Responses to the Evaluation of Reviewer #1

General Comments

This study presents a Bayesian data-driven framework that incorporates multiple physical constraints to correct structural errors in groundwater contaminant transport models. The two case studies suggest that, compared with the unconstrained and single-constraint approaches, incorporating multiple constraints can improve parameter identifiability, reduce predictive uncertainty, and enhance physical consistency. However, several issues need to be clarified or further addressed before the manuscript can be considered for publication. Please see the comments below.

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: In Bayesian inference, MCMC sampling generally requires the target posterior distribution to remain fixed during the sampling process. In this study, however, the weights of the sub-likelihood functions are dynamically updated during sampling, which may cause the target posterior distribution to vary with iterations and thus affect the validity of the Bayesian inference. The authors are therefore encouraged to discuss whether the proposed dynamic weight-updating strategy affects MCMC convergence. Appropriate convergence diagnostics should also be provided.

Response: Thanks for pointing out this important issue. We apologize for not clearly describing the implementation of the dynamic weight-updating strategy during MCMC sampling. It should be noted that the dynamic weight-updating strategy is applied only during the burn-in period. During this period, the weights of the sub-likelihood functions are updated separately for each Markov chain based on its own evolution. Once the stopping criterion for weight updating is satisfied, the weights are fixed, and only the subsequent samples generated using these fixed weights are used for posterior inference. Therefore, the dynamic updating of the sub-likelihood weights does not alter the target posterior distribution used for posterior inference. We have added further details on the implementation of the dynamic weighting strategy in the revised manuscript (Line 275).

In addition, we employed the Gelman–Rubin diagnostic to assess the convergence of the MCMC chains. This diagnostic is widely used to evaluate MCMC convergence, with convergence considered to be achieved when the calculated R_{hat} values (Gelman–Rubin statistics) are below 1.2 (Kwon et al., 2025; Vats & Knudson, 2021; Roy, 2020). Figures S1 and S2 show the evolution of R_{hat} with the number of iterations for all parameters in the two case studies under multiple physical constraints. As shown in these figures, the R_{hat} values for all parameters decrease below 1.2 before the end of the burn-in period (iteration = 4000), indicating convergence of the MCMC sampling. In addition, we have also marked the iterations at which the weights become stable in Figures 10 and 11.

Line 275: Since multiple Markov chains are employed, the weights of the sub-likelihood functions are updated separately for each chain during the burn-in period based on its own evolution. Once the stopping criterion is satisfied, a common set of fixed sub-likelihood weights is used for subsequent sampling across all chains.

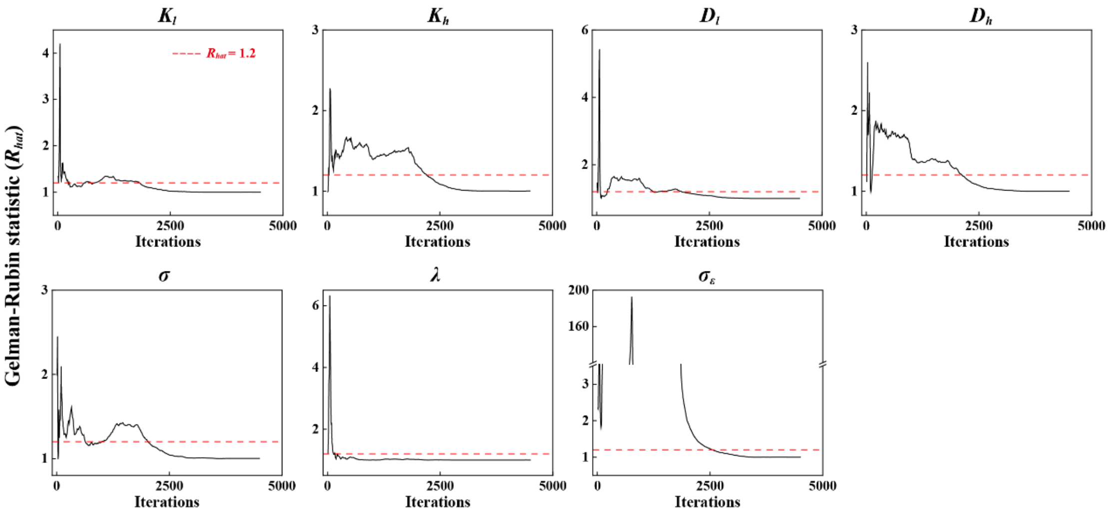

Figure S1. Gelman-Rubin diagnostics of Markov chains for the synthetic groundwater PCE reactive transport simulation under the multiple-constraint scenario.

