TU Delft Delft University of Technology

La Ode Marzujriban Masfara
Delft University of Technology,
Department of Geoscience and Engineering
Stevinweg 1
2628 CN Delft, The Netherlands
Phone:+31633442776
Email: l.o.m.masfara@tudelft.nl

Delft, June 17, 2022

To: The editor of Solid Earth

Dear editor and reviewers,

Please find our replies to the reviewers' comments on our manuscript se-2021-156, titled "Efficient probabilistic inversion for induced earthquake parameters in 3D heterogeneous media". We thank the reviewers for their detailed and constructive reviews. We have prepared two additional files to complement these replies. One is the "clean" version, and another is the "track changes" version of the revised manuscript. Below, we address the reviewers' comments point by point. In addition, we group each reviewer's comments into two categories, major and minor. Major comments are related to the main concerns of both reviewers, whereas minor comments are comments related to typos and those in the annotated pdf (attached by reviewer 1). We also added colors in the track changes version and our response below to make every change easy to track. The colors are used in the following way:

- Green: We use this color to point to the initial draft (e.g., to specify line number)
- Orange: We use this color to point to the "clean version" of the revised manuscript (e.g., to specify line number)
- Blue: This color is used to highlight parts **added** to the initial draft and point to the line number in the "track changes" version of the revised manuscript.
- Red: This color is used to highlight **deleted** parts from the initial draft and point to the line number in the "track changes" version of the revised manuscript.

Reviewer 1

Prof. Andreas Fichtner (Reviewer 1)

Major comments

Reviewer Point R 1.1 — Given that this is a purely synthetic study that could in principle be translated to a broad range of applications, the introduction seems unnecessarily focused on Groningen. This may confuse some readers, and it artificially limits the scope of this work. It may make more sense to centre the introduction around induced seismicity and earthquake moment tensor inversion in general.

Reply: We are highly aware of this concern. We have tried to limit the focus on Groningen and make it more general by dedicating only 2 out of 8 paragraphs related to Groningen's events in the introduction. We did this to accommodate the future use of the workflow as mentioned in Line 62 and to bring details about what has been done for the Groningen earthquakes. However, we admit that the discussion of these two paragraphs might be too focused and not capture the "general" use of this workflow for induced seismicity events. Therefore, we agree with the reviewer's point of view about the focus, and we address this by rearranging and condensing the two paragraphs. The arrangement for our introduction is now as follows:

- Paragraph 1 & 2: We keep these two paragraphs as in the initial manuscript, except we deleted the last two sentences of the second paragraph.
- Paragraph 3: We take paragraph 5 from the initial draft, except we alter the first two sentences to address point R1.5.
- Paragraph 4: We take paragraph 6 from the initial draft, and we alter the text to partially address points R1.2, R1.6, R1.7, and R1.8.
- Paragraph 5: We added a new paragraph to address points R1.2, R1.6, R1.7, and R1.8 and explain the relevance of using the Groningen subsurface model for our study.
- Paragraph 6: We used paragraph 7 from the initial manuscript, but we modified the first and the last two sentences. These modifications emphasize the choice for the Groningen gas field as our study case.
- Paragraph 7 & 8: We take paragraphs 3 and 4 from the initial draft, except now we deleted the last three sentences of paragraph 3 to have better coherence and make it more concise.

- Paragraph 9: For this part, we use paragraph 8 from the original draft without alteration.

Reviewer Point R 1.2 — The reasons for stronger nonlinearity and weaker priors given on page 3 are not so obvious and need more substance. In general, arguing in terms of frequency may have limited meaning because what ultimately counts is the number of wavelengths between source and receiver.

