Small-scale effects on global seismic waves

Marta Pieńkowska-Côte
DPhil Earth Sciences
Wolfson College
University of Oxford

Trinity 2019

To my beloved sister, Olga.

Acknowledgements

Thank you.

To my supervisor Tarje for introducing me to geophysics, for guiding me along the way, and for showing support and understanding. To Stuart and David for their patience and for the countless hours they spent explaining the painfully obvious things.

To my husband Pierre who loves and always loved, who found the right words when I was going to give up, and who somehow managed to put up with my personal roller-coaster of the last few years. To my parents, Barbara and Ryszard, for never doubting and for always being present, and for the support we had for each other in the difficult times. To my beloved sister Olga who was also my best friend that I always unconsciously imitated; I have still not re-defined myself in a world without her. To her two extraordinary children, Marta & Paweł, for their limitless joy and unconditional love.

To my most amazing housemates Nomi, Brad and Tom with whom I created a community like no other. Especially to Nomi for being the friend-turned-into-family who made Oxford feel like home. To the many friends scattered around the world who have been part of my story since way before the PhD, especially to Iza, Ewa, Marta, Ania, Vanessa, and Adina, wonderful women who always care, listen, share and remember. To fellow earth scientists Arola, Dave, Sean, Tim, Pete and Jo who made my lost soul of a mathematician find its place in the department and who became great friends. To the Israeli gang, Smadar and Daniel (a.k.a. Smadaniel), and Michal and Nir, for feeding me on most Friday nights during the past three years, and for making me part of the family. To Baptiste who never asked but always encouraged, and whose climbing enthusiasm became the foundation of my final year in Oxford. To Joey for showing support when I needed it, and to the whole Seymour family who entrusted me with Nemo – the absolute highlight of my final weeks of PhD writing.

To colleagues and friends from the Oxford seismology group – Tarje, Karin, Kuangdai, Maria, Kasra, Afsaneh, Harriet, Paula, Kara, Ben×2, Alex, Eoghan, Tom, Andy, Claudia, Laura, Matthew, Lucia – for the many scientific and not-so-scientific discussions. To all "WAVES fellows" for the good times and the readiness for beer. Especially to Gil for his support on Corsica. And finally to the office – Maria, Harriet, Brook, Eoghan and Alex – for staying tuned whenever I needed to vent my frustrations and complain. Especially to Harriet for sharing that nasty walk through The Valley with me.

I also gratefully acknowledge funding received towards my PhD from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement (No 641943), from the Earth Sciences Burdett Coutts fund in Oxford, as well as from Funds for Women Graduates (FfWG). Thanks also to the NERC Innovation Placement with AWE Blacknest which initiated a collaboration that formed a basis for some of the presented work.

Abstract

Earth structure is multi-scale, yet capturing such a broad range of complex heterogeneities with seismic wave propagation is computationally prohibitive. In this thesis we present an Instaseis-based (van Driel et al., 2015) framework for injection type hybrid methods that addresses this challenge and provides 5,000-100,000 speed-up relative to full 3-D global solvers without compromising on the full physics. The modified Instaseis interface allows to couple the global wave-propagation solver, AxiSEM (Nissen-Meyer et al., 2014), with an arbitrary local 3-D solver of choice, and thus embed any heterogeneous 3-D domain within a spherically symmetric Earth model. The global 1-D medium accounts for source and paths effects, while the local 3-D domain optimally honours complexities in the model to resolve local scattering effects where necessary. The method can reach the highest frequencies observed teleseismically (1-4 Hz) and opens global wave propagation to a new parameter space.

The Instaseis-based method is applied to study the 1960s-1980s nuclear explosions from the USSR test site in the Degelen mountains. Observed waveforms up to 2 Hz bear a strong signature of near-source effects, yet the relative importance of contributing factors – such as explosion depth and yield, scattering off topography and geological heterogeneities, as well as non-linear effects – is not well understood. An analysis of the dataset suggests that waveform features are dependent on the source location in the Degelen mountain range, though simple measures do not indicate a straightforward relationship with topography. We therefore turn to deterministic hybrid modelling to evaluate waveform variations stemming from topographic interactions in the vicinity of the source. Amplitudes are in good agreement and many observed features are reproduced by synthetic seismograms at 2 Hz, showing that topography significantly contributes to waveform complexity at teleseismic distances.

Extended Abstract

Numerical wave propagation is one of the primary means of studying wave physics and complex scattering effects, helping us to interpret teleseismic recordings and to infer structural and dynamic properties of the Earth. High-resolution synthetic wavefields can capture the effects of multi-scale three-dimensional structures, yet they are computationally prohibitive and hence pose a fundamental challenge. Full 3-D solvers that can represent any geometry scale with frequency to the fourth power (three in space, one in time) and thus incur large computational costs for high-frequency simulations. We therefore often rely on methods that include the full physics of wave propagation but reduce media complexity by assuming 1-D spherically symmetric models. Such structural simplification facilitates wave propagation for realistic Earth models and provides a compromise between broad frequency ranges and computational costs, but cannot capture the complexity of 3-D scattering.

Hybrid methods that couple two solvers can to some extent bridge that gap between high structural complexity and computational cost, though they have not been routinely used in global scale applications. Such hybrid techniques represent the bulk 'background' medium as a sparse or smooth structure that accounts for source and path effects, and optimally honour complexities in the model to resolve local 3-D scattering in areas of interest. The methodology put forward in this thesis offers a global scale hybrid coupling that includes the full physics, targets the highest observed frequencies of body waves (in the 1-4 Hz range) and caters for both wavefield injection and extrapolation. With a framework built around 1-D spherically symmetric Instaseis databases (van Driel et al., 2015) the method provides 5,000-100,000 speed-up relative to full 3-D global solvers and boasts an unprecedented level of adaptability. The databases act as a once-and-for-all solution to global wave propagation and can therefore be reused for a broad range of applications. Each database can be coupled with different local solvers, and offers flexibility with respect to parameter alterations, such as modifications in source properties (radiation patterns, source-time function), in the source-receiver geometry, and in local domain dimensions and location (including regions around the source or receivers, and at depth along the propagation path).

The method is applied to a case study of the teleseismic signature of near-source scattering in the Degelen nuclear explosions test site where the USSR conducted 340 underground detonations between the early 1960s and the late 1980s. The large and unique dataset provides an ideal starting point for a systematic analysis that can be related to synthetic data – the source locations, site topography and site geology are well known, and waveforms have been extensively processed. Such prior knowledge reduces the level of uncertainty in the subsequent numerical set-up and allows for more straightforward comparisons between the observed data and synthetics.

The first part of the study focuses on a broad analysis of the digitised waveforms that bear a strong signature of near-source effects. Despite many shared features – associated with the stations and the propagation along the path that are common between all events – there are significant differences in the P wave and its coda within the Degelen dataset. Given that the sources are at most 15 km apart, the observed variations are attributed primarily to the source region. The degree of variability changes between the arrays and clearly depends on the take-off angle and azimuth, suggesting significant directional properties of the generation of the P wave and of the scattered phases.

The relative importance of factors that may be contributing to variations in P waves – such as explosion depth and yield, scattering off topography and geological heterogeneities, as well as non-linear effects – is not well understood. An analysis of the dataset suggests that waveform features are dependent on the source location in the Degelen test site, while depths and yields do not show a consistent effect. We therefore propose that the change in signal characteristics on all arrays is related to the mountainous topography in the source region. The hypothesis, however, cannot be validated from data alone, as capturing seismogram complexity and topographic features in meaningful measures is not straightforward. The chosen measures prove largely inconclusive – although they suggest that P wave complexity is related to source location – and in the second part of the study we turn to deterministic hybrid modelling to test the imprint of Degelen topography at teleseismic distances up to 2 Hz.

We generate synthetic seismograms for 24 explosion locations across both Degelen mountain ridges, selected to span various peaks, and to account for different distances between events. We then choose to approximate the 'average' small- to medium-yield Degelen event of 15 kt by an isotropic source with a scalar moment of 10^{16} Nm and a Gaussian moment rate function with a 'dominant period' of 2 Hz. The global elastic

1-D model (IASP91) of the Instaseis database includes attenuation, and NASA's SRTM topography is implemented in the local domain around the test site.

Despite the simplistic assumptions and uncertainties in real source locations, features from observed recordings are present in the synthetics, suggesting a significant effect of the Degelen topography on teleseismic P waves. Many parameters have not been accounted for in the modelling – such as realistic source models, geological heterogeneities, and non-linearity of wave propagation – and yet we found a very good overall qualitative fit of shape pulses, and amplitudes at each station were well reproduced. To the best of the author's knowledge, it is the first time that teleseismic 2 Hz waveforms have been compared to synthetics that account for 3-D structures, and the excellent fit achieved highlights the importance of the imprint of local 3-D effects on long-range wave propagation.

Waveform complexity has not been fully reproduced, however, and further work is needed to refine and improve the synthetics in order to reliably quantify topographic scattering. The simulated seismograms have been generated to contain most of the energy up to 2 Hz with a minimal computational effort ('dominant frequency' of 2 Hz, i.e. corner frequency of 1 Hz), and thus have a lower corner frequency than the data. Exact source locations were not considered, and the same approximate source model was used for all simulations regardless of the yield and depth. Waveforms at higher resolution that incorporate realistic source models are therefore necessary to extend the case study and to carry out a systematic quantitative investigation of the data.

Statement of Authorship

Work in Chapter 3 and Chapter 4 was carried out in collaboration with Stuart Nippress and David Bowers from AWE Blacknest who provided the waveforms dataset from the Degelen test site, as well as their expertise on nuclear explosions monitoring. All data analysis and modelling work has been carried out by the author of this thesis, Marta Pieńkowska-Côte, except Figures 3.24, 3.25 and 3.26 that were provided by Stuart Nippress. All text describing and analysing the figures, including figure captions, was written by Marta Pieńkowska-Côte.

Contents

1 Introduction **1**

1.1 Overview of existing numerical methods 2

1.2 Motivation for the proposed method 4

1.3 Thesis outline 5

2 Hybrid method: global modelling of strong 3D scattering **7**

2.1 Introduction 7

2.2 Theory 9

2.2.1 Wavefield injection 9

2.2.2 Extrapolation 13

2.3 Implementation 15

2.3.1 AxiSEM: 2.5-D global wavefields 15

2.3.2 Instaseis: a database approach 17

2.3.3 Hybrid: a flexible framework 18

2.3.4 Local solvers: SPECFEM3D Cartesian and Wave Propagation Program . 20

2.4 Validation 22

2.4.1 1-D benchmarks: Specfem-Instaseis wavefield reconstruction 23

2.4.2 3-D benchmarks: Specfem-Instaseis local scattering 23

2.4.3 Short period WPP-Instaseis wavefield reconstruction 30

2.5 Performance 32

2.5.1 Database generation and storage 32

2.5.2 Hybrid injection and extrapolation 33

2.6 Technical considerations 37

2.6.1 Limitation: higher order long range scattering 37

2.6.2 Implementation: local-global model mismatch 38

2.6.3 Extensions: 3-D background model, multiple local domains, non-linear coupling 38

2.7 Wave-propagation applications 41

2.8 Summary . . . 42

3 A nuclear explosions case study,
Part I: A manifestation of source complexity in observed waveforms 43
3.1 Introduction . . . 43
3.1.1 Motivation for the nuclear explosions case study . . . 43
3.1.2 The Degelen test site . . . 45
3.2 The AWE Blacknest Data . . . 47
3.3 The complexity of the Degelen waveforms . . . 53
3.3.1 Previous AWE Blacknest data analysis: the P and pP arrivals . . . 53
3.3.2 A closer look at waveform differences . . . 54
3.4 An analysis of the Degelen dataset . . . 56
3.4.1 The general signal characteristics . . . 57
3.4.2 Waveform variations with depth and body-wave magnitude . . . 61
3.4.3 Waveform variations with location in the test site . . . 65
3.4.4 Waveform complexity through a topographic lens . . . 73
3.5 Summary and discussion of results . . . 85

4 A nuclear explosions case study,
Part II: Modelling the effects of topography at teleseismic distances 87
4.1 Introduction . . . 87
4.1.1 Modelling of near-source scattering . . . 87
4.1.2 The Degelen waveforms . . . 90
4.2 Model description and setup . . . 91
4.2.1 Defining the local topography . . . 91
4.2.2 The 2 Hz database and the IASP91 approximation . . . 94
4.2.3 Choosing the moment rate function . . . 97
4.3 Results of simulations . . . 98
4.3.1 Teleseismic 1-D waveforms at 2 Hz . . . 98
4.3.2 Topographic effects of Degelen at teleseismic distances . . . 109
4.3.3 Local source-topography interactions at Degelen . . . 117
4.4 Data v.s. synthetics . . . 120
4.4.1 Amplitudes . . . 120
4.4.2 Waveform features . . . 120
4.5 Summary and discussion of results . . . 129

5 Discussions **133**
5.1 Hybrid framework . . . 133
5.1.1 Overview . . . 133
5.1.2 Method developments . . . 134
5.2 Nuclear explosions case study . . . 135
5.2.1 Overview . . . 135
5.2.2 Further work . . . 136

Appendices **137**

A The modified Instaseis package **138**
A.1 Instaseis: wavefield injection . . . 138
A.1.1 Injection requirements . . . 138
A.1.2 Structure of interfacing HDF5 files . . . 139
A.1.3 Sample code . . . 141
A.2 Instaseis: wavefield extrapolation . . . 143
A.2.1 Extrapolation requirements . . . 143
A.2.2 Structure of interfacing HDF5 files . . . 143
A.2.3 Sample code . . . 145

B WPP-Instaseis: the spherical Earth **147**

C The Degelen test site data **149**

D Cross-correlation, envelope and phase misfit coefficients of the Degelen seismograms **152**

E The resolution limit of the mesh: choosing a moment rate function **158**

F Degelen simulations: 1-D benchmarks **160**

G Overview maps of synthetic results **167**

H Local source-topography interactions: location B3 **171**

I Data v.s. synthetics **174**

List of Figures

2.1 Three hybrid setups based on wavefield injection and extrapolation. 8
2.2 Definition of domains involved in wavefield injection. 10
2.3 Definition of domains involved in wavefield extrapolation. 14
2.4 Sketch of an Earth model with axisymmetric geometry. 16
2.5 Comparison of computational costs of global wavefield simulations. 17
2.6 1-D benchmarks of the hybrid Specfem-Instaseis coupling. 24
2.6 (continued). 25
2.7 Sketch of benchmark set-ups of the Specfem-Instaseis hybrid coupling. 27
2.8 3-D benchmarks of the hybrid Specfem-Instaseis coupling. 28
2.8 (continued). 29
2.9 1-D benchmark of hybrid WPP-Instaseis coupling around the source region. . . . 31
2.10 Comparison of storage requirements of different reciprocal Instaseis databases. . . 33
2.11 Sketch of missing interactions in a hybrid one-way injection coupling. 37
2.12 Illustration of topography mismatch on a boundary of WPP-Instaseis coupling. . 39
2.12 (continued). 40

3.1 Location of the Degelen mountain test site. 45
3.2 Topography of the Degelen mountain complex. 46
3.3 Locations of the UK-designed arrays and the main nuclear test sites used for underground nuclear tests by the USA, USSR, UK, France and China. 48
3.4 Layout of four UK-designed arrays used in this study. 49
3.5 Instrument response of a UK-type array short-period seismometer to an impulse of ground displacement 50
3.6 Map of the Degelen Mountain tunnel complex with adit portals and ends. 51
3.7 Example of observed waveform differences for two Degelen events. 54
3.8 Examples of observed waveforms at the Konystan test site with a flat surface. . . 55
3.9 Example of noisy signals in the Degelen dataset. 58
3.10 Illustration of imperfect threshold-based rejection of low-magnitude events. . . . 59

3.11 Distributions of cross-correlation, phase misfit, and normalised envelope misfit coefficients in the Degelen dataset at EKA. 60
3.12 Distributions of CC, PM, and normalised EM coefficients at EKA for events with depths between 180 m and 200 m. 61
3.13 Example of observed waveform differences for two Degelen events at the same depth. 62
3.14 Example of similar waveforms for different depths of firing. 63
3.15 Distributions of cross-correlation, phase misfit, and normalised envelope misfit coefficients at EKA for events with body-wave magnitudes between 5.3 and 5.5. . 64
3.16 Example of similar waveforms for different body-wave magnitudes. 64
3.17 Illustration of the failure of simple threshold-based clustering of close lying events. 66
3.18 Location-dependent seismogram features at the GBA array. 69
3.19 Location-dependent seismogram features at the EKA array. 70
3.20 Location-dependent seismogram features at the YKA array. 71
3.21 Location-dependent seismogram features at the WRA array. 72
3.22 Change in perception of seismograms due to normalisation. 73
3.23 Illustration of RMS and envelope measures of complexity. 75
3.24 Values of envelope and RMS complexity as a function of event location. 76
3.24 (continued). 77
3.25 Values of envelope and RMS complexity as a function of topographic skewness. . 79
3.25 (continued). 80
3.26 Values of envelope and RMS complexity as a function of the difference of the station magnitude and the network magnitude. 82
3.26 (continued). 83

4.1 Topography downloaded from the NASA's Shuttle Radar Topography Mission. . 92
4.2 Topography after removing underlying elevation of the Degelen mountain complex. 93
4.3 Topography after removing underlying elevation and topographic trend of the Degelen mountain complex. 93
4.4 Sketch of proportions of the local Degelen domain with respect to global IASP91 discontinuities. 95
4.5 Comparison of database wavefields at 85.3 degrees epicentral distance. 96
4.6 Comparison of frequency characteristics of Gaussian and triangle moment rate functions. 97
4.7 1-D benchmark of the hybrid seismogram at YKA. 99
4.8 Illustration of 1-D seismic phases. 102
4.9 Arrival times of phases present in the first 15 seconds of IASP91 1-D seismograms. 103

4.9 (continued). . . . 104
4.10 Evolution of 1-D waveforms at YKA with increasing source depth. . . . 106
4.11 Peak-to-peak amplitude of the first arrival as a function of source depth. . . . 107
4.12 Illustration of the effect of take-off angle on the path of pP. . . . 108
4.13 Overview of synthetic waveforms from the Degelen mountain at YKA. . . . 110
4.14 Detailed comparison of waveform differences caused by the Degelen topography at the YKA array. . . . 111
4.14 (continued). . . . 112
4.15 Factor 2 difference in peak-to-peak amplitudes of first arrivals can be observed in the synthetics. . . . 114
4.16 Comparison of explosions simulated at different depths. . . . 116
4.17 Displacement cross sections for $y = 19$ km for a source located at C4 at $(16.5, 19.0)$. 118
4.18 Displacement ross sections for $x = 16.5$ km for a source located at C4 at $(16.5, 19.0)$. 119
4.19 Illustration of the effect of normalisation on waveform comparison at YKA. . . . 121
4.20 Map of Degelen topography with synthetic and observed event locations. . . . 122
4.21 Illustration of dependence of waveform similarity on source proximity in observed and synthetic datasets at YKA. . . . 123
4.22 Small relative shift of coda with respect to location is present in the synthetics at GBA. . . . 124
4.23 Comparison of synthetic and observed seismograms at GBA for explosions in close lateral proximity. . . . 125
4.24 Comparison of synthetic and observed seismograms at EKA for explosions in close lateral proximity. . . . 126
4.25 Comparison of synthetic and observed seismograms at YKA for explosions in close lateral proximity. . . . 127
4.26 Comparison of synthetic and observed seismograms at WRA for explosions in close lateral proximity. . . . 128
4.27 Fitting waveform features could be used to constrain explosion locations. . . . 132

B.1 Sketch of the effect of imposing the Earth's curvature as topography in WPP-Instaseis coupling. . . . 148

D.1 Distributions of CC, PM, and EM coefficients in the Degelen dataset at GBA. . . . 153
D.2 Distributions of CC, PM, and EM coefficients in the Degelen dataset at YKA. . . . 154
D.3 Distributions of CC, PM, and EM coefficients in the Degelen dataset at WRA. . . . 155
D.4 Distributions of CC, PM, and EM coefficients at GBA, YKA and WRA for events with depths between 180 and 200 metres. . . . 156

D.5 Distributions of CC, PM, and EM coefficients at GBA, YKA and WRA for events with body-wave magnitudes between 5.3 and 5.5. 157

E.1 Verification of effects of numerical noise for Gaussian functions above the resolution of the global AxiSEM mesh. 159

F.1 1-D benchmark of the hybrid seismogram for the $30 \times 30 \times 10$ km WPP domain at GBA, EKA and WRA. 161
F.1 (continued). 162
F.2 1-D benchmark of the hybrid seismogram for the $10 \times 10 \times 10$ km WPP domain at GBA, EKA and WRA. 163
F.2 (continued). 164
F.3 Example of the effect of grid refinement on the quality of 1-D benchmarks. 165
F.3 (continued). 166

G.1 Overview of synthetic waveforms from the Degelen mountain at GBA. 168
G.2 Overview of synthetic waveforms from the Degelen mountain at EKA. 169
G.3 Overview of synthetic waveforms from the Degelen mountain at WRA. 170

H.1 Displacement cross sections for $y = 13.5$ km for a source located at B3 at $(13.6, 13.5)$. 172
H.2 Displacement cross sections for $x = 13.6$ km for a source located at B3 at $(13.6, 13.5)$. 173

I.1 Illustration of dependence of waveform similarity on source proximity in observed and synthetic datasets at GBA. 175
I.2 Illustration of dependence of waveform similarity on source proximity in observed and synthetic datasets at WRA. 176

List of Tables

2.1 Computational and storage requirements of Instaseis databases used for benchmarking. 33
2.2 Summary of total hybrid costs of benchmark runs. 36

3.1 Locations of four arrays used in this study and their distance and azimuth to the Degelen test site. 47
3.2 Summary of the Degelen dataset available at AWE Blacknest. 52
3.3 Statistics for the least squares analysis of complexity. 84
3.4 Station terms s from the least squares analysis of complexity. 84

4.1 Values of t^* resulting from PREM Q values used in the global simulation and the amplitude correction factor. 94
4.2 Correction of synthetic peak-to-peak amplitudes for t^* representing more realistic attenuation for high frequencies. 120

Chapter 1

Introduction

The past few decades have seen a significant improvement in our capacity to extract valuable information from global seismic datasets and to image the Earth's interior. Rapid hardware and software developments, along with increasing and more accessible high-quality broadband data, have driven progress in our attempts to infer the Earth's structure and dynamics (Becker and Boschi, 2002; Dziewonski and Romanowicz, 2007). Such advances have taken global seismology from the widely accepted one-dimensional (1-D) Earth models – such as PREM (Dziewonski and Anderson, 1981), IASP91 (Kennett and Engdahl, 1991) and AK135 (Kennett et al., 1995) – to three-dimensional (3-D) models that incorporate lateral heterogenieties at various scales (for global and regional examples see the Earth Model Collaboration (EMC) database on IRIS at ds.iris.edu/ds/producs/emc-earthmodels). While the discontinuities of spherically symmetric 1-D models account for a large portion of the observed seismic records, the remaining unexplained differences carry the signature both of the finite extent of the source and of the complex structures in the Earth's interior and on its surface. Such 3-D wavefield effects, therefore, conceal information about the source itself and about the Earth's thermal, chemical and topographic structures, and studying their influences on global seismic recordings is crucial to broadening our understanding of seismic sources and of the Earth's dynamic processes.

Numerical wave propagation that generates synthetic waveforms is the primary means of studying wave physics and the complex scattering effects. However, even though we theoretically understand wave propagation and its numerical solution, it remains computationally prohibitive to generate high-frequency synthetic waves in a realistic, multi-scale, heterogeneous Earth – a challenge especially relevant when considering global scale problems. The bandwidth of teleseismic recordings spans about four orders of magnitude, from the longest oscillations of a few hours (Dahlen and Tromp, 1999) up to around 8 Hz for teleseismic P-waves (Sharrock et al., 1995), with the high-frequency content limited by attenuation. Accounting for such a broad range of frequencies numerically and fully capturing the scattering effects remains one of the major challenges of computational seismology today.

Theoretically, with full 3-D numerical solvers at hand, such as finite difference (Igel et al., 2002; Moczo et al., 2014), discontinuous Galerkin (Dumbser and Käser, 2006) and spectral element (Komatitsch and Tromp, 2002b; Komatitsch and Tromp, 2002a; Chaljub et al., 2007) methods, we can tackle media of various geometries and desired complexities. The computational cost of those methods, however, scales with frequency to the 4^{th} power (three dimensions in space, one in time), decidedly restricting their applicability to global scale problems for the observable frequency band. At present, frequencies around 1 Hz can be considered the practical limit for a few full 3-D global simulations on large clusters (Leng et al., 2016), while iterative inverse problems or systematic studies of synthetic waveforms require large numbers of such simulations.

In what follows, in Section 1.1 we discuss the existing approaches to computing teleseismic high-frequency body waves. In Section 1.2 we take a closer look at their shortcomings that motivate the new hybrid method delivered in this thesis, and in Section 1.3 we outline the remainder of the presented work that includes the hybrid framework itself and applications to nuclear explosions datasets.

1.1 Overview of existing numerical methods

Due to high computational costs of full 3-D numerical solvers, to generate high-frequency body waves we often rely on methods that include the full physics of wave propagation but reduce media complexity by assuming 1-D spherically symmetric models, such as AxiSEM (Nissen-Meyer et al., 2014), DSM (Kawai et al., 2006), GEMINI (Friederich and Dalkolmo, 1995), or Yspec (Al-Attar and Woodhouse, 2008). Such structural simplification facilitates wave propagation for realistic Earth models and provides a compromise between broad frequency ranges and computational costs, allowing for direct comparisons of synthetics with observed waveform data (Nissen-Meyer et al., 2014). Leng et al. (2019), however, show in a synthetic study that 3-D scatterers of a couple of wavelengths in size have a recognisable impact on high-frequency global wave propagation, an effect that is currently largely ignored. Along with insufficient data coverage, such structural modelling assumptions resulted in global tomographic models that are in reasonable agreement for large-scale structures (Becker and Boschi, 2002; Auer et al., 2014), but diverge at smaller length scales. Improving seismic Earth models and our understanding of waveforms must therefore incorporate numerical methods that not only capture the relevant physics and provide a high-frequency resolution, but also account in a computationally feasible manner for the complexity that affects global wave propagation in 3-D.

More recently, AxiSEM3D (Leng et al., 2016) has drastically improved our capability to generate high-frequency synthetic wavefields on a global scale. The method requires about 100,000 CPU hours (CPUh) for 1 Hz wavefields and thus is about 1,000 times faster than a community-

standard full 3-D solver Specfem3D Globe (Komatitsch and Tromp, 2002b). AxiSEM3D exploits the smoothness of global wavefields, and thus its computational efficiency is model dependent – it is cheapest for 1-D models, and increasing model complexity increases numerical costs. Even though the method can be applied to highly heterogeneous structures, with a tenfold increase in computational costs for 2 Hz wavefields (about 1,000,000 CPUh) relative to 1 Hz it remains prohibitive for routine use at resolutions above 1 Hz.

Hybrid methods that couple two solvers to some extent bridge that gap between high structural complexity and computational cost. Such techniques represent the bulk 'background' medium as a sparse or smooth structure that accounts for source and path effects, and optimally honour complexities in the model to resolve local 3-D scattering in areas of interest. The wavefields are coupled on the boundary between the subdomains, effectively embedding smaller-scale structures in the larger model. Each method provides accurate solutions in its domain of validity, while the reduced complexity (e.g. 1-D model) or approximate physics (e.g. ray theory or the frequency-wavenumber (FK) method) in the background medium decrease the overall computational cost of the problem.

A large proportion of coupling methods proposed to date in seismic wave propagation, often referred to as wavefield injection or domain reduction methods, stem from the work by Alterman and Karal (1968) addressing the source singularity problem in finite difference schemes. Such methods are effectively a two-step procedure with no dynamic exchange of information between the solvers. First, the wavefield is computed in the smooth background medium and recorded on the boundaries of a subvolume of interest (i.e. where we wish to incorporate heterogeneities). This background field is then imposed on the boundaries of the region to drive the local simulation. The method accounts for all interactions in the local subvolume, but neglects the second order long-range effects, as the scattered wavefield cannot re-enter the subvolume following interactions with the background structures. Such injection-type hybrid methods – both for regional and global applications – have been evolving over the years as numerical solvers themselves evolved, striving to efficiently obtain ever better wavefield estimates for both the incident background field and the local scattered field (e.g., Bielak and Christiano, 1984; Moczo et al., 1997; Wen and Helmberger, 1998; Robertsson and Chapman, 2000; Opršal and Zahradnik, 2002; Bielak, 2003; Chevrot et al., 2004; Zhao et al., 2008; Opršal et al., 2009; Roecker et al., 2010; Monteiller et al., 2013; Masson et al., 2014; Tong et al., 2014). For local domains away from receiver arrays, injection can be combined with wavefield extrapolation to re-propagate the scattered wavefields to the desired receiver locations (Robertsson and Chapman, 2000; Masson et al., 2014; Masson and Romanowicz, 2017b; Masson and Romanowicz, 2017a; Clouzet et al., 2018; Wu et al., 2018).

Other two-way coupling methods have also been proposed to account for higher order long range scattering between the two domains (Capdeville et al., 2003b; Capdeville et al., 2003a;

van Manen et al., 2007; Galis et al., 2008; Willemsen et al., 2016; Broggini et al., 2017; Masson and Romanowicz, 2017a), but dynamic boundary conditions on the coupling interface between the subdomains require a significant computational effort and prove to be expensive for applications requiring large numbers of high-frequency simulations. Masson and Romanowicz (2017a), however, examined the higher order scattering that is neglected by the one-way coupling and concluded that it does not significantly influence global waveforms, making the injection-type approximation to scattering a computationally efficient choice for numerous applications.

Offering advantages not provided by either of the methods on their own, hybrid methods have recently been used to study the effects of Moho and surface topography on teleseismic wavefields (Monteiller et al., 2013), to carry out parametric studies of inversion methods (Yuan et al., 2017; Beller et al., 2018), and to locally improve the resolution of tomographic models (Masson and Romanowicz, 2017b; Clouzet et al., 2018). Although a coupling of two solvers does not provide the ultimate solution to high-frequency simulations on global scale, it is well-suited for a wide range of applications and can bring us closer to understanding the multi-scale Earth structure and the imprints of small heterogeneities – around sources, receivers, or at depth – on the global wavefield.

1.2 Motivation for the proposed method

Among the one-way hybrid methodologies that provide a significant computational advantage relative to full 3-D methods, a vast majority offers to couple the local domain directly under the receiver arrays on the surface. Only a few of such injection methods are suited for full 3-D global rather than regional applications, with Monteiller et al. (2013) coupling 1-D AxiSEM with Specfem3D Cartesian, Tong et al. (2014) coupling 1-D frequency-wavenumber (FK) wavefields with Specfem3D Cartesian, and Masson and Romanowicz (2017b) and Masson and Romanowicz (2017a) coupling Specfem3D Globe with regSEM (Cupillard et al., 2012). To the best of the author's knowledge, only two global-scale methods include domains away from the receivers – around the source (Masson and Romanowicz, 2017a; Wu et al., 2018), or at depth (Masson and Romanowicz, 2017b; Masson and Romanowicz, 2017a; Clouzet et al., 2018) – that extrapolate local wavefields to stations at distance. All of the above methods are specific to a selected local solver, and only Masson and Romanowicz (2017b) and Masson and Romanowicz (2017a) offer both the wavefield injection and extrapolation set-up that allows buried local subdomains away from the free surface. Masson and Romanowicz (2017b) and Masson and Romanowicz (2017a), however, focus on fully 3-D structures in both the background and the local domains, and the high computational requirements of the global background 3-D model restrict the applicability of their method for the uppermost frequencies observed teleseismically.

The methodology put forward in this thesis fills the gap for a global scale hybrid coupling that

includes the full physics, targets the highest observed frequencies and caters for both wavefield injection and extrapolation. It provides an unprecedented level of flexibility, with a framework built around 1-D spherically symmetric Earth models that allows for coupling with an arbitrary local 3-D solver of choice. The framework employs Instaseis (van Driel et al., 2015), a tool that allows a near-instantaneous extraction of global Green's functions from pre-computed AxiSEM (Nissen-Meyer et al., 2014) databases. Such databases act as a once-and-for-all solution to wave propagation in spherically symmetric background models, leaving the flexibility with respect to parameter alterations, such as modifications in source properties (radiation patterns, source-time function), in the source-receiver geometry, and in local domain dimensions and location (including regions around the source or receivers, and at depth) to subsequent hybrid applications without the need of re-running the global simulation. Thanks to reasonable computational costs (about 10,000 CPU hours for a re-usable database) and storage requirements (a few TB for global 1 Hz waveforms) of Instaseis databases, the hybrid method can reach the highest observable body-wave frequencies in global seismology and opens numerical simulations to new parameter regimes.

1.3 Thesis outline

Chapter 2 provides the theoretical basis behind the hybrid method, including an overview of both wavefield injection and extrapolation, and discusses the implementation of the framework based on Instaseis (van Driel et al., 2015) databases and global AxiSEM (Nissen-Meyer et al., 2014) wavefields. The coupling of two very different local solvers – Specfem3D Cartesian (Komatitsch and Tromp, 2002b) and WPP (Nilsson et al., 2007; Petersson and Sjogreen, 2013) – is validated, and its limitations, performance and potential applications are considered.

Chapters 3 and 4 present a nuclear explosions case study, comprised of two parts, where for the first time teleseismic P wave recordings at 2 Hz are compared to synthetic waveforms with near-soure 3-D structural signature. Part I in Chapter 3 focuses on an analysis of a well constrained explosions dataset, with waveform features showing dependence on the location of the source within the test site. Part II in Chapter 4 isolates the mountainous topography as one of the potential parameters affecting the location-dependence of the waveforms, and attempts to evaluate through hybrid modelling the waveform variations stemming from such topographic interactions in the vicinity of the source. P wave amplitudes and many observed features are reproduced by the synthetics at 2 Hz, showing that topography significantly contributes to waveform complexity at teleseismic distances and can perturb or mask the signature of the source.

Finally, Chapter 5 reviews the hybrid method proposed in this work, summarises its benefits, and discusses future developments. It also concludes the nuclear explosions case study and

explores possible directions for further research within the context of near-source topographic scattering.

Chapter 2

Hybrid method: global modelling of strong 3D scattering

2.1 Introduction

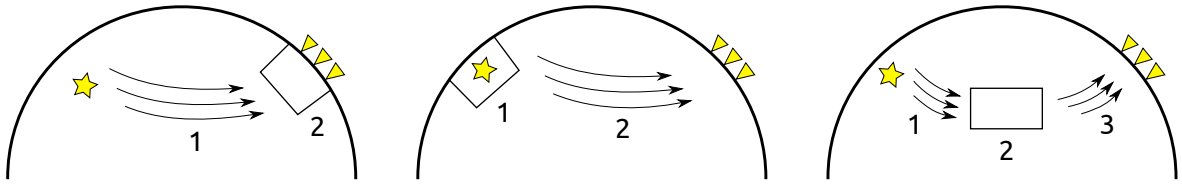
Hybrid methods, as understood in this thesis, decompose the computational domain into subdomains – often referred to as the background and local domain – in order to exploit advantages of individual solvers, increase the overall algorithmic efficiency and decrease time to solution. As discussed in Section 1.1 in Chapter 1, a variety of such methods have been proposed in seismology. Due to high computational costs of dynamically linking two solutions in time, the majority have been implemented as one-way couplings, and the methodology proposed in this thesis also falls into that category.

The uniqueness of the approach outlined in this chapter lies within the flexibility of Instaseis (van Driel et al., 2015) databases of Green's functions generated by AxiSEM (Nissen-Meyer et al., 2014) that provide a global framework for the coupling:

- The databases act as a once-and-for-all solution to wave propagation in spherically symmetric background models. A single global computation for a chosen model allows for subsequent parameter alterations, such as modifications in source properties (radiation patterns, source-time function), in the source-receiver geometry, and in local domain dimensions and location (including regions around the source or receivers, and at depth). Thus, changing problem parameters for most cases requires only accessing the databases anew and re-running only the local simulations.

- The databases cater for local domains located both on the Earth's surface – around the source region and under the receivers – as well as at depth.

- The framework is centred around the global solver and leaves the flexibility to integrate



(a) (b) (c)

Figure 2.1: Wavefield injection and wavefield extrapolation methods that couple solutions in two domains provide a framework for three set-ups: (a) With the local domain around the receivers via wavefield injection. The wavefield is propagated through the background medium (step 1) and then injected into the local medium (step 2). (b) With the local domain around the source region via wavefield extrapolation. The source is simulated in the local domain (step 1) and the wavefield is then extrapolated to the receivers at distance (step 2). (c) With the local domain at depth, via a combination of injection and extrapolation. The wavefield is propagated into the local domain (step 1), interactions at depth are accounted for locally (step 2), and finally repropagated to the receivers at distance (step 3).

any local solver of choice, provided that the necessary modifications are implemented in the local code.

The principal aim of the framework is to efficiently generate high-frequency teleseismic body waves. Given our approximation of the Earth as a spherically symmetric layered medium with a flat free surface, the method is not well suited for surface-wave applications that require a realistic 3-D representation of the crust.

In what follows, in Section 2.2 we first review the principles of wavefield injection and of wavefield extrapolation (via the Helmholz-Kirchhoff integral in the representation theorem) which form the basis of the one-way methods and connect the solution in two subdomains. We then look in Section 2.3 at the specificities of the implementation of the coupling with global AxiSEM wavefields and Instaseis databases, and in Section 2.4 we validate the framework with two very different local solvers. In Section 2.5 we indicate the subtleties of assessing the performance of hybrid methods and discuss the key parameters defining efficiency. We also compare the hybrid runs to full global runs, with the coupled framework performing up to 100,000 times faster than 3-D global solvers. In Section 2.6 we discuss the method with its limitations, further developments and extensions, and in Section 2.7 we consider possible applications. Finally, in Section 2.8 we summarise the Instaseis-based hybrid approach.

2.2 Theory

Wavefield injection caters for cases when the source is in the background domain that carries the bulk of the computation, and the wavefield needs to be imposed in the local domain that is embedded therein. The Helmholz-Kirchhoff integrals, on the other hand, allow to extrapolate the wavefield from the local domain around the source region to receivers at distance in the background medium. Together the two provide a framework for three distinct set-ups: with the local domain around the receiver (injection, Figure 2.1a), around the source (extrapolation, Figure 2.1b), and at depth (both injection and extrapolation, Figure 2.1c). Note that the concepts in this section are presented in their most general form and have to be appropriately tailored for implementation with a local solver of choice – be it a spectral element solver, a finite difference solver, or other.

2.2.1 Wavefield injection

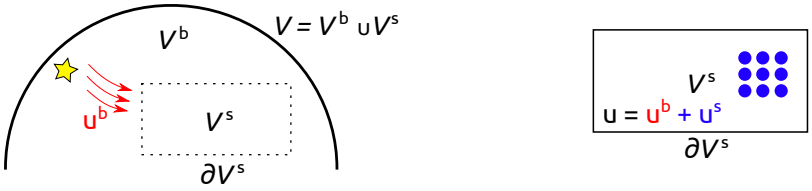
Wavefield injection introduces an incident wavefield generated by a source in a background medium into a local domain embedded in that medium. The source and the majority of the path effects are accounted for in the background domain and are carried over to a regional solver via the local boundary: either via imposing appropriate boundary conditions that drive the local simulation (Robertsson and Chapman, 2000) or by introducing a distribution of body forces on the local boundary (Masson et al., 2014). The method therefore consists of two steps. First, the interactions in the background domain are simulated and recorded on a fictitious boundary of an area of interest. Next, that area is represented in a local solver, with the boundary introducing the background wavefield.

To define the two domains rigorously, let V be a linearly elastic background model subdivided by a fictitious interface ∂V^s such that V^s is the region of the local truncated volume embedded in V, and V^b the part of the domain including the seismic source, i.e. $V = V^b \cup V^s$ (Figure 2.2). The domain V^s, along with its bounding interface ∂V^s, has two representations, one in each solver:

- In the global solver, $V = V^b \cup V^s$ is represented, and ∂V^s is a *fictitious*, non-physical boundary that surrounds the region of interest V^s within the global domain V. ∂V^s serves only as a coupling interface.

- In the local solver only V^s is represented, while ∂V^s is an *artificial*, non-physical boundary that serves as a coupling interface *and* requires absorbing boundary conditions.

In the first step, the background wavefield $\mathbf{u}^b(\mathbf{x}, t)$ is recorded on ∂V^s within the global domain V (Figure 2.2a). In the second step, the volume V^s is represented in a different solver and the wavefield $\mathbf{u}^b(\mathbf{x}, t)$ is imposed on the boundary ∂V^s (Figure 2.2b). If the model in the



(a) step 1 of wavefield injection (b) step 2 of wavefield injection

Figure 2.2: Definition of domains involved in wavefield injection. Background volume $V = V^b \cup V^s$ is subdivided by a fictitious interface ∂V^s. Subdomain V^b includes the seismic source, while subdomain V^s defines the truncated volume later represented in a local solver. (a) The background wavefield $\mathbf{u}^b(\mathbf{x}, t)$ is recorded on the fictitious boundary ∂V^s (step 1). (b) The volume V^s is represented in a local solver and the wavefield $\mathbf{u}^b(\mathbf{x}, t)$ is imposed on the boundary (step 2). Modifying the structure in V^s in the local solver, hybrid coupling accounts for local interactions and the wavefield $\mathbf{u}(\mathbf{x}, t)$ in V^s is a linear combination of the background wavefield $\mathbf{u}^b(\mathbf{x}, t)$ and the scattered wavefield $\mathbf{u}^s(\mathbf{x}, t)$. Note that the structure of the coupling interface ∂V^s itself must remain unchanged relative to its representation in the background volume V. With V^s the same as its representation in the background volume V there is no residual scattered wavefield relative to the background model and $\mathbf{u}^s(\mathbf{x}, t) = 0$, i.e. hybrid coupling reproduces the original wavefield $\mathbf{u}^b(\mathbf{x}, t)$.

local domain remains identical to the representation of V^s in the background model, there is no resudual scattered wavefield relative to the background model and the coupling should reproduce the original background wavefields in V^s, i.e. $\mathbf{u}^s(\mathbf{x}, t) = 0$ and $\mathbf{u}(\mathbf{x}, t) = \mathbf{u}^b(\mathbf{x}, t)$. When the model in V^s is modified to account for complex local structures, the total wavefield $\mathbf{u}(\mathbf{x}, t)$ in the local solver becomes $\mathbf{u}(\mathbf{x}, t) = \mathbf{u}^b(\mathbf{x}, t) + \mathbf{u}^s(\mathbf{x}, t)$, where $\mathbf{u}^s(\mathbf{x}, t) \neq 0$ represents local scattering off heterogeneities. In what follows, we discuss two ways of imposing the background wavefield on the boundary of the local domain to drive the local simulation.

Boundary condition

The background wavefield $\mathbf{u}^b(\mathbf{x}, t)$ can be imposed via a Dirichlet condition on the boundary of the local domain. Let's first consider the total displacement wavefield $\mathbf{u}(\mathbf{x}, t)$ (Figure 2.2b) that

is a solution to the elastodynamic equations of motion

$$\begin{cases} \rho\ddot{u}_n - \partial_j \sigma_{nj} = f_n & \text{in } V^s \cup V^b \\ \sigma_{nj} = C_{njkl}\partial_k u_l & \text{in } V^b \\ \sigma_{nj} = \mathcal{H}[\mathbf{u}(\mathbf{x},t)] & \text{in } V^s \\ \sigma_{nj} n_n = 0 & \text{on } \partial V \\ u_n^+ = u_n^- & \text{on } \partial V^s \\ \sigma_{nj}^+ n_n = \sigma_{nj}^- n_n & \text{on } \partial V^s \end{cases} \tag{2.1}$$

where $\mathcal{H}[\mathbf{u}(\mathbf{x},t)]$ represents a general constitutive relation in V^s, and where we assume ∂V to be the Earth's free surface, as we are primarily concerned with global scale applications. The plus and minus superscripts denote limits from V^b and V^s, respectively, to warrant continuity of displacement and traction across any internal boundaries. The wavefields and constitutive relations are required to be continuous on the boundary ∂V^s. While the background model is elastic in V^b, the general relation $\mathcal{H}[\mathbf{u}(\mathbf{x},t)]$ in V^s does not require an elastic behaviour in V^s. Hence V^s can encompass, for example, non-linear interactions or kinematic ruptures provided that ∂V^s remains elastic. Given the linearly elastic model in V^b and on ∂V^s, we can define hybrid variables $\mathbf{u}^h(\mathbf{x},t), \boldsymbol{\sigma}^h(\mathbf{x},t)$ such that

$$\mathbf{u}^h(\mathbf{x},t) = \begin{cases} \mathbf{u}(\mathbf{x},t) \\ \mathbf{u}(\mathbf{x},t) - \mathbf{u}^b(\mathbf{x},t) \end{cases} = \begin{cases} \mathbf{u}(\mathbf{x},t) & \text{in } V^s \\ \mathbf{u}^s(\mathbf{x},t) & \text{in } V^b \end{cases} \tag{2.2}$$

(and similarly for $\boldsymbol{\sigma}^h(\mathbf{x},t)$) where $\mathbf{u}^b(\mathbf{x},t)$ is the background wavefield computed in the first step, $\mathbf{u}^s(\mathbf{x},t)$ is the scattered wavefield due to 3-D model heterogeneities in V^s, and $\mathbf{u}(\mathbf{x},t)$ is the total wavefield that solves equation (2.1). We then want to find the hybrid field $\mathbf{u}^h(\mathbf{x},t)$ such that

$$\begin{cases} \rho\ddot{u}_n^h - \partial_j \sigma_{nj}^h = 0 & \text{in } V^s \cup V^b \\ \sigma_{nj}^h = C_{njkl}\partial_k u_l^h & \text{in } V^b \\ \sigma_{nj}^h = \mathcal{H}[\mathbf{u}^h(\mathbf{x},t)] & \text{in } V^s \\ \sigma_{nj}^h n_n = 0 & \text{on } \partial V \\ u_n^{h^+} = u_n^{h^-} - u_n^b & \text{on } \partial V^s \\ \sigma_{nj}^{h^+} n_n = \sigma_{nj}^{h^-} n_n - \sigma_{nj}^b n_n & \text{on } \partial V^s \end{cases} \tag{2.3}$$

In the new problem, unlike in equation (2.1), the information about the excitation is given by a Dirichlet condition on ∂V^s that imposes a discontinuity in the wavefield rather than by the body force $\mathbf{f}(\mathbf{x},t)$. As previously mentioned, this result could be extended to problems with a non-linear constitutive relation in V^b as long as $\mathbf{u}^h(\mathbf{x},t)$ remains a linear perturbation around $\mathbf{u}^b(\mathbf{x},t)$ on ∂V^s.

Posing this problem as above, we can see that the boundary ∂V^s not only generates the complete field $\mathbf{u}(\mathbf{x},t)$ in V^s, but also the scattered field $\mathbf{u}^s(\mathbf{x},t)$ propagating out of ∂V^s. In practice, however, the second computation happens only in the truncated subdomain V^s with absorbing boundary conditions applied on ∂V^s to the outgoing field $\mathbf{u}^s(\mathbf{x},t)$. The background wavefield $\mathbf{u}^b(\mathbf{x},t)$ can be computed with any numerical or analytic method provided that the wavefield can be interpolated to the grid points of the local solver.

Distribution of point sources

Wavefield injection can also be implemented via a discretisation of the surface integrals in the representation theorem into multiple point sources. Recall that the representation theorem, without assumptions regarding the surface S, reads:

$$\begin{aligned} u_i(\mathbf{x},t) = &\int_{-\infty}^{\infty} d\tau \int_V f_i(\boldsymbol{\xi},\tau) G_{in}(\boldsymbol{\xi}, t-\tau; \mathbf{x}, 0) \mathrm{d}V \\ &+ \int_{-\infty}^{\infty} d\tau \int_S n_j C_{ijkl} \partial_k u_l(\boldsymbol{\xi},\tau) G_{in}(\boldsymbol{\xi}, t-\tau; \mathbf{x}, 0) \mathrm{d}S \\ &- \int_{-\infty}^{\infty} d\tau \int_S n_j C_{ijkl} u_i(\boldsymbol{\xi},\tau) \partial_l G_{kn}(\boldsymbol{\xi}, t-\tau; \mathbf{x}, 0) \mathrm{d}S. \end{aligned} \tag{2.4}$$

In hybrid modelling, the surface S consists of the outer boundary ∂V of the background model and of the boundary of the local region of interest ∂V^s, i.e. $S = \partial V \cup \partial V^s$. With a global background model we let ∂V be the Earth's stress-free surface, so that integrals over the surface S reduce to integrals over ∂V^s:

$$\begin{aligned} u_i(\mathbf{x},t) = &\int_{-\infty}^{\infty} d\tau \int_V f_i(\boldsymbol{\xi},\tau) G_{in}(\boldsymbol{\xi}, t-\tau; \mathbf{x}, 0) \mathrm{d}V \\ &+ \int_{-\infty}^{\infty} d\tau \int_{\partial V^s} n_j C_{ijkl} \partial_k u_l(\boldsymbol{\xi},\tau) G_{in}(\boldsymbol{\xi}, t-\tau; \mathbf{x}, 0) \mathrm{d}(\partial V^s) \\ &- \int_{-\infty}^{\infty} d\tau \int_{\partial V^s} n_j C_{ijkl} u_i(\boldsymbol{\xi},\tau) \partial_l G_{kn}(\boldsymbol{\xi}, t-\tau; \mathbf{x}, 0) \mathrm{d}(\partial V^s). \end{aligned} \tag{2.5}$$

Recall now that we aim to impose the background wavefield on the boundary in order to represent that background wavefield in the subdomain V^s. The wavefield $\mathbf{u}^b(\mathbf{x},t)$ does not have sources in V^s, so $\mathbf{f}^b(\mathbf{x},t) = 0 \; \forall \; \mathbf{x} \in V^s$ and the first term of equation (2.4) vanishes. Thus equation (2.4) reduces to

$$\begin{aligned} u_i^b(\mathbf{x},t) = &\int_{-\infty}^{\infty} d\tau \int_{\partial V^s} n_j C_{ijkl} \partial_k u_l^b(\boldsymbol{\xi},\tau) G_{in}(\boldsymbol{\xi}, t-\tau; \mathbf{x}, 0) \mathrm{d}(\partial V^s) \\ &- \int_{-\infty}^{\infty} d\tau \int_{\partial V^s} n_j C_{ijkl} u_i^b(\boldsymbol{\xi},\tau) \partial_l G_{kn}(\boldsymbol{\xi}, t-\tau; \mathbf{x}, 0) \mathrm{d}(\partial V^s), \end{aligned} \tag{2.6}$$

where $\mathbf{n}$ is the outward pointing normal. This result can be rigorously derived by considering a surface-source type of problem of scattering by an arbitrary inhomogeneous object (Fokkema and van den Berg, 1993). Approximating the surface integrals by a numerical quadrature

$\int_S f(\mathbf{x})\mathrm{dS} \approx \sum_{m=1}^{M} \alpha_m f(\mathbf{x}_m)$ and using the Dirac delta function property $f(x) = \int_V f(x')\delta(x' - x)\mathrm{dV}$ we get

$$\begin{aligned} u_i^b(\mathbf{x},t) \approx & \sum_{m=1}^{M} \int_{-\infty}^{\infty} d\tau \int_{V^s} \alpha_m n_j C_{ijkl} \partial_k u_l^b(\boldsymbol{\xi}_m,\tau) G_{in}(\boldsymbol{\xi}_m, t-\tau; \mathbf{x}, 0)\delta(\boldsymbol{\xi} - \mathbf{x}_m)\mathrm{d}(V^s) \\ & - \sum_{m=1}^{M} \int_{-\infty}^{\infty} d\tau \int_{V^s} \alpha_m n_j C_{ijkl} u_i^b(\boldsymbol{\xi}_m,\tau) G_{kn}(\boldsymbol{\xi}_m, t-\tau; \mathbf{x}, 0)\partial_l \delta(\boldsymbol{\xi} - \mathbf{x}_m)\mathrm{d}(V^s) \end{aligned} \tag{2.7}$$

for $\mathbf{x}_m$ on ∂V^s.