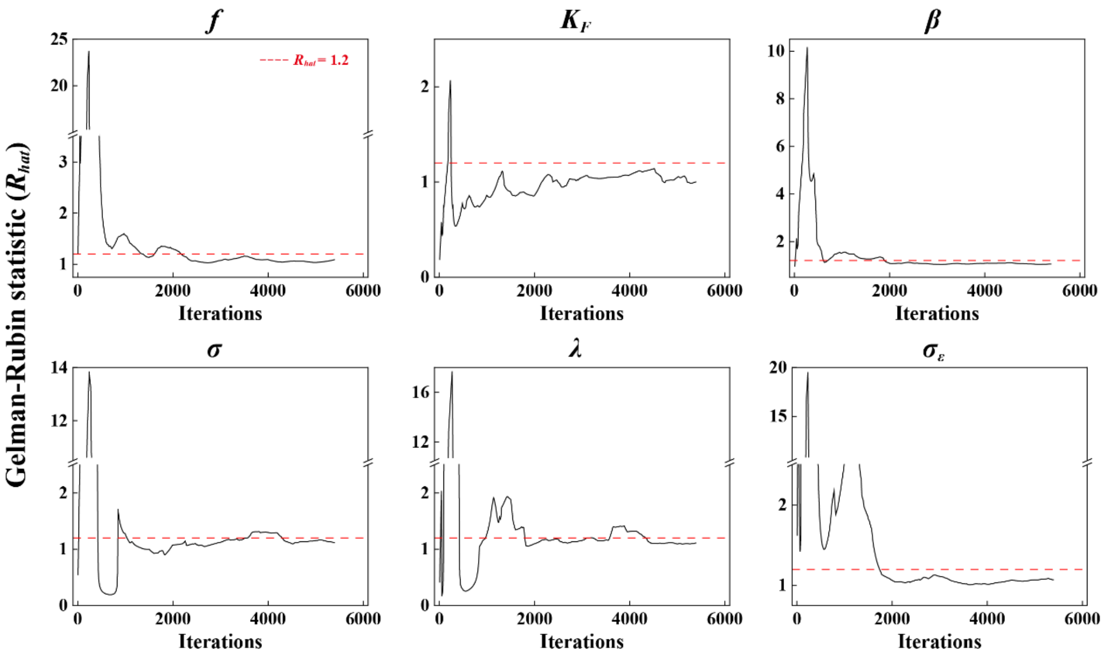

Figure S2. Gelman-Rubin diagnostics of Markov chains for the sand column Cd-phosphate cotransport simulation under the multiple-constraint scenario.

<u>**Comment 2**</u>: The physical constraint strictness, sliding window length, and minimum tolerance variance may directly influence the posterior parameter distributions and predictive results. The authors should provide a clearer justification for the selected

values.

Response: We thank the reviewer for pointing out this issue. Following this suggestion, we conducted a sensitivity analysis of the constraint variance, σ_c^2. For both case studies, we evaluated prediction performance and mass conservation errors under different constraint variances, i.e., $\sigma_c^2 = 0.01$, 0.1, 0.2, and 0.5 (Figures S3 and S4).

The results show that when σ_c^2 is relatively small or large, such as $\sigma_c^2 = 0.01$ or 0.5, the agreement between the predictions and observations deteriorates, resulting in relatively large RMSE values and mass conservation errors. When $\sigma_c^2 = 0.2$, both RMSE and mass conservation error are minimized. Therefore, $\sigma_c^2 = 0.2$ was adopted in this study, consistent with the findings of our previous study (Tian et al., 2026).

