Estimating induced seismic source parameters using an efficient Hamiltonian Monte Carlo algorithm

ESTIMATING INDUCED SEISMIC SOURCE PARAMETERS USING AN EFFICIENT HAMILTONIAN MONTE CARLO ALGORITHM

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on Tuesday 1 July at 15:00 o'clock

by

La Ode Marzujriban MASFARA RACHMAN

Master of Science in Applied Geophysics,
TU Delft, ETH Zürich, RWTH Aachen
Born in Bau-Bau, Indonesia.

This dissertation has been approved by the promotor and copromotor.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. C. P. A. Wapenaar	Delft University of Technology, *promotor*
Dr. C.Weemstra	Delft University of Technology *copromotor*

Independent members:

Prof. dr. L. G. Evers	Delft University of Technology
Prof. dr. T. van Leeuwen	Utrecht University
Prof. dr. M. van der Meijde	University of Twente
Dr. ir. D. J. Verschuur	Delft University of Technology
Dr. ir. E. N. Ruigrok	Utrecht University
Prof. dr. W. A. Mulder	Delft University of Technology, reserve member

This research has been supported by the project DeepNL, funded by the Netherlands Organization for Scientific Research (NWO) (grant no. DEEP.NL.2018.048).

Keywords: Induced seismicity, probabilistic inversion, Hamiltonian Monte Carlo, Groningen gas field, centroid moment tensor, rupture directivity

Printed by: Gildeprint

Front & Back: Minimalist cover art with abstract contours on a blue background.

ISBN 978-94-6518-089-2

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

"To the teenager I once was, and the son or daughter I may someday raise"

CONTENTS

SUMMARY

Induced seismicity, referring to earthquakes triggered by human activities, has become a pressing concern, particularly in regions with extensive subsurface resource extraction, such as the Groningen gas field in the Netherlands. Decades of natural gas production have altered regional subsurface stress and pressure, reactivating faults and leading to frequent earthquakes, which pose significant structural and societal concerns. To better understand the nucleation and mechanisms behind these earthquakes, source characterization has become a fundamental scientific approach. Accurate characterization not only provides insights into earthquake dynamics but also plays a crucial role in estimating subsurface models, optimizing extraction techniques, and helping mitigate future seismic risks.

In this thesis, we delve into the characterization methods that is by developing an efficient probabilistic workflow for estimating induced earthquake source parameters along with its validation and extensive application. By integrating advanced probabilistic inversion techniques with geological knowledge, the workflow aims to improve characterization accuracy while reducing computational costs. To establish a foundation for this work, we leverage a variant of the Hamiltonian Monte Carlo (HMC) algorithms into the workflow. This algorithm offers significant advantages over traditional Markov Chain Monte Carlo (MCMC) techniques, particularly in handling high-dimensional parameter spaces. The workflow is designed to estimate key earthquake source parameters, including the centroid, moment tensor, and origin time, even in cases where prior information on those parameters is limited or imprecise. The effectiveness of this approach is first demonstrated through a synthetic earthquake scenario, confirming its ability to recover source characteristics with high confidence.

Following the synthetic validation, the study explores how modeling assumptions influence inversion accuracy, including its application to real events in Groningen. First, a geomechanical simulation of the 2018 M_L 3.4 Zeerijp event, is used to investigate whether the commonly applied point source assumption sufficiently represents induced earthquakes or if more complex finite fault models are necessary. Despite the simplifications involved in point source modeling, the results indicate that the estimated earthquake parameters remain robust, even when using simplified velocity models. This suggests that computationally expensive finite fault assumptions may not always be necessary for reliable earthquake characterization. Building on this, the workflow is then applied to real earthquake data from the 2019 M_L 3.4 event below Westerwijtwerd village, where geological prior knowledge is incorporated to enhance computational efficiency. Instead of using a uniform prior distribution for the earthquake's location, fault-based prior information is employed, significantly reducing computational time while maintaining estimation accuracy. This highlights the practical benefits of integrating geological knowledge into probabilistic inversion frameworks.

Expanding the scope of the workflow applicability, the study then examines multi-

ple earthquakes in Groningen, which are ten additional events with magnitudes exceeding M_L 2. Here, the analysis of the workflow outputs is paired with rupture directivity analysis to provide deeper insights into the rupture dynamics of these earthquakes. The findings show that the seismic events are predominantly governed by normal faulting mechanisms, aligning with Groningen's geological structures. Meanwhile, the rupture directivity effects, although minor, show a small correlated trend with known fault orientations. The research concludes with a discussion of the broader implications of these findings and the challenges encountered throughout the study.

While the proposed workflow proves to be both efficient and accurate, two key areas for future improvement are identified. First, incorporating S-wave analysis into the inversion process could enhance the resolution of earthquake parameter estimation, as S-waves carry stronger seismic signals than P-waves. However, this comes with computational challenges due to the complex velocity structure of the Groningen subsurface. Second, a more rigorous treatment of uncertainty is recommended, particularly in both observed and synthetic seismograms. Accounting for these uncertainties would provide more robust and reliable probabilistic estimates, improving the overall accuracy of induced earthquake characterization. By addressing these aspects, this research contributes to the ongoing development of advanced seismic inversion techniques, offering valuable tools for assessing and mitigating the risks associated with induced seismicity. The proposed workflow provides a framework that can be applied beyond Groningen, aiding in the characterization of induced earthquakes in other regions with significant subsurface activities.

SAMENVATTING

Geïnduceerde seismiciteit, oftewel seismiciteit als gevolg van menselijke activiteiten, is een urgent probleem geworden. Dit is onder andere het geval in regio's met uitgebreide olie of gasproductie, zoals het Groningen-gasveld in Nederland. Decennia van aardgasproductie hebben de ondergrondse spanningen en druk in de regio veranderd, waardoor breuken opnieuw zijn geactiveerd en er frequent aardbevingen optreden. Dit brengt aanzienlijke structurele en maatschappelijke zorgen met zich mee. Om de nucleatie van, en mechanismen achter deze aardbevingen beter te begrijpen, is bronkarakterisering een fundamentele wetenschappelijke benadering geworden. Deze karakterisering biedt niet alleen inzicht in de dynamiek van aardbevingen, maar speelt ook een cruciale rol bij het verbeteren van ondergrondmodellen, het optimaliseren van winningstechnieken en het beperken van toekomstige seismische risico's.

In dit thesis richten we ons op het ontwikkelen van een efficiënte probabilistische workflow voor het schatten van bronparameters van geïnduceerde aardbevingen, samen met validatie en uitgebreide toepassing ervan. Door geavanceerde probabilistische inversietechnieken te integreren met geologische kennis, is het doel van de workflow om de nauwkeurigheid van de karakterisering te verbeteren en tegelijkertijd de rekentijd te verminderen. De basis voor dit werk is een variant van het Hamiltonian Monte Carlo (HMC) algoritme. Dit algoritme biedt aanzienlijke voordelen ten opzichte van traditionele Markov Chain Monte Carlo (MCMC) technieken, met name als het gaat om hoogdimensionale parameter ruimten. De workflow is ontworpen om belangrijke aardbevingsbronparameters te schatten, waaronder het hypocentrum, het haardmechanisme en de oorsprongstijd, zelfs in gevallen waarin a priori informatie over deze parameters beperkt of onnauwkeurig is. De effectiviteit van deze aanpak wordt eerst aangetoond aan de hand van een synthetisch aardbevingsscenario, waarbij wordt bevestigd dat de bronkenmerken met hoge betrouwbaarheid kunnen worden verkregen.

Na de synthetische validatie onderzoeken we hoe bepaalde aannames de nauwkeurigheid van de inversie beïnvloeden. In dit geval specifiek voor de toepassing op bevingen in Groningen. Eerst wordt een geomechanische simulatie van de M_L 3,4 Zeerijp-aardbeving in 2018 gebruikt om te onderzoeken of de gangbare puntbron-aanname voldoende representatief is voor de geïnduceerde aardbevingen in Groningen, of dat complexere, eindige bronmodellen nodig zijn. Ondanks de vereenvoudigingen in puntbronmodellering, geven de resultaten aan dat de geschatte aardbevingsparameters robuust blijven, zelfs bij gebruik van vereenvoudigde snelheidsmodellen. Dit suggereert dat (rekenkundig dure) eindige bronmodellen niet altijd nodig zijn voor betrouwbare bronkarakterisering. Vervolgens wordt de workflow toegepast op gegevens van de M_L 3,4 aardbeving in 2019 onder het dorp Westerwijtwerd, waarbij bestaande geologische kennis wordt geïntegreerd in de workflow om de efficiëntie hiervan te verhogen. In plaats van een uniforme (a priori) verdeling voor de locatie van de aardbeving, wordt op breuken gebaseerde voorkennis gebruikt, wat de rekentijd aanzienlijk vermindert zonder afbreuk

te doen aan de nauwkeurigheid van de schattingen. Dit benadrukt de praktische voordelen van het integreren van geologische (a priori) kennis in probabilistische inversies.

Om de toepasbaarheid van de workflow verder aan te tonen, wordt de studie vervolgens toegepast op meerdere aardbevingen in Groningen; namelijk tien additionele geinduceerde bevingen met een magnitude groter dan M_L 2. De analyse van de workflowresultaten wordt gecombineerd met eenanalyse van de voortplantingsrichting langs de breuk om diepgaander inzicht te verkrijgen in de dynamiek van deze aardbevingen. De bevindingen tonen aan dat de geinduceerde aardbevingen in Groningen worden gedomineerd door afschuivingen, in overeenstemming met zowel de geologische structuren in Groningen als de drukverlaging in het reservoir. De gereconstrueerde voortplantingsrichtingen van de bevingen vertonen, hoewel klein, een lichte correlatie met de reeds bekende breukoriëntaties. De studie eindigt met een bespreking van de bredere implicaties van de bevindingen en de uitdagingen die tijdens het onderzoek zijn ondervonden.

Hoewel de voorgestelde workflow zowel efficiënt als nauwkeurig blijkt te zijn, worden twee belangrijke verbeterpunten voor toekomstig onderzoek geïdentificeerd. Ten eerste zou het opnemen van S-golven in het inversieproces de resolutie van de parameterbepaling kunnen verbeteren, aangezien S-golven sterkere seismische signalen dragen dan P-golven. Dit brengt echter rekenkundige uitdagingen met zich mee vanwege de complexe snelheidsstructuur van de Groningse ondergrond. Ten tweede wordt een strengere behandeling van onzekerheid aanbevolen, zowel in geobserveerde als synthetische seismogrammen. Het meenemen van deze onzekerheden zou robuustere en betrouwbaardere probabilistische schattingen opleveren, wat de algemene nauwkeurigheid van de karakterisering van geïnduceerde aardbevingen ten goede zou komen. Door deze aspecten aan te pakken, draagt dit onderzoek bij aan de verdere ontwikkeling van geavanceerde seismische inversietechnieken en biedt het waardevolle instrumenten voor het beoordelen en beperken van de risico's van geïnduceerde seismiciteit. De voorgestelde workflow biedt een raamwerk dat niet alleen toepasbaar is op Groningen, maar ook op andere regio's met significante ondergrondse seismische activiteit.

1

INTRODUCTION

Induced seismicity, a term continually used to describe earthquakes that result from human activities, contrasts with natural earthquakes (Cesca et al., 2013). These human-induced earthquakes are commonly linked with various industrial practices. Among the practices are mining operations. For example, when a mine collapses, it can cause seismic disturbances in the surrounding areas (Caputa et al., 2021). Another example is the impoundment of large dams, which is often categorized as reservoir-induced seismicity (RIS) (Basbous et al., 2022; Ruiz-Barajas et al., 2019). Lastly, hydrocarbon exploitation might alter in-situ stress conditions, affecting the stability of nearby faults and increasing the possibility of triggering earthquakes.

In the context of the oil and gas industry, hydraulic fracturing is intentionally used to induce small-scale seismicity by creating fractures in low-permeability rock formations to enhance hydrocarbon extraction (Hui et al., 2022). While this microseismic activity is a desired outcome of the process, concerns arise when hydraulic fracturing unintentionally triggers larger seismic events, such as fault reactivations or felt earthquakes. Similarly, hydrocarbon extraction can also lead to fault reactivation. The depletion in the reservoir may also result in observed subsidence that, in extreme conditions, can alter the landscape significantly. This subsidence occurs due to the compaction of the reservoir rock as fluids are extracted, reducing the pore pressure that supports the overlying strata. Over time, this compaction can potentially cause the ground surface to sink, leading to regional subsidence (van Thienen-Visser and Fokker, 2017). Another recent trend in the oil and gas industry is carbon capture and storage (CCS). CCS involves capturing CO2 emissions produced by industrial activities and depositing it underground in rock formations so it won't enter the atmosphere. However, without proper monitoring and mitigation efforts, this activity could also lead to induced seismicity (White and Foxall, 2016).

All in all, the underlying mechanism of induced seismicity involves altering stress conditions in the subsurface and is directly attributed to human activities. These alterations can destabilize existing faults or create new pathways for seismic energy release, ultimately resulting in earthquakes (Candela et al., 2019). Although these induced seis-

mic events are often less powerful than their natural counterparts, they can still have substantial impacts. This is particularly true in regions not typically prone to seismic activity, where infrastructure and populations are unprepared for such events. As such, understanding induced seismicity and mitigating its risks is becoming increasingly important.

In this thesis, we focus on understanding induced seismicity by estimating source characteristics of induced earthquakes due to gas extraction. Specifically, we present case studies from the Groningen gas field in the Netherlands. Further down, we begin by summarizing induced seismicity from the area, including efforts that have been made to mitigate such seismicity. We then continue to show the general formulation of what we refer to as the "forward problem," which connects the field observation with an ideal representation of the earthquake sources. Following that, we explain the concept of "inverse problems," which allows us to estimate source characteristics given field observations. Finally, we lay down our objectives and the outline of this thesis, summarizing each following chapter.

1.1. GRONINGEN INDUCED SEISMICITY

Groningen, a province in the Netherlands, has become a notable case study for induced seismicity due to natural gas extraction. The Groningen gas field is located in the northeast part of the province, discovered in 1959, and is regarded as the largest natural gas field in mainland Europe (van Thienen-Visser and Breunese, 2015). Initially, gas extraction was seen as a boon for the Dutch economy, but by the 1980s, small earthquake events were observed in the area. The intensity and frequency of the events then escalated over the years. The year 2012 was a turning point when a 3.6 magnitude earthquake occurred near the town of Huizinge in August of that year. This led to widespread attention to the issue and prompted significant scientific research and policy responses (Jansen and Herber, 2017).

Multifaceted effects have resulted from induced seismicity in Groningen. The most direct effects are physical damage to buildings and infrastructure. Thousands of homes and historical buildings have experienced structural impairment, ultimately introducing safety concerns and decreasing the values of property. Beyond the physical damage, there are also considerable social and psychological effects on the local population, including stress and anxiety due to the unpredictability of earthquakes. To address the effects, the Dutch government has made several efforts to support and reassure the affected communities. This includes not only financial compensation for damages but also initiatives aimed at improving the mental well-being of residents (Vlek, 2019).

Beyond financial, social, and policy aid, the government has also established several research initiatives. These initiatives include strengthening the capability of the national seismological agency, KNMI, which stands for "Koninklijk Nederlands Meteorologisch Instituut" to monitor earthquakes in the area. That is by collaborating with NAM, "Nederlandse Aardolie Maatschappij", a major oil and gas operator in the Netherlands, which provided the extension of the seismology network array installed in the gas field. This extension enables more accurate field monitoring, effectively reducing the uncertainty in estimating the induced earthquake characteristics (Dost et al., 2017). Furthermore, the government established a research program named DeepNL, funded by NWO ("Neder-

landse Organisatie voor Wetenschappelijk"), which is the Dutch national research council. The initial main focus of DeepNL pertains to studying the subsurface of the Groningen area. The sub-focus of the projects includes investigating the response of the shallow part of the Groningen subsurface to earthquakes (Sen et al., 2013), modeling fault occurrences (Ruan et al., 2023) in the area, and estimating surface subsidence for different gas extraction scenarios (Kim et al., 2020).

Under the DeepNL program, this thesis falls under the project of Deep Image, with one of the main focuses on developing new techniques to monitor the gas field. More specifically, we seek to design a workflow to estimate earthquake characteristics in an efficient but probabilistic manner. Beyond that, we try to incorporate available geological information into the estimation. And ultimately applying the workflow in both synthetic and field datasets to showcase its capability.

1.2. INDUCED SEISMICITY CHARACTERIZATION

Earthquake characterization requires understanding various parameters of earthquake sources, such as their magnitude, depth, and epicenter. This process is essential in assessing the potential impact and developing appropriate response strategies. In order to estimate those parameters, seismologists use seismic networks to measure ground vibrations due to the seismic waves generated by the induced earthquakes. The analysis of the seismograms typically requires the application of inversion algorithms. These algorithms help find optimum earthquake parameters that allow forward-modeled seismograms to fit observed seismograms.

In general, induced earthquakes exhibit different characteristics compared to natural earthquakes. In particular, the magnitude of induced earthquakes is relatively smaller than that of natural earthquakes. These low-magnitude earthquakes can then exacerbate the characterization process, thus adding uncertainty to the estimation. Because of the low magnitudes, the frequency content of the induced earthquake seismograms shifts towards higher frequencies. In that sense, in order to accurately forward-model field-observed seismograms, more detailed subsurface models are required, in turn implying higher computational costs.

In the context of the Groningen gas field, much research has been conducted to characterize events in the area. The researchers used different parts of the recorded seismograms and paired them with different types of algorithms and subsurface models depending on the sought-for parameters. For estimating hypocenters, for example, the first arrival of field-observed seismograms is adequate. The study by Spetzler and Dost (2017) uses a deterministic algorithm paired with 1D velocity models, whereas Smith et al. (2020) utilize a probabilistic algorithm paired with 3D velocity models, respectively. In order to include magnitude in the estimation, first arrival information is inadequate, and the seismograms/waveforms themselves need to be utilized. In that case, Willacy et al. (2019) make use of a deterministic algorithm paired with 3D velocity models, while Dost et al. (2020) utilize a probabilistic algorithm paired with 1D velocity models.

Comparing deterministic and probabilistic algorithms, the main benefit of probabilistic algorithms is that they provide the uncertainty of the estimation despite having higher computational costs. In terms of models, 1D velocity suffers from inaccuracies compared to 3D models. However, 1D models are relatively easy to utilize, particularly

when using them to generate forward-modeled seismograms with low computational resources. Given the benefits and drawbacks of performing earthquake characterization using different techniques and models, this thesis will focus on developing an efficient probabilistic inversion workflow while making use of 3D velocity models. This efficiency will be discussed in detail in the following chapter of this thesis, including a comparison with a generic probabilistic algorithm. The workflow will then be tested on one of the largest events in the Groningen gas field, and its estimates will be compared against available research results. Also, integration with available geological knowledge will be linked to the workflow to increase its efficiency and accuracy. Finally, we apply the workflow to more available earthquake datasets and use its output to analyze further rupture propagation along the faults where those earthquakes were predicted to nucleate.

1.3. THE FORWARD PROBLEM

In order to obtain earthquake characteristics, we first have to define a mathematical expression of our forward problem. This mathematical expression links observed seismograms with earthquake source parameters. In this case, the moment tensor source is the earthquake source, representing the point-source equivalent of fault movement in the subsurface. Both observed and synthetic seismograms used in this thesis are in the form of displacements. Assuming that the earthquake source is small enough that it can be represented as a point source, the mathematical expression to compute the synthetic seismograms can be written as (Aki and Richards, 2002):

$$u_i\left(\mathbf{x}^{(\mathrm{r})}, t\right) = \sum_{j=1}^{3}\sum_{k=1}^{3} M_{jk}\left(t, T_0; \mathbf{x}^{(\mathrm{a})}\right) * G_{ij,k}\left(\mathbf{x}^{(\mathrm{r})}, t; \mathbf{x}^{(\mathrm{a})}\right), \tag{1.1}$$

with u_i denoting the displacement u along direction i observed at point $\mathbf{x}^{(\mathrm{r})}$, and considering the source at $\mathbf{x}^{(\mathrm{a})}$ and the propagation beginning at T_0. East, north, and downward displacements are represented by u_1, u_2, and u_3, respectively. M_{jk} is an element of the moment tensor, a symmetric 3x3 matrix with six independent components. Here, j and k signify the force's action and orientation direction. The Green's function, denoted as G, involves components i and j, indicating the recording components and the direction of the moment tensor force's action, respectively. The notation of a comma following the second subscript implies a spatial derivative in the k direction. Finally, the symbol $*$ signifies a time-domain convolution. In essence, displacement $\mathbf{u}$ is a result of earthquake parameters encoded in moment tensor source $\mathbf{M}$ located at $\mathbf{x}^{(\mathrm{a})}$, and which starts to propagate (or 'nucleates') at T_0. Estimating the value of those parameters such that the resulting displacement fits the observed seismograms requires inversion algorithms. In this thesis, we will explore the use of probabilistic inversion algorithms.

1.4. INVERSE PROBLEMS

The probabilistic inversion algorithm used in this study relies on Bayes' theorem. In general, the theorem describes how the probability of a model that explains the data or the "posterior" $\rho(\mathbf{m}|\mathbf{d}^{\mathrm{obs}})$ can be obtained by combining the "prior knowledge" of the model $\rho(\mathbf{m})$ with the likelihood of the data given that model $\rho(\mathbf{d}^{\mathrm{obs}}|\mathbf{m})$. Mathematically,

the theorem can be written as:

$$\rho(\mathbf{m}|\mathbf{d}^{\text{obs}}) \propto \rho(\mathbf{d}^{\text{obs}}|\mathbf{m})\rho(\mathbf{m}). \tag{1.2}$$

Linking equation 1.1 to equation 1.2, the displacement u is represented as $\mathbf{d}^{\text{obs}}$. Hence, $\mathbf{m}$ represents the earthquake parameters: moment tensors ($\mathbf{M}$) , hypocenter ($\mathbf{x}^{(a)}$), and origin time (T_0). In order to calculate the left-hand side of equation 1.2, multiple probabilistic algorithms can be employed. Among them are the Hamiltonian Monte Carlo (HMC) algorithms. The algorithms are known to be efficient in exploring multi-dimensional model spaces, which, in our case, contain the solution or the estimates of our earthquake parameters. Because of this efficiency, HMC algorithms could estimate our posterior with less computational burden and without compromising the accuracy of the estimation compared to more generic probabilistic algorithms.

1.5. RESEARCH QUESTIONS AND OUTLINE

In this thesis, we focus on developing a workflow capable of efficiently estimating induced earthquake parameters in the Groningen gas field in the Netherlands. Specifically, we integrate existing HMC algorithms into the workflow and aim to answer the following research questions:

1. How can we modify existing efficient HMC probabilistic algorithms to estimate induced earthquake parameters without hampering the accuracy of the estimation?
2. What geological information can be incorporated to speed up the workflow/algorithm?
3. Can the modified algorithm be applied to real observed seismograms?
4. What information can we draw in regard to the estimated parameters and the geological knowledge in the area (e.g., reservoir layer and major geological faults)?

To answer the above questions, we divided this thesis into the following chapters:

- ***Chapter 2*** : An efficient probabilistic inversion for induced earthquake parameters in 3D heterogeneous media

  In this chapter, we discuss our modification to the existing variant of the HMC algorithm for the purpose of estimating induced earthquake parameters. We merely use synthetic earthquake data. Specifically, we exploit the HMC algorithm that has been successfully tested on natural earthquakes with larger magnitudes and at greater depths. The algorithm is then integrated into a workflow, which we use in conjunction with 3D subsurface models of the Groningen gas field.

- ***Chapter 3*** : Probabilistic centroid-moment tensor inversions of geomechanically simulated waveforms

  After the successful application of the workflow on the synthetic data in Chapter 2, we now apply it to synthetic data resulting from a geomechanical simulation. The simulation emulates dynamic rupture at a finite fault in both space and time. The resulting waveform from this simulation is more realistic since it does not limit

the source to just the point source as referred to in equation 1.1. In the end, this chapter will validate the use of the point source assumptions we use in our workflow in estimating earthquake parameters, given dynamic rupture simulation data as the simulated "observed" data. In addition, our workflow output can then be used to validate and benchmark geomechanical simulation for real events in the Groningen gas field.

- ***Chapter 4***: Hamiltonian Monte Carlo to characterize induced earthquakes: application to a M_L 3.4 event in the Groningen gas field and the role of prior

  Here, we test our workflow to one of the largest events ever recorded in the Groningen gas field. That is, the 2018 earthquake below the village of Westerwijtwerd with a magnitude of M_L 3.4. Furthermore, we detail the comparison of our workflow with a generic probabilistic algorithm in terms of computational efficiency. In addition to that, we investigate the use of geological prior knowledge of the area to obtain our estimated parameters and reduce the required computation resources. Finally, we compare our estimated values with the results from available research in the area.

- ***Chapter 5***: Moment tensor and rupture directivity analysis of induced earthquakes in Groningen gas field, the Netherlands

  In this chapter, we focus on applying the workflow to more events in Groningen. Specifically, we apply it to ten earthquakes with local magnitudes higher than 2. We then continue by investigating the rupture propagation of these earthquakes. That is, by deriving a quantity commonly referred to as "fault directivity" using the Empirical Green's function (EGF) method.

- ***Chapter 6*** : Conclusions and recommendations

  In this chapter, the conclusions from each of the preceding chapters (Chapters 2 to 5) are summarized, highlighting how the research questions posed at the beginning of the thesis have been addressed. A detailed reflection on the key findings and insights from each chapter is provided, linking the theoretical developments, algorithm modifications, and practical applications. In the end, we add a discussion regarding the challenges we faced, followed by recommendations for future relevant research.

REFERENCES

Aki, K. and Richards, P. G. (2002). *Quantitative Seismology*. University Science Books, 2^{nd} edition.

Basbous, J., Nemer, T., Yehya, A., and Maalouf, E. (2022). Assessing the potential for reservoir induced seismicity from the Bisri dam project in Lebanon. *Engineering Geology*, 304:106679.

Candela, T., Osinga, S., Ampuero, J.-P., Wassing, B., Pluymaekers, M., Fokker, P. A., van Wees, J.-D., de Waal, H. A., and Muntendam-Bos, A. G. (2019). Depletion-induced seis-

micity at the Groningen gas field: Coulomb rate-and-state models including differential compaction effect. *Journal of Geophysical Research: Solid Earth*, 124(7):7081–7104.

Caputa, A., Rudziński, Ł., and Cesca, S. (2021). How to assess the moment tensor inversion resolution for mining induced seismicity: A case study for the Rudna mine, Poland. *Frontiers in Earth Science*, 9:671207.

Cesca, S., Rohr, A., and Dahm, T. (2013). Discrimination of induced seismicity by full moment tensor inversion and decomposition. *Journal of seismology*, 17:147–163.

Dost, B., Ruigrok, E., and Spetzler, J. (2017). Development of seismicity and probabilistic hazard assessment for the Groningen gas field. *Netherlands Journal of Geosciences*, 96(5):s235–s245.

Dost, B., van Stiphout, A., Kühn, D., Kortekaas, M., Ruigrok, E., and Heimann, S. (2020). Probabilistic moment tensor inversion for hydrocarbon-induced seismicity in the Groningen gas field, the Netherlands, part 2: Application. *Bulletin of the Seismological Society of America*, 110(5):2112–2123.

Hui, G., Chen, Z., Wang, P., Gu, F., Kong, X., and Zhang, W. (2022). Mitigating risks from hydraulic fracturing-induced seismicity in unconventional reservoirs: Case study. *Scientific Reports*, 12(1):12537.

Jansen, J.-D. and Herber, R. (2017). Research into induced seismicity in the Groningen field – further studies. *Netherlands Journal of Geosciences*, 96(5):s279–s284.

Kim, S., Vossepoel, F., Hanssen, R. F., Wouters, M., Govers, R., and Stouthamer, E. (2020). A particle method strategy to estimate subsidence induced by a high-dimensional disc-strain model for reservoir compaction. In *EGU General Assembly Conference Abstracts*, page 22117.

Ruan, J., Ghose, R., and Mulder, W. (2023). Fault intersection and induced seismicity: the effects on the induced stress field and the dynamic rupture, and their implications. In *SEG Technical Program Expanded Abstracts*, volume 2023, pages 909–913. Society of Exploration Geophysicists.

Ruiz-Barajas, S., Santoyo, M., Benito Oterino, M., Alvarado, G., and Climent, A. (2019). Stress transfer patterns and local seismicity related to reservoir water-level variations: a case study in central Costa Rica. *Scientific Reports*, 9(1):5600.

Sen, A. T., Cesca, S., Bischoff, M., Meier, T., and Dahm, T. (2013). Automated full moment tensor inversion of coal mining-induced seismicity. *Geophysical Journal International*, 195(2):1267–1281.

Smith, J. D., White, R. S., Avouac, J.-P., and Bourne, S. (2020). Probabilistic earthquake locations of induced seismicity in the Groningen region, the Netherlands. *Geophysical Journal International*, 222(1):507–516.

Spetzler, J. and Dost, B. (2017). Hypocentre estimation of induced earthquakes in Groningen. *Geophysical Journal International*, 209(1):453–465.

van Thienen-Visser, K. and Breunese, J. (2015). Induced seismicity of the Groningen gas field: History and recent developments. *The Leading Edge*, 34(6):664–671.

van Thienen-Visser, K. and Fokker, P. A. (2017). The future of subsidence modelling: compaction and subsidence due to gas depletion of the Groningen gas field in the Netherlands. *Netherlands Journal of Geosciences*, 96(5):s105–s116.

Vlek, C. (2019). Rise and reduction of induced earthquakes in the Groningen gas field, 1991–2018: statistical trends, social impacts, and policy change. *Environmental earth sciences*, 78(3):59.

White, J. A. and Foxall, W. (2016). Assessing induced seismicity risk at CO2 storage projects: Recent progress and remaining challenges. *International Journal of Greenhouse Gas Control*, 49:413–424.

Willacy, C., van Dedem, E., Minisini, S., Li, J., Blokland, J.-W., Das, I., and Droujinine, A. (2019). Full-waveform event location and moment tensor inversion for induced seismicity. *Geophysics*, 84(2):KS39–KS57.

2

AN EFFICIENT PROBABILISTIC INVERSION FOR INDUCED EARTHQUAKE PARAMETERS IN 3D HETEROGENEOUS MEDIA

We present an efficient probabilistic workflow for the estimation of source parameters of induced seismic events in three-dimensional heterogeneous media. Our workflow exploits a linearized variant of the Hamiltonian Monte Carlo (HMC) algorithm. Compared to traditional Markov-Chain Monte Carlo (MCMC) algorithms, HMC is highly efficient in sampling high-dimensional model spaces. Through a linearization of the forward problem around the prior mean (i.e., the "best" initial model), this efficiency can be further improved. We show, however, that this linearization leads to a performance in which the output of an HMC chain strongly depends on the quality of the prior, in particular, because not all (induced) earthquake model parameters have a linear relationship with the recordings observed at the surface. To mitigate the importance of an accurate prior, we integrate the linearized HMC scheme into a workflow that (i) allows for a weak prior through linearization around various (initial) centroid locations, (ii) is able to converge to the mode containing the model with the (global) minimum misfit by means of an iterative HMC approach, and (iii) uses variance reduction as a criterion to include the output of individual Markov chains in the estimation of the posterior probability. Using a three-dimensional heterogeneous subsurface model of the Groningen gas field, we simulate an induced earthquake to test our workflow. We then demonstrate the virtue of our workflow by estimating

Parts of this chapter have been previously published as :

Masfara, L. O. M., Cullison, T., and Weemstra, C. *(2022). An efficient probabilistic workflow for estimating induced earthquake parameters in 3D heterogeneous media. Solid Earth, 13(8):1309–1325.*

The text has been extended and modified in several ways to ensure consistency within this thesis.

the event's centroid (three parameters), moment tensor (six parameters), and the earthquake's origin time. Using the synthetic case, we find that our proposed workflow is able to recover the posterior probability of these source parameters rather well, even when the prior model information is inaccurate, imprecise, or both inaccurate and imprecise.

2.1. INTRODUCTION

The need to understand earthquake source mechanisms is an essential aspect in fields as diverse as global seismology (Ekström et al., 2005), oil and gas exploration (Gu et al., 2018), hazard mitigation (Pinar et al., 2003) and space exploration (Brinkman et al., 2021). In its simplest form, an earthquake source can be described, from a physics point of view, by means of a moment tensor (MT) (Aki and Richards, 2002). An MT captures displacement, (potential) fault orientation, and the energy released during an earthquake. In a regional seismology context, MT inversions can provide insight into seismic afterslip patterns of megathrust earthquakes (e.g., Agurto et al., 2012). In the case that seismic activity is induced by anthropogenic subsurface operations, characterizing seismic sources may also prove essential (e.g., Sen et al., 2013; Langenbruch et al., 2018). With regard to oil and gas exploration, earthquake source mechanisms are often monitored when hydrocarbons are extracted or when fluids are injected into the subsurface (e.g., for fracking). In fact, such monitoring can be used to assess and mitigate the risk of ongoing injection processes activating existing faults (Clarke et al., 2019).

For the purpose of monitoring induced seismicity, arrays of seismometers can be installed over the exploration area. The waveforms recorded by these seismometers can subsequently be exploited to characterize the induced events. For example, the time of the first-arrival (typically the direct P-wave) is sensitive to the earthquake hypocenter and origin time. There are many inversion algorithms that exploit first-arrivals to obtain estimates of earthquake hypocenters and origin times, such as the double-difference (Waldhauser and Ellsworth, 2000) and equal differential time (EDT) (Lomax, 2005) algorithms. However, to estimate MTs, it is insufficient to use only (first-arrival) travel times. In this study, we then develop a workflow that utilizes full waveforms as input. Importantly, we pair the workflow with a probabilistic inversion algorithm.

In terms of computational efficiency, each combination of a specific inversion algorithm and a specific subsurface model has both its advantages and disadvantages. In general, the main advantage of using a probabilistic approach is that 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. Probabilistic approaches, however, are computationally significantly more expensive than deterministic ones. One way to reduce the computational expense is using 1D subsurface models instead of 3D velocity models to model seismograms. Unfortunately, this can adversely affect the reliability of the obtained posterior because part of the heterogeneity of the subsurface is not accounted for (Hingee et al., 2011; Hejrani et al., 2017). In our workflow, we, therefore, deploy a computationally efficient probabilistic algorithm to invert for centroid (three coordinate components), origin time, and MT (six independent MT components) while at the same time utilizing a detailed 3D subsurface model.

The algorithm used in our workflow is the Hamiltonian Monte Carlo (HMC) algo-

rithm, which, for sampling high-dimensional posterior distributions, has shown to be significantly more efficient than the conventional probabilistic Metropolis-Hasting family of algorithms (Betancourt, 2017). Using frequencies lower than 0.1 Hz and available prior information, Fichtner and Simutė (2018) developed a variant of the HMC and demonstrated its efficiency to invert for the source parameters of a tectonic earthquake. More recently, Simute et al. (2022) demonstrated the variant's ability to estimate earthquake parameters of tectonic earthquakes while employing 3D subsurface models of the Japanese islands. In contrast to tectonic earthquakes, where prior information regarding the event's MT, centroid, and origin time is often available, such prior information is usually absent for induced earthquakes. An insufficiently constrained prior reduces the ability and efficiency of sampling algorithms to properly sample the posterior distribution and increases the chance of the sampler getting trapped in local minima (Sen and Stoffa, 2013). In addition, compared to tectonic events, the frequency content of induced earthquake waveforms is usually significantly higher. This is because tectonic events usually occur at greater depths than induced events, and hence, the higher frequencies have been attenuated more. Also, most of the studied induced events are of lower magnitudes than tectonic events (e.g., below Mw 3) and, therefore, do not excite frequencies below 1 Hz that effectively.

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. The presence of shorter wavelengths matters because other things are equal and therefore 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 Simutė, 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 Simutė, 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).

In this study, the absence of a well-constrained prior and an increase in non-linearity receives significant attention. First, the challenge of a weaker prior is met by means of a workflow in which the initial prior is updated before running the HMC algorithm. In addition, multiple chains of the HMC variant are run sequentially, where the results of the current chain serve as priors for the next chain. This iterative HMC is meant to provide improved prior information resulting in an adequate linear approximation. We demonstrate the validity of our workflow using data from a synthetically generated induced earthquake, which was simulated using the velocity model of the Groningen subsurface.

It should be understood that the proposed workflow is of interest for the characterization of induced seismic events in general. The Groningen case is merely chosen because of the quality and density at which the induced wavefields are sampled and the relatively high resolution of the available velocity model.

The Groningen gas field is one of the largest gas reservoirs in Europe. Since production began in 1963, more than 2115 billion cubic meters of natural gas have been produced from the field (van Thienen-Visser and Breunese, 2015). Due to this gas production, the reservoir layer has compacted over time, causing earthquakes that have, in some cases, caused damage to buildings in the Groningen province (van Eck et al., 2006) and led to several protests against further gas extraction in the area (Verdoes and Boin, 2021). To investigate these earthquakes, an extensive seismometer array was installed, which is operated by the KNMI (The Royal Netherlands Meteorological Institute) on behalf of Nederlandse Aardolie Maatschappij (NAM) (Ntinalexis et al., 2019). Event recordings collected over the Groningen field have been used as input for several inversion algorithms. Spetzler and Dost (2017) use the EDT algorithm to invert for the hypocenters of many Groningen earthquakes. They invert arrival times of 87 events and found that all earthquakes occurred within a depth interval of 2300 to 3500 m, with most of the events originating from the reservoir layer (approximately 3000 m depth). These findings are in line with the results of Smith et al. (2020), who use the envelopes of the seismic arrivals as input to their probabilistic algorithm. To invert for both hypocenter (or centroid) and MT, Willacy et al. (2018) take a different approach. Contrary to Spetzler and Dost (2017), who use a 1D model to represent Groningen's subsurface, they utilize a 3D heterogeneous model similar to Smith et al. (2020), and use the model to generate synthetic waveforms to perform a full-waveform deterministic MT inversion. The results of Willacy et al. (2018), however, only focus on pure double-couple sources, which might not capture the true source dynamics. In fact, Dost et al. (2020) recently follow a probabilistic approach to invert event centroids and MTs of a selected number of events and consistently found the (non double-couple) isotropic component of the MT to be dominant and negative. The latter is in agreement with expectations for a compacting medium. Similar to Willacy et al. (2018), they invert waveforms but employ 1D local subsurface models to generate the modeled seismograms.

In what follows, we first introduce the forward problem of obtaining surface displacements (recorded wavefields) due to induced seismic source activity, including the description of a seismic source in terms of elementary moment tensors. Subsequently, we introduce the Bayesian formulation and detail the linearized HMC algorithm. Afterwards, we proceed with the description and implementation of our workflow, which involves several steps that are specific to the characterization of induced seismic sources. We then test the proposed workflow using synthetic recordings of an induced earthquake source. We end by giving a perspective discussion of our results, including an outlook of applying our workflow to actual field recordings of induced earthquakes from the Groningen gas field.

2.2. FORWARD PROBLEM

As with all Markov Chain Monte Carlo algorithms, HMC involves an evaluation of forward-modeled data against observed data. In our case, this evaluation is between (forward)

modeled surface displacement and observed displacement. Specifically, we compute synthetic displacement seismograms $\mathbf{u}$ due to a moment tensor source $\mathbf{M}$ (with each of their components denoted as u and M) (Aki and Richards, 2002)

$$u_i(\mathbf{x}_r, t) = \sum_{j,k} M_{jk}(\mathbf{x}_a, t) * G_{ij,k}(\mathbf{x}_r; \mathbf{x}_a, t) \tag{2.1}$$

with $\mathbf{x}_r$ the location at which $\mathbf{u}$ is recorded, $\mathbf{x}_a$ the source location, and $*$ representing temporal convolution. Subscripts i, j, and k take on values 1,2 and 3 such that a vector can be decomposed in three Cartesian components, associated with the x_1, x_2, and x_3 axis, respectively. $\mathbf{G}$ is the Green's function, and its first subscript represents its recorded component. The second subscript indicates the direction in which an impulsive (delta function) force couple is acting. The comma after the second subscript represents a spatial derivative at $\mathbf{x}_a$ with respect to the direction represented by the subscript after the comma. Each component of $\mathbf{M}$ represents the strength of a force couple. Together, the 9 constants M_{jk} constitute the second order seismic moment tensor $\mathbf{M}$. The MT effectively approximates a seismic source by collapsing it into a single point. Furthermore, due to the conservation of angular momentum, the MT has only six independent components (e.g., Aki and Richards, 2002; Jost and Herrmann, 1989).

Instead of repeatedly computing $\mathbf{u}$ for each source-receiver pair location, it is convenient to exploit source-receiver reciprocity. That is, we exploit the fact that $G_{ij}(\mathbf{x}_r; \mathbf{x}_a, t) = G_{ji}(\mathbf{x}_a; \mathbf{x}_r, t)$ (Aki and Richards, 2002; Wapenaar and Fokkema, 2006), which yields

$$u_i(\mathbf{x}_r, t) = \sum_{j,k} M_{jk}(\mathbf{x}_a, t) * G_{ji,k}(\mathbf{x}_a; \mathbf{x}_r, t)\,. \tag{2.2}$$

To facilitate the computation of seismograms for a specific $\mathbf{M}$, we follow the work of Mustać and Tkalčić (2016) who use six independent tensors that they call elementary moment tensors as decomposed by Kikuchi and Kanamori (1991)

$$\mathbf{M}^1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \mathbf{M}^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \mathbf{M}^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

$$\mathbf{M}^4 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{M}^5 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{M}^6 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.3}$$

Under the assumption that each of these elementary moment tensors have the same time dependence (e.g., in case of pure shear, this would imply faulting to occur along a straight "trajectory"), a specific $\mathbf{M}$ can be described as a linear combination of these elementary moment tensors, i.e.,

$$\mathbf{M} = \sum_{n=1}^{6} a_n \mathbf{M}^n = \begin{pmatrix} -a_4 + a_6 & a_1 & a_2 \\ a_1 & -a_5 + a_6 & -a_3 \\ a_2 & -a_3 & a_4 + a_5 + a_6 \end{pmatrix}. \tag{2.4}$$

where the coefficients a_n $(n = 1, 2, ..., 6)$ are usually referred to as expansion coefficients. In this study, we assume instantaneous rupturing of the source. This is not an uncommon assumption for (relatively small) induced seismic events. This assumption implies that the time dependence of an MT is modeled using a Heaviside function. Using the decomposition above and source-receiver reciprocity, we compute elementary seismograms S_i^n

$$S_i^n(\mathbf{x}_r;\mathbf{x}_a,t) = \sum_{j,k} G_{ji,k}(\mathbf{x}_a;\mathbf{x}_r,t) * M_{jk}^n(t). \tag{2.5}$$

Consequently, we obtain

$$u_i(\mathbf{x}_r;\mathbf{x}_a,t) = \sum_{n=1}^{6} a_n S_i^n(\mathbf{x}_r;\mathbf{x}_a,t). \tag{2.6}$$

In practice, all $S_i^n(\mathbf{x}_r;\mathbf{x}_a,t)$ are computed for a finite number of $\mathbf{x}_a$ on a predetermined subsurface grid with a specific grid spacing. We detail the numerical implementation of computing the $S_i^n(\mathbf{x}_r;\mathbf{x}_a,t)$ further below.