Reply: We agree with this comment. We address this by making a new paragraph in the introduction (paragraph 5). This paragraph clarifies the criterion for the degree of "non-linearity" (as suggested) and presents the comparison of the number of wavelengths between a tectonic event against an induced event, particularly in the case of the Groningen events. In detail we have added the following:

Line 57/Line 62 – Due to the higher frequencies present in recordings of induced events, the wavelengths are significantly shorter. Layers of sediment/basin infill close to the Earth's surface may exacerbate this, since velocities usually decrease rapidly in this case. Shorter wavelengths matter, because, other things being equal, it increases non-linearity. In essence, however, the degree to which the relation between the source parameters and the recorded waveforms is non-linear depends on the ratio between the nominal event-receiver separation and the wavelength. For example, consider (i) an induced seismic event at 3 km depth, an average P-wave velocity of 2.5 km/s, periods that range between 1 and 0.33 seconds, and event-receiver distances of 4 to 11 km (this study), and (ii) a tectonic event at 50 km depth, an average P-wave velocity of 5 km/s, periods between 100 and 15 seconds, and event-receiver distances of 200 to 1100 km (e.g., Fichtner and Simute, 2018). These values correspond to ratios between event-receiver separation and wavelength that vary (approximately) between 2 and 14 (this study), and 1 to 14 (Fichtner and Simute, 2018). As soon as shear waves are used to perform centroid-moment tensor inversions, however, the non-linearity in the induced seismic setting considered in this study increases relative to the tectonic case considered. This is due to the fact that Vp/Vs ratios are typically significantly higher in the near surface (i.e., the top 1 to 2 km) than at greater depth. This is particularly the case in Groningen (e.g., Spetzler and Dost, 2017).

Reviewer Point R 1.3 — While the derivation of the representation theorem for moment tensor sources has educational value, I doubt that this well-known result needs to be repeated in a research paper. My suggestion would be to leave this out and thereby shorten this already rather long manuscript.

Reply: We agree with this point, and therefore, we have deleted some parts of the derivation. Specifically, we delete the derivation of equations 1 - 5. We keep the rest of the equations except now for equation 6, we write it in two forms. One is to explain the generated displacements

Line 109/Line 123

$$u_i\left(\mathbf{x}_r, t\right) = \sum_{j,k} M_{jk}(\mathbf{x}_a, t) * G_{ij,k}\left(\mathbf{x}_r; \mathbf{x}_a, t\right)$$

and another to highlight the concept of reciprocity

Line 122/Line 165

$$u_i\left(\mathbf{x}_r, t\right) = \sum_{j,k} M_{jk}(\mathbf{x}_a, t) * G_{ji,k}\left(\mathbf{x}_a; \mathbf{x}_r, t\right).$$

To accommodate the changes in the equation, we also alter the explanation of each equation. Apart from these changes, everything remains the same as in the original manuscript.

Reviewer Point R 1.4 — The conclusion that the proposed algorithm works may be a bit too strong, given that it is only run in a synthetic scenario. So, what is actually shown is that the algorithm satisfies some sort of minimum requirement, which is the delivery of reasonable results under idealised conditions. This is still am important achievement!

Reply: This is indeed a fair point. We thank the reviewer for emphasizing this. To address this, and point R1.20, we have added the caveats regarding our use of synthetic data, including the minimum requirements for the workflow to work. In addition to that, we add caveats for the future use of this workflow with real data. More specifically, we have added the following:

Line 398/Line 444 - A number of caveats need to be made though. First, the synthetic recordings used to test our probabilistic workflow are the result of propagating a wavefield through the very same velocity model as the one used to estimate the posterior (i.e., the velocity model in our probabilistic workflow). In application to field data, this would obviously not be the case. Part of the misfit between modeled recordings and observed recordings would then be the result of discrepancies between the true velocity model and the employed numerical velocity model. Second, and in the same vein, we employed the same code (SPECFEM3D-Cartesian) for generating the synthetic recordings as for modelling the wavefield in the probabilistic workflow. And although this code is known to be rather accurate (Komatitsch and Tromp, 2002), undoubtedly some of the physics describing the actual wavefield propagation is not fully captured by SPECFEM3D-Cartesian. Third, this study does not include an application to field data. This is intentional as our objective is to present a stand-alone workflow that can be applied in any induced seismic setting. Applying a methodology to field recordings of induced seismic events (e.g., in Groningen) would require numerous processing details, which we consider to be beyond the scope of this paper. We are currently drafting a follow-up manuscript in which we apply the proposed HMC workflow to field recordings of induced seismic events in Groningen.

