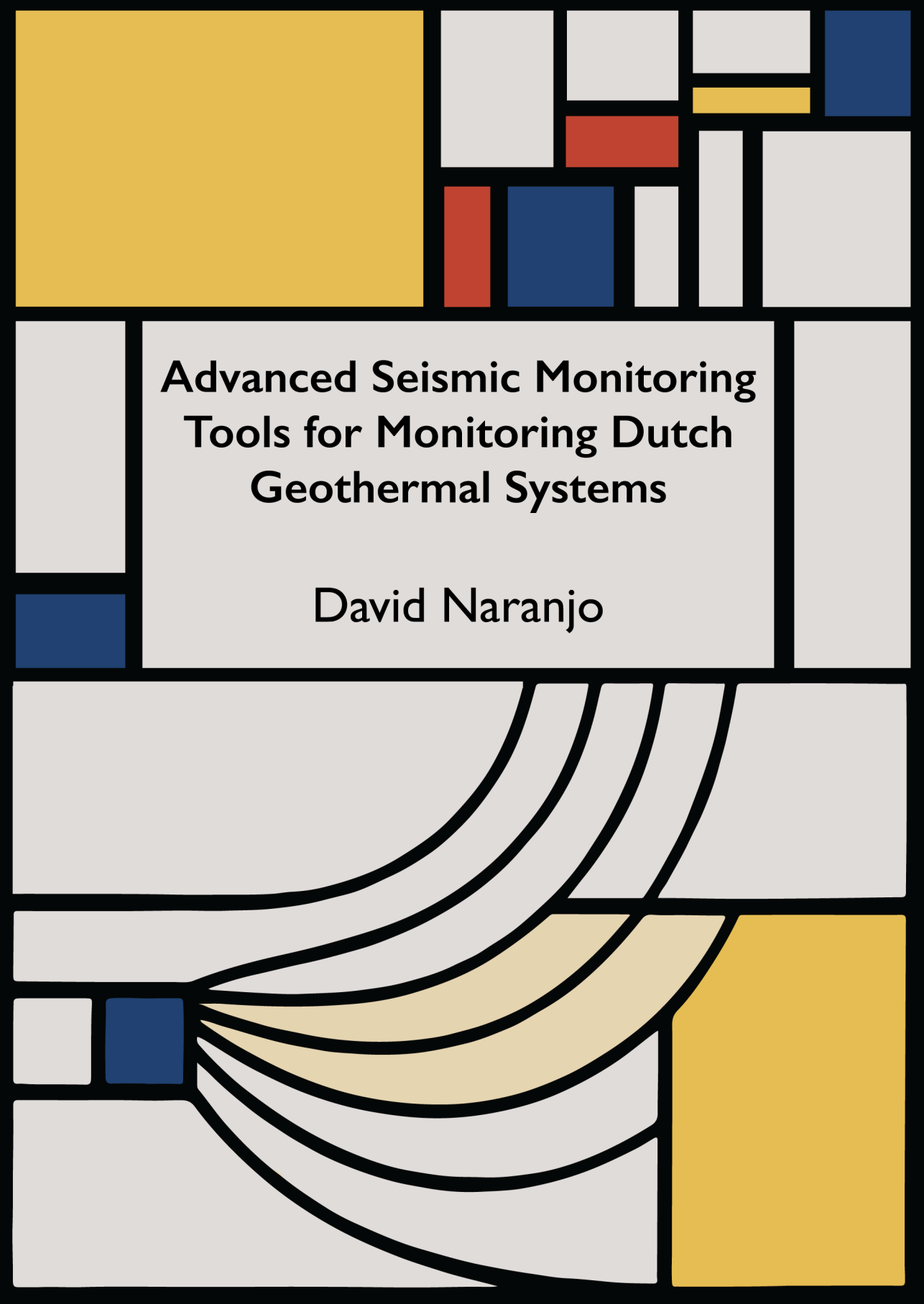
Advanced Seismic Monitoring Tools for Monitoring Dutch Geothermal Systems

David Naranjo



Advanced Seismic Monitoring Tools for Monitoring Dutch Geothermal Systems

Addressing Event Detection in Noisy Environments, Hypocenter Inversion, Velocity-Model Validation, Experimental Network Design, and Correction of Clock Errors

Advanced Seismic Monitoring Tools for Monitoring Dutch Geothermal Systems

Addressing Event Detection in Noisy Environments, Hypocenter Inversion, Velocity-Model Validation, Experimental Network Design, and Correction of Clock Errors

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. H. Bijl, chair of the Board for
Doctorates to be defended publicly on Wednesday, 4 March 2026 at 17:30

by

David Francisco Naranjo Hernandez

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof.dr. D.S. Draganov,	Delft University of Technology, promotor
Prof.dr. F. Wellmann,	Rheinisch-Westfälische Technische Hochschule Aachen University, Germany, promotor
Dr. C. Weemstra,	Royal Netherlands Meteorological Institute, copromotor
Independent members:	
Prof.dr. P.J. Vardon,	Delft University of Technology
Prof.dr. L.G. Evers,	Royal Netherlands Meteorological Institute and Delft University of Technology
Dr. A.G. Muntendam-Bos,	Delft University of Technology
Prof.dr. P. Martinez Garzon,	Rheinisch-Westfälische Technische Hochschule Aachen University, Germany, and Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences Potsdam, Germany
Reserve members:	
Prof.dr. D.F. Bruhn,	Delft University of Technology

The doctoral research has been carried out in the context of an agreement on joint doctoral supervision between Rheinisch-Westfälische Technische Hochschule Aachen University, Germany and Delft University of Technology, the Netherlands.
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956965.

Keywords: Geothermal Energy, Induced Seismicity Monitoring, Seismic Interferometry, Passive Seismic Data Processing, Probabilistic Inversion, Network Design, Ocean-Bottom Seismometers

Cover design by D. Naranjo based on Figure A.1 and inspired by De Stijl.

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

This thesis is dedicated to the memory of Kemâl Erbas.
His passion, kindness, and support are a source of inspiration.

CONTENTS

Summary xiii

Samenvatting xv

Acknowledgements xvii

1 Introduction 1
- 1.1 Passive Seismic Monitoring . . . 3
- 1.2 Advancing seismic monitoring . . . 5
- 1.3 Offshore Geothermal Energy and Carbon Capture and Storage . . . 5
- 1.4 Research Questions . . . 6
- 1.5 Thesis Aim and Objectives . . . 7
- 1.6 Outline . . . 7
- References . . . 9

2 A Workflow for Low-Magnitude Event Detection, Hypocenter Estimation & Uncertainty Quantification 13
- 2.1 Introduction . . . 14
- 2.2 Geological Context and Seismicity in the Netherlands . . . 14
- 2.3 Workflow for seismic monitoring . . . 16
  - 2.3.1 Data Management . . . 16
  - 2.3.2 Event detection . . . 17
  - 2.3.3 Hypocentre Estimation . . . 18
  - 2.3.4 Local Magnitude Estimation . . . 23
- 2.4 Data . . . 24
  - 2.4.1 Waveform data . . . 24
  - 2.4.2 Seismic-velocity model . . . 24
- 2.5 Results . . . 25
  - 2.5.1 Event detections . . . 25
  - 2.5.2 Waveform characterization . . . 27
  - 2.5.3 Hypocentres and magnitudes . . . 28
- 2.6 Discussion . . . 33
- 2.7 Conclusions . . . 34
- References . . . 34

3 Velocity-Model Validation Through Body-wave Seismic Interferometry 39
- 3.1 Introduction . . . 39
- 3.2 Theory . . . 40
  - 3.2.1 Body-wave Seismic Interferometry . . . 40

3.3 Workflow . . . 42
3.4 Data . . . 44
3.4.1 Seismic velocity model and expected two-way travel times . . . 44
3.4.2 Waveform data . . . 45
3.5 Noise analysis and applied filtering . . . 50
3.6 Waveform analysis . . . 51
3.7 Results . . . 52
3.7.1 Vertical ZO-VRR trace analysis . . . 52
3.7.2 Horizontal ZO-VRR trace analysis . . . 52
3.8 Discussion . . . 55
3.9 Conclusions . . . 55
References . . . 56

4 Seismic Network Design for Monitoring Offshore 59
4.1 Introduction . . . 60
4.2 Theory . . . 61
4.2.1 Non-linear DN-optimisation . . . 61
4.3 Workflow . . . 63
4.3.1 Step 1: Prior distribution of seismicity . . . 64
4.3.2 Step 2: Design Space . . . 65
4.3.3 Step 3: Travel-time estimation . . . 65
4.3.4 Step 4: Travel Time Uncertainty . . . 66
4.3.5 Step 5: Event Detectability . . . 67
4.3.6 Step 6: Likelihood estimation . . . 67
4.3.7 Step 7: Optimisation . . . 67
4.3.8 Step 8: Validation—Posterior distribution . . . 68
4.4 Data . . . 69
4.4.1 Site information—Aramis CCS field . . . 69
4.4.2 Fault geometries . . . 69
4.4.3 Seismic-velocity model . . . 70
4.4.4 Observed travel-time uncertainties . . . 70
4.4.5 Ground-motion prediction equation . . . 70
4.5 Results . . . 71
4.6 Discussion . . . 75
4.7 Conclusions . . . 76
References . . . 76

5 Clock error correction with Ambient Seismic Noise: Application to Reykjanes Geothermal Field 79
5.1 Introduction . . . 79
5.2 Theory . . . 81
5.2.1 Seismic interferometry . . . 82
5.2.2 A model to account for clock drift . . . 82
5.2.3 Additional arrival time shifts . . . 85
5.2.4 Determination of $t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})}$ . . . 86
5.2.5 Matrix formulation . . . 87

5.2.6 Inverting for clock drift . . . 88
5.3 Implementation & application to data . . . 89
5.3.1 The IMAGE data set . . . 90
5.3.2 Workflow . . . 91
5.3.3 OCloC . . . 96
5.3.4 Bootstrap re-sampling . . . 98
5.4 Results . . . 99
5.4.1 Clock drift rates . . . 99
5.4.2 Comparing inversion strategies . . . 104
5.5 Discussion . . . 105
5.5.1 The effect of the surface wave (noise) illumination pattern . . . 105
5.5.2 Validation using only land stations . . . 105
5.5.3 On the validity of the assumption of linear clock drift and an initial clock error . . . 105
5.5.4 Performance in the absence of land stations? . . . 107
5.5.5 Which projects can benefit from OCloC? . . . 107
5.6 Conclusions . . . 108
References . . . 109

6 Conclusions **115**
6.1 Main conclusions and novel contributions . . . 115
6.2 Limitations and future work . . . 116

A Appendix **119**
A.1 Influence of station 030 on hypocentre uncertainty . . . 119
A.2 Example of cross-correlation pick correction . . . 120
A.3 Summary of source-receiver pairs selected for body-wave seismic interferometry . . . 121
A.4 Spectrogram analysis . . . 123
A.5 Matrix formulation . . . 127
A.6 Detection of outliers . . . 128
A.7 Clock drifts of each OBS station . . . 129
References . . . 131

Curriculum Vitæ **133**

List of Publications **137**

Glossary and List of Abbreviations

M_L Local magnitude. A magnitude scale based on the maximum amplitude of seismic waves recorded at local distances.. 23, 29, 68

M_c Magnitude of completeness: the lowest earthquake magnitude that a seismic network can detect.. 4, 5, 15

M_d Duration magnitude. A magnitude scale derived from the signal duration of a seismic event.. 14, 25

CCS Carbon Capture and Storage. 6, 8, 60, 69, 75

CK Chalk Group.. 55

ED Experimental design: the method of planning an experiment to make sure that the data to be collected is appropriate and sufficient to answer the research questions of the experiment.. 59, 60, 64

EHE East–West horizontal ground-motion component recordings of the geophone used for S-wave analysis.. 43, 50, 52, 54

EHN North–South horizontal ground-motion component recordings of the geophone used for S-wave analysis.. 43, 50, 52, 54

EHZ Vertical ground-motion component recordings of the geophone used for P-wave analysis.. 43, 50

EIG Expected Information Gain: The average Information Gain that would result from performing an experiment. It quantifies, before collecting data, how much we expect the uncertainty about model parameters would decrease.. 61

GHP Geothermal-heat production from deep subsurface reservoirs in the Netherlands, typically at depths greater than 500 m, using open-loop systems (e.g., doublets) to supply heat for industrial and district heating applications.. 1, 3–6

GMPE Ground Motion Prediction Equations.. 67, 70

GNSS Global Navigation Satellite System: a satellite-based system that provides geospatial positioning with global coverage, enabling time synchronization and navigation.. 6, 79, 108

KNMI Royal Netherlands Meteorological Institute (Dutch: Koninklijk Nederlands Meteorologisch Instituut).. 3–5, 7, 8, 15, 33

NLNM Lower North Sea Group and Middle North Sea Group.. 52, 55

NU Upper North Sea Group.. 52

OBS Ocean-Bottom Seismometer.. 6, 7, 79–84, 86, 88–96, 98–108, 110, 112

PSM Passive seismic monitoring: the continuous recording and analysis of ambient and induced seismic signals using a network of sensors, typically without active source input, commonly used to detect, locate, and characterize seismic events.. 3, 7, 8, 14

RVG Roer Valley Graben.. 14, 15

SHRA Seismic Hazard and Risk analysis, the English term for *Seismische Dreigings- en Risico Analyse* (SDRA), a mandatory analysis required for Dutch geothermal project permit applications when production is deeper than 500 m.. 3

SI Seismic Interferometry: A method used to retrieve the Green's function between two seismic receivers by cross-correlating ambient noise or recorded seismic wavefields, enabling applications such as the study of subsurface structures without active sources.. 39–42, 45, 55, 80–83

SNR Signal-to-noise ratio.. 92–94, 98, 100, 105, 107, 108

TLS Traffic Light System: a real-time operational risk management tool used to monitor and mitigate induced seismicity during subsurface activities such as geothermal energy production.. 3

UTC Coordinated Universal Time.. 30, 81, 88, 89, 107

Velmod The regional velocity model VELMOD-4 is the successor of VELMOD-3.1 and is built in a Joint Industry project in the time period 2017-2019 (partners: Total and NAM). For the first time seismic stacking velocities have been used as input to the VELMOD velocity model in an effort to better capture the lateral velocity variation in between wells. The model integrates velocities measured in boreholes (sonic logs and check shot data) with the stacking velocities derived from seismic data.. 16

WNB West Netherlands Basin.. 15

ZO-VRR Zero-offset virtual reflection response: a seismic response retrieved by SI to approximate the reflection signal that would be recorded at a receiver colocated with a virtual source, with no source-receiver offset.. 40–43, 52–55

SUMMARY

One of the major challenges of modern society is the goal of reducing the output of anthropogenic greenhouse gases to the atmosphere. To contribute to this goal, the Netherlands is scaling up the use of geothermal energy (GE), a low-carbon technology that provides heating for infrastructure. Geothermal energy requires the extraction of geothermal fluids from a geologic reservoir to provide heat at the surface, followed by re-injection of the fluid into the subsurface. The re-injection, circulation, and extraction of the fluids affect the local stress conditions and can lead to the generation of low-magnitude seismic events. The detection, characterisation, and interpretation of these events improve the efficiency and safety of geothermal operations. This thesis aims to contribute to the Dutch government's efforts to upscale the use of geothermal energy by enhancing and extending the existing passive seismic tools for monitoring the safety and efficiency of geothermal energy.

In Chapter 2, we introduce a seismic monitoring workflow to detect and characterise low-magnitude seismic events in noisy environments. We incorporate uncertainties from the open-access regional seismic velocity model into the hypocentre estimations. We apply the workflow to the recorded data from a temporary passive network deployed around the Kwintsheul geothermal operation in South Holland, where one low-magnitude seismic event had previously been reported. The network is located in a high-noise environment, characteristic of most geothermal operations in the Netherlands. Despite the high noise levels, we identify five additional low-magnitude seismic events that occurred close to a local fault and the injection well. These are the first events ever recorded in the region. However, large hypocentre uncertainties—due to limitations in the seismic velocity model and sparse azimuthal coverage—prevent a clear interpretation of the underlying processes. From these two limiting factors, only the velocity model can be refined after an event has been recorded. However, refining the available seismic velocity model implies significant costs, as active seismic sources are usually used.

In Chapter 3, we introduce a workflow for validating the seismic velocity model based on body-wave seismic interferometry as a cost-effective alternative. Our workflow is motivated by the possibility of retrieving virtual-offset reflection responses when seismic energy arrives with near-vertical incidence to the receivers. We apply our workflow using the low-magnitude seismic events that we detected. We find that the P-wave velocity model effectively explains the observed retrieved reflections at shallow depths. In contrast, the available S-wave models do not match the data. We conclude that the P-wave model is reliable for hypocentre studies, but that the S-wave model requires refinement.

In Chapter 4, we address how the network geometry influences the detectability and hypocentre resolution of seismic events and implement a workflow for designing seismic networks. In our workflow, we integrate open-access subsurface information to generate a synthetic earthquake catalogue using knowledge of faults and areas of expected higher seismicity risk. We then apply a non-linear design strategy and a global search algorithm to ensure approximately optimal configurations. Finally, we validate the network designs

through synthetic hypocentre inversions. We identify the Dutch North Sea as the area in most need of seismic receivers, due to (i) upcoming carbon capture and storage (CCS) initiatives, (ii) the lowest existing network coverage, and (iii) the potential future use of existing oil-and-gas infrastructure for offshore geothermal-energy developments. We apply our workflow to the K-14 offshore field, where carbon capture and storage is planned. The results show that the optimised networks provide sufficient azimuthal coverage and location accuracy, even under simplified assumptions. This workflow can guide the design of cost-effective networks in both onshore and offshore environments.

In Chapter 5, we focus on accurate time synchronization of seismic networks. We introduce a data-driven method to detect and correct clock errors using the time-symmetry of ambient-noise correlations. We apply our method to the IMAGE network in Reykjanes, Iceland, deployed to monitor offshore geothermal activity. Offshore geothermal-energy operations introduce additional challenges due to the need for ocean-bottom seismometers (OBS), which lack direct access to GNSS signals, leading to clock-drift errors that affect event timing and localisation. We show that most OBS in the network experienced clock drift, and some had large initial time offsets. We provide an open-source Python package (OCloC) that implements this method, enabling better timing accuracy and improved hypocentre estimation in future offshore monitoring, which can be applied in future offshore geothermal energy and the upcoming carbon capture and storage operations in the Dutch North Sea.

Together, in this thesis we introduce new or adapted workflows to tackle specific limitations in current low-magnitude seismic monitoring practices. By addressing these challenges, this thesis advances the capabilities of seismic monitoring in both onshore and offshore settings. By improving detection, location, velocity model validation, network design, and timing correction, this thesis contributes to the development of robust and cost-effective seismic monitoring systems. These tools support operators and regulators in making informed decisions for the safe and sustainable scaling of geothermal energy and carbon storage in the Netherlands and beyond.

SAMENVATTING

Het terugdringen van de uitstoot van broeikasgassen is een van de grootste uitdagingen van de moderne maatschappij. Als onderdeel van het klimaatbeleid schaalt Nederland het gebruik van energie uit aardwarmte op als een warmtebron met een lage uitstoot van broeikasgassen. Aardwarmte wordt aan de grond onttrokken door de circulatie van vloeistoffen door diepe aardlagen, waarbij de spanningscondities in de ondergrond veranderen en wat seismische activiteit kan opwekken. Het detecteren en nauwkeurig lokaliseren van deze zwakke signalen is lastig vanwege hoge omgevingsruis, beperkte kennis van de ondergrond en een dun of suboptimaal netwerk van meetstations. Dit proefschrift richt zich op het verbeteren van passieve seismische monitoring om de veiligheid en efficiëntie van geothermie in Nederland te ondersteunen.

In Hoofdstuk 2 introduceren we een workflow voor het detecteren en karakteriseren van lage-magnitude aardbevingen in een ruisrijke omgeving. Deze wordt toegepast op een tijdelijk netwerk rondom de geothermische installatie van Kwintsheul (Zuid-Holland), waar eerder slechts één seismisch evenement werd gerapporteerd. Met behulp van template matching en probabilistische locatiebepaling detecteren we vijf extra lage-magnitude seismische gebeurtenissen in de buurt van een breuk en het injectiepunt. Dit zijn de eerste geregistreerde bevingen in deze regio. Door onzekerheden in het snelheidsmodel en beperkte dekking kunnen we de aard van deze gebeurtenissen echter niet eenduidig interpreteren.

In Hoofdstuk 3 evalueren we het gebruikte snelheidsmodel voor hypocentra. We ontwikkelen een workflow op basis van seismische interferometrie met ruimtegolven om virtuele reflecties te verkrijgen bij verticale invalshoeken. Toegepast op de dataset van Kwintsheul bevestigt deze methode dat het P-golvenmodel de gemeten ondiepe reflecties juist kan verklaren, terwijl het S-golvenmodel niet aansluit bij de data. Het P-golvenmodel is dus geschikt voor locatiebepaling, maar het S-golvenmodel heeft verfijning nodig voor een nauwkeurigere analyse.

In Hoofdstuk 4 onderzoeken we hoe de geometrie van het meetnetwerk de detecteerbaarheid en hypocentrum resolutie van aardbevingen beïnvloedt. We presenteren een workflow voor het ontwerpen van seismische netwerken. Hierin integreren we openbaar beschikbare data van de ondergrond om een synthetische aardbevingscatalogus te maken gebaseerd op kennis van breuken en seismische risicogebieden. We passen niet-lineaire optimalisatiestrategieën en een globaal zoekalgoritme toe om een optimale configuratie te verzekeren. Uiteindelijk valideren we de meetnetwerken met synthetische hypocentrum inversies. We identificeren dat het Nederlandse Noordzeegebied het vooral meer seismische meetapparaten nodig heeft, omdat (i) er daar meerdere initiatieven voor de afvang en opslag van CO2 (CCS) plaatsvinden, (ii) er op dit moment de laagste dekkingsgraad van netwerken is en (iii) het potentieel voor het ontwikkelen van geothermische energie-installaties groot is door de bestaande olie- en gas-infrastructuur. We passen onze workflow toe op het K-14 veld, waar CCS onder de Noordzee is gepland. De resultaten tonen aan dat het geoptimaliseerde netwerk voldoende azimutale dekking en locatienauwkeurigheid biedt, zelfs onder

vereenvoudigde aannames. Deze workflow kan het ontwerp van kost-effectieve netwerken helpen, zowel op land als ter zee.

In Hoofdstuk 5 richten we ons op tijdsynchronisatie van seismische netwerken. We introduceren een datagestuurde methode om klokfouten te detecteren en te corrigeren met de tijdsymmetrie van correlaties van omgevingsruis. Deze methode passen we toe op het IMAGE-netwerk in Reykjanes, IJsland, dat werd ingezet voor geothermische monitoring. Bij het monitoren van het opwekken van geothermische energie ter zee worden zogeheten ocean-bottom seismometers (OBS) gebruikt, die geen verbinding met GNSS-signaal hebben. Dit heeft opbouwende klokfouten tot gevolg, die leiden tot problemen met timing en localisatie. De resultaten tonen aan dat vrijwel alle OBS-stations in het netwerk zulke fouten ondervonden en dat sommige stations een grote initiële fout hadden. We hebben een open-source Python-pakket ontwikkeld—OCloC—om deze correcties uit te voeren, zodat de timing en locatiebepaling van microseismiciteit ter zee in toekomstige monitoringsprojecten verbeterd kan worden, waaronder de komende CCS-operaties in de Nederlandse Noordzee.

In dit proefschrift introduceren we nieuwe en aangepaste workflows om specifieke beperkingen in de huidige praktijk van het monitoren van seismiciteit met lage magnitudes aan te pakken. Hierdoor verbetert de scriptie de capabiliteiten van seismische monitoring zowel ter land als ter zee. Door detectie, lokalisatie, modelvalidatie, netwerkontwerp en tijdscorrectie draagt deze scriptie bij aan het ontwerp van robuuste en kosteneffectieve seismische monitoringssystemen. Deze hulpmiddelen ondersteunen de exploitanten en toezichthouders om verstandige keuzes te maken bij het veilig en duurzaam opschalen van geothermische energie en de ondergrondse opslag van CO2 in Nederland en het buitenland.

ACKNOWLEDGEMENTS

It has been a long-time dream to study abroad, and I could not have anticipated ending up at TU Delft doing a PhD. I was very lucky to receive the Colfuturo and Colciencias scholarship that enabled my arrival in Europe in 2018. Over the past ≈ 7 years, I have been fortunate to be welcomed, supported, and motivated by many people, without whom this thesis would not have been possible. This chapter is dedicated to them.

Greatness is defined by the ability to inspire those around you. With this in mind, I want to express my deepest gratitude to **Deyan Draganov**. In moments when I doubted myself or my progress, your encouragement reminded me that I could reach the finish line. Your guidance through scientific advice, thoughtful planning, and the art of asking the right question at the right time helped me move forward. You are a leader who "knows the way, goes the way, and shows the way" (C. Maxwell).

As part of becoming an independent researcher, I had to learn how to test the limits of our ideas. In this regard, I am deeply grateful to **Kees Weemstra** for his trust and meticulous scientific input. Your feedback pushed me to sharpen my ideas and grow into a more rigorous researcher. I am grateful for the many opportunities that you and **Deyan** provided. This includes some of the most life-changing experiences, such as presenting at conferences worldwide and receiving training in the subjects I was most eager to learn.

I want to thank Nature's Heat company and **Paul van Schie** for allowing us to conduct seismic monitoring experiments on-site and for their assistance in maintaining the network. I appreciate the continued freedom that we were given to explore and investigate. More companies should strive for this level of openness, willingness to share, and commitment to safety.

My PhD was made possible by the EASYGO-ITN program under the Marie Skldowska-Curie grant agreement No 956965. I am grateful to Maren Brehme for forming this ITN and being an excellent leader. Thank you, Maren and Paromita Deb, for always welcoming everyone with a warm smile and helping us fix all sorts of issues. I would also like to thank my cohort: **Andrea**, **Anna**, **Esteban**, **Hagen**, **Josselin**, **Lorenzo**, **Mahmoud**, **Nicolas**, **Ryan**, **Selvican**, **Tristan**, and **Wen**. It was great fun to share our training weeks with you. I always felt supported by our easygoing network.

My research was conducted at the Section of Applied Geophysics and Petrophysics at TU Delft. It was a pleasure being part of this section, and I would like to thank all my colleagues, past and present, with whom I worked. Thank you, **Barbara**, and **Marlijn** for your patience during our short Dutch conversations and for your continuous support. I would also like to extend a warm appreciation to my PhD colleagues/friends from the Geoscience and Engineering Department. Thank you, **Aleks, AliReza, Andrea, Aukje, Aydin, Billy, Camille, Chris, Cocho, Diether, Eddy, Emilio, Faezeh, Florencia, Jasper, Jingming, Joeri, Johno, Liliana, Maria, Menno, Milad, Parvin, Shihao, Siamak, Sverre, Valeria,** and **Zhenja**.

Even in busy times, some people still make time to help others. During our fieldwork, I was genuinely encouraged by how readily some of my colleagues/friends offered their support. Thank you, **Boris Boullenguer**, **Consuelo Rebolledo**, **Deyan**, **Diana Arias**, **Eddy Revelo**, **Edoardo Navarra**, **Fan Wang**, **Flavio Manara**, **Jens**, **Joeri Brackenhof**, and **Sverre Hassing**, for helping out during our fieldwork campaign. A special thanks to you, **Jens van den Berg**, for your tireless work during fieldwork—your dedication and work ethic are truly exceptional. I would also like to thank **Joeri** for his selfless willingness to help others in our department.

Being part of the geophysics student chapter opened many opportunities for me. I want to thank my fellow DOGS board members—**Aukje**, **Billy**, **Eddy**, **Entela**, **Florencia**, **Iban**, **Johnno**, **Maria**, and **Sverre**. It was a pleasure working with you. A big thank you to **Evert Slob**, who not only guided the chapter but also continually strives to create a better work environment for everyone in the faculty. I would also like to thank the **SEG** for the travel grant to attend the leadership symposium in Houston, which was a truly remarkable experience.

Part of my research was conducted at RWTH Aachen. Thank you, **Florian Wellmann**, for receiving me in your group. This was truly an important part of my PhD. Not only did I get fresh ideas on uncertainty, but we also enjoyed wonderful team outings, company visits, and weekly seminars. Thank you, **Sandra** and **Lukas**, for helping me out with all my documents and our lunch breaks. Thank you, **Andrea**, **Elisa**, **Esteban**, the **Florians (Wellmann and Wagner)**, **Hagen**, and **Sofia**, for making my time in Aachen truly enjoyable.

Another part of my research was also done at EBN. Thank you, **Marc Hettema, Sjoukje de Vries, Jorien Schaaf,** and **Mara van Eck**, for making my secondment at EBN a wonderful experience and for welcoming me into your group.

The International Training Course (ITC) in Seismology and Seismic Risk at the GFZ was instrumental in both deepening my understanding of the field and motivating me to pursue a career in seismology. Thank you, **Claus Milkereit**, **Dorina Kroll**, **Simone Cesca**, and all the lecturers of the ITC. I cannot express how much the program helped me grow. Thank you, **Marius Isken**, for welcoming me into your office well after the ITC had ended, for your selfless support, and for enabling a smooth and productive collaboration—the way open science should work. A special thank you to my dear ITC seismo-friends: Francisco, Sol, and Andrés. You truly motivated me—scientifically, personally, and musically.

Part of my PhD research started from my internship at KAUST. This internship led to my first journal publication and my first experience supervising an MSc thesis. Thank you, **Sigurjón Jónsson** and **Laura Parisi**, for providing me with this invaluable opportunity. **Laura**, thank you especially for continuing to check in with me long after the internship was over and for guiding me in making important career decisions.

I would also like to thank **Sören Johnson** and the Centre for Languages and Academic Skills at TU Delft. The one-on-one sessions greatly improved my academic writing, and the courses provided many useful tools as well as a supportive environment in which fellow PhD candidates could share their experiences of writing their theses.

An important part of every PhD journey is what happens outside of work. Thank you, **Katrin Loer**, for your constant guidance and inspiring conversations, and for encouraging me to start playing the drums again. For the wonderful times performing at the faculty and the many great moments we shared, I would also like to thank the Geosirens band

members: **Eric Verschuur**, **Joep Storms**, **Justin Pittman**, **Katrin**, **Natascha Libanon**, and **Sian Jones**.

I would also like to thank the people who supported me throughout my PhD outside of the office. Thank you, **Maria**, and **Rigo**, for being my second family since the moment I stepped into Europe. I would like to thank my dear Potsdamer friends and almost family **Andy**, **Mauricio**, **Alice**, **Matija**, and **Olka**. Thank you, **Eddy**, for being such a great friend and flatmate. Thank you **Hamed**, **Nilgun**, and **Rutger**, for your unconditional friendship. Thank you also to Los Sobrevivientes **Diego**, **Marisol**, **Hector**, **Rafa**, **Male**, **Diana**, **Sebas**, y **Lore**. I also want to thank **Jos de Neve**, my taalcoach, for being one of the greatest teachers that I have had, and for her patience in meeting up with me every week to teach me Dutch. "Say what you mean and mean what you say". Thank you **Jean-Claire de Ruwe** for being a great teacher.

I would also like to thank those who inspired in me a passion for science: **Jean-Baptiste Tary**, **Kemal Erbas**, **Philippe Jousset**, **Tania Toledo**, and **Deyan Draganov**.

I am deeply grateful for the unconditional love of my parents, **Francisco** and **Astrid**, and my brother, **Federico**. I would not have been able to complete this PhD without your unwavering support. I hope one day I will be able to explain what I do in simple terms and the difference between an essay and a scientific article. Finally, I want to thank **Anastasia**, who put up with me during so many weekends and evening hours. Thank you for lifting me when I needed you the most, and for bringing so many colours into my life.

1

INTRODUCTION

A very critical attitude is required to avoid reading phases where they are expected to be

Inge Lehmann, 1987, Seismology in the Days of Old

The release of large amounts of CO_2 into the atmosphere is one of the main factors driving climate change (Raupach et al., 2007). To reduce CO_2 emissions, the Dutch government signed the National Climate Agreement in 2019. The agreement sets a target of a 49% reduction in CO_2 emissions by 2030, relative to 1990 levels (Government of the Netherlands, 2019). The agreement highlights the importance of sector-specific strategies. The heating sector is expected to reduce its CO_2-equivalent emissions by 40% in 2030 compared to 2015 (Schoof et al., 2018). To support this reduction, geothermal energy has been recognized as a reliable low-carbon option for the heating sector. In this context, the Dutch Geothermal Master Plan aims to scale up geothermal-heat production from 3 to 50 PJ by 2030 (Schoof et al., 2018).

The most common configuration for industrial-scale geothermal-heat production in the Netherlands is the doublet system, in which water circulates between an injection and a production well through a deep subsurface reservoir, typically $1.6-3$ km deep (Buijze et al., 2019). The wells form a production loop in which hot formation water is extracted through a production well, passes through a heat exchanger at the surface, and the cooled water is reinjected into the reservoir via an injection well. By re-injecting cooled water back into the reservoir, these systems help maintain subsurface pressure and extend the reservoir lifespan (Kamila et al., 2020).

The majority of Dutch geothermal-heat-production licenses are being granted in South Holland, as shown by the yellow polygons in Figure 1.1. Note that these licenses correspond to geothermal projects targeting reservoirs at depths greater than 500 m. Throughout this study, we refer to this type of operation as geothermal-heat production (GHP). Most of these licenses are for horticultural companies that are changing their heating systems from gas to geothermal energy. The advantage of geothermal energy is that it can be harnessed next to the infrastructure it serves and provides constant heat that is unaffected by daily or seasonal variations.

The injection and extraction of fluids in deep geological formations can lead to stress changes and fault reactivation (Eaton, 2018). Changes in temperature and pressure are

1

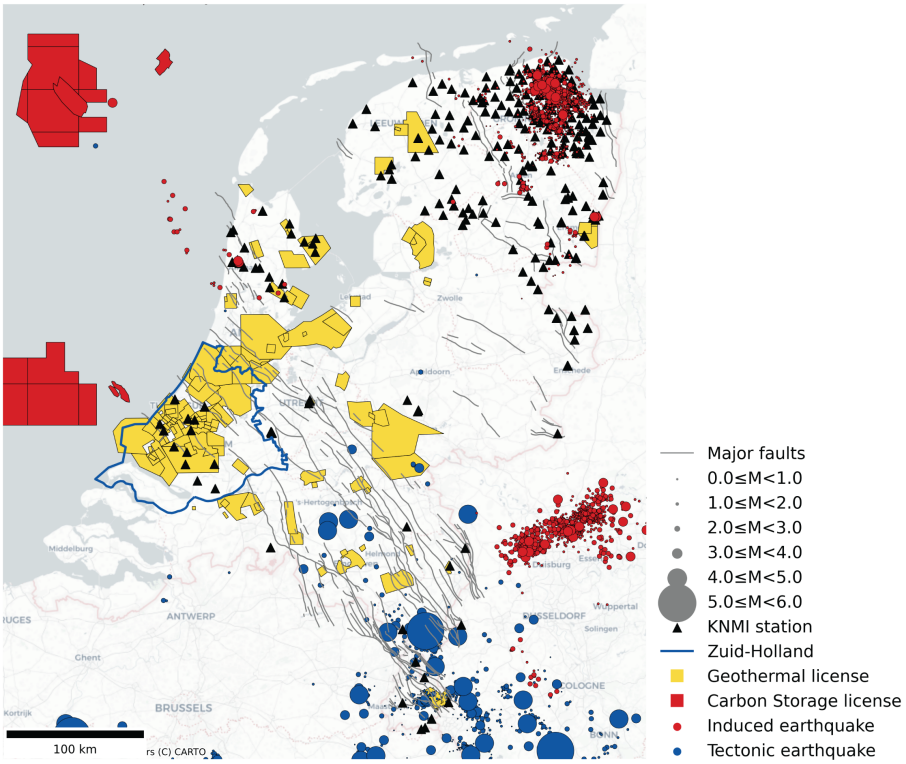


Figure 1.1: Map of the Netherlands showing the distribution of active geothermal-heat and CO_2-storage (CCS) licences alongside seismic and geological context. Geothermal-heat licences are shown as yellow polygons, while red polygons indicate CO_2 storage (CCS) licences (www.nlog.nl). The complete earthquake catalogue is shown as red (induced) and blue (tectonic) earthquakes (KNMI, 2023). The seismic receivers of the national seismic network are shown as black triangles (KNMI, 1993). The province of Zuid-Holland is outlined in blue.

the main factors affecting stress changes (Buijze et al., 2021). Unlike depletion-induced seismicity (e.g., in gas production), geothermal operations in permeable sedimentary reservoirs primarily can cause fault reactivation through cooling effects (Buijze, 2020; Mathur et al., 2024). Such stress changes are most pronounced around the injection well. Other key factors that could lead to fault reactivation are the presence of critically stressed faults, the magnitude of pressure changes, and the mechanical properties of the reservoir and surrounding formations (e.g., Evans et al., 2011; Foulger et al., 2017). Assessing these factors is essential for understanding the seismic risks associated with geothermal operations.

To manage seismic risks, GHP operators must adhere to the Seismic Hazard and Risk Assessment (SHRA), a step-wise procedure. The SHRA includes multiple factors to evaluate the safety of geothermal operations (Mijnlieff et al., 2023). Among these, the following are relevant to this thesis:

- **Fault presence and reactivation potential:** Induced earthquakes occur when changes in stress destabilize existing faults. Critically stressed faults are more susceptible, but industrial operations can also reactivate faults that were not initially critically stressed (Moein et al., 2023).

- **Historical seismicity:** The maximum earthquake magnitude is likely controlled by regional tectonics (Moein et al., 2023). The tectonic source of strain often controls the extent of the rupture. The timing, location, and magnitude of earlier events help constrain the evolution of stress and the likelihood of larger earthquakes.

- **Evolution of seismicity:** a Traffic Light System is implemented. The system requires investigation and mitigation when seismic events above 0.3 mm/s PGV are detected.

To assess these factors reliably, regulators require accurate subsurface models and continuous monitoring systems. In this thesis, we focus on the role of passive seismic monitoring in supporting this assessment.

1.1. PASSIVE SEISMIC MONITORING

Passive-seismic-monitoring (PSM) involves continuously recording ground motions using a series of receivers (e.g., geophones). The recorded ground-motion data are analysed to identify natural and/or induced seismic events, identify seismically-active faults, and support compliance with regulatory traffic-light systems (TLS).

The Royal Netherlands Meteorological Institute (KNMI) operates the Dutch PSM network and provides continuous monitoring of tectonic and induced earthquakes (KNMI, 1993). Figure 1.1 shows the network and the associated earthquake catalogue up to 2023 (KNMI, 2023). Tectonic earthquakes (blue circles) are concentrated in the south of the Netherlands, where they are linked to active faulting in the Roer Valley Graben (Houtgast and van Balen, 2000). Induced earthquakes (red circles) are concentrated in the north of the Netherlands, where they are primarily associated with gas extraction from the Groningen field and other hydrocarbon reservoirs (Bourne et al., 2014). The catalogue shows no earthquakes recorded in the province of South Holland, where the majority of geothermal

licenses have been granted[1]. The complete earthquake catalogue of the Netherlands is used to establish the historical seismicity, which is utilised in the SHRA analysis.

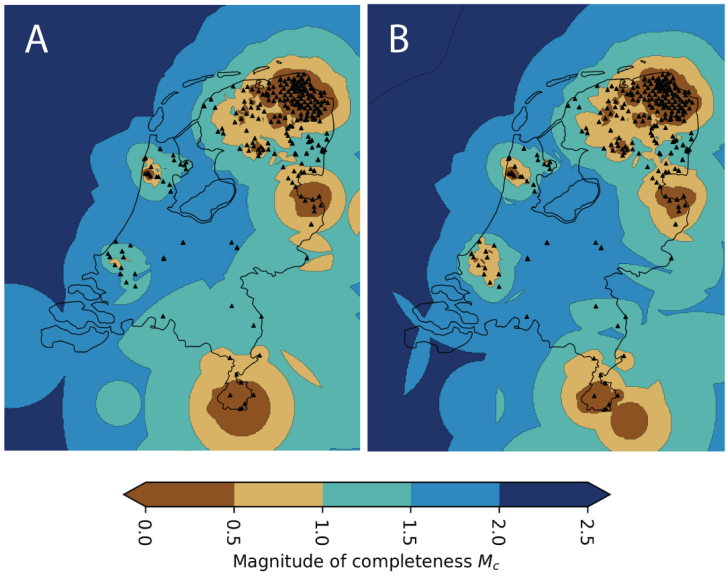


Figure 1.2: Magnitude of completeness of KNMI's seismic network (Ruigrok et al., 2023). A. Magnitude of completeness in 2021. B. Magnitude of completeness in 2024. The contour of the Netherlands is shown in black. The receivers of KNMI's network are shown as black triangles (KNMI, 1993).

In recent years, KNMI has increased the number of seismic stations installed in areas where mining activities have taken place[2]. A denser network allows for the detection of earthquakes of lower magnitudes. The minimum magnitude that a network can detect is defined by the magnitude of completeness (M_c). Figure 1.2 shows the M_c in 2021 (A) and 2024 (B). In 2021, the magnitude of completeness was mostly 1-1.5 in areas where most geothermal licenses are located. The magnitude of completeness in the same areas was reduced to values between 0.5-1.0. The high sensitivity of KNMI's network makes it ideal for monitoring the traffic light system that regulates geothermal projects.

An important aspect for monitoring GHP is the subsurface model. The Netherlands' subsurface models, which include lithostratigraphic units, faults, and seismic velocities, are compiled from active seismic data acquired under the Mining Act and are publicly available (Pluymaekers et al., 2017). These are used for estimating parameters of the seismicity recorded by the KNMI network.

[1] Muntendam-Bos et al. (2022) reported a seismic event that is not included in KNMI's catalogue). This event is analysed in Chapter 2.

[2] In the Netherlands, the term "mining" includes other uses of the deep subsurface, such as oil and gas production, geothermal energy, carbon dioxide storage, subsurface storage of other substances, and salt mining.

1.2. ADVANCING SEISMIC MONITORING

The traffic light system implemented under the SHRA applies conservative thresholds that guarantee safe geothermal operations. The SRIMA model of induced seismicity, used by the SHRA, simplifies the fault-reactivation process as a single large slip event during the project's lifetime (Fokker et al., 2023). These thresholds are effectively monitored by the KNMI's network, with M_c ranging from 0.5 to 1.5 in areas with geothermal licenses.

The detection of seismic events below the traffic light system thresholds provides valuable information for geothermal research. Low-magnitude events may indicate processes such as fluid migration or stress accumulation during injection/production activities (Eaton, 2018). Stress accumulation is a gradual process, and fault reactivation is likely to occur through multiple low-magnitude seismic events over time (Fokker et al., 2023).

To effectively identify and interpret the underlying process that low-magnitude seismicity indicates, several challenges need to be addressed. These include:

1. **Event detection**

   Low-magnitude seismic events (e.g., those with magnitudes below the magnitude of completeness) are challenging to detect because their signals have comparable amplitudes to the ambient seismic noise. Additionally, most GHP projects are often located in urban environments where ambient seismic noise is higher than in other areas.

2. **Event characterization**

   If seismic events are detected, we need to estimate their hypocentres to identify their underlying cause. Accurate event-hypocentre estimations depend on the quality of the available seismic velocity model. This model describes how seismic waves propagate through the Earth's subsurface. The seismic velocity of the Netherlands (Pluymaekers et al., 2017) contains information about its uncertainty, but this information is rarely incorporated into hypocentre-estimation methods. To advance geothermal monitoring, we need to develop methods to quantify and integrate velocity-model uncertainties into hypocentre estimations.

3. **Velocity-model validation**

   Beyond uncertainty quantification, validating the seismic velocity models remains a key challenge. If a velocity model is incorrect, event locations will be systematically biased, reducing the confidence in the monitoring results. Direct validation methods, such as active-source surveys, are costly. Alternative cost-effective methods need to be developed for the geothermal industry.

Addressing these challenges is critical for improving our understanding of the subsurface response to GHP and enhancing the reliability of seismic monitoring.

1.3. OFFSHORE GEOTHERMAL ENERGY AND CARBON CAPTURE AND STORAGE

Advancing our understanding of the subsurface response to geothermal-energy operations may facilitate the transition from onshore to offshore operations. Offshore geothermal en-

ergy represents a largely untapped renewable resource, where the seismic risk for the general population is minimal. The transition can be achieved by combining knowledge from other subsurface production/injection technologies.

The North Sea has a long history of oil and gas production that leaves behind wells and infrastructure that can be repurposed for offshore geothermal-energy production (Sircar et al., 2022). The Netherlands is currently exploring the usage of this infrastructure for offshore Carbon Capture and Storage (CCS). These projects present an opportunity to test new technologies that could be used in offshore geothermal energy.

CCS is part of the Dutch strategy to reduce industrial CO_2 emissions. CCS captures CO_2 from industry and injects it into deep formations such as depleted reservoirs and saline aquifers. Figure 1.1 shows the location of several exploratory licenses in the Dutch North Sea (Krevor et al., 2023; TNO – Advisory Group for Economic Affairs, 2022).

Safe storage of CO_2 has similar challenges as GHP. The effective storage depends on maintaining caprock integrity to prevent it from escaping to the surface (Verdon et al., 2010). The risk of induced seismicity is also present since injection of fluids changes local stress conditions and may reactivate faults (Verdon et al., 2010).

To monitor offshore injection operations, we need tools for:

4. **Offshore network design**

   Offshore, the Dutch seismic network has the lowest magnitude of completeness. To monitor offshore CCS operations, we need to design and install a seismic network with ocean-bottom seismometers (OBS). This type of seismic sensor can increase the coverage of the network in offshore settings.

5. **Ocean-bottom seismometer clock-error detection and correction**

   OBS stations do not have access to a Global Navigation Satellite System (GNSS) as the signal can not reach the sea bottom. This means that their internal clocks are not synchronised with an absolute reference time and can potentially drift. Clock-drift errors affect the accuracy of hypocentre estimations.

1.4. RESEARCH QUESTIONS

Several questions arise from this context:

- Since most geothermal doublets are located near industrial or urban areas, how can microseismic events be reliably detected in high-noise geothermal environments?
- How can the hypocentres of such events be located with quantified uncertainty?
- If we aim to expand geothermal operations offshore, how should offshore monitoring networks be designed for geothermal operations?
- How can clock errors in ocean-bottom seismometers be corrected to ensure accurate monitoring of offshore geothermal energy?

1.5. THESIS AIM AND OBJECTIVES

The aim of this thesis is to contribute to the efforts of the Dutch Climate Agreement (Government of the Netherlands, 2019) by applying and extending passive seismic methods to support the monitoring of geothermal-energy operations. To achieve this aim, the goals of this thesis are:

- Detection of low-magnitude seismicity beyond the thresholds of regulatory traffic light systems, to reveal signals of underlying geomechanical processes, such as fluid migration and cold-front propagation.
- Characterisation of hypocentre uncertainties by assessing the influence of velocity models and array configurations on event-location accuracy.
- Design and evaluation of monitoring strategies for offshore operations, with specific attention to the synchronisation of OBS in the absence of GNSS signals.
- Development of open-source tools that automate key steps in seismic monitoring, ensuring reproducibility and applicability in both onshore and offshore geothermal projects.

This thesis is structured into four core chapters that address the research questions presented above.

1.6. OUTLINE

This thesis is composed of the following chapters:

- **Chapter 2: A Workflow for Low-Magnitude Event Detection, Hypocenter Estimation & Uncertainty Quantification**
  We introduce a comprehensive workflow for detecting and characterising low-magnitude seismic events in urban geothermal settings. We address two challenges: (i) the detection of weak seismic events in high-noise-level environments and (ii) the estimation of their hypocentres accounting for velocity-model uncertainties. We apply the workflow in Kwintsheul (Netherlands), an area where the KNMI seismic network has not detected seismic activity.
- **Chapter 3: Velocity-Model Validation Using Seismic Interferometry**
  We explore Seismic Interferometry (SI) as a tool to validate seismic velocity models. We leverage the geometry of Kwintsheul's PSM network and the hypocentre estimations to retrieve zero-offset seismic sections. We analyse the zero-offset sections and validate the seismic velocity model used for hypocentre estimations. Our approach is a cost-effective way to analyse and validate seismic velocity models using body waves of seismic events.
- **Chapter 4: Experimental Network Design for Seismic Monitoring Offshore**
  We implement a framework for designing PSM networks for monitoring geothermal-energy projects. We use a Bayesian Optimal Experimental Design (OED) framework to reduce the expected hypocentre uncertainties before any data has been collected.

