**Comments to the Author**  
The reviewers suggest major revisions. Some major ones include: i) strengthening the contribution of the work by intensive literature review and comparing to similar work as suggested by reviewers; ii) more description of the model/framework, demonstration of the method, and adding inputs for reservoir simulations; 3) adding data or analysis on uncertainty/risk assessment 4) consolidating the conclusions through providing the comparison with Chen et al (2020); I also concur with other comments from reviewers.  
  
  
Technical Reviewer: 1  
  
**Comments to the Author**  
This manuscript presents a concise and clear study of the usage of ES-MDA-GEO algorithm for assimilating 2D saturation maps. I'd suggest publication with minor revisions. My specific comments are:  
  
1: Looking at the updated models shown in Figure 3., it seems there is some level of uncertainty collapse (i.e.: all posterior models look very similar), and none of the updated models in Figure 3 has a "high-perm channel" around the top left corner as shown in Ground truth model.

It is possible that the workflow is suffering from uncertainty collapse. Our ongoing work is studying the effect of spatial data assimilation on uncertainty quantification and will be in the scope of our future work.  
  
2: Please provide more discussions regarding the use of ES-MDA-GEO such as showing the evolution of objective function w.r.t iterations. In addition, other than visually checking the history-matching quality, providing some metrics such as RMSE will be useful.

We show the evolution of ESMDA-GEO with respect to each assimilation step in Figure 5, as the predicted saturation maps based on prior-vs-updated models. Ongoing work on uncertainty quantification will address different metrics (such as RMSE and SSIM) on the responses.  
  
3: The results in Figure 7 are a bit weird and may not be very representative of a practical case. All the updated models look identical, which could be due to uncertainty collapse. In addition, I assume here the 10% error is a relative error. So even with such a level of error, the CO2 front will still be captured very well since 0 CO2 saturation will still be 0 CO2 saturation with added noise. While in practice, the seismic survey is unlikely to capture the CO2 front precisely like what you have shown in Figures 5 and 6. So instead of looking at the error level, it might be more useful to perturb the CO2 saturation front and look at the corresponding impact.

Thank you for this comment. It is definitely true that perturbing the saturation front might have significant effects on the parameter estimation, and this is part of our ongoing work and future results. For this publication, we focus on adding random noise to the CO2 saturation values, but will study the effects of perturbing the front in the future.

Technical Reviewer: 2  
  
**Comments to the Author**  
This paper deals with updating permeability by history matching based on CO2 saturation map from 4D seismic data. The manuscript consists of the following two sensitivity analysis:  
  
a) available 4D seismic data: 1 year vs 1 and 3 years vs 1, 3, and 5 years

Correct, we have 4D seismic data, used to estimate saturation maps, and recorded at years 1, 3 and 5. At each assimilation step (3) we assimilate the 1st year, 1st and 3rd year, and 1st, 3rd, and 5th year.  
  
b) noise of CO2 saturation map: 1%, 3%, 5%, and 10%  
Overall workflow is obvious and the manuscript was written well. However, it requires few considerations, which should be fixed before the publication to meet the quality of SPE Reservoir Evaluation & Engineering. Please find the comment as below.

Correct, we add relative errors of 1%, 3%, 5%, and 10% to the saturation maps and perform data assimilation to obtain the corresponding estimated permeability maps.  
  
1) Lines 44-45, Page 2: There are few research papers, which dealt with 4D seismic data as dynamic data for ensemble-based methods. Please find the following references. To secure the originality of this manuscript, I recommend that the author should describe the differences compared to previous studies.  
- [https://doi.org/10.1016/j.petrol.2020.107961](https://urldefense.com/v3/__https:/doi.org/10.1016/j.petrol.2020.107961__;!!Bt8fGhp8LhKGRg!DPgP3Fz6Qx0VEcwF6Czs1PoAyG8MQlF9yi9uVfxNrtfEtTshUsF7tjTwIcRyX-8NXAqQVmdEpSO0gw09PLqRk_zH2A3SC-I$)  
- [https://doi.org/10.2118/180025-PA](https://urldefense.com/v3/__https:/doi.org/10.2118/180025-PA__;!!Bt8fGhp8LhKGRg!DPgP3Fz6Qx0VEcwF6Czs1PoAyG8MQlF9yi9uVfxNrtfEtTshUsF7tjTwIcRyX-8NXAqQVmdEpSO0gw09PLqRk_zHH1JCva8$)  
- [https://doi.org/10.1016/j.petrol.2021.109728](https://urldefense.com/v3/__https:/doi.org/10.1016/j.petrol.2021.109728__;!!Bt8fGhp8LhKGRg!DPgP3Fz6Qx0VEcwF6Czs1PoAyG8MQlF9yi9uVfxNrtfEtTshUsF7tjTwIcRyX-8NXAqQVmdEpSO0gw09PLqRk_zHWJZrP_M$)

Included these references in introduction.  
  