Minor comments

Reviewer Point R 1.5 — Line 62 - The existence of observational errors and a null space totally excludes the existence of anything like a "best model". Please drop this term..

Reply: We drop the term and modify:

Line 39 - ... the output does not consist of the best (source) model parameters but the posterior distribution of the desired earthquake parameters.
to

Line 35/Line 39 - ... the output does not consist of a single set of (source) model parameters that minimizes an objective function, but the posterior distribution (see, e.g., Tarantola, 2006) of the desired earthquake parameters.

Reviewer Point R 1.6 — Line 71 - I think this needs some context. In fact, compared to the size of the domain we worked with, these frequencies were pretty high. What counts is not the frequency, but the number of wavelengths between source and receiver!

Reply: We address this point with our answer for point P1.2.

Reviewer Point R 1.7 — Line 74 - I would contest this! Without a doubt, a large tectonic earthquake has a much broader frequency range, especially at the lower frequency end. The high frequencies are simply attenuated with increasing distance. See also previous comment.

Reply: We address this point with our answer for point P1.2. Specifically, when we compare the number of wavelengths between two events, a tectonic and an induced seismicity event.

Reviewer Point R 1.8 — Line 77 - The same here. What counts is the number of wavelengths.

Reply: We address this point with our answer for point P1.2.

Reviewer Point R 1.9 — Line 97 - Please mention that you imply the summation convention.

Reply: Since we deleted this part, we address this point by making sure that all equations have the correct summation sign.

Reviewer Point R 1.10 — Line 107 - I do not quite understand what you mean here.

Reply: We want to explain that we apply reciprocity using sources with identical characteristics/amplitudes to the original sources. Hence, the displacement we obtain is identical to one without applying reciprocity. Furthermore, similar to our response in point P1.9, we have decided to delete this part in order to address point P1.3.

Reviewer Point R 1.11 — Line 134 - Of course, this is a very common assumption. I have always been wondering where this is actually coming from and how reasonable it is.

Reply: In our understanding, and after reading several literatures, we believe that this general assumption is backed by the assumption that the rupture process is homogeneous. With the homogeneous rupture process, each co-located sub-event will have similar time history. Hence, similar moment rate function. And the smaller the fault the more accurate this approximation, as simulated by Melgar et al., 2019 (Differences between heterogeneous and homogeneous slip in regional tsunami hazards modeling: Figure 2) (e.g., in case of pure shear, this would imply faulting to occur along a straight "trajectory"). The assumption of the rupture process can then be addressed by the use of far-field recordings as described by Dahm and Krüger, 2014 (Resolving the time dependency of moment tensors: section 3.1) which ultimately valid for "small faults" as in the most cases of induced seismicity.

Reviewer Point R 1.12 — Line 148 - This sounds like the algorithm has undergone an evolution with successive improvements, which is not that much the case.

Reply: To answer this point, we modify

> Line 184 - Over time, the HMC algorithm has become one ...

to

> Line 141/Line 184 - It is known to be one ...

Reviewer Point R 1.13 — Line 160 - Why recursive? It is a purely sequential algorithm.

Reply: We change the word from (Line 196) recursive to (Line 153)/(Line 196) sequential .

Reviewer Point R 1.14 — Line 168 - Maybe add references. "Hamiltonian Monte Carlo solution of tomographic inverse problems" actually contains a simple explanation and illustration of this. "Autotuning Hamiltonian Monte Carlo" contains more details.