To test the framework, we design a PSM network for monitoring the K-14 Offshore CCS project. The CCS field is located in the Dutch North Sea, an area without coverage of the KNMI's seismic network.

- **Chapter 5: Clock Error Correction with Ambient Seismic Noise: Application to the Reykjanes Geothermal Field**
  We address the challenge of synchronising OBS clocks for offshore geothermal monitoring. We introduce a method that synchronises OBS clocks using the time symmetry of noise cross-correlations. We apply this method to data from the IMAGE network, which monitored geothermal activity on the Reykjanes peninsula in Iceland.

- **Chapter 6: Conclusions and Future Work**
  We summarise our findings and their implications for seismic monitoring of geothermal-energy projects. We discuss the broader impact of the developed methodologies and highlight potential future research directions.

The thesis presents a complete PSM workflow for monitoring geothermal energy that can be applied in onshore and offshore environments. The workflow is summarised in Figure 1.3.

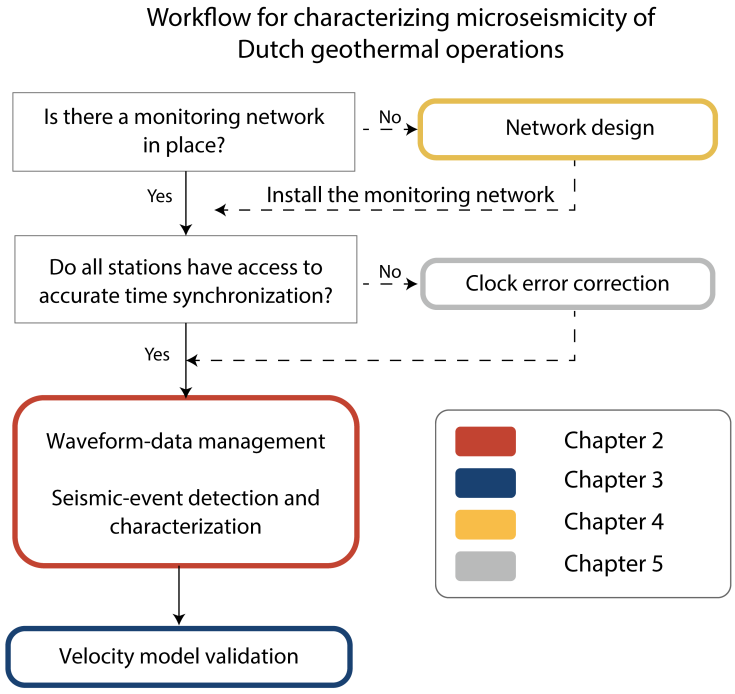


Figure 1.3: Workflow for characterising microseismicity in Dutch geothermal operations. The main sequence consists of waveform-data management, seismic-event detection, seismic-event characterisation, and velocity-model validation, forming the basis for interpreting microseismicity. If no monitoring network is in place or if stations lack accurate time synchronisation, additional preparatory steps such as network design and clock correction are required. These supporting steps are addressed in later chapters, as indicated.

REFERENCES

Bourne, S. J., Oates, S. J., Van Elk, J., and Doornhof, D. (2014). A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir. *Journal of Geophysical Research Solid Earth*, 119(12):8991–9015. https://doi.org/10.1002/2014jb011663.

Buijze, A. J. L. (2020). Numerical and experimental simulation of fault reactivation and earthquake rupture applied to induced seismicity in the Groningen gas field. [Doctoral dissertation, Utrecht University]. https://doi.org/10.33540/60.

Buijze, L., Fokker, P. A., and Wassing, B. (2021). Quantification of induced seismicity potential of geothermal operations. Technical Report 060.43190, Nederlandse Organisatie voor toegepast-natuurwetenschappelijk onderzoek. [Report]. https://resolver.tno.nl/uuid:2f977092-f026-482d-860e-19a1a7184c16.

Buijze, L., Van Bijsterveldt, L., Cremer, H., Paap, B., Veldkamp, H., Wassing, B. B.,

Van Wees, J.-D., Van Yperen, G. C., Ter Heege, J. H., and Jaarsma, B. (2019). Review of induced seismicity in geothermal systems worldwide and implications for geothermal systems in the Netherlands. *Netherlands Journal of Geosciences – Geologie en Mijnbouw*, 98. https://doi.org/10.1017/njg.2019.6.

Eaton, D. W. (2018). *Passive Seismic Monitoring of Induced Seismicity: Fundamental Principles and Application to Energy Technologies*. Cambridge University Press. [Book]. https://doi.org/10.1017/9781316535547.

Evans, K. F., Zappone, A., Kraft, T., Deichmann, N., and Moia, F. (2011). A survey of the induced seismic responses to fluid injection in geothermal and CO2 reservoirs in Europe. *Geothermics*, 41:30–54. https://doi.org/10.1016/j.geothermics.2011.08.002.

Fokker, P., Buijze, L., and Pluymaekers, M. (2023). SRIMA: Background information of the Python tool. Technical Report R11303, TNO Public report. [Report]. https://www.nlog.nl/sites/default/files/2023-11/tno2023_r11303_srima_background_information_of_the_python_tool.pdf.

Foulger, G. R., Wilson, M. P., Gluyas, J. G., Julian, B. R., and Davies, R. J. (2017). Global review of human-induced earthquakes. *Earth-Science Reviews*, 178:438–514. https://doi.org/10.1016/j.earscirev.2017.07.008.

Government of the Netherlands (2019). Climate agreement. Technical report, Ministry of Economic Affairs. https://www.government.nl/documents/reports/2019/06/28/climate-agreement.

Houtgast, R. and van Balen, R. (2000). Neotectonics of the roer valley rift system, the netherlands. *Global and Planetary Change*, 27(1):131–146. https://doi.org/10.1016/S0921-8181(01)00063-7.

Kamila, Z., Kaya, E., and Zarrouk, S. J. (2020). Reinjection in geothermal fields: An updated worldwide review 2020. *Geothermics*, 89:101970. https://doi.org/10.1016/j.geothermics.2020.101970.

KNMI (1993). Netherlands seismic and acoustic network. Royal Netherlands Meteorological Institute (KNMI). Other/Seismic Network. https://doi.org/10.21944/e970fd34-23b9-3411-b366-e4f72877d2c5.

KNMI (2023). Earthquakes - complete catalogue for the netherlands and near surrounding. Retrieved from https://dataplatform.knmi.nl/dataset/aardbevingen-catalogus-1.

Krevor, S., De Coninck, H., Gasda, S. E., Ghaleigh, N. S., De Gooyert, V., Hajibeygi, H., Juanes, R., Neufeld, J., Roberts, J. J., and Swennenhuis, F. (2023). Subsurface carbon dioxide and hydrogen storage for a sustainable energy future. *Nature Reviews Earth & Environment*, 4(2):102–118. https://www.nature.com/articles/s43017-022-00376-8.

Mathur, B., Hofmann, H., Cacace, M., Hutka, G. A., and Zang, A. (2024). Thermo-hydro-mechanical simulation of cooling-induced fault reactivation in Dutch geothermal reservoirs. *Netherlands Journal of Geosciences – Geologie en Mijnbouw*, 103. 10.1017/njg.2023.12.

Mijnlieff, H., Vries, S., Jaarsma, B., and Vogelaar, B. (2023). Seismic hazard and risk assessment for geothermal projects in the Netherlands. Technical report, TNO-AGE and EBN. https://www.nlog.nl/sites/default/files/2023-11/20230922_hoofdrapport_sdra_aardwarmteprojecten_-_shra_for_geothermal_projects.pdf.

Moein, M. J. A., Langenbruch, C., Schultz, R., Grigoli, F., Ellsworth, W. L., Wang, R., Rinaldi, A. P., and Shapiro, S. (2023). The physical mechanisms of induced earthquakes. *Nature Reviews Earth & Environment*, 4(12):847–863. https://www.nature.com/articles/s43017-023-00497-8.

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 – Geologie en Mijnbouw*, 101. https://doi.org/10.1017/njg.2021.14.

Pluymaekers, M., Doornenbal, J., and Middelburg, H. (2017). TNO 2017 R11014 with erratum page 67 Final, Velmod-3.1. *TNO report*. https://www.nlog.nl/sites/default/files/2018-11/060.26839%20R11014%20with%20erratum%20page%2067%20Doornenbal-final.sec_.pdf.

Raupach, M. R., Marland, G., Ciais, P., Quéré, C. L., Canadell, J. G., Klepper, G., and Field, C. B. (2007). Global and regional drivers of accelerating co2 emissions. *Proceedings of the National Academy of Sciences*, 104(24):10288–10293. 10.1073/pnas.0700609104.

Ruigrok, E., Kruiver, P., and Dos, B. (2023). Construction of earthquake location uncertainty maps for the netherlands. KNMI number: TR-405, Year: 2023, Pages: 158.

Schoof, F., van der Hout, M., van Zanten, J., and van Hoogstraten, J. W. (2018). Master Plan Geothermal energy in the Netherlands. Technical report, Stichting Platfor, Geothermie, DAGO, Stichting Warmtenetwerk, EBN. https://www.ebn.nl/wp-content/uploads/2023/05/108-LR_114337_EBN-Focus-Geothermal-Energy-ENG-V3.pdf.

Sircar, A., Bist, N., and Yadav, K. (2022). A comprehensive review on exploration and exploitation of offshore geothermal energy. *Marine Systems & Ocean Technology*, 17(3-4):135–146. https://doi.org/10.1007/s40868-022-00120-3.

TNO – Advisory Group for Economic Affairs (2022). Natural Resources and Geothermal Energy in the Netherlands. Annual Review 2022. Technical Report Annual review 2022, Directorate General of Groningen and Subsurface of the Dutch Ministry of Economic Affairs and Climate Policy, nl. https://www.nlog.nl/sites/default/

files/2023-09/annual_review_2022_-_natural_resources_and_geothermal_energy_in_the_netherlands.pdf.

Verdon, J. P., Kendall, J.-M., White, D. J., Angus, D. A., Fisher, Q. J., and Urbancic, T. (2010). Passive seismic monitoring of carbon dioxide storage at weyburn. *The Leading Edge*, 29(2):200–206. https://doi.org/10.1190/1.3304825.

2

A Workflow for Low-Magnitude Event Detection, Hypocenter Estimation & Uncertainty Quantification

If one believes in Bayes' theorem, scientific progress is inevitable as predictions are made and as beliefs are tested and refined.

Nate Silver

The detection and hypocentre estimation of seismic events determines how well we can monitor seismicity related to geothermal-energy operations. Their detection can provide valuable insights into reservoir behaviour, whereas their hypocentres indicate the presence of faults or induced seismicity. There are two significant challenges when monitoring the seismicity in Dutch geothermal operations: i) detecting signals from seismic events as noise levels are typically high in regions hosting geothermal operations, and ii) accurately estimating their corresponding hypocentre and uncertainty. In this chapter, we present a comprehensive workflow for the detection and characterisation of low-magnitude seismic events. We integrate data preparation, template-matching, and machine-learning-based event detection, probabilistic hypocentre estimation, and local magnitude calculation. We apply our workflow to four months of recordings in Kwintsheul, Netherlands, where a low-magnitude seismic event was reported in a previous study. With our workflow, we detect six weak seismic events (local magnitude $M_L < 0.0$) near a local fault and a geothermal injection well. These events suggest the presence of a recurring microseismic sequence. However, the nature of these events is unclear because of the uncertainties related to the seismic velocity model, the short monitoring period, and the limited azimuthal coverage of

This chapter has been published as a journal paper in the Netherlands Journal of Geosciences (Naranjo et al., 2025). Note that we changed parts of the original publication to make the text consistent with this thesis.

the PSM network. Our workflow offers a scalable solution for enhancing seismic monitoring, particularly in urban geothermal settings.

2.1. INTRODUCTION

The Netherlands is expanding the use of geothermal heat to reduce carbon emissions and support the transition to low-carbon heating systems. Understanding the subsurface processes associated with geothermal-heat production is essential to optimise resource extraction and ensure long-term sustainability. These processes, such as fluid migration and fault reactivation, are often accompanied by low-magnitude seismic events. Therefore, seismic monitoring is a key tool for characterising geothermal reservoirs.

In 2018, a geothermal doublet became operational in Kwintsheul, South Holland, to heat 64 hectares of greenhouses. At that time, the coverage of the Dutch Meteorological Institute's seismological network (KNMI, 1993) was still relatively sparse in the province of South Holland, as shown in Figure 2.1. In 2019, a temporary passive seismic network was installed to study the interaction between the geothermal operation and the underlying geological structures. During this monitoring period, a seismic event with a duration magnitude (M_d) of 0.16 was detected on July 14, 2019 (Muntendam-Bos et al., 2022). This event suggested the presence of more low-magnitude seismicity in the region, motivating us to investigate the feasibility of detecting additional events in the recorded data.

Monitoring low-magnitude seismicity around geothermal operations is challenging in the Netherlands. Many of these operations are located in noisy urban environments, where low-magnitude events are difficult to detect due to strong background noise (Groos and Ritter, 2009). This raises a fundamental question: how can we effectively monitor low-magnitude seismicity in such settings to better characterise subsurface processes?

We address this question by presenting a comprehensive workflow for detecting and characterising low-magnitude seismic events in urban areas. The workflow integrates data management, machine-learning-based event detection, and probabilistic hypocentre estimation, providing a portable approach to seismic monitoring. We apply this workflow to a four-month monitoring period in Kwintsheul and demonstrate its effectiveness in detecting previously unreported seismicity. Our findings highlight the importance of dense seismic arrays to enhance redundancy and azimuthal coverage, which is crucial for detecting and locating low-magnitude seismicity. The results provide insights into ongoing low-magnitude (M<0) seismic activity.

2.2. GEOLOGICAL CONTEXT AND SEISMICITY IN THE NETHERLANDS

The complete earthquake catalogue for the Netherlands and surrounding regions (KNMI, 2023) indicates that natural seismicity is concentrated in the Roer Valley Graben (RVG), depicted as red circles in Figure 2.1. These events are associated with the northeastern extension of the active faults bounding the Lower Rhine Graben (Houtgast and van Balen, 2000). Induced seismic activity is concentrated in the northern Netherlands, particularly in the Groningen region, linked to gas extraction (Muntendam-Bos et al., 2022). These induced events, shown as blue circles in Figure 2.1, are typically low in magnitude ($M_L \leq 4.0$).

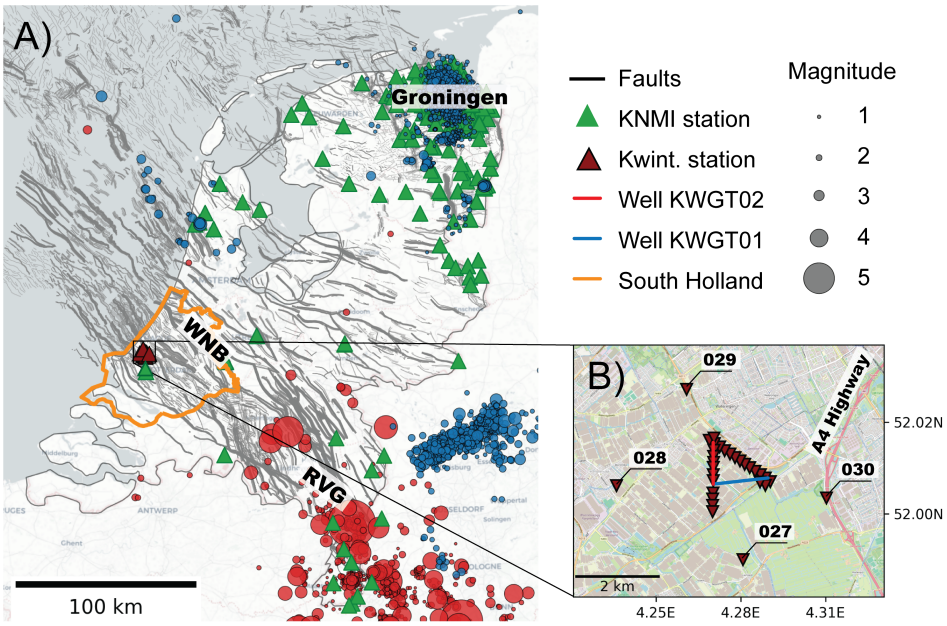


Figure 2.1: A) Regional distribution of seismicity in the Netherlands, main structural features, and location of the Kwintsheul geothermal doublet. The epicentres from the complete KNMI earthquake catalogue (1920–2021) (KNMI, 2023) are shown in red (natural seismicity) and blue (induced seismicity) circles. The stations of KNMI's permanent seismic network (operational from July 22, 2019, to November 9, 2019) are shown in green (KNMI, 1993). The West Netherlands Basin (WNB), Groningen region, and Roer Valley Graben (RVG) are highlighted, and the individual faults (v. Gessel et al., 2021) are shown in grey. The orange polygon delineates the province of South Holland, where the Kwintsheul geothermal doublet is located (black square). B) Enlarged view of the Kwintsheul geothermal-heat-production site. The injector well (KW-GT-01) is shown in blue, and the producer well (KW-GT-02) is shown in red. The temporal monitoring seismic stations are shown in dark red triangles. These were operational from July 22 to November 9, 2019. The background map tiles are provided by OpenStreetMap (OpenStreetMap contributors, 2017)

Kwintsheul is located on top of the West Netherlands Basin (WNB) in the province of South Holland. To this date, no seismic activity has been detected in this region by the Dutch seismic-monitoring network (KNMI, 2023). The WNB is a 60-km-wide transtensional basin that forms part of a failed rift system (Boersma et al., 2021). It contains NW-SE-oriented normal faults and consists of Permian to Tertiary deposits that reach thicknesses of up to 5 km, with a well-connected fault system throughout (Duin et al., 2006; Worum et al., 2005; Boersma et al., 2021).

The location of the geothermal doublet is depicted in Figure 2.1.B. The injection and production wells reach a depth of approximately 2300 m, targeting the Delft Sandstone Member. The separation between the two wells is approximately 1500 m at reservoir depth. Until 2020, the Royal Netherlands Meteorological Institute (KNMI) reported a magnitude of completeness (M_c) of 1.0 around Kwintsheul. Earthquakes below this threshold are likely to remain undetected (Muntendam-Bos et al., 2022; Ruigrok et al., 2023)[1].

[1]Following a reanalysis, KNMI identified that the magnitude-of-completeness estimate was based on detections

2.3. WORKFLOW FOR SEISMIC MONITORING

To characterise seismicity around Kwintsheul's geothermal operation, we adopted a multi-stage workflow designed to manage seismic data, detect seismic events, and characterise their hypocentres and magnitudes (Figure 2.2). The workflow begins with data preparation, which includes managing seismic-waveform data, station information, and retrieving local velocity models. Next, event detections are performed using a combination of template-matching and machine-learning-based detection techniques. Finally, the detected events are characterised by estimating their hypocentres and magnitudes, accounting for uncertainties in both observed and theoretical phase arrival times. The following sections provide a detailed description of each stage in the workflow.

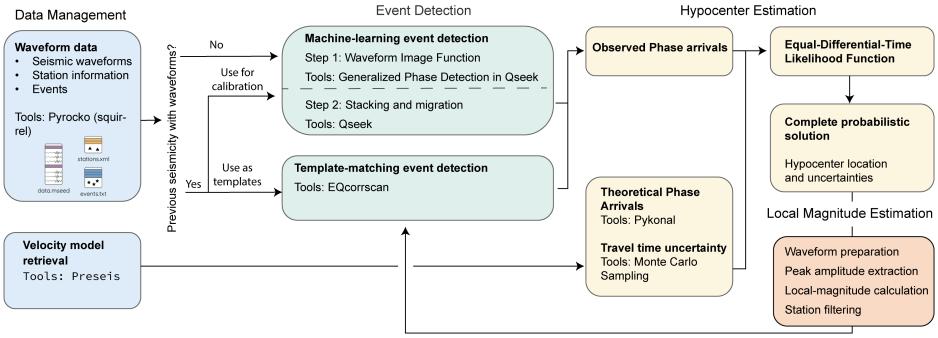


Figure 2.2: Workflow for seismic-event detection, location, and characterisation. The process is divided into four main stages: Data management, Event Detection, Hypocenter Estimation, and Local Magnitude Estimation.

2.3.1. DATA MANAGEMENT

WAVEFORM DATA

Our workflow begins with preparing seismic waveform data and station metadata. We use the `Squirrel` package from the Pyrocko toolbox (Heimann et al., 2017) to manage these datasets. Its metadata caching allows quick inspection, helping us detect and resolve issues such as gaps or timing errors.

VELOCITY-MODEL RETRIEVAL

As part of the data management stage, we also retrieve the velocity model required for hypocentre inversion. In this study, we use Velmod (3.1), the Dutch regional seismic velocity model (Pluymaekers et al., 2017). Velmod is a 3D velocity model that integrates velocities measured in boreholes (sonic logs and check-shot data) with stacking velocities derived from seismic surveys. It provides a 3D distribution of seismic velocities and includes estimates of their uncertainties. Further details regarding the parameterisation of Velmod 3.1 are provided in Section 2.3.3. To retrieve the local velocity cube from Velmod, we use `PRESEIS` (Kraaijpoel, 2025), which extracts the region of interest and outputs the

at only three stations, while their operational software required at least five stations for reliable event detection. As a result, the actual detection threshold was higher, likely above $M_L = 1.5$.

model as a multidimensional array in `Xarray` format (Hoyer and Hamman, 2017). The next stage of the workflow involves event detection and preliminary locations.

2.3.2. EVENT DETECTION

TEMPLATE-MATCHING DETECTION—EQCORRSCAN

As the waveforms of a previously detected seismic event are available, we use template matching (TM) to identify additional events with similar waveform characteristics. In this method, previously detected events serve as templates, which are cross-correlated with continuous waveform recordings from multiple seismic stations to find matching events. Events identified through this process are known as repeating microearthquakes, as they exhibit high waveform similarity across different stations (e.g., Menke, 1999; Ellsworth, 1995). These events often originate from a concentrated volume of hypocentres and provide valuable insights for monitoring induced seismicity.

Template matching was performed using the open-source Python package `EQcorrscan` (Chamberlain et al., 2017). The preprocessing parameters and threshold values for the template-matching are presented in Section 2.5.1.

MACHINE-LEARNING EVENT DETECTION

Machine-learning event detection involves two steps. First, we identify the onset times of seismic phases, such as P- and S-waves, by distinguishing their unique signal attributes from background noise. Second, we associate these seismic phases across multiple stations with a single seismic event. These steps are explained in detail below, and the preprocessing parameters are introduced in Section 2.5.

STEP 1: WAVEFORM IMAGE FUNCTION

Deep-learning phase-picking models are trained on large datasets of manually labelled seismic phases (Soto and Schurr, 2021; Mousavi et al., 2020). These models convert seismic waveforms into time series known as image functions (or characteristic functions), which quantify the probability of a P-phase, S-phase, or noise being present in the waveform (Zhu and Beroza, 2019). An example image function generated from single-station waveforms is shown in Figure 2.6 in Section 2.5.1. A phase detection is declared if the probability is above a certain threshold.

We select the Generalised Phase Detection (GPD) model, which was trained to detect P- and S-phases from events with magnitudes between -0.81 and 5.7 and epicentral distances of less than 100 km (Ross et al., 2018). The model was trained using hand-labelled data archives from the Southern California Seismic Network (California Institute of Technology and United States Geological Survey Pasadena, 1926), which included 1.5 million P- and S-wave seismograms and an equal number of 4-second noise windows.

STEP 2: STACKING AND MIGRATION—QSEEK

To associate the neural network's phase annotations with a coherent seismic source, we apply the stacking and migration approach implemented in the `Qseek` software (Isken et al., 2025). Figure 3 illustrates the stacking and migration approach, where the seismograms are annotated with the first arrivals of the P- and S-phases using a pre-trained neural network. The magnitude of each annotation represents the model's certainty in the phase arrival.

These annotations are backprojected onto a subsurface grid according to theoretical travel times and stacked at each node. The node with the most constructive stack corresponds to the maximum semblance (i.e., coherence of arrivals). If this value exceeds a predefined threshold, the algorithm indicates a detection and assigns the corresponding node as the most likely location of the seismic event.

We describe the selection of preprocessing parameters and detection thresholds in Section 2.5.1.

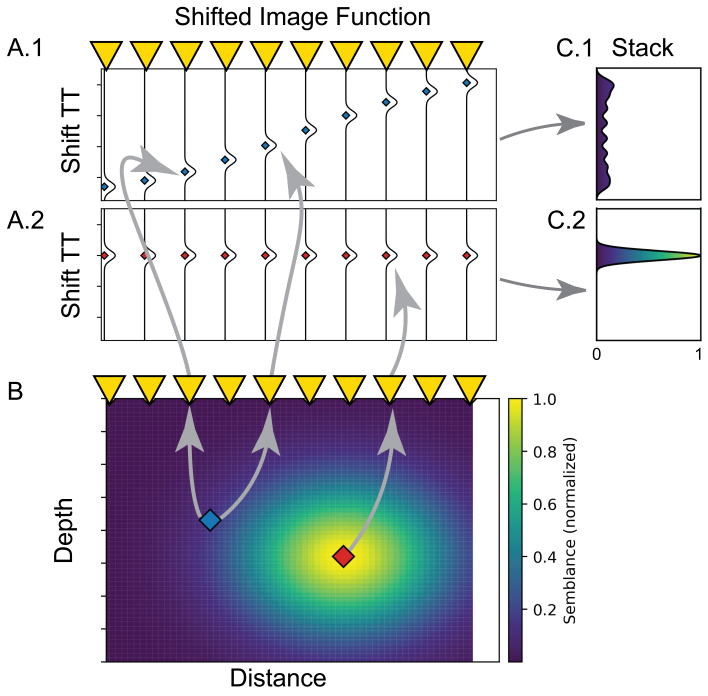


Figure 2.3: Conceptual 2D illustration of the stacking-and-migration procedure used in Qseek (Isken et al., 2025). A.1 Positive image functions recorded at the receiver array (yellow triangles) after shifting each trace by the predicted P-wave travel time from the blue candidate node shown in B. The pulses remain misaligned, and the resulting stack (C.1) displays low coherence. A.2 Image functions shifted by the travel times from the red candidate node in B. The shifts align the pulses coherently, producing a sharp, high-amplitude peak in the corresponding stack (C.2). B. Two-dimensional search grid beneath the array. The colourmap denotes the maximum normalised semblance obtained at each node. The blue and red squares indicate the two candidate nodes evaluated in panels A.1 and A.2.

2.3.3. HYPOCENTRE ESTIMATION

Hypocentre inversion aims to estimate the spatial location $\mathbf{x_s}$ (often-times together with the origin time T_0) at which energy is released during a seismic event using the arrival times of seismic phases at various seismic stations. It involves solving both the forward and the inverse problem. In the forward problem, theoretical seismic-phase arrival times are computed, simulating the travel times of seismic waves originating from a (potential)

hypocentre to various receiver stations. The inverse problem seeks to identify the solution(s) that minimise the differences between theoretical phase arrivals (t_{calc}) and observed phase arrivals (t_{obs}). Note that different hypocentres may satisfy the fitting criterion, and small perturbations in the input data (e.g., arrival times) can lead to significant changes in the inferred hypocentre. Therefore, we are interested not only in the model parameters that best fit the observed data but also in the overall volume of solutions and their associated uncertainties.

OBSERVED PHASE ARRIVALS AND ASSOCIATED UNCERTAINTY

An observed phase arrival corresponds to the (picked) onset time of the first measurable seismic energy recorded on a seismogram. Due to potential errors in picking the onset time, it is more accurate to represent a phase arrival as a probability density function rather than a scalar value. Following Lomax et al. (2009), we use a normal distribution to describe an observed phase arrival. The key parameters for the hypocentre estimation are the mean of the distribution, t_{obs}, and the corresponding standard deviation, σ_{obs}. Figure 2.4 illustrates how these terms are derived from the picking process and the corresponding observed phase-arrival distribution. We manually reviewed and selected the phase arrival times and their associated uncertainty.

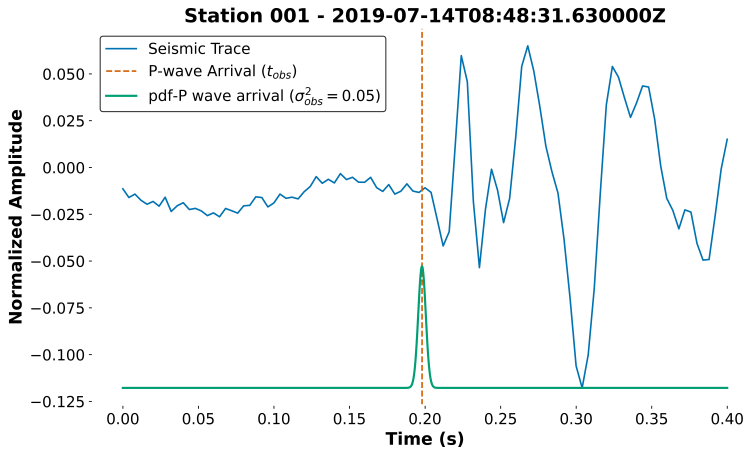


Figure 2.4: Illustration of a p-phase arrival time and associated picking uncertainty. The blue line represents a seismic trace, while the green line is a normal distribution that describes the uncertainty of the phase arrival time. Note that the normal distribution is scaled for visual clarity, and its integral equals one. The label on the vertical axis only applies to the seismic trace amplitude.

THEORETICAL PHASE ARRIVAL TIMES

Estimating the hypocentre requires a model to compute theoretical phase arrival times, which are then compared with the observed phase arrival times. The theoretical phase arrival, t_{calc}, results from the solution of the forward problem. To compute t_{calc}, we require the receiver's location, a hypocentre location (our guess), and the forward function u describing the propagation time from the hypocentre's location to the receiver's location.

The theoretical phase arrival time is defined as

$$t_{calc}(\mathbf{x_s}, \mathbf{x_r}, \mathbf{t_0}, \mathbf{v_{model}}) = t_0 + u(\mathbf{x_s}, \mathbf{x}_r, \mathbf{v_{model}}). \tag{2.1}$$

In this study, we adopt an infinite frequency approximation (Lin and Ritzwoller, 2011). Specifically, u exploits the fast Marching Method (FMM) (Sethian and Popovici, 1999), implemented in the `PyKonal` Python Package (White et al., 2020). Besides being contingent on the validity of the infinite frequency approximation, the accuracy of the estimated travel time depends strongly on the accuracy of the velocity model $\mathbf{v_{model}}$.

TRAVEL-TIME UNCERTAINTY

The accuracy of the velocity model $\mathbf{v_{model}}$ directly affects the accuracy of the theoretical phase arrival times, which in turn impacts the reliability of the hypocentre inversion. The velocity model should closely represent accurate velocities in the subsurface and, as such, correctly account for the interaction of the phases with the different lithostratigraphic units and their geometry. Variations in subsurface properties, such as layer thicknesses or velocity gradients, can lead to deviations between the theoretical and observed phase arrivals. To account for these variations, the velocity model must incorporate both mean velocity values and measures of variability, such as standard deviations or probability distributions. These can be used as inputs to estimate the uncertainties of the theoretical phase arrivals.

The following steps detail the procedure we use to estimate the standard deviation of σ_{calc}, which is subsequently used in the likelihood function:

1. **Retrieve the velocity profile**: Select a 1D velocity profile with well-defined mean velocity values and standard deviations for each geological unit (see Section 2.4.2). For this study, the profile is taken from the surface location of Kwintsheul's geothermal doublet.

2. **Monte Carlo Sampling**: Use the mean velocity values and their standard deviations to generate multiple realisations of the velocity model, perturbing the mean velocities by sampling from their respective distributions.

3. **Recalculate theoretical phase arrivals**: For each perturbed velocity model, recompute the t_{calc} at the maximum likelihood hypocentre.

4. **Quantify variability**: Compute the standard deviation of the recalculated t_{calc} values across all realizations. This standard deviation represents σ_{calc}.

The resulting σ_{calc} quantifies the uncertainty in the theoretical phase arrival times caused by variability in the velocity model. This value is subsequently used in the likelihood function to account for travel-time uncertainties in hypocentre estimation. It is important to note that we did not consider potential correlations of the calculated travel times between different stations. Ideally, this would have been the case, but to limit the computational burden, we chose not to do so.

EQUAL-DIFFERENTIAL-TIME LIKELIHOOD FUNCTION

A likelihood function is used to quantify how likely it is that a specific hypocentre could explain the observed phase arrival times. The Equal-Differential-Time formulation (Font et al., 2004) is an approximation that allows removing the origin time from the unknown parameters. It is also robust in case of outliers (Lomax et al., 2009). If a hypocentre $\mathbf{x_s}$ is perfectly determined within the Earth model, and in the absence of noise, the time difference between the calculated phase arrivals at two seismic stations a and b should be equal to the difference between their corresponding observed phase arrivals, i.e.,

$$t^a_{calc} - t^b_{calc} = t^a_{obs} - t^b_{obs}. \tag{2.2}$$

A reliable hypocentre estimation requires accounting for the uncertainties related to the theoretical phase arrivals and observed phase arrivals. Accounting for these terms, the likelihood of the hypocentre becomes

$$L(\mathbf{x_s}) = \left[\sum_{a,b} \frac{1}{\sqrt{\sigma_a^2 + \sigma_b^2}} \cdot \exp\left(- \frac{([t^a_{obs} - t^b_{obs}] - [t^a_{calc}(\mathbf{x_s}) - t^b_{calc}(\mathbf{x_s})])^2}{\sigma_a^2 + \sigma_b^2} \right) \right]^N, \tag{2.3}$$

where $L(\mathbf{x_s})$ estimates the Gaussian likelihood of the hypocentre $\mathbf{x_s}$. The terms t^a_{obs} and t^b_{obs} are the observed arrival times at seismic stations a and b, while $t^a_{calc}(\mathbf{x_s})$ and $t^b_{calc}(\mathbf{x_s})$ are the corresponding modelled arrival times. The variances of the observed and theoretical phase arrivals of stations a and b are denoted by σ_a^2 and σ_b^2, respectively.

For stations a and b, the total uncertainty is expressed as

$$\sigma_a^2 = \sigma^2_{\text{obs},a} + \sigma^2_{\text{calc},a}, \quad \sigma_b^2 = \sigma^2_{\text{obs},b} + \sigma^2_{\text{calc},b}, \tag{2.4}$$

where $\sigma^2_{\text{obs},a}$ and $\sigma^2_{\text{obs},b}$ represent the uncertainties in the observed phase arrivals at stations a and b, respectively, and $\sigma^2_{\text{calc},a}$ and $\sigma^2_{\text{calc},b}$ correspond to the uncertainties in the theoretical phase arrivals. These combined uncertainties are used in calculating the likelihood function $L(\mathbf{x_s})$, which measures the agreement between the observed and theoretical arrival times for a given hypocentre $\mathbf{x_s}$.

$L(\mathbf{x_s})$ reaches its maximum value when the differentials of the observed and theoretical phase arrivals are equal, hence the term Equal Differential Time (EDT). Since the summation over observations occurs outside the exponential, the EDT probability density function (PDF) attains its highest values at locations where the most observation pairs are satisfied, making it robust against outliers (Lomax et al., 2009). Additionally, the EDT PDF is independent of the earthquake's origin time.

COMPLETE PROBABILISTIC SOLUTION—POSTERIOR SAMPLING

Tarantola and Valette (1982) introduced a general methodology to obtain a complete probabilistic solution of a seismic event's hypocentre by computing the posterior probability distribution (PPD)

$$\pi(\mathbf{m}) = \kappa \rho(\mathbf{m}) L(\mathbf{m}). \tag{2.5}$$

Here, $\mathbf{m}$ is the model parameter vector, which includes the hypocentre's spatial coordinates (x, y, z). In this equation, κ is a normalisation constant that ensures the posterior integrates to one, $\rho(\mathbf{m})$ is the prior PDF representing prior knowledge of the model parameters, and $L(\mathbf{m})$ is the likelihood function representing how well the observed data fits a given set of model parameters. In our case, the likelihood is estimated through equation (2.3), which effectively removes the origin time from $\mathbf{m}$ (as explained in Section 2.3.3). It should be understood that this likelihood is an approximation of the true Gaussian likelihood function in the sense that the correlation between different arrival time differences is ignored. In principle, these can be accounted for using a covariance matrix (Spetzler et al., 2024). This, however, would require us to limit the number of arrival-time differences to $N-1$ (where N is the number of stations for which a P-wave arrival-time pick is available). In this study, we choose to adopt the likelihood proposed by Lomax et al. (2009), which is based on all arrival-time differences $(N(N-1)/2)$.

In principle, any set of model parameters can be arbitrarily chosen to calculate theoretical phase arrivals. However, prior information, often based on fundamental laws or physical constraints, can indicate whether a set of parameters is feasible. Here, we use a uniform prior probability density that assumes that any model parameter has an equal probability of explaining the observed data:

$$\rho(\mathbf{m}) = const. \tag{2.6}$$

I.e., we assign equal a priori probabilities to equal volumes. Note that the prior is not completely uninformative, as we still choose a search area.

The spatial domain must be discretised into a grid to perform numerical computations. Each grid cell represents a surrounding region $(\Delta x \Delta y \Delta z)$, which must be sufficiently small to ensure the probability distribution remains approximately constant within the cell (Mosegaard and Tarantola, 1995). Consequently, the probability associated with a point in the grid denotes the probability that the hypocentre falls within the surrounding region represented by the grid cell.

The marginal PDF of the hypocentre can be estimated by integrating over specific parameters in the model parameter space. The marginal PDF of the epicentral location can be obtained by integrating over the depth parameter z as:

$$\pi_{xy}(x, y) = \int_{z_{\min}}^{z_{\max}} \rho(\mathbf{m}) L(\mathbf{m})\, dz, \tag{2.7}$$

where the limits of integration represent the plausible range of hypocentral depths.

The marginal PDF for the depth, $\pi_z(z)$, can be obtained by integrating over the epicentral coordinates x and y from the full posterior distribution. This process effectively reduces the three-dimensional posterior PDF to a one-dimensional distribution along the depth axis, which allows us to understand the uncertainty specifically related to the depth of the hypocentre. The depth PDF is computed as:

$$\pi_z(z) = \int_{x_{\min}}^{x_{\max}} \int_{y_{\min}}^{y_{\max}} \rho(\mathbf{m}) L(\mathbf{m})\, dx\, dy, \tag{2.8}$$

where the integration bounds represent the plausible spatial extent of the epicentre.

The spatial uncertainty of the seismic events is quantified by analysing the depth's marginal PDF π_z. We define the uncertainty range as the interval where the likelihood

values exceed 95% of the maximum likelihood, ensuring it encompasses the most probable depth values. We refer to this range as the 95% confidence interval.

To estimate this range, we follow these steps:

1. Identify the maximum likelihood value, $L_{\max}$, from the 3D posterior distribution.

2. Define a threshold at 95% of $L_{\max}$, above which depth solutions are considered significant.

3. Calculate the mean likelihood over x and y. The depth uncertainty range is determined as the interval $[z_{\min}, z_{\max}]$ where the mean likelihood exceeds the threshold.

4. The depth uncertainty is given by $\Delta z = z_{\max} - z_{\min}$.

The probabilistic hypocentre determination provides the spatial coordinates and depth of each event, along with their associated uncertainties. These results serve as essential inputs for the next step: estimating event magnitudes.

2.3.4. LOCAL MAGNITUDE ESTIMATION

We estimate the local magnitude (M_L) of the events following the methodology described by Dost et al. (2004). This approach uses the observed peak amplitudes from seismic waveforms, corrected for distance-dependent attenuation. The attenuation correction factor was calibrated using a set of induced earthquakes at approximately 3 km depth in the Groningen area (Dost et al., 2004).

The procedure for estimating the M_L involves the following steps:

1. **Waveform preparation**: Seismic waveforms are detrended and preprocessed using instrument-response corrections to obtain Wood-Anderson-equivalent amplitudes. We use only horizontal components of the waveform, corresponding to S-wave arrivals, for magnitude estimation.

2. **Peak amplitude extraction**: For each horizontal trace, the absolute peak amplitude (A_{WA}) is identified. Here, A_{WA} is the maximum averaged horizontal-displacement amplitude of a simulated Wood-Anderson instrument, expressed in millimetres.

3. **Local-magnitude calculation**: The local magnitude for each station is computed as

$$M_L = \log_{10}(A_{WA}) + 1.33\log_{10}(R) + 0.00139R + 0.424 \tag{2.9}$$

   (Dost et al., 2004), where A_{WA} is the observed peak amplitude of the simulated Wood-Anderson displacement, and R is the hypocentral-distance correction term.

4. **Station filtering**: Magnitudes are calculated for all stations that recorded the event, and the median magnitude is used as the event's M_L. Stations with anomalous deviations from the event-wise mean are excluded.

2.4. DATA

2.4.1. WAVEFORM DATA

The temporary seismic network in Kwintsheul was operational from July 22, 2019, to November 9, 2019 (Muntendam-Bos et al., 2022; Naranjo et al., 2022). The network consisted of 30 three-component force-balance Seismotech geophone sensors, which recorded at a bandwidth of 0.2 to 100 Hz at 250 sps. These were installed on the surface. . The geometry of the network consisted of two intersecting lines, each composed of 13 stations, covering an area of approximately 3.8 km^2. These stations were installed with an average in-line spacing of 150 meters. The layout was designed to record ambient seismic noise for consecutive illumination analysis and application of body-wave seismic interferometry (e.g., Panea et al., 2014). Additionally, an outer ring of four peripheral stations surrounded the array, encompassing an 18 km^2 area around the injection point of the geothermal doublet. The goal of the peripheral stations was to increase the azimuthal coverage for the location of events and a more robust estimation of hypocentral depths. The network's layout and the location of the geothermal doublet are shown in the inset of Figure 2.1.

As explained in Section 2.3.1, we assess the completeness of the available seismic waveform data. Figure 2.5 provides an overview of the data, with vertical lines indicating gaps in the data that directly affect the detection results and hypocentre estimations (see Section 2.5).

The most significant data gaps are observed at stations 027, 029, and 030. These stations are part of the outer ring designed to improve azimuthal coverage and extend the array's reach (Figure 2.1). Specific details regarding these gaps include:

- Station 029 stopped recording on July 18, 2019.
- Station 027 stopped recording on August 16, 2019.
- Station 030 experienced intermittent gaps and stopped recording on September 29, 2019.

The effects of these gaps on detection capabilities and hypocentre estimations are further detailed in Section 2.5.

2.4.2. SEISMIC-VELOCITY MODEL

We compute theoretical phase arrivals and travel-time uncertainties (Sections 2.3.3 and 2.3.3) using Velmod 3.1 (Pluymaekers et al., 2017). Velmod provides both the mean velocities and corresponding uncertainties for each geological unit. It is parameterised as follows:

$$V_{inst}^{unit}(z) = \mathcal{N}(V_0^{unit}, \sigma_V^{unit}) + k \cdot z, \tag{2.10}$$

where $V_{inst}^{unit}(z)$ is the P-wave velocity at depth z, V_0^{unit} is the mean velocity at the top of the unit, σ_V^{unit} is the standard deviation of the velocity, and κ is the velocity gradient within the unit. For additional details on retrieving the velocity model, refer to Section 2.3.1.

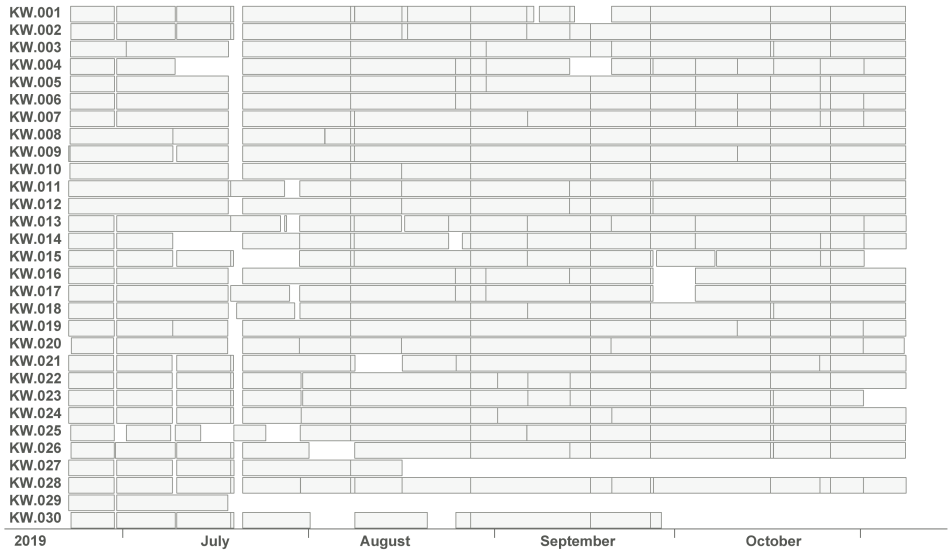


Figure 2.5: Overview of seismic waveform data visualized using `Snuffler` (Heimann et al., 2017). Continuous colours indicate sections without gaps, while colour changes highlight interruptions in the data.

2.5. RESULTS

2.5.1. EVENT DETECTIONS

The magnitude M_d 0.16 event reported in Muntendam-Bos et al. (2022) raised the question of whether more seismicity occurred below Kwintsheul. We, therefore, applied the template-matching detection routine (see Section 2.3.2) using the waveforms from the reported event as templates. To enhance the signal, we apply a band-pass filter between 4 Hz and 20 Hz and use templates with a length of 1 s following the P- and S-wave phase arrival time. We use templates for each station channel (i.e., P-wave for vertical components and S-wave for horizontal components). As a result, we identified five additional seismic events with near-identical waveforms. Together with the event reported in Muntendam-Bos et al. (2022), this gives six events.

Building on the six identified seismic events, we calibrate the machine-learning detection routine. Among several deep-learning arrival-time picking models, we select the Generalised Phase Detection (GPD) model (Ross et al., 2018), whose training dataset includes low-magnitude events (see Section 2.3.2 for details). To ensure consistency with the features learned by the GPD model, we apply the same preprocessing: a high-pass filter at 2 Hz and resampling to 100 Hz. Note that these processing parameters deviate from those of the template-matching routine. We select P- and S-pick thresholds of 0.3 based on the analysis of image functions for the identified events. Figure 2.6 shows an example

image function generated using these threshold values. When the image function exceeds the threshold, a phase arrival is annotated in the time series.

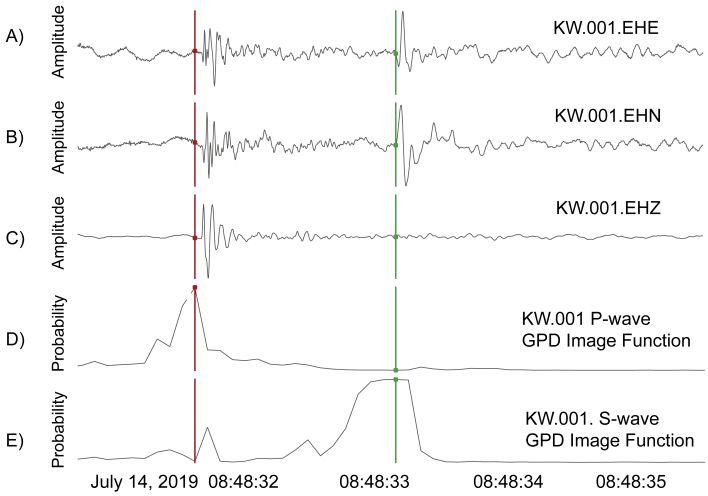


Figure 2.6: Seismic waveforms of the A) East, B) North, and C) vertical components from station 001 of the temporary seismic array installed in Kwintsheul. (D, E) Corresponding annotated image functions of P- and S-waves using the GPD model with P- and S-pick thresholds of 0.3.

