To: Prof. Heng Dai MARCH 2, 2026

Editor, Hydrology and Earth System Sciences

Re: Revised manuscript "Joint characterization of heterogeneous conductivity fields and pumping well attributes through iterative ensemble smoother with a reduced-order modeling strategy for solute transport" (Paper EGUSPHERE-2025-5320) by Chuan-An Xia, Jiayun Li, Bill X. Hu, Alberto Guadagnini, Monica Riva.

Dear Editor:

We truly appreciate the efforts you and the Referees have invested in our manuscript. Following your recommendation, we are submitting a revised work addressing the Referees' comments.

Please, find in the following an itemized list of the Referees' comments together with our response to each. Comments are listed in black font and our responses in blue font. Modifications implemented in the Revised Manuscript are indicated in the "Author's track-changes" document.

Sincerely,

Chuan-An Xia, Jiayun Li, Bill X. Hu, Alberto Guadagnini, Monica Riva.

Referee#3

This paper proposes a ROM-based iES method and applies it to the joint estimation of aquifer hydraulic parameters and pumping-well parameters. The authors systematically test the factors that influence the performance of the algorithm. The conclusions of the paper are generally reliable. Moreover, to the best of my knowledge, there have been relatively few previous studies on the joint inversion of hydraulic conductivity (K) fields and pumping-well characteristics; most existing studies have focused on the joint inversion of K fields and contaminant source parameters. However, the main methodology adopted in this paper and the key factors investigated, such as the choice of n and N_MC, as well as the quantity and quality of observation data, have already been extensively discussed in prior studies. Therefore, I believe that this work mainly explores a new application scenario based on existing methods, rather than introducing fundamentally new methodological developments.

The authors have responded positively to the comments raised by previous reviewers.

Below, I provide several additional suggestions for the authors' consideration.

Answer: While in line with our overall research strategy, integration of an iterative Ensemble Smoother (iES) with a Reduced-Order Model (ROM) for solute transport in randomly heterogeneous aquifers is proposed and evaluated here for the first time, to the best of our knowledge. Similar to any other reduced-order modeling approach, it is associated with approximation errors. The latter depend on the selected reduced dimension (n). As the effect of the number of Monte Carlo simulations always depends on the scenario considered and (in our extensive experience) there is only limited theoretical basis according to which it can be quantified in a general way, one always needs to address it. Our detailed point-by-point responses to the Referee's comments are provided below.

Major Comments

1. The primary issue addressed in this paper is computational cost. However, based on my experience, the computational burden of the test models considered in this study is not particularly high. I have reviewed several of the authors' previous papers, many of which focus on ROM techniques, and I myself conducted substantial research on surrogate models earlier in my career. With the continuous improvement in computational performance, computational cost has become a relatively less critical issue. In groundwater inverse modeling, more fundamental challenges arise from factors such as the non-Gaussianity of parameter fields and the equifinality caused by insufficient observational data. I encourage the authors, in future studies, to apply these methods (or to develop new ones) to more challenging problems that better reflect these intrinsic difficulties.

Answer: We agree that, beyond computational efficiency, inverse modeling in subsurface hydrology settings involves fundamental challenges such as, e.g., non-Gaussian parameter distributions and equifinality arising from limited observational data. In response to the Referee's comments, we have expanded the discussion on future research directions to include these intrinsic difficulties and to clarify the broader applicability of the proposed framework. The revised text (lines: 773-783 of the revised manuscript) now reads: "Additional elements of interest for future studies on coupling iES with ROM can include analyses of transient saturated/unsaturated flows (in conjunction with, e.g., time-dependent pumping strategy), reactive transport processes, and density-dependent flow/transport scenarios. Future efforts should also address characterization of aquifer heterogeneity upon relying on the theoretical framework associated with generalized sub-Gaussian random fields (Riva et al., 2015; Xia et al., 2024), which may aptly represent non-Gaussian features and statistical scaling of subsurface properties. When considering nonlinear systems, reliance on discrete matrix interpolation schemes (Negri et al., 2015; Bonomi et al., 2017) appears to be a promising strategy to further enhance the computational efficiency and robustness of ROM-based approaches.”.

