Dear Stephen Hicks,
Thank you and the reviewers for the comments back on our previously submitted manuscript. We appreciate the time and effort it takes for the reviewers to read through this and suggest feedback on our current submission. We have addressed all the comments made by both reviewers in the paper, and have included a list below here on the changes that we have made for each point, and a reason for anything we felt did not need to be changed that may have been previously suggested. Please reach out to me if anything is not clear, or if anything needs further clarifying.
Best Wishes,
Jade Eyles

Stephen Hicks:
Based on reviews I have received, your manuscript may be suitable for publication after some revisions. Both reviewers find that your approach is quite a novel one and offers a more accessible and understandable advance over conventional ways of quantifying earthquake location uncertainty (e.g., root-mean-square residuals). The manuscript is well-written and the figures are pretty clear. However, the reviewers have concerns over the naming of the "trusted" parameter and its benchmarking. When reading the manuscript myself, I was initially confused by whether the "trusted" parameter relates to real individual earthquakes - or to the velocity model and location set-up - I think it is the latter, but this should be made clearer. Maybe then, for example, the title should say "earthquake location methods" instead of just "earthquake locations"? I also agree with Reviewer A about the treatment of station correction terms. To make your study more generally applicable (as per Reviewer B's comment), I would be interested to read a concise summary of the different location approaches taken by volcano observatories around the world - not just MVO - I should imagine that there is some existing literature on these.

I have added a paragraph in the introduction of different methodologies currently used at observatories (line 53-62), and addressed the comment on why we did not use station corrections on lines 156-158. I don't believe the title should change to earthquake location methods being improved – this is because we are aiming to improve the earthquake hypocenters themselves using different methods/velocity models. And we are not trying to make a new location method per say, but use existing methods to help refine previous seismicity, and hence have not changed this title to reflect that.

Reviewer A:
In this paper the authors present a methodology for assessing the reliability ("trusted") of earthquake locations based on whether or not, for synthetic locations, the spatial location error is less than the calculated error of the hypocenter location. The authors use this trusted criteria to select events in order compare the standard location algorithm setup for Montserrat with 3 setups with other algorithms. Using relocations from the standard procedure and the optimal algorithm, the authors re-analyze and re-interpret Montserrat volcano-tectonic seismicity in 1995. I have a few comments and also indicate minor issues.
1. I find this work to be notable and valuable in basing the "trusted" criteria on a comparison of reported location error with true error for a thorough set of synthetic

locations. Reported error and true error are fairly independent, output quality measures of a location algorithm, and not input-dependent measures such as azimuth gap and station distance. I also think use of such output measures is also more objective than analysis of RMS since RMS is explicitly or implicitly the value minimized in most location algorithms and so not an independent, output measure of location quality. But it does seem possible that the configuration which as the most trusted events is not necessarily that with the most hypocenters closest to the synthetic hypocenters. It would be informative to have analysis, discussion and figures of the performance of the different configurations with respect to the distance of the calculated hypocenter from the original synthetic location, independently from the computed location error.

I have added an additional image into the Electronic Supplement (S9). This image shows the vector representing the change in location of the earthquake (from its original synthetic location to its relocated location) by each location method. This is in additional to the tables which show the data of the changes in location, and errors for each method independent of the trusted locations. I have added a sentence on this additional figure in the Results section (line 260), and added some more information on changes in location/error making sure this is clearly linked to the tables in the supplementary.

2. Another general issue is that there are other well established ways to improve locations of seismicity, in particular 1) station corrections, 2) inversion for "optimal" 1D velocity model 3) inversion for 3D velocity model, amongst others. You cover 2) and 3) by examining 4 different models, but only allude to 1) in the Conclusions. I would suggest adding some material to the Introduction on these issues as background, and elsewhere (Methods? Discussion?) on how you used them, or why you did not, and how they might help with future locations study at Montserrat or elsewhere.