2.3. HAMILTONIAN MONTE CARLO

The HMC algorithm originated from the field of classical mechanics and its application to statistical mechanics (Betancourt, 2017). It is known to be one of the most efficient probabilistic algorithms within the Markov Chain Monte Carlo (MCMC) family. For our workflow, we apply a variant of the HMC algorithm that utilizes a linearization of the forward problem. Therefore, we include several initial steps in our workflow to obtain priors that enable meaningful linearization. In total, our workflow estimates ten source parameters. These are the centroid $\mathbf{x}_a$ (three components), the origin time T_0, and the MT (6 independent MT components).

Similar to other probabilistic algorithms, HMC is deployed in the context of Bayesian inference. The objective of Bayesian inference is to obtain an estimate of the posterior probability distribution $\rho(\mathbf{m}|\mathbf{d})$ that approaches the true posterior probability distribution (from here on, we will refer to $\rho(\mathbf{m}|\mathbf{d})$ as being "the posterior"). This approach combines the likelihood $\rho(\mathbf{d}|\mathbf{m})$ of the observed data given the modeled data with the simultaneous assimilation of the distribution of prior knowledge $\rho(\mathbf{m})$, i.e.,

$$\rho(\mathbf{m}|\mathbf{d}) \propto \rho(\mathbf{d}|\mathbf{m})\rho(\mathbf{m}), \tag{2.7}$$

where $\mathbf{m}$ is a vector that contains model parameters and $\mathbf{d}$ a vector containing the observed data. The likelihood evaluates a model $\mathbf{m}$ against the observed data $\mathbf{d}$ by evaluating the misfit between the latter and forward modeled data associated with $\mathbf{m}$.

The HMC algorithm relies on the sequential calculation of two quantities. These are the potential energy U, which explicitly quantifies $\rho(\mathbf{m}|\mathbf{d})$, and the kinetic energy K, which is a function of momentum vector $\mathbf{p}$. Together, they make up the Hamiltonian $H(\mathbf{m},\mathbf{p})$, which represents the total energy of a system (Neal, 2011), and is written as follows:

$$H = K(\mathbf{p}) + U(\mathbf{m}), \quad \text{with} \quad U(\mathbf{m}) = -\ln\rho(\mathbf{m}\,|\,\mathbf{d}) \quad \text{and} \quad K(\mathbf{p}) = \mathbf{p}^T\mathscr{M}^{-1}\mathbf{p}/2. \tag{2.8}$$

A model $\mathbf{m}$ can be interpreted as the position of a particle within *phase space*. The phase space has a dimension that is twice the dimension of the model space (i.e., this dimension coincides with the length of the vector $\mathbf{m}$ multiplied by two). By having the same dimension as $\mathbf{m}$, the elements of the auxiliary momentum vector $\mathbf{p}$ are therefore needed to complement each dimension of the model space (Betancourt, 2017). The movement of the particle is highly dependent on the mass matrix $\mathcal{M}$, which therefore often acts as a tuning parameter (Fichtner et al., 2019, 2021). The mass matrix affects the "distance" a particle travels and ideally coincides with the posterior covariance matrix. Given a certain initial momentum $\mathbf{p}$ drawn from $exp(\mathbf{p}^T\mathcal{M}^{-1}\mathbf{p}/2)$, the particle is allowed to travel for a certain (artificial) time τ while in conjunction fulfilling Hamilton's equations:

$$\frac{d\mathbf{m}}{d\tau} = \frac{\partial K}{\partial \mathbf{p}}, \quad \frac{d\mathbf{p}}{d\tau} = -\frac{\partial U}{\partial \mathbf{m}}. \tag{2.9}$$

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 a new location representing a new model $\mathbf{m}(\tau)$. The new model and momentum vectors are associated with updated potential and kinetic energies, respectively, and hence a higher or lower Hamiltonian $H(\mathbf{p}(\tau),\mathbf{m}(\tau))$. Given the probability θ, that the particle will stay at the new location, the acceptance probability is given by,

$$\theta = \min\left[1, \frac{\exp[-H(\mathbf{p}(\tau),\mathbf{m}(\tau))]}{\exp[-H(\mathbf{p},\mathbf{m})]}\right]. \tag{2.10}$$

By sequentially evaluating Equations (2.8) to (2.10) in an iterative manner, we collect all locations (models) visited by the particle, except for a number of initial models (representing the burn-in period). The density of the collected models asymptotically approaches the posterior probability distribution.

In Figure 2.1, 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). Both algorithms start with the same starting model, which is represented by the red ball. The low a posteriori probability of this initial model corresponds to a high U. The question marks in Figure 2.1(a) represent randomly selected models by the Metropolis algorithm, which were not accepted due to their relatively low acceptance probability. Hence, each of these question marks involves a (computationally expensive) solution to the forward problem. Instead of using random sampling, in the HMC algorithm, the particle within phase space moves along trajectories obtained by solving Equation (2.9) leading the particle being exerted towards areas with low U, as illustrated in Figure 2.1(b). Furthermore, in Figure 2.1(b), the result of solving Equation (2.9) (i.e., the HMC trajectory) is represented by the brown dashed lines, and the pointing finger represents the momentum vector $\mathbf{p}$. For both the HMC and Metropolis algorithms, an accepted model serves as starting model for the next sample. Although being probabilistic in terms of acceptance probabilities, the trajectories of the HMC algorithm are deterministically guided by $\partial U/\partial \mathbf{m}$ as shown in Equation (2.9). Therefore, the algorithm is also known as the Hybrid Monte Carlo algorithm (Duane et al., 1987). Thus, after proper tuning, the HMC algorithm requires less sampling

than the Metropolis algorithm to converge, which makes the HMC algorithm computationally more efficient.

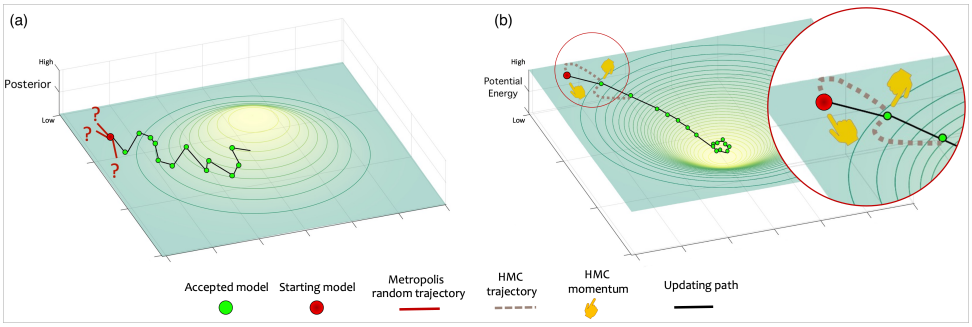

Figure 2.1: *Comparison between the sampling strategy of* (*a*) *Metropolis algorithm and* (*b*) *Hamiltonian Monte Carlo algorithm* .

Assuming Gaussian distributed, uncorrelated, and coinciding data variance σ_d^2, we can write U as (Fichtner and Simutė, 2018)

$$U(\mathbf{m}) = \frac{1}{2T}\sum_{r=1}^{N_r}\sum_{i=1}^{3}\int_0^T \sigma_d^{-2}\left[u_i(\mathbf{x}_r,t;\mathbf{m}) - u_i^{\text{obs}}(\mathbf{x}_r,t)\right]^2 \mathrm{d}t + \frac{1}{2N_m}\left(\mathbf{m}-\mathbf{m}^0\right)^{\mathsf{T}}\mathbf{C}_m^{-1}\left(\mathbf{m}-\mathbf{m}^0\right). \tag{2.11}$$

In our context, the $\mathbf{x}_r$ are the locations of the N_r three-component KNMI seismometers ($r = 1,2,3,..,N_r$). Furthermore, T is the length of observed and forward modeled seismograms in time, N_m the number of model parameters (ten in our case), $\mathbf{m}^0$ a vector containing prior means, and $\mathbf{C}_m$ the prior covariance matrix. In application to field data, the u_i^{obs} would be field recordings by seismometers, but in this study, we restrict ourselves to a numerically simulated induced event.

2.4. LINEARIZATION OF THE FORWARD PROBLEM

In our workflow, most of the computational burden in running HMC involves the evaluation of Equation (2.9). This is because for each $d\tau$ we have to evaluate $\partial U/\partial\mathbf{m}$. To speed up the process, we use a variant of the HMC algorithm introduced by Fichtner and Simutė (2018), in which $\mathbf{u}(\mathbf{x}_r,t;\mathbf{m})$ is approximated by means of an expansion around the prior mean, i.e., around $\mathbf{m}^0$:

$$u_i(\mathbf{x}_r,t;\mathbf{m}) = u_i\left(\mathbf{x}_r,t;\mathbf{m}^0\right) + \sum_{p=1}^{N_m}\frac{\partial}{\partial m_p}u_i\left(\mathbf{x}_r,t;\mathbf{m}^0\right)\left(m_p - m_p^0\right). \tag{2.12}$$

Substituting this linearized expression in Equation (2.11) gives,

$$U(\mathbf{m}) = \frac{1}{2}\sum_{p,q=1}^{N_m}\left(m_p - m_p^0\right)A_{pq}\left(m_q - m_q^0\right) + \sum_{p=1}^{N_m} b_p\left(m_p - m_p^0\right) + \frac{1}{2}c \tag{2.13}$$

where A_{pq}, b_p, and c , given model uncertainty σ_m^2 read

$$A_{pq} = \frac{1}{T\sigma_d^2}\sum_{r=1}^{N_r}\sum_{i=1}^{3}\int_0^T [\frac{\partial}{\partial m_p}u_i\left(\mathbf{x}_r,t;\mathbf{m}^0\right)\frac{\partial}{\partial m_q}u_i\left(\mathbf{x}_r,t;\mathbf{m}^0\right)]\mathrm{d}t + \frac{1}{N_m\sigma_m^2}, \tag{2.14}$$

$$b_p = \frac{1}{T\sigma_d^2}\sum_{r=1}^{N_r}\sum_{i=1}^{3}\int_0^T \left[u_i\left(\mathbf{x}_r,t;\mathbf{m}^0\right) - u_i^{\mathrm{obs}}\left(\mathbf{x}_r,t\right)\right]\frac{\partial}{\partial m_p}u_i\left(\mathbf{x}_r,t;\mathbf{m}^0\right)\mathrm{d}t, \tag{2.15}$$

and

$$c = \frac{1}{T\sigma_d^2}\sum_{r=1}^{N_r}\sum_{i=1}^{3}\int_0^T \left[u_i\left(\mathbf{x}_r,t;\mathbf{m}^0\right) - u_i^{\mathrm{obs}}\left(x_r,t\right)\right]^2\mathrm{d}t. \tag{2.16}$$

Differentiating Equation (2.13) with respect to m_p, we have (Fichtner and Simutė, 2018)

$$\frac{\partial U}{\partial m_p} = \sum_{q=1}^{N_m} A_{pq}\left(m_q - m_q^0\right) + b_p, \tag{2.17}$$

which, together with the random momentum vector, determines the HMC trajectory.

Because the displacement depends linearly on the moment tensor components (see Equations (2.5) and (2.6)), Equation (2.12) is exact with respect to these parameters. The dependence on the other parameters is non-linear, and this non-linearity increases as the frequency of the input data increases. Therefore, in the case of induced events, which usually generate higher frequencies than stronger, regional events, the non-linearity is considerably higher. Hence, to have a tolerable linearization, accurate priors are required when it comes to the centroid and origin time. Without sufficiently accurate priors, the above HMC variant will struggle to sample the mode containing the global minimum of the potential energy. Therefore, we propose an approach that involves an initial estimation of the prior mean in order to permit this linearization. This is detailed further below.

2.5. NUMERICAL IMPLEMENTATION

In practice, the elementary seismograms discussed in Section 2.2 are computed for a finite number of possible centroid locations. That is, prior to our probabilistic inversion, we generate a database of these seismograms. This database contains, for each possible source location $\mathbf{x}_a$ and receiver location $\mathbf{x}_r$ (r=1,...,N_r), a total of 3 x 6 = 18 elementary seismograms (three components for each of the six elementary moment tensors). In our case, each $\mathbf{x}_r$ corresponds to a (KNMI) seismometer location that recorded the induced event. The elementary seismograms are computed using the spectral element software SPECFEM3D-Cartesian (Komatitsch and Tromp, 2002) and we exploited source-receiver reciprocity while doing so. We use an existing detailed Groningen velocity model (Romijn, 2017) for this purpose from which we construct a regular grid of the model using gnam and PyAspect Python packages that are available at `https://github.com/code-cullison/gnam` and `https://github.com/code-cullison/pyaspect`.

To confirm the successful implementation of source-receiver reciprocity, we simulate a scenario of an induced event in the Groningen gas reservoir (Figure 2.2). The centroid is indicated with a red star, and the receivers are depicted as white triangles.

At each location, the wavefield is "recorded" at 200 meters depth by the deepest of a series of four borehole geophones (Ruigrok and Dost, 2019). The elementary seismograms computed at the location of KNMI station G094 are shown in Figure 2.3 (green), and superimposed on top (yellow) are the waveforms resulting from the application of source-receiver reciprocity. All seismograms are bandpass filtered between 1 and 3 Hz, similar to the passband used by Dost et al. (2020).

We integrate the above HMC variant into our workflow by implementing a leapfrog algorithm for evaluating Equation (2.9). Furthermore, we define $d\tau$ as suggested by Neal (2011) to ensure numerical stability and set a fixed value for τ for all chains. The construction of the mass matrix $\mathscr{M}$ is discussed in the next section.

2.6. AN ITERATIVE APPROACH

The performance of the linearized HMC variant strongly depends on the prior means (see Equation (2.12)). For that reason, we propose a workflow in which the algorithm is run iteratively, with each iteration involving an update of the priors to allow for an updated linearization. Specifically, instead of evaluating Equation (2.14) - (2.16) once, we run a sequence of HMC chains. For each successive chain, the posterior means and standard deviations from the previous chain act as prior means and entries for $\mathscr{M}$ in the new chain (i.e., the next iteration). For the first chain in the sequence, the "initial" prior means (i.e., $\mathbf{m}^0$) are obtained via a specific scheme integrated into the workflow. In what follows, we will first introduce the setup of our synthetic earthquake, followed by an approach to estimate required prior means before wrapping up by summarizing the full workflow, which runs iteratively.

2.6.1. SYNTHETIC EXPERIMENT SETUP

We test our workflow for an induced event shown in Figure 2.2. We set the MT components to 9×10^{13} Nm, -1×10^{13} Nm, -3×10^{13} Nm, 8×10^{13} Nm, 5×10^{13} Nm, and 4×10^{13} Nm for M_{11}, M_{22}, M_{33}, M_{12}, M_{13}, and M_{23}, respectively. Using the moment-magnitude relation given by Gutenberg (1956) and Kanamori (1977), this moment tensor can be shown to correspond to an earthquake of magnitude 3.28 Mw. We also add noise to our synthetic seismograms in order to make our experiment more realistic. This noise is added in the frequency domain by multiplying the (complex) spectrum of each synthetic seismogram with a bivariate normal distribution that has a zero mean and a standard deviation of 15% of the amplitude of the seismogram at the dominant frequency. As a result, this noise will not only give amplitude variations but also varying time-shifts with respect to the true synthetic seismograms. 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. Admittedly, this is rather arbitrary, and in the application to field data, the data uncertainty has to be estimated from the obtained seismograms themselves. Finally, we set the origin time to 14 s.

2.6.2. PRIOR MEAN ESTIMATION

Before running the first HMC chain, we need to estimate the initial prior means and variances. In short, we propose an approach in which a first-arrival based algorithm is used

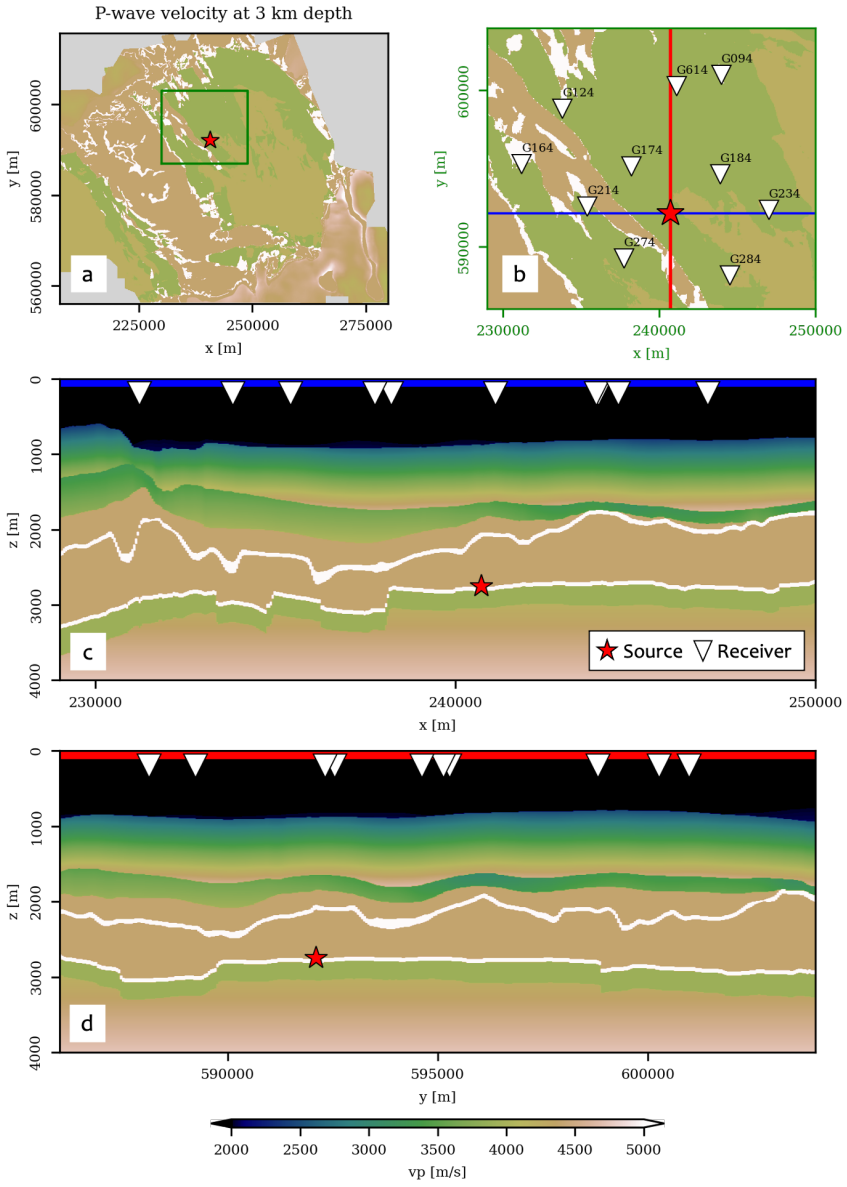

Figure 2.2: *Scenario of an induced earthquake in Groningen area. a. Horizontal slice of the Groningen P-wave velocity model. b. Zoom of the area indicated by the green rectangle in a. 'Inverted triangles indicate locations of KNMI stations (i.e., the $\mathbf{x}_r$)'. c. Vertical slice along the blue line in b. d. Vertical slice along the red line in b. Spatial coordinates in the images are given in the Rijksdriehoekscoördinaten (RD) coordinate system.*

to estimate the centroid. Subsequently, the origin time can be estimated, after which Equation (2.17) can be used to compute the prior means for the individual moment ten-

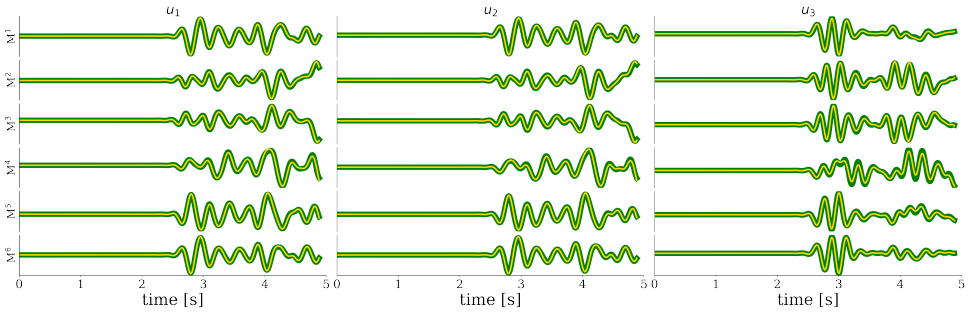

Figure 2.3: *Comparison between elementary seismograms due to a source at the actual location (red star) and the receiver at G094 (green) and the elementary seismograms resulting from the implementation of source-receiver reciprocity (yellow). The equality of the traces confirms the successful numerical implementation of source-receiver reciprocity. Along the vertical axis, all six (independent) elementary seismograms are depicted. Left, center, and right plots show particle displacement in the x_1, x_2, and x_3 direction, respectively.*

sor components. Each of these steps is now discussed in more detail.

Numerous algorithms exist that allow one to estimate an earthquake's hypocenter and/or centroid. Here we propose using first-arrival based algorithms for this purpose since these are computationally more efficient than waveform-based algorithms. First-arrival based algorithms only require the computation of the P and S-wave arrival times, and by adopting a high-frequency approximation (e.g., Aki and Richards, 2002), these arrivals can be found by running one of the various Eikonal solvers (e.g., Noble et al., 2014). For example, the EDT method detailed in Lomax (2005) can be used for this purpose (Masfara and Weemstra, 2021).

As an alternative to using a first-arrival based algorithm, the prior means of the centroid can instead be retrieved from existing literature if it exists. For example, in the case of the induced seismicity in the Groningen field, Smith et al. (2020) have shown that they could resolve hypocenters with maximum uncertainties of 150 m and 300 m for epicenter and depth, respectively. Their results could be considered as priors. Another option is to use the epicenters from the KNMI earthquake database, which by default has all depths set to 3 km.

Given a centroid prior mean that was either calculated or acquired from literature, the prior mean of the origin time can be estimated by computing the P-wave traveltimes from a centroid prior to each of the receivers. These travel times can be computed using the same Eikonal solver that was used to obtain the centroid prior (e.g., the Fast Marching Method; Sethian and Popovici, 1999). By subtracting the computed travel times from the observed (picked) first-arrival times and averaging across receivers, an initial origin time prior mean can be obtained.

To refine the initial origin time prior estimate, we cross-correlate the envelope of the observed seismograms $env[u_i^{obs}(\mathbf{x}_r, t)]$ with the envelope of the forward modeled seismograms $env[u(\mathbf{x}_r, t, \mathbf{m}^0)]$. We do this for each component of each receiver location individually. The forward-modeled seismograms are computed using full-waveform modeling (detailed in Section 2.5) using the initial prior means for centroid and origin time,

and given arbitrary MT components. Specifically, we compute

$$T_{0shift} = \underset{t}{\operatorname{argmax}} \sum_{r=1}^{N_r} \sum_{i=1}^{3} env[u_i(\mathbf{x}_r, t; \mathbf{m}^0)] \star env[u_i^{obs}(\mathbf{x}_r, t)], \tag{2.18}$$

where T_{0shift} is the additional time shift that needs to be added to the initial origin time prior mean to obtain the refined origin time prior.

We test Equation (2.18) using the synthetic earthquake shown in Figure 2.2. For this test, we add 600 m to the true x_1, x_2, and x_3 centroid components, and we impose a (rather aggressive) 9 s time shift with respect to the true origin time. Note, this implies that we did not employ the aforementioned procedure to obtain initial centroid and origin time priors because this would result in a centroid and origin time estimate that would be too close to the true centroid and origin time – essentially rendering the use of Equation (2.18) unnecessary. In other words, we deliberately impose large deviations from the true values to show the merit of using Equation (2.18).

Given arbitrary MT components, we show in Figure 2.4 the result of applying Equation (2.18) to vertical surface displacements. In Figure 2.4(a), we depict the envelopes of modeled seismograms given available prior means (i.e., 600 m deviation from the true x_1, x_2, and x_3 values, and 9 s from the true T_0). Figure 2.4(b) shows the noisy synthetic "observed" seismograms and Figure 2.4(c) is the result of applying Equation (2.18) to each of the displacement envelopes. In Figure 2.4(d), we show the result of stacking all signals in Figure 2.4(c). The vertical blue line indicates the time at which the stack of the cross-correlated envelopes attains its maximum value, i.e., T_{0shift}; the vertical red line represents the deviation of the initial origin time prior from the true origin time (i.e., 9 seconds in this example).

Having sufficiently accurate prior means for the centroid and origin time, we then estimate the prior mean of the MT. For this purpose, we keep the centroid and origin time constant but solve for the remaining six parameters (the independent MT elements). In Section 2.4, we showed that because equation (2.13) is a quadratic function of $\mathbf{m}$, its derivative is linear in $\mathbf{m}$ (see Equation (2.17)). This first derivative hence coincides with zero for that model for which $U(\mathbf{m})$ attains its (global) minimum value. As such, setting this derivative to zero allows us to obtain a first estimate (i.e., prior means) of the moment tensor components. Setting the left-hand side of Equation (2.17) to zero yields

$$\mathbf{m} = \mathbf{A}^{-1}\left(\mathbf{A}\mathbf{m}^0 - \mathbf{b}\right) = \mathbf{m}^0 - \mathbf{A}^{-1}\mathbf{b}. \tag{2.19}$$

It should be understood that Equation (2.19) is implemented with T_0 and centroid fixed. Hence, the model vector has only six elements, and $\mathbf{A}$ is a six by six matrix. The quadratic nature of U in Equation (2.13) furthermore implies that arbitrary values can be chosen for the initial moment tensor components in $\mathbf{m}^0$. In fact, in the absence of noise and the correct prior means for the centroid and origin time, the MT priors estimated using Equation (2.19) will coincide with the true MT components.

In practice, the prior means resulting from Equation (2.19) may still deviate significantly from the true values due to the inaccuracy of the initial centroid and origin time priors. Solving Equation (2.19) nevertheless provides sufficiently accurate prior information regarding the magnitude of the induced event. Finally, it is useful to draw a parallel with typical least-squares optimization problems (e.g., Virieux and Operto, 2009).

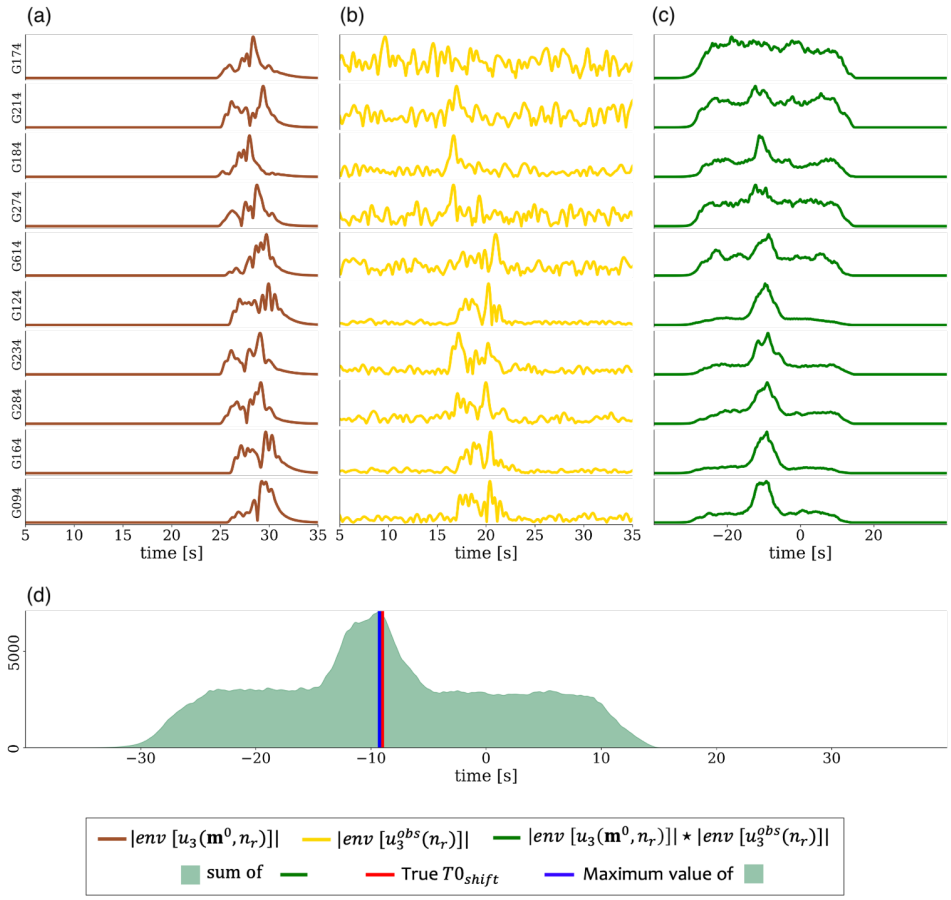

Figure 2.4: *The results of estimating the prior mean of origin time using Equation (2.18) (d), given the envelopes of modeled displacements (a), noisy synthetic "observed" seismograms (b), and the convolution between (a) and (b) in (c).*

In such context, $\mathbf{A}$ is analogous to the Hessian and the difference between $\mathbf{m}$ and $\mathbf{m}^0$ in Equation (2.19) can be considered the model update vector.

2.6.3. FULL WORKFLOW

In Figure 2.5, we illustrate our entire workflow. The main component of the workflow is the iterative HMC procedure, which is preceded by the (just-described) determination of the initial prior and succeeded by the evaluation of the posteriors.
The determination of the initial prior consist of the following four steps:

1. Estimate the initial prior mean for the centroid, either by running a first-arrival based probabilistic inversion algorithm or by extracting it from existing literature.

2. Estimate the initial prior mean of the origin time using (P-wave) travel times from

the centroid obtained in step 1 to the receiver locations. This estimate is refined by evaluating Equation (2.18) using an arbitrary MT.

3. Estimate the initial prior mean of the MT by fixing centroid and origin time to their prior means (steps 1 & 2) and solving Equation (2.19). The sought-after MT prior means are contained in $\mathbf{m}$ upon substitution of arbitrary MT components in $\mathbf{m}^0$.

4. Determine the standard deviation for each of the ten model parameters: centroid (3), origin time, and moment tensor (6). These standard deviations are needed to construct our first mass matrix $\mathscr{M}$. Ideally, $\mathscr{M}$ is the posterior covariance matrix. Here we approximate it by a (10x10) diagonal matrix with the following entries for the diagonal. For the first three entries (representing the centroid), we take the standard deviation of the centroid prior mean obtained in step 1. For the entry representing origin time, we use half the period of the dominant frequency in the recordings. For the MT components, we use 5% of the minimum absolute value of the MT prior means obtained by solving Equation (2.19).

Now that the (initial) prior means and standard deviations are determined, the HMC variant is run iteratively up to n chains (yellow box in Figure 2.5). A test for chain convergence might be required to determine the number of chains needed, and it is highly dependent on the quality of the prior means, data uncertainty, model uncertainty, initial model, and the dominant frequency of the observed recordings. In our example (detailed below), approximately ten chains are sufficient when the distance between the initial estimation of the centroid and the true centroid is less than 700 m. The separate steps of the iterative HMC procedure are:

5. Collect the prior means and associated standard deviations, and construct the mass matrix $\mathscr{M}$. For the first chain, the output from steps 1 to 4 is used as input. In subsequent chains, they are extracted from the posterior of the previous HMC chain. In this step, also $\mathbf{A}$ (Equation (2.14)), $\mathbf{b}$ (Equation (2.15)) and c (Equation (2.16)) are recomputed.

6. Run a new HMC chain with a preset number of iterations and burn-in period. Note, that for each chain, the results are stored for latter use.

7. Collect the results. The means and standard deviations will serve as input of for the next iteration (see step 5).

After a total of n HMC chains, we evaluate the posteriors (dark blue box in Figure 2.5). This involves:

8. For each of the n posteriors, compute the means $\mathbf{m}^s$ $(s = 1, ..., n)$. We use these means to generate synthetic recordings and evaluate them against the observed data through determination of the variance reduction (VR)

$$VR = 1 - \sqrt{\frac{\sum_{r=1}^{N_r}\sum_{i=1}^{3}\left(u_i(\mathbf{x}_r, t; \mathbf{m}^s) - u_i^{\text{obs}}(\mathbf{x}_r, t)\right)^2}{\sum_{r=1}^{N_r}\sum_{i=1}^{3}\left(u_i^{\text{obs}}(\mathbf{x}_r, t)\right)^2}} \tag{2.20}$$

9. Define a VR threshold. Posteriors associated with a $\mathbf{m}^s$ for which the VR exceeds this threshold are used to compute the final posterior distribution.

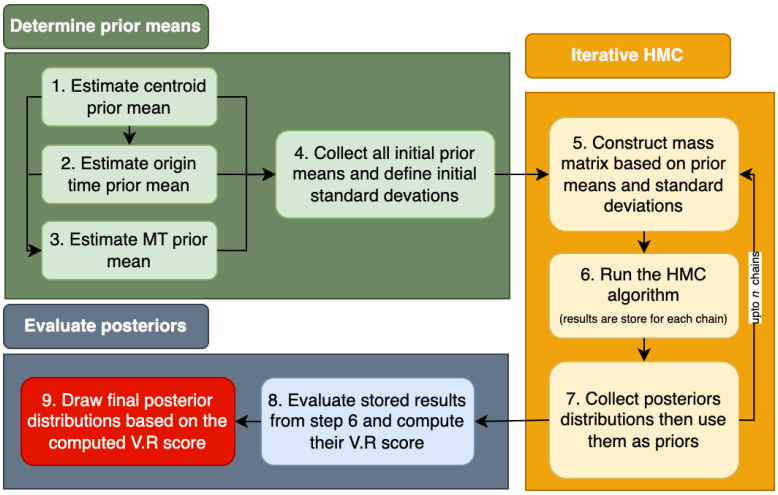

Figure 2.5: *Full workflow of our iterative HMC scheme.*

We use the above workflow to estimate the parameters of the synthetic event shown in Figure 2.2. In step 1, we assume a suitable prior of the centroid can be retrieved from literature (e.g., from Smith et al. (2020) in case of the seismicity in Groningen). To simulate the fact that this prior may well deviate from the true centroid, we shift this initial centroid prior mean by 600 m in all directions (i.e., with respect to the correct event location). Having the prior mean for the centroid, we follow steps 2 and 3 in the workflow to obtain the other prior means. To encode for a state of ignorance, we set the standard deviation σ_m of each model parameter to infinity, which implies that the last term of Equation (2.14) evaluates to 0. The elements of our initial mass matrix are taken from the results of steps 1-3 as explained in the full workflow (step 4), except for those elements that correspond to the centroid; these we set to 300 m. Using the initial prior means and the initial mass matrix, we run 20 chains of the HMC variant. Furthermore, we run 2500 iterations (step 6) for every chain, with the first 500 samples discarded as burn-in samples. After finishing all iterations, the results of each current chain are then used to update the prior means and mass matrix for the next HMC chain (the actual iterative HMC procedure). For each of the ten model parameters, all 40,000 samples (20 iterations × 200 samples) are depicted in Figure 2.6. To obtain our final posterior, we take the results of chains for which the means are associated with seismograms yielding a VR ≥ 85% of the maximum VR. In detail, the samples from all chains (black dots) and the selected chains (green dots) are depicted in Figure 2.7. For the samples of the selected chains, the one-dimensional marginal probability distributions of each of the ten source parameters are shown in Figure 2.8. In Figure 2.9, we show the synthetic (observed) seismograms (green), seismograms generated from the final posterior means (blue), and the true noise-free seismograms (red). With the use of a database containing pre-computed

elementary seismograms and using the python code we developed, the entire workflow takes approximately one minute to finish on a single-core CPU system.

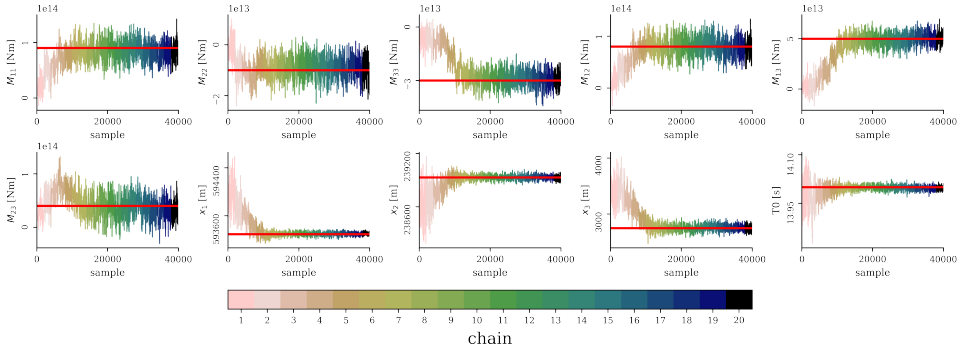

Figure 2.6: *Results of our iterative HMC scheme for a total of 20 chains, each involving 2500 steps of which the first 500 are discarded as burn-in samples (not shown). The red lines are the true values.*

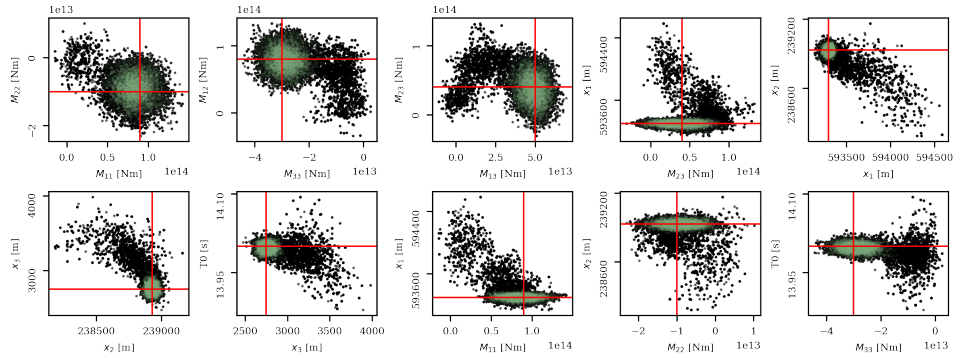

Figure 2.7: *Ten two-dimensional marginal probability densities of the inverted model parameters. Black dots are all the samples given the results from all chains (Figure 2.6), whereas the green dots represent the samples from chains that give VR≥ 85% of the maximum VR, which then represent our final posterior. The red lines are the true values.*

2.7. THE IMPORTANCE OF THE PRIOR

The above workflow might not be optimal if the initial prior information is "weak" in the sense that the initial centroid prior mean deviates significantly from the true value. This is due to the fact that our forward problem is in essence a non-linear problem, whereas the adopted linearization (Section 2.4) relies on the assumption of it being only weakly non-linear. In other words, poor initial centroid priors imply linearization around a location $\mathbf{x}$ that deviates too much from the true source location $\mathbf{x}_a$, which may result in the HMC algorithm "getting stuck" in local minima. This problem can be mitigated by running the workflow with multiple initial prior means. Depending on how close each of

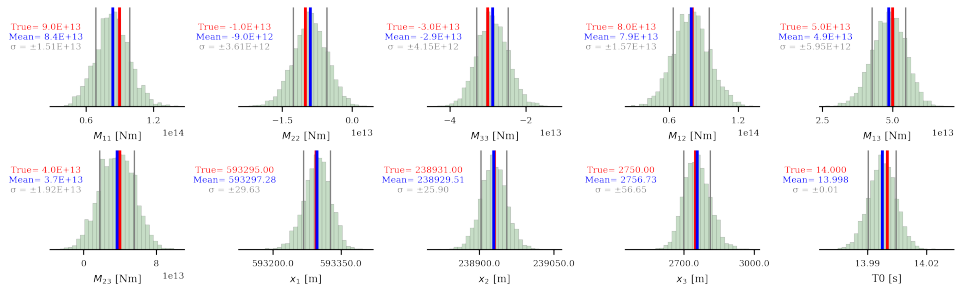

Figure 2.8: *The final marginal posterior distributions (green samples in Figure 2.7). The means are represented by the blue lines and the gray lines are the standard deviations. Red lines are the true values.*

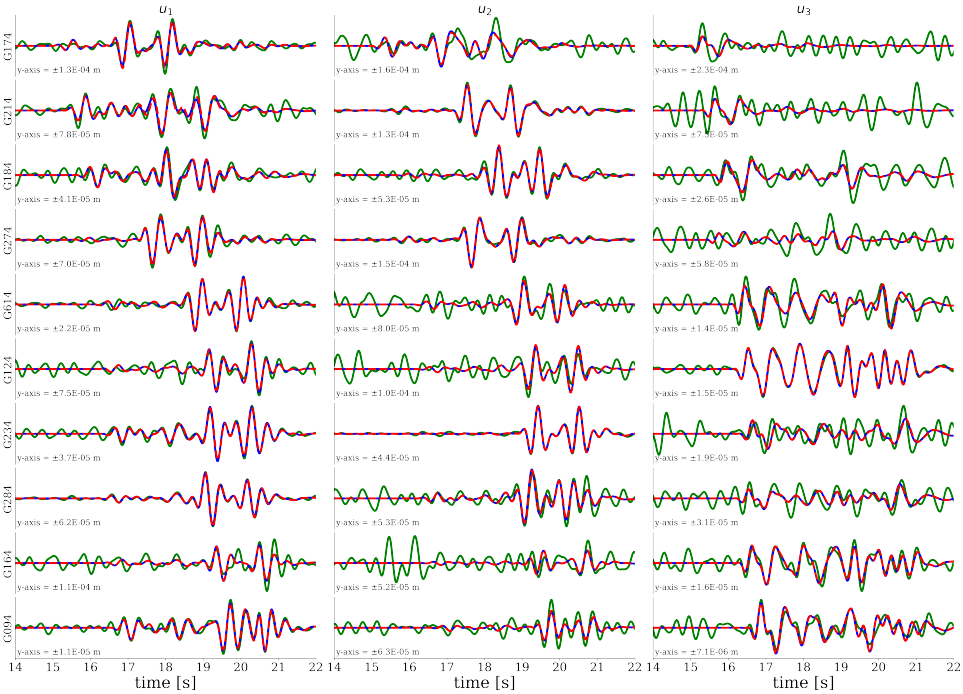

Figure 2.9: *Comparison between the true seismograms (red), observed (true+noise) seismograms (green), and seismograms generated using the final posterior mean (blue).*

the initial prior means is to the true values, some chains might get stuck in a local (minimum) mode while others correctly sample the mode containing the global minimum (or global maximum if one considers $\rho(\mathbf{m}|\mathbf{d})$). In the end, the final posterior can be drawn by combining the results of all chains given multiple initial prior means.

To showcase the effect of weak prior information in the context of induced seismicity in the Groningen gas field, we re-use the synthetic earthquake in Figure 2.2. However, instead of shifting the initial centroid prior mean by 600 m for all coordinate components,

we rigorously shift it by 1 km for each horizontal coordinate. Meanwhile, the depth is set to 3 km, corresponding to the default depth in the KNMI database because this database will be our primary source to obtain our priors. To get additional initial centroid prior means, we construct a 2.8 x 2.8 km 2D grid, at a depth of 3 km, with a spacing of 700 m, centered around the initial centroid prior mean. A pre-test can help in determining the grid spacing. We previously demonstrated that our workflow performs well when the centroid prior means are shifted by 600 m (in all directions) from their true values. This shift corresponds to an absolute deviation of about 1 km. Given the spacing of the constructed grid, assuming that the depth could be around ± 500 m, the maximum total distance is about 700 m, which is then considered safe for the HMC algorithm to sample the mode containing the global minimum.

