Round 1

Reviewer A

Reviewer Comments

For author and editor

Thank you very much for your paper, I enjoyed reading it and like you believe Bayesian OED is an important and under-utilized methodology in seismic monitoring. Overall, I believe this is a great paper, but I have some comments:

1) I would be interested to see further discussion around data correlations and their impact on sensor deployment. For example, if the earth model has significant errors e.g. assuming a constant velocity instead of a true 3D heterogenous velocity model, these model errors induce correlation in the data. For small sensor deployments where sensors are far apart this might not be important but as sensors are added so the average distance between stations decreases this will likely be more important. I understand that ignoring correlation is much simpler, so for developing and demonstrating this framework it is not a bad assumption but a discussion of how it could be important to consider would be useful in the text.

2) Using a discrete grid instead of the full continuous prior distribution makes sense as it really speeds up computation, but it would be good to evaluate whether your choice of the number of grids impacts the OED results at all? Unless I am mistaken, but using a grid the max KL you can get is ln(# grid points). Therefore, for networks with more sensors, your EIG might be biased to be lower than expected.

3) I felt that it would be nice to have the inference problem explicitly written out somewhere in the main text or the appendix e.g. bayes theorem and defining all the prior and likelihood terms in one place so it is easy to follow how all the pieces fit together.

4) For a minor correction, the statement "Calculating the EIG using the NMC method, while unbiased, is computationally expensive." in line 126 is not technically correct as the second term in eq 5 is only asymptotically unbiased.

5) The statement "The DN method is computationally far more efficient than the NMC method but offers an upper bound for the true EIG and therefore overestimates the EIG" starting on line 135 would probably benefit from a citation of it being an upper bound (I assume it's actually asymptotically an upper-bound since you don't have the actual covariance of the data only an estimate). Just for clarification, is it an upper bound because if you know the mean and covariance of the data distribution, the maximum entropy distribution is a Gaussian? Also, I would emphasize that you are not assuming a Gaussian distribution in the parameter space but in the data space as I originally misunderstood this as something like a Laplace approximation.

6) In Section 2.2.1 I would make it clear that even when you assume a homogenous velocity model that your model still includes the topography.

7) In Section 2.2.2 where you discuss the amplitude based method, I was wondering how you do this without having the source amplitude as one of the model parameters? For the arrival time and array data I see how you could ignore the amplitude but for this I didn't understand that. As an aside, ideally there would be some source amplitude dependance in the arrival time and array data as it will affect whether the signal is detected at all and how much measurement error there can be in the pick.

8) In the paragraph starting on line 218, you mention coherence-based methods. As future work I think it would be really interesting to think about how Bayesian OED methods can be used there. I think it is more difficult as typically these methods don't have a simple Bayesian interpretation.

9) The statement starting on line 297: "In experimental design problems, we average over many samples of the prior distribution, which makes the impact of these assumptions less severe than in a typical inference problem." I find this hard to justify, so either provide citations or remove this. From my experience improper assumptions in Bayesian OED are more impactful than in standard inference. OED guides how you collect the data so your prior and likelihood both bias the models you learn, and the data used to update the model. In some cases, this double bias causes you to never collect data that would falsify your model.

10) In the main text or appendix, I would like to see some validation of the DN method vs NMC. This could be plotting the EIG surface for each so we can visually compare them or plotting a scatter plot of the DN EIG vs NMC EIG. In the text you say it basically gives the same results because there is just an offset in the EIG computation, but it would be good to show evidence of this.

Reviewer B

Reviewer Comments

For author and editor

Thank you to the editor for chosing me as a reviewer and to the authors for an excellent manuscript. I have little domain knowledge on volcanic seismic source identification but was able to easily follow the paper, method and results. I am also impressed at the software accompanying the manuscript I have been running the code with ease.

