Thank you so much for the valuable comments. We will try our best to take account of your directions. Here are your comments in Italic followed by our replies.

The inversion method itself seems improperly chosen. Much computationally simpler approach providing mathematically the same result can be formulated considering that the Bayesian approach is "additive" in terms of applying constraints. In particular, the authors could use the ISOLA-ObsPy code (which they use anyway) and omit all results dissatisfying the polarities and DC percentage limit. This way, they could avoid the nested MCMC approach to sample the posterior. My proposed approach assumes that the centroid location is sought on a predefined grid of points for which the GFs are precalculated.

Reply: We believe the inversion method is worthy of attention in the framework Bayesian analysis, especially in combining Metropolis-Gibbs sampler of Lomax et al. (2000) and incorporating the variation in location (Gu et al. 2018); and it is useful in sparse broadband networks with available short-period stations mainly for small earthquakes. Please also note the results with one- or two-station configurations. The method applies the constraints to each trial moment tensor during the random walk to explore likelihood regions which, to the best of our knowledge, is different from checking the final closed form least squares solution of ISOLA-ObsPy and other methods with polarities and DC percent. We have only used the uncertainty analysis part of ISOLA-ObsPy code. A part of the Python code was converted to C language because the code is in C and also in Fortran. The assumption about the precalculated GFs is right for the manuscript computations but the code only reads random GFs during its search for optimum location. On the fly calculations can also be simply employed to reduce the time of computing GFs.

The structure of the paper should be simplified. All methodological details should be given in a single (Method) section. In the present version, the coarsening method (also called tempering) is introduced in the middle of the discussion of the results, which is confusing. Moreover, many things regarding the calculations are missing. For example: what is the sampling step of the centroid location, are the Green's functions precalculated or calculated on the fly? how are the frequency ranges determined?, what Cd is used when synthetic tests with no noise are considered?, is the coarsening used only for determining the centroid location?, how long is the burn-in period?

Reply: We corrected the manuscript according to your directions and added the parts you mentioned. Sampling step of the location is 1 km and the frequency ranges are determined by trial to obtain maximum variance reduction for Malard earthquake. For Sargans event we used the same frequency range as in the example of ISOLA-ObsPy so that the results would be comparable, yet increasing the range did not improve the variance reduction significantly. In the case of no noise calculations, we used the same Cd as in the noised tests, and set SNR to a very high value. The coarsening is just used for the location and we didn't use any burn-in period in the calculations, although it could be used in the code.

The results and conclusions of the synthetic tests are not laid out clearly. How do the polarities help to the inversion? Are there any critical lessons learned that are then considered in the real data applications?

Reply: We will try to explain them more clearly in the manuscript. Polarity or waveform inversion alone cannot led to reliable moment tensor solutions for small events, so waveforms and polarities should be inverted together. For small earthquakes the recording stations with high SNR are limited, so polarities are used as added information.

At line 236 (just before the Conclusions section), something about "one polarity" is mentioned, which is not discussed in the synthetic tests at all. Generally speaking, the synthetic tests and the real data applications must be made more coherent.

Reply: That is explained in relation to real earthquake data inversion. In the case of using single or two-station configuration, sometimes we get reverse polarity for the mechanism when using no polarity data. We mentioned that in those cases, using only one polarity can lead to similar estimate as in the all-station solution. In the synthetic tests we only showed a two station configuration (LIENZ + SGT04) with different SNRs as an example.

Replies to the annotations in the manuscript

We will correct the manuscript and explain what is missing in more detail. Here are the replies to some of the questions raised in the annotations.

Line 85: "… -1.5 to 1.5 …"

What are those values? Moreover, units are missing?

Reply: They are the uniform prior range for moment tensor and the unit is Nm. According to the other reviewer's comment we should have used ±sqrt(2) in the trace and ±1 in the off-diagonal elements. We will recalculate and revise the manuscript using these numbers.

Line 109: "… $d\boldsymbol{m}$ ..."