Reply: We added the suggested references in the text.

Reviewer Point R 1.15 — Line 173 - To not confuse readers who are not familiar with this, maybe mention that this is an artificial time. In fact, the whole system is artificial.

Reply: To address this point we modify

> Line 208 - At time τ, the particle arrives at a new location representing a new model $\mathbf{m}(\tau)$.

to

Line 165 / Line 208 - We parenthetically coined τ an artificial time because is shouldn't be confused with physical time t. It is this artificial time with which the model moves through phase space: at time τ, the particle arrives at new location representing a new model $\mathbf{m}(\tau)$.

Reviewer Point R 1.16 — Line 176 - Again, there is no recursion here. These steps are evaluated sequentially and iteratively.

Reply: To address this point we modify

Line 214 - By recursively evaluating equations (9) to (11)...

to

Line 171/Line 214 - By sequentially evaluating equations (8) to (10) in an iterative manner...

Reviewer Point R 1.17 — Line 179 - There is actually no such thing as The Metropolis-Hastings algorithm. It is a whole family, the members of which are distinguished by their respective proposal distributions. So, which proposal distribution do you consider here?

Reply: Here, we consider symmetric proposal distribution, in this case, Gaussian distribution. Therefore for clarity, we modify :

Line 217 - In Figure 1 we exemplify the sampling behavior of both the HMC algorithm (a) and the more generic Metropolis-Hasting (MH) algorithm (b) for a 2D joint probability distribution...

to

Line 174/Line 217 - In Figure 1, we visualize the sampling behavior of both the Metropolis algorithm (a) and the HMC algorithm (b) for a 2D joint probability distribution. Note that he metropolis algorithm is a special case of the Metropolis-Hastings algorithm in the sense that the proposal distribution is symmetric (Hoff, 2009)....

Reviewer Point R 1.18 — Line 241 - I can see this, but I wonder if any other reader would find this as obvious. Maybe you can present some more evidence for this?

Reply: We showcased the evidence of this in Figure 10. Here, we evaluate the variance reduction (VR) given the posterior means of each chain from different sets of prior (bottom panel). These prior sets have varying hypocenter locations. Prior sets containing hypocenter priors closer to the true location (see the distance in the top panel) ended up around the highest VR value implying they ended up in the global minimum. Whereas in the opposite case, the sampler ended up in the lower VR value indicating that they "got stuck" in a local minimum.

Reviewer Point R 1.19 — Line 255 - What exactly is this? Is it some standard deviation? Do you actually include correlated observational errors?

Reply: In this case, we do not include correlated observational errors. We assuming Gaussian distributed, uncorrelated noise. Thus, for each seismogram we simply took 30% of the maximum amplitude of the seismogram as its standard deviation (σ_d). To clarify this we modify the following:

Line 233 - Assuming Gaussian distributed errors σ_d^2 in the observed data,

Line 296 - When running the Markov chains, we assume the data uncertainty (σ_d) to be 30% of the maximum amplitude of each seismogram.

to

Line 188/Line 233 - Assuming Gaussian distributed, uncorrelated, and coinciding data variance σ_d^2

Line 251/Line 296 - When running the Markov chains, we assume the square root of the data variance (σ_d) to be 30% of the maximum amplitude of each seismogram.

Furthermore, we mention in the original text "Admittedly, this is rather arbitrary, and in the application to field data, the data uncertainty has to be estimated from the obtained seismograms themselves." to inform the reader about this choice and our plan when inverting real data.

Reviewer Point R 1.20 — Line 401 - This statement may be a bit too strong. After all, you consider an inverse crime. Hence, you do not demonstrate that the algorithm really works with actual data. What you do show is that the algorithm satisfies the requirement of working under idealised conditions. If that were not the case, there would be no hope for real-data inversions.

Reply: We agree with this point and we answer this in our reply for point R1.4.