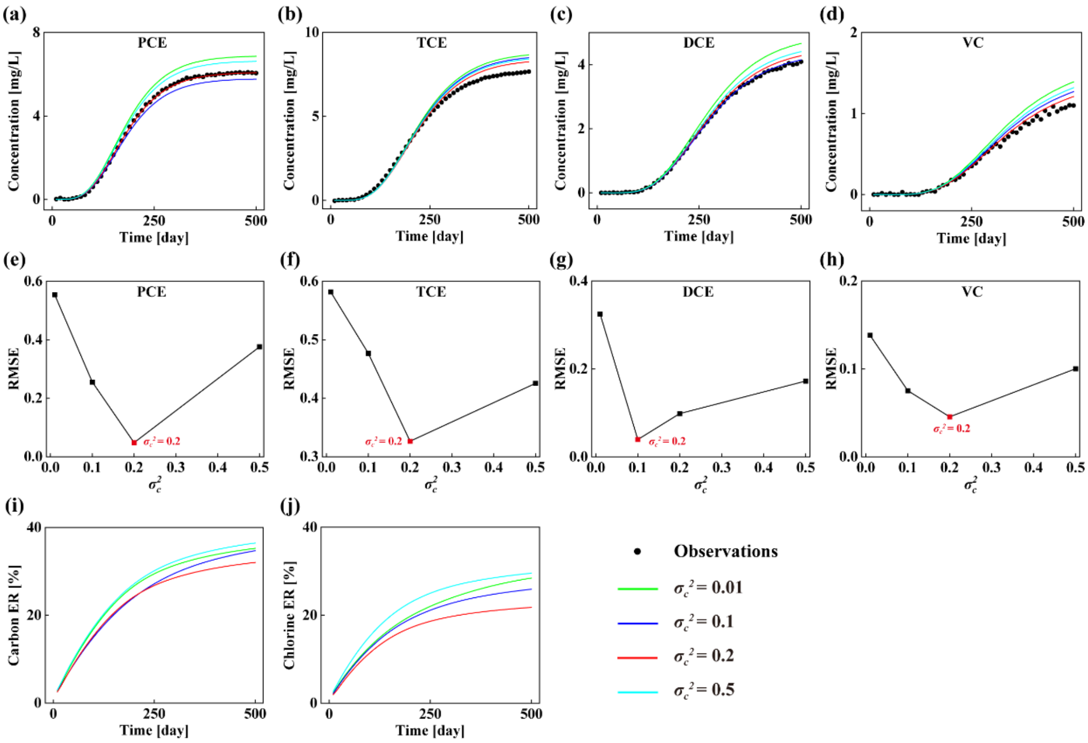

Figure S3. Predictions of synthetic groundwater PCE reactive transport simulation under different constraint variances (σ_c^2).

