

COMMENTS FROM Referee #3

General comments

The paper presents a very interesting topic and can be useful for the community. I think it needs some clarification in the writing as it is difficult to follow the full discussion. There is the need to provide some clarification for when the authors reference to the truth case, the reference and the benchmark scenarios. Consistency on how the different conditions are referred to will help the reader.

Thanks for your suggestion. We have provided a detailed description of the benchmarking scenarios and synthetic cases in Sections 4.2 and 4.3, respectively. The clarity of the reference solution may be improved, and as such, we have included the following explanation of the reference solutions in the abstract (lines 6-9):

“However, in past studies until now, comparisons were made among approximate methods without firm reference solutions. Note that the reference solutions are the best possible solutions with best estimate, and posterior standard deviation and so forth.”

One question is about the added information that would come from the two synthetic truths: they only have a different standard deviation and I am not sure there is much to gain from comparing the two of them. I wonder if it would be more useful to compare different hydrogeological conditions. Something like bookend scenarios that look at very different assumptions would be very beneficial for the reader to understand the power of the methods presented.

Thank you for your comment. The rationale for utilizing two synthetic truths with varying standard deviations is to assess the efficacy of methods in handling fields with low and high heterogeneity.

We greatly value your suggestion and believe that it would be interesting to conduct a comparative analysis of diverse hydrogeological conditions. We will certainly take this into consideration and encourage the community to explore alternative hydrogeological conditions in future research endeavors. This has been added in the discussion and conclusion section in lines 705-707:

“Furthermore, it would be intriguing to undertake a comparative analysis of various hydrogeological conditions to gain a comprehensive understanding of the efficacy of inverse methods.”

Looking at the scenarios in table 3, the regular well distribution scenario does not seem to add more information and S0, S2 and S3 do not really represent a variety of conditions.

Thanks for your suggestion. As we mentioned in lines 419-421, Scenario S1 with the regular monitoring network is a fallback scenario for the comparison with irregular monitoring networks. The close spacing of some monitoring wells in the irregular networks may present challenges for certain methods due to their high autocorrelation. The standard deviations of InK synthetic truths and measurement errors are different for S0, S2, and S3, and the impact of these two conditions is significant in reality.

My comments below are mostly about Figures: the paper would really benefit from a more consistent representation and discussion of the results and more consistency in the figures with results.

I would also suggest to provide some more informative conclusions, maybe summarized in a bullet point lists.

Thank you for your comments and suggestions on our manuscript. Comments and suggestions are all valuable and very helpful in revising and improving our paper. We have carefully considered the comments and have made corrections that we hope will meet with your approval.

Specific comments

Line: suggest to change to “across multiple scales”

Replaced as suggested

Line 37: I would not say that data scarcity and subsurface heterogeneity are the sole responsible for uncertainty. Depending on the model, we need to add uncertainty on other terms of the water budget.

Thanks for your comment and suggestion. We acknowledge this limitation in the manuscript in lines 39-42:

“There are multiple contributing factors to the uncertainty, including cognitive limitations of models, recharge values, lateral inflows, data scarcity, and subsurface heterogeneity. These factors are responsible for the persistence of uncertainty even after calibration. The latter type of uncertainty is crucial and must be quantified to provide robust decision support in engineering and management practice.”

Line 413: why S1 is a fallback scenario? I wonder if S1 is needed, and if yes I think there is the need to have at least to scenarios with regular well distribution to be able to make a comparison

Thanks for your comment. We believe that including S1 in the comparison is essential for evaluating the performance of irregular monitoring wells (S0) versus regular monitoring (S1). As mentioned in our paper, the close spacing of some

monitoring wells in the irregular networks may present challenges for certain methods due to their high autocorrelation.

Figure 2: as mentioned earlier: is it a limiting factor having two truths that are so similar and only different for standard deviation?

Thank you for your comment. Although the two K fields share the same correlation, they have different variances: one with a standard deviation of 2 and the other with a standard deviation of 1. The rationale for utilizing two synthetic truths with varying standard deviations is to assess the efficacy of methods in handling fields with low and high heterogeneity. There are many hyperparameters used for generating the K fields, and it is not realistic to compare the effect of all hyperparameters in one paper. However, we welcome the community to test other hyperparameters in the future.

Figure 3: I think that the use of a different scale for S2 is misleading, but if the authors believe that it is important I would suggest to add explanation

Thank you for your suggestion. We have added the following explanation in the caption of the figure:

“Please note that S2 exhibits the same standard deviation as synthetic truth 2, while the others exhibit the same standard deviation as synthetic truth 1. Therefore, the scale of S2 should align with that of synthetic truth 2, and the scale of other scenarios should align with that of synthetic truth 1.”

Besides, we also add the following explanation in the caption of Figure 2:

“Please note that to maintain the primary characteristics of the lnK fields given their diverse standard deviations, the scale for these two synthetic truths has been adjusted individually.”

Figure 4: same comment on the scale

Thank you for your suggestion. We have added the following explanation in the caption of the figure:

“Please note that S2 exhibits the same standard deviation of 1 as synthetic truth 2, while the others exhibit a standard deviation of 2 similar to synthetic truth 1. Therefore, to ensure the distinct features of the figures, the scale setting for S2 is adjusted differently from the others.”

Figure 5: same comment: it would be good to add an explanation

Thank you for your suggestion. We have added the following explanation in the caption of the figure:

“Please note that in order to maintain the primary characteristics of the evolution of the RMSE, the scale for these RMSE has been adjusted individually.”

Line 514: it is really hard to see the sharp decrease

Yes, as indicated in the paper, there is a sharp decrease in RMSE within a brief initial period depicted in the plots. In comparison to the overall extended period, this rapid change during the initial phase is scarcely discernible.

Figure 6: the legend for S0 says “reference” and for the other scenarios “Synthetic truth”. Is this correct? Also, are there some conclusion to be derived from the mean always being above the reference/synthetic truth?

Thanks for your comment. The term "Synthetic" should also be used, and the figure for S0 has been adjusted. Despite the minor discrepancy, it is still unclear why the mean consistently remains higher than the synthetic truth.

Figure 9: is the reference solution, the true solution?

Thank you so much for your comment. The reference solutions are the best possible solutions with best estimate and posterior standard deviation (as mentioned in lines 443-444 of the original version of the manuscript). We have now included the following explanation of the reference solutions in the abstract (lines 6-9).

“However, in past studies until now, comparisons were made among approximate methods without firm reference solutions. Note that the reference solutions are the best possible solutions with best estimate, and posterior standard deviation and so forth.”

Technical corrections

No technical corrections.