of the Hybrid model relies solely on temperature and is decoupled from the water column; therefore, in long-term simulations, this limitation can gradually undermine the model's accuracy. In contrast, the SD model may be more appropriate when the goal is to explore system-wide feedbacks and temporal dynamics over extended periods—especially those involving sediment accumulation and nutrient cycling—where it may provide valuable insight into underlying processes, provided that sufficient observational data become available to support its additional state variables. Moreover, a model's effectiveness heavily depends on the user's familiarity with its structure and their skill in calibration. Yet, it is unrealistic to expect researchers to master the implementation of every available modeling approach. As such, comparisons between models should be interpreted carefully, acknowledging the influence of user expertise on performance outcomes (Piccolroaz et al. 2024). Overall, to strengthen the analysis, it is recommended that users apply all available SOD modeling approaches in the case of the CE-QUAL-W2 model and assess the model's behavior. This comprehensive evaluation provides a solid foundation for further modeling efforts and helps ensure that the chosen approach is well-suited to the system's specific conditions and objectives.

Finally, there were numerous editorial errors throughout the manuscript that need addressing; a few examples below, although there are more:

1) Discrepancies in the citations and the bibliography. Examples include:

Line 54: Should be just 'Zoubabi-Aloui'

Line 73: I believe this should be 'Wells 2021'

Line 139: 'Adelena et al. 2015', does not appear in the bibliography

Line: 142: Should be 'Berger and Wells 2014'

Etc.

Author response: Thank you for bringing this to our attention. We have carefully reviewed the entire manuscript and addressed the editorial issues you noted, including correcting the discrepancies between in-text citations and the bibliography. Specifically:

Line 65: Corrected to 'Zouabi-Aloui'

Line 73: The sentence with this reference was removed.

Line 139: Removed 'Adelena et al. 2015' as it does not appear in the bibliography

Line 126: Corrected to 'Berger and Wells 2014'

In addition, we have conducted a thorough review to identify and fix any remaining citation and formatting inconsistencies throughout the manuscript and reference list. We appreciate your careful reading and helpful comments.

2) Also seems to be some discrepancies in the Section number cross-refs (for example Lines 106 and 109, refer to Section 1.2.3 and 1.2.4, respectively, with other instances throughout the document).

Author response: Thank you for pointing this out. The section numbers have been corrected accordingly.

3) Line 285 .. for DO, "...the W2_zero-order model performed slightly better according to all metrics, with the exception of PBIAS". I am wondering if the authors mean R² (which is marginally worse than the SD model)? Perhaps it is me that is mistaken, but for PBIAS it seems the zero-order model performs better for DO than the SD model, with the assumption the goal is a low-bias model. This should be clarified.

Author response: Thank you for pointing this out. You are correct to note the inconsistency. Following the inclusion of three additional SOD models and a recalculation of the performance metrics, we have revised the sentence in question to reflect the updated results more accurately. The original statement has been replaced with the following text to clarify the comparative performance of the models with respect to DO, including a corrected interpretation of PBIAS and R² values.

PAGE 13 LINE 286 to PAGE 14 LINE 312

"Tables A2 through A8 display the most significant CE-QUAL-W2 coefficients obtained after the calibration process. The results of the calibration process for all models, are presented in Table 4 and Table A9 and illustrated in figures 3 to 6 and figures 8 and 9. The performance metrics for water temperature across the different sediment models show consistent accuracy, with NSE and R² values ranging from 0.95 to 0.96 and minimal variation across models. The RMSE and MAE for temperature also remain low, indicating reliable thermal performance regardless of the sediment model applied. In contrast, DO predictions show more variability. The Hybrid model achieved the best overall DO performance, with the highest NSE (0.76 ± 0.30) and R² (0.76 ± 0.31), as well as the lowest RMSE (1.87 ± 0.72) and MAE (1.22 ± 0.55), while maintaining a near-zero PBIAS (-0.55 ± 11.14), indicating minimal systemic bias. The Zero-order model also performed reasonably well, with slightly lower error metrics than the SD model. The First-order model, however, showed the weakest DO performance, with a lower NSE (0.68 ± 0.22), higher RMSE (2.15 ± 0.82), and a significant negative PBIAS (-12.17 ± 15.44), suggesting an underestimation of oxygen concentrations. Overall, the results suggest that while temperature simulation is robust across all models, DO dynamics are better captured using the Hybrid or Zero-order models, with the Hybrid model offering the most balanced and accurate representation under the tested conditions. However, the differences in performance metrics for DO among the models are relatively small and often fall within overlapping standard deviations, with the exception of the First-order model, which consistently shows lower accuracy and higher bias, suggesting that while the Hybrid model offers slightly better overall performance, the improvements over the SD and Zero-order models are modest and should be interpreted with caution. In terms of nutrient dynamics, the Hybrid and Zero-order models improve TN and TP predictions relative to the SD and First-order models. The Hybrid model, for example, improves TN R² to 0.31 and TP to 0.27, although the associated biases remain significant (e.g., -18.75% for TN and $+36.49\%$ for TP). BOD₅ and Chl-a remain poorly simulated across all models, with R² values consistently low (≤ 0.06 for Chl-a and ≤ 0.03 for BOD₅), and large PBIAS values, particularly in the SD and First-order configurations. The Zero-order model slightly reduces bias in Chl-

a and Total N compared to the SD model but performs poorly for TP due to a large overestimation (PBIAS = 103.43%) (Fig.4D). Notably, the SD and First-order models failed to reproduce observed phosphorus release events from sediments on 2018-09-18, 2020-09-08, and 2021-08-31 (Figures 3D and 5D). In contrast, the Hybrid model successfully captured these events by modeling phosphorus release as a linear function of SOD, providing a more realistic representation of sediment–water nutrient interactions (Fig.6D). Overall, while no model fully captures the complexity of all constituents, the Hybrid model consistently provides the most balanced and improved representation, particularly for DO and nutrient parameters.”

4) Line 312: It should read Fig4b after NSE. **Author response:** Thank you for point this out. The sentence was corrected.

PAGE 20 LINE 351-354

“The SOD values strongly influence the water column DO; therefore, this parameter was considered to support this analysis. Figure 7 shows the SOD values from the reservoir bottom layer, predicted by the SD model for Runs 1 to 6, compared with the RMSE (Fig7A) and the NSE (Fig7B) values obtained between the predicted water column DO profiles and the mean initial POC values (across all sites values) for each run.”

Reviewer #2

Comments on Evaluating the performance of CE-QUAL-W2 version 4.5 sediment diagenesis model Manuel Almeida, Pedro Coelho

Overall, this is a useful evaluation of the sediment diagenesis model in CE-QUAL-W2 model. The next logical step would be to compare first order and zero order model with sediment diagenesis. The MAE for temperature simulations seems high compared to other systems and this can drastically affect dissolved oxygen profiles. This may be the result of inflow temperatures as well as outflow dynamics. It would be useful to work on improving temperature predictions (if there is a path forward) and to see how that affects the results in this study. The dissolved oxygen profiles are very complex in this reservoir and often the model reproduced the correct shape of the profiles.

Author response: We appreciate the time and effort that the reviewer has invested in evaluating our manuscript. Their insightful comments and constructive suggestions have been invaluable in helping us improve the quality and clarity of our work. We have addressed the reviewer's suggestion, and the revised manuscript now includes four distinct modeling approaches: (i) a user-defined zero-order model, (ii) a simple predictive first-order model, (iii) a hybrid approach combining the zero- and first-order models, and (iv) the sediment diagenesis model. While revising the manuscript, we discovered that the calibration metrics had not been properly applied. After correcting this, the mean absolute error (MAE) for water temperature across all simulations is now $0.88\text{ }^{\circ}\text{C} \pm 0.02\text{ }^{\circ}\text{C}$, which can be considered a very reasonable value. A new figure—Figure 8—was included to show the observed and predicted water temperature profiles, allowing for a clearer comparison of model performance across depths and time. We agree with the reviewer that the dissolved oxygen profiles in the reservoir are quite complex. However, the discrepancies between the modeled and observed dissolved oxygen concentrations are primarily driven by factors other than water temperature—namely, the inflow of organic matter and algal biomass. Since the boundary conditions are the same across all models and the models reproduce the dissolved oxygen profiles reasonably well, we believe that the modeling approach is both sound and well-substantiated.