Overall, given our 5x5 horizontal grid, we have 25 initial centroid prior means, each of them being subjected to our workflow. For each workflow run, we use the same model and data uncertainty as in the initial synthetic case. The same applies to the number of chains (20), samples per chain (2500), and burn-in period (500 samples). To reduce the computational time, we run the 25 workflows (associated with 25 initial centroid prior mean) simultaneously by parallelizing our code. We subsequently collect the results of each workflow to obtain an estimate of our posterior distribution. The results of this parallelization are summarized in Figure 2.10, which highlights the effect of the separation between the centroid prior mean and the true centroid. Using the same threshold as in our initial experiment (VR≥ 85% of maximum VR), in Figure 2.11, we show all (non burn-in) samples associated with the selected chains and samples from all chains given the calculated VR. For the selected chains, the marginal probability distribution of each parameter is presented in Figure 2.12. As expected, chains with an initial centroid prior mean relatively close to the true centroid converge to the true mode (containing the global minimum). At the same time, chains starting from a centroid further away from the true centroid, "get stuck" in a local mode. Fortunately, our VR strategy is still successful in picking appropriate chains, allowing us to obtain an estimate of the posterior distributions.

2.8. DISCUSSION AND CONCLUSIONS

Using synthetic events, we demonstrate that the proposed probabilistic workflow is able to efficiently estimate the posterior probability of the various parameters describing induced seismic events. 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.

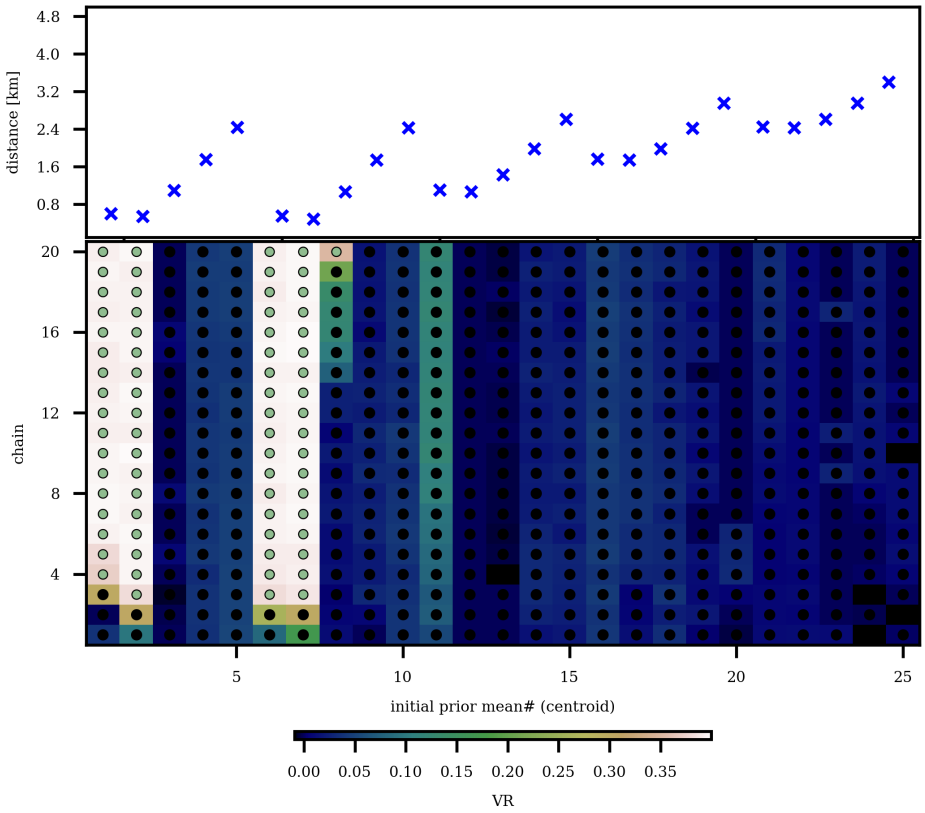

Figure 2.10: *Summary of running the workflow using the 25 initial prior means. The distance between each of the initial centroid prior means and the true centroid is indicated on top. Below, we show for each of these initial prior means the VR as a function of chain number (vertical axis). Chains associated with a VR*≥ 85% *of the maximum VR (0.4) are labeled with green dots, whereas chains with a posterior mean yielding seismograms for which the VR does not exceed* 85%*, are labeled with black dots.*

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 chapter.

The aforementioned deviation of the available numerical velocity model from the true subsurface velocities will pose a number of challenges. First, the estimated posterior probability would give a lower bound in terms of the variability of the source parameters: inaccuracies in the velocity model necessarily imply broader posterior probabilities. Second, in the presence of strong anisotropy, the posterior could be adversely affected. In particular, in the case of non-pure shear mechanisms, this effect could be significant (Ma et al., 2022). Third, cycle skipping will be particularly hard to tackle in case the velocity model is rather inaccurate.

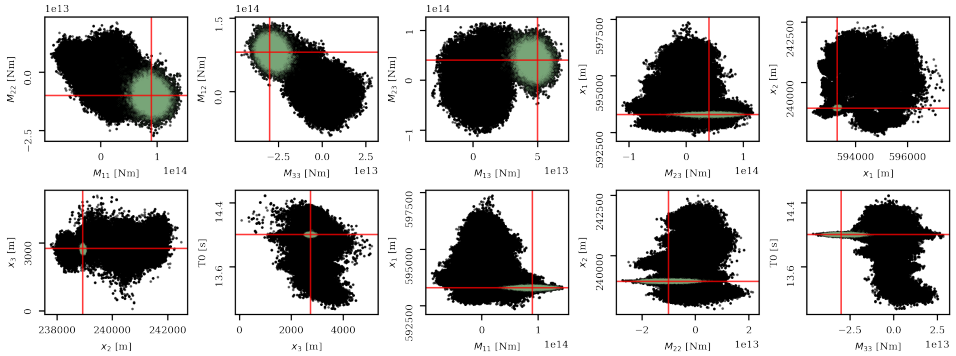

Figure 2.11: *Same ten two-dimensional marginal probability densities as in Figure 2.7. Note that scales on horizontal and vertical axes differ. Black dots represent the samples from all chains, whereas the green dots represent the samples from chains with a posterior mean that yields a VR that is higher than* 85% *of the maximum VR (the detail of the chain can be seen in Figure 2.10). The red lines represent the true values.*

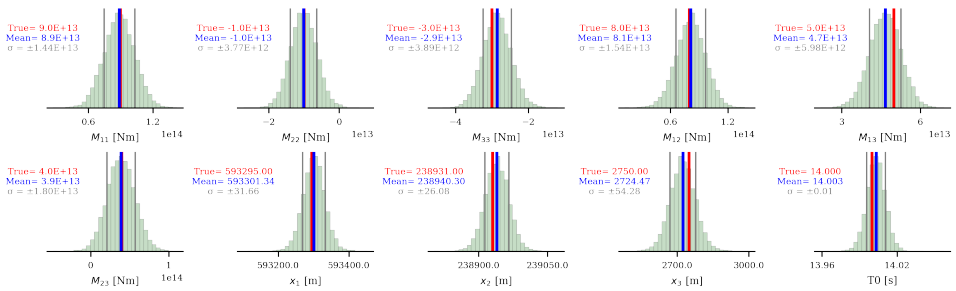

Figure 2.12: *Marginal posterior distributions for each model parameter given the selected chains (depicted as green dots in Figure 2.10). The blue lines represent the means, and the standard deviations are represented by the grey lines. The red lines represent the true values.*

Our workflow includes a systematic approach to obtain meaningful initial priors, which is particularly important for the employed HMC variant: the linearization of the forward problem around the prior mean requires the initial priors to be sufficiently close to the true event location. Furthermore, we show that by using an iterative scheme, we can update the prior mean such that convergence is obtained to a centroid location that allows the estimation of a meaningful posterior. The iterative scheme involves sequentially updating the prior mean of each new HMC chain using the posterior estimate obtained from the previous HMC chain. This approach is based on the suggestion of Fichtner and Simutė (2018) to repeat the Taylor expansion (of the forward problem) for each new sample of the Markov chain. However, we (only) do this every 2500 samples in our case. A brute-force approach to perform the expansion at each step of every chain would render computational costs prohibitively large.

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. Using a single-core CPU system, a single run of our workflow with 20 sequential HMC chains takes about one minute to finish, where each chain consists of 2500 iterations. In contrast, the computational costs of the Metropolis algorithm (to get the same results) would be much higher, as previously shown by Fichtner and Simutė (2018); Fichtner et al. (2019). Furthermore, various modifications could be applied to the workflow, such as adding simulated annealing and tempering (Tarantola, 2006), including a step to quantify the error in the input seismograms (Mustać and Tkalčić, 2016), and/or applying a scheme that is able to tune $d\tau$ and τ for each HMC run (Hoffman and Gelman, 2014). These modifications could be beneficial, especially when dealing with field observations, which is the subject of the next chapters. We also show that the workflow can be adapted to account for scenarios in which the initial centroid prior mean is rather inaccurate and/or the initial prior is weak. If that is the case, an approach can be adopted in which various iterative HMC workflows, each using a different centroid prior mean as a starting point, are run. Subsequently, using the variance reduction associated with the posterior means of the individual chains as a binary criterion for selecting a chain's samples, a final estimate of the posterior probability can be obtained.

We would like to emphasize that our workflow is, in principle, not limited to inversions of the parameters we use here. We could extend our probabilistic inversion to parameters such as stress drop, velocity, or to invert for finite fault source parameters. Furthermore, it is important to mention that our workflow aims to invert seismic source parameters using seismic surface recordings in a specific frequency range. That is, it is specifically geared towards inverting for induced seismic events. We found that the workflow works well when applied to data with frequencies between 1 and 3 Hz. For higher frequencies, however, some testing might be needed because the non-linearity between the input data and model parameters increases with increasing frequency.

REFERENCES

Agurto, H., Rietbrock, A., Ryder, I., and Miller, M. (2012). Seismic-afterslip characterization of the 2010 MW 8.8 Maule, Chile, earthquake based on moment tensor inversion. *Geophysical Research Letters*, 39(20):1–6.

Aki, K. and Richards, P. G. (2002). *Quantitative Seismology*. University Science Books, 2^{nd} edition.

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*.

Brinkman, N., Stähler, S. C., Giardini, D., Schmelzbach, C., Khan, A., Jacob, A., Fuji, N., Perrin, C., Lognonné, P., Beucler, E., et al. (2021). First focal mechanisms of marsquakes. *Journal of Geophysical Research: Planets*, 126(4):e2020JE006546.

Clarke, H., Verdon, J. P., Kettlety, T., Baird, A. F., and Kendall, J.-M. (2019). Real-time imaging, forecasting, and management of human-induced seismicity at Preston New Road, Lancashire, England. *Seismological Research Letters*, 90(5):1902–1915.

Dost, B., van Stiphout, A., Kühn, D., Kortekaas, M., Ruigrok, E., and Heimann, S. (2020). Probabilistic moment tensor inversion for hydrocarbon-induced seismicity in the Groningen gas field, the Netherlands, part 2: Application. *Bulletin of the Seismological Society of America*, 110(5):2112–2123.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics letters B*, 195(2):216–222.

Ekström, G., Dziewoński, A., Maternovskaya, N., and Nettles, M. (2005). Global seismicity of 2003: Centroid–moment-tensor solutions for 1087 earthquakes. *Physics of the Earth and Planetary Interiors*, 148(2-4):327–351.

Fichtner, A. and Simutė, S. (2018). Hamiltonian Monte Carlo inversion of seismic sources in complex media. *Journal of Geophysical Research: Solid Earth*, 123(4):2984–2999.

Fichtner, A., Zunino, A., and Gebraad, L. (2019). Hamiltonian Monte Carlo solution of tomographic inverse problems. *Geophysical Journal International*, 216(2):1344–1363.

Fichtner, A., Zunino, A., Gebraad, L., and Boehm, C. (2021). Autotuning Hamiltonian Monte Carlo for efficient generalized nullspace exploration. *Geophysical Journal International*, 227(2):941–968.

Gu, C., Marzouk, Y. M., and Toksöz, M. N. (2018). Waveform-based Bayesian full moment tensor inversion and uncertainty determination for the induced seismicity in an oil/gas field. *Geophysical Journal International*, 212(3):1963–1985.

Gutenberg, B. (1956). The energy of earthquakes. *Quarterly Journal of the Geological Society*, 112(1-4):1–14.

Hejrani, B., Tkalčić, H., and Fichtner, A. (2017). Centroid moment tensor catalogue using a 3-D continental scale Earth model: Application to earthquakes in Papua New Guinea and the Solomon Islands. *Journal of Geophysical Research: Solid Earth*, 122(7):5517–5543.

Hingee, M., Tkalčić, H., Fichtner, A., and Sambridge, M. (2011). Seismic moment tensor inversion using a 3-D structural model: Applications for the Australian region. *Geophysical Journal International*, 184(2):949–964.

Hoff, P. D. (2009). *A first course in Bayesian statistical methods*, volume 580. Springer.

Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623.

Jost, M. u. and Herrmann, R. (1989). A student's guide to and review of moment tensors. *Seismological Research Letters*, 60(2):37–57.

Kanamori, H. (1977). The energy release in great earthquakes. *Journal of geophysical research*, 82(20):2981–2987.

Kikuchi, M. and Kanamori, H. (1991). Inversion of complex body waves III. *Bulletin of the Seismological Society of America*, 81(6):2335–2350.

Komatitsch, D. and Tromp, J. (2002). Spectral-element simulations of global seismic wave propagation—I. Validation. *Geophysical Journal International*, 149(2):390–412.

Langenbruch, C., Weingarten, M., and Zoback, M. D. (2018). Physics-based forecasting of man-made earthquake hazards in Oklahoma and Kansas. *Nature Communications*, 9(1):1–10.

Lomax, A. (2005). A reanalysis of the hypocentral location and related observations for the great 1906 California earthquake. *Bulletin of the Seismological Society of America*, 95(3):861–877.

Ma, J., Wu, S., Zhao, Y., and Zhao, G. (2022). Cooperative P-Wave velocity measurement with full waveform moment tensor inversion in transversely anisotropic media. *Sensors*, 22(5):1935.

Masfara, L. and Weemstra, K. (2021). Towards efficient probabilistic characterisation of induced seismic sources in the Groningen gas field. In *EAGE GeoTech 2021 First EAGE Workshop on Induced Seismicity*, volume 2021, pages 1–5. European Association of Geoscientists & Engineers.

Mustać, M. and Tkalčić, H. (2016). Point source moment tensor inversion through a Bayesian hierarchical model. *Geophysical Journal International*, 204(1):311–323.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov chain Monte Carlo*, 2(11):2.

Noble, M., Gesret, A., and Belayouni, N. (2014). Accurate 3-D finite difference computation of traveltimes in strongly heterogeneous media. *Geophysical Journal International*, 199(3):1572–1585.

Ntinalexis, M., Bommer, J. J., Ruigrok, E., Edwards, B., Pinho, R., Dost, B., Correia, A. A., Uilenreef, J., Stafford, P. J., and van Elk, J. (2019). Ground-motion networks in the Groningen field: usability and consistency of surface recordings. *Journal of Seismology*, 23(6):1233–1253.

Pinar, A., Kuge, K., and Honkura, Y. (2003). Moment tensor inversion of recent small to moderate sized earthquakes: implications for seismic hazard and active tectonics beneath the Sea of Marmara. *Geophysical Journal International*, 153(1):133–145.

Romijn, R. (2017). Groningen velocity model 2017—Groningen full elastic velocity model September 2017 (available at https://nam-feitenencijfers.data-app.nl/download/rapport/9a5751d9-2ff5-4b6a-9c25-e37e76976bc1?open=true). *Nederlandse Aardolie Maatschappij (NAM)*.

Ruigrok, E. and Dost, B. (2019). Seismic monitoring and site-characterization with near-surface vertical arrays. In *Near Surface Geoscience Conference and Exhibition*, pages 1–5. European Association of Geoscientists Engineers.

Sen, A. T., Cesca, S., Bischoff, M., Meier, T., and Dahm, T. (2013). Automated full moment tensor inversion of coal mining-induced seismicity. *Geophysical Journal International*, 195(2):1267–1281.

Sen, M. K. and Stoffa, P. L. (2013). *Global optimization methods in geophysical inversion*. Cambridge University Press.

Sethian, J. A. and Popovici, A. M. (1999). 3-D traveltime computation using the fast marching method. *Geophysics*, 64(2):516–523.

Simute, S., Boehm, C., Krischer, L., Gokhberg, A., Vallée, M., and Fichtner, A. (2022). Bayesian seismic source inversion with a 3-D Earth model of the Japanese Islands. *Earth and Space Science Open Archive*, page 42.

Smith, J. D., White, R. S., Avouac, J.-P., and Bourne, S. (2020). Probabilistic earthquake locations of induced seismicity in the Groningen region, the Netherlands. *Geophysical Journal International*, 222(1):507–516.

Spetzler, J. and Dost, B. (2017). Hypocentre estimation of induced earthquakes in Groningen. *Geophysical Journal International*, 209(1):453–465.

Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature physics*, 2(8):492–494.

van Eck, T., Goutbeek, F., Haak, H., and Dost, B. (2006). Seismic hazard due to small-magnitude, shallow-source, induced earthquakes in The Netherlands. *Engineering geology*, 87(1-2):105–121.

van Thienen-Visser, K. and Breunese, J. (2015). Induced seismicity of the Groningen gas field: History and recent developments. *The Leading Edge*, 34(6):664–671.

Verdoes, A. and Boin, A. (2021). *Earthquakes in Groningen: Organized suppression of a creeping crisis*, pages 149–164. Springer International Publishing, Cham.

Virieux, J. and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC127–WCC152.

Waldhauser, F. and Ellsworth, W. L. (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California. *Bulletin of the seismological society of America*, 90(6):1353–1368.

Wapenaar, K. and Fokkema, J. (2006). Green's function representations for seismic interferometry. *Geophysics*, 71(4):SI33–SI46.

Willacy, C., van Dedem, E., Minisini, S., Li, J., Blokland, J. W., Das, I., and Droujinine, A. (2018). Application of full-waveform event location and moment-tensor inversion for Groningen induced seismicity. *The Leading Edge*, 37(2):92–99.

3

PROBABILISTIC CENTROID-MOMENT TENSOR INVERSIONS OF GEOMECHANICALLY SIMULATED WAVEFORMS

Understanding earthquake processes often begins with estimating the source parameters of the earthquake. This requires the use of inversion techniques in which the observed field data is systematically compared with synthetic data. In the process, however, multiple simplifications are used to generate the synthetic, such as the used point source assumption. A more realistic assumption is to use a finite fault assumption. Despite being more realistic, this will significantly increase computational needs when performing the inversion. To better understand how the point source assumption affects our inversion, we perform a synthetic experiment where we generate noise-free and noisy synthetic data by simulating the 2018 M_L 3.4 Zeerijp event in the Groningen gas field, the Netherlands. Specifically, we perform geomechanical simulations that consider the regional stress field and the Groningen field 3D subsurface model. The simulation then mimics the earthquake nucleation process of the Zeerijp event, which also satisfies the finite fault assumption. With the geomechanically generated synthetic observations, we use the inversion algorithm we developed in Chapter 2 to estimate earthquake parameters that correspond to the earthquake simulation. In the inversion, we compare both the noisy and noise-free synthetic observations with other synthetically generated data from 1D generalized Groningen ´s velocity models computed with the point source assumption. With the aim of understanding the limitation of the point source assumption and, in addition, the use of subsurface model simplification (via 1D velocity models for generating synthetic data), we find that the use of the point source assumption has little to no significant effect on

Parts of this chapter are currently being prepared for a journal article.

the inversion results. The estimated parameters when inverting using noisy and noise-free data are similar. We also compare those estimated parameters with what is observed during the geomechanical simulation. For example, we showcase how the estimated centroid locations match the distribution of displacements observed in the fault during the simulation as well as the estimated moment tensor, which corresponds to the fault orientation and movement that triggers the earthquake nucleation.

3.1. INTRODUCTION

Induced earthquakes, which occur due to human activities, usually receive less attention than natural earthquakes because of their smaller size and lower impact and because they are less frequently observed. In some cases, however, induced earthquakes may have the same destructive impact as natural earthquakes (Foulger et al., 2018). Primary factors that drive induced earthquakes are the extraction of subsurface resources (e.g., oil and gas) and the extraction of heat from the subsurface (i.e., geothermal energy production). The Groningen gas field is a notorious example of an area where induced earthquakes are triggered by fault reactivation driven by prolonged gas extraction. The first detected earthquake in the area occurred in 1986, and a rapid increase in seismicity was observed between 2002 and 2013 (Muntendam-Bos et al., 2022). Although seismic reactivation continues to take place, seismic activity decreased recently NAM (2023). The prolonged seismic activity has caused significant devastation, including the destruction of residences, which in turn had, and still has, profound effects on local communities (Bakema et al., 2018). Consequently, it is essential to study the mechanisms behind induced seismicity to assess seismic risks thoroughly and devise efficient strategies for mitigating these risks in impacted regions.

Geomechanical simulation is an indispensable tool while studying induced earthquakes. It leverages subsurface geological and physical properties such as fault geometry, stress distribution, and extraction scenarios to understand how an earthquake nucleates and propagates until it is observed at the surface (Van Wees et al., 2017; Buijze et al., 2017). It is occasionally paired with earthquake source inversion algorithms, which harness the observed seismic signals to estimate specific characteristics of (induced) seismic sources. In this study, we integrate both tools, with geomechanical simulation being used to solve the forward problem and the source inversion algorithm to solve the inverse problem. The application of earthquake source inversion algorithms is often contingent on the validity of specific assumptions regarding the earthquake source model and/or subsurface properties. A notable example is the frequently adopted point source approximation, which parametrizes the displacement along the fault by means of a nine-component tensor (the *moment tensor*; Aki and Richards, 2002). The approximation involves a simplification of the rupture process in the sense that an earthquake is collapsed to an infinitesimal point (the *centroid*). Mathematically, however, a rupture process may also be formulated to be a summation of multiple contributions, each associated with a different point on the fault plane and having its own displacement and time history. Rupturing processes along a finite fault can be simulated by geomechanical simulation. Seismograms obtained from such a simulation better represent true field observations. In this study, we will exploit these more "realistic" synthetic observations to test the validity of the point-source assumption as well as our inversion algorithm

itself.

In this study, we perform a centroid-moment tensor inversion of *simulated* recordings (or *synthetic* recordings) resulting from geomechanical simulation of a finite rupture represented by multiple, separately moving slip patches. The inversion itself is based on a Hamiltonian Monte Carlo (HMC) algorithm tailored to induced earthquake sources, which is explained in detail in Chapter 2. The geomechanical simulation, which uses the Groningen reservoir models (NAM, 2020), is designed to simulate induced seismicity below the village of Zeerijp. Specifically, the simulation aims to replicate the M_L 3.4 induced earthquake that rocked the village of Zeerijp in 2018. It is one of the largest events recorded in the Groningen area (the largest being a M_L 3.6 event near Huizinge in 2012). Our simulation also incorporates the geometry of the seismological network operated by the seismology division of the Royal Netherlands Meteorological Institute (KNMI) in the sense that the locations of 7 KNMI borehole seismometers define the positions where the waveforms generated by geomechanically simulated fault slip are modeled (see Figure 3.1).

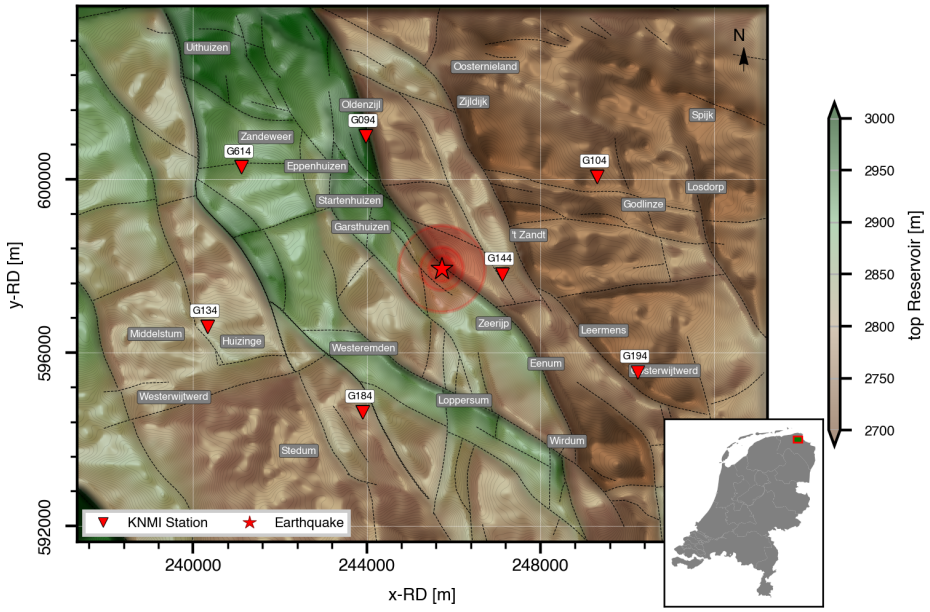

Figure 3.1: Epicenter of the M_L 3.4 Zeerijp earthquake estimated by KNMI (red star) with the top reservoir and fault map of the Groningen gas field below it. Locations of 7 vertical borehole arrays consisting of three-component geophones at depths of 50, 100, 150, and 200 meters are also depicted. Locations of local villages are shown in grey boxes.

In what follows, we will start by explaining the methodology used in this chapter. Particularly, we detail our approach for simulating an earthquake geomechanically. This is then followed by a subsection detailing our technique to make use of the waveforms from the simulation to obtain earthquake parameters corresponding to the simulated

earthquake. The parameter estimation is made possible by applying the HMC-based inversion algorithm to the modeled waveforms, which is discussed in the subsequent subsection. Finishing up the methodology, we delve into the result of the inversion given two cases; the modeled recordings are (i) noise-free and (ii) noisy. We then compare the results of the inversion for both scenarios with what has been observed in the geomechanical simulation before proceeding with the final subsection, highlighting our conclusions.

3.2. METHODOLOGY

We separately describe (i) the methodology associated with the (geomechanical) generation of the synthetic recordings and (ii) the methodology associated with the probabilistic centroid-moment tensor inversion. These are described in Sections 3.2.1 and 3.2.2, respectively. The geomechanical simulation provides insight into the displacement within our subsurface model, particularly along fault surfaces, enabling investigation into the nucleation process of earthquakes. In addition, it generates synthetic waveforms by propagating these displacements using waveform simulation techniques. The theory and approach detailed in Section 3.2.2 describes how these synthetic recordings are subsequently inverted, resulting in a posterior probability density of both the centroid and the moment tensor.

3.2.1. GEOMECHANICAL MODELING

To generate synthetic recordings via geomechanical simulation, we first create, down to a depth of 6 km, a 16 by 12 km one-dimensional finite difference subsurface model below our area of interest (16 km east-west and 12 km north-south). The one-dimensional velocity structure of this model coincides with the velocity structure of the NAM's velocity mode at the epicenter (Romijn, 2017). Within this subsurface model, a three-dimensional block of 2×2×1 km surrounding the hypocenter is embedded. The latter block is sliced from the larger reservoir model made available by the NAM (2020). This block extends 1 km down from a depth of 2350 m and includes several major faults. The block is discretized using a finite element mesh to capture the complexity of the included faults. We depict the sliced 3D block in Figure 3.2. In (a), we show the map representing the top reservoir, including the fault delineation of our model. Also, we define and mark four reference points to ease the interpretation. In (b), we show in 3D the delineation of the major faults within the selected block. In (c), we depict the geometry of the model/block, including the faults in (b), for a depth range of 1 km. Below, we summarize the adopted approach. For details, we refer to Ruan et al. (2023).

We utilize the open-source finite-element package `Defmod` developed by Meng (2017) to simulate the 2018 M_L 3.4 Zeerijp event in the Groningen gas field in our 2×2×1 km finite element model. In detail, our model is divided into four vertical regions, each with different distinct properties (i.e., P-wave velocity, S-wave velocity, and density). These are the overburden, the top seal, the reservoir, and the underburden. We then follow the work of Wentinck (2018) to define the materials and elastic parameters, as well as the regional stress setting applied to each vertical region of our model. Once the material properties are assigned, we initiate boundary reactions on the side and top walls, which

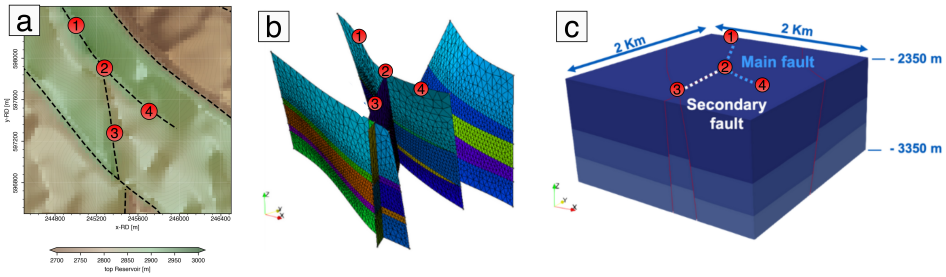

Figure 3.2: a) Top reservoir map of the research area, which is located near the M_L 3.4 Zeerijp event (see the area near the red star in Figure 3.1). Given the research area, the corresponding fault geometry and reservoir model extracted from NAM (2020) are shown in b) and c), respectively.

are based on the assigned regional stress regime. At the bottom, during initialization, we apply the roller boundary condition to prevent the occurrence of vertical displacement at the bottom of our model (similar to the approach used by Buijze et al. (2017, 2019) in a 2D medium). We assume the pore pressure to be 35 MPa at a depth of 3000 m. The hydrostatic pressure in each material is then calculated based on the depth and the density of gas and water. After initializing the model with pore pressure and boundary traction, we calculate the initial stress on the faults. To simulate the gas extraction process, we apply a uniform depletion scenario inside the reservoir until the nucleation of seismic events. The depletion phase is implemented in a quasi-static way as we consider the faults to remain stable. During the depletion, we calculate the incremental stresses and evaluate these on the fault.

After each quasi-static time step, we simulate dynamic fault slip if the fault becomes critical, that is, when the shear stress exceeds the friction. Friction at the fault is calculated using the linear slip-weakening law, with an initial friction coefficient of 0.6, a residual coefficient of 0.45, and a maximum weakening distance of 0.005 m. These values are slight modifications of those used in the 2D geomechanical simulations by Buijze et al. (2017), optimized here for our 3D simulations to generate seismic events. Next, the simulation outputs the fault slip until it is re-stabilized. During the dynamic simulation, which lasts for 10 seconds, two outcomes are possible. If the simulated rupture expands from the original slip patch as a self-sustained rupture, the simulated rupture is considered seismic; otherwise, it is aseismic.

When the depletion reaches 26 Mpa (i.e., at a pore fluid pressure of 9 Mpa), a seismic event with a magnitude equivalent to M_L 3.0 (assuming that $M_w = M_L$ (Dost et al., 2018)) nucleates at the main fault near the intersection with the secondary fault. The seismic slip starts at the top of the reservoir juxtaposition and then propagates along the fault in both strike and dip directions. In the end, the seismic slip patch is limited by the juxtaposition with a size of approximately 200 m × 200 m. The maximum slip is located in the middle of the reservoir interval.

To reduce the computational costs of propagating the resulting seismic wavefield, we combine the geomechanical simulation mesh (finite-element) with the aforementioned coarser regular mesh that supports finite-difference computations (the 16x12x6 one-dimensional velocity model). The level of grid coarseness in the finite-difference

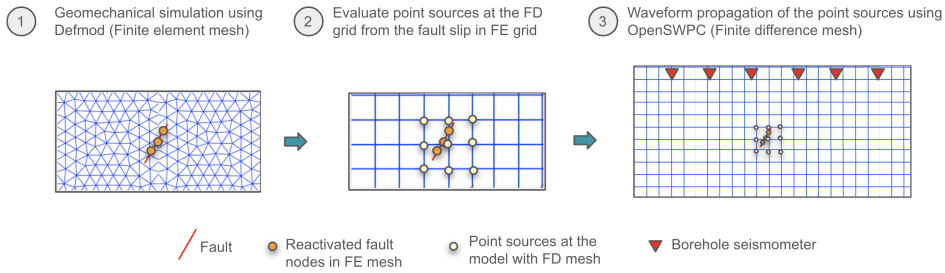

Figure 3.3: Schematic of our forward simulations to geomechanically generate synthetic seismograms. It uses a combination of geomechanical modeling taking place in the finite-element mesh (1) and then propagated into the nearby finite-difference mesh (2), which is then further propagated through the entire medium discretized with the finite-difference mesh until it gets recorded by our choice of receiver setup (3).

mesh is designed such that the recorded wavefield is not aliased at the maximum expected frequency. In figure 3.3, we illustrate the transition between the two meshes. The geomechanical simulation takes place in the finite element mesh. This way, the structure and shape of the fault can be better represented, as illustrated in Figure 3.3(1). When the fault becomes reactivated (orange dots), the displacement on the finite element node(s) is then assigned to adjacent finite difference nodes (yellow dots in Figure 3.3(2)) due to the binding and overlapping of both meshes (details can be found in Meng and Wang (2018)). Each of the yellow points will then act as a point source, which is then propagated in the finite-difference mesh using waveform simulation code OpenSWPC developed by Maeda et al. (2017) as depicted in Figure 3.3(3). The propagated waveforms are then recorded at specific locations, which, in our case, coincide with the coordinates of our selected KNMI borehole seismometers situated at 200 m depth.

3.2.2. A PROBABILISTIC CENTROID-MOMENT TENSOR INVERSION

In this Section, we describe the methodology associated with the probabilistic centroid-moment tensor inversion. Our inversion is based on a Hamiltonian Monte Carlo (HMC) algorithm and is tailored to induced earthquake sources; the description here effectively summarizes the theory detailed in Chapter 2.

THE FORWARD PROBLEM

The moment tensor is a point-source representation of an earthquake, where this point is usually referred to as the centroid. In practice, it depends on the strength of the (finite) seismic source (the strength of the earthquake being quantified by its seismic moment in Newton meter), the fault's strike and dip, and the rake describing the direction of relative rupture along the fault. Seismograms due to a point source described by a moment tensor read (Aki and Richards, 2002),

$$u_i(\mathbf{x}_r, t) = \sum_{j=1}^{3} \sum_{k=1}^{3} M_{jk}(t, \mathbf{x}_a, T_0) * G_{ij,k}(\mathbf{x}_r, t; \mathbf{x}_a). \tag{3.1}$$

Here, u_i represents the displacement u in the x_i direction at the observation point $\mathbf{x}_r$,

given the source located at $\mathbf{x}_a$ and with a rupture propagation starting at time T_0. Axes x_1, x_2, and x_3 correspond to east, north, and vertical (positive downward), respectively. The moment tensor $\mathbf{M}$, which is a symmetric 3×3 matrix, has six independent components. Each component M_{jk} represents a moment due to the force in the x_j direction and an arm in the x_k direction. The Green's tensor, denoted by $\mathbf{G}$, has indices i and j, where the first index indicates the axis along which the displacement is recorded (i.e., it yields u_i) and the second index the direction in which the (moment tensor) force is acting. The comma following the second subscript implies a spatial derivative in the x_k direction with respect to $\mathbf{x}_a$. The asterisk $*$ symbolizes convolution in the time domain.

It should be understood that the displacement at the left-hand side of Equation 3.1 is the displacement due to the activation of a single point (one orange dot) of our mesh in Figure 3.3, whereas, in our geomechanical simulation, multiple points are of course activated. The displacement at $\mathbf{x}_r$ for such a (extended) source is effectively the result of an integration over the different contributions along the fault plane and can be written (Aki and Richards, 2002, ; equation 3.20)

$$u_i(\mathbf{x}_r, t) = \sum_{j=1}^{3} \sum_{k=1}^{3} \int_A M_{jk}(t; \mathbf{x}_a, T_0) * G_{ij,k}(\mathbf{x}_r, t; \mathbf{x}_a)\, dA \tag{3.2}$$

with u_i now representing displacements due to a rupture along a fault plane with a surface area A and $\mathbf{x}_a$ the position on that fault plane; $M_{jk}(t; \mathbf{x}_a, T_0)$ varies as a function of $\mathbf{x}_a$. By using Equation 3.1 to invert the synthetic recordings due to geomechanical modeled (finite) fault slip and comparing the obtained source characteristics to the geomechanical parameters describing the rupture along the finite fault, we effectively investigate the validity of replacing Equation 3.2 by Equation 3.1 in the inversion. The fact that, for the inversion, Equation 3.1 is implemented assuming a simplified one-dimensional velocity model of the subsurface below Groningen, which lacks detail in comparison to the 3D model we use for the Geomechanical simulation (within the 2x2x1 km finite element block), gives rise to some additional discrepancies between the recovered parameters and the ones associated with the geomechanically simulated rupture.

INVERSE PROBLEM

Here, we summarize the approach we use to estimate the posterior probability density of the centroid moment tensor using Bayesian inference (detailed in Chapter 2). The posterior is obtained (up to a constant of proportionality) by multiplying two quantities: the likelihood $\rho(\mathbf{m}|\mathbf{d}^{\text{obs}})$ and the prior information $\rho(\mathbf{m})$. When computing $\rho(\mathbf{m}|\mathbf{d}^{\text{obs}})$, we essentially quantify the discrepancy (misfit) between synthetic data resulting from the geomechanical simulation (here being the *observed recordings* u_i^{obs}) and recordings u_i resulting from forward modeling using Equation 3.1 given parameters concatenated in the vector $\mathbf{m}$. Assuming that the noise on the synthetic recordings u_i^{obs} is uncorrelated and having a Gaussian distribution, $\rho(\mathbf{m}|\mathbf{d}^{\text{obs}})$ can be formulated as

$$\rho(\mathbf{m}|\mathbf{u}^{obs}) = \exp\Big[-\frac{1}{2T} \sum_{r=1}^{N_r} \sum_{i=1}^{3} \int_0^T \sigma_d^{-2} \Big[u_i(\mathbf{x}_r, t; \mathbf{m}) - u_i^{\text{obs}}(\mathbf{x}_r, t) \Big]^2 \mathrm{d}t \Big] \tag{3.3}$$

where N_r represents the number of receivers and σ_d^2 the data variance. Solving an inverse problem involves the identification of model parameters $\mathbf{m}$ that best approximate

the source characteristics. In this case, the prior information of $\mathbf{m}$ can be computed as

$$\rho(\mathbf{m}) = \exp\Big[-\frac{1}{2N_m} (\mathbf{m}-\mathbf{m}^0)^\top \mathbf{C}_m^{-1} (\mathbf{m}-\mathbf{m}^0) \Big] \tag{3.4}$$

where model vector $\mathbf{m}$ contains ten earthquake source parameters, with $\mathbf{m}^0$ the vector containing the mean of the prior of those parameters, whereas $\mathbf{C}_m$ is their covariance matrix. In search for the optimum $\mathbf{m}$, inversion algorithms are employed, which could be either deterministic or probabilistic. While deterministic algorithms offer lower computational costs in comparison to probabilistic ones, probabilistic algorithms are known for their ability to encode and output uncertainties of their estimation.

In the context of earthquake source inversion, a probabilistic algorithm (or sampler) with relatively low computational requirements is the variant of the Hamiltonian Monte Carlo (HMC) algorithm introduced by Fichtner and Simutė (2018), which involves a linearization of the forward problem (about the model vector $\mathbf{m}^0$). This modified sampler, however, requires a relatively accurate *a priori* estimate of this model vector, which means that the $\mathbf{m}^0$ should be relatively close to the true $\mathbf{m}$. This is typically the case for (larger) natural earthquakes, which generate lower-frequency signals. When $\mathbf{m}^0$ is inaccurate, however, the modified sampler would not only be efficient but also inaccurate in its output. To overcome this, we make use of the inversion algorithm explained in Chapter 2, which is (yet another) modification of the modified HMC variant. All in all, this new algorithm aims to improve the efficiency of the inversion process without losing the accuracy of the estimated posterior probability.

INVERSION PARAMETERS

Given the forward problem and the inverse problem described above, we will now detail the setup and inversion parameters we use to estimate the earthquake parameters. In general, we adopt a previously tested setup, which was validated using the Groningen subsurface characteristics in Chapter 2. First, we adopt a frequency range similar to the one used by Dost et al. (2020), which implies we apply a 1-4 Hz bandpass filter. We also adhere to their tapering recommendations, leading us to select a 3 s measurement window with both ends tapered for 0.9 seconds. This approach ensures we capture not only the P-waves but also certain portions of the S-waves, especially from receivers in close proximity to the initial epicenter estimate. This is vital because our S-wave velocity model, used for generating point source waveforms u_i, lacks accuracy compared to the P-wave model (Romijn, 2017). The further the receiver, the higher the uncertainty of the modeled S-wave segments and the more uncertain the posterior. Nevertheless, the S-wave carries more energy associated with higher amplitude and higher signal-to-noise ratio compared to the P-wave. Hence, having an accurately modeled S-wave will better constrain the inversion when dealing with field data, which, in this synthetic study, is captured by the taper exclusively to receivers near the epicenter.

When it comes to the synthetic recordings $u_i^{\text{obs}}(\mathbf{x}_r, t)$, we distinguish between two cases. One is noise-free, and the other has some additive random noise. This noise is obtained by generating white noise with 10 % of the maximum amplitude of the noise-free data for each receiver and each component in the frequency domain (later, these values will also be used as entries for σ_d). As such, the noise not only results in ampli-

tude variations but also introduces (small) time shifts affecting the origin time and location accuracy in the inversion process. For the model uncertainties, which are encoded by $\mathbf{C}_m$, we use several sources. For the hypocenter, we comprehend a study delivered by Smith et al. (2020) regarding the probabilistic estimation of earthquake hypocenters. Here, given 5% of velocity model error, they obtain ± 100 m of deviation in the hypocenter. For the moment tensor, we set it to infinity to encode the state of ignorance. As for the origin time, we use an arbitrary value of 0.1 s.

Another vital element is the choice of $\mathbf{m}^0$. In this study, similar to what is showcased in Chapter 2, we generate multiple realizations of $\mathbf{m}^0$ surrounding the hypocenter estimated by the KNMI. Specifically, we take the epicenter from KNMI (see the red star in Figure 3.1) and draw a circle with a radius of 700 m. Fault segments that are captured inside this circle are then discretized laterally every 200 m. In total, we obtained 14 points, leading to the generation of 14 sets $\mathbf{m}^0$. Each of the sets will have 10 entries representing the parameters we seek to estimate. The entries of the epicenter (first two coordinates of the centroid) will be taken from the discretized fault points. The depth of all $\mathbf{m}^0$ is set to 3 km, and the origin time is set to 0 s. Finally, the entries to the moment tensor are taken from the fault geometries at each discretized fault point. Specifically, we convert the strike and dip at each point to the moment tensor with the assumption that the earthquake is driven by pure normal fault movement (the rake is set to 90°). Sampling the posterior by means of multiple (multi-stage) HMC chains is described in detail in Chapter 4.

3.3. RESULTS AND DISCUSSION

This section presents and discusses the results of our inversion. Specifically, we highlight the differences between the estimated parameters of the inversions of the noise-free and noisy synthetic seismograms. Our analysis is divided into three parts. First, we examine the posterior distribution of our estimated earthquake parameters for these two different datasets. We then convert those distributions to earthquake mechanisms and the properties of the moving fault in terms of strike, dip, and rake. Second, we showcase the seismograms generated using the estimated parameters (with the point-source assumption) and compare them with the synthetic seismograms. Third, we compare the estimated centroids with the (planar) locations where displacements occurred during the geomechanical simulation.