My recommendation is to accept manuscript but I would ask for the singular revision of adding a subplot to Figure 8 where some true error metric is shown. By this I mean: how well does the posterior distribution capture the ground truth? I would suggest simply measuring the l2 norm between the posterior mean and the ground truth source location. I believe that under the assumptions of the paper, we should see Bayesian contraction of the posterior to the ground truth as the number of receivers increases. Regardless, I am interested in seeing a demonstration of the performance of this algorithm beyond the EIG and reduction in uncertainty. Both of which do not neccesarily attest to the ability of the design in accurately locating the seismic source. If indeed the algorithm is locating the sources more accurately than the two baseline algorithms then the message of the paper is made stronger.

Small comments:
-There were some small spelling error throughout the paper but I trust that the final revision of the paper will not include these.

Response to Reviewers

We thank both the editor and the reviewers for their valuable comments and suggestions, the revised manuscript is significantly improved thanks to your feedback. We have addressed all the comments and suggestions in the revised manuscript. Below we provide a point-by-point response to the comments and suggestions (The outer enumeration corresponds to the reviewers comments and the inner enumerations corresponds to the authors responses).

Editor

1. In particular, I agree with the reviewer that further exploration of how errors in the velocity model affect the results would be highly valuable.
   (a) See Reviewer A point 1 for how this was addressed in the revised manuscript.

Reviewer A

1. I would be interested to see further discussion around data correlations and their impact on sensor deployment. For example, if the earth model has significant errors e.g. assuming a constant velocity instead of a true 3D heterogeneous velocity model, these model errors induce correlation in the data. For small sensor deployments where sensors are far apart this might not be important but as sensors are added so the average distance between stations decreases this will likely be more important. I understand that ignoring correlation is much simpler, so for developing and demonstrating this framework it is not a bad assumption but a discussion of how it could be important to consider would be useful in the text.
   (a) This is a good point, and we have studied the impact of data correlations in a separate source location project and found that the impact of data correlations is not significant. This is most likely due to the diminishing return for networks with many sensors, which is where data correlations would actually make a difference. This is in line with this new paper:

      Callahan, Jake, Kevin Monogue, Ruben Villarreal, and Tommie Catanach. 2024. "Analysis and Optimization of Seismic Monitoring Networks with Bayesian Optimal Experiment Design." arXiv [Stat.AP]. arXiv. http://arxiv.org/abs/2410.07215.

   which was not available at the time of submission. We include a discussion and this reference in the revised manuscript.

Addition
Even for designs with a larger number of receivers (small inter-station dis- tances), pre- vious tests have shown that the perfor- mance of the design is not significantly impacted by the assumption of independent data (Callahan et al., 2024)

2. Using a discrete grid instead of the full continuous prior distribution makes sense as it really speeds up computation, but it would be good to evaluate whether your choice of the number of grids impacts the OED results at all? Unless I am mistaken, but using a grid the max KL you can get is ln(#grid points). Therefore, for networks with more sensors, your EIG might be biased to be lower than expected.
   (a) While we use a discrete grid to describe the prior, the prior samples themselves are not on the discrete grid points since we add an offset that is sampled from a uniform distribution between the grid boundaries.
   (b) You raise an interesting point, but even if the prior samples were sampled from the grid, there are several orders of magnitude more grid cells than prior samples needed to calculate the EIG, so it is very unlikely for this to be a problem.

(c) We have added a clarification on how the prior is sampled in the revised manuscript.

old text	**new text**
In section 3.0.2, we show how this grid can be defined and refined using readily available information about any volcano of interest.	To sample from the prior distribution, we first choose a grid cell according to its probability, and then sample a source location within that grid cell according to a uniform distribution. In section 3.0.2, we show how this prior probability distribution across the grid can be defined and refined using readily available information about any volcano of interest.

3. I felt that it would be nice to have the inference problem explicitly written out somewhere in the main text or the appendix e.g. Bayes theorem and defining all the prior and likelihood terms in one place so it is easy to follow how all the pieces fit together.

   (a) We have added Appendix D that describes the inference problem explicitly in the revised manuscript.

4. For a minor correction, the statement "Calculating the EIG using the NMC method, while unbiased, is computationally expensive." in line 126 is not technically correct as the second term in eq 5 is only asymptotically unbiased.