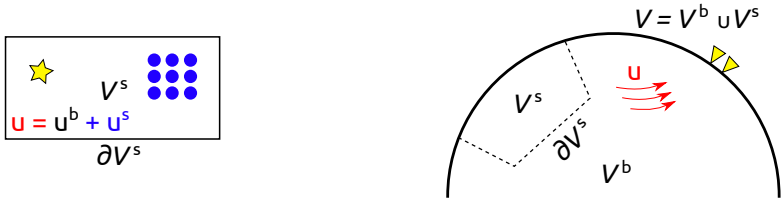
Expression 2.7 has the form of a body-force contribution with a discrete distribution of moment tensor and force sources on ∂V^s, and it completely represents the background wavefield in an *unmodified model* in subdomain V^s. It is similar to the so called "time reversal mirrors" in acoustics (Fink et al., 1989), although with reconstructing the wavefield forward in time rather than refocussing it backward in time. Modifying the structure in V^s and imposing the same moment tensor and force sources on the boundary results in wavefield reconstruction with additional scattering terms superimposed on the background field: $\mathbf{u}(\mathbf{x},t) = \mathbf{u}^b(\mathbf{x},t) + \mathbf{u}^s(\mathbf{x},t)$. As before, absorbing boundary conditions are necessary on the representation of ∂V^s in the local solver to absorb $\mathbf{u}^s(\mathbf{x},t)$ and avoid non-physical reflections off model boundaries.

The equivalence between the distribution of body forces discussed in this section and introducing a discontinuity on the surface ∂V^s in equation 2.3 is discussed in Chapter 3 of Aki and Richards (2002) in the context of representing finite faults.

2.2.2 Extrapolation

Wavefield extrapolation propagates the wavefield from a local domain through an encompassing background domain. The near-source interactions are accounted for in the local solver and carried over to receivers at distance. Like wavefield injection, extrapolation consists of two steps. First, the source region is simulated in the 3-D solver and the fields are recorded on the boundary. Next, that local boundary is fictitiously represented in the background medium where it introduces the local wavefields.

As before, let V be a linearly elastic background model subdivided by a fictitious interface ∂V^s such that V^s is the region of the local truncated volume embedded in V, and $V = V^b \cup V^s$ (Figure 2.3). This time, however, the representation of V^s in the local solver includes the seismic source. In the first step, the local wavefield is recorded on ∂V^s (Figure 2.3a). In the second step, the boundary ∂V^s is fictitiously represented in the background medium and allows for extrapolation to receivers (Figure 2.3b). If the model in the local domain remains unchanged relative to the representation of V^s in the background model, the coupling should reproduce the original background wavefields in V^s, as $\mathbf{u}^s(\mathbf{x},t) = 0$ and $\mathbf{u}(\mathbf{x},t) = \mathbf{u}^b(\mathbf{x},t)$. With the structure in the local representation of V^s modified to include heterogeneities, the total wavefield $\mathbf{u}(\mathbf{x},t)$



(a) step 1 of wavefield extrapolation (b) step 2 of wavefield extrapolation

Figure 2.3: Definition of domains involved in wavefield extrapolation. Background volume $V = V^b \cup V^s$ is subdivided by a fictitious interface ∂V^s. Subdomain V^s defines the truncated volume represented in a local solver where it contains the source. (a) The volume V^s is represented in a local solver that includes the seismic source, and the local wavefields are recorded on ∂V^s. Modifying the structure in the volume V^s in the local solver, hybrid coupling accounts for the local interactions, and the wavefield in V^s is a linear combination of the background wavefield $\mathbf{u}^b(\mathbf{x}, t)$ and the scattered wavefield $\mathbf{u}^s(\mathbf{x}, t)$. With V^s identical to its representation in the background volume V, however, there is no residual scattered wavefield relative to the background model, as $\mathbf{u}^s(\mathbf{x}, t) = 0$ and $\mathbf{u}(\mathbf{x}, t) = \mathbf{u}^b(\mathbf{x}, t)$. Note that the structure on the coupling interface ∂V^s itself must remain unchanged relative to its representation in the background volume V. (b) The fictitious boundary ∂V^s is used to repropagate the wavefield $\mathbf{u}(\mathbf{x}, t)$ from the local solver to receivers at distance (step 2).

in the local solver and at receivers after extrapolation becomes $\mathbf{u}(\mathbf{x}, t) = \mathbf{u}^b(\mathbf{x}, t) + \mathbf{u}^s(\mathbf{x}, t)$, where $\mathbf{u}^s(\mathbf{x}, t)$ is the locally scattered wavefield.

We can repropagate the scattered wavefield out of the local domain V^s to the receivers in the background domain V^b by discretising surface integrals in the representation theorem like in Section 2.2.1, and imposing moment tensor and force sources on the fictitious representation of ∂V^s in the background model. The total wavefield at the receivers at distance is the sum of the background and scattered wavefields, $\mathbf{u}(\mathbf{x}, t) = \mathbf{u}^b(\mathbf{x}, t) + \mathbf{u}^s(\mathbf{x}, t)$. We know the background wavefield $\mathbf{u}^b(\mathbf{x}, t)$ in V^b and on ∂V^s, and hence we wish to represent the scattered wavefield $\mathbf{u}^s(\mathbf{x}, t)$ in V^b in order to define the total displacement $\mathbf{u}(\mathbf{x}, t)$ at the receivers in V^b. $\mathbf{u}^s(\mathbf{x}, t)$ does not have sources in V^b, so $\mathbf{f}^s(\mathbf{x}, t) = 0 \; \forall \; \mathbf{x} \in V^b$ and the first term of equation (2.4) vanishes. As mentioned above, integrals over the surface reduce to integrals over ∂V^s. Repeating the procedure from Section 2.2.1 for the scattered field in V^b, with the numerical quadrature and the delta property of the Dirac delta, we get

$$\begin{aligned} u_n^s(\mathbf{x}, t) \approx & \sum_{m=1}^{M} \int_{-\infty}^{\infty} d\tau \int_{V^b} \alpha_m n_j C_{ijkl} \partial_k u_l^s(\boldsymbol{\xi}_m, \tau) G_{in}(\boldsymbol{\xi}_m, t - \tau; \mathbf{x}, 0) \delta(\boldsymbol{\xi} - \mathbf{x}_m) \mathrm{d}(V^b) \\ & - \sum_{m=1}^{M} \int_{-\infty}^{\infty} d\tau \int_{V^b} \alpha_m n_j C_{ijkl} u_i^s(\boldsymbol{\xi}_m, \tau) G_{kn}(\boldsymbol{\xi}_m, t - \tau; \mathbf{x}, 0) \partial_l \delta(\boldsymbol{\xi} - \mathbf{x}_m) \mathrm{d}(V^b) \end{aligned} \tag{2.8}$$

for $\mathbf{x}_m$ on ∂V^s. Note that for a domain at the surface around the source region, the total wavefield from the local solver can be considered as the locally scattered wavefield $\mathbf{u}^s(\mathbf{x}, t)$, since all path effects are accounted for in the extrapolation step. For the more general case of a local domain along the propagation path, however, the locally scattered wavefield needs to be separated from the background medium in order to account for the propagation effects of the background wavefield outside of the local region.

Expression 2.8 has the form of a body-force contribution with a discrete distribution of moment tensor and force sources on ∂V^s. In other words, we can regenerate the scattered field $\mathbf{u}^s$ at any point $\mathbf{x} \in V^b$ if we know the scattered field on ∂V^s and the Green's functions between $\mathbf{x}$ and ∂V^s.

2.3 Implementation

This section describes details of our implementation of the hybrid coupling which embeds any local 3-D domain in a 1-D spherically symmetric Earth model. We treat the global background medium using AxiSEM (Nissen-Meyer et al., 2014) and Instaseis (van Driel et al., 2015) both to extract the background wavefields necessary for injection in the local domain and to extrapolate the scattered wavefields back to receivers on the Earth's surface. The approach is centred around precomputed databases that provide the general framework of the method and leave flexibility for coupling with an arbitrary local solver of choice.

In what follows, we discuss both AxiSEM and Instaseis, and consider in detail how Instaseis databases allow for coupling with any local 3-D solver in three distinct set-ups presented in Figure 2.1: with the local solver on the surface under the receivers, on the surface around the source, or 'buried' at depth along the propagation path. We then take a closer look at the specificities of two 3-D solvers, SPECFEM3D Cartesian (Komatitsch and Tromp, 2002b) and Wave Propagation Program (WPP) (Nilsson et al., 2007; Petersson and Sjogreen, 2013), that have currently been tied into the framework.

2.3.1 AxiSEM: 2.5-D global wavefields

AxiSEM is a parallel spectral element axisymmetric solver based on the variational form of the wave equation. Simulations are run in two-dimensional (2-D) domains, while the third azimuthal dimension is tackled analytically to generate 3-D wavefields in 1-D Earth models, as shown in Figure 2.4. As described by Nissen-Meyer et al. (2007), the 3-D equations of motion are re-cast into a suite of 2-D problems on a D-shaped slice, which results in a new system of either two or four independent equations for a single force source or a moment tensor source, respectively. Together, the system of two equations represents the three elements of a single force, whereas the system of four represents the six independent elements of a moment tensor. This collapse

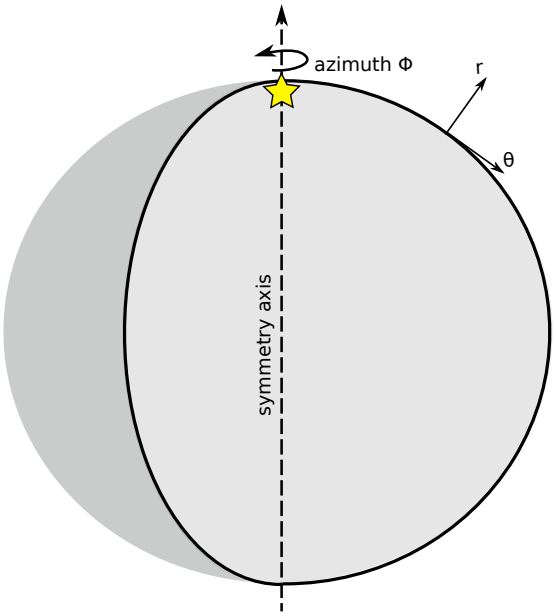


Figure 2.4: Sketch of an Earth model with axisymmetric geometry used in AxiSEM. The sphere is collapsed to a 2-D disk, while the azimuthal dimension is tackled analytically to generate 3-D wavefields in a 1-D Earth. The 3-D equations of motion are re-cast into a suite of 2-D problems with the source on the symmetry axis.

from three- to two- and from six- to four-equation systems honours the azimuthal redundancy that results from the assumed model symmetry. Therefore, two 2-D simulations are required for a full force source, and four 2-D simulations for a full moment tensor source, as a single 2-D AxiSEM simulation is designed only for an individual source element. The separate simulations have to be summed along the azimuth a posteriori to obtain a full radiation pattern.

Such an axisymmetric setting reduces the computational cost by orders of magnitude relative to full 3-D modelling. With AxiSEM simulations scaling with frequency to the third power, it allows for global wave propagation at frequencies above 1 Hz with a reasonable computational effort (see Figure 2.5). Should only source-receiver pairs with distances under 180 degrees be of interest, the numerical cost can be further limited by truncating the 2-D slice to a given distance (which adds an absorbing boundary to the system). For example, limiting the domain to 90 degrees would halve the computational requirements relative to a full Earth simulation. Section 2.5 discusses the overall performance of the hybrid method including further details on the generation of background wavefields in AxiSEM.

Apart from a compromise on model symmetry, the solver does not make any limiting assumptions regarding short-period wave propagation physics (the lack of rotation and gravity is not relevant for periods below 100 s) or on the source radiation pattern. Transverse anisotropy – the most complex type of anisotropy for a spherically symmetric model (van Driel and Nissen-

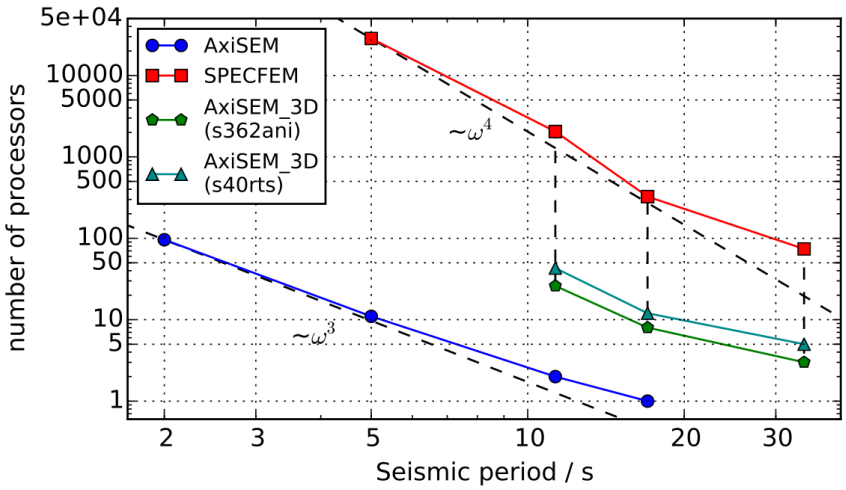


Figure 2.5: Comparison of computational costs of global wavefield simulations in 'real time' (i.e. seismogram length equals wall-clock simulation time) for different methods. The cost of running AxiSEM scales with frequency (inverse of the period) to the third power, and allows for global wave propagation at frequencies above 1 Hz with reasonable computational effort. (Source: Leng et al. (2016))

Meyer, 2014a) – and attenuation (van Driel and Nissen-Meyer, 2014b) can also be accounted for.

2.3.2 Instaseis: a database approach

The AxiSEM-generated Instaseis databases of 2-D Green's functions act as a once-and-for-all solution to wave propagation in spherically symmetric models and store the basis coefficients of Lagrange polynomials, typically of degree 4. As discussed in Section 2.3.1, given the axisymmetric nature of AxiSEM simulations where the source is always rotated to the north pole and the third dimension is calculated analytically, the databases require well-tailored post-processing to result in meaningful 3-D seismograms.

Instaseis – the backbone of our method – is a python package that does exactly that: it efficiently accesses, processes, and extracts wavefields from the pre-computed AxiSEM databases of Green's functions. The tool boasts a wide variety of user-friendly features, and a few should be highlighted with the flexibility of the hybrid coupling in mind. First of all, Instaseis integrates the coordinate systems and rotates the wavefields from a cylindrical system native to the AxiSEM D-slice to the geographic coordinates. Moreover, the spherical symmetry of the model leaves flexibility with respect to the source-receiver geometry that can be extracted from a single database. Given the 4^{th} order spatial accuracy of the Green's functions, the tool can also

interpolate the rotated wavefields to a location independent of the AxiSEM mesh. Changing the source radiation pattern or the source-time function, or defining a finite fault does not require a new database generation, as Instaseis can convolve the source parameters into the Green's functions. Finally, the package provides Lanczos resampling, a high-order temporal interpolation that allows to retrieve the final seismograms at any sampling rate. For more details on the implementation and functionality of Instaseis refer to van Driel (2015).

There are two types of AxiSEM-generated Instaseis databases, and both are crucial for the proposed hybrid framework:

1. The *forward database* is generated by a moment tensor point source at a fixed depth. For such Green's functions receivers exist throughout the medium, whereas the source can be rotated from the north pole to any longitude and latitude. However, it remains at the depth prescribed at the time of the simulation.

2. The *reciprocal database* is generated by a single force point source at a fixed depth and recorded throughout the medium. Due to the reciprocity of Green's functions, the locations of the single force source can be treated as three component receivers. Therefore, in such databases the sources exist throughout the medium, whereas receivers can be rotated to any longitude and latitude, but must remain at the depth prescribed at the time of the simulation. Typically, reciprocal databases are generated by a single force on the Earth's surface which mimics real-world receivers.

Recall that a single 2-D AxiSEM simulation is designed only for an individual source element, with two 2-D simulations required for a full force source and four 2-D simulations for a full moment tensor source. Therefore, the computational effort to generate a reciprocal database is half the one required to generate a forward database. Given that Green's functions are then saved per source element, a similar relationship applies to storage requirements, as reciprocal databases generally take half the space of forward databases with otherwise the same model and simulation parameters.

In the following section we are going to discuss how the two distinct classes of databases serve as a basis for the global hybrid framework. The computational and storage requirements are further discussed in Section 2.4.3 where we look at the overall performance of the method.

2.3.3 Hybrid: a flexible framework

The Instaseis-based approach to global hybrid modelling exploits the reciprocity of Green's functions to accommodate both injection and extrapolation of wavefields discussed in Section 2.2. Each of the database types described in Section 2.3.2 imposes different geometrical source-receiver constraints, but nonetheless both represent the same response of the same medium (for a

given spherically symmetric model and a given dominant frequency). A forward database allows to place receivers anywhere in the domain and requires a fixed source depth, and therefore lends itself to wavefield injection that necessitates the response of the background medium on the boundary of the local domain. A reciprocal database, on the other hand, allows to place sources anywhere in the domain and fixes receivers at the surface, and is appropriate for wavefield extrapolation.

As already discussed, three coupling scenarios are possible, with the local domain at the surface on the receiver side (injection, Figure 2.1a), at the surface on the source side (extrapolation, Figure 2.1b), or buried under the surface along the propagation path (both injection and extrapolation, Figure 2.1c). In the most general case of a complex subdomain at depth, we can think of the proposed global hybrid method as a three-step approach (with the other cases reducing to two steps, steps 2 & 3 and steps 1 & 2 for local domains around the source and receiver, respectively):

1: [**Instaseis**] The forward database is used to extract the background wavefield on an artificial boundary which is defined by the mesh of the local method chosen for the coupling. The incoming wavefield is then imposed on those boundary points of the 3-D solver and drives the local wave propagation. To cater for different local implementations of injection, the modified Instaseis interface provides any combination of displacement, velocity, stress, strain or traction.

2: [**Local solver**] The total wavefield from the simulation is recorded on the bounding box. The locally scattered wavefield is equal to the difference between the total local wavefield and the background wavefield ($\mathbf{u}^s(\mathbf{x}, t) = \mathbf{u}(\mathbf{x}, t) - \mathbf{u}^b$).

3: [**Instaseis**] The locally scattered field is convolved with the reciprocal database of Green's functions and repropagated back to the receivers on the surface. It is then superimposed on the pre-computed background wavefield to render the total global wavefield that incorporates global propagation effects as well as local 3-D scattering.

The availability of both forward and reciprocal databases of Green's functions for a global model, along with the Instaseis features discussed in Section 2.3.2, renders the method particularly attractive. The spherically symmetric set-up caters for a broad range of possible geometries and sources with only two database simulations (i.e. a total of six AxiSEM simulations, two for a reciprocal database and four for a forward database as described in Section 2.3.1) – the same databases can be queried for a variety of source-receiver configurations. For many applications, the heavy computational burden of simulating the background field is therefore transposed to storage requirements and to the embarrassingly parallel procedure of accessing a relevant database to extract and process wavefields.

The modified Instaseis interface provides the processing infrastructure, with rotations from local to global coordinate systems, interpolations of wavefields to points defined by the mesh or quadrature of the local method used for the coupling, temporal resampling of the local and global wavefields, as well as changes in the source mechanism. Communication between Instaseis and the local solver happens via HDF5 files with a specifically designed structure. What the local method needs to provide depends on the coupling case. For all three cases, Instaseis requires from the local 3-D method:

- the coordinates of mesh/gll points on the boundary where the coupling takes place;
- should the coordinates be in a local Cartesian coordinate system, a rotation matrix from the local Cartesian to a global Cartesian coordinate system (Instaseis uses this information to bridge between the local coordinate system and the AxiSEM cylindrical coordinate system).

Moreover, for source-side and depth coupling that make use of the reciprocal database for extrapolation Instaseis requires:

- the quadrature weights for each coordinate point of the mesh (α_m in equation 2.8);
- the unit normal vector to the boundary surface of the local method at each coordinate point (n_j in equation 2.8);
- the three components of displacement and six components of strain at each coordinate point (u_i^s and $C_{ijkl}\partial_k u_l^s$ in equation 2.8), as well as their sampling rate.

Finally, to correctly rotate the coordinates of a domain located at depth, the radius of the top boundary needs to be specified. For more details on the implementation and examples on how to use the modified Instaseis package, see Appendix A.

2.3.4 Local solvers: SPECFEM3D Cartesian and Wave Propagation Program

Currently the framework has been coupled with two local 3-D codes: Specfem3D Cartesian (Komatitsch and Tromp, 2002b), a spectral-element (SEM) solver, and Wave Propagation Program (WPP Nilsson et al., 2007; Petersson and Sjogreen, 2013) , a second order finite-difference (FD) solver. The Specfem3D implementation includes all three coupling cases, with the choice of the local domain on the source side, on the receiver side, and at depth. The WPP implementation, however, has been developed for the specific source-side application in Chapters 3 and 4 and has not yet been extended to the receiver and depth cases.

Specfem3D Cartesian

Specfem3D Cartesian is a software that is mostly used for seismic wave propagation at local and regional scales, although it can be applied to other acoustic and elastic wave propagation problems. It is well-suited for handling various geometries and allows to account for 3-D complexities affecting the seismic wavefield, such as lateral variations of elastic parameters and density, anelasticity, full 21-parameter anisotropy, and topography of internal discontinuities and of the free surface. It also includes coupled fluid-solid domains, but non-linear or visco-plastic behaviours cannot be accounted for. As an open-source software with a long-standing history and a wide user-base in the research community, Specfem3D Cartesian was a natural choice for a local 3-D solver for the hybrid framework.

We inject the global wavefield into a Specfem3D simulation via imposing the Dirichlet boundary condition on all points of the boundary of the SEM mesh. As described in Section 2.2.1, we subtract the background wavefield from the 'total' local wavefield, effectively imposing the residual scattered wavefield on the boundary. In order to avoid spurious reflections, we then apply the paraxial boundary condition (Stacey, 1988) to absorb the residual outgoing scattered wavefield. Applying the boundary condition to the outgoing scattered wavefield requires the knowledge of background tractions on the boundary of the SEM mesh – the known background tractions are subtracted from the locally computed tractions, while the residual (i.e. the difference not known a priori) is absorbed. Therefore, we query a forward Instaseis database for velocities and tractions on the GLL points on the mesh boundary. The values are stored in HDF5 files and later accessed by Specfem3D during the simulation – a modification added to the code for coupling purposes. Note that absorbing only the lower energy residual scattered wavefield renders the boundary condition more effective.

For extrapolation steps necessary for depth and source-side cases, we save displacements and stresses from the local simulation into interfacing HDF5 files. Given the mesh construction, there is no need to define a new quadrature to discretise the integral in the representation theorem and we directly use the weights of the GLL points in the extrapolation equation 2.8. The integration points are therefore well-defined for a high order quadrature which renders the extrapolation step of the coupling efficient and easy to implement. Moreover, the mesh construction respects the curvature of the Earth's surface as well as of the 1-D interfaces.

Wave Propagation Program

The Wave Propagation Program (WPP) is a second order (in time and space) open source FD solver developed at Lawrence Livermore National Laboratory for local and regional seismic wave propagation. WPP solves the elastodynamic equations of motion for elastic or anelastic materials with fully 3-D heterogeneous material model specification, including anisotropy and anelastic

attenuation with P- and S- wave quality factors. Surface topography can be represented through a boundary conforming curvilinear mesh created automatically with a built-in mesh generator from user-supplied files. Supergrid absorbing boundary conditions are implemented to dampen the outgoing wavefields, and depth-dependent mesh refinement can account for the increase in seismic wavespeeds.

For the extrapolation step necessary for the source-side coupling, we save displacements and strains from the local simulation into interfacing HDF5 files. Given the Cartesian mesh construction, we need to define a quadrature to discretise the integral in the representation theorem. With a dense regular FD grid, we choose to use the mid-point rule with the areas between grid points as weights in equation 2.8. With the current implementation, the mesh construction does not respect the curvature of the Earth's surface or of the 1-D interfaces – an issue further discussed in Appendix B.

2.4 Validation

In this section, we benchmark the proposed hybrid scheme. We first consider the hybrid Specfem3D Cartesian implementation against Instaseis and SEM solver Specfem3D Globe (Komatitsch and Tromp, 2002b). In Section 2.4.1 we keep the local Specfem domain unmodified, aiming to reconstruct the background wavefield and benchmark the hybrid solution for a 1-D model against the original Instaseis database. We then introduce a smooth Gaussian perturbation in the local Specfem region in Section 2.4.2 and benchmark hybrid seismograms against a 3-D reference solution computed by Specfem3D Globe. Finally, in Section 2.4.3 we turn to the WPP-Instaseis source-side implementation at short periods and discuss the issues of a Cartesian FD mesh in a global coupling. All 1-D simulations that reconstruct the background wavefield were validated for multiple receiver locations and components. The 3-D validation, however, was constrained to receiver locations at 30 degrees due to computational costs of global Specfem3D Globe simulations. We chose 30 degrees as a compromise between the simulation time of Specfem3D Globe and our aim to validate waveforms at teleseismic distances.

It should be highlighted that the accuracy of the benchmarks, and more generally of the hybrid method, depends on the meshing of the local solver relative to the 'dominant period' of the database. The parameters of the local simulation need to be considered carefully both relative to the requirements of the local solver itself, as well as relative to the discretisation of the surface integrals of the representation theorem for extrapolation and for injection via a distribution of point sources. Each use case needs to be considered separately, and it is recommended to test every problem-specific coupling set-up with 1-D benchmarks before introducing 3-D structures.

Note that the dominant period is the inverse of the dominant frequency of a simulation, and in this chapter it is understood according to the AxiSEM definition. The dominant frequency is

the frequency corresponding to the half-maximum of the spectral amplitude of the source-time function, ensuring that the bulk of the energy up to the dominant frequency is contained. In a standard Specfem3D Globe nomenclature, on the other hand, the frequency corresponding to the maximal frequency with non-zero spectral amplitude is considered as the 'dominant frequency'. The 10 s dominant period runs discussed below, for example, would be referred to as '5 s simulations' in Specfem3D Globe.

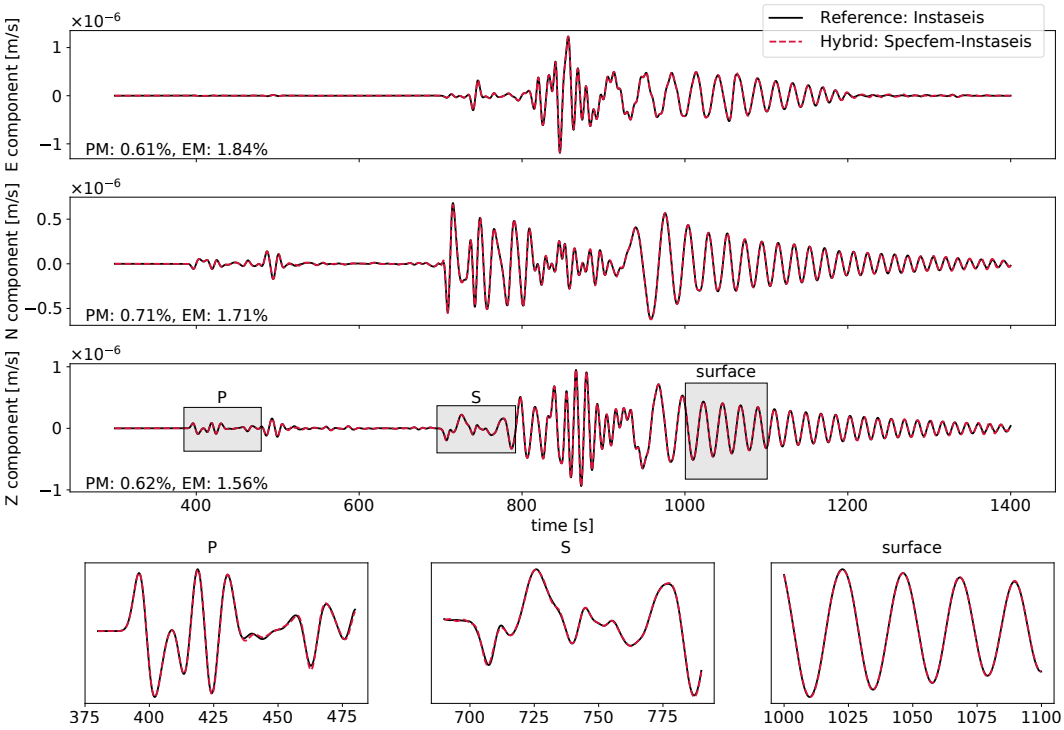
2.4.1 1-D benchmarks: Specfem-Instaseis wavefield reconstruction

As discussed in Section 2.2, the hybrid methodology should reproduce the original background wavefields if the local domain remains unmodified relative to the background model. With the local domain at the surface, either around the source or under the receivers, the wavefields can be reconstructed at teleseismic distances. However, for the local domain at depth along the propagation path, only the "residual" scattered wavefield ($\mathbf{u}^s(\mathbf{x}, t)$ in Section 2.2) is extrapolated to receivers at distance, while unchanged 1-D models only regenerate the 1-D interactions. Hence, the 1-D benchmark for the local domain at depth can only prove a correct reconstruction of the 1-D wavefield inside of a buried domain without a free surface. It tests a correct implementation of wavefield injection at depth, as extrapolation for the depth case cannot be validated with a 1-D model.

The accuracy is quantified in terms of phase and envelope misfits as defined by Kristeková et al. (2009) (PM and EM, respectively). For each pair of synthetics (hybrid Specfem-Instaseis versus 1-D), we compute the globally normalized EM and PM. Figure 2.6 presents sample velocity benchmark waveforms for a 1-D PREM model with a Gaussian source-time function of a 5 second half-width (i.e. 10 s dominant period) and a full moment tensor source. The local Specfem domains are 16 wavelengths in x, y and z – dimensions chosen to encompass a reasonable Gaussian perturbation of the 1-D model in subsequent 3-D benchmarks in Section 2.4.2 – and the mesh is constructed with 1.5 elements per S wavelength. For 4^{th} order spectral elements, this results in about 7 points per wavelength (PPW). The 1-D PREM model does not include attenuation for later ease of 3-D benchmarking against Specfem3D Globe, as different implementations of attenuating properties of the medium can add to waveform discrepancies. The 1-D PM and EM turn out invariably small ($<2\%$) with the 1-D hybrid waveforms indistinguishable from the original background waveforms.

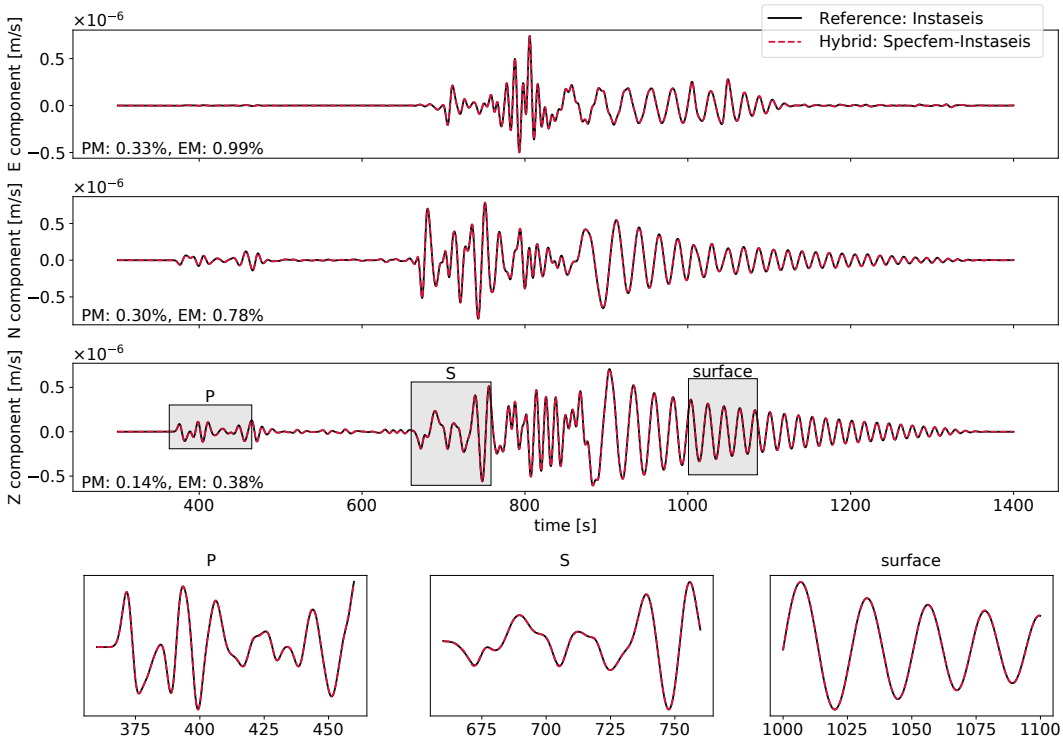
2.4.2 3-D benchmarks: Specfem-Instaseis local scattering

To benchmark the hybrid framework for 3-D structures in the local domain, we introduce a 40% Gaussian perturbation of seismic velocities v_p, v_s, and of density ρ over PREM into the Specfem3D Cartesian domains at depth and on the surface, around the source and under the

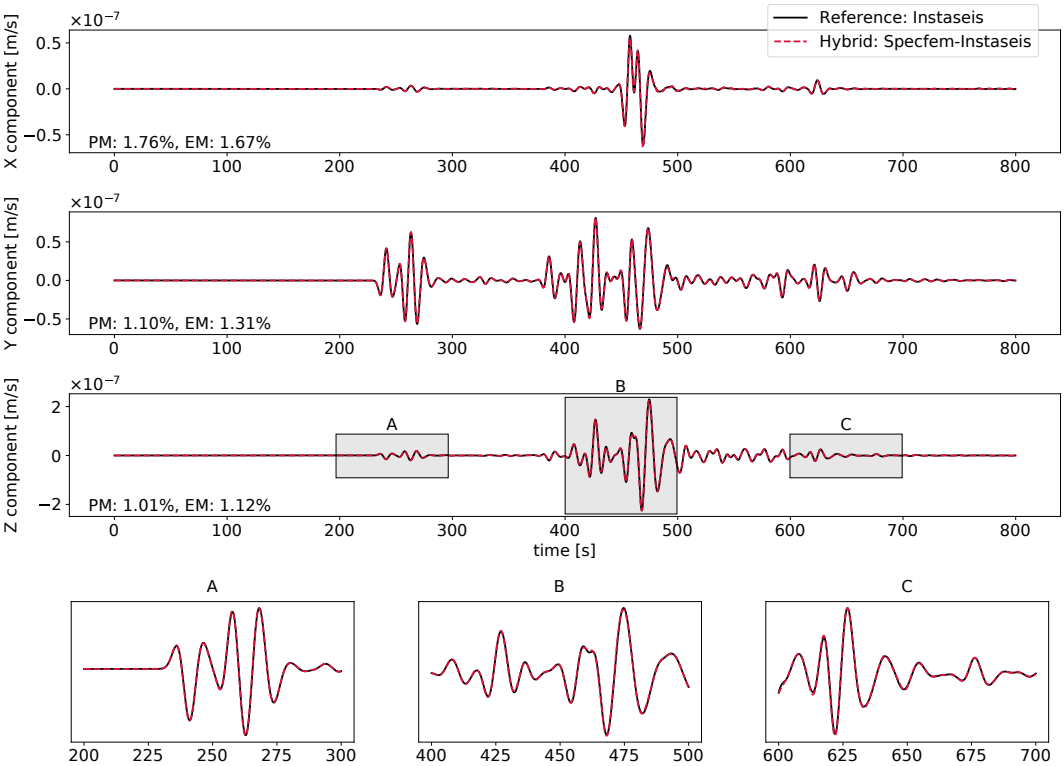


(a) local domain around the source

Figure 2.6: Examples of velocity benchmark waveforms for a 1-D PREM model with a local Specfem3D Cartesian domain (a) at the surface around the source, (b) at the surface under the receivers, and (c) at depth along the propagation path. All benchmarks include three component seismograms of the original 1-D waveform from the Instaseis database (black line) and a Specfem-Instaseis hybrid reconstruction (dashed red line). Benchmark in (c) reconstructs wavefields on the boundary of the local domain at depth and validates only the injection step for a buried domain. Extrapolation to receivers at distance cannot be tested for this set-up with a 1-D model (see main text). Presented waveforms are for a full moment tensor source with a Gaussian source-time function of a 5-second half-width (10 s dominant period of seismograms), and PREM model does not include attenuation.



(b) local domain under the receivers



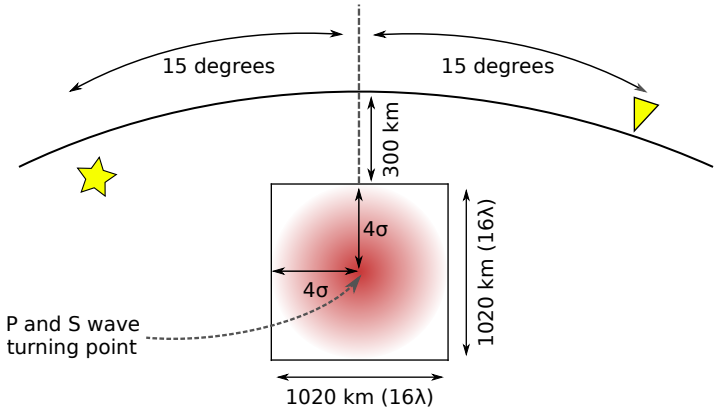
(c) local domain at depth

Figure 2.6: (continued).

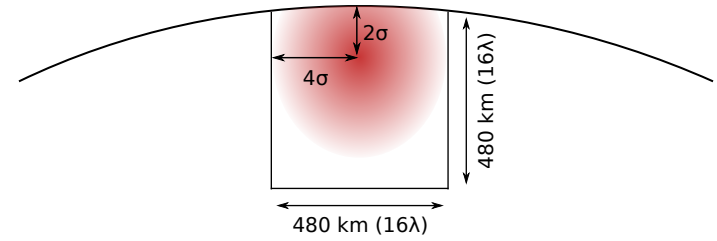
receivers. As before, a full moment tensor source and a Gaussian source-time function of a 5 second half-width are used, resulting in wavefields with a dominant period of 10 s. The Gaussian perturbation is defined with a standard deviation equal to two S wave wavelengths. For the domain at depth, the perturbation is located in the middle of the domain in x, y and z, and the domain itself is centred around the turning point of the P and S wave for the presented source-receiver configuration of 30 degrees epicentral distance (see Figure 2.7a). For the domains at the free surface – around the source and under the receivers – the centre of the Gaussian is located in the middle of the domain in x and y), and at a depth equal to two standard deviations, resulting in a strong perturbation on the surface (see Figure 2.7b). The domains are $16 \times 16 \times 16$ wavelengths in size, which is equivalent to $8 \times 8 \times 8$ standard deviations of the Gaussian perturbation. This ensures that the centre of the perturbation is 4 standard deviations from each boundary, leaving the boundary unaltered relative to the 1-D PREM model. A for the 1-D benchmarks, the mesh is constructed with 1.5 elements per S wavelength (i.e. 7 PPW).

Figure 2.8 presents sample velocity benchmark waveforms for the aforementioned 1-D PREM model with a 40% Gaussian perturbation (hybrid Specfem-Instaseis versus Specfem3D Globe). While the 1-D benchmark for a buried domain along the propagation path is not complete, as extrapolation to receivers cannot be tested by construction, for the 3-D case it presents the most complete setting that validates both injection (i.e. coupling with a forward database) and extrapolation (i.e. coupling with a reciprocal database). The global wavefield for the depth perturbation has therefore been simulated for 23 minutes to obtain a complete seismogram at 30 degrees, and Figure 2.8a presents the resulting benchmark of the full synthetic signal. Given the high computational cost of Specfem3D Globe simulations (271,200 CPUh for 23 minute seismograms at 10 s dominant period) and the completeness of the depth benchmark for injection and extrapolation, we choose to validate the source and receiver cases without surface waves and compute only 15 minutes of Specfem3D Globe seismograms. Moreover, since a correct generation of realistic surface waves requires a full 3-D representation of the crust between the source and the receiver, our hybrid methodology that approximates the Earth as a spherically symmetric layered medium with a flat free surface is not well suited for surface-wave applications. Since the hybrid benchmark at depth is complete and fully sufficient, and given our focus on high-frequency body waves, a validation of 1-D surface waves for all three coupling cases is of little interest.

The model perturbation has a comparable effect in both solvers, with differences arising due to numerical reasons as well as limitations of the hybrid method. The numerical issues, such as distinct meshing schemes, source and receiver mislocations, and accumulative round-off errors, are a consequence of solver differences. According to Leng et al. (2016), the misfits between Instaseis databases (that correspond to AxiSEM3D solution to 1-D models) and Specfem3D Globe simulations can generally be expected to remain below 2% for 1-D models, although

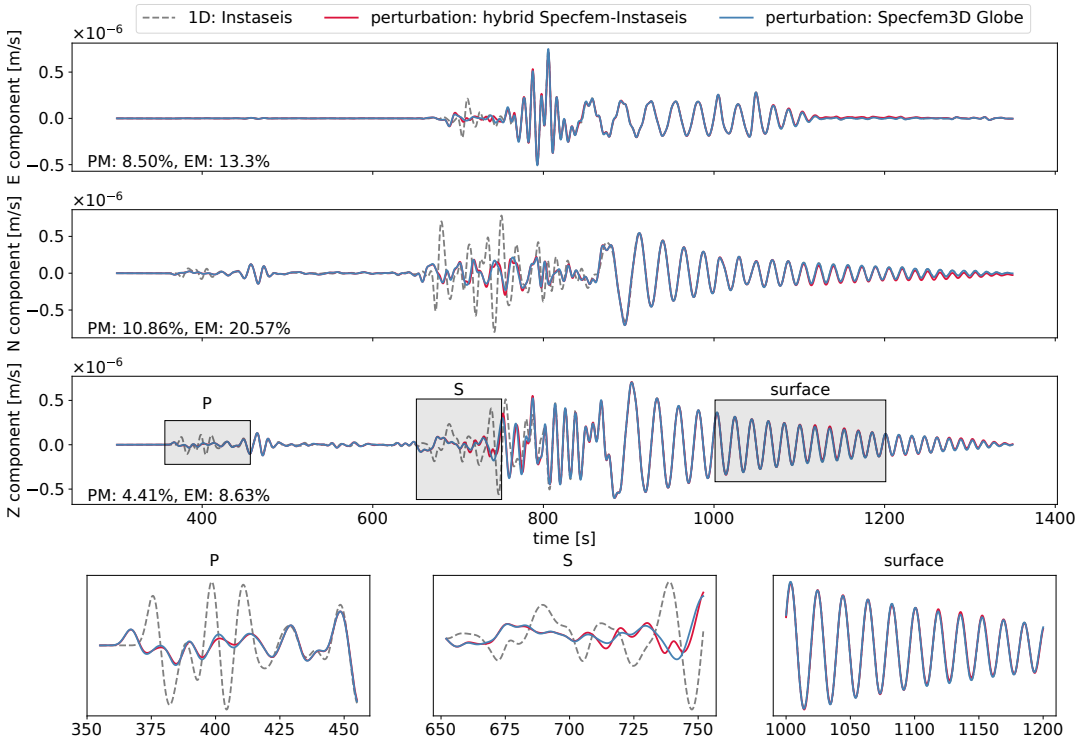


(a) local domain at depth



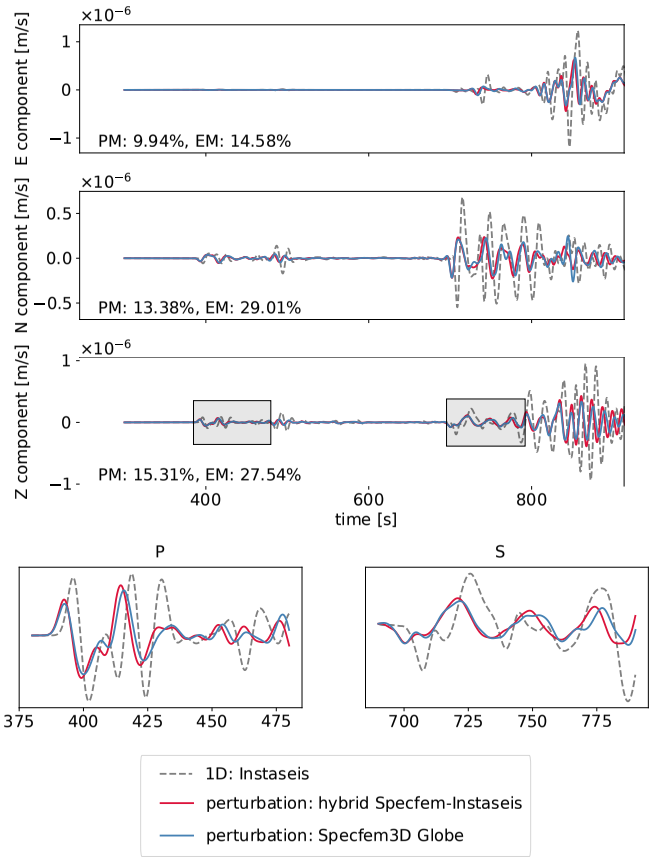
(b) local domain under the surface (around the source and under the receivers)

Figure 2.7: Sketch of benchmark set-ups of the Specfem-Instaseis hybrid coupling with a 40% Gaussian perturbation of seismic velocities v_p, v_s, and of density ρ. The domains are $16 \times 16 \times 16$ wavelengths in size, which is equivalent to $8 \times 8 \times 8$ standard deviations of the Gaussian perturbation. This ensures that the centre of the perturbation is 4 standard deviations from each boundary, so that the boundary matches the 1-D background model.
(a) The domain along the propagation path is centred around the turning point of the direct P and S waves for the source-receiver configuration of 30 degrees epicentral distance. The top of the domain is 300 km under the Earth's surface. The standard deviation of the Gaussian perturbation is equal to approximately two S wavelengths in the domain, i.e. 130 km.
(b) For domains at the free surface – around the source and under the receivers – the centre of the Gaussian is located in the middle of the domain in x and y, and at a depth equal to two standard deviations, resulting in a strong perturbation on the surface. The standard deviation of the Gaussian perturbation is equal to approximately two S wavelengths in the domain, i.e. 60 km.

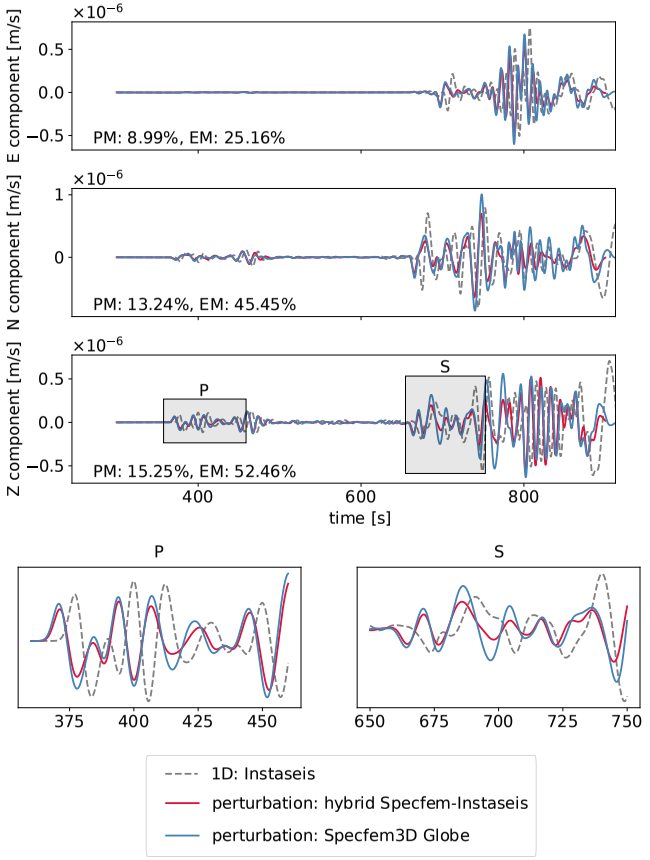


(a) local domain at depth (ID 1)

Figure 2.8: Examples of velocity benchmark waveforms of a local Specfem3D Cartesian coupling for a 1-D PREM model with a 40% Gaussian perturbation (a) at depth along the propagation path, (b) at the surface around the source, and (c) at the surface under the receivers. All benchmarks include three component seismograms of the original 1-D waveform from the Instaseis database (dashed grey line), a Specfem-Instaseis hybrid perturbation (red line), and a Specfem3D Globe perturbation (blue line). Benchmark in (a) validates both injection and extrapolation and presents a full seismogram including surface waves, while (b) and (c) are limited to body waves (see main text for an explanation). Presented waveforms are for a full moment tensor source with a Gaussian source-time function of a 5-second half-width (10 s dominant period of seismograms), and PREM model does not include attenuation. Run ID in each subfigure caption corresponds to IDs in Table 2.2 with a summary of computational costs.



(b) local domain around the source (ID 2)



(c) local domain under the receivers (ID 3)

Figure 2.8: (continued).

envelope misfits can sometimes be very large for unidentified reasons. For 3-D tomographic models, the misfits between AxiSEM3D and Specfem3D Globe are more significant and vary between receiver locations. Thus, with a locally perturbed 1-D model we can expect a mismatch between the hybrid and full 3-D seismograms, as meshing differences alone can result in different representations of the perturbation. Moreover, in the local method of the hybrid scheme, it is necessary to use absorbing boundary conditions which are known to be imperfect and can result in often significant spurious reflections. On top of such numerical issues, the limitations of a one-way hybrid scheme are expected to contribute to the visible waveform discrepancies. As later discussed in Section 2.6.1, higher order long range scattering is not accounted for in the coupling, while the Specfem3D Globe waveform represents all interactions between the perturbation, the free surface, and the internal 1-D model discontinuities.

The benchmark for the perturbation at depth along the propagation path shows the best fit, with PM of 10% and less and EM of 20% or less – values considered to represent an 'excellent fit' by Kristeková et al. (2009). With the local domain of 16 wavelengths in size (1020 km in x, y and z) and centred around the turning point of P and S waves, most contributing long range scattering is accounted for due to the size of the domain, and the resulting misfits can be mostly attributed to numerical discrepancies discussed above. For the source and receiver cases, however, the local domain is significantly smaller (480 km in x, y and z), as the decrease in wave velocities results in shorter wavelengths close to the Earth's surface. The local domain, therefore, does not encompass all contributing long range scattering and results in larger differences, illustrating the limitation discussed in Section 2.6.1. Filtering all components at 20 seconds reduces the misfits for all three cases, though the visible amplitude differences for the source and receiver coupling settings remain. Taking all contributing factors into consideration, we deem the results acceptable with the PM and EM under 20% for the complete coupling case that incorporates a perturbation at depth, with the larger misfits for the coupling set-ups under the free surface attributed to the limitations of the method.

2.4.3 Short period WPP-Instaseis wavefield reconstruction

As mentioned in Section 2.3.4 and discussed in Appendix B, the current implementation of the WPP-Instaseis source-side coupling does not respect the curvature of the Earth's surface or of internal discontinuities. A regular Cartesian WPP mesh is used, which for large domains results in a significant difference in altitude between the spherical Earth and the flat Cartesian surface. The method, however, requires that the coupling surface (∂V^i from Section 2.2) remains unchanged relative to the background model – an issue further discussed in Section 2.6.2. For large WPP Cartesian domains, this condition is not met and numerical errors arise. We have found that the magnitude of resulting errors is also dependent on the size of the altitude gap relative to the wavelength.