The annotated image functions are the input for the stacking and migration approach explained in Section 2.3.2. To improve the accuracy of stacking and migration, we define a 1D velocity model retrieved at the location of the injection well. This velocity model is used by `Qseek` to estimate the travel-time shifts (see Figure 2.3.A). We select a maximum semblance threshold of 0.8, which enabled us to detect five of the six events used for calibration.

In total, the machine-learning detection routine identified 65 events that exhibit coherent seismic signals across the seismic array. We classified these events based on their waveform characteristics and spatial distribution. The detection statistics are summarised in Figure 2.7, and further details on each category can be found in Section 2.5.2 below.

A limitation of the machine-learning detection routine is the disproportionately large number of detections produced when fewer sensors are recording. After the association step outlined in Section 2.3.2, the detection workflow yielded 17,108 detections. However, 16,812 occurred on just six specific dates when station coverage was particularly low (< 50%). These dates and the number of sensors operating at those times are summarised in Table 2.1.

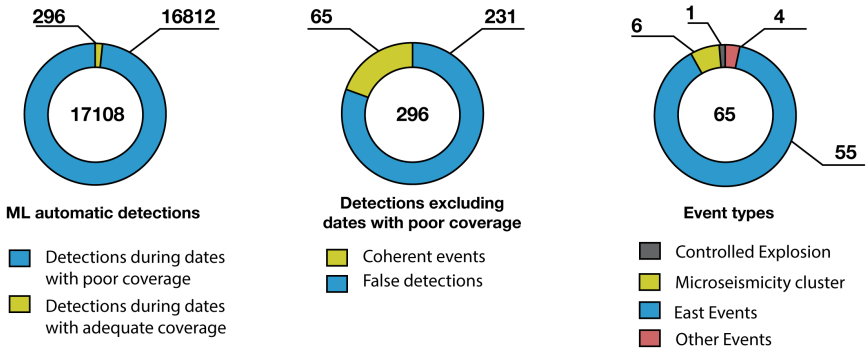


Figure 2.7: Summary of machine-learning detection outcomes. The left chart categorises the trigger detections based on station coverage. In blue, detections during periods of poor station coverage (<50%); in green, detections during adequate station coverage (>50%). The middle chart highlights the filtered dataset (after discarding the events on poor-station-coverage dates), displaying coherent events (green) versus false (incoherent) detections (blue). The right chart classifies these coherent events by event characteristics (green = microseismic cluster at 2.3 km depth, blue = events to the east, grey = controlled explosion, and red = other events). These groups are explained in Section 2.5.2.

Table 2.1: Summary of days with fewer than 16 operational stations and elevated detection numbers.

Date	Number of available stations	False Detections
June 29, 2019	6	2786
July 9, 2019	15	36
July 18, 2019	13	119
July 19, 2019	4	4386
July 20, 2019	5	8409
November 8, 2019	7	1076

2.5.2. WAVEFORM CHARACTERIZATION

The 65 events detected in Kwintsheul can be categorised based on their waveform attributes into i) microseismic cluster, ii) eastern events, iii) other events, and iv) a controlled explosion.

1. **Microseismicity cluster:** corresponds to six events that exhibit clear, impulsive P- and S-wave arrivals with frequencies between 5 and 50 Hz (see Figure 2.8) and P to S-wave delays of approximately 1.8 seconds. The character of the phase arrivals is impulsive, with a clear, sudden onset. P-waves are most prominent on the vertical component recordings, whereas the S-waves are most pronounced on the horizontal components.

2. **Eastern Events:** correspond to 55 coherent events coming from the East of the array. These events exhibit frequencies from 1 to 25 Hz for P-waves and 1 to 12 Hz for the

later arriving waves (delayed by approximately 4 seconds with respect to the impulsive P waves; see Figure 2.9.A and 2.9.B). In an attempt to estimate their hypocentres, we found that they appear to originate at or close to the Earth's surface. Since there is no mechanical argument for having earthquakes originate at depths less than approximately 200 m (unconsolidated sediments do not allow for seismic stress release), we believe that they, in fact, originated at the Earth's surface. The lack of a clear S-wave arrival supports this conclusion. We did not include the hypocentre inversion results because the high azimuthal gap (>250) renders the posterior very broad (although it is clear that they come from the East).

3. **Other Events:** correspond to three events that exhibit three pulses of low-frequency waves in the vertical component. The spectrogram analysis (Figure 2.9.C and 2.9.D) reveals that the energy is predominantly concentrated along the vertical component, with frequencies ranging between 5 and 10 Hz. The energy travels primarily vertically as the signals arrive almost simultaneously at all stations. There is, furthermore, no clear evidence of an S-wave associated with these events.

4. **Controlled Explosion:** corresponds to a controlled explosion in the North Sea on 2019-10-15 at 17:47:11, also reported in the International Seismological Centre's On-Line Bulletin (ISC, 2025) with Event ID 618929881.

2.5.3. HYPOCENTRES AND MAGNITUDES

To investigate the spatial distribution of the microseismicity cluster, we derived their maximum-likelihood hypocentre solutions with 95% confidence intervals as error bars shown in Figures 2.10.A, 2.10.B, and 2.10.C. The hypocentres of the microseismicity cluster are located at approximately 2.3 km depth, with vertical uncertainties ranging from 267 to 735 m (Table 2.2).

Events on 2019-09-11, 2019-10-02 and 2019-10-06 have different depth hypocentre values (2.1 km), as shown in Table 2.2. These events lack an observed phase arrival recorded at any peripheral station. This increased their azimuthal gap, resulting in a higher depth uncertainty. The influence of peripheral stations on hypocentre confidence intervals is illustrated in Appendix A.1 (Figure A.1).

To contextualise the hypocentre solutions with the known faults in the Kwintsheul area, we compare them in Figure 2.10.D against the mapped faults, the geothermal doublet, and the seismic stations. The interpretation of these mapped faults is derived from a regional 3D seismic dataset reprocessed in 2012, covering approximately 1200 km^2 in the West Netherlands Basin (Merrifield, 2012), and was provided by Peeters (2025). The six seismic events lie close to a local fault as well as the injection well, as shown in Figure 2.10.D. We estimated the minimum distance from each event to the nearest point on the closest fault, finding values ranging from 265 m to 434 m. Since the uncertainties in the X, Y, and Z coordinates are high, it is not possible to confirm whether the events originated on the fault. However, their spatial proximity makes an origin on the fault plausible. Activation of smaller fractures near the injection well is also feasible, although this is considered less likely in permeable sandstones, where pre-existing fractures are less common than in more brittle lithologies such as carbonates or granites. Nevertheless, thermally-induced stresses could potentially trigger such minor fractures.

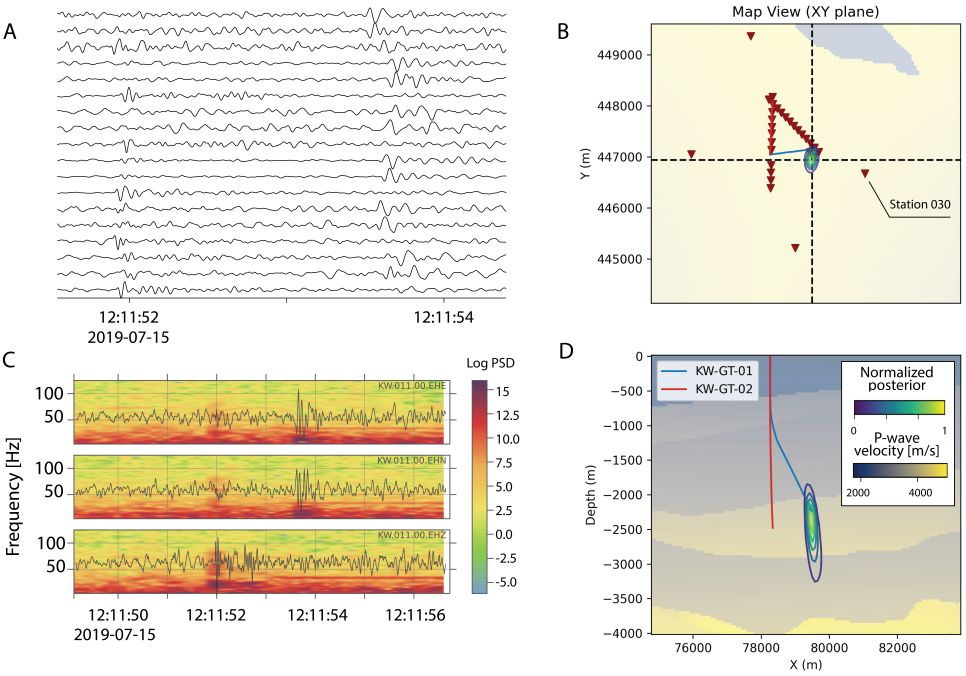


Figure 2.8: Analysis of the seismic event on 2019-07-15T12:11:51. (A) Waveforms of the event band-pass filtered between 4 and 20 Hz. (B, D) Map views and cross-sections showing the probability distribution of the event location are displayed on the Viridis colour scale, with green indicating the maximum likelihood location. (C) Spectrograms representative of other stations, exemplified here for station 023, show the vertical, north, and east components. The colour scale represents the logarithm of the power spectral density, with red indicating higher values.

The local magnitudes of the microseismicity cluster range from -1.60 to -2.52, computed with the site-specific local-magnitude formula given in Equation 2.9. The deviation of station-specific M_L's with respect to the median varies from 0.13 to 0.18 M_L, which shows a relatively low variability among the different stations. This is not surprising, given that the stations are relatively close to each other.

Event UTC	**x [m]**	**y [m]**	**z [m]**	**Az. gap**	M_L	Var(M_L)	σ_x **[m]**	σ_y **[m]**	σ_z **[m]**
2019-06-23 08:02:08	79483.25	447004.24	-2314.38	93	-2.04	0.13	602	363	267
2019-07-14 08:48:31	79483.30	446967.93	-2354.52	143	-1.60	0.21	662	399	294
2019-07-15 12:11:51	79543.51	446986.10	-2314.38	95	-1.79	0.18	632	381	280
2019-09-11 08:51:06	79212.04	446986.08	-2274.25	149	-2.00	0.13	1627	980	722
2019-10-02 11:46:50	79392.84	447004.24	-2100.33	150	-2.11	0.13	1446	871	642
2019-10-06 01:20:01	79392.90	446986.08	-2153.85	170	-2.52	0.18	1657	998	735

Table 2.2: Characterization of the detected microseismic events. Origin times are given in Coordinated Universal Time (UTC). Hypocentre coordinates are reported in the Dutch national reference system (EPSG: 28992). M_L denotes the local magnitude, and **Var($\mathbf{M_L}$)** represents the variance of the local-magnitude estimates across multiple stations. The parameters σ_x, σ_y, and σ_z correspond to the 95% confidence intervals of the hypocentre location in each coordinate direction. The last two events were not recorded by station 030 or 028, significantly increasing their location uncertainties.

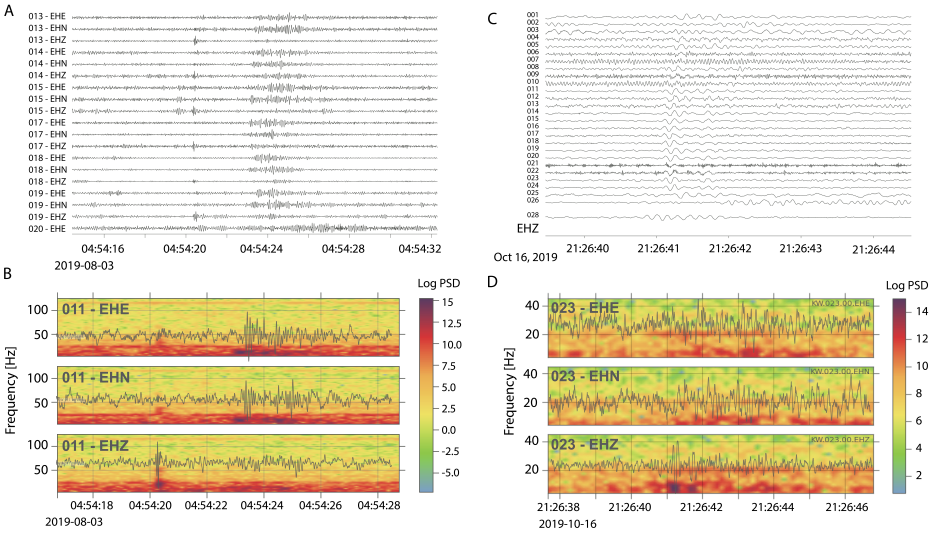


Figure 2.9: Analysis of Eastern Events and Non-Seismic Events. A. Waveforms of one of the Eastern Events recorded on 2019-08-03T04:54:18. B. Spectrogram of station 011 showing the vertical, north, and east components of the Eastern Event. C. Waveforms of one of the non-seismic events recorded on 2019-10-16 at 21:26:40. D. Spectrogram of station 023 showing the vertical, north, and east components of the non-seismic event.

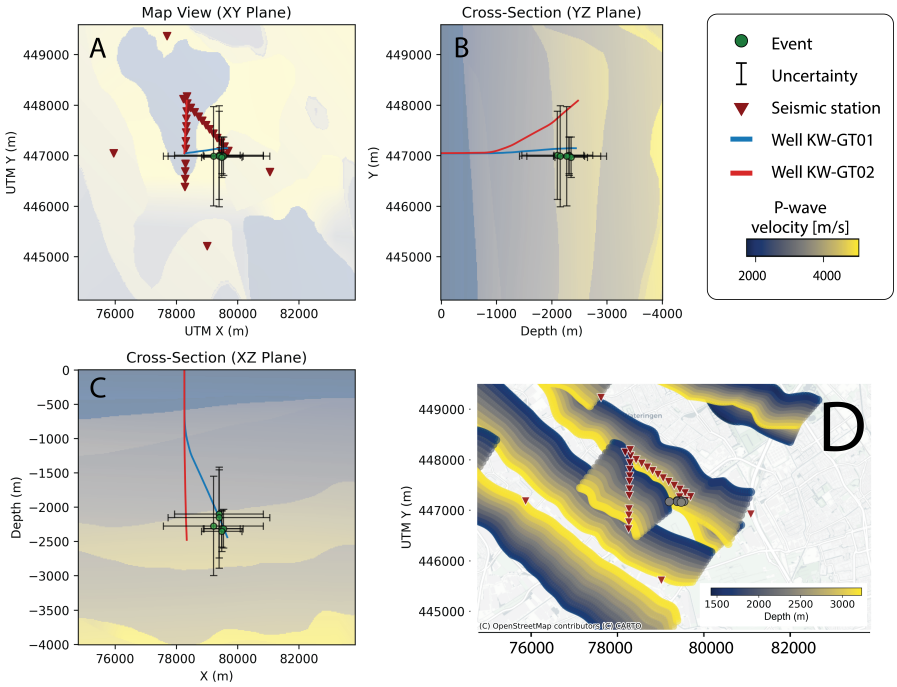


Figure 2.10: Hypocentres that belong to the microseismic cluster. (A, B, C) Map showing the spatial distribution of event hypocentres with associated 95% confidence interval shown as error bars. The background image corresponds to the 3D velocity model used for the inversion, and its corresponding colorscale is in the legend on the right side. D. Projection in 2D showing the localised events, previously mapped faults (Peeters, 2025), and the seismic stations. The colour scale indicates the depth at which the events and faults are located.

2.6. DISCUSSION

A primary motivation for this study was to develop a workflow to monitor low-magnitude seismic events in the Netherlands, particularly those around areas with geothermal operations. Although the nationwide KNMI network effectively detects events above the regional and spatially varying magnitude of completeness (Ruigrok et al., 2023), lower magnitude events will go unnoticed. Detecting and accurately locating these low-magnitude events is crucial for understanding seismicity patterns, informing network designs, and ensuring the safe expansion of geothermal operations. Our results confirm that such low-magnitude seismic events occur in the study area.

We demonstrated that local dense seismic arrays are essential for microseismic-event detectability. False detections increased sharply when the array was partially or poorly configured, highlighting the importance of robust station coverage. Moreover, noisy urban environments benefit from a multi-method approach that includes deep-learning pickers and template matching (Diaferia et al., 2024; Sugan et al., 2023; Panebianco et al., 2023). Using the multi-method approach introduced in Section 2.3.2, we identified 65 events that exhibited coherent signals during the 4-month monitoring period.

We found that the seismic event, reported in Muntendam-Bos et al. (2022), was not an isolated occurrence but part of a repetitive microseismic sequence. This follows from applying the template-matching method, which identified five similar events. The nearly identical waveform patterns indicate a common source region and mechanism. The six seismic events occur at a depth of approximately 2.3 km near both a mapped fault system and the injection well KW-GT-01. The spatial uncertainties, however, prevent a definitive interpretation of whether these events are fault-related or induced by the injection. The limited duration of the monitoring period further complicates this distinction. Longer-term monitoring would help clarify whether these events are truly linked to seismic stress release in the geothermal reservoir, activities at the well, or natural tectonic activity.

In addition, we detected seismicity originating to the east of the array, referred to as Eastern Events (or cluster 2 events). These events exhibit impulsive P-wave arrivals, but they lack an impulsive S-wave. The attenuation of high-frequency S-waves suggests a more distant source for Cluster 2 events. The high azimuthal gap in the current array configuration prevented us from properly estimating their hypocentres, but an origin at the Earth's surface is likely (though not confirmed). Clearly, low detection thresholds are needed to identify low-magnitude seismicity. These low thresholds result in a higher number of false detections, which requires manual inspection. In regions with high seismicity rates, such manual inspection may be infeasible. Therefore, we recommend retraining the machine-learning pickers on site-specific datasets to improve pick accuracy, enable lower thresholds, and, thereby, reduce false detections. This strategy is particularly suitable where permanent seismic stations and high seismicity rates are available.

Our findings highlight a critical limitation of the current array design: excessive reliance on the outer ring of stations. Although the two lines of aligned sensors provide redundancy in signal detection, they do little to improve azimuthal coverage or depth resolution. This array design is, therefore, suboptimal for passive seismic studies that require accurate event locations (at reservoir depths of two or three kilometres). An optimised network targeting the already localised seismic events is recommended for future campaigns (Maurer et al., 2010; Toledo et al., 2020; Esquivel-Mendiola et al., 2022, e.g.,). For large-scale operations,

including borehole geophones would aid in estimating the depths of hypocentres.

Beyond the need for better station coverage, our study underscores the importance of refining the velocity model used for event location. Although the publicly available P-wave velocity model (Velmod 3.1) was instrumental for this study, an S-wave velocity model is needed to enhance the depth resolution and reduce location uncertainties (e.g., Spetzler et al., 2024).

Our analysis also highlights the challenges of magnitude estimation for low-magnitude events. The Groningen-calibrated magnitude formula (Equation 2.9), used here as a first-order approximation, may result in systematic under- or overestimation in South Holland, where the subsurface properties are different. In addition, Equation 2.9 is calibrated using recordings by borehole geophones at 200 m depth. It should, in principle, be used for particle motions recorded at that (or close to that) depth. Establishing a specialised local magnitude scale for this region would require a dedicated network design, knowledge of site-specific attenuation properties, and additional data collected over a longer monitoring period (i.e., more earthquakes).

Taken together, our findings address an important gap between large-scale seismic detection frameworks and the finer resolution needed for local hazard assessment. Moving forward, efforts to refine velocity models, improve monitoring networks, and establish regional magnitude scales will enhance our ability to capture and interpret low-magnitude seismicity.

2.7. CONCLUSIONS

We presented a comprehensive workflow for monitoring seismicity, specifically designed to detect and characterise low-magnitude events in urban areas. By integrating data preparation, machine-learning-based event detection, and probabilistic hypocentre estimation, we addressed key challenges in seismic monitoring. Applying this workflow to the Kwintsheul area, we detected six seismic events near a local fault and near the bottom of injection well KW-GT-01, although spatial uncertainties remain in the order of hundreds of meters. Additionally, we detected 59 events with coherent signals, but inadequate azimuthal coverage hindered their accurate characterisation. Our findings highlight the need for more spatially distributed networks, refined local velocity models, and region-specific magnitude calibrations to enhance the accuracy and reliability of seismic monitoring. Our workflow provides a scalable and adaptable solution for improving seismic monitoring in urban environments.

REFERENCES

Boersma, Q. D., Bruna, P. O., de Hoop, S., Vinci, F., Moradi Tehrani, A., and Bertotti, G. (2021). The impact of natural fractures on heat extraction from tight triassic sandstones in the west netherlands basin: a case study combining well, seismic and numerical data. *Netherlands Journal of Geosciences*, 100:e6. https://doi.org/10.1017/njg.2020.21.

California Institute of Technology and United States Geological Survey Pasadena (1926). Southern california seismic network. https://www.fdsn.org/networks/detail/CI.

Chamberlain, C. J., Hopp, C. J., Boese, C. M., Warren-Smith, E., Chambers, D., Chu, S. X., Michailos, K., and Townend, J. (2017). EQcorrscan: Repeating and Near-Repeating Earthquake Detection and Analysis in Python. *Seismological Research Letters*, 89(1):173–181. https://doi.org/10.1785/0220170151.

Diaferia, G., Valoroso, L., Improta, L., and Piccinini, D. (2024). A high-resolution seismic catalog for the southern apennines (italy) built through template-matching. *Geochemistry, Geophysics, Geosystems*, 25(3):e2023GC011160. https://doi.org/10.1029/2023GC011160.

Dost, B., van Eck, T., and Haak, H. W. (2004). Scaling of peak ground acceleration and peak ground velocity recorded in the netherlands. *Bollettino Di Geofisica Teorica Ed Applicata*, 45:153–168.

Duin, E., Doornenbal, J., Rijkers, R., Verbeek, J., and Wong, T. (2006). Subsurface structure of the netherlands - results of recent onshore and offshore mapping. *Netherlands Journal of Geosciences - Geologie en Mijnbouw*, 85(4):245–276. https://doi.org/10.1017/S0016774600023064.

Ellsworth, W. (1995). - characteristic earthquakes and long-term earthquake forecasts: Implications of central california seismicity. In Cheng, F. and Sheu, M.-S., editors, *Urban Disaster Mitigation: The Role of Engineering and Technology*, pages 1–14. Pergamon, Oxford.

Esquivel-Mendiola, L. I., Calò, M., Tramelli, A., and Figueroa-Soto, A. (2022). Optimization of local scale seismic networks applied to geothermal fields. the case of the acoculco caldera, mexico. *Journal of South American Earth Sciences*, 119:103995. https://doi.org/10.1016/j.jsames.2022.103995.

Font, Y., Kao, H., Lallemand, S., Liu, C.-S., and Chiao, L.-Y. (2004). Hypocentre determination offshore of eastern Taiwan using the Maximum Intersection method. *Geophysical Journal International*, 158(2):655–675.

Groos, J. C. and Ritter, J. R. R. (2009). Time domain classification and quantification of seismic noise in an urban environment. *Geophysical Journal International*, 179(2):1213–1231. https://doi.org/10.1111/j.1365-246X.2009.04343.x.

Heimann, S., Kriegerowski, M., Isken, M., Cesca, S., Daout, S., Grigoli, F., Juretzek, C., Megies, T., Nooshiri, N., Steinberg, A., Sudhaus, H., Vasyura-Bathke, H., Willey, T., and Dahm, T. (2017). Pyrocko - an open-source seismology toolbox and library.

Houtgast, R. and van Balen, R. (2000). Neotectonics of the roer valley rift system, the netherlands. *Global and Planetary Change*, 27(1):131–146. Environmental Tectonics and Climate: The Netherlands Environmental Earth System Dynamics Initiative.

Hoyer, S. and Hamman, J. (2017). xarray: N-D labeled arrays and datasets in Python. *In revision, J. Open Res. Software*.

ISC (2025). International seismological centre. on-line bulletin. https://doi.org/10.31905/D808B830.

Isken, M., Heimann, S., Niemz, P., Münchmeyer, J., Cesca, S., Vasyura-Bathke, H., and Dahm, T. (2025). Qseek: A data-driven Framework for Automated Earthquake Detection, Localization and Characterization. *Seismica*, 4(1). https://doi.org/10.26443/seismica.v4i1.1283.

KNMI (1993). Netherlands seismic and acoustic network. Royal Netherlands Meteorological Institute (KNMI). Other/Seismic Network. https://doi.org/10.21944/e970fd34-23b9-3411-b366-e4f72877d2c5.

KNMI (2023). Earthquakes - complete catalogue for the netherlands and near surrounding. Retrieved from https://dataplatform.knmi.nl/dataset/aardbevingen-catalogus-1.

Kraaijpoel, D. (2025). TNO/PRESEIS-dgm-velmod-sampler: v2025.1 (Software). https://doi.org/10.5281/zenodo.14698789.

Lin, F.-C. and Ritzwoller, M. H. (2011). Helmholtz surface wave tomography for isotropic and azimuthally anisotropic structure. *Geophysical Journal International*, 186(3):1104–1120. https://doi.org/10.1111/j.1365-246X.2011.05070.x.

Lomax, A., Michelini, A., and Curtis, A. (2009). *Earthquake Location, Direct, Global-Search Methods*, pages 1–33. Springer New York, New York, NY. https://doi.org/10.1007/978-3-642-27737-5_150-2.

Maurer, H., Curtis, A., and Boerner, D. E. (2010). Recent advances in optimized geophysical survey design. *Geophysics*, 75(5):75A177–75A194. https://doi.org/10.1190/1.3484194.

Menke, W. (1999). Using waveform similarity to constrain earthquake locations. *Bulletin of the Seismological Society of America*, 89(4):1143–1146. https://doi.org/10.1785/BSSA0890041143.

Merrifield, T. (2012). R-2823 Rotterdam Donkersloot Pre-SDM. Technical report, Shell UIE. Internal report.

Mosegaard, K. and Tarantola, A. (1995). Monte carlo sampling of solutions to inverse problems. *Journal of Geophysical Research: Solid Earth*, 100(B7):12431–12447. https://doi.org/10.1029/94JB03097.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., and Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11:3952. https://doi.org/10.1038/s41467-020-17591-w.

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. https://doi.org/10.1017/njg.2021.14.

Naranjo, D., Draganov, D., Polychronopoulou, K., de Bas, M., and Weemstra, C. (2022). Seismic monitoring of Nature's Heat Geothermal Reservoir in Kwintsheul, Netherlands. In *Proceedings European Geothermal Congress 2022*, Berlin, Germany. https://research.tudelft.nl/files/174035345/Naranjo_Microseismic_monitoring_of_Natures_Heat_Geothermal_Project_432_ExtAbstract.pdf.

Naranjo, D., Isken, M., Boullenger, B., Toledo, T., Weemstra, C., and Draganov, D. (2025). Urban challenges in seismology: seismic monitoring of Kwintsheul's geothermal operation (the Netherlands). *Netherlands Journal of Geosciences*, 104. https://njgjournal.nl/index.php/njg/article/view/12188.

OpenStreetMap contributors (2017). Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org.

Panea, I., Draganov, D., Vidal, C. A., and Mocanu, V. (2014). Retrieval of reflections from ambient noise recorded in the mizil area, romania. *GEOPHYSICS*, 79(3):Q31–Q42. https://doi.org/10.1190/geo2013-0292.1.

Panebianco, S., Serlenga, V., Satriano, C., Cavalcante, F., and Stabile, T. A. (2023). Semi-automated template matching and machine-learning based analysis of the august 2020 castelsaraceno microearthquake sequence (southern italy). *Geomatics, Natural Hazards and Risk*, 14(1):2207715. 10.1080/19475705.2023.2207715.

Peeters, S. (2025). Fault interpretation based on merrifield (2012) 3d seismic data. Technical report, TNO. Personal communication.

Pluymaekers, M., Doornenbal, J., and Middelburg, H. (2017). TNO 2017 R11014 with erratum page 67 Final, Velmod-3.1. *TNO report*. https://www.nlog.nl/sites/default/files/2018-11/060.26839%20R11014%20with%20erratum%20page%2067%20Doornenbal-final.sec_.pdf.

Ross, Z. E., Meier, M., Hauksson, E., and Heaton, T. H. (2018). Generalized Seismic Phase Detection with Deep Learning. *Bulletin of the Seismological Society of America*, 108(5A):2894–2901. https://doi.org/10.1785/0120180080.

Ruigrok, E., Kruiver, P., and Dos, B. (2023). Construction of earthquake location uncertainty maps for the netherlands. KNMI number: TR-405, Year: 2023, Pages: 158.

Sethian, J. A. and Popovici, A. M. (1999). 3-d traveltime computation using the fast marching method. *Geophysics*, 64(2):516–523. https://doi.org/10.1190/1.1444558.

Soto, H. and Schurr, B. (2021). DeepPhasePick: a method for detecting and picking seismic phases from local earthquakes based on highly optimized convolutional and recurrent deep neural networks. *Geophysical Journal International*, 227(2):1268–1294. https://doi.org/10.1093/gji/ggab266.

Spetzler, J., Ruigrok, E., and Bouwman, D. (2024). Hypocenter uncertainty analysis of induced and tectonic earthquakes in the Netherlands. *Journal of Seismology*. https://doi.org/10.1007/s10950-024-10205-8.

Sugan, M., Peruzza, L., Romano, M. A., Guidarelli, M., Moratto, L., Sandron, D., Plasencia Linares, M. P., and Romanelli, M. (2023). Machine learning versus manual earthquake location workflow: testing loc-flow on an unusually productive microseismic sequence in northeastern italy. *Geomatics, Natural Hazards and Risk*, 14(1):2284120. 10.1080/19475705.2023.2284120.

Tarantola, A. and Valette, B. (1982). Inverse problems = Quest for information. *Journal of geophysics*, 50:159–170.

Toledo, T., Jousset, P., Maurer, H., and Krawczyk, C. (2020). Optimized experimental network design for earthquake location problems: Applications to geothermal and volcanic field seismic networks. *Journal of Volcanology and Geothermal Research*, 391:106433. https://doi.org/10.1016/j.jvolgeores.2018.08.011.

v. Gessel, S., Hintersberger, E., v. Ede, R., ten Veen, J., Doornenbal, H., Diepolder, G., den Dulk, M., Hamiti, S., Vukzaj, N., Çako, R., Prendi, E., Ceroni, M., Mara, A., Barros, R., Tovar, A., Britze, P., Baudin, T., Stück, H., Jähne-Klingberg, F., Jahnke, C., Höding, T., Malz, A., Kristjánsdóttir, S., Þorbergsson, A., Di Manna, P., D'Ambrogi, C., Congi, M., Lazauskienė, J., Andriuškevičienė, G., Baliukevičius, A., Jarosiński, M., Gogołek, T., Stępień, U., Krzemińska, E., Salwa, S., Habryn, R., Aleksandrowski, P., Szynkaruk, E., Konieczyńska, M., Ressurreição, R., Machado, S., Moniz, C., Sampaio, J., Dias, R., Carvalho, J., Fernandes, J., Ramalho, E., Filipe, A., Celarc, B., Atanackov, J., Jamšek Rupnik, P., Shevchenko, A., Melnyk, I., and Lapshyna, A. (2021). The HIKE european fault database (EFDB) compiled in the framework of the GeoERA project HIKE.(2018-2021). *Deliverable GeoERA project HIKE, European Geological Data Infrastructure (EGDI)*. https://egdi.geology.cz/record/basic/5edf7bd4-9270-4188-b69d-7ddd0a010833.

White, M. C. A., Fang, H., Nakata, N., and Ben-Zion, Y. (2020). PyKonal: A Python Package for Solving the Eikonal Equation in Spherical and Cartesian Coordinates Using the Fast Marching Method. *Seismological Research Letters*, 91(4):2378–2389. https://doi.org/10.1785/0220190318.

Worum, G., Michon, L., van Balen, R. T., van Wees, J.-D., Cloetingh, S., and Pagnier, H. (2005). Pre-neogene controls on present-day fault activity in the west netherlands basin and roer valley rift system (southern netherlands): role of variations in fault orientation in a uniform low-stress regime. *Quaternary Science Reviews*, 24(3):473–488. https://doi.org/10.1016/j.quascirev.2004.02.020.

Zhu, W. and Beroza, G. C. (2019). Phasenet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273. https://doi.org/10.1093/gji/ggy423.

3

Velocity-Model Validation Through Body-wave Seismic Interferometry

The medium is the message.

Marshall McLuhan

Seismic velocity models determine the accuracy of hypocentre estimations and remain a significant source of uncertainty. Developing a new seismic velocity model requires a significant investment. The geothermal-heat-production industry continuously seeks innovative and cost-effective approaches to support such decisions. In this chapter, we present a methodology to validate a seismic velocity model using Seismic Interferometry (SI) applied to six microseismic events recorded during geothermal operations in Kwintsheul. Some of these events exhibit near-vertical-incidence angles at specific stations, enabling the retrieval of zero-offset virtual reflection responses beneath the stations. We compare these SI-derived reflections with synthetic responses generated from the available P- and S-wave velocity models. While the P-wave model agrees well with the observed responses, neither of the tested S-wave models matches the data. Based on this discrepancy, we excluded the S-wave model from the hypocentre inversion presented in the previous chapter. The workflow demonstrates that SI can provide independent, reflection-based validation of velocity models under favourable conditions, offering a cost-effective method.

3.1. Introduction

Accurate estimation of hypocentre locations is essential for characterising microseismicity and understanding subsurface processes in geothermal reservoirs. These estimations rely heavily on the quality of the seismic velocity model, which is used to determine the travel times of seismic waves through the subsurface. However, it is challenging to decide whether to improve an existing velocity model. Direct methods, such as active-source surveys, are costly. This issue raises a central question: can we use local passive seismicity to validate the velocity model in a cost-effective manner?

In Chapter 2, we identified six microseismic events recorded during a temporary seismic monitoring campaign in 2019 in Kwintsheul, South Holland. Four stations in the local network were positioned close to the epicentres, which means that the seismic energy from the events arrives with near-vertical incidence. The wavefield includes not only direct arrivals but also multiple-order reflections between the free surface and subsurface seismic discontinuities, among others. These reflected signals arrive after the direct wave and contribute to the coda. As a result, the coda waves carry information about the reflection response and the seismic velocities beneath the recording stations.

This geometry presents an opportunity to assess the existing velocity model. We apply body-wave Seismic Interferometry (SI) to the coda waves of the recorded microseismic events. SI enables us to retrieve zero-offset virtual reflection responses (ZO-VRR) beneath the stations, which approximate the reflection wavefield that would be recorded if an impulsive source, co-located with the receiver, were placed at the surface. Similar approaches have been successfully used in various settings, such as mineral exploration (Polychronopoulou et al., 2019), for imaging the reinjection of CO_2 (Hassing et al., 2024), and crustal-scale imaging (Nishitsuji et al., 2016). We adopt the body-wave approach, focusing on the vertical-incidence geometry of the recorded events to extract local reflection responses. We then compare the retrieved responses with synthetic ones generated from the available P- and S-wave velocity models to evaluate their validity.

3.2. THEORY

3.2.1. BODY-WAVE SEISMIC INTERFEROMETRY

Initially derived by Claerbout (1968) for 1D media, SI relies on the principle that cross-correlation, or autocorrelation, of a transmission response approximates a reflection response. This principle has been demonstrated in various seismological contexts (Scherbaum, 1987; Daneshvar et al., 1995; Rickett and Claerbout, 1999; Van IJsseldijk et al., 2019) and has been mathematically formalised by Wapenaar (2003) and Wapenaar and Fokkema (2006) for 3D media. Here, we focus on the application of body-wave SI using local microseismic sources.

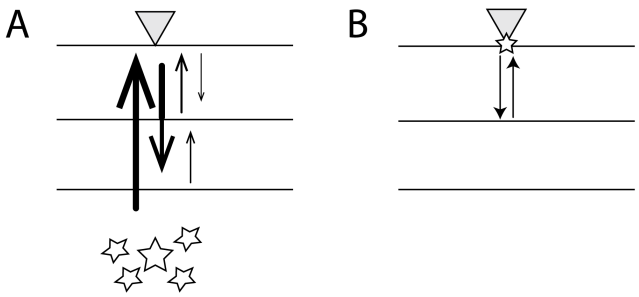


Figure 3.1: Schematic illustration of seismic interferometry by autocorrelation. A. Seismic sources located within the stationary-phase zone, generating direct arrivals and multiple reflections from subsurface interfaces. B. Schematic of a virtual source co-located with a receiver after estimating the ZO-VRR from the configuration in A. The direct path is removed, and the multiple-order reflections are captured in the ZO-VRR.

Body-wave SI can be used to retrieve ZO-VRR traces by calculating the autocorrelation of the wavefield recorded at receivers with seismic energy arriving with vertical incidence. These ZO-VRR traces approximate the wavefield that would be recorded if an impulsive source had been placed at the free surface, illuminating the medium from above with vertically downward emitted energy (Galetti and Curtis, 2012), as shown in Figure 3.1. In a horizontally layered subsurface, the global reflection response (including all orders of internal multiples) of a vertically propagating plane wave below a seismic station can be expressed as

$$R(t) + R(-t) = \delta(t) - T(t) * T(-t), \tag{3.1}$$

where $R(t)$ and $R(-t)$ are the global reflection response and its time-reversed version, respectively, $T(t)$ is the global transmission response, and $\delta(t)$ is a delta function that can be interpreted as an impulsive source illuminating the subsurface from above (Wapenaar et al., 2010; Claerbout, 1968). Equation 3.1 shows that the global reflection response can be obtained from the autocorrelation of the transmission response.

For the application to field data, the autocorrelation function retrieves the (near-)vertical reflection response of the medium from non-plane-wave subsurface sources like the local microseismicity if I) the medium is lossless, ii) the medium enclosing the propagation path between source and receiver does not scatter energy back to the receivers, and iii) the medium parameters (velocity and density) change smoothly along the source boundary. Additionally, a high-frequency assumption is considered (Wapenaar and Fokkema, 2006).

The retrieved reflection response, obtained from the autocorrelation function, originates from sources located in the stationary phase along the source boundary. For those sources, the wave energy propagates from the source to a receiver, reflects at the Earth's surface, and after reflecting from a target reflector, is recorded again at the receiver position. The autocorrelation effectively removes the propagation path from source to receiver, leaving only the reflection path from the Earth's surface to the different reflectors.

However, when applied to field data, the autocorrelation function may yield non-physical events due to energy scattered back from inhomogeneities outside (below) the source boundary. The effect of spurious signals in the autocorrelation function may be reduced by using responses from sources at different (random) depth levels and stacking their results (Draganov et al., 2004). The stacking process increases the signal of physical events, such as reflections, while cancelling spurious events. Following Polychronopoulou et al. (2019), the reflection response below receiver i can be expressed as

$$R^i(t) + R^i(-t) = \delta^i(t) - \sum_n T_n^i(t) x T_n^i(-t), \tag{3.2}$$

where the term $\sum_n T_n^i(t) x T_n^i(-t)$ represents the stacking of the autocorrelations from different vertically propagating plane waves of n seismic sources.

In an elastic medium, the autocorrelation function may retrieve cross-component artefacts from the conversion between P- and S-waves (Polychronopoulou et al., 2019). Figure 3.2 illustrates a simplified example of phase conversions at the free surface of a solid half-space. The cross-component interference effect can be reduced by selecting receivers where the body-wave energy arrives with near-vertical incidence. We provide details in Section 3.3.

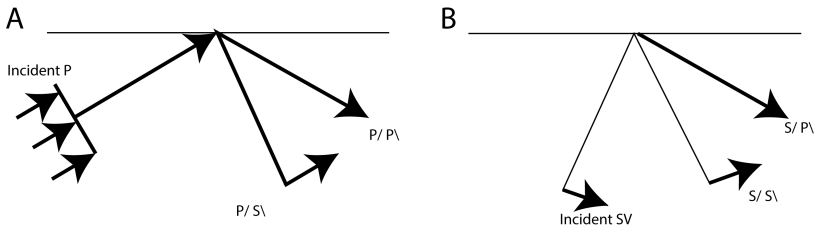


Figure 3.2: Schematic illustration of wavefield conversions at the free surface of a solid half-space, adapted from Aki and Richards (1980). A. An incident P-wave generates reflected P and converted SV phases. B. An incident SV-wave generates reflected SV and converted P phases. These conversions can affect the retrieval of ZO-VRR traces and must be minimised in SI applications.

In summary, the following conditions are required for the application of body-wave interferometry for 1-D velocity model validation:

1. The wavefield is composed of vertically propagating body waves from sources that lie within the stationary-phase zone.

2. Mode-conversion effects are negligible for the contributing paths. This condition can be met by selecting receivers located directly above the microseismic sources, thereby reducing P–S conversions along the propagation path.

3. The waveforms contain sufficiently high-frequency energy to justify the high-frequency approximation.

4. The medium enclosing the propagation path between source and receiver does not scatter energy back to the receiver. Because this cannot be directly verified, we must stack different ZO-VRR traces from multiple sources to enhance reflections and suppress non-physical events in the autocorrelation function.

3.3. WORKFLOW

The goal of this workflow is to validate seismic velocity models using ZO-VRR traces retrieved from seismic event recordings. These traces represent the wavefield recorded by a receiver co-located at the surface with a source, and their reflections correspond to two-way travel times from such a configuration. We compare these travel times with synthetic ones generated using candidate velocity models. The difference in arrival times between observed and synthetic data indicates the model's accuracy. Based on this approach, we implement the workflow described in this section.

1. **Source-receiver selection**

   We analyse all available receiver–source pairs and select those for which the source lies within the stationary-phase region. To do this, we select pairs where the wave arrives with an incidence angle below 10°. We also select pairs where phase conversion (P-to-S and S-to-P) is minimal. To quantify the converted energy, we use scattering

matrices from Aki and Richards (1980), which are implemented in the Pyrocko toolbox (Heimann et al., 2019). These matrices provide analytical expressions for the fraction of energy that is transmitted, reflected, or converted at interfaces.

2. **Signal extraction**

   For retrieving ZO-VRR traces, we need to identify the P- and S-wave signals and their corresponding coda-wave signal. We manually pick the P-phase arrival time (T_p) and S-phase arrival times (T_s) of each source-receiver pair. We use the vertical component (EHZ) for P-waves and the horizontal components (EHN and EHE) for S-waves.

   We compute a correlation window of 3 s starting from the phase-onset time. However, the interpretation is limited to the coda-wave time window. We select a P-wave coda time window with a length of $T_p - T_s = 1.56s$ following Polychronopoulou et al. (2019). Although the S-wave coda time window can be estimated by including a term related to the source duration, for events of $M < 5$ this term can be neglected. While Polychronopoulou et al. (2019) suggested the same duration for the S-wave coda, we extend it to 3 s to include deeper reflected phases. We choose this longer window despite the expected attenuation and proceed with careful interpretation.

3. **Pick correction**

   We apply the cross-correlation pick correction method of Deichmann and Garcia-Fernandez (1992) to correct bias in the manual picks and to verify waveform alignment within the selected windows. An example of the waveform alignment before and after correction is shown in Appendix A.2.

4. **Preprocessing**

   We apply detrending, tapering, and band-pass filtering to the recorded three-component waveforms to suppress ambient noise while preserving body-wave energy. The selected frequency band is discussed in Section 3.5.

5. **Autocorrelation**

   We normalise each extracted window to ensure consistent amplitude levels across all windows. We autocorrelate the normalised traces for each event and each component separately.

6. **Stacking per receiver component**

   We compute the ZO-VRR by stacking normalised autocorrelation traces for each receiver and component separately. Normalisation ensures that each input trace contributes equally, regardless of its absolute amplitude. The resulting stacked autocorrelation traces represent the ZO-VRR for that specific receiver and component. We then compare the ZO-VRRs with expected reflection arrivals based on the seismic velocity model.

3.4. DATA

3.4.1. SEISMIC VELOCITY MODEL AND EXPECTED TWO-WAY TRAVEL TIMES

We evaluate the following three 1D seismic velocity profiles at the location of the Kwintsheul geothermal doublet (see Figure 3.3):

1. We use a P-wave velocity model from Velmod 3.2, the regional model for the Netherlands (Pluymaekers et al., 2017). We extract a 1D profile at the average epicentral location of the six microseismic events introduced in Chapter 2.
2. We use an S-wave velocity model derived from Vp/Vs ratios obtained from well data in the Zeerijp and Borgsweer areas of the northeastern Groningen gas field, where such data are available (J. Spetzler, personal communication, January 24, 2022).
3. We use an S-wave velocity model obtained by inverting three-component beamforming and cross-correlation dispersion curves at the Kwintsheul site (Simonet, 2024). This profile covers the upper 500 m. For the deeper layers, we use the same Vp/Vs ratios as those in Model 2.

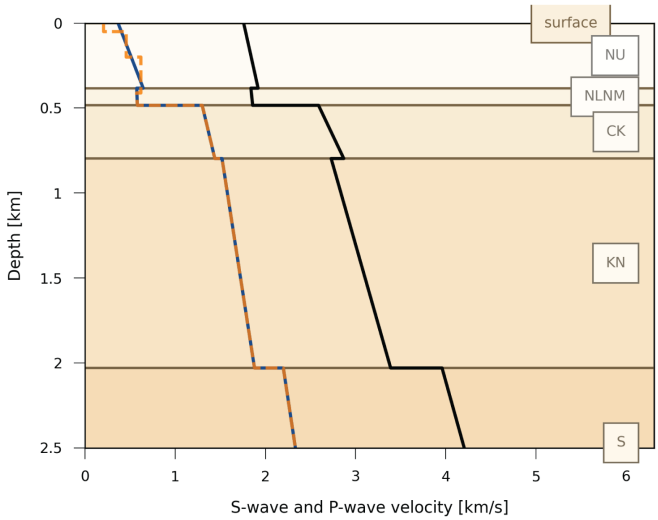


Figure 3.3: Available seismic velocity models. The P-wave velocity model is shown in black. Two S-wave velocity models are available: the first (blue) is derived from Vp/Vs ratios from the northern Netherlands, and the second (orange, dashed), down to 500 m, is based on surface-wave dispersion inversion at Kwintsheul (Simonet, 2024).

From these models, we calculate the expected two-way travel times (TWTTs) for reflected body waves, assuming a source co-located with the receiver at the surface (Figure 3.1.B). Table 3.1 summarises the calculated TWTTs and corresponding lithostratigraphic layers.

Strat. layer	Depth [m]	P TWTT [s]	S^1 TWTT [s]	S^2 TWTT [s]
NU	383	0.440	1.552	1.732
NLNM	484	0.526	1.902	2.075
CK	797	0.775	2.360	2.533
KN	2030	1.564	3.816	3.989
S	2910	1.985	4.573	4.747

Table 3.1: Two-way travel times (TWTT) of a signal from a source at the surface reflected at the different lithostratigraphic units. TWTTs are computed using the P-wave and two S-wave velocity models introduced in Section 3.4.1. The lithostratigraphic units correspond to Upper North Sea Group (NU), Lower and Middle North Sea Groups (NLNM), Chalk Group (CK), Rijnland Group (KN), and Scruff Groups (S). Depth, two-way travel times, and corresponding lithostratigraphic layers from Velmod-3.2 (Pluymaekers et al., 2017).

3.4.2. WAVEFORM DATA

Chapter 2 provides details of the seismic network and the microseismic events we use as input. We estimate incidence angles and phase-conversion percentages to identify suitable source–receiver pairs for the SI analysis. Figure 3.4 shows their spatial distribution and associated waveforms. Appendix A.3 provides a complete overview of the selected pairs and their corresponding parameters. Among all receivers, 026 and 025 exhibit the lowest expected pS conversion, with average values below 0.02%.

After manually picking the events, we select the waveforms shown in Figures 3.5–3.7 for further processing. Each trace includes 1 s before and 3 s after the P-wave onset time. The vertical-component recordings (Figure 3.5) exhibit a clear, impulsive P-wave signal. Receivers 002 and 003 show prominent long-period noise, visible as broad red and blue lobes in the noise window (0-1 s) and continue in the P-wave coda.

The horizontal-component recordings, shown in Figures 3.7 and 3.6, cover the same time window as Figure 3.5. The S-wave onset is around 1.56 s after the P-wave arrival time. The event on 2019-07-14 is well recorded by most receivers in both horizontal components. Receiver 001 shows a clear S-wave arrival on the North–South component across five microseismic events. In the remaining receivers, S-wave amplitudes are comparable to those in the noise window (0-1s).