2) Methodology section, Page 3: Please present basic equations and concepts how to determine adequate attenuation of variations in ES-MDA-GEO by Rafiee and Reynolds (2017).

Page 3, lines 15-16: we describe how to obtain the geometric inflation factors through TSVD.  
  
3) Line 36, page 3: Please present structure map (i.e., cross section map) of synthetic model of the reference model.

We only assume the existence of seismic model/data, from which we can obtain saturation maps through fluid substitution equations.  
  
4) Most figures in the manuscript shows individual models but please present the mean permeability model together. They can be seen in Appendix section if the authors wish.

Included mean and standard deviation maps of ensemble permeability in Figure 1.

5) Figure 4, page 5: how the measurement error covariance, Cd, was set for the 4D seismic data from the reference field in ES-MDA.

The measurement error covariance matrix, Cd, was updated based on the Gaspari-Cohn method.

6) Figure 6, page7: clarify the results are from the R1 model.

Clarified in Page 7, lines 7-8, and also in Figure 6 caption.  
  
7) Lines 30-31: Please explain in detail for “the level of data noise from 1% to 10%”. The x% of CO2 saturation for each simulation grid, which were simulated from the reference field were changed directly or the authors set the measurement error covariance in ES-MDA.

Clarified in Page 7, lines 24-25.  
  
8) Please provides reservoir simulation conditions as table and figures including fluids and rock characteristics including injection condition of CO2 (e.g., constant injection rate during 15 years) and boundary condition at the edge of reservoir area

Reservoir simulation conditions are further explained in Page 3, lines 41-47.

9) Figure 7, page7: clarify whether the history matching results include the 5th year of 4D seismic data.

Clarified in Page 7, lines 26-27  
  
10) Lines 34-37, page 8 in conclusions: in the manuscript, there is no comparison with the previous work in Chen et al. (2020). Therefore, I cannot confirm that “Compared to our previous work (Chen et al, 2020) based on data assimilation of point measurements from monitoring wells, the spatial data appear to have more value of information than point measurements from monitoring wells to reduce the uncertainty in the prediction of CO2 plume area or distribution.”

We refer readers to our previous work with point measurements of monitoring wells and use this paper to describe the assimilation of spatial data instead. Though we do not provide quantitative comparisons, we intend to point out the advantage of spatial data assimilation compared to point measurements.  
  
11) Others  
- title of Figure 2: clarify the results are from the reference model.

Clarified in figure caption  
- there are two Chen et al. (2022) in References section. Please distinguish them as 2020a and 2020b.  
Clarified 2022a and 2022b for Chen et al. in Page 2, lines 13, 24, and 26.

Technical Reviewer: 3  
  
**Comments to the Author**  
The paper by Chen et al presents results from using an established data assimilation framework for the purposes of mapping CO2 plumes in CO2 geologic storage for CO2 sequestration applications. The paper presents one case study and concludes that using ES-MDA-GEO with spatial CO2 saturation data, provides better estimates of the CO2 plume compared to not using the spatial CO2 saturation data. The paper itself is clearly written, with proper writing and the data shown support the claim of the authors. However, there are several improvements that need to be made to the paper before considering publications. For this reason, I recommend publication of this article after major revision. The revisions I am suggesting include additional writing, and further demonstration of the method. Comments are listed below in the order of appearance in the manuscript, not in order of importance.  
  
1. The title of the paper claims that the study involves risk forecasting. The authors did not present any measures of uncertainty for their case study, they only did CO2 mapping. IF the authors want to keep Risk Forecasting in their title, they need to include relevant results in their paper.

Removed “Risk forecasting” from the title, focus is mainly on spatial data assimilation.  
  
2. Introduction: The authors do not introduce the topic sufficiently. ES-MDA is mentioned without defining the acronym. While the method is established in the DA field, it still needs to be named as an ensemble smoother and the multiple data assimilation version of an ES needs to be spelled out.

Expanded the acronym in Page 2, lines 7-8  
  
3. This is definitely not the first paper using spatial information. There are many papers that have used spatial information, one that comes to mind is Li et al., 2015 (10.1002/2015WR017203) , but there are many data assimilation in geophysics that utilizes spatial data. Please provide a more thorough literature review.

Included prior literature on this in Page 1 lines 28-29 and Page 2 line 2.  
  
4. The authors claim to present a framework, but no framework is presented, only the results are shown. In addition, the authors claim integration into NRAP (which is neither defined nor explained), but do not mention what his integration consists of.

NRAP is defined and explained in the Abstract (Page 1, lines 8-9) as well as Page 2, lines 17-18. The integration is mentioned in Page 2, lines 6-7 and lines 17-18.   
  
5. Page 3 line 14: The authors claim that ESMDAGEO is recognized as the most effective ES variant. While I do not question that, the authors need to cite literature for this claim.

Removed text in parenthesis “(the most advanced version of ES-MDA)”.

6. Page 3 line 29: Typo in citation. Parenthesis needs to be deleted.

Parenthesis removed in citation.  
  