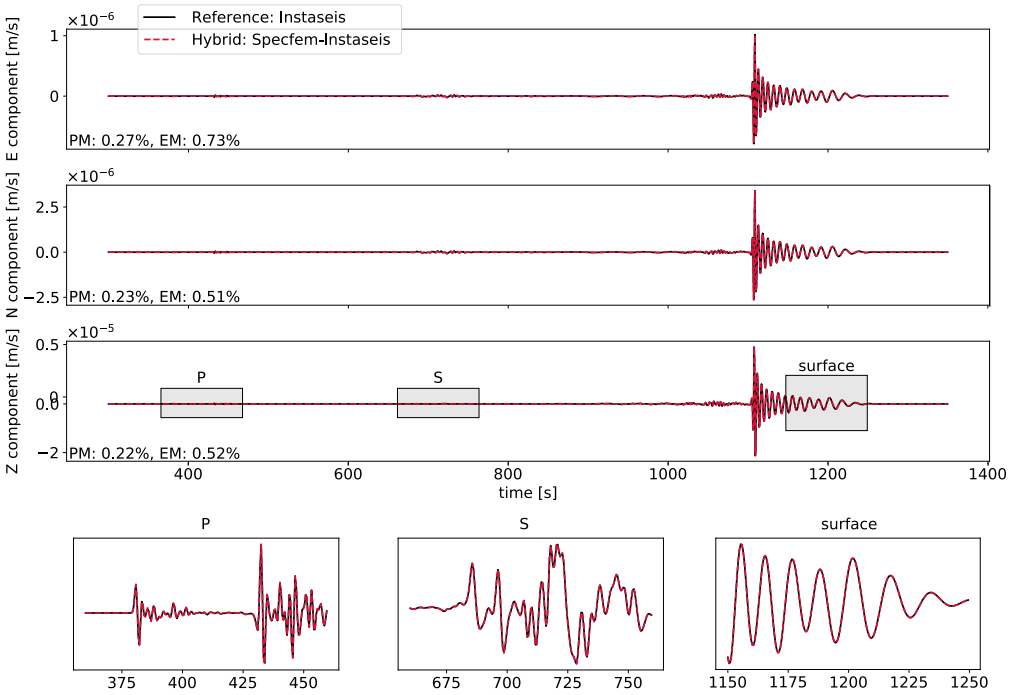


Figure 2.9: Example of velocity benchmark waveforms for a 1-D IASP91 model with a local WPP domain at the surface around the source (ID 4). The three component seismograms of a WPP-Instaseis hybrid reconstruction (dashed red line) are shown against the original 1-D waveform from the Instaseis database (black line). Presented waveforms are for an isotropic moment tensor source (explosion) with a Gaussian source-time function of a 1-second half-width (2 s dominant period of seismograms), and IASP91 model accounts for attenuation.

For small domains on the order of tens of kilometres, the errors remain small. Given the cost of global Specfem 3-D simulations, however, our 3-D benchmarks are limited to a dominant period of 10 seconds. The local domain needs to be large enough to encompass a smooth Gaussian perturbation of two wavelengths (see Section 2.4.2), which for a 10-second simulation means 500 km in x, y, and z. In a domain of this size the model mismatch is large, as both the free surface and five discontinuities that are within the top 500 km of a 1-D model are misrepresented on a Cartesian mesh relative to a spherical background model.

We therefore choose to benchmark the WPP-Instaseis wavefield extrapolation for a short period 1-D reconstruction at 2 seconds with a local domain of size $30 \times 30 \times 30$ km. According to the WPP manual (Petersson and Sjogreen, 2013), the number of points per wavelength (PPW), the largest significant frequency of the source time function (f_{max}), the minimum wavelength (v_{min}) and the FD grid spacing h_{min} are related according to

$$h_{min} = \frac{v_{min}}{\text{PPW} * f_{max}}. \tag{2.9}$$

With a minimum wavespeed v_s of 3360 m/s, 16 PPW recommended for best accuracy, and the required f_{max} of 1 Hz (1 s, for a 2 s dominant period), the WPP simulation should be sampled with at least 210 m FD grid spacing in x, y and z.

For the sake of variety, we choose an isotropic moment tensor source (explosion) and IASP91 as the 1-D background model, and we include attenuation to show the validity of the scheme for attenuating databases. As for the Specfem-Instaseis coupling in Section 2.4.1, the PM and EM are below 1% (Figure 2.9) for the chosen FD grid spacing of 200 m. With such 1-D extrapolation benchmarks demonstrating successful coupling of WPP with the source-side framework, and with the Specfem-Instaseis 3-D benchmarks proving the validity of the method for 3-D structures in the local domain, a locally modified model can be considered accurate for WPP-Instaseis as long as the WPP implementation of 3-D model complexities is accurate.

2.5 Performance

2.5.1 Database generation and storage

The bulk of the computational effort of the hybrid methodology lies in database generation. As illustrated in Figure 2.5, the cost of generating an AxiSEM database scales with frequency to the 3$^{\text{rd}}$ power. The actual cost of each run depends both on the machine and on the chosen simulation parameters (e.g. attenuation, number of elements per wavelength), and a forward database is approximately twice as expensive as a reciprocal database for the same set of parameters. It should be noted that the cost can be significantly reduced for specialised studies with limited epicentral distances, as the 2-D disk discussed in Figure 2.4 can be cropped to encompass only the required latitudes (with absorbing boundaries applied to the additional non-physical boundary).

The advantage of database generation is the once-and-for-all approach. Storing the Green's functions from a single AxiSEM run, one can freely access and re-use the 1-D background wavefields for a variety of hybrid applications – the location and size of the local domain, and more generally of the source-receiver geometry can vary, and the source parameters can be modified. Therefore, the cost of generating a database can be considered a one-off investment for a large spectrum of problems, shifting the focus from regular access to substantial computational resources to the necessary database storage requirements. Figure 2.10 shows the dependence of the required storage space of reciprocal databases on frequency, maximum source depth, components, seismogram length and the range of epicentral distances. A forward database takes approximately twice as much space as a reciprocal database for the same set of parameters. The computational and storage requirements for the databases for benchmarks in Section 2.4 are

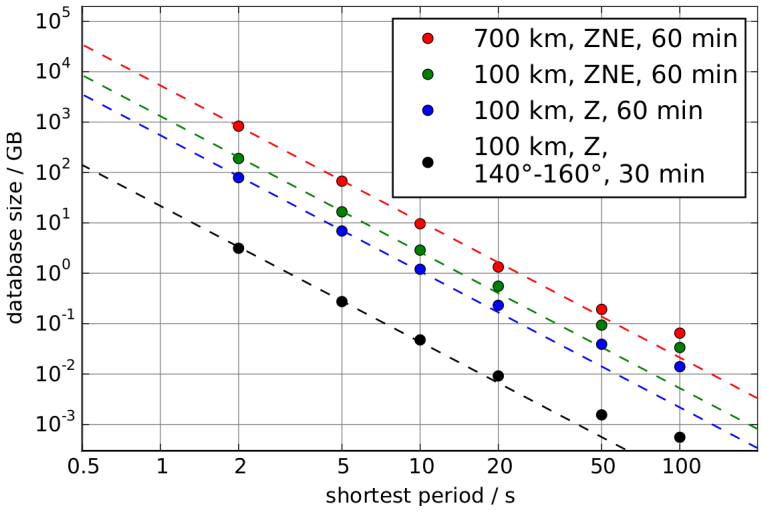


Figure 2.10: Comparison of storage requirements of reciprocal Instaseis databases with different storage parameters after zip compression. Maximum source depth, components, seismogram length and the range of epicentral distances define the final database size. Dashed lines are fitted functions $g(f) = af^{2.7}$. Given that zip compression is more efficient for longer time series, the exponent is slightly smaller than the expected 3. (Source: van Driel (2015))

DB type	model	CPUh	Storage	Length	Components	Depths	Distances
BWD 10 s	PREM	27	54 GB	1500 s	3	0-1000 km	all
FWD 10 s	PREM	57	114 GB	1500 s	3	0-1000 km	all
BWD 2 s	IASP91 (attenuating)	n/a	1.38 TB	3700 s	3	0-700 km	all

Table 2.1: Computational and storage requirements of Instaseis databases used for benchmarking, where FWD stands for forward, and BWD for reciprocal database, and database type also indicates the dominant period of the run. Maximum source depth, components, seismogram length and the range of epicentral distances define the final database size. Exact CPU hours of the 2 s reciprocal database are unknown, as it has been generated for other purposes prior to this study.

listed in Table 2.1.

2.5.2 Hybrid injection and extrapolation

There is no general rule for the costs of injection and extrapolation, as both depend on the the length of the local simulation, as well as on mesh discretisation and time step of the local

method:

- The number of points on the boundary of the local domain influences both injection and extrapolation. It depends on the size of the local domain, as well as on the points chosen for the coupling. For injection, the points must correspond to the mesh of the local method. For extrapolation, one could define a quadrature independent of the mesh points, provided that the local method can accurately interpolate the local wavefields to any point. In both Specfem and WPP implementations, however, the extrapolation points are chosen to be the local mesh points, and in WPP using only every other point on the boundary proves sufficiently accurate for extrapolation and reduces the cost.

- For injection, background wavefields extracted from the database are required to have the same time length and time step as defined in the local simulation, and both factors influence the cost. Longer time series require more time to save to HDF5. Moreover, the costs of interpolating wavefields to the right time step vary and depend both on the sampling of the database and on the new sampling required.

- For extrapolation, the time length of the local wavefields influences the computational cost. However, the local wavefields are downsampled for extrapolation to avoid the high costs of interpolation, and hence the time step of the local simulation has less of an effect on the overall cost than in case of injection.

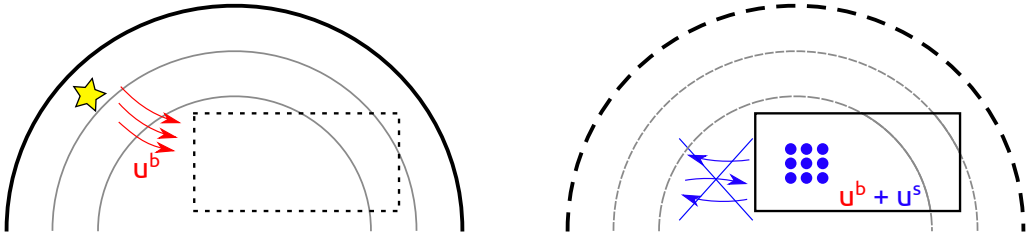
Table 2.2 summarises the injection and extrapolation costs and parameters (number of points, time steps, lengths of wavefields) for benchmark runs in Section 2.4.2, as well as CPU hours (CPUh) of respective local simulations and one-off database runs. Note that the cost of injection includes all receivers defined in the local domain, while the cost of extrapolation is a per receiver cost for a three-component seismogram. Therefore, the injection cost for a single application that requires multiple realisations of the same local domain is a one-off cost. Wavefields are saved into HDF5 files according to the requirements of the local mesh and remain available for multiple runs of the local solver, further reducing the costs of injection solely to the cumulative cost of required local runs. However, modifications in the local domain for hybrid extrapolation require re-running both the local simulation and the extrapolation to receivers.

Depending on the parameters of the set-up, the computational cost of a full standard hybrid run – including the cost of the local simulation that is method dependent, but excluding database generation – is reduced by a factor of 5,000-100,000 relative to a full 3-D run, provided that the local domain is of the order of tens of wavelengths in size. A Specfem3D Globe simulation for the 3-D benchmark at depth (run ID 1) cost 271,200 CPUh for 23 minutes of wave propagation, while the source and receiver cases (run ID 1 and 2, respectively) required 227,500 CPUh for 15 minutes of wave propagation. The corresponding hybrid coupling costs are on the order of tens of CPUh.

The database runs, even though included in the comparison, are a one-off cost available for multiple hybrid applications and settings. Note also that the costs of local Specfem runs in Table 2.2 (i.e. for runs 1-3) might seem inconsistent between the three cases. This is a result of an inefficient implementation of parallel HDF5 save in Specfem that is not native to the code and was added for the purpose of the coupling. Improving the parallel HDF5 save would further reduce the costs of the simulations and make the cost of the runs less dependent on the number of points – a necessary further development discussed in Section 5.1.2.

run ID	DB type	domain type	domain size	# of points	length of in/outputs	dt of local run	dt of DB	CPUh inject.	CPUh extrap.	seismogram length	CPUh of DB	CPUh of local run	CPUh w/o DB	CPUh with DB
1.	fwd&bwd 10s	depth	1,020 km	153,000	800 s	0.2 s	1.257 s	21.8	3.6	1,350 s	84	33.6	59.0	143.0
2.	bwd 10s	src	480 km	50,400	600 s	0.1 s	1.257 s	n/a	1.4	900 s	27	2.2	3.6	30.6
3.	fwd 10s	rec	480 km	50,400	600 s	0.1 s	1.257 s	11.5	n/a	900 s	57	0.6	12.1	69.1
4.	bwd 2s	src	30 km	13,276	100 s	0.0427 s	0.483 s	n/a	0.6	1,350 s	n/a	2.7	3.3	n/a

Table 2.2: Summary of total hybrid costs of benchmark runs in Section 2.4.2. Injection CPUh is a one-off cost for a single application that requires multiple realisations of the local domain. Extrapolation CPUh is a cost of a single three component seismogram. Run ID corresponds to IDs in Figures 2.8 and 2.9. Excluding costs of database generation, a full standard hybrid run is 5,000-100,000 times cheaper than a full 3-D run by Specfem3D Globe, and exact costs depend on the local method, on the chosen parameters, and on the location of the local domain. A Specfem3D Globe simulation for the 3-D benchmark at depth (run ID 1) cost 271,200 CPUh for 23 minutes of wave propagation, while the source and receiver cases (run ID 1 and 2, respectively) required 227,500 CPUh for 15 minutes of wave propagation. The corresponding hybrid coupling costs presented in the table are on the order of tens of CPUh. Benchmark database indicates whether a forward (fwd) or a reciprocal (bwd) database was used, and specifies the dominant period, and columns 3-8 summarise injection and extrapolation parameters that affect the costs. Length of inputs/outputs defines the time length of wavefields in interfacing HDF5 files with spacing between samples equal to the time step (dt) of a given local run. Recall that the database cost is a one-off cost, as a single database is available for a multitude of hybrid applications.



(a) (b)

Figure 2.11: Sketch of missing long range higher order interactions in a hybrid one-way injection coupling. (a) Background wavefield accounts for all global interactions in a background model. (b) In the 3-D local domain, all interactions are accounted for *within* the domain itself. However, the residual scattered wavefield $\mathbf{u}^s(\mathbf{x}, t)$ can only leave the domain, but cannot subsequently return after a reflection off a 1-D discontinuity in order to interact with 3-D structures once again. (a) and (b) illustrate missing interactions for an injection scheme, but the same can be said for extrapolation to receivers at distance.

2.6 Technical considerations

2.6.1 Limitation: higher order long range scattering

Not all wavefield interactions are accounted for with a one-way approach to coupling given the lack of a dynamic exchange of information on the boundary. All scattering effects are present within the local domain both for wavefield injection and extrapolation, and the scattered wavefield can leave the local domain. However, it can never return into the local domain in order to interact with embedded 3-D structures again, as the global and local solvers do not act simultaneously or iteratively, but sequentially. Figure 2.11 illustrates such neglected higher order long range interactions for an injection scheme, and the same principle applies to wavefield extrapolation. The 3-D benchmarks in Section 2.4.2 illustrate the effect relative to a full 3-D solution for an exaggerated heterogeneity of 40% in seismic velocities v_p, v_s and density ρ.

Two-step type methods, therefore, are not suitable for problems where scattering is not localised. For example, they are often deemed inappropriate for near-surface receiver-side problems in exploration geophysics, where a dynamic exchange of information is necessary to capture all interactions of interest (Broggini et al., 2017). In global seismology, on the other hand, the effects of higher order long range scattering at teleseismic distances are often small relative to primary interactions that the method can help understand. When designing a problem, however, it is of paramount importance to understand this limitation and the potential effect in can have on resulting waveforms.

2.6.2 Implementation: local-global model mismatch

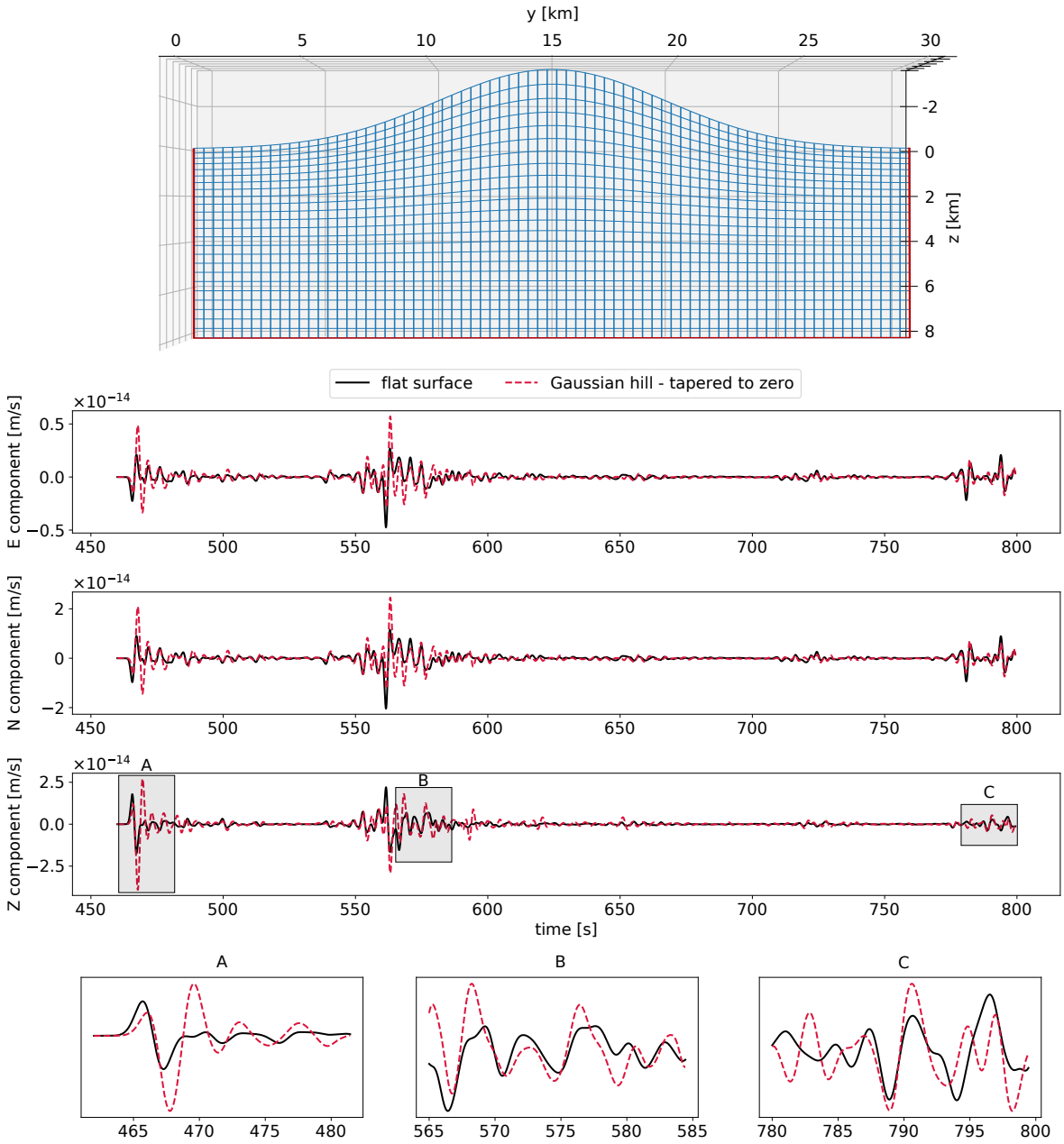
As discussed in Section 2.2, for exact results the boundary of the local domain should remain unmodified relative to the background model. One can, however, envisage applications for which an exact set-up is not achievable, as tapering a heterogeneity of interest to a 1-D background model might generate non-physical structures – for example, when considering topography or oceans – while defining a large domain fully encompassing the heterogeneity carries a high computational cost. Provided an in-depth understanding of the consequences, one can choose to violate the condition of a perfect model match to avoid tapering. Such approach necessitates a careful coupling design and problem-specific benchmarking.

We illustrate such violation of the matching boundary in Figure 2.12 with an example of a local topography that diverges from the 1-D topography. First, a Gaussian hill is added in the local domain, with correct tapering to the level of the free surface of the 1-D model. The hill is 20.2 km wide and 3.4 km high, which for a simulation of 2 seconds period corresponds to three wavelengths and half a wavelength, respectively. The effect of such correct implementation of the hill is shown in Figure 2.12a relative to a 1-D background seismogram. Next, the local domain is progressively cropped around the hill, leading to ever larger topography mismatches between the local and the global model. As can be seen in Figure 2.12b, the effect on the 2-second waveform is negligible relative to the overall effect of the hill, and shows that violating the boundary condition – although it might – does not necessarily lead to drastic errors and can be an intentional choice of an implementation.

2.6.3 Extensions: 3-D background model, multiple local domains, non-linear coupling

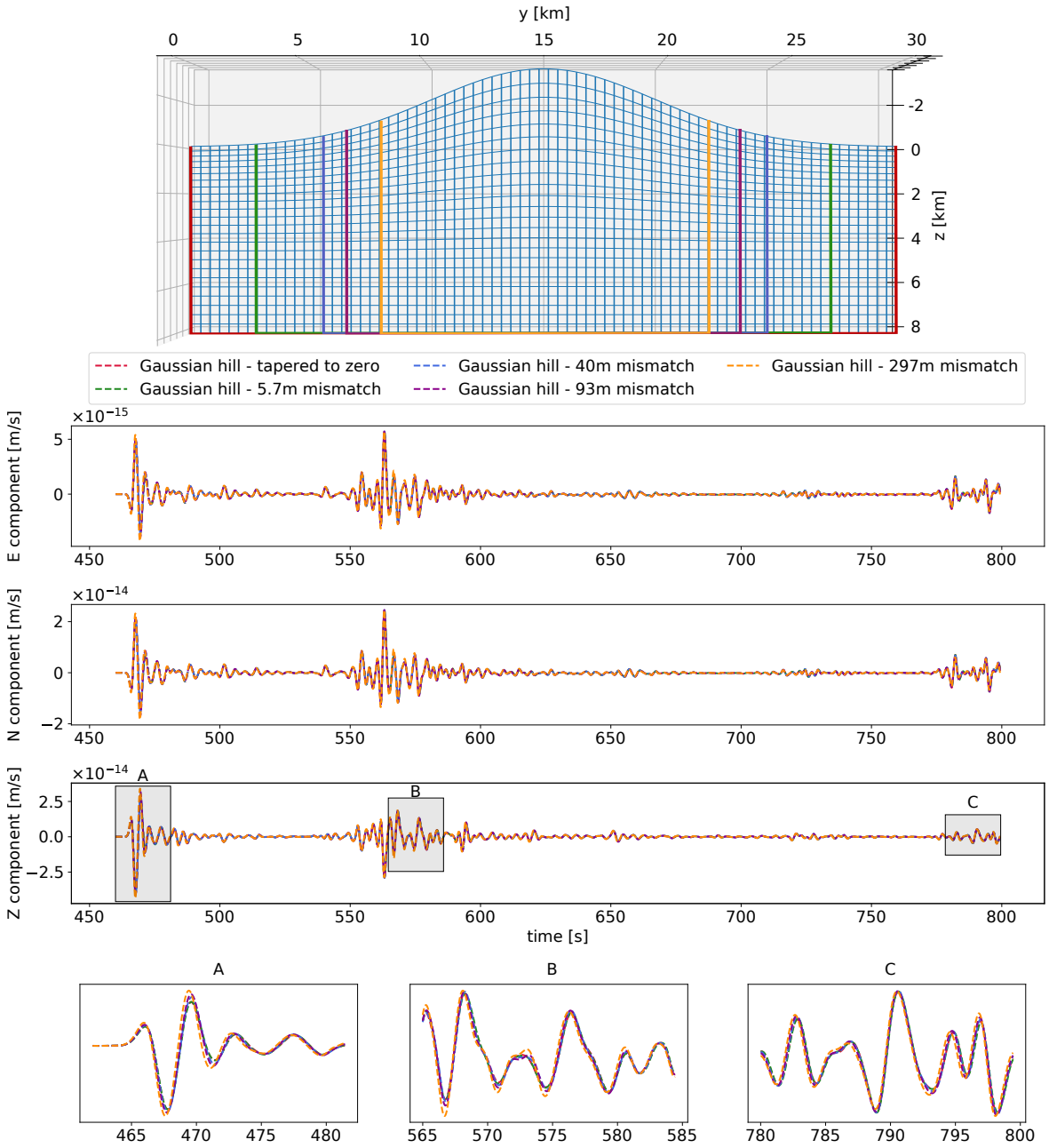
The theory in Section 2.2 was presented in its most general form, applicable to any background and local solvers. The global scale implementation with Instaseis databases defines the background medium as a spherically symmetric 1-D Earth model, a structural compromise that allows to compute high-frequency teleseismic wavefields with reasonable costs. A tomographic 3-D background model is a natural extension that would bring hybrid modelling closer to realistic scenarios. AxiSEM3D (Leng et al., 2016) – a method exploiting the smoothness of global wavefields – has drastically improved our capability to simulate high-frequency teleseismic waves, and is currently developed towards integration with Instaseis with a new generation of 3-D databases. An AxiSEM3D extension to Instaseis will provide the hybrid framework with 3-D to 3-D coupling at resolutions above 1 Hz that retains the once-and-for-all property of re-usable databases.

One can also envisage incorporating local solvers that include more complex physics than presented in this chapter. Xu et al. (2013), for example, propose a local-to-local hydrodynamic-



(a)

Figure 2.12: Illustration of topography mismatch on the boundary coupling WPP with Instaseis. A 20.2 km wide and 3.4 km high Gaussian hill is introduced in WPP, which for a simulation of 2 seconds period corresponds to three wavelengths and half a wavelength, respectively. Only 350 seconds after the P wave onset are shown to increase seismogram legibility. (a) presents the seismic signature of a correctly implemented Gaussian hill that is tapered off to zero to match the 1-D background model, while (b) shows the negligible effects of increasing topography mismatches between the local and the global model. For such cases, violating the boundary condition of hybrid coupling can be an intentional design choice for a specific problem.



(b)

Figure 2.12: (continued). In top panels of (a) and (b), a cross-section of the mesh of the correctly implemented local domain is shown where the Gaussian hill tapers off to zero. In (b), the different local domain sizes are sketched in their respective colours.

to-elastic coupling to analyse the far-field signature of non-linear near-source effects generated by nuclear explosions. Embedding hydrodynamic interactions in a global database framework, one could study the contributions of such effects at teleseismic distances. Incorporating dynamic rupture solvers (e.g. SeisSol, `http://www.seissol.org/`), on the other hand, would give insights into the long range effects of complex earthquake fault ruptures.

Finally, the flexible implementation of the framework – with HDF5 files that provide an interface between Instaseis databases and local methods – allows for set-ups that include multiple local domains. Multiple localised structures at depth can readily be incorporated, although interactions between structures in individual domains are not accounted for (similarly to second order long range interactions, see Figure 2.11).

2.7 Wave-propagation applications

Given its ability to generate high-frequency seismograms at teleseismic distances, our hybrid method allows to study the influence of localised 3-D structures on global propagation of body waves with frequencies over 1 Hz. Purely synthetic studies in multi-scale media are made possible, providing insights into the long-range effects of complex scattering and a better understanding of uncertainties in existing models. Realistic simulations of structure-induced anomalies, on the other hand, are essential to interpretation and analysis of seismic recordings, and can now be achieved for highest frequencies observed teleseismically.

With strong topography prevalent both in regions of high seismicity and in nuclear test sites, accounting for topographic scattering in a local 3-D model around the source region would allow to capture its teleseismic effects on body waves – a topic that is the focus of the remainder of this thesis in the context of forensic seismology. The data-driven study in Chapters 3 and 4 compares observations with synthetics at 2 Hz, helping to interpret waveform complexity. Source-side modelling can also account for heterogeneous small-scale geologic structures, including random media, or for realistic finite faults (Earthquake Hazards Program, 2017) or dynamic ruptures (Harris et al., 2018; Wollherr et al., 2018). In all such applications, trade-offs between various structural and source parameters – and how such interactions translate to teleseismic signals – could be evaluated. Moreover, applying backprojection (Ishii et al., 2005; Xu et al., 2009) or time-reversal (Boschi et al., 2018) to synthetics could help to better map the time and space history of high-frequency body waves and their source-time evolution, highlighting interactions with specific scatterers in the source region.

Embedding 3-D solvers at depth and modelling localised small-scale structures along the propagation path could improve our understanding of teleseismic signatures – or lack thereof – of deep mantle structures and provide important constraints on the evolution of the Earth's interior. Global tomographic models are in reasonable agreement for large-scale 3-D structures

(Becker and Boschi, 2002; Auer et al., 2014), but diverge at smaller length scales as a result of structural modelling assumptions and insufficient data coverage. A multi-scale view of thermal and chemical heterogeneities in the interior of the Earth, however, carries implications for global mantle convection and the Earth's heat budget, including core-mantle exchanges and geochemical as well as mineralogical constraints (Helffrich and Wood, 2001; Dziewonski et al., 2010; Steinberger and Torsvik, 2012). Synthetic waveforms at high frequencies could provide insights into seismic properties of structures in the mid-mantle and around the CMB, such as velocities, densities, anisotropy and attenuation, as well as topography and sharpness of the lateral (e.g. D", 660 and Moho) and vertical (e.g. LLSVP) interfaces. Hybrid methods can also be directly adapted to localised tomographic studies based on full waveforms in order to improve the resolution of global Earth models at depth (Masson and Romanowicz, 2017b).

Finally, receiver-side hybrid modelling could provide a numerical means of exploring the ever larger high quality datasets from seismic arrays, such as USArray, AlpArray or Hi-net. Initially installed to improve the signal-to-noise ratio of teleseismic body waves from nuclear explosions, seismic arrays – with their regular and closely spaced instrument configurations – have since been used to study the Earth's fine-scale structure. Mirroring such dense datasets with high-frequency hybrid modelling would help not only in the interpretation of seismic recordings, but also in a systematic analysis of various array techniques, such as beamforming and its derivatives, different slant stacking techniques, and frequency-wave number analysis (for a review of array methods, see for example Rost and Thomas (2002)).

2.8 Summary

Applying the principles of wavefield injection and wavefield extrapolation, we propose a global hybrid framework built around Instaseis databases. The methodology boasts an unprecedented level of flexibility, and can provide up to a 100,000 fold speed-up relative to full 3-D global solvers. With the databases serving as a once-and-for-all solution to wave propagation in spherically symmetric Earth models, the global wavefields are applicable to a variety of problems, shifting the heavy computational costs to long-term storage requirements. A single database can not only accommodate for different geometries and locations of the local domain within the background model, but also can be used for coupling with multiple local solvers. Moreover, it allows alterations in source properties, such as locations, source-time functions and radiation patterns, leaving such choices to later hybrid simulations. With the ability to generate and store wavefields in the 1-4 Hz frequency range, and with the flexibility of the subsequent local modelling, the hybrid method can reach the highest frequencies observed teleseismically and opens global wave propagation to a new parameter space.

Chapter 3

A nuclear explosions case study, Part I: A manifestation of source complexity in observed waveforms

3.1 Introduction

3.1.1 Motivation for the nuclear explosions case study

An in-depth understanding of teleseismic signatures of source-structure interactions is of paramount importance in forensic seismology. In realistic 3-D Earth models, factors such as depth of burial, explosion yield, near-source scattering from surrounding topography and from strong structural heterogeneities can all impact the complexity of observed waveforms. Non-linear effects in close vicinity of the source (e.g. cavity creation and elastic radius) can further alter the wavefield radiating from the source region. An explosion-induced tectonic strain release, a change in rock properties, a collapse of the cavity, as well as spall (when the material above the explosion falls down in a "slapdown" after an initial upward acceleration) are thought to contribute to seismic signal characteristics (Rodean, 1981). The relative importance of such effects at teleseismic distances, however, is not well understood. Quantifying how these near-source interactions contribute to waveform characteristics and how they may obscure the signature of the source could lead to improvements in event detection and identification, as well as help refine existing nuclear discrimination methods. Both changes in amplitudes of the first P wave and of the wave-train that follows the direct arrival impact our ability to accurately estimate event depths and magnitudes, and then extrapolate this information to yields which are crucial in nuclear explosions monitoring. Moreover, the main teleseismic discriminants of nuclear explosions – such as P wave complexity (simplicity), ratio of body- to surface-wave magnitudes, and depth estimates (Bowers and Selby, 2009) – all can be affected by a strong variability in

the source-structure interactions. Complexity discriminants have been shown to perform better when measurements over multiple stations are considered (Bowers, 1996), and a combination of discriminants tends to generate more robust results (Taylor and Anderson, 2009), suggesting that a broad set of measurements over a range of distances can to some extent correct for direction- and azimuth-dependent near-source and path effects. However, only understanding and quantifying such effects will render discrimination methods more effective and reliable.

During the late 1980s and early 1990s, near-source contributions to teleseismic P waves and early coda waves have been analysed for a number of underground explosion sites. Lay (1987) studied events from Pahute Mesa (at the Nevada Test Site, NTS), Amchatka and Novaya Zemlya that exhibit different patterns of P coda scattering. They concluded that the overall variations are a combination of near-source heterogeneities and deep-mantle propagation effects that are difficult to disentangle through data analysis and should be addressed via deterministic modelling. Lynnes and Lay (1988) analysed events at the Yucca Flat (NTS) and demonstrated a strong correlation of P wave coda with location, suggesting contributions of the near-source region. McLaughlin and Jih (1988) discussed common dominant characteristics of P-coda from the French Southern Sahara explosions for stations at similar take-off angles and azimuths and noted that they may be attributed to the direction-dependent scattering in the vicinity of the source. Lynnes and Lay (1989) inverted P wave data from Yucca Flat (NTS) explosions in an attempt to determine isotropic point scatterers near the source and suggested that topographic irregularities might play a role in coda generation. Lay and Zhang (1992) examined the spectral content of the P wave and its early coda for NTS and Novaya Zemlya events. The two sites show different trends, and they concluded that further quantification of the observed behaviour is necessary to draw any conclusions regarding the near-source structures.

With new processing techniques and numerical methods at hand, revisiting the now digitised data from underground nuclear explosions of the 1960s-1980s could bring valuable insights into the complexities of near-source interactions. Those large datasets provide an ideal starting point for a systematic and rigorous analysis that can be related to synthetic data – the source locations, site topography and site geology are often well known, and waveforms have been extensively processed and analysed at frequencies up to 8 Hz. Such prior knowledge reduces the level of uncertainty in the subsequent numerical set-up and allows for more straightforward comparisons between the observed data and synthetics.

In this study – comprised of two parts – we analyse both the observed waveforms and the hybrid synthetics from explosions at the Degelen test site and investigate the effects of the source region on teleseismic wavefields from nuclear explosions. The first part of the study discussed in this chapter focuses on an analysis of recorded waveforms, while the second part in Chapter 4 attempts to evaluate through hybrid modelling waveform variations stemming from interactions with topography in the vicinity of the source.

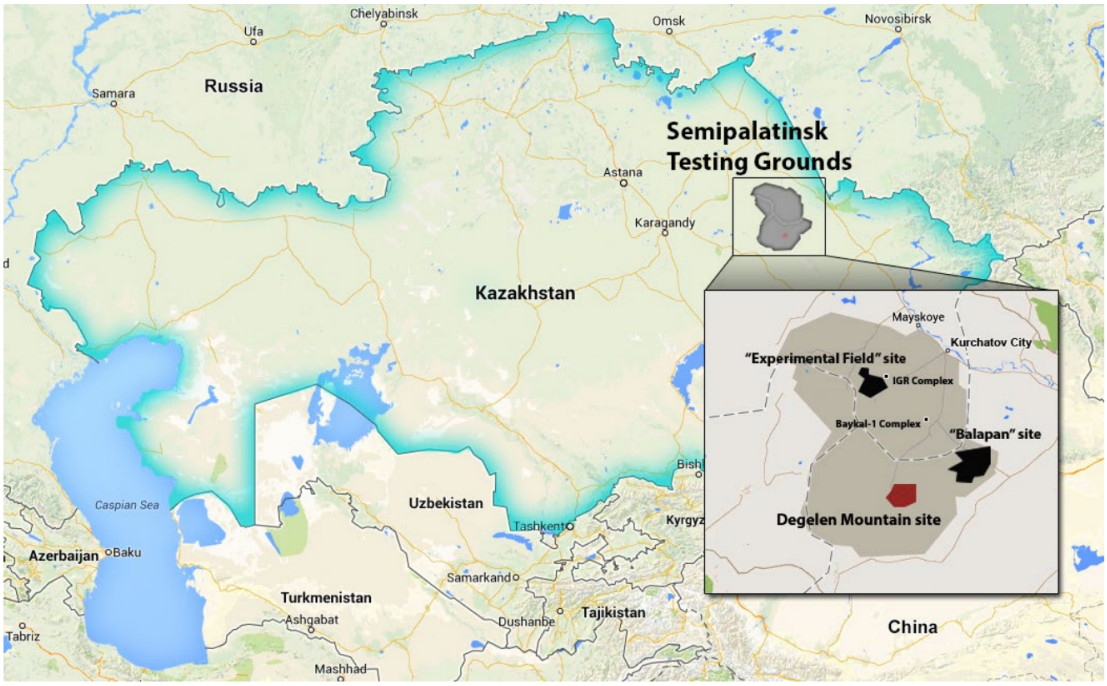


Figure 3.1: Location of the Degelen mountain test site, one of the three main test areas of the USSR nuclear test site Semipalatinsk (STS) in north-east Kazakhstan. (Source: Harrell and Hoffman (2013))

3.1.2 The Degelen test site

The Semipalatinsk test site (STS), located in the north-east of the Republic of Kazakhstan, was the primary location for nuclear testing in the USSR with 116 detonations conducted in the atmosphere and 340 underground in years 1949-1989 (Harrell and Hoffman, 2013). 214 of the underground tests took place in the Degelen Mountan test site, a small granite mountain range that spans an area of about 220 km^2 in the south of the STS (see Figure 3.1). Most explosions were small- or medium-sized with charges between 4 to 20 kt TNT, with a number of explosions in the 20-125 kt range (Marshall et al., 1992; Böttger et al., 1998). All tests have been conducted in adits (sub-horizontal tunnels) drilled into the slopes of the Degelen massif in various locations, with charges usually placed at the adit end (Böttger et al., 1998).

Given the sheer quantity of Degelen detonations, as well as the locations, depths and yields officially released in 1989 by the USSR for the pre-1972 tests (Bocharov et al., 1989), the recorded waveforms constitute a large and unique data set. The quality of many of the waveforms is outstanding, as the UK-design seismic arrays – further discussed in Section 3.2 – have been installed specifically to monitor the USSR explosions, and their directional set-up maximises the signal-to-noise ratio for the STS. Moreover, plans of the entire adit complex are available (Trabant et

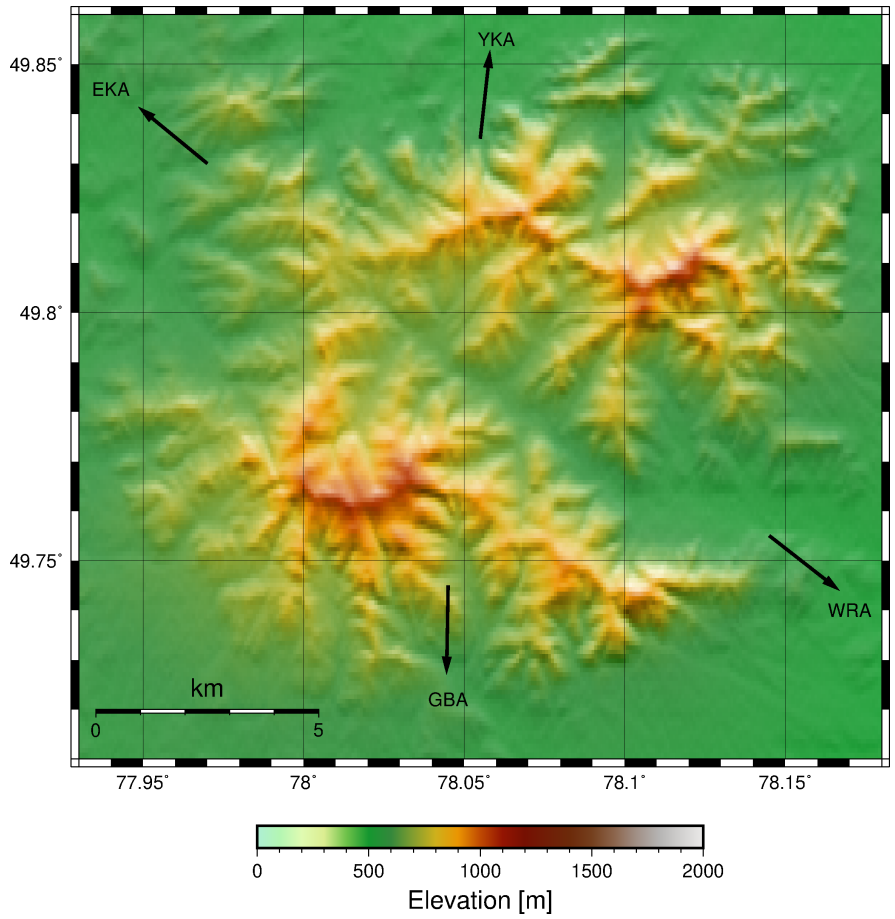


Figure 3.2: The Degelen mountain complex constitutes two ridges running from north-west to south-east and separated by a valley. The ridges come together in the north-west. The flat steppe that surrounds Degelen has altitudes of about 500-600 m above sea-level, while the mountains themselves attain a maximum altitude of 1084 m. Generally the slopes are not very steep. The arrows indicate the direction to the four UK-design arrays used in this study.

Array	Latitude in degrees	Longitude in degrees	Distance in degrees	Azimuth source-receiver degrees E of N	Backbearing receiver-source degrees E of N
Gauribidanur (GBA)	13.6042 N	77.4361 E	36.4	182	2
Eskdalemuir (EKA)	55.3317 N	3.1592 W	47.1	310	60
Yellowknife (YKA)	62.4932 N	114.6053 W	67.0	7	351
Warramunga (WRA)	19.9426 S	134.3394 E	85.3	129	328

Table 3.1: Locations of four arrays used in this study and their distance and azimuth to the Degelen test site. The quoted latitudes and longitudes correspond to the point of intersection of the two lines of each array as shown in Figure 3.4.

al., 2002), and topography of the two ridges that constitute the site is well mapped out on a 90 metre scale by the NASA's Shuttle Radar Topography Mission (SRTM) (see jpl.nasa.gov/srtm for details). A bird's eye view of the mountain complex is shown in Figure 3.2.

As well-constrained as the Degelen data set is, there is, nonetheless, a level of uncertainty related to varying explosion emplacements, as the Semipalatinsk test site was home to many 'non-standard' underground configurations (Adushkin and Leith, 2001). For example, nuclear detonations were either set up for an intentional release of radiation to the surface or were over-buried (i.e. placed deeper than the charge size required), or multiple-explosion scenarios were tested where two detonations happened simultaneously in either a single tunnel or in two diverging tunnels (Adushkin and Leith, 2001). These emplacement uncertainties translate to observed variability in waveforms, however the Degelen test site provides a well-defined setting for a systematic study of the effects of near-source interactions on teleseismic wavefields.

3.2 The AWE Blacknest Data

The Atomic Weapons Establishment (AWE) Blacknest holds a library of high-quality seismograms from four global UK-designed arrays installed in the early 1960s: the Eskdalemuir array (EKA) in Scotland, the Gauribidanur array (GBA) in India, the Warramunga array (WRA) in Australia and the Yellowknife array (YKA) in Canada. They were among the first experimental seismic arrays set up by the United Kingdom Atomic Energy Authority to study the feasibility of teleseismic detection and identification of nuclear explosions. Unlike many temporary arrays, they have been in a continuous operation for almost the entire period of underground nuclear testing, providing a valuable dataset of recordings with a high signal-to-noise ratio. Locations of the array stations, as well as their distance and azimuth to the Degelen test site are given in Table 3.1, while Figure 3.3 shows both the array and the Degelen test site locations on a map.

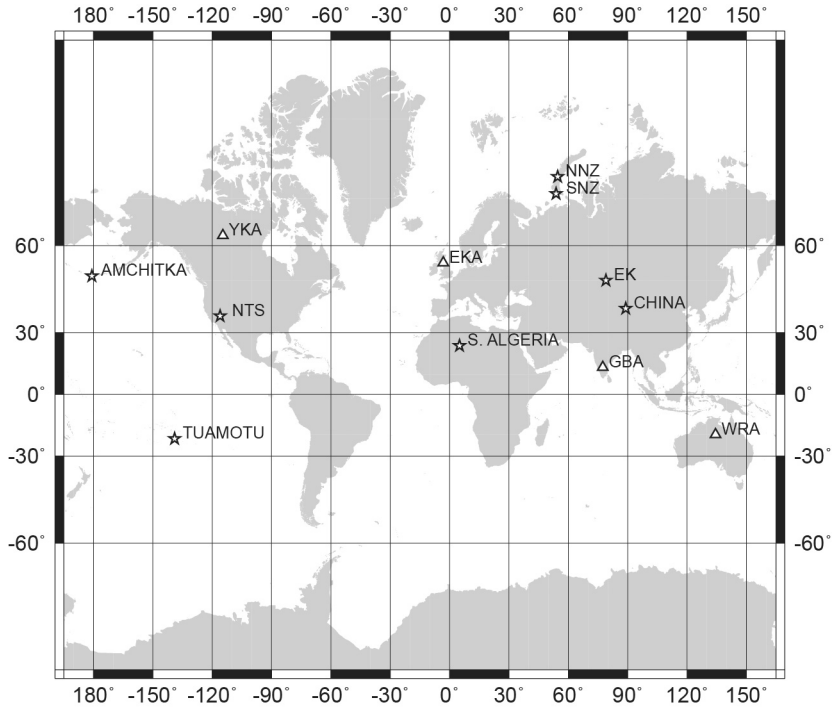


Figure 3.3: Locations of the four UK-designed arrays (triangles) and the main nuclear test sites (stars) used for underground nuclear tests by the USA, USSR, UK, France and China. EK stands for the 'East Kazakhstan' test sites that include the Degelen mountain complex. (Source: Douglas (2013), p. xli)

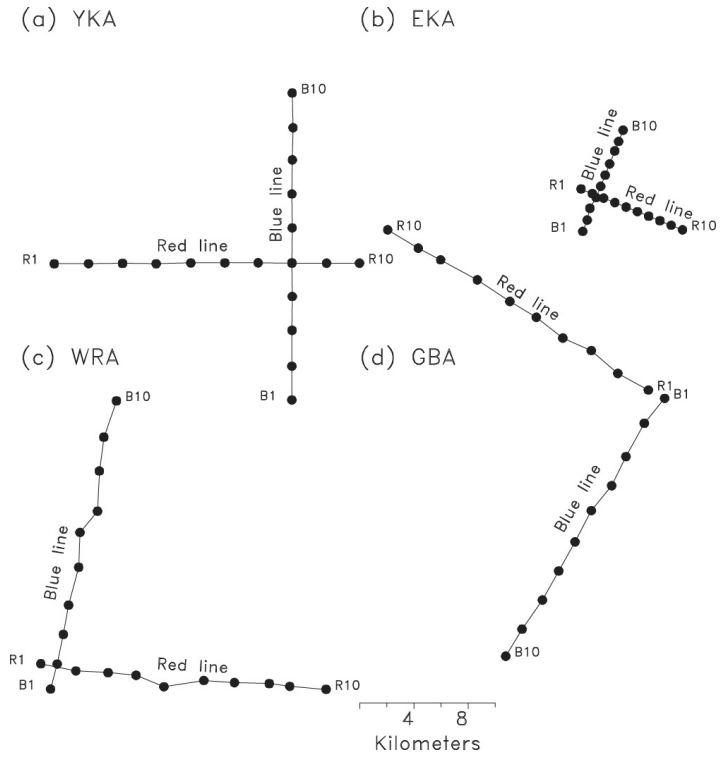


Figure 3.4: Layout of four UK-designed arrays used in this study: (a) Yellowknife (YKA); (b) Eskdalemuir (EKA); (c) Warramunga (WRA); (d) Gauribidanur (GBA). The line nearest to N-S is called the 'blue line', while the other one is called the 'red line'. The arrays have been installed specifically to monitor Semipalatinsk, and hence they all point in the direction of the site which reduces the signal-to-noise ratio. Such UK-designed arrays are usually described as medium-aperture. (Source: Douglas (2013), p.20)

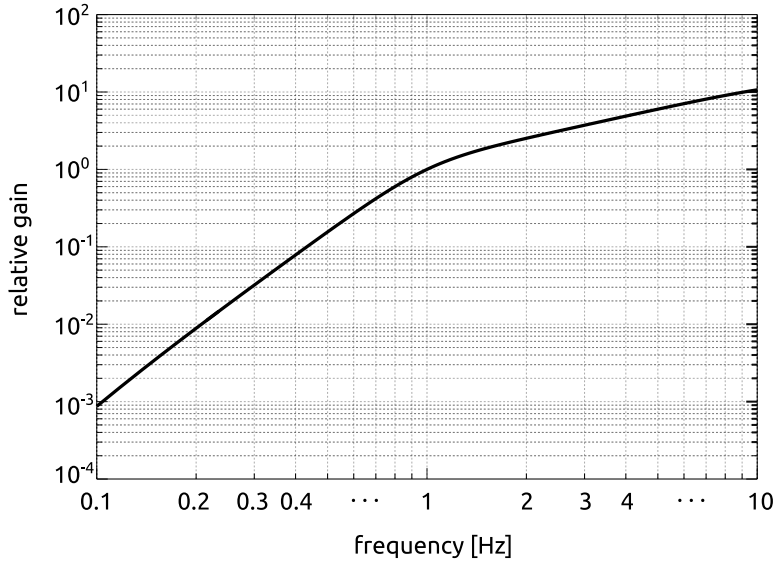


Figure 3.5: Instrument response of a UK-type array short-period seismometer to an impulse of ground displacement. The relative gain has been set to unity at 1 Hz. Above 1 Hz, the response is approximately proportional to velocity, and falls off roughly proportional to frequency to the third power below 1 Hz.

Figure 3.4 depicts typical orthogonal and L-shaped layouts of the UK array design.

Some of the more recent explosions were recorded directly as digital data, but the majority of signals were registered on analogue magnetic tapes and then transcribed onto library tapes. The data was later digitised at 20 samples per second, the sampling rate used for digital recordings. Recently AWE Blacknest started anew the processing of its raw data in an effort to standardise its databases and eliminate the variability introduced by various analysts over the years. For each explosion, the seismograms are beam-formed by delay-and-sum processing of individual channels using slowness derived from the IASP91 travel times. The channels were manually examined and data with inconsistencies – such as dropouts, excessive system noise, inverted signals, mislabelling or analogue-to-digital conversion problem, as well as recordings for simultaneous double explosions – were discarded (J Young 2019, personal communication, 25 July). As a result, good-quality data is available for all four arrays only for some of the events, with a large majority on three or fewer.

A conversion factor from digital counts to ground motion – in units of nm at 1 Hz – has been applied to all array seismograms. The seismograms, however, resemble the response of an impulse in ground velocity, as above 1 Hz the seismometer is effectively acting as a velocity transducer. Figure 3.5 shows the amplitude-frequency response of a UK-type array short-period seismometer to an impulse of ground displacement, with the relative gain set to unity at 1

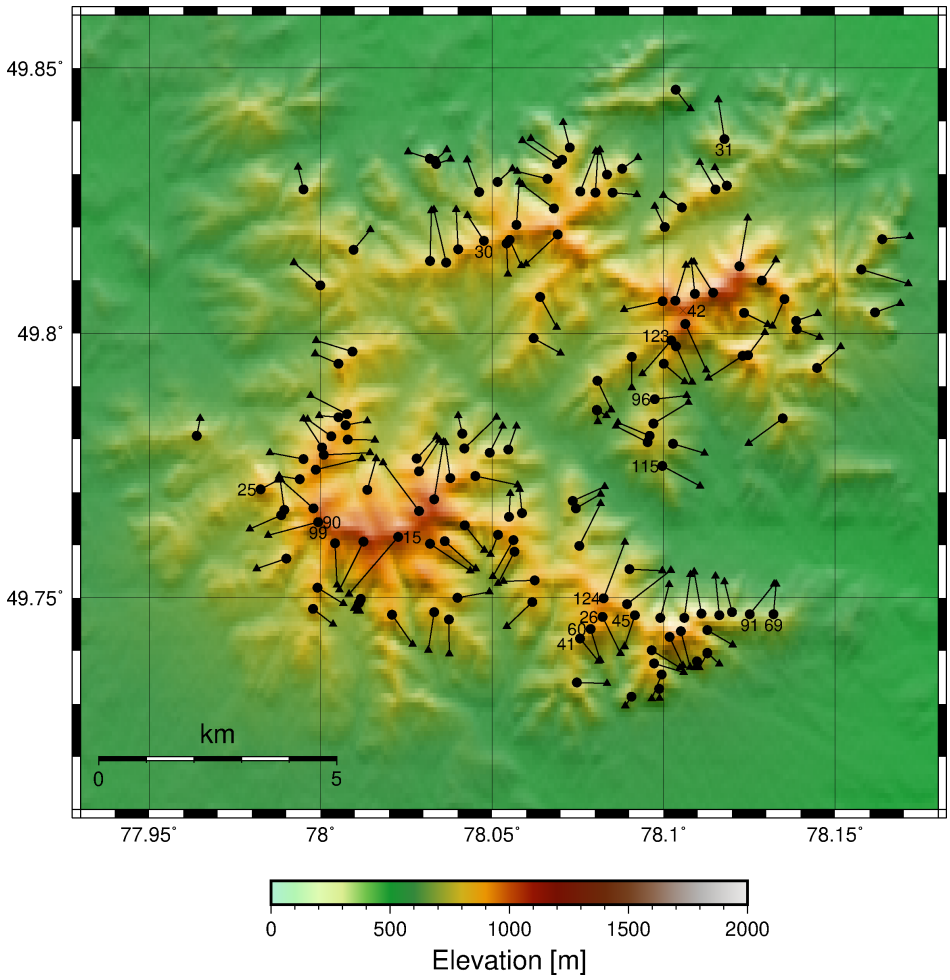


Figure 3.6: Map of the Degelen Mountain tunnel complex with adit portals (triangles) and ends (circles) for data available at AWE Blacknest. For events without adit portal and end locations, the available GPS coordinates are shown (crosses). All event locations referenced in the text and figures are labelled with their event ID as listed in Appendix C. Events labelled with IDs correspond to seismograms shown later in this chapter.