I have added a comment in the introduction on lines 72 -78 on how these can be used to improve locations, and added a comment in the methods section on why we did not (lines 156-158)

3. Line 116: "Synthetic P and S wave arrival times were calculated from each synthetic point in the grid to all active seismic stations using each velocity model tested. Synthetic arrival times were calculated using the Eikonal finite-difference scheme of Podvin et al. (1991)." – Does this choice give any advantage to NLL over Hypocenter in your comparisons since NLL uses Podvin Lecomte as a travel-time calculator and, presumably, Hypocenter uses a different, analytic calculation valid for layered models?

My understanding is that Hypocenter follows the method from Hypoinverse (Hypocenter is itself a combination of HypoInverse and Hypo71). Hypoinverse takes in a travel-time table that is independently created from the program, but they suggest using TTGEN in the supporting documentation. An issue with this is that it only works up to zero depth, and therefore the velocity model and outputted earthquakes need to be shifted to represent this. But the documentation also has said the user is able to upload their own travel time grids, and so I believe that this should still be fine to upload the same travel time grid from both. I have added a sentence to highlight this in the text on lines 141-143.

4. Line 135: "The calculated arrival times included a gaussian pick error of 0.1 seconds" – do you mean the synthetic times had a random gaussian shift of 0.1 seconds added before being used as picks? Or the calculated travel-time error used by the location procedure is 0.1?

Yes, I meant that the arrival times (synthetic) had a random gaussian shift of 0.1 seconds is incorporated into the travel time calculation when generating arrival times. I have updated this to make it clearer on lines 174-176.

5. Line 139: "A range of pick errors were tested" – do you mean here the nominal pick uncertainties specified to the location programs?

No, this is the same as above, in that a random gaussian shift of 0.1-5 seconds is incorporated into the generation of the arrival times. I have updated the text to make this clearer on lines 178-182.

6. Line 236: "Method 2 and 4 outputted consistently low hypocenter errors." – Do you mean the errors reported by the location algorithm? NLL GAU gives generally lower location error than NLL EDT since EDT finds a larger and more complicated PDF when there are outlier data. EDT should usually find a less biased location but give larger formal location errors than GAU when there are outlier data.

I have hopefully made this clearer by adding the additional lines 263-266.

7. Line 332: "The MVO setup does produce a greater number of hypocenters when filtered for events less than 3 km, but synthetic testing has proven these to be less trustworthy." – since the absolute error statistics between the two setups is not comparable, and there is indication the MVO locations are less trustworthy, perhaps it is valid to select the same number of events from each setup for comparison, after ordering the events by some error measure(s).

This is a good idea, but we felt it was not necessary to include this at this point, as the synthetic testing had shown that most MVO locations had low confidence, and therefore it is better to use the new locations with the optimal setup.

8. Figure 4: The filtered MVO locations show greater concentration in depth (though along a wider and more acuate zone) than do the optimal setup hypocenters. Could this arcuate zone of seismicity reflect a real volcano-tectonic feature such as a caprock surface, or is it more likely an artefact of error/bias in the original MVO location procedure? Is it possible to do an analysis as in Figure 4 but with many more events over a longer period of time? This could perhaps show more convincingly better performance of the optimal setup.

We did test this over the 25 year catalogue and found that the optimal setup preformed better over this whole time frame (in terms of errors and number of events location). The overall trend of earthquakes looked similar, although we did note false artefacts of clustering of events for locations with few picks (we hope to publish this paper later).

The events that were removed for the optimal setup due to too large hypocenter errors generally included events with picks less than 8 stations. Comparatively, the MVO setup found these locations to have smaller errors and so were included in the plot, but we believe these are less trusted due to the testing, and small number of picks. I have added a sentence into the text on this on lines 368-376.