What exactly is ment by this?

Reply: It is the vector difference of $\boldsymbol{m}_{new}$ and $\boldsymbol{m}_{curr}$. But as you said the symbol is confusing so we removed it.

Line 128: *It is unclear whether CT by Hallo and Gallovic (2016) is actulally used or not. Any specific settings apply?*

Reply: CT was used and calculated using saxcf subroutine. For Sargans earthquake the width of the uniform time shift distribution, L1, is 3 samples (1.6 sec) for all stations. In the case of Malard earthquake that is between 2 and 6 samples (2.4 and 7.2 sec) for different stations. The width is computed by using a linear dependence on source–receiver distance, following values for standard deviation of velocity model perturbations equal to 10 percent. We added the explanations to the manuscript.

Line 161: "… for two-station cases we fix the location …"

why? If generally suggested, it should be explicitly stated.

Reply: When we used two stations, the target location could not be retrieved, although the solution was near to the target one in case of high SNR.

Line 170: "… the accepted Metropolis locations are drawn by green cubes …"

Completely unclear. I guess all the locations are at some point accepted by the Metropolis criterion. This applies also to other appearances of the term "accepted".

Reply: We will elaborate on that in the manuscript. The next Metropolis location sample is obtained by the adaptive Metropolis algorithm of Haario, et al. (2001), then the Metropolis acceptance criterion is applied on its likelihood, at this stage the location is either accepted or not.

Line 178: "… Equation 3…"

1) This is not compatible with Eq. (1). Refer to a particular equation which is this related to, and use the same symbols - L_A in this case?

Reply: It seems to us that, this definition is the same as the definition on page 1969 of Gu et al. (2018), i.e. the coarsened marginal likelihood given x. That is used in the first chain to search for the location. Tempering is just used in the first chain to find the optimum location, so $\boldsymbol{m}$ i.e. elements of elementary seismogram are absent in the definition. For example please compare Equation 15 and Equation 23 of Gu et al. (2018), they have used the same symbol P for both of the equations which is actually our σ.

2) This approach, also called tempering, if applied to the likelihood function corresponds to changing the data covariance matrix (Cd is multiplied by gamma). This in turn means that by changing gamma one tests various values of the data error.

Reply: Yes, larger γ and lower SNR both causes higher false "error" in the location. That does not affect the moment tensor.

Line 192: "… For this case this optimum value is 50 …"

Does this means sqrt(50) times larger data error should be considered? Please elaborate more on this.

Reply: We don't understand your meaning of sqrt(50), but yes, higher γ leads to larger scatter of location points, so the "error" spuriously increases.

Line 204: "… There are six accepted locations with the increasing likelihood out of 1000 tested locations.…"

Again why just six? All others were not visited anytime?

Reply: Yes, Not all, but many are visited, they are proposed sources, but after the Metropolis acceptance criterion, we have the accepted locations and then there is a conditional statement to only pass the locations with higher likelihood than the previous high likelihood (in this case, they are six in number), then the GFs of that location are used in the next chain to retrieve moment tensor.

Line 219: "… and inverted in the displacement domain …"

Needs rewrite - perhaps the authors want to say that they integrate to displacement, or use displacement waveforms in their inversion.

Reply: Yes, we use displacement waveforms in the inversion. We will clarify this.

Line 220: "… Figure 5 shows the tested locations for 1000 iterations. …"

It does not agree with the numbers in the figure caption.

Reply: Yes, here there should be 2000 iterations.

Line 230: "… As an example, all visited focal angles and accepted solutions with higher likelihood for LIENZ and SGT04 stations inversion are shown in Fig. 8. In two-station calculations, the location is fixed to the estimated value of all-station result, therefore Fig. 8 contains less visited sites. …"

What is the purpose of this text?

Reply: We deemed it good to show the random walk for a case of two-station configuration because the method is especially beneficial when only small number of broadband stations are available mainly for small earthquakes. However we will remove all random walk figures in the space of focal angles as your direction.