Hz. Above 1 Hz, the response is approximately proportional to velocity, and falls off roughly proportional to frequency to the third power below 1 Hz.

Official absolute source locations, origin times and yields were published by Bocharov et al. (1989) for 68 of the Degelen explosions from years 1961-1972. A later technical report by Leith (1998) consolidated the known ground truth information, including GPS-determined locations of adit entrances and correlations between explosions and the adits used. In an effort to refine the location capabilities at Degelen and determine more accurate ground truth, Trabant et al. (2002) analysed the information available for the post-1972 explosions not included in Bocharov et al. (1989). They estimated the source locations from detailed maps of the Degelen adit complex

Array	Total # of events	# of events with end-of-adit locations	# of events with GPS locations	# of events with known depths & yields
GBA	103	95	8	40
EKA	100	83	17	54
YKA	119	101	18	59
WRA	82	73	9	39

Table 3.2: Summary of the data for the Degelen mountain events available at AWE Blacknest and used in this study. The table lists the total number of recorded events available per array, the number of events with end-of-adit locations available, the number of events where only GPS locations are available, as well as the number of events with depths published by Bocharov et al. (1989).

(including portals and galleries, see Figure 3.6) provided to the Defense Threat Reduction Agency by the National Nuclear Center of Kazakhstan (KNNC) as part of the Degelen Mountain Tunnel Closure Program. They compared the GPS-determined portals in Leith (1998) to their portal-location estimates and noted a mean difference of 88 m between the two datasets. They have also reported a mean distance of 81 m between their inferred adit ends and the known absolute locations from Bocharov et al. (1989) for the first detonation in a given tunnel. When multiple tests took place in the same adit, the secondary charges were placed in galleries at varying distances from the adit end. Based on these observations, Trabant et al. (2002) defined the adit ends as the ground truth for events without official absolute locations. For primary detonations the error is likely to be around 100 m, whereas for secondary explosions it is limited by the length of the given adit (685 m on average).

Data from all four arrays span a total of 137 events for years 1964-1989. Table 3.2 summarises the events available per array and specifies the number of explosions with known depths and yields for each array, as well as the number of events for which the end-of-adit locations from Trabant et al. (2002) are available. The remaining events have the GPS coordinates form Leith (1998) as the most accurate location estimate. More detailed information, including the explosion dates, latitudes, longitudes, depths and yields from Bocharov et al. (1989), as well as the body-wave magnitude (m_b) estimates by Peacock et al. (2017) can be found in Appendix C. Peacock et al. (2017) provides values for all events considered in this study, and the m_b values have been re-processed using the maximum-likelihood method which removes any network bias.

3.3 The complexity of the Degelen waveforms

In Sections 3.1.2 and 3.2 discussing the test site itself and the data available at AWE Blacknest, respectively, we have highlighted the suitability of Degelen for a systematic study of the near-source effects on teleseismically recorded waveforms. In this section we take a first look at the seismograms, starting from a previous AWE analysis in Section 3.3.1, followed in Section 3.3.2 by an initial inspection of waveform differences present in the dataset.

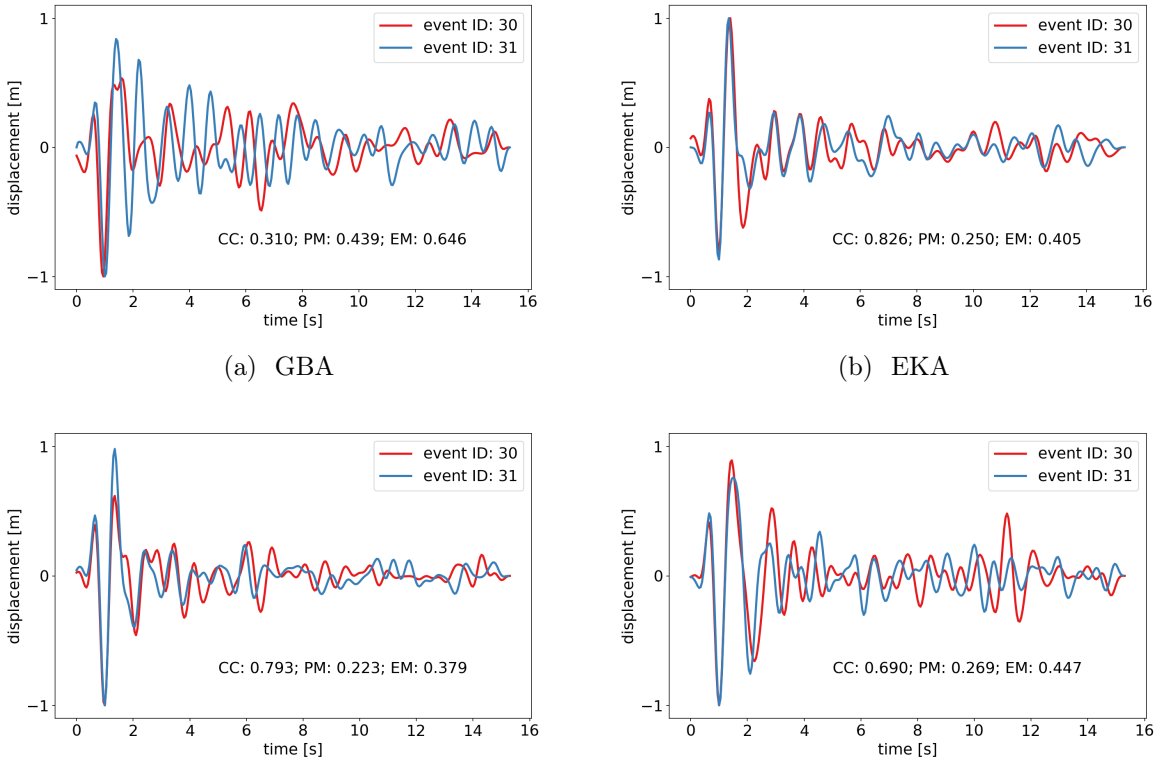
Since P wave complexity (simplicity) and the ratio of body- to surface-wave magnitudes are the main teleseismic discriminants of nuclear explosions (Bowers and Selby, 2009), we focus in the following sections on the Z components of the recorded signals. The cited discussions from McLaughlin and Jih (1988) and Marshall et al. (1992) also refer to the Z components only.

3.3.1 Previous AWE Blacknest data analysis: the P and pP arrivals

The data from the Degelen site explosions has been routinely analysed at AWE Blacknest. Initially, when little official information was available, the efforts focused on estimating the origin times, epicentres, depths and magnitudes of events (Marshall et al., 1984; Lilwall and Farthing, 1990). Later, after locations and yields for some of the 1961-72 test were published (Bocharov et al., 1989), AWE Blacknest released an extensive report (Marshall et al., 1992) where it reviewed its previous estimates, calibrated the relationship between the moment and yield for the Degelen site, and discussed the features and the general complexity of waveforms. Previously the Degelen site received little attention due to the relatively small sizes of tested charges in comparison with the Shagan River area of the STS (Marshall et al., 1992).

In their waveform analysis of 8 explosions with known source data, Marshall et al. (1992) look at the effects of depth of burial on the P wave signature and examine the broadening of the pulse as a function of yield. They note that at the GBA, EKA and YKA arrays, the broadband P pulse appears to broaden with increasing yield, but 'there is little, if any, evidence' of this effect at WRA. Seismograms from all four arrays exhibit a negative pulse following P which appears to get more pronounced with increasing source depth. The free surface reflection pP is discussed as a possible explanation for this change. However, given the high P wave velocities in granite, what appears to be the pP reflection arrives too late after P given the shallow explosion sources.

Comparing explosions from the Degelen mountain with those from the Konystan area of the STS with a flat surface, the authors of the report conclude that the sharp topography may be the reason for the Degelen seismogram complexity, producing more scattering and P wave focussing. Konystan explosions exhibit lower P wave amplitudes and noticeably less coda variation as a function of azimuth.



(a) GBA (b) EKA

(c) WRA (d) YKA

Figure 3.7: Example of observed waveform differences for two Degelen events on each of the arrays: (a) GBA, (b) EKA, (c) WRA and (d) YKA. Both events have an estimated m_b of 5.4 and are buried at 161 m (event 31) and 195 m (event 30). They are located over 5 km apart, both on the northern edge of the mountain complex. The event IDs correspond to those listed in Appendix C and the corresponding locations are shown in Figure 3.6. All waveforms are filtered at 2 Hz, and the P wave onset is at $t = 0$ s. CC, PM and EM stand for cross-correlation, phase misfit and envelope misfit, respectively.

3.3.2 A closer look at waveform differences

As reported by Marshall et al. (1992), the complexity of the Degelen seismograms is remarkable not only in comparison with the simple signals recorded from Konystan. Despite many common features – associated with the stations themselves and the propagation along the path that are common between all events – there are significant differences in the P wave and its coda within the Degelen dataset itself. Given that the sources are at most 15 km apart, Marshall et al. (1992) attribute the observed variations primarily to the source region. The degree of variability changes between the arrays and clearly depends on the take-off angle and azimuth, suggesting significant directional properties of the generation of the P wave and of subsequent scattering.

Figure 3.7 shows one example of two events of similar magnitude and illustrates waveform

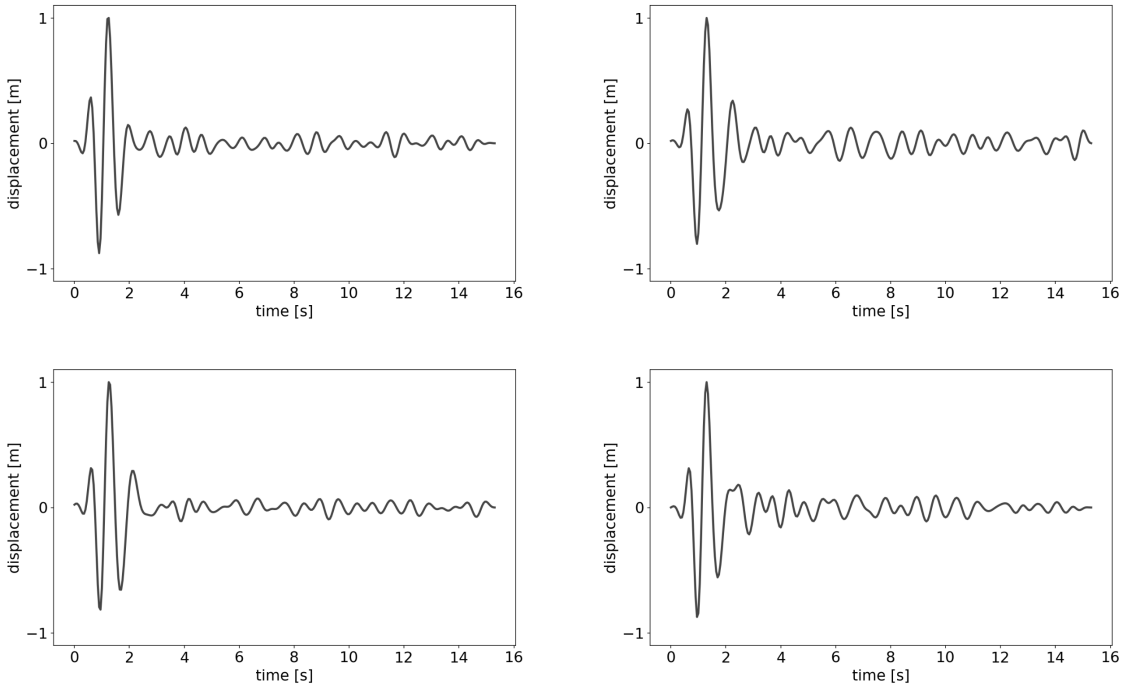


Figure 3.8: Examples of waveforms recorded at GBA from explosions at the Konystan test site with a flat surface. Konystan is less than 40 km away from the Degelen mountain and the path and receiver effects are expected to be very similar, yet the Konystan waveforms remain simple. All waveforms are normalised by the maximum amplitude, filtered at 2 Hz, and the P wave onset is at $t = 0$ s.

differences observed at the GBA, EKA, YKA and WRA arrays for frequencies up to 2 Hz. The GBA waveforms for the shortest epicentral distance (Figure 3.7a) show the highest level of complexity and, as will be shown later, it is true for the entire dataset. The wavefields are thought to be influenced by depth of burial, explosion yield, scattering off surrounding topography and off strong structural heterogeneities, as well as by non-linear effects. The latter can include an explosion-induced tectonic strain release, a collapse of the cavity, as well as spall, which takes place when the layers of rock above the explosion are separated and sent in free-fall to then reconnect in a "slapdown" (Rodean, 1981). Moreover, Ben-Zion and Ampuero (2009) have recently suggested that non-linear changes in the properties of the material in close vicinity of the source can also lead to seismic radiation.

For comparison, Figure 3.8 shows four examples of explosions located at the Konystan area of the STS with a flat surface and recorded at GBA. Given the proximity of the two test sites (< 40 km), the path and receiver effects are expected to be very similar. However, even the GBA waveforms – most complex for the Degelen test site – remain significantly simpler for Konystan explosions.

It should be noted that McLaughlin and Jih (1988) observe a similar variation in waveforms from events at the French southern Sahara test site. They discuss the common dominant characteristics of P-coda for stations at similar take-off angles and azimuths that may be attributed to direction-dependent scattering in the source region. They also analyse the variability of the 'pP arrivals' – for most stations and events there is no negative polarity pulse, while some waveforms 'suggest greatly delayed or multi-pathed pP pulses'. From 2-D numerical experiments McLaughlin and Jih (1988) conclude that a rough surface topography may be responsible for these observations, causing wave scattering, focussing and defocussing. They remark that non-linear interactions with the free surface near the source region – such as aforementioned spall and slapdown – could also generate such multiple, broadened and delayed 'pP' arrivals, as well as other secondary phases.

3.4 An analysis of the Degelen dataset

Given difficulties in modelling non-linear interactions, understanding and quantifying their contribution to generation of teleseismic waves is challenging to test. The large Degelen dataset, however, can be explored for wavefield variations as a function of depth of burial, explosion magnitude and source location in the test site, all of which can be further tested with hybrid modelling.

In this section we first look at the waveform noise levels and discard those for which the signal is obscured by noise. We then consider the global characteristics of the dataset and evaluate differences between the arrays to understand the overall variations. Following this general anal-

ysis, we investigate the degree of signal similarity as a function of depths and magnitudes, and examine particular waveform features that vary as source locations 'move' along the two ridges of the Degelen mountain complex. With the results suggesting that the most significant signal variations depend on the source location, we then attempt to measure the signal complexity as a function of location and topographic skewness.

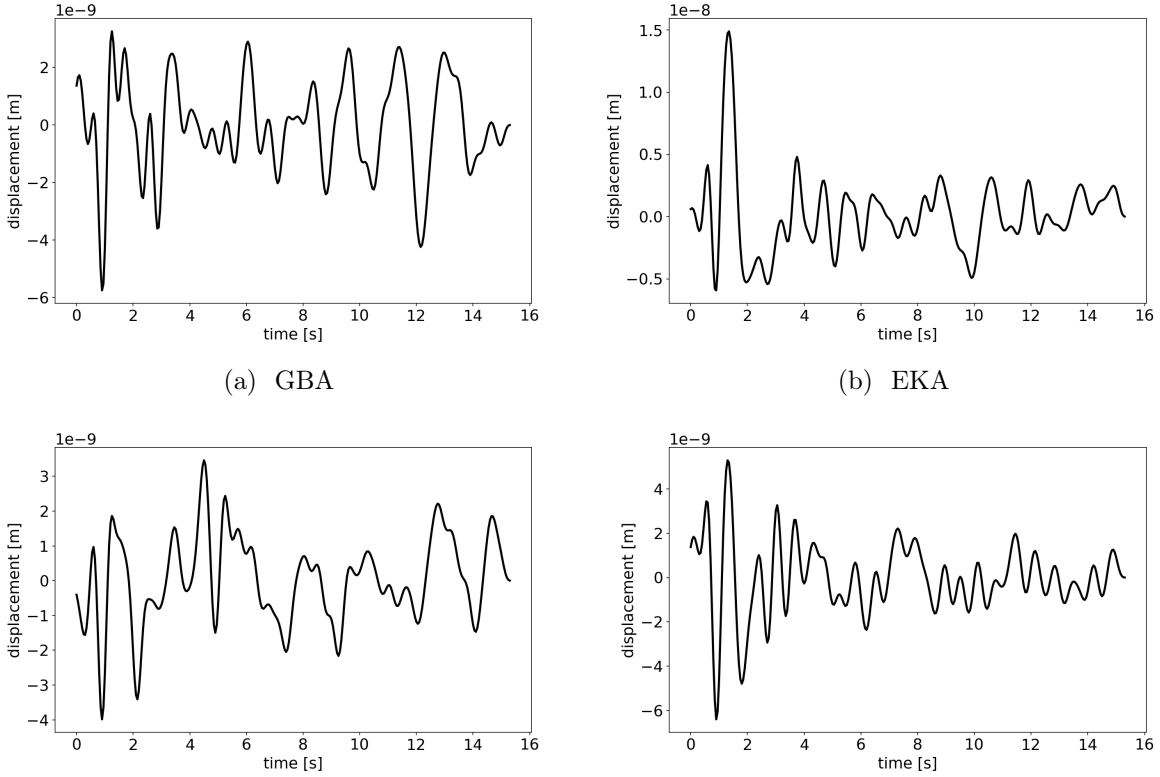
In the following, the explosions are referred to by their 'event ID' from Appendix C. As stated previously, given the focus on teleseismic P waves we consider only the Z components of the seismograms. Since array signal processing techniques are steered to the P wave energy leaving the source region, and later arriving waves are suppressed, we consider only the first 15 seconds after the first arrival. We refer to the entire wave-train after the direct P arrival as the P wave coda – our definition thus includes both single and multiple scattering off heterogeneities and discontinuities around the source, along the path, and under the receivers. All waveforms are cut with the P wave onset at time $t = 0$ s and are filtered with a zero-phase low-pass Butterworth filter at 2 Hz. According to the Mueller-Murphy model (Mueller and Murphy, 1971), a 15 kt explosion in granite at 170 m depth – an event representative of those at Degelen – has a moment of 10^{16} Nm and a corner frequency of 1.5 Hz, and hence we consider the 2 Hz filter frequency sufficiently high for all events.

3.4.1 The general signal characteristics

Filtering out low magnitude events to reduce noise

Due to our interest in the subtle similarities and differences of the P wave and its complex scattering, it is of paramount importance to consider signals where noise does not significantly obstruct lower energy arrivals. The signal-to-noise ratio can be defined in a multitude of ways and is difficult to quantify accurately, but generally the smaller magnitude events exhibit higher relative noise levels. Figure 3.9 shows an example of a signal for a small explosion of body-wave magnitude $m_b = 4.64$ at all four stations where the noise significantly distorts the arrivals despite the noise-reducing array processing.

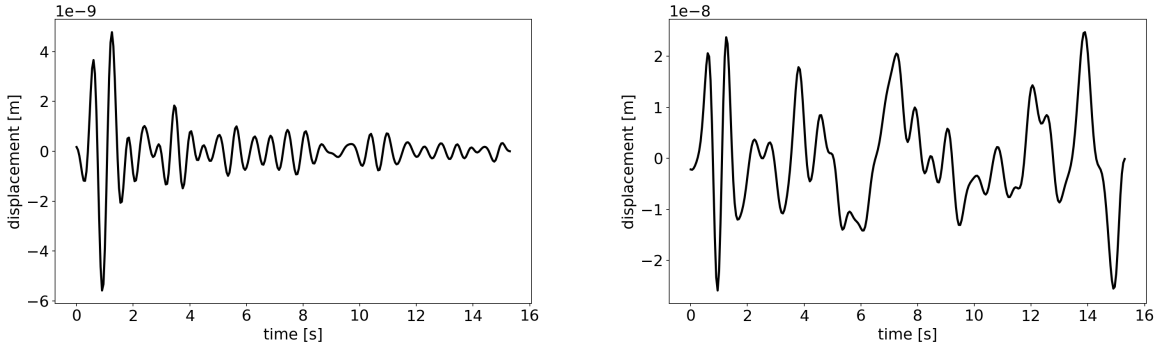
Given that the event magnitude is a good proxy for noise levels, we consider 94 (out of 137) events with m_b estimates above 4.95 in order to reduce the effect of noise in the general analysis. The threshold has been picked after a detailed visual examination of seismograms at all arrays. As may be expected, such threshold-based filtering is not flawless, since waveforms for the same event may exhibit varying noise levels at different stations. The signal quality depends not only on the source magnitude, but also on the array itself and on the conditions on the day of the recording. Hence, some low-magnitude and high-quality recordings are rejected, while certain higher magnitude and lower signal quality signals are retained (see Figure 3.10 for examples). Generally, however, the threshold of $m_b = 4.95$ significantly reduces the overall noise level in the



(a) GBA (b) EKA

(c) WRA (d) YKA

Figure 3.9: Example of noisy signals for a Degelen event of $m_b = 4.64$ (event ID: 115) on each of the arrays: (a) GBA, (b) EKA, (c) WRA and (d) YKA. The noise masks the P wave onset at $t = 0$ s, significantly perturbs the subsequent arrivals, and thus complicates direct comparisons of waveform features. The corresponding event location is shown in Figure 3.6.



(a) YKA, event ID: 25, $m_b = 4.67$ (b) EKA, event ID: 45, $m_b = 5.07$

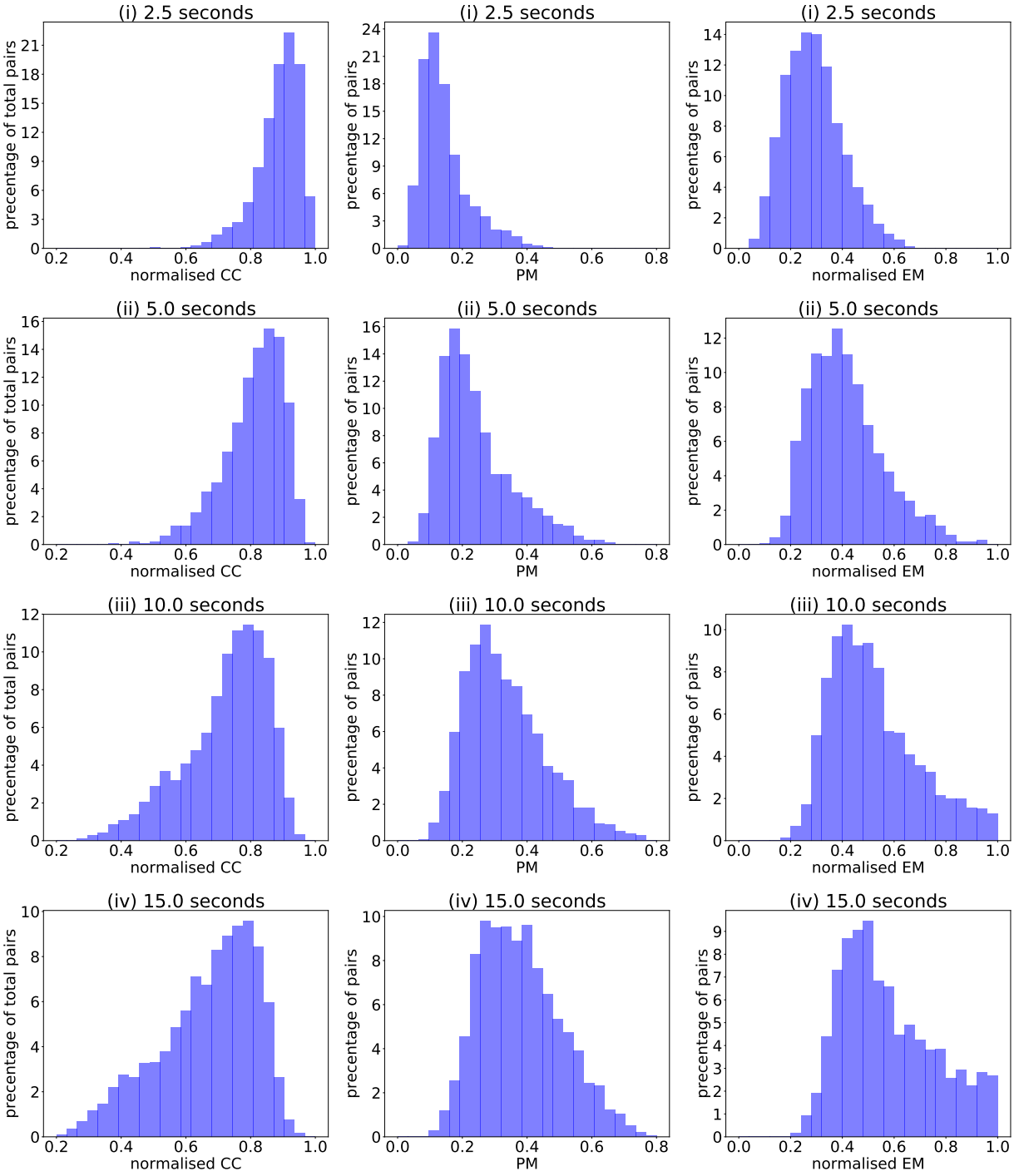
Figure 3.10: Like any threshold-based classification, selecting events with moment-magnitude $m_b > 4.95$ is not flawless. Signal quality depends not only on the size of the explosion, but also on the array conditions on the day of the recording. (a) illustrates a good-quality recording of an event of m_b below the threshold of 4.95 that has been rejected, while (b) shows a noisy recording of an event of $m_b > 4.95$ that has been retained. Overall, however, a visual examination proved that the m_b is a good proxy for signal quality. The threshold of 4.95 removes the majority of the noisy recordings as in (b) and significantly improves the quality of the dataset. The corresponding event locations are shown in Figure 3.6.

dataset.

An array-by-array degree of waveform similarity

To better understand the overall degree of waveform variability in the dataset, we analyse the normalised cross-correlation (CC) and envelope and phase misfit (EM and PM, respectively, as defined by Kristeková et al. (2009)) coefficients for all pairs of events on every array. By considering time windows of increasing length after the first P arrival, we examine the general effects of the P wave coda on waveform similarity. The normalised CC explores the general properties of the signal without the magnitude-dependent amplitude differences, while the PM and EM decouple the phase and envelope effects and characterise the nature of the differences. The PM looks at the phase coherence of traces of original amplitudes, while the EM reflects the envelope similarity of normalised traces. The misfit coefficients indicate levels of disagreement between two signals and are designed for comparisons of seismograms of similar nature.

Figure 3.11 shows distributions of the CC, PM and EM coefficients for the EKA array for time windows extending to 2.5, 5, 10 and 15 seconds after the onset of the P wave. The level of similarity between waveforms decreases with an increasing time window, suggesting that the first P arrival is highly consistent. The coda waves, on the other hand, exhibit a larger variation that outweighs the similarity of the first P wave arrival for the longer time intervals. However, it



(a) CC (b) PM (c) EM

Figure 3.11: Normalised cross-correlation (CC), phase misfit (PM), and normalised envelope misfit (EM) coefficients as a percentage of the total number of pairs for events at EKA for time windows of (i) 2.5, (ii) 5, (iii) 10 and (iv) 15 seconds after the onset of the P wave. See Figures D.1-D.3 in Appendix D for analogous figures for the other arrays.

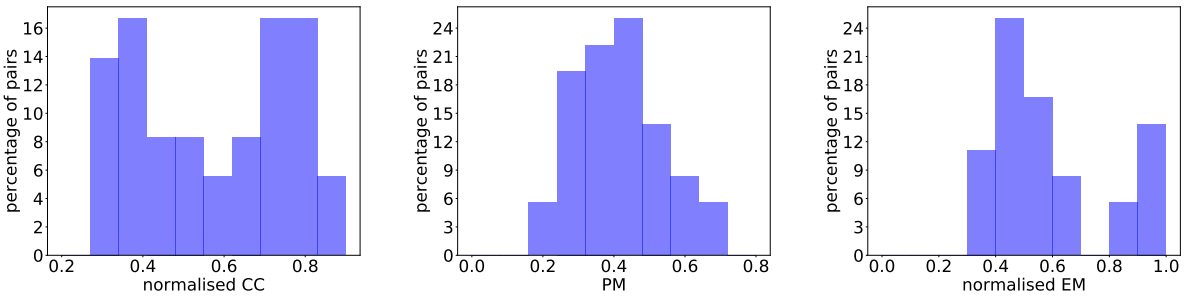


Figure 3.12: The normalised CC, PM, and normalised EM coefficients at the EKA array as a percentage of the total number of pairs for events with depths between 180 and 200 metres. Comparing the distributions of the chosen subset to the distributions of the total dataset (Figure 3.11), we conclude that depth is not the primary factor explaining similarities and differences in the Degelen dataset. See Figure D.4 in Appendix D for analogous figures for the other arrays.

is also true that the CC coefficients and PM & EM coefficients remain high and low, respectively, for a small number of pairs for all time windows considered, indicating a high level of similarity in the P wave coda for those seismograms. Similar properties can be observed for the coefficients of the GBA, YKA and WRA seismogram pairs (see corresponding Figures D.1-D.3 in Appendix D).

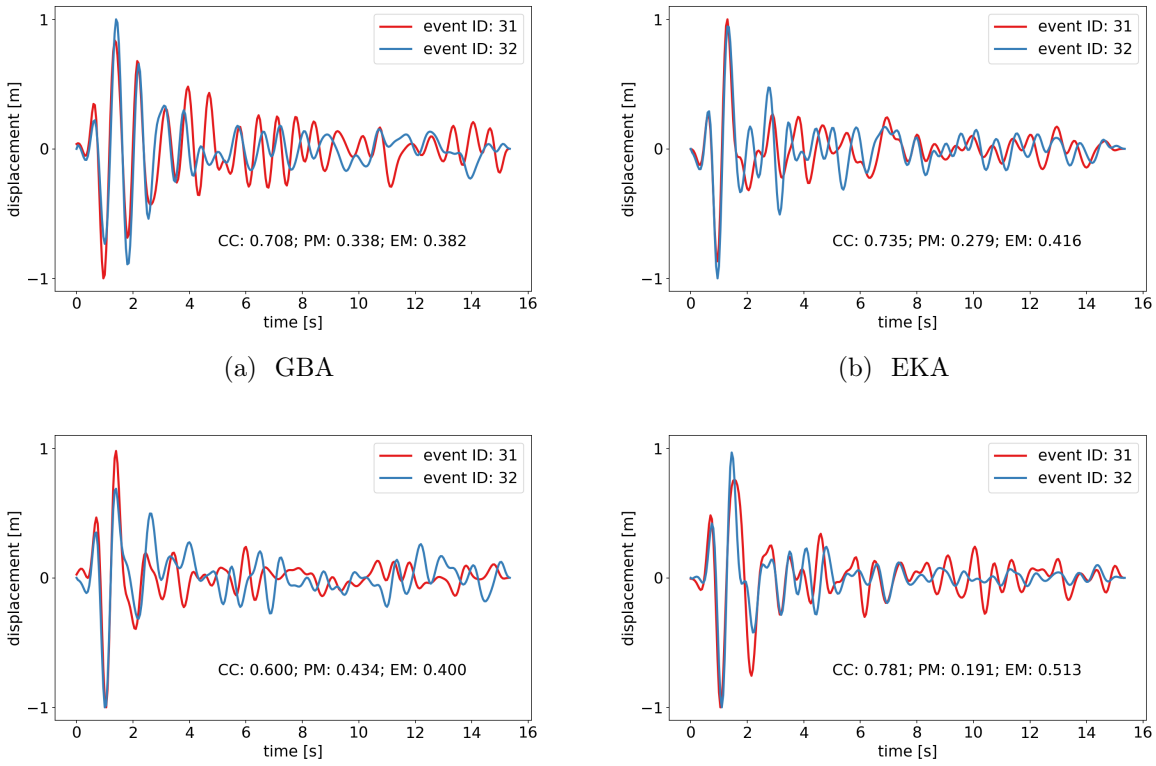
Distributions of coefficients also hint at varying noise levels at the arrays that are consistent with our manual analysis of seismogram quality mentioned in section above. The YKA array is known to be the quietest and most reliable of the four, and generally higher CC and lower PM and EM (for all time windows) relative to the other stations confirm that.

3.4.2 Waveform variations with depth and body-wave magnitude

We have first analysed waveform similarities and complexities with respect to depth of burial and magnitude of explosions. Except for unusual testing circumstances, the depth of burial depends on the size of the charge (Murphy, 1977), and the body-wave magnitude (m_b) values also tend to reflect that linear relationship with depth. Nonetheless, we have decided to consider the two separately, as the depth-yield relationship is only approximate. Moreover, the official depths and yields are known for only a fraction of all events, while the body-wave magnitude values have been calculated for the majority.

Depths

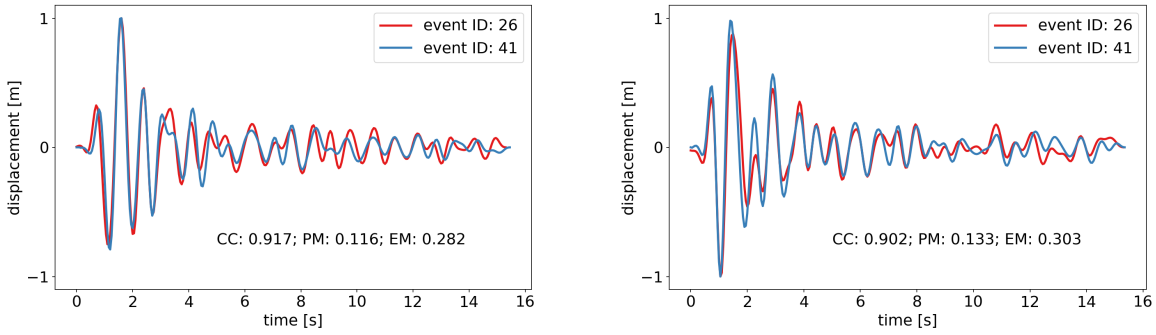
Figure 3.12 shows normalised CC, normalised EM, and PM at the EKA array for a 15 second time window for pairs of events with known depths in the range of 180–200 m. In case of a strong dependence of waveform features on depth, we would expect those events to represent



(a) GBA (b) EKA

(c) WRA (d) YKA

Figure 3.13: Example of observed waveform differences for two Degelen events at the same depth located over 10 km apart on two sides of the valley (160 m for event number 31 ($m_b = 5.40$) and 161 m for event number 32 ($m_b = 5.34$)). The first arrivals are relatively similar on all stations, but subsequent arrivals vary significantly. The signals have been normalised to correct for the difference in magnitude. The corresponding locations are shown in Figure 3.6.



(a) GBA (b) EKA

Figure 3.14: Example of similar waveforms for different depths of firing at (a) GBA and (b) EKA. Explosion 26 ($m_b = 6.08$) was buried at 241 m and explosion 41 ($m_b = 5.47$) at 162 m, and the two were located within 500 m from each other. Despite a reported 80 m difference in depth, the coda waves are highly consistent. The signals have been normalised to correct for the difference in magnitude. The corresponding locations are shown in Figure 3.6.

the pairs with:

- the highest values of CC coefficients in Figure 3.11a(iv);
- the lowest values of PM coefficients in Figure 3.11b(iv);
- the lowest values of EM coefficients in Figure 3.11c(iv).

This, however, is not the case, and limiting depths to a specific range does not seem to drastically improve the overall degree of similarity in the subset: many of the CC coefficients in Figure 3.12 are below 0.7, and most of the PM and EM coefficients are above 0.3. Choosing different depth ranges produces similar inconclusive outcomes. The same is valid for the other three arrays (see Figure D.4 in Appendix D).

A visual examination of waveforms confirms this lack of a straightforward dependence on depth. Figure 3.13 shows an example of two signals from events of the same depth (160 and 161 meters as reported by Bocharov et al. (1989)), but located on two sides of the valley, about 10 km apart. Figure 3.14, on the other hand, shows events at different depths – of 241 m and 162 m – located within 500 m from each other. The same-depth events manifest significant differences after the first arrival, while events with different burial depths located in close proximity can generate nearly identical coda waves. Since the dataset is not complete (official depths are known only for a fraction of events and most events are available on three stations or fewer), examples of the same-location events with significant differences in the depth of firing were scarce and thus only two arrays are shown in Figure 3.14.

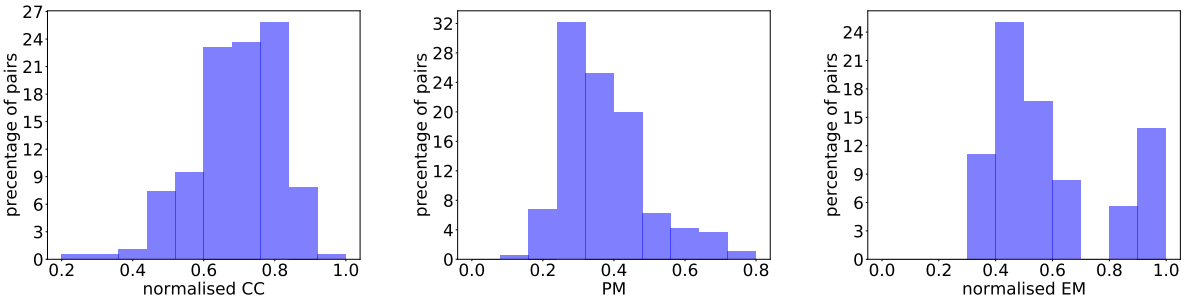
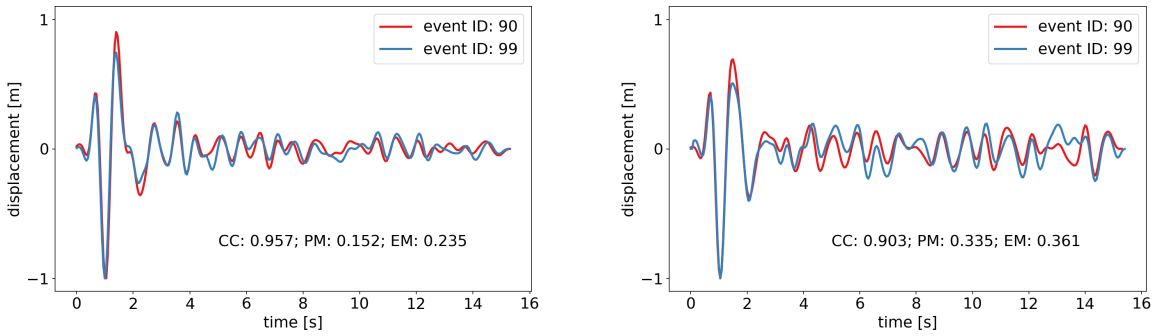


Figure 3.15: Normalised CC, PM, and normalised EM coefficients at the EKA array as a percentage of the total number of pairs for events with body-wave magnitudes between 5.3 and 5.5. Comparing the distributions of the chosen subset to the distributions of the total dataset (Figure 3.11), we conclude that m_b is not the primary factor explaining similarities and differences in the Degelen dataset. See Figure D.5 in Appendix D for analogous figures for the other arrays.



(a) YKA (b) WRA

Figure 3.16: Example of similar waveforms for different body-wave magnitudes at (a) YKA and (b) WRA. Explosion 90 has m_b estimated at 5.73, while explosion 99 at 5.23 (depths are unknown for both). Both explosions took place in the same tunnel, with the location difference of 500 m or less. Despite the difference of 0.5 in m_b, the coda waves are highly consistent. The signals have been normalised to correct for the difference in magnitude. The corresponding locations are shown in Figure 3.6.

Body-wave magnitudes

We now consider events with values of body-wave magnitudes between 5.3 and 5.5 as computed by Peacock et al. (2017). As before, the normalised CC, normalised EM, and PM for pairs of signals in the subset do not suggest a high level of waveform similarity (Figure 3.15 for the EKA array and Figure D.5 in Appendix D for the other arrays), and choosing different m_b ranges does not improve the results.

Again, visually examining the waveforms verifies this conclusion. Figure 3.7 in Section 3.3.2 shows an example of two signals from events of the same m_b, but at different locations, over 5 km apart. Figure 3.16, on the other hand, shows events of a different m_b in the same tunnel, likely significantly less than 500 m apart. The same-m_b events manifest significant differences, while the events of different m_b's can generate similar arrivals. Since the dataset is not complete, no pair of same-location events of different m_b's was available at all four arrays, and the pair shown has been chosen to complement the stations from the different-depths example in Figure 3.14.

3.4.3 Waveform variations with location in the test site

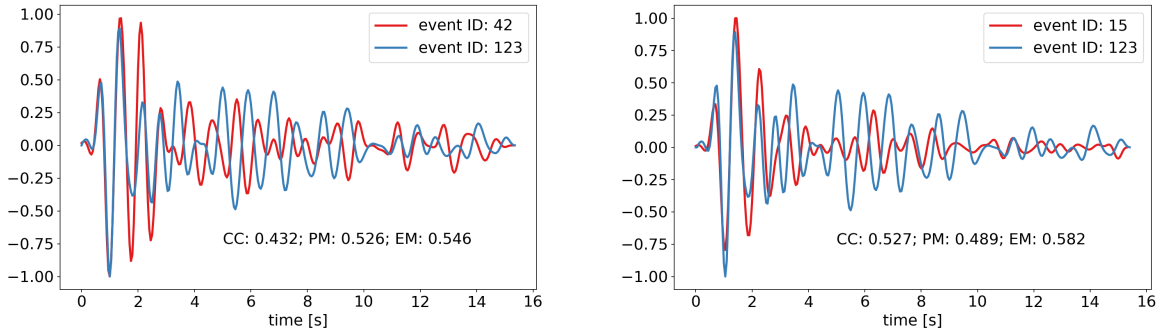
In the previous section we have not been able to establish a clear dependence of Degelen waveforms on explosion depths and body-wave magnitudes. Our visual examination of the data, however, suggests that wave characteristics change with respect to the source location in the massif. As shown in Figure 3.14 and in Figure 3.16, events of different depths or of different body-wave magnitudes that lay in close proximity can exhibit similar waveform features. We have, therefore, thoroughly analysed the events with respect to their location in the test site.

Failure of automatic spatial clustering

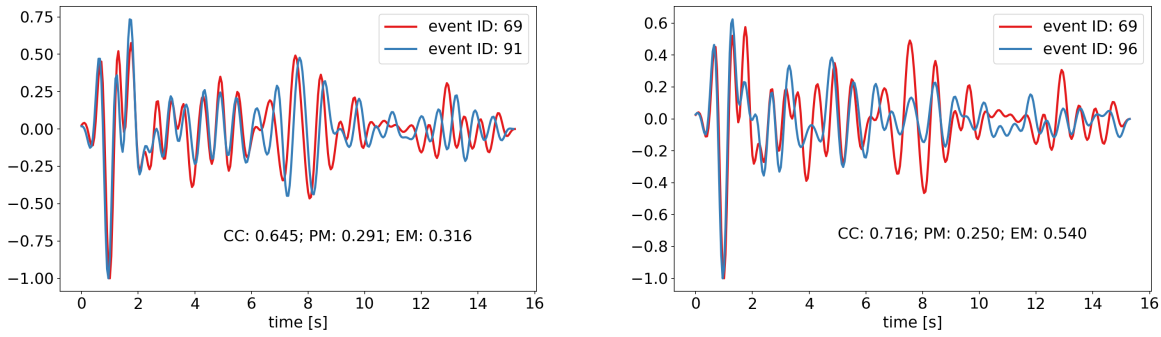
Given the overall high similarity of signals and varying noise levels, threshold-based spatial clustering on cross-correlations and phase and envelope misfits proved unsuccessful. Rather than identifying and quantifying shared features, the calculated coefficients average out the variations in a given time window and a single coefficient often masks the characteristic complexity of a pair of signals. Moreover, a small lateral shift in the location of the source can result in perturbations of the phase, or of the envelope, or both. Devising meaningful clustering that extracts spatial variations would need to account for irregular changes in multiple coefficients and hence proved ineffective.

Figure 3.17 illustrates the issue on an example of two sets of waveforms at GBA, although the same effect is observed at all arrays. Even though the close lying events (under 1 km in the given examples) generate coda waves with related features, the coefficients do not immediately reveal that high similarity relative to other pairs of events and simple thresholding on specific values leads to misleading event clusters.

One could argue that time- and frequency-dependent PM and EM would highlight the subtle differences in phase and amplitude. However, such misfits demand a thorough analysis for every event pair. Given the number of pairs of events for each station – ranging from 3321 for WRA to 7021 for YKA – to benefit from the additional time- and frequency-dependent information one would require highly sophisticated clustering methods beyond the scope of this work.



(a) 42 and 123: 1 km apart; 15 and 123: 15 km apart



(b) 69 and 91: 500 m apart; 69 and 96: 5 km apart

Figure 3.17: Example of CC, PM and EM coefficients at the GBA array which illustrate that clear-cut thresholding is inadequate and hence simple threshold-based clustering of close lying events with similar features fails. The same effect can be observed at the other three arrays. The seismograms have been normalised to correct for magnitude differences. The corresponding locations are shown in Figure 3.6.

(a) Events 42 (m_b = 5.89, depth = 290 m) and 123 (m_b = 5.53, depth unknown) lie 1 km apart, close by on the north-eastern peak of the Degelen complex. Event 15 (m_b = 6.07, depth = 294 m), on the other hand, is located 8 km away from 123 in the south-western corner on the opposite side of the valley. The values suggest an overall low level of similarity for both pairs, with the CC lower and the PM higher for the close lying pair 42-123 (left figure of (a)) relative to the pair 15-123 (right figure of (a)). Upon visual examination, however, it becomes evident that events 42 and 123 have remarkably related coda waves, although strongly shifted.

(b) Events 69 (m_b = 5.27, depth unknown) and 91 (m_b = 5.37, depth unknown) lie on the eastern end of the southern ridge of the massif less than 500 m apart. Event 96 (m_b = 5.28, depth unknown) is located 5 km away, close to the highest peak of the northern ridge. The features of the close lying pair 69-91 (left figure of (b)) are in extraordinary agreement, but the coefficient values do not reflect that similarity due to a slight shift of later arriving waves. Seismograms 69 and 96 (right figure of (b)) have fewer common features and exhibit strong amplitude differences that result in a high EM, but the CC value remains high and the PM low.

Manually extracted spatial variations

Due to the subtleties discussed in Figure 3.17 that render automated spatial clustering on coefficient thresholds inefficient, the waveforms for each array have been visually analysed to extract the location-dependent features. Figures 3.18-3.21 illustrate variations of the signals at each array as the source locations 'move' along the two ridges of the Degelen massif – the change is nearly continuous, with specific features progressively emerging and fading away as location changes. Examining the spatial pattern in Figures 3.18-3.21 we observe the following:

- As previously noted, events in close lateral proximity are related and often exhibit similar features both in the first arrivals (0–2 s) and in later arriving coda waves (2–15 s).
- However, waveforms appear very sensitive to spatial variations and moving the source location by just a few kilometres can result in visible differences. The general characteristics of the signal are preserved, but the coda waves shift, some arrivals disappear or new arrivals appear, or parts get (de)amplified. In other words, some of the energy seems to be delayed or arrive earlier as a result of the location change. It should also be noted that subtle changes in features due to small variations in source location may be amplified by the different noise levels at a given station on different explosion dates.
- For events in two distinct locations (e.g. at 5 km or further), coda waves after the first 3 seconds can look drastically different with nearly no common features. However, the opposite can also be true, and is mostly the case in regions where topography is less pronounced.
- The degree of variation differs between the stations, suggesting a dependence of the teleseismic signature of the location in the massif on the azimuth and on the take-off angle. However, the change in features as sources move along the ridges is visible on all stations. The varying waveform shapes between the stations suggest that this location-dependent near-source scattering is strongly directional.
- Normalisation of traces is necessary to correct for differences in magnitudes, but it should be noted that such scaling could be exacerbating or masking differences between seismogram pairs. This issue is further discussed at the end of this section (see Figure 3.22).

Given the complexity of the patterns illustrated in Figures 3.18-3.21 and the multitude of emerging features, in what follows we are going to discuss a single observation per station that highlights the intricacy of what we believe to be an influence of the source region on the Degelen waveforms. Notice that the effects discussed for a given station are not exclusive to that station and the same or similar effects may be present on the other arrays.

GBA: the delayed/shifted coda (Figure 3.18)

Plots marked A-D on Figure 3.18 depict events for which a change in location on the order of 0.4-0.8 km results mainly in a shift in the arrival of coda waves and no other major perturbations of waveform characteristics are present. On all four plots one can compare the seismograms nearly wiggle-by-wiggle, but the coda of one of the traces accumulates a delay relative to the other. For the pairs on plots A and B, the delay is relatively small and all arrivals are present in both signals. For the pairs on plots C and D, on the other hand, the delay in energy is so large that an additional arrival becomes apparent on one of the traces around 3.5 seconds after the onset and causes an effect resembling a cycle skip for the subsequent arrivals in the coda.

EKA: the re-emerging feature (Figure 3.19)

Certain features emerge and vanish as source locations move along the ridge, and plots A-D illustrate such behaviour. A prominent fourth positive peak is present on all of the traces in A-D, and this strong amplification is not observed in other locations. That feature is not constrained to a single source location, hinting that the source position relative to the slope could be of paramount importance in the generation of certain arrivals.

YKA: the splitting of the second arrival (Figure 3.20)

Changes in location strongly affect the second positive arrival on all presented traces at YKA. In plot A all energy arrives as a single impulse on both seismograms, though one of the pulses is broader and of lower amplitude. Plot B indicates a further delay of some of the energy, and the second peak seems to nearly 'split' into two separate arrivals. This is also visible in plots C and D, where the events have coda in phase, but despite being in close proximity the character of the second arrival varies significantly. Finally, plot E shows events along a relatively long slope that have an identical – single and broadened – second positive arrival even though they are relatively far apart. It should also be noted that plots B-E show a very high waveform similarity in the coda.

WRA: close proximity, large variations (Figure 3.21)

For regions with sharp topography, a small shift in source locations can cause large variations in waveforms. All events on plots A-D are located within the same mountain, but the waveforms change significantly after the first arrival for traces on plots A-C – the second peak sometimes splits into a 'double peak' similar to the split discussed for the YKA array, and the later arrivals vary in amplitudes and wavelengths. The change, however, is not always so drastic and depends on the locations of events relative to each other and to the slope of the mountain. Plot D, for example, shows two events in the same mountain which exhibit a coda shift (similar to what is discussed for the GBA array) and amplitude variations only very late in the trace.