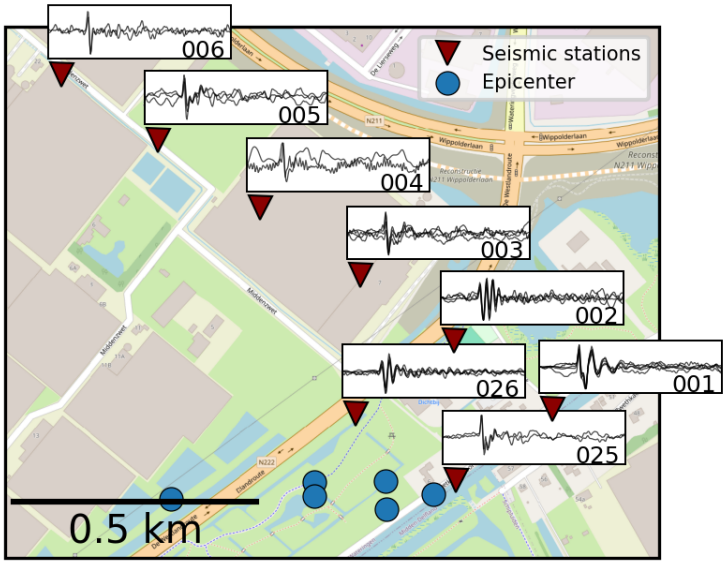


Figure 3.4: Map view of the selected sources and receivers. Each vertical-component waveform is plotted at its corresponding receiver location, showing the P-wave arrival. Each inset shows the superposition of waveforms from all detected microseismic events, after detrending and normalisation.

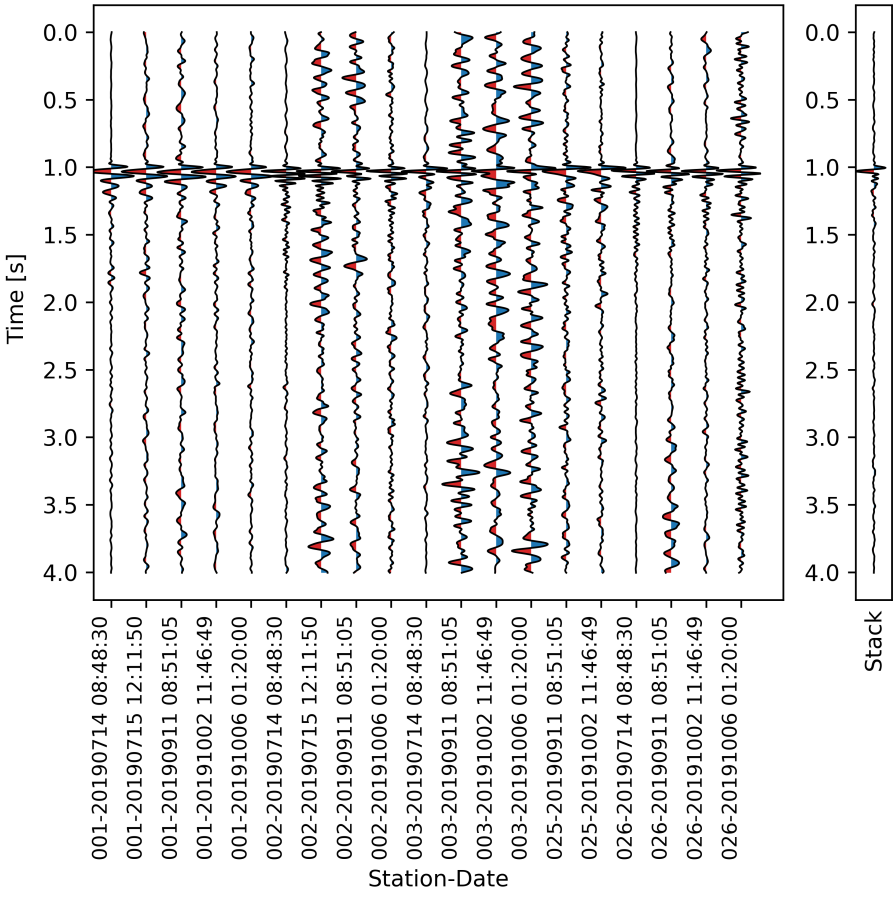


Figure 3.5: Vertical-component waveforms selected for further processing. The traces are detrended, normalised, band-pass filtered between 4 and 30 Hz, pick-corrected, and trimmed from 1.0 s before to 3.0 s after the P-wave arrival. Blue and red fills indicate positive and negative polarities, respectively.

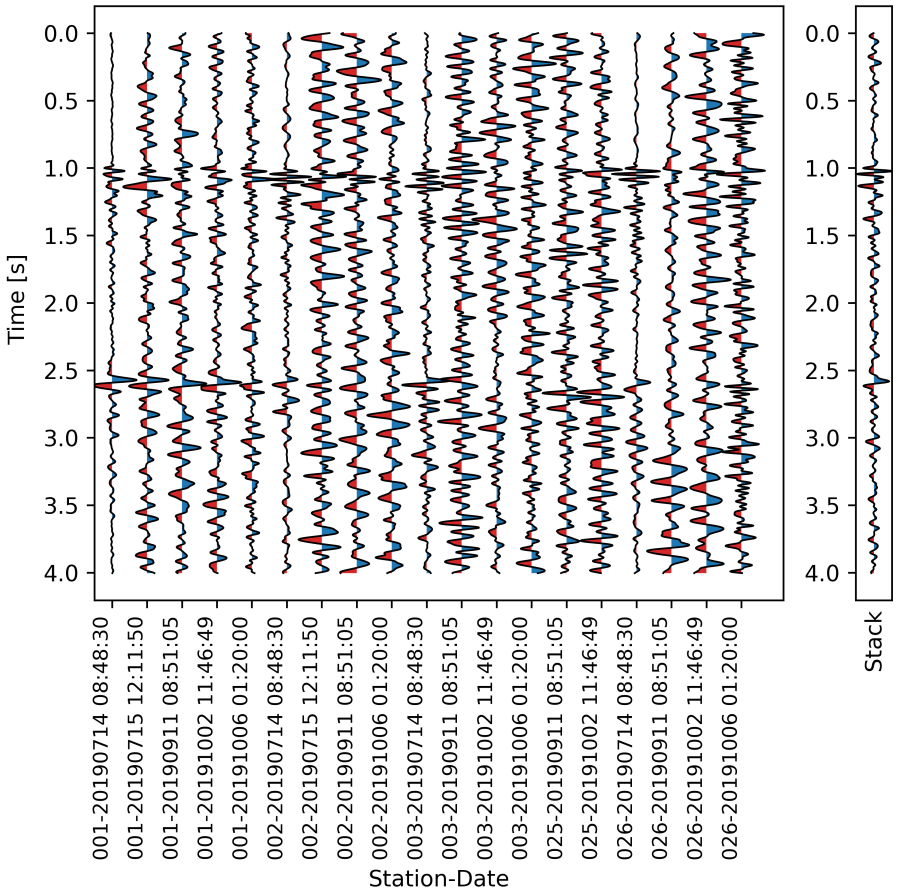


Figure 3.6: North-South-component waveforms selected for further processing. The traces are detrended, normalised, band-pass filtered between 4 and 30 Hz, pick-corrected, and trimmed from 1.0 s before to 3.0 s after the P-wave arrival (same time window as in Figure 3.5). Blue and red fills indicate positive and negative polarities, respectively.

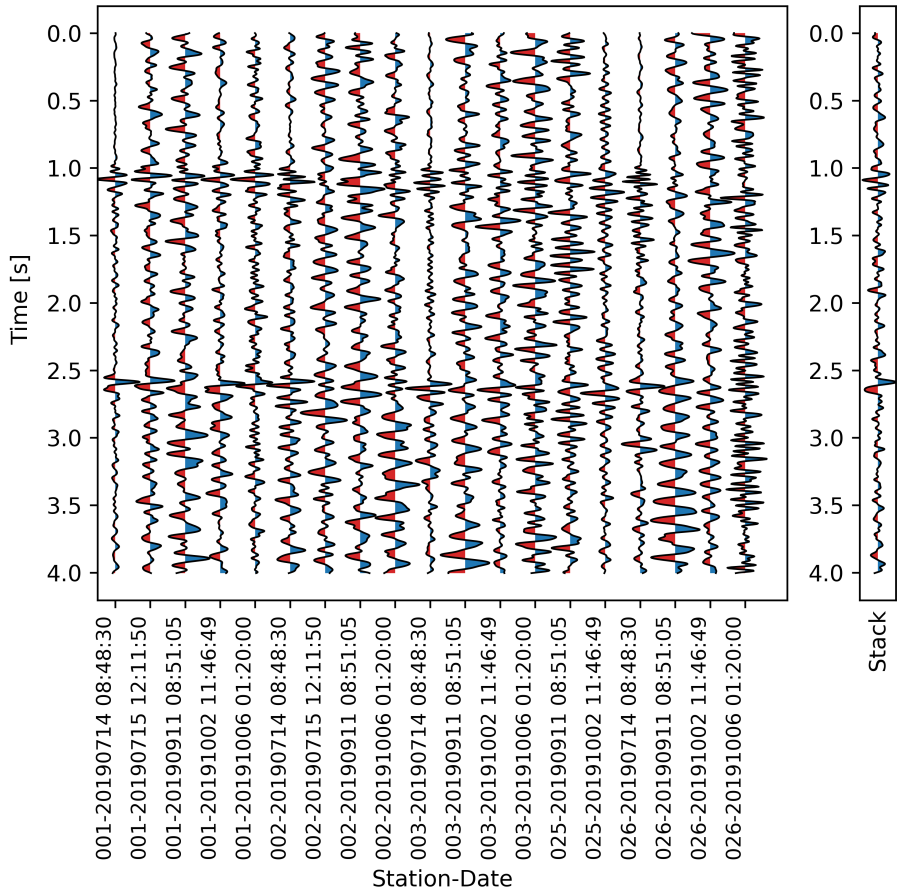


Figure 3.7: East-West-component waveforms selected for further processing. The traces are detrended, normalised, band-pass filtered between 4 and 30 Hz, pick-corrected, and trimmed from 1.0 s before to 3.0 s after the P-wave arrival (same time window as in Figure 3.5). Blue and red fills indicate positive and negative polarities, respectively.

3.5. NOISE ANALYSIS AND APPLIED FILTERING

We analyse the frequency content of the microseismic events recorded by receiver 001 in Figure 3.8. The spectrograms of the vertical component (left column) reveal that the P-wave arrival has most of its energy concentrated between 8 and 25 Hz. Low-frequency noise (below 4 Hz) is visible before and after the arrival. Other receiver recordings exhibit low-frequency noise up to 10 Hz (see Appendix A.4).

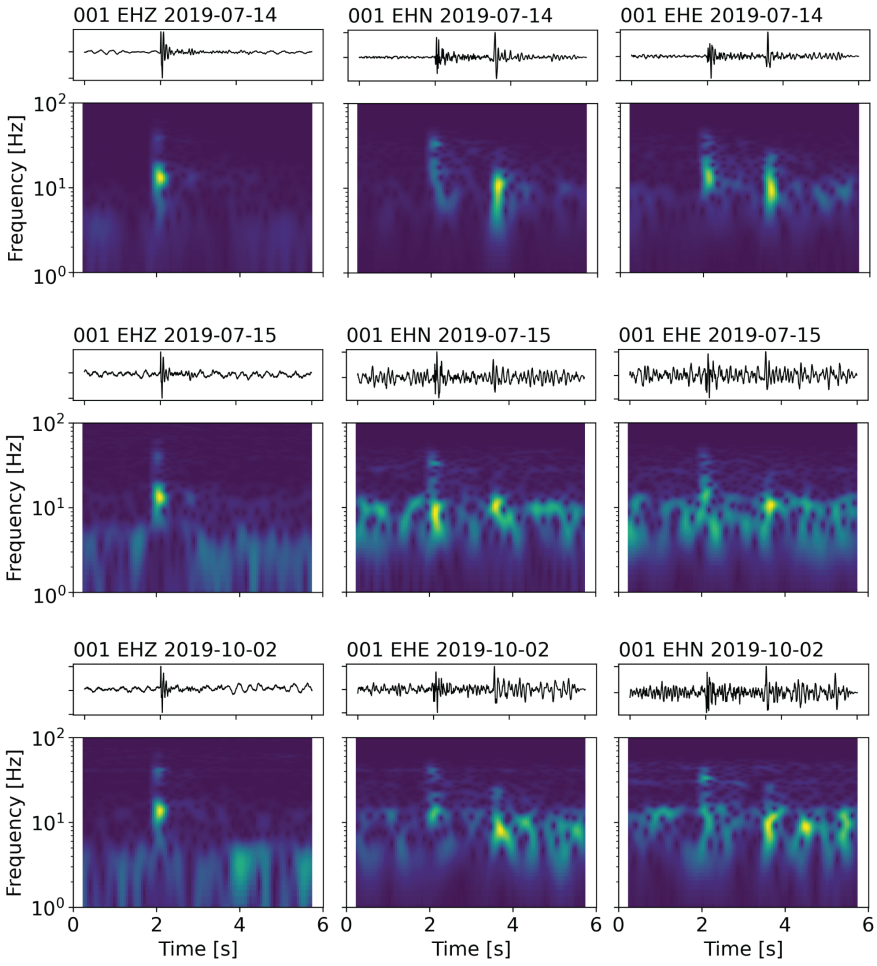


Figure 3.8: Seismic waveforms (top panels) and spectrograms (bottom panels) for three events recorded at receiver 001 on the vertical (EHZ), North–South (EHN), and East–West (EHE) components. Each column corresponds to a component, and each row to a different event (2019-07-14, 2019-07-15, and 2019-10-02). The traces are detrended and normalised. In the spectrograms, brighter colours represent higher amplitudes at each frequency.

The horizontal components exhibit stronger low-frequency noise, as shown in the second and third columns of Figure 3.8. We apply a band-pass filter between 4 and 45 Hz to

isolate the P- and S-wave signals and suppress low- and high-frequency noise that dominates the raw waveforms. The event shown in the first row of Figure 3.8 displays a clear S-wave arrival with energy between 8 and 15 Hz. In contrast, the second and third events exhibit noise contamination in this frequency range, which is already visible in the pre-arrival noise window (0–2 s). The relatively stronger low-frequency content of the S waves compared to the P waves is likely due to the greater attenuation of high-frequency S-wave energy, which commonly occurs in the near-surface.

Based on these observations, we apply a band-pass filter between **4 and 30 Hz** to the traces before estimating their autocorrelation functions.

3.6. WAVEFORM ANALYSIS

As a diagnostic step before the autocorrelation, we analyse the raw traces of the strongest event and the expected reflection signals. In Figure 3.9A, we analyse the P-wave signal and its coda at the station with the highest signal-to-noise ratio. The coda contains three distinct events, labelled e1, e2, and e3. The first event, e1, appears to consist of two interfering wavelets. The second event, e2, shows the highest amplitude and appears at 0.75 s. After event e3, the signal becomes highly attenuated.

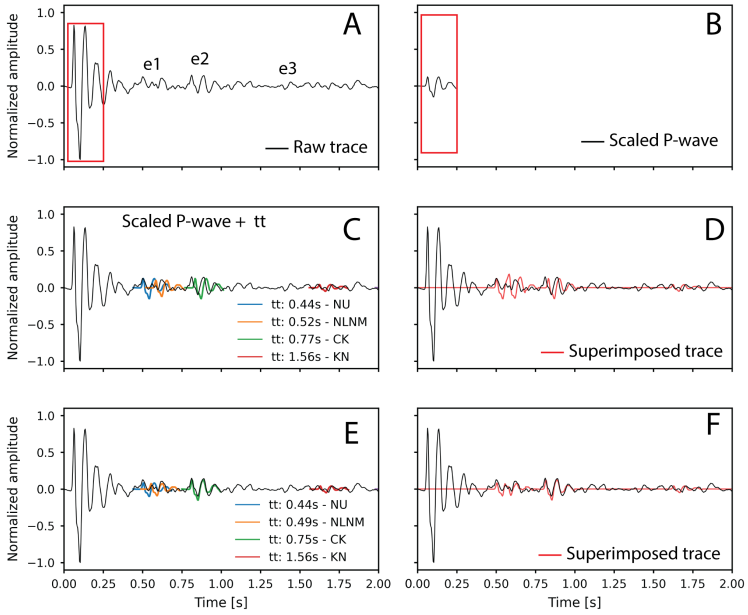


Figure 3.9: P-wave coda analysis for the event on 2019-07-14 at 08:48:31 recorded on station 001. A. Vertical-component trace with the P-wave signal marked by a red box. B. P-wave signal scaled to 15 % of its original amplitude. C. Scaled P-wave signal shifted by the predicted two-way travel times of reflectors from the velocity model. D. Summed shifted traces in red compared to the raw signal in black. E. Example with manually adjusted travel times. F. Summed trace using the adjusted shifts, showing improved alignment with coda reflections.

We then evaluate the P-wave signal after scaling it to 10% of its original amplitude (Fig-

ure 3.9B). We shift this scaled P-wave signal by the expected two-way travel times from a source and receiver co-located at the surface, which reflects at the different lithostratigraphic units of the velocity model. The shifted signals, colour-coded by lithostratigraphic unit, are shown in Figure 3.9C, while the summed trace is shown in red in Figure 3.9D.

We identify two expected reflectors with nearly overlapping arrivals: NU at 0.44 s and NLNM at 0.52 s. When we plot the shifted signals (Figure 3.9C), these reflectors interfere with each other, creating a tuning effect, as seen in the red summed trace in Figure 3.9D. This signal arrives almost simultaneously with the event we previously labelled as e1. Shifting the P-wave signal by the two-way travel time of reflector CK (0.77 s) matches the event labelled e2 in the P-wave coda. Event e3 does not match any of the arrivals from the reflections.

3.7. RESULTS

Among all receivers, only station 001 recorded more than three microseismic events with a sufficiently high signal-to-noise ratio (SNR). The other receivers showed either strong ambient noise or weak arrivals, which prevented the retrieval of reliable autocorrelations. For this reason, we focus the analysis of both P- and S-wave reflections on receiver 001.

3.7.1. VERTICAL ZO-VRR TRACE ANALYSIS

In Figure 3.10.A, we show the filtered traces of the vertical component of station 001 before estimating their autocorrelation. The P-wave coda shows alignment of the three clear events highlighted as e1, e2, and e3. After computing the individual autocorrelation functions, we observe clear alignments of the expected reflection events, as shown in Figure 3.10B. The ZO-VRR trace highlights these events more clearly.

The expected two-way travel times of the NU and NLNM reflectors are 0.44 s and 0.52 s, respectively. A distinct event is observed at this time in the ZO-VRR trace shown in Figure 3.10D. The event appears as a tuned wavelet.

Having validated at least some reflectors at depth, the next step is to analyse the ZO-VRR traces from the horizontal components to determine whether the empirical S-wave velocity relation, derived for the northern Netherlands, also holds in this region.

3.7.2. HORIZONTAL ZO-VRR TRACE ANALYSIS

Figure 3.11 shows the horizontal ZO-VRR traces at receiver 001, with the N–S component (EHN) displayed on the left and the E–W component (EHE) on the right. The horizontal ZO-VRR traces show several coherent arrivals that align with the expected two-way travel times from the S-wave models (Figure 3.11). On both components, we observe a clear energy packet between 1.5 and 1.7 s. Stacking enhances the visibility of this packet, as shown in Figure 3.11.C and 3.11.F, where it appears between the predicted arrival times of the two S-wave models for the NU reflector. In Figure 3.11.F, we see two retrieved signals that coincide with each of the estimated arrival times of the NLNM reflection (orange) predicted by V_S^1 and V_S^2.

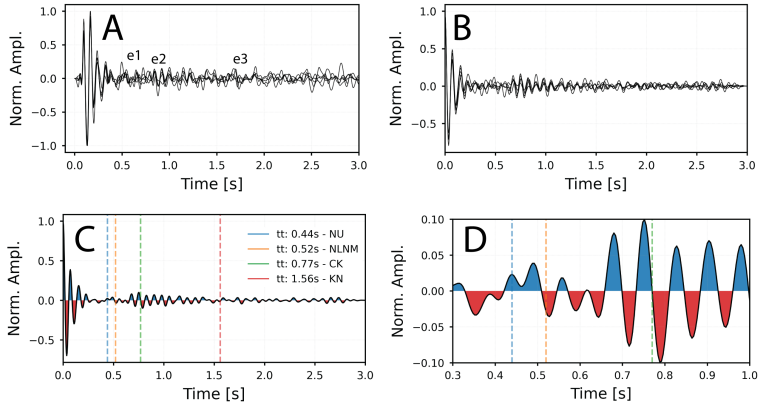


Figure 3.10: Analysis of ZO-VRR retrieved from the vertical component of receiver 001. A. Filtered traces showing the time window that is autocorrelated. B. Causal part of the autocorrelation function of each seismic event at station 001 overlaid on each other. C. Causal part of the ZO-VRR trace after stacking. D. Trimmed ZO-VRR trace window highlighting the first three expected reflection arrivals. Vertical lines denote predicted two-way travel times to the top of NU (blue), NLNM (orange), CK (green), and KN (red) for a source and receiver co-located on the surface.

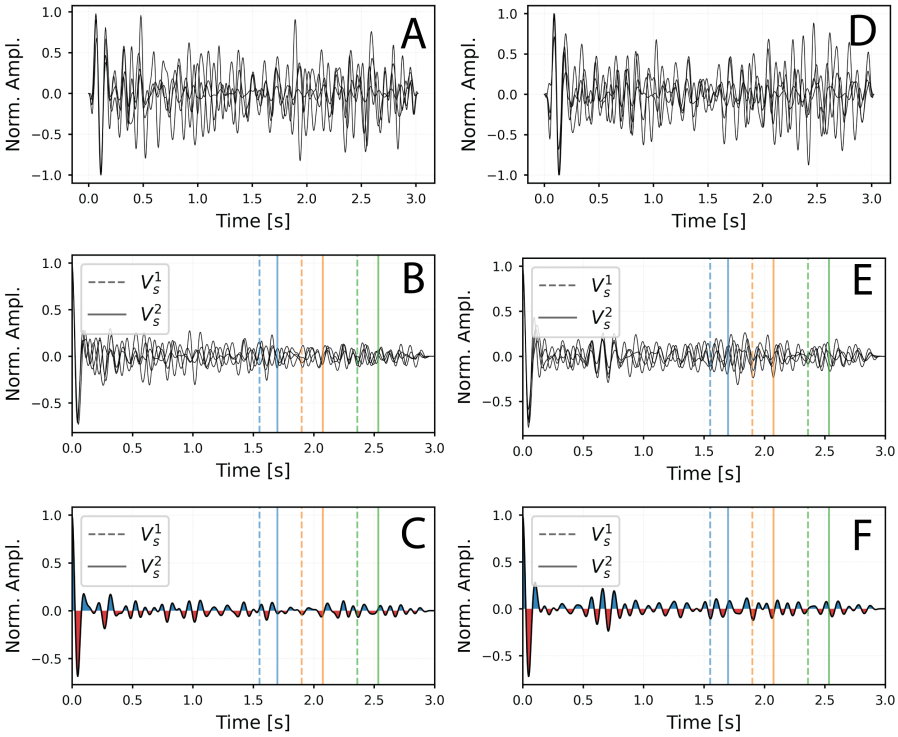


Figure 3.11: Analysis of ZO-VRR retrieved at receiver 001. The left column shows the N–S component (EHN), and the right column shows the E–W component (EHE). A. EHN-component filtered traces showing the time window that is autocorrelated. B. EHN-component causal part of the autocorrelation function of each seismic event overlaid on each other. C. EHN-component causal part of the ZO-VRR trace after stacking. D. EHE-component filtered traces showing the time window that is autocorrelated. E. EHE-component causal part of the autocorrelation function of each seismic event overlaid on each other. F. EHE-component causal part of the ZO-VRR trace after stacking. Vertical lines denote predicted two-way travel times to the top of NU (blue), NLNM (orange), and CK (green) for a source and receiver co-located on the surface.

3.8. DISCUSSION

Validating seismic velocity models is essential for reliable hypocentre estimation, yet direct approaches such as active-source surveys are often too costly for monitoring geothermal-heat-production projects. In this chapter, we demonstrate how body-wave seismic interferometry (SI) applied to local microseismic events can provide an independent measurement for assessing seismic-velocity models. Using six low-magnitude seismic events recorded during geothermal-heat production in Kwintsheul, we retrieve zero-offset virtual reflection responses (ZO-VRRs), which are then compared with the predicted times of the available seismic-velocity models. Our results show that the retrieved reflections strongly support the validity of the P-wave velocity model for the shallow subsurface. In contrast, both S-wave models fail to reproduce the observed arrivals.

Despite the limited number of seismic events, the vertical ZO-VRR trace shows clear reflections at the expected P-wave arrival times of the NU and NLNM reflectors. These reflections appear as a tuned wavelet, which is consistent with the slight depth separation between the two units. The CK reflector also coincides with a distinct retrieved arrival. Note that these arrivals appear to form part of a longer energy package. These arrivals might be a result of the SI retrieval process, i.e., due to insufficient constructive and destructive interference to form a separate event with a high SNR. On the other hand, if we trust the retrieved result, it means that there are more reflectors in the subsurface, which are not present in the velocity model. We see a weak retrieved event at the expected arrival time of the KN reflector, which may be due to attenuation, limited energy at depth, or a weak impedance contrast. The good match between retrieved and predicted arrivals confirms that the P-wave velocity model is reasonably reliable for the shallow subsurface. However, the number of available events is very limited, and longer monitoring periods are required to obtain more stable results. Deeper reflectors remain unconstrained, but with more events, this method could provide a cost-effective way to improve the P-wave velocity model at depth.

The horizontal-component ZO-VRR traces show a retrieved S-wave arrival for the NU reflector that falls between the predicted arrival times of the two S-wave velocity models. This suggests that neither model accurately captures the true arrival time. For the NLNM and CK reflectors, each model predicts an arrival time that coincides with a different retrieved event on both horizontal components. These observations suggest that neither the S-wave velocity model consistently matches the data, and are likely unreliable.

A longer monitoring period, with more detected seismic events, would allow improved stacking, reduce uncertainties, and provide more reliable constraints on the velocity structure. To complement the longer monitoring period, the sensors should be deployed at depth, as instrument burial significantly reduces surface-related noise. For example, Ruigrok and Dost (2019) showed that installing a sensor at 200 m depth reduced noise power by 27 dB. Such a setup could improve the signal-to-noise ratio and enhance the reliability of the retrieved P- and S-wave ZO-VRR traces.

3.9. CONCLUSIONS

We demonstrated the potential for validating seismic-velocity models using seismic interferometry in geothermal-heat production projects where microseismic events have been

recorded with near-vertical source–receiver ray paths. Using six microseismic events that were detected at the Kwintsheul geothermal project, we retrieved zero-offset virtual reflection responses separately for recordings on the vertical and two horizontal components. The vertical-component recordings correspond to the retrieved P-wave reflections, while the results from the horizontal-component recordings correspond to retrieved S-wave reflections. The retrieved P-wave reflections show good agreement with several predicted arrivals, confirming that the P-wave velocity model is more or less reliable for the shallow subsurface. However, deeper reflectors remain unconstrained, as only a limited number of events were available.

The predicted arrival times from the two available S-wave velocity models coincide with different retrieved S-wave reflections. This showed that the S-wave velocity models are not sufficiently reliable, as both could be valid. On the other hand, the comparison means that the S-wave results have also retrieved reflections. The limited number of events prevents sufficient stacking to suppress noise and non-physical arrivals, making robust S-wave model validation challenging.

A longer monitoring period is needed to detect more microseismic events. More events would improve the stability of the interferometric results, reduce uncertainties, and provide a stronger basis for refining both P- and S-wave velocity models. A future study based on a larger dataset could also explore the potential to assess anisotropy. Finally, the quality of the retrieved reflections was strongly influenced by high noise levels at the surface. Future monitoring campaigns should consider burying receivers to reduce ambient noise and improve the signal-to-noise ratio. Overall, we showed that seismic interferometry with microseismic events is a cost-effective and promising approach for local velocity-model validation and structural imaging, but its full potential requires longer-term monitoring, higher event counts, and the detection of local microseismicity.

REFERENCES

Aki, K. and Richards, P. G. (1980). *Quantitative seismology : theory and methods*. University Science Books Mill Valley, California. http://ci.nii.ac.jp/ncid/BA00498583.

Claerbout, J. F. (1968). Synthesis of a layered medium from its acoustic transmission response. *GEOPHYSICS*, 33(2):264–269. https://doi.org/10.1190/1.1439927.

Daneshvar, M. R., Clay, C. S., and Savage, M. K. (1995). Passive seismic imaging using microearthquakes. *Geophysics*, 60(4):1178–1186. https://doi.org/10.1190/1.1443846.

Deichmann, N. and Garcia-Fernandez, M. (1992). Rupture geometry from high-precision relative hypocentre locations of microearthquake clusters. *Geophysical Journal International*, 110(3):501–517.

Draganov, D., Wapenaar, K., and Thorbecke, J. (2004). Passive seismic imaging in the presence of white noise sources. *The Leading Edge*, 23(9):889–892. https://doi.org/10.1190/1.1803498.

Galetti, E. and Curtis, A. (2012). Generalised receiver functions and seismic interferometry. *Tectonophysics*, 532-535:1–26. https://doi.org/10.1016/j.tecto.2011.12.004.

Hassing, S. H. W., Draganov, D., Janssen, M., Barnhoorn, A., Wolf, K. A. A., Van Den Berg, J., Friebel, M., Van Otten, G., Poletto, F., Bellezza, C., Barison, E., Brynjarsson, B., Hjörleifsdóttir, V., Obermann, A., Sánchez-Pastor, P., and Durucan, S. (2024). Imaging $CO2_2$ reinjection into basalts at the CarbFix2 reinjection reservoir (Hellisheiði, Iceland) with body-wave seismic interferometry. *Geophysical Prospecting*, 72(5):1919–1933. https://doi.org/10.1111/1365-2478.13472.

Heimann, S., Vasyura-Bathke, H., Sudhaus, H., Isken, M. P., Kriegerowski, M., Steinberg, A., and Dahm, T. (2019). A Python framework for efficient use of pre-computed Green's functions in seismological and other physical forward and inverse source problems. *Solid Earth*, 10(6):1921–1935. https://se.copernicus.org/articles/10/1921/2019/se-10-1921-2019.html.

Nishitsuji, Y., Minato, S., Boullenger, B., Gomez, M., Wapenaar, K., and Draganov, D. (2016). Crustal-scale reflection imaging and interpretation by passive seismic interferometry using local earthquakes. *Interpretation*, 4(3):SJ29–SJ53. https://doi.org/10.1190/int-2015-0226.1.

Pluymaekers, M., Doornenbal, J., and Middelburg, H. (2017). TNO 2017 R11014 with erratum page 67 Final, Velmod-3.1. *TNO report*. https://www.nlog.nl/sites/default/files/2018-11/060.26839%20R11014%20with%20erratum%20page%2067%20Doornenbal-final.sec_.pdf.

Polychronopoulou, K., Lois, A., and Draganov, D. (2019). Body-wave passive seismic interferometry revisited: mining exploration using the body waves of local microearthquakes. *Geophysical Prospecting*, 68(1):232–253.

Rickett, J. and Claerbout, J. (1999). Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring. *The Leading Edge*, 18(8):957–960. https://doi.org/10.1190/1.1438420.

Ruigrok, E. and Dost, B. (2019). Seismic Monitoring and Site-Characterization with Near-Surface Vertical Arrays. *25th European Meeting of Environmental and Engineering Geophysics*, pages 1–5. https://doi.org/10.3997/2214-4609.201902455.

Scherbaum, F. (1987). Seismic imaging of the site response using microearthquake recordings. Part II. Application to the Swabian Jura, southwest Germany, Seismic network. *Bulletin of the Seismological Society of America*, 77(6):1924–1944. https://doi.org/10.1785/bssa0770061924.

Simonet, G. (2024). S-wave velocity profile from inversion of 3-C beam forming and noise cross-correlation dispersion curves at the Kwintsheul geothermal field, Netherlands. *Bachelor's Thesis. Universite Grenoble Alpes, Delft University of Technology, and ISTerre*.

Van IJsseldijk, J., Ruigrok, E., Verdel, A., and Weemstra, C. (2019). Shallow crustal imaging using distant, high-magnitude earthquakes. *Geophysical Journal International*, 219(2):1082–1091. https://doi.org/10.1093/gji/ggz343.

Wapenaar, K. (2003). Synthesis of an inhomogeneous medium from its acoustic transmission response. *GEOPHYSICS*, 68(5):1756–1759. https://doi.org/10.1190/1.1620649.

Wapenaar, K., Draganov, D., Snieder, R., Campman, X., and Verdel, A. (2010). Tutorial on seismic interferometry: Part 1 — basic principles and applications. *GEOPHYSICS*, 75(5):75A195–75A209. https://doi.org/10.1190/1.3457445.

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

4

SEISMIC NETWORK DESIGN FOR MONITORING OFFSHORE

Information is the resolution of uncertainty.

Claude Shannon

The design of seismic networks is critical for monitoring geothermal energy as it determines the detectability of seismic events and the accuracy of their hypocentre estimations. To improve monitoring performance, we use Experimental Design (ED) to optimise the number, type, and placement of receivers for effective monitoring. Compared to onshore operations, offshore network design demands particular attention due to higher deployment costs, logistical constraints, and limited accessibility. To address these challenges, we explore a non-linear ED approach for monitoring offshore geothermal energy. We apply the network design approach to the K14-FA Carbon Capture and Storage site, an analogous injection operation offshore with similar monitoring goals as for geothermal energy. We first present the theoretical foundations of ED, followed by a description of the optimisation approach. Next, we describe the open-access data we use for designing the monitoring network. We then outline the workflow for network optimisation, along with the assumptions made about the expected data. Finally, we present the results of the optimised network design, discuss the limitations of the method, and outline future considerations for its application as more observational data become available.

4.1. INTRODUCTION

Passive seismic networks are installed around geothermal energy operations to characterise the seismic activity in the areas where they operate. The seismic activity indicates features such as stress changes, the presence of faults, or the opening of cracks caused by the injection of cold water into the subsurface (e.g., Okamoto et al., 2018). The hypocentre locations of the microseismic events help identify such features in the subsurface (Guo et al., 2025; Yu et al., 2023).

The accuracy of hypocentre estimations depends on several factors, including the distribution of seismic receivers, the accuracy of picking the onset of the phase arrivals, and the velocity model used to predict seismic phase arrivals. While velocity models can often be refined after data acquisition, poor azimuthal coverage of an event cannot be corrected post hoc. Therefore, it is important to optimise the attributes of the seismic network, including the number, type, and location of receivers, cost-effectively.

Experimental design (ED) aims to select network geometries that maximise the information about specific model parameters while accounting for practical constraints such as site accessibility, ambient-noise conditions, and deployment logistics (e.g., Kijko, 1977; Steinberg et al., 1995). The main advantage of ED is that it allows for optimising the experimental setup before any data are collected to answer specific scientific questions. For instance, can we distinguish between injection-related seismicity? Or: is the seismicity located on the fault plane?

There are two main approaches for ED. The first approach assumes linear relationships between the observed data and model parameters (e.g., Toledo et al., 2020; Curtis et al., 2004). The second approach is to assume the relationships between the observed data and model parameters are fully non-linear (e.g., Coles and Curtis, 2011; Callahan et al., 2025). Non-linear approaches are particularly appealing because they account for the full complexity of the relationship between observed data and model parameters. Coles and Curtis (2011) introduced a non-linear ED called the D_N-optimisation for geophysical problems. Bloem et al. (2020) compared the D_N-optimisation to other linear and non-linear design approaches. Their findings showed that while linear approaches offer predictable computation times per added receiver, they result in less effective configurations and require substantially more total computation time. In a recent paper, Strutz and Curtis (2025) introduced a D_N-optimisation method for designing seismic networks for monitoring volcanoes. Their method is ideal for scenarios where scarce prior data about hypocentre locations is available.

In this chapter, we extend the D_N-optimisation method for monitoring geothermal energy offshore. Specifically, we account for travel-time uncertainties from velocity models, magnitude-dependent detectability based on ground-motion prediction equations, and geological priors reflecting injection wells and mapped faults. We apply this workflow to the Dutch offshore K14-FA CCS site, used here as an analogue for offshore geothermal operations. Our workflow uses open-access geophysical data from the Netherlands. By doing so, we demonstrate how Experimental Design can produce operationally feasible network designs that improve the resolution of low-magnitude seismicity.

4.2. THEORY

4.2.1. NON-LINEAR DN-OPTIMISATION

Independent of prior knowledge, the design of an experiment determines how much information can be extracted from the observations. Lindley (1956) introduced a measure of the information provided by an experiment based on Shannon entropy[1]. The key idea behind this is that if the model parameters $\mathbf{m}$ are finite, the amount of information contained in the prior distribution can be measured by how much information would be needed before the value of $\mathbf{m}$ becomes known.

The expected information gain (EIG) from collecting data at specific locations is defined as the expected reduction in uncertainty, measured by the average difference in information content between the posterior and prior distributions (Lindley, 1956). Following Strutz and Curtis (2025), the EIG is expressed as:

$$\mathrm{EIG} = \mathbb{E}_{p(\mathbf{d}|\boldsymbol{\xi})}\left\{ I\left[p(\mathbf{m} \mid \mathbf{d}, \boldsymbol{\xi})\right] - I\left[p(\mathbf{m})\right]\right\}, \tag{4.1}$$

where I is Shannon's information content (Shannon, 1948), and $\mathbb{E}_{p(\mathbf{d}|\boldsymbol{\xi})}$ is the expectation over all possible data that is likely to be observed following the prior information. This measure allows us to compare the state of knowledge before and after an experiment, and to quantify the information gained from the data.

Evaluating experimental designs based on posterior distributions enables direct quantification of uncertainty in the inferred hypocentre locations. However, this approach is computationally expensive, as it requires repeated forward modelling and inversion for many candidate network designs and possible source locations. This limitation motivates the use of surrogate objective functions that capture the information content of a design without explicitly computing posterior distributions.

One such approach is to evaluate the separation of predicted observations in data space. Informative measurement configurations produce predicted data that are well separated for different source locations, thereby increasing the ability to distinguish between competing models. This can be quantified by maximising the entropy of the predicted data (Coles and Curtis, 2011). Shewry and Wynn (1987) showed that maximising the entropy of predicted data is mathematically equivalent to minimising the entropy of the posterior model distribution. As a result, optimising data-space entropy provides an efficient proxy for reducing posterior uncertainty in hypocentre inversions.

Figure 4.1 illustrates the principle underlying entropy-based network design using a simplified hypocentre example. Each row corresponds to a different receiver geometry, indicated by the white triangles in the left column. The blue and orange density contours show the prior distribution of possible hypocentre locations, which defines the region where seismicity may occur. From this prior, a subset of four representative hypocentres (black symbols) is selected to illustrate how information propagates through the predicted arrival times (second column) and the posterior (third column).

In the second column of Figure 4.1, we plot the predicted arrival times at receiver 1 (x-axis) and receiver 2 (y-axis), assuming a homogeneous seismic-velocity model. Black

[1] Shannon quantified the *average surprise* of a discrete random variable X as its entropy, given by $H(X) = -\sum_i p_i \log_2 p_i$ (Shannon, 1948, Eq. 9). In short, frequent events carry little information and inject redundancy, while rare events are highly informative.

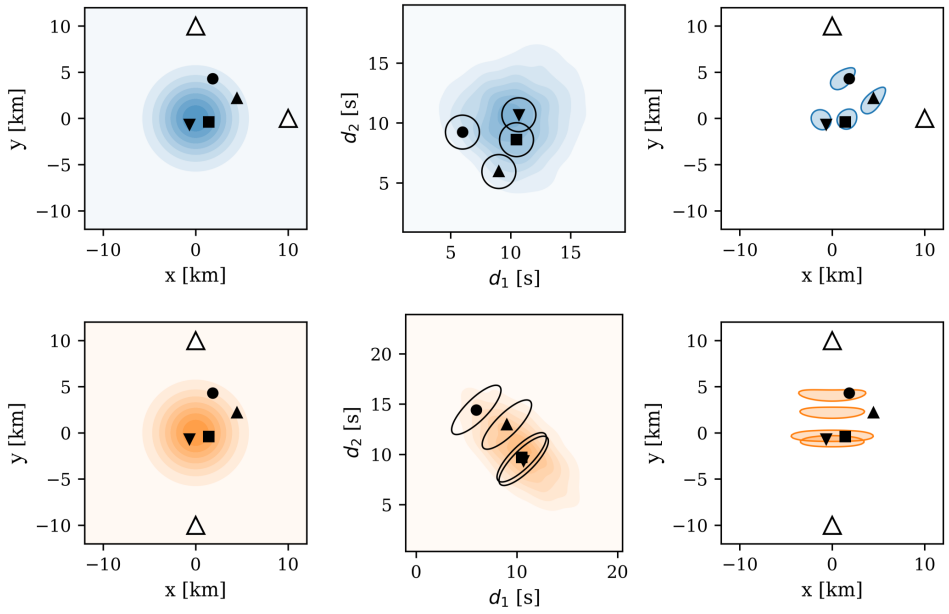


Figure 4.1: Conceptual illustration of the relationship between entropy in modelled data and in the posterior model. The left column shows the prior distribution of hypocentres. Each column shows a different network design where white triangles represent the receiver locations. From the prior distribution, a subset of candidate hypocentres (shown in black) is evaluated. The middle column shows the predicted arrival times (centre of each marker) and their corresponding uncertainty (black ellipse) of the subset of candidate hypocentres. The right column shows the resulting posterior distribution of the subset of candidate hypocentres. A location is erroneous when the marker is not inside its coloured posterior distribution. Modified after Strutz and Curtis (2025).

markers indicate the arrival times associated with the sampled hypocentre locations. Black ellipses represent the corresponding measurement uncertainty. The coloured density background shows the distribution of predicted data obtained by propagating all samples from the prior through the forward model. The first receiver geometry produces a broader spread of predicted data, corresponding to higher entropy in data space. This indicates a larger separation between the travel-time observations of different candidate source locations. In contrast, the second receiver geometry yields overlapping predicted arrival times, which reduces the separation between observed arrival times and limits the ability to distinguish between competing hypocentres.

In the third column of Figure 4.1, we show the posterior distributions, which represent the estimated hypocentre locations after accounting for uncertainty in the travel-time measurements. The receiver geometry in the first row produces a compact and well-constrained posterior distribution. Consistent with this behaviour, the same geometry also yields high entropy in the predicted data. By contrast, the low-entropy geometry produces elongated and overlapping posterior regions, indicating poorer constraint on the hypocentre location. This relationship demonstrates that maximising the entropy of the predicted data directly reduces posterior model uncertainty and motivates the network-design optimisation strategy adopted in this study.

Following Strutz and Curtis (2025), we rearrange the expression for the EIG to depend explicitly on the likelihood and the evidence $p(\mathbf{d} \mid \boldsymbol{\xi})$, rather than on the posterior distribution. This reformulation shifts the optimisation into data space, avoiding the need for repeated posterior evaluations. The derivation from Strutz and Curtis (2025) is reproduced below for clarity:

$$\begin{aligned}
\mathrm{EIG}(\boldsymbol{\xi}) &= \mathbb{E}_{p(\mathbf{d}|\boldsymbol{\xi})}\left[I\left(p(\mathbf{m} \mid \mathbf{d}, \boldsymbol{\xi})\right) - I\left(p(\mathbf{m})\right)\right] && (4.2)\\
&= \mathbb{E}_{p(\mathbf{d},\mathbf{m}|\boldsymbol{\xi})}\left[\log \frac{p(\mathbf{m} \mid \mathbf{d}, \boldsymbol{\xi})}{p(\mathbf{m})}\right] && (4.3)\\
&= \mathbb{E}_{p(\mathbf{d},\mathbf{m}|\boldsymbol{\xi})}\left[\log \frac{p(\mathbf{m}, \mathbf{d} \mid \boldsymbol{\xi})}{p(\mathbf{m})\, p(\mathbf{d} \mid \boldsymbol{\xi})}\right] && (4.4)\\
&= \mathbb{E}_{p(\mathbf{d},\mathbf{m}|\boldsymbol{\xi})}\left[\log \frac{p(\mathbf{d} \mid \mathbf{m}, \boldsymbol{\xi})}{p(\mathbf{d} \mid \boldsymbol{\xi})}\right] && (4.5)
\end{aligned}$$

$$\mathrm{EIG}(\boldsymbol{\xi}) = \mathbb{E}_{p(\mathbf{m})}\left[I\left(p(\mathbf{d} \mid \mathbf{m}, \boldsymbol{\xi})\right) - I\left(p(\mathbf{d} \mid \boldsymbol{\xi})\right)\right] \tag{4.6}$$

The optimal network design $\boldsymbol{\xi}^*$ is the configuration that maximises the EIG. This optimisation problem can be expressed mathematically as

$$\boldsymbol{\xi}^* = \underset{\boldsymbol{\xi} \in \Xi}{\arg\max}\, \mathrm{EIG}(\boldsymbol{\xi}), \tag{4.7}$$

D_N-optimisation was introduced for hypocentre-location problems by Coles and Curtis (2011). This optimisation is inspired by the determinant-based D-optimality criterion from linear design theory and extended to fully non-linear problems (Mitchell, 2000). The D_N approach approximates the evidence term by a multivariate Gaussian distribution and results in the following expression:

$$\mathrm{EIG}_{DN} = \frac{1}{N}\sum_{i=1}^{N} \log\left[p(d_i \mid m_i, \xi)\right] - \frac{3}{2}\left(1 + \log(2\pi)\right) + \frac{1}{2}\log|C_d|, \tag{4.8}$$

where N is the number of models of the prior distribution that are sampled, and the evidence term becomes

$$I\left[p(d_i \mid \xi)\right] = \frac{k}{2}\left(1 + \log(2\pi)\right) + \frac{1}{2}\log\left|C_d\right|, \tag{4.9}$$

where C_d is the covariance matrix of the data samples and k is the dimension of the data space (Strutz and Curtis, 2025).

4.3. WORKFLOW

Seismic experiments are designed to answer specific questions about the state of the subsurface. For injection-related operations, we aim to design a seismic network that can detect and characterise seismicity with sufficient resolution to (i) identify active faults, (ii) track fluid migration pathways, and (iii) monitor reservoir seismicity to ensure safe and sustainable operations. Therefore, this workflow has the following goals:

Goals of the Workflow

i) increase the detectability of seismic events within the reservoir or from faults near the reservoir, and ii) reduce the hypocentre uncertainties of such events.

While delineating fluid migration pathways is not addressed explicitly as a separate goal, it is inherently accounted for by increasing the hypocentre resolution within the reservoir bounds and near the injection wells. In this way, our workflow indirectly supports the detection of migration pathways through microseismicity patterns associated with fluid movement.

ED provides a framework to accomplish these goals with the optimal placement of seismic receivers by combining prior information and expected data to maximise the value of monitoring (Arnold and Curtis, 2018). In the following subsections, we introduce the steps of the workflow.

4.3.1. STEP 1: PRIOR DISTRIBUTION OF SEISMICITY

The workflow begins by defining a synthetic earthquake catalogue, which serves as our prior knowledge of expected seismicity. This step identifies the areas where monitoring must be most sensitive: (i) regions with a higher risk of induced seismicity, (ii) potential escape pathways where fluids could migrate to unwanted formations, (iii) critically stressed faults, and (iv) the reservoir itself.

Prior information can be derived from historical seismicity, mapped geological structures, or fault databases. In synthetic studies, priors are often constructed from reservoir geometry or anticipated injection zones, assuming either a uniform distribution within the target volume or weighted probabilities along faults or other critical features.

Here we define two types of prior using the fault and reservoir data described in Sections 4.4.1 and 4.4.2.

- **Prior 1—Microseismic cloud around the injection well:** We assume that induced events cluster around the planned injection point (see Section 4.4.1). Their synthetic hypocentres follow a Gaussian distribution centred on the injection point of the well, with a one standard-deviation uncertainty of approximately 100 m in the x, y, and z directions.