Reviewer 2

Dr Tom Kettlety (Reviewer 2)

Major comments

Reviewer Point R 2.1 — The majority of induced seismicity microseismic monitoring takes place with relatively dense local arrays, some of which are downhole. The dominant frequency of recorded signals can be larger than the 1-3 Hz data the study limits itself to. This is especially true of downhole data, where high ($>$ 1000 Hz) sampling rates enables events with corner frequencies in excess of 100 Hz to be routinely detected. The paper claims to be more effective at "higher" frequencies, but would still require significant loss of signal through filtering. How badly are the results affected by this loss of signal? How badly does the inversion perform when, say, 10 Hz data is used?

Reply: This indeed a highly relevant point. The "loss of signal" certainly affects the inversion results. However, it should be understood that these results strongly depend on the quality of the subsurface velocity models and computational power. Having a very detailed and accurate velocity model supported by high computational power enables the inversion to use data with higher frequency, especially if the high frequencies did not attenuate. In this case, we might expect the posterior density to be narrower. In the absence of a detailed velocity model and/or limited computational power, sacrifices are often made, for example, by using 1D instead of 3D velocity models to reduce computational costs. This 1D approximation can be adequate if the heterogeneity is relatively weak. At the same time, another justification for using lower frequency data is to average the error of the modeled seismogram due to the model's impreciseness which applies when using either 1D or 3D models. In conclusion, extension of our approach to higher frequencies is first and foremost contingent on the presence of a sufficiently detailed velocity model at those (higher) frequencies.

The efficiency of our proposed workflow lies in its ability to sample the posterior density at less computational costs (less forward modeling) while the priors are imprecise. In an identical setup (e.g., velocity models and source-receiver configurations), inverting data with a higher upper frequency (say 10Hz) implies a higher degree of non-linearity than 1-3Hz data (see also the paragraph we added to the introduction in reply to Reviewer 1). Using imprecise initial prior can lead the sampler towards local minima for both cases. Because of the employed linearization of the forward problem, this is particularly a problem. As shown in section 7 ("The Importance of Prior"), this be addressed by using multiple starting models, with each starting model having different hypocenter locations within a grid. With 10Hz data, the distance between each posterior minimum would be shorter (due to relatively shorter wavelenghts). Therefore, the grid spacing for the hypocenter priors must be shorter than when inverting 1-3Hz data. In this case, the grid spacing can be shorter than the wavelength given the average velocity and upper frequency of the data used in the inversion. Consequently, the inversion would require more starting models than when using 1-3Hz data. In the manuscript (section 7), we use 25 initial starting models, and time-wise it takes 3-4 minutes to finish using four cores CPU of the 2017 MacbookPro (using more cores and better parallelization technique may reduce the run time). Therefore, with more initial models, one can expect to spend more running time. Although result-wise, assuming the velocity model is accurate, the results of inverting using 10Hz data will have higher precision and accuracy.

Reviewer Point R 2.2 — Without testing any real data, even for a small sample of events from Groningen, it is difficult to justify the conclusion stated in lines 15 and 401. It certainly seems to work for this single synthetic event, but maybe the conclusions shouldn't be overstated. I certainly look forward to the upcoming paper where this is used on Groningen data, but do wish this manuscript included at least a couple of tests of real data to validate the inversion scheme and compare to conventional methods used to locate and invert for focal mechanism.

Reply: We fully agree with the reviewer. We do, however, prefer not to add the real data inversion here because we want the separated manuscript to have an in-depth analysis and discussion on multiple aspects, such as data pre-processing, data uncertainty estimations, and interpretation of the results. And as reviewer 1 remarked, our manuscript is already rather lengthy (we have therefore shortened the manuscript for this reason). Nevertheless, we have changed the sentences of the commented lines. For the comment regarding line 15, we add a description that we employ the algorithm specifically on synthetic data and therefore we modify

Line 14 - We find that our workflow is able to ...

to

Line 13/Line 14 - Using the synthetic case, we find that our proposed workflow is able to ....