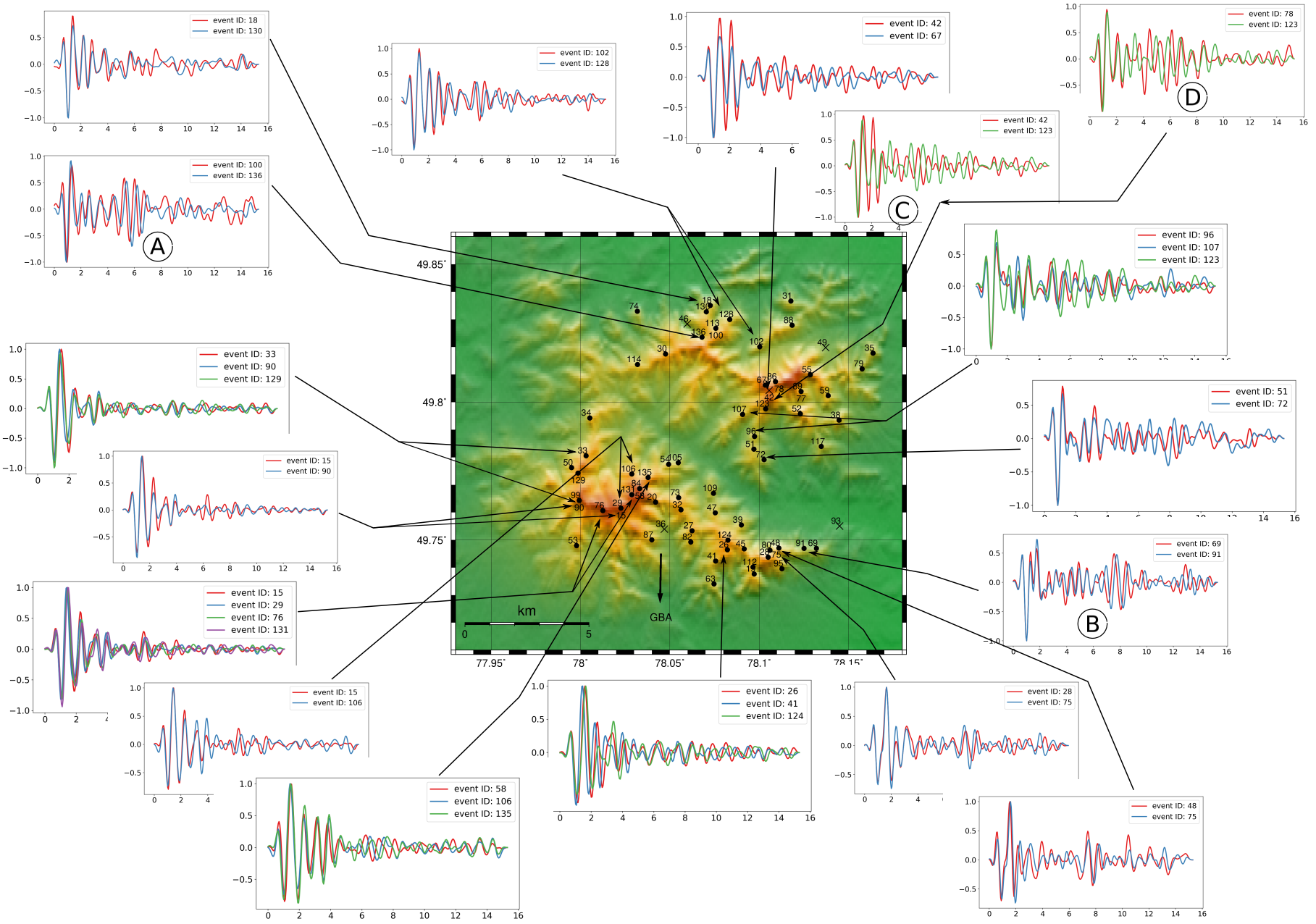


Figure 3.18: Location-dependent seismogram features at the GBA array. The direction from Degelen to GBA is indicated with an arrow. The adit ends corresponding to each event are plotted over the Degelen topography (circles). For events without adit-end locations, the GPS locations are shown (crosses).

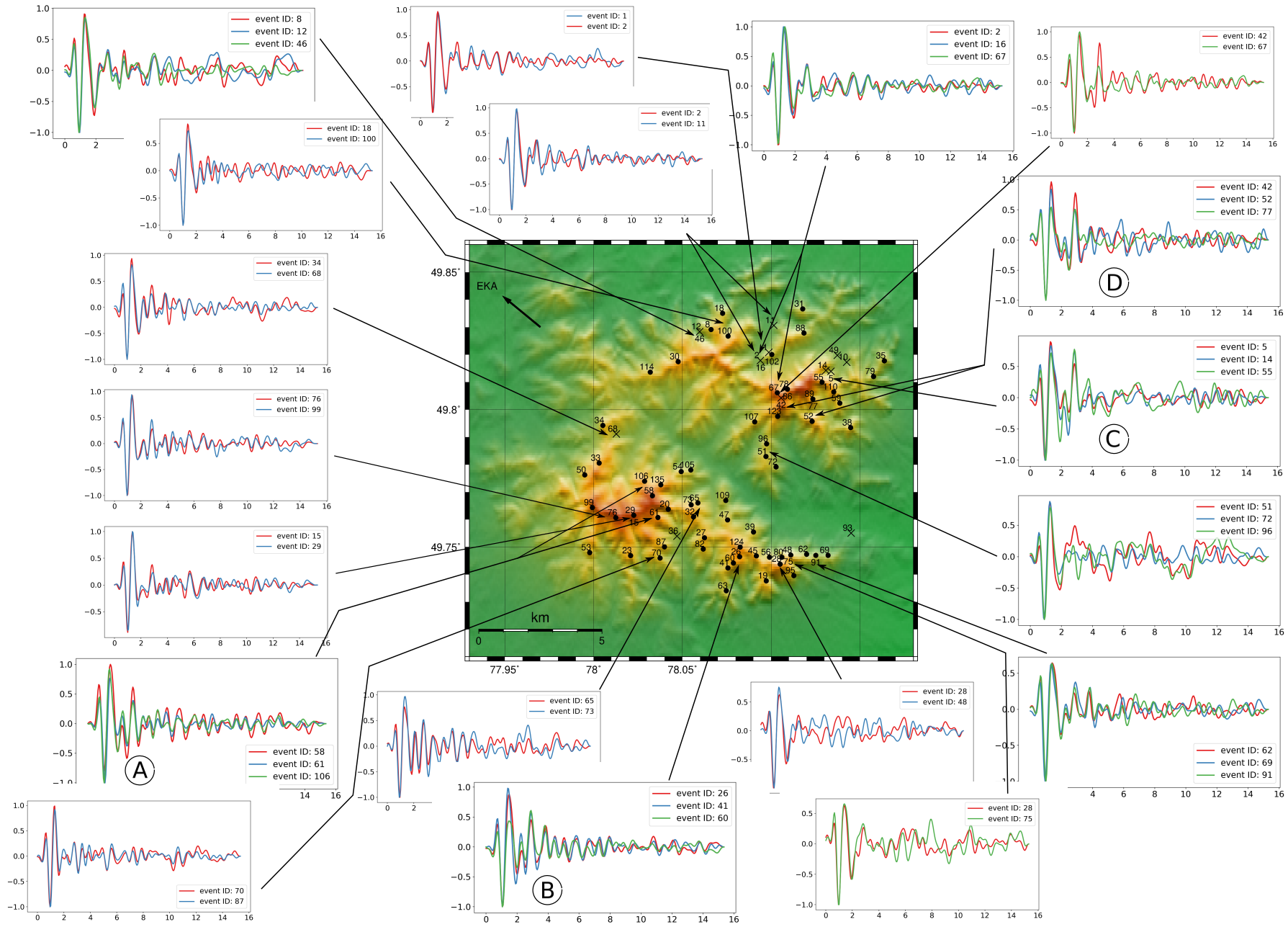


Figure 3.19: Location-dependent seismogram features at the EKA array. The direction from Degelen to EKA is indicated with an arrow. The adit ends corresponding to each event are plotted over the Degelen topography (circles). For events without adit-end locations, the GPS locations are shown (crosses).

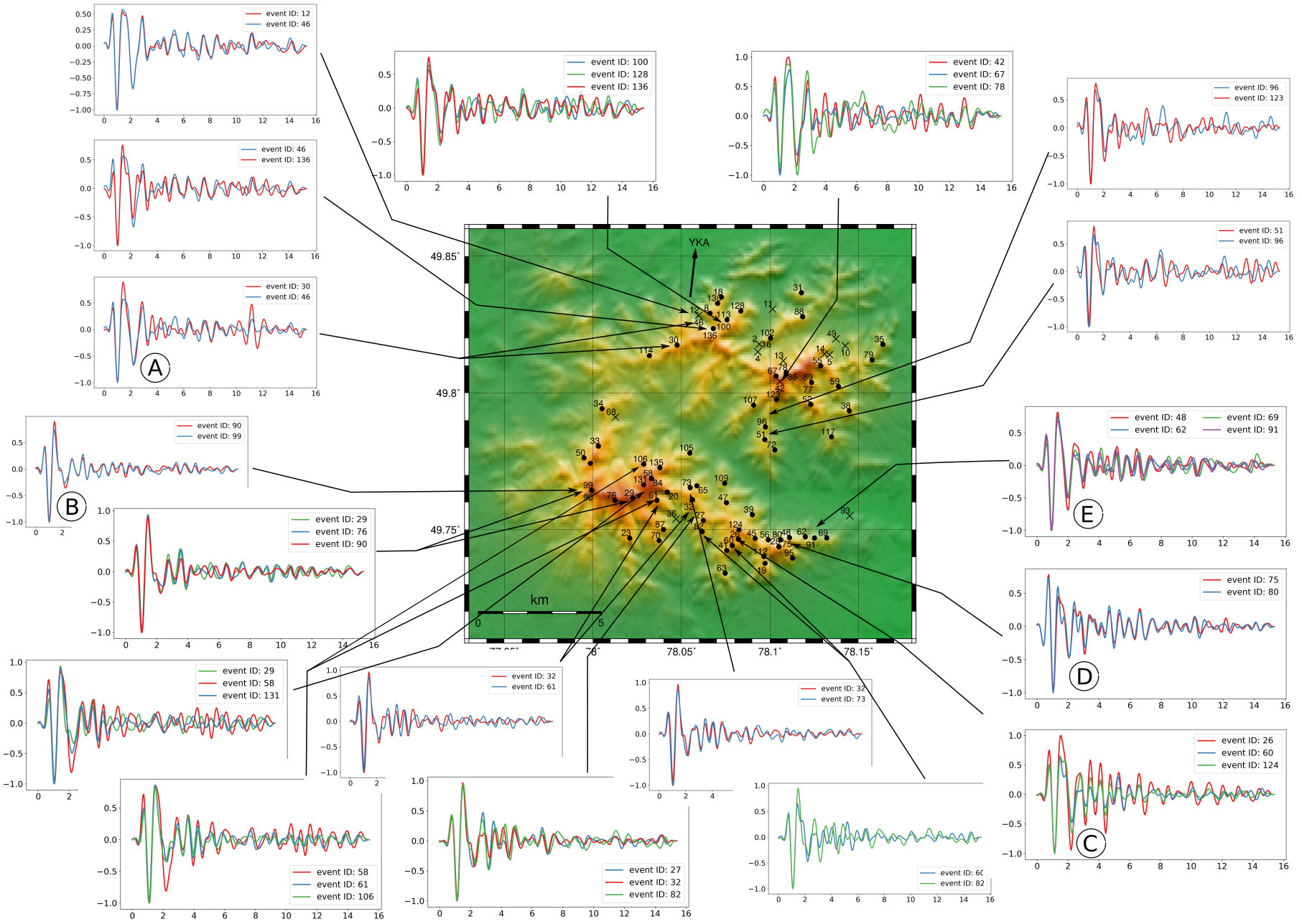


Figure 3.20: Location-dependent seismogram features at the YKA array. The direction from Degelen to YKA is indicated with an arrow. The adit ends corresponding to each event are plotted over the Degelen topography (circles). For events without adit-end locations, the GPS locations are shown (crosses).

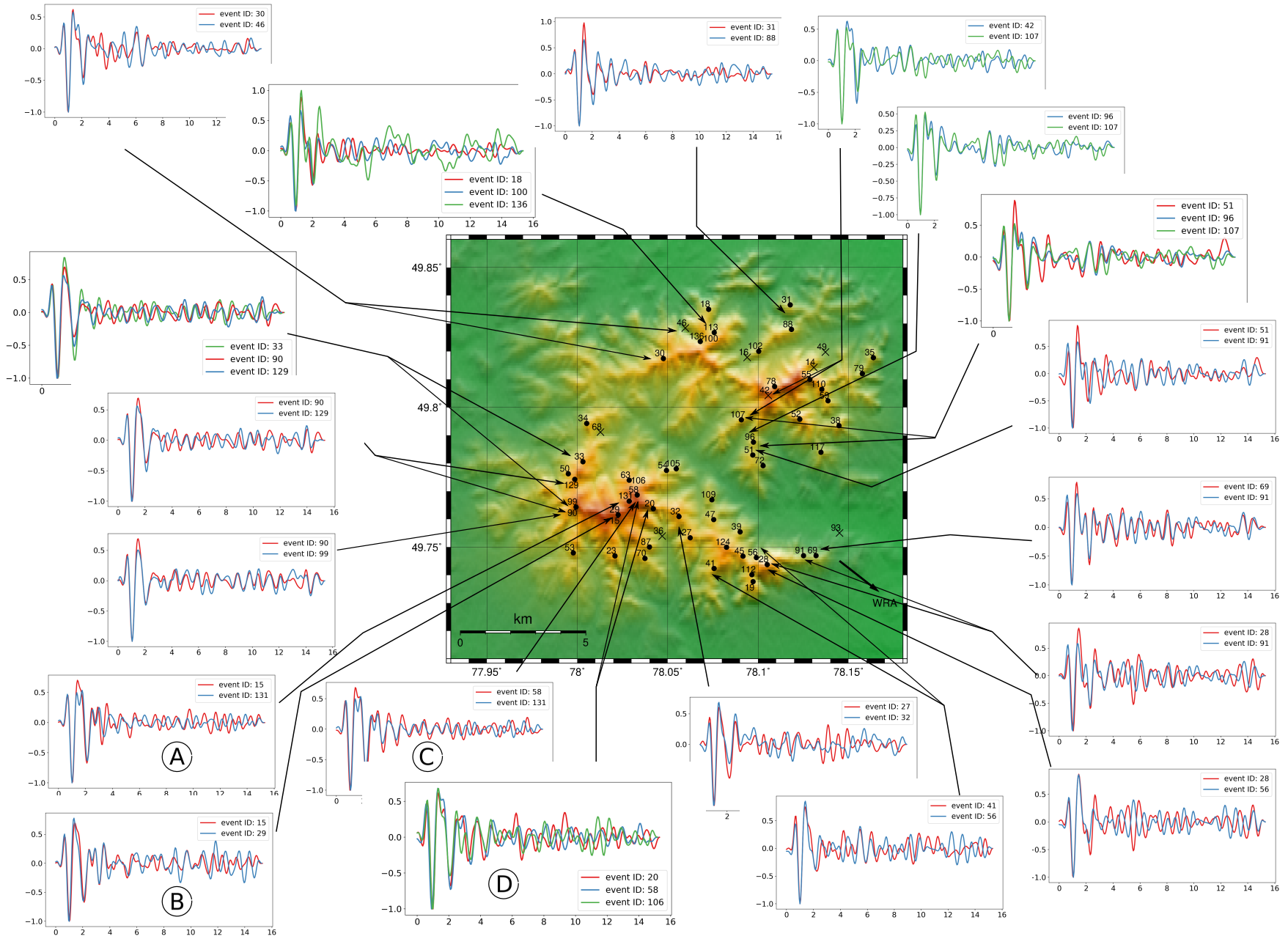
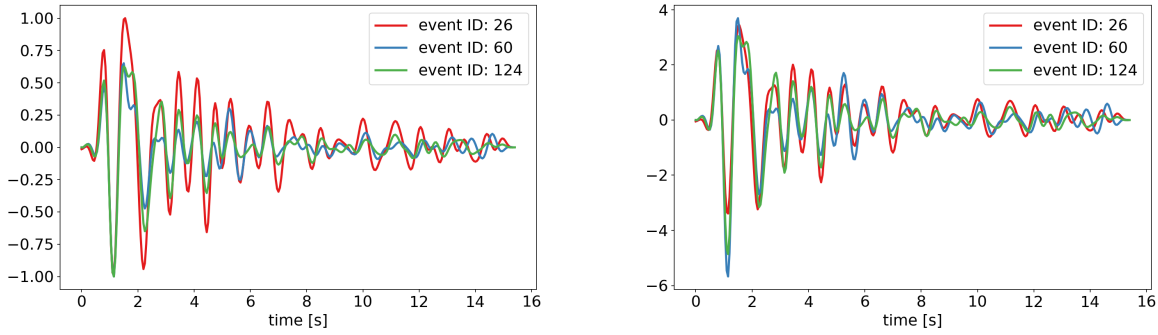


Figure 3.21: Location-dependent seismogram features at the WRA array. The direction from Degelen to WRA is indicated with an arrow. The adit ends corresponding to each event are plotted over the Degelen topography (circles). For events without adit-end locations, the GPS locations are shown (crosses).



(a) (b)

Figure 3.22: Three events, located within 400-600 m, recorded at YKA and (a) normalised by the maximum amplitude and (b) standardised, i.e. normalised by the standard deviation of amplitudes. Both normalisations remove the effect of source magnitudes and allow for a direct comparison of seismograms. The perception of seismogram similarities and differences, however, changes with the normalisation used and might hint that by normalising we exacerbate or alleviate some relative topographic effects on the traces. It should be noted that for most event pairs the differences were minor and this is an example of one of the more extreme variations. The corresponding locations are shown in Figure 3.6.

A final note on normalisation

To eliminate magnitude variations, all plots in Figures 3.18-3.21 (as well as in some of the other figures presented in the text) have been normalised by the maximum amplitude. As previously noted, however, this could perturb our analysis of seismogram similarities and differences. Figure 3.22 explores this by comparing traces normalised by the maximum amplitude to those normalised by the standard deviation of amplitudes. Given that the standard deviation expresses the dispersion of the amplitudes around the mean (equal to zero, as all data is demeaned), normalising by the standard deviation takes into account all amplitudes, and not just the maximum. For the majority of traces this does not result in strong differences in characteristics, but in some cases – like the one shown in Figure 3.22 – differences are immediately apparent and could indicate that normalisation not only collapses the effects of magnitudes, but also modifies our perception of waveforms relative to each other.

3.4.4 Waveform complexity through a topographic lens

The visual analysis of seismograms in the previous section suggests that the source location influences the waveform shape, with events in close lateral proximity exhibiting high similarities. It appears that this location-dependent near-source scattering is strongly directional, as waveform

shapes vary between the stations. Previous studies have suggested interactions with topographic irregularities in the source region as one of possible factors contributing to such variations (e.g., McLaughlin and Jih, 1988; Lynnes and Lay, 1989; Marshall et al., 1992). With the Degelen explosions located in a mountain range, we postulate that the spatial pattern uncovered in Figures 3.18-3.21 – with its progressively changing waveforms along the ridges – is related to the sharp topography. In the following, we attempt to verify this hypothesis by measuring the complexity of waveforms as a function of lateral location in the test site and of topographic skewness. We also test it against the weak-signal hypothesis, and finally we run least-squares statistical tests to verify our hypothesis of the relationship between complexity and stations and events.

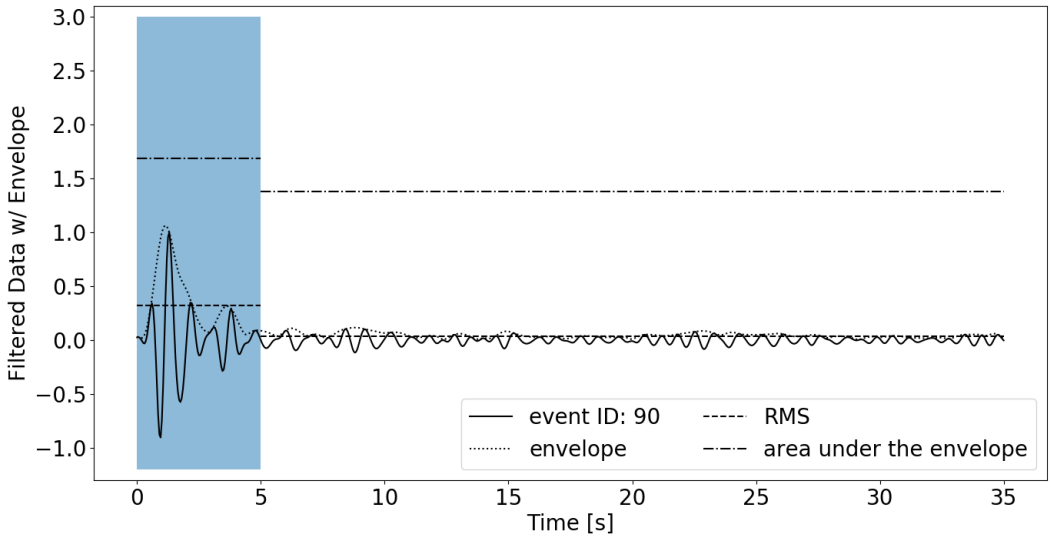
In the analysis we use two dimensionless measures of the variation of waveform complexity: the area under the envelope and the Root Mean Square (RMS, also known as the quadratic mean). Both are calculated for bandpass-filtered seismograms (0.5-2 Hz, Butterworth zero-phase filter) and normalised with respect to the maximum amplitude. The area under the envelope is calculated via integrating the seismogram envelope over a given time window. The RMS is the square root of the mean of squares of each sample, i.e. $\sqrt{(\frac{\sum_{i=1}^{n} {y_i}^2}{n})}$, where y_i is the waveform amplitude at time sample i in a given time window.

Given that the most prominent location-dependent features are present within the first 5 seconds of the signals (Figures 3.18-3.21), we choose to focus on the 0-5 s time window for the majority of the analysis, with 0 s at the onset of the P wave. Many previous studies have also observed such changes in complexity of nearby events – where the first 5 seconds are most coherent and visibly change with location, while the later coda tends to exhibit less coherence – and analysed the 0-5 s and 5-35 s time widows separately (e.g., Douglas, 1967; Lay, 1987). To test against the weak-signal hypothesis, however, we also consider the ratios between the two time windows. Figure 3.23 illustrates the complexity measures.

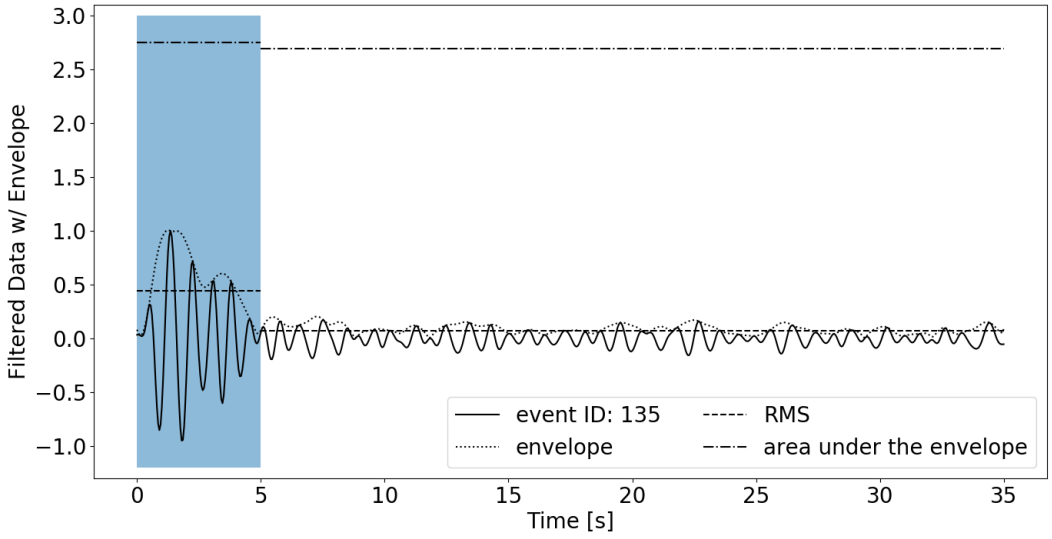
Complexity v.s. event location

The first proxy for topography is the event location itself. Figure 3.24 shows plots of the values of the area under the envelope and of the RMS for each location in the Degelen massif. For all four stations, both measurements of complexity are in agreement and appear to reproduce the progressive waveform changes along the ridges that can be seen in Figures 3.18-3.21. Moreover, they also hint at varying levels of complexity between stations that can be observed in Figures 3.18-3.21 – seismograms at the nearer stations (GBA and EKA) have more high-envelope and high-RMS events than those further away (YKA and WRA).

- GBA: The measures appear to be progressively changing in the mountain range, especially along the southern ridge. Looking from North-West towards the South-East, the values increase along the southern ridge and clusters of similar complexity levels can be identi-



(a)



(b)

Figure 3.23: RMS and envelope measures of complexity for the 0-5 s (blue area) and 5-35 s (white area) time windows for two GBA events: (a) event 90 and (b) event 135. Both measures are calculated for seismograms filtered with a Butterworth zero-phase filter to 0.5-2 Hz and normalised with respect to the maximum amplitude. The area under the envelope is calculated via integrating the envelope (dotted line) over the interval of 0-5 s and 5-35 s. The RMS is the square root of the mean of squares of each sample, i.e. $\sqrt{(\frac{\sum_{i=1}^{n} {y_i}^2}{n})}$, where y_i is the amplitude at time sample i in a given time window.

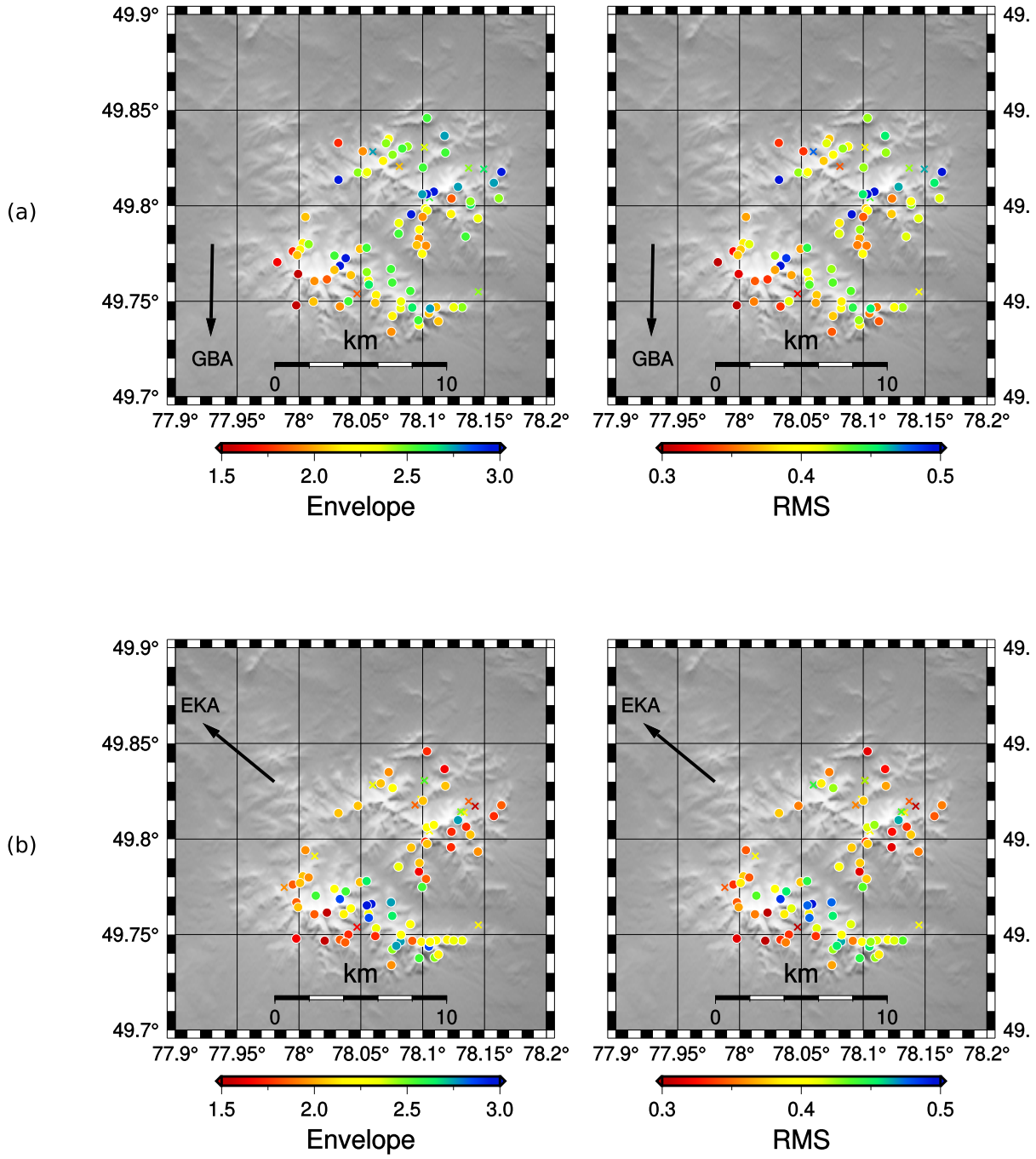
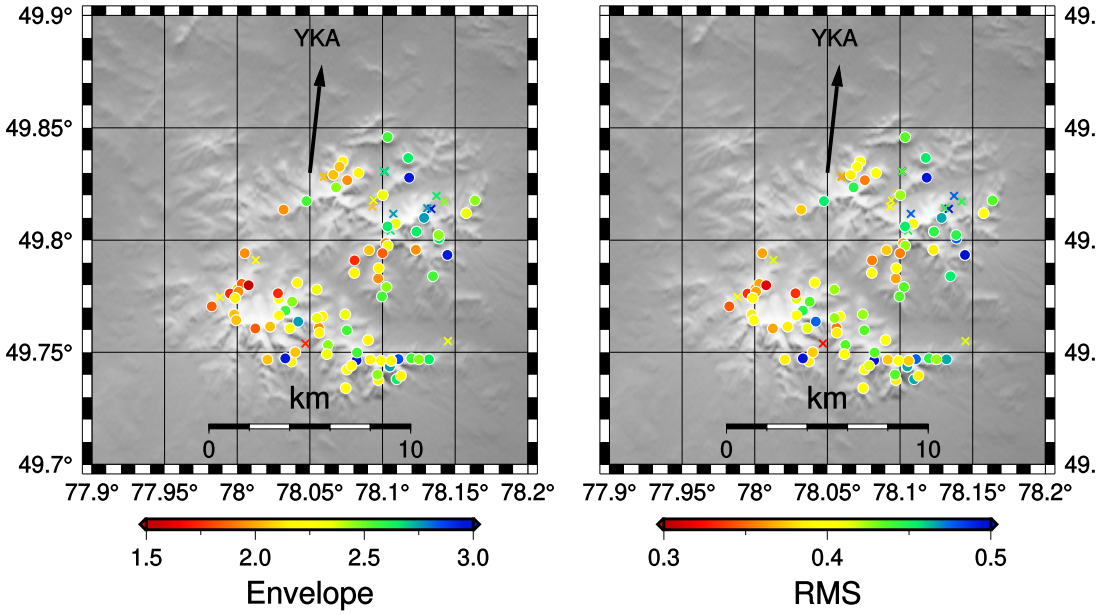


Figure 3.24: Values of envelope and RMS complexity measured for each event in the Degelen massif at (a) GBA, (b) EKA, (c) YKA, and (d) WRA. The spatial distribution of complexity measurements appears to reproduce the progressive waveform changes along the ridges observed in Figures 3.18-3.21.

(c)



(d)

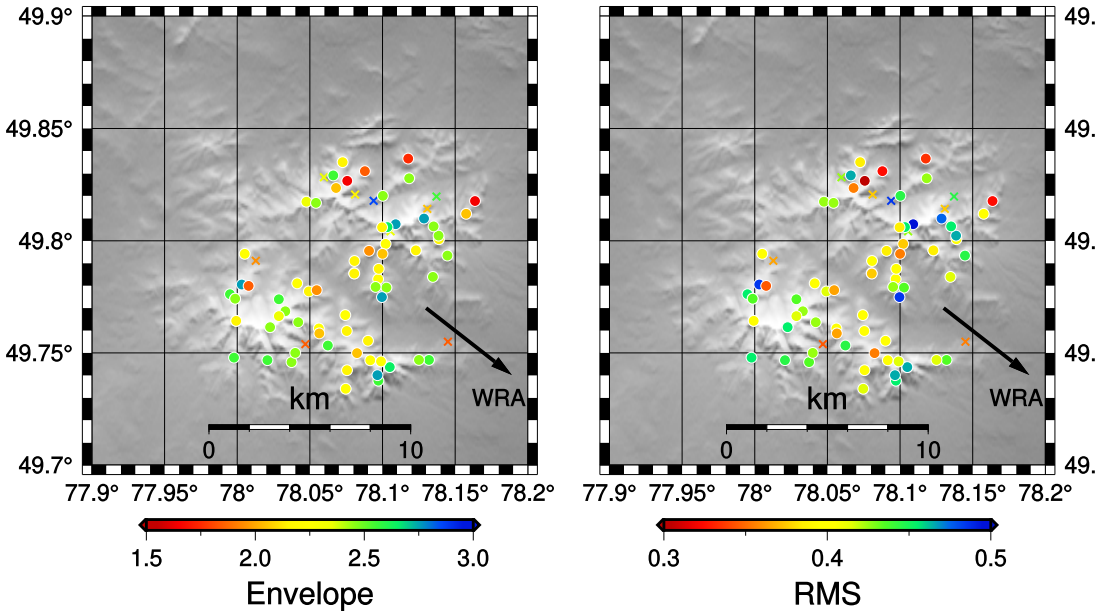


Figure 3.24: (continued).

fied. Moreover, the measures vary not only along, but also across the ridge – with lower complexity on the side of the valley, and higher complexity on the outward-facing slopes. (3.24a)

- EKA: The measures appear to be progressively changing in the mountain range, with relatively clear clusters of similar complexity levels. Again, the North-West to South-East pattern emerges on the southern ridge. The pattern on the northern ridge is less clear, yet still present – with more complex waveforms towards the centre of the highest northern peak. (3.24b)

- YKA: The measures suggest simpler events in the North-West, while events get more complex as source locations move South-East. The pattern is present on both the northern and southern ridge. (3.24c)

- WRA: The measures appear to be progressively changing in the southern ridge, but more scatter is present on the northern ridge. (3.24d)

Complexity v.s. topographic skewness

To examine a more direct relationship between seismogram complexity and topography right above the source, we calculate topographic skewness as a measure of topographic roughness in the direction of the great circle path to each array. The topography of the area from the NASA's Shuttle Radar Topography Mission is used (SRTM, jpl.nasa.gov/srtm). A topographic profile in the direction of the great circle path is then extracted (similar to the profiles shown in Figure 4.1 in Chapter 4) and a Gaussian weighting function with a half-width of 1.5 km is calculated in order to center the sensitivity to topography right above and around the source. Finally, weighted skewness (Stuart and Ord, 1994) is computed for each event at every array according to:

$$\frac{\sum_i \omega_i \left(\frac{x_i - \overline{x}}{\sigma^2}\right)^3)}{\sum_i \omega_i} \tag{3.1}$$

where ω_i is the value of the Gaussian weight at i, $\overline{x}$ is the topographic mean along the profile, σ is the topographic standard deviation along the profile, and x_i is the value of topography itself at i.

Figure 3.25 summarises the results and shows envelope and RMS measures plotted against topographic skewness for each event at every array. Both complexity measures show similar clustering. However, no relationships can be identified, and we hypothesize that scattering in the source region is sensitive to topography beyond the direction of propagation and beyond the scope limited by the Gaussian weighting function.

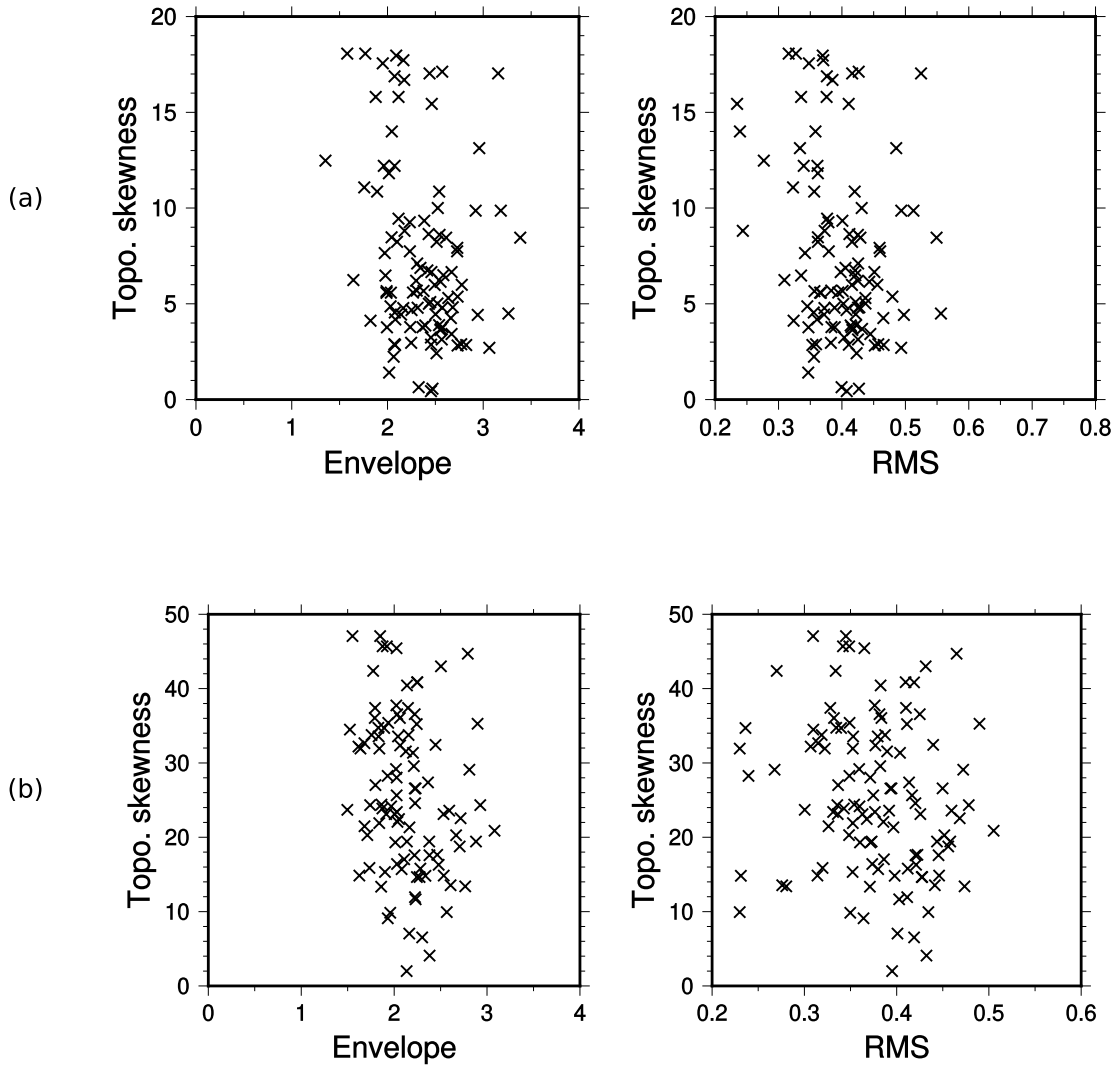
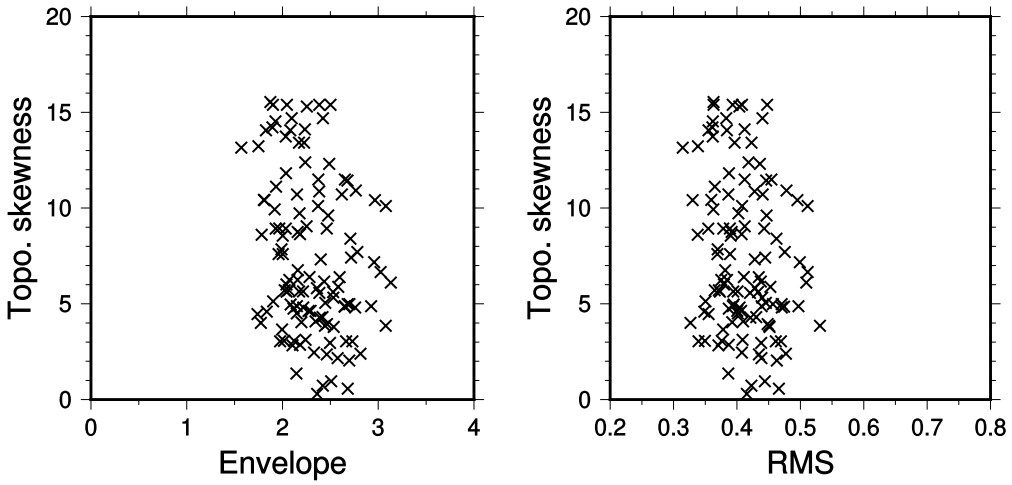


Figure 3.25: Values of envelope and RMS complexity measured for each event in the Degelen massif as a function of topographic skewness at (a) GBA, (b) EKA, (c) YKA, and (d) WRA. No relationships can be identified for any of the stations.

(c)



(d)

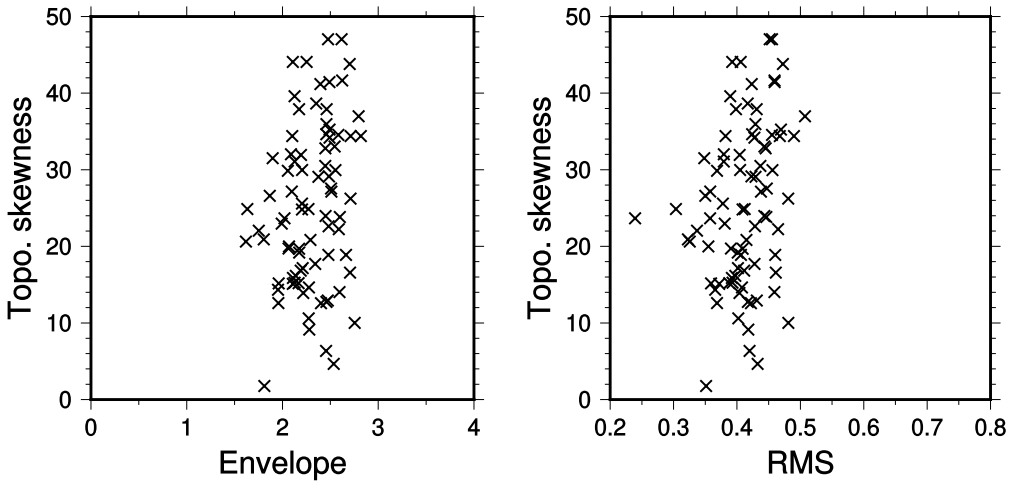


Figure 3.25: (continued).

Complexity v.s. weak signal hypothesis

Douglas (1967) suggests that complexity can be a result of low-amplitude fist P wave arrivals (0-5 s) relative to regular-amplitude coda (5-35 s) rather than of a coda wave-train of large amplitudes relative to regular fist P wave arrivals – a hypothesis later referred to as *the weak signal hypothesis*. Douglas et al. (1971) and Douglas et al. (1973) apply the weak signal hypothesis to explosion seismograms and show evidence that complex seismograms are recorded when the direct P waves have travelled a path of a lower average Q than the paths followed by the later arrivals. Simpson and Cleary (1977), however, argue that the weak signal hypothesis – at least as applied by Douglas et al. (1971) and Douglas et al. (1973) – does not give a satisfactory explanation of complex explosion seismograms. They suggest that complexity depends on the degree of heterogeneity of the scattering regions in the crust and upper mantle between the source and the receiver. Douglas et al. (1981) defend the weak signal hypothesis and show evidence that complex waveforms are recorded when the P amplitude is relatively small rather than the coda relatively large. Douglas et al. (1981) reject scattering, including interactions with topography, as the main contribution to complexity of explosion seismograms.

As previously stated, given that the Degelen sources are at most 15 km apart, Marshall et al. (1992) attribute the variations observed in the dataset primarily to the source region. However, in our analysis we have only found an indication of a relationship between the chosen complexity measures and event location, and neither of the measures could be related to topographic skewness. We thus choose to verify whether the weak signal hypothesis holds for our dataset, and whether differences in the average Q along the propagation paths of different phases could account for the observed complexity of Degelen waveforms.

One feature of complex explosion seismograms that fit the weak signal hypothesis is that the corresponding station m_b (i.e. m_b calculated from the recording at the given station) is below the average m_b by up to one magnitude unit. This suggests that one should expect a relationship between seismogram complexity (as defined by Douglas (1967), i.e. relating complexity of the 0-5 s and 5-15 s time windows) and the difference between the station m_b and the averaged network m_b – the simpler the seismogram, the smaller the difference. In order to test the hypothesis on the Degelen events, we have calculated the difference between the station magnitude (STNmb) and the maximum likelihood estimates (MLmb) of the network magnitude (Peacock et al., 2017) for every event at each array. In Figure 3.26, the m_b difference is plotted against the ratio of envelope and RMS measures of complexity for time windows of 0-5 s and 5-35 s. There is no relationship between the ratios for the chosen measures of complexity and the difference between the station and network m_b values. The results suggest that the weak signal hypothesis does not stand for Degelen waveforms, and observed complexity does not seem to result from a lower average Q along the propagation path of the direct P wave.

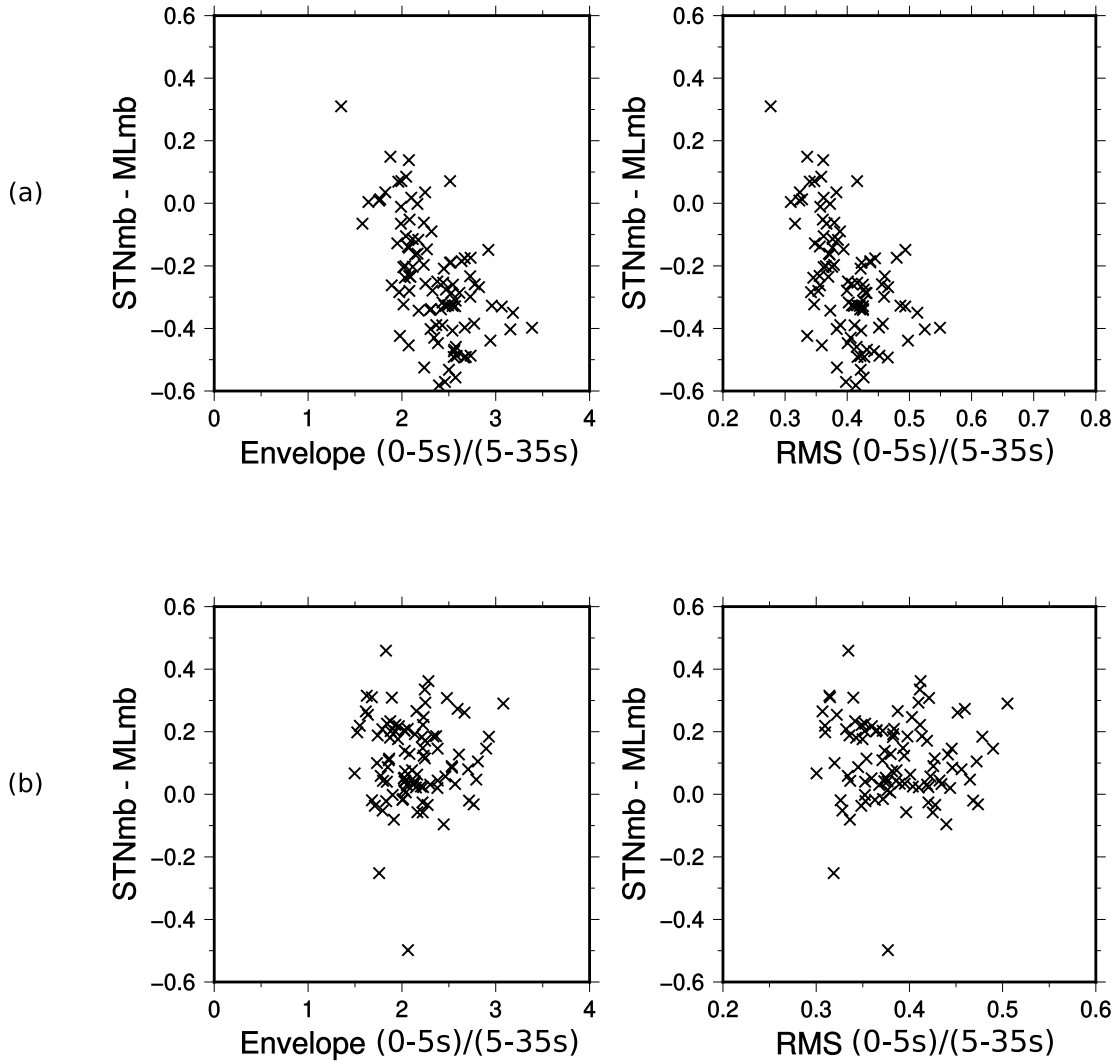
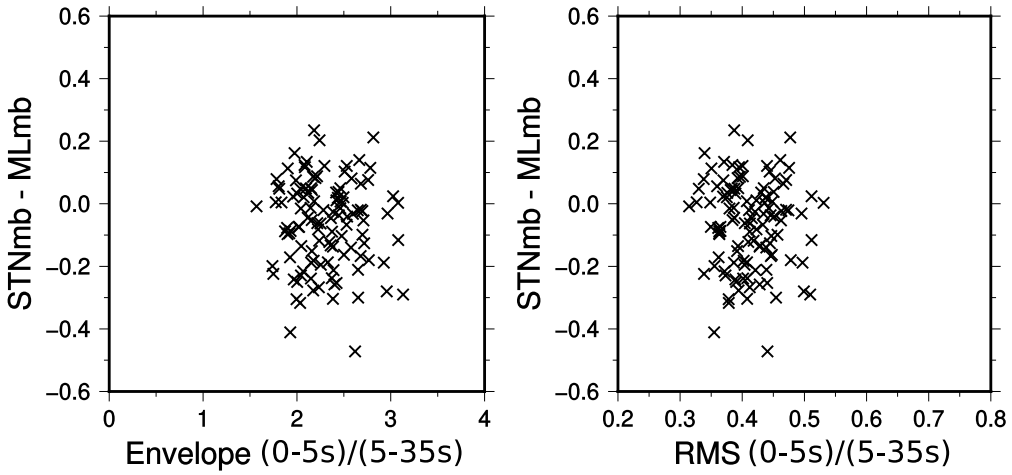


Figure 3.26: The m_b difference between the station magnitude (STNmb) and the maximum likelihood estimates (MLmb) of the network magnitude plotted against the ratios of envelope and RMS measures of complexity between 0-5 s and 5-35 s time windows at (a) GBA, (b) EKA, (c) YKA, and (d) WRA. The energy ratios of complexity show no relationship with the m_b differences, suggesting that the weak signal hypothesis does not stand. One may argue that there is a trend for both measures at GBA in (a), but it is very weak and scattered.

(c)



(d)

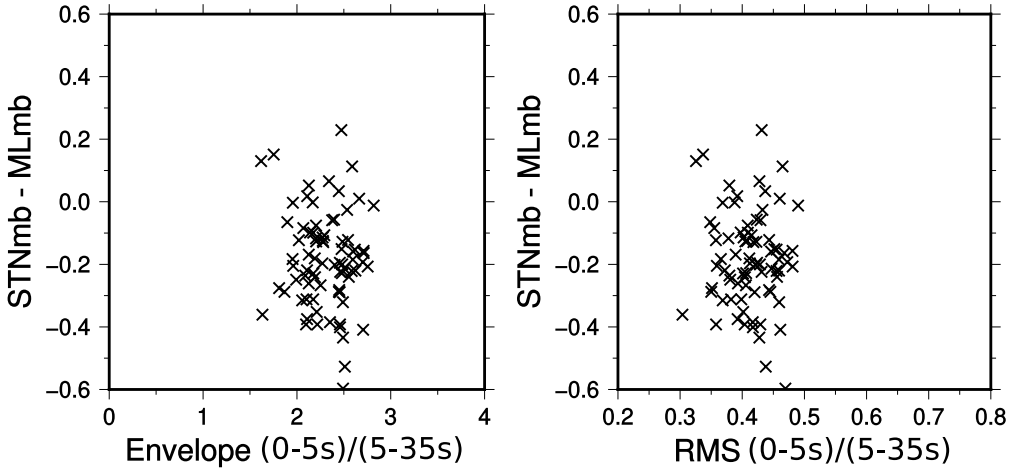


Figure 3.26: (continued).

Mean square of residuals:	0.089954
Degrees of freedom:	209
Mean square due to stations:	0.9411
Station degrees of freedom:	3
Mean square due to events:	0.1465
Events degrees of freedom:	90

Table 3.3: Statistics for the least squares analysis of complexity discussed in the text.