- **Prior 2—seismicity along a fault plane:** We create a synthetic earthquake catalogue by placing hypocentres along mapped faults (data described in Section 4.4.2). To represent the spatial uncertainty, we discretise each fault into N points $\{\boldsymbol{\mu}_i\}_{i=1}^{N}$ in 3-D space. Around each point, earthquakes are distributed using anisotropic Gaussian kernels that spread more within the fault plane than perpendicular to it. This anisotropy is encoded in the diagonal covariance matrix

$$\mathbf{C} = \mathrm{diag}\left(\sigma_{xy}^2, \sigma_{xy}^2, \sigma_z^2\right), \tag{4.10}$$

  where σ_{xy} controls the in-plane uncertainty and σ_z the off-plane uncertainty.

  The synthetic hypocentres $\mathbf{x}$ are then generated using a Gaussian-mixture model

(GMM) with uniform weights,

$$p(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{N}(\mathbf{x}\,|\,\boldsymbol{\mu}_i, \mathbf{C}), \tag{4.11}$$

which concentrates probability along the fault while admitting realistic spatial uncertainty. Figure 4.2.A shows a schematic of the GMM and the synthetic hypocentres in a 3D plane. The final distribution is shown in Figure 4.2.B.

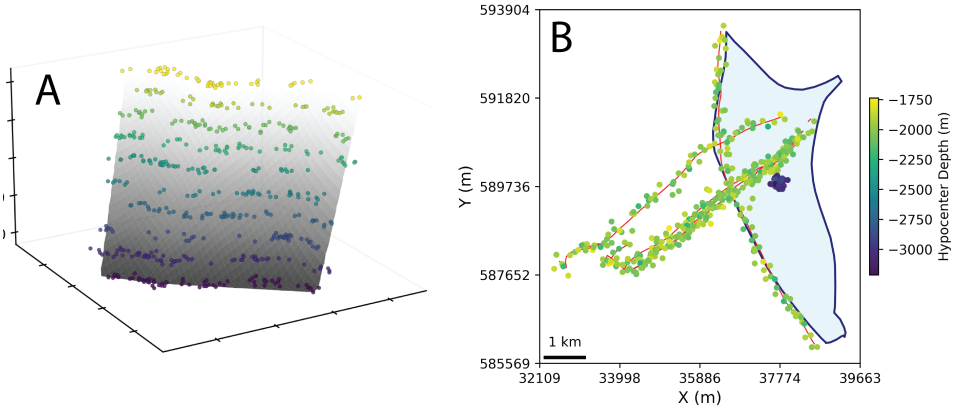


Figure 4.2: Synthetic seismicity used as prior information. A. Example of a Gaussian-mixture representation of a mapped fault where the fault is shown in grey with synthetic hypocentres colored by depth. B. Final catalogue of synthetic events coloured by depth. Faults are shown in red and the reservoir outline in blue.

4.3.2. STEP 2: DESIGN SPACE

The next step defines the design space, that is, the set of plausible locations where seismic receivers can be deployed. We define it by first selecting the smallest rectangle that could fit all the events in the prior distribution. Then we add 10 km distance away from the rectangle in all directions. Inside the buffered rectangle we place 400 trial sites with a Sobol low-discrepancy sequence. A Sobol pattern gives an even spread without the rigid look of a grid. All candidates are set at the surface level ($z = 0$).

The receiver grid can be seen in Figure 4.3.

4.3.3. STEP 3: TRAVEL-TIME ESTIMATION

After defining the design space and prior seismicity, the next step is to calculate the theoretical phase arrivals ($\mathbf{t_{calc}}$) from the synthetic hypocentres ($\mathbf{x_s}$) to the candidate receivers (ξ). $\mathbf{t_{calc}}$ is expressed as

$$\mathbf{t_{calc}}(\mathbf{x_s}, \xi) = u(\mathbf{x_s},\ \xi) + \boldsymbol{\epsilon}(\xi). \tag{4.12}$$

The forward function u estimates the propagation time of a seismic phase from a hypocentre to a receiver, or in this case, to an array of receivers ξ. $\boldsymbol{\epsilon}(\xi)$ is the noise affecting the observed data.

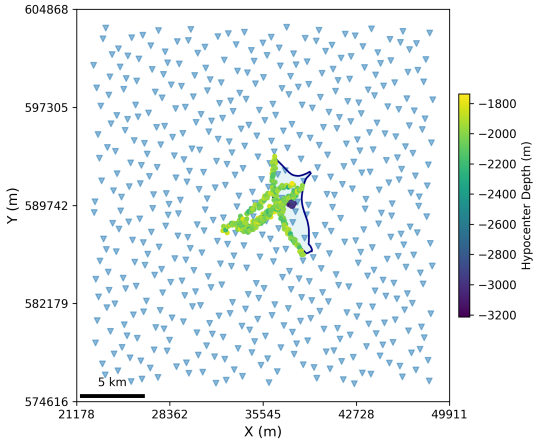


Figure 4.3: Design space showing candidate receiver locations generated with a Sobol-scrambled sequence (blue triangles) and the prior seismicity distribution coloured by depth.

As done in Chapter 2, we calculate $\mathbf{t_{calc}}$ using the fast Marching Method (FMM) (Sethian and Popovici, 1999) implemented in the `PyKonal` Python Package (White et al., 2020). All possible $\mathbf{t_{calc}}$ are precomputed and stored in a travel time table. The velocity model is described in Section 4.4.3.

4.3.4. STEP 4: TRAVEL TIME UNCERTAINTY

After estimating the theoretical phase arrivals, we need a measurement of their uncertainty. The covariance matrix of the travel-time observations accounts for two sources of uncertainty: the picking uncertainty and the uncertainty associated with the velocity model. Following Strutz and Curtis (2025), the combined variance $\sigma_t(t)^2$ for each travel-time observation is expressed as

$$\sigma(t)^2 = \sigma_{tobs}^2 + t \cdot \sigma_v^2, \tag{4.13}$$

where σ_{tobs} represents the picking uncertainty and σ_v is the relative uncertainty in the velocity model.

The picking uncertainty is presented in Section 4.4.4, which depends on the distance between each hypocentre-receiver pair.

The velocity-model uncertainty scales with travel time, following the approach of Strutz and Curtis (2025), who described this behaviour as analogous to a Gaussian random walk, where the mean-squared error increases with distance.

4.3.5. STEP 5: EVENT DETECTABILITY

We provide two different detectability thresholds for the study. The first case has no limitation, meaning the events signal would always reach the receivers within the array.

For the second case, we focus on the detectability of low-magnitude events with a target local magnitude $M_L = 2.0$. To estimate the distance range within which such events can be reliably detected, we use the ground motion prediction equation (GMPE) for P-wave peak ground velocity presented in 4.4.5.

We incorporate the detectability by making σ_{tobs} large when the detectability threshold is passed. Therefore, each receiver has a different pick uncertainty and blows up when the event is not detectable.

4.3.6. STEP 6: LIKELIHOOD ESTIMATION

The likelihood quantifies the probability of observing the recorded data given the model parameter values and a specific network design ξ. In our case, the likelihood is the probability of observing the arrival times of seismic phases from a given source and network configuration.

To account for measurement uncertainty, we model the observed arrival times as perturbed with random Gaussian errors. Consequently, the likelihood is represented as a multivariate Gaussian distribution centred on the predicted data, with a covariance matrix reflecting the expected measurement uncertainties.

Following Strutz and Curtis (2025), the likelihood of a source (sampled from the prior) using a specific network design is estimated following these steps:

1. Calculate the predicted arrival times t_{calc} of seismic phases that propagate from the source to each receiver using a forward model explained in Section 4.3.3.
2. Compute the covariance matrix, where each diagonal term σ_i combines picking uncertainty and travel-time-dependent velocity-model uncertainty (see Section 4.3.3).
3. Simulate observational noise by adding Gaussian random noise to the predicted arrival times, where the corresponding entry of the covariance matrix gives the variance for each observation.
4. Compute the likelihood over a network design ξ with i receivers as

$$\log p(\mathbf{d} \mid \mathbf{m},\ \xi) = -\frac{1}{2}\sum_i \left(\log(2\pi\sigma_i^2) + \frac{(t_i^{noisy} - t_i^{calc})^2}{\sigma_i^2} \right). \tag{4.14}$$

Note that we diverge from the likelihood term introduced in Chapter 2, which used the EDT formulation. Although this formulation is effective in reducing the effects of outlier data, it is computationally expensive.

4.3.7. STEP 7: OPTIMISATION

DN CRITERIA

At this point, the priors, design space, velocity-model uncertainties, and detectability constraints have already been defined. The objective is to select the configuration of M receivers that maximises the EIG. A detailed derivation of the D_N-optimisation criterion was

given in Section 4.2.1. For clarity, we reintroduce here the expression used in the optimisation:

$$\mathrm{EIG}_{DN} = \frac{1}{N}\sum_{i=1}^{N} \log\big[p(d_i \mid m_i, \xi)\big] - \frac{3}{2}\big(1 + \log(2\pi)\big) + \frac{1}{2}\log|C_d|, \tag{4.15}$$

where N is the number of prior samples, $p(d_i \mid m_i, \xi)$ is the likelihood of the travel-time data for model m_i and design ξ, and C_d is the covariance matrix of the predicted data.

To solve this optimisation problem, we use a discrete variant of the *Differential Evolution* (DE) algorithm implemented in the `Scipy` library. DE is a global, population-based search method that evaluates many network designs simultaneously, making it well-suited for stochastic objective functions with multiple local optima. Each iteration proposes trial receiver sets, evaluates their EIG_{DN}, and retains the best-performing candidates until convergence.

We constrain the search to select $M = 13$ receivers from the design space for two detectability scenarios: (i) unlimited detection, where all events are observable by all receivers, and (ii) magnitude-limited detection, where we simulate a catalogue of events with a local magnitude (M_L) value of 2.0. The resulting network designs are presented in Section 4.5.

NMC CRITERIA

After convergence, we validate the final design with the unbiased but more expensive nested Monte Carlo (NMC) estimator,

$$\mathrm{EIG}_{\mathrm{NMC}}(\xi) = \frac{1}{N}\sum_{i=1}^{N}\left\{\log p(\mathbf{d}_i \mid \mathbf{m}_i, \xi) - \log\left[\tfrac{1}{N}\sum_{j=1}^{N} p(\mathbf{d}_i \mid \mathbf{m}_j, \xi)\right]\right\}, \tag{4.16}$$

to confirm robustness to the Gaussian-evidence approximation (Strutz and Curtis, 2025). The optimization is performed in the same way as in Section 4.3.7, but using the *NMC* estimator.

4.3.8. STEP 8: VALIDATION—POSTERIOR DISTRIBUTION

The posterior distribution $p(\mathbf{m} \mid \mathbf{d}, \xi)$ combines the prior information and the likelihood of the observed data to update our understanding of the model parameters $\mathbf{m}$. The posterior fully characterises the uncertainties, correlations, and most likely values of the model parameters given the observed data and network design. Additionally, it can be used to describe higher-order properties, such as the mean, spread, shape, skewness, and potential multimodality of the parameter distributions.

To assess whether the optimised network delivers the expected precision, we ran a synthetic mislocation experiment that mirrors the standard EDT processing introduced in Chapter 2. Each hypocentre is treated as a “true" origin; its theoretical P–arrival times at the selected receivers are extracted from the heterogeneous velocity cube, perturbed with zero-mean Gaussian noise ($\sigma = 0.015\,\mathrm{s}$), and re-inverted on the same 3D grid with the EDT likelihood normalised over the full volume. Every inversion returned (i) the maximum-likelihood grid node, (ii) its 3D mislocation error, (iii) marginal 1σ uncertainties in x, y, and z, and (iv) the azimuthal gap at the estimated epicentre.

4.4. DATA

In this section, we introduce the open-access geological and geophysical datasets from the Netherlands to design and test seismic networks for seismic monitoring.

4.4.1. SITE INFORMATION—ARAMIS CCS FIELD

As a case study, we consider the K14-FA field, part of the Aramis CCS project in the Dutch North Sea (Fig. 4.4). In this field, the goal is to inject up to 2.5 million tonnes of CO_2 annually, with operations expected to commence in 2028 and continue for approximately 15 years (Ministerie van Economische Zaken en Klimaat, 2024).

The field lies approximately 90 km northwest of Den Helder. The target reservoir is located in the Upper Rotliegend Group at a depth of approximately 2.8 km, with a gas-water contact at 3.1 km. The reservoir is overlain by claystone and rock-salt units of the Zechstein Group, which provide an effective seal.

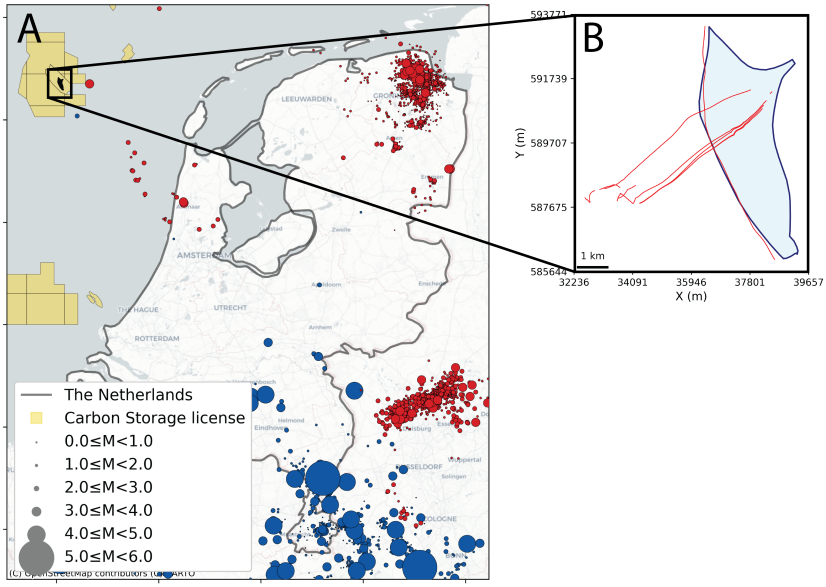


Figure 4.4: Location of the K14-FA CO_2 sequestration site. A. Regional map showing the location of all CCS licenses (yellow), natural seismicity (blue), and induced seismicity (red). B. Geometry of the K-14 field in blue and local faults in orange.

4.4.2. FAULT GEOMETRIES

Our network design strategy focuses on monitoring seismic activity related to faults in the vicinity of the fluid injection points. We simulate seismicity along faults that could lead to the leakage of CO_2 into the atmosphere or that could trigger unwanted seismic activity. Here we use mapped faults in the Dutch offshore from the European Fault Database (v. Gessel et al., 2021). In section 4.3.1, we explain how we construct the prior distribution

of seismicity along the fault planes. These data delineate regions of elevated seismic hazard where enhanced monitoring sensitivity is most desirable.

4.4.3. SEISMIC-VELOCITY MODEL

We use Velmod 3.1 (Pluymaekers et al., 2017) as the reference seismic velocity model. This model provides P- and S-wave velocities, together with their uncertainties, for each stratigraphic layer across the Dutch subsurface. It serves as the basis for travel-time calculations in our workflow.

4.4.4. OBSERVED TRAVEL-TIME UNCERTAINTIES

Observed P- and S-wave travel-time uncertainties are taken from Ruigrok et al. (2022). These values were determined from Dutch seismic recordings for distances up to 160 km and are used to parametrise the covariance matrix of the predicted arrivals. Standard deviations are provided for local (0–20 km), regional (20–60 km), and national (60–160 km) ranges (Table 4.1).

Naming	Distance range [km]	σ_P [s]	σ_S [s]
Local	0 – 20	0.115	0.186
Regional	20 – 60	0.162	0.322
National	60 – 160	0.295	0.568
All	0 – 160	0.137	0.248

Table 4.1: Travel-time uncertainties σ_P and σ_S as a function of distance range, adapted from Ruigrok et al. (2022).

4.4.5. GROUND-MOTION PREDICTION EQUATION

We use the GMPE of Ruigrok et al. (2022) to determine the maximum epicentral distance at which a receiver can detect a seismic event. This GMPE describes the attenuation of P-wave ground-motion amplitudes with distance and was calibrated using induced earthquakes from the Groningen region. It provides magnitude-dependent estimates of amplitude decay with distance and allows us to define realistic detection thresholds. Figure 4.5 shows the predicted decay of peak ground velocity with epicentral distance for low-magnitude events ($M_L = 2.0$). The red line indicates the minimum epicentral distance at which an event is expected to be detectable.

The GMPE is given by:

$$\ln(Y_{\text{mod}}) = c_1 + c_2 M + g(R^*), \tag{4.17}$$

where Y_{mod} is the modelled modelled peak ground velocity (in mm/s), M is the local magnitude, and $g(R^*)$ accounts for the geometric spreading and attenuation effects. The effective distance R^* is defined as

$$R^* = \sqrt{R^2 + D^2 + \left(e^{e_1 M + e_2}\right)^2}, \tag{4.18}$$

where R is the epicentral distance, D is the event depth, and the last term accounts for

magnitude-dependent saturation. The function $g(R^*)$ takes a piecewise form:

$$g(R^*) = \begin{cases} c_4 \ln(R^*) & \text{if } R^* \leq d, \\ c_4 \ln(d) + c_{4a} \ln\left(\frac{R^*}{d}\right) & \text{if } R^* > d, \end{cases} \tag{4.19}$$

with d representing a transition distance beyond which the decay rate changes. The numerical values of the GMPE coefficients are listed in Table 4.2.

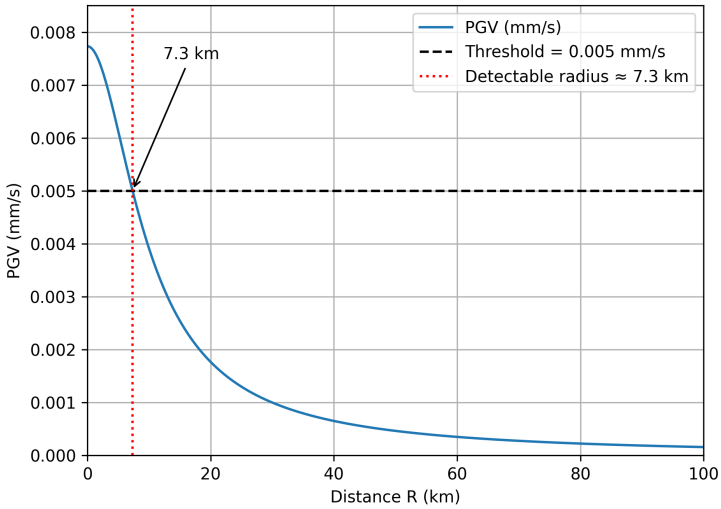


Figure 4.5: Predicted peak-ground-velocity decay with epicentral distance based on the adopted GMPE. The red line marks the detection threshold for $M_L = 2.0$ events.

Coefficient	Value
c_1	-0.20
c_2	1.96
c_4	-3.44
c_{4a}	-1.62
e_1	0.45
e_2	-0.80
d $[km]$	8.0

Table 4.2: GMPE coefficients from Ruigrok et al. (2022) used in this study.

4.5. RESULTS

The resulting network designs are shown in Figure 4.6. The top and middle rows show the optimisation without setting a magnitude-detection threshold, which assumes that all of the receivers can detect any of the seismic phases from the events in the prior. The top row shows the optimization using the D_N-optimization criteria. We can see that most of the

selected receivers are distributed following the orientation of the mapped faults. A similar pattern, but with receivers closer to the centre of the field, can be seen in the middle row, which corresponds to the design that is optimized using the NMC-criteria. The bottom row of Figure 4.6 shows the DN-EIG optimisation results obtained when assuming a uniform event magnitude of $M_L = 2$. As expected, networks optimized for detecting events with $M_L = 2$ are concentrated closer to the centre of the array, because signal amplitudes attenuate more rapidly with distance. Consequently, configurations with larger apertures are less effective at detecting such low-magnitude events.

The validation of the network hypocentre resolution is shown in Figure 4.7. The network design optimised using the EIG–D_N criterion (left column) yields hypocentral depth uncertainties ranging from approximately 200 m to 1.2 km. Most events located at reservoir depth exhibit depth uncertainties of around 800 m. Depth uncertainties are estimated using the 95% confidence interval (see Section 2). The corresponding azimuthal gaps (middle row) range between 80° and 110°, with a small number of events reaching values up to approximately 130°. The average Euclidean distance between the true and recovered hypocentre locations is approximately 120 m, as shown in the bottom row.

The network design optimised using the EIG–NMC criterion (right column) shows overall improved performance. Hypocentral depth uncertainties are reduced to values between approximately 200 m and 900 m, while most reservoir-level events remain resolved with uncertainties close to 800 m. Azimuthal gaps are generally smaller, ranging between 50° and 110°. This improvement is also reflected in the reduced average Euclidean mislocation error of approximately 100 m.

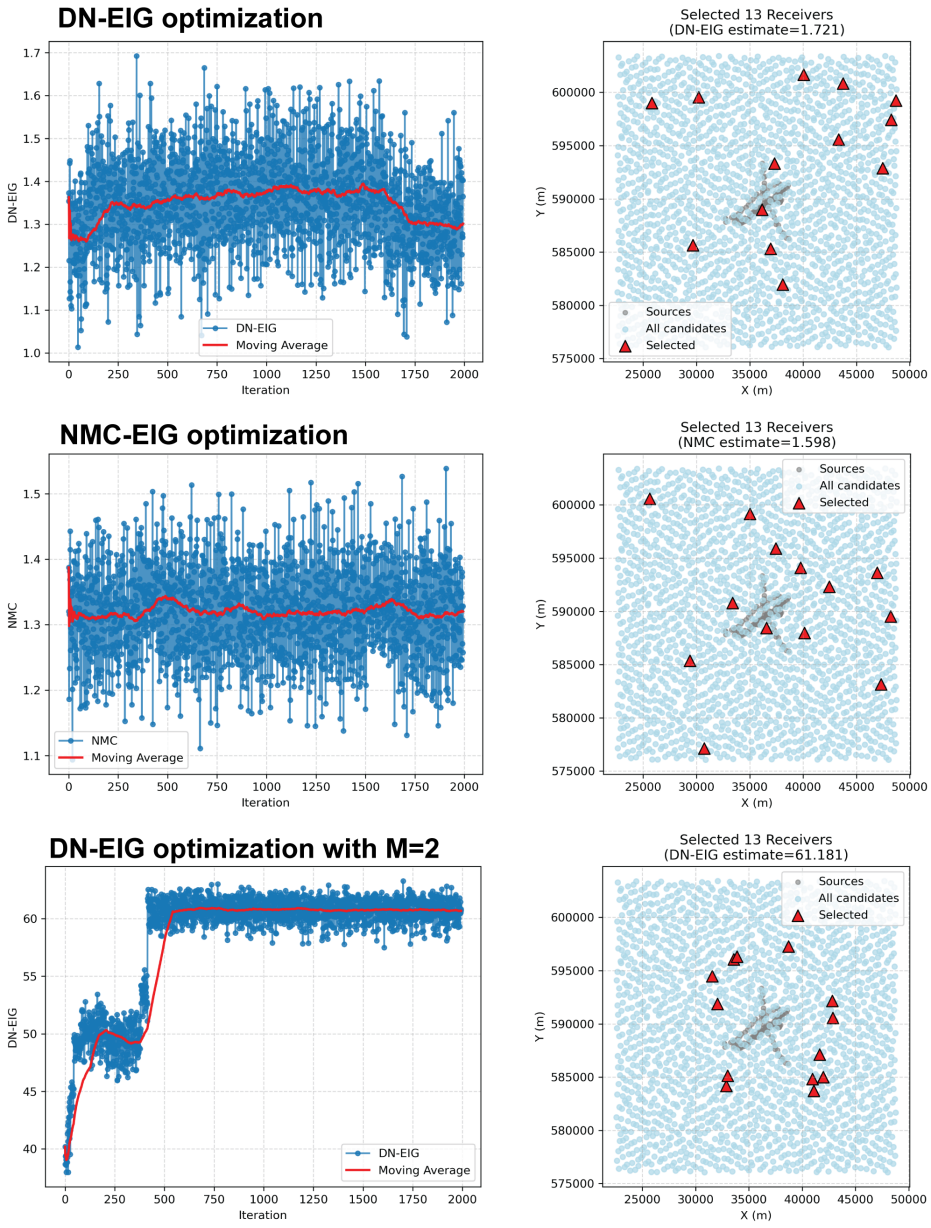


Figure 4.6: Results of the expected-information-gain (EIG) optimization. Left: evolution of the objective function over iterations, with the moving average shown in red. Right: resulting receiver layouts with blue circles marking all candidate receiver sites, red triangles marking the selected receivers, and grey points marking the simulated sources. The first row shows the design using the $EIG - D_N$ criteria. The middle row shows the design using the NMC optimization criteria. The bottom row shows the magnitude-limited case ($M_L = 2$) using the $EIG - D_N$ criteria.

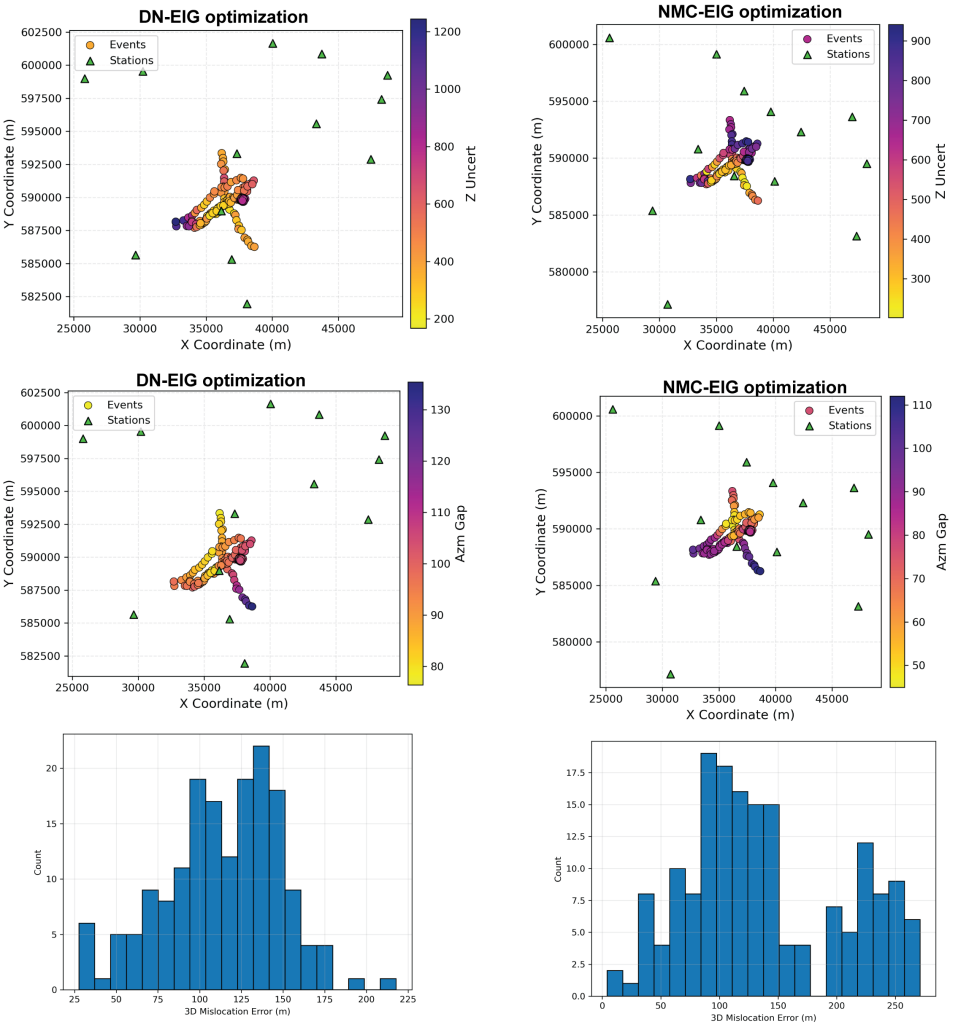


Figure 4.7: Validation of the optimized network designs. The Depth uncertainties (top row), azimuthal gaps (middle row), and 3D mislocation errors computed as Euclidean distances between true and recovered hypocentres (bottom row) are shown for the D_N-optimization (left column) and the NMC-optimization (right column) schemes.

4.6. DISCUSSION

We have presented a seismic-network-design workflow for monitoring geothermal-energy or CCS operations. We applied the workflow to design a network in the K-14 offshore field for monitoring a CCS operation that is in the implementation phase. The workflow integrates open-access subsurface information from the Netherlands (e.g., fault structures, seismic-velocity models, reservoir geometry, and borehole locations). We used this information to define the monitoring targets and simulate a prior distribution of expected seismicity. For the design process, we incorporated two non-linear design strategies that are rooted in well-established linearised design strategies, namely, the D_N- and NMC-optimisation. We also carried out the optimisation using a global search based on a sequential evolution algorithm to find approximately globally-optimal network designs. We included a validation of the designs based on the synthetic estimation of hypocentres from events in the prior.

Applying this approach to the K14-FA field, we found that the optimal D_N-optimised configuration yielded hypocentral depth uncertainties ranging from approximately 200 m to 1.2 km and mean 3-D mislocation errors of about 120 m. The NMC-optimised configuration showed improved performance, with depth uncertainties between 200 m and 900 m, and a lower mean 3-D mislocation error of about 100 m. The similar spatial receiver distributions obtained from both optimisation strategies indicate a stable network solution. These uncertainty levels define the expected monitoring performance of the proposed network design using 12 receivers placed on the sea bottom. Additional instrumentation, such as borehole stations (geophones or distributed acoustic sensing placed along a borehole), should be considered when lower spatial uncertainties are required.

The array optimised for detecting events with $M_L < 2$ places more stations closer to the expected epicentres. While this configuration improves the detectability of low-magnitude events, the smaller array aperture limits its ability to improve hypocentral resolution. Combining both design strategies—the unconstrained in magnitude for resolution and the constrained for detectability—can enhance overall network performance.

The proposed workflow requires extension in specific scenarios. First, we optimized and validated our workflow for areas defined as high-risk or high-importance. This approach overlooks the sensitivity of the network in other areas of the subsurface. If the sensitivity in other areas of the subsurface needs to be evaluated, we recommend implementing the simulated magnitude of completeness approach of (Ruigrok et al., 2022). Second, if the projects are located around seismically-active zones, the prior-seismicity should simulate magnitudes according to the Gutenberg-Richter magnitude–frequency distribution (Gutenberg and Richter, 1944) as done in Kraft et al. (2013). Third, travel-time errors can be correlated between receivers when their inter-station separation is short or if ambient noise is correlated. The influence of travel-time-correlated errors can be modelled using ensembles of velocity models, as demonstrated by Callahan et al. (2025). Lastly, ambient-seismic-noise conditions influence the detectability of seismic events. Noise constraints can be incorporated using offshore hydrophone-based noise models from the JOMOPANS database (MacGillivray and De Jong, 2021) and onshore noise power spectral density estimates derived from existing monitoring stations.

4.7. CONCLUSIONS

Seismic-network design optimizes the placement of sensors to improve the resolution required to monitor seismicity associated with geothermal or carbon-capture-and-storage operations. We presented a network-design workflow that uses public information about the Dutch subsurface. The workflow is based on well-established experimental design approaches. We demonstrated the workflow by designing the network for a carbon-capture-and-storage project in the North Sea. We validated the resulting designs to locate microseismicity down to $M_L \approx 2$ in the zones defined as high-risk or high-interest. The validation step in our workflow provides the limitations of the hypocentre resolution. If a higher hypocentre resolution is required, the addition of borehole seismic stations should be considered. Our workflow can be extended to incorporate updated velocity, noise, and attenuation models. We hope that our workflow will reduce the cost barrier associated with monitoring geothermal operations and provide optimised networks to meet specific monitoring goals.

REFERENCES

Arnold, R. and Curtis, A. (2018). Interrogation theory. *Geophysical Journal International*, 214(3):1830–1846.

Bloem, H., Curtis, A., and Maurer, H. (2020). Experimental design for fully nonlinear source location problems: Which method should I choose? *Geophysical Journal International*, 223(2):944–958. 10.1093/gji/ggaa358.

Callahan, J., Monogue, K., Villarreal, R., and Catanach, T. (2025). Analysis and optimization of seismic monitoring networks with Bayesian optimal experimental design. *Geophysical Journal International*, 240(3):1802–1824.

Coles, D. and Curtis, A. (2011). Efficient nonlinear Bayesian survey design using DN optimization. *Geophysics*, 76(2):Q1–Q8. https://doi.org/10.1190/1.3552645.

Curtis, A., Michelini, A., Leslie, D., and Lomax, A. (2004). A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys. *Geophysical Journal International*, 157(2):595–606. https://doi.org/10.1111/j.1365-246x.2004.02114.x.

Guo, H., Thurber, C., Cunningham, E., Cardiff, M., Lord, N., Sobol, P., Wang, H., and Feigl, K. L. (2025). Microseismicity modulation due to changes in geothermal production at San Emidio, Nevada, USA. *Geophysical Research Letters*, 52(4). 10.1029/2024gl112063.

Gutenberg, B. and Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, 34(4):185–188. https://doi.org/10.1785/bssa0340040185.

Kijko, A. (1977). An algorithm for the optimum distribution of a regional seismic network—I. *pure and applied geophysics*, 115(4):999–1009. 10.1007/BF00881222.

Kraft, T., Mignan, A., and Giardini, D. (2013). Optimization of a large-scale microseismic monitoring network in northern Switzerland. *Geophysical Journal International*, 195(1):474–490. https://doi.org/10.1093/gji/ggt225.

Lindley, D. V. (1956). On a Measure of the Information Provided by an Experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005. 10.1214/aoms/1177728069.

MacGillivray, A. and De Jong, C. (2021). A reference spectrum model for estimating source levels of marine shipping based on automated identification system data. *Journal of Marine Science and Engineering*, 9(4):369. https://doi.org/10.3390/jmse9040369.

Ministerie van Economische Zaken en Klimaat (2024). Ontwerp-omgevingsvergunning mijnbouwinstallatie k14-fa-2: Aramis - fase 1. https://www.rvo.nl/onderwerpen/bureau-energieprojecten/lopende-projecten/aramis. Letter addressed to Shell Gas & Power Development B.V., including advice from Staatstoezicht op de Mijnen.

Mitchell, T. J. (2000). An Algorithm for the Construction of "D-Optimal" Experimental Designs. *Technometrics*, 42(1):48. 10.2307/1271431.

Okamoto, K., Yi, L., Asanuma, H., Okabe, T., Abe, Y., and Tsuzuki, M. (2018). Triggering processes of microseismic events associated with water injection in Okuaizu Geothermal Field, Japan. *Earth Planets and Space*, 70(1). 10.1186/s40623-018-0787-7.

Pluymaekers, M., Doornenbal, J., and Middelburg, H. (2017). TNO 2017 R11014 with erratum page 67 Final, Velmod-3.1. *TNO report*. https://www.nlog.nl/sites/default/files/2018-11/060.26839%20R11014%20with%20erratum%20page%2067%20Doornenbal-final.sec_.pdf.

Ruigrok, E., Kruiver, P., and Dost, B. (2022). Construction of earthquake location uncertainty maps for the Netherlands. Technical Report Technical report TR-405, Royal Netherlands Meteorological Institute. Ministry of Infrastructure and Water Management.

Sethian, J. A. and Popovici, A. M. (1999). 3-d traveltime computation using the fast marching method. *Geophysics*, 64(2):516–523. https://doi.org/10.1190/1.1444558.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423. 10.1002/j.1538-7305.1948.tb01338.x.

Shewry, M. C. and Wynn, H. P. (1987). Maximum entropy sampling. *Journal of Applied Statistics*, 14(2):165–170. 10.1080/02664768700000020.

Steinberg, D. M., Rabinowitz, N., Shimshoni, Y., and Mizrachi, D. (1995). Configuring a seismographic network for optimal monitoring of fault lines and multiple sources. *Bulletin of the Seismological Society of America*, 85(6):1847–1857. 10.1785/bssa0850061847.

Strutz, D. and Curtis, A. (2025). Near-real-time design of experiments for seismic monitoring of volcanoes. *Seismica*, 4(1). https://seismica.library.mcgill.ca/article/view/1452.

Toledo, T., Jousset, P., Maurer, H., and Krawczyk, C. (2020). Optimized experimental network design for earthquake location problems: Applications to geothermal and volcanic field seismic networks. *Journal of Volcanology and Geothermal Research*, 391:106433. 10.1016/j.jvolgeores.2018.08.011.

v. Gessel, S., Hintersberger, E., v. Ede, R., ten Veen, J., Doornenbal, H., Diepolder, G., den Dulk, M., Hamiti, S., Vukzaj, N., Çako, R., Prendi, E., Ceroni, M., Mara, A., Barros, R., Tovar, A., Britze, P., Baudin, T., Stück, H., Jähne-Klingberg, F., Jahnke, C., Höding, T., Malz, A., Kristjánsdóttir, S., Þorbergsson, A., Di Manna, P., D'Ambrogi, C., Congi, M., Lazauskienė, J., Andriuškevičienė, G., Baliukevičius, A., Jarosiński, M., Gogołek, T., Stępień, U., Krzemińska, E., Salwa, S., Habryn, R., Aleksandrowski, P., Szynkaruk, E., Konieczyńska, M., Ressurreição, R., Machado, S., Moniz, C., Sampaio, J., Dias, R., Carvalho, J., Fernandes, J., Ramalho, E., Filipe, A., Celarc, B., Atanackov, J., Jamšek Rupnik, P., Shevchenko, A., Melnyk, I., and Lapshyna, A. (2021). The HIKE european fault database (EFDB) compiled in the framework of the GeoERA project HIKE.(2018-2021). *Deliverable GeoERA project HIKE, European Geological Data Infrastructure (EGDI)*. https://egdi.geology.cz/record/basic/5edf7bd4-9270-4188-b69d-7ddd0a010833.

White, M. C. A., Fang, H., Nakata, N., and Ben-Zion, Y. (2020). PyKonal: A Python Package for Solving the Eikonal Equation in Spherical and Cartesian Coordinates Using the Fast Marching Method. *Seismological Research Letters*, 91(4):2378–2389. https://doi.org/10.1785/0220190318.

Yu, P., Dempsey, D., Rinaldi, A. P., Calibugan, A., Ritz, V. A., and Archer, R. (2023). Association between injection and microseismicity in geothermal fields with multiple wells: Data-Driven modeling of Rotokawa, New Zealand, and Húsmúli, Iceland. *Journal of Geophysical Research Solid Earth*, 128(4). 10.1029/2022jb025952.

5

Clock error correction with Ambient Seismic Noise: Application to Reykjanes Geothermal Field

Hypocentre estimation relies on the arrival times of seismic waves measured by a network of receivers. The receivers need to be correctly synchronised in time to avoid erroneous hypocentre locations and misinterpretation of the underlying seismic processes. It is common to synchronise the receivers with the absolute time provided by a Global Navigation Satellite System (GNSS). When the receivers are Ocean-Bottom Seismometers (OBS), synchronisation is not possible as GNSS signals are not accessible from the ocean bottom. As a result, OBS experiences clock drift, which introduces receiver-specific timing errors. In this chapter, we introduce a new method that identifies and corrects clock errors in OBS. Our method leverages the theoretical time symmetry of time-averaged cross-correlations of ambient seismic noise; broken time symmetry is attributed to clock drift. We apply our method to the IMAGE seismic network, a two-year seismological experiment that monitored geothermal areas and seismic activity in the Reykjanes Peninsula (Iceland). Our results indicate that all IMAGE OBS experienced clock drift, and a significant number of them were subject to an initial clock error at the time of deployment. Based on these results, we release `OCloC`, an open-source Python package of our method. With our method and package, we expect to facilitate the monitoring of offshore geothermal energy projects.

5.1. Introduction

Over the past few decades, the use of ocean-bottom seismometers (OBS) has increased significantly. OBS readings enable the identification of remarkable features, such as undersea

This chapter has been published as a journal paper in the Seismica Journal (Naranjo et al., 2024). Note that minor changes have been introduced to make the text consistent with the other chapters of this thesis.

volcanic eruptions (Matsumoto et al., 2019) or seismic activity linked to tectonic strain and gas emissions through fault conduits (Tary et al., 2011). In particular, OBS readings are frequently used for imaging of the crust and/or mantle (e.g., Dongmo Wamba et al., 2023). Despite these successes, a key challenge in using OBS remains the accurate (time) synchronisation of the instruments' recordings. In fact, most OBS clocks drift, meaning they do not run at the same rate as a reference clock. This issue might be overcome by using atomic clocks instead of the traditional microprocessor-compensated crystal oscillator clocks that most OBS have (Gardner and Collins, 2012). This, however, would increase the inventory costs and power consumption, implying fewer instruments and less monitoring time, respectively. If the network is not properly synchronised, the incorrectly timed recordings may result in biased earthquake locations and Earth structure models.

One simple approach to identify clock drift is to measure the time difference between the instrument's internal clock and a GNSS signal before deployment and after recovery. This time difference is commonly referred to as the instrument's 'skew'. Assuming the instrument's clock drifted at a linear rate, a time correction can then be applied (e.g., Geissler et al., 2010). The skew, however, is not always possible to retrieve (e.g., when the instrument's battery dies before recovery). For this reason, several authors have proposed alternative methods for correcting clock errors; many of these exploit the presumed temporal stability of time-averaged cross-correlations of ambient seismic noise (e.g., Sens-Schönfelder, 2008; Loviknes et al., 2020; Hannemann et al., 2014; Jousset et al., 2013). These approaches, however, ignore errors that could arise if the initial synchronisation with a GNSS signal is either lacking or erroneous, or if there is an "initial" clock error resulting from the temperature shock during deployment (Zhang et al., 2023).

In theory, time-averaged cross-correlations of recordings of ambient seismic noise (henceforth 'noise cross-correlations') result in a signal that is symmetric around $t = 0$ (e.g., Stehly et al., 2006). In fact, under favourable conditions, the signals at positive and negative time lag coincide with the medium's Green's function (between the positions of the two seismic stations) and its time reverse, respectively. As such, it is referred to as 'seismic interferometry' (SI) (Wapenaar and Fokkema, 2006). In practice, these conditions are often not entirely fulfilled. Notwithstanding, provided the illumination is sufficiently uniform, the operation of averaging noise cross-correlations over time still yields two interferometric surface wave responses: one at the positive and one at the negative time lag(s). Violation of the noise cross-correlations' time symmetry may indicate the presence of clock errors (e.g., Hannemann et al., 2014).

Currently, two distinct approaches use noise cross-correlations to detect and correct clock errors (Gouédard et al., 2014). The first approach is based on the presumed temporal stability of the noise cross-correlations (Hable et al., 2018; Loviknes et al., 2020). In this approach, cross-correlation functions of ambient noise are calculated over different periods. The drift is then estimated as the time shift that maximises the Pearson correlation coefficient between each cross-correlation function and a reference correlation function (Hable et al., 2018). However, this method ignores the possibility of an initial clock error at the time of deployment due to a temperature shock during the OBS' descent to the ocean floor (Gardner and Collins, 2012; Zhang et al., 2023). The second approach exploits the above-mentioned time symmetry between the retrieved interferometric responses (Sens-Schönfelder, 2008; Weemstra et al., 2021). Contrary to the first approach, both direct

surface wave arrivals (i.e., at positive and negative time lag(s) need to be retrieved successfully in this case. Low signal-to-noise ratios or stations that are too close to each other (in terms of wavelength) prohibit this.

Although existing approaches for correcting clock errors have proven successful, a few challenges remain. First, the symmetry of ambient noise cross-correlation, while a valuable theoretical concept, is rarely realised in practice. A non-uniform illumination pattern may cause shifts in the arrival time of the interferometric responses with respect to the true arrival time (a challenge that is often overlooked). Second, current methods ignore the possibility of the aforementioned initial clock error during deployment. This clock error, introduced during the OBS' descent, is not expected given the mechanism causing clock drift (e.g., Shariat-Panahi et al., 2009), but it would nonetheless be good to rule out; in particular because the first approach mentioned above (Hable et al., 2018; Loviknes et al., 2020), does not allow such initial clock error to be detected. Finally, many of the current methods rely on land seismometers that are considered to be devoid of clock errors, ideally in the vicinity of the OBS deployment. This, however, will not be the case when the OBS network is located in oceanic regions far from the coast.

In this chapter, we present a versatile method that addresses all these challenges. Our approach (i) uses a weighted least-squares inversion to minimise the detrimental effect of non-uniform illumination patterns, (ii) allows for a potential initial clock error at deployment time, and (iii) does not require land stations to be included in the network to synchronise the recordings. Regarding the third claim, although our approach allows the OBS network's recordings to be synchronised, the combined set of recordings cannot be synchronised with Coordinated Universal Time (UTC). To achieve that, a land station (with a UTC-synchronised clock) needs to be included in the network. The presented method is implemented in an open-source Python package named `OCloC` (OBS Clock Correction), which accompanies this chapter. It combines the two aforementioned techniques for clock error detection (i.e., the one relying on the presumed temporal stability of noise cross-correlations and the one relying on their presumed time symmetry). Our method (and hence the package) is particularly useful in application to large-N seismic arrays.

To show the validity of our method, we use data from a seismic network deployed on and around the Reykjanes peninsula, SW Iceland (Jousset et al., 2020a). This seismic network was deployed in the context of the geothermal project IMAGE (Integrated Methods for Advanced Geothermal Exploration, see also Jousset et al., 2020b; Blanck et al., 2020). The data set used consists of recordings by 30 on-land stations and 17 OBS (this is a subset of the stations used in Weemstra et al., 2021). In the following sections, we detail the theory underlying our approach (Section 5.2), discuss and exemplify the implementation of this theory (Section 5.3), present and discuss our findings (Section 5.4), and list the most important conclusions (Section 5.6). A more detailed description of the data is included in Section 5.3 (Section 5.3.1). In addition, a brief description of `OCloC` is given in this section (Section 5.3.3).

5.2. THEORY

In this section, the theory is introduced step-wise. First, we briefly highlight the most important theoretical aspects of Seismic Interferometry (SI). Second, we introduce a model adequate for determining clock drift, which is an extension of the model introduced by

Weemstra et al. (2021). Third, we introduce potential additional time shifts (i.e., in addition to clock drift) affecting the arrival times of the interferometric responses. Fourth, we describe how a single noise cross-correlation's drift, and deviation from symmetry, can be retrieved. Fifth, we present the matrix notation of the introduced model. Finally, we briefly describe two different inversion approaches.

5.2.1. SEISMIC INTERFEROMETRY

Early types of seismic interferometry (SI) were introduced to the geophysics community by Aki (1957) and Claerbout (1968). Over the last two decades, the theory underlying SI has been established (Lobkis and Weaver, 2001; Wapenaar and Fokkema, 2006; Snieder, 2004; Shapiro and Campillo, 2004), and the method has been exploited in numerous applications. Examples include subsurface characterisation (Draganov et al., 2007; Jousset et al., 2016), reservoir monitoring (Sánchez-Pastor et al., 2019), and glaciology (Lindner et al., 2018). In this chapter, SI is used as an independent method to recover clock errors without needing skew measurements.

Applying SI to recordings of ambient seismic noise allows one to retrieve new seismic responses between pairs of stations by means of simple cross-correlations (Wapenaar and Fokkema, 2006; Stehly et al., 2006). Under specific conditions, the time-averaged cross-correlation contains the response to two 'virtual sources': one at negative lag times (usually referred to as the 'acausal part') and another at positive lag times (referred to as the 'causal part'), and with the virtual sources coinciding with the receiver locations. Time averaging is required to suppress spurious travel time delays that arise from constructive interference of signals coming from different sources. The time-averaged noise cross-correlation is proportional to the medium's Green's function if: (i) the noise sources illuminate the station pairs uniformly from all angles, (ii) the noise sources are uncorrelated, (iii) the medium is lossless, and (iv) sources have coinciding amplitude spectra (Wapenaar and Fokkema, 2006). Under these assumptions, the time-averaged cross-correlation of noise recorded by stations at $\mathbf{x}_i$ and $\mathbf{x}_j$, which we denote by $C_{i,j}(t)$, is proportional to the Green's function $G(\mathbf{x}_j, \mathbf{x}_i, t)$ and its time-reversed version, convolved with the autocorrelation of the signal emitted by the (noise) sources, i.e.,

$$C_{i,j}(t) \propto \left[G(\mathbf{x}_j, \mathbf{x}_i, t) + G(\mathbf{x}_j, \mathbf{x}_i, -t)\right] * P(t), \tag{5.1}$$

where $P(t)$ denotes the signal's autocorrelation generated by noise sources. In this chapter, we focus on the direct surface wave part of the Green's functions, ignoring the scattered signal. We refer to Wapenaar and Fokkema (2006) and Halliday and Curtis (2008) for a more detailed discussion of the assumptions underlying SI.