We present the estimated posterior probability by means of (1D) marginal probability density functions in Figure 3.4. For all ten parameters, the marginals associated with the noisy and noise-free synthetic recordings pretty much coincide. In general, we find no significant difference in both mean values and standard deviation. We then continue with estimating the parameters of the simulated rupture by converting the six probability distributions representing moment tensors in Figure 3.4 into probability densities of strike, dip, and rake (Figure 3.5). Converting those, we obtain two sets of solutions. To narrow down our analysis, we only select solutions that give strike values between 300 and 320° which is the range of strike from the closest fault segment given the mean value of the estimated centroid locations. Obviously, such selection is justified only when prior information regarding subsurface fault orientations is available. In addition, we decompose the six moment tensor distributions into three (end-member) types, which are ISO

(Isotropic), DC (Double-coupled), and CLVD (Compensated linear vector dipole) (e.g., Mustać and Tkalčić, 2016; Cesca et al., 2013; Vavryčuk, 2015). We display these earthquake mechanisms in two forms. First, we plot these as 1D marginal probability density, and second, we project those densities in the Hudson plot to better see the distribution of the source mechanism/type (Hudson et al., 1989); both are presented in Figure 3.5. As depicted, the mean value of the strike is aligned with the strike of the nearby fault segment (which was used for the geomechanical simulation). Similarly, for the dip, the estimated mean value is 78°, which is close to the dip of the closest fault segment, which is in the range of 75° to 80°. As for the rake, we found that both probabilistic inversion results yield, on average, values close to 90°.

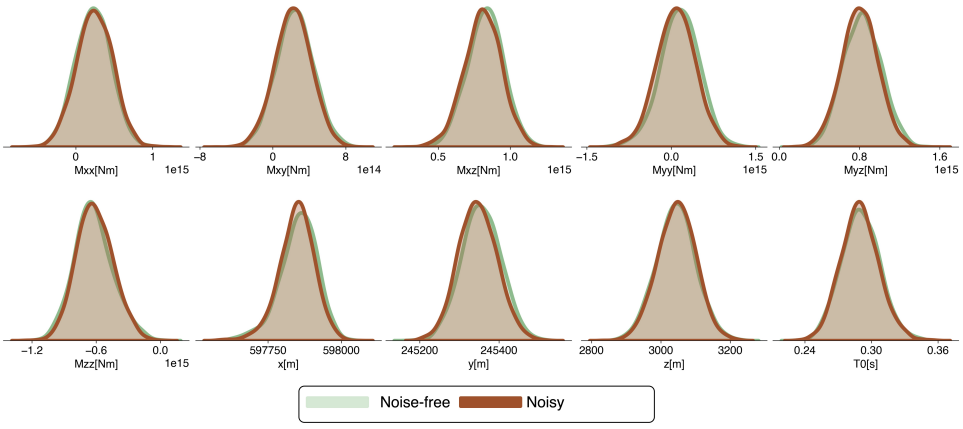

Figure 3.4: Marginal probability densities of the HMC-derived posterior probability density based on both the noise-free and noisy synthetic data (i.e., resulting from the geomechanical simulations).

Decomposing the recovered moment tensors, we find that the simulation is mainly driven by DC mechanisms for both datasets. In the Hudson plot, this manifests itself in the higher probability density at the center of the axis. The CLVD percentages, on the other hand, are different for the noisy and noise-free recordings, with one slightly positive and another slightly negative. Finally, the posterior means estimated from both datasets indicate negative isotropic components. In the geomechanical calculation, we do not allow for any volume changes, and hence, the found dominant DC is expected. The presence of the ISO and CLVD components is the result of the workflow trying to best fit the seismogram using multiple combinations of moment tensors (and other parameters). This includes adding slight ISO and CLVD components to the solutions. In real events, the presence of ISO components could be the result of volume loss due to reservoir depletion (negative ISO). As for the CLVD, this is often attributed to a complex seismic process. The positive CLVD physically represents a situation where a portion of the subsurface is being compressed along the vertical axis and expanding in the two lateral axes, such as in the activation of ring fault (Shuler et al., 2013). Conversely, negative CLVD represents the opposite, that is, where a portion of the subsurface is expanded in the vertical axis but compressed in the two lateral axes, such as in the case of collapse

mining (Rudziński et al., 2016). However, the presence of CLVD is sometimes also attributed to the presence of noise in the data or high seismic anisotropy in the subsurface (Stierle et al., 2014).

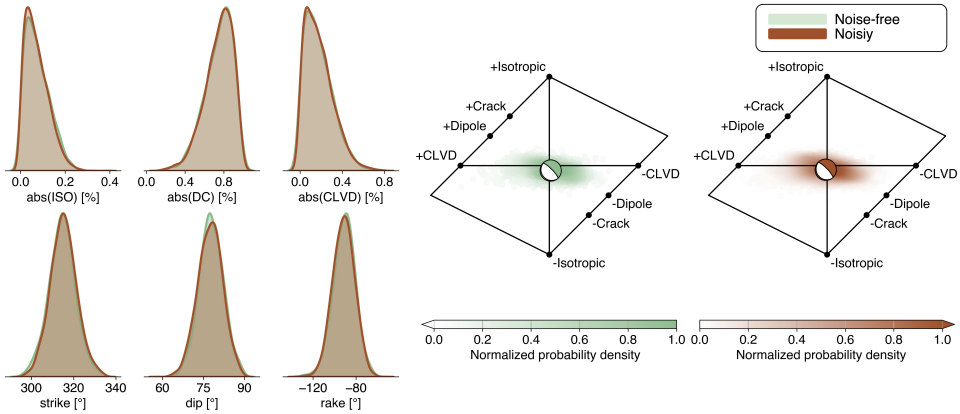

Figure 3.5: Results of converting and decomposing 1D probability functions of Moment tensor solutions in Figure 3.4 (the first six density functions) into the strike, dip, rake solutions, and earthquake mechanisms which are the Isotropic (ISO), Compensated linear vector dipole (CLVD), and Doubled-coupled (DC).

In Figure 3.6, we examine the fit of the synthetic data (i.e., resulting from the geomechanical simulation described in Section 3.2.1) with synthetic recordings generated using the mean of the estimated posterior probability density (associated with the noisy synthetic dataset). Not surprisingly, the generated synthetic recordings match both simulated observed datasets. The match is preserved not only in the measured area where the taper is being applied but also beyond it. Note that the green shaded area in Figure 3.6 is the part where all datasets are being measured fully (i.e., effectively the u_i^{obs} and u_i serving as input to the probabilistic inversion) and the brown areas are where cosine tapers are being applied. Despite the good match, we still observe significant discrepancies between the generated synthetic data (using point source assumption) and the noise-free synthetic datasets. These discrepancies are most likely not caused by inaccurate source parameters; rather, they stem from differences in the underlying velocity models used to generate the datasets and the fact that a centroid moment tensor source yield deviates from rupture along a finite fault. The difference between the waveforms essentially reflects the limitations inherent in simplifying complex physical systems into idealized models. Focusing on the noisy recordings, we show that the introduction of frequency domain noise not only varies the amplitude but also causes phase shifts, further complicating the inversion process. This variation, however, is not enough to deviate the estimated posteriors from the inversion using the noise-free dataset. This stability in the inversion results indicates that the algorithm is robust against moderate noise levels, and the variation is captured by the standard deviation fed in the inversion.

We now show the comparison of centroids against the observed displacement in the fault plane in Figure 3.7. On the left side, we show the means and covariances of our estimated epicenter on top of the major fault map of the Groningen gas field. Note that the

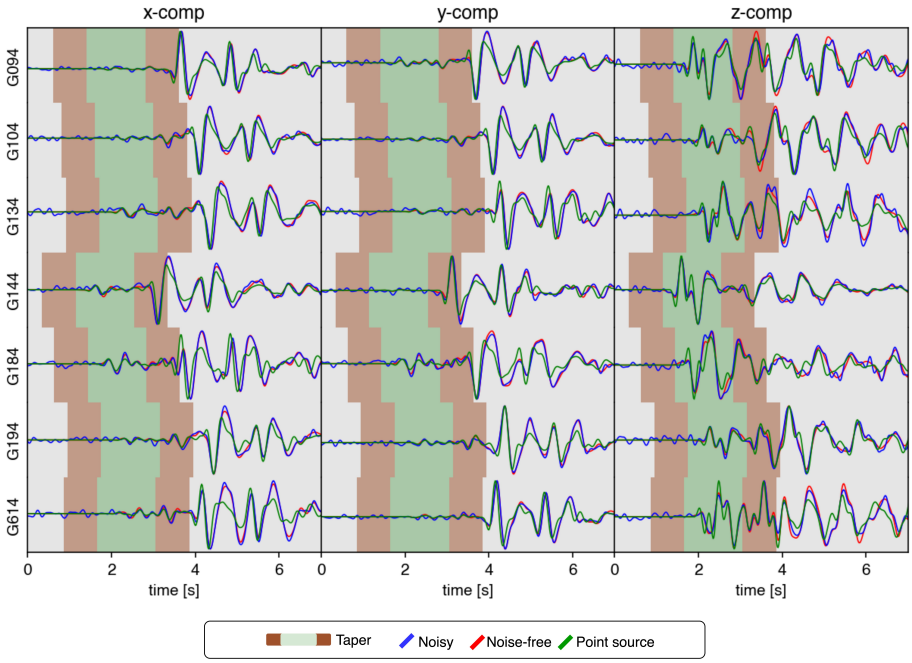

Figure 3.6: Comparison of seismograms. Green seismograms are generated using the point source assumption given the mean values of the estimated posteriors using the noisy geomechanically generated synthetic data (blue seismograms). Red seismograms are the noise-free geomechanically generated synthetic seismograms. The shaded area represents the measuring window, with the green shade representing the fully measured area (no taper) and the brown area representing the area where cosine tapers are applied to the seismograms.

fault is dipping towards the northeast, and as expected, the estimated values sit on the simulated fault. In the right panels, we show the snapshot of the observed displacement at 0.1, 0.29, and 0.8 s. The shown displacement is the projection of the along-fault displacement on the x-axis. As presented, the estimated covariances encircle the location with the maximum displacement. The mean origin time of the posterior probability density is 0.29; it is the same for both the noisy and noise-free synthetic recordings. These values approximately coincide with the time of maximum displacement.

3.4. CONCLUSION

In this study, we estimate the posterior probability density of ten earthquake source parameters describing (a point-source approximation of) a geomechanically simulated induced seismic rupture. That is, we effectively investigate the validity of using a point source approximation for (induced) slip along a finite fault in the context of seismic source inversion of recorded seismograms at the Earth's surface. In addition, however, the seismograms resulting from the solution of the forward model in the Hamiltonian Monte Carlo sampler are generated using a more simplistic 1D model compared to the velocity model used to propagate the geomechanically displacements along the finite

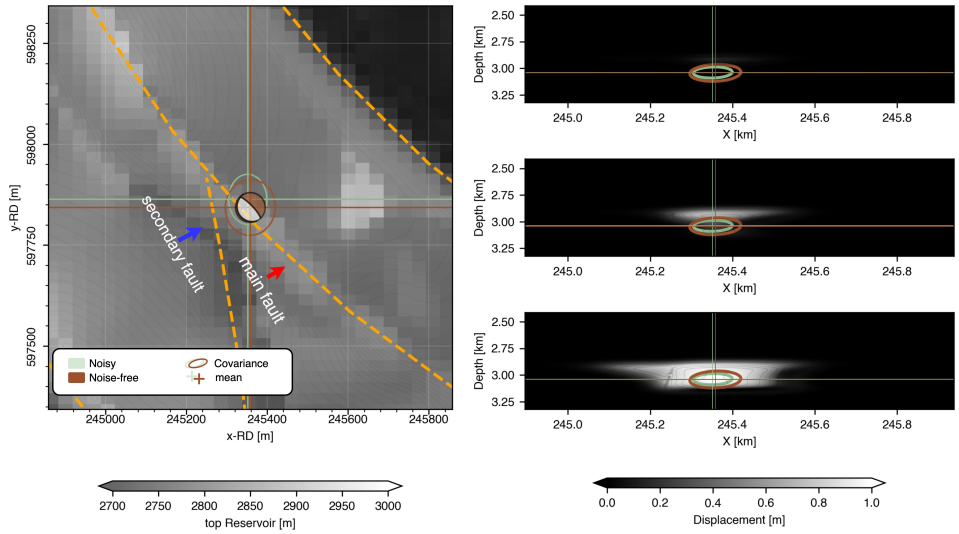

Figure 3.7: Left: Estimated centroid means, and covariances overlaid with the Groningen fault map. the main fault and the secondary faults are parts of the geomechanical simulation. Right panels: Seismic slip snapshot during the geomechanical simulation on the main fault projected in the x-axis; from top to bottom: 0.1, 0.29, and 0.8 s.

fault. Despite these differences, the inversion algorithm appears to be robust in the sense that the centroid, the moment tensor, and the origin time well approximate the (finite fault) rupture, even in the presence of moderate noise. Our inversion reveals that the (geomechanically) simulated event is dominated by the DC component, with slight variations in the CLVD, likely reflecting noise interference. In addition, we also observed negative ISO components, which are the result of the workflow trying to fit the waveforms from simulated events with multiple combinations of moment tensors (hence mechanisms). In application to field data, a significant negative ISO may comply with scenarios involving volume loss, such as reservoir depletion. Next, the centroid location estimations aligned well with the observed displacement on the fault plane, particularly at the moment when maximum displacement is observed, providing further validation of the accuracy of the inversion process. The ability of the inversion method to handle noise and maintain accurate estimations underscores its potential application in real seismic monitoring, where data is often contaminated with noise and other uncertainties.

REFERENCES

Aki, K. and Richards, P. G. (2002). *Quantitative Seismology*. University Science Books, 2^{nd} edition.

Bakema, M. M., Parra, C., and McCann, P. (2018). Analyzing the social lead-up to a

human-induced disaster: The gas extraction-earthquake nexus in Groningen, The Netherlands. *Sustainability*, 10(10):3621.

Buijze, L., van den Bogert, P. A., Wassing, B. B., and Orlic, B. (2019). Nucleation and arrest of dynamic rupture induced by reservoir depletion. *Journal of Geophysical Research: Solid Earth*, 124(4):3620–3645.

Buijze, L., Van Den Bogert, P. A., Wassing, B. B., Orlic, B., and Ten Veen, J. (2017). Fault reactivation mechanisms and dynamic rupture modelling of depletion-induced seismic events in a Rotliegend gas reservoir. *Netherlands Journal of Geosciences*, 96(5):s131–s148.

Cesca, S., Rohr, A., and Dahm, T. (2013). Discrimination of induced seismicity by full moment tensor inversion and decomposition. *Journal of seismology*, 17:147–163.

Dost, B., Edwards, B., and Bommer, J. J. (2018). The relationship between M and ML: a review and application to induced seismicity in the Groningen Gas Field, The Netherlands. *Seismological Research Letters*, 89(3):1062–1074.

Dost, B., van Stiphout, A., Kühn, D., Kortekaas, M., Ruigrok, E., and Heimann, S. (2020). Probabilistic moment tensor inversion for hydrocarbon-induced seismicity in the Groningen gas field, the Netherlands, part 2: Application. *Bulletin of the Seismological Society of America*, 110(5):2112–2123.

Fichtner, A. and Simutė, S. (2018). Hamiltonian Monte Carlo inversion of seismic sources in complex media. *Journal of Geophysical Research: Solid Earth*, 123(4):2984–2999.

Foulger, G. R., Wilson, M. P., Gluyas, J. G., Julian, B. R., and Davies, R. J. (2018). Global review of human-induced earthquakes. *Earth-Science Reviews*, 178:438–514.

Hudson, J., Pearce, R., and Rogers, R. (1989). Source type plot for inversion of the moment tensor. *Journal of Geophysical Research: Solid Earth*, 94(B1):765–774.

Maeda, T., Takemura, S., and Furumura, T. (2017). OpenSWPC: an open-source integrated parallel simulation code for modeling seismic wave propagation in 3D heterogeneous viscoelastic media. *Earth, Planets and Space*, 69(1):102.

Meng, C. (2017). Benchmarking Defmod, an open source FEM code for modeling episodic fault rupture. *Computers & Geosciences*, 100:10–26.

Meng, C. and Wang, H. (2018). A finite element and finite difference mixed approach for modeling fault rupture and ground motion. *Computers & Geosciences*, 113:54–69.

Muntendam-Bos, A. G., Hoedeman, G., Polychronopoulou, K., Draganov, D., Weemstra, C., van der Zee, W., Bakker, R. R., and Roest, H. (2022). An overview of induced seismicity in the Netherlands. *Netherlands Journal of Geosciences*, 101:e1.

Mustać, M. and Tkalčić, H. (2016). Point source moment tensor inversion through a Bayesian hierarchical model. *Geophysical Journal International*, 204(1):311–323.

NAM (2020). Petrel geological model of the Groningen gas field, the Netherlands (available at: https://public.yoda.uu.nl/geo/UU01/1QH0MW.html). *Open access through EPOS-NL*.

NAM (2023). Rapportage seismiciteit Groningen - November 2023 (available at: https://open.overheid.nl/documenten/18073003-36c5-4b73-9f98-aa976e34aa11/file). *NAM (Nederlandse Aardolie Maatschappij) Annual Report*.

Romijn, R. (2017). Groningen velocity model 2017—Groningen full elastic velocity model September 2017 (available at https://nam-feitenencijfers.data-app.nl/download/rapport/9a5751d9-2ff5-4b6a-9c25-e37e76976bc1?open=true). *Nederlandse Aardolie Maatschappij (NAM)*.

Ruan, J., Ghose, R., and Mulder, W. (2023). 3D geomechanical modelling of induced seismicity including intersecting faults and reservoir compartments. In *84th EAGE Annual Conference & Exhibition*, volume 2023, pages 1–5. European Association of Geoscientists & Engineers.

Rudziński, Ł., Cesca, S., and Lizurek, G. (2016). Complex rupture process of the 19 March 2013, Rudna Mine (Poland) induced seismic event and collapse in the light of local and regional moment tensor inversion. *Seismological Research Letters*, 87(2A):274–284.

Shuler, A., Ekström, G., and Nettles, M. (2013). Physical mechanisms for vertical-CLVD earthquakes at active volcanoes. *Journal of Geophysical Research: Solid Earth*, 118(4):1569–1586.

Smith, J. D., White, R. S., Avouac, J.-P., and Bourne, S. (2020). Probabilistic earthquake locations of induced seismicity in the Groningen region, the Netherlands. *Geophysical Journal International*, 222(1):507–516.

Stierle, E., Vavryčuk, V., Šílenỳ, J., and Bohnhoff, M. (2014). Resolution of non-double-couple components in the seismic moment tensor using regional networks—I: a synthetic case study. *Geophysical Journal International*, 196(3):1869–1877.

Van Wees, J.-D., Fokker, P. A., Van Thienen-Visser, K., Wassing, B. B., Osinga, S., Orlic, B., Ghouri, S. A., Buijze, L., and Pluymaekers, M. (2017). Geomechanical models for induced seismicity in the Netherlands: Inferences from simplified analytical, finite element and rupture model approaches. *Netherlands Journal of Geosciences*, 96(5):s183–s202.

Vavryčuk, V. (2015). Moment tensor decompositions revisited. *Journal of Seismology*, 19:231–252.

Wentinck, H. (2018). Dynamic modelling of large tremors in the Groningen field using extended seismic sources (available at: https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/063425b3-f3e0-4c6f-9d85-c5fe6f0eca07). Technical report, Nederlandse Aardolie Maatschappij BV, Assen, The Netherlands.

4

HAMILTONIAN MONTE CARLO TO CHARACTERIZE INDUCED EARTHQUAKES: APPLICATION TO A M_L 3.4 EVENT IN THE GRONINGEN GAS FIELD AND THE ROLE OF PRIOR

The Hamiltonian Monte Carlo (HMC) algorithm is known to be highly efficient when sampling high-dimensional model spaces due to Hamilton's equations guiding the sampling process. For weakly non-linear problems, linearizing the forward problem enhances this efficiency. This study integrates this linearization with geological prior knowledge for optimal results. We test this approach to estimate the source parameters of a 3.4 magnitude induced event that originated in the Groningen gas field in 2019. The source parameters are the event's centroid (three components), its moment tensor (six components), and its origin time. In terms of prior knowledge, we tested two sets of centroid priors. The first set exploits the known fault geometry of the Groningen gas field, whereas the second set is generated by placing initial centroid priors on a uniform horizontal grid at a depth of 3 km (the approximate depth of the gas reservoir). As for the forward problem linearization, we use an approach in which the linearization is run iteratively in tandem with updates of the centroid prior. We demonstrate that, in the absence of a sufficiently accurate initial centroid prior, the linearization of the forward model necessitates multiple initial centroid

Parts of this chapter have been previously published as :

***Masfara, L. O. M. and Weemstra, C. (2024)**. Hamiltonian Monte Carlo to characterize induced earthquakes: Application to a M_L 3.4 event in the Groningen gas field and the role of prior. Earth and Space Science, 11(1):e2023EA00318.*

The text has been extended and modified in several ways to ensure consistency within this thesis.

priors. Eventually, both prior sets yield similar posteriors. Most importantly, however, they agree with the geological knowledge of the area: the posterior peaks for model vectors containing a centroid near a major fault and a moment tensor that corresponds to normal faulting along a plane with a strike almost aligning with that of the major fault.

4.1. INTRODUCTION

Characterizing an earthquake is essential for a number of reasons. First, its source parameters (centroid, magnitude, slip direction, etc.) determine, to a large extent, the damage it may cause (Lui et al., 2016). This is because the depth, size, and type of rupture all affect the amount of shaking produced (Trippetta et al., 2019). Secondly, source characterization may help to improve our understanding of an event's nucleation, which is essential for developing reliable earthquake hazard models (Ellsworth et al., 2015). In addition, an increased understanding of source characteristics can potentially be used to improve earthquake early warning systems by providing (additional) information that can be used to generate alerts before strong shaking takes place (Peng et al., 2021).

Seismologists distinguish between 'natural' and 'induced' earthquakes. Induced earthquakes usually emit shorter period signals compared to tectonic earthquakes (Li et al., 2020). This is because, on average, induced events have relatively low magnitudes compared to (stronger) tectonic earthquakes, although some induced events are reported to be as high as 5.8 (Foulger et al., 2018). In addition, induced events usually occur at relatively shallow depths. Combined, shallower depths and higher frequencies imply that induced events may still cause significant damage to buildings and infrastructure (Vlek, 2018). In addition, ground motions are exacerbated by high amplification factors in some areas (Bommer et al., 2017).

A notable example of induced seismicity is the events occurring in the Groningen gas field, the Netherlands (Sarhosis et al., 2019). The Groningen gas field, located in the northern part of the Netherlands, is the largest gas field in Europe. Since the first reported induced earthquake in 1986, there has been a gradual increase in seismic activity in the field (van Thienen-Visser and Breunese, 2015). Because of the societal unrest associated with the earthquakes (Nepveu et al., 2016), the Dutch government has recently taken steps to reduce the extraction of natural gas from the Groningen gas field. The field closed down permanently on October 1, 2024, with production halted on October 1, 2023. Concurrent with the production reduction, an extensive array of seismometers was installed by the Dutch Meteorological Institute (KNMI, which stands for Koninklijk Nederlands Meteorologisch Instituut), funded by NAM (Nederlandse Aardolie Maatschappij), the major operator in the Groningen gas field (Ntinalexis et al., 2019). The array also includes borehole seismometers, enabling improved source characterization in the area (Smith et al., 2020), i.e., due to a significant increase of the signal-to-noise ratio (SNR) at depth (Ruigrok and Dost, 2019).

An earthquake source can be parameterized in several ways (Aki and Richards, 2002). In this study, we consider a moment tensor (MT) representation (Jost and Herrmann, 1989). This implies that the seismic event is collapsed to a single position (point-source representation), which is usually referred to as 'the centroid'. Such a representation is justified in case the waveform data is analyzed at periods for which the seismic source is effectively a point source (Aki and Richards, 2002). Additionally, assuming instanta-

neous rupturing, we end up with ten source parameters. The first six are the moment tensor components, where the MT's magnitude is a measure of the amount of energy released. This MT can be decomposed into isotropic (ISO), double-couple (DC), and compensated linear vector dipole (CLVD) components (Jost and Herrmann, 1989). The other four parameters are the event's east, north, and depth coordinates and the origin time.

Various datasets and techniques have been utilized to estimate the source characteristics of Groningen earthquakes. Willacy et al. (2018) adopt a deterministic approach to estimate moment tensors and centroids. These authors employ a detailed 3D subsurface model of Groningen but restrict the search space to DC sources. In contrast, Dost et al. (2020) use a probabilistic approach to estimate the centroid and full moment tensor (implying that they allowed for the ISO and CLVD components as well) but employed (locally) 1D models. Deterministic approaches often provide faster computations compared to probabilistic approaches. However, probabilistic approaches quantify the uncertainty of the different parameters; in this case, these are the uncertainties of the ten earthquake source parameters. Also, the use of 3D subsurface models has a clear advantage over 1D subsurface models. This is because 3D models take into account the subsurface lateral heterogeneity that will affect the shape (amplitude and phase) of the seismogram generated from simulating an earthquake event using those 3D models.

In this study, we investigate the combination of a probabilistic approach with 3D subsurface models to estimate the source parameters of a real event in Groningen. To mitigate the aforementioned "inefficiency" of probabilistic approaches, we modify the workflow described in Chapter 2. This workflow relies on a variant of the Hamiltonian Monte Carlo (HMC) algorithm and has previously been tested using synthetic recordings generated using the 3D Groningen subsurface velocity model. For this study, we consider the 2019 3.4 local magnitude earthquake below the village of Westerwijtwerd (Figure 4.1). Since we estimate the full moment tensor, our estimation does not limit the search space to just DC components but includes the ISO and CLVD components. Also, the inclusion of origin time in the estimation quantifies the trade-off between origin time and estimated depth. In what follows, we first describe the theory underlying the workflow. We subsequently introduce and discuss the (retrieval of the) recordings used to estimate the parameters, including the prior information that is used to increase the computational efficiency of the workflow. Finally, we compare our results to results obtained in other studies and draw conclusions, including the outlook of applying the same approach to a larger set of events in the Groningen area.

4.2. METHODOLOGY

To enable source characterization, the formal relationship between the observed data and the source (model) parameters is introduced and detailed in the first subchapter. Subsequently, we introduce Bayes' theorem and, assuming Gaussianity, cast it in a form allowing us to utilize it. In Subchapter 4.2.3, we then introduce the HMC algorithm. Finally, in Subchapters 4.2.4 and 4.2.5, we describe how the algorithm's efficiency can be enhanced via linearization of the forward problem and by choosing meaningful prior information, respectively.

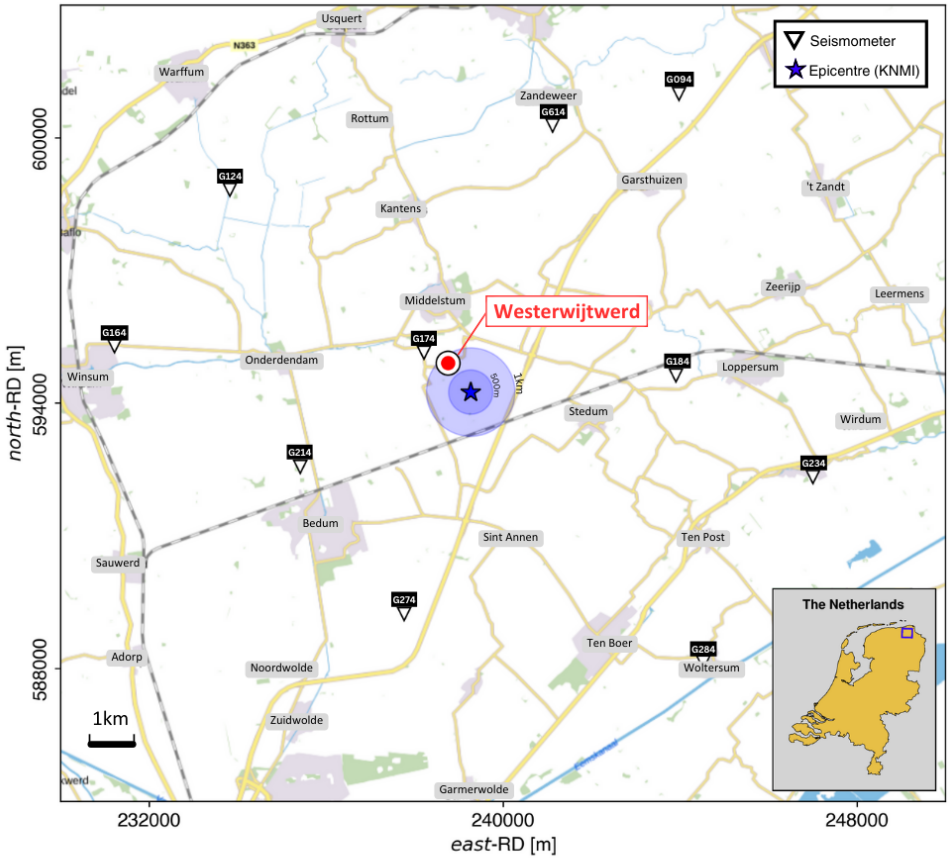

Figure 4.1: Map of the research area. The inverted triangles indicate the location of the KNMI seismometers, and the blue star is the epicenter of the 2019 Westerwijtwerd earthquake, as estimated by the KNMI. Axes indicate location using the Dutch RD coordinate system. This specific coordinate system gives the geodetic coordinates for European Netherlands and is used in official national maps. The inset at the bottom right shows the location of the study area.

4.2.1. THE FORWARD PROBLEM

In this study, the posterior probability of the model parameters is estimated by means of a Markov process. The generation of such a Markov chain is detailed further below (Subchapter 4.2.3), but, at this point, it should be understood that for each sample in the chain, forward-modeled data is compared against measured data. In the context of our problem, a specific model $\mathbf{m}$ (or sample) implies assigning a specific value to each of the ten aforementioned source parameters (MT, centroid, and origin time). The measured data $\mathbf{d}^{\text{obs}}$ consists of the induced event's waveform data, which, in our case, are recordings of particle displacement recorded by KNMI instruments. Computation of the likeli-

hood $\rho(\mathbf{d}^{\text{obs}}|\mathbf{m})$ yields the probability of these recordings given a model $\mathbf{m}$ and involves quantification of the misfit between the recorded particle displacements and numerically modeled particle displacements. The latter is computed by numerically solving the wave equation, i.e., they are the result of solving (what is usually referred to as) 'the forward problem'. Mathematically, the forward problem can be written as

$$u_i\left(\mathbf{x}^{(\text{r})}, t\right) = \sum_{j=1}^{3}\sum_{k=1}^{3} M_{jk}(t, T_0) * G_{ij,k}\left(\mathbf{x}^{(\text{r})}, t; \mathbf{x}^{(\text{a})}\right), \tag{4.1}$$

where u_i is the i^{th} component of the particle displacement vector ($\mathbf{u} = (u_1, u_2, u_3)$ where 1, 2, 3, correspond to the east, north, and down direction, respectively), M_{jk} representing an element of the 3×3 moment tensor $\mathbf{M}$ at position $\mathbf{x}^{(\text{a})}$, i.e., the centroid. Note that j and k indicate the axis along which the force is acting and the direction in which the arm is pointing, respectively (Aki and Richards, 2002). Furthermore, $\mathbf{x}^{(\text{r})}$ denotes the position where the displacement is recorded, $G_{ij,k}$ represents the elements of $\mathbf{G}$, the Green's tensor with the first two subscripts representing its recorded component and the direction in which the source force is pointing. Furthermore, the comma after the second subscript implies a spatial derivative in the k direction with respect to $\mathbf{x}^{(\text{a})}$ and $*$ represents a temporal convolution, and T_0 denotes the origin time. To make the computation of $\mathbf{u}\left(\mathbf{x}^{(\text{r})}, t\right)$ for a large number of potential centroids (i.e., a large number of $\mathbf{x}^{(\text{a})}$) more efficient, we invoke reciprocity (Aki and Richards, 2002). In this study, the numerically modeled particle displacements are generated using SPECFEM3D (Komatitsch and Tromp, 2002). For this purpose, we use the 3D subsurface models of the Groningen gas field by Romijn (2017).

4.2.2. BAYES' THEOREM

The probabilistic workflow used in this study relies on Bayes' theorem (or rule). In general, Bayes' theorem describes how, in the presence of prior knowledge, the probability of a hypothesis (or model) $\mathbf{m}$ depends on the available data $\mathbf{d}^{\text{obs}}$. The prior knowledge is accounted for by the prior probability distribution (often simply referred to as 'the prior'). Ignoring the marginal probability (or 'evidence'), Bayes' theorem can be written as

$$\rho(\mathbf{m}|\mathbf{d}^{\text{obs}}) \propto \rho(\mathbf{d}^{\text{obs}}|\mathbf{m})\rho(\mathbf{m}), \tag{4.2}$$

where $\rho(\mathbf{m}|\mathbf{d}^{\text{obs}})$ is the posterior probability distribution (or simply 'the posterior'), $\rho(\mathbf{d}^{\text{obs}}|\mathbf{m})$ the likelihood, and $\rho(\mathbf{m})$ the prior probability distribution. The model vector $\mathbf{m}$ is a ten-component vector containing the centroid $\mathbf{x}^{(\text{a})}$ (where a Cartesian east-north-down coordinate system implies that $\mathbf{x}^{(\text{a})} = (x_1^{(\text{a})}, x_2^{(\text{a})}, x_3^{(\text{a})})$; hence three model parameters), the moment tensor $\mathbf{M}$ (six independent elements and hence six model parameters), and the origin time T_0 (one model parameter). This implies that $\rho(\mathbf{m})$ represents the prior probability of these ten parameters.

Assuming Gaussian observational errors and a Gaussian distributed prior probability, the posterior in Equation 4.2 can be written as (Fichtner and Simutė, 2018; Masfara

et al., 2022):

$$\rho\left(\mathbf{m} \mid \mathbf{d}^{\text{obs}}\right) \propto \exp\left(-\frac{1}{2}\left(\mathbf{d}(\mathbf{m})-\mathbf{d}^{\text{obs}}\right)^{\top}\mathbf{C}_d^{-1}\left(\mathbf{d}(\mathbf{m})-\mathbf{d}^{\text{obs}}\right)-\frac{1}{2}\left(\mathbf{m}-\mathbf{m}^{(0)}\right)^{\top}\mathbf{C}_m^{-1}\left(\mathbf{m}-\mathbf{m}^{(0)}\right)\right). \tag{4.3}$$

Here, $\mathbf{d}(\mathbf{m})$ contains the numerically modeled displacement recordings (solution of Equation 4.1) and $\mathbf{d}^{\text{obs}}$ the observed ones. Explicitly, for a total of N_r three-component instruments, $\mathbf{d}(\mathbf{m})$ is a concatenation of all $3 \times N_r$ modeled seismograms and $\mathbf{d}^{\text{obs}}$ a concatenation of all $3 \times N_r$ recorded seismograms. $\mathbf{C}_d$, $\mathbf{C}_m$, and $\mathbf{m}^{(0)}$ are the data covariance matrix, prior covariance matrix, and prior mean, respectively. Evaluating Equation 4.3 results in the (a posteriori) probability of the model parameters, i.e., their probability given observations and prior knowledge of the system (Tarantola, 2006).

4

4.2.3. HAMILTONIAN MONTE CARLO

Although Bayes' theorem describes how the posterior probability distribution depends on the available data $\mathbf{d}^{\text{obs}}$ (through the likelihood) and prior knowledge $\rho(\mathbf{m})$, that posterior can usually not be estimated directly (Tarantola and Valette, 1981). In particular, a large number of model parameters and non-linearity prohibit this. To overcome this, we generate a sequence of specific models (often called 'samples') in what is referred to as a 'Markov chain.' The density of these samples reflects the density of the posterior distribution we seek to find.

Numerous sampling algorithms are available to estimate $\rho(\mathbf{m}|\mathbf{d}^{\text{obs}})$, all with their own advantages and disadvantages. In this study, we implement a workflow that relies on the Hamiltonian Monte Carlo (HMC) algorithm. HMC was derived from classical mechanics, applied to statistical mechanics (Betancourt, 2017), and considered one of the most efficient probabilistic algorithms for exploring high-dimensional model spaces. HMC relies on the sequential calculation of two quantities. These are the 'potential energy' U, which is a function of the model vector $\mathbf{m}$, and the 'kinetic energy' K, which, in our framework, is solely a function of the momentum vector $\mathbf{p}$. This momentum vector is an auxiliary vector that has the same dimension as $\mathbf{m}$ (ten in our case). Together, $\mathbf{m}$ and $\mathbf{p}$ make up what is often referred to as the 'phase space,' and their joint probability is described by the 'canonical distribution' $\rho\left(\mathbf{p},\mathbf{m}\right)$.

The canonical distribution can be written in terms of an invariant function $H\left(\mathbf{p},\mathbf{m}\right)$, i.e.,

$$\rho\left(\mathbf{p},\mathbf{m}\right) = e^{-H(\mathbf{p},\mathbf{m})}. \tag{4.4}$$

Here, $H\left(\mathbf{p},\mathbf{m}\right)$ is referred to as 'the Hamiltonian,' and its value in phase space is usually called 'the energy' at that point (Neal, 2011). As such, a model $\mathbf{m}$ can be looked upon as the position of a "particle" (Betancourt, 2017).

Rewriting Equation 4.4, and substituting the posterior probability (i.e., $\rho\left(\mathbf{p},\mathbf{m}\right) \rightarrow \rho(\mathbf{p},\mathbf{m} \mid \mathbf{d}^{\text{obs}})$), we have

$$\begin{aligned} H\left(\mathbf{p},\mathbf{m} \mid \mathbf{d}^{\text{obs}}\right) &\equiv -\ln\left(\rho\left(\mathbf{p},\mathbf{m} \mid \mathbf{d}^{\text{obs}}\right)\right) \\ &= -\ln[\rho(\mathbf{p} \mid \mathbf{m})] - \ln\left[\rho\left(\mathbf{m} \mid \mathbf{d}^{\text{obs}}\right)\right] \\ &= K(\mathbf{p},\mathbf{m}) + U(\mathbf{m}). \end{aligned} \tag{4.5}$$

Here, $U(\mathbf{m}) \equiv -\ln \rho(\mathbf{m} \mid \mathbf{d}^{\text{obs}})$.

Equation 4.5 describes the more general case; in our implementation, $K(\mathbf{p},\mathbf{m})$ is merely a function of the momentum vector and hence $K(\mathbf{p},\mathbf{m}) \rightarrow K(\mathbf{p})$. Specifically, it is given by (Fichtner and Simutė, 2018; Masfara et al., 2022)

$$K(\mathbf{p}) = \mathbf{p}^T \mathscr{M}^{-1} \mathbf{p}/2, \tag{4.6}$$

where the mass matrix $\mathscr{M}$ acts as a tuning parameter (Fichtner et al., 2019, 2021), allowing the particle to move through the desired areas of phase space with corresponding potential and kinetic energy (Betancourt, 2017).

Starting from an initial estimate of $\mathbf{m}$ with some prescribed initial momentum, Hamilton's equations, which read

$$\frac{d\mathbf{m}}{d\tau} = \frac{\partial K}{\partial \mathbf{p}}, \quad \frac{d\mathbf{p}}{d\tau} = -\frac{\partial U}{\partial \mathbf{m}}, \tag{4.7}$$

will efficiently explore areas with relatively low potential energies (corresponding to the a posteriori more probable areas of the model space; see Equation 4.5). Here, the quantity τ is the 'artificial time' that is used to propagate (the particle) from the initial model along trajectories of constant H. This propagation occurs for some (to-be-determined) time τ_{lp}, where the subscript 'lp' stems from 'leap' as we use the leapfrog algorithm to evaluate 4.7. The model reached at τ_{lp}, i.e., $\mathbf{m}(\tau_{\text{lp}})$, is subsequently accepted with probability

$$\theta = \min\left[1, \frac{\rho\left(\mathbf{p}(\tau_{\text{lp}}), \mathbf{m}(\tau_{\text{lp}})\right)}{\rho\left(\mathbf{p},\mathbf{m}\right)}\right], \tag{4.8}$$

which is usually referred to as the 'metropolis rule' (Tarantola, 2005). If the model $\mathbf{m}(\tau_{\text{lp}})$ is not accepted, the process will be repeated by introducing a new (different) momentum vector to the initial model. If accepted, the model $\mathbf{m}(\tau_{\text{lp}})$ will serve as the starting point for a new deterministic trajectory after being endowed with momentum.

One of the main advantages of using HMC over generic probabilistic sampling algorithms such as Metropolis-Hasting (MH) algorithms is its ability to sample the posterior distribution more efficiently, which is illustrated in Figure 4.2. In MH (left figure), new samples are randomly drawn based on specified proposal distributions. This behaviour is represented by the red balls 'around' the current sample, which for the first iteration are close to the starting model. In HMC (right figure), the iterative endowment with the momentum of the current sample results in a combination of short and long trajectories, reducing the correlation between samples. Furthermore, the inclusion of the gradient of the target distribution via the computation of Hamilton's equations enables HMC to stay in the 'typical set', constituting areas where the posterior probability is elevated. Therefore, with a combination of short and long trajectories, and the ability to slide along the typical set, HMC can explore the posterior distribution more efficiently with fewer iterations/samples. This type of exploration is particularly useful when sampling complex posterior distributions in high-dimensional model spaces. Computationally, the main burden of HMC is the computation of the derivatives of the potential energy with respect to the model parameters. This computational burden increases as the number of model parameters increases. To mitigate this, we adopt an approach in which we linearize the forward problem, which we will now briefly describe.

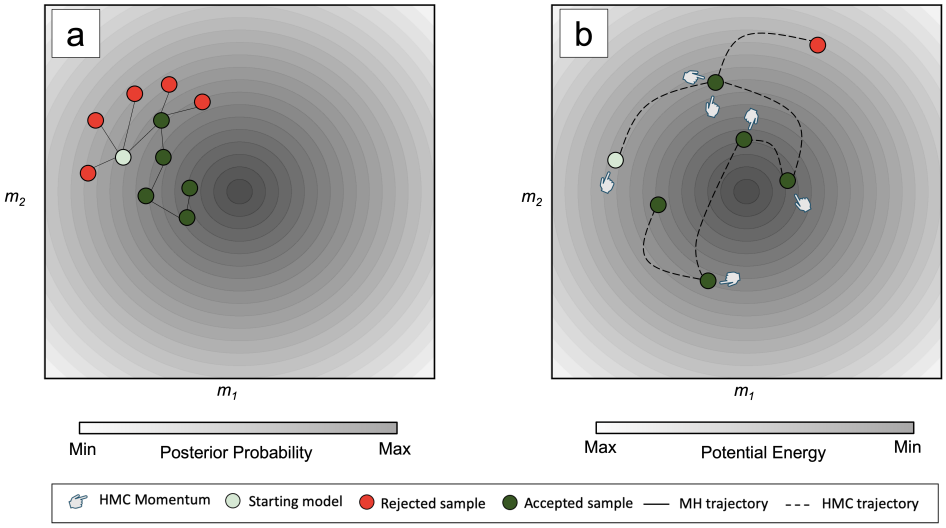

Figure 4.2: Illustration of model space exploration using Metropolis-Hastings (a) and Hamiltonian Monte Carlo (b) algorithms. Note that with a similar number of accepted samples, HMC explores the distribution more efficiently via a combination of iterative short and long trajectories. This is achieved by prescribing a different momentum for each trajectory and iterative computation of Hamilton's equations. Note that we only show the rejected samples of the first two moves/accepted samples for both algorithms.

