Responses to the Evaluation of Reviewer #3

General Comments

This manuscript presents a physics-informed data-driven framework for correcting structural errors in groundwater reactive transport models. The topic is timely and relevant, and the proposed framework has the potential to improve predictive accuracy in groundwater models. I recommend publication after the following points are addressed.

Response: We thank the reviewer for the positive evaluation and insightful comments to this manuscript, and we have carefully addressed these comments in below responses and revised the manuscript.

Specific Comments

Comment 1: (Section 2.3) The authors introduced a dynamic weight-updating strategy to improve parameter identification and prediction accuracy when coupling multiple physical constraints. However, this is somewhat expected because additional physical constraints introduce more information into the inverse problem. Would similar improvements be obtained if fixed weights were assigned to multiple constraints? A comparison between constant-weight likelihood formulation may better demonstrate the necessity of the proposed adaptive weighting scheme.

Response: We thank the reviewer for this important suggestion. We agree that it is necessary to compare the simulation results obtained using the dynamic and fixed weighting strategies. Following this suggestion, we have added the simulation results using the fixed weighting strategy when multiple physical constraints are simultaneously coupled (Figures S5–S8), together with the corresponding performance metrics (Tables S5 and S6), and discussed these results in the revised manuscript (Line 655).

Compared with the dynamic weight-updating strategy, the fixed weighting strategy results in relatively large carbon and chlorine mass conservation errors in Case Study 1 (Figure S5), and substantially poorer predictive performance for all four chlorinated hydrocarbons (Figure S6). Similarly, in Case Study 2, the fixed weighting strategy results in a substantially larger Cd mass conservation error (Figure S7) and greater bias in the predicted Cd concentrations (Figure S8). These results demonstrate that although multiple physical constraints can provide additional information, appropriately balancing the contributions of different physical constraints is critical for improving predictive performance and physical consistency, further supporting the effectiveness and necessity of the dynamic weight-updating strategy.

Line 655: In contrast, the proposed dynamic weight-updating strategy is well suited to combined likelihood functions that integrate multiple sub-likelihoods associated with distinct physical mechanisms, resulting in substantially improved predictive accuracy and physical consistency compared with the fixed weighting strategy (Figure S5-S8 in

Supporting Information).

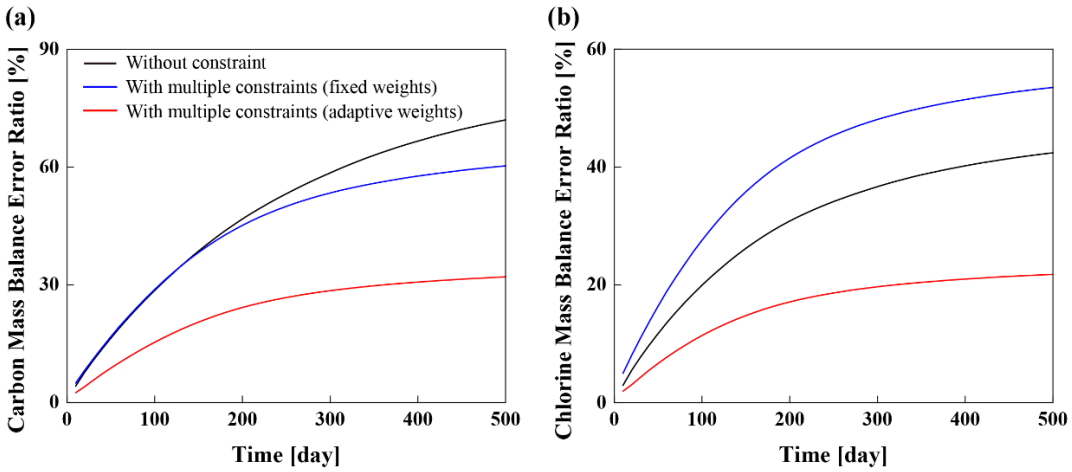

Figure S5. Variation in the carbon/chlorine mass balance error rate for the synthetic groundwater PCE reactive transport model under the multiple-constraint scenario with fixed weights.