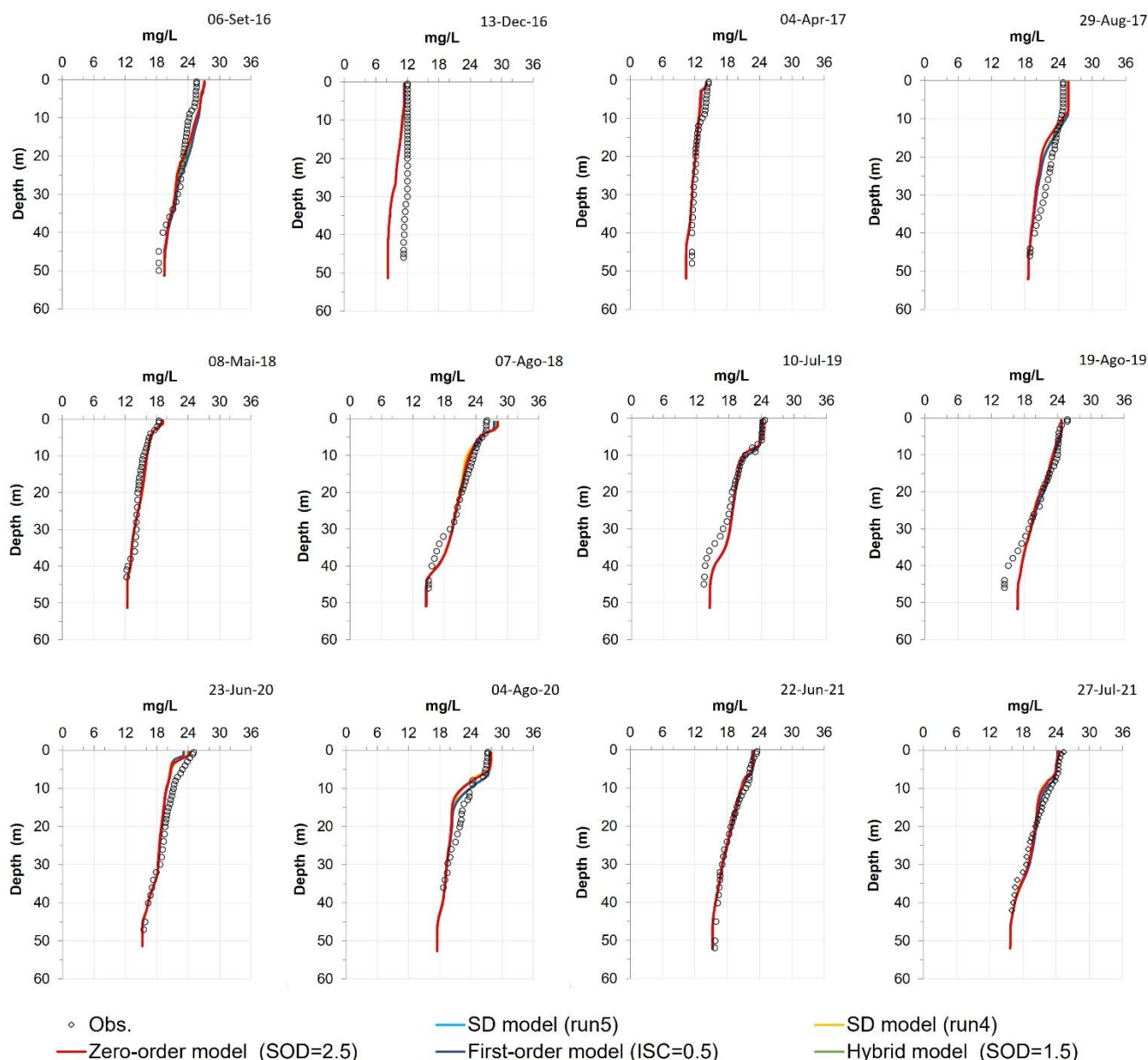


Figure 8: Observed water temperature profiles (300 m from the dam) compared to predicted profiles using the SD model (Run 4) and (Run 5 - baseline), Zero-order model (zero-order SOD = 2.5 g O₂/m²/day - baseline); First-order model (ISC= 0.5 g/m² - baseline) and the Hybrid model (zero order SOD= 1.0 g O₂/m²/day - baseline).

There were a few comments on the text which are summarized below:

Line 42-43: “if the SOD is not accurately computed the waterbody phosphorous balance will, in turn, be incorrect.” This expression needs further explanation. If the zero order SOD model is used, then the anoxic release of PO₄ is a linear function of the SOD in the CE-QUAL-W2 model, in other words SOD[g O₂/m²/day]*PO₄release rate [g P/g O₂]. If one uses a predictive model, like sediment diagenesis, then the SOD and P release from the sediments will be a function of the organic and nutrient loading of particulate matter from the water column.

Author response: Thank you for pointing this out. We agree with the reviewer that this section was unclear and have revised the text accordingly, as follows:

PAGE 2 LINE 42-60

“The main challenge with these modeling approaches is that the sources of DO depletion—such as the inflow of organic matter or algal mortality—can significantly influence DO dynamics, and these sources must be well characterized to ensure accurate predictions. While the baseline model can reproduce observed DO profiles with reasonable accuracy, its predictive reliability may be compromised if key DO sinks and sources are not well defined.

For example, the model’s response to a reduction in external phosphorus loading is influenced by internal phosphorus release from sediments during anoxic periods. In CE-QUAL-W2, when a zero-order SOD model is used, the anoxic release of phosphate (PO_4) is modeled as a linear function of SOD: $\text{SOD} [\text{g O}_2/\text{m}^2\text{day}] \times \text{PO}_4 \text{ release rate} [\text{g P/g O}_2]$. Thus, any error in the estimation of SOD will directly affect the predicted internal phosphorus loading, and by extension, the overall phosphorus balance in the waterbody. In contrast, when using the predictive sediment diagenesis model, internal phosphorus loading depends on the organic and nutrient inputs from particulate matter in the water column and the sediment’s biogeochemical response, which is highly influenced by the initial value of particulate organic carbon (POC). As a result, this approach introduces additional uncertainty when key particulate components are not adequately measured or constrained in both the water column and sediments. Calibrating other constituents, such as orthophosphate (P-PO_4), can help reduce uncertainty. P-PO_4 is released from sediments under anaerobic conditions, and its calibration can enhance the accuracy of DO modeling. Still, this release is influenced by multiple factors, including the initial sediment P-PO_4 concentration and the release rate (in the zero-order model), or the mineralization of POP (in the diagenesis model). In both cases, significant uncertainty remains without observed data for POC, PON, and POP in both the water column and sediments. Of these, POC has the most significant influence on SOD, making access to sediment POC data essential for improving model accuracy, even when PON and POP measurements are lacking.”

Line 48: “In other words, the modeling uncertainty may diminish but will persist without observed POC, PON and POP” – it is unclear, is this a discussion about water column POC, PON and POP or sediment POC, PON, and POP?

Author response: We appreciate the reviewer’s comment and acknowledge that the original sentence lacked clarity. We believe that the previous revised sentence (PAGE 2 LINE 43-61), addresses this concern by clearly referring to the need for observed POC, PON, and POP data in both the water column and sediments.

Line 73: “dissolved oxygen uptake rates in the water column (Wells, 2011).” – reference to Wells, 2011 not found in references~

Author response: Thank you for pointing this out. The reference was corrected to Wells, 2021.

Line 114-115: “This is not, however, a predictive approach, as, other than variations resulting from the temperature dependence of the decay rate, the rates remain constant over time (Wells, 2021).” – Note that also when there is anoxia in the water column SOD is turned OFF.

Author response: Thank you for your comment. The following sentence was included in the manuscript.

PAGE 4 LINE 111-113

“The zero-order model is not a predictive approach, as, other than variations resulting from the temperature dependence of the decay rate, the rates remain constant over time (Wells, 2021). Additionally, under anoxic conditions in the water column, SOD is disabled in the model.”

Line 142: “model has been elaborated in works by Prakash et al. (2014), Berg and Wells (2014), and Vandenberg et al. (2015)” – change ‘Berg’ to ‘Berger’. Also, the V4.5 model had many enhancements to the sediment diagenesis model as outlined in the User Manual. The initial V4 model is much different and limited compared to the V4.5 model. **Author response:** Thank you for pointing this out. The following sentence was included in the manuscript:

PAGE 4 LINE 123 to PAGE 5 LINE 126

“The conceptual framework of the model has been elaborated in works by Prakash et al. (2014), Berger and Wells (2014), and Vandenberg et al. (2015). It is important to note that significant enhancements to the sediment diagenesis module were introduced in version 4.5 of the model, as detailed in the User Manual (Wells, 2021).”

Line 157-158: “The meteorological data used to drive the model, including hourly air temperature, dew point, solar radiation, cloud cover, and wind characteristics, were sourced from ERA5-Land...” – Was there an effort to ground-truth this ERA5-Land dataset with on-site meteorological measurements in the area as a check?

Author response: Thank you for pointing this out. Unfortunately, there are no on-site meteorological stations within the study region available to directly validate the ERA5-Land dataset. However, at the initial stage of the study, we referred to the findings of Almeida and Coelho (2023b), “A First Assessment of ERA5 and ERA5-Land Reanalysis Air Temperature in Portugal,” and Barbosa et al. (2022), “Extreme Heat Events in the Iberian Peninsula from Extreme Value Mixture Modeling of ERA5-Land Air Temperature.” Their analyses demonstrated a strong correlation between observed and reanalysis air temperature data at both daily and seasonal timescales, supporting the reliability of ERA5-Land data in this region. Moreover, the model's performance metrics for water temperature prediction further support the adequacy of the meteorological forcing, indicating that it was appropriately captured and contributed to the accurate simulation results. The following sentence was added to the manuscript to reflect this clarification:

PAGE 7 LINE 165-171

“The meteorological data used to drive the model, including hourly air temperature, dew point, solar radiation, cloud cover, and wind characteristics, were sourced from ERA5-Land, a high-resolution reanalysis dataset optimized for land applications.

Although no on-site meteorological stations are available in the study area for direct validation, studies by Almeida and Coelho (2023b) and Barbosa et al. (2022) have demonstrated a strong correlation between ERA5-Land air temperature data and observed measurements at regional scales, supporting the reliability of this dataset for our modeling purposes. Furthermore, the accuracy of water temperature predictions in our simulations indicates that the meteorological forcing was well represented, confirming the suitability of ERA5-Land data for driving the model.”

Line 235: “six state variables was evaluated with five different metrics (vide section 1.2.6).” – note sure what ‘(vide section’ means – typo?

Author response: Thank you for pointing this out. We intended “vide” to direct the reader to Section 1.2.6 for further details. However, we recognize that this usage may be unclear or unfamiliar to some readers. To improve clarity, we have replaced “vide” with “see” in the revised manuscript.

Line 245: “two parameters retained their default values shown in Table 1.” – I think Table 1 is an incorrect table reference.

Author response: Thank you for pointing this out. You are correct—the correct table reference is Table 3, not Table 1. We have updated the manuscript accordingly.

Line 305: Figure 3 is very hard to see data and model. Figure needs to be broken up or redone to allow others to see model vs data clearly.

Author response: Thank you for pointing this out. We have included one figure per model.