4.2.4. LINEARIZATION OF THE FORWARD PROBLEM

To ease the computation of the gradient of the potential energy in the model space, Fichtner and Simutė (2018) linearize Equation 4.1 by means of a Taylor expansion around the prior mean $\mathbf{m}^{(0)}$ (see Appendix A). Simutė et al. (2022) use this same modification and 3D Earth models to characterize tectonic earthquakes below the Japanese peninsula. In these studies, $\mathbf{m}^{(0)}$ is obtained from an earthquake catalog, which is not always directly available for induced earthquakes. Replacing, in $\mathbf{d}(\mathbf{m})$, the numerically modeled displacements $\mathbf{u}(\mathbf{x}^{(r)}, t)$ by numerically modeled displacements resulting from a linear approximation of Equation 4.1 implies that we assume $\mathbf{m}^{(0)}$ to be "sufficiently close" to the true model parameters. This merely applies for the centroid $\mathbf{x}^{(a)}$ and origin time T_0. That is, since the particle displacement depends linearly on the moment tensor components, the linearization does not impose an approximation when it comes to the moment tensor components. Importantly, "sufficiently close" means that the centroid $\mathbf{x}^{(a)}$ and origin time T_0 should be at sub-wavelength and sub-period distance from the true centroid and origin time, respectively.

In our case, the assumption that $\mathbf{m}^{(0)}$ is sufficiently close to the true model parameters is usually not met. This will render the application of HMC ineffective (to state the least). In order to apply HMC (including a linear approximation of Equation 4.1) to induced earthquakes, two main challenges, therefore, need to be addressed. First, the recorded seismograms are often dominated by high-frequency signals (>1 Hz), increasing the non-linearity of the forward problem. Second, as mentioned earlier, the prior information is often unavailable or rather inaccurate. To address these challenges, in

this study, we use the multi-stage workflow introduced in Chapter 2. This means that we iteratively update $\mathbf{m}^{(0)}$, which is detailed in the remainder of this section. In addition, we run this workflow multiple times (in parallel), each starting from a different $\mathbf{m}^{(0)}$. This is explained in Section 4.2.5. In the remainder of this chapter, we will refer to the HMC variant that involves a linearization of the forward problems as 'linearized HMC'. It should be understood, however, that this does not involve a linearization of Hamilton's equations itself.

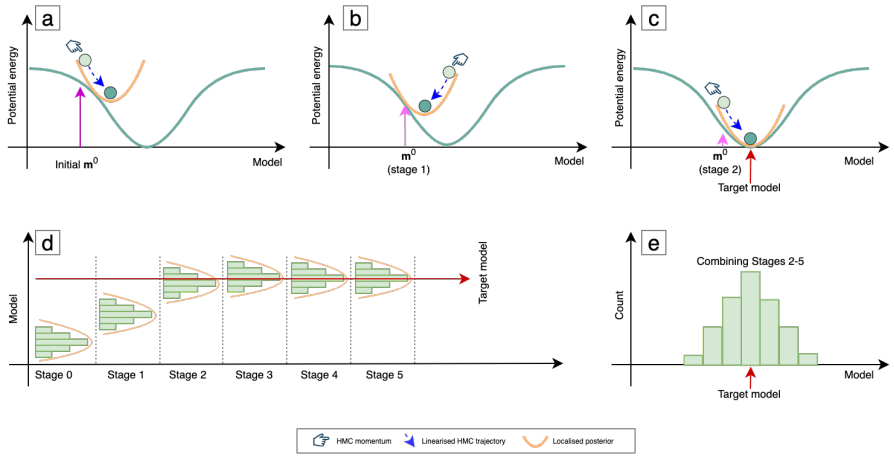

Figure 4.3: Illustration of linearized HMC embedded in the multi-stage workflow detailed in Chapter 2. Panels (a) to (c) depict the sampling of a local posterior associated with different $\mathbf{m}^{(0)}$. In (a), $\mathbf{m}^{(0)}$ is the initial prior mean. In the next stage (b), $\mathbf{m}^{(0)}$ is updated using the results of the exploration of the local posterior associated with this initial prior mean until $\mathbf{m}^{(0)}$ (almost) coincides with the most likely model (c). The workflow's progression up to five stages is shown in (d). (e) is the final posterior composed using variance reduction criterion, which discriminates the first two stages from stages 2 to 5.

Figure 4.3 illustrates the embedding of linearized HMC in the proposed multi-stage workflow. Iteratively updating $\mathbf{m}^{(0)}$ partly overcomes deviations of the estimated posterior from the true posterior, thus addressing the first challenge. Given a first $\mathbf{m}^{(0)}$, the three quantities in Equations 4.12-4.14 need to be computed only once in order to sample a "local posterior" around that $\mathbf{m}^{(0)}$. These quantities are used to compute the gradient of the potential energy and hence evaluate Hamilton's equations and the Hamiltonian itself (Equations 4.7 and 4.5, respectively). Importantly, in the absence of a linearization of the forward problem, the computation of Equations 4.7 and 4.5 requires the forward problem to be evaluated during each deterministic trajectory. Linearization of Equation 4.1, resulting in the three aforementioned quantities, renders this unnecessary for each individual stage (Chapter 2).

When $\mathbf{m}^{(0)}$ does not coincide with the true model parameters, the linearized HMC algorithm will explore a "local posterior" that deviates from the true posterior distribution despite being computationally efficient. This is illustrated in Figure 4.3(a), where the linearized HMC can only explore the area above the orange curve. To obtain a better approximation of the posterior, the workflow uses the result of exploring the local posterior in Figure 4.3(a) to obtain a new $\mathbf{m}^{(0)}$ (essentially taking the mean of the local pos-

terior and using that as $\mathbf{m}^{(0)}$). Linearization of the forward problem about the updated $\mathbf{m}^{(0)}$ and re-computation of the aforementioned quantities allows for a new exploration of the model space in Figure 4.3(b) and (c). After five Taylor expansions about the new $\mathbf{m}^{(0)}$, six local posteriors are estimated. The associated distributions are, for each stage, depicted in Figure 4.3(d). Having the results from all stages in (d), the workflow then uses variance reduction (e.g., Mustać and Tkalčić, 2016; Masfara et al., 2022) as a criterion to select stages that should be included in the estimate of the final posterior. This is depicted in Figure 4.3(e).

4.2.5. THE IMPORTANCE OF THE PRIOR

Having an inaccurate $\mathbf{m}^{(0)}$ can only partly be overcome by updating $\mathbf{m}^{(0)}$ in progressive stages. That is, the multi-stage workflow will still be ineffective when the initial $\mathbf{m}^{(0)}$ is located in a "local mode" of the posterior distribution (i.e., associated with a local minimum of the potential energy). The chance of this happening increases with an increase in the non-linearity between the model parameters and the observed displacement recordings (i.e., higher frequencies). In practice, this happens when the centroid $\mathbf{x}^{(a)}$ and origin time T_0 in $\mathbf{m}^{(0)}$ are separated from the true centroid and true origin time by more than (approximately) half a wavelength or half a period, respectively. To address this, we additionally use multiple initial vectors $\mathbf{m}^{(0)}$ concatenated in a list which we denote by $\mathbf{m}^{(0)}_{\text{list}}$ (the list consists of $\mathbf{m}^{(0)}_1$, $\mathbf{m}^{(0)}_2$, ... $\mathbf{m}^{(0)}_N$ with N being the total number of vectors $\mathbf{m}^{(0)}$). These initial $\mathbf{m}^{(0)}_i$ differ to the extent that the centroid position is different for each of them. The use of $\mathbf{m}^{(0)}_{\text{list}}$ is to ensure some of the individual $\mathbf{m}^{(0)}_i$ are contained in the global minimum. The same criterion is used to select which (local posterior) distributions can be included in the final posterior (i.e., which stages). That is, the variance reduction is now computed for all stages associated with the individual (initial) $\mathbf{m}^{(0)}_i$ in $\mathbf{m}^{(0)}_{\text{list}}$. We illustrate the process of using multiple $\mathbf{m}^{(0)}$ in Figure 4.4. We depict three initial $\mathbf{m}^{(0)}$, with one located in the "correct" lobe, that is, $\mathbf{m}^{(0)}_2$. Each of the $\mathbf{m}^{(0)}$ will then be updated in a similar fashion as shown in Figure 4.3. While $\mathbf{m}^{(0)}_1$ and $\mathbf{m}^{(0)}_3$ end up sampling the wrong lobe, the updated $\mathbf{m}^{(0)}_2$ enables the linearized HMC algorithm to sample the correct lobe. In Figure 4.4(b), we detail the last stage of the multi-stage workflow that started with $\mathbf{m}^{(0)}_2$ in the red circle.

We end this section by emphasizing that although being very efficient in sampling the posterior distribution (through the potential energy), the proposed multi-stage workflow (including the use of multiple initial priors $\mathbf{m}^{(0)}_i$) ultimately only results in an approximate posterior distribution. This is because the true observational errors are not necessarily Gaussian and uncorrelated (which we assume in this study) and because we linearized the relation between observed particle displacement and model parameters. In addition, the 3D velocity model used to model (numerically) displacement recordings (according to Equation 4.1) is assumed to coincide with the true velocity model. Since this will not be the case, another "source of error" is introduced, which in practice will result in a deviation of the estimated posterior from the true posterior. Moreover, since a Markov process only approaches the true posterior asymptotically, a Markov-chain-based estimate of the posterior is, by definition, an approximation. Whereas the latter two cannot be circumvented (we don't have the exact subsurface model and also cannot

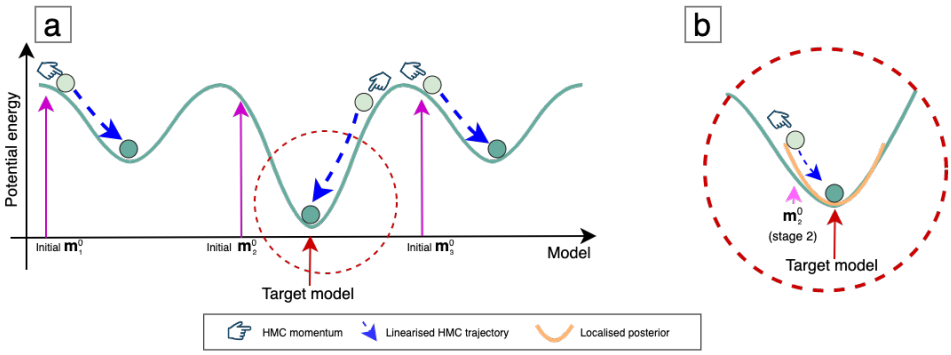

Figure 4.4: (a) Illustration of using multiple initial $\mathbf{m}^{(0)}$ while sampling a complex/multimodal posterior distribution using linearized HMC. (b) zoom of the last stage of the multiple stages associated with the initial model prior $\mathbf{m}_2^{(0)}$.

run a Markov chain for an infinite amount of time), the linearization is, in principle, not necessary, and also Gaussian observational errors do not need to be assumed. Not doing so, however, would make the computational demands prohibitively large.

4.3. DATA

In this study, $\mathbf{d}^{\text{obs}}$ contains the $3 \times N_r$ recordings of displacements ($\mathbf{u}^{\text{obs}}$) due to an induced event that occurred close to the village of Westerwijtwerd in 2019, the province of Groningen (see Figure 4.1). The KNMI estimates the magnitude of the earthquake to be 3.4 local magnitude. We collected $\mathbf{u}^{\text{obs}}$ from ten G-network seismometers. These seismometers are selected based on their distance and azimuthal coverage with respect to the estimated epicenter. In Figure 4.5(a), we depict the ten seismometers as white inverted triangles and the location of the KNMI-estimated epicenter by a blue star. The seismometers are part of the KNMI borehole network: each borehole contains four vertically-separated seismometers. The number at the end of their ID indicates their depth, i.e., their IDs run from ..1 to ..4, with the instruments numbered ..1 being at 50 m depth and the instruments numbered ..4 being at 200 m depth. We illustrate the configuration of a string of borehole seismometers in Figure 4.5(b).

From the four seismometers in each borehole, we solely use the seismograms recorded by the deepest seismometers: they have a higher signal-to-noise ratio than the shallower seismometers (Dost et al., 2012). Furthermore, all seismometers experience a horizontal rotation while lowering them in the borehole. Consequently, a rotation needs to be carried out for projecting the horizontal recordings to specific preferred orientations, which in our case are to the east-west(x_1-axis) and north-south(x_2-axis) orientations, respectively. In Figure 4.5(b), we illustrate the orientation of the deepest borehole seismometer. The axes H1 and H2 are proxies of east and north. We then rotate the data to the true east and north using the angles given in Ruigrok et al. (2019). We depict the original seismograms (obtained from the KNMI) and the rotated seismograms of the selected seismometers in Figure 4.6.

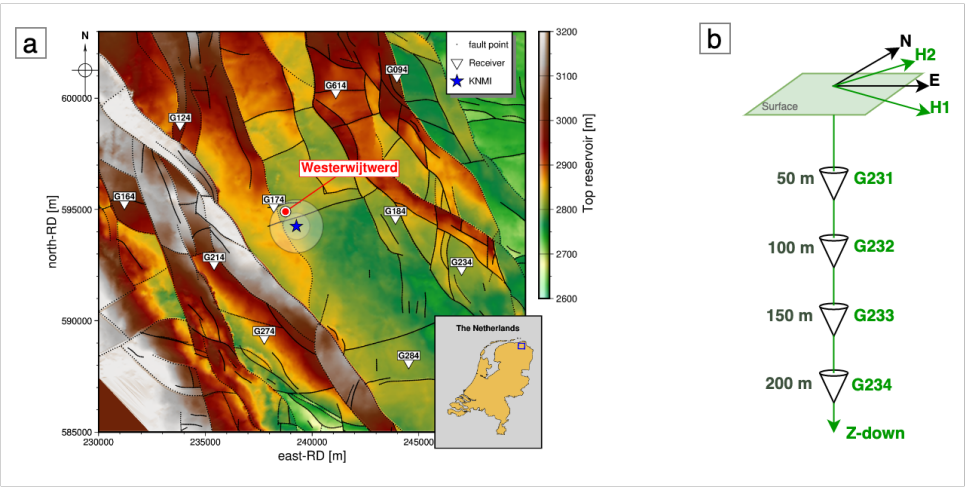

Figure 4.5: (a) Depth of top Rotliegend (reservoir) in area of interest. Solid black dots delineate mapped faults (Bourne and Oates, 2017). The inset at the bottom right shows the location of the study area. (b) Illustration of borehole seismometers in the G-network.

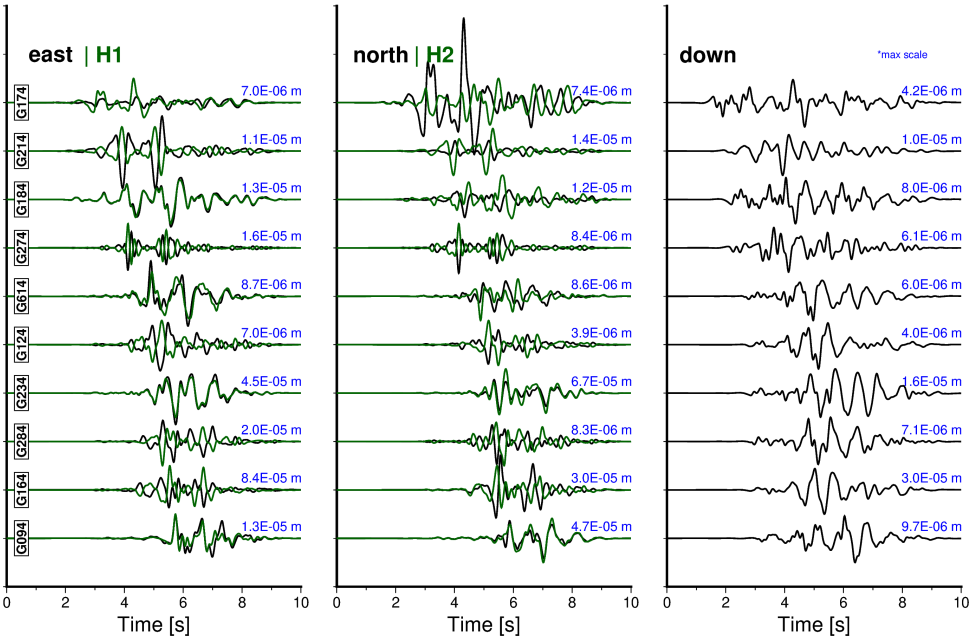

Figure 4.6: Observed seismograms before (green) and after rotation/polarity switch (black). Recordings are normalized (individually) with respect to maximum particle displacement (written in blue).

Dost et al. (2020) have used the same recordings to characterize the Westerwijtwerd event probabilistically. These authors, however, use local 1D velocity models to solve the forward problem. Furthermore, they separately use 0.5 and 1 s windows of P and S waves, respectively, where the P-wave is given more weight and evaluated at higher frequencies (i.e., 2-4 Hz for P and 1-3 Hz for S-wave). The P-wave waveform is given a higher weight because of the higher accuracy of the employed P-wave velocity models (compared to the S-wave velocity models). Also, these authors only use the vertical components of the recorded P-wave and the transverse component of the recorded S-waves. To account for inaccuracy in the velocity models, they allow individual, station-specific shifts of 0.1 s for both wave types. Another study in the area is by Smith et al. (2020), which uses a coherence method. This study focuses on determining the hypocenter. They find most Groningen earthquakes to systematically originate approximately 200 m above the reservoir layer. In this study, we exclusively use P-wave seismograms due to the significantly higher accuracy of the P-wave model. Furthermore, we use both the vertical and horizontal components and filter the recordings using a passband of 1-4 Hz, similar to the frequency range used by Dost et al. (2020). As for the length of the measurement window, we use 2.5 s for all components and taper both ends with a 0.5 s cosine taper. For the data covariance, we use a diagonal matrix representing uncorrelated noise and estimate this to be 5% of each component's maximum amplitude. By taking a certain fraction of the maximum amplitude, we overestimate the 'true noise'. The reason for this is that we want to account for (part) of the waveform misfit arising from the deviation of the employed velocity model (Romijn, 2017) from the true (unknown) velocity model. Before applying it to the field data, we perform a synthetic experiment, detailed in the next section.

4.4. SYNTHETIC EXPERIMENT

In this section, we test the validity of the proposed workflow and data processing parameters (i.e., frequency band, length of the measurement window, and noise criteria) on a synthetic event. For this, we first generate synthetic data using the KNMI-estimated hypocenter as the centroid of our synthetic earthquake. We then set T_0 to 3 s, and for the MT, we use the values of 0.2×10^{13} Nm, 2.86×10^{13} Nm, -3.07×10^{13} Nm, 0.76×10^{13} Nm, -0.45×10^{13} Nm, -1.71×10^{13} Nm for M_{nn}, M_{ee} M_{dd}, M_{ne}, M_{nd}, and M_{ed} respectively. These values represent pure shear normal faulting (rake of -90°) along a geological fault with a strike of 165°, a dip of 60°, and a moment magnitude of 3. We then corrupt the data in the frequency domain to simulate the presence of uncorrelated noise. This is implemented using the same approach as Mustać and Tkalčić (2016). In the time domain, the uncorrelated noise results in amplitude variations that affect the estimation of our centroid and MT, and shift the observed recordings in time (resulting in uncertainty in T_0). To effectively test the workflow, we first choose a (single) $\mathbf{m}^{(0)}$ that significantly deviates from the actual value (i.e., the synthetic earthquake parameters). For the centroid, we impose a shift of 200 m along each axis, i.e., the centroid in $\mathbf{m}^{(0)}$ deviates 200 m from $x^{(a)}_{east}$, $x^{(a)}_{north}$, and $x^{(a)}_{depth}$. For the MT, we simply assign a uniform value to each MT component, and for T_0, we impose a shift of 0.5 s. We then run our workflow for 20 stages (i.e., the prior mean $\mathbf{m}^{(0)}$ is updated twenty times). The results are presented in

Figure 4.7.

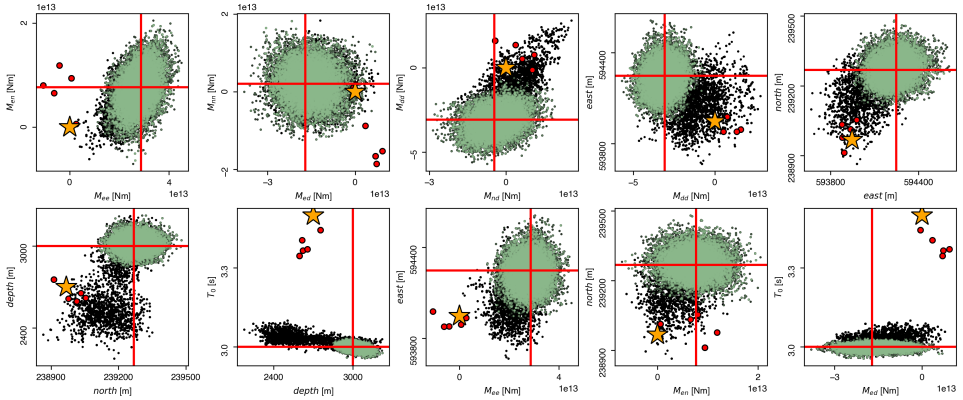

Figure 4.7: Marginal posterior probabilities obtained through applying the proposed linearized HMC workflow to synthetic recordings. The stars represent the initial $\mathbf{m}^{(0)}$. The black and green dots represent all accepted samples from all 20 stages and samples from selected stages (i.e., the VR-score exceeds 0.95) used to compose the final posterior, respectively. The red lines represent the true (synthetic) model parameters, and the red dots are the samples generated running the generic (non-linearized) HMC algorithm.

The yellow stars represent the initial $\mathbf{m}^{(0)}$, and the red lines represent the true synthetic earthquake parameters. The black dots are the samples generated from all 20 stages, which are equivalent to samples used to build all the histograms from exploring local posteriors in Figure 4.3(d). The green dots are the samples from selected stages based on a VR criterion, equivalent to the samples from the selected stages in Figure 4.3(e). The red dots represent samples resulting from a generic HMC run (i.e., HMC without linearizing the forward problem). This run was terminated as soon as the number of times for which the forward problem needed to be solved coincided with the number of times the forward problem was solved while running the multi-stage workflow in which the forward problem was linearized. Note that each solution of the forward problem involves the computation of 3×10 seismograms (recall from Section 4.3 that $N_r = 10$).

Let us demonstrate the computational benefit of the multi-stage workflow (in conjunction with a linearization of the forward problem) over generic HMC (which does not involve this linearization). The number of times the forward problem needs to be solved in order to generate four model samples using generic HMC (represented by the star and the red dots in Figure 4.7) is 404. Here, each 'solution of the forward problem' in practice involves a separate computation of the $u_i\left(\mathbf{x}^{(\mathrm{r})}, t\right)$ in Equation 4.1. We arrive at 404 as follows: it depends on the number of generated samples N_s (4 in this case), the number of leaps N_{lp} to arrive at $\mathbf{m}(\tau_{lp})$ (here we use 5), and the number of model parameters N_m (10 in our case). First, with the prescribed five leaps to arrive at a new model started from the current model, we evaluate Equation 4.7 five times. Second, the evaluation of Equation 4.7 requires the computation of $\frac{\partial U}{\partial \mathbf{m}}$. For that, we use a central difference approximation, which means that for each of the ten parameters in $\mathbf{m}$, we must evaluate U twice. Additionally, after the five leaps, we still have to compute $\rho\left(\mathbf{p}(\tau_{\mathrm{lp}}), \mathbf{m}(\tau_{\mathrm{lp}})\right)$ to eval-

uate Equation 4.8, which requires one additional solution of the forward problem per sample. Consequently, the total number of forward problem solutions coincides with $N_s \times N_{lp} \times 2N_m + N_s = 404$. Linearization of the forward problem reduces this number dramatically. In fact, for every stage of the multi-stage workflow, the number of samples that can be generated is unlimited in the sense that it does not require additional solutions to the forward problem. The forward problem just needs to be run $2 \times N_m = 20$ times per stage. This number stems from the (one-time) computation of the derivatives of U. These derivatives are included in the A_{pq}, b_p, and c (Equations 4.12, 4.13, and 4.14 in appendix 4.7, respectively). Therefore, to generate all samples for a total of 20 stages (i.e., 20 updates of $\mathbf{m}^{(0)}$), the number of times the solution to the forward problem needs to be computed is just 400. We use the mean of the approximate posterior resulting from

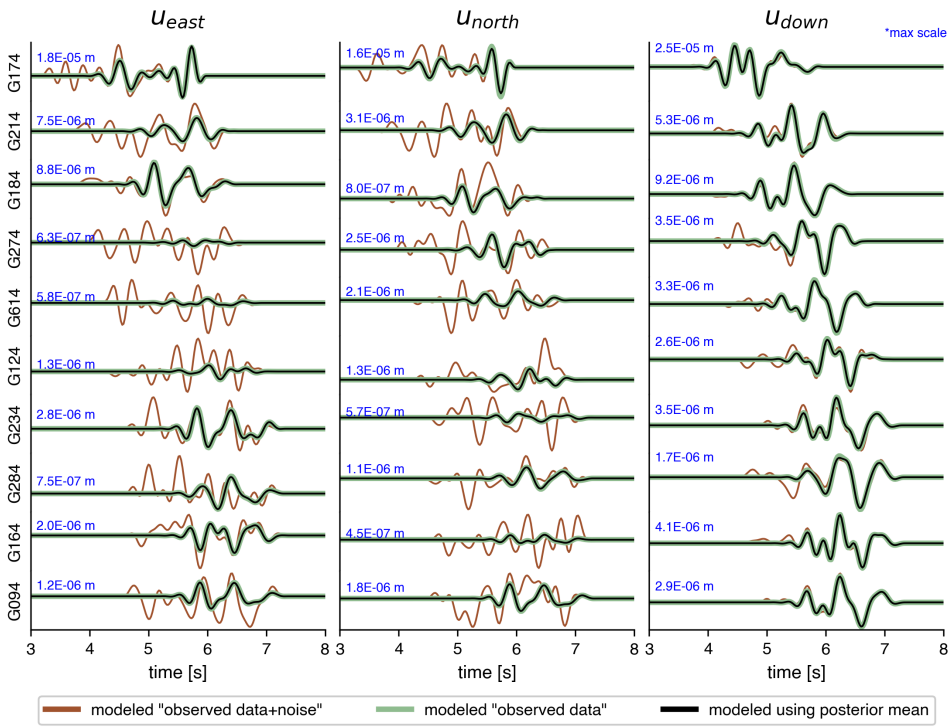

Figure 4.8: Seismograms modeled using the posterior mean (gray) compared to the modeled observed recordings with noise added (brown) and the modeled observed recordings without noise (green).

our multi-stage workflow to generate displacement recordings. In Figure 4.8, we compare these recordings with the observed (synthetic) recordings. The observed recordings are depicted in brown (recall that noise is added to these seismograms). The recordings associated with the mean values of our estimated posterior are depicted in grey and align well with the noise-free recordings associated with the true source parameters (depicted

in green).

4.5. PRIOR KNOWLEDGE

In Subchapter 4.2.5, we discussed the importance of using $\mathbf{m}_{\text{list}}^{(0)}$ to avoid getting trapped in a local mode. For the purpose of generating $\mathbf{m}_{\text{list}}^{(0)}$, we make use of the available fault map of Groningen's subsurface by Bourne and Oates (2017). This is inspired by research that reveals a strong correlation between hypocenters and major faults in Groningen's subsurface (Pickering, 2015; Spetzler and Dost, 2017; Willacy et al., 2018). In this context, we also evaluate the importance of the displacement along the horizontal components for the estimated posterior. The reason for this is potential errors arising from possible incorrect rotations of the horizontal displacements (see Section 4.3). Combined, we, therefore, investigate three different cases: two centroid prior configurations (i.e., with different $\mathbf{m}_{\text{list}}^{(0)}$) of which one is used in conjunction with both the vertical component recordings and the three-component recordings. The configuration that uses known faults in the reservoir as a basis to generate the $\mathbf{m}_{\text{list}}^{(0)}$, in conjunction with the vertical component recordings only, is referred to as '1C-fault'. The same configuration, but used to estimate the posterior based on the recordings by all three components, is referred to as '3C-fault'. The other centroid prior configuration we consider consists of a square grid that covers not just the fault but also the surrounding area. This configuration of $\mathbf{m}_{\text{list}}^{(0)}$ is only used in conjunction with the recordings by all three components and is referred to as the '3C-grid'. This centroid prior configuration is considered to evaluate whether the recovered posterior might peak at a centroid position that deviates from the known fault geometry. The two different centroid prior configurations are depicted in Figure 4.9.

To generate the entries (individual $\mathbf{m}^{(0)}$) in $\mathbf{m}_{\text{list}}^{(0)}$ of the two considered prior configurations, we first draw a circle with a 1 km radius around the epicenter estimated by the KNMI. The enclosed area is colored dark green in Figure 4.9. Next, we discretize the fault inside the circle using a spatial sampling criterion based on the approximate seismic P-wave velocity within the circle and the highest frequency we use while fitting the waveforms. This criterion provides a rough estimate of the minimum "wavelength" of the posterior distribution. By discretizing the fault such that the individual centroids (associated with individual $\mathbf{m}_i^{(0)}$) in $\mathbf{m}_{\text{list}}^{(0)}$ are separated by less than half this wavelength, we therefore, ensure that at least one of the initial priors is located in the "correct" lobe, i.e., similar to what we have illustrated in Figure 4.4. Given the P-wave velocities at reservoir depth and a maximum frequency of 4 Hz (recall that we filter the recordings using a passband of 1-4 Hz), we arrive at a value of 200 m for this criterion. This is hence the separation along the fault at which individual centroid priors are placed. We depict these initial centroid priors in Figure 4.9 as yellow stars. At the same time, the fault orientations at these positions are used to determine the six moment tensor entries in the initial priors. As for the depth and origin time T_0 in the $\mathbf{m}_i^{(0)}$, we use the values estimated by the KNMI for both configurations (i.e., 3 km for the depth and 2019-05-22T03:49:00.075s for the origin time). In total, 19 individual vectors $\mathbf{m}^{(0)}$ are concatenated in $\mathbf{m}_{\text{list}}^{(0)}$ for 1C-fault and 3C-fault. For the third case, we consider a centroid prior configuration consisting of a square grid of 2 km × 2 km, with the center again being the epicenter estimated by the KNMI. We use the same criterion (200 m) to determine the horizontal spacing be-

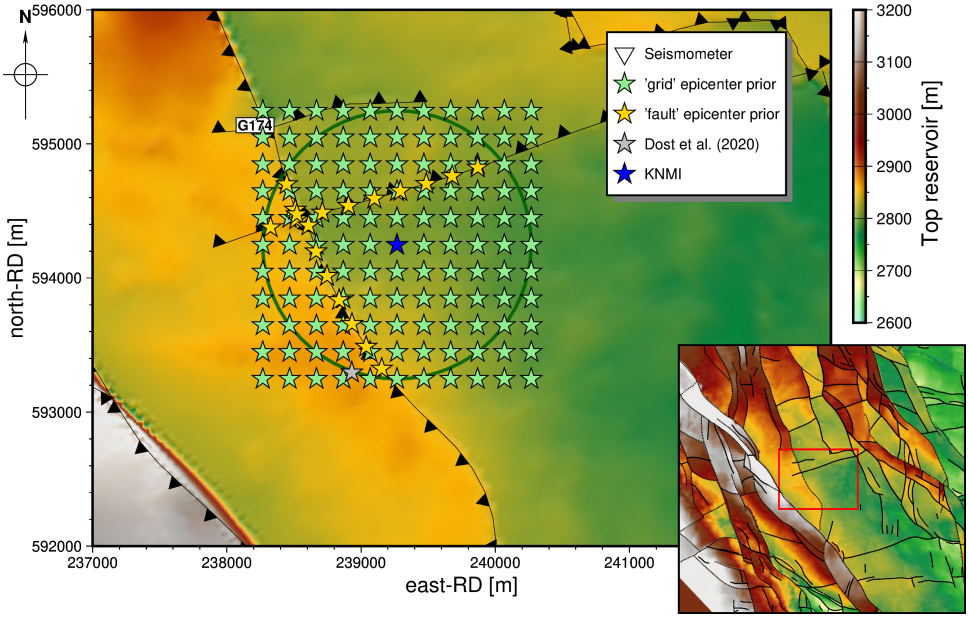

Figure 4.9: Horizontal positions of the different centroid priors for the two different prior configurations considered. The first configuration is guided by the known fault geometry inside the green circle and is represented by the yellow stars. This circle has a 1km radius and is centered at the epicenter estimated by the KNMI (blue star). The second centroid prior configuration uses a 2 km x 2 km grid with the KNMI-estimated epicenter at its center. These centroid priors are depicted as green stars.

tween the individual centroid priors. In Figure 4.9, we depict these as green stars. For the depth and origin time, we use identical values. Furthermore, for the MT, we assign a uniform value to each MT component for each individual $\mathbf{m}^{(0)}$. In total, we obtain 121 initial vectors $\mathbf{m}^{(0)}$ for this configuration.

4.6. APPLICATION TO FIELD DATA

For all cases described above (1C-fault, 3C-fault, 3C-grid), our multi-stage workflow consists of 20 stages. For the centroid prior configuration derived from the geometry of the known faults within the reservoir, $\mathbf{m}^{(0)}_{\text{list}}$ contains $N = 19$ $\mathbf{m}^{(0)}$, which implies a total of 380 stages. For the 3C-grid, a total of 121 initial priors serve as the starting model of the 121 multi-stage workflows (see Figure 4.9), resulting in a total of 2420 stages for this configuration. For each stage, we then compute the VR score based on the recordings $\mathbf{u}\left(\mathbf{x}^{(r)}, t; \overline{\mathbf{m}}\right)$ associated with the mean model $\overline{\mathbf{m}}$ of all 3000 individual models within that stage. Stages for which the VR score exceeds 0.95 are subsequently used to build our final posterior distribution. For each of the three cases considered, and for each of the initial centroid prior means, we show in Figure 4.10 the VR score associated with that $\overline{\mathbf{m}}^{(0)}_i$ of the 20 $\mathbf{m}^{(0)}$ in $\mathbf{m}^{(0)}_{\text{list}}$ for which the VR score attains its maximum. Note that this model's centroid is usually not at the location of the initial centroid prior mean (i.e., the centroid

in $\mathbf{m}_0^{(0)}$) because the models for which the waveforms best fit the observed recordings are often found in one of the later stages; see also Figure 4.3. For all three cases considered here, the highest VR scores are obtained in those chains for which the initial centroid prior mean is close to a fault.

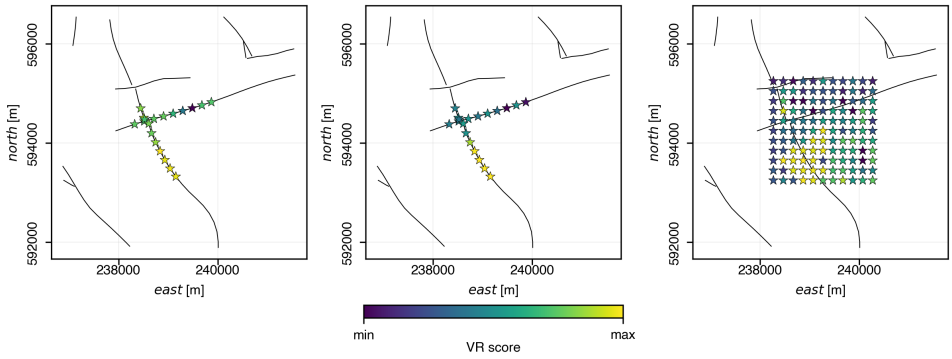

Figure 4.10: Maximum VR score in each of the chains associated with the different initial $\mathbf{m}^{(0)}$ for the three different cases considered (from left to right: 1C-fault, 3C-fault, and 3C-grid, respectively). Note that here we represent them by plotting the initial prior means (of the lateral positions) of the centroid.

4.6.1. ESTIMATED POSTERIOR

In Figure 4.11, we display the 1D marginal posterior distributions obtained from the selected stages of each configuration. In general, the mean value of these posteriors is fairly consistent across configurations, especially for 3C-fault and 3C-grid. For the 1C-fault case, the mean of the posteriors slightly deviates while at the same time having a slightly broader distribution compared to the other two cases. We attribute this to the fact that, for 3C-fault and 3C-grid, the additional data reduces the uncertainty of the estimates. In Figure 4.12, and for 3C-fault, we also plot the progression of the different stages associated with one of the individual centroid priors included in one of the $\mathbf{m}^{(0)}$ in $\mathbf{m}_{\text{list}}^{(0)}$. Specifically, we show the progression of that workflow (i.e., starting from that $\mathbf{m}_{\text{i}}^{(0)}$) that contains the stage that results in the overall maximum VR score. The vertical lines represent the start of different stages, and the red horizontal lines are the posterior means computed using the selected stages (after evaluating the VR scores for all stages). The progression follows a trend identical to the illustration in Figure 4.3(d), especially for the origin time T_0 with a slight variation for some others, such as for the depth and M_{nd} that shift monotonically to lower values. It is important to add that an initial estimate of T_0 was obtained using the envelope of the traces. This is described in detail in Section 2.6.2.

4.6.2. TRACES ASSOCIATED WITH THE POSTERIOR DISTRIBUTION

Using the posterior mean in Figure 4.11, we generate synthetic data and compare these with the observed data in Figure 4.13. In our workflow, the misfit in Equation 4.9 is based on 2.5 seconds of the observed particle displacement, bandpass filtered between 1 and

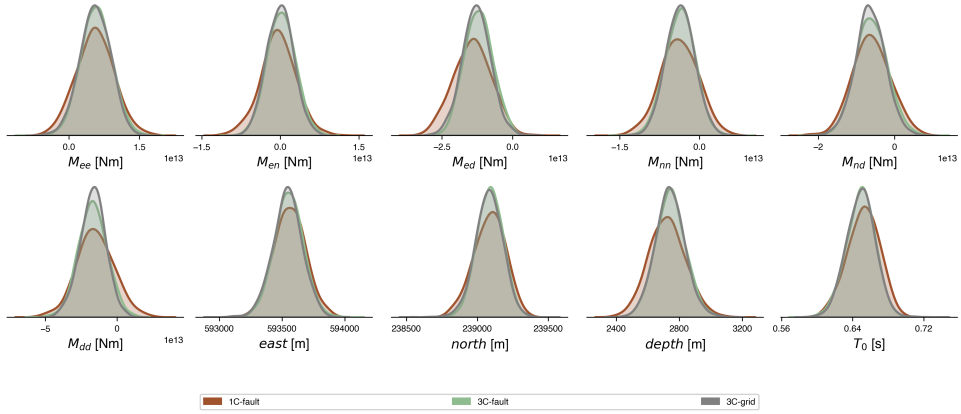

Figure 4.11: 1D marginal posterior distributions for the three different cases considered. '1C-fault': initial centroid prior configuration derived from the geometry of the known faults within the reservoir, and only the vertical particle displacement recordings are used. '3C-fault': initial centroid prior configuration derived from the geometry of the known faults within the reservoir, but both horizontal and vertical particle displacement recordings are used. '3C-grid': initial centroid prior located on a regular grid in a horizontal plane at the approximate (expected) depth of the event, and both horizontal and vertical particle displacement recordings are again used.

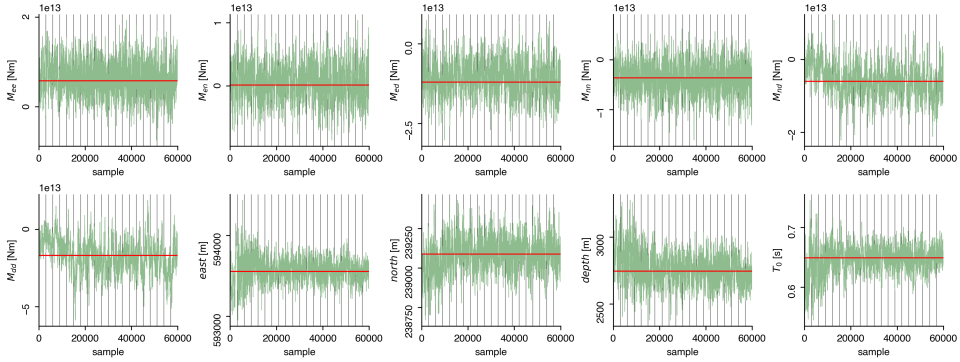

Figure 4.12: Progression of 20 stages from using one of the $\mathbf{m}^{(0)}$ in the 3C-fault configuration. The vertical lines represent different stages, whereas the red lines are the posterior mean (i.e., the mean of the green distributions in Figure 4.11) obtained from the selected stages based on the VR criteria for 3C-fault configuration.

4 Hz. Here, for consistency, we adopted the same values for these parameters. Additionally, we show in Figure 4.13 the maximum and minimum bounds using synthetic data generated from 1000 models drawn from the posterior distribution. We depict those bounds as a shaded area in Figure 4.13.

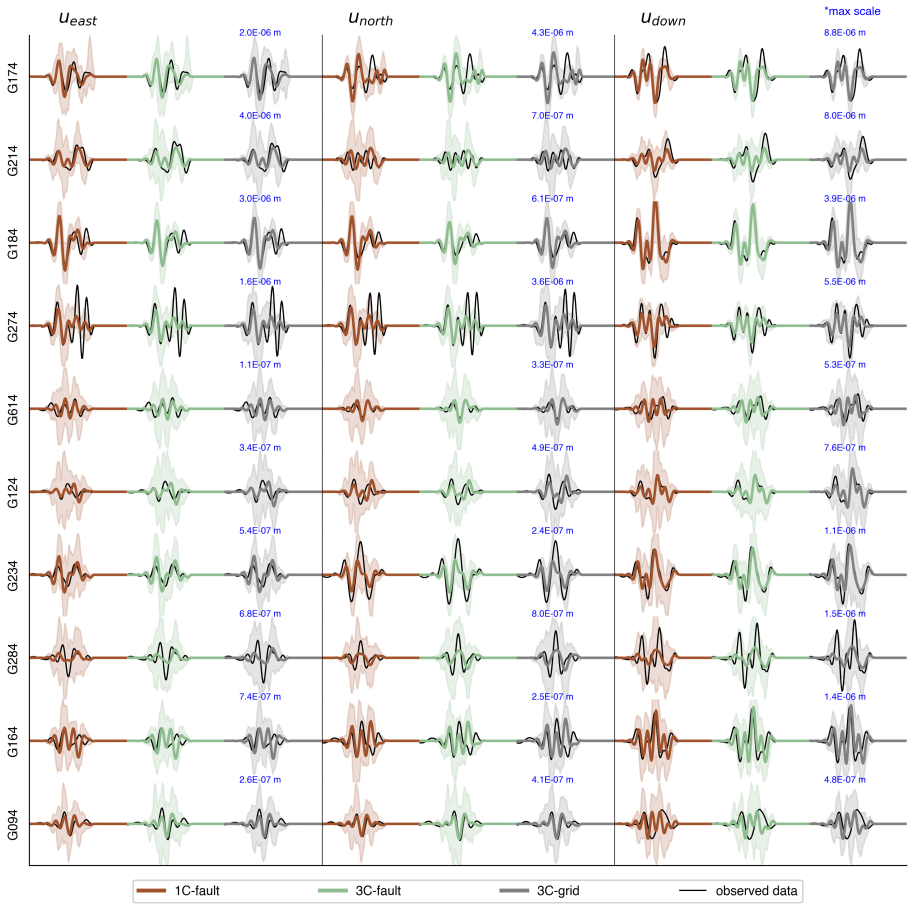

Figure 4.13: The comparison between observed and numerically modeled seismograms. The modeled seismograms are generated given the posterior mean estimated for each of the considered cases (see Figure 4.11). The shaded area is within the maximum and minimum bounds of a total of 1000 waveforms generated using 1000 models drawn from the posterior distributions in Figure 4.11. Each seismogram is filtered and tapered using the same parameters used in the multi-stage HMC workflow. The duration of each trace plotted here is 3.25s.