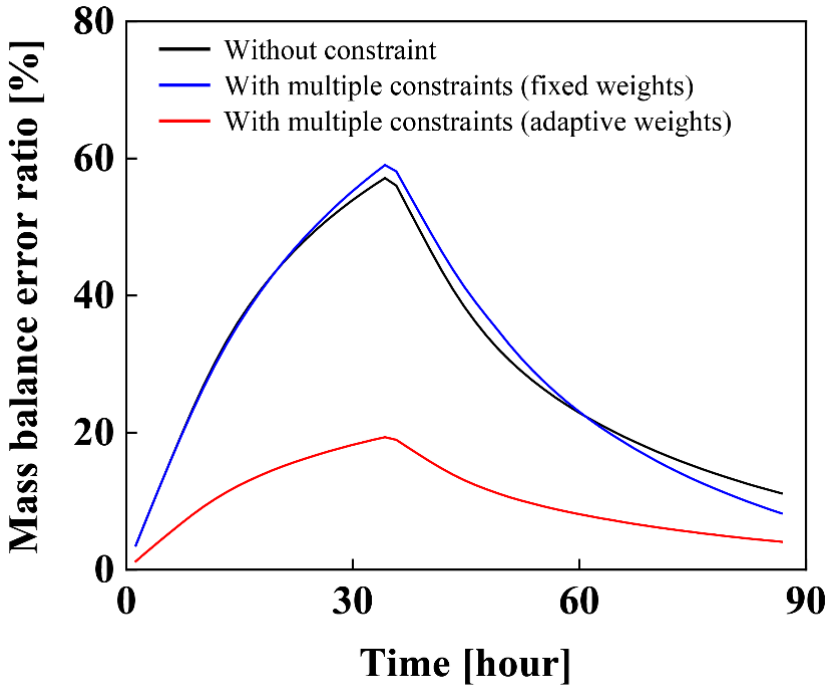

Figure S6. Variation in the cadmium mass balance error rate for the sand column Cd-phosphate cotransport model under the multiple-constraint scenario with fixed weights.

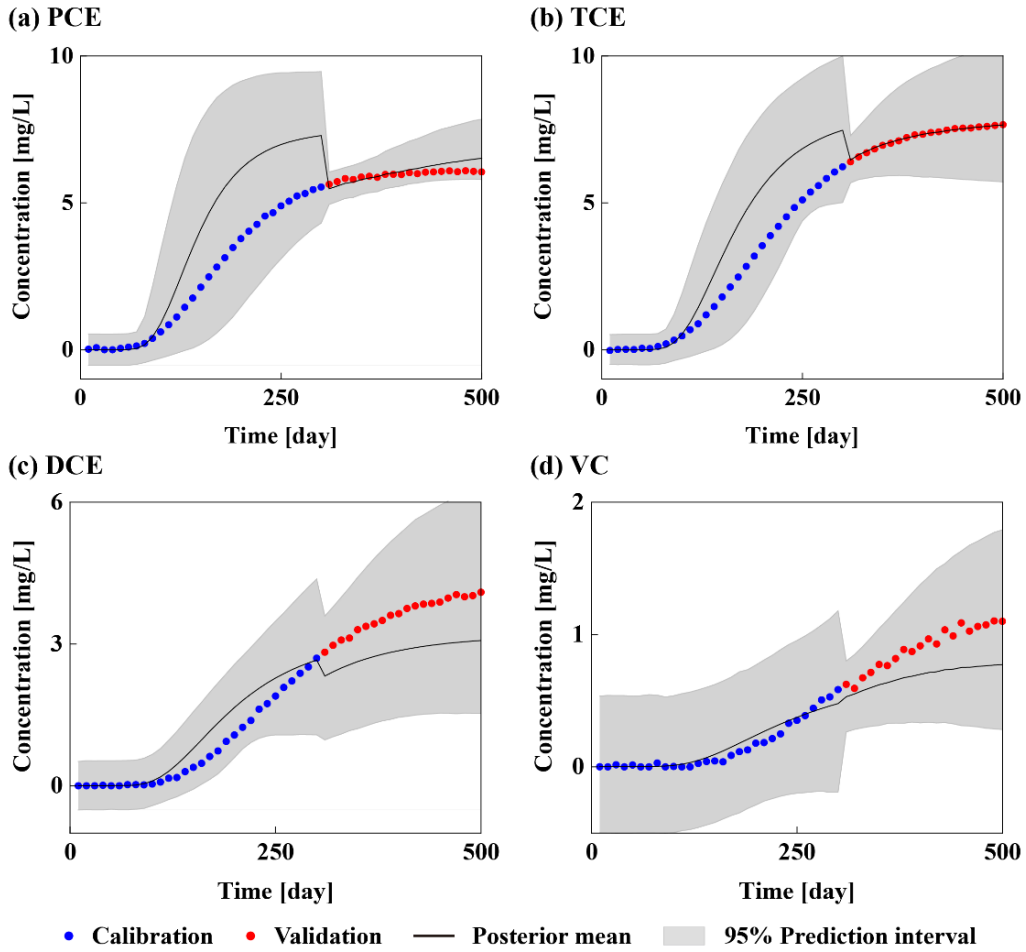

Figure S7. DDM predictions of chlorinated ethenes concentrations under the multiple-constraint scenario with fixed weights.