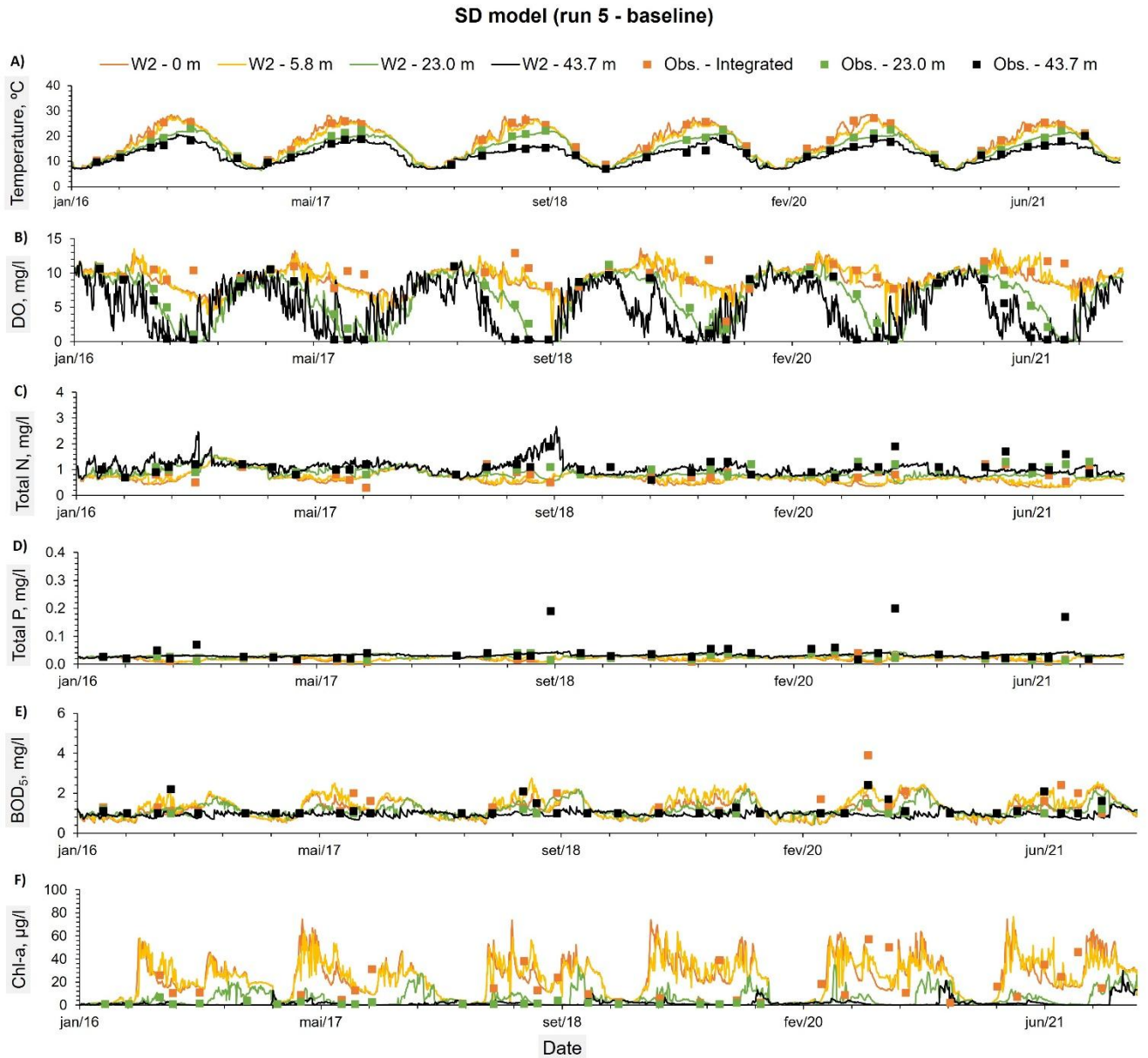


Figure 3: Constituents observed values at three different depths: (a) an integrated sample between the reservoir surface and an average depth of 5.8 meters, (b) an average depth of 23 meters, and (c) an average depth of 43.7 meters. These observed values were compared with the predicted time series from the SD model (run 5 - baseline) (A to F) for the same depths.

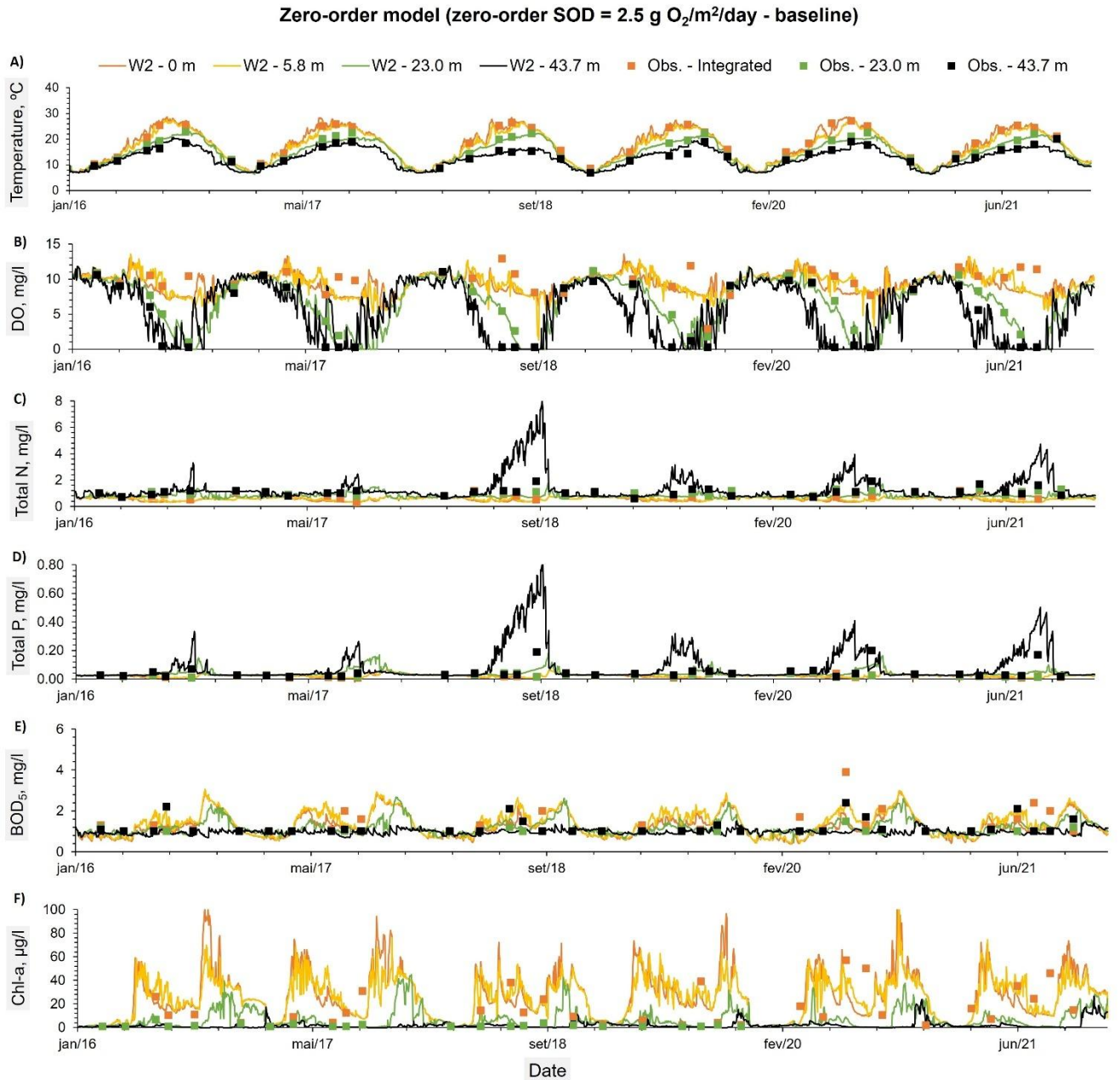


Figure 4: Constituents observed values at three different depths: (a) an integrated sample between the reservoir surface and an average depth of 5.8 meters, (b) an average depth of 23 meters, and (c) an average depth of 43.7 meters. These observed values were compared with the predicted time series from the Zero-order model (zero order SOD = 2.5 g O₂/m²/day - baseline) (A to F) for the same depths.

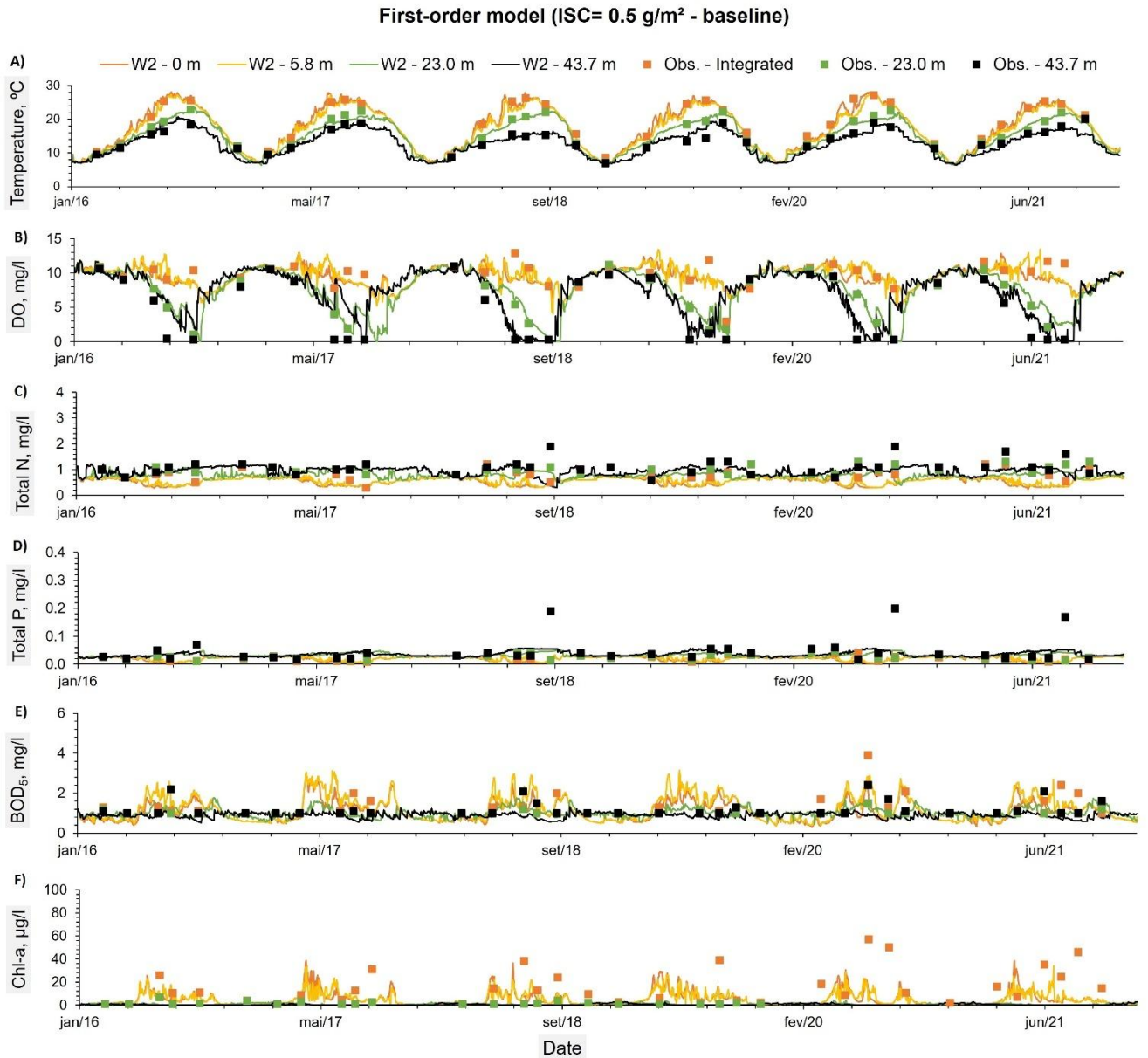


Figure 5: Constituents observed values at three different depths: (a) an integrated sample between the reservoir surface and an average depth of 5.8 meters, (b) an average depth of 23 meters, and (c) an average depth of 43.7 meters. These observed values were compared with the predicted time series from the First-order model (ISC=0.5 g/m² - baseline) (A to F) for the same depths.