References

Riva, M., Neuman, S.P., Guadagnini, A., 2015. New scaling model for variables and increments with heavy-tailed distributions. Water Resour. Res. 51, 4623-4634.

2. The authors should report and compare the computational time required by different approaches. In addition, quantitative error metrics between the ROM and FES outputs, such as R2 and RMSE, should be provided to better assess the accuracy of the ROM.

Answer: We now write that (lines: 686-700): “A single forward simulation for TC28 requires approximately 13 minutes of CPU time on the hardware platform used in this study (13th Gen Intel® Core^{TM} i7-13700K 3.40 GHz with 32 GB RAM). The total CPU time required to complete TC6 upon relying on iES_FSM is 122 minutes, whereas the corresponding CPU time required to complete TC28 through iES_ROM is 28 minutes, thus representing a speedup of approximately a factor of 9. Percentage differences associated with E_Y and S_Y are equal to 0.50% and 0.21%, respectively, suggesting that the computational gain is achieved with negligible loss of accuracy. To further quantify the approximation error introduced by the ROM, we evaluate the residual mean square errors between the concentration field obtained from the first realization of the initial ensemble using the full-scale model and the corresponding ROM solutions with reduced dimensions n = 5, 10, 15, 20, 25, and 30. The ensuing error values are 0.8630, 0.4156, 0.2699, 0.1909, 0.1312, and 0.1336, respectively, demonstrating systematic error reduction as the reduced dimension increases. We further note that all of the associated coefficients of determination are higher than 0.99."

3. The study considers a steady-state scenario with a scalar pumping rate, which is somewhat overly simplified. I recommend extending the analysis to include time-varying pumping rates. In my own tests of joint inversion problems involving groundwater contamination sources and K fields, the contaminant release parameters were time-dependent.

Answer: At this stage, the assumption of steady-state groundwater flow serves two main purposes: (*i*) to reduce the complexity of the joint inversion problem and (*ii*) to ensure interpretability of the results, thus providing a clear and tractable foundation for the analysis. Considering the already broad scope of the current manuscript, we are strongly convinced that a detailed treatment of transient flow conditions would be more appropriately addressed in a future dedicated research. Accordingly, we have incorporated this suggestion into the "Future Work" section of our Conclusions.

The revised text (lines: 773-783) now reads: "Additional elements of interest for future studies on coupling iES with ROM can include analyses of transient saturated/unsaturated flows (in conjunction with, e.g., time-dependent pumping strategy), reactive transport processes, and density-dependent flow/transport scenarios. Future efforts should also address characterization of aquifer heterogeneity upon relying on the theoretical framework associated with generalized sub-Gaussian random fields (Riva et al., 2015; Xia et al., 2024), which may aptly represent non-Gaussian features and statistical scaling of subsurface properties. When considering nonlinear systems, reliance on discrete matrix interpolation schemes (Negri et al., 2015; Bonomi et al., 2017) appears to be a promising strategy to further enhance the computational efficiency and robustness of ROM-based approaches.".

References

Riva, M., Neuman, S.P., Guadagnini, A., 2015. New scaling model for variables and increments with heavy-tailed distributions. Water Resour. Res. 51, 4623-4634.

Minor Comments

1. If an abbreviation has already been defined earlier in the text, there is no need to restate the full term or redefine it later; this is the purpose of introducing abbreviations.

Answer: We intentionally chose to redefine certain abbreviations within the text. This contributes, in our view, to enhance readability for a broader audience. For this reason, we prefer to retain the current presentation. We will naturally abide to the Editor's decision on this point.

2. When introducing previous studies, the authors use the present tense; however, since these studies describe past work, the simple past tense is more appropriate. For example, at L106: Ju et al. (2018) relied on ...

Answer: We acknowledge that both past and present tense are generally acceptable in scientific journals and are commonly used in the literature, including our previous works. Consistent with our preference, we respectfully prefer to maintain the current formulation. We will of course abide to the Editor's final decision on this point.

3. L87: replace location with localization.

Answer: We have revised the text (lines: 86-87) which now reads: "Main advantages associated with localization approaches are related to the observation that".