4.6.3. SOURCE CHARACTERISTICS

To investigate the source characteristics of the analyzed induced event, we first decompose the MTs of the posteriors shown in Figure 4.11. In this study, we do not limit our solutions to a single mechanism. We, therefore, decompose our moment tensor solutions into their ISO, DC, and CLVD components. We do this for each case (1C-fault, 3C-fault, and 3C-grid) and depict the decompositions in the Hudson plots in Figure 4.14. The mean MT for each case is represented by the beachball with the red outline. For all cases, the DC “region" is densely clustered (i.e., the center of the plot), with negative ISO components clearly outnumbering positive ISO components. This is often attributed to

the compaction due to the gas extraction (Dost et al., 2020). We show the posterior distributions of the different MT components in Figure 4.15 (bottom row). Furthermore, in the top row, we depict the translation of the MT solutions in Figure 4.14 to distributions of strike, dip, and rake. Here, we only show solutions with strikes between 90° and 180°, which are in accordance with the orientation of the fault close by (given the centroid posterior distributions).

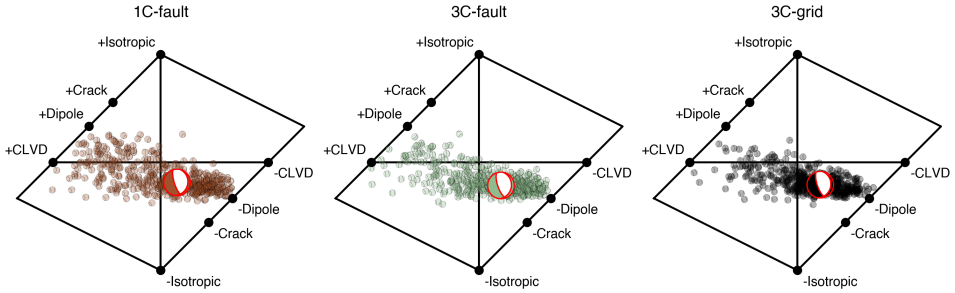

Figure 4.14: Hudson plot that shows the decomposition of the source mechanisms given the posterior distributions in Figure 4.11. The beachball with the red outlines represents the mean MT.

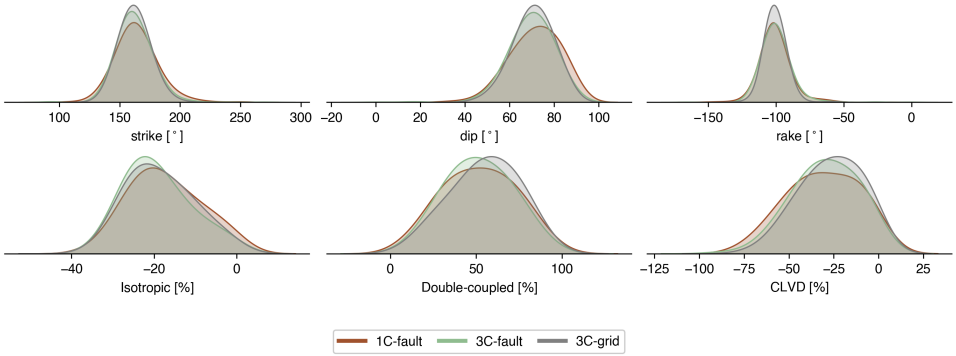

Figure 4.15: Top: The distributions of strike, dip, and rake solutions given the beachballs in Figure 4.14. Here we only show one part of the solutions closer to the orientation of the nearby major faults. Bottom: The marginal posterior distributions for different earthquake mechanisms given the decomposition in Figure 4.14.

We visualize the centroid posterior distributions using horizontal and vertical slices of the Groningen subsurface (Figure 4.16). In the top row, we show the depth of the top reservoir as a contour map, including the location of faults from Bourne and Oates (2017) at that depth. On top of these contour maps, we show the samples used to generate the 2D marginal posterior distributions of the lateral position of the centroids. We also plot the result from Dost et al. (2020) and the KNMI as the black beachball and blue star, respectively. The red beachball represents the mean MT which is also depicted in Figure 4.14 (beachball with red outlines). Not only do the posterior means of the (lateral)

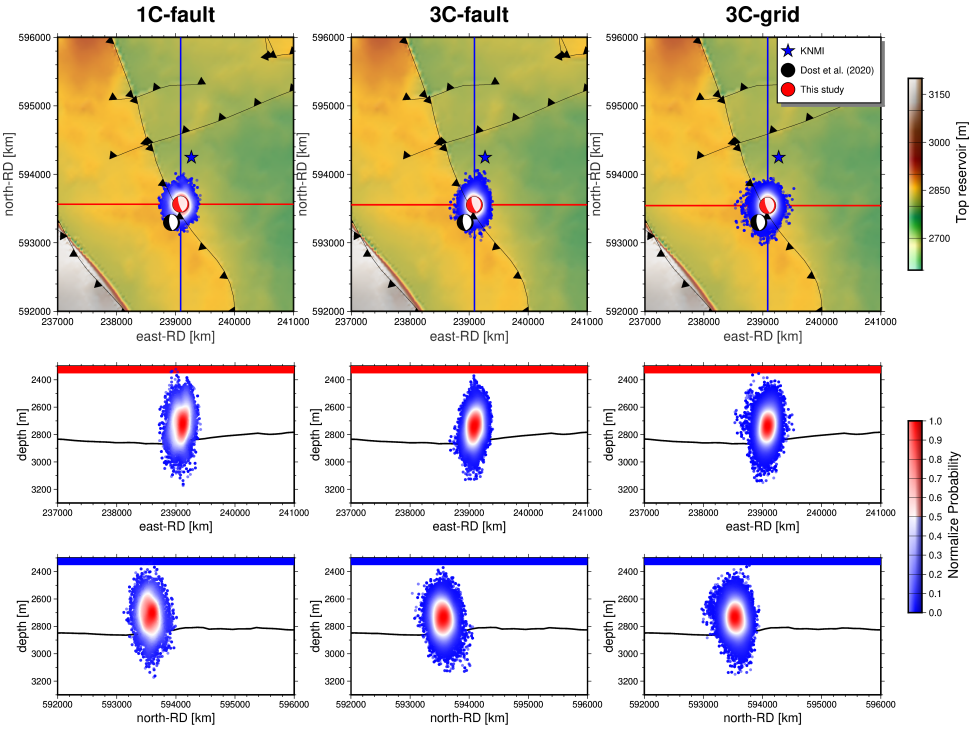

Figure 4.16: Comparison of samples used to generate centroid posterior distributions in Figure 4.11 (east, north, and depth) with the centroid estimated by Dost et al. (2020) and the KNMI. The samples are color coded with the density of centroid posteriors. The black line in the last two rows represent the top reservoir obtained from slicing the top reservoir map based on the red and blue line in the top row.

centroid positions coincide with the known fault, but also, the moment tensor solution agrees quite well with the strike of the nearby fault. On the vertical slices (middle and bottom rows), we depict the depth of the top reservoir as solid black lines. The location of the east-west vertical cross section and the north-south vertical cross section are shown as red and blue lines in the contour maps, respectively. For this specific earthquake, we find the posterior mean of the centroid to be slightly shallower than the centroid estimated by Dost et al. (2020). In fact, instead of being within the reservoir, we find the probability of having the earthquake nucleated above the reservoir is higher. The earthquake (model) parameter that has the strongest trade-off with depth is origin time. This is because an earlier origin time can be translated to an earthquake occurring at greater depth and vice versa. In this study, origin time uncertainty is considered, and the result shows that the estimated T_0 from the KNMI is lagging by a few milliseconds. As a caveat, however, we do not consider the uncertainty in the 3D velocity models, which may not only introduce amplitude variations but also affect the origin time and/or depth. For a more detailed comparison, in Table 4.1, we list the mean and standard deviation of our estimated parameters (for the MTs, we convert these into strike, dip, and rake solutions)

and compare them with the result of Dost et al. (2020) and the KNMI (hypocenter only).

4.7. DISCUSSION AND CONCLUSION

Using a probabilistic workflow incorporating the HMC algorithm, we estimate the source characteristics of a 3.4 M_L induced earthquake associated with gas extraction from the Groningen gas field. Specifically, we estimate the posterior probability density of ten earthquake parameters using two different sets of initial prior probabilities, of which one is used in conjunction with two sets of data: one consisting only of vertical component displacement recordings and a second one composed of the particle displacement in all three directions (east, north, down). We find that the posteriors estimated using both horizontal and vertical components of the seismograms (i.e., the latter data set) have similar shapes. At the same time, the one that only depends on the vertical component recordings yields a posterior that deviates (slightly) from the results of the other two cases while simultaneously being slightly broader. However, we find no substantial difference in the modeled seismograms associated with the different posterior means. In terms of runtime, using an 8-core MacBook Pro (2018 version), it took us a maximum of 3 minutes to run the 19 multi-stage workflows of the 1C-fault and 3C-fault case, and 12 minutes for the 121 multi-stage workflows of 3C-grid.

The main factor that affects the shape of the posteriors is uncertainty, which, in this case, is formulated as data and model uncertainty. In our study, we choose a uniform distribution for the model parameters to encode a state of ignorance (i.e., $\sigma_m \rightarrow \infty$). Whereas the data uncertainty is estimated individually for each component on each seismometer (and hence captured by σ_{ri} in A_{pq}, see Equation 4.12, where the indices r and i are associated with a specific receiver and component, respectively). It is assumed that the noise is uncorrelated. Prescribing the noise to be correlated will make the workflow more complex and computationally more costly and require us to estimate data covariance matrices. In addition, a study by Gu et al. (2018) reveals that in the case of induced seismicity, accounting for (potentially) correlated noise has relatively little effect compared to the uncertainty arising from the inaccuracy of the velocity model. Ideally, the latter is also formally included. The relation between a specific source model (i.e., a specific set of model parameters) and the particle displacement at the surface will, in that case, be quantified by means of a probability density function (Tarantola and Valette, 1981). Due to limited computational resources, however, we disregard the uncertainty associated with the velocity model. Including it (for our 3D velocity model) will require enormous computational effort as each 'cell' in the model must be varied according to their variance when computing the forward problem represented by Equation 4.1 (effectively, the Green's functions will become probability density functions). While using 1D velocity models, lateral heterogeneity is not considered, and therefore, the number of cells will be exponentially reduced, hence the computational burden. In general, using 3D models has improved the characterization of earthquake sources since they better represent the subsurface compared to 1D models (Hingee et al., 2011; Hejrani et al., 2017; Wang and Zhan, 2020).

Many studies involving MT inversions limit the model space to purely double-couple sources. Often, this limitation is justified by (presumed) a priori information of the source type. For example, a DC mechanism is usually sufficient to explain faulting in tec-

Configuration	Strike [°]	Dip [°]	Rake [°]	North [m]	East [m]	Depth [m]	T_0 [s]
1C-fault	163.9± 16.9	71.5 ± 10.6	-100.7 ± 11.4	593562 ± 125	239086 ± 111	2722 ± 115	0.652± 0.018
3C-fault	162.5 ± 15.0	69.8 ± 10.0	-99.8 ± 11.7	593554 ± 116	239088 ± 91	2744 ± 96	0.649± 0.016
3C-grid	162.2 ± 12.4	70.6 ± 8.6	-100.7 ± 8.0	593543± 117	239081± 97	2742 ± 94	0.649± 0.016
Dost et al. (2020)	163 ± 15	70 ± 10	-100 ± 12	593295	238931	2953 ± 94	-
KNMI	-	-	-	594247	239268	3000	-

Table 4.1: Comparison between earthquake parameters estimated in this study and estimation from Dost et al. (2020) and the KNMI. Specifically for T_0, the value is relative to the origin time estimated by the KNMI.

tonically active areas where volumetric components can be expected to be negligible. In the context of induced seismicity, however, numerous studies have found that non-DC components explain a substantive part of the observed recordings (Caputa et al., 2021; Cesca et al., 2013; Šílenỳ and Milev, 2008). In the context of the gas extraction below Groningen, a study by Willacy et al. (2019) uses waveform data to obtain moment tensor solutions, assuming that the earthquakes can be explained by DC mechanisms. Hence, they only estimated the best DC mechanisms of each observed earthquake. Meanwhile, another study by Kühn et al. (2020) (also focusing on the events in Groningen) reveals that ignoring non-DC components significantly affects the solution and data fit. In this study, we find the DC component to be dominant but still need the ISO and CLVD components to be non-zero in order to explain the data.

As for the centroid, we find that it is likely that the earthquake nucleated above the reservoir. In our case, the posterior mean is located a bit above 2.8 km depth. This is a small shift from the estimate by Dost et al. (2020), who estimate the earthquake to be located inside the reservoir. A recent study by Smith et al. (2020), however, finds that most of the Groningen earthquakes nucleated just above the reservoir, although this study does not include the event we are using here. Considering both the centroid and MT solution, we find that the models that best explain the recorded particle displacements correlate well with the nearby fault (see Figure 4.16).

For the workflow to be applied to a larger number of induced earthquakes, we believe a couple of additions would be beneficial. The first is related to the estimation of the data uncertainty. Since the workflow relies on Bayesian inference, the data uncertainty is rather critical while shaping reliable final posterior distributions. A second addition would be to allow for correlated noise. Particularly for the Groningen earthquakes, the effect of correlated noise for source characterization is not considered in any of the publications cited in this manuscript. Quantifying its effect on source parameter estimations would, therefore, be relevant.

APPENDIX: LINEARIZATION OF THE FORWARD PROBLEM

In the context of Hamiltonian Monte Carlo, a model $\mathbf{m}$ can be interpreted as the position of a particle in the $2N_m$-dimensional phase space (Betancourt, 2017). Using Equation 4.3, this particle's potential energy U, which is defined as $U(\mathbf{m}) \equiv -\ln \rho(\mathbf{m} \mid \mathbf{d}^{\text{obs}})$, therefore reads (Fichtner and Simutė, 2018; Masfara et al., 2022)

$$U(\mathbf{m}) = \frac{1}{2}\left(\mathbf{d}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\right)^{\top} \mathbf{C}_d^{-1}\left(\mathbf{d}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\right) + \frac{1}{2}\left(\mathbf{m} - \mathbf{m}^{(0)}\right)^{\top} \mathbf{C}_m^{-1}\left(\mathbf{m} - \mathbf{m}^{(0)}\right). \tag{4.9}$$

where we have, for convenience, ignored the proportionality constant (this does not affect our results as it is independent of $\mathbf{m}$.)

In this study, the HMC variant that involves a linearization of the forward problems is referred to as 'linearized HMC.' Linearization of the forward model implies a Taylor expansion of Equation 4.1 about the prior mean $\mathbf{m}^{(0)}$, and subsequently dropping higher order terms. The linear approximation of u_i reads (Fichtner and Simutė, 2018)

$$u_i\left(\mathbf{x}^{(\mathrm{r})}, t; \mathbf{m}\right) = u_i\left(\mathbf{x}^{(\mathrm{r})}, t; \mathbf{m}^{(0)}\right) + \sum_{p=1}^{N_m} \frac{\partial}{\partial m_p} u_i\left(\mathbf{x}^{(\mathrm{r})}, t; \mathbf{m}^{(0)}\right)\left(m_p - m_p^{(0)}\right). \tag{4.10}$$

Substituting this approximation in Equation 4.9, $U(\mathbf{m})$ can be written as follows:

$$U(\mathbf{m}) = \frac{1}{2}\sum_{p,q=1}^{N_m}\left(m_p - m_p^{(0)}\right)A_{pq}\left(m_q - m_q^{(0)}\right) + \sum_{p=1}^{N_m} b_p\left(m_p - m_p^{(0)}\right) + \frac{1}{2}c, \tag{4.11}$$

where A_{pq}, b_p, and c read

$$A_{pq} \equiv \frac{1}{T}\sum_{r=1}^{N_r}\sum_{i=1}^{3}\sigma_{ri}^{-2}\int_0^T \frac{\partial}{\partial m_p}u_i\left(\mathbf{x}^{(\mathrm{r})}, t; \mathbf{m}^{(0)}\right)\frac{\partial}{\partial m_q}u_i\left(\mathbf{x}^{(\mathrm{r})}, t; \mathbf{m}^{(0)}\right)dt + \frac{1}{N_m\sigma_m^2}, \tag{4.12}$$

$$b_p \equiv \frac{1}{T}\sum_{r=1}^{N_r}\sum_{i=1}^{3}\sigma_{ri}^{-2}\int_0^T \left[u_i\left(\mathbf{x}^{(\mathrm{r})}, t; \mathbf{m}^{(0)}\right) - u_i^{\mathrm{obs}}\left(\mathbf{x}^{(\mathrm{r})}, t\right)\right]\frac{\partial}{\partial m_p}u_i\left(\mathbf{x}^{(\mathrm{r})}, t; \mathbf{m}^{(0)}\right)dt, \tag{4.13}$$

and

$$c \equiv \frac{1}{T}\sum_{r=1}^{N_r}\sum_{i=1}^{3}\sigma_{ri}^{-2}\int_0^T \left[u_i\left(\mathbf{x}^{(\mathrm{r})}, t; \mathbf{m}^{(0)}\right) - u_i^{\mathrm{obs}}\left(\mathbf{x}^{(\mathrm{r})}, t\right)\right]^2 dt, \tag{4.14}$$

respectively. Here, σ_{ri}^{-2} encodes the data uncertainty for receiver number r and component i. This formulation implies that the noise is assumed to be uncorrelated. Similarly, a pairwise uncorrelated prior probability of the model parameters is considered. (i.e., $\mathbf{C}_m$ is diagonal). Using the three quantities above, $\frac{\partial U}{\partial \mathbf{m}}$ in Equation 4.7 can be replaced by

$$\frac{\partial U}{\partial m_p} = \sum_{q=1}^{N_m} A_{pq}\left(m_q - m_q^{(0)}\right) + b_p. \tag{4.15}$$

REFERENCES

Aki, K. and Richards, P. G. (2002). *Quantitative Seismology*. University Science Books, 2^{nd} edition.

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*.

Bommer, J. J., Stafford, P. J., Edwards, B., Dost, B., van Dedem, E., Rodriguez-Marek, A., Kruiver, P., van Elk, J., Doornhof, D., and Ntinalexis, M. (2017). Framework for a ground-motion model for induced seismic hazard and risk analysis in the Groningen gas field, the Netherlands. *Earthquake Spectra*, 33(2):481–498.

Bourne, S. and Oates, S. (2017). Induced seismicity within the Groningen gas field (v1.0.0) (available at https://zenodo.org/records/1035226). *Zenodo*.

Caputa, A., Rudziński, Ł., and Cesca, S. (2021). How to assess the moment tensor inversion resolution for mining induced seismicity: A case study for the Rudna mine, Poland. *Frontiers in Earth Science*, 9:671207.

Cesca, S., Rohr, A., and Dahm, T. (2013). Discrimination of induced seismicity by full moment tensor inversion and decomposition. *Journal of seismology*, 17:147–163.

Dost, B., Goutbeek, F., van Eck, T., and Kraaijpoel, D. (2012). Monitoring induced seismicity in the north of the Netherlands: status report 2010 (available at https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubWR/WR2012-03.pdf). *KNMI Scientific report*.

Dost, B., van Stiphout, A., Kühn, D., Kortekaas, M., Ruigrok, E., and Heimann, S. (2020). Probabilistic moment tensor inversion for hydrocarbon-induced seismicity in the Groningen gas field, the Netherlands, part 2: Application. *Bulletin of the Seismological Society of America*, 110(5):2112–2123.

Ellsworth, W. L., Llenos, A. L., McGarr, A. F., Michael, A. J., Rubinstein, J. L., Mueller, C. S., Petersen, M. D., and Calais, E. (2015). Increasing seismicity in the US midcontinent: Implications for earthquake hazard. *The Leading Edge*, 34(6):618–626.

Fichtner, A. and Simutė, S. (2018). Hamiltonian Monte Carlo inversion of seismic sources in complex media. *Journal of Geophysical Research: Solid Earth*, 123(4):2984–2999.

Fichtner, A., Zunino, A., and Gebraad, L. (2019). Hamiltonian Monte Carlo solution of tomographic inverse problems. *Geophysical Journal International*, 216(2):1344–1363.

Fichtner, A., Zunino, A., Gebraad, L., and Boehm, C. (2021). Autotuning Hamiltonian Monte Carlo for efficient generalized nullspace exploration. *Geophysical Journal International*, 227(2):941–968.

Foulger, G. R., Wilson, M. P., Gluyas, J. G., Julian, B. R., and Davies, R. J. (2018). Global review of human-induced earthquakes. *Earth-Science Reviews*, 178:438–514.

Gu, C., Marzouk, Y. M., and Toksöz, M. N. (2018). Waveform-based Bayesian full moment tensor inversion and uncertainty determination for the induced seismicity in an oil/gas field. *Geophysical Journal International*, 212(3):1963–1985.

Hejrani, B., Tkalčić, H., and Fichtner, A. (2017). Centroid moment tensor catalogue using a 3-D continental scale Earth model: Application to earthquakes in Papua New Guinea and the Solomon Islands. *Journal of Geophysical Research: Solid Earth*, 122(7):5517–5543.

Hingee, M., Tkalčić, H., Fichtner, A., and Sambridge, M. (2011). Seismic moment tensor inversion using a 3-D structural model: Applications for the Australian region. *Geophysical Journal International*, 184(2):949–964.

Jost, M. u. and Herrmann, R. (1989). A student's guide to and review of moment tensors. *Seismological Research Letters*, 60(2):37–57.

Komatitsch, D. and Tromp, J. (2002). Spectral-element simulations of global seismic wave propagation—I. Validation. *Geophysical Journal International*, 149(2):390–412.

Kühn, D., Heimann, S., Isken, M. P., Ruigrok, E., and Dost, B. (2020). Probabilistic moment tensor inversion for hydrocarbon-induced seismicity in the Groningen gas field, the Netherlands, part 1: Testing. *Bulletin of the Seismological Society of America*, 110(5):2095–2111.

Li, N., Li, B., Chen, D., Wang, E., Tan, Y., Qian, J., and Jia, H. (2020). Waveform characteristics of earthquakes induced by hydraulic fracturing and mining activities: Comparison with those of natural earthquakes. *Natural Resources Research*, 29:3653–3674.

Lui, S. K., Helmberger, D., Yu, J., and Wei, S. (2016). Rapid assessment of earthquake source characteristics. *Bulletin of the Seismological Society of America*, 106(6):2490–2499.

Masfara, L. O. M., Cullison, T., and Weemstra, C. (2022). An efficient probabilistic workflow for estimating induced earthquake parameters in 3D heterogeneous media. *Solid Earth*, 13(8):1309–1325.

Mustać, M. and Tkalčić, H. (2016). Point source moment tensor inversion through a Bayesian hierarchical model. *Geophysical Journal International*, 204(1):311–323.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov chain Monte Carlo*, 2(11):2.

Nepveu, M., van Thienen-Visser, K., and Sijacic, D. (2016). Statistics of seismic events at the Groningen field. *Bulletin of Earthquake Engineering*, 14:3343–3362.

Ntinalexis, M., Bommer, J. J., Ruigrok, E., Edwards, B., Pinho, R., Dost, B., Correia, A. A., Uilenreef, J., Stafford, P. J., and van Elk, J. (2019). Ground-motion networks in the Groningen field: Usability and consistency of surface recordings. *Journal of Seismology*, 23(6):1233–1253.

Peng, C., Jiang, P., Ma, Q., Wu, P., Su, J., Zheng, Y., and Yang, J. (2021). Performance evaluation of an Earthquake early warning system in the 2019–2020 M 6.0 Changning, Sichuan, China, seismic Sequence. *Frontiers in Earth Science*, 9:699941.

Pickering, M. (2015). An estimate of the earthquake hypocenter locations in the Groningen gas field (available at https://nam-feitenencijfers.data-app.nl/download/rapport/8b16a9eb-a607-494a-baa5-56e349b6ed0c?open=true). *Nederlandse Aardolie Maatschappij (NAM)*.

Romijn, R. (2017). Groningen velocity model 2017—Groningen full elastic velocity model September 2017 (available at https://nam-feitenencijfers.data-app.nl/download/rapport/9a5751d9-2ff5-4b6a-9c25-e37e76976bc1?open=true). *Nederlandse Aardolie Maatschappij (NAM)*.

Ruigrok, E., Domingo-Ballesta, J., van den Hazel, G.-J., Dost, B., and Evers, L. (2019). Groningen explosion database. *First Break*, 37(8):37–41.

Ruigrok, E. and Dost, B. (2019). Seismic monitoring and site-characterization with near-surface vertical arrays. In *Near Surface Geoscience Conference and Exhibition*, pages 1–5.

Sarhosis, V., Dais, D., Smyrou, E., and Bal, İ. E. (2019). Evaluation of modelling strategies for estimating cumulative damage on Groningen masonry buildings due to recursive induced earthquakes. *Bulletin of Earthquake Engineering*, 17(8):4689–4710.

Šílenỳ, J. and Milev, A. (2008). Source mechanism of mining induced seismic events—Resolution of double couple and non double couple models. *Tectonophysics*, 456(1-2):3–15.

Simutė, S., Boehm, C., Krischer, L., Gokhberg, A., Vallée, M., and Fichtner, A. (2022). Bayesian seismic source inversion with a 3-D Earth model of the Japanese islands. *Journal of Geophysical Research: Solid Earth*, page e2022JB024231.

Smith, J. D., White, R. S., Avouac, J.-P., and Bourne, S. (2020). Probabilistic earthquake locations of induced seismicity in the Groningen region, the Netherlands. *Geophysical Journal International*, 222(1):507–516.

Spetzler, J. and Dost, B. (2017). Hypocentre estimation of induced earthquakes in Groningen. *Geophysical Journal International*, 209(1):453–465.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. EngineeringPro collection. Society of industrial and applied mathematics, Philadelphia, PA, USA.

Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature physics*, 2(8):492–494.

Tarantola, A. and Valette, B. (1981). Inverse problems = Quest for information. *Journal of Geophysics*, 50(1):159–170.

Trippetta, F., Petricca, P., Billi, A., Collettini, C., Cuffaro, M., Lombardi, A. M., Scrocca, D., Ventura, G., Morgante, A., and Doglioni, C. (2019). From mapped faults to fault-length earthquake magnitude (FLEM): A test on Italy with methodological implications. *Solid Earth*, 10(5):1555–1579.

van Thienen-Visser, K. and Breunese, J. (2015). Induced seismicity of the Groningen gas field: History and recent developments. *The Leading Edge*, 34(6):664–671.

Vlek, C. (2018). Induced earthquakes from long-term gas extraction in Groningen, the Netherlands: Statistical analysis and prognosis for acceptable-risk regulation. *Risk analysis*, 38(7):1455–1473.

Wang, X. and Zhan, Z. (2020). Moving from 1-D to 3-D velocity model: Automated waveform-based earthquake moment tensor inversion in the Los Angeles region. *Geophysical Journal International*, 220(1):218–234.

Willacy, C., van Dedem, E., Minisini, S., Li, J., Blokland, J. W., Das, I., and Droujinine, A. (2018). Application of full-waveform event location and moment-tensor inversion for Groningen induced seismicity. *The Leading Edge*, 37(2):92–99.

Willacy, C., van Dedem, E., Minisini, S., Li, J., Blokland, J.-W., Das, I., and Droujinine, A. (2019). Full-waveform event location and moment tensor inversion for induced seismicity. *Geophysics*, 84(2):KS39–KS57.

5

MOMENT TENSOR AND RUPTURE DIRECTIVITY ANALYSIS OF INDUCED EARTHQUAKES IN GRONINGEN GAS FIELD, THE NETHERLANDS

In Groningen, the Netherlands, earthquakes are solely induced by the extraction of natural gas. Over the past decade, the induced events have emerged as a significant concern. Their occurrence necessitates not only the importance of understanding the earthquakes' source characteristics but also the development of more sophisticated characterization methods. The objective of this study is twofold. First, we estimate the induced earthquakes' source characteristics using a probabilistic method paired with geological knowledge of the Groningen subsurface. Second, we investigate the rupture characteristics of the earthquakes. Specifically, we aim to estimate the centroid coordinates and moment tensors (MTs) of ten major induced earthquakes (magnitude > 2 M_L) in the area. Accompanying these probabilistic MT estimates is an analysis of the rupture directivity of the selected earthquakes to gain insight into the dynamics of the earthquake's ruptures. For the probabilistic centroid moment tensor (CMT), we use an iterative workflow based on the Hamiltonian Monte Carlo (HMC) algorithm paired with available Groningen subsurface knowledge. For the estimation of rupture directivity, we employ the Empirical Green's Function (EGF) method and benefit from the dense and large seismic monitoring network in the Groningen area. The recovered (most likely) CMTs reveal that the lateral centroid coordinates of the selected earthquakes align with the available Groningen fault map. Furthermore, the depths are chiefly distributed in the vicinity of either the top or the

Parts of this chapter are currently being prepared for a journal article.

bottom of the reservoir layer. In terms of source mechanisms, earthquakes are predominantly governed by double-couple mechanisms featuring normal faulting. Converting the MT results to strike and dip, we obtain values comparable with the strike and dip of the (known) existing faults intersecting the reservoir (at the mean centroid location). When it comes to the rupture directivity, our results suggest that the events exhibit a minor directivity effect. Despite this minimal effect, the rupture directions consistently align with strike values from both CMT inversions and the available fault map.

5.1. INTRODUCTION

Induced seismicity in the Groningen gas field, the Netherlands, has become a significant concern. Since the first recorded event in December 1991, the magnitudes of the induced events in the area have gradually increased, reaching a maximum (local) magnitude of 3.6 in 2013 (Muntendam-Bos et al., 2022; Dost and Kraaijpoel, 2013). As these magnitudes increased, measured surface accelerations also rose, causing damage to buildings near the epicenter. To better understand the physics of these earthquakes, a number of measures have been taken, initiatives launched, and developments initiated. A notable example is the expansion of the Groningen seismic network (Dost et al., 2017). Not only was this expansion initiated to lower the magnitude threshold of earthquakes, but also to improve the quality of the characterization of those earthquakes (Ntinalexis et al., 2019).

In general, studying earthquake characteristics is important to better understand the key factors driving destructive earthquakes. Also, earthquake characterization facilitates the classification of events; for example, it may clarify whether those events are natural or anthropogenic (van Thienen-Visser and Breunese, 2015). In addition, the study of earthquakes may validate or augment available subsurface data such as velocity models and fault maps or challenge the existing subsurface knowledge (Paulssen et al., 2022). Once the earthquake is well-characterized, one can perform multiple analyses, including seismic hazard assessment (Lui et al., 2016). For example, by predicting (peak) ground motions, simulating different types of fault behavior, or investigating possible hazard mitigation scenarios. Characterization of the induced events in Groningen has been performed using various methods and data, depending on the estimated parameters. When it comes to seismology, typical parameters are the centroid, which consists of three coordinate components, the moment tensor (MT), which has six independent components, and the origin time of the event. The following authors have presented several important findings regarding the induced events in Groningen: Spetzler and Dost (2017) estimate the earthquake centroid using a one-dimensional (1D) velocity model of Groningen and find the depth of the induced earthquakes to be at the reservoir level (i.e., about 3 km depth). Also, in estimating centroid locations, Smith et al. (2020) use a three-dimensional (3D) velocity model of Groningen and pair it with a probabilistic algorithm. They find that, instead of being within the reservoir layers, there is a significant probability that many of the earthquakes nucleate just above the reservoir layer. Other authors, such as Dost et al. (2020) and Willacy et al. (2018), include the MT in their analysis. Both reveal that the earthquakes appear to be mainly driven by normal faulting, which can be expected in the Groningen subsurface (Buijze et al., 2019), and with depths mainly within the reservoir. As for the centroid lateral coordinates, all authors show that most

of the earthquakes align with known major faults.

In this study, we focus on analyzing two earthquake parameter categories. The first set of parameters pertains to the centroid moment tensors (CMTs), whereas the second involves rupture directivity. In essence, the estimated CMT parameters are similar to the parameters considered by the aforementioned authors. In this study, however, we use an efficient probabilistic inversion workflow based on the Hamiltonian Monte Carlo (HMC) algorithm paired with the *a priori* available geological knowledge of the area for optimum results. We furthermore use the available detailed 3D subsurface velocity models of Groningen to generate our modeled data/seismograms. The purpose of the 3D models is to account for the imprint of lateral heterogeneity in the Groningen subsurface on the recorded waveforms. In general, using 3D subsurface models has been shown to improve the quality of the estimated parameters since they better represent the true subsurface (Wang and Zhan, 2020; Zhu and Zhou, 2016). We apply the workflow to estimate the CMTs of ten of the largest events recorded in Groningen to date (exceeding a local magnitude of 2). We show the epicenter of the ten earthquakes in Figure 5.1A, overlaid with the map of major subsurface faults and the depth of the top of the reservoir (Bourne and Oates, 2017). In Figure 5.1B, we depict for each of the ten events the azimuthal distribution of the geophones used in this study (i.e., borehole geophones at a depth of 200 m whose seismograms are (quantitatively) compared against the modeled seismograms in the probabilistic inversion workflow).

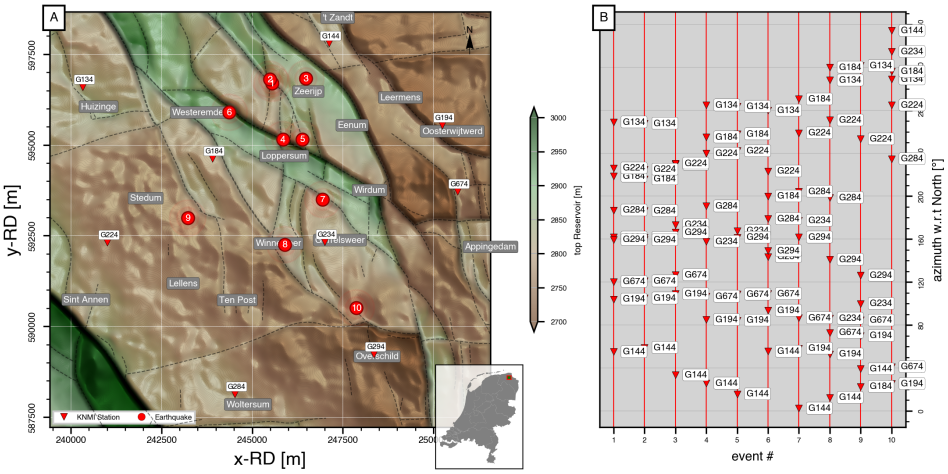

Figure 5.1: (A) Study area with the labeled locations representing local villages, the inverted triangles representing the KNMI seismometers, and the red circles representing the earthquake events studied in this chapter. (B) Seismometer coverage with the inverted triangles now representing seismometer azimuth with respect to the north for each earthquake event shown in (A).

The second category of earthquake parameters we investigate involves the direction and speed of rupture propagation. Depending on its directivity, an earthquake's impact may be different in different directions and hence locations. For example, in the direction of propagation, the ground motion is expected to be higher, resulting in more de-

structive motion and vice versa (Cesca et al., 2011; Courboulex et al., 2013). An analysis by Oates et al. (2024) reveals how several earthquakes in Groningen exhibit a clear directivity pattern. Therefore, the second focus of this chapter is to analyze the directivity of the ten events for which we estimate their CMTs. Following that, the output of both analyses, CMT and directivity, is compared with the available subsurface data (i.e., local fault strike) to evaluate their consistency. In what follows, we first focus on estimating the CMTs and hence start by introducing the theory underlying the HMC workflow. We then continue with the results of the CMT estimation and the decomposition of the moment tensors, which allows one to extract specific characteristics of the source mechanisms. In the second part, we quantify the rupture directivity. We again start by introducing the methodology, usually referred to as the Empirical Green's function (EGF) method (e.g., Abercrombie, 2015). This is followed by a discussion of the selection of the small (*child*) events that are later used to deconvolve the recordings of the ten larger earthquakes. Finally, the results of both analyses are interpreted and discussed in the discussion section of this article, including recommendations to optimize the results further.

5

5.2. PROBABILISTIC CENTROID MOMENT TENSOR INVERSION

Estimating the CMT of a seismic event requires an inversion technique that minimizes the difference between the observed seismograms and the modeled seismograms. While the observed data are collected from selected seismic stations (in our case, these are shown in Figure 5.1A), the modeled seismograms are generated using a specific set of CMT parameters via the solution of what is referred to as the 'forward problem.' By minimizing the difference (or *misfit*) between the observed and modeled seismograms, one can infer those CMT parameters that most likely resulted in the observed waveforms. In a Bayesian context, this is slightly different in the sense that not a single, optimum solution is sought but an ensemble of solutions (i.e., an ensemble of CMTs) whose density is proportional to the posterior probability. Below, we start by introducing the forward problem we use to generate our modeled data, followed by a description of Bayes formulation and some details of the HMC implementation. We end this section with a presentation of our findings.

5.2.1. THEORY

Both observed and modeled seismograms in this study are in the form of particle displacement. Assuming the extent of a seismic source is sufficiently small (in comparison to the wavelength of the recordings used), it can be considered a point source. This implies that particle displacement may mathematically be written as (Aki and Richards, 2002):

$$u_i\left(\mathbf{x}^{(\mathrm{r})}, t\right) = \sum_{j=1}^{3}\sum_{k=1}^{3} M_{jk}(t, T_0) * G_{ij,k}\left(\mathbf{x}^{(\mathrm{r})}, t; \mathbf{x}^{(\mathrm{a})}\right), \tag{5.1}$$

where u_i represents the displacement in the x_i direction at position $\mathbf{x}^{(\mathrm{r})}$ (geophone location), and $\mathbf{x}^{(\mathrm{a})}$ the position of the source that starts to radiate seismic energy at T_0. The index i can take on values 1, 2, and 3, and hence the u_i combine to form the displacement vector $\mathbf{u} \equiv (u_1, u_2, u_3)$. These components correspond to the particle displacement in the east, north, and down directions, respectively. The moment tensor element M_{jk} repre-

sents one of the nine components of the moment tensor $\mathbf{M}$, a 3×3 symmetric matrix with six independent components. Meanwhile, the subscripts j and k represent the direction in which a force is acting and pointing, respectively (hence *moment* tensor; $\mathbf{M}$ has units of Newton-meters). The comma after the second subscript of the Green's function $G_{ij,k}$ implies a spatial derivative in the k direction with respect to $\mathbf{x}^{(a)}$. Finally, $*$ represents a time-domain convolution. In general, the moment tensor $\mathbf{M}$ is time-dependent. Usually, all moment tensor components are assumed to have the same time dependence, i.e., a synchronous source. Here, we furthermore assume instantaneous rupturing. This is not an uncommon assumption for (relatively small) induced seismic events. This assumption implies that the time dependence of $\mathbf{M}$ is modeled using a Heaviside function.

In this study, we estimate the CMTs using a probabilistic algorithm that relies on Bayes' theorem. In general, Bayes' theorem quantifies the probability of a model explaining the observed data given prior knowledge and available observed data. Mathematically, it can be written as

$$\rho(\mathbf{m}|\mathbf{d}^{\text{obs}}) \propto \rho(\mathbf{d}^{\text{obs}}|\mathbf{m})\rho(\mathbf{m}), \tag{5.2}$$

where $\rho(\mathbf{m}|\mathbf{d}^{\text{obs}})$ represents the posterior probability distribution, $\rho(\mathbf{d}^{\text{obs}}|\mathbf{m})$ the likelihood, and $\rho(\mathbf{m})$ the prior probability distribution. In our case, the model vector $\mathbf{m}$ contains the ten CMT parameters we want to infer, with $\rho(\mathbf{m})$ the prior knowledge we possess of these parameters. The observed data is contained in the data vector $\mathbf{d}^{\text{obs}}$. The likelihood $\rho(\mathbf{d}^{\text{obs}}|\mathbf{m})$ quantifies the difference between $\mathbf{d}^{\text{obs}}$ and the modeled data (numerical solution of Equation 5.1) computed for a specific set of model parameters (i.e., a specific $\mathbf{m}$). In other words, the product in the right-hand side of Equation 5.2 measures the overall probability of a model $\mathbf{m}$, given the observed data $\mathbf{d}^{\text{obs}}$.

In order to estimate the posterior, many different sampling algorithms exist. In this study, we use the Hamiltonian Monte Carlo (HMC) algorithm. The HMC algorithm is described extensively in Neal (2011), and proven to be highly efficient in comparison to other (more generic) types of probabilistic inversion algorithms (e.g., Metropolis-Hastings algorithms) when tuned optimally. This is because HMC uses a deterministic process based on Hamiltonian dynamics, which guides the exploration of the model space. This is contrary to Metropolis-Hastings (MH) algorithms, for example, which randomly sample the model space. In the context of CMT parameters, Fichtner and Simutė (2018) enhanced the efficiency of the HMC sampler by linearizing the forward problem. This approach was later adopted by Simutė et al. (2023) to estimate the CMTs of several major earthquakes on the Japanese peninsula. In application to induced seismic events, in Chapter 2, we embed the same linearization and HMC sampling approach in an efficient probabilistic workflow. This workflow relies on the process of iteratively linearizing the forward problem around the centroid, and updating this centroid after estimation of the ("local") posterior probability. We refer to this process as a *multi-stage workflow*. To further reduce computational costs, in Chapter 4, we pair the multi-stage workflow with the available geological prior knowledge. They showcase this workflow by estimating the CMT parameters of a M_L 3.4 event, which occurred in 2019 in Groningen. In this study, we also adopt that approach and include a-priori geological knowledge in the form of a fault map to estimate the CMTs of the ten selected events.

5.2.2. IMPLEMENTATION

In this section, we summarize the most important aspects of our approach, including the use of an iterative probabilistic workflow. For technical details regarding the workflow, we refer to Chapter 2 and 4. For this study, we feed the aforementioned multi-stage workflow with (initial) prior information from the fault geometry in the vicinity of the studied earthquakes. This implies that, for each of the ten events, we generate n initial model vectors denoted $\mathbf{m}_1^0$, $\mathbf{m}_2^0$,...,$\mathbf{m}_n^0$. Each initial model $\mathbf{m}_i^0$ is associated with a different initial estimate of the event's centroid and moment tensor. These are derived from the fault geometry and serve as the starting model of a separate multi-stage workflow. In practice, these separate multi-stage workflows are run in parallel. Being the starting model of an individual multi-stage workflow, the first linearization is performed around the centroid $\mathbf{x}^{(a)}$ and T_0 contained in $\mathbf{m}_i^0$ (note that the particle displacement in Equation (5.1) already depends linearly on the moment tensor components, and hence this relationship does not need to be linearized). In our case, the centroid entries of the $\mathbf{m}_i^0$ are determined by the geometry of the faults close to the event's hypocenter, as it was estimated by the Royal Netherlands Meteorological Institute (KNMI) (depicted in Figure 5.1A). Around these hypocenters, we draw a circle with a diameter of 2 km and sample the fault segments inside this circle (see, for example, the center plot of Figure 4.10 in Chapter 4). Each sample is then contained in a separate $\mathbf{m}_i^0$, and serves as the starting model of one of the n separate multi-stage workflows. The density of the sampling along the faults is based on the maximum frequency of the waveforms (in $\mathbf{d}^{\text{obs}}$). In our case, we use P-wave arrivals bandpass filtered between 1 and 4 Hz contained in cosine-tapered windows with a duration of 3 seconds. Given the upper frequency of 4 Hz and a P-wave velocity of approximately 3500 m/s (the lowest P-wave velocity between 2800 and 3000 m depth in the region of interest), the shortest wavelength (at reservoir depth) is approximately 875 m.

