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

In general, the article is hard to understand due to countless language style and formulation issues and needs significant copy-editing. Thus, maybe many important aspects of the work got "lost in translation", but for me the authors did not present clear enough what is the novelty and strength of their method. To do so, they could compare test results of no-constraints inferences with results where they used these constraints.

Reply: We will try to improve the readability of the manuscript. The work does not present a big novelty; we use Bayesian formulation but methodical details differ from similar previous studies. In particular we combine Metropolis-Gibbs sampler similar to that of Lomax et al. (2000) applying the constraints; and incorporate the variation in location as in Gu et al. (2018). As to the comparing the no-constrained with constrained results, we think, we have done that in the tables 2 and 3 for the synthetic tests and the tables 4 and 5 for the applications.

The overall structure of the article is unfortunate and the authors jump back and forth between describing sampling algorithms in the results section and the methods. Also previous (published) work is discussed in the method section as well as in the introduction.

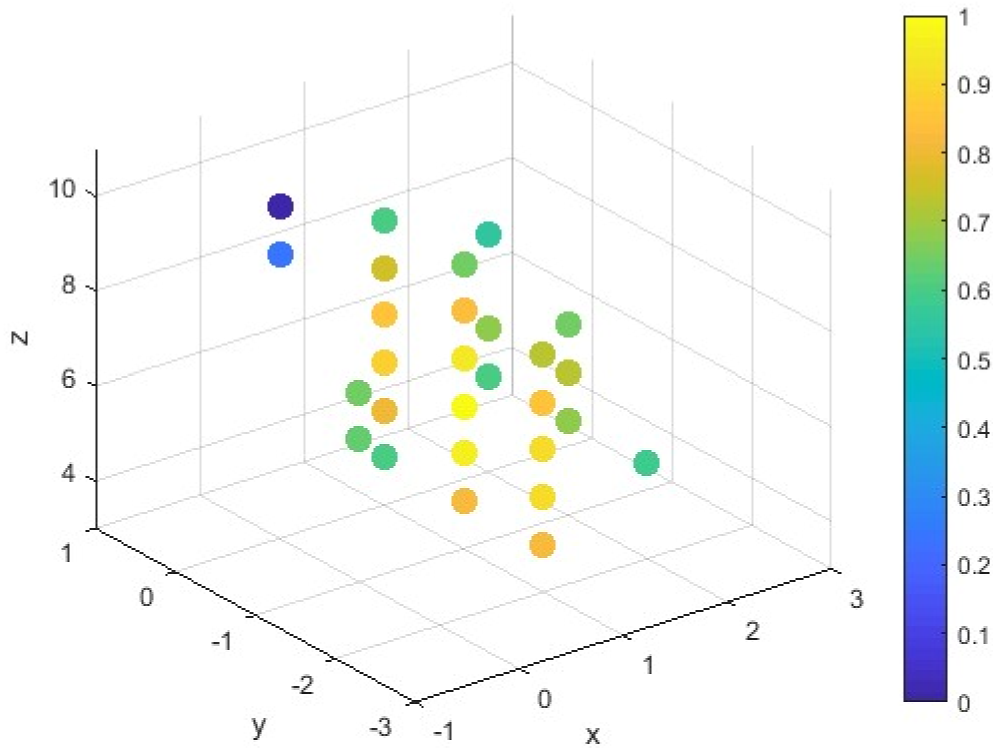
Reply: We transferred all methodological discussions to the method section. We will make the manuscript complete and readable.

Also the authors have to be careful with established terminology in the literature (e.g. Metropolis test- is called the Metropolis acceptance criterion) as well as mixing of data and theory errors and how they map to parameter uncertainties. This causes many sentences in the manuscript to make no or little sense in their current form.

Reply: The terminology was changed. About the errors we tried to make it clear that we have used theoretical errors and explained the settings applied.

One BIG major issue is in the section when the authors introduce the "coarsening" concept- also called tempering in published literature (used by several MCMC sampling algorithms such as Parallel Tempering/ Replica Exchange- or Sequential Monte Carlo). The tempering parameter (gamma) in the manuscript has to be = 1 if the obtained samples are supposed to be from the posterior probability density (PPD)- if this is not the case the samples are not from the PPD and estimated parameters and their marginals are meaningless. In published literature the tempering is used in order to help exploration of the Markov Chain(s) to avoid getting stuck in local minima in the solution space, but MUST be switched of once one wants to sample from the PPD. Thus, the presented results of the synthetic test and the real data cases needs to be revisited and revised- apparently the authors also used Gamma = 1 as indicated by figures 3 and 6, but maybe there the Markov Chains need to sample longer.