Table S5. Evaluation of predictions for the synthetic groundwater PCE reactive transport model under the multiple-constraint scenario with fixed weights.

Constraints	Calibration/ Validation	Species	NSE	MAE	RMSE
With multiple physical constraints (fixed weights)	Calibration	PCE	0.3562	1.3212	1.6362
		TCE	0.6518	1.0199	1.2937
		DCE	0.8649	0.2452	0.3297
		VC	0.9274	0.0381	0.0492
	Validation	PCE	-2.1390	0.1848	0.2252
		TCE	0.9921	0.0253	0.0328
		DCE	-3.5710	0.7924	0.8078
		VC	-1.0849	0.2150	0.2324

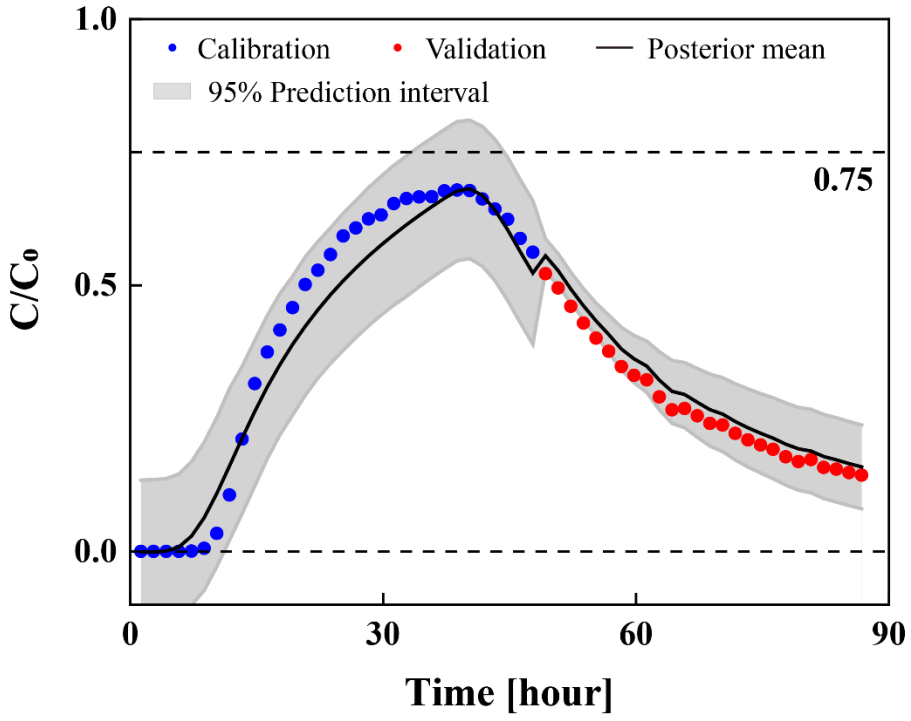

Figure S8. DDM predictions of cadmium concentration under the multiple-constraint scenario with fixed weights.

Table S6. Evaluation of predictions for the sand column Cd-phosphate cotransport model under the multiple-constraint scenario with fixed weights.

Constraints	Calibration/ Validation	NSE	MAE	RMSE
With multiple physical constraints (fixed weights)	Calibration	0.9112	0.0408	0.0499

Comment 2: (Section 4.1) The physical constraints considered in the synthetic case are based on carbon and chlorine mass conservation, which can be accurately quantified in the example. In practical applications, however, physical constraints are often subject to uncertainty, e.g., boundary fluxes or source terms. Can the authors discuss how the proposed framework would handle situations when the physical constraints themselves are uncertain?