5.2.2. A MODEL TO ACCOUNT FOR CLOCK DRIFT

When it comes to the recovery of clock errors, an essential feature of the noise cross-correlation is its presumed time symmetry: under the assumptions listed in the previous section, the direct surface waves in $C_{i,j}(t)$ arrive at time lags of equal magnitude but opposite signs (Figure 5.1a). A violation of this time symmetry, such as the one in Figure 5.1b, indicates the presence of clock errors. To infer these clock errors from noise cross-correlations, Weemstra et al. (2021, Section 4) recently introduced an appropriate model. These authors, however, did not include clock drift in their model as they assumed the instrumental clock

errors to be time-independent (or constant). We extend the model introduced by Weemstra et al. (2021) to account for time-dependent clock errors such as clock drift.

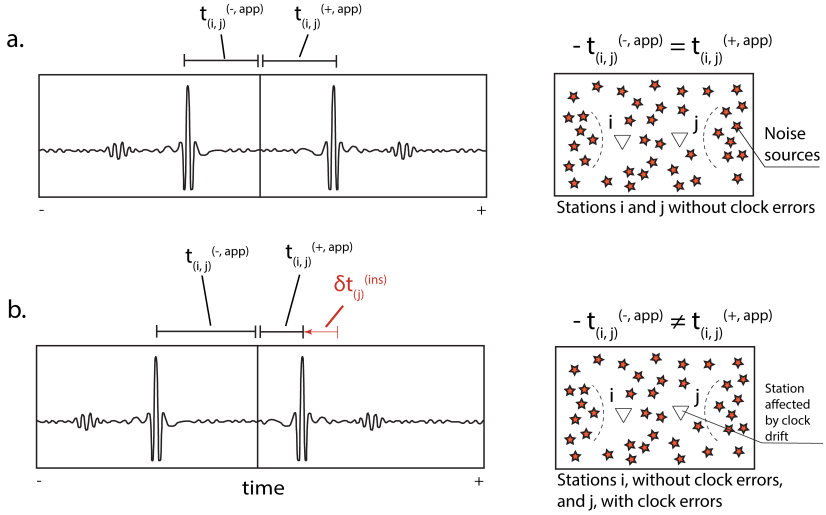


Figure 5.1: Schematic illustration of SI for the detection of clock errors. a. Noise cross-correlations computed using two stations without clock errors. The noise cross-correlation is almost symmetric in this case (for a relatively uniform illumination), and $t_{i,j}^{(+,\mathrm{app})} = -t_{i,j}^{(-,\mathrm{app})}$. b. Noise cross-correlations computed when one of the two stations is subject to clock errors (e.g., due to clock drift of one or both instruments). The noise cross-correlation is asymmetric (even for a relatively uniform illumination), and $t_{i,j}^{(+,\mathrm{app})} \neq -t_{i,j}^{(-,\mathrm{app})}$. In b, station j is subject to a clock error of $\delta t_j^{(\mathrm{ins})}$, which causes the noise cross-correlation to shift to negative time by that amount.

Here we assume the (potential) OBS clock drift to be linear. This is based on the fact that (i) the drift rate should be steady at constant temperature and (ii) the ambient temperature tends to be rather stable in deep water (note that the drift rate at a certain temperature is dictated by the frequency of the quartz oscillators in seismic clocks; Shariat-Panahi et al., 2009). The validity of this assumption has been demonstrated for OBS at larger depths in previous studies (Hable et al., 2018; Loviknes et al., 2020).

To estimate clock drift, we compute time-lapse cross-correlations $C_{i,j}(t, t^{(\mathrm{lps})})$, where $t^{(\mathrm{lps})}$ is the timing of the time-lapse cross-correlation. We refer to $C_{i,j}(t, t^{(\mathrm{lps})})$ as the 'lapse cross-correlation'. Note that $t^{(\mathrm{lps})}$ is the average time of all time windows contributing to the lapse cross-correlation. Therefore, $t^{(\mathrm{lps})}$ is not necessarily the time exactly in between the time of the first and last time window contributing to $C_{i,j}(t, t^{(\mathrm{lps})})$: in case the recordings by one of the two stations (or both) contain gaps, $t^{(\mathrm{lps})}$ may be skewed towards the beginning or end of the entire period over which individual cross-correlations are averaged.

For the considered linear parametrisation, the time-dependent clock error of station i, denoted by $\delta t_i^{(\mathrm{ins})}$, is written as

$$\delta t_i^{(\mathrm{ins})}\left(t^{(\mathrm{lps})}\right) = a_i\, t^{(\mathrm{lps})} + b_i, \tag{5.2}$$

where $\delta t_i^{(\mathrm{ins})}$ is the clock error of station i at $t^{(\mathrm{lps})}$, $t^{(\mathrm{lps})}$ is the average time of the time-lapse cross-correlation, a_i is the clock drift rate of station i, and b_i is the incurred clock error of station i at $t^{(\mathrm{lps})} = 0$.

Note that $t^{(\mathrm{lps})}$ is a continuous variable and that it is conveniently (but arbitrarily) set to 0 at the 21^{st} of August 2014. This is the approximate time of deployment of the OBS considered in this study (the OBS have been deployed over the course of a number of days around that date). Furthermore, $\delta t_i^{(\mathrm{ins})}$ is defined such that negative values imply that the recordings by station i are subject to a time delay. The rate at which the clock of station i is drifting is given by a_i, whereas b_i represents a possible clock error of station i at $t^{(\mathrm{lps})} = 0$. These are the two unknown parameters that we want to recover in this study (for all the OBS). A different parametrisation of $\delta t_i^{(\mathrm{ins})}$ in terms of, for example, cubic splines or trigonometric basis functions (i.e., Fourier series) is relatively straightforward.

A deviation from time symmetry can result from clock errors in either one or both stations involved in the noise cross-correlation. Similar to Weemstra et al. (2021), we denote the arrival time of the causal direct surface wave in $C_{i,j}\left(t, t^{(\mathrm{lps})}\right)$ by $t_{i,j}^{(+,\mathrm{app})}$ and the arrival time of the acausal direct surface wave by $t_{i,j}^{(-,\mathrm{app})}$. Accounting for the time-dependent clock errors above, we obtain the following expression for the apparent arrival time of the causal direct surface wave:

$$t_{i,j}^{(+,\mathrm{app})}\left(t^{(\mathrm{lps})}\right) = t_{i,j}^{(+)} + \delta t_i^{(\mathrm{ins})}\left(t^{(\mathrm{lps})}\right) - \delta t_j^{(\mathrm{ins})}\left(t^{(\mathrm{lps})}\right). \tag{5.3}$$

Similarly, the apparent arrival time of the acausal direct surface wave is given by

$$t_{i,j}^{(-,\mathrm{app})}\left(t^{(\mathrm{lps})}\right) = t_{i,j}^{(-)} + \delta t_i^{(\mathrm{ins})}\left(t^{(\mathrm{lps})}\right) - \delta t_j^{(\mathrm{ins})}\left(t^{(\mathrm{lps})}\right). \tag{5.4}$$

Here, $t_{i,j}^{(+)}$ and $t_{i,j}^{(-)}$ are the true arrival times of the direct surface waves, i.e., the direct surface waves in $G\left(\mathbf{x}_j, \mathbf{x}_i, t\right)$ and $G\left(\mathbf{x}_j, \mathbf{x}_i, -t\right)$, respectively. Consequently, by definition, $t_{i,j}^{(+)} = -t_{i,j}^{(-)}$. It is useful to note that a temporal change in the medium (e.g. Lindner et al., 2018) does not affect the equality between $t_{i,j}^{(+)}$ and $-t_{i,j}^{(-)}$, as it merely modifies the Green's function.

Summing the left-hand and right-hand sides of equations (5.3) and (5.4), and subsequently substituting the linear parametrisation defined in Equation (5.2), we find

$$\begin{aligned}
&\left(t_{i,j}^{(+,\mathrm{app})} + t_{i,j}^{(-,\mathrm{app})}\right)\left(t^{(\mathrm{lps})}\right) \\
&\quad = 2\delta t_i^{(\mathrm{ins})}\left(t^{(\mathrm{lps})}\right) - 2\delta t_j^{(\mathrm{ins})}\left(t^{(\mathrm{lps})}\right) \\
&\quad = 2a_i t^{(\mathrm{lps})} + 2b_i - 2a_j t^{(\mathrm{lps})} - 2b_j.
\end{aligned}$$

The variables here are shown schematically in Figure 5.1. In the ideal case that (i) the station couple is illuminated uniformly from all angles, (ii) spurious energy has effectively been stacked out in the time-averaging process, and (iii) the recordings are not subject to clock errors and/or drift, the right-hand side of Equation (5.5) evaluates to zero. If this is the case, then $t_{i,j}^{(+,app)} = -t_{i,j}^{(-,app)} = -t_{i,j}^{(-)} = t_{i,j}^{(+)}$. If, however, the measured $t_{i,j}^{(+,\mathrm{app})}$ and $t_{i,j}^{(-,\mathrm{app})}$ are such that the left-hand side of Equation (5.5) does not coincide with zero (and

the aforementioned conditions are fulfilled), this indicates a clock error at either one or both stations. The associated broken time symmetry is illustrated in Figure 5.1b.

Assuming the number of lapse cross-correlations $N^{(\mathrm{lps})}$ to coincide for all cross-correlation pairs, $t^{(\mathrm{lps})}$ can be discretised as $t_k^{(\mathrm{lps})}$, where $k = 1, 2, \ldots, N^{(\mathrm{lps})}$. In that case, Equation (5.5) can be written as

$$t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})} = 2a_i t_k^{(\mathrm{lps})} + 2b_i - 2a_j t_k^{(\mathrm{lps})} - 2b_j. \tag{5.5}$$

where the indices k in $t_{i,j,k}^{(+,\mathrm{app})}$ indicate that the arrival times of the direct surface waves are associated with lapse time $t_k^{(\mathrm{lps})}$. The procedure involving the determination of the $t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})}$ is based on the study by Weemstra et al. (2021), and detailed in Section 5.2.4. The associated practical implementation is explained in Section 5.3.2. Finally, it is useful to note that we merely assume the number of lapse cross-correlations per station couple to coincide for notational convenience. In practice, both the number of lapse cross-correlations and their timing (i.e., the values of the $t_k^{(\mathrm{lps})}$) may (and will) vary from one station couple to another.

5

5.2.3. ADDITIONAL ARRIVAL TIME SHIFTS

Differences in amplitude between the causal and acausal arrivals occur if the noise intensity is larger in one stationary-phase direction than in the other (Stehly et al., 2006). Importantly, a non-uniform illumination pattern may also introduce (small) deviations, or time shifts, from the correct arrival time of the causal and acausal surface waves. We denote these additional time shifts by $\delta t_{i,j,k}^{(\mathrm{src})}$ (the superscript 'src' implies that the time shift is associated with the source distribution). This time shift depends on all three indices since the (noise) illumination pattern usually varies as a function of both time (hence the index k) and station couple (hence the indices i and j). The time dependence of this term is related to the illumination pattern, which is usually non-stationary (e.g., Yang and Ritzwoller, 2008; Weemstra et al., 2013). The i, j dependence of this term is explained by the fact that the retrieved causal and acausal direct surface wave responses are associated with opposite stationary-phase regions (e.g., Snieder, 2004; Boschi and Weemstra, 2015). Azimuthal variations of the noise intensity in the two directions along the line connecting a station pair i and j, determine the magnitude of this arrival time shift. We therefore distinguish between $\delta t_{i,j,k}^{(+,\mathrm{src})}$ and $\delta t_{i,j,k}^{(-,\mathrm{src})}$, which represent (illumination related) arrival time shifts of the direct waves at positive (causal) and negative (acausal) time lag(s), respectively. In other words, the illumination-induced (additional) arrival time shifts of the causal and acausal direct surface waves can be expected to differ from each other (Weaver et al., 2009; Froment et al., 2010). We parenthetically note that the medium appears to be slower for a positive $\delta t_{i,j,k}^{(+,\mathrm{src})}$, whereas a positive $\delta t_{i,j,k}^{(-,\mathrm{src})}$ makes the medium appear to be faster than the actual medium.

In addition to the illumination-related arrival time shifts, we account for the presence of spurious energy by defining the additional time shifts $\delta t_{i,j,k}^{(+,\mathrm{spur})}$ and $\delta t_{i,j,k}^{(-,\mathrm{spur})}$, which, similar to $\delta t_{i,j,k}^{(+,\mathrm{src})}$ and $\delta t_{i,j,k}^{(-,\mathrm{src})}$, represent shifts in the arrival times of the causal and acausal direct surface waves, respectively (for details we refer to Weemstra et al., 2021). Including these

time shifts in our model, Equation (5.5) reads:

$$\begin{aligned} t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})} \\ = 2a_i t_k^{(\mathrm{lps})} + 2b_i - 2a_j t_k^{(\mathrm{lps})} - 2b_j \\ + \delta t_{i,j,k}^{(+,\mathrm{src})} + \delta t_{i,j,k}^{(-,\mathrm{src})} \\ + \delta t_{i,j,k}^{(+,\mathrm{spur})} + \delta t_{i,j,k}^{(-,\mathrm{spur})}. \end{aligned} \tag{5.6}$$

5.2.4. DETERMINATION OF $t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})}$

As explained in Section 5.2.2, clock errors manifest themselves by breaking the time-symmetry of the lapse cross-correlations. In order to solve for a large number of a_i and b_i (i.e., to determine clock drift for large OBS arrays), time shifts of individual lapse cross-correlations need to be extracted in an automated fashion.

The $t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})}$ (for all i, j, and k) are the entries of the data vector $\mathbf{t}^{(\mathrm{app})}$. Our procedure starts by computing *a priori* estimates of these $t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})}$. This estimate is based on the assumption that, for an individual station couple i, j, the drift accumulated over the interval from $t_1^{(\mathrm{lps})}$ to $t_{N^{(\mathrm{lps})}}^{(\mathrm{lps})}$ is the combined result of a_i and a_j (i.e., that it is linear). Based on the presumed stability of both the medium and the noise illumination, the accumulated drift is estimated by cross-correlating the earliest lapse cross-correlation with the latest lapse cross-correlations: ($C_{i,j}(t, t_1^{(\mathrm{lps})})$ is cross-correlated with $C_{i,j}(t, t_{N^{(\mathrm{lps})}}^{(\mathrm{lps})})$). Assuming the drift to be linear and clock errors to coincide with zero at $t_k^{(\mathrm{lps})} = 0$ then results in the sought-after *a priori* estimates of $t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})}$, which we denote by $t_{i,j,k}^{(\mathrm{a\,priori})}$. Note that, as such, the *a priori* estimate of $2b_i - 2b_j$ is assumed to be zero (see Equation 5.5). Clearly, this is a rather strong assumption. If an initial screening reveals that this assumption is not justified, it may be necessary to combine the procedure here with the procedure described in Section 5 of Weemstra et al. (2021). Finally, it is useful to note that instead of station-couple-specific *a priori* estimates, Weemstra et al. (2021) use station-specific *a priori* estimates to obtain $t_{i,j}^{(\mathrm{a\,priori})}$ (without an index k because the analysis by Weemstra et al. (2021) does not account for clock drift, but merely allows one to determine time-independent clock errors).

The $t_{i,j,k}^{(\mathrm{a\,priori})}$ are used to fill an initial estimate of the data vector $\mathbf{t}^{(\mathrm{app})}$. By solving the inverse problem (explained below in Section 5.2.6), we recover *a priori* estimates of the a_i and b_i. As soon as these estimates are obtained, we apply the procedure described in Section 5 of Weemstra et al. (2021). In summary, this involves determining the time windows in which the causal and acausal direct surface waves are expected using (i) a reference surface wave velocity (which can be station-couple specific), (ii) the station-to-station distance, and (iii) *a priori* estimates of a_i and b_j. Knowing the approximate time windows in which the direct causal and acausal surface waves can be expected, the envelopes of the lapse cross-correlations are subsequently computed. The envelopes are used to determine the arrival time of the direct surface wave (either causal or acausal) with the largest amplitude difference between the top and bottom envelope (denoted by t^{est} in Weemstra et al., 2021). Finally, after interpolating the lapse cross-correlation for a time window

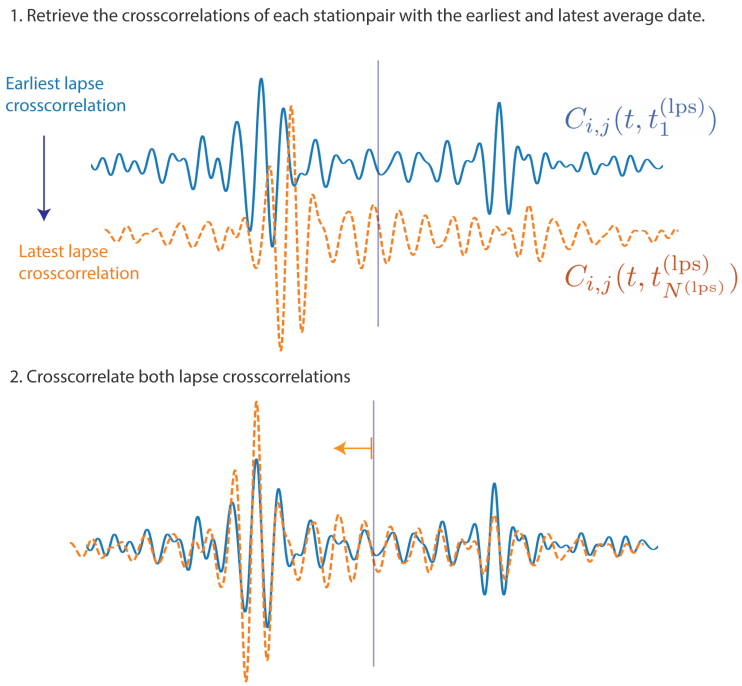


Figure 5.2: Processing steps for calculating an *a priori* estimate of the combined clock drift of a given station pair.

(with a length of about one period) centered around the *a priori* estimates of $t_{i,j,k}^{(+,\mathrm{app})}$ and $t_{i,j,k}^{(-,\mathrm{app})}$, and cross-correlating the signals in these two-time windows, the desired measurement $t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})}$ can be obtained. For a detailed description of the entire process, we refer to Weemstra et al. (2021, Section 5).

5.2.5. MATRIX FORMULATION

Assuming we possess synchronous noise recordings by a total of N seismic stations and we compute a total of $N^{(\mathrm{lps})}$ lapse cross-correlations between each station pair, a maximum of $N^{(\mathrm{lps})}$ times $N(N-1)/2$ lapse cross-correlations can be obtained. The set of equations governing the $t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})}$ can in that case be written as

$$\mathbf{A}\mathbf{t}^{(\mathrm{ins})} + \mathbf{n}^{(\mathrm{src})} + \mathbf{n}^{(\mathrm{spur})} = \mathbf{t}^{(\mathrm{app})}, \tag{5.7}$$

where the vector $\mathbf{t}^{(\mathrm{ins})}$ contains the sought-for clock drift rates a_i and initial clock errors b_i. This vector has a length of $2N$. The rows of $\mathbf{A}$ relate to different station pairs and lapse times $t_k^{(\mathrm{lps})}$, i.e., they are associated with different $C_{i,j}\left(t, t_k^{(\mathrm{lps})}\right)$. Each column of $\mathbf{A}$ is associated with either an a_i or a b_i. Consequently, $\mathbf{A}$ has dimension $N^{(\mathrm{lps})}(N(N-1)/2) \times 2N$. The length of the vectors $\mathbf{t}^{(\mathrm{app})}$, $\mathbf{n}^{(\mathrm{src})}$, and $\mathbf{n}^{(\mathrm{spur})}$ obviously coincides with the number of rows of $\mathbf{A}$. The vector $\mathbf{t}^{(\mathrm{app})}$ contains the measurements and is often referred to as the 'data vector'.

For the sake of clarity, we have detailed these vectors and matrices in Appendix A.5. Note that throughout this work, both matrices and vectors are indicated in bold; matrices are also capitalised, vectors not.

5.2.6. INVERTING FOR CLOCK DRIFT

In the model introduced above, we considered the number of lapse cross-correlations $N^{(\mathrm{lps})}$ to coincide for all station pairs. In addition, we assumed these lapse cross-correlations to exist for all possible combinations of stations, i.e., $N(N-1)/2$. In application to field data, however, $t^{(+,\mathrm{app})}$ and/or $t^{(-,\mathrm{app})}$ often cannot be determined for all lapse cross-correlations (i.e., all combinations of i, j and k). This implies that the number of rows M of the matrix $\mathbf{A}$ (and hence the number of elements of $\mathbf{t}^{(\mathrm{app})}$, $\mathbf{n}^{(\mathrm{src})}$, and $\mathbf{n}^{(\mathrm{spur})}$) will in practice often be smaller than $N^{(\mathrm{lps})}N(N-1)/2$.

The inability to accurately determine $t^{(+,\mathrm{app})}$ and/or $t^{(-,\mathrm{app})}$ can be due to a number of reasons. First, if two stations are too close to each other with respect to the wavelengths considered, the direct surface-wave response at a positive time will overlap with the direct surface-wave response at a negative time. Second, the absence of sources in one of the two stationary phase directions will prevent the retrieval of the corresponding direct surface-wave response (e.g., Snieder, 2004). Clearly, this also prevents determining the associated arrival time. Third, gaps in the recordings by one or more stations may lead to fewer lapse cross-correlations.

Before we explain the two inversion approaches, we clarify the relation between matrix $\mathbf{A}$ and the ability to obtain a unique (least-squares) estimate of $\mathbf{t}^{(\mathrm{ins})}$. Because, as defined in Appendix A.5, the rank of $\mathbf{A}$ is two lower than the number of unknowns $2N$ (having a matrix with a rank that is lower than the number of unknowns is often referred to as 'rank deficient'). This indicates that the system of equations is effectively underdetermined. In other words, a unique estimate of $\mathbf{t}^{(\mathrm{ins})}$ does not exist for the system of equations defined in Equation (5.7). We distinguish between two cases: a land station is included in the network, or no land station is included in the network. In the first case, a unique estimate of $\mathbf{t}^{(\mathrm{ins})}$ exists if a number of conditions are fulfilled. We will detail these in the paragraph below. If the network consists solely of OBS, a unique estimate of $\mathbf{t}^{(\mathrm{ins})}$ does not exist. We discuss that further below. Finally, an intuitive explanation of the rank deficiency is provided. Consider 10 lapse cross-correlations, each associated with a different $t_k^{(\mathrm{lps})}$ but with the same two OBS. The matrix $\mathbf{A}$ would be a 10×4 matrix in that case (see Equation 5.5). Clearly, an infinite number of (least-squares) solutions exist for b_1 and b_2 since adding any (arbitrary) value to both b_1 and b_2 would result in the same left-hand side. In other words, a unique solution for b_1 and b_2 does not exist. The same applies to a_1 and a_2.

In case a station with a UTC-synchronised clock is included in the network (i.e., a land station), the entries of that station can be eliminated from $\mathbf{t}^{(\mathrm{ins})}$ and the associated columns eliminated from $\mathbf{A}$ (see also the discussion in Section 6 and Appendix A in Weemstra et al., 2021). Subsequently, a number of conditions need to be fulfilled for a unique estimate of $\mathbf{t}^{(\mathrm{ins})}$ to exist. First, the system of equations (as defined in Equation 5.7) needs to contain at least two lines associated with lapse cross-correlations involving that station. These two lapse cross-correlations should be associated with a different lapse time $t_k^{(\mathrm{lps})}$. The land station in the first of the (at least) two lapse cross-correlations may, in fact, be a different land station from the land station associated with the second lapse cross-correlation, as long

as the two lapse cross-correlations are associated with different $t_k^{(\mathrm{lps})}$. Second, each of the OBS needs to be part of at least two lapse cross-correlations: there need to be two rows in $\mathbf{A}$ for which the entries associated with that OBS are non-zero. And again, these entries should be associated with different $t_k^{(\mathrm{lps})}$. In case these two conditions are fulfilled, the rank of $\mathbf{A}$ coincides with the number of unknowns ($2N$), and a unique least-squares estimate of the a_i and b_i in $\mathbf{t}^{(\mathrm{ins})}$ exists. Finally, the larger the difference in time between the various lapse cross-correlations of an OBS, the more accurate the estimates of its a_i and b_i.

If the network consists solely of OBS, a unique estimate of $\mathbf{t}^{(\mathrm{ins})}$ does not exist. In that case, that least-squares estimate of $\mathbf{t}^{(\mathrm{ins})}$ is chosen that has the lowest norm, i.e., that minimizes $||\tilde{\mathbf{t}}^{(\mathrm{ins})}||$, where $\tilde{\mathbf{t}}^{(\mathrm{ins})}$ is any least-squares solution (or *least-squares estimator*) of the underdetermined system of equations. This solution is usually referred to as the *minimum norm* solution. The second condition above, which needed to be fulfilled to obtain a unique estimate of $\mathbf{t}^{(\mathrm{ins})}$, still applies in this case. That is, each of the OBS still needs to be "part of" at least two lapse cross-correlations. The minimum-norm solution yields an estimator of $\mathbf{t}^{(\mathrm{ins})}$ that allows the OBS recordings to be synchronised with respect to each other, but not with respect to UTC. This is, of course, still useful as it would enable tomographic studies using only the OBS or the localisation of seismic events (earthquake hypocentres) below the OBS array.

We consider two estimators of $\mathbf{t}^{(\mathrm{ins})}$. These are the ordinary-least-squares estimator $\tilde{\mathbf{t}}^{(\mathrm{ins})}_{(\mathrm{ols})}$, and the weighted-least-squares estimator $\tilde{\mathbf{t}}^{(\mathrm{ins})}_{(\mathrm{wls})}$. We refer to Weemstra et al. (2021) for a detailed description (and derivation) of these estimators and will only provide a brief explanation of these two estimators here. The ordinary least-squares estimator minimizes the misfit function $||\mathbf{t}^{(\mathrm{app})} - \mathbf{A}\mathbf{t}^{(\mathrm{ins})}||$ and hence does not account for (potential) variations in the $\delta t^{(\mathrm{src})}_{i,j,k}$ and/or $\delta t^{(\mathrm{spur})}_{i,j,k}$ for different i, j, k. The weighted least-squares estimator, instead, exploits the inverse proportionality of the illumination-related arrival time shifts (i.e., the inverse proportionality of $\delta t^{(\mathrm{src})}_{i,j,k}$) to the true station-to-station travel time $t_{i,j}$ (as derived by Weaver et al., 2009). But since this travel time is usually not known, it uses the station-to-station distances $|\mathbf{x}_j - \mathbf{x}_i|$ as a proxy for the $t_{i,j}$. Measurements (i.e., individual $t^{(+,\mathrm{app})}_{i,j,k} + t^{(-,\mathrm{app})}_{i,j,k}$) associated with lapse cross-correlations between stations (i and j) that are further apart are hence assigned larger weights in the inversion.

5.3. IMPLEMENTATION & APPLICATION TO DATA

In this section, we describe the workflow that allows the estimators of $\mathbf{t}^{(\mathrm{ins})}$ to be computed (Section 5.3.2). Although predominantly methodological aspects of the workflow are discussed (results are presented in Section 5.4), some examples with field data are shown. We therefore start by introducing the IMAGE's seismic network and its lapse cross-correlations (Section 5.3.1). After describing the workflow, we dedicate one subsection to our package `OCloC` (Section 5.3.3). We finish this section with a description of a bootstrapping procedure that allows the stability of the recovered clock drift values to be assessed (Section 5.3.4).

5.3.1. THE IMAGE DATA SET

For heuristic purposes, the explanation of some processing steps of our workflow includes these steps' application to a set of lapse cross-correlations. These lapse cross-correlations are retrieved from recordings of ambient seismic noise acquired on and around the Reykjanes peninsula, SW Iceland (Jousset et al., 2020a). It concerns lapse cross-correlations between a subset of the stations considered by Weemstra et al. (2021). Specifically, about one year of noise recorded by 30 land stations and 17 OBS is used (see Figure 5.3 for the station locations). The OBS in this experiment are equipped with Seascan clocks (SEASCAN microcomputer compensated crystal oscillators), which are temperature-compensated.

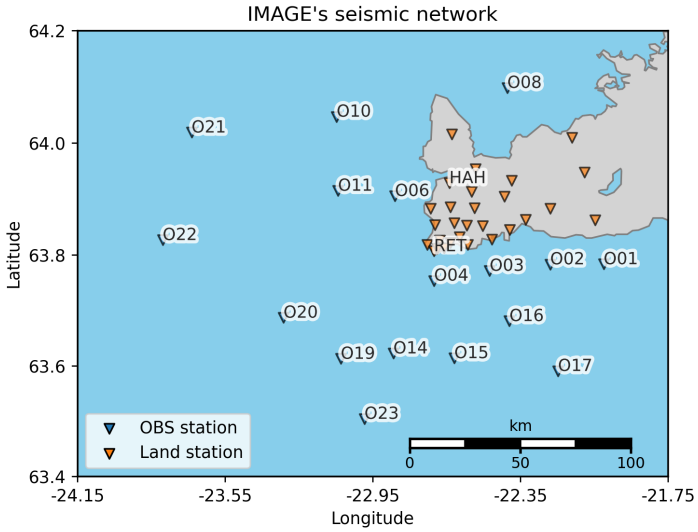


Figure 5.3: On- and off-shore stations of IMAGE's seismic network, SW Iceland, whose lapse cross-correlations were used in this chapter. Note that the numbering of the OBS runs up to 23. In contrast, only 17 OBS are included in our set of lapse cross-correlations (some stations did not sample the ambient seismic field sufficiently long and were hence excluded from our analysis; see also Figure S1 in Weemstra et al. (2021)). Only the land stations 'HAH' and 'RET', which are analysed in Sections 5.3.2 and 5.4, are labelled due to space constraints.

The lapse cross-correlations are computed by averaging individual station-to-station cross-correlations over a 100-day period. These individual cross-correlations are computed per hour with a 50% overlap. We refer to Weemstra et al. (2021) for a detailed description of the computation of the hourly cross-correlations. Averaging individual (hourly) cross-correlations is performed in a two-step process. First, daily cross-correlations are computed based on a maximum of 47 hourly cross-correlations ($24 \times 2 - 1$). Subsequently, these daily cross-correlations are averaged. Importantly, gaps in the recordings by one or both stations are accounted for in the sense that the timing of a lapse cross-correlation, i.e., its $t_k^{(\mathrm{lps})}$, is defined as the average time of the individual cross-correlations. Gaps in the data can cause the average time of the correlations ($t_k^{(\mathrm{lps})}$) to deviate from the centre of the 100-day period. Note that the $t_k^{(\mathrm{lps})}$ are allowed to differ between different station couples, as they

are explicitly included in **A** (see also Appendix A.5). In case the number of individual cross-correlations contributing to a lapse cross-correlation does not exceed 75% of the maximum number of individual cross-correlations (which is 100×47), that lapse cross-correlation is discarded. An overview of the data availability is given in Figure S1 of Weemstra et al. (2021).

5.3.2. WORKFLOW

To determine and correct clock drift using lapse cross-correlations of ambient seismic noise, we adopt the processing sequence in Figure 5.4. It is this workflow that is implemented in `OCloC`. The workflow comprises five steps. We now dedicate one subsection to explain and discuss each of these steps.

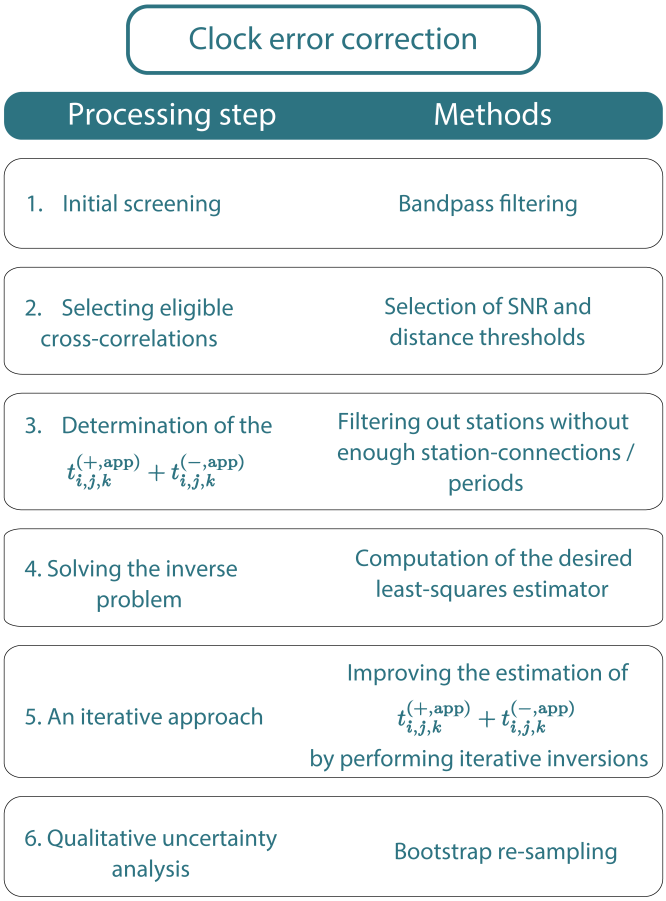


Figure 5.4: Workflow for the determination of OBS clock drift using lapse cross-correlations of ambient seismic noise between a large number of OBS (computed from large-N ocean-bottom seismometer deployments).

INITIAL SCREENING

In Figure 5.5a, all stations and ray paths associated with the available lapse cross-correlations are shown. To get a first impression of whether or not the OBS recordings are subject to clock drift, one can plot the different lapse cross-correlations in a single plot (i.e., time-averaged cross-correlations associated with different $t_k^{(\mathrm{lps})}$). In Figure 5.5b, we depict lapse cross-correlations between stations 020 and HAH (land station) for 5 different lapse times. Potential clock drift of an OBS manifests itself as a shift in time of the lapse cross-correlations: for this specific station couple, the lapse cross-correlations associated with larger $t_k^{(\mathrm{lps})}$ are shifted to a later time.

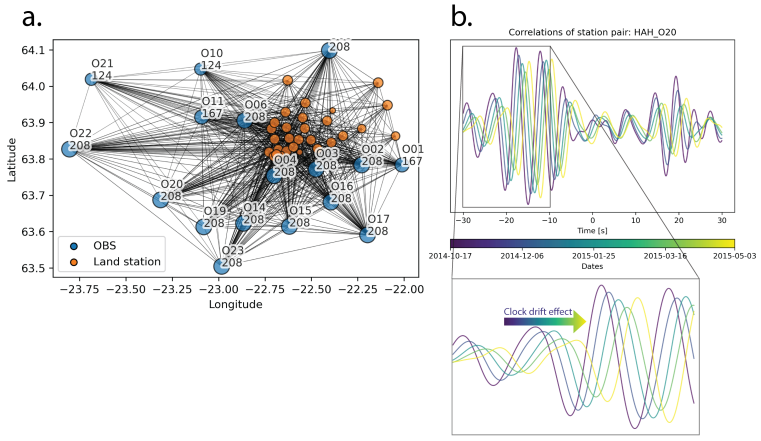


Figure 5.5: a. All seismic stations and ray paths; blue and orange circles correspond to OBS and land seismometers, respectively. Only the station names of the OBS are indicated. Below each station name, the number of available lapse cross-correlations involving that specific station is depicted. b. All lapse cross-correlations for a given station pair. The colours indicate the average timing ($t_k^{(\mathrm{lps})}$) of the lapse cross-correlation.

Prior to the determination of clock drift, it is important to choose an adequate bandpass filter. For the IMAGE data, the surface waves in the retrieved interferometric responses have the highest signal-to-noise ratio (SNR) between 0.1 and 0.4 Hz. In general, however, the pass band depends on parameters such as the nominal station-to-station distance, the amplitude of the noise sources, the illumination pattern, and the geographical location of the OBS array (e.g., Yang and Ritzwoller, 2008). Note that, due to surface-wave dispersion, lower frequency bands usually result in smaller separations in time of the causal and acausal surface wave peaks. Importantly, the choice of frequency band also strongly affects the capability to determine the $t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})}$ of individual lapse cross-correlations.

SELECTING ELIGIBLE LAPSE CROSS-CORRELATIONS

There are two parameters that determine a lapse cross-correlation's eligibility to be included in the clock error estimation process: the SNR threshold and the station-to-station distance threshold. Together, these parameters determine which lapse cross-correlations are included in the inversion and which are not (i.e., whether their $t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})}$ will be determined and added to data vector $\mathbf{t}^{(\mathrm{app})}$ or not).

In general, the quality of the measurements (i.e., the $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$) strongly depends on the signal-to-noise ratio (SNR). If the SNR is too low, the algorithm experiences difficulties determining the arrival times of the interferometric responses. Low SNRs are mainly due to low-intensity illumination from (one of) the stationary-phase regions (Snieder, 2004; Weaver et al., 2009). Consequently, the measurements may be inaccurate, or even subject to cycle skipping (Weemstra et al., 2021). Obviously, inaccurate entries in the data vector $\mathbf{t}^{(\text{app})}$ (i.e., inaccurate $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$) adversely affect the inversion results. A clear example is shown in Figure 5.6a, where the causal peaks of the lapse cross-correlations between stations O08 and O21 have low SNR. In this case, the determination of the arrival time of the causal peak is not straightforward and hence may result in inaccurate $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$. Both SNR, of the causal and acausal interferometric direct surface waves, need to exceed the SNR threshold for the lapse cross-correlations to be included in the inversion. For details regarding the computation of the SNR, we refer to Weemstra et al. (2021).

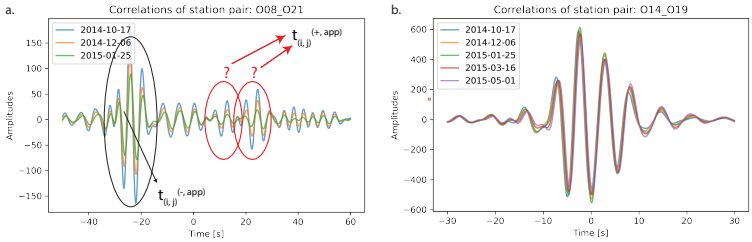


Figure 5.6: a. Lapse cross-correlations between OBS O08 and O21: the signal-to-noise ratio of the causal wave is rather low, complicating the determination of $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$. b. Lapse cross-correlations between stations O14 and O19: The station-to-station distance of these stations is so small (10.6 km) that the causal and acausal surface waves overlap (note that for surface waves with a period of 5 seconds that propagate at 3000 m/s, 10.6 km corresponds to only 2/3 of a wavelength).

The second important parameter when it comes to the accuracy of the $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$ is the station-to-station distance. If two stations are too close to each other, the direct surface-wave response at a positive time (i.e., the causal arrival) will overlap with the direct surface-wave response at a negative time (i.e., the acausal arrival). Consequently, our algorithm will not be able to correctly determine the $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$ for those station couples. To prevent the inclusion of such measurements in the system of equations, the user must set a station-to-station distance threshold. This threshold is expressed in terms of wavelengths since the ability to distinguish the causal from the acausal arrival does not merely depend on the surface wave travel time, but on the ratio between the travel time and the (dominant) period of the interferometric surface waves. This threshold needs to be set at the start of the workflow (for further details regarding the station-to-station distance threshold we refer to Weemstra et al., 2021). A lapse cross-correlation's station-to-station distance needs to exceed the distance threshold for that lapse cross-correlation to be included in the inversion (i.e., for the $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$ to be determined and added to data vector $\mathbf{t}^{(\text{app})}$).

Using the IMAGE lapse cross-correlations, we investigate how different thresholds affect the number of eligible lapse cross-correlations. If the thresholds are set too high, there will not be sufficient lapse cross-correlations to (accurately) determine the clock drift of all

OBS (i.e., the vector $\mathbf{t}^{(\mathrm{app})}$ will be relatively short). Conversely, if the thresholds are too low, we add too many inaccurate data points to the data vector, in turn leading to less accurate a_i and b_i (and hence less accurate clock drift estimates). Figure 5.7 depicts the number of eligible station pairs exceeding a specific combination of thresholds. Obviously, lower thresholds result in a higher number of eligible lapse cross-correlations. Although a higher number of lapse cross-correlations implies a larger number of measurements, it has been shown that station-to-station distance thresholds in the range of 2 to 4 wavelengths and SNR thresholds of about 15 yield the most accurate clock errors (Weemstra et al., 2021). The latter values, however, are based on synthetic data. Here, we therefore choose a slightly more conservative SNR threshold of 30, while setting the station-to-station distance threshold to 2.5. The lapse cross-correlations fulfilling these criteria (i.e., exceeding these thresholds) are added to $\mathbf{t}^{(\mathrm{app})}$ and hence enter the inversion.

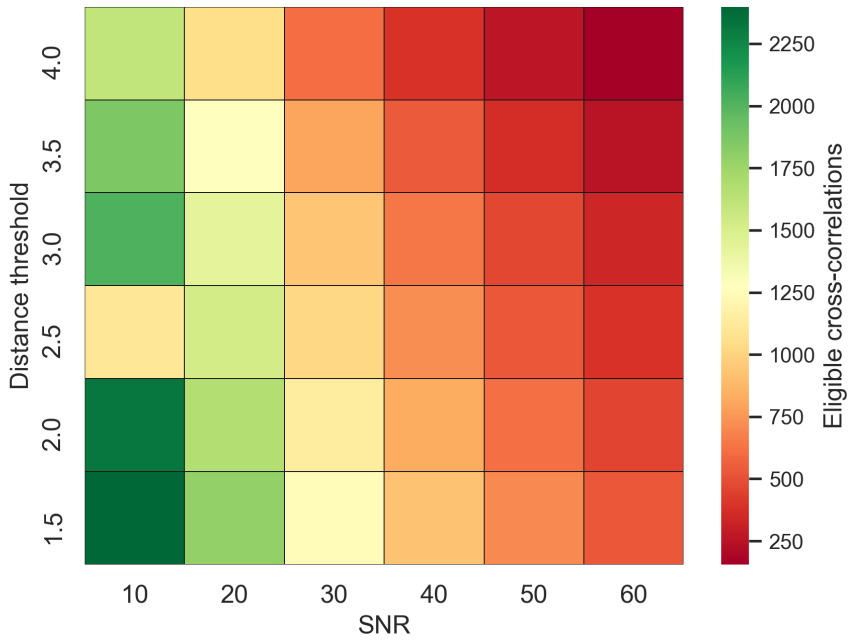


Figure 5.7: Number of eligible lapse cross-correlations for different station-to-station distances and SNR thresholds.

DETERMINATION OF THE $t_{i,j,k}^{(+,\mathrm{app})} + t_{i,j,k}^{(-,\mathrm{app})}$ FOR ALL SELECTED COMBINATIONS i, j, k

The algorithm computes the *a priori* clock-drift estimate only for lapse cross-correlations that exceed the SNR and station-to-station distance thresholds. This results in some stations having fewer connections with other stations. To avoid including stations without enough lapse cross-correlations, we set a minimum-number-of-connections threshold and a

minimum-number-of-total-lapse-cross-correlations threshold. We also remove the remaining lapse cross-correlations that are associated with a station that does not exceed these thresholds (i.e., the data vector $\mathbf{t}^{(\text{app})}$ will be shortened, and the number of rows and columns of the matrix $\mathbf{A}$ decreases).

To recover a unique clock drift estimate (a_i), the station needs to be associated with lapse cross-correlations at various lapse times $t_k^{(\text{lps})}$ (recall the discussion in Section 5.2.6). By defining (i) a minimum number of correlation periods, (ii) the number of different lapse times required, and (iii) the minimum separation in days between an OBS' lapse cross-correlations, a unique solution can be guaranteed, provided a land station is present in the system of equations, which will then not be rank deficient.

The accuracy of clock-drift inversion is affected by cycle skipping (Weemstra et al., 2021). That is, a measurement deviates from the true $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$ by approximately one period. Needless to say, the inclusion of these measurements in the inversion leads to incorrect a_i and b_i. In Appendix A.6, we describe a procedure allowing one to detect such outliers and discard them.

SOLVING THE INVERSE PROBLEM

We implement two inversion approaches using i) the ordinary least squares estimator $\tilde{\mathbf{t}}_{(\text{ols})}^{(\text{ins})}$ and ii) the weighted least-squares estimator $\tilde{\mathbf{t}}_{(\text{wls})}^{(\text{ins})}$.

The ordinary least-squares estimator can be used if noise sources uniformly illuminate the stations from all directions. In that case, the vector $\mathbf{n}^{(\text{src})}$ in Equation (5.7) coincides with $\mathbf{0}$ and the only source of noise is $\mathbf{n}^{(\text{spur})}$. Assuming that the entries of the latter vector have coinciding variance, the ordinary least-squares estimator $\tilde{\mathbf{t}}_{(\text{ols})}^{(\text{ins})}$ will give the most accurate estimate of $\mathbf{t}^{(\text{ins})}$ (in a least-squares sense).

When the surface-wave illumination pattern is not uniform (as is often the case in practice; Yang and Ritzwoller, 2008; Stehly et al., 2006), $\mathbf{n}^{(\text{src})}$ does not coincide with zero. In this case, it is more appropriate to compute the weighted least-squares estimator $\tilde{\mathbf{t}}_{(\text{wls})}^{(\text{ins})}$, where the station-to-station distances $|\mathbf{x}_j - \mathbf{x}_i|$ act as weights (see Section 5.2.6, and, for further details, Weemstra et al., 2021). In Section 5.5.1, we demonstrate the superiority of the weighted least-squares estimator, which was previously shown using synthetic noise cross-correlations. Finally, in the absence of lapse cross-correlations with recordings by a land station, the minimum-norm solution is computed. In this case, the recovered b_i differs from the true (unknown) b_i by a common time shift.

AN ITERATIVE APPROACH

Upon solving the inverse problem using the *a priori* estimates $t_{i,j,k}^{(\text{a priori})}$, we obtain an initial estimate of the a_i and b_i values of each station. The latter can subsequently be used to improve the estimation of $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$ as they can be used to predict the arrival time of the interferometric surface wave responses (see also Weemstra et al., 2021). It is recommended to perform several inversions, each iteration using the previously obtained a_i and b_i to guide the estimation of the $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$ resulting in an updated data vector $\mathbf{t}^{(\text{app})}$, until the recovered a_i and b_i do not change. By simply plotting the evolution of the recovered a_i and b_i, it is possible to determine when this is the case.

By plotting all lapse cross-correlations associated with a single station couple in one frame, and doing this separately for the corrected (using the obtained a_i and b_i) and uncorrected set of lapse cross-correlations, a (qualitative) impression of the result is obtained. If the lapse cross-correlations associated with different lapse times align, then the clock drift is successfully removed. An example of a successful clock drift removal is shown in Figure 5.8. It is clear that the lapse cross-correlations of OBS O20 had a clock drift (Figure 5.8a). Once the clock drift is removed, the lapse cross-correlations associated with different lapse times align, as shown in Figure 5.8b.

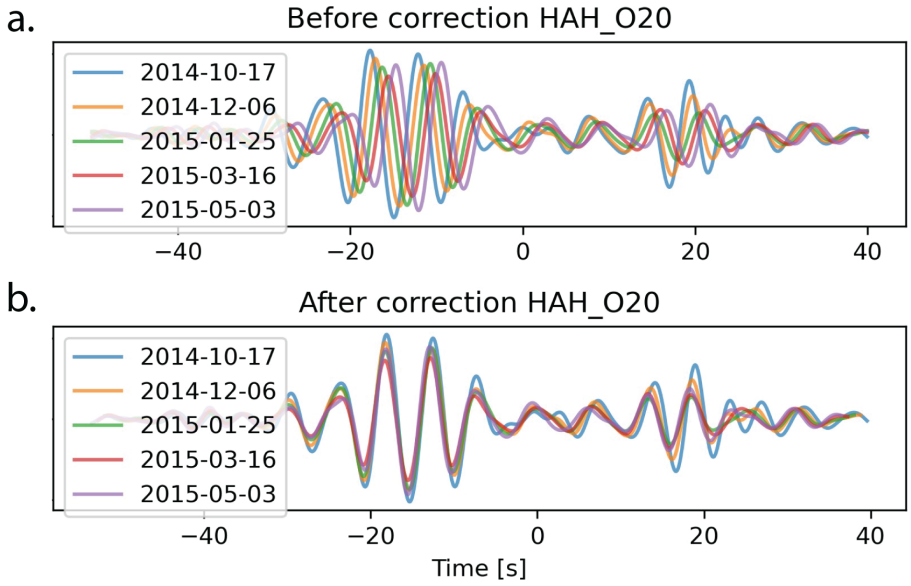


Figure 5.8: Lapse cross-correlations between land station HAH and OBS O20 for different lapse times. a. Cross-correlations before applying time corrections b. Lapse cross-correlations after correcting the clock drift of O20 using the a_i and b_i recovered by means of the iterative weighted least-squares inversion.