To ensure that the HMC algorithm samples the global (potential energy) minimum, the along-strike separation between the centroids of the separate $\mathbf{m}_i^0$ is set to less than a quarter of a wavelength, i.e., 200 m. Note that we only sample the faults horizontally and not vertically. This is because the KNMI, by default, sets the centroids at a depth of 3 km, whereas the reservoir thickness is, on average, only 300 m. The latter is in the same range as one-quarter of a wavelength: having the initial $\mathbf{m}_i^0$ all at 3 km depth should hence suffice to ensure sampling of the global minimum. Finally, the strike and dip of the faults at the positions of the $\mathbf{m}_i^0$ are converted to moment tensor elements (assuming the rake is -90 °). Since we employ a probabilistic algorithm, we include uncertainties associated with both the data and the model parameters in the form of data and model (co-)variance. We set the data variance to 10% of the maximum amplitude of the recorded seismograms for each receiver's component individually. This is an overestimation of the actual data variance, but by overestimating this value, we aim to accommodate simultaneously the inaccuracy of our velocity model (which leads to disparity in the synthetic seismograms) and the noise present in the recorded seismograms, although alternatively, both can be estimated individually (Duputel et al., 2011).

For the uncertainty of the model parameters, we use a different approach, which depends on the model parameter. First, for the epicenter, we extract latitude and longitude values from a number of previous studies in the area (Smith et al., 2020; Dost et al., 2020;

Spetzler and Dost, 2017). Based on the values published in those studies, we set the variance to be 300 m in both horizontal directions. The uncertainty of the depth we set to 100 m. We also account for uncertainty in the origin time, which additionally reflects the inaccuracy of the velocity model. For this, we set the uncertainty to be 0.1 s, which is slightly lower than half of the wavelength duration, given the upper frequency we use is 4 Hz. For the moment tensor components, we set the uncertainty of all components to infinity to encode the state of ignorance, which means that all values for each parameter are equally weighted as a solution despite the mean values being pre-determined (concatenated in $\mathbf{m}^0$).

5.2.3. RESULTS

We present the marginal posterior probabilities for the centroids of the ten selected events in Figure 5.2. In the main box of Figure 5.2, these centroids are shown as scatter points overlaid on the top reservoir map. The red beach ball represents the mean moment tensor solution, plotted at the lateral position of the mean centroid (i.e., the average location of the event's scatter dots). Both the moment tensors and the epicenters demonstrate good agreement with major faults in the area. Furthermore, to detail the depth profiles of the models constituting the posterior probability, we collapse the centroids onto cross-sections along several profiles slicing through the Groningen subsurface. These are displayed as smaller insets on the right side (excluding the one with the red frame). The depths are not uniformly distributed in the middle of the reservoir but are concentrated at either the top or the bottom of the reservoir layer (Rotliegend sandstone). The red box at the bottom right magnifies the area around event 9 and serves to highlight the "size" of the posterior with respect to a circle with a diameter of 2 km around the KNMI-reported epicenter (the red-shaded area). The green dots represent samples of the faults within this circle. These samples correspond to the initial centroids associated with the separate $\mathbf{m}_i^0$.

We then decompose the moment tensor solutions from each event into their isotropic (ISO), double-couple (DC), and compensated linear vector dipole (CLVD) components. These decompositions are shown in the Hudson plots in Figure 5.3. The mean solutions are predominantly double-couple (i.e., close to the middle of the Hudson plot), which is attributed to slip on a fault. However, they all have (slightly) negative ISO components, which points to a decrease in volume. In Groningen, this can be attributed to compaction resulting from the gas extraction in the area (e.g., Smith et al., 2019). While the mean solution is close to double-couple, the optimum solutions are skewed towards a higher percentage of the negative CLVD components. Negative CLVD components in themselves can be seen as the movement of materials from the lateral axes in combination with compression in the vertical axis, both without volume change. In the context of mining, it is often associated with the collapse of voids.

To evaluate how our solutions explain the observed data, we compare the observed data against the modeled data generated from our mean CMT solution. We plot this comparison in Figure 5.4 (red and black seismograms) with the horizontal axis corresponding to 4 seconds of seismograms. In running the workflow, we set the uncertainty of the observed data to 10% of the maximum amplitude of the seismogram on each station of each component. This addition of uncertainty allows the workflow to accommo-

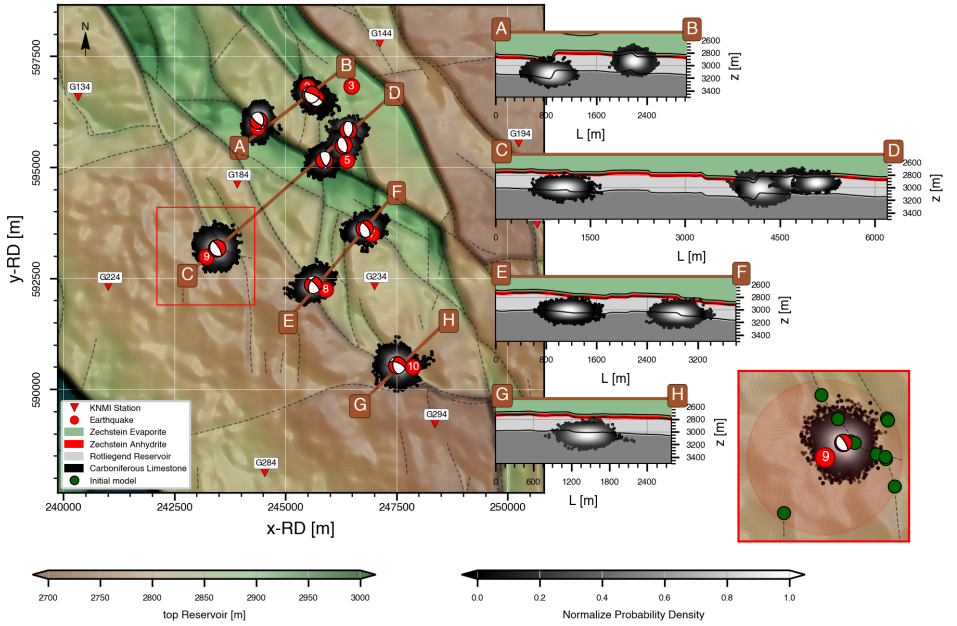

Figure 5.2: Main frame: Marginal posteriors of the centroid for each of the ten events (scatter clouds) plotted on top of a contour of the depth of the top reservoir. Black dashed lines delineate known faults cutting through the top reservoir (Rotliegend). The numbered red circles represent the (KNMI-estimated) epicenters. The beachballs represent the DC component of the mean MT of the posterior probability. Cross-sections along a number of profiles (solid brown lines in the main frame) are shown as four vertically separated insets to the right of the main frame. The plot at the bottom right (red colored box) is a magnification of the area of event #9. It shows a circle with a diameter of 2km as a red-shaded area surrounding the KNMI's epicenter, including samples of faults (green dots), which served as initial centroid priors (i.e., these are associated with individual $\mathbf{m}_i^0$).

date solutions that explain modeled data within that range. To examine how the uncertainty affects our solutions and, ultimately, the generated modeled data, in Figure 5.4, we plot the upper and lower bounds delineating the ensemble of waveforms generated from a total of 100 samples from our posterior probability (dark grey solid lines bounding the gray area in Figure 5.4). We summarize our CMT inversion results in table 5.1.

5.3. RUPTURE PROPAGATION ANALYSIS

A seismic rupture's energy is usually not radiated uniformly in all directions. Consequently, different regions on the Earth's surface are affected differently (Courboulex et al., 2013). This irregular distribution of seismic energy can be attributed to several factors, including the direction in which the rupture propagates along a fault. When the rupture propagates in a particular direction, the seismic waves, and hence the released energy, tend to be focused along that path. This phenomenon is known as the *directivity effect* (Cesca et al., 2011). Seismic waves propagating in the direction of rupture propagation

#	Date	Strike	Dip	Rake	M_w	ISO	CLVD	Lat	Lon	Depth
		[°]				[%]		[m]		
1	04/10/2021	295±8	74±9	-109±16	2.25±0.08	27±8	25±17	596563±111	245664±112	2935±62
2	04/10/2021	292±7	77±7	-115±13	2.02±0.07	28±8	23±17	596639±103	245572±108	2932±58
3	11/03/2017	18±9	57±12	-71±20	1.99±0.09	21±11	31±20	595836±108	246413±156	2917±63
4	08/02/2018	152±8	64±12	-114±20	1.90±0.07	31±9	30±17	595158±110	245877±112	3022±79
5	11/06/2021	348±7	59±13	-81±24	1.95±0.08	16±9	31±21	595489±128	246310±132	2906±59
6	14/07/2020	127±8	63±13	-120±20	2.51±0.07	27±11	30±19	596049±157	244393±116	3131±70
7	09/06/2019	323±8	60±13	-113±27	2.16±0.09	28±11	25±17	593612±106	246782±152	3063±62
8	16/11/2021	312±8	67±17	-134±20	2.71±0.09	22±11	27±19	592351±129	245621±139	3019±60
9	19/06/2021	326±8	69±13	-86±30	2.24±0.10	15±9	32±21	593171±166	243462±167	3005±62
10	11/02/2018	292±9	62±17	-121±31	1.97±0.08	17±10	32±21	590539±159	247512±160	3050±63

Table 5.1: Mean and standard deviations of the marginal probability densities for both the centroid and the decomposed moment tensors.

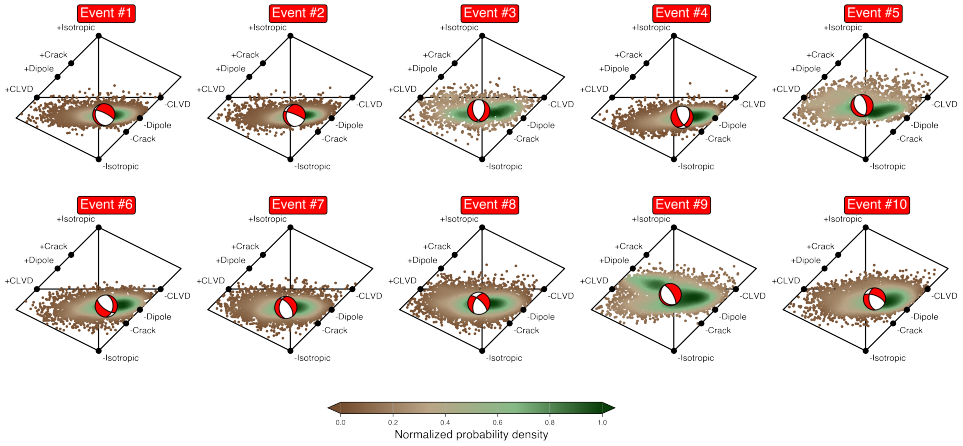

Figure 5.3: Hudson plots generated by decomposing the MT solutions (i.e., the individual samples of our posterior probability density). The samples are additionally color-coded based on the 'normalized' probability density. The beachball represents the DC component of the mean solution, including its position on the Hudson plot.

may cause more damage and ground shaking compared to areas in the opposite direction. Local geological conditions may further modify the particle acceleration, amplifying or damping the shaking experienced at the surface (Garini et al., 2017). In other words, two areas equidistant from the earthquake's epicenter might experience different levels of shaking based on the path of the rupture and the local geological conditions. Therefore, understanding the radiation pattern resulting from rupture propagation provides crucial insights for earthquake preparedness and risk mitigation, aiding in the development of strategies to protect communities in earthquake-prone areas. In our case, these are the citizens of the province of Groningen. In this study, we analyze the rupture propagation of the ten events investigated in the previous section. More precisely, we estimate the horizontal direction in which those ruptures propagate, the horizontal components of the event's rupture lengths, and their rupture velocity. In the next subsection, we first describe the theory underlying our analysis, after which we explain the selection of the *child* events that are needed to estimate rupture direction and distance. This is followed by the results of our analysis.

5.3.1. THEORY

The seismic energy of an earthquake rupture can propagate along a fault line. Depending on the rupture parameters, such as propagation speed, size, and preferential rupture direction, different seismogram characteristics can be observed when measuring in different azimuthal directions. These differences are often referred to as *rupture directivity* effect (Cesca et al., 2011). One way to evaluate this effect is by analyzing the amplitude and frequency spectra of the recorded seismograms. Assuming that the rupture is propagating unilaterally in a homogeneous medium, the amplitude of seismograms recorded at the azimuth parallel to the rupture direction will be higher and will consist of more

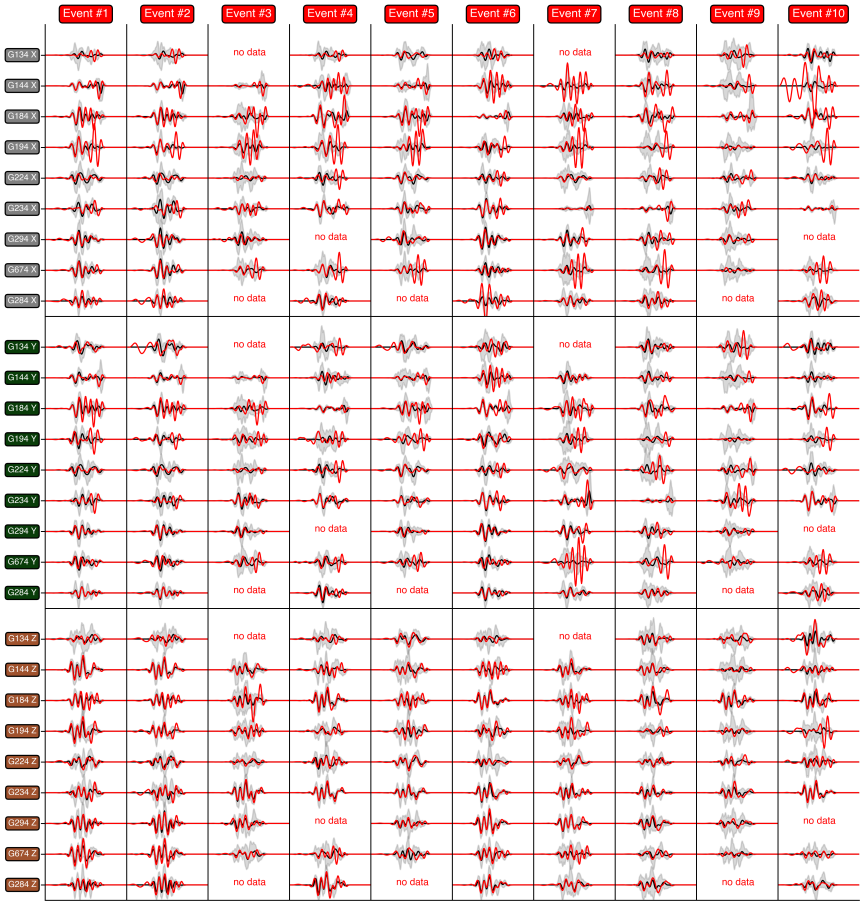

Figure 5.4: Comparison between observed data (red) and modeled data (black) generated from the mean CMT of the posterior for each event. The gray shaded area marks the maximum and minimum bound of a set of 100 (forward) modeled wavefields generated from 100 samples drawn from the posterior probability density. The horizontal axis corresponds to 4s of seismograms.

high-frequency signals. Conversely, seismograms recorded on the opposite side will display lower amplitudes with a predominance of lower-frequency signals. These differences in amplitude and frequency lead to varying levels of ground motion in different areas, which in turn affects the risk posed to buildings and infrastructure differently.

In a more complex scenario, when the medium is heterogeneous and/or the rupture is propagating bilaterally, resolving rupture directivity will be more intricate. One way to address this complexity is by extracting the source time function of a rupture. The observed seismogram is effectively the temporal convolution of a complex Green's function and a source time function (STF). While the Green's function captures the Earth's intrinsic properties, the STF represents the excited signal. In the case of a small rup-

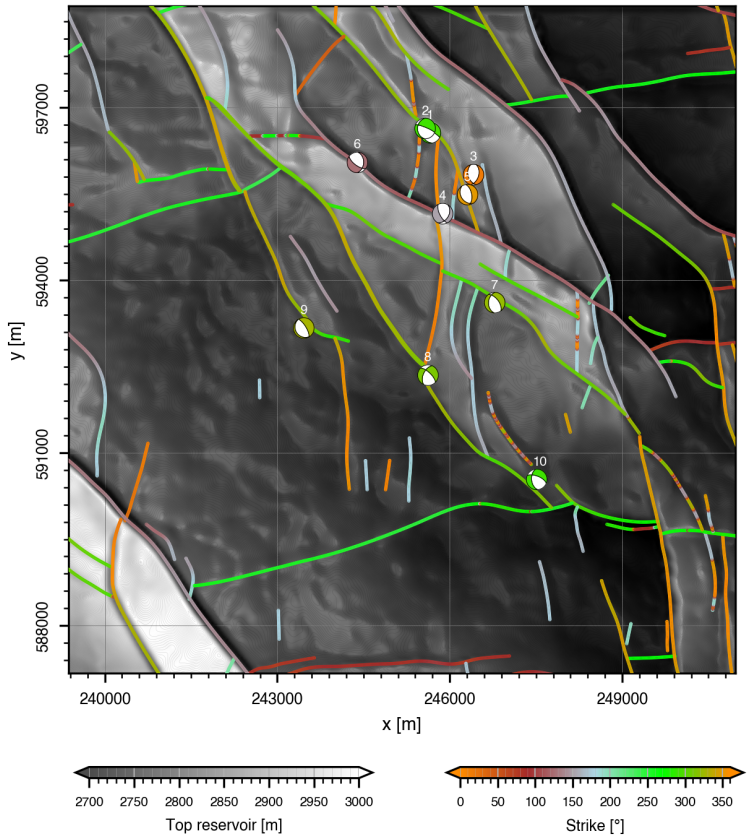

Figure 5.5: Preferred strike, dip, and rake solutions given the mean of CMT solutions for each event.

ture, a simple point source assumption can be used to model the observed displacement (Equation 5.1), but in the case of a propagating rupture, the recorded displacement can be seen as the result of summing displacements due to multiple point sources. We illustrate these multiple point sources in Figure 5.6 (in top view). The solid light green line represents the fault line, and the rupture is propagating in the direction towards receiver B. As depicted, each point source represented by the red ball will result in an individual observed seismogram, which is represented by the red traces. Because the rupture propagates at a specific speed and in a specific direction, depending on which receiver we are looking at, there will be differences in the arrival times. At receiver B, the wavefront of the propagated waves will be compressed; hence, the observed seismograms (the black trace obtained by summing over the red traces) at B will be close to being in phase with the individual black traces in comparison to when looking at receiver A. This results in observed seismograms that have relatively higher amplitude and consist of more high-frequency signals at receiver B in comparison to seismograms at receiver A. Deconvolving the black traces with a single point source results in a relative source time function

(RSTF). In that case, the RSTF in receiver A will likely be broader with smaller amplitude compared to the RSTF in receiver B as represented by the green traces in Figure 5.6, a phenomenon referred to as *Doppler broadening* [1] (Oates et al., 2024). The method of deconvolving observed seismograms from the large earthquakes with the signals from the smaller earthquakes (which represent the point source) to obtain the RSTF is often referred to as the empirical Green's function (EGF) method (Hartzell, 1978; Courboulex et al., 1996; Hutchings and Viegas, 2012).

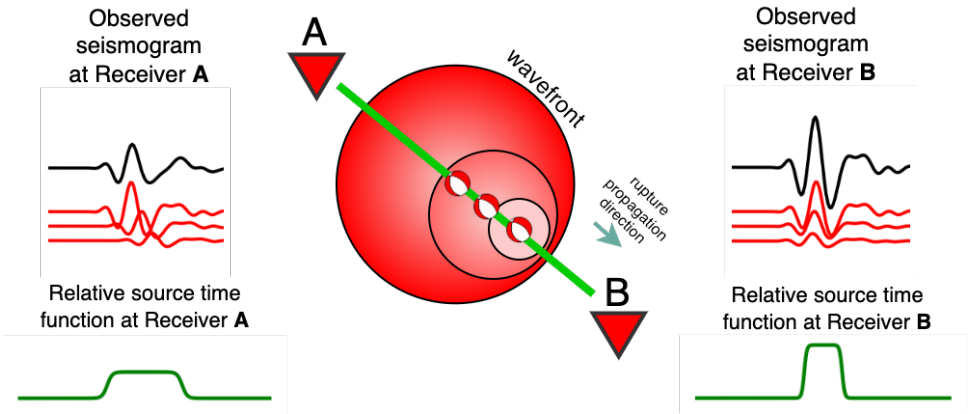

Figure 5.6: (Modified after Poiata et al. (2017)) Illustration of observed seismograms at two opposite receivers, A and B, given the source is a propagating rupture towards receiver B (multiple point sources). The point sources are represented by the red beachball, which is due to the movement of fault represented by the green line. Individual seismogram from each point source is represented by the red trace, and the observed seismogram (black trace) is obtained by summing the red traces. Deconvolving the black traces with the same single red trace results in a relative source time function represented by the green traces.

In application, similar to what we illustrated previously, RSTF is achieved by deconvolving seismograms from two events that are estimated to be co-located and governed by the same mechanism (i.e., fault movements). An important condition is that the latter seismogram should be due to a 'relatively' small event in terms of event magnitude. This will ensure that it exhibits less to no directivity compared to the larger event, which then closely resembles a point source relative to the larger event. Mathematically, the deconvolution of the seismic recording of the stronger event with the seismogram of the weaker event can be written as:

$$u_i^{\mathrm{P}} * (u_i^{\mathrm{C}})^{-1} = \sum_{j=1}^{3}\sum_{k=1}^{3}(M_{jk}^{\mathrm{P}} \cdot S^{\mathrm{P}}) * G_{ij,k} * \left(\sum_{j=1}^{3}\sum_{k=1}^{3}(M_{jk}^{\mathrm{C}} \cdot S^{\mathrm{C}}) * G_{ij,k}\right)^{-1} \tag{5.3}$$

$$= \Delta M \cdot S^{\mathrm{P}} * (S^{\mathrm{C}})^{-1} \tag{5.4}$$

$$= S^{\mathrm{R}} \tag{5.5}$$

with u_i^{P} being the seismogram of the larger event (often also referred to as the *parent*

[1] As a caveat, Doppler broadening mentioned here is a different phenomenon than the Doppler effect where the frequency shift is due to a moving source.

event; hence the P), and u_i^C being the recording of the *child* event. The STF of the Parent event is denoted by the S^P and the STF of the child event by the S^C. Note that this implies that the moment tensors M of both events are not time-dependent because this time dependence is now captured by S^P and S^C. Furthermore, ΔM represents the magnitude ratio between both events.

By assuming that the child event is an impulsive point source (as in the previous section), the relative source time function (RSTF) S^R is obtained through deconvolution of the parent seismogram with the child seismogram. The assumption that the child event is an impulsive point source means that S^C is effectively a delta function, which in turn implies that S^R is a scaled version of S^P. In our application, we deconvolve the seismograms in the frequency domain using additive noise to stabilize the deconvolution (Clayton and Wiggins, 1976). More specifically, we deconvolve the horizontal components (north-south and east-west recordings) of the S-wave in the seismograms. The S-wave is preferred over the P-wave because it exhibits a higher signal-to-noise ratio, particularly on the horizontal components, rendering the deconvolution results more stable. We then use the duration of the RSTF Δt (the width of the pulse) at different seismic stations to recover rupture characteristics. Specifically, we parameterize the duration of the RSTF (measured between zero crossings) using rupture propagation velocity c_R, shear wave velocity v_S, rupture length L, rise time t_r, and azimuth φ (clockwise with respect to strike and being 0 in the direction of propagation) as follows (Schubert, 2015; Ben-Menahem and Singh, 2012):

$$\Delta t(\varphi) = t_r + \frac{L}{c_R} - \frac{L}{v_S}\cos(\varphi). \tag{5.6}$$

Equation 5.6 is valid for unilateral rupture, that is, the situation when a rupture propagates (predominantly) in a single direction. Other variations would be asymmetrical and pure bilateral rupture. In that case, Equation 5.6 can be reformulated as (Oates et al., 2024)

$$\Delta t(\varphi) = \mathrm{Max}\left[t_r + \frac{L_1}{c_R} - \left(\frac{L_1}{v_S}\right)\cos(\varphi), t_r + \frac{L_2}{c_R} + \left(\frac{L_2}{v_S}\right)\cos(\varphi)\right], \tag{5.7}$$

with L_1 and L_2 now being fault segments where an earthquake propagates in the direction in which L_1 and L_2 are pointing; ideally, strike and 180° + strike or the other way around.

It should be noted that both Equations 5.6 and 5.7 are, kinematically, contingent on the assumption that the earthquake nucleates and slips in the same horizontal plane as the receiver. In reality, the receivers, of course (are close) to the surface and the earthquake at depth. This may seem like a violation of the mentioned assumption, but this violation is minimal because (i) we only use stations that are at least 1 km away from the events and (ii) the propagation close to the rupture is sub-horizontal for propagation to stations at such distances (Kraaijpoel and Dost, 2013). As for t_r, since induced earthquakes usually generate higher frequency seismograms, which corresponds to short rise times, t_r and should often be smaller than RSTF, this parameter is often assumed to be negligible even in events larger than induced earthquakes (Oates et al., 2024; Park and Ishii, 2015; Paulssen et al., 2022). Furthermore, φ can be reformulated in terms of the azimuth of the recording receivers with respect to the north. In the following sections,

we estimate L and c_R for both rupture models by minimizing the sum of the squares of the difference between the observed $\Delta t(\varphi)$ and the $\Delta t(\varphi)$ predicted by Equations 5.6 and 5.7. The goodness of fit is furthermore computed to indicate which model (unilateral or bilateral) better explains the rupturing process for each event.

5.3.2. SELECTING CHILD EVENTS

For each parent event, the successful application of Equations 5.6 and 5.7 is contingent on the selection of appropriate child events. For this, we determine, based on the KNMI catalog, the 10 events closest to the location of each parent event. We depict these candidate child events in Figure 5.7 by means of solid orange circles. Since our main events are all above $M_L = 2$, the candidate child events are required to have magnitudes below 2. The usual approach for selecting a single child event per parent event involves estimating the CMTs of the candidate child events (Paulssen et al., 2022; López-Comino et al., 2012). The child event that has a CMT solution that is most similar to the parent event is then selected. The low signal-to-noise ratio (SNR) of the seismograms of the potential child events, however, makes it difficult to follow this procedure, in particular, because P-waves are used in the CMT inversion (whereas the S-wave amplitudes are often higher, resulting in higher SNRs). Therefore, in this study, we simply apply, for each parent event, the EGF method to all candidate child events and examine the results. These results are shown in Appendix A.

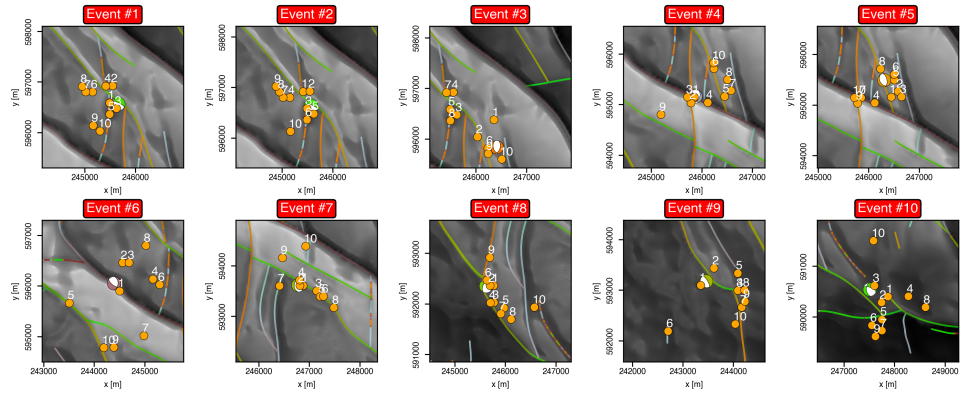

Figure 5.7: Epicenter of each main event (beachball) with the ten closest candidate child events (based on their horizontal distance).

After applying the EGF method to all potential child events (Appendix A), we proceed by examining the results in terms of the azimuthal variation of the amplitudes. Specifically, we aim for deconvolution results that yield relatively high amplitudes around time delay = 0 and/or have a distinct sinusoidal pattern. The high amplitude not only indicates that the child event nucleated at or close to the location of the parent event but also that it is driven by a mechanism similar to the parent event. Furthermore, the sinusoidal pattern indicates that the relative orientation of a child event is located along strikes but at some lateral distance away from the centroid of its main event. That is,

we expect the maximum and minimum of the sinusoidal pattern to coincide with the “strike" derived from both the CMT inversion and strike of the available (known) fault geometry, as illustrated in Figure 5.8. In the left panel of this figure, assuming that the strike direction is parallel to receiver A, the child event (the orange beachball) is located northwest of the parent event and along the strike. Consequently, at receiver A, the travel time of the signal originating from the parent events (the red beachballs) is larger compared to the child event and vice versa for receiver B. Application of the EGF method, therefore, yields a positive delay time in the direction of station A. If the child event is located southeast of the parent event, we observe an opposite sinusoidal pattern, which is shown in the right panel of Figure 5.8. Finally, if the child event and the parent event are co-located, we expect to see the pattern shown in the middle panel of Figure 5.8 (i.e., no to minimum sinusoidal pattern is observed). It should be understood that the patterns depicted in Figure 5.8 are ideal cases in the sense that parent and child events are located at the same depth level as the event and in the absence of lateral heterogeneity in the subsurface. In reality, these patterns will often (slightly) deviate from this ideal situation. What is more, the width/temporal duration of the extracted RSTF for receiver B will likely be shorter in comparison to what is extracted at receiver A in all three scenarios due to Doppler broadening explained earlier.

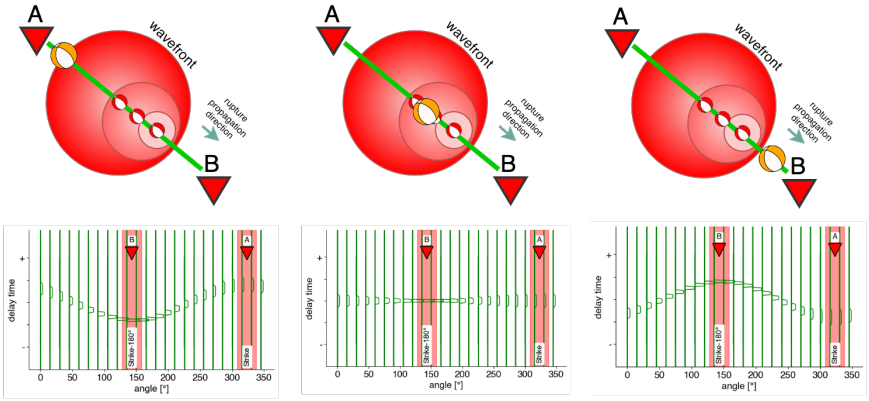

Figure 5.8: (Illustration of how the relative along-strike location of the child event affects the pattern of the EGFs as a function of azimuth. (Left) Illustrates the pattern if the child event is located closer to station (depicted as being station A here) in the along-strike direction than the parent event; the EGFs exhibit a sinusoidal pattern where station A has a positive delay time and station B a negative delay time. (Right) Illustrates the opposite case. (Middle) Illustrates the EGFs if the small and main events are co-located, resulting in no sinusoidal pattern. Given that the rupture propagates towards receiver B, the duration of the extracted RSTF (the width of the green seismograms) recorded at receiver B is shorter than at receiver A.

The selection process of a child event is exemplified for parent event #7 in Figure 5.9, using only four candidate child events. The mean strike obtained from our CMT inversion for this event is 323°. To determine which receivers to use when applying the EGF method to all parent events, including this one, we draw a 35 km diameter circle from the event’s epicenter and use active receivers within that radius, as represented by the blue inverted triangles in Figure 5.9. However, some receivers, such as receiver G134, were

not active during event #7. In a zoomed view over the epicenter of event #7 (see red box in the left corner), we show that candidate child events #1,#2, and #4 are clustered near the mean centroid of the parent event, while candidate child event #3 is located approximately 100 meters southeast, in the opposite direction of the strike (see also Figure 5.7). For the three clustered events, the observed pattern is similar to that in the left panel of Figure 5.8, and the opposite is true for the smaller event #3. Considering the assumptions underlying Figure 5.8 and the EGF method, the three clustered candidate child events are most likely located laterally northwest and along the strike. However, based on the KNMI earthquake catalog used to plot the child coordinates, they are mapped slightly in the opposite direction. This discrepancy could be due to several factors: the parent event might actually be located further southeast than indicated (note that the presented CMT solution and its location represent the posterior mean; standard deviations are provided in Table 5.1), the epicenters of child events #1, #2, and #3 might not have been accurately located by the KNMI (a reasonable assumption considering that a 1D velocity model was used), or a combination of both factors.

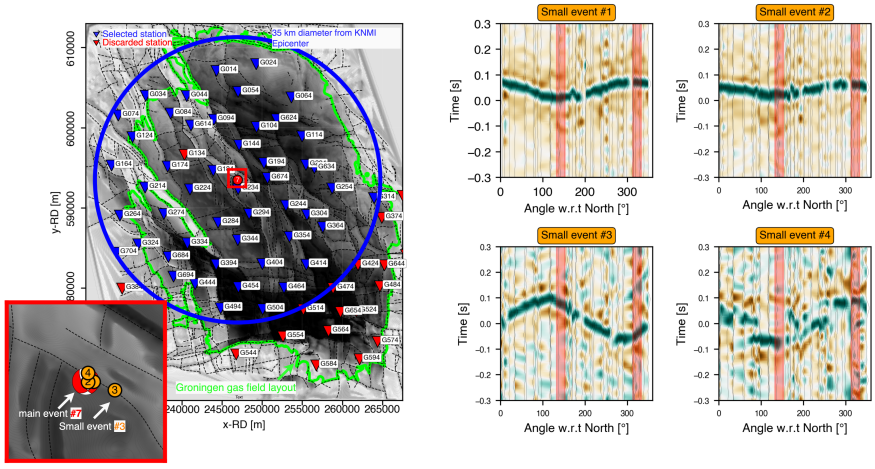

Figure 5.9: Application of the EGF method to parent event #7 and four potential child events. The left panel shows the receivers selected to compute the EGFs. At the bottom left is the zoom over the area where the main event epicenter is located, as well as the position of the four potential child events (estimated by the KNMI). The CMT-derived mean strike of the parent event is 323°; hence, its opposite is 143°. The four panels on the right side depict the EGFs for the four different potential child events, with the red vertical bands representing the strike direction and its opposite with ±10°.

After visual inspection, we use, for parent event #7, the deconvolutions with child event #2 for further analysis. The reason is that these give the most consistent high amplitudes, in particular in comparison to the deconvolutions with child event #4. In addition, it has the *flattest* sinusoidal pattern, which means that child event #2 is most likely closest to the parent event. A similar visual inspection has been performed for all 10 (parent) events considered in this study. That is, from each of the ten sets of stabilized deconvolutions, each associated with a different candidate child event, one is chosen for further analysis (indicated by means of the blue frames in Figure 5.12). The ten selected

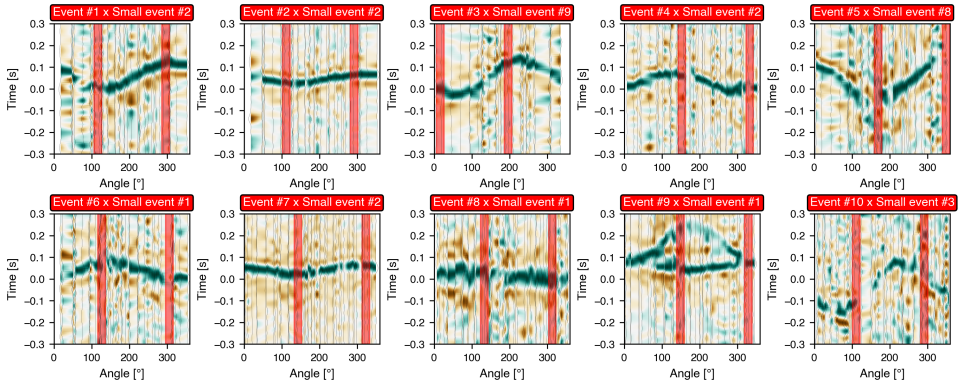

Figure 5.10: Selected sets of stabilized parent-child deconvolutions (indicated by blue frames in Figure 5.12). The vertical red bands correspond to the strike direction ($\pm 10^{\circ}$) estimated from the posterior mean CMT solutions (see Table 5.1).

sets of stabilized deconvolutions are shown in Figure 5.10. The vertical red bands again indicate the strike directions and its opposite (either + or - 180°) estimated from the posterior mean CMT solutions (see Table 5.1). In most cases, the maximum and minimum of the sinusoidal pattern coincide with the estimated strike. In the following Section, we use the azimuthal variation of the RSTFs, i.e, the measured $\Delta t(\varphi)$, to invert for rupture length, rupture velocity, and rupture direction, considering both unilateral and bilateral rupture scenarios (Equations 5.6 and 5.7, respectively).

5.3.3. RESULTS

Given the selected child events, we proceed by examining the RSTFs for each main event. Specifically, we focus on identifying the zero crossings that limit the dominant pulse. For each selected parent-child pair, the dominant pulses, filled with a green hue, are displayed in Figure 5.11. The solid orange circles in both rectangular and polar frames represent the RSTF duration as estimated based on the zero crossings. We then use the estimated RSTF durations $\Delta t(\varphi)$ to fit Equations 5.6 and 5.7. While doing so, the shear wave velocity (v_S) is set to 2453 m/s, similar to the study by Oates et al. (2024), which represents the shear wave velocity of the Groningen reservoir. The best fittings $\Delta t(\varphi)$ for a unilateral rupture are depicted as solid blue lines in the rectangular and polar frames in Figure 5.11 and as red solid lines for a bilateral rupture. Estimated rupture parameters are colored accordingly. The (accordingly colored) straight solid lines in the polar plots indicate the estimated rupture direction, with blue being the direction of unilateral rupture and red being the direction of bilateral rupture.

We compare the rupture directions estimated for unilateral and bilateral scenarios with the fault geometries and (posterior mean) CMT estimates in Figure 5.12. The rupture direction for the unilateral case is depicted by means of blue wedges in Figure 5.12. For the bilateral case, we depict the main rupture direction using a red wedge and its opposite as a green wedge. Note that the size of the wedges is not on the same scale as the map.

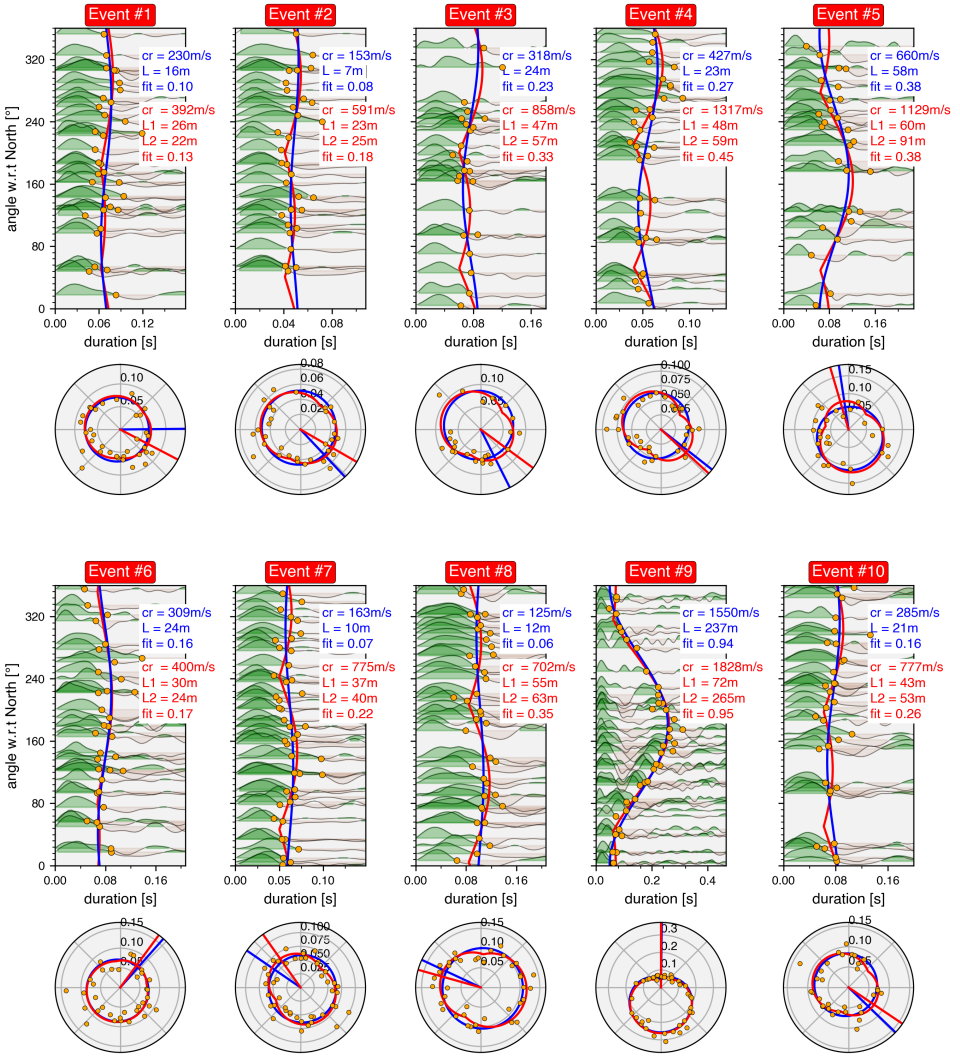

Figure 5.11: Result of the rupture analysis of each parent event. The solid orange dots in the rectangular panels delimit the zero crossing, which provides the apparent duration for each relative source time function. Best fitting rupture models, as described by Equations 5.6 and 5.7 are depicted as blue and red solid curves, respectively. Associated parameters are listed using the same color; the rupture speed (cr), rupture length (L), and their goodness of fit. The polar plot shows the RSTF durations and the models using polar coordinates, with the straight-colored solid lines corresponding to the rupture direction associated with the best-fitting rupture models.