4. L126: remove "of the".

Answer: We have revised the text (lines: 125-127) which now reads: "ROMs are typically constructed upon projecting the governing equations and boundary conditions onto a lower-dimensional subspace spanned by a set of basis functions.”.

5. The paper contains a large number of variables, and better consistency is recommended throughout the manuscript (this issue was also raised by a previous reviewer). In particular, N and N_MC both appear to denote ensemble size. In addition, when variables appear in figure captions, brief explanations should be provided.

Answer: We have checked very carefully throughout the text and make sure the consistency of variable definition. We find that adding explanations for variables in the captions of Tables and Figures increase the redundancy of the text of the captions. We prefer to maintain the current wording, unless there are strong elements to change it.

6. Section 2.2.1: additional explanation is needed. The physical meanings and roles of A, c, and F are not sufficiently clear.

Answer: We now have revised the description to these variables. The text (lines: 243-247) now reads: “**A** is the full-system stiffness matrix (of size $N\times N$), which embeds information on spatial velocity, dispersion, and effective porosity; **c** is the vector (of size $N\times 1$) of solute concentration values; and **F** is a vector (of size $N\times 1$) whose entries encompass source/sink terms and initial and boundary conditions.”.

7. L281: please explain why a logarithmic transformation is applied to q_s.

Answer: We have revised the text (lines: 282-289) which now reads: “We denote by $\mathbf{m}=\left[Y_1,Y_2,\ldots,Y_N,\ln q_s,x_{1,q_s},x_{2,q_s}\right]^T$ the vector (of size P = N+3) whose entries correspond to the uncertain model parameters (i.e., the log-conductivity, $Y = \ln K$, field) and flow rate and location of a pumping well. The pumping rate is parameterized in logarithmic form to ensure consistent scaling with the log-conductivity field and to reduce disparities in parameter magnitudes. Such a transformation is consistent with common practice in inverse modeling to improve numerical conditioning, enhance stability of the estimation process, and mitigate potential bias arising from large disparities in parameter magnitudes.”.

8. Equation (7): if a Kalman update is used, gamma should represent the covariance matrix of observation errors. Furthermore, if the ESMDA framework is adopted, inflation of the observation error covariance matrix should also be considered. Based on the formulation, the EnRML framework does not appear to be used. In addition, the innovation vector seems to be written incorrectly; it should be the observation data plus random perturbations minus the model output.

Answer: Our revised text (lines: 290-300) now reads: “We further denote by $\mathbf{d}=\left[d_1,d_2,\ldots,d_O\right]^T$ the vector (of size O) of the randomly perturbed observations (i.e., measured head and concentration values). To estimate **m**, we implement the iES (Luo and Bhakta, 2020; Xia et al., 2024):

$$\begin{cases} \mathbf{m}^{k+1}=\mathbf{m}^{k}+\underline{\underline{\mathbf{K}}}_{Gain}^{k}\Delta\mathbf{d}^{k} \\ \underline{\underline{\mathbf{K}}}_{Gain}^{k}=\underline{\underline{\mathbf{S}}}_{m}^{k}\left(\underline{\underline{\mathbf{S}}}_{d}^{k}\right)^{T}\left(\underline{\underline{\mathbf{S}}}_{d}^{k}\left(\underline{\underline{\mathbf{S}}}_{d}^{k}\right)^{T}+\gamma^{k}\underline{\underline{\mathbf{C}}}_{d}\right)^{-1} \quad \text{with} \quad \gamma^{i}=\xi^{i}\,\mathrm{trace}\left(\underline{\underline{\mathbf{S}}}_{d}^{i}\left(\underline{\underline{\mathbf{S}}}_{d}^{i}\right)^{T}\right)\Big/O\,. \\ \Delta\mathbf{d}^{k}=\mathbf{d}-g\left(\mathbf{m}^{k}\right) \end{cases} \tag{7}$$