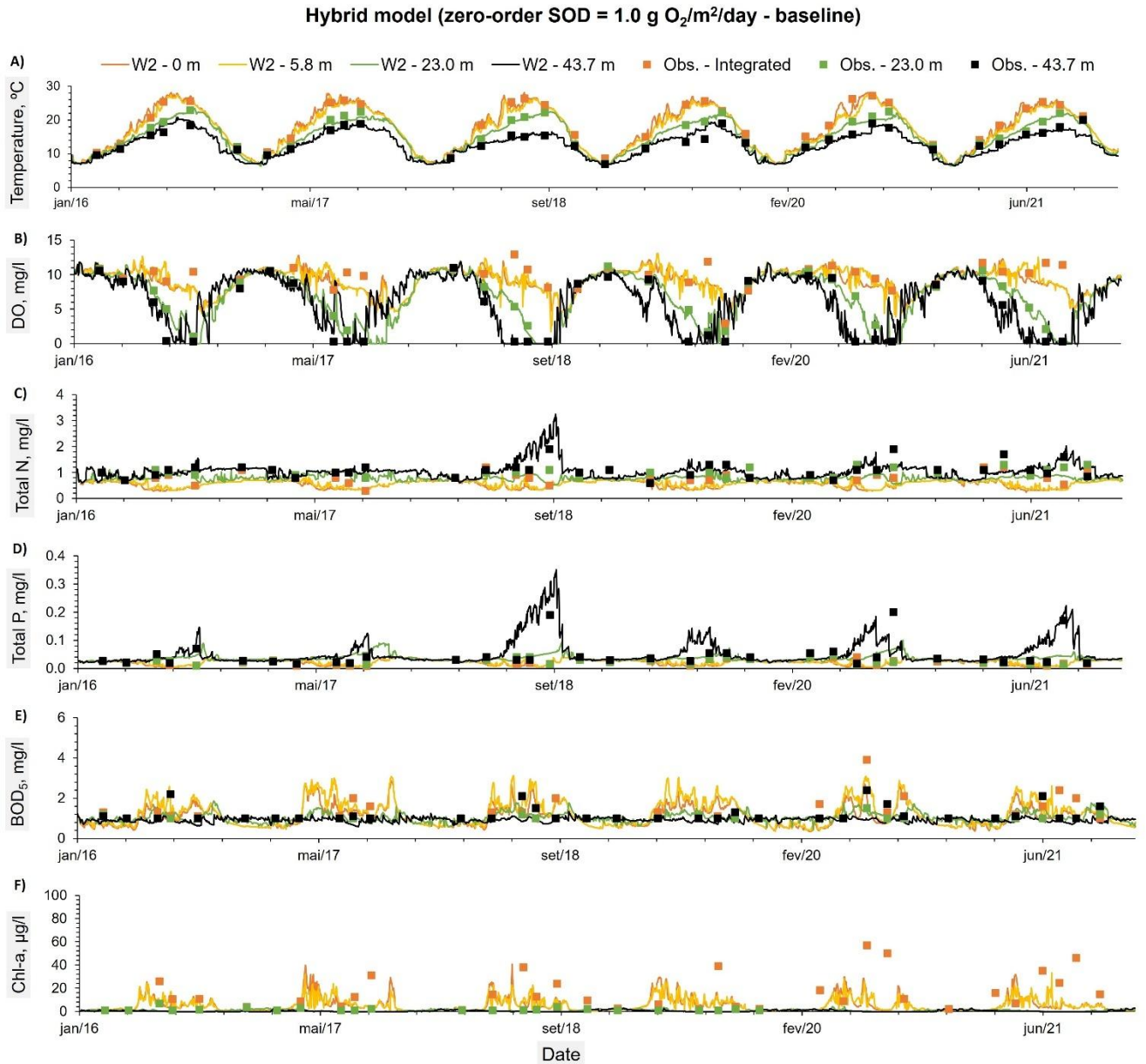


Figure 6: Constituents observed values at three different depths: (a) an integrated sample between the reservoir surface and an average depth of 5.8 meters, (b) an average depth of 23 meters, and (c) an average depth of 43.7 meters. These observed values were compared with the predicted time series from the Hybrid model (zero order SOD= 1.0 g O₂/m²/day - baseline) (A to F) for the same depths

Line 391: “W2_zero-order model ($2.50 \text{ gO}_2/\text{m}^2/\text{day}$) was significantly higher than the mean SOD computed with the best W2_SD model (Run 2) ($0.810 \text{ gO}_2/\text{m}^2/\text{day}$).” – This is not a correct comparison since the Zero order SOD was at 20oC (or at its maximum) and the SD model result is actual SOD at the temperature at the bottom of each segment. Looking at the temperature near the bottom in Fig 3 a year-round average is probably around 10-12oC – hence much lower year-round than the 20oC maximum rate.

Author response: Thank you for pointing this out. The reviewer is correct—we inadvertently used fixed maximum rates for zero-order SOD instead of representing it as a function of temperature. In the revised version we compare the temperature-corrected zero-order SOD value (using bottom water temperature) with the SOD flux from the sediment diagenesis model as applied to the bottom water layer, since this reflects the total sediment oxygen demand at the sediment-water interface.

Line 392: “This can be explained by the fact that the W2_zero-order model SOD represents all of the reservoir’s DO uptake rate in the water column and not just the sediment uptake.” – See comment above – it is related to the temperature. The zero order model only is for sediment demand, not water column demand.

Author response: Thank you for pointing this out. The reviewer is correct—this sentence, as written, is inaccurate. We acknowledge that the zero-order SOD model specifically represents oxygen consumption at the sediment-water interface, and does not account for other oxygen-demanding processes such as BOD decay or nitrification in the water column, which are modeled separately. Our original intention was to suggest that, conceptually, a higher SOD value might reflect the overall oxygen demand, including contributions from other sources influencing oxygen uptake. However, as previously mentioned, the zero-order SOD was initially not computed using bottom-layer temperature, making the interpretation misleading. Therefore, this sentence is no longer valid and has been removed from the revised manuscript.

Line 400: “The zero-order model employs a constant SOD value that only varies with water temperature and does not account for organic matter decay or its impact on SOD values.” – Why did you not use the zero order model with the first order model as reported in your introduction?

Author response: Thank you for the comment. In this study, we compared the performance of the zero-order sediment model and the sediment diagenesis model in simulating observed dissolved oxygen profiles. While CE-QUAL-W2 allows for the simultaneous use of zero-order and first-order sediment compartments, we initially chose not to include the first-order model, as it was not essential to our original research objective. Our goal was to evaluate the performance of two contrasting modeling approaches: the zero-order model, which is the simplest representation of sediment oxygen demand in CE-QUAL-W2, and the sediment diagenesis model, which is the most detailed. This choice allowed us to assess model behavior across the spectrum of complexity—from a highly simplified empirical approach to a more process-based, predictive framework. The first-order model, although it provides a dynamic response to increased organic matter flux to the sediments, does not simulate nutrient release processes such as phosphorus release. Representing such processes would require coupling it with the zero-order model, introducing additional

complexity and interaction effects that were beyond the scope of our initial comparison. However, following the reviewer's suggestion, we have revised the manuscript to include two additional modeling approaches involving the first-order model. The revised manuscript now evaluates four distinct sediment modeling configurations: (i) a user-defined zero-order formulation decoupled from the water column, (ii) a simple predictive first-order model, (iii) a hybrid approach combining zero- and first-order models, and (iv) the sediment diagenesis model. The manuscript has been updated accordingly to reflect these additions and all models were made available in Almeida, M., and Coelho, P., 2025. Furthermore, a new section—3.4 Inflow Organic Matter and Phosphorus Load Reduction Scenarios—was added to assess the sensitivity of each model to reductions in external inputs of organic matter (OM) and phosphorus ($\text{PO}_4\text{-P}$). Two separate scenario analyses were conducted: the first involved an 80% reduction in OM inflow, and the second applied an 80% reduction in both OM and $\text{PO}_4\text{-P}$ inflow loads. These reductions were implemented specifically in the main reservoir branch (Branch 1 – Tâmega River), which receives the highest nutrient and organic inputs. Accordingly, the Methods section was updated to reflect these new scenarios.

PAGE 1 LINE 7-23

Abstract

“This research evaluates the performance of the CE-QUAL-W2 v4.5 sediment diagenesis model in simulating water temperature, dissolved oxygen, total phosphorus, total nitrogen, chlorophyll-a, and biochemical oxygen demand in a Portuguese reservoir over a six-year period (2016–2021). The model was calibrated using 35 observed profiles of temperature and dissolved oxygen, as well as six annual measurements of total nitrogen, total phosphorus, chlorophyll-a, and biochemical oxygen demand at multiple depths. To benchmark performance, three alternative sediment oxygen demand formulations—a Zero-order, First-order, and a Hybrid model combining both approaches—were also implemented and compared. All models achieved NSE and RMSE values within or near the ranges reported in the literature, effectively capturing the system's water quality dynamics. Among them, the Hybrid model yielded the best overall performance while maintaining a simpler structure (Water temperature - NSE: 0.96 ± 0.18 ; RMSE: 1.09 ± 0.23 °C; Dissolved oxygen - NSE: 0.76 ± 0.30 ; RMSE: 1.87 ± 0.72 mg/L). The sediment diagenesis model exhibited similar performance metrics (Water temperature - NSE: 0.95 ± 0.18 ; RMSE: 1.13 ± 0.28 °C; Dissolved oxygen - NSE: 0.71 ± 0.14 ; RMSE: 2.01 ± 0.59 mg/L). Overall, the results suggest that the diagenesis model may be better suited for capturing detailed process-based dynamics over extended timeframes, whereas simpler models, such as the Hybrid model, are more appropriate for short- to medium-term applications or situations with limited data availability. Hopefully, the results of this study will help improve water management strategies by supporting more informed model selection tailored to the temporal scope and data constraints of reservoir monitoring programs.”