Suggested minor issues to address:
1. Line 43: "combination of arrival time pick error, velocity model and location algorithm" → "combination of arrival time picks and uncertainties, velocity model and travel-time uncertainties, and location algorithm" or similar.
Completed change above
2. Figure 1: The box in the thumbnail is clearly much larger than the area of the main Montserrat map panel, maybe reduce this to a small box that better corresponds to the position and area of the main Montserrat map panel?
Thank you for noticing this! It has been updated in Figure 1.
3. Lines 119, 122: "Podvin et al." → "Podvin and Lecomte"
This has been updated
4. Line 130: "10 different seismic configurations from 1995 to February 2020" → "10 different seismic station configurations available from 1995 to February 2020" or similar.
This has been rewritten on line 151
5. Paragraph starting Line 145: "Trusted" is used in the preceding paragraph, it would be better to put this paragraph before the preceding one.
I have updated the paragraph order in the methods to account for this
6. Line 199: "4D probability density function" → "3D, spatial probability density function"
This has been updated
7. Figures S3 and S4: I suppose these results are obtained with only the standard MVO configuration, since the others have not yet been introduced in the main text. Perhaps best to confirm the configuration(s) in the main text and captions.
The text has been updated in the caption – this was calculated using NonLinLoc with the Rowe velocity model.

Recommendation: Revisions Required

Reviewer B
In this manuscript, authors compute synthetic earthquakes and compare them to the seismic catalog from Monserrat Volcano Observatory to test for several location methods and seismic velocity models. They then introduce a parameter of confidence and chose the best couple of location method / velocity model to relocate a part of the MVO seismic catalog. From these relocations, they re-analyze the previous propositions about dyke orientations that could impact the understanding of the volcano dynamic during the 1995 eruption.
This is an interesting paper, but it seems to miss some important discussions about the limits of the proposed method, and maybe a comparison with other relocation methods used at other volcanoes. In my point of view, this work is an important first step for the improvement of operational catalog, but not enough global to propose a new

methodology. Overall, I am not convinced by the dyke propagation interpretations, as lineation are really unclear on the maps and the dataset used is quite small.

I have few suggestions and questions that are listed below:
1) Title: I think that the first part of the title is unclear, and the paper does not provide the answer. It is still a big deal for seismologist to improve the earthquake locations, and without a perfect knowledge of the internal structure of a volcano it is difficult to prefer one location to another since they are based on models. I suggest to use 'improving' earthquake locations.

Thank you for this suggestion – I have updated the title to better reflect the paper: *Can Earthquake Locations be Improved?* ..

2) Abstract:
Line 19: I suggest to replace the term 'trusted' in the manuscript by 'preferred' or to introduce a 'confidence parameter' instead of a parameter 'trusted'.

I liked the idea of calling this a confidence parameter, but felt this did not fit overall in the whole text when referring to a subset of located synthetic events. Instead I hope I have made this clearer in the abstract by saying that we introduce a confidence parameter that we then use to attain trusted events.

3) Lines 99-112: There are repetitions with the introduction. Maybe it would be interesting to have an overview of other existing methods and the presentation of the actual method used by MVO.

I have added a paragraph on different existing methods used at observatories on lines 53-62. I have also added a sentence about the current methodology used at MVO on lines 89-91. I have removed repetitions between the introduction and the method: this included the sentence on MVO not updating its methodology, and on trialing different stations from 1995-2018.

4) Line 115: Why 5.8 km depth? Is there any known structural interface?

This was done to test a velocity model being used which only extended to 6km and for which we wanted to test how it operated near the edge of the model (added to the text on lines 130-131).

5) Line 118: 'using each velocity model tested'. Existing velocity models for the area were not presented yet, so it is difficult to understand what you are really doing to calculate the synthetic travel paths. Did you compute a set of synthetic travel times for each velocity model? Did you use the topography of the volcano?

I agree that the velocity models used have not been presented yet, but from the abstract and preceding paragraph I believe it is clear that I will be testing different methods and velocity models. I have tried to make this clearer by updating the sentence on lines 131-135.

6) Lines 130-133: Maybe it should be necessary to have more materials to illustrate these tests (maps). What does 'performed similar to each other' mean?
Which parameters or results are compared?

I have added another figure in the supplementary (S3) to show an example of changes in location for X and Y locations for the synthetic earthquakes for all station configurations tested. We have added the following text on lines 151-156.

7) Line 141: The definition of what is a 'trusted' earthquake is not yet defined.

I have shuffled the paragraphs around so this is no longer an issue – thank you for noticing.

8) Line 150: typo error (calculated).
Updated – thanks for noticing.

9) Lines 189-191: I think that other observatories will be use models available for their interest areas as each volcano have a specific structure.