Here, superscript k is the index of the iteration step; matrices $\underline{\underline{\mathbf{S}}}_{m}^{k}=\left[\mathbf{m}_1^k-\bar{\mathbf{m}}^k,\cdots,\mathbf{m}_{N_{MC}}^k-\bar{\mathbf{m}}^k\right]\Big/\sqrt{N_{MC}-1}$ (of size $P\times N$, where $\bar{\mathbf{m}}^k=\sum_{j=1}^{N}\mathbf{m}_j^k\Big/N_{MC}$) and $\underline{\underline{\mathbf{S}}}_{d}^{k}=\left[g\left(\mathbf{m}_1^k\right)-g\left(\bar{\mathbf{m}}^k\right),\cdots,g\left(\mathbf{m}_{N_{MC}}^k\right)-g\left(\bar{\mathbf{m}}^k\right)\right]\Big/\sqrt{N_{MC}-1}$ (of size $O\times N$, where $g(\bullet)$ represents model operator, which is either FSM or ROM) collect the ensemble anomalies of parameters and simulated observations associated with the k^{th} iteration step; $\underline{\underline{\mathbf{C}}}_{d}$ is the covariance matrix of observation errors”.

9. L313–314: please clarify the meanings of the inner and outer iterations.

Answer: Our revised text (lines: 320-323) now reads: “If an outer iteration does not lead to reduction in data misfit, an inner iteration is triggered to progressively decrease the time step until a lower misfit is obtained. When implementing the LM optimization scheme, we set the maximum number of both inner and outer iteration to 10 (see also Luo and Bhakta, 2020).”.

10. L327: what does N represent here? Earlier, N appears to denote ensemble size. Based on my understanding, the number of iES model evaluations should be (N + 1) * N_iter.

Answer: Here, N indicates the number of nodes in the numerical grid associated with the finite element solution. We refer to the computational cost in a single forecast step. To improve clarity, our revised text (lines: 335-340) now reads: “The steady-state groundwater flow is solved through the FSM in both iES_FSM and iES_ROM, with a computational cost of order $O\left(N^3\left(N_{MC}+1\right)\right)$ for each forecast step. The main computational cost for the N_{MC} FSM-based MC realizations of solute transport at a single time step in iES_FSM is $O\left(N^3\left(N_{MC}+1\right)\right)$, while being $O\left(\left(sN+N^2\right)\left(N_{MC}+1\right)\right)$ (where $s \approx 7$ or ≈ 15 in two and three dimensions, respectively) for iES_ROM.”.

11. Figure 3 caption: Delta E_Y represents a percentage; it is recommended to indicate (%) on the y-axis.

Answer: We have revised it.