5.4. DISCUSSION

In this section, we discuss the implications and significance of the estimated CMTs and rupture parameters. Our probabilistic CMT inversion results provide useful insights into

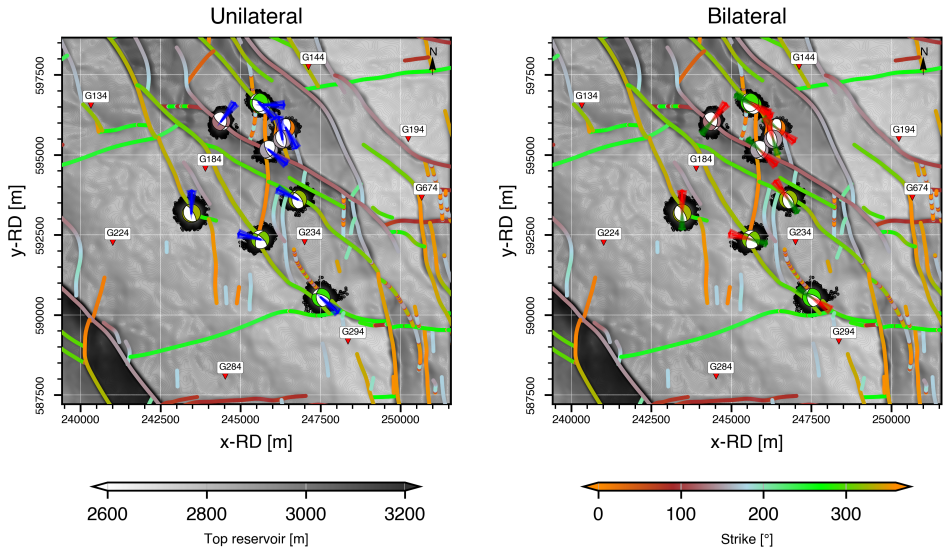

Figure 5.12: Comparison between rupture directions estimated assuming unilateral and bilateral rupture scenarios, the available fault map, and the (mean) CMT solution for each event. The left frame depicts the estimated unilateral rupture direction using blue wedges ($\pm 10^{\circ}$). The right frame depicts the estimated bilateral rupture directions, where the red wedges represent the rupture direction of the larger fault segment and the green wedge the rupture direction of the smaller fault segment (i.e., in the opposite direction), with both triangles being $\pm 10^{\circ}$ of their corresponding directions. Both the CMT-derived beachballs and the faults are color-coded based on their strike (according to the color bar at the bottom right).

the centroid and source mechanisms associated with the studied events. Notably, the centroid is consistent with the available fault map for all parent events. As for the source mechanisms, we find that the events are predominantly explained by DC mechanisms (normal faulting) but with the consistent presence of negative non-DC mechanisms (ISO and CLVD). This is in line with the study by Dost et al. (2020). In general, negative ISO components are associated with a loss of volume in a medium. In our case, this loss is most probably due to gas extraction in the area. As for the negative CLVD components, it represent the combination of compression in the vertical axis and dilatation in the lateral axes (Vavryčuk, 2014, 2015). Examples of induced events that exhibit a significant amount of negative CLVD mechanisms are often found in mining activities driven by collapse events (Rudziński et al., 2016; Šílenỳ and Milev, 2008; Caputa et al., 2021). And naturally, it can be found in volcanic-related earthquakes (Shuler et al., 2013; Fichtner and Tkalčić, 2010). In many cases, however, the existence of CLVD components could also be due to the presence of noise in the seismograms and/or the poorly known subsurface models, such as the presence of high anisotropy (Mustać and Tkalčić, 2016; Stierle et al., 2014). Therefore, in most cases, when dealing with earthquakes, seismologists often assume that the CLVD components are minimum or close to zero, particularly when the prior knowledge suggests it.

An ideal natural geological setting that results in negative CLVDs is a ring fault. In a

volcanic environment, for example, this complies with the form of a caldera (Sandanbata et al., 2021). Other settings that could result in CLVD-type source mechanisms are graben-like structures and complex fault systems (Irmak, 2013). The subsurface of Groningen is known for its complex fault systems (De Jager and Visser, 2017). Assuming that the CLVD mechanisms are unfavorable solutions, one can suppress their presence by putting a strong prior for the DC mechanisms, which are distributed within the center of the Hudson plot (Cesca et al., 2013). Nevertheless, in our application, we put a uniform prior on the earthquake mechanisms to encode the state of ignorance. Notwithstanding all this, the CMT results indicate that the earthquakes are predominantly governed by double-coupled mechanisms.

The estimated widths of the RTSFs in this study constrain the rupture directions of the source events. We find that the goodness of fit, in general, is relatively poor (see 'fit' in Figure 5.11) except for some events. These low values could be due to most (parent) events having normal faulting mechanisms with only relatively small rupture lengths. If this is the case, then there is simply little directivity. Therefore, from a model perspective, rather than having a circle with either one lobe for unilateral cases or two lobes for bilateral cases Cesca et al. (2011), a circle centered at 0 would be appropriate. Nevertheless, looking at the goodness of fit, we find that in some events, shifting to the bilateral rupture model improves the fitness value significantly. Hence, we might argue that the bilateral rupture assumption slightly better explains the data (RSTF duration) except for event #5 (where the goodness of fit has similar values for both models). This is potentially because a bilateral rupture model with equivalent rupture length better mimics the little to no directivity effect, which is a circle centered at 0 (although having two lobes). This then explains events with a jump in fit value have similar values for L_1 and L_2. However, in general, the increase we found does not reach the same fitness where the directivity is clearly visible, for example, in event #9. Therefore, further analysis and perhaps a more complex model are required to capture the nuance of the data. As a caveat, we did not include the noise analysis present in the dataset when performing directivity analysis. Hence, the better fit in the bilateral rupture model could be due to the overfitting of the available data points.

In terms of rupture direction, assuming bilateral rupture for most events does not improve its alignment with the available fault map except for events #1, #2, and #3. However, it should be understood that the value on the map is the (global) trend of a fault segment. In more detail, the geometry might change with depth and lateral position, as shown in the study (Kortekaas and Jaarsma, 2017). As for the rupture speed, we found that our results varied significantly between the two rupture models. That is, the bilateral rupture assumption gives consistently higher values. Sometimes even as much as twice the rupture speed compared to the rupture speed estimated for a unilateral rupture model. Typically, the rupture speed varies between 0.7 and 0.9 times the shear wave velocity (Ammon et al., 2020), which are categorized as subshear. In this study, however, we find that the values are often smaller (with the exception of event #9). At the same time, a study by Weng and Ampuero (2020) suggests that rupture velocities might be smaller than those typical values in case the rake angle is heading towards -90° (i.e., pure normal faulting).

A number of measures and/or additions could potentially improve the presented

analysis. For the CMT inversion, for example, including the direct S-wave could be beneficial. This is because its amplitude is significantly stronger than the P-wave, resulting in a better signal-to-noise ratio and, therefore, sometimes leading to more accurate results (Xu et al., 2021). A significant challenge while including the S-wave, however, involves the separation of the direct wave and (slightly) later arriving converted waves and/or multiply reflected phases. Often, therefore, the horizontally polarized S-wave (or SH-wave) is used (Spetzler et al., 2024). Another challenge involves the solution of the forward problem (Equation 5.1), in particular in the case of numerical modeling rather than analytical solutions. The necessity to generate accurate (direct) S-waves requires the numerical model to have a denser mesh size. This is exacerbated in areas with very low-velocity layers, such as the shallow part of Groningen's subsurface. Hence, including the direct S-waves in the probabilistic inversion requires extensive computational power. Another potential improvement would be the inclusion of anisotropy in the moment tensor inversion. This is because the inclusion of anisotropy affects the results of the CMT inversion to a certain degree, as shown by Rößler et al. (2007); Vavryčuk (2005). As for the directivity analysis, one beneficial addition would be assessing the obtained directivity pattern by means of forward modeling. This can be optimally done by performing geomechanical simulations given Groningen subsurface models. These simulations can then be constrained by the source parameters obtained from CMT inversions (Ruan et al., 2022).

5.5. CONCLUSION

We performed a comprehensive analysis of ten of the largest induced seismic events in the Groningen reservoir to date. Specifically, we ran both a probabilistic CMT inversion and conducted a rupture directivity analysis to gain insights into source mechanisms and rupture characteristics of the selected events. Based on our analyses, we conclude the following: (i) In general, our CMT inversion reveals that the selected induced earthquakes can chiefly be explained by double-coupled mechanisms. In terms of faulting, normal faulting appears to be dominant. This is to be expected, given the differential compaction along the existing faults intersecting the gas reservoir. Furthermore, the isotropic and CLVD mechanisms both show negative values. These are explained by a loss of volume synchronous with normal faulting and are also consistent with the subsidence in the area. (ii) The rupture directivity pattern obtained using the EGF method generally shows low goodness of fit while testing using both unilateral and bilateral rupture scenarios. However, there is a small increase in goodness of fit when switching from unilateral to bilateral rupture. One reason could be the bilateral rupture can better represent the situation where the directivity effect of a rupture is minimal. However, the increase of fit in the scenario does not pose the same magnitude as in one event (event #9) where the directivity is highly visible. Therefore, we suggest a more detailed look at the validity of the rupture model. This includes, for example, noise analysis, as the noise can highly affect the extracted RSTF data points, which in turn can either underfit or overfit the data. Another improvement can also be made by comparing the obtained rupture orientation with a more detailed fault model. In addition, improvement of the EGF analysis could be achieved further by coupling the EGF method with geomechanical simulations. This way, we can get into more detail on the intrinsic nature of the rupture

process taking place in the Groningen subsurface.

APPENDIX: SELECTION OF CHILD EVENTS

Below, we show the result of applying the EGF method (i.e., the stabilized parent-child deconvolution) to each parent event and all candidate child events. Prior to applying this method, we taper the horizontal components of the observed seismograms to isolate the S-wave using a cosine taper with a taper length of 4s. We add one percent of additive noise when applying the deconvolution to stabilize the results.

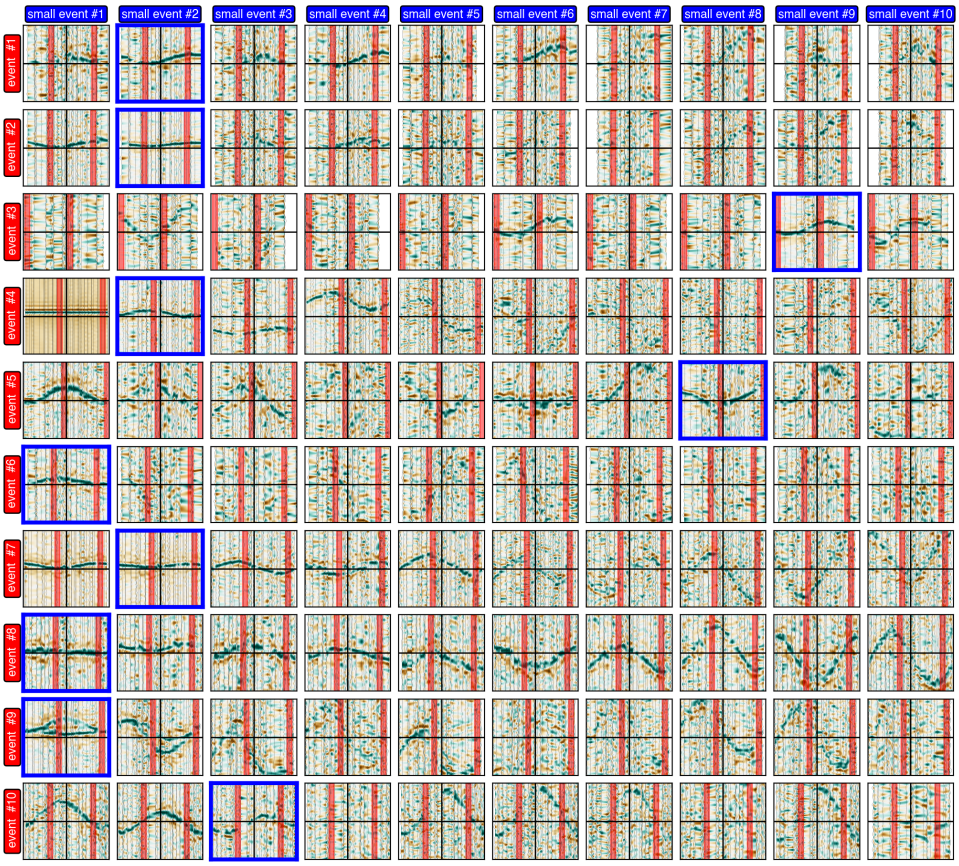

Figure 5.13: Application of the EGF method for each parent event to all candidate child events. The child event associated with the deconvolved traces in the blue frame is selected for further analysis and estimation of the rupture parameters. The horizontal and vertical black lines represent delay times of 0s and an azimuth of 0 degrees with respect to the north, respectively. The limit along the vertical and horizontal axes are ±0.5 s and± 180°, respectively (see the annotation along the horizontal and vertical axes of Figure 5.10). The vertical red bands indicate the strike directions obtained from the CMT inversion shown in Table 5.1 (with ±10°).

REFERENCES

Abercrombie, R. E. (2015). Investigating uncertainties in empirical Green's function analysis of earthquake source parameters. *Journal of Geophysical Research: Solid Earth*, 120(6):4263–4277.

Aki, K. and Richards, P. G. (2002). *Quantitative Seismology*. University Science Books, 2^{nd} edition.

Ammon, C. J., Velasco, A. A., Lay, T., and Wallace, T. C. (2020). *Foundations of modern global seismology*. Academic Press.

Ben-Menahem, A. and Singh, S. J. (2012). *Seismic waves and sources*. Springer Science & Business Media.

Bourne, S. and Oates, S. (2017). Induced seismicity within the Groningen gas field (v1.0.0) (available at https://zenodo.org/records/1035226). *Zenodo*.

Buijze, L., van den Bogert, P. A., Wassing, B. B., and Orlic, B. (2019). Nucleation and arrest of dynamic rupture induced by reservoir depletion. *Journal of Geophysical Research: Solid Earth*, 124(4):3620–3645.

Caputa, A., Rudziński, Ł., and Cesca, S. (2021). How to assess the moment tensor inversion resolution for mining induced seismicity: A case study for the Rudna mine, Poland. *Frontiers in Earth Science*, 9:671207.

Cesca, S., Heimann, S., and Dahm, T. (2011). Rapid directivity detection by azimuthal amplitude spectra inversion. *Journal of seismology*, 15:147–164.

Cesca, S., Rohr, A., and Dahm, T. (2013). Discrimination of induced seismicity by full moment tensor inversion and decomposition. *Journal of seismology*, 17:147–163.

Clayton, R. W. and Wiggins, R. A. (1976). Source shape estimation and deconvolution of teleseismic bodywaves. *Geophysical Journal International*, 47(1):151–177.

Courboulex, F., Dujardin, A., Vallée, M., Delouis, B., Sira, C., Deschamps, A., Honoré, L., and Thouvenot, F. (2013). High-frequency directivity effect for an Mw 4.1 earthquake, widely felt by the population in southeastern France. *Bulletin of the Seismological Society of America*, 103(6):3347–3353.

Courboulex, F., Virieux, J., Deschamps, A., Gibert, D., and Zollo, A. (1996). Source investigation of a small event using empirical Green's functions and simulated annealing. *Geophysical Journal International*, 125(3):768–780.

De Jager, J. and Visser, C. (2017). Geology of the Groningen field–an overview. *Netherlands Journal of Geosciences*, 96(5):s3–s15.

Dost, B. and Kraaijpoel, D. (2013). The August 16, 2012 earthquake near Huizinge (Groningen)(KNMI Scientific report available at https://www.eumonitor.nl/9353000/1/j4nvgs5kjg27kof_j9vvik7m1c3gyxp/vj6rovcbvxyl/f=/blg205498.pdf). *KNMI Scientific report*.

Dost, B., Ruigrok, E., and Spetzler, J. (2017). Development of seismicity and probabilistic hazard assessment for the Groningen gas field. *Netherlands Journal of Geosciences*, 96(5):s235–s245.

Dost, B., van Stiphout, A., Kühn, D., Kortekaas, M., Ruigrok, E., and Heimann, S. (2020). Probabilistic moment tensor inversion for hydrocarbon-induced seismicity in the Groningen gas field, the Netherlands, part 2: Application. *Bulletin of the Seismological Society of America*, 110(5):2112–2123.

Duputel, Z., Rivera, L., Fukahata, Y., and Kanamori, H. (2011). Uncertainty estimations for seismic source inversions. *Geophysical Journal International*, 190:1243–1256.

Fichtner, A. and Simutė, S. (2018). Hamiltonian Monte Carlo inversion of seismic sources in complex media. *Journal of Geophysical Research: Solid Earth*, 123(4):2984–2999.

Fichtner, A. and Tkalčić, H. (2010). Insights into the kinematics of a volcanic caldera drop: Probabilistic finite-source inversion of the 1996 Bardarbunga, Iceland, earthquake. *Earth and Planetary Science Letters*, 297(3-4):607–615.

Garini, E., Gazetas, G., and Anastasopoulos, I. (2017). Evidence of significant forward rupture directivity aggravated by soil response in an Mw6 earthquake and the effects on monuments. *Earthquake Engineering & Structural Dynamics*, 46(13):2103–2120.

Hartzell, S. H. (1978). Earthquake aftershocks as Green's functions. *Geophysical Research Letters*, 5(1):1–4.

Hutchings, L. and Viegas, G. (2012). Application of empirical green's functions in earthquake source, wave propagation and strong ground motion studies. *Earthquake research and analysis-new frontiers in seismology*, pages 87–140.

Irmak, T. S. (2013). Focal mechanisms of small-moderate earthquakes in Denizli Graben (SW Turkey). *Earth, Planets and Space*, 65(9):943–955.

Kortekaas, M. and Jaarsma, B. (2017). Improved definition of faults in the Groningen field using seismic attributes. *Netherlands Journal of Geosciences*, 96(5):s71–s85.

Kraaijpoel, D. and Dost, B. (2013). Implications of salt-related propagation and mode conversion effects on the analysis of induced seismicity. *Journal of Seismology*, 17(1):95–107.

López-Comino, J.-Á., Mancilla, F. d. L., Morales, J., and Stich, D. (2012). Rupture directivity of the 2011, Mw 5.2 Lorca earthquake (Spain). *Geophysical Research Letters*, 39(3).

Lui, S. K., Helmberger, D., Yu, J., and Wei, S. (2016). Rapid assessment of earthquake source characteristics. *Bulletin of the Seismological Society of America*, 106(6):2490–2499.

Muntendam-Bos, A. G., Hoedeman, G., Polychronopoulou, K., Draganov, D., Weemstra, C., van der Zee, W., Bakker, R. R., and Roest, H. (2022). An overview of induced seismicity in the Netherlands. *Netherlands Journal of Geosciences*, 101:e1.

Mustać, M. and Tkalčić, H. (2016). Point source moment tensor inversion through a Bayesian hierarchical model. *Geophysical Journal International*, 204(1):311–323.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov chain Monte Carlo*, 2(11):2.

Ntinalexis, M., Bommer, J. J., Ruigrok, E., Edwards, B., Pinho, R., Dost, B., Correia, A. A., Uilenreef, J., Stafford, P. J., and van Elk, J. (2019). Ground-motion networks in the Groningen field: Usability and consistency of surface recordings. *Journal of Seismology*, 23(6):1233–1253.

Oates, S., Schmitz, J., Zurek, B., Piesold, T., and van Dedem, E. (2024). Empirical Green's function analysis of some induced earthquake pairs from the Groningen gas field. *Journal of Seismology*, pages 1–35.

Park, S. and Ishii, M. (2015). Inversion for rupture properties based upon 3-D directivity effect and application to deep earthquakes in the Sea of Okhotsk region. *Geophysical Journal International*, 203(2):1011–1025.

Paulssen, H., Micallef, T., Bouwman, D., Ruigrok, E., Herman, M., Fadel, I., van der Meijde, M., Kwadiba, M., Maritinkole, J., and Ntibinyane, O. (2022). Rifting of the Kalahari Craton through Botswana? New seismic evidence. *Journal of Geophysical Research: Solid Earth*, 127(4):e2021JB023524.

Poiata, N., Miyake, H., and Koketsu, K. (2017). Mechanisms for generation of near-fault ground motion pulses for dip-slip faulting. *Pure and Applied Geophysics*, 174:3521–3536.

Rößler, D., Krüger, F., and Rümpker, G. (2007). Retrieval of moment tensors due to dislocation point sources in anisotropic media using standard techniques. *Geophysical Journal International*, 169(1):136–148.

Ruan, J., Masfara, L. O. M., Ghose, R., and Mulder, W. (2022). Linking geomechanical simulation of induced seismicity to surface seismic observations: Simulated finite fault rupture to moment tensor inversion. In *SEG International Exposition and Annual Meeting*, page D011S081R003. SEG.

Rudziński, Ł., Cesca, S., and Lizurek, G. (2016). Complex rupture process of the 19 March 2013, Rudna Mine (Poland) induced seismic event and collapse in the light of local and regional moment tensor inversion. *Seismological Research Letters*, 87(2A):274–284.

Sandanbata, O., Kanamori, H., Rivera, L., Zhan, Z., Watada, S., and Satake, K. (2021). Moment tensors of ring-faulting at active volcanoes: Insights into vertical-CLVD earthquakes at the Sierra Negra caldera, Galápagos Islands. *Journal of Geophysical Research: Solid Earth*, 126(6):e2021JB021693.

Schubert, G. (2015). *Treatise on geophysics*. Elsevier.

Shuler, A., Ekström, G., and Nettles, M. (2013). Physical mechanisms for vertical-clvd earthquakes at active volcanoes. *Journal of Geophysical Research: Solid Earth*, 118(4):1569–1586.

Šílenỳ, J. and Milev, A. (2008). Source mechanism of mining induced seismic events—Resolution of double couple and non double couple models. *Tectonophysics*, 456(1-2):3–15.

Simutė, S., Boehm, C., Krischer, L., Gokhberg, A., Vallée, M., and Fichtner, A. (2023). Bayesian seismic source inversion with a 3-D Earth model of the Japanese Islands. *Journal of Geophysical Research: Solid Earth*, 128(1):e2022JB024231.

Smith, J. D., Avouac, J.-P., White, R. S., Copley, A., Gualandi, A., and Bourne, S. (2019). Reconciling the long-term relationship between reservoir pore pressure depletion and compaction in the Groningen region. *Journal of Geophysical Research: Solid Earth*, 124(6):6165–6178.

Smith, J. D., White, R. S., Avouac, J.-P., and Bourne, S. (2020). Probabilistic earthquake locations of induced seismicity in the Groningen region, the Netherlands. *Geophysical Journal International*, 222(1):507–516.

Spetzler, J. and Dost, B. (2017). Hypocentre estimation of induced earthquakes in Groningen. *Geophysical Journal International*, 209(1):453–465.

Spetzler, J., Ruigrok, E., and Bouwman, D. (2024). Hypocenter uncertainty analysis of induced and tectonic earthquakes in the Netherlands. *Journal of Seismology*.

Stierle, E., Vavryčuk, V., Šílenỳ, J., and Bohnhoff, M. (2014). Resolution of non-double-couple components in the seismic moment tensor using regional networks—I: a synthetic case study. *Geophysical Journal International*, 196(3):1869–1877.

van Thienen-Visser, K. and Breunese, J. (2015). Induced seismicity of the Groningen gas field: History and recent developments. *The Leading Edge*, 34(6):664–671.

Vavryčuk, V. (2005). Focal mechanisms in anisotropic media. *Geophysical Journal International*, 161(2):334–346.

Vavryčuk, V. (2014). Moment tensors: Decomposition and visualization. *Encyclopedia of earthquake engineering*, pages 1–16.

Vavryčuk, V. (2015). Moment tensor decompositions revisited. *Journal of Seismology*, 19:231–252.

Wang, X. and Zhan, Z. (2020). Moving from 1-D to 3-D velocity model: Automated waveform-based earthquake moment tensor inversion in the Los Angeles region. *Geophysical Journal International*, 220(1):218–234.

Weng, H. and Ampuero, J.-P. (2020). Continuum of earthquake rupture speeds enabled by oblique slip. *Nature Geoscience*, 13(12):817–821.

Willacy, C., van Dedem, E., Minisini, S., Li, J., Blokland, J. W., Das, I., and Droujinine, A. (2018). Application of full-waveform event location and moment-tensor inversion for Groningen induced seismicity. *The Leading Edge*, 37(2):92–99.

Xu, J., Zhang, W., Liang, X., Rong, J., and Li, J. (2021). Joint microseismic moment-tensor inversion and location using P-and S-wave diffraction stacking. *Geophysics*, 86(6):KS137–KS150.

Zhu, L. and Zhou, X. (2016). Seismic moment tensor inversion using 3D velocity model and its application to the 2013 Lushan earthquake sequence. *Physics and Chemistry of the Earth, Parts A/B/C*, 95:10–18.

6

CONCLUSIONS AND RECOMMENDATIONS

In the following subsections, we list the key conclusions of each chapter, starting with Chapter 2. In particular, we examine and link the conclusions with the research questions presented in Chapter 1. Following the conclusions, we present recommendations based on the challenges in the concluded chapters, including proposed outlooks for relevant future research.

6.1. CONCLUSIONS

In Chapter 2, a workflow backed by a variant of the Hamiltonian Monte Carlo algorithms aims to answer the first research question, which reads "*How can we modify existing efficient HMC probabilistic algorithms to estimate induced earthquake parameters without hampering the accuracy of the estimation?*". Our analysis concluded that the proposed workflow is capable of efficiently estimating the posterior probability of 10 parameters describing a synthetic earthquake source. However, because the event is a synthetic one, several caveats must be emphasized here. First, the modeled seismograms used to solve the forward problem in the workflow and the seismograms of the synthetic event are both generated using the same velocity model and code. This does give us better control in testing the workflow's efficiency while assessing its output against the true parameters of the synthetic event. Second, the study intentionally excludes an application to field data, as the goal is to present a stand-alone workflow that can be applied in any induced seismic setting. Applying the methodology to field recordings of induced seismic events requires numerous processing details that fall beyond the scope of this chapter. For example, the mismatch between the field seismograms and synthetic seismograms might come from the lack of accuracy in the subsurface velocity model, leading to higher uncertainty in the workflow's outputs.

Chapter 3 also attempts to answer the first research question by detailing the application of the inversion workflow in Chapter 2. However, this time, we increase the

complexity of the model and the process used to obtain synthetic observed recordings to make it more realistic. Particularly, we perform a geomechanical simulation in a 3D subsurface model and use the finite fault assumption in generating the recordings instead of the point source assumption (which is used in Chapter 2). At the same time, we use a more simplistic velocity model (1D) and the point source assumption to model the seismograms resulting from the solution of the forward problem in the HMC chain. This simplification reflects our inability to accurately represent true subsurface models and the limitations in our assumptions on generating seismograms for fitting actual field observations. One of our key conclusions is that the seismograms generated using different assumptions and models exhibit exceptional waveform fitting at the selected borehole receivers, highlighting the effectiveness of the workflow. Even when synthetic field recordings are corrupted with noise, the accuracy of the fitting process remains stable. This robustness is further demonstrated by the minimal differences in estimated earthquake parameters between noisy and noise-free synthetic recordings and by the close alignment of these parameters (e.g., centroid locations) with those observed in the geomechanical simulation.

The second and the third research questions highlighting "*the inclusion of geological information to the inversion*" and "*the implementation of the HMC workflow to field data*" are addressed in Chapter 4. Here, we use the workflow introduced in Chapter 2 to estimate the source characteristics of a 3.4 M_L induced earthquake in the Groningen gas field. This estimation is performed using two different sets of initial prior probabilities, which vary in how the epicenter locations are derived. The first set is based on a uniform grid, while the second set is informed by the location and orientation of subsurface faults. Our key finding is that the overall shape of the posterior probability distribution is largely similar between the two prior sets, except when limiting the seismogram input. In such cases where only the vertical component of the seismogram is used, the posterior distribution becomes slightly broader, indicating less precision. Moreover, using the fault-derived prior set resulted in significantly higher computational efficiency, reducing the computation time by a factor of four compared to the grid-based approach. When comparing our findings with existing research, we observe that our posterior estimates place the earthquake at a slightly shallower depth than the reservoir level, around 3 km deep, which is consistent with previous studies. However, the estimated fault orientation responsible for the earthquake aligns more closely with the available fault map than the results from prior research.

In Chapter 5, we aim to answer the last research question, which is to "*draw a link between the estimated parameters and the geological knowledge of the area.*" We achieve that by taking what we have learned from the previous chapters (e.g., workflow and geological prior knowledge) and applying it to 10 more earthquakes and including an additional analysis of the rupture directivity. Our analysis concludes that, in general, all earthquakes are mainly dominated by double-coupled mechanisms driven by normal faulting in the area. This is expected, given the extensive graben-like structures in the Groningen subsurface. In addition, we also find that the earthquakes have negative isotropic components, which most likely represent the loss of volume due to the extensive gas extraction from the reservoir. Further analysis of the directivity patterns reveals that the majority of the earthquakes exhibit minimal or negligible signs of directivity,

with the exception of one notable event. Overall, although the fit is relatively low, they generally correspond to the fault lines mapped in the region. This alignment suggests that the observed directivity, though weak, may still be influenced by the underlying geological structures, highlighting the importance of considering fault orientations in seismic assessments.

6.2. RECOMMENDATION

Building on the conclusions above, we highlight two recommendations that not only potentially address the challenges we face but could also be relevant to future earthquake characterization-related research. In summary, our recommendations are (i) to include S-wave on the field data inversion and (ii) to include the uncertainty of both the observed and synthetic seismograms.

To break it down, our first recommendation will plainly be beneficial for elevating the results obtained in Chapter 4 and 5, where we deal with field data inversion. In those chapters, we do not optimally utilize other wave segments except for the P-wave. It is unfortunate because other segments, such as S-waves and surface waves, have more dominant amplitudes in the seismograms. If included, those high amplitudes will ease the fitting process as they have better signal-to-noise ratios with potentially lower data uncertainty (assuming that the noise is uncorrelated and applied for the whole earthquake recordings), giving more robust results on the estimated parameters. In the context of Groningen earthquakes, Willacy et al. (2019) perform a detailed 3D waveform modeling, simulating an earthquake that nucleates from the Groningen reservoir due to a double couple source, which is the dominant earthquake mechanism in the area. One of the highlights they present is the difference in amplitudes of several wave segments generated by the simulated earthquake. Here, the weakest are the direct P-waves, followed by the guided P-waves and eventually the S-waves and guided S-waves. Despite the domination of the S-wave amplitudes, including it in the inversion of Groningen field data imposes two challenges: one is the computational challenge, and the other is the accuracy of the modeled S-waves, which stems from the accuracy of the S-wave velocity model itself.

On the first challenge, the value of Vp/Vs at the very shallow part of Groningen´s subsurface is relatively high (Kruiver et al., 2017a,b). This high ratio, which corresponds to slower S-wave velocities, results in a denser computational grid required for accurately modeling S-waves compared to P-waves at the same upper frequency, making simulations computationally expensive. Another challenge lies in the accuracy of the S-wave velocity model itself. While testing the workflow to fit the P-waves of field data, we also evaluate how the estimated parameters model S-waves and compare them to field data. Although the P-wave segments align well with the observations (as they fall within the measured window), the modeled S-waves often show discrepancies, sometimes even cycle-skipping in certain parts. These discrepancies typically stem from differences in arrival times rather than mismatches in the shape of the S-wave segments and are more pronounced at distant receivers. This inconsistency poses a challenge in our workflow, as it assumes the same origin time for all receivers and wave segments. Consequently, the arrival times for both P and S-waves are expected to be aligned. Our hypothesis is that the S-wave velocity models might lack sufficient accuracy since they are derived

mainly from the Vp/Vs relation extracted from a limited number of wells in comparison to the P-wave velocity, which is constrained by numerous available sonic information in the area (Romijn, 2017). One possibility to (partly) circumvent this inaccuracy of the S-wave velocity model would be to perform the inversion with each P and S-wave segment used individually, as was done by Dost et al. (2020).

As for the second recommendation, we argue that incorporating data uncertainty analysis is essential for improving the reliability of the estimated parameters in this thesis (and also future relevant studies). Here, the focus on data uncertainty is primarily on field data in the assumption of uncorrelated noise. In general, this uncertainty is typically linked to measurement errors, instrument noise, and environmental factors. While the inclusion of this type of data uncertainty is important, it does not fully capture the other sources of uncertainty that are related to the modeling of the synthetic seismograms. In that case, the uncertainty is influenced by inaccuracies in velocity models, numerical methods, and assumptions made during simulations. Typically, data uncertainty on field seismograms can be estimated directly from the recordings itself (e.g., Mustać and Tkalčić, 2016), while quantifying the data uncertainty in synthetic seismograms requires additional computational effort and, therefore, it is often disregarded despite its significance. In the context of the field data example, Duputel et al. (2011) highlights that incorporating modeling errors not only enhances the posterior distributions of estimated moment tensors but also improves the accuracy of these estimates. Similarly, Gu et al. (2018) demonstrates that inaccuracies in 1D velocity models can significantly affect moment tensor estimations. Furthermore, on synthetic data, Pham et al. (2024) show that incorporating errors from modeling using 3D velocity models yields more reliable estimated parameters. All in all, accounting for the uncertainty of both field and synthetic data is crucial for achieving a holistic understanding of the limitations and errors inherent in seismic studies. By addressing these uncertainties, the reliability and robustness of estimated parameters can be significantly improved.

REFERENCES

Dost, B., van Stiphout, A., Kühn, D., Kortekaas, M., Ruigrok, E., and Heimann, S. (2020). Probabilistic moment tensor inversion for hydrocarbon-induced seismicity in the Groningen gas field, the Netherlands, part 2: Application. *Bulletin of the Seismological Society of America*, 110(5):2112–2123.

Duputel, Z., Rivera, L., Fukahata, Y., and Kanamori, H. (2011). Uncertainty estimations for seismic source inversions. *Geophysical Journal International*, 190:1243–1256.

Gu, C., Marzouk, Y. M., and Toksöz, M. N. (2018). Waveform-based Bayesian full moment tensor inversion and uncertainty determination for the induced seismicity in an oil/gas field. *Geophysical Journal International*, 212(3):1963–1985.

Kruiver, P. P., van Dedem, E., Romijn, R., de Lange, G., Korff, M., Stafleu, J., Gunnink, J. L., Rodriguez-Marek, A., Bommer, J. J., van Elk, J., et al. (2017a). An integrated shear-wave velocity model for the Groningen gas field, The Netherlands. *Bulletin of Earthquake Engineering*, 15:3555–3580.

Kruiver, P. P., Wiersma, A., Kloosterman, F. H., de Lange, G., Korff, M., Stafleu, J., Busschers, F. S., Harting, R., Gunnink, J. L., Green, R. A., et al. (2017b). Characterisation of the Groningen subsurface for seismic hazard and risk modelling. *Netherlands Journal of Geosciences*, 96(5):s215–s233.

Mustać, M. and Tkalčić, H. (2016). Point source moment tensor inversion through a Bayesian hierarchical model. *Geophysical Journal International*, 204(1):311–323.

Pham, T.-S., Tkalčić, H., Hu, J., and Kim, S. (2024). Towards a new standard for seismic moment tensor inversion containing 3-D earth structure uncertainty. *Geophysical Journal International*, 238(3):1840–1853.

Romijn, R. (2017). Groningen velocity model 2017—Groningen full elastic velocity model September 2017 (available at https://nam-feitenencijfers.data-app.nl/download/rapport/9a5751d9-2ff5-4b6a-9c25-e37e76976bc1?open=true). *Nederlandse Aardolie Maatschappij (NAM)*.

Willacy, C., van Dedem, E., Minisini, S., Li, J., Blokland, J.-W., Das, I., and Droujinine, A. (2019). Full-waveform event location and moment tensor inversion for induced seismicity. *Geophysics*, 84(2):KS39–KS57.

ACKNOWLEDGEMENTS

Whenever people ask me about my hometown, I always say *Baubau*, a small city on a tiny island in eastern Indonesia. But I also add that my second hometown is **Delft**. This city and the TU have been incredibly kind to me. They've given me opportunities, friendships, and countless good memories. My PhD journey hasn't always been easy, but it has been filled with meaningful experiences and unforgettable adventures.

Accomplishing this journey would not have been possible without the support of my promotors, **Kees Weemstra** and **Kees Wapenaar**. Thank you both for giving me this opportunity and for guiding me with patience, clarity, and trust, helping me grow not only as a researcher but also as an independent thinker. **Kees Weemstra**, I still remember our early meetings when you told me your PhD years were some of the best times of your life. I was skeptical at the time, but looking back, I can now say the same. Your calm, constructive, and thoughtful mentoring helped me through both technical challenges and moments of self-doubt. As the oldest in my family, I rarely had someone to lean on. Your support often felt like what I imagine having an encouraging older brother would be like. I've learned so much from you about science, academic integrity, and most importantly, work and life balance. **Kees Wapenaar**, thank you for your deep insights, sharp scientific intuition, and for always encouraging me to look beyond the equations and consider the bigger picture. I remember reading one of your papers as a bachelor student. It was complex and well beyond my grasp at the time, and it made me dream that one day I could attend your lectures. That dream came true in 2017. Your ability to simplify difficult concepts and connect them to fundamental principles has been truly inspiring. I'm also deeply grateful to **Evert Slob**. The letter you wrote for me helped me secure the scholarship that brought me to Europe, an opportunity that changed the course of my life and opened doors I never imagined.

Over the past four years, I've had the privilege of working alongside many wonderful individuals with thoughtful minds. In particular, I shared my time at TU Delft with eight remarkable office mates: **Jingming**, **Aukje**, **Samantha**, **Faezeh**, **Hamed**, **Aleks**, **Sara**, and **Milad**. Thank you for the insightful discussions, constructive feedback, and for contributing to a supportive and collegial atmosphere. Being part of such an encouraging environment made a meaningful difference throughout this journey. To all the colleagues I met at the campus during this PhD, whether we exchanged ideas in meetings, shared coffee during breaks, or sat together in the canteen, thank you, you made this journey not only possible but also meaningful. To all the **great people of D.O.G.S** thank you for the fun small parties during gaming night. To all my **Indonesian colleagues** in Delft, thank you for being a second family to me. Your presence makes me feel closer to home, and I am deeply grateful for the laughter we shared and the support you offered throughout this journey. I am deeply grateful for the laughter we shared and the support you offered throughout this journey. Special thanks to **Bang George**, who has been a constant and supportive friend throughout my PhD years.

To my friends from the **Applied Geophysics Master's** program, thank you. You guys are amazing and have been truly inspirational. To all my **teachers** throughout my academic life, I am deeply grateful to have been taught and guided by you. Your knowledge and dedication shaped the path that brought me here. If someone had asked me sixteen years ago what I wanted to be, I wouldn't have had an answer. As a teenager, I wished for nothing more than a simple life. Growing up in a remote part of the Indonesian coast, I had little sense of how powerful education can be, how vast the world is, and how diverse other cultures are. And yet, here I am, standing in a place I never imagined I would reach.

To everyone who has shown me kindness along the way, there are truly so many of you. You may not even realize the impact you have made on me. Your words, gestures, and presence made all the difference. Even now, I sometimes catch myself looking up at the sky and asking, *Why am I here? How did I get here?* In those quiet moments, I am grateful to have your support. You reminded me that I was not alone. You inspired me to keep going and to grow not just as a researcher but as a human being. Because of you, I dared to imagine a future where I could also give back, where I could become someone who uplifts others just as you uplifted me. For that, and so much more, I am profoundly grateful. To my family back home, thank you for supporting me from the very beginning. The journey was not easy at the start, and there were many uncertainties along the way. Your support, in both small and significant ways, gave me the foundation I needed to pursue this path.

I dedicate this thesis mainly to my **teenage self**, who lived through difficult times and often doubted whether something like this would ever be possible. I would like to make it a personal reminder that steady effort and patience can take us further than we imagine. And to my **future son or daughter**, I hope this work serves as an example that distant goals are within reach with persistence and courage. More importantly, I hope to offer you the kind of support and guidance I once wished for myself.

And once again, to all the **kind people** who helped me through difficult times, **may God bless you, and may life be kind to you**.

CURRICULUM VITÆ

La Ode Marzujriban MASFARA RACHMAN

24-09-1993 Born in Bau-Bau, Indonesia.

EDUCATION

2011–2016 Bachelor of Science in Geophysics study program
Hasanuddin University, Makassar

2017–2019 IDEA League Joint Master in Applied Geophysics
TU Delft, ETH Zürich, RWTH Aachen

2019–2025 PhD in Applied Geophysics
TU Delft
Thesis: Estimating induced seismic source parameters using an efficient Hamiltonian Monte Carlo algorithm
Promotor: Prof. dr. ir. C.P.A Wapenaar
Copromotor: Dr. C. Weemstra

EXPERIENCE

2023–Present Geophysicist
Viridien, Oslo

List of Publications

Journal Articles

3. **Masfara, L. O. M. and Weemstra, C.** (In preparation). *Moment tensor and rupture directivity analysis of induced earthquakes in the Groningen gas field, the Netherlands.*

2. **Masfara, L. O. M. and Weemstra, C**. (2024). *Hamiltonian Monte Carlo to characterize induced earthquakes: Application to a ML 3.4 event in the Groningen gas field and the role of prior*. Earth and Space Science, 11(1):e2023EA003184.

1. **Masfara, L. O. M., Cullison, T., and Weemstra, C**.(2022). *An efficient probabilistic workflow for estimating induced earthquake parameters in 3D heterogeneous media*, Solid Earth, 13, 1309–1325.

Conference abstracts

7. **Masfara, L. O. M. and Weemstra, C**. (2023). *Hamiltonian Monte Carlo inversion of Groningen earthquake sources*, in: XXVIII General Assembly of the International Union of Geodesy and Geophysics (IUGG), GFZ German Research Centre for Geosciences.

6. **Masfara, L. O. M., Weemstra, C., and Cullison, T**. (2022).*Characterizing induced seismic events in the Groningen gas field using an efficient Hamiltonian Monte Carlo sampler: a case study*, in: EGU General Assembly Conference Abstracts, pp. EGU22–5398.

5. **Ruan, J., Masfara, L., Ghose, R., and Mulder, W**. (2022).*3D geomechanical modelling of induced seismicity: Finite source wavefield simulation to moment tensor inversion*, in: NAC Nederlands Aardwetenschappelijk Congres 2022.

4. **Ruan, J., Masfara, L. O. M., Ghose, R., and Mulder, W**. (2022b). *Linking geomechanical simulation of induced seismicity to surface seismic observations: Simulated finite fault rupture to moment tensor inversion*, in: SEG International Exposition and Annual Meeting, p. D011S081R003, SEG.

3. **Masfara, L. O., Cullison, T., Ruan, J., and Weemstra, C.** (2022). *Induced earthquake source characterization using multi-stage of Hamiltonian Monte Carlo algorithm*, in: AGU Fall Meeting Abstracts, vol. 2022, pp. S35E–0309.

2. **Masfara, L. O. M. and Weemstra, K**. (2021). *Towards efficient probabilistic characterisation of induced seismic sources in the Groningen gas field*, in: EAGE GeoTech 2021 First EAGE Workshop on Induced Seismicity, vol. 2021, pp. 1–5, European Association of Geoscientists & Engineers.

1. **Masfara, L. O. M., Weemstra, C., and Cullison, T**. (2021). *Efficient probabilistic inversion of induced earthquake parameters in 3D heterogeneous media*, in: 82nd EAGE Annual Conference & Exhibition, vol. 2021, pp. 1–5, European Association of Geoscientists & Engineers.

CODES

2. **Cullison, T. and Masfara, L. O. M**. (2022). *code-cullison/pyaspect: First Release*, https://doi.org/10.5281/zenodo.6987368.

1. **Cullison, T., Masfara, L. O. M., and Hawkins, R**. (2022). *code-cullison/gnam: First Release*, https://doi.org/10.5281/zenodo.6987375.