7. Page 3 line 36: The authors claim they developed a workflow. The workflow of ES-MDA is specific and it does not seem that there was any change made to it. Using spatial data can be handled exactly the same way that point data can be used. There is no discussion of how the spatial data were assimilated, if it was done any differently than in the authors’ previous study. Authors should both tone down the novelty of the work, and explain in much more detail the way they assimilated spatial data, since that is the contribution of the work, spatial versus well data, as the authors’ themselves mention on line 31, 32 of page 4.

The previous work consisted on assimilating point measurements (monitoring well data) instead of spatial map data. The workflow presented consists of assimilation spatial data from geologic CO2 storage project for the estimation of uncertain geology.   
  
8. Page 3 line 41: depth and thickness correspond to y and z? Did the authors mean width and thickness?

Reworded this sentence. We mean that the reservoir is 100 meters thick (z-direction), and it is 1 km deep in the subsurface.  
  
9. Page 3 line 44: was there any vertical correlation or was it independent 2D random permeability maps with given statistics?

Page 3, lines 40-41 state that we assume the reservoir to be homogeneous in the z-direction, meaning that there is no vertical correlation along the 11 layers in the z-direction (“… the remaining ten layers possess an equivalent permeability distribution to that of the uppermost layer.”)  
  
10. Figure 2: Authors need to explicitly state if they used all data points on the CO2 maps, some coarser grid, or some other spatial metric of the CO2 fields. As with all Figures of the paper, the font of the axes needs to be increased.

Page 4, lines 13-14. We include a statement to explicitly state we use the entire CO2 maps, not a coarser grid or other spatial metric.

11. Figure 3: Focus on the top right corner of all the estimated fields. For all realizations, the smoother was able to update the field and get that low blow point on the top left corner. Given that there is no data to support that small low k point, how do the authors justify that the smoother was able to update it to the correct value? It would be more convincing if the authors also shows the updates for each field, or if they attempted to explain the smoother’s success in these low sensitivity areas. What would happen if the authors started with a prior that was not so informative with the exactly correct statistics? Font of all axes needs to be matching text font.

We include Figure 5 to show the (mean and standard deviation) of the evolution of the updates at each assimilation step. With this, we attempt to demonstrate and support the claim that the smoother is able to update these values correctly over the assimilation steps. For future work we are considering using different geostatistical simulation parameters to generate more/less informative priors for the estimation.  
  
12. Page 7 line 27: Authors talk about uncertainty and risk quantification, but there is no data presented on these metrics. I think the authors should include uncertainty maps. The lack of uncertainty maps makes me wonder what exactly was integrated with NRAPs framework.

Page 6, we include (new) Figure 5 to show the mean and standard deviation evolution of the permeability ensemble after each assimilation step.  
  
13. Page 7 line 34: increasing the error by that match and not seeing an impact one estimation, means that the data is not really what drives the estimation or that the noise added was not really that impactful. Did the authors add random noise at 10% of saturation? That means that a saturation that was truly 0.4, would have been 0.44 instead or was it a random iid noise? More explanation needed. The “noisy dataset” should be shown in an appendix.

We added random iid noise by x% to the observed data (saturation maps), and used those to assimilate the model ensemble. We believe that the addition of noise to the observed data is not impactful toward the estimation since it is sufficiently robust to predict the permeability map from the spatial saturation maps data.  
  
14. Figure 7: it is impossible to discern differences. A quantitative metric of data match should be used. The lack of difference makes me think there is an issue with the implementation.

Included Figure 8 to show the average MSE of the ensemble vs. ground truth model at each assimilation step. Also included sentence (Page 7, lines 19-21) to show the average mse of ensemble vs ground truth.  
  
15. Page 8 line 47: As noted previously, the authors should tone down the novelty of their approach and more thoroughly review and acknowledge existing literature, as well as strengthen the contribution of their work by adding more information on their case study.  
Thank you for this comment. We have included more literature on previous work for spatial data assimilation, as well as including more information about the data assimilation and uncertainty quantification in our work.

Final Comments from the Executive Editor: Thanks for your submission. The subject manuscript has been evaluated by three technical reviewers (TRs). Two of these reviewers (TR#2 and TR#3) consensually recommended "Major Revisions", while the other reviewer (TR#1) suggested "Minor Revisions". Associate editor (AE) has also provided several major comments in his recommendation to significantly improve the quality of your manuscript. Please refer to the detailed comments from TRs and AE above.   
  
In summary, I concur with AE's "Major Revision" decision recommendation. Therefore you should carefully revise this manuscript to address all the comments/concerns raised by TRs together with those important suggestions provided in the AE recommendation. Also make sure to provide a point-by-point response to these comments, without skipping any, while you are submitting the revised manuscript.  
  
We look forward to seeing the highest quality version of your revised manuscript for further consideration in SPE Reservoir Evaluation & Engineering-Reservoir Engineering.