Response: We thank the reviewer for this important suggestion. We agree that physical constraints are often difficult to define appropriately in real-world groundwater systems. In the proposed framework, the confidence on the accuracy of each physical constraint is represented by the constraint variance of each physical constraint (σ_c^2). A smaller σ_c^2 indicates lower uncertainty in the physical constraint and therefore imposes a stricter constraint. In addition, the physical constraints incorporated into the proposed framework must be expressed in a quantifiable mathematical form, such as equality or

inequality constraints.

Comment 3: (Section 5.1) It is expected that the posterior distributions contract compared to the case with single constraint or without physical constraint. I suggest plotting the reference distributions in Figures 4 and 5. This would allow readers to assess whether they converge toward the correct parameter values.

Response: We thank the reviewer for this important suggestion. We consider the reviewer's suggestion to verify whether the true parameter values fall within the corresponding posterior intervals to be highly valuable. However, the present study adopts an external correction approach for model structural error rather than an internal correction approach. Therefore, the identified posterior distributions of the parameters should be compared with the true parameter values of the simplified model.

In the first case study, the preferential flow regions differ in shape between the true and simplified models. Consequently, the parameter values of the true model (hydraulic conductivity and dispersivity) cannot be regarded as the true values of the corresponding parameters in the simplified model, and the true parameter values of the simplified model are therefore unavailable. The second case study is a real case based on a laboratory sand column experiment, in which the physical parameters to be identified (the proportion coefficient of equilibrium sorption f, the Freundlich isotherm coefficient k_d and the isotherm shape parameter β) cannot be directly measured. In addition, the GPR hyperparameters in both case studies are statistical parameters used to characterize the structural error and lack direct physical meaning. Therefore, their posterior distributions cannot be validated against true values.

It is worth noting that, in the second case study, if hydraulic conductivity were selected as an unknown parameter, its experimentally measured value could be used as the true value to validate the posterior distribution of the identified parameter.

Comment 4: (Section 5.4.1) The proposed framework iteratively updates the weights of the sub-likelihood functions. I have two questions regarding this procedure: (1) In case 1, the sub-likelihood values differ by several orders of magnitude. Does this imply that certain physical constraints are inherently more informative than the others? (2) Since one of the motivations of the framework is its application to practical problems, the computational cost of the adaptive weighting procedure needs to be addressed.

Response: We thank the reviewer for raising this important issue.

(1) We consider that differences in the orders of magnitude among the sub-likelihood functions do not necessarily indicate that certain physical constraints contain more information. Rather, these differences mainly arise from the different formulations of the physical constraints. Because different physical constraints are constructed based on different physical quantities and physical relationships, their corresponding constraint error terms generally differ in numerical scale and range of variation, resulting in different orders of magnitude among the sub-likelihood functions. For example, in Case Study 1, carbon and chlorine account for different mass fractions

in the four chlorinated hydrocarbons. Consequently, the mass conservation error terms for carbon and chlorine differ in numerical scales, leading to different orders of magnitude in their corresponding sub-likelihood functions. This further highlights the importance of the proposed dynamic weight-updating strategy.

(2) Following the reviewer's suggestion, we have added a discussion of the computational cost of the proposed method in the Discussion section (Line 659).

Line 659: However, the dynamic weight-updating strategy introduces additional computational overhead. Compared with the DDM without physical constraints or with fixed weights, the proposed method requires the weights of sub-likelihood functions to be dynamically calculated and updated during MCMC sampling, with the computational burden increases as more physical constraints are introduced.

Comment 5: (Section 5.4.3) This study does not compare the developed framework against other physics-informed machine learning approaches. As such, it is difficult to evaluate how much improvement comes from this model compared with existing approaches, such as PINNs. I understand implementing these methods is beyond the scope of this work; can the authors include a discussion comparing the proposed framework with existing physics-informed approaches?

Response: We thank the reviewer for this important suggestion. Following this suggestion, we have added a comparative discussion of the proposed method and PIML approaches in the Discussion section (Line 734).

Line 734: In addition, the proposed method primarily focuses on structural error correction and can adaptively correct structural error based on the discrepancy between physical model predictions and observations, providing a reasonable mechanism for structural error compensation. Unlike PIML approaches, which typically require the calculation of partial differential equation (PDE) residuals or modification of machine learning model architectures for specific problems, the proposed method can flexibly incorporate multiple physical constraints within a Bayesian framework. Moreover, it avoids the limitation of applying PIML to structural error correction, where physical constraints are imposed only on the error correction model rather than on the entire corrected prediction.