Array	N	Station term
GBA	73	0.072 ± 0.062
EKA	81	-0.170 ± 0.060
YKA	88	0.032 ± 0.067
WRA	61	0.067 ± 0.058

Table 3.4: Station terms s from the analysis of observed complexity with 95% confidence limits. The terms display no distance dependence.

Statistical hypothesis testing

Given that both representations of complexity appear to change according to the event location in the massif in nearly identical ways (see Figure 3.24), we choose to statistically analyse one of the measures – the area under the envelope – in relation to stations and event locations. We use an approach based on the least-squares method as described by Douglas (1966). Originally formulated for magnitudes, the method can be used to solve any problem that can be expressed as a sum of event and station terms. For our application, we assume that the complexity measurement c_{ij} of the j^{th} explosion at the i^{th} station can be expressed as

$$c_{ij} = b_j + s_i + \overline{c} + \epsilon_{ij} \tag{3.2}$$

where b_j is the explosion term, s_i is the station term, $\overline{c}$ is a constant that can be thought of as the complexity of the average source at the average station (a mean explosion-station term), and ϵ_{ij} is the error of each observation of c_{ij}. By assuming that $\sum_j b_j = 0$, $\sum_i s_i = 0$, and $\sum_{ij} \epsilon_{ij} = 0$, equation 3.2 is equivalent to a two-way analysis of variance for a problem where all stations record all explosions, i.e. c_{ij} is known for all event-station combinations. Seismic records, however, are often incomplete – either because signal amplitudes are below detection thresholds or because a station was not operating for a period of time – and the method of least squares does not require all of c_{ij} to be known (Douglas, 1966). Moreover, by assuming a normal distribution of ϵ_{ij}, the confidence limits can be estimated (Douglas, 1966).

The statistical parameters from the analysis are summarised in Table 3.3, and the calculated station terms of complexity are given in Table 3.4. We can use the values in Table 3.3 and the F-test to test a number of hypotheses. Firstly, Table 3.4 suggests that there is no relationship between station and complexity. Using mean squares of residuals and stations to calculate the variance ratio F and associated degrees of freedom we obtain a probability of 0 (significance level 0), suggesting that there is indeed no relationship between station and complexity. Secondly, we test whether there is a relationship between event and complexity (using mean squares of residuals and events and associated degrees of freedom). For this test we obtain probability: 0.00243, significance level: 0.25%, suggesting there is a relationship between event and the envelope measure of complexity.

3.5 Summary and discussion of results

The Degelen waveforms constitute a large and unique dataset that bears a strong signature of source complexity. The high-quality recordings from four UK-designed arrays provide an ideal starting point for a systematic study of relative importance of near-source effects, as sub-surface geology and topography of the test site are known, and explosion locations, depths and yields are well-constrained. Such prior knowledge reduces uncertainties in numerical set-ups and allows for direct comparisons of observed seismograms with synthetic results. While it is difficult to numerically test contributions of non-linear effects – such as explosion-induced tectonic strain release, change in rock properties, cavity creation and collapse, or spall – the teleseismic effects of depths and magnitudes, as well as of elastic scattering from topography and from geological heterogeneities can be investigated with hybrid modelling.

We therefore explore the Degelen dataset for teleseismic manifestations of the effects of depth, yield and location in the test site. Distributions of CC, PM and EM coefficients suggest that neither depths nor magnitudes are the primary factor explaining similarities and differences in waveform features. The characteristics of signals, however, appear to be dependent on event location. Waveforms of close-lying explosions of different depths or body-wave magnitudes are highly similar, while same-depth or same-m_b explosions in different locations result in less coherent arrivals. A manual analysis of seismograms on each array with respect to source location in the test site revealed a nearly-continuous change of waveforms, with features progressively emerging and disappearing as locations 'move' along Degelen mountain ridges. The degree of variation differs between the stations, suggesting directionality of near-source effects that translate to dependence of teleseismic recordings on take-off angle and azimuth.

Given the mountainous nature of the Degelen test site, we postulate that strong topography is one of possible contributions to variations in the recorded waveforms. We attempt to quantitatively relate signal complexity to Degelen topography, but simple measures do not indicate

a straightforward relationship. The spatial distribution of values of RMS and of the area under the envelope of the 0-5 s time window hint at progressive waveform changes along the ridges. Topographic skewness along the great circle path to each array, however, does not relate to the chosen measures of complexity. We therefore test whether the weak signal hypothesis – a competing theory that proposes differences in average Q along the propagation paths of different phases as the primary source of seismogram complexity – explains differences observed in the dataset. The hypothesis does not seem to hold for Degelen waveforms at any of the arrays. We finally turn to statistical hypothesis testing of the envelope representation of complexity that appears to change according to the event location in the massif, and dependence of complexity on event location is statistically significant.

Identifying the origins of complex waveform features proves difficult without a good understanding of the relative importance and influence of contributing factors. Although consistent with each other, the chosen complexity measures average out waveform variations and collapse the defining features to a single number, masking characteristic details and subtleties. Topographic skewness along the great circle path also only approximates the mountainous features, and does not capture the majority of the three-dimensional terrain that influences the waveforms. The manual analysis of the dataset along with envelopes of the 0-5 s time window suggest that location in the test site influences the seismograms, yet relating emerging features to topography or other contributing factors requires deterministic modelling that includes three-dimensional wave propagation effects. In the second part of this study presented in Chapter 4 we therefore turn to hybrid modelling to numerically test the effects of near-source topographic scattering on teleseismic signals.

Chapter 4

A nuclear explosions case study, Part II: Modelling the effects of topography at teleseismic distances

4.1 Introduction

4.1.1 Modelling of near-source scattering

The first attempts at modelling various effects of near-source scattering at teleseismic distances were made in the late 1970s. However, due to high computational costs limiting such long-range simulations at high frequencies of interest, the efforts till date have been limited. In one of first such studies, Bouchon (1976) combines the representation theorem with the flat layer theory in order to model body waves from explosive sources and earthquakes at teleseismic distances. The method accounts for the free surface and the heterogeneous 1-D layers around the source, and repropagates the local scattering effects to distant receivers. Bouchon (1976) demonstrates that for a range of distances separated from triplications and caustics (30-100 degrees), the recorded seismograms mostly reflect the influence of the free surface and further crustal reverberations in the source region, bearing little resemblance to the emitted signal. The effect of the Yucca Flat (at the Nevada Test Site, NTS) geology is also studied in detail, revealing that the detonation medium and source depth strongly affect the m_b-M relationship. Vidale (1987) couples a local acoustic two-dimensional (2-D) finite difference (FD) scheme with the representation theorem that propagates the locally modelled energy through a half-space to a receiver 10,000 km from the source. Through the coupling, Vidale (1987) attempts to examine the effect of a deep-penetrating heterogeneous slab on body waves from deep earthquakes – with the 2-D FD plane defined as the plane perpendicular to the strike of the slab – and concludes that the faster material in the slab advances some arrivals, and defocuses or distorts others. A similar method

coupling a local 2-D FD scheme with the representation theorem was later proposed by Stead and Helmberger (1988) for the elastic case. The authors apply it to the Yucca Flat site, and show that interactions with the basin boundary cause local wavefield variations that can be observed at teleseismic distances shortly after P and pP phases. McLaughlin and Jih (1988) use a 2-D FD method to explore the effects of topography on underground nuclear tests at the southern Sahara French test site in Algeria, and generate synthetic seismograms with energy in the 0.5-3 Hz range to match the frequency content of the data. The simulations suggest that the free surface reflections (pP) could be strongly affected by the Taourirt Tan Afella massif, producing a 0.3 magnitude variation in the network m_b. McLaughlin and Jih (1988) expect such near-source topographic scattering to depend both on the azimuth and on the take-off angle. Also with the aim of studying nuclear explosions, Stevens and Xu (2008) present another coupling method based on the representation theorem – this time embedding a full three-dimensional (3-D) calculation in the source region within a layered Earth model. Both linear and non-linear local simulations are discussed, but only benchmarks are demonstrated. In a similar approach, Avants (2014) couples a 3-D FD scheme with 3-D teleseismic P wave Green's functions. Despite the fact that the method can account for frequencies up to 5 Hz, the coupling proved unsuccessful in reproducing features in the observed data. The coupling surface used was limited only to the bottom surface of the 3-D method, and hence much of the local 3-D wavefield was unaccounted for in the subsequent repropagation to receivers. Nonetheless, applying the hybrid scheme to underground nuclear tests in North Korea, Avants (2014) reports that rough surface topography can strongly affect the timing of the depth phases (pP, sP) relative to the direct phases from shallow sources. Most recently, Wu et al. (2018) proposed a hybrid coupling of a 3-D spectral element method (SEM) around the source region, with a global one-dimensional (1-D) direct solution method (DSM). Wu et al. (2018) apply the method to a number of case studies, reaching frequencies up to 0.1 Hz in the examination of effects of water, topography and bathymetry at teleseismic distances. A distortion of direct P waves and a generation of some coda waves is reported.

Given the lower computational requirements of local and regional simulations, a wide variety of studies focused on the effects of near-source structures on high-frequency wavefields at shorter distances, ranging from one to a few hundred kilometres. In particular, the influence of topography has been extensively studied in the context of nuclear explosion monitoring. Frankel and Leith (1992) perform 3-D FD simulations of a hypothetical explosion at the northern Nova Zemlya test site and show azimuthal variations in the amplitudes of downgoing P waves due to the variations in the surface reflection pP, as well as substantial SH motion for certain azimuths. Gaffet and Massinon (1997) use a 2-D discrete wavenumber-indirect boundary integral method to study contributions of the Taourirt Tan Afella massif to the southern Sahara French nuclear tests. They report increases in time duration and amplification of pP, new phases that

appear in the coda of the pP phase on the radial component, and a great variability of the transverse maximum amplitudes with take-off angle. Myers et al. (2007) explore the generation of S waves from explosion sources, and via 2-D FD calculations show that the primary S wave pulse originates near the free surface and at high-impedance geologic interfaces that are close to the explosion. Rodgers et al. (2010) consider a realistic topography of the North Korean peninsula to investigate elastic propagation effects, with an emphasis on the theoretical aspects of S wave generation. Simulations up to 8 Hz show that motions at near-vertical take-off angles (corresponding to teleseismic propagation) are increased or decreased, with the variations caused by pP surface reflections. Both SV- and SH- polarised motions are amplified simultaneously, and the shear wave energy is equipartitioned on all three components. Pitarka et al. (2015) aim to numerically reproduce the SPE-3 (Source Physics Experiment 3) explosion and look at wavefields in the 0.01-10 Hz frequency range, at epicentral distances of up to 1 km, via two approaches: 3-D anelastic wave propagation modelling, and coupled hydro-dynamic modelling of a chemical explosion source with anelastic wave propagation modelling. The anelasic approach, despite an isotropic source, partially replicates observed azimuthal variations of the P wave amplitudes as well as increased waveform complexity with distance. Moreover, the simulations show that a large proportion of the generated shear motion is propagated downwards. Gaebler et al. (2019) consider a 2-D model of Mt. Mantap to estimate the effect of topography on peak amplitudes and assess the possible range of explosive yields. Large differences in the downwards propagating pP and pS waves are observed, with the increases in amplitude by a factor of about 1.5 due to topography. Given that at teleseismic distances the P and pP phases are considered a signle onset, Gaebler et al. (2019) suggest that topography might lead to an overestimation of the explosion yield.

The effects of topography on local and regional wavefields have also been studied in the context of crustal earthquakes (e.g. Takemura et al., 2015), volcano seismology (e.g. Ripperger et al., 2003; Kumagai et al., 2011), and ground motions for seismic hazard assessment (e.g. Komatitsch et al., 2004; Ma et al., 2007; Lee et al., 2009; Chaljub et al., 2010; Imperatori and Mai, 2013; Magnoni et al., 2014). Other studies of near-source scattering on local and regional scales have also considered the effects of random heterogeneous 3-D structures (e.g. Xie and Lay, 1994; Nielsen et al., 2003; Pitarka et al., 2007; Burgos et al., 2016), and of deterministic structures such as slabs (e.g. Igel et al., 2002). To take advantage of such prominent topographic scattering, Wang et al. (2016) recently explored the use of P-coda waves generated by surface topography to obtain high-resolution location of near-surface seismic events like nuclear explosions and shallow earthquakes. They show that the uncertainty in the vertical source location is reduced when P coda is considered, provided that the P coda can be accurately modelled and is sufficiently strong relative to random noise.

All of the cited studies agree that near-source scattering off topography and complex struc-

tures has an effect on waves of wavelengths corresponding to the length-scales of heterogeneities. None of them, however, can provide meaningful comparisons of the synthetically generated signals with data recorded teleseismically. The long-range studies from the 70s and 80s reduce the dimensionality of the problem to either 1-D or 2-D, compromising on the 3-D propagation effects – and in some cases on the full physics – in favour of computational feasibility. More recently, Avants (2014) attempted a full-physics 3-D extension to the representation theorem approach to extrapolating wavefields to distant receivers, though the method implementation did not consider all effects radiating from the 3-D volume and teleseismic waveforms were not well reproduced. Many of local high-frequency simulations, on the other hand, account for full physics and 3-D effects, and thus provide valuable theoretical insights into the complexities of realistic wave scattering, but cannot be compared to any recorded signals at distance. The altered downgoing waves, however, hint at teleseismic effects of the local structures in the source region.

Meaningful comparisons of synthetic and real data require numerical methods that do not compromise on the physics or on the 3-D effects, and are nonetheless computationally feasible for high frequencies that capture the structural length-scales of the heterogeneities of interest. The hybrid Instaseis framework proposed in Chapter 2 honours both these requirements: it accounts for high-resolution full waveform effects within a reasonable computational cost. Despite its simplifications of a 1-D layered Earth in the background domain and an anelastic wave propagation both globally and locally, it provides the most complete setting to date for generating realistic high-resolution teleseismic waveforms.

4.1.2 The Degelen waveforms

The P waves in the Degelen dataset carry a strong signature of near-source interactions. Although the degree of the effects varies between the stations (GBA, EKA, YKA and WRA arrays, see Figure 3.3 and Table 3.2), in Chapter 3 we established a remarkable dependence of waveform features on location within the Degelen site for every array. Events in close lateral proximity exhibit noteworthy similarities, and given the mountainous nature of the test site we postulate that waveform variations are strongly influenced by rough topography. Due to the complexity of such interactions and a multitude of other factors and source effects that alter waves propagating teleseismically, however, it is difficult to derive meaningful data analyses that would relate specific signal properties to topography. Measuring the complexity of the waveforms through RMS and the area under the envelope, we have established a systematic – although weak – variation of complexity along the Degelen ridges (see Figure 3.24), though other chosen measures remained inconclusive.

Therefore, we turn to global hybrid modelling to examine if and how topography influences high-frequency teleseismic P waves up to 2 Hz, and whether any of the distinct waveform varia-

tions can be attributed to topography alone. Factors such as heterogeneous surrounding geology, dependence of moment rate on explosion depth and yield, non-linear effects in the vicinity of the source (for further details, refer to Chapter 3), and complex path and receiver effects are excluded. Given the simple assumptions, and the uncertainty of locations of real events, we do not attempt to reconstruct particular events or fit the waveforms exactly, but instead we focus on the extent of the topographic effect and a qualitative assessment of the two datasets. Amplitudes and general waveform characteristics of 3-D synthetics and observables at teleseismic distances can for the first time be directly compared and, despite the simplicity of the model, are in good agreement. We show that teleseismic P waves carry a clear signature of the Degelen mountain range, implying that topographic effects should not be ignored.

4.2 Model description and setup

In order to propagate the effects of topography to teleseismic distances, we define a local 3-D domain that includes the Degelen mountain ridges, and via the hybrid methodology we embed it in a spherically symmetric Earth. In this application we use the Wave Propagation Program (WPP Nilsson et al., 2007; Petersson and Sjogreen, 2013) that has been specifically coupled for this purpose to the source side of the hybrid framework (refer back to Section 2.3.4). Below we discuss the details of our modelling choices as well as the specifics of their implementation.

4.2.1 Defining the local topography

We use the topographic variations of Degelen on a 90 metres latitude-longitude grid from the NASA's Shuttle Radar Topography Mission (SRTM) (see jpl.nasa.gov/srtm for details). As illustrated on cross sections in Figure 4.1, the two Degelen mountain ridges stretch over an area of approximately 30 by 30 kilometres, and the flat steppe that surrounds them has altitudes of about 500-600 m above sea-level. Therefore, defining the local simulation domain to encompass only the sharp topography of interest results in a region with boundaries at an elevation of 500-600 m. The hybrid method, however, requires that the local 3-D domain matches the global background model on the coupling surfaces, and the Degelen topography needs to be tapered off to zero altitudes of the spherically symmetric Earth.

Artificial tapering from 500 m to zero, even though smooth, could result in significant undesirable effects on wave propagation. Instead, given that we aim to isolate the influence of sharp topographic features on high-frequency global wavefields, we choose to remove both the surrounding elevation and the underlying topographic trend, as shown in Figure 4.2 and in Figure 4.3, respectively. It should be noted that by leaving only the peaks of the Degelen mountain complex we marginally change the angle of the slopes relative to the source-receiver propagation path, but we avoid introducing additional non-existent structures through tapering.

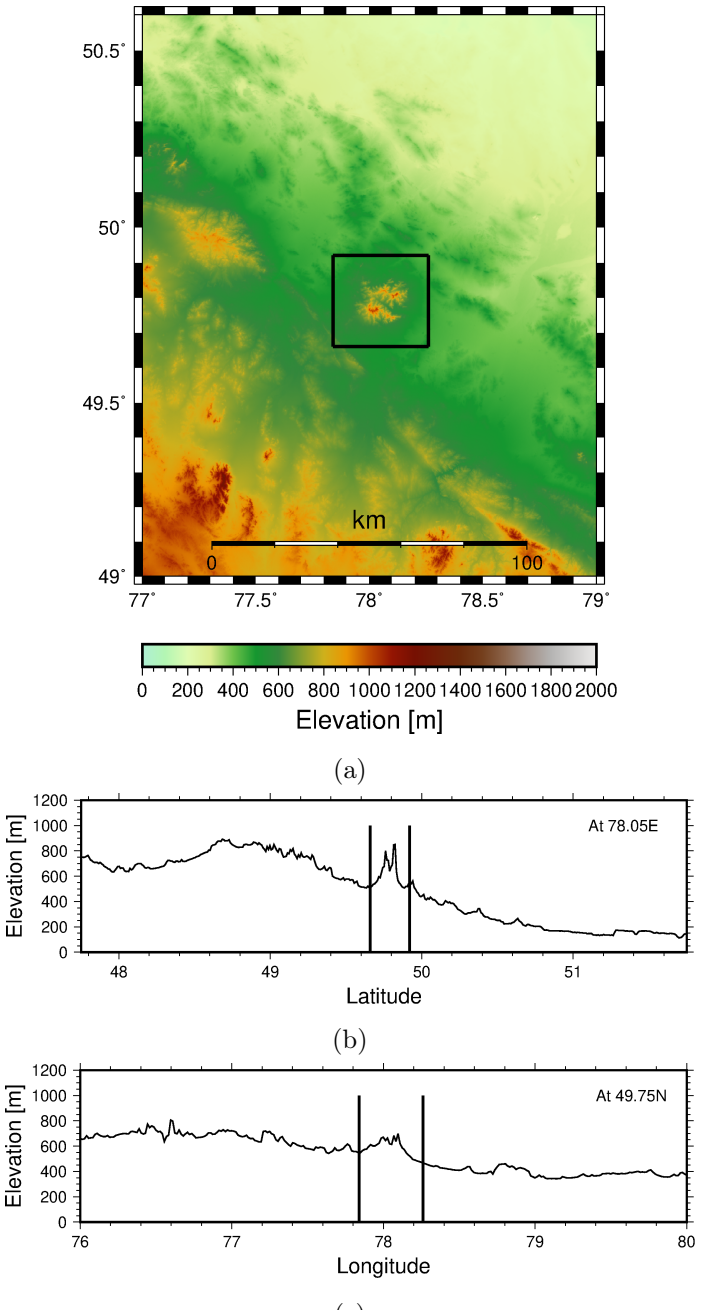


Figure 4.1: Degelen topography from the NASA's Shuttle Radar Topography Mission (SRTM, jpl.nasa.gov/srtm). (a) The Degelen site, indicated by the black square of 30×30 km, and its surroundings. (b) A cross section of Degelen topography at a longitude of 78.05 degrees East. (c) A cross section of Degelen topography at a latitude of 49.75 degrees North.

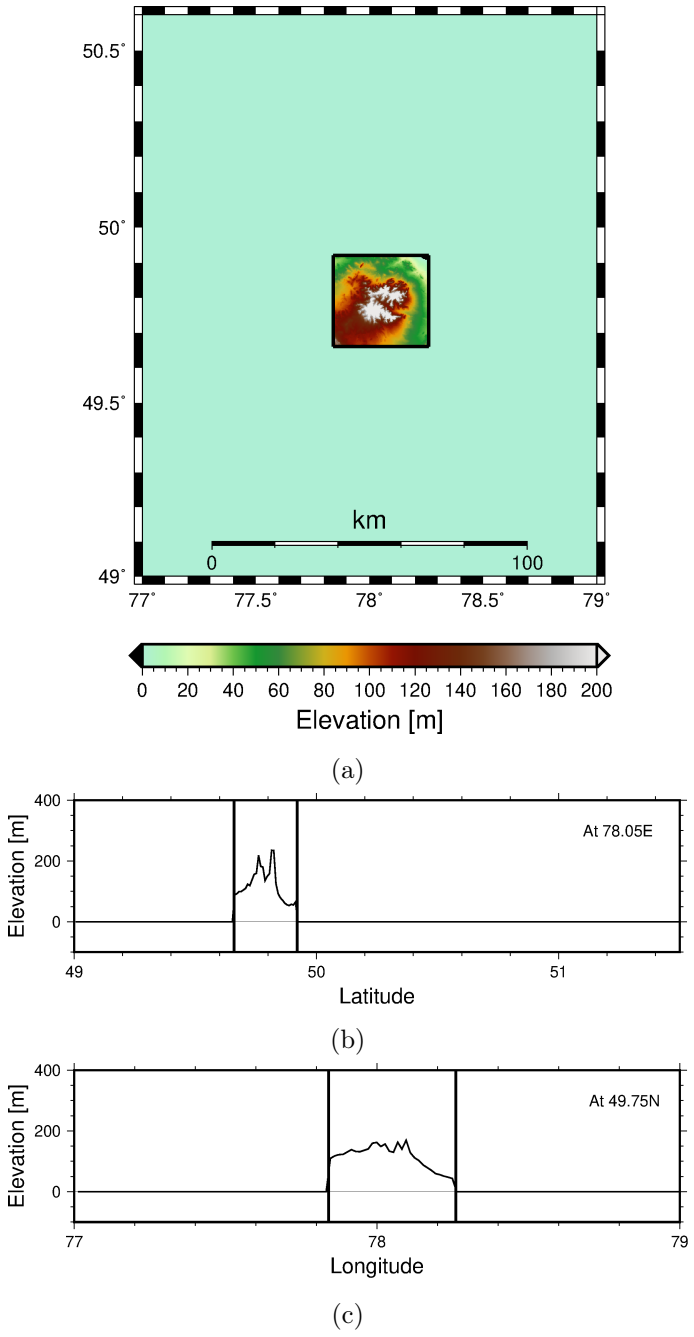


Figure 4.2: Topography after removing the underlying elevation of the Degelen mountain complex. The remaining topographic trend and its discrepancy from zero elevation is clearly visible. (a) The Degelen site, indicated by the black square of 30×30 km, and its surroundings. Only the N-E corner of the site matches the zero elevation of the surroundings. (b) A cross section of Degelen topography at a longitude of 78.05 degrees East. (c) A cross section of Degelen topography at a latitude of 49.75 degrees North.

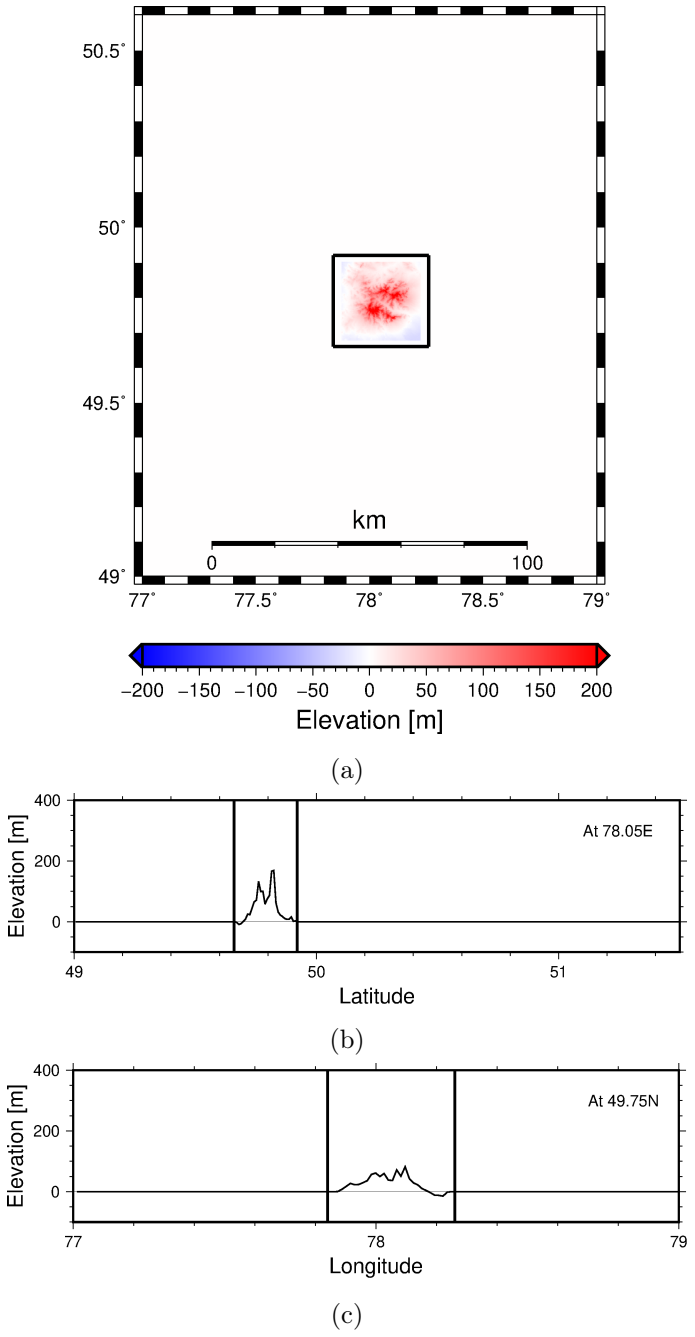


Figure 4.3: Topography after removing the underlying elevation and the topographic trend of the Degelen mountain complex in order to match the zero elevation of the background model. (a) The Degelen site, indicated by the black square of 30 × 30 km, and its surroundings. The topography along all edges of the site is equal to zero. (b) A cross section of Degelen topography at a longitude of 78.05 degrees East. (c) A cross section of Degelen topography at a latitude of 49.75 degrees North.

Array	station distance in degrees	value of t^* in the 1-D model	amplitude correction factor
GBA	36.4	0.75 s	6.5
EKA	47.1	0.83 s	8.5
YKA	67.0	0.99 s	14.0
WRA	85.3	1.12 s	21.0

Table 4.1: Values of t^* resulting from the PREM Q values used in the global simulation of the 2 Hz database, as well as the amplitude correction factors relative to the assumed $t^* = 0.15$ s. The longer the wavefield travels through the mantle, the higher the t^* operator and the higher the correction factor.

4.2.2 The 2 Hz database and the IASP91 approximation

2 Hz Instaseis database

We have generated a 2 Hz Instaseis database specifically for this study. Given that data discussed in Chapter 3 is available at four arrays – the Eskdalemuir array (EKA) in Scotland, the Gauribidanur array (GBA) in India, the Yellowknife array (YKA) in Canada, and the Warramunga array (WRA) in Australia – with a minimum and maximum epicentral distances of 36.4 and 85.3 degrees from the Degelen test site, respectively, we have limited the AxiSEM simulation to the upper hemisphere with an absorbing boundary at 90 degrees. The mesh of the semi-disk consists of 18,603,552 spectral elements and takes 17 GB of disk space. The simulation took 114,500 CPU hours for 1500 seconds of wave propagation, and the database of Green's functions with epicentral distances limited between 34 and 88 degrees requires 308 GB.

IASP91 and attenuation for high frequencies

Given the velocity structure under the Degelen massif (Belyashova et al., 2001), IASP91 (Kennett and Engdahl, 1991) was chosen as a suitable background Earth model for this study. The local region of $30 \times 30 \times 10$ km that includes the tapered topography (see Section 4.2.1) is then embedded in the global IASP91 model, as shown in Figure 4.4. The local domain remains above the 20 km discontinuity of the IASP91 model and therefore the local velocity model is homogeneous.

Since Q values are not available for IASP91, the attenuation structure of PREM (Dziewonski and Anderson, 1981) was used in the simulations (reference frequency 1 Hz). It has been suggested that the Q values as obtained for longer period waves do not accurately represent attenuation of higher frequencies in the range of 1-8 Hz (Sharrock et al., 1995), but we expect

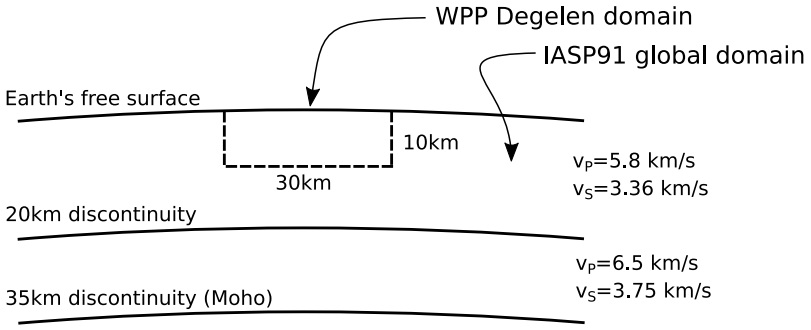


Figure 4.4: Sketch of proportions of the local Degelen domain with respect to global IASP91 discontinuities. The domain defined in WPP is 30×30 km wide and 10 km deep, and therefore remains in the homogeneous layer above the 20 km discontinuity of the IASP91 background model. The local domain dimensions and the discontinuity depths are shown to scale.

the differences in amplitudes to be small for wavefields up to 2 Hz. Moreover, we are not aware of global Q models constrained by high-frequency body waves above 1 Hz, so a more suitable model is simply not available.

We can, however, estimate the undesirable effects of the PREM Q values on amplitudes. The Q values used result in high t^* operators for every array, ranging from 0.75 seconds to 1.12 seconds (see Table 4.1). For a simple constant Q model, however, the t^* operator has been estimated to approximately 0.15 seconds for the Degelen explosions for the reference frequency of 1 Hz (Bache et al., 1985). With Douglas and Mandel (1992) suggesting that t^* is independent of frequency for short period body waves, the "Blacknest view" is to accept the simplest model that explains the Degelen observations, and we therefore choose $t^* = 0.15$ s to correct for the undesirable effects of the PREM Q values on synthetic amplitudes.

The amplitude reduction term due to attenuation is given by

$$A(f) = A_0 e^{(-\pi t^* f)}$$

where f is the reference frequency (Shearer, 2009). Thus, assuming $t^* = 0.15$ s, the amplitude correction factor can be estimated for every station as

$$\frac{e^{(-0.15\pi)}}{e^{(-\pi t^*)}}$$

for the reference frequency of 1 Hz. The values of the correction factor are given in Table 4.1, and are later used to compare the amplitudes of the synthetic seismograms to observable data in Section 4.4.1.

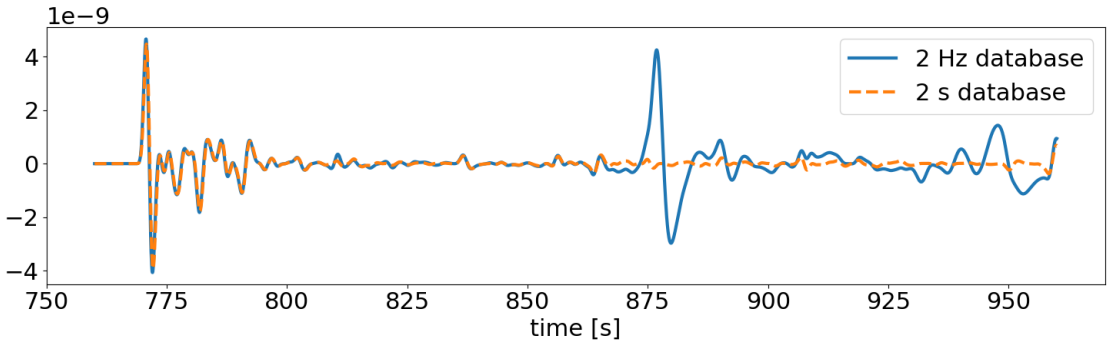
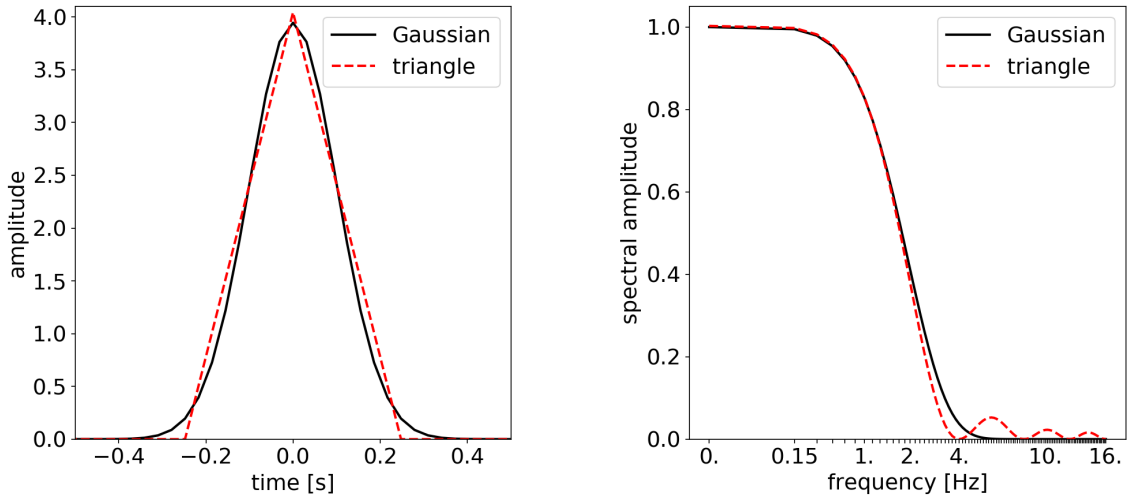


Figure 4.5: Comparison of database wavefields at 85.3 degrees epicentral distance: the 2 Hz database generated on a half-disk with an absorbing boundary condition at 90 degrees (blue line) is plotted against a 0.5 Hz database generated on the entire 2-D disk without absorbing boundaries (dotted orange line). The Green's functions from each database have been convolved with a Gaussian source-time function of 2 s half-width. Artificial reflections off the boundary at 90 degrees in the 2 Hz database appear only over 35 seconds after the onset. Synthetic seismograms for the WRA array (at 85.3 degrees from the source region) are therefore reliable in the 0–15 s time-window of interest.

A note on AxiSEM absorbing boundary conditions

Generating the database on the upper hemisphere halves the mesh size and the computation time, but imperfections of the absorbing sponge-layer boundary condition (Cerjan et al., 1985; Kosloff and Kosloff, 1986) need careful consideration. With the WRA array at 85.3 degrees epicentral distance from the source region, we need to make sure that artificial reflections off the boundary at 90 degrees do not pollute the Warramunga seismograms.

We compared 1-D seismograms at 85.3 degrees from the 2 Hz database with seismograms from a 0.5 Hz database generated on the entire 2-D disk without absorbing boundaries. Both Green's functions have been convolved with a Gaussian source-time function of 2 s half-width. From Figure 4.5 it is clear that the first unphysical boundary reflection in the 2 Hz database arrives over 35 seconds after the onset of the P wave. This is not problematic in this study, as we are concerned only with approximately the first 15 seconds of the seismograms due to the limitations of the observed data. The array signal processing techniques are steered to the P wave energy leaving the source region, and later arriving waves are suppressed in the Degelen dataset with coda amplitudes at the level of the noise.



(a) (b)

Figure 4.6: Comparison of frequency characteristics of Gaussian and triangle moment rate functions. (a) A Gaussian (solid black line) is very similar to a triangle with a half-duration of 0.24 s (dotted red line). (b) Spectral amplitude of the source. Unlike the triangle, the Gaussian moment rate function has no significant energy at frequencies higher than 4 Hz.

4.2.3 Choosing the moment rate function

The Mueller-Murphy model (Mueller and Murphy, 1971) – a nuclear explosions source model standard in forensic seismology – estimates the source spectrum characteristics of the moment rate function according to surrounding geology, and explosion depth and yield. For small-to-medium explosions in granite, the moment rate stemming from the Mueller-Murphy model (with its flat spectrum followed by a sharp drop that is depth- and yield-dependent) can be well approximated by either a triangle or a Gaussian moment rate function. As shown in Figure 4.6, a Gaussian manifests better frequency characteristics than a triangle, and thus we choose a Gaussian moment rate function for this study. Moreover, given a tight range of yields of the Degelen explosions and uncertainties of those estimates (see table in Appendix C), we decide to use the same width of the Gaussian moment rate for all simulations. We therefore neglect variations in topographic interactions with changes in the moment rate. Such an analysis remains beyond this study and ought to be the scope of future work.

According to the Mueller-Murphy model, explosion source spectra scale with cube root to the yield. Thus, events in the 1-100 kt range – a range representative of Degelen – have similar spectra and corner frequencies of about 1.3-2 Hz. Given our interest in waveforms with energy up to 2 Hz, we require a Gaussian half-width of 0.25 seconds or less. As can be seen in Figure 4.6b, a Gaussian moment rate function of half-duration of 0.25 s has a flat spectrum up to 1 Hz

and reaches the half-maximum of the spectral amplitude at 2 Hz. For such a half-width, the bulk of the energy up to 2 Hz is still contained, and – following the AxiSEM definition – we call the half-maximum of the Gaussian the 'dominant frequency' of the global simulation.

The half-width of the Gaussian determines the mesh resolution of the global AxiSEM simulation used in the coupling, which in turn decides the computational cost of the global simulation. Due to limited computational resources, the database for this study was generated with a 'dominant frequency' of 2 Hz, and captures a high proportion of the global interactions up to 2 Hz. By construction, a Gaussian moment rate function of 0.25 s half-width is on the resolution limit of the chosen global mesh, as some energy is present in the 2-4 Hz band (see Figure 4.6b). Doubling the half-width would ensure that all energy is contained within the 2 Hz limit of the mesh, with the half-maximum of the spectral amplitude at 1 Hz; however, most energy in the 1-2 Hz range would be lost. We thus choose the Gaussian of 0.25 s half-width for all coupled simulations, bearing in mind that some of the energy in the 2-4 Hz range will not be well resolved. In Appendix E we explore and discuss the numerical effects of this choice.

The frequency parameter $freq$ that defines the WPP moment rate function specifies the angular frequency. In order to match the Gaussian of 0.25 s half-width used for AxiSEM database generation, we require $freq = 9.89$, and therefore the largest significant frequency in the time function is

$$f_{max} = \frac{2.5 freq}{2\pi} = 3.97 \tag{4.1}$$

as per the WPP manual (Petersson and Sjogreen, 2013).

4.3 Results of simulations

4.3.1 Teleseismic 1-D waveforms at 2 Hz

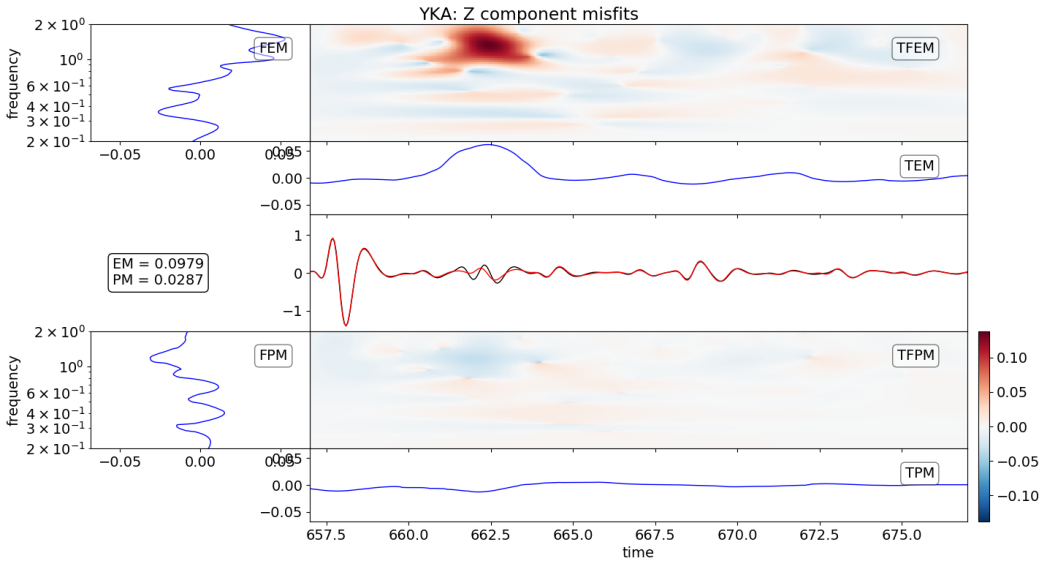
Coupling configuration and benchmarking

As discussed in Section 2.4, repropagating local wavefields to teleseismic distances requires fine-tuning, as an interplay between two factors determines both computational costs and accuracy of final seismograms:

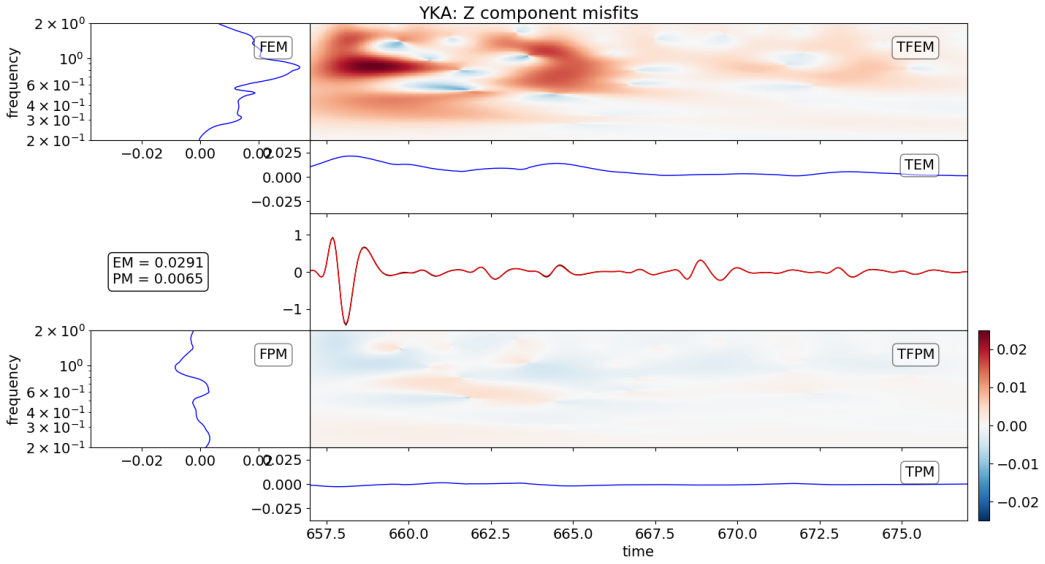
(i) grid spacing of the mesh for local simulations in WPP, and

(ii) density of points on the boundary of the local domain used for repropagation in Instaseis.

In order to embed the Degelen topography in simulations, the local WPP domain needs to be 30×30 km in x and y, and we set the depth to 10 km. Testing this WPP-Instaseis set-up with the 2 Hz database for a 1-D model with a flat surface, we have determined a maximal grid spacing both for the WPP mesh and for the repropagation surface that result in high quality benchmarks relative to the original 1-D Instaseis wavefields. Simulations are performed on a



(a)



(b)

Figure 4.7: 1-D benchmark of the hybrid Z-component seismogram for the WPP-Instaseis coupling (black line) against the AxiSEM 1-D seismogram (red line) at YKA for (a) $30 \times 30 \times 10$ km WPP domain and (b) $10 \times 10 \times 10$ km WPP domain. The WPP domain is discretised with 100 m spacing in x, y and z, while a coarser grid of 200 m is used for Instaseis repropagation. The plot details the frequency-envelope misfit (FEM), the time-envelope misfit (TEM), the time-frequency-envelope misfit (TFEM), the frequency-phase misfit (FPM), the time-phase misfit (TPM), and the time-frequency-phase misfit (TFPM). Misfits in (a) are highest around 30.5–32.5 s and result mostly from the mismatch between the local and global models due to the lack of spherical surface in the Cartesian WPP domain. Misfits are significantly reduced in (b), as the mismatch between the local and global models is smaller.

regular finite difference grid of 100 metres in x, y and z, which for the moment rate function with a maximal frequency of 3.97 Hz (as chosen in Section 4.2.3) results in 8.5 points per wavelength according to

$$\mathrm{PPW} = \frac{v_{min}}{h_{min} * f_{max}} = \frac{3360\mathrm{m/s}}{100\mathrm{m} * 3.97\mathrm{Hz}} = 8.5 \tag{4.2}$$

where PPW is the number of points per wavelength, f_{max} the largest significant frequency of the source time function, v_{min} the minimum wavelength, and h_{min} the FD grid spacing (Petersson and Sjogreen, 2013).

A coarser spacing of 200 metres on all four boundary surfaces is used for Instaseis repropagation of local wavefields to receivers. Such coarse sampling on the boundary reduces the number of points – and thus storage space and repropagation CPU time – by a factor of 4, and results in a high quality of final seismograms at teleseismic distances. An example benchmark at the YKA station location, along with phase and envelope misfits (Kristeková et al., 2009) (PM and EM, respectively), is shown in Figure 4.7a. The chosen simulation and repropagation discretisations result in PM of 2.9% and EM of 9.8% at YKA. Benchmarks for the remaining three stations are shown in Figure F.1 in Appendix F.

A major limitation of the WPP-Instaseis coupling discussed in Section 2.4.3 is WPP Cartesian mesh that does not incorporate the spherical surface of the Earth. For 30×30 km box, the difference in altitude between the spherical Earth and the flat Cartesian surface reaches nearly 145 m at the point furthest from the centre of the local coordinate system. While this might not be problematic for lower frequency simulations, as the "gap" remains small relative to the wavelength, we have found that for the 2 Hz case it is a major source of misfits in Figures 4.7a and F.1. For the same grid parameters used on a smaller domain with horizontal dimensions of 10×10 km (resulting in a maximal altitude difference of 15 m), PM and EM at the YKA array location are reduced to 0.011 and 0.042, respectively (Figure 4.7b). Further refinement tests (e.g. spacing of 50 m in the local simulation resulting in 15 points per wavelength, a configuration recommended for best accuracy in WPP) decrease PM and EM at YKA below 0.7% and 2.5%, respectively (see Figure F.3 in Appendix F for details).

With such refinements increasing the computational costs, and given the lack of spherical surface in the Cartesian box that affects the 30×30 km domain, we choose to accept the marginally lower precision and run local WPP simulations on 100 m mesh. A single WPP simulation with the chosen specification and for a duration of 10 seconds costs 46 CPUh (11.5 minutes on 240 cores). Subsequent repropagation to a single receiver costs about 2 CPUh (10 minutes on 12 cores). A full run for a single source location and four receivers takes 54 CPUh.

All waveforms presented in this chapter are generated with an isotropic moment tensor source, low-pass filtered at 2 Hz with a Butterworth zero-phase filter, and convolved with an instrument response corresponding to each array. The instrument response is convolved into syn-

thetic seismograms to allow for comparisons with the Degelen observations discussed in Chapter 3. While modern feedback seismometers have a more stable response, for old short-period digitized datasets like Degelen deconvolving the instrument response is often unstable (Douglas, 2013), and hence we choose to convolve the response into synthetic waveforms rather than deconvolving it out of real data.

1-D seismic phases

As illustrated by the benchmarks, 1-D seismograms at all four arrays are highly complex, with multiple arrivals following the P wave onset. To better understand the signals and the common phases between the stations, we have calculated their respective travel times using the obspy TauPy package (Beyreuther et al., 2010) that serves as a port of the Java TauP Toolkit (Crotwell et al., 1999). The phase labelling used in the remainder of this chapter follows their convention:

- P and S – upgoing or downgoing compressional and shear waves, respectively; p (s) is a strictly upgoing P wave (S wave).
- Numbers represent depths of interactions. A number that appears between two phase legs, e.g. P20s, represents a transmitted phase conversion at the given depth, not a reflection.
- ˆ and v represent bottom-side and top-side reflections, respectively. They are followed by a number to represent the reflection depth.

Figure 4.8 illustrates propagation paths of phases S35P, pS35P, P20S35P20s and Pv35pˆ20P around the source and under the receiver (described below). Figure 4.9 labels phases present in the first 15 seconds of 1-D seismograms at all four arrays and shows the multitude of contributions that result in the final waveform. At 661.8 seconds of the YKA waveform (Figure 4.9c), for example, twelve phases contribute to the visible arrival:

- P35s, a downgoing P wave converted at the 35 km discontinuity under the receiver to an upgoing S wave, and its corresponding surface reflections pP35s and sP35s.
- S35P, symmetric to P35s, a downgoing S wave converted under the source to a downgoing P wave at the 35 km discontinuity, and its corresponding surface reflections pS35P and sS35P.
- P20S35P20s, a downgoing P wave converted at the 20 km discontinuity under the receiver to a downgoing S wave, which is almost immediately converted back to a downgoing P wave at the 35 km discontinuity. Under the receiver, it is then converted at the 20 km discontinuity to finish as an upgoing S wave. Along with its corresponding surface reflections pP20S35P20s and sP20S35P20s.