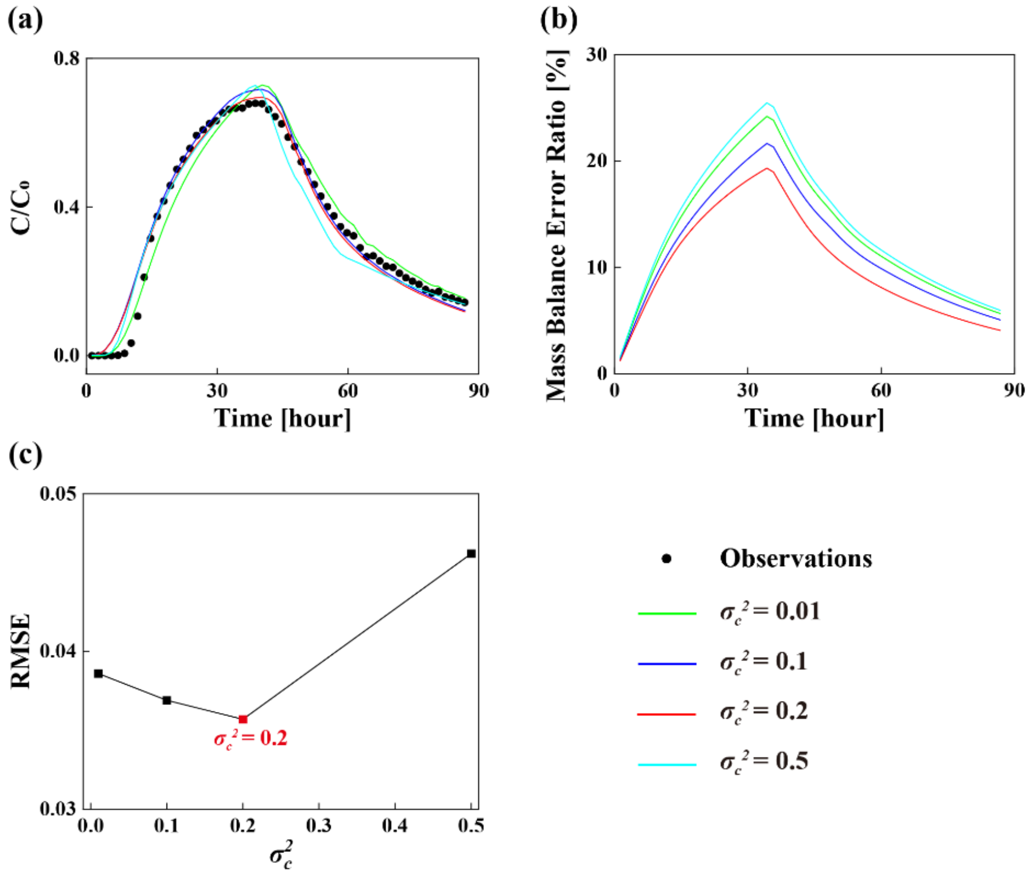

Figure S4. Predictions of sand column Cd-phosphate cotransport simulation under different constraint variances.

Comment 3: The manuscript states that incorporating multiple physical constraints makes the structural error simpler and easier to characterize using GPR. However, this conclusion is currently supported mainly by qualitative interpretation. Quantitative statistics of the structural-error term should be provided to support this claim.

Response: We greatly appreciate the reviewer's important suggestion and fully agree with it. We have added statistical information on the structural error terms for the two case studies in Tables S1 and S2, respectively, including the mean, standard deviation, range, and RMSE of the GPR predictions. We have also discussed these results in the revised manuscript (Line 712). The statistical results show that, after coupling multiple physical constraints, the standard deviation and range of the structural error, as well as the RMSE of the GPR predictions, are substantially reduced. These results indicate that coupling multiple physical constraints simplifies the structural error forms, making them easier for GPR to characterize.

Line 712: In contrast, when multiple physical constraints are coupled simultaneously, the posterior distributions of model parameters are further constrained, leading to substantially smaller and more stable structural error fluctuations, characterized by smaller standard deviation and range (Table S1 and S2 in Supporting Information).

Table S1. Quantitative statistics of structural errors and GPR prediction performance for the synthetic groundwater PCE reactive transport simulation under the three scenarios.

Constraints	Species	GPR prediction error (RMSE)	Mean of structural error	Standard deviation of structural error	Range of structural error
Without physical constraint	PCE	0.6653	-1.1629	0.5854	2.2019
	TCE	0.7688	0.0967	0.8372	2.4760
	DCE	0.3763	0.4085	0.5429	1.5274
	VC	0.1118	0.1455	0.1675	0.4985
With single physical constraint	PCE	0.3538	-0.9716	0.3812	1.4644
	TCE	0.3326	-0.0809	0.5454	1.6375
	DCE	0.2832	0.2941	0.4461	1.2877
	VC	0.0502	0.0992	0.1201	0.3857
With multiple physical constraints	PCE	0.1567	0.0744	0.1774	0.7876
	TCE	0.1444	0.2350	0.2224	0.8359
	DCE	0.0738	0.1262	0.1536	0.5336
	VC	0.0242	0.0248	0.0387	0.1399

Table S2. Quantitative statistics of l structural errors and GPR prediction performance for the sand column Cd-phosphate cotransport simulation under the three scenarios.

Constraints	GPR prediction error (RMSE)	Mean of structural error	Standard deviation of structural error	Range of structural error
Without physical constraint	0.0984	-0.0050	0.1061	0.4113
With single physical constraint	0.0426	-0.0010	0.0485	0.2104
With multiple physical constraints	0.0396	-0.0021	0.0380	0.1677

Comment 4: The posterior parameter distributions show that some parameters still exhibit pronounced multimodality under the multiple-constraint scenario. This suggests that parameter equifinality or limited parameter identifiability may still remain. The

authors should discuss the possible causes of this phenomenon and its implications for the reliability of parameter estimation.

Response: We thank the reviewer for this important suggestion. We agree that some parameters still exhibit multimodal posterior distributions even after coupling multiple physical constraints.