Line 246: "… Although the two-station no-constraints DC solutions are better in terms of Kagan angle but deviatoric solutions deteriorate. …".

Not clear. How can the authors say which solution is better in case of real data?

Reply: We will clarify the sentence. The comparison is made with all-station solution, with the thought that the all-station configuration gives better result due to containing more information.

Line 264: "… stations are filtered to frequency ranges 0.04-0.17, 0.04-0.08, 0.055-0.085, 0.055-0.085 and 0.055-0.08 to gain better waveform fit …"

how were those ranges determined? For the Sargans earthquake a single frequency range was sufficient?

Reply: We determined them by trial to gain higher variance reduction. For the Sargans earthquake we tried different values for different stations, but variance reduction didn't improve substantially, so we kept the range used in ISOLA-ObsPy so that the results would be comparable also.

Line 266: "… We also determine two- and one-station solutions for this event. The results for this event show the advantage of the constraints of polarity and DC% again. Of course, for all the cases, only one polarity is enough to constrain the solution to the optimum solution. That is except in the case of using the single station of VRN, in which more polarity constraints are needed for better compatibility with all station solution. …"

One polarity is sufficient? This is neither shown nor discussed!

Reply: Our paragraph was not clear. For those cases only the compressional and dilatational quadrants of the bechballs are reversed and one polarity is enough for correcting the place of the quadrants. We changed the paragraph to "We also determine two- and one-station solutions for this event with and without 21 first motion polarities (Table 5). The results for this event show the advantage of the constraints of polarity and DC% again. For all the cases of not using the polarities, only one polarity is enough to constrain the solution to the optimum solution. That is except in the case of using the single station of VRN, in which more polarity constraints are needed for better compatibility with all station solution."

Line 375: "Table 1"

- True (target) parameters should be added to the first row of the table. - Location error should be added.- Do the focal mechanisms include the uncertainty of the location? - The same applies also to Tabs 2 and 3.

Reply: We have added the first row. About location error we could not think of a way to set γ. The selected γs discussed in the manuscript is just to make sure about obtaining better estimate for the source location, i.e. where the curves of source location range versus γ does not show trends as illustrated for example in Figure 12 of the manuscript.

Line 433: "… Cubes and squares show proposed locations colored according to their iteration number and sized in keeping with the likelihood value, that is, largest cubes and squares indicate greater than 1% maximum a posteriori location, etc. …"

What does it mean?

Reply: We will clarify, it is confusing. The likelihood values are shown by square sizes, that is, maximum aposteriori of greater than 0.01 distance from 1 have the largest symbols and maximum aposteriori of 0.01 to 0.1 distance from 1 have smaller symbols, etc.

Line 435: "… There are seven accepted solutions with the increasing likelihood that their paths are shown by green lines, reaching to the input location. …"

Unclear sentence? What is input? Which seven? What about the others? Note that the same applies to Fig. 5.

Reply: By "input location" we meant target location of the synthetic test, the input to create synthetic seismograms. We will change it to "target location". We hope that by the reply to line 204, we could have answered to your other questions. We will add more descriptions to the manuscript.

Line 447: "… Source location and shaded error bars versus γ…"

This is averaged over Markov Chains? After the burn in? More explanation in the Method section is needed.

Reply: We will do that. Yes, we calculate all accepted sources for each γ, eliminate burn-in samples and compute the average.

Line 454: "… Figure 4 …"

- What is the purpose of this figure? - It seems the the accepted solutions include both left-lateral, right-lateral and normal solutions, which however does not correspond to what is shown in Tab. 1. This needs to be resolved. - If there is no strong need for Fig. 4, I suggest its removal, together with Figs. 7 and 8. I think all is said by the beach balls.

Reply: We wanted to show how random walk behave in the method, especially comparing the number of visited points with all points in the space visually. Sampling points can be any point in the space of focal angles. MCMC tries to find an optimum one which is shown by large green circle and is present in Table 1. We will remove the figures as you suggested.