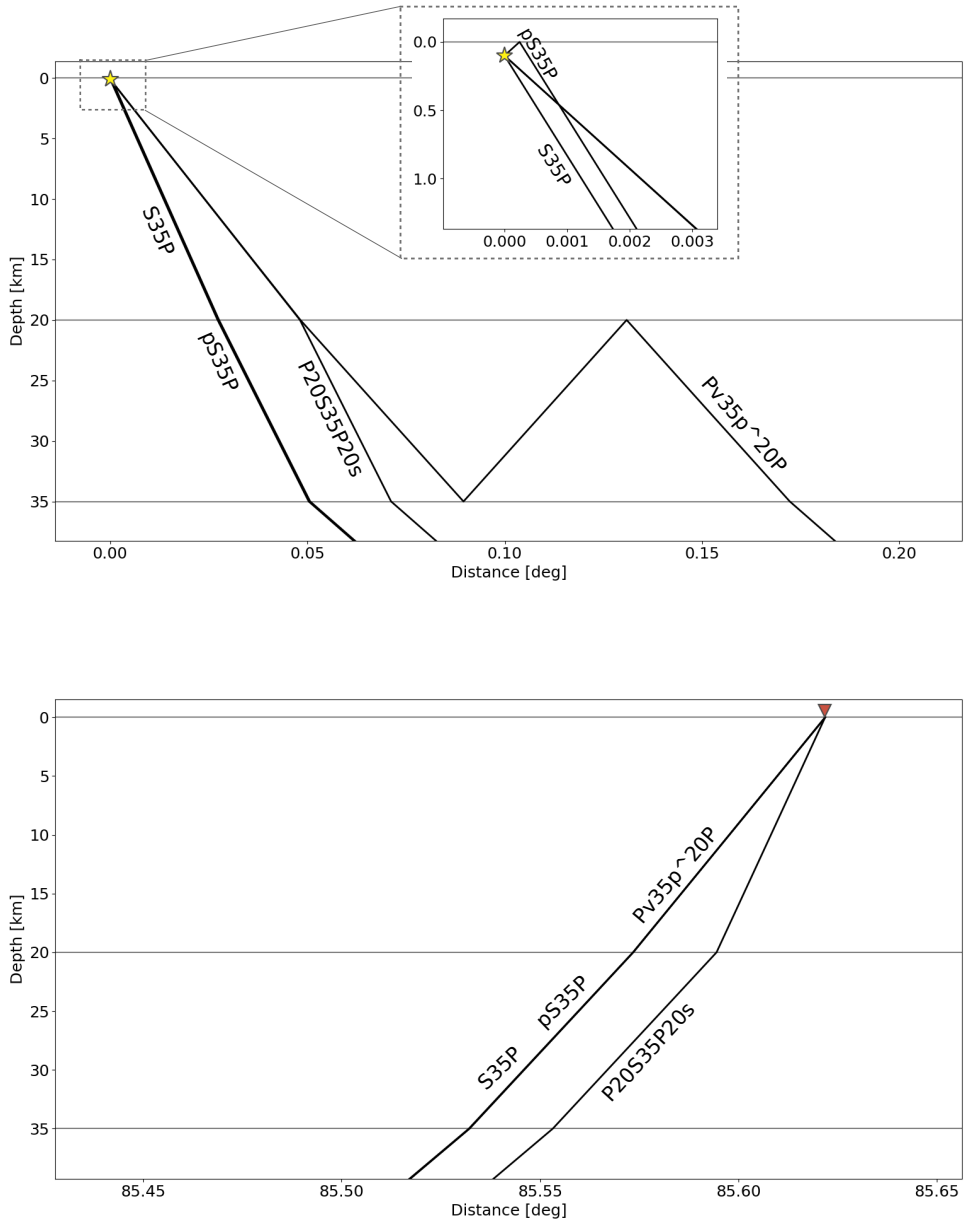
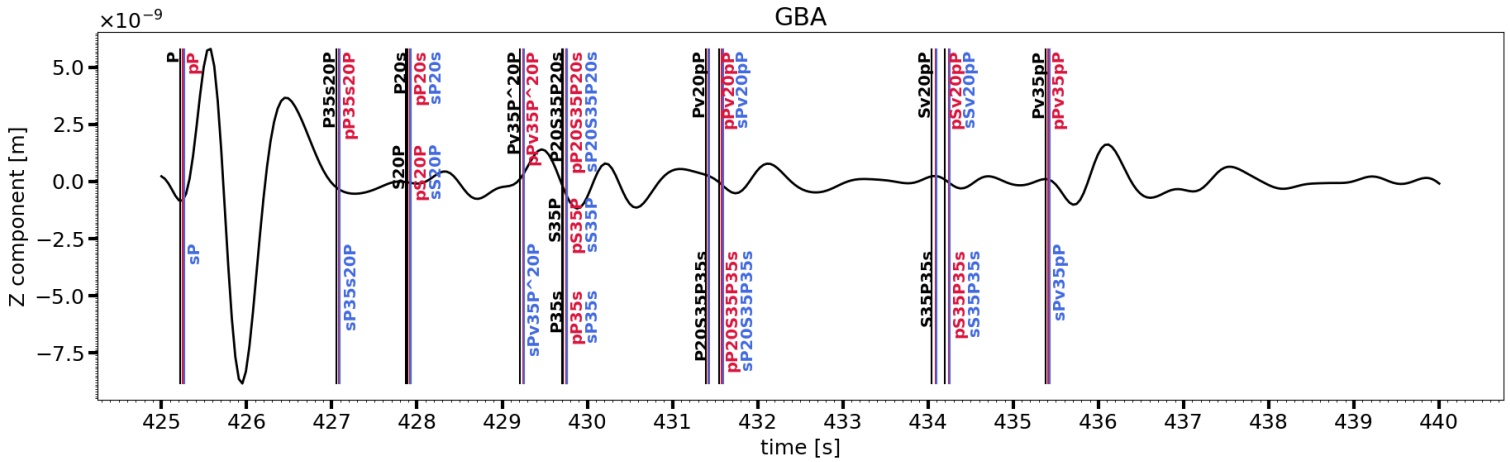
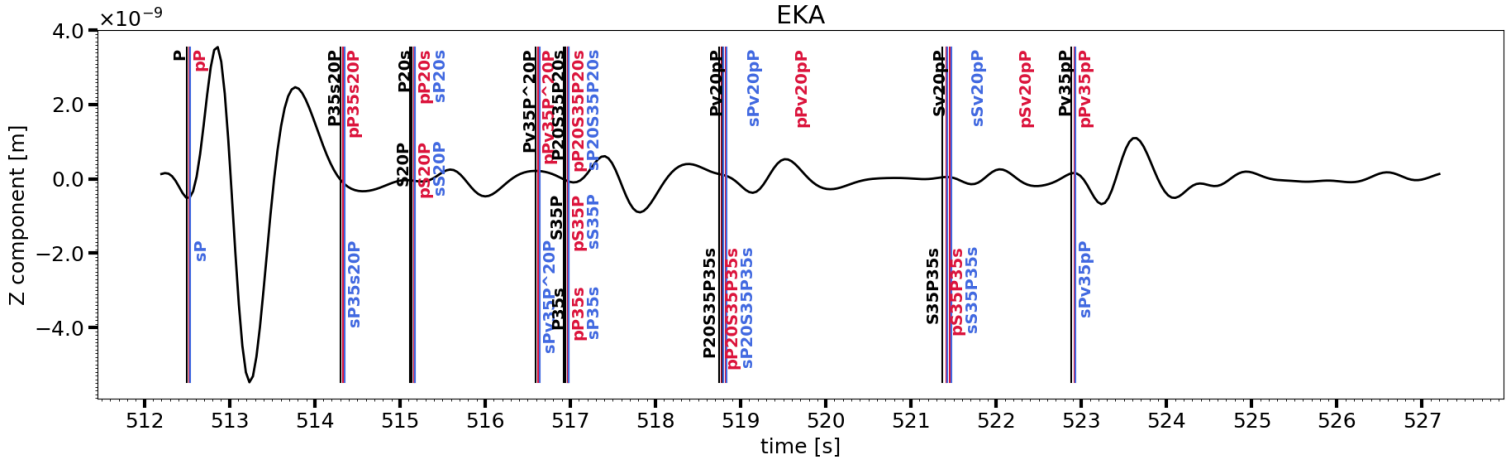


Figure 4.8: Illustration of four seismic phases present in the first 15 seconds of 1-D seismograms: S35P, pS35P, P20S35P20s and Pv35pˆ20P. The top and bottom panels show the paths up to 40 km under the source and under the receiver, respectively. A detailed description of the interaction of the four phases with the 20 km and 35 km discontinuities can be found in the main text.

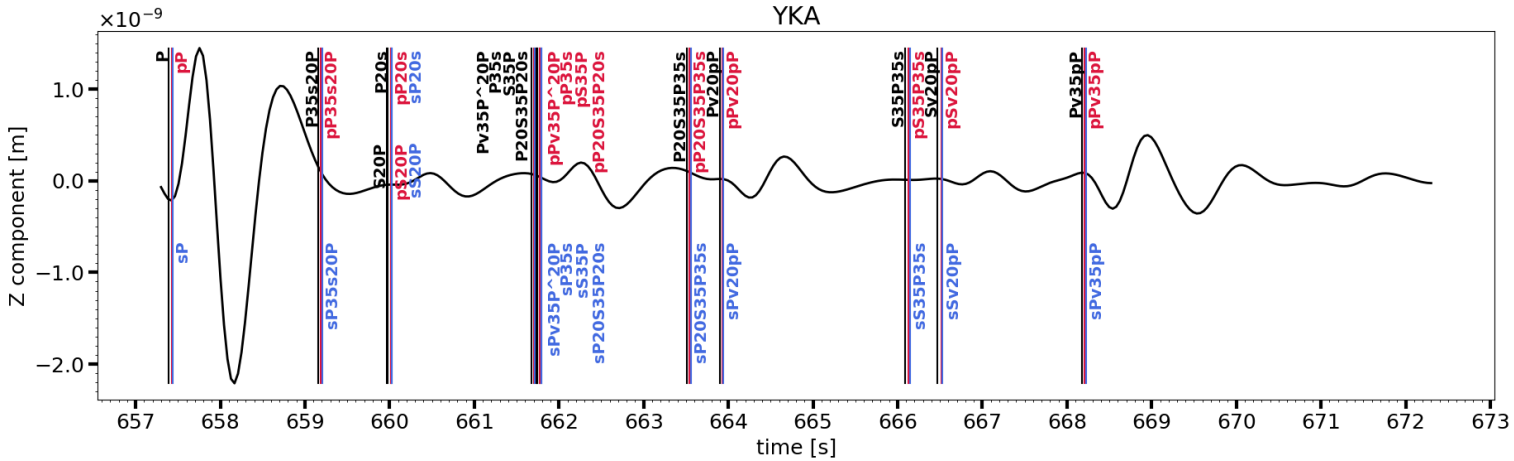


(a)

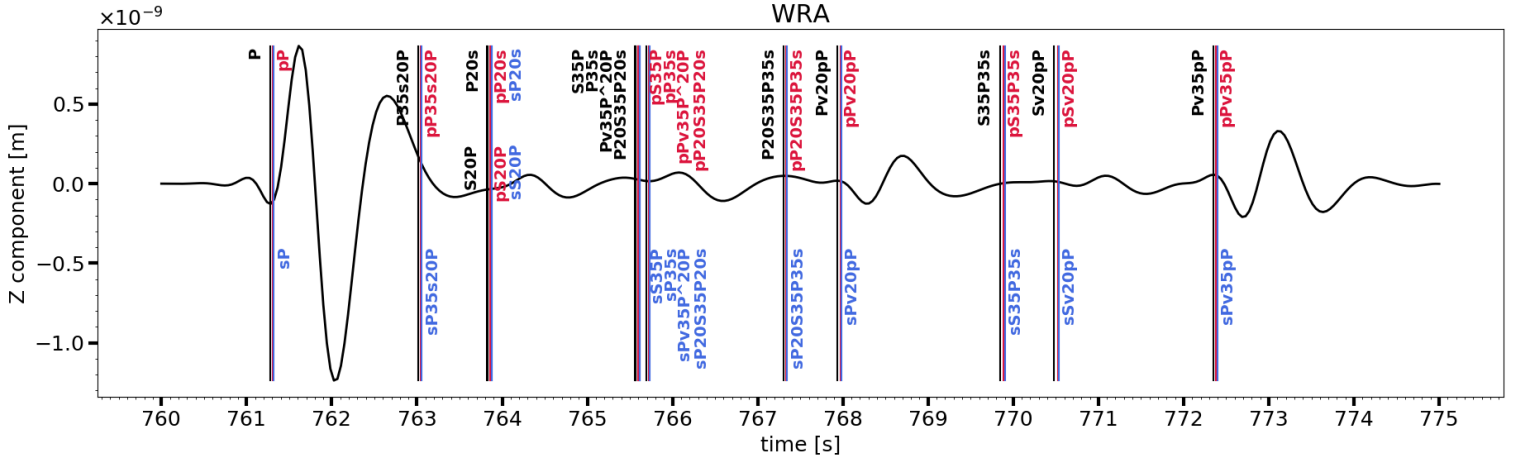


(b)

Figure 4.9: Arrival times of phases present in the first 15 seconds of IASP91 1-D seismograms at (a) GBA, (b) EKA, (c) YKA, and (d) WRA. Phases leaving the source as downgoing P or S waves are marked in black, phases leaving the source as upgoing P waves are marked in red, and phases leaving the source as upgoing S waves are marked in blue. Refer to Section 4.3.1 in the main text for a full description of the labelling convention.



(c)



(d)

Figure 4.9: (continued).

- Pv35pˆ20P, a downgoing P wave reflected back up as a P wave at the Moho, then reflected down as a P wave at the 20 km discontinuity towards the receiver. Along with corresponding surface reflections pPv35pˆ20P and sPv35pˆ20P.

Arrivals corresponding to the same phases can also be identified on the other arrays: at 429.2 and 429.7 seconds at GBA (Figure 4.9a), at 5176.6 and 517 seconds at EKA (Figure 4.9b), and at 765.7 seconds at WRA (Figure 4.9d).

Given that the model is spherically symmetric, depths of discontinuities under the source and under the receiver are the same and some phases arrive simultaneously. For example, the P35s phase is symmetric to S35P. Both waves travel 35 km as an S wave and cover the rest of the path as a P wave, with P35s travelling as an upgoing S wave under the source, while S35P as a downgoing S wave under the receiver. In a real Earth, however, the discontinuity depths might vary between the source and the receiver, breaking such symmetries. As a result, rather than culminating in single peaks of higher amplitudes, the phases would appear as separate lower-energy arrivals at distinct times.

Depth-dependent amplitudes for shallow sources: the P and pP interference

An analysis of 1-D teleseismic waveforms at the four array locations revealed a strong depth dependence of the peak-to-peak amplitudes of the first P wave arrival for shallow sources. Explosions located within a wavelength from the free surface exhibit large variations due to the interference of the P and pP phases. For depths larger than a wavelength, the pP phase travels an extra distance of over two wavelengths: at least a wavelength from the source to the free surface, and then that same additional distance back down towards the receiver. It therefore appears on the seismograms as a distinct arrival after the direct P. As a result, for sources deeper than a wavelength the peak-to-peak amplitude of the first arrival contains only the energy of the direct P phase and stabilises at a constant value. An example of the evolution of the waveform shape at YKA relative to the source depth is shown in Figure 4.10.

Figure 4.11 further explores the effect of the destructive and constructive interference of P and pP phases on amplitudes of 2 Hz waveforms at GBA, EKA, YKA and WRA locations. Absolute amplitudes vary with epicentral distance due to attenuation, but the trend is nearly identical. The maximum value of the peak-to-peak amplitude is attained for depths close to half a wavelength where the strongest constructive interference takes place, while amplitudes stabilise for sources deeper than a wavelength. The exact source depth corresponding to the maximum peak-to-peak value at each station depends on the take-off angle that affects the distance travelled by the pP phase to the surface. As shown in Figure 4.12, for steeper take-off angles the pP phase travels more vertically and covers a distance closest to that of the source depth, so the maximal positive interference happens for sources half a wavelength under the free

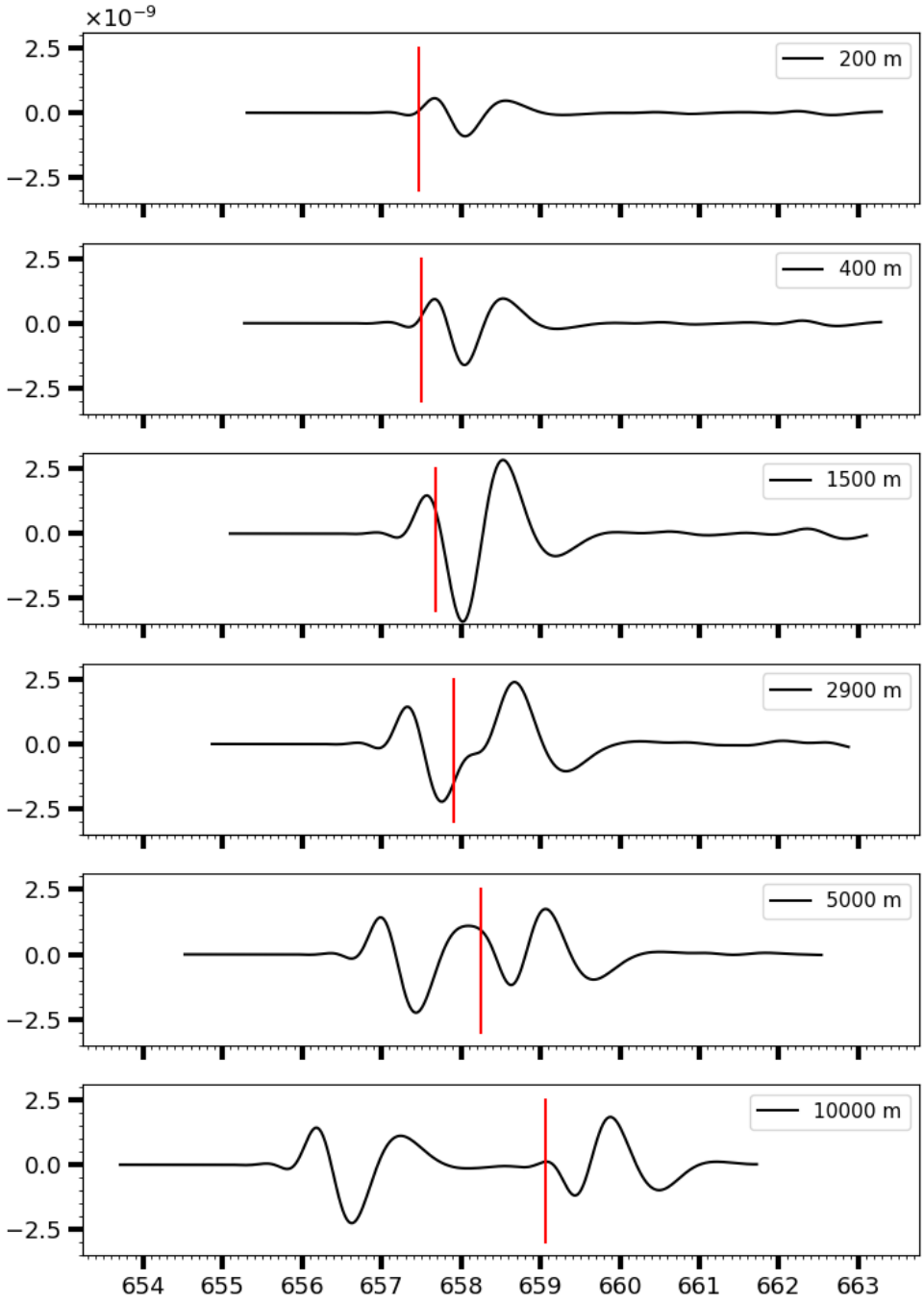
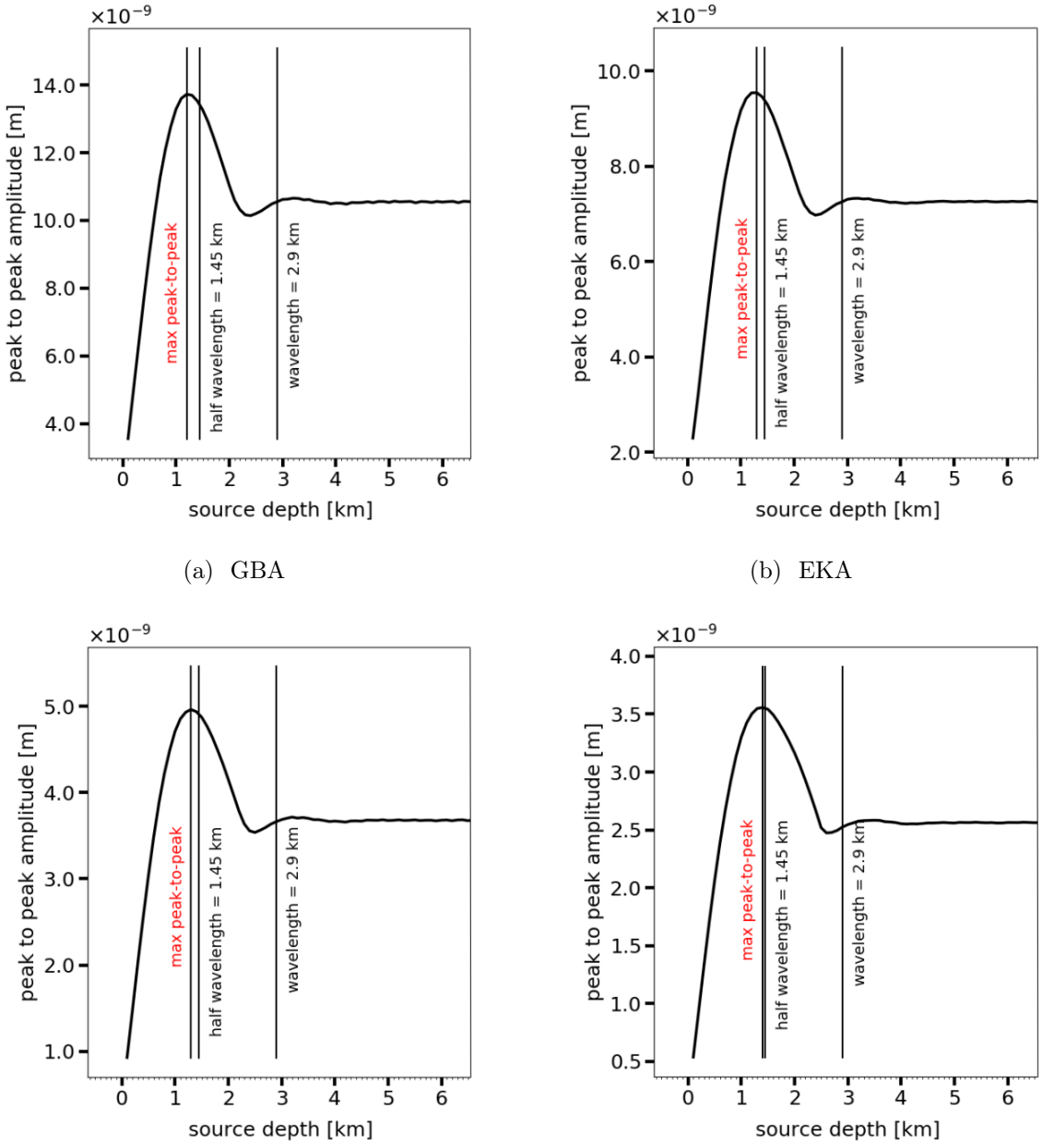


Figure 4.10: Evolution of 1-D waveforms at YKA with increasing source depth. In the top two panels (for depths of 200 m and 400 m), a negative interference of P and pP significantly reduces the peak-to-peak amplitude of the first arrival. With the source at 1500 m – close to half of the wavelength at 1450 m – the negative arrival is strongly amplified due to positive interference. For depths below a wavelength (bottom three panels), the pP phase appears later until it eventually separates into a distinct arrival and the peak-to-peak amplitude measures only the energy of the direct P wave. For each source depth, the arrival time of pP is indicated with a red line.



(a) GBA

(b) EKA

(c) YKA

(d) WRA

Figure 4.11: Peak-to-peak amplitude of the first arrival as a function of source depth at (a) GBA, (b) EKA, (c) YKA, and (d) WRA. Absolute amplitudes vary with epicentral distance due to attenuation, but the trend is nearly identical. The maximum value of peak-to-peak amplitudes is attained for depths close to half a wavelength where the strongest constructive interference takes place, while amplitudes stabilise for sources deeper than a wavelength. The exact source depth corresponding to the maximum peak-to-peak value at each station depends on the take-off angle which affects the distance travelled by the pP phase to the surface (see Figure 4.12). The stable trend continues beyond the depths shown.

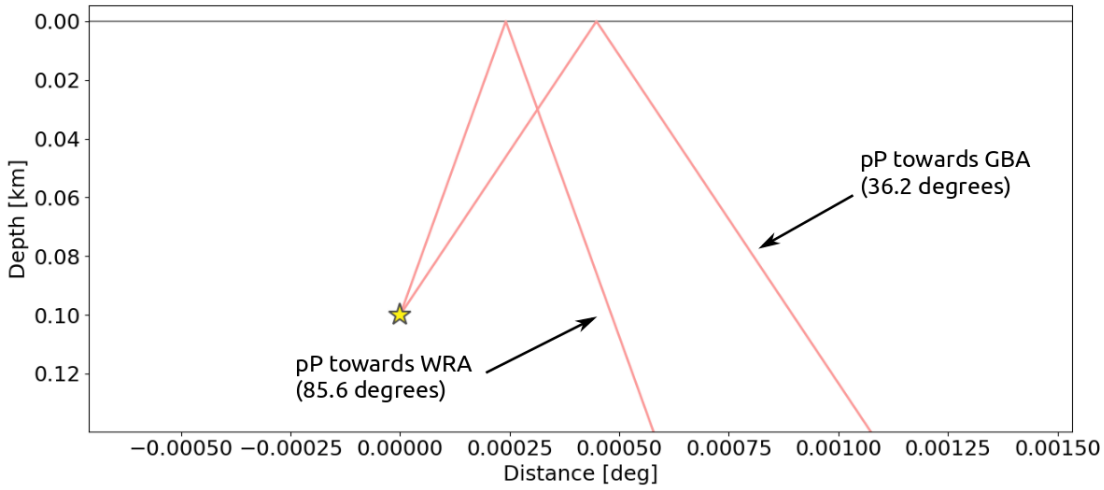


Figure 4.12: Illustration of the effect of take-off angle on the path of pP. For a steeper take-off angle towards WRA, the pP phase travels more vertically and covers a distance closer to that of the distance between the source and the free surface. For a larger take-off angle towards GBA, the path of the pP phase towards the surface is further from the vertical. Thus, pP covers a distance marginally larger when travelling towards GBA than when travelling towards WRA.

surface (as in 4.11d). For shorter epicentral distances, the path of the pP phase towards the surface is further from the vertical, and hence pP covers a distance marginally larger than the vertical distance between the source and the free surface. Therefore, for stations closer to the explosion, the maximum interference takes place for sources marginally shallower than half a wavelength (as in 4.11a-4.11c).

It should be noted that for shallow sources this positive and negative interference affects amplitudes of other arrivals within the 15 second time window of interest. Re-examining Figure 4.9 where the arrival times for a source at 100 m depth are identified, we can see that Pv35pP arrives with the surface reflections sPv35pP and pPv35pP. Given the negative interference of P and pP, the Pv35pP and pPv35pP phases (and the like) undergo the same effect. Thus, in a 1-D model, the amplification or reduction of the amplitude of the first arrival is going to be carried over to subsequent phases.

The results presented here are for an isotropic moment tensor source, have been low-pass filtered at 2 Hz (Butterworth zero-phase filter) and an instrument response has been applied at each station. The exact trend of amplitude reduction and amplification relative to the source depth depends on the choice of the frequency band and of the source radiation pattern. For example, a vertical single force shows a different behaviour with depth than an isotropic moment tensor, but a detailed analysis is beyond the scope of this chapter.

4.3.2 Topographic effects of Degelen at teleseismic distances

To explore the effects of the Degelen mountain topography on teleseismic waveforms, we have selected 24 event locations. Location groups A (6 locations), B (5 locations) and D (3 locations) are on the southern ridge, while groups C (6 locations) and E (4 locations) are on the northern ridge. They span the highest Degelen mountain peaks, and have been selected to include locations under the main peaks and along their slopes. All events have been simulated with an isotropic moment tensor source with a scalar moment of 10^{16} Nm buried 100 m under the topography. Locations in group C have also been simulated with sources at 150 m and 250 m under the topography to explore the effects of depth on waveforms. Distances between locations in each group vary from 100 m in group E to nearly 1000 m in group D. Exact locations of events, as well as an overview of the resulting waveforms at the YKA array, are shown in Figure 4.13. Analogous plots for GBA, EKA and WRA can be found in Figures G.1-G.3 in Appendix G.

Note that the locations have been selected to span various peaks, and to account for different distances between events. Given the uncertainty in exact locations and in depths of the Degelen explosions, and considering the simplicity of our model, we choose in the first instance to focus on examining the general teleseismic effects of topography in an elastic and attenuating medium, and do not intend to reproduce specific waveforms from the observed dataset.

Figures 4.13 and G.1-G.3 illustrate the extent to which the Degelen topography affects teleseismic synthetic signals. First arrivals are either amplified or deamplified, while coda waves change with location on the mountain ridges: distinct phases appear and disappear, some arrivals shift, and some change amplitudes. As illustrated by events in group E, however, a change in location of only 100 m – on the order of our mesh resolution – seems to bear little effect, suggesting that closely located events exhibit similar features. In the following, we explore in detail the properties of synthetic YKA seismograms with respect to the Degelen topography, with a similar analysis valid for the other arrays.

Emerging phases and a location-dependent coda wave-train

The most prominent feature of all synthetic YKA waveforms – highlighted in red in Figure 4.14 – is the drastic change in the second positive arrival at around 658.5 seconds (the exact time depends on the specific location). The topography influences the pP and sP free surface reflections, which in turn lead to changes in arrival times of the pP35s20P and sP35s20P phases relative to P35s20P. In some locations the second positive peak only broadens (locations A, Figure 4.14b), in some it deforms into a step-like double-peak (locations B, Figure 4.14c and Figure 4.14d), in some a more pronounced "split" double-peak forms (locations E, Figure 4.14g), while yet in others a distinct arrival emerges (e.g. location D1, Figure 4.14f).

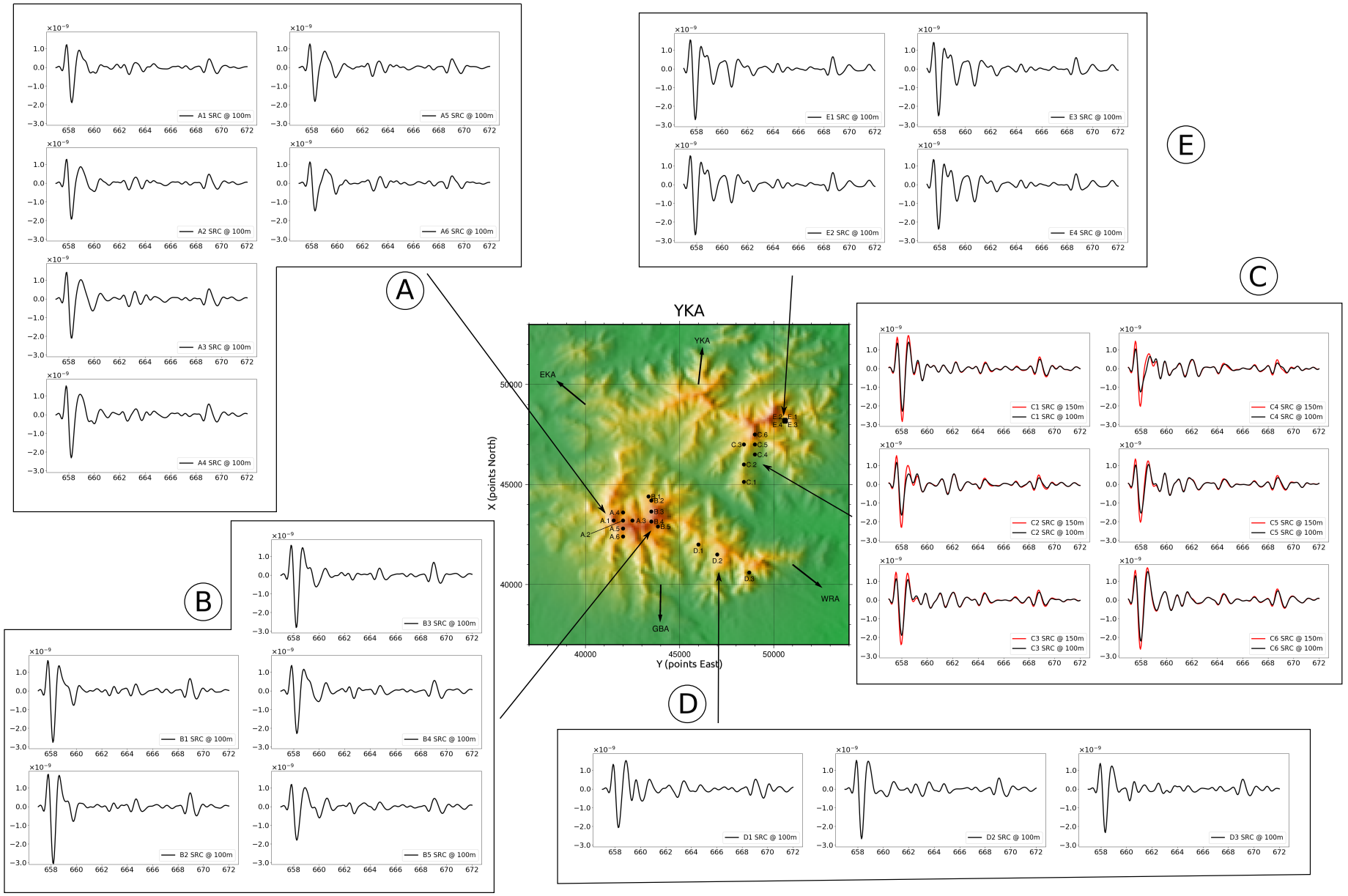


Figure 4.13: Overview of synthetic waveforms from the Degelen mountain at the YKA array. First arrivals are either amplified or reduced, and coda waves change with location – new arrivals appear and disappear, some shift, and some change amplitudes. The direction from Degelen to YKA is indicated with an arrow.

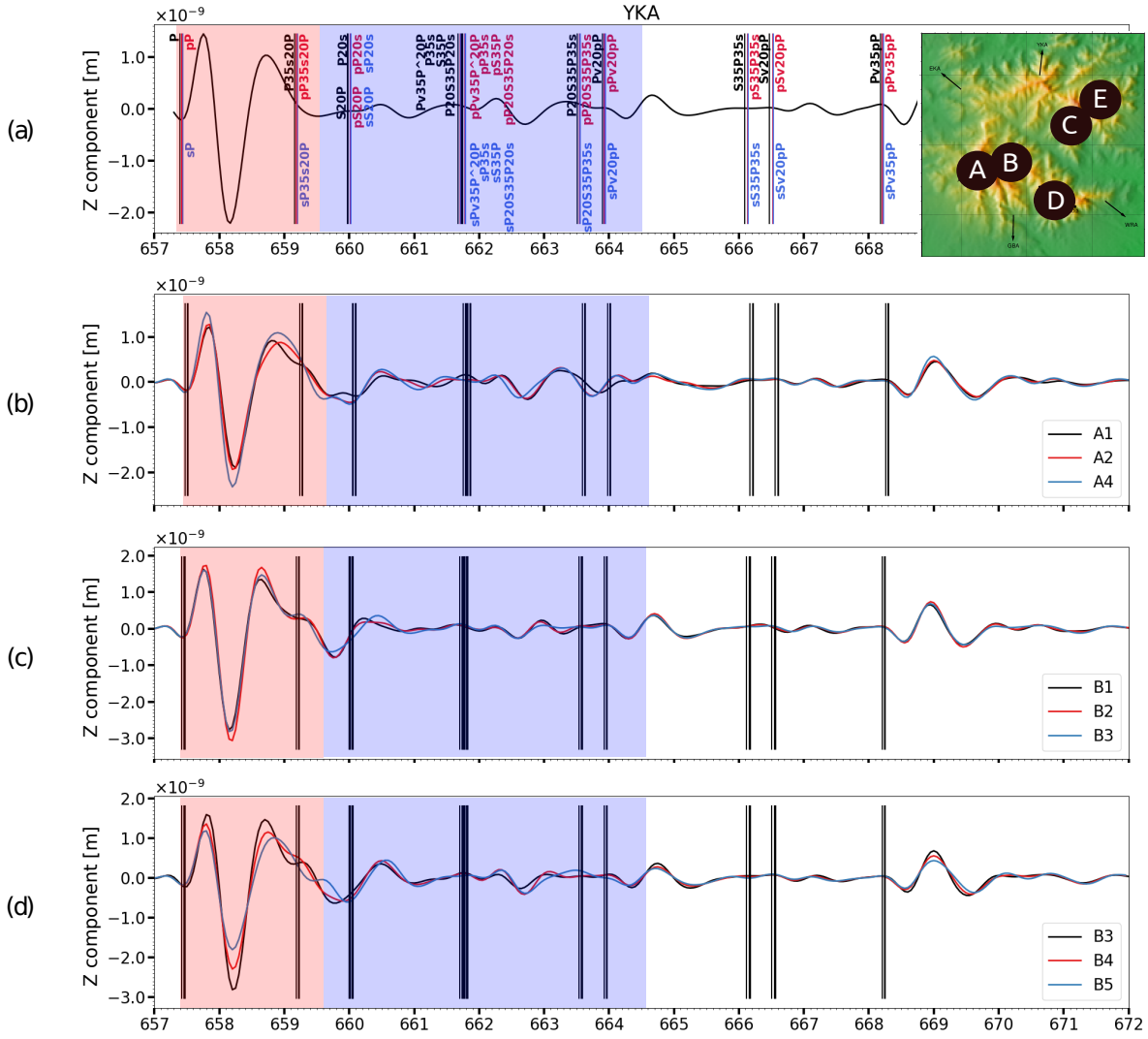


Figure 4.14: Detailed comparison of waveform differences caused by the Degelen topography at the YKA array. The first two positive peaks are highlighted in red, while the following five seconds are highlighted in blue. Seismograms are aligned to the black trace on each plot. (a) 1-D YKA waveform and arrival times of phases, as well as a map highlighting locations of source groups used in simulations. (b) A1, A2 and A4 located at most 500 m apart. The first arrival is broadened (red area), while subsequent phases are shifted relative to each other (blue area). (c) B1, B2 and B3 located at most 500 m apart, traversing a peak southwards towards the centre at B3. The waveforms are remarkably similar. (d) B3, B4 and B5 located at most 500 m apart, traversing a peak southwards away from the centre at B3. A progressive evolution in the shape of the second positive peak (red area) can be observed.

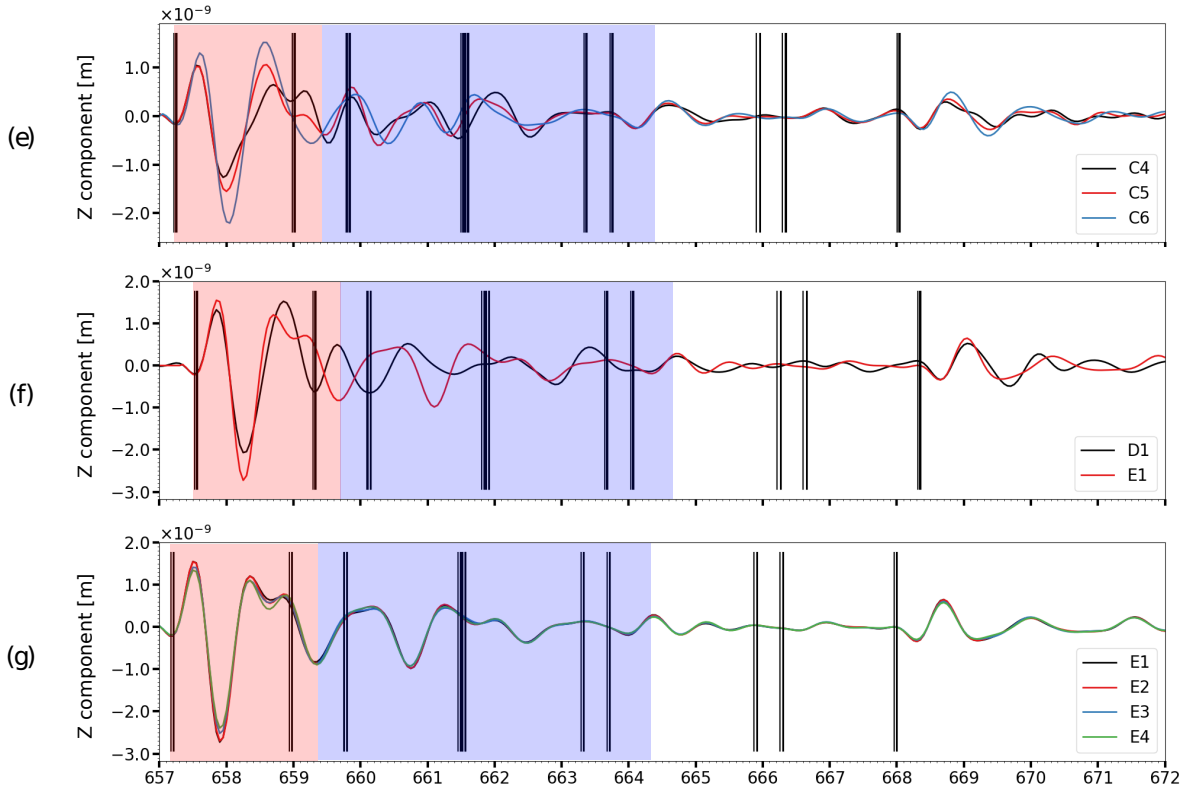


Figure 4.14: (continued). (e) C4, C5 and C6, located at most 400 m apart. The location shift causes prominent changes in the first 2 seconds (red area), but subsequent scattering has more similar features. (f) D1 and E1 located 8 km apart, shown against each other to highlight waveform differences for distant events. (g) E1-E4, located within a 100 m of each other, with nearly identical waveforms.

Generally, events in the same location group – i.e. within the same mountain – tend to have similar shapes of the second positive arrival (groups A, B and E). Events in group E are located within 100 m of each other and result in nearly identical seismograms (Figure 4.14g). Events in group B are located across one of the peaks on the southern ridge, with a maximum distance of 450 m between the neighbouring points: B1 and B2 are on the northern slope, B3 is in the centre, and B4 and B5 are on the southern slope. Figures 4.14c and 4.14d illustrate the progressive change of the second positive arrival as locations in group B move across the mountain peak. Events in group C, on the other hand, show that a location shift of only 400 m can lead to more pronounced changes than those present in group B. As can be seen in Figure 4.14e, the shape changes from a prominent double-peak at C4, through a step-like double peak at C5, to a single narrow peak at C6.

Figure 4.14 also highlights in blue the four seconds in the 659.5-664.5 s range – following the second peak or the double-peak – that exhibit a large variety of waveform shapes and amplitudes. Most notably, events in location group E bear two high-amplitude arrivals at 659.5 s and 661 s (Figure 4.14g). In other groups, a larger number of distinct arrivals of lower amplitudes is present. We postulate that those arrivals emerge as a superposition of primary and secondary (multiple) surface reflections of the P wave. As indicated in Figure 4.14a, many primary reflections arrive between 659.5 and 664.5 seconds (e.g. sSv20pP, pP35s). Topography alters their propagation paths and arrival times, affecting final seismograms. Moreover, apart from changing take-off angles of such primary reflections, the mountains cause multiple reverberations of the wavefield between their slopes. As waves reverberate, various multiple reflections progressively leave the source region towards the receiver. Reverberations reaching the stations shortly after the P, pP and sP phases have the highest energy and hence variations are the most prominent in the the 659.5 - 664.5 s range. Later arriving coda is also affected, as multiply scattered waves are present throughout the trace, but reverberations lose energy with time and the effect weakens. The idea of reverberations of mountain peaks is further discussed in Section 4.3.3, where we examine the evolution of the local wavefield.

The agreement both of the first P wave arrival and its coda between seismograms generally appears to be a good indicator of the proximity of the sources. In Figure 4.14f, events D1 and E1 located 8 km apart are shown against each other to highlight waveform differences for distant events. Events within 100 m, like E1-E4, result in very small differences (Figure 4.14f), while slightly larger distances can introduce consistent changes. Notice, for example, the clear coda shift in the blue region of Figure 4.14b between A1 and A4 that are about 500 m apart. We attribute such consistent shifts to wavefield reverberations within the mountain – the location change under the same summit excites reverberations in a slightly different place and causes the shift, but the wavefield nonetheless interacts with the same slopes in a similar manner.

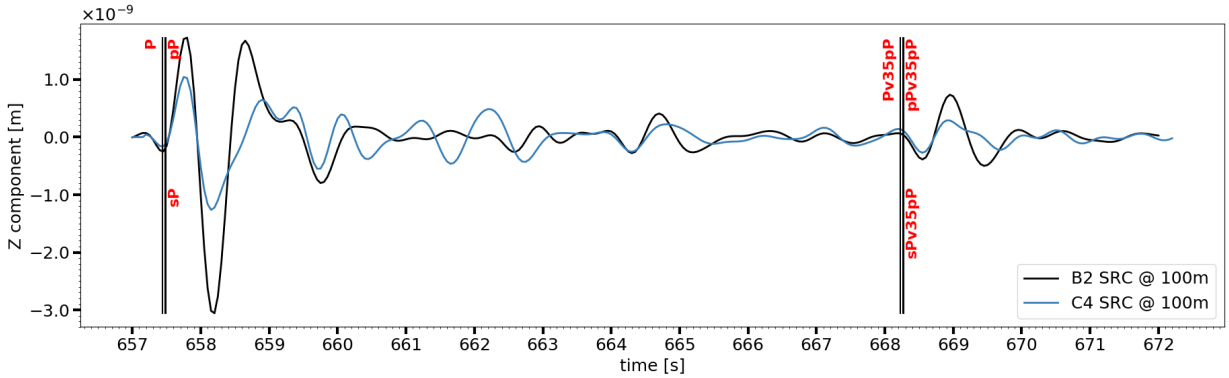


Figure 4.15: Event B2 (black line) experiences the largest amplification of the first arrival at YKA, while event C4 (blue line) the largest reduction among the synthetics at 100 m depth. We conclude that topography influences pP and sP phases, and thus the interference of the direct P with its primary surface reflections changes drastically. As a result, a factor 2 difference in peak-to-peak amplitudes of the first arrival can be observed. Notice that this amplification and reduction also strongly influences amplitudes of the Moho reflections Pv35pP, pPv35pP and sPv35pP.

Location-dependent amplitude changes

Depending on the source location, topographic variations can result either in a reduction or in an amplification of the first arrival. The peak-to-peak amplitude of the 1-D YKA first arrival for a source at 100 m is 3.67×10^{-9} m (maximum positive amplitude: 1.46×10^{-9} m; maximum negative amplitude: -2.21×10^{-9} m; see Figure 4.9c). Some of the events with sources buried 100 m under the topography retain a similar peak-to-peak amplitude at YKA, while others are strongly amplified or reduced. Among all simulations 100 m under the topography, the strongest amplification at the YKA array is at location B2, with the peak-to-peak amplitude increased by a factor 1.3 relative to the 1-D waveform. The strongest reduction at YKA comes from location C4, with the peak-to-peak amplitude a factor of 1.6 smaller. Figure 4.15 compares the B2 and C4 YKA seismograms to illustrate the striking factor 2 difference between the largest and the smallest amplitudes of the first arrival.

Figure 4.15 also shows that amplification and reduction strongly influence the arrival of the Moho reflections Pv35pP, pPv35pP and sPv35pP at 668.2 seconds. The effect is clearly present when visually comparing seismograms from all locations – see for example plots in Figure 4.14 – although the interactions are too complex to be captured by a ratio of the peak-to-peak amplitude of the first arrival to the peak-to-peak amplitude of the Moho reflection. The ratio oscillates between 3.22 and 4.31 for YKA synthetics at 100 m depth.

Following observations in Section 4.3.1 on constructive and destructive interference of P, pP

and sP phases in the 1-D model, we conclude that topographic variations influence the arrival of the primary surface reflections of the P wave and change the interference pattern. It is difficult to classify, however, how any specific topographic features modify the waveforms. Some of YKA's largest first arrivals come from locations right within the mountain, beneath its peak (e.g. B2, see Figure 4.13), while some from locations on the mountain side, under the slope (e.g. E1, see Figure 4.13). The same can be said about seismograms where the first arrival amplitudes are reduced. Unsurprisingly, amplifications and de-amplifications also vary with take-off angle and azimuth, and hence the effects differ between the arrays. For example, at EKA the B2 location also results in a strong amplification of the first arrival, while at GBA and WRA the first arrival from B2 is de-amplified (see Figures G.1-G.3). We postulate that slope steepness and orientation play a major role in the pP and sP reflections which change the interference with the direct P wave, but this hypothesis remains to be systematically tested.

Depth-dependent amplitude changes

Locations in group C have been simulated with explosions 100 m, 150 m and 250 m under the mountain range to test how the topographic influence changes with depth. As shown in Figure 4.16, increasing the source depth increases the peak-to-peak amplitude of the first arrival (highlighted in blue), as well as of the arrivals corresponding to the reflections of the P, pP and sP phases off the 20 km discontinuity (at 664.5 s) and off the Moho (at 668.2 s). As in the 1-D case discussed in Section 4.3.1, the varying interference of the direct P wave with its primary surface reflections is the cause of such strong amplitude variations.

The other arrivals – most notably those in the 659.5-664.5 s window (highlighted in blue) that vary significantly with location in the mountain complex – remain nearly unaffected by the change in source depth. Some differences are present, although they are barely visible on the seismograms. When discussing the phases that emerge due to the topography, we have suggested that the variation in P-coda that follows the first arrival is a combination of two effects: the topography alters the propagation paths and arrival times of the primary reflections of upgoing P and S waves, and it also generates multiple reflections while the wavefield reverberates within the mountain peaks.

If the primary reflections were the principal source of the energy arriving in the 659.5-664.5 s window, a change in source depth should result in an amplification effect similar to the one on the first arrival and the discontinuity reflections (at 664.5 s at 668.2 s). We therefore conclude that the coda energy mainly comes from reverberations in the mountain peaks. Increasing the source depth by a mere 50 m or 100 m – similarly to moving the source by a 100 m horizontally, like in location group E (Figure 4.14g) – results in reverberation patterns indistinguishable from each other at teleseismic distances.

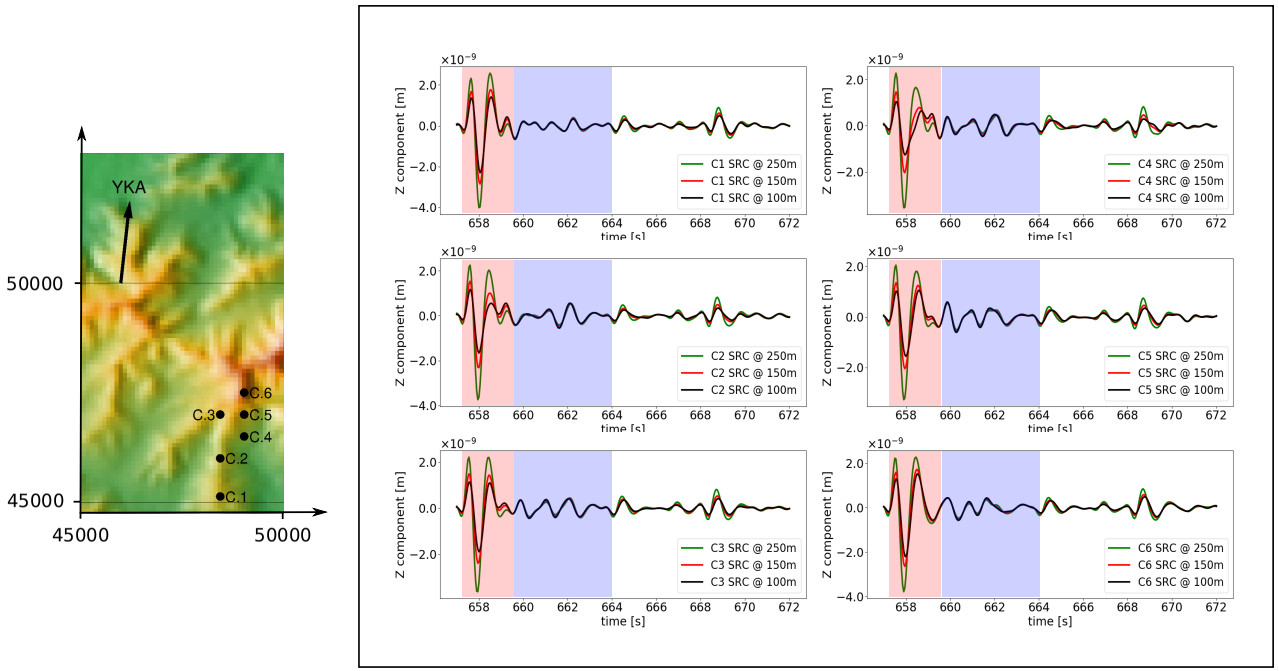


Figure 4.16: Comparison of explosions simulated at locations in group C with depths of 100 m, 150 m and 200 m. Increasing the source depth increases the peak-to-peak amplitude of the first arrival (highlighted in blue), as well as of arrivals corresponding to reflections of the P, pP and sP phases off the 20 km discontinuity (at 664.5 s) and off the Moho (at 668.2 s). Arrivals highlighted in blue – that vary significantly with location in the mountain complex – remain nearly unaffected by changes in source depths.

4.3.3 Local source-topography interactions at Degelen

In the previous section we have suggested that the coda waves (after the first 2 s) are mainly affected by the reverberations of the wavefield within the Degelen peaks. Figures 4.17 and 4.18 illustrate such interactions with the topography at location C4 and demonstrate how the mountains affect the wavefield. The cross-sections intersect the source location (16.5 km, 19.0 km): Figure 4.17 is the cross section for y fixed at 19 km, and Figure 4.18 for x fixed at 16.5 km. Both cross-sections are crucial to understanding the structures that capture the wavefield. For example, at $t = 0.168$ s and $t = 0.185$ in the y cross-section in Figure 4.17, the wavefield is clearly retained by one slope. Only looking at corresponding time steps in Figure 4.18 reveals the two reverberating peaks.