PAGE 3 LINE 75-79

“To achieve this, the water quality of a highly productive reservoir was simulated over a six-year period (2016–2021) using the CE-QUAL-W2 v4.5 model. The simulation incorporated a Zero-order sediment model, a First-order model, a Hybrid model combining both approaches, and a sediment diagenesis model. The Zero-order, First-order, and Hybrid models were included to provide alternative representations of

sediment oxygen demand, enabling comparative analysis and supporting the calibration and evaluation of the more complex sediment diagenesis model.”

PAGE 4 LINE 108-121

“This model represents SOD through four distinct approaches: (i) a user-defined zero-order formulation that is decoupled from the water column, (ii) a simple predictive first-order model, (iii) a hybrid approach combining the zero- and first-order methods, and (iv) a comprehensive sediment diagenesis model. The zero-order model is not a predictive approach, as, other than variations resulting from the temperature dependence of the decay rate, the rates remain constant over time (Wells, 2021). Additionally, under anoxic conditions in the water column, SOD is disabled in the model. The first-order sediment model does not function as a full sediment diagenesis model, as it lacks the capability to track the fate of organic nutrients delivered to the sediments, their breakdown, and the release of byproducts into the water column under low-oxygen conditions. However, it does represent the deposition of particulate organic matter and dead algal biomass, along with the resulting oxygen demand imposed on the water column. By including this first-order sediment process, the model becomes sensitive to increased organic loading to the sediment, which in turn influences sediment oxygen demand. A combination of the zero and first order model can be considered where organic materials accumulate and decay in the sediments under aerobic conditions and are released based on the SOD zero-order decay rate under anaerobic conditions. In contrast, the sediment diagenesis model simulates kinetic processes occurring within the sediment and at the sediment–water interface.”

PAGE 13 LINE 286 to PAGE 15 LINE 327

“Tables A2 through A8 display the most significant CE-QUAL-W2 coefficients obtained after the calibration process. The results of the calibration process for all models, are presented in Table 4 and Table A9 and illustrated in figures 3 to 6 and figures 8 and 9. The performance metrics for water temperature across the different sediment models show consistent accuracy, with NSE and R^2 values ranging from 0.95 to 0.96 and minimal variation across models. The RMSE and MAE for temperature also remain low, indicating reliable thermal performance regardless of the sediment model applied. In contrast, DO predictions show more variability. The Hybrid model achieved the best overall DO performance, with the highest NSE (0.76 ± 0.30) and R^2 (0.76 ± 0.31), as well as the lowest RMSE (1.87 ± 0.72) and MAE (1.22 ± 0.55), while maintaining a near-zero PBIAS (-0.55 ± 11.14), indicating minimal systemic bias. The Zero-order model also performed reasonably well, with slightly lower error metrics than the SD model. The First-order model, however, showed the weakest DO performance, with a lower NSE (0.68 ± 0.22), higher RMSE (2.15 ± 0.82), and a significant negative PBIAS (-12.17 ± 15.44), suggesting an underestimation of oxygen concentrations. Overall, the results suggest that while temperature simulation is robust across all models, DO dynamics are better captured using the Hybrid or Zero-order models, with the Hybrid model offering the most balanced and accurate representation under the tested conditions. However, the differences in performance metrics for DO among the models are relatively small and often fall within overlapping standard deviations, with the exception of the First-order model, which consistently shows lower accuracy and higher bias, suggesting that while the Hybrid model offers slightly better overall performance, the improvements over the SD and Zero-order models are modest and should be interpreted with caution. In terms of nutrient dynamics, the Hybrid and Zero-order models improve TN and TP predictions relative to the SD and First-order models.

The Hybrid model, for example, improves TN R^2 to 0.31 and TP to 0.27, although the associated biases remain significant (e.g., -18.75% for TN and +36.49% for TP). BOD₅ and Chl-a remain poorly simulated across all models, with R^2 values consistently low (≤ 0.06 for Chl-a and ≤ 0.03 for BOD₅), and large PBIAS values, particularly in the SD and First-order configurations. The Zero-order model slightly reduces bias in Chl-a and Total N compared to the SD model but performs poorly for TP due to a large overestimation (PBIAS = 103.43%) (Fig.4D). Notably, the SD and First-order models failed to reproduce observed phosphorus release events from sediments on 2018-09-18, 2020-09-08, and 2021-08-31 (Figures 3D and 5D). In contrast, the Hybrid model successfully captured these events by modeling phosphorus release as a linear function of SOD, providing a more realistic representation of sediment–water nutrient interactions (Fig.6D). Overall, while no model fully captures the complexity of all constituents, the Hybrid model consistently provides the most balanced and improved representation, particularly for DO and nutrient parameters.”

Table 4: Metrics between observed and predicted values for all models. Water temperature and DO metrics were obtained from 36 observed and predicted profiles.

Constituent	SD model (run 5 - baseline)				
	NSE	R^2	PBIAS	RMSE	MAE
Water temperature	0.95±0.18	0.96±0.07	1.96±3.08	1.13±0.28	0.89±0.26
DO	0.71±0.14	0.73±0.29	4.43±15.06	2.01±0.59	1.38±0.46
Constituent	Zero-order model (zero-order SOD = 2.5 g O ₂ /m ² /day - baseline)				
	NSE	R^2	PBIAS	RMSE	MAE
Water temperature	0.95±0.19	0.96±0.07	1.91±3.09	1.13±0.28	0.89±0.25
DO	0.73±0.20	0.74±0.30	1.75±15.87	1.97±0.74	1.29±0.57
Constituent	First-order model (ISC= 0.5 g/m ² - baseline)				
	NSE	R^2	PBIAS	RMSE	MAE
Water temperature	0.96±0.19	0.96±0.08	1.46±2.97	1.09±0.23	0.85±0.20
DO	0.68±0.22	0.73±0.27	-12.17±15.44	2.15±0.82	1.50±0.65
Constituent	Hybrid model (zero-order SOD= 1.0 g O ₂ /m ² /day - baseline)				
	NSE	R^2	PBIAS	RMSE	MAE
Water temperature	0.96±0.18	0.96±0.08	1.45±2.97	1.09±0.23	0.85±0.20
DO	0.76±0.30	0.76±0.31	-0.55±11.14	1.87±0.72	1.22±0.55

PAGE 24 LINE 404 to PAGE 25 LINE 417

Figure 11 shows the RMSE (Fig. 11A) and the NSE (Fig. 11B) values between observed and predicted water column DO profiles for all models: SD model (Runs 1 to 6), Zero-order model and Hybrid model, each with six different SOD values ranging from 0.5 to 3.0 g/m²/day, along with the corresponding reservoir SOD values. Additionally, this figure illustrates how the First-order model varies with the initial sediment concentration. Among the four models evaluated, the Hybrid model

demonstrated the best overall performance in predicting DO concentrations in the reservoir. With an average SOD of 1.49 g O₂/m²day, the hybrid model achieved the lowest RMSE (1.87 mg/L) and highest NSE (0.76), demonstrating superior predictive accuracy. The Zero-order model followed closely, reaching optimal performance at an average zero-order SOD of 1.43 g O₂/m²day, with an RMSE of 1.965 mg/L and an NSE of 0.732. The SD model also performed well, attaining its best accuracy at an average SOD of 1.07 g O₂/m²day, where the RMSE decreased to 2.011 mg/L and the NSE peaked at 0.716; however, further improvements plateaued beyond this point. In contrast, the First-order model consistently exhibited higher RMSE values (ranging from 2.15 mg/L to 2.22 mg/L) and lower NSE values (between 0.66 and 0.68), regardless of the initial sediment concentration. Moreover, its SOD at the bottom layer remained relatively stable, indicating limited sensitivity to input variations. Overall, these results underscore the hybrid model's robustness and accuracy, followed by the Zero-order and SD models, while the First-order model demonstrated the weakest performance in this context.

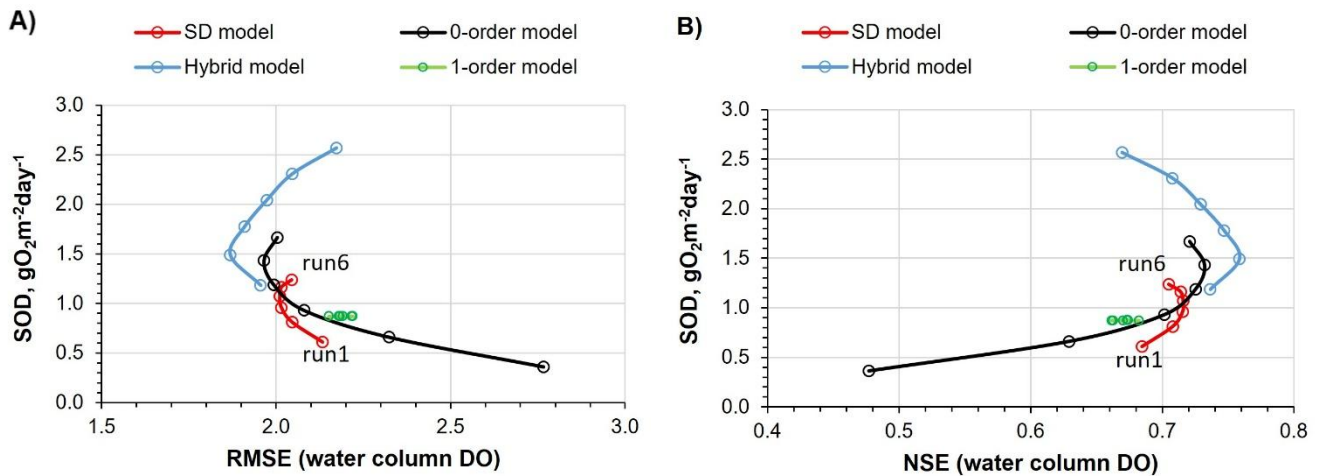


Figure 11. (A) RMSE between observed and simulated DO profiles in the water column for all models: the SD model (Runs 1–6), the Zero-order model, the Hybrid model with six SOD values ranging from 0.5 to 3.0 g O₂/m²day, and the First-order model with initial sediment organic matter concentrations from 0.0 to 3.0 g m⁻². (B) Same as (A), but using the Nash–Sutcliffe Efficiency (NSE) as the performance metric.