5.3.3. OCLoC

The methodology presented in this chapter has been implemented in OCloC. In particular, OCloC allows the workflow detailed in the previous subsection to be executed. OCloC is an open-source Python package that has been tested for the operating systems Linux and macOS. We chose Python as OCloC's main programming language for its open-source, versatile, and cross-platform compatible nature, which is widely used in the Earth sciences (e.g., Werthmüller et al., 2021; Rücker et al., 2017). In the case of OCloC, the portability of Python enabled us to outsource specific computational aspects to a pre-compiled Fortran module.

Through the application of seismic interferometry, the proposed correction of clock errors is contingent on the availability of synchronous noise recordings. The computation of the lapse cross-correlations, however, is deliberately left out of OCloC. The reason is

that it will be nearly impossible to account for the plethora of different (pre-)processing approaches (Seats et al., 2012; Groos et al., 2012; Weemstra et al., 2014; Fichtner, 2014). This implies that users of the package have complete freedom regarding pre-processing (e.g., one-bit normalisation, spectral whitening, etc.) and potential filter settings while computing the lapse cross-correlations, and that they are expected to do this themselves prior to the application of `OCloC`. The lapse cross-correlations can subsequently be imported as `OCloC` objects.

`OCloC`'s functionality includes loading lapse cross-correlation files, storing and accessing station metadata, and solving the linear systems of equations in Section 5.2.5 in a (weighted) least-squares sense. It also has some other supporting functions. To keep the use of `OCloC` simple, a hierarchical object-oriented design has been adopted. This kind of architecture breaks down the whole process of determining and correcting clock errors into solvable chunks while letting the user know when an error occurred and how to prevent it.

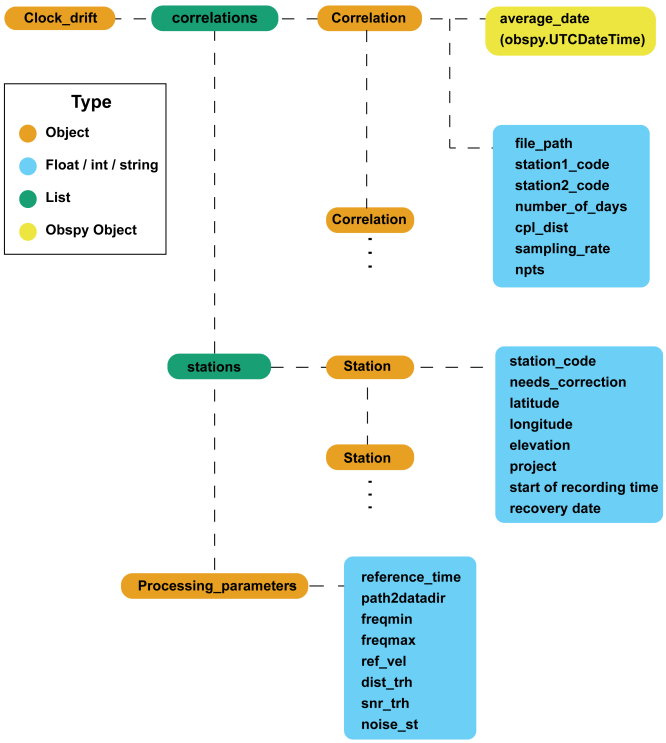


Figure 5.9: Object hierarchy of `OCloC`.

The main object types of `OCloC` are: `ClockDrift`, `ProcessingParameters`, `Correlation`, and `Station`. Figure 5.9 depicts, schematically, the algorithm's object hierarchy. These objects need some clarification:

1. `ClockDrift`: The outermost layer of the hierarchical structure. The user deals with this object for the main processing steps described in Section 5.3.2. This object stores

the different `Station` and `Correlation` objects in the form of lists. This object also stores the system of equations, described in Section 5.2, in the form of a `Pandas` dataframe (Wes McKinney, 2010a). The different methods of `ClockDrift` provide access to correlation files, station metadata, plotting functions, and different processing tools required for the algorithm's usage.

2. `ProcessingParameters`: The recovery of clock errors depends on the adequate selection of some pre-processing steps. `ProcessingParameters` object stores the value of these parameters. These parameters are the band-pass filter's corner frequencies, the SNR threshold, and the station-to-station distance threshold. These parameters are detailed in Section 5.3.2.

3. `Correlation`: Stores the metadata of each cross-correlation file such as the station names, the lapse time $t_k^{(lps)}$, and station-to-station distance, among others. Additionally, this object has functions to compute $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$, together with the signal-to-noise ratios of the causal and acausal surface wave arrivals.

4. `Station`: Contains metadata such as location, code, and timestamp when the station started recording. Moreover, after solving the linear system of equations, the recovered clock errors, i.e., the a_i and b_i, can be retrieved through these objects.

In addition to the core module, `OCloC` incorporates third-party dependencies that yield advanced functionality, namely, the `Numpy` programming library (Harris et al., 2020), several signal processing functions from `Obspy` (Krischer et al., 2015), and the data visualisation tools of `Pandas` (Wes McKinney, 2010b) and `Matplotlib` (Hunter, 2007). For specific details regarding the package installation and usage, please refer to the online documentation available at `https://ocloc.readthedocs.io`.

5.3.4. Bootstrap re-sampling

To verify the robustness of the obtained results, we repeat the inversion several times using different sets of measurements $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$. By repeating the inversion multiple times, mean values and confidence intervals of the sought-after parameters are obtained. One way to artificially generate different sets of measurements is using bootstrap resampling (Efron, 1982). Bootstrapping is a statistical method that falls under the broader class of resampling methods. It allows one to estimate statistical properties of interest, such as sample averages and variances (Schnaidt and Heinson, 2015). Effectively, it gives an indication of which results are likely and which are less likely without computing new lapse cross-correlations. Here, we seek to obtain an estimate of the variance of the estimated a_i and b_i. Instead of using all the measured data points, we sample with replacement (Efron, 1982, 1992). In practice, we generate a large number of data vectors $\mathbf{t}^{(\text{app})}$ (usually referred to as 'realisations'), each with the same length as the original data vector, but with values that are drawn from the original data vector, allowing duplicates. Specifically, we perform the following steps:

I. An initial estimate of clock drift is obtained following steps one to five of Figure 5.4. A SNR threshold of 30 and a station-to-station distance threshold of 2.5 wavelengths

are applied. It is necessary to check that the recovered a_i and b_i values are no longer changing after several iterations. This results in the data vector $\mathbf{t}^{(\mathrm{app})}$ that serves as the input of our bootstrapping procedure.

II. Allow sampling with replacement by randomly selecting measurements of $t_{i,j}^{(+,\mathrm{app})} + t_{i,j}^{(-,\mathrm{app})}$ (bootstrapped samples).

III. Once having re-sampled the measurements, perform the inversion and store the recovered a_i^* and b_i^* values of each station.

IV. Repeat steps II and III one thousand times. By doing so, we store 1000 possible realisations of the recovered a_i^* and b_i^* values.

V. Based on all the a_i^* and b_i^* realisations, compute a statistical measure, such as 95% confidence intervals, for each of the stations.

To identify stations with relatively uncertain a and b values (either due to a limited number of data points, or due to a lot of noise on the lapse cross-correlations associated with that specific station), we estimate the standard deviation and 95% confidence intervals (CI) from the 1000 realisations. The CI represents the range in which 95% of the a_i^* and b_i^* values lie.

The bootstrap approach allows one to identify OBS with narrow or large confidence intervals. Narrower confidence intervals suggest the recovered a and b are well-determined, whereas stations with larger confidence intervals point to larger uncertainties in the recovered clock errors. In the absence of noise, i.e., $\mathbf{n}^{(\mathrm{src})}$ and $\mathbf{n}^{(\mathrm{spur})}$ both coinciding with $\mathbf{0}$, all a_i^* and b_i^* of a given station should coincide and hence the 95% confidence interval would be zero.

5.4. RESULTS

5.4.1. CLOCK DRIFT RATES

We computed the weighted least-squares estimator of $\mathbf{t}^{(\mathrm{ins})}$ for the OBS in the IMAGE's network. Our findings indicate that all OBS stations experienced clock drift. Compared to the other OBS, the clock drift of OBS O20 was particularly large. Table 5.1 summarises the estimated clock drift rates (i.e., the a_i) and incurred clock errors at the time of deployment. The latter may deviate slightly from the b_i because the b_i represents the clock errors on August 21, 2014 ($t^{(\mathrm{lps})} = 0$), whereas most stations were not deployed exactly on that date. In addition, we list the measured skews in the last column. To compare these skew values, we provide in the fifth column the clock error at the time of recovery computed using the estimated a_i and b_i. Note that most OBS recordings end prior to that date due to full disks. We also obtained a drift estimate for OBS O21, which had no skew value documented due to a dead battery at the time of recovery. The incurred initial clock errors at the time of deployment ranged from a minimum of -0.404 s to a maximum of 0.037 s.

The bootstrap re-sampling introduced in Section 5.3.4 allows us to estimate the variance of the recovered a_i and b_i. By generating 1000 different data vectors (realisations) and subsequently performing a separate inversion for each of the generated data vectors,

Table 5.1: Estimated clock drift rates (a_i) of the OBS (2nd column), and their corresponding standard deviation (3rd column). The clock drift rates in column four are based on the measured skew values, assuming a linear drift and no clock errors at $t^{(\mathrm{lps})} = 0$. The estimated clock errors at deployment and recovery time in the fifth and sixth columns, respectively, are computed by substituting the estimated a_i and b_i in Equation (5.2) with $t^{(\mathrm{lps})}$ set to each OBS' day of deployment and recovery. OBS O21 had no skew value reported, as the battery died before recovery. Station O23 has no standard deviation because it was associated with too few data points in $\mathbf{t}^{(\mathrm{app})}$ to be successfully included in the bootstrapping procedure. This is probably due to the relatively low SNR of the lapse cross-correlations involving this station.

Station name	Clock drift rate based on `OCloC` [s/year]	σ [s/year]	Clock drift rate based on skew values [s/year]	Clock error at deployment time (`OCloC`) [s]	Clock error at recovery time [s]	Measured skew [s]
O01	-0.734739	0.042760	-1.011740	-0.312892	-1.055899	-1.023125
O02	-1.055136	0.097011	-0.782087	-0.115908	-1.182942	-0.790906
O03	-0.401807	0.251438	-0.136396	-0.109015	-0.515270	-0.137906
O04	-0.770560	0.153252	-0.765888	-0.189565	-0.968726	-0.774437
O06	-0.172589	0.152249	-0.126476	-0.167177	-0.343849	-0.129468
O08	-0.104288	0.245437	-0.096861	-0.326683	-0.433947	-0.099625
O10	-1.095582	0.073045	-0.789557	-0.225825	-1.354066	-0.813093
O11	-0.667440	0.179920	-0.457559	-0.404513	-1.089599	-0.469656
O14	-0.304885	0.147440	-0.326255	-0.211671	-0.462262	-0.268156
O15	-1.465134	0.340357	-1.633493	-0.342048	-1.839114	-1.669093
O16	-0.712585	0.131879	-0.635126	-0.216183	-0.944088	-0.648781
O17	-0.547051	0.034074	-0.350884	-0.161655	-0.720530	-0.358468
O19	-0.985476	0.115891	-1.119642	-0.378849	-1.388873	-1.147531
O20	-4.652023	0.176057	-4.324439	0.038416	-4.744142	-4.445781
O21	-1.234367	0.141809	N/A	-0.185787	-1.456213	N/A
O22	-0.415065	0.077167	-0.643822	-0.315925	-0.743071	-0.662562
O23	-0.312865	N/A	-0.289709	-0.300249	-0.620852	-0.296875

1000 weighted least-squares estimators of $\mathbf{t}^{(\mathrm{ins})}$) are obtained. The standard deviation of the recovered a_i^* from the a_i (recovered using the original $\mathbf{t}^{(\mathrm{app})}$) is listed in column 3 of Table 5.1. In Figure 5.10, we visualise the recovered a_i and b_i, including the bootstrap-

derived uncertainties.

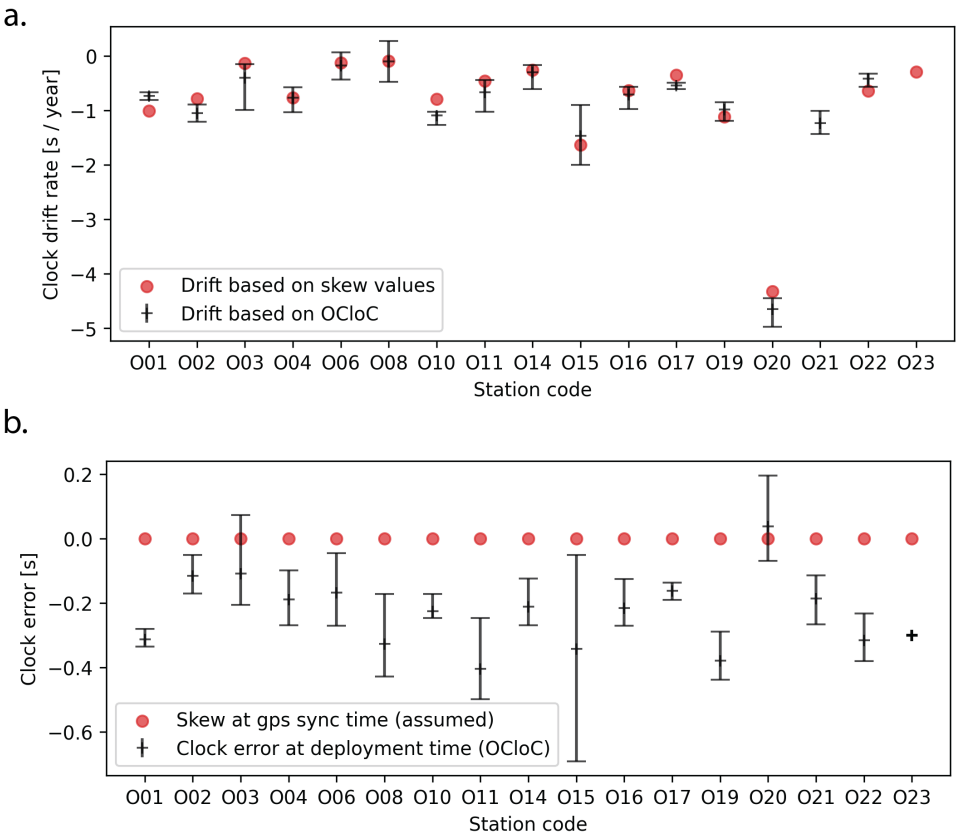


Figure 5.10: Comparison between clock drifts obtained in this study, and the measured skews. a. Comparison of clock drift rates estimated based on the skew values and the total recording time (red dots) and the a_i obtained from our weighted least-squares inversion (black crosses). The error bars correspond to the 95% confidence intervals resulting from the bootstrap re-sampling. Note that no skew value was reported for OBS O21 as this instrument's battery died before recovery. b. Comparison of the initial clock error at the OBS' deployment time. In both a and b, OBS O23 has no error bars as this OBS was associated with too few data points in $\mathbf{t}^{(\text{app})}$ to be successfully included in the bootstrapping procedure.

For all station pairs, we evaluate the waveform (mis)alignment of the lapse cross-correlations to verify the removal of clock errors. Figure 5.11 shows the time-lapse cross-correlations between OBS O01 and land station RET (which is devoid of clock errors) in three states: a) uncorrected clock errors, b) corrected using skew-derived drift rates, and c) corrected using the weighted-least-squares inversion (accounting for clock drift and an initial clock error). In Figure 5.11a, before clock correction, the later lapse cross-correlations shift monotonically to an earlier time, which can be explained by clock drift. In Figure 5.11b, after applying the skew-derived clock corrections, the later lapse cross-correlations shift monotonically to a later time. In contrast, after correcting the drift using the weighted least-squares inversion, the waveforms align properly as shown in Fig-

ure 5.11c.

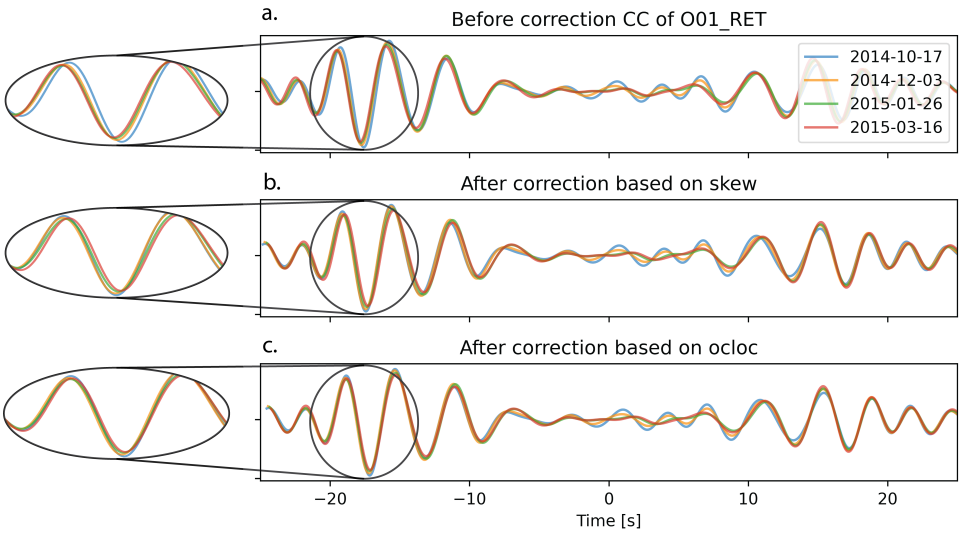


Figure 5.11: Lapse cross-correlations between OBS O01 and land-station RET in the frequency range of 0.2 to 0.4 Hz. Colours indicate the average time of all time windows that contribute to the lapse cross-correlation, and this colour scheme is consistent across all three sub-figures (legend provided in the upper right corner of figure a). a. Original lapse cross-correlations prior to any corrections. b. Lapse cross-correlations after clock drift correction using skew values. c. Lapse cross-correlations after clock drift correction using the (OCloC-derived) drift rates (a_i) and initial clock errors (the b_i) estimated in this study.

In Figure 5.12, a more systematic and quantitative comparison of the linear drift based on our code ('OCloC-drift') and the skew values ('skew-drift') is presented for three OBS. The drifts of all the other OBS are shown in Appendix A.7. Figure 5.12 also shows the time offsets between the lapse cross-correlations and a reference lapse cross-correlation. We only use the cross-correlations between the OBS and land stations. For each station pair, the highest signal-to-noise ratio cross-correlation is selected as the reference cross-correlation. The time offsets correspond to the time shift that maximises the Pearson correlation coefficient between the reference cross-correlation and each cross-correlation. The skew-derived drift in the top figures assumes that there is no initial clock error at the onset of deployment (i.e., $b = 0$), and all subsequent time offsets are linearly interpolated based on this assumption. For the bottom figures, the time offset at deployment time corresponds to the OBS' b value (or initial clock error) estimated from the weighted least-squares inversion (again, all subsequent offset times are interpolated accordingly).

The OCloC-drift corrections of OBS01 and OBS02 (Figures 5.12a and 5.12b) seem to align better with the time offset between the cross-correlations and the reference cross-correlation. This is not the case for OBS10 (Figure 5.12c), where the skew-based clock drift aligns better with the time offsets. Station O10, however, is also one of the stations with the shortest recording time, which results in fewer time-lapses. This highlights one limitation of our approach: the need for longer monitoring time to include more lapse cross-correlations. In Section 5.5.3, we further discuss the implications of the OBS not monitoring for a full

year.

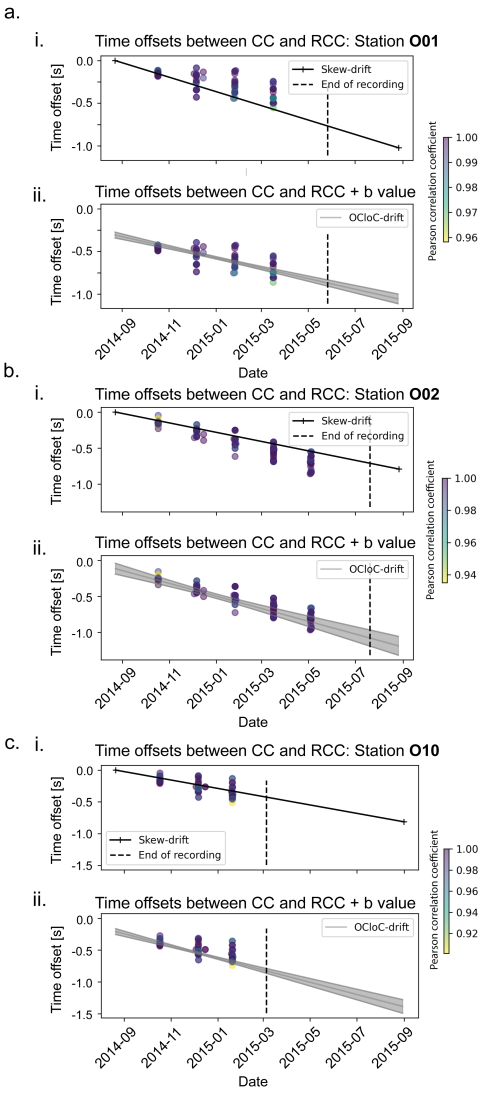


Figure 5.12: Comparison between (i) the skew-derived linear clock drift and (ii) the linear clock drift recovered using the weighted least-squares inversion for three selected OBS. **Top**: Time offsets between cross-correlations and a reference cross-correlation assuming no initial clock error at the onset of deployment. **Bottom**: Time offsets considering the initial clock error (b value) at deployment time. The drift based on our code (weighted least-squares inversion) and the confidence intervals are dubbed '`OCloC`-drift', while the drift based on the skew values is termed 'skew-drift'. The highest signal-to-noise ratio cross-correlation for each station pair is chosen as the reference cross-correlation. The depicted time offsets result from maximising the Pearson correlation coefficient between the reference cross-correlation and the other lapse cross-correlations, plus a correction based on the b value to the skew correction.

5.4.2. COMPARING INVERSION STRATEGIES

The fact that a non-uniform illumination pattern can break the time symmetry of the retrieved surface-wave responses is detrimental to the method presented in this chapter. Weemstra et al. (2021) showed that applying a weighted least-squares inversion based on station-to-station distances decreases the adverse effects of a non-uniform surface-wave-illumination pattern. Using synthetic recordings of ambient seismic noise, these authors demonstrated the advantage of the weighted least-squares estimate over the ordinary least-squares estimate. To evaluate the accuracy of the weighted least-squares estimator $\tilde{\mathbf{t}}^{(\mathrm{ins})}_{(\mathrm{wls})}$ in the presence of a non-uniform surface wave illumination, we compare it to the ordinary least-squares estimator $\tilde{\mathbf{t}}^{(\mathrm{ins})}_{(\mathrm{ols})}$. To do so, we used the bootstrap re-sampling approach introduced in Section 5.3.4. Figure 5.13 shows the histogram and cumulative distribution of 1000 bootstrap realisations of the a^* values of both inverse strategies. We used the same starting parameters and data vectors in both cases. The weighted inversions are shown in red, whereas the ordinary least-squares inversions are shown in blue. Figure 5.13a shows the distribution of the bootstrap realisations for all stations centred around 0 (mean values have been subtracted for each OBS individually). Figure 5.13b shows the cumulative distribution of the bootstrap realisations, with the 5th and 95th quantiles marked as vertical lines. The weighted least-squares distribution has narrower confidence intervals than the ordinary least-squares distribution.

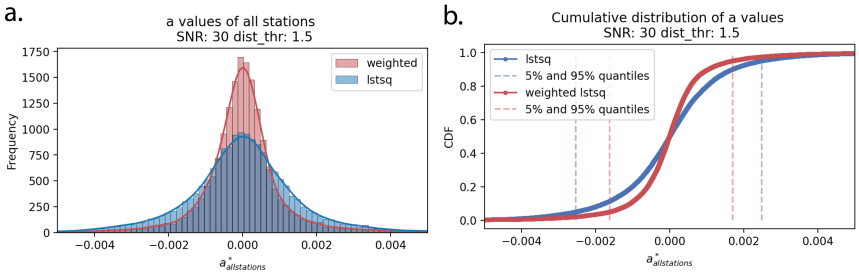


Figure 5.13: Bootstrap analysis of a^* values for all stations showing the 1000 realisations stacked for all stations. The mean values have been removed. The results of the weighted least-squares inversion are shown in red, and the results of the ordinary least-squares inversion are depicted in blue. a. Frequency histogram of the recovered values for all stations, with the probability density function overlaid. b. Cumulative distribution of the bootstrap realisations, with the 5th and 95th quantiles marked.

For data vectors associated with large station-to-station distance thresholds, we do not find significant differences between the two inversion strategies. This is expected because the threshold removes measurements associated with station couples that are closer to each other, which hence removes those lapse cross-correlations that are susceptible to larger illumination-related noise errors. For data vectors resulting from decreasing station-to-station distance thresholds, however, we find that the weighted inversion results in narrower bootstrap confidence intervals.

5.5. DISCUSSION

5.5.1. THE EFFECT OF THE SURFACE WAVE (NOISE) ILLUMINATION PATTERN

A limitation of the presented method is the fact that a non-uniform illumination pattern can lead to deviations of the retrieved surface-wave responses from the true surface-wave responses (e.g., Tsai, 2009; Weaver et al., 2009). As such, timing errors due to a non-uniform illumination pattern (captured in $\mathbf{n}^{(\mathrm{src})}$) lead to deviations of the recovered drift from the true clock drift. Consequently, bootstrap confidence intervals can be expected to be larger for more pronounced non-uniform illumination patterns. The bootstrapping results presented in Section 5.4.2 show that the distance-based weighted least-squares inversion results in both a lower spread of the distribution of the a^* values (Figure 5.13a) and a narrower range between the 5th and 95th quantiles (Figure 5.13b).

The bootstrapping results presented in Section 5.4.2 confirm the earlier, synthetic-data-based findings by Weemstra et al. (2021). Compared to the ordinary least-squares inversion, the weighted least-squares inversion decreases the adverse effects of a non-uniform illumination pattern. Note that the reasoning above can also be turned around: the fact that the weighted least-squares inversion results in more accurate clock drift estimates strongly suggests a (time-varying) non-uniform surface wave illumination. Given the available literature (Stehly et al., 2006; Mulargia, 2012; Weemstra et al., 2013) in general, and the large differences in SNR between (some of) the causal and acausal direct surface waves in particular, this can hardly be surprising.

5.5.2. VALIDATION USING ONLY LAND STATIONS

We run a separate test using only the lapse cross-correlations between the land stations. Lapse cross-correlations involving OBS are discarded. Apart from two stations, we pretend these land stations to be suffering from clock errors and hence neither eliminate the columns associated with any of them from $\mathbf{A}$ (in reality, those stations' a and b coincide with zero of course) nor any of its two entries from $\mathbf{t}^{(\mathrm{ins})}$. We subsequently compute the weighted least-squares estimator of $\mathbf{t}^{(\mathrm{ins})}$. The inversion yields drift rates (i.e., a_i) of maximum 0.1 s/year, which demonstrates that (i) noise on the data (i.e., non-zero $\mathbf{n}^{(\mathrm{src})}$ and $\mathbf{n}^{(\mathrm{spur})}$) prevents the recovery of drift rates of 0 s/day and (ii) that drift rates lower than 10^{-4} s/day cannot be recovered unambiguously (for our specific station configuration, noise illumination, and frequency band). The maximum b_i that is recovered has a value of 0.12 s, but this is an outlier in the sense that for most of the stations, the estimated initial clock error at $t^{(\mathrm{lps})} = 0$ does not exceed 0.05 s. Although the recovered a_i and b_i do not coincide with zero, we know that, in practice, the land stations do not suffer from clock drift and/or initial clock errors. Effectively, this experiment tells us that our approach allows us to successfully recover a seismic station's clock drift with an uncertainty of approximately 0.1 s/year.

5.5.3. ON THE VALIDITY OF THE ASSUMPTION OF LINEAR CLOCK DRIFT AND AN INITIAL CLOCK ERROR

While introducing our model (Section 5.2.2), we assumed the clock drift rates to be constant. Specifically, we formulated a time-dependent clock error $\delta t_i^{(\mathrm{ins})}(t^{(\mathrm{lps})})$ which drifts at a constant rate a_i, while allowing for a possible clock error b_i at $t^{(\mathrm{lps})} = 0$. The latter is in-

troduced to allow for an initial clock error at deployment time. This could, for example, be invoked by the temperature shock while the OBS is sunk (Zhang et al., 2023). We discuss in this section (i) the differences between the skew-derived drift rates and the recovered drift rates (i.e., the a_i), (ii) the fact that the b_i are non-zero, and (iii) the relation between these two observations.

The maximum difference between clock-error estimates obtained from the skew-derived drift rates and the weighted least-squares parameters is 0.62 s (Table 5.1), with most values being considerably smaller. Upon comparison with the time offsets between the individual lapse cross-correlations and the reference cross-correlation, the clock errors estimated using the weighted least-squares inversion align better than the skew-derived clock errors for most OBS. The weighted least-squares inversion considers an initial clock error at the deployment time. Figure 5.11 shows the better waveform alignment with an initial clock error (b) of -0.31 s. The estimated clock error at the time of recovery coincides with the skew for this station (see Table 5.1). However, if we would ignore the initial clock error, the skew-based corrections overcompensate the observed clock drift. This is evident from the shift of later lapse cross-correlations to positive times in Figure 5.11b. In contrast, `OCloC`-based corrections do not yield any (visible) residual drift (Figure 5.11c). This implies that an initial clock error at the time of deployment (i.e., a non-zero b) is indeed needed to explain the observed clock errors. Note that OBS O01 is used as an example because this station has one of the largest initial clock errors at deployment time, whereas its estimated clock error at recovery time almost coincides with the measured skew.

The fact that the `OCloC`-corrected lapse cross-correlations align better than the skew-corrected lapse cross-correlations can hardly be surprising. The drift rate estimates and the initial clock errors at the time of deployment are based on these very lapse cross-correlations. Therefore, it can be misleading to conclude from this observation that our approach yields more accurate drift rates than the skew-derived drift rates. This is because, although the weighted least-squares inversion mitigates the effect of arrival time shifts resulting from a non-uniform surface wave illumination, it will not undo it entirely. Illumination-related arrival time shifts (i.e., non-zero $\delta t_{i,j,k}^{(+,\mathrm{src})}$) may still have some effect. However, given the fact that (i) we averaged hourly cross-correlations over a period of 100 days, (ii) an SNR-threshold of 30 was imposed, and (iii) a station-to-station separation threshold of 2.5 wavelengths needed to be exceeded, we do not expect that these illumination-related arrival time shifts to be the cause of b_i as high as 0.3 or 0.4 s. The experiment discussed in Section 5.5.2 supports this claim.

Considering the above, we identify two possible explanations for the fact that the initial clock errors at the time of deployment are found to be non-zero and have values as high as (minus) 0.4 s. One explanation is that they result from the temperature shock during the OBS' descent to the ocean floor (see e.g., Zhang et al., 2023). In other words, they are real. This would not be surprising considering the experimental results by Gardner and Collins (2012), who find that the drift rates of the SEASCAN clocks may change significantly in the weeks after a temperature shock (in practice: after deployment). A second possible explanation for their deviation from zero stems from the fact that the OBS experience seasonal temperature variations during their deployment at (relatively) shallow depths. The study by Jochumsen et al. (2016), for example, reports on seasonal seawater-temperature variations on the order of five degrees centigrade at those depths. This is consistent with

the temperature variations within the data logger, which reveal annual temperature variations of about four degrees centigrade (these temperature sensors have a resolution of one degree only). In general, an OBS' drift rate is temperature dependent (Shariat-Panahi et al., 2009). However, we do not expect the drift rate of the SEASCAN clocks to suffer from such temperature variations: the SEASCAN clocks are temperature compensated (Gardner and Collins, 2012). Nonetheless, if such a seasonally varying drift would exist, it may be more appropriate to have our drift model (Equation 5.2) include a sinusoid with a period of one year. This may be the topic of future work.

Of all recovered drift rates, the drift rate by OBS O20 stands out (see Figure 5.10). This may well be explained by the fact that, compared to the other OBS, the logger and hence SEASCAN crystal oscillator of OBS O20 was newer. It was only two years old at the time of deployment, whereas the loggers (and hence clocks) of the other OBS were approximately 8 years old (Alfred Wegener Institute, personal communication, 2023). This matters because of a natural process in the crystal oscillator, which is referred to as ageing. Ageing implies that the drift rate of an oscillator slowly changes with time. Essentially, it is the time derivative of the drift rate (Gardner and Collins, 2012). The ageing of the crystal is a very important factor when it comes to the drift rate of the SEASCAN clocks, with younger crystals usually ageing faster. And even though ageing can be mitigated by regular recalibration of the SEASCAN clocks, it could well have been the cause of the larger drift rate of the SEASCAN clock of OBS O20.

5.5.4. PERFORMANCE IN THE ABSENCE OF LAND STATIONS?

OBS arrays in remote oceanic regions will not have the benefit of land stations in their near vicinity. In that case, lapse cross-correlations between the OBS and a station with a correct UTC reference time do not contribute to the data vector $\mathbf{t}^{(\text{app})}$. The system of equations will, in that case, be underdetermined (the rank of $\mathbf{A}$ being lower than the number of unknowns) and that weighted least-squares estimator $\tilde{\mathbf{t}}^{(\text{ins})}_{(\text{wls})}$ is chosen that has the lowest norm (see Section 5.2.6). The minimum-norm solution yields an estimator of $\mathbf{t}^{(\text{ins})}$ that allows the OBS recordings to be synchronised with respect to each other, but not with respect to UTC. In other words, the recovered b_i differs from the true (unknown) b_i by a common time shift, but the drift rates (i.e., the a_i) can still be recovered (with some uncertainty, of course). This is still useful as it would enable tomographic studies using only the OBS, or the localisation of seismic events (earthquake hypocentres) below the OBS array.

The accuracy of the recovered clock drift parameters depends linearly on the wave frequency. That is, lapse cross-correlations at higher frequencies will hence result in more accurate estimates of clock drift than lapse cross-correlations at lower frequencies, provided the illumination pattern and the SNR at both frequencies coincide. In practice, the latter is often not the case: lapse cross-correlations at lower frequencies usually benefit from more uniform noise illumination patterns (e.g., Yang and Ritzwoller, 2008). It may therefore be beneficial to include measurements associated with different frequency bands in $\mathbf{t}^{(\text{app})}$. It is beyond the scope of this work to investigate this here.

5.5.5. WHICH PROJECTS CAN BENEFIT FROM OCLOC?

Several methods can be used for correcting OBS clock errors. The fastest to implement is simply using the recovered skew values and assuming a linear drift rate. Here, however, we

show that this method may not be reliable. Moreover, it may not be possible because the battery has died before recovery. Other methods require correcting each OBS one by one by simply evaluating cross-correlations of ambient seismic noise in a non-automated manner. This requires a level of inspection that is not attractive (time-wise) for large-N OBS arrays. `OCloC` is suitable for such a type of deployment as it automatically and simultaneously computes clock drift rates of all OBS. Other cases where a GPS clock is lost, particularly with only on-land-station deployments, can significantly benefit from `OCloC`.

Projects that do not benefit from our approach are those with a limited deployment time. The reason is that `OCloC` requires the retrieval of interferometric surface wave responses at positive and negative times. In addition, lapse cross-correlations need to be computed at different lapse times $t_k^{(\mathrm{lps})}$ (at least two). To retrieve both responses, noise cross-correlations need to be averaged over a sufficiently long time. Here "sufficiently long" is location, processing, and frequency dependent (e.g., Yang and Ritzwoller, 2008; Seats et al., 2012; Snieder, 2004). In our case (Reykjanes peninsula, spectral whiting prior to cross-correlation, and 0.2–0.4 Hz frequency band), individual noise cross-correlations were averaged over 100 days to obtain surface waves with sufficiently high SNR at both positive (causal peak) and negative (acausal peak) time.

For projects that might not be suitable for `OCloC`, alternative solutions exist, such as the methodologies proposed by Sens-Schönfelder (2008); Hable et al. (2018); Loviknes et al. (2020); Jousset et al. (2013); Gouédard et al. (2014), among others.

5.6. CONCLUSIONS

To address the timing errors commonly affecting ocean-bottom seismometers (OBS), we introduce a new method that simultaneously estimates clock drift rates for large OBS networks. Our method relies predominantly on the time symmetry of retrieved interferometric surface-wave responses to estimate clock drift. Unlike existing approaches, our method enables recovery of the initial clock error at the time of deployment alongside the drift rate. Our method supports two deployment scenarios: i) OBS deployments including stations devoid of clock errors, such as land stations with constant access to a GNSS signal, and ii) OBS deployments without such stations. In both cases, drift rates can be recovered. However, when no stations devoid of clock errors are available, the absolute timing cannot be recovered because the network lacks an external time reference to anchor the clock corrections. We implement the proposed method in `OCloC`, an open-source Python package, and validate it using ambient seismic-noise data acquired during IMAGE's seismic campaign for monitoring geothermal energy in and around the Reykjanes Peninsula (Iceland). Our results show that all OBS in the network had clock drift. Furthermore, we find that skew-based correction methods, which are commonly applied for clock-error estimation, fail to accurately recover OBS drift rates. Finally, we demonstrate that a weighted least-squares inversion, in which receiver pairs are weighted by inter-station distance, significantly reduces errors arising from non-uniform noise illumination.

CODE AVAILABILITY

Name of the package: `OCloC` (OBS Clock Correction)
Contact: d.f.naranjohernandez@tudelft.nl

Program language: Python
The documentation can be found at: https://ocloc.readthedocs.io
The source codes are available for download at: https://github.com/davidn182/ocloc

REFERENCES

Aki, K. (1957). Space and time spectra of stationary stochastic waves, with special reference to microtremors. *Bulletin of the Earthquake Research Institute*, 35:415–456. https://doi.org/10.15083/0000033938.

Blanck, H., Jousset, P., Hersir, G. P., Ágústsson, K., and Ólafur G. Flóvenz (2020). Analysis of 2014–2015 on- and off-shore passive seismic data on the reykjanes peninsula, sw iceland. *Journal of Volcanology and Geothermal Research*, 391:106548. https://doi.org/10.1016/j.jvolgeores.2019.02.001.

Boschi, L. and Weemstra, C. (2015). Stationary-phase integrals in the cross correlation of ambient noise. *Reviews of Geophysics*, 53(2):411–451. https://doi.org/10.1002/2014RG000455.

Claerbout, J. F. (1968). Synthesis of a layered medium from its acoustic transmission response. *Geophysics*, 33(2):264–269. https://doi.org/10.1190/1.1439927.

Dongmo Wamba, M., Montagner, J.-P., and Romanowicz, B. (2023). Imaging deep-mantle plumbing beneath La Réunion and Comores hot spots: Vertical plume conduits and horizontal ponding zones. *Science Advances*, 9(4):eade3723. https://www.science.org/doi/full/10.1126/sciadv.ade3723.

Draganov, D., Wapenaar, K., Mulder, W., Singer, J., and Verdel, A. (2007). Retrieval of reflections from seismic background-noise measurements. *Geophysical Research Letters*, 34(4). https://doi.org/10.1029/2006GL028735.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics. https://epubs.siam.org/doi/abs/10.1137/1.9781611970319.

Efron, B. (1992). *Bootstrap Methods: Another Look at the Jackknife*. Springer New York, New York, NY.

Fichtner, A. (2014). Source and processing effects on noise correlations. *Geophysical Journal International*, 197(3):1527–1531.

Froment, B., Campillo, M., Roux, P., Gouédard, P., Verdel, A., and Weaver, R. (2010). Estimation of the effect of nonisotropically distributed energy on the apparent arrival time in correlations. *Geophysics*, 75. 10.1190/1.3483102.

Gardner, A. T. and Collins, J. A. (2012). Advancements in high-performance timing for long term underwater experiments: A comparison of chip scale atomic clocks to traditional microprocessor-compensated crystal oscillators. In *2012 Oceans*, pages 1–8. 10.1109/OCEANS.2012.6404847.

Geissler, W. H., Matias, L., Stich, D., Carrilho, F., Jokat, W., Monna, S., IbenBrahim, A., Mancilla, F., Gutscher, M.-A., Sallarès, V., and Zitellini, N. (2010). Focal mechanisms for sub-crustal earthquakes in the gulf of cadiz from a dense obs deployment. *Geophysical Research Letters*, 37(18). https://doi.org/10.1029/2010GL044289.

Gouédard, P., Seher, T., McGuire, J., Collins, J., and Hilst, R. D. (2014). Correction of ocean-bottom seismometer instrumental clock errors using ambient seismic noise. *Bulletin of the Seismological Society of America*, 104(3):1276–1288. https://doi.org/10.1785/0120130157.

Groos, J., Bussat, S., and Ritter, J. (2012). Performance of different processing schemes in seismic noise cross-correlations. *Geophysical Journal International*, 188(2):498–512. http://gji.oxfordjournals.org/cgi/doi/10.1111/j.1365-246X.2011.05288.x.

Hable, S., Sigloch, K., Barruol, G., Stähler, S. C., and Hadziioannou, C. (2018). Clock errors in land and ocean bottom seismograms: high-accuracy estimates from multiple-component noise cross-correlations. *Geophysical Journal International*, 214(3):2014–2034. https://doi.org/10.1093/gji/ggy236.

Halliday, D. and Curtis, A. (2008). Seismic interferometry, surface waves and source distribution. *Geophysical Journal International*, 175(3):1067–1087. https://doi.org/10.1111/j.1365-246X.2008.03918.x.

Hannemann, K., Krueger, F., and Dahm, T. (2014). Measuring of clock drift rates and static time offsets of ocean bottom stations by means of ambient noise. *Geophysical Journal International*, 196:1034–1042. 10.1093/gji/ggt434.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585:357–362. 10.1038/s41586-020-2649-2.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science Engineering*, 9(3):90–95. 10.1109/MCSE.2007.55.

Jochumsen, K., Schnurr, S. M., and Quadfasel, D. (2016). Bottom temperature and salinity distribution and its variability around iceland. *Deep Sea Research Part I: Oceanographic Research Papers*, 111:79–90. https://doi.org/10.1016/j.dsr.2016.02.009.

Jousset, P., Blanck, H., Franke, S., Metz, M., Águstsson, K., Verdel, A., Ryberg, T., Hersir, G., Weemstra, K., Bruhn, D., and Flovenz, O. (2016). "seismic tomography in reykjanes , sw iceland". In *European geothermal congress 2016*, pages 1–10. European Geothermal Congress 2016, EGC 2016 ; Conference date: 19-09-2016 Through 24-09-2016.

Jousset, P., Budi-Santoso, A., Jolly, A. D., Boichu, M., Surono, Dwiyono, S., Sumarti, S., Hidayati, S., and Thierry, P. (2013). Signs of magma ascent in lp and vlp seismic events and link to degassing: An example from the 2010 explosive eruption at merapi volcano, indonesia. *Journal of Volcanology and Geothermal Research*, 261:171–192. https://doi.org/10.1016/j.jvolgeores.2013.03.014.

Jousset, P., Hersir, G. P., Blanck, H., Kirk, H., Erbas, K., Hensch, M., Franke, S., Bruhn, D., and Huenges, E. (2020a). IMAGE (integrated methods for advanced geothermal exploration). *GFZ Data Services. Other/Seismic Network.* https://geofon.gfz-potsdam.de/doi/network/4L/2014.

Jousset, P., Mortensen, A. K., Ómar Fridleifsson, G., Ágústsson, K., and Gudmundsson, M. T. (2020b). Reykjanes, iceland: Structure and dynamics of mid-oceanic ridge geo/hydrothermal systems. *Journal of Volcanology and Geothermal Research*, 391:106692. https://doi.org/10.1016/j.jvolgeores.2019.106692.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C., and Wassermann, J. (2015). Obspy: A bridge for seismology into the scientific python ecosystem. *Computational Science and Discovery*, 8:014003. 10.1088/1749-4699/8/1/014003.

Lindner, F., Weemstra, C., Walter, F., and Hadziioannou, C. (2018). Towards monitoring the englacial fracture state using virtual-reflector seismology. *Geophysical Journal International*, 214(2):825–844. https://doi.org/10.1093/gji/ggy156.

Lobkis, O. and Weaver, R. (2001). On the emergence of the green's function in the correlations of a diffuse field. *Journal of the Acoustical Society of America*, 110(6):3011–3017.

Loviknes, K., Jeddi, Z., Ottemöller, L., and Barreyre, T. (2020). When Clocks Are Not Working: OBS Time Correction. *Seismological Research Letters*, 91(4):2247–2258. https://doi.org/10.1785/0220190342.

Matsumoto, H., Zampolli, M., Haralabus, G., Stanley, J., Mattila, J., and Meral Ozel, N. (2019). Interpretation of detections of volcanic activity at ioto island obtained from in situ seismometers and remote hydrophones of the international monitoring system. *Scientific Reports*, 9:19519. 10.1038/s41598-019-55918-w.

Mulargia, F. (2012). The seismic noise wavefield is not diffuse. *The Journal of the Acoustical Society of America*, 131(4):2853. 10.1121/1.3689551.

Naranjo, D., Parisi, L., Jónsson, S., Jousset, P., Werthmüller, D., and Weemstra, C. (2024). Ocean Bottom Seismometer Clock Correction using Ambient Seismic Noise. *Seismica*, 3(1). https://doi.org/10.26443/seismica.v3i1.367.

Rücker, C., Günther, T., and Wagner, F. M. (2017). pyGIMLi: An open-source library for modelling and inversion in geophysics. *Computers and Geosciences*, 109:106–123. 10.1016/j.cageo.2017.07.011.

Schnaidt, S. and Heinson, G. (2015). Bootstrap resampling as a tool for uncertainty analysis in 2-D magnetotelluric inversion modelling. *Geophysical Journal International*, 203(1):92–106. https://doi.org/10.1093/gji/ggv264.

Seats, K. J., Lawrence, J. F., and Prieto, G. A. (2012). Improved ambient noise correlation functions using Welch's method. *Geophysical Journal International*, 188(2):513–523. http://dx.doi.org/10.1111/j.1365-246X.2011.05263.x.

Sens-Schönfelder, C. (2008). Synchronizing seismic networks with ambient noise. *Geophysical Journal International*, 174(3):966–970. https://doi.org/10.1111/j.1365-246X.2008.03842.x.

Shapiro, N. M. and Campillo, M. (2004). Emergence of broadband rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, 31(7). https://doi.org/10.1029/2004GL019491.

Shariat-Panahi, S., Alegria, F. C., Làzaro, A. M., and del Rio, J. (2009). Time Drift of Ocean Bottom Seismometers (OBS). In *Proceedings of the 19th IMEKO World Congress on Fundamental and Applied Metrology*, pages 2548–2553.

Snieder, R. (2004). Extracting the green's function from the correlation of coda waves: A derivation based on stationary phase. *Phys. Rev. E*, 69:046610. 10.1103/PhysRevE.69.046610.

Stehly, L., Campillo, M., and Shapiro, N. M. (2006). A study of the seismic noise from its long-range correlation properties. *Journal of Geophysical Research: Solid Earth*, 111(B10). https://doi.org/10.1029/2005JB004237.

Sánchez-Pastor, P., Obermann, A., Schimmel, M., Weemstra, C., Verdel, A., and Jousset, P. (2019). Short- and long-term variations in the reykjanes geothermal reservoir from seismic noise interferometry. *Geophysical Research Letters*, 46(11):5788–5798. https://doi.org/10.1029/2019GL082352.