Changed observatories to MVO – this was phrased incorrectly, and had meant that observatories would choose velocity models that were openly available. Thank you for highlighting the confusion in this sentence.

10) Line 296: Replace S10 by S11. 'Similar distribution between both dataset': It seems on the graphs that the RMS with the Roman et al. picks are slightly better. I think you should discuss that.

Updated wording to show that the RMS is better with Roman et al and that this is likely due to higher pick quality – lines 326-328.

11) Lines 363-364: Add a reference.
Added two references for this on line 423

12) Line 394-395: In which direction? It is not obvious on the Fig. 6 that there is a possible migration through time. The points are scattered both in depth and laterally.

This would be to the WNW. We agree that this is not as clear as we would like for this paper when only using high quality hypocenters, and as such have removed that there is a migration through time, and that this would benefit from further analysis. We believe there is still a clear lineation of seismicity to the WNW which could be representative of dyke propagation.

Figure 1: The inset does not correspond to the area of the main map. For the plate motions arrows, please add the velocities. For the colored regions, it will be easier to insert a legend directly on the main map. What about purple and brown areas?
Updated figure 1. I added in description of coloured regions in legend as felt this was not needed in the main map.

Figure 4: Add the number of events in each panel.
Added these into the figure 4

Figures 5 and 6: Add errors bars for your relocations.
I have added error bars for the relocations in supplementary S15 as I felt this was too messy to include in the main figure.

Figure 7: The black dashed line has not the same orientation than the one presented on Fig. 6.
Thank you for noting – I have updated this so that they are now the same.

Figure S1: The map lacks a grid, a scale bar and the depth information in the legend is missing. Maybe you can merge Fig. S1 and S2.
I decided not to merge these together as it looked quite messy – but have remade Figure S1 to be more clearer with the grid, and also the depth information included in the legend.

Figure S3: A lot of informations are missing in the legend (error bars, black circles, blue boxes, etc.). The RMS is in seconds, not in km.
Which method and velocity model were used for this comparison?
I have updated the legend to explain the diagram, updated RMS to seconds, and added description of which velocity model/method was used for comparison.

Figure S4: It is clearly not obvious that the number of 'trusted' events increase with the pick error (green line). You should discuss that point.
We decided to remove this as felt the graph was not representative of what we were a)trying to show and b) the results. We believe that the outputted errors shown in Figure S4 better represent this. The reason for an increase in depth trusted events is because the depth has a smaller range (in km) than in latitude and longitude location – hence it is easier to incorporate the original location with a larger error.

Figure S5: You should try to present the 3D Paulatto model.
I did not want to include an image of the 3D model, as this can be found easily with the related paper. I could represent this as a 1D model, but felt this would raise confusion as it would not be fully representative over the whole island.

Figure S6: Method 3 with V3 seems to be the best combination. The majority of the events used here are above 5-6 km, why do not try this one? Also, you could merge V3 and V2 for the deeper part.
I agree that this is the case that for under 5km, V3 was better, but produced very similar percentages to V2 when using method 3. V2 was higher 6 times, and V3 9 times so there is not a clear difference. We also looked at changes in location, calculated errors, RMS etc as well as the number of trusted events. Overall, because V3 did not extend greater than 6km, we used V2. Most of this seismicity is under 5km during this time frame, but the overall aim was to make this applicable to MVO to use moving forward, where greater depth events could be recorded. We did think about combining two velocity models, and agree this is something that could be useful for future work. However, in this paper, we

wanted to highlight how a user could trial pre-existing velocity models, and did not want to make this process complicated by creating additional velocity models, or extending current ones.

Figure S7: In the RMS table, the method 3 produced 0.08 in all cases. How did you choose the best?
Values are rounded to 2 decimal places, but are different values, and so were picked on the lowest before being rounded. I have added this into the legend.

Figures S9 and S10: Describe DX, DY, DZ and EX, EY, EZ in the legend.
Updated legend to explain this.

Figure S12: Why seismic network is different from Fig 1? Please check both figures and legends.
Thank you for noticing this – the missing station has now been added

Recommendation: Resubmit for Review