PAGE 29 LINE 476 to PAGE 32 LINE 595

4 Discussion

Overall, the temperature and DO predictions for the reservoir boundary conditions (Tâmega river) were quite good: PBIAS: 0.76% and 0.92%, respectively. When a significant number of samples and forcing variables are available the accuracy of machine learning algorithms can be greatly enhanced. This was demonstrated in the studies by Lu et al. (2020), Rajesh and Rehana (2021), and Feigl et al. (2021), where the RMSE for river water temperature prediction reached 1.04°C, 1.03°C, and 0.58°C, respectively. The results obtained for alkalinity, conductivity and TSS were also good: Alkalinity-PBIAS: 17.44%; Conductivity - PBIAS: 8.23%; TSS - PBIAS: 11.86%. However, as expected, the PBIAS values obtained for the remaining constituents were not as favorable (Total P- PBIAS: 7.11%; N-NOX- PBIAS: 3.92%; BOD5- PBIAS: 6.93%; Chla- PBIAS: 30%). The modeling of these constituents involves complex

biological, chemical, and physical processes that are harder to model accurately. However, except for Chla, the PBIAS values were generally less than 10%, reflecting acceptable levels of bias. Ammonium (N-NH₄) was the only parameter for which performance was significantly lower, generating a PBIAS of 28.27%. Moriasi et al. (2015) suggest that $\pm 10 \leq \text{PBIAS} \leq \pm 25$ is indicative of a satisfactory model performance.

Based on the RMSE, the overall reservoir calibration results obtained for all constituents with all models for the 2016-2021 period were consistent with the results seen in other studies (see Table A11). The mean RMSE values for Chl-a obtained with all models (SD model (run 5 - baseline): 17.72 µg/L; Zero-order model (zero-order SOD = 2.5 g O₂/m² day - baseline): 17.78 µg/L; First-order model (ISC= 0.5 g/m² - baseline): 14.88 µg/L and the Hybrid model (zero-order SOD=1.0 g O₂/m² day - baseline): 14.88 µg/L) are aligned with the results of other modeling studies (Brito et al., 2018: 62.9 µg/L; Kim et al., 2019: 6.7 to 13.2 µg/L; Tasnim et al., 2021: 0.6 to 27.6 µg/L; Almeida et al., 2023: 19.36 to 25.57 µg/L). For TP, the mean RMSE values were 0.03 mg/L for both the SD model (Run 5 – baseline) and the First-order model (ISC = 0.5 g/m² – baseline), while the Hybrid model (zero-order SOD = 1.0 g O₂/m²day – baseline) showed a slightly higher value of 0.04 mg/L. These results fall within the range reported in previous studies, including Brett et al. (2016) at 0.012 mg/L, Kim et al. (2019) between 0.014 and 0.068 mg/L, Tasnim et al. (2021) from 0.005 to 0.036 mg/L, and Almeida et al. (2023) ranging from 0.07 to 0.09 mg/L. The only exception was the Zero-order model (SOD = 2.5 g O₂/m²day – baseline), which overestimated phosphorus export from sediments during the summer months (July to September) of 2018 to 2021, resulting in a notably higher RMSE of 0.1 mg/L. Even with a very low phosphorus release rate from the sediments—representing a fraction of the SOD (0.001)—the Zero-order model still overestimated phosphorus concentrations, particularly during periods of elevated sediment oxygen demand. This suggests that the model may lack the sensitivity needed to accurately simulate low-level sediment-phosphorus interactions under such conditions. The mean RMSE values obtained for TN were lower than the only reference value available in the literature—0.77 mg/L reported by Deliman et al. (2002). Specifically, the SD model (Run 5 – baseline) yielded an RMSE of 0.33 mg/L, the First-order model (ISC = 0.5 g/m² – baseline) produced 0.36 mg/L, and the Hybrid model (zero-order SOD = 1.0 g O₂/m²day – baseline) resulted in 0.35 mg/L. The only exception was the Zero-order model (SOD = 2.5 g O₂/m²day – baseline), which had a significantly higher RMSE of 0.79 mg/L—slightly exceeding the value reported by Deliman et al., yet still within a comparable range. The RMSE obtained with the SD model (Run 5 - baseline), Zero-order model (zero-order SOD = 2.5 g O₂/m²/day - baseline); First-order model (ISC= 0.5 g/m² - baseline) and the Hybrid model (zero order SOD= 1.0 g O₂/m²/day - baseline) for DO, 2.01 mg/L, 1.97 mg/L, 2.15 mg/L and 1.87 mg/L respectively) are also in line with the results obtained in other studies (e.g., Deliman et al., 2002: 1.34 mg/L; Brett et al., 2016: 1.2 mg/L; Brito et al., 2018: 7.6 mg/L; Luo et al., 2018: 1.78 mg/L; Tasnim et al., 2021: 2.33 mg/L). In the SD model (Run 5 – baseline), bottom-layer SOD values ranged from 0.015 to 5.152 g O₂/m²day ($\mu = 1.162$; $\sigma = 0.823$), reflecting moderate variability driven by seasonal biogeochemical processes. In comparison, the Zero-order model (SOD = 2.5 g O₂/m²day - baseline) showed a broader but more temperature-driven range, from 0.000 to 15.640 g O₂/m²day ($\mu = 1.432$; $\sigma = 2.122$). The First-order model (ISC = 0.5 g/m² - baseline) yielded values between 0.000 and 20.000 g O₂/m²day, with a much lower mean ($\mu = 0.870$) and relatively high variability ($\sigma = 1.920$), consistent with its sensitivity to organic matter loading. The Hybrid model

(zero-order SOD = 1.0 g O₂/m²day - baseline) incorporated both zero- and first-order processes and produced the widest overall range, from 0.000 to 21.938 g O₂/m²day ($\mu = 1.491$; $\sigma = 2.024$), highlighting its enhanced responsiveness to both physical (e.g., temperature) and biogeochemical (e.g., organic matter) drivers. The monthly variation in SOD across the four models reveals distinct seasonal patterns influenced by their underlying formulations (Fig. A2). All models show notable peaks in May and October, corresponding to periods of elevated organic matter inflow, while a consistent decline is observed during the summer months (June to August), when external organic inputs are comparatively low. The Zero-order model (baseline SOD = 2.5 g O₂/m²·day) exhibits a sharp rise from winter to a peak of 1.919 g O₂/m²·day in May, then gradually declines over the summer, before increasing again in October (1.910 g O₂/m²·day). A similar double-peak pattern is observed in the Hybrid model (zero-order SOD = 1.0 g O₂/m²·day, baseline), with SOD reaching 1.715 g O₂/m²·day in May and a more pronounced maximum of 2.338 g O₂/m²·day in October, reflecting the combined effects of temperature and organic matter availability. The SD model (Run 5 – baseline) shows more moderate seasonal variation, with values dipping to 0.679 g O₂/m²·day in August, then rising to 1.501 g O₂/m²·day in November, consistent with internal sediment dynamics. The First-order model (ISC = 0.5 g/m², baseline), which is most sensitive to organic matter loading, also mirrors this seasonal structure, peaking in October (1.235 g O₂/m²·day) after a gradual summer decline. Collectively, these patterns underscore the importance of organic matter availability—particularly in spring and autumn—as a key driver of SOD across the different modeling approaches. This pattern indicates the model's responsiveness to both organic matter inputs and temperature, leading to a more nuanced representation of seasonal variation compared to the other models. These values are consistent with the SOD values obtained in other studies, such as those of Schnoor and Fruh (1979), which concluded that the SOD values of Lake Lydon B. Johnson (located in the U.S.) ranged from 1.7 to 5.8 g O₂/m²·day, and of Beutel (2015), which measured SOD values in different locations around Lake Hodges (located in the U.S.) ranging from 0.6 to 2.3 g O₂/m²·day. It would be useful to be able to compare these results with SOD values measured at different sites within the Torrão reservoir.

It is important to emphasize that this study was primarily designed to evaluate the performance of the sediment diagenesis model. However, by incorporating alternative SOD modeling approaches, it inevitably allowed for a comparative ranking of model performance, highlighting the relative strengths and limitations of each formulation. The performance limitations of the Zero-order and First-order models can be attributed to their structural simplifications. Specifically, the Zero-order model's strong temperature dependence, coupled with its disregard for the dynamics of organic matter loading, reduces its ability to capture temporal variability driven by external inputs. Similarly, the lower accuracy of the First-order model likely stems from its exclusion of anaerobic decay processes and limited representation of sediment biogeochemistry, which becomes especially relevant under low-oxygen conditions. The Hybrid model outperformed all other approaches. Considering the principle of parsimony (Occam's razor) (Burnham and Henderson, 2002), the simpler Hybrid model proved more effective than the complex SD model, making it the preferred choice for simulating SOD dynamics in the reservoir. These findings underscore the importance of selecting models that align with the specific characteristics of the system being studied. Simpler models, such as the Hybrid model, may be adequate for steady-state conditions, short- to medium-term forecasts, or scenarios with limited data. The zero-order SOD component of the Hybrid model relies solely on temperature

and is decoupled from the water column; therefore, in long-term simulations, this limitation can gradually undermine the model's accuracy. In contrast, the SD model may be more appropriate when the goal is to explore system-wide feedbacks and temporal dynamics over extended periods—especially those involving sediment accumulation and nutrient cycling—where it may provide valuable insight into underlying processes, provided that sufficient observational data become available to support its additional state variables. Moreover, a model's effectiveness heavily depends on the user's familiarity with its structure and their skill in calibration. Yet, it is unrealistic to expect researchers to master the implementation of every available modeling approach. As such, comparisons between models should be interpreted carefully, acknowledging the influence of user expertise on performance outcomes (Piccolroaz et al. 2024). Overall, to strengthen the analysis, it is recommended that users apply all available SOD modeling approaches in the case of the CE-QUAL-W2 model and assess the model's behavior. This comprehensive evaluation provides a solid foundation for further modeling efforts and helps ensure that the chosen approach is well-suited to the system's specific conditions and objectives.