Tary, J. B., Géli, L., Henry, P., Natalin, B., Gasperini, L., Çomoğlu, M., Çağatay, N., and Bardainne, T. (2011). Sea-Bottom Observations from the Western Escarpment of the Sea of Marmara. *Bulletin of the Seismological Society of America*, 101(2):775–791. https://doi.org/10.1785/0120100014.

Tsai, V. C. (2009). On establishing the accuracy of noise tomography travel-time measurements in a realistic medium. *Geophysical Journal International*, 178(3):1555–1564. http://dx.doi.org/10.1111/j.1365-246X.2009.04239.x.

Wapenaar, K. and Fokkema, J. (2006). Green's function representations for seismic interferometry. *GEOPHYSICS*, 71(4):SI33–SI46. https://doi.org/10.1190/1.2213955.

Weaver, R., Froment, B., and Campillo, M. (2009). On the correlation of non-isotropically distributed ballistic scalar diffuse waves. *The Journal of the Acoustical Society of America*, 126(4):1817–1826. 10.1121/1.3203359.

Weemstra, C., Boschi, L., Goertz, A., and Artman, B. (2013). Seismic attenuation from recordings of ambient noise. *GEOPHYSICS*, 78(1):Q1–Q14. https://doi.org/10.1190/geo2012-0132.1.

Weemstra, C., de Laat, J. I., Verdel, A., and Smets, P. (2021). Systematic recovery of instrumental timing and phase errors using interferometric surface-waves retrieved from large-N seismic arrays. *Geophysical Journal International*, 224(2):1028–1055. https://doi.org/10.1093/gji/ggaa504.

Weemstra, C., Westra, W., Snieder, R., and Boschi, L. (2014). On estimating attenuation from the amplitude of the spectrally whitened ambient seismic field. *Geophysical Journal International*, 197:1770–1788. 10.1093/gji/ggu088.

Werthmüller, D., Rochlitz, R., Castillo-Reyes, O., and Heagy, L. (2021). Towards an open-source landscape for 3-D CSEM modelling. *Geophysical Journal International*, 227(1):644–659. https://doi.org/10.1093/gji/ggab238.

Wes McKinney (2010a). Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61. 10.25080/Majora-92bf1922-00a.

Wes McKinney (2010b). Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61. 10.25080/Majora-92bf1922-00a.

Yang, Y. and Ritzwoller, M. H. (2008). Characteristics of ambient seismic noise as a source for surface wave tomography. *Geochemistry, Geophysics, Geosystems*, 9(2). https://doi.org/10.1029/2007GC001814.

Zhang, J., Zhao, M., Yao, Y., and Qiu, X. (2023). Analysis of clock drift based on a three-dimensional controlled-source ocean bottom seismometer experiment. *Geophysical Prospecting*. https://doi.org/10.1111/1365-2478.13384.

6

CONCLUSIONS

In this thesis, we set out to advance seismic techniques for monitoring geothermal-heat-production projects in the Netherlands, supporting the safe expansion of the geothermal-energy sector in accordance with the Dutch Climate Agreement. This chapter summarises the main conclusions and outlines their relevance for seismic monitoring of geothermal operations.

6.1. MAIN CONCLUSIONS AND NOVEL CONTRIBUTIONS

In Chapter 2, we introduce a complete workflow for identifying, characterising, and interpreting seismicity related to geothermal-heat-production projects in urban environments. We apply our workflow to a temporary seismic array deployed to monitor a geothermal operation in South Holland province, where a low-magnitude seismic event had previously been identified. Despite high ambient-noise levels, we detected five additional microseismic events. These results indicate that low-magnitude seismicity may remain undetected when relying solely on sparse regional seismic networks. Our workflow shows that machine-learning phase pickers and stacking-and-migration techniques, combined with knowledge of the seismic-velocity model, can help detect seismic events that would otherwise be missed. We also demonstrate that applying template-matching detection is essential for identifying additional events that share similar waveform characteristics. Our detection results, which include the only seismic events recorded to date in South Holland, demonstrate the potential for monitoring microseismicity below traffic-light-system thresholds. These findings highlight the need for detection strategies tailored to improve the resolution of seismic monitoring of geothermal-heat-production projects in urban areas.

An important challenge that follows seismic-event detection is the estimation of hypocentres, along with a quantification of their associated uncertainties. In Chapter 2, we introduce a method that explicitly accounts for velocity-model uncertainties for hypocentre estimations. This approach is well-suited for analysing seismicity in the Netherlands, as it directly incorporates the uncertainties provided in the open-access Dutch regional seismic-velocity model. Using this method, we estimate the hypocentres and associated uncertainties of the six detected seismic events. The hypocentres occur near a local fault and close to the injection well. However, the hypocentre uncertainties prevented us from interpreting the geomechanical processes that triggered the seismicity. Both an origin on the mapped fault and activation of minor fractures near the injection well are plausible

within the uncertainty bounds.

Uncertainty in seismic-velocity models is one of the factors limiting the accuracy of hypocentre estimations. In Chapter 3, we demonstrate that microseismic events recorded during geothermal operations can be used to validate seismic-velocity models through body-wave seismic interferometry. By leveraging near-vertical incidence angles, we retrieved zero-offset virtual reflection responses (ZO-VRRs) beneath selected stations. We compared them to the estimated two-way travel times from existing P- and S-wave velocity models. The P-wave model showed strong agreement with the retrieved reflections, while the two available S-wave models failed to reproduce the observed arrival times. Our findings showed that under favourable conditions, such as dense receiver arrays and appropriate source–receiver geometries, body-wave seismic interferometry provides a cost-effective and independent method for velocity-model validation. This approach can support improved event characterization in geothermal monitoring workflows and is particularly valuable in urban environments, where conventional active-source surveys are often impractical.

Hypocentre resolution is limited by the design of the seismic network used for its detection. In Chapter 4, we implement a seismic-network-design workflow tailored for offshore monitoring of microseismicity associated with geothermal-energy and carbon-capture-and-storage operations. Using the Dutch K14-FA Carbon Capture and Storage field as a case study, we applied well-established Bayesian experimental design principles to identify receiver configurations that maximize the expected information gain about earthquake hypocentres. Our approach incorporates realistic subsurface information, travel-time uncertainties, and detectability constraints for low-magnitude events. We demonstrated that both optimisation strategies can produce cost-effective network designs capable of resolving microseismicity down to $M_L = 2.0$ in regions of interest, such as near faults and prospective injection wells. However, we did not test the addition of borehole stations (such as fibre-optic sensing or geophones in a borehole), which could potentially improve the hypocentre-depth resolution. Overall, our workflow optimizes receiver placement costs by optimizing station number and placement, while providing a quantitative estimate of the achievable hypocentre resolution of a given network design.

A major technical challenge for offshore seismic monitoring is the synchronisation of ocean-bottom seismometers, which lack access to an absolute-time reference such as the Global Navigation Satellite System. In Chapter 5, we introduced a method to correct clock errors in offshore seismic networks. Our method accounts for initial clock errors at deployment time and for the effect of non-uniform noise-illumination patterns when synchronizing OBS recordings. We applied it to the IMAGE network on the Reykjanes Peninsula, a network designed to monitor offshore geothermal energy. We show that all OBSs experienced clock drift and that some had initial time offsets at deployment. To support offshore geothermal monitoring, we released the open-source Python package `OCloC` to automate time corrections for large-N arrays.

6.2. LIMITATIONS AND FUTURE WORK

We identified the first indications of microseismicity in South Holland, but uncertainties in the hypocentre estimations limited our ability to confidently interpret what caused the events. These uncertainties can be reduced by conducting active-source surveys, which

would improve seismic-velocity models and hypocentre-estimation accuracy. Additionally, permanent stations could help identify additional events.

Although body-wave interferometry has the potential to improve the seismic-velocity model itself, its success is restricted by the availability of seismic events located within the stationary phase recorded at near-vertical incidence. The short monitoring period in Kwintsheul limited the number of such events. We recommend longer monitoring campaigns, ideally with permanent stations close to the epicentres of the events we reported. This method should be used in combination with other available geophysical or well data. This could enable a cost-effective way to validate and improve the seismic-velocity models for monitoring geothermal energy.

The network-design workflow that we proposed relies on available open data from the K14-FA field. This dataset did not include key details such as the precise reservoir boundaries, injection-well locations, or a high-resolution fault characterization. The absence of these inputs limited our network design. We therefore recommend extending the workflow with more detailed geological and operational information as it becomes available.

There is an urgent need to establish a reliable pre-injection seismicity baseline for geothermal-energy and carbon-capture-and-storage operations. Without a reliable seismicity baseline, it is not possible to differentiate between natural and injection-triggered seismicity. We recommend deploying seismic networks designed to increase the detectability and resolution of microseismic events in areas where geothermal-energy and carbon-capture-and-storage operations are planned.

A

APPENDIX

A.1. INFLUENCE OF STATION 030 ON HYPOCENTRE UNCERTAINTY

The influence of Station 030 on hypocentre location accuracy is illustrated in Figure A.1, which compares two inversions: one excluding data from Station 030 (Panels A and C) and one including it (Panels B and D). The likelihood probability density function (PDF) is visualised using a Viridis colour scale, where higher values indicate greater likelihood.

In the inversion without Station 030 (Figure A.1.A and C), the maximum likelihood location is estimated at x = 79,355.37, y = 446,982.12, and z = -2,084.03 m. The uncertainties for this estimate are: azimuthal gap of 168.68 °, depth uncertainty of 470.59 m, x uncertainty of 1,060.01 m, and y uncertainty of 638.47 m.

Including Station 030 in the inversion (Figure A.1.B and D) significantly reduces uncertainties and shifts the maximum likelihood location. The new estimate is at x = 79,506.80, y = 446,936.51, and z = -2,386.55 m. The uncertainties for this solution are: azimuthal gap of 146.07°, depth uncertainty of 201.68 m, x uncertainty of 454.29 m, and y uncertainty of 273.63 m.

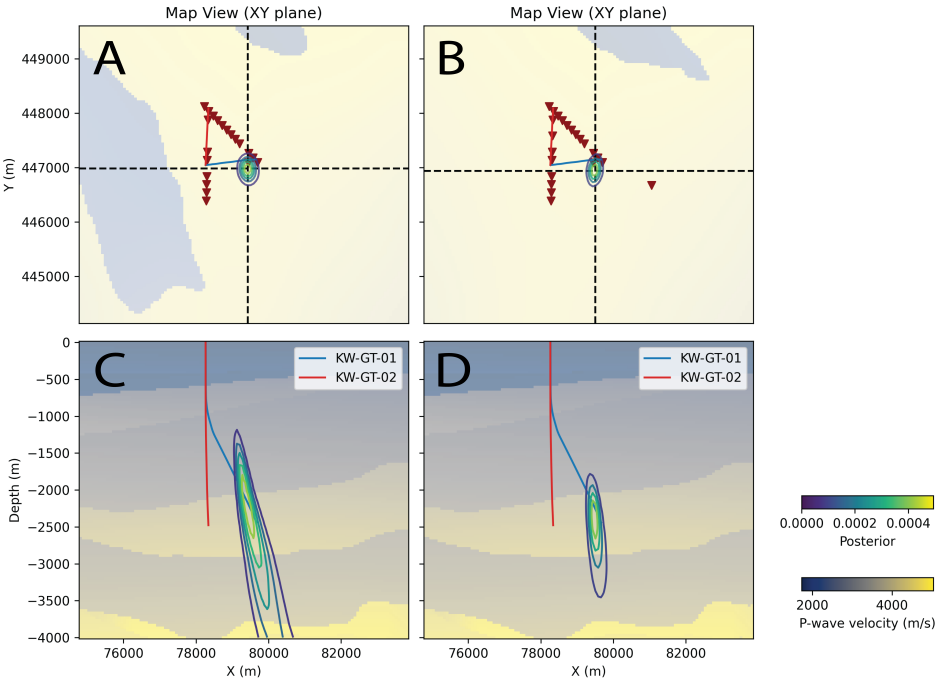


Figure A.1: Probability density function (PDF) projections for two seismic events. A and C show the map view and cross-section of the event on July 14, 2019, while B and D show the event on October 6, 2019. Red and blue lines represent the geothermal wells KW-GT01 and KW-GT02. Dark red triangles indicate seismic stations, and dotted lines mark the cross-section positions. PDF contours (Viridis scale) are overlaid on the velocity model used for the inversion.

A.2. EXAMPLE OF CROSS-CORRELATION PICK CORRECTION

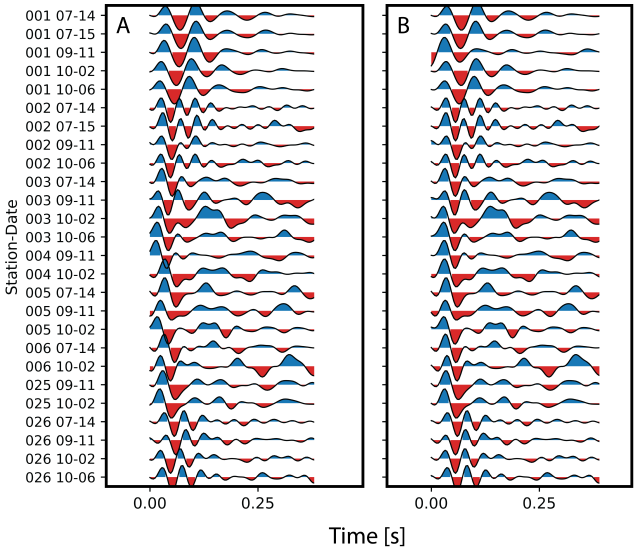


Figure A.2: Waveform alignment before (A) and after (B) applying the cross-correlation pick correction method of Deichmann et al., (1992). We use this method to correct bias from manual picks and to ensure consistent alignment across waveforms.

A.3. SUMMARY OF SOURCE-RECEIVER PAIRS SELECTED FOR BODY-WAVE SEISMIC INTERFEROMETRY

Table A.1: Incidence angles and conversion efficiencies by source and station

Source Date	Station	P inc. (°)	S inc. (°)	pS conv. (%)	sP conv. (%)
2019-06-23	001	3.5	1.4	0.1	0.2
2019-07-15	001	2.8	1.2	0.0	0.1
2019-10-02	001	5.4	2.3	0.1	0.4
2019-10-06	001	5.4	2.2	0.1	0.4
2019-07-14	001	3.6	1.5	0.1	0.2
2019-09-11	001	7.6	3.1	0.2	0.8
2019-06-23	002	3.0	1.2	0.0	0.1
2019-07-15	002	3.0	1.2	0.0	0.1
2019-10-02	002	4.4	1.8	0.1	0.3
2019-10-06	002	4.4	1.9	0.1	0.3
2019-07-14	002	3.4	1.4	0.0	0.2
2019-09-11	002	6.3	2.6	0.2	0.5
2019-06-23	003	4.0	1.6	0.1	0.2
2019-07-15	003	4.4	1.8	0.1	0.3

Continued on next page

Source Date	Station	P inc. (°)	S inc. (°)	pS conv. (%)	sP conv. (%)
2019-10-02	003	4.6	1.9	0.1	0.3
2019-10-06	003	4.8	2.0	0.1	0.3
2019-07-14	003	4.4	1.8	0.1	0.3
2019-09-11	003	5.7	2.4	0.1	0.5
2019-06-23	004	5.7	2.4	0.1	0.4
2019-07-15	004	6.3	2.6	0.2	0.6
2019-10-02	004	6.1	2.6	0.2	0.5
2019-10-06	004	6.2	2.6	0.2	0.5
2019-07-14	004	6.1	2.5	0.2	0.5
2019-09-11	004	5.9	2.4	0.1	0.5
2019-06-23	005	7.7	3.2	0.2	0.8
2019-07-15	005	8.4	3.4	0.3	0.9
2019-10-02	005	8.1	3.4	0.3	0.9
2019-10-06	005	8.1	3.4	0.3	0.9
2019-07-14	005	7.9	3.3	0.3	0.8
2019-09-11	005	6.9	2.9	0.2	0.7
2019-09-11	006	8.4	3.5	0.3	0.9
2019-06-23	025	1.3	0.6	0.0	0.0
2019-07-15	025	0.5	0.2	0.0	0.0
2019-10-02	025	3.1	1.3	0.0	0.1
2019-10-06	025	3.0	1.3	0.0	0.1
2019-07-14	025	1.4	0.6	0.0	0.0
2019-09-11	025	5.6	2.3	0.1	0.4
2019-06-23	026	1.4	0.6	0.0	0.0
2019-07-15	026	2.2	0.9	0.0	0.1
2019-10-02	026	1.8	0.7	0.0	0.0
2019-10-06	026	2.0	0.8	0.0	0.1
2019-07-14	026	1.9	0.8	0.0	0.1
2019-09-11	026	4.0	1.7	0.1	0.2

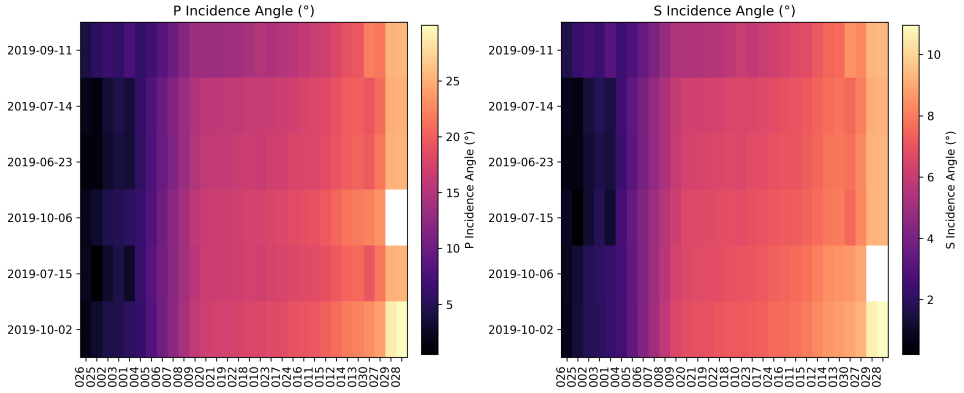


Figure A.3: Heatmap showing incidence angles from the microseismic events to the different receivers of the Kwintsheul network.

A.4. SPECTROGRAM ANALYSIS

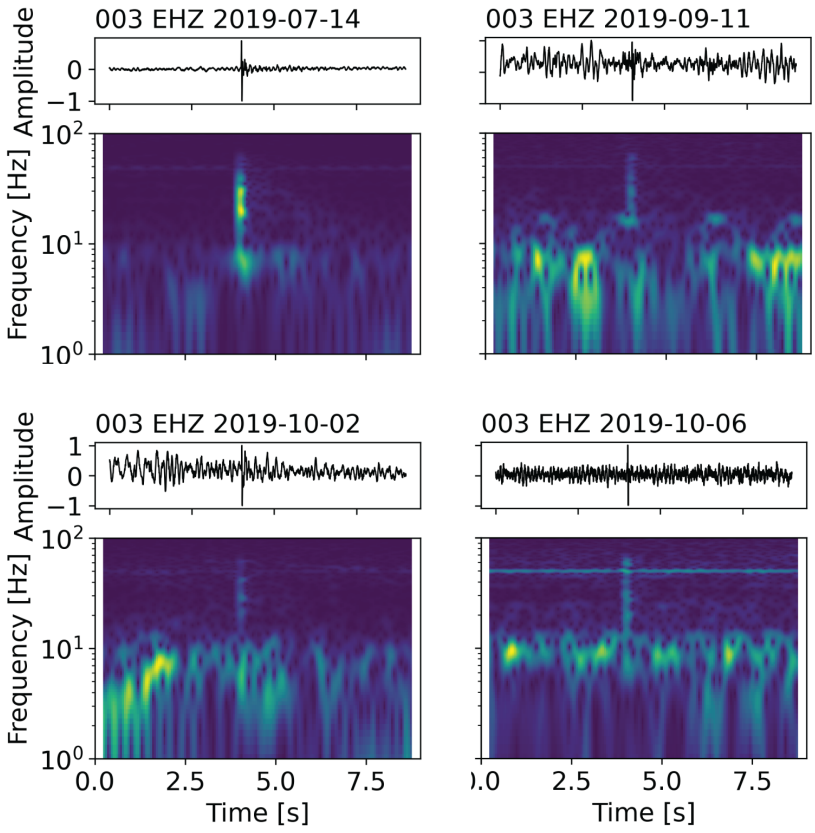


Figure A.4: Normalized, detrended vertical-component waveforms (top) and corresponding spectrograms (bottom) for four microseismic events recorded at receiver 003. Brighter colors in the spectrograms indicate higher amplitudes at each frequency.

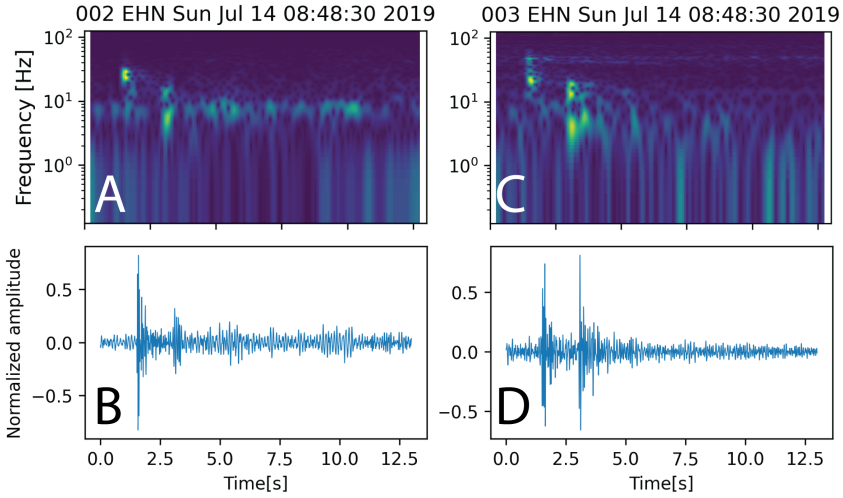


Figure A.5: Spectrograms and normalized waveforms of the North component (EHN) showing the P- and S-wave arrivals for two receivers. B. Spectrogram at receiver 002. C. Corresponding seismic trace. D. Spectrogram of receiver 003. D. Corresponding seismic trace. Brighter colors indicate higher amplitude at each frequency.

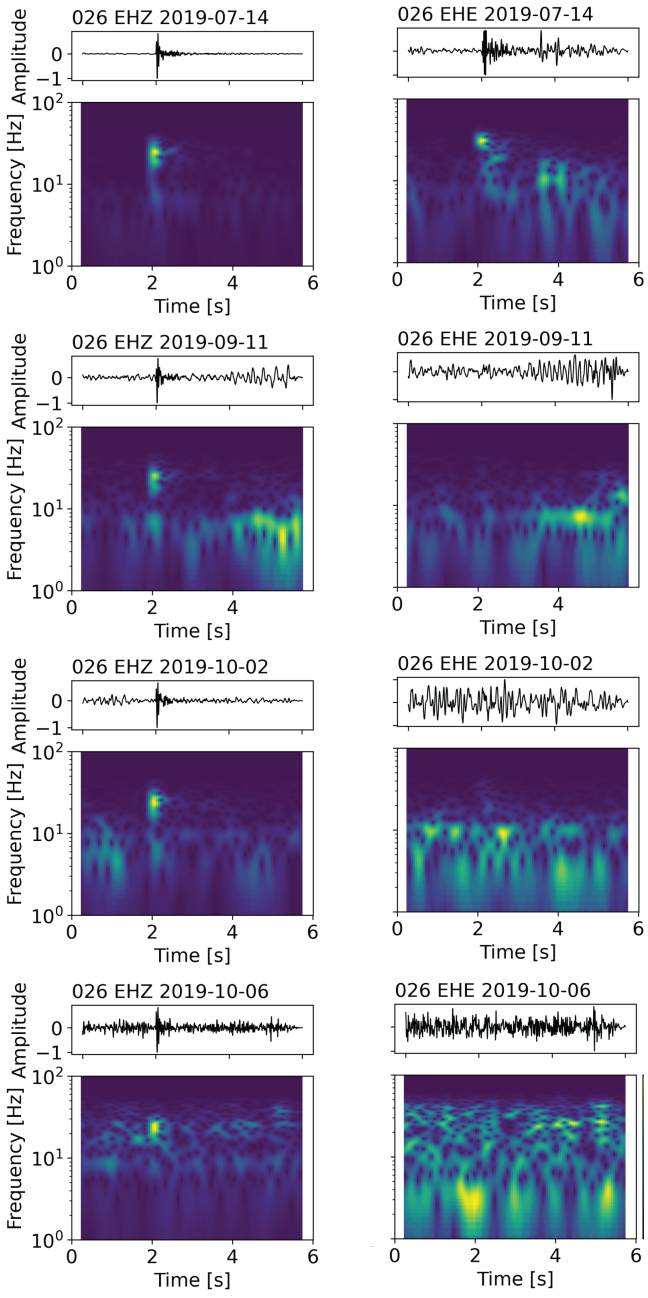


Figure A.6: Normalized, detrended vertical-component waveforms (top) and corresponding spectrograms (bottom) for four microseismic events recorded at receiver 006. Brighter colors in the spectrograms indicate higher amplitudes at each frequency.

A.5. MATRIX FORMULATION

To clarify the rather mathematical description of the inverse problem, let's consider the following example. If one would compute monthly time-averaged cross-correlations for an OBS deployment of 10 stations that would last one full year, $N^{(\mathrm{lps})}$ would be 12 and N (obviously) 10. This would imply the number of rows of the matrix $\mathbf{A}$ (and the length of the vectors $\mathbf{t}^{(\mathrm{app})}$, $\mathbf{n}^{(\mathrm{src})}$, and $\mathbf{n}^{(\mathrm{spur})}$) would coincide with $12 \times (10 \times 9)/2 = 540$. The length of $\mathbf{t}^{(\mathrm{ins})}$ would coincide with 20 (2×10) and so would the number of columns of $\mathbf{A}$. Expressing then $t_k^{(\mathrm{lps})}$ in terms of days (instead of seconds, which is the customary unit of time) and setting it to zero at the onset of the OBS deployment, this would imply $t_1^{(\mathrm{lps})} \approx 15$, $t_2^{(\mathrm{lps})} \approx 46$, and so on, and so forth.

For N stations, vector $\mathbf{t}^{(\mathrm{ins})}$ can be written as:

$$\mathbf{t}^{(\mathrm{ins})} \equiv \begin{pmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \\ \vdots \\ a_N \\ b_N \end{pmatrix}, \tag{A.1}$$

To aid in the interpretation of Equation (5.7), we depict below (Figure A.7) the rows associated with the first lapse cross-correlations (i.e., the lapse cross-correlations associated with $t_1^{(\mathrm{lps})}$) are shown in light blue. In addition, we have depicted in purple (for $t_1^{(\mathrm{lps})}$ only) the elements of the matrix associated with the lapse cross-correlation between stations 1 and 2, and in yellow the elements of the matrix associated with the lapse cross-correlation between stations 2 and 3. Note that, as it stands, the matrix in Figure A.7 is rank deficient. This implies that the system of equations is underdetermined, and a unique solution does not exist. If one of the 10 stations is a land station devoid of clock errors, the two columns associated with that station could be eliminated from $\mathbf{A}$ (that OBS' a and b would coincide with zero), and its two entries eliminated from $\mathbf{t}^{(\mathrm{ins})}$. The resulting matrix $\mathbf{A}$ would be full rank, and a unique estimator of $\mathbf{t}^{(\mathrm{ins})}$ would exist.

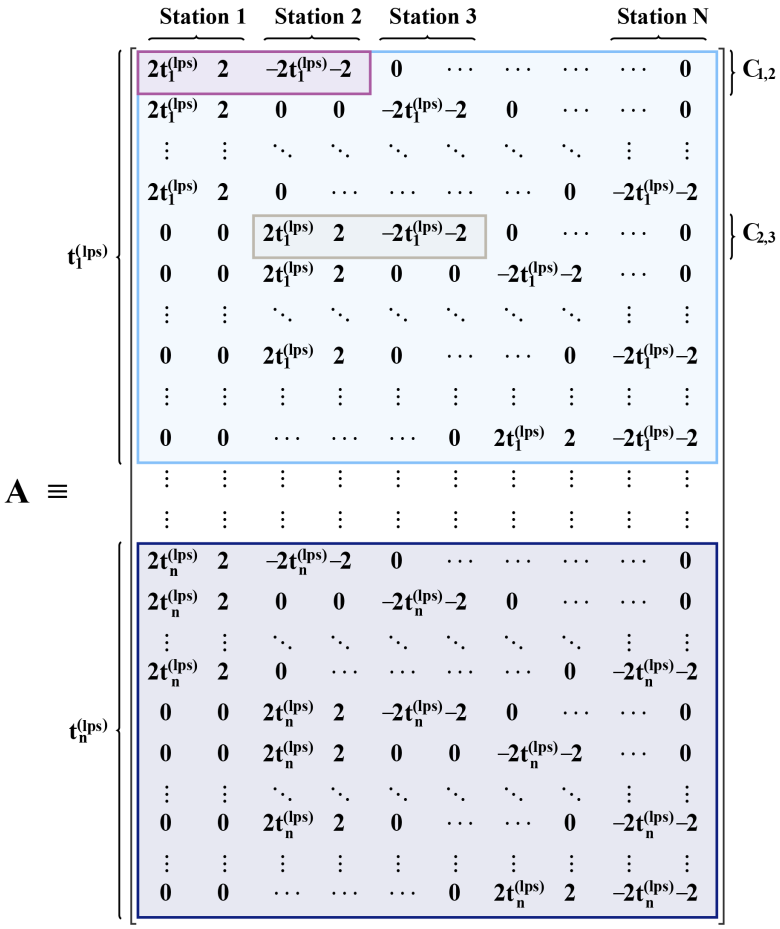


Figure A.7: Example of matrix **A** when using N stations and n lapse times

A.6. DETECTION OF OUTLIERS

When measuring the $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$ it might be possible to get an erroneous measurement due to a phenomenon that is similar in nature to what is referred to as 'cycle skipping' in full-waveform inversion (e.g., Warner and Guasch, 2014). That is, the measured $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$ deviates from the "true value" by approximately one period (see also Weemstra et al., 2021). Needless to say, inclusion of these measurements in the inversion leads to incorrect a_i and b_i. To prevent such measurements, we implemented a method that compares the measured $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$ with the expected $t_{i,j,k}^{(+,\text{app})} + t_{i,j,k}^{(-,\text{app})}$. The latter is computed using the a's and b's recovered during a first inversion. After identifying the outliers, i.e., points that do not follow the overall trend (blue areas in Figure A.8), we set a certain threshold for removing or keeping measurements. Repeating this process multiple times allows us to"clean" the data vector from such measurements.

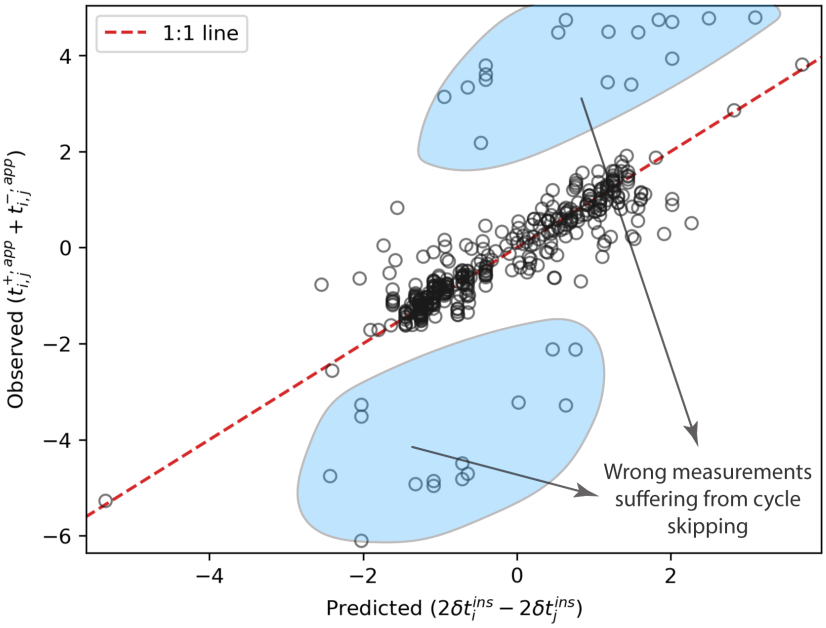


Figure A.8: Observed time symmetry shifts plotted against the estimated time symmetry shifts after inversion. The clusters in blue might indicate inaccurate measurements product of cycle skipping.

A.7. CLOCK DRIFTS OF EACH OBS STATION

In this appendix section, we provide an extended comparison of linear corrections from our code against the skew values for each OBS. The figures display the time offsets between the cross-correlations and a chosen reference cross-correlation, offering a detailed view of our approach's alignment with standard skew value corrections across different OBSs. Stations with very high uncertainty (for example O03, O06, and O08) yielded fewer data points ($t_{i,j,k}^{+,app} + t_{i,j,k}^{-,app}$) as their SNR and distance separation did not meet the required thresholds (see Figure 5.6 for examples of those cross-correlations).

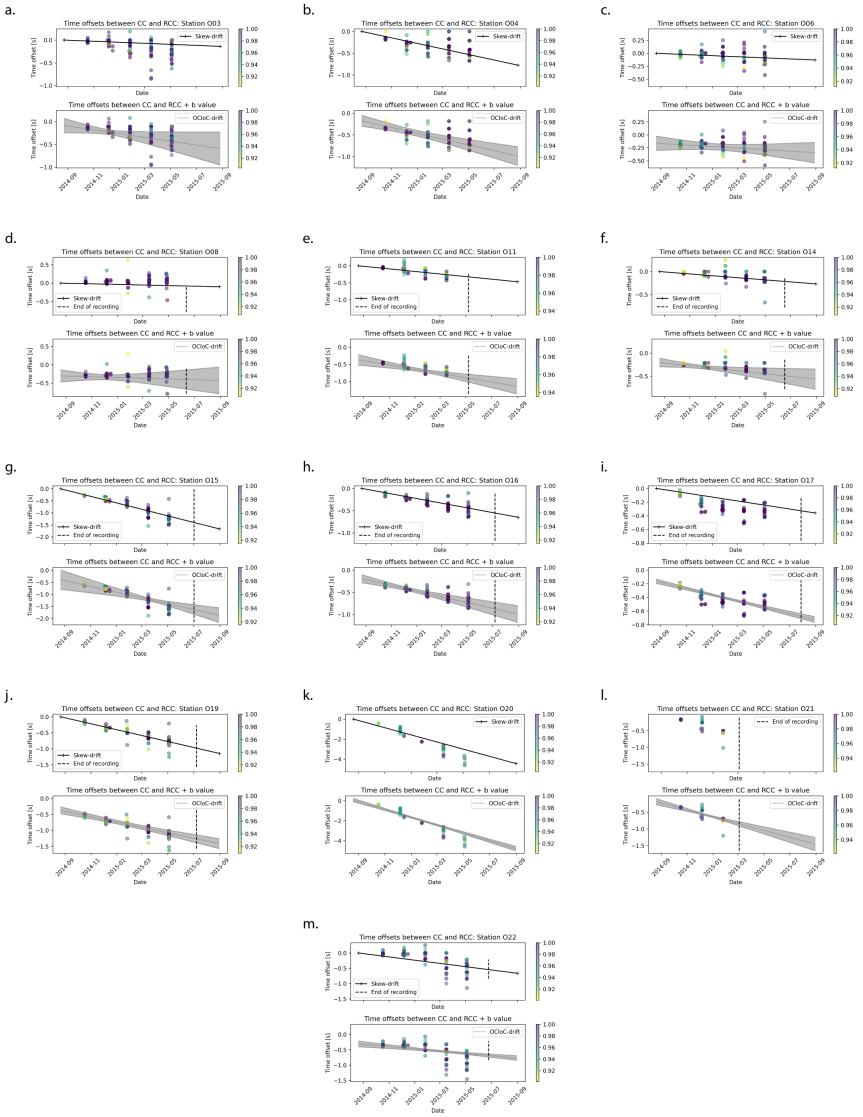


Figure A.9: Comparison between the observed clock drift, (i) the skew-derived linear clock drift, and (ii) the linear clock drift recovered using the weighted least-squares inversion of each OBSs (except O01, O02, and O10, which are in Section 5.4). **Top**: Time offsets between cross-correlations and a reference cross-correlation (RCF) assuming no initial clock error at the onset of deployment. **Bottom**: Time offsets considering the initial clock error (b value) at deployment time. The drift based on our code (weighted least-squares inversion) and the confidence intervals is dubbed '`OCloC`-drift', while the drift based on the skew values is termed 'skew- drift'. The highest signal-to-noise ratio cross-correlation for each station pair is chosen as the RCF. The depicted time offsets result from maximizing the Pearson correlation coefficient between the RCF and the other lapse cross-correlations, plus a correction based on the value of b in the skew correction.

REFERENCES

Warner, M. and Guasch, L. (2014). Adaptive waveform inversion - FWI without cycle skipping - theory. *Proceedings*. https://doi.org/10.3997/2214-4609.20141092.

Weemstra, C., de Laat, J. I., Verdel, A., and Smets, P. (2021). Systematic recovery of instrumental timing and phase errors using interferometric surface-waves retrieved from large-N seismic arrays. *Geophysical Journal International*, 224(2):1028–1055. https://doi.org/10.1093/gji/ggaa504.

CURRICULUM VITÆ

David Francisco NARANJO HERNANDEZ

EDUCATION

2021 - 2026 **Ph.D. in Applied Geophysics and Seismology**
Delft University of Technology, the Netherlands, and RWTH Aachen, Germany
Thesis: *Advanced Seismic Monitoring Tools for Monitoring Dutch Geothermal Systems*
Supervised by Prof.dr. D.S. Draganov, Prof.dr. F. Wellmann, and Dr. C. Weemstra

2022 **Industry Secondment**
EBN B.V., Utrecht, The Netherlands
Supervised by Dr. Sjoukje de Vries, Dr. Marc Hettema, M.Sc. Jorien Schaaf, and M.Sc. Mara van Eck van der Sluijs

2023 **Research Secondment**
RWTH Aachen, Germany
Supervised by Prof. Florian Wellmann

2018 - 2020 **M.Sc. in Applied Geophysics**
Delft University of Technology, the Netherlands
ETH Zurich, Switzerland
RWTH Aachen, Germany

2013 - 2018 **B.Sc. in Geosciences**
Universidad de Los Andes, Colombia

1999 - 2013 **Secondary School**
Buckingham School, Bogotá, Colombia

WORK EXPERIENCE

10.2020 - 02.2021	**Visiting Student Research Program** Ocean-bottom seismometer clock drift correction KAUST, Saudi Arabia
07.2019 - 10.2019	**Research assistant** Microseismic monitoring of the Theistarykir Geothermal power plant, Iceland. GFZ, Potsdam, Germany
07.2018 - 08.2018	**Research Assistant** Coseismic slip distribution inversion of the Mw 6 Mutatá earthquake. Universidad de Los Andes, Colombia

HONORS, SCHOLARSHIPS, GRANTS AND FUNDING

2026 Travel Grant from the Seismological Society of America Annual Meeting (2026), Pasadena, California, (US). Abstract title: Seismic Interferometry as a Cost-Effective Tool for Seismic Velocity-Model Validation in Geothermal Fields

2024 Travel Grant from the SEG/Chevron Student Leadership Symposium (SLS) and IMAGE, Houston, TX, USA, Aug 25–30.

2024 Outstanding performance as EAGE Student Chapter 2023-2024 in the Europe Region.

2023 Travel Grant for International Training Course on Seismology and Risk Mitigation at Helmholtz Centre Potsdam GFZ, Germany, June 21-July 19.

2018 Holland Scholarship by the Dutch Ministry of Education.

2018 Colfuturo Scholarship.

2018 Travel Grant from the Seismological Society of America for the SSA Annual Meeting, Miami, FL, May 14-17. Abstract title: Co-Seismic Slip Distribution Inversion Applied to the Mutata Earthquake, Colombia

REVIEWER

2025 Geophysical Journal International
2025 McGill University - Seismica
2026 Southern African Geophysical Association 2026 Conference

VOLUNTEER WORK

2023 President of the Delft Organization of Geophysics Students (DOGS), TU Delft, Netherlands.

2022 Conference organizer for the 13th European Geothermal PhD Days, Aachen, Germany, April 27-29.

2021 PhD committee representative of the Applied Geophysics & Petrophysics section

TEACHING AND SUPERVISION

2025 **Teaching**
Course: Monitoring microseismicity at geothermal projects.
Program: EASYGO-ON Training Week on Geophysical Monitoring in Geothermal Systems.
University: ETH Zurich (Switzerland).

2022 **M.Sc. thesis supervision**
Guus Hoogewerf (2022). Clock error correction for ocean bottom seismometers with the new OCloC package, using ambient seismic noise from the Saudi Arabian Red Sea coastal area.

2022 **B.Sc. thesis supervision**
Carleyn M. den Ouden (2022). Using beamforming to maximize the detection capability of Kwintsheul's seismic array.

2014-2018 **Teaching Assistant**
University: Universidad de Los Andes (Colombia) Courses: Seismic and Seismology, Exploration Geophysics, Integral Calculus and Differential Equations, Geosciences, and Paleontology.

List of Publications

Journal papers

1. **Naranjo, D.**, Isken, M., Boullenguer, B., Toledo, T., Weemstra, C., & Draganov, D.S. (2025). Urban Challenges in Seismology: Seismic Monitoring of Kwintsheul Geothermal Operation (The Netherlands). Netherlands Journal of Geosciences (accepted for publication).

2. **Naranjo, D.**, Parisi, L., Jónsson, S., Jousset, P., Werthmüller, D., & Weemstra, C. (2024). Ocean Bottom Seismometer Clock Correction using Ambient Seismic Noise. Seismica, 3(1). doi:10.26443/seismica.v3i1.367

3. Tary, J.-B., Mojica Boada, M. J., Vargas, C. A., Montaña Monoga, A. M., **Naranjo-Hernandez, D. F.**, & Quiroga, D. E. (2022). Source characteristics of the Mw 6 Mutatá earthquake, Murindo seismic cluster, northwestern Colombia. Journal of South American Earth Sciences, 115, 1-10. doi:10.1016/j.jsames.2022.103728

Extended abstracts

2. **Naranjo, D.**, Draganov, D., Polychronopoulou, K., De Bas, M., & Weemstra, C. (2022). Seismic monitoring of Nature's Heat Geothermal Reservoir in Kwintsheul (Netherlands). Paper presented at European Geothermal Congress 2022, Berlin. Download abstract (PDF)

1. Erbaş, K., Schäfer, F., Guþmundson, Á., Júlíusson, E., Páll Hersir, G., Warburton, R.J., Bernard, J-D., Portier, N., Hinderer, J., Drouin, V., Sigmundsson, F., Ágústsson, K., Männel, B., Güntner, A. Voigt, C., Schöne, T., Jolly, A, Hjartasson, H., **Naranjo, D.**, Jousset, P. (2020). Continuous Microgravity Monitoring in a Volcanic Geothermal Field: Integrated Observational Approach in Þeistareykir, NE Iceland. In Proceedings World Geothermal Congress 2020+1, paper 13124, Reykjavik, Iceland, April-October 2021. Download abstract (PDF)

Conference proceedings (first author)

8. **Naranjo, D.**, & Draganov D.S. (2026). Seismic Interferometry as a Cost-effective Tool for Seismic Velocity-Model Validation in Geothermal Fields. Accepted abstract, Seismological Society of America Annual Meeting, Pasadena, USA, April 14–17, 2026.

7. **Naranjo, D.**, Isken, M., Boullenguer, B., Toledo, T., Weemstra, C., & Draganov, D.S. (2025). Workflow for Seismic Monitoring of Dutch Geothermal Operations: Motivation for the Kwintsheul-Geo4All Active-Source Seismic Survey. Dedicated

session GE: Heating up the low countries: the development of geothermal methods in the Netherlands. Invited presentation at *Sixth EAGE Global Energy Transition Conference & Exhibition (GET 2025)*, Rotterdam, The Netherlands.

6. **Naranjo, D.**, Boullenger, B., and Draganov, D. (2024). Overcoming Urban Noise and Model Uncertainty: Induced Seismicity Monitoring in Dutch Geothermal Fields. 587. Abstract from 39th General Assembly of the European Seismological Commission (ESC2024), Corfu, Greece. Download Abstract (PDF)

5. **Naranjo, D.**, Weemstra C., and Draganov D. Probabilistic hypocentre determination – velocity model uncertainties (2023). International Training Course on Seismology, Seismic Data Analysis, Hazard Assessment, and Risk Mitigation. Organized and sponsored by Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences

4. **Naranjo, D.**, Parisi, L., Jonsson, S., Jousset, P., Werthmüller, D., & Weemstra, C. (2023). Determining clock errors of ocean-bottom seismometers: an ambient-noise based method designed for large-scale ocean bottom deployments. Abstract from EGU General Assembly 2023, Vienna. Austria. doi:10.5194/egusphere-egu23-15012

3. **Naranjo, D.**, Draganov, D. S., & Weemstra, C. (2022). Seismic monitoring of Nature's Heat Geothermal Project in Kwintsheul (Netherlands). 54-54. Abstract from 13th European Geothermal PhD Days - 2022, Aachen, North Rhine-Westphalia, Germany. Download abstract (PDF)

2. **Naranjo, D.**, Parisi, L., Jonsson, S., Jousset, P., Werthmüller, D., & Weemstra, C. (2021). Determining OBS clock drift using ambient seismic noise. Paper presented at the Nederlands Aardwetenschappelijk Congres (NACGEO) 2021.

1. **Naranjo, D.F.**, Tary, J. B., (2018). Co-seismic Slip Distribution Inversion Applied to the Mutata Earthquake, Colombia. Poster presented at the 2018 Seismology of the Americas Meeting, Miami, Florida, USA, 14-17 May 2018. doi:10.1785/0220180082

CONFERENCE PROCEEDINGS (CO-AUTHOR)

5. Martuganova, E., **Naranjo, D.**, Kühn, D., and Barnhoorn, A. (2025). Ensuring safe North Sea CO2 storage: the design of robust seismic networks to enable focal mechanism analyses for stress field orientation . EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-615, doi:10.5194/egusphere-egu25-615

4. Martuganova, E., **Naranjo Hernandez, D. F.**, Kühn, D., and Barnhoorn, A. (2024) Assessing earthquake focal mechanisms in the North Sea for risk mitigation of large-scale CO2 injections, EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-8973, doi:10.5194/egusphere-egu24-8973

3. Draganov, D., **Naranjo, D.**, Polychronopoulou, K., & Weemstra, C. (2023). The potential of probabilistic moment-tensor inversions for the characterization of geothermal reservoirs in urban environments. Abstract from EGU General Assembly 2023, Vienna, Austria. doi:10.5194/egusphere-egu23-14534

2. Schäfer, F., Jousset, P., Toledo, T., Güntner, A., Schöne, T., Naranjo, D., Erbas, K., Júlíusson, E., & Warburton, R. (2020). Integrated microgravimetric and seismic monitoring approach in the Theistareykir volcanic geothermal field (North Iceland). *EGU General Assembly 2020*, Online, 4–8 May 2020, EGU2020-11801. https://doi.org/10.5194/egusphere-egu2020-11801

1. Tary, J. B. and **Naranjo Hernández, D. F.** (2018). Finite-fault slip inversion for the Mw 6 Mutata earthquake (2016), Colombia. Poster presented at the AGU Fall Meeting Dec. 2018. doi:2018AGUFM.T43E0457T

DATA PUBLICATIONS

2. Martuganova, E. M., Naranjo Hernandez, D. F., Barnhoorn, A. (2025). A hands-on guide for computing and exploring focal mechanisms in the North Sea for risk mitigation of large-scale CO2 injections. TU Delft - 4TU.ResearchData. 10.4121/E1254235-C512-4310-9731-9499C2D2FA2F

1. Erbaş, Kemal, Toledo, T., **Naranjo, D.**, Metz, M., Rietz, F., Forster, F., and Jousset, P. (2021). Dataset of the 3P seismic network at the Þeistareykir geothermal area, Iceland, 2017–2021. GFZ Data Services. Dataset/Seismic Network. doi:10.14470/8C7562967492