We consider this phenomenon to reflect the inherent parameter uncertainty in complex inverse problems. First, although multiple physical constraints provide additional information, they primarily constrain the corrected predictions rather than individual parameters directly. Therefore, different parameter combinations may still produce predictions that simultaneously satisfy the observational data and multiple physical constraints, and parameter uncertainty consequently remains. In addition, when no physical constraints or only a single physical constraint is coupled, the information provided by the observational data and physical constraints may be insufficient to adequately constrain parameter uncertainty, resulting in relatively broad posterior distributions. After multiple physical constraints are coupled, more effective information becomes available for identifying the parameter posterior distributions, allowing multiple local regions with high probability density to be more effectively identified and thereby resulting in more pronounced multimodal posterior distributions.

These results indicate that coupling multiple physical constraints does not increase parameter equifinality or uncertainty, but instead further improves parameter identifiability. Nevertheless, some parameter uncertainty remains because of insufficient direct data for parameter identification.

Comment 5: Increasing the number of physical constraints does not necessarily improve prediction, and additional constraints may be redundant or even conflict with existing constraints. The authors are encouraged to discuss the principles for selecting appropriate physical constraints.

Response: We are grateful to the reviewer for this important recommendation. We agree that increasing the number of physical constraints does not necessarily lead to continuous improvements in predictive performance, and that inappropriate constraints may even introduce redundant information or conflict with existing constraints. Following this suggestion, we have added a discussion on the principles for selecting physical constraints in the Discussion section (Line 749).

Line 749: In addition, the effectiveness of physical constraints depends on the information they provide rather than simply on their number. When multiple physical constraints are introduced, emphasis should not be placed solely on increasing their number. Instead, the complementarity and potential competition among different physical constraints should be carefully evaluated to prevent redundant or conflicting constraints from adversely affecting parameter identification, thereby ensuring the physical consistency of the predictions.

Comment 6: The proposed method is tested only using a laboratory sand column experiment and an idealized synthetic model. The authors should clarify in the Discussion the conditions under which the proposed method can be extended to real contaminated sites or watershed-scale applications, as well as its potential limitations.

<u>Response:</u> We appreciate the reviewer's insightful comment. We agree that the laboratory and synthetic case studies cannot fully represent the complexity of real-world conditions. Following this suggestion, we have added a discussion of the potential limitations and applicability of the proposed method to real contaminated sites and watershed-scale applications (Line 754).

Line 754: Notably, although the effectiveness and generalizability of the proposed method have been demonstrated through two case studies, some limitations remain when applying the method to real contaminated sites and watershed-scale applications. Effective application of the method requires sufficient and representative observational data, reliable identification of the sources of model structural error, and strong relationships between the introduced physical constraints and structural error.

<u>Comment 7</u>: Since the proposed method introduces additional sub-likelihood functions and dynamic weight updating, the computational cost should be discussed.

<u>Response:</u> We are grateful to the reviewer for highlighting this important concern. Following this 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.

<u>Comment 8</u>: The values of porosity and bulk density appear to be inconsistent between the text and the table in Case Study 2. Please check and clarify these values.

<u>Response:</u> We thank the reviewer for pointing out this issue. We apologize for the inconsistency in the reported porosity and bulk density values in Case Study 2. We have corrected the erroneous values (Line 352) and carefully checked the manuscript to ensure that all parameter values are now reported correctly.

Line 352: The column was uniformly packed with dry soil, and the average porosity (0.4) and bulk density (1.24 g/cm^3) were determined using the gravimetric method.

References

Kwon, S., Zhang, S., Koehn, H., & Zhang, B. (2025). MCMC stopping rules in latent variable modelling. *British Journal of Mathematical & Statistical Psychology*, *78*(1), 225–257. https://doi.org/10.1111/bmsp.12357

Roy, V. (2020). Convergence Diagnostics for Markov Chain Monte Carlo. In N. Reid, S. Stigler, & T. Louis (Eds.), *Annual Review of Statistics and its Application, Vol 7, 2020* (Vol. 7, pp. 387–412). https://doi.org/10.1146/annurev-statistics-031219-041300

Vats, D., & Knudson, C. (2021). Revisiting the Gelman-Rubin Diagnostic. *Statistical Science*, *36*(4), 518–529. https://doi.org/10.1214/20-STS812

Tian, J., Zeng, X., Wang, D., & Wu, J. (2026). A data-driven approach coupled with physical constraints to improve groundwater models with structural error. Water Resources Research, 62, e2025WR040247. https://doi.org/10.1029/2025WR040247