The results also revealed that the particulate fraction of organic carbon in the reservoir sediments corresponded to 80% of the TOC. This value is small compared to the results obtained for Taihu Lake by Yu et al. (2022), where the ratio of POP to TOC varied from 97.85% to 89.53%. However, this value (80%) was obtained indirectly through the analysis of the reservoir's predicted SOD values as a function of different initial POC values and may, therefore, reflect other sources of uncertainty, such as inflow organic matter characterization. Given the fact that the magnitude of TOC in the sediment can be affected by numerous factors, including water column productivity, terrestrial inputs of organic materials, sediment properties, and microbial activity rates (Gireeshkumar et al., 2013), and that, partly due to differences in reservoir productivity and morphology, the spatial distribution and sources of organic carbon vary greatly across regions (Anderson et al., 2009), it is reasonable to assume that the only way to accurately assess the POC prediction is by monitoring the reservoir POC content. Furthermore, this study has highlighted the need to expand research to additional waterbodies across diverse regions to improve our understanding of the CE-QUAL-W2 diagenesis model's performance under varying environmental conditions. This includes evaluating its applicability in long-term scenarios, which are essential for capturing cumulative sediment dynamics and climate-driven trends. Additional SOD monitoring studies need to be conducted in lakes and reservoirs and extended to other latitudes, with particular focus on the chemical characterization of sediments and the definition of sediment burial rates.

5 Conclusions

This research evaluates the performance of the CE-QUAL-W2 v4.5 sediment diagenesis model in simulating water temperature, dissolved oxygen, total phosphorus, total nitrogen, chlorophyll-a, and biochemical oxygen demand in a Portuguese reservoir over the period from 2016 to 2021. Calibration was based on 35 sets of observed temperature and dissolved oxygen profiles, supplemented by six annual measurements of total nitrogen, total phosphorus, chlorophyll-a, and biochemical oxygen demand collected at various depths. To evaluate model accuracy, three alternative sediment oxygen demand formulations — a Zero-order model, a First-order model, and a Hybrid approach combining features of both — were also applied and compared. The Hybrid model consistently outperformed the other formulations, striking an effective balance between accuracy and simplicity. It

therefore represents the most suitable choice for modeling the reservoir. In contrast, the Zero- and First-order models exhibited limitations related to temperature dependence and inadequate sediment process representation, respectively. Simpler models, such as the Hybrid model, may be adequate for steady-state conditions, short- to medium-term forecasts, or scenarios with limited data. In contrast, the SD model — despite its good performance — may be more appropriate when the goal is to explore system-wide feedbacks and temporal dynamics over extended periods, especially in cases involving sediment accumulation and nutrient cycling. In such contexts, it may offer valuable insights, provided that sufficient observational data are available to support its additional state variables. Overall, the study reinforces the importance of choosing models based on site characteristics, available data, and simulation goals. Future work should broaden the evaluation of these models across various waterbodies and extended timeframes, while highlighting the need for enhanced sediment monitoring to support detailed process-based modelling.

Appendix A

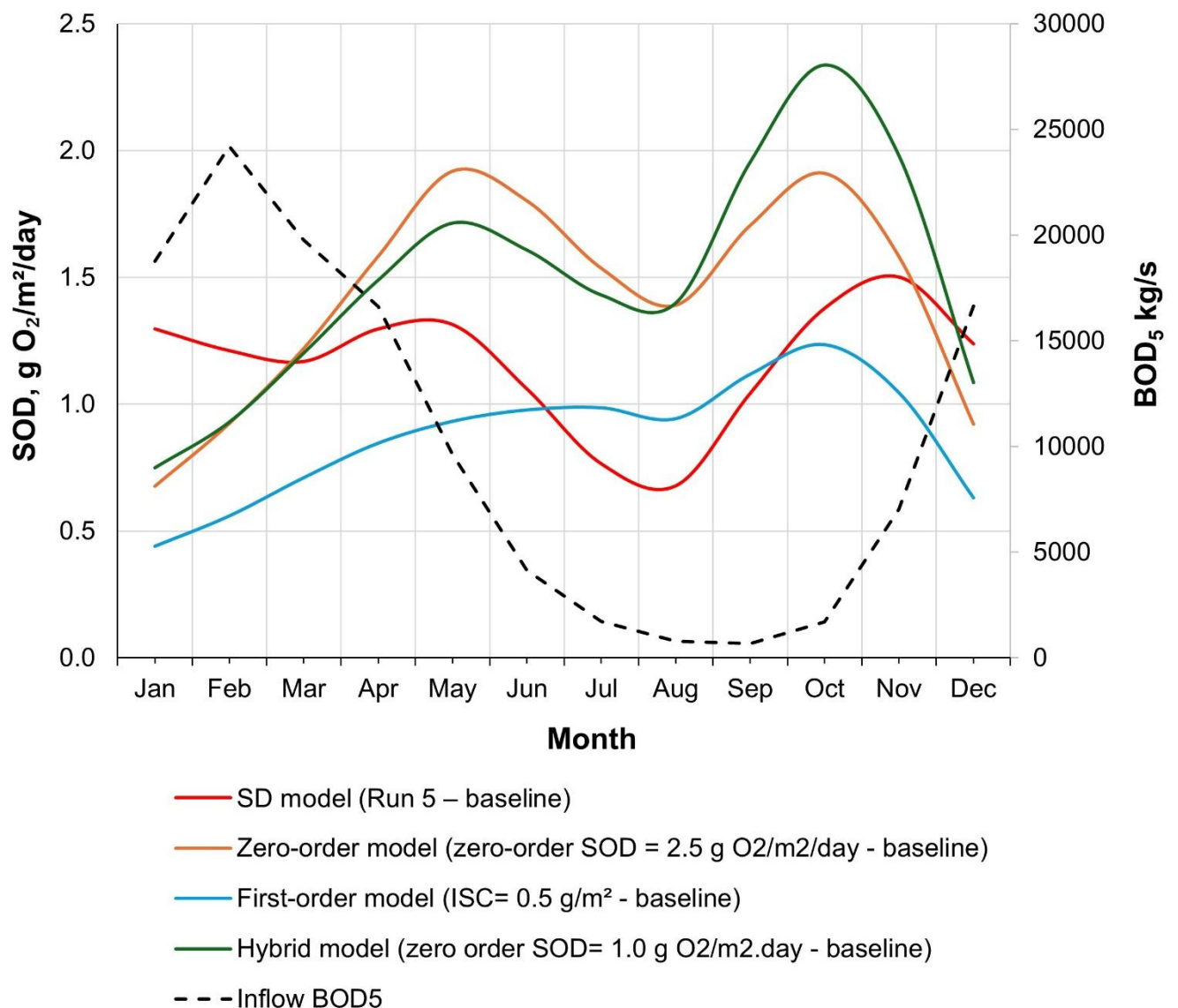


Figure A2: Average monthly sediment oxygen demand (SOD) at the reservoir bottom layer predicted by the SD model (Run 5 - baseline), Zero-order model (baseline: SOD = 2.5 g O₂/m²day), First-order model (baseline: ISC = 0.5 g/m²), and Hybrid model (baseline: zero-order SOD = 1.0 g O₂/m²day). Also shown is the inflow BOD₅ load from the reservoir's main branch.

PAGE 7 LINE 154-163

“To assess the sensitivity of each model to reductions in external organic matter (OM) and phosphorus (PO₄-P) inputs, two separate scenario analyses were conducted. The first scenario involved an 80% reduction in OM inflow load, while the second applied an 80% reduction in both OM and PO₄-P inflow loads. These reductions were implemented specifically in the main reservoir branch (Branch 1 – Tâmega River), where the majority of nutrient and organic inputs occur. Each sediment model—SD, Zero-order, First-order, and Hybrid—was run under baseline conditions and under both reduction scenarios. The impact on DO dynamics was evaluated using time series of depth- and segment-averaged DO concentrations. Each model—SD, Zero-order, First-order, and Hybrid—was run under baseline conditions and then under this reduced-loading scenario. The evaluation of model performance, along with the results of the sensitivity analysis, provided deeper insights into simulating SOD dynamics using the sediment diagenesis approach in comparison to the other SOD formulations.

PAGE 26 LINE 440-453

“3.4 Inflow Organic Matter and Phosphorus Load Reduction Scenarios

The results reveal clear differences in model sensitivity to inflow load reductions, with the First-order and Hybrid models exhibiting a stronger response compared to the SD and Zero-order models (Figures 12 and 13). The SD model showed minimal change, indicating limited sensitivity to external loading (Figures 12a and 13a), likely due to strong internal loading feedback from legacy phosphorus and organic matter stored in sediments. The Zero-order model demonstrated limited utility for management scenarios because it is decoupled from the water column, reducing its responsiveness to external changes. The First-order model may overestimate sensitivity as it tends to underestimate internal loading contributions. The Hybrid model, which combines both approaches, is less reactive than the First-order model due to the influence of the Zero-order component, offering a more balanced response. However, the Zero-order SOD component in the Hybrid model depends solely on temperature and remains decoupled from water column conditions; this limitation may gradually reduce the model's accuracy in long-term simulations. These differences in model sensitivity are further reflected in the evolution of average SOD across scenarios (Table 5). While the Zero-order and SD models show virtually no change in bottom-layer SOD under reduced loading conditions, the First-order and Hybrid models register clear declines. The First-order model's SOD drops from 0.87 g O₂/m²day to 0.42 g O₂/m²day (80% OM reduction) and 0.29 g O₂/m²day (80% OM and P reduction) and the Hybrid model from 1.49 g O₂/m²day to 1.07 g O₂/m²day (80% OM reduction) and 0.94 g O₂/m²day (80% OM and P reduction).”