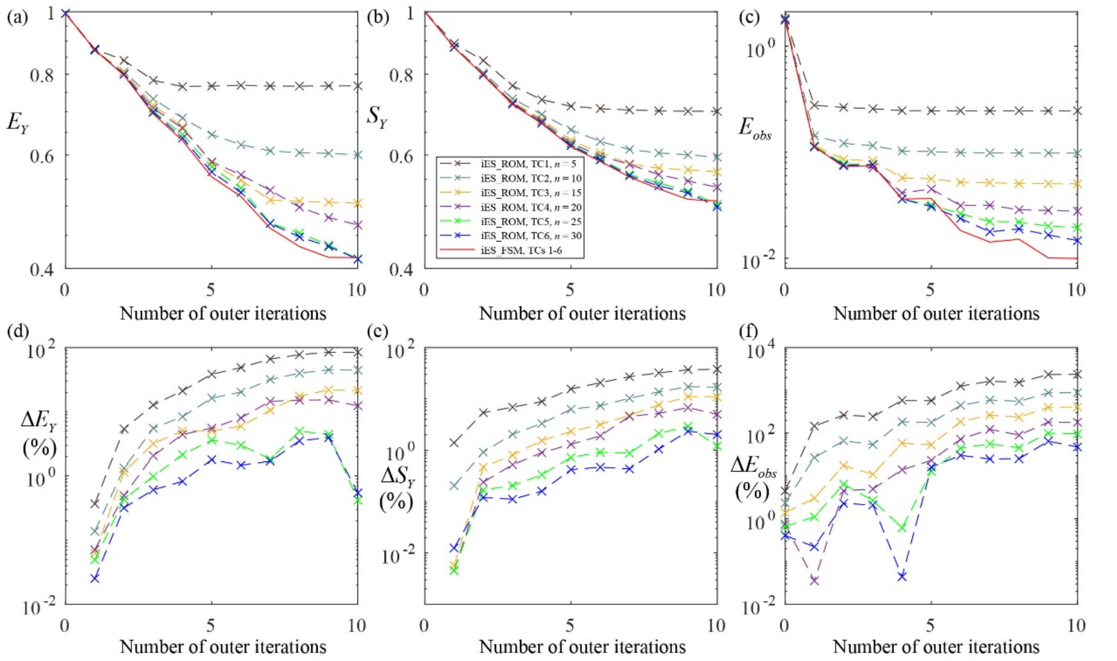

Fig. 3 Values of (a) E_Y, (b) S_Y, and (c) E_{obs} versus the number of outer iterations obtained through iES_ROM considering various dimensions of reduced-order model (with n = 5, 10, 15, 20, 25, and 30 for TCs 1-6, respectively) and iES_FSM (which provides identical results for TCs 1-6) for ensemble size N_{MC} = 10,000; corresponding percentage differences between the values of (d) E_Y (ΔE_Y), (e) S_Y (ΔS_Y), and (f) E_{obs} (ΔE_{obs}) evaluated through iES_ROM and iES_FSM.

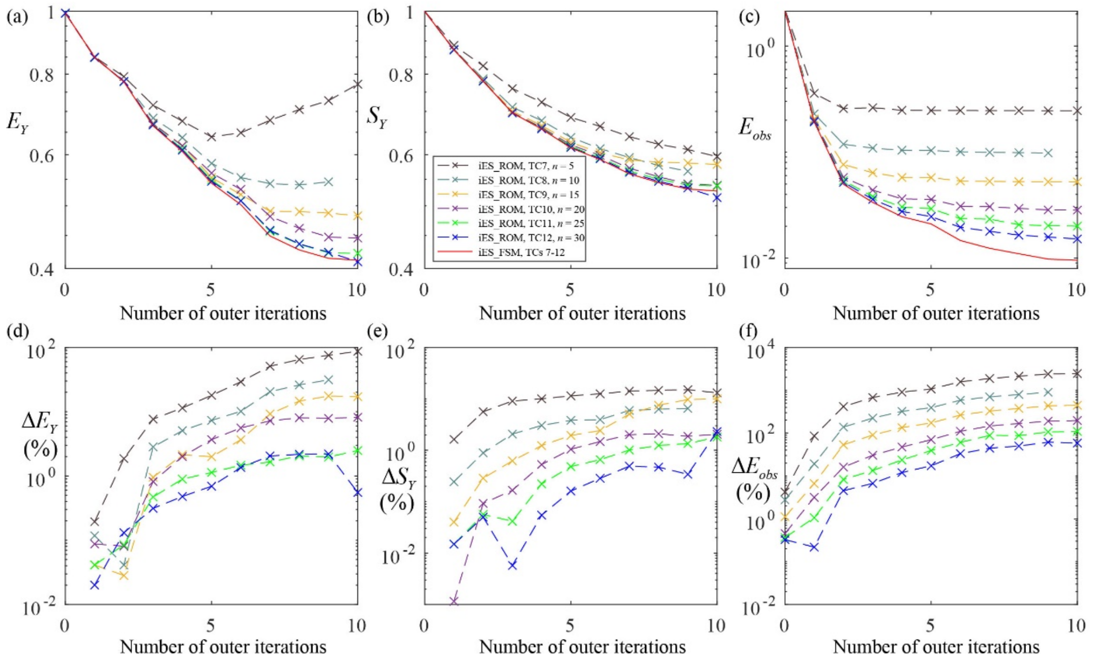

Fig. 4 Values of (a) E_Y, (b) S_Y, and (c) E_{obs} versus the number of outer iterations obtained through iES_ROM considering various dimensions of reduced-order model (with n = 5, 10, 15, 20, 25, and 30 for TCs 7-12, respectively) and iES_FSM (which provides identical results for TCs 7-12), when the pumping rate and location are previously known and for an ensemble size N_{MC} = 10,000; corresponding percentage differences between the values of (d) E_Y (ΔE_Y), (e) S_Y (ΔS_Y), and (f) E_{obs} (ΔE_{obs}) evaluated through iES_ROM and iES_FSM.