   (a) This has been corrected

5. The statement "The DN method is computationally far more efficient than the NMC method but offers an upper bound for the true EIG and therefore overestimates the EIG" starting on line 135 would probably benefit from a citation of it being an upper bound (I assume it's actually asymptotically an upper-bound since you don't have the actual covariance of the data only an estimate). Just for clarification, is it an upper bound because if you know the mean and covariance of the data distribution, the maximum entropy distribution is a Gaussian? Also, I would emphasize that you are not assuming a Gaussian distribution in the parameter space but in the data space as I originally misunderstood this as something like a Laplace approximation.

   (a) We have added the relevant citation and clarified that it is an asymptotic upper bound in the revised manuscript.

   (b) Yes, more generally, if the information of the evidence is estimated using any functional approximation to the evidence then a lower bound to the EIG is obtained. See:

      Foster, Adam, Martin Jankowiak, Elias Bingham, Paul Horsfall, Yee Whye Teh, Thomas Rainforth, and Noah Goodman. 2019. "Variational Bayesian Optimal Experimental Design." Advances in Neural Information Processing Systems 32 (March). https://proceedings.neurips.cc/paper/2019/file/d55cbf210f175f4a37916eafe6c04f0d-Paper.pdf.

   (c) We have added a clarification that the Gaussian approximation is in the data space in the revised manuscript.

addition
See Appendix C for a benchmark of the DN method against the NMC method. It is important to note that the DN method assumes a Gaussian distribution in the data space, not in the parameter space such as in the Laplace approximation (Long et al., 2013), and avoids using local gradients of the likelihood to approximate the posterior distribution.

6. In Section 2.2.1 I would make it clear that even when you assume a homogeneous velocity model that your model still includes the topography.

   (a) While the receivers will be restricted to the surface, the homogeneous velocity model does not take the topography into account, since this would require a 3D ray tracer to cover all cases. As long as the topography is close enough to a convex set this should not be a problem, but we will clarify this in the revised manuscript.

addition
The methods used for heterogeneous velocity models are also necessary if the topography deviates substantially from a convex set, in which straight rays between source and receiver might intersect the topographic surface, which is not accounted for in the homogeneous velocity model.

7. In Section 2.2.2 where you discuss the amplitude based method, I was wondering how you do this without having the source amplitude as one of the model parameters? For the arrival time and array data I see how you could ignore the amplitude but for this I didn't understand that. As an aside, ideally there would be some source amplitude dependence in the arrival time and array data as it will affect whether the signal is detected at all and how much measurement error there can be in the pick.

   (a) We can ignore the source amplitude since:

      i. It will be the same factor for all receivers, so it will not affect the relative data measured at the receivers.

      ii. We do not have an "observation uncertainty" for the amplitude, but instead assume that the uncertainty in Q and the travel time/distance are the main contributors to the uncertainty in the amplitude. The propagated uncertainty term is then again scaled by the source amplitude, so the relative uncertainty in the amplitude is the same for all receivers.

   (b) It would, of course, be ideal to include the source amplitude, and also include an observation uncertainty but that would then also involve the need to estimate a "receiver term" which takes the local geology into account. While we could include all of those quite straightforwardly in theory, getting accurate estimates on all those terms would be quite difficult in practice.

   (c) Yes there is a source amplitude dependence in the arrival time uncertainty. For an example of how to model this see for example this paper:

      Fuggi, Antonio, Simone Re, Giorgio Tango, Sergio Del Gaudio, Alessandro Brovelli, and Giorgio Cassiani. 2024. "Assessment of Earthquake Location Uncertainties for the Design of Local Seismic Networks." Earthquake Science 37 (5): 415–33.

which would be quite straightforward to include in a regional scale study, where heuristic formulas mapping magnitude to amplitude and noise levels as a function of station location might be available. In a volcano setting, however, there are a number of problems

   i. No heuristic formulas mapping magnitude to amplitude are typically available.
   ii. The source mechanisms and therefore magnitude distribution might be quite different from volcano to volcano.
   iii. The noise levels at stations might change rapidly with station location due to exposure to many environmental factors.