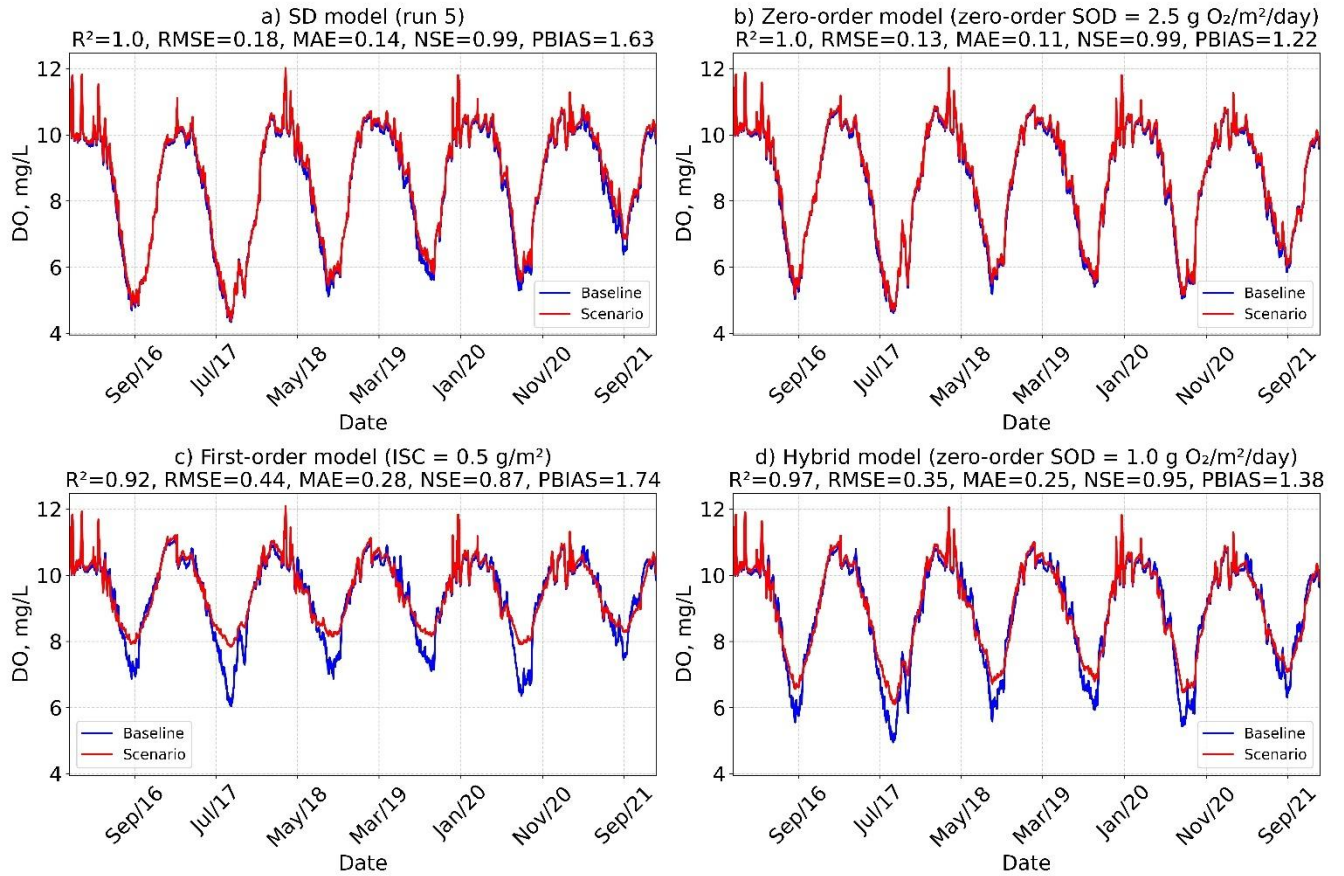


Figure 12. Time series of DO, averaged across all model layers and segments, for each baseline model scenario: SD model (Run 5), Zero-order model (SOD = 2.5 g O₂/m²/day), First-order model (initial sediment concentration = 0.5 g/m²), and Hybrid model (Zero-order SOD = 1.0 g O₂/m²/day). The figure compares baseline conditions with an 80% reduction in organic matter inflow load in the main reservoir branch (Branch 1 – Tâmega River). Performance metrics (R^2 , RMSE, MAE, NSE, and PBIAS) are also shown for each case.

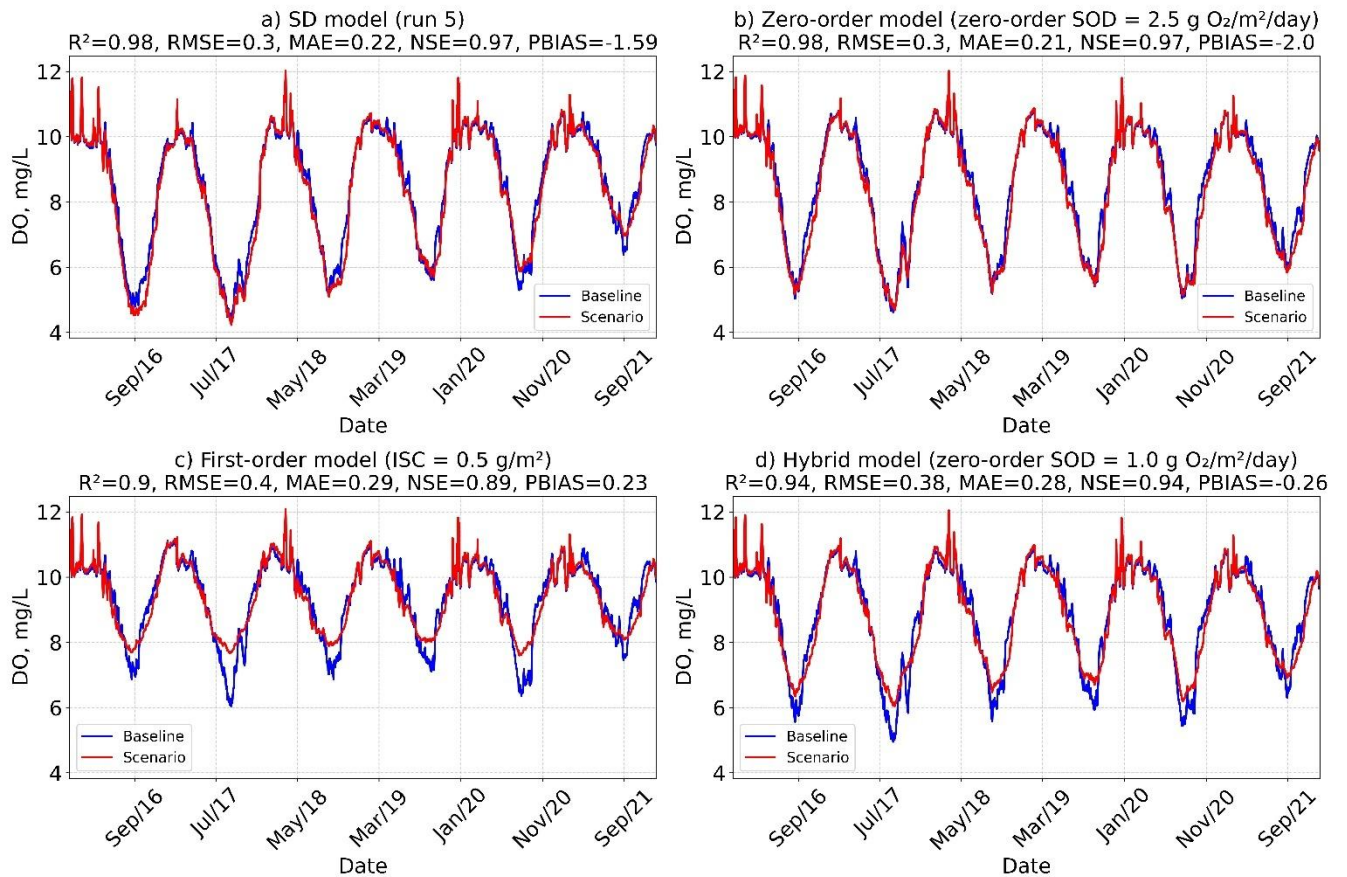


Figure 13. Time series of DO, averaged across all model layers and segments, for each baseline model scenario: SD model (Run 5), Zero-order model (SOD = $2.5 \text{ g O}_2/\text{m}^2/\text{day}$), First-order model (initial sediment concentration = $0.5 \text{ g}/\text{m}^2$), and Hybrid model (Zero-order SOD = $1.0 \text{ g O}_2/\text{m}^2/\text{day}$). The figure compares baseline conditions with an 80% reduction in organic matter and P-PO₄ inflow loads in the main reservoir branch (Branch 1 – Tâmega River). Performance metrics (R^2 , RMSE, MAE, NSE, and PBIAS) are also shown for each case.

Table 5. Average sediment oxygen demand (SOD) in the bottom layers of the reservoir, calculated across all segments, for each model under three scenarios: Reference (baseline conditions), 80% reduction in organic matter inflow (OM 80%), and combined 80% reduction in organic matter and phosphorus inflow (OM and P%) in the in the main reservoir branch (Branch 1 – Tâmega River)

Scenario	SD model (Run 5)	Zero-order model (SOD = $2.5 \text{ g O}_2/\text{m}^2/\text{day}$)	First-order model (ISC = $0.5 \text{ g}/\text{m}^2$)	Hybrid model (Zero-order SOD = $1.0 \text{ g O}_2/\text{m}^2/\text{day}$)		
				Aggregate	Zero- order term	First- order term
Baseline	1.16 ± 0.82	1.43 ± 2.12	0.87 ± 1.19	1.49 ± 2.02	0.59 ± 0.85	0.90 ± 0.75
OM 80% reduction	1.13 ± 0.83	1.44 ± 2.12	0.42 ± 0.61	1.07 ± 1.43	0.61 ± 0.85	0.46 ± 0.75
OM and P% 80% reduction	1.13 ± 0.82	1.46 ± 2.11	0.29 ± 0.37	0.94 ± 1.19	0.64 ± 0.84	0.30 ± 0.42

References:

Almeida, M., and Coelho, P.: Evaluating the performance of CE-QUAL-W2 version 4.5 sediment diagenesis model (Manuscript related material: input data and source code) (1.0.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.14606105>, 2025.

Almeida, M., Coelho, P.: A first assessment of ERA5 and ERA5-Land reanalysis air temperature in Portugal, *International Journal of Climatology*, 43(14), 6643–6663, doi.org/10.1002/JOC.8225, 2023.

Barbosa, S., Scotto, M.G.: Extreme heat events in the Iberia Peninsula from extreme value mixture modeling of ERA5-Land air temperature, *Weather. Clim. Extrem.*, 36, 100448, doi.org/10.1016/j.wace.2022.100448, 2022.