For line 401, we have made a description of how we achieved the results using synthetic data. Importantly, and as per request of reviewer 1, we have included a paragraph with caveats (Point R 1.20 and R1.4)

Reviewer Point R 2.3 — In the Groninegen velocity model, is anisotropy included? It can make the MT inversion highly degenerate, especially when it comes to inverting for volumetric components, and should be brought up in the introduction.

Reply: The NAM report includes anisotropy estimations, and the values are assigned based on formations. On average, the anisotropy is in the range of 1.6% measured by the Vh/Vv ratio. Looking at several publications, such as Ma et al. (2022) (Cooperative P-Wave Velocity Measurement with Full Waveform Moment Tensor Inversion in Transversely Anisotropic

Media), this value is relatively small. Their findings implicitly show that the inversion for full moment tensor is still accurate even above the range, but this is of course not a general rule. Nevertheless, we perform the forward modeling without taking into account the anisotropy to reduce high computational costs. However, we do agree with the reviewer in this matter, and therefore we have added the following:

Line 413/Line 460 - ... Second, in the presence of strong anisotropy, the posterior could be adversely affected. In particular, in case of non pure shear mechanisms this effect could be significant (Ma et al. 2022). ....

However, instead of adding these in the introduction, we think that it might be more suitable if we put them in the discussion section. The consideration is that the manuscript focuses on cases where the anisotropy is relatively weak. Yet, we might consider the effect of anisotropy in future use, especially when using real data.

Minor comments

Reviewer Point R 2.4 — Line 111: Is the semi-colon in the wrong place at the start of line 111?

Reply: The semi-colon is indeed in the wrong place. Instead of

$$G_{ij}\left(\mathbf{x}_r, t; \mathbf{x}_a + d\hat{\mathbf{k}}\right)$$

it should be

$$G_{ij}\left(\mathbf{x}_r; \mathbf{x}_a + d\hat{\mathbf{k}}, t\right).$$

To address several comments regarding the length of the manuscript, we have decided to adjust and delete some parts of the equation. That includes one in this comment. Nevertheless, we ensure that our revision does not contain similar typos.

Reviewer Point R 2.5 — Line 179: the references to Figure 1 (a) and (b) appear to be the wrong way round, and (b) is misquoted on line 185. This made Figure 1 more difficult to interpret.

Reply: We admit that the references are misplaced. We fixed this by changing the following:

Line 217 - ... we exemplify the sampling behavior of both the HMC algorithm (a) and the more generic Metropolis-Hasting (MH) algorithm (b) for a 2D joint probability distribution ...

to

Line 174/Line 217 - ... we visualize the sampling behavior of both the Metropolis algorithm (a) and the HMC algorithm (b) for a 2D joint probability distribution ...

Reviewer Point R 2.6 — Line 331: What fraction (%?) of the MT do you use for MT prior stated on line 331

Reply: In this example we use 5%. In that regards, we change the following:

> Line 373 - ...we use a fraction of the minimum absolute value...
>
> to
>
> Line 327/Line 373 - ...we use 5 % of the minimum absolute value...

Reviewer Point R 2.7 — Line 412: You say on line 412 "significant computing power" to calculate reference seismograms, but how long did SPECFEM take in this case? Would be good to reference in relation to the one minute quoted to run the workflow.

Reply: We thank the reviewer for bringing up this point. This point is indeed essential, and therefore, we have added

Line 426/Line 473 - Prior to executing the workflow, one needs to compile a database of the elementary seismograms, which often requires significant computing power. In our case, it took about one day to generate the database using one node of our computer cluster that consists of 24 CPU cores (Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz) with a total ram of 503GiB. Once compiled, our workflow can be run efficiently.

If you have any questions, I would happily answer them. In anticipation of your reply, I remain,

Yours sincerely

La Ode Marzujriban