(d) In light of this we implicitly take the approximation here that all events we observe have a sufficient signal-to-noise ratio in which the picking uncertainty is more or less constant from station to station. Even if this is not fully accurate, the governing term for the travel time uncertainty in the absence of a (detailed) 3D model is typically the dominant contribution of the uncertainty in the velocity model. The picking uncertainty mostly determines the cut-off at which no information can be gained when the signal-to-noise ratio is below 1, which we do not cover here.

(e) We have added a clarification about this in the revised manuscript.

addition
By using a fixed picking uncertainty term, we implicitly assume that all arrivals have a sufficiently high signal-to-noise ratio, and every event can be observed at every station with the same uncertainty. Typically the velocity uncertainty term is much larger than the picking uncertainty term in a volcano setting, making the exact value of the picking uncertainty term relatively less important.

8. In the paragraph starting on line 218, you mention coherence-based methods. As future work I think it would be really interesting to think about how Bayesian OED methods can be used there. I think it is more difficult as typically these methods don't have a simple Bayesian interpretation.

   (a) While the data processing is more involved than classical travel time picking, the coherence-based methods are still basically a way of calculating (differential) travel times. As long as the velocity model uncertainty term is the dominant term in the uncertainty, the Bayesian OED methods should work just as well as the arrival time-based methods. Of course, if the data processing is more involved and leads to a larger uncertainty in the data, the uncertainty term might need to be adjusted.

9. The statement starting on line 297: "In experimental design problems, we average over many samples of the prior distribution, which makes the impact of these assumptions less severe than in a typical inference problem." I find this hard to justify, so either provide citations or remove this. From my experience improper assumptions in Bayesian OED are more impactful than in standard inference. OED guides how you collect the data so your prior and likelihood both bias the models you learn, and the data used to update the model. In some cases, this double bias causes you to never collect data that would falsify your model.

   (a) We have removed the statement in the revised manuscript.

   (b) While we fully agree that the prior will heavily affect the resulting designs, this statement was written mostly thinking about mis-specifications in the covariance of the observed data and the velocity model, which doesn't seem to affect the optimal design as much as they do the EIG and posterior solutions. But we fully agree that this statement was too strong.

10. In the main text or appendix, I would like to see some validation of the DN method vs NMC. This could be plotting the EIG surface for each so we can visually compare them or plotting a scatter plot of the DN EIG vs NMC EIG. In the text you say it basically gives the same results because there is just an offset in the EIG computation, but it would be good to show evidence of this.

    (a) Appendix C in the revised manuscript shows a comparison of the DN and NMC methods.

Reviewer B

1. My recommendation is to accept manuscript but I would ask for the singular revision of adding a subplot to Figure 8 where some true error metric is shown. By this I mean: how well does the posterior distribution capture the ground truth? I would suggest simply measuring the L_2 norm between the posterior mean and the ground truth source location. I believe that under the assumptions of the paper, we should see Bayesian contraction of the posterior to the ground truth as the number of receivers increases. Regardless, I am interested in seeing a demonstration of the performance of this algorithm beyond the EIG and reduction in uncertainty. Both of which do not necessarily attest to the ability of the design in accurately locating the seismic source. If indeed the algorithm is locating the sources more accurately than the two baseline algorithms then the message of the paper is made stronger.

    (a) We have added a subplot to Figure 8 in the revised manuscript showing the L_2 norm between the posterior mean and the ground truth source location. We have also added a discussion on this in the revised manuscript.

addition
In addition we can also examine the average L2-norm distance between the mean of the posterior distribution and the true event location as a function of the number of receivers. This shows that the mean moves close to the true event location after only three receivers have been added and improves only slightly with the addition of more receivers.

2. There were some small spelling error throughout the paper but I trust that the final revision of the paper will not include these.

    (a) We have gone through the manuscript with an eye on this. Please see the revised manuscript for the corrected version.

Round 2

Reviewer A

Reviewer Comments

For author and editor

Thank you very much for your thoughtful additions to the manuscript in response to my comments. I think this is an excellent paper. As an aside, I am really interested in learning more now about the Dn method. It looks like a great approximation from what we see in Appendix C.

Reviewer B

Reviewer Comments

For author and editor

Hello, the authors have responded to my requests for revision. I recommend the acceptance of this manuscript.