Response to the comments of Editor and Reviewers

Manuscript ID: hess-2024-315

Title: Enhancing Inverse Modeling in Groundwater Systems through Machine Learning: A Comprehensive Comparative Study

Dear Editor and Reviewers,

We sincerely appreciate your valuable feedback and the opportunity to revise our manuscript. We have carefully considered each comment and made significant revisions to enhance the methodological depth, case study design, noise robustness analysis, and practical relevance of our work. Below, we provide a detailed, point-by-point response to the comments.

Comment 1. A third case study with a different spatial correlation should be added.

Response:

We have included an additional case study focusing on a non-Gaussian random field. The details of this new case study are presented in Section 3.3.

Additionally, for the dimensionality reduction of the non-Gaussian random field, the Karhunen-Loève Expansion (KLE) method is no longer applicable. Hence, a new parameterization method based on the Octave Convolution Adversarial Autoencoder (OCAAE) has been introduced in Section 2.2.2.

The surrogate model for this case was constructed using deep residual network (ResNet) with 2000 training samples, and its prediction accuracy is discussed in the last paragraph of Section 4.1. ("*We further extended the ResNet for the surrogate model construction of both Gaussian and non-Gaussian random field scenarios, ...*")

The inversion results are presented in Section 4.2.3.

Comment 2. Noisy data should be used for all the tests, and I would appreciate very much an analysis of the stability of the methods, i.e., which is the behavior of the applied methods, for different noise magnitude.

Response:

In the second paragraph of the Section 3, we have supplemented the description of the noise settings for each scenario. Specifically: In the original two cases, we used 1% Gaussian noise, introduced by multiplying the standardized numerical simulation data with a random variable $\varepsilon \sim N(1, 0.01^2)$ to generate noisy observational data. For the newly added Case 3, we have also implemented a 1% noise scenario to ensure consistency.

Additionally, we fully agree with the editor and reviewer's suggestion to explore the impact of different noise magnitudes on the stability of inversion results. To address this, we specifically designed a 10% observational noise scenario for Case 3, as this case involves a high-dimensional parameter space and a more complex heterogeneous condition. Regarding low-dimensional parameter settings, we previously examined the effects of higher observational noise levels (σ = 0.05 and 0.1) and real-world noise conditions on inversion accuracy in our earlier study (Chen et al., 2021). Therefore, this study focuses on extending the noise analysis to more complex high-dimensional cases. ("*The observation data for model parameter inversion are generated by adding*

Gaussian noise perturbations to the numerical model simulation results. Specifically, oobservational noise is introduced by multiplying the simulated data by a random noise factor $\varepsilon \sim N(1, \sigma^2)$*,⋯⋯⋯we conducted an extended analysis on Case 3—the most complex scenario—by increasing the noise level to 10% (σ=0.1)*")

For the results in the high-dimensional noise setting, given that our study assumes Gaussian noise and that observational constraints are sufficiently dense, the impact on the inversion results of model parameters is not significant ("*Nevertheless, it is important to note that while the inversion accuracy under a 10% noise level remains comparable to that in the 1% noise scenario, increasing observational noise inevitably raises the convergence value of the least-squares loss function. ⋯⋯⋯In such cases, incorporating additional system information as regularization constraints is essential to enhance the robustness of the objective function and mitigate ill-posedness*."). However, we acknowledge that real-world scenarios often involve more complex observational noise distributions and practical constraints, such as limited monitoring networks due to cost restrictions. Hence, we have expanded the discussion in the last two paragraphs of Section 4.3, highlighting potential challenges in real-world studies and possible strategies. This issue is also further addressed in our response to Comment 3.

Comment 3. The discussion about the practical relevance of the test cases must be improved. The authors claim that "the model conditions of the two synthetic cases in this study are primarily based on previous studies, as well as large-scale sandbox". Their discussion of the relevance of the results for practical applications is not convincing: the remarks in section 4.3 and in the replies to the reviewers are not well supported from physical arguments and should be deeply revised.

Response:

We followed these suggestions to reformat and revise our manuscript deeply. The description of the similarities between our model setup and real-world hydrogeological conditions has been moved to the first paragraph of Section 3, including detail discussions on the newly added Case 3. Specifically, Case 3 now features a lower hydraulic gradient and a model scale of 0.8km. Thus, the three cases collectively cover high, medium, and low hydraulic gradient scenarios, providing a more comprehensive representation of different hydrogeological conditions. ("*This study designed three synthetic cases based on previous research, covering different model scales and hydraulic gradient combinations (Jose et al., 2004; Zhang et al., 2018; Mo et al., 2019) to evaluate the performance of the TNNA algorithm against conventional metaheuristic algorithms. Both Case 1 and Case 2 are small-scale scenarios, with simulation time measured in days.⋯⋯⋯ase 3 simulates contaminant plume migration at a sub-regional scale (approximately 1 km), with simulation time measured in years. It uses a hydraulic gradient of 0.00625, representing a smaller natural gradient typically found in plain aquifers.⋯⋯*.")

In Section 4.3, we have revised and expanded the discussion on practical applications. In the supplemented descriptions, we explain that well-performing optimization algorithms are generally applicable across different optimization tasks, provided that the constraints for inversion are properly defined. However, the primary challenge in real-world studies lies in establishing robust constraint conditions tailored to specific scenarios. We specifically provided a detailed discussion on three potential challenges associated with constructing nonlinear optimization models for real-world applications: 1) Representation of complex heterogeneous model parameter fields; 2)

Maximizing the effective observational information while optimizing monitoring costs; and 3) Integrating multi-source data and accounting for uncertainties in model process to better address complex observational noise scenarios and uncertainties in physical mechanisms.

("*4.3 Parameter inversion method comparison results*

This study validates the computational efficiency and inversion reliability of the TNNA algorithm under three different heterogeneous conditions. In optimization-based inversion studies, the primary challenge is to establish nonlinear inversion constraints and design efficient algorithms to find optimal parameter solutions………..")