Reply: γ is used in the context of the paper by Gu et al. (2018). From what we understand, as their Figure 3 shows, log likelihood function for location is unimodal. The following figure is our log likelihood as a function of source location for a synthetic test with SNR = 2 and γ = 50.



x and y are the horizontal distances and z is the depth and the target location is (1, -1, 6) km. The points show the accepted locations passed through the Metropolis acceptance criterion and the distribution is unimodal. Low γ causes posteriors to underestimate uncertainty, so they use a tuned higher one. We could not think of a way to set the parameter to estimate location errors, so setting it 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.

In general, the authors need to explain more throughout the manuscript for example: 1. the parameterisation of the moment tensor 2. how the Greens functions are calculated 3. demonstrate influence of the polarities

Reply: We will explain them in the manuscript.

Replies to the annotations in the manuscript

Line 47: "… Marginal-then-conditional sampling…"

- This is not an established term, but a specific sampling strategy of Gu et al. 2018, so it cannot be stated like that.

Reply: We changed the sentences to "In their study, the marginal posterior probability distribution for any given source location and velocity model is obtained; then for each sample of these parameters, they directly sample the MT from its Gaussian full conditional distribution. [Then in the method section we inserted:] The algorithm is called marginal-then-conditional sampling or marginal algorithm in statistics."

Line 52: "… For the details about the polarity likelihood refer to Brillinger (1980); Walsh et al. (2009) and Wéber (2018). …"

- Thats not a sentence one would put in the introduction but rather in a method section. Or its structure could be changed to keep in introduction..

Reply: We transferred the sentence to the method section where we explained the polarity likelihood function.

Line 55: "… Velocity distributions …"

What is that? Do you mean velocity model?

Reply: Yes, we changed that to velocity model.

Line 85: "… the coefficients of elementary seismograms are set to -1.5 and 1.5. …"

this is a strong assumption, but unclear from the text and needs better description. how are these bounds determined? need to introduce the source parameterisation first. A common parameterization is +-sqrt(2) in the trace and+-1 in the off-diagonals (e.g. Vavrycuk 2015 Moment tensor decompositions revisited, or Staehler et al 2014 Fully probabilistic seismic source inversion – Part 1 :Efficient parameterisation

Reply: Thank you so much for your direction. We actually parametrized coefficients of the elementary seismograms (a_i) belonging to six elementary tensors. a_i relations with the elements of moment tensor is as follows:

$a_1 = M_{xy}$

$a_2 = M_{xz}$

$a_3 = -M_{yz}$

$a_4 = (-2M_{xx}+M_{yy}+M_{zz})/3$

$a_5 = (M_{xx}-2M_{yy}+M_{zz})/3$

$a_6 = (M_{xx}+M_{yy}+M_{zz})/3$,

so according to what you said we should have used the following boundary values for a_i: a_1, a_2 and a_3 between -1 and 1; a_4 and a_5 between -1.88561 and -1.88561 and a_6 between –sqrt(2) and sqrt(2). We will correct and revise the manuscript.

Line 90:

how is G calculated? why spatial derivative?

Reply: G is constructed by concatenating the synthetic seismograms of the six elementary moment tensors and is calculated by frequency wavenumber code AXITRA (Cotton and Coutant, 1997) based on Bouchon (1981) method. The far field displacement due to a point source is represented by the equations 3.18 and 3.23 of Aki and Richards (2009) which contains the spatial derivatives of Green's functions. We will include details about calculating G in the manuscript.

Line 118:

The polarity constraint needs to be stated with an equation. If it has been done before (stated in introduction) it needs a short appendix or supplement. The polarity constraint is claimed to be an important part of the work and the authors cannot rely on the readers to have access to the other publications.

Reply: The constraint does not have any equation and its implementation is simple. After performing the metropolis acceptance criterion in the second chain, we reject any sample which does not satisfy polarity constraint.

Line 131: ... CT is dominant for stronger (uncorrelated noise free) earthquakes...

That is again a strong statement to make, but there is no reference given. Whether the theory errors are dominant depends mostly on how well the Earth structure i.e. the velocity model is known (for seismic waveform data). It is thus often very complicated to even seperate these two components.

Reply: We removed the assertion.