As expected, in the first 0.3 seconds the strongest interactions take place under the peaks right above the source, and the initial symmetry of the wavefield is broken by irregular structures. As the wavefield travels outwards, it encounters further topographic features that reverberate. For example, such interactions are clearly visible on the x component in Figure 4.17 after $t = 0.437$ s. Unsurprisingly, the z component retains most of its symmetry, although upon closer inspection the wavefield is clearly affected by the topography as it propagates away to the receiver. For completeness, similar plots are included for location B3 in Appendix H. A single location demonstrates the properties of the local interactions, but only comparing two locations highlights differences in emerging wavefields. Note the horizontal exaggeration of the plots that visually deforms the propagation of the circular wavefront.

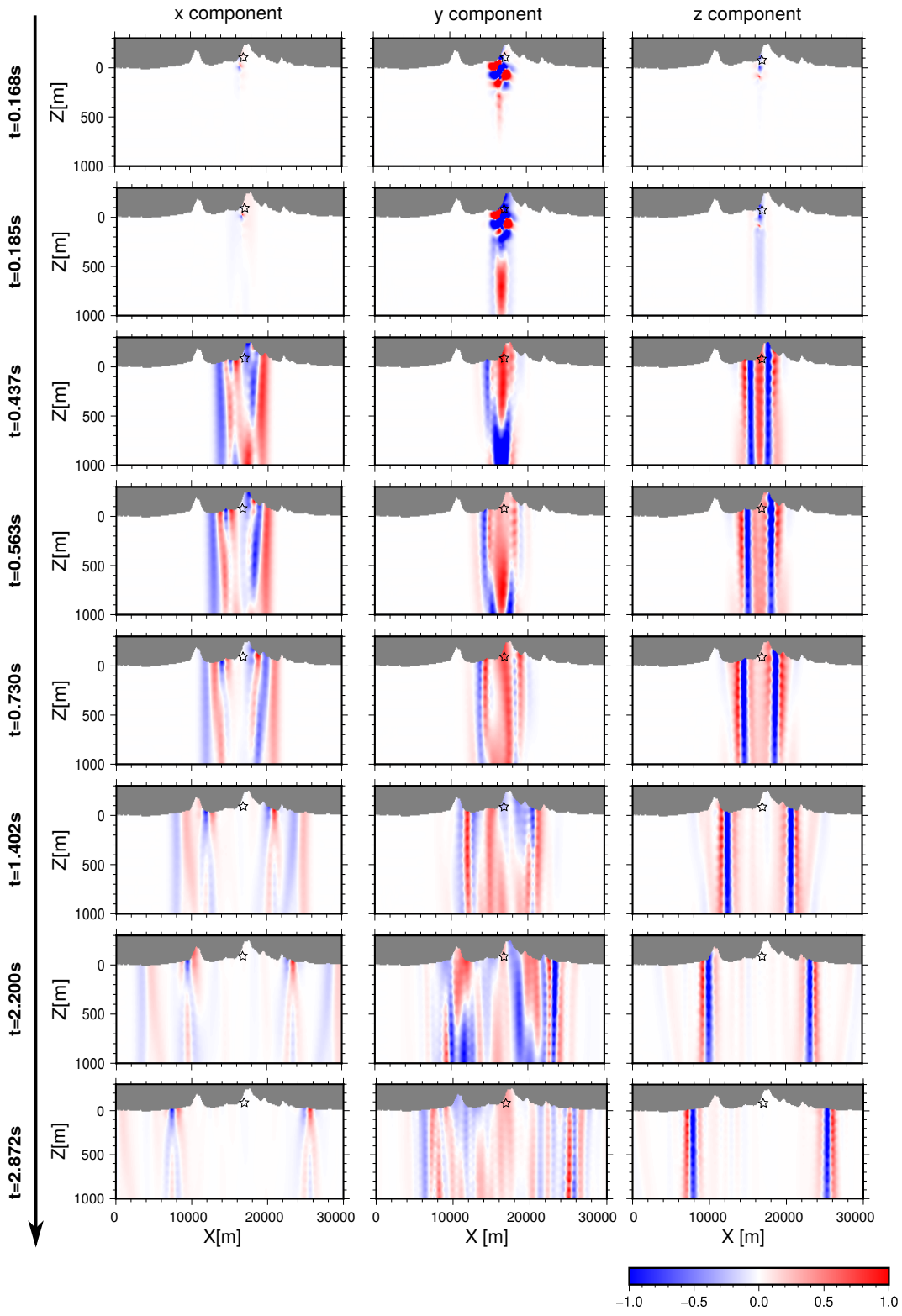


Figure 4.17: Cross sections of displacement for $y = 19$ km for a source located at C4 at $(16.5, 19.0)$ km. All plots have been normalised by the maximum displacement amplitude at the given time step. Due to the domain size, the wavefield is horizontally exaggerated. The visible ringing is a result of a coarse WPP mesh of 8.5 elements per wavelength. As discussed in Section 4.3.1, refining the WPP mesh showed only a marginal improvement teleseismically.

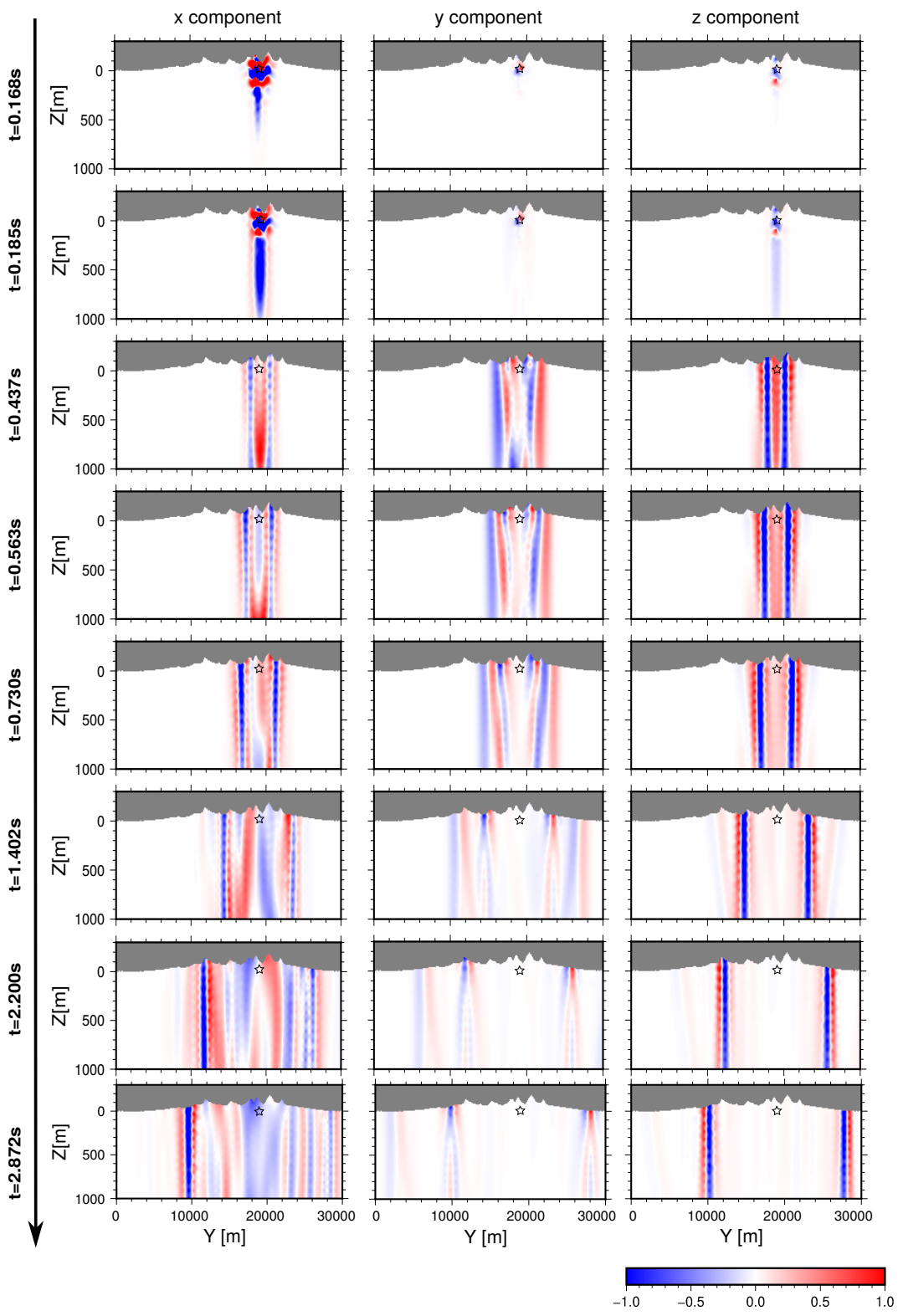


Figure 4.18: Cross sections of displacement for $x = 16.5$ km for a source located at C4 at $(16.5, 19.0)$ km. All plots have been normalised by the maximum displacement amplitude at the given time step. Due to the domain size, the wavefield is horizontally exaggerated. The visible ringing is a result of a coarse WPP mesh of 8.5 elements per wavelength. As discussed in Section 4.3.1, refining the WPP mesh showed only a marginal improvement teleseismically.

array	station distance in degrees	observed amplitudes	synthetic amplitudes	correction factor	adjusted synthetic amplitudes
GBA	36.4	2.89 – 12.5 E-08	1.10 – 2.45 E-08	6.5	7.15 – 15.93 E-08
EKA	47.1	8.47 – 15.9 E-08	6.36 – 14.3 E-09	8.5	5.41 – 12.16 E-08
YKA	67.0	3.90 – 12.6 E-08	2.31 – 6.58 E-09	14.0	3.23 – 9.21 E-08
WRA	85.3	2.87 – 8.96 E-08	1.62 – 4.78 E-09	21.0	3.40 – 10.0 E-08

Table 4.2: Correction of synthetic peak-to-peak amplitudes for t^* representing more realistic attenuation for high frequencies. The ranges of peak-to-peak amplitudes of simulations fall within or close to the ranges of peak-to-peak amplitudes of the Degelen events of m_b between 5.35 and 5.45. According to the Mueller-Murphy model for granite (Mueller and Murphy, 1971), $m_b = 5.4$ corresponds to the moment of 10^{16} Nm used in all simulations.

4.4 Data v.s. synthetics

4.4.1 Amplitudes

All events have been simulated with an isotropic moment tensor source of a scalar moment of 10^{16} Nm. According to the Mueller-Murphy model (Mueller and Murphy, 1971), it corresponds to explosions of $m_b = 5.4$ in granite. Thus, to compare data to synthetics we are going to consider the peak-to-peak amplitudes of the Degelen events of m_b in the 5.35–5.45 range to account for the uncertainty in m_b.

As discussed in Section 4.2.2, the Q values used in the global background model result in high t^* parameters for each station. We have estimated an amplitude correction factor per station for a lower, more realistic t^* of 0.15 seconds. Correcting for the t^*, the range of peak-to-peak amplitudes of the simulations falls within or close to the ranges of peak-to-peak amplitudes of the Degelen events of m_b between 5.35 and 5.45. The exact results per station are summarised in Table 4.2, with adjusted synthetic peak-to-peak amplitudes reflecting the observed amplitude ranges. The synthetic amplitude ranges reflect both changes in amplitude due to location in the Degelen massif as well as due to depth, as discussed in Section 4.3.2.

4.4.2 Waveform features

In what follows, we are going to discuss general similarities and differences between observed and synthetic seismograms. With a multitude of factors not accounted for in the simulations – such as complex 3-D geological features under the mountain ridges, variations in the moment rate functions for events of different amplitudes, or non-linear effects such as spall, tectonic strain release and change in rock properties – we do not attempt study the waveform fit. In-

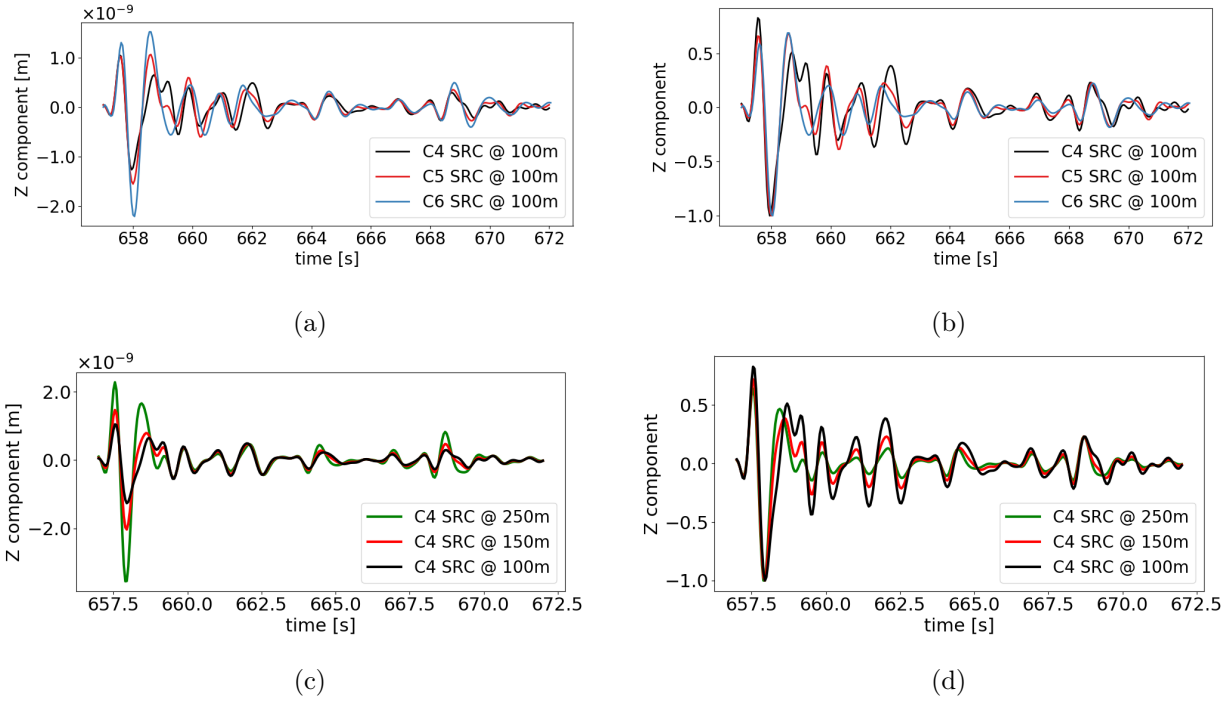


Figure 4.19: Illustration of the effect of normalisation on the comparison of waveforms at YKA. (a) and (b) look at perceived differences due to location for raw and normalised seismograms, respectively, while (c) and (d) look at perceived differences due to depth for raw and normalised seismograms, respectively. In both cases, normalising by the maximum amplitude masks variations in the first arrival and exacerbates differences in the following coda. The increase in perceived differences is particularly strong in the coda immediately following the first arrival, in the 659.5-664.5 s time window.

stead, we concentrate on general seismogram properties and particular features that emerge from interactions of the wavefield with the Degelen topography.

Recall from Chapter 3 that the analysis of the data required normalisation to correct for varying event magnitudes. However, in Figure 3.22 we point out that normalisation also changes the perception of seismogram similarities and differences, and it might exacerbate or alleviate some topography-related features in the Degelen dataset. As further illustrated in Figure 4.19, the effect becomes more apparent on synthetic seismograms. Both the location around the mountain range and the depth of burial affect the amplitudes of the first arrivals for explosions of the same magnitude. In such cases, normalisation by the maximum amplitude changes the perception of the coda waves when comparing waveforms. Thus, in order to compare the features of synthetic seismograms to those of the Degelen observations discussed in Chapter 3, we need to look at normalised waveforms. Moreover, the P wave onset in the synthetics is shifted to $t = 0$ s to match the onset in the data. Locations of all synthetic events, as well as locations of

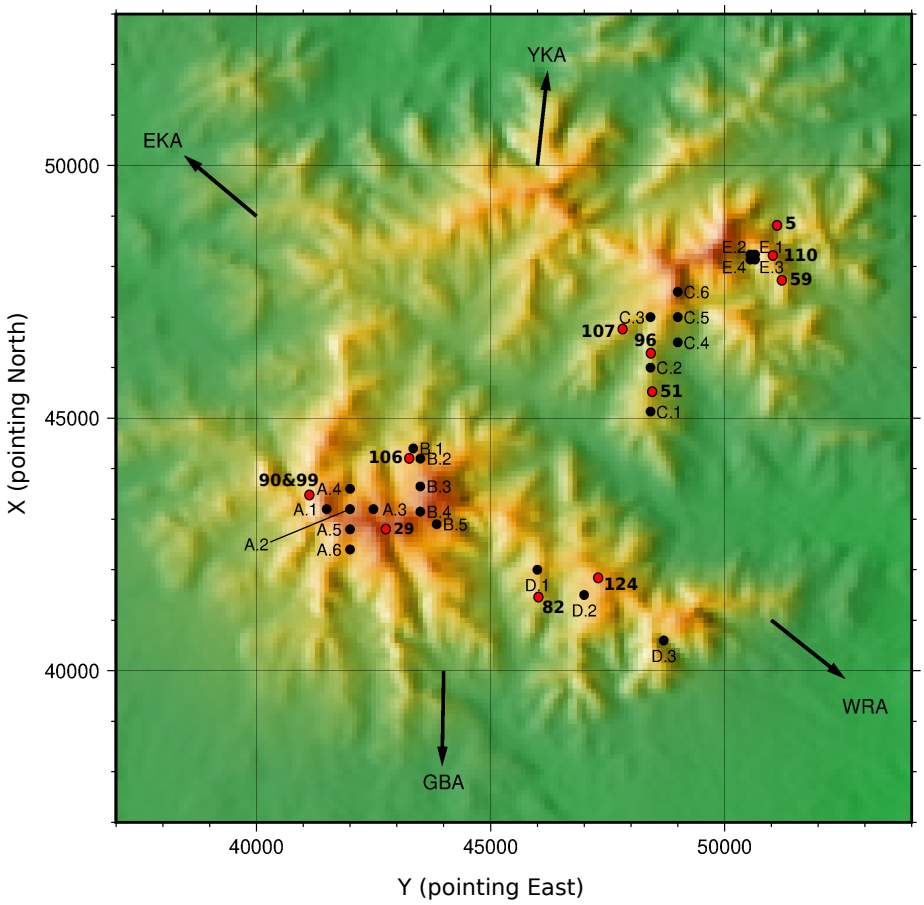


Figure 4.20: Map of Degelen topography with synthetic event locations (black circles) and locations of real Degelen events used for waveform comparisons in Section 4.4.2. Event IDs of observed Degelen events correspond to those listed in Appendix C.

real events presented in this chapter, are shown in Figure 4.20.

Location-dependent waveform similarities

In Section 3.4.3, we discuss the location-dependent waveform differences in the Degelen dataset. We uncover a progressive change in waveform features along the Degelen ridges, with high signal similarities for events in close lateral proximity. The generated synthetics exhibit a similar behaviour, and an example comparing waveforms at the YKA array for close-lying events is shown in Figure 4.21. When events are within 200 m, the first arrival and the coda (after 1.5-2 s after the onset) are usually nearly identical for event pairs both in synthetics and in real data (Figure 4.21a, synthetics B1&B2, observations 90&99). When events are located 500 m apart, differences in coda start to appear in both datasets (Figure 4.21b, synthetics

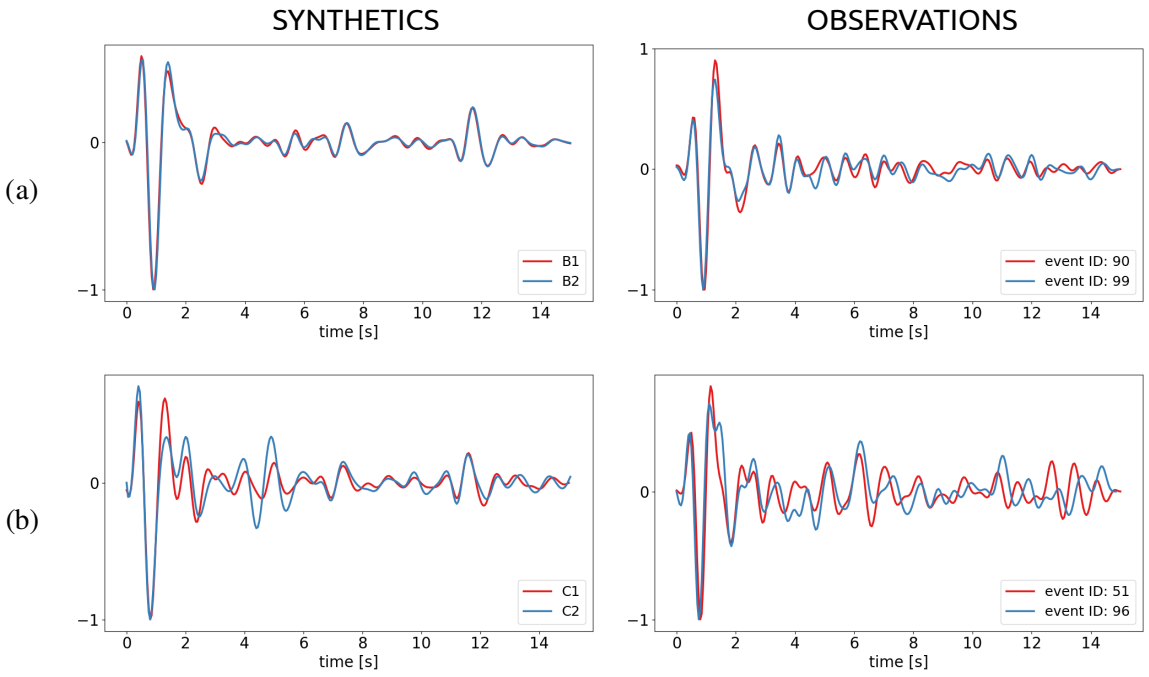
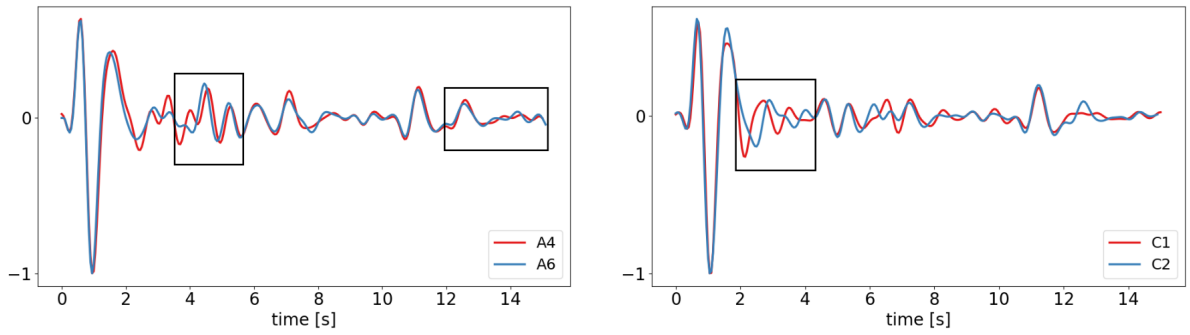


Figure 4.21: Illustration of dependence of waveform similarity on source proximity in observed and synthetic datasets at YKA. Event IDs of observed Degelen events correspond to those listed in Appendix C. Both synthetic and real locations are shown in Figure 4.20.
(a) When events are within 200 m, the first arrival and the coda (after 1.5-2 s after the onset) are nearly identical for event pairs both in synthetics (B1&B2) and in real data (90&99).
(b) When events are located 500 m apart, differences in the coda start to appear in both datasets (synthetics C1&C2, observations 51&96), although an overall high level of coherence is preserved.



(a) (b)

Figure 4.22: Small relative shift of coda with respect to location is present in the synthetics at GBA and is indicated here with a black frame on two pairs of close-laying events. Simulations, however, do not reproduce shifts as significant as in observed data. It should be noted that the coda can be very sensitive to location, so a stronger shift effect could potentially be generated with further simulations. Synthetic locations are shown in Figure 4.22.

C1&C2, observations 51&96), although an overall high level of coherence is preserved. Appendix I includes corresponding plots for the other arrays.

Apart from illustrating waveform similarities for close-laying events, Figure 4.21 also highlights differences in the general coda complexity. The synthetic waveforms are simpler, with less variation after the first arrival, and with a visibly lower corner frequency. Recall from Section 4.2.3 that the numerical requirements of the global mesh imposed a constraint on the half-width of the moment rate function, and the spectral amplitude at 2 Hz was at the half-maximum. As a result, some of the energy at 2 Hz and above is not reproduced in the synthetics, while it is present in the observed dataset. Moreover, the 20 km and Moho reflections at 7 and 11 seconds in the synthetics, respectively, are generally less prominent and less discernible in the data. This is to be expected, as discontinuities in the real Earth are not at a constant depth, and the source-receiver symmetry with respect to the discontinuities is broken.

In Section 3.4.3 we also discuss shifts in the coda waves – that are most prominent at the GBA array – due to changes in event location on the order of a few hundred metres. Simulations did not reproduce shifts as significant as in the data, though a small effect is present in the GBA synthetics (Figure 4.22). It should be noted that the coda can be very sensitive to location, so a stronger shift effect could be uncovered with further simulations.

Waveform variations in first arrivals

As mentioned above, the data are more complex and shows stronger variations in the coda. Synthetic seismograms have a visibly lower corner frequency than the Degelen observations,

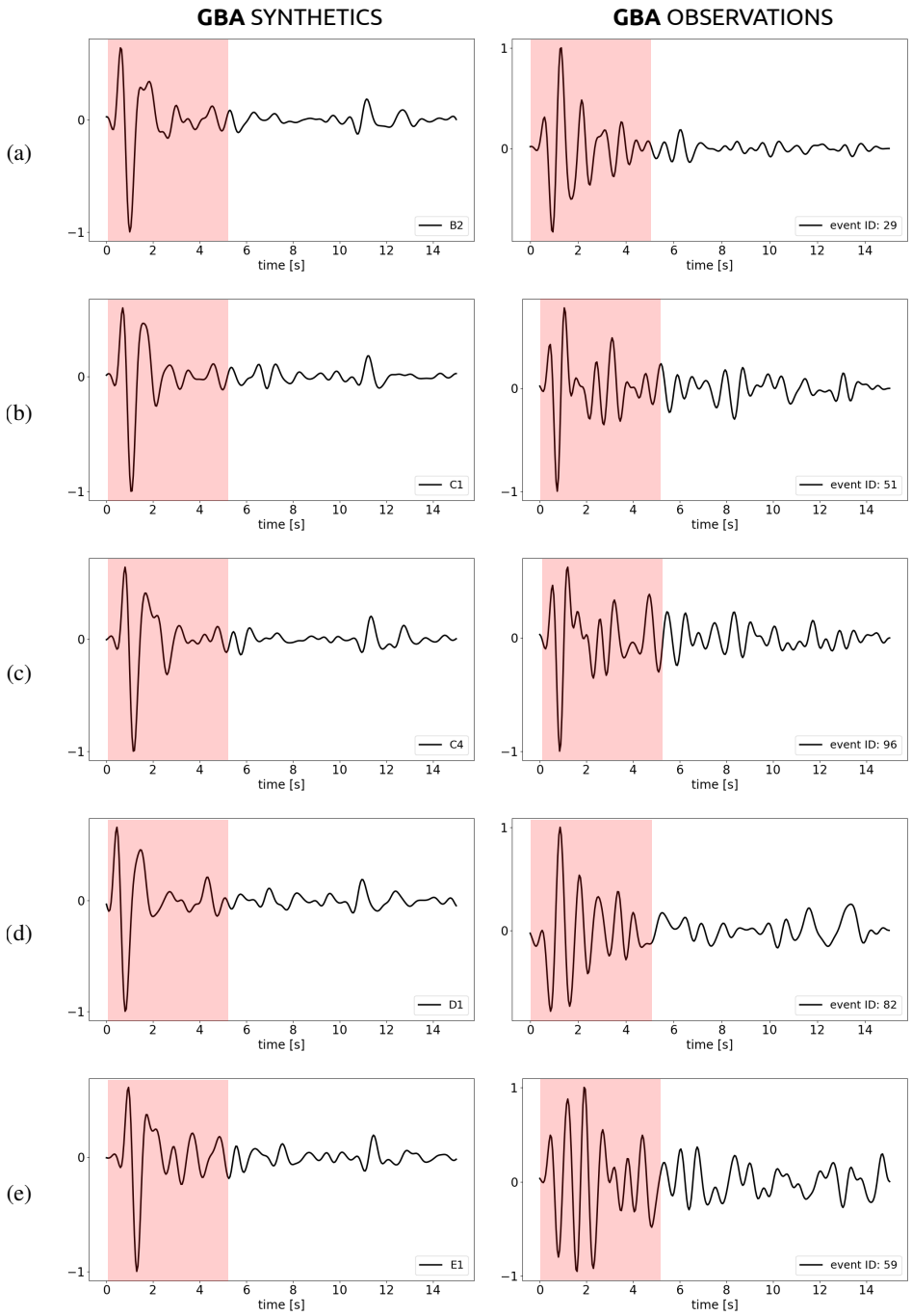


Figure 4.23: Comparison of synthetic and observed seismograms at GBA (36.4 degrees epicentral distance) for explosions located in close lateral proximity within the Degelen massif. The longer high-energy arrivals are not reproduced by the synthetics in (a), (d) and (e). However, there is more similarity in (b) and (c) with observations where energy is lower in the 2-5 s window. Event IDs of observed Degelen events correspond to those listed in Appendix C. Both synthetic and real locations are shown in Figure 4.20.

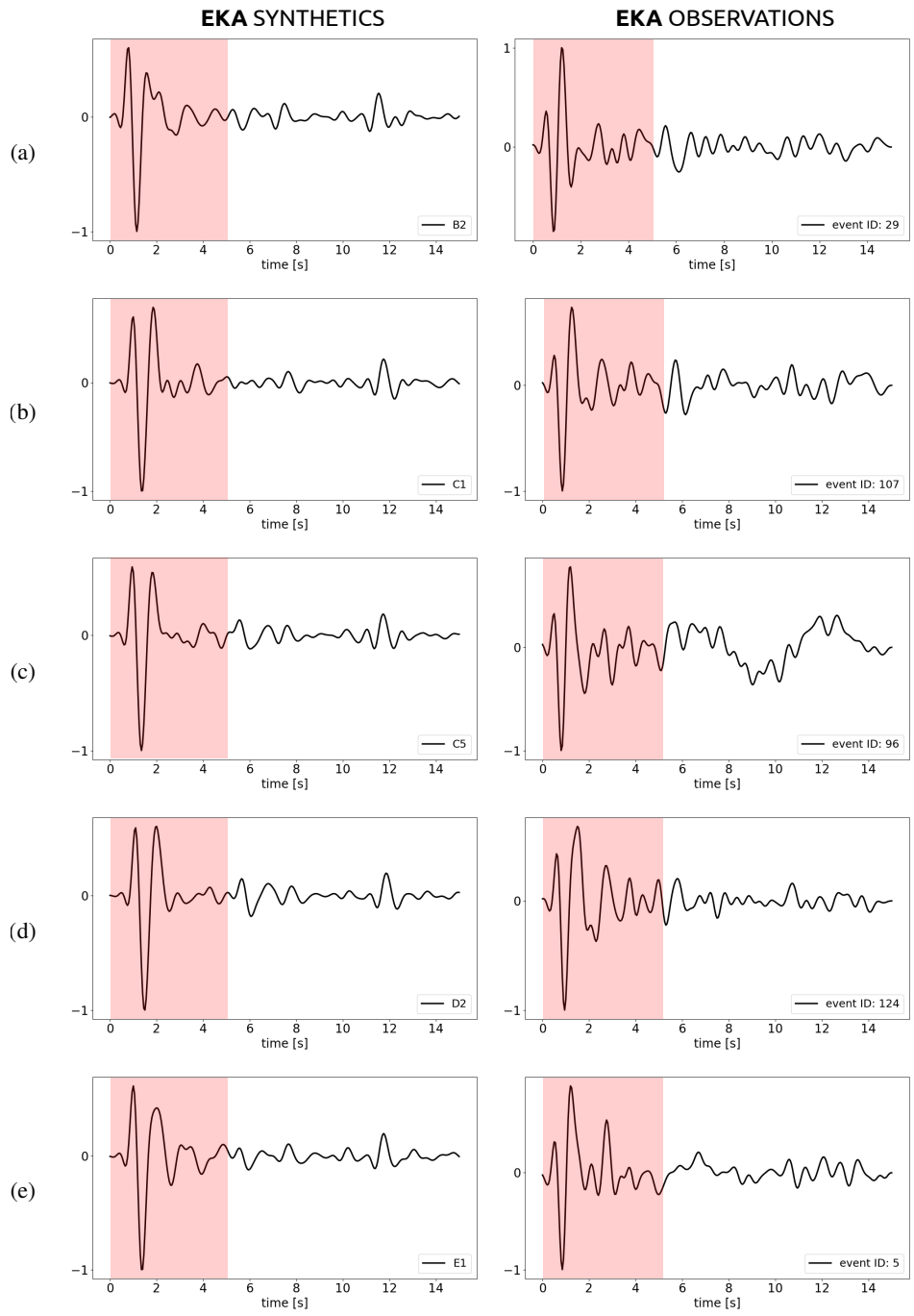


Figure 4.24: Comparison of synthetic and observed seismograms at EKA (46.1 degrees epicentral distance) for explosions located in close lateral proximity within the Degelen massif. The simplicity of the pulse shapes in the highlighted time window is well represented by the synthetics despite the lower corner frequency. Event IDs of observed Degelen events correspond to those listed in Appendix C. Both synthetic and real locations are shown in Figure 4.20.

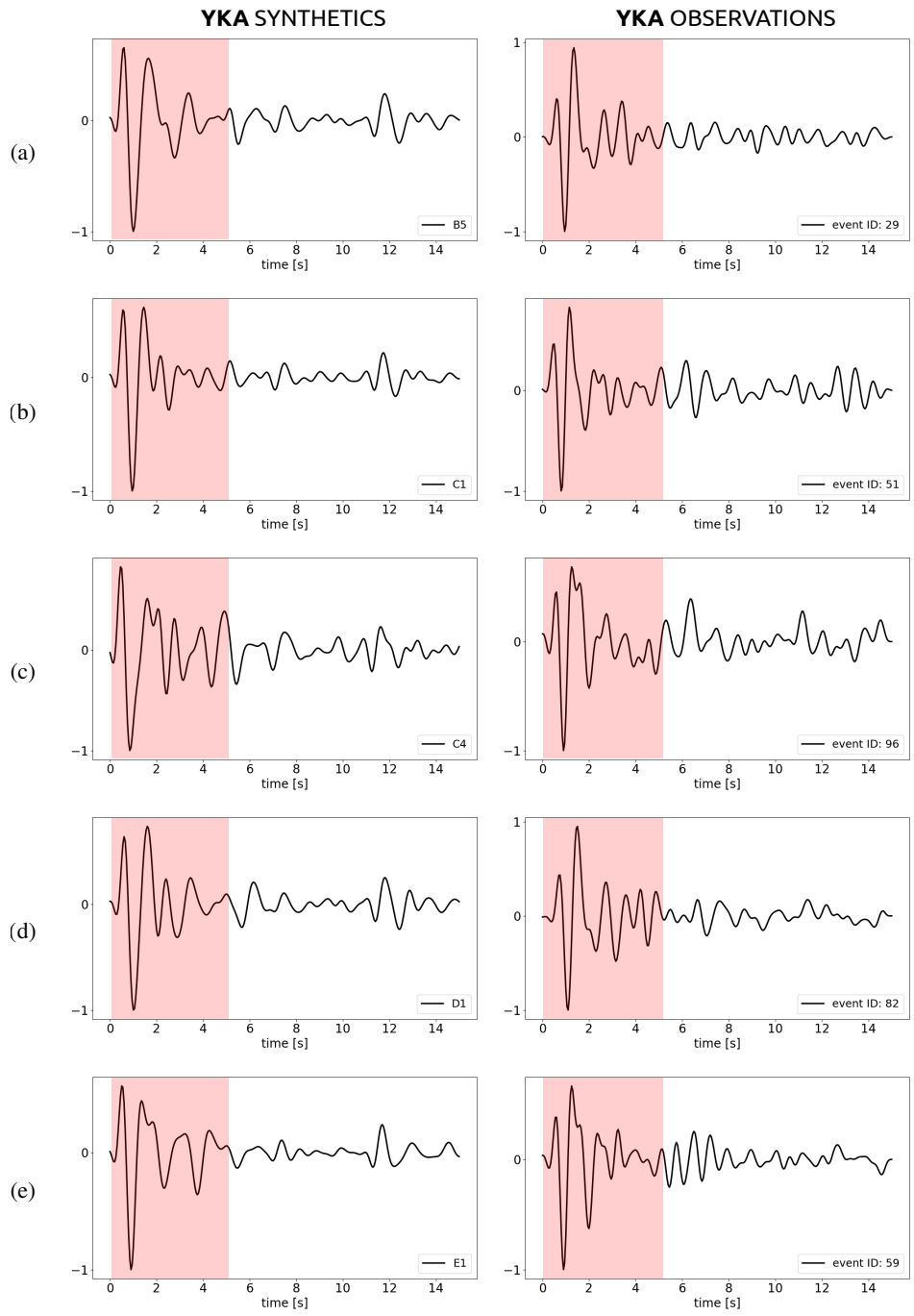


Figure 4.25: Comparison of synthetic and observed seismograms at YKA (67.0 degrees epicentral distance) for explosions located in close lateral proximity within the Degelen massif. The match of the pulse shapes in the highlighted time window is remarkable despite the lower corner frequency of the synthetics. Event IDs of observed Degelen events correspond to those listed in Appendix C. Both synthetic and real locations are shown in Figure 4.20.

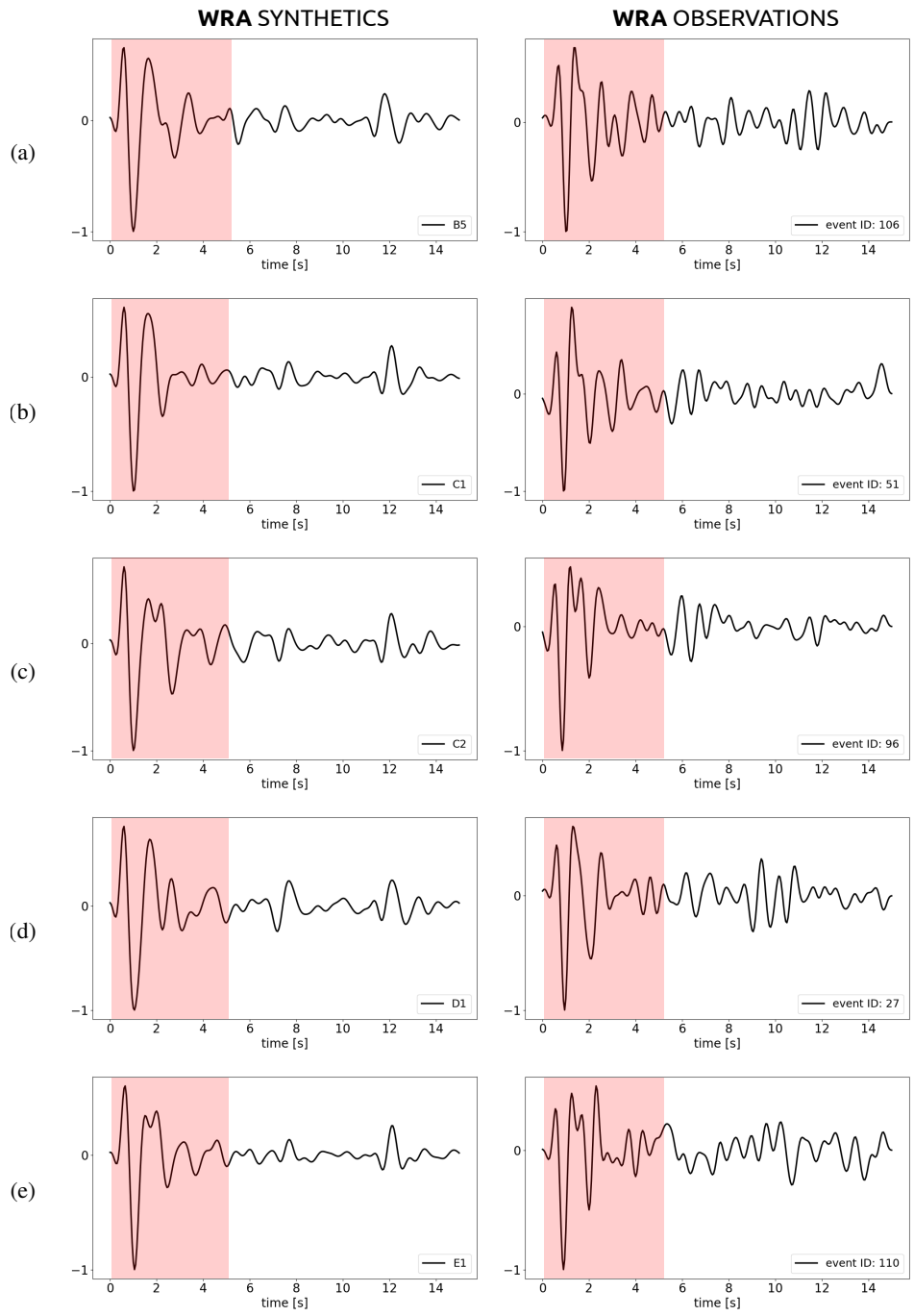


Figure 4.26: Comparison of synthetic and observed seismograms at WRA (85.3 degrees epicentral distance) for explosions located in close lateral proximity within the Degelen massif. The match of the pulse shapes in the highlighted time window is remarkable despite the lower corner frequency of the synthetics. Event IDs of observed Degelen events correspond to those listed in Appendix C. Both synthetic and real locations are shown in Figure 4.20.

rendering irrelevant any quantitative comparisons beyond the amplitudes discussed in Section 4.4.1. One striking feature of the synthetics, however, is the extent of the topographic effects present in the first 5 seconds of the seismograms. Such prominent variations are also observed in the recorded Degelen signals and appeared location-dependent, yet our efforts in Section 3.4.4 to relate these specifically to topography were inconclusive.

Figures 4.23-4.26 present synthetic and observed seismograms for explosions located in close lateral proximity within the Degelen massif, and highlight the similarities in the first arrivals on all four arrays. A visual qualitative comparison is sufficient to conclude that many complex features – in particular the splitting of the second peak discussed in Section 3.4.3, and the second negative peak – are reproduced by the synthetics and some of the variation in the data can be attributed to the mountainous topography. The match of the pulse shapes is remarkable for YKA (Figure 4.25) and WRA (Figure 4.26) arrays that are at 67.0 and 85.3 degrees epicentral distance, respectively. The simplicity of the first pulses at EKA at 46.1 degrees is also well represented (Figure 4.24), while the longer high-energy arrivals at GBA at 36.4 degrees are not reproduced by the synthetics (Figure 4.23). For the GBA observations where the energy is lower in the 2-5 seconds window, however, the synthetics show more similarity (Figures 4.23b and 4.23c). Overall, despite a simplified model that accounts only for anelastic wave propagation with an isotropic moment tensor source and a Gaussian moment rate function, and given the uncertainties in the actual event locations, the qualitative match of the pulse shapes is exceptional. However, not all signal variations are captured, and the dependence on take-off angle and azimuth is less prominent in the synthetics than in the data. Note also that not all synthetic-observed event pairs in Figures 4.23-4.26 are the same, as data is not available for each event on every array. The corresponding real and synthetic event locations are shown on a map of Degelen in Figure 4.20.

4.5 Summary and discussion of results

Waveforms from Degelen explosions bear a strong signature of near-source effects, and a visual analysis of the dataset in Chapter 3 suggests a dependence of the features on the location in the mountain range. We have attempted to quantitatively relate complexity of recorded signals to Degelen topography, but simple measures did not indicate a straightforward relationship and proved inconclusive. In this chapter, therefore, we turned to hybrid numerical modelling to explore the extent to which rough topographic features can influence high-frequency teleseismic P wave signals from the Degelen test site.

We have generated synthetic seismograms for 24 explosion locations across both mountain ridges. We have chosen to approximate the 'average' small- to medium-yield Degelen event of 15 kt by an isotropic source with a scalar moment of 10^{16} Nm and a Gaussian moment rate

function with a 'dominant period' of 2 Hz. The global elastic 1-D model (IASP91) included attenuation, and NASA's SRTM topography was implemented in the local domain around the test site.

Despite the simplistic assumptions and uncertainties in true source locations, many features from observed recordings are present in the synthetics, suggesting a significant effect of the Degelen topography on teleseismic P waves. Many parameters have not been accounted for in the modelling – such as realistic source models, geological heterogeneities, and non-linearity of wave propagation – and yet we found a very good overall qualitative fit of shape pulses, and amplitudes at each station were well reproduced. To the best of the author's knowledge, it is the first time that teleseismic 2 Hz waveforms have been compared to synthetics that account for 3-D structures, and the excellent fit achieved highlights the importance of the imprint of local 3-D effects on long-range wave propagation.

A brief examination of local topographic interactions at Degelen in Section 4.3.3 has shown that the strongest effects are present on the horizontal components of the wavefield. This suggests that waveforms at shorter distances (i.e. with shallower take-off angles) carry a stronger imprint of topography, while less of the horizontally modified energy will radiate towards receivers further away – a conclusion supported by the data, where highest complexity is observed at GBA, the array closest to Degelen. The lower-frequency synthetics, however, do not reproduce the highest-complexity GBA waveforms well, and further studies of the relationship of complexity with take-off angle are necessary. This importance of contributions of horizontally propagating waves also explains why studies attempting to repropagate only the downgoing waves to receivers at distance did not successfully replicate teleseismic waveforms (e.g. Avants, 2014).

By proving the existence of strong topographic scattering at 2 Hz, this chapter lays a foundation for further studies that are quantitative in nature and that would allow more accurate waveform comparisons. Such work requires simulations that:

1. match the corner frequency of the real data (1.5-2 Hz), requiring global Green's functions with a 'dominant frequency' up to 3-4 Hz;

2. include more accurate source models; and

3. reproduce the actual instances of all locations and constitute a full dataset directly comparable to the observations.

More detailed studies would allow both probabilistic and deterministic approaches to analysing the spectral content, complexity, and other features of both datasets. Further simulations could examine the source-topography interactions of a variety of proposed explosion source models, determine the relationship between complexity and take-off angle and azimuth, and perhaps even relate particular aspects of topography to specific features of waveforms at given azimuths and

distances. As illustrated in Figure 4.27, improving the modelling and generating synthetics that better mimic the real dataset could ultimately lead to waveform-based constraints on explosion locations, for example via grid-search methods or more advanced learning algorithms.

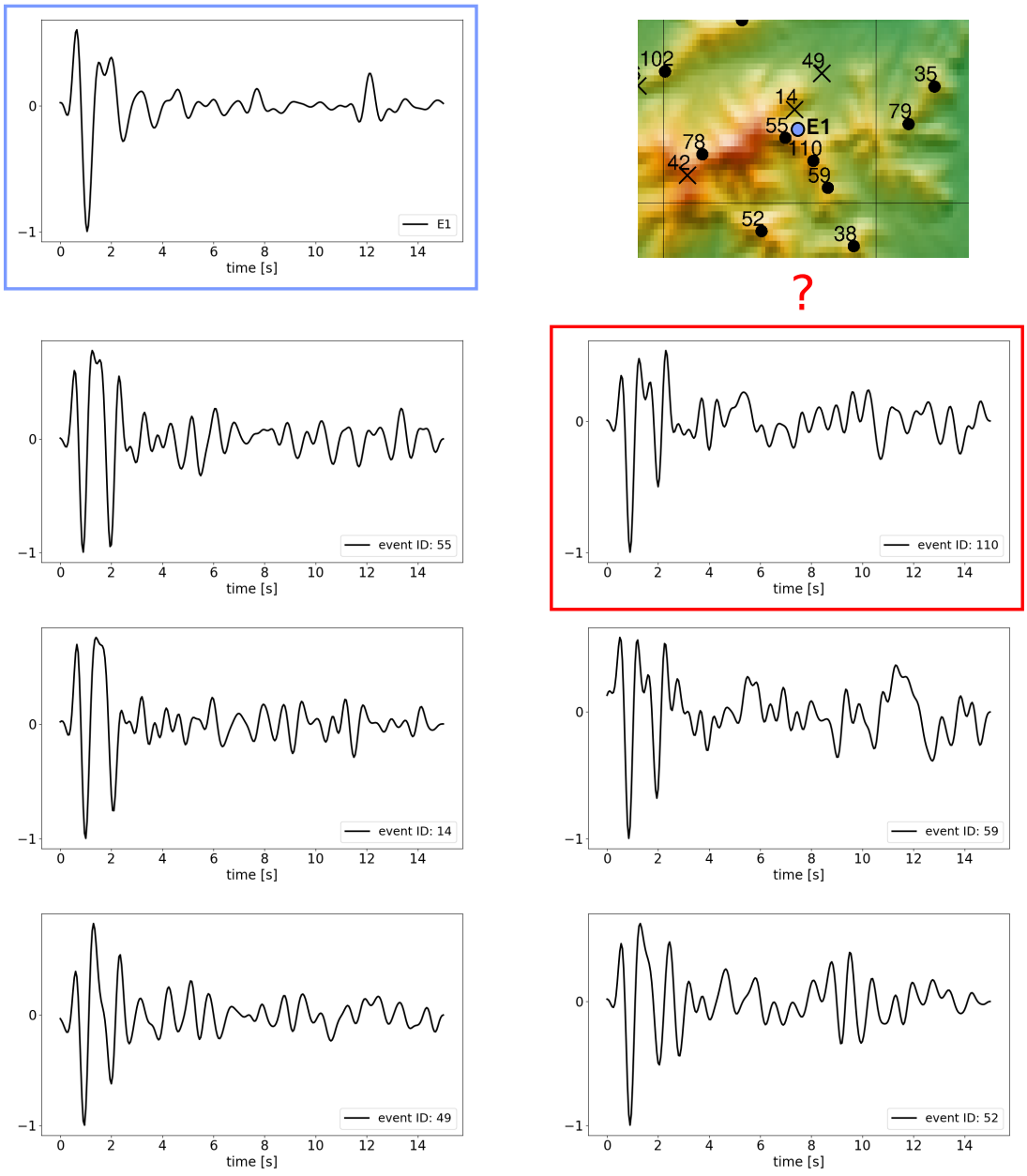


Figure 4.27: With higher frequency modelling and more realistic source parameters, fitting waveform features could be used to constrain locations of numerical explosions. The idea is illustrated on an example at WRA array, where a synthetic waveform at location E1 is compared to six observations. Despite a difference in the frequency content and a simplified moment rate function, waveform at E1 bears most similarity to seismograms from closest real explosions.

Chapter 5

Discussions

5.1 Hybrid framework

5.1.1 Overview

Numerical wave propagation is one of the primary means of studying wave physics and complex scattering effects, helping us to interpret teleseismic recordings and to infer structural and dynamic properties of the Earth. A number of global scale three-dimensional methods – able to tackle various geometries and complexities, and accounting for the full physics – have been developed, yet high computational requirements of such solvers decidedly limit their applicability and render high-frequency simulations prohibitive. Improving our understanding of the Earth's structures and processes, and how they interact with seismic events, must therefore incorporate numerical methods that not only capture the relevant physics and account for the complexity that affects global wave propagation in 3-D, but also provide a high-frequency resolution in a computationally feasible manner.

In this thesis we propose an Instaseis-based hybrid method that couples two solvers and provides a compromise between structural complexity and computational cost. The framework embeds a local 3-D solver of choice in spherically symmetric global wavefields, and reaches the highest body-wave frequencies observed teleseismically (1-4 Hz) with a factor 5,000-100,000 speed-up relative to full 3-D global solvers. Accounting for arbitrary 3-D structures and strong localised scattering, the hybrid method opens global wave propagation to new parameter regimes and thus to a range of outstanding problems in seismology.

Based on pre-computed Instaseis databases of Green's functions, the framework brings a new approach to hybrid methods and offers an unprecedented level of flexibility. The databases can:

- be used with an unlimited number of local 3-D solvers, provided that necessary adaptations are implemented in the local code;
- accommodate for a range of parameters, acting as a once-and-for-all solution to global

wave propagation in spherically symmetric Earth models, as source properties (radiation patterns, source-time function), the source-receiver geometry, and the local domain dimensions can all be modified;

– cater for 3-D domains located both on the Earth's surface – around the source region and under the receivers – as well as at depth along the propagation path.

A single database can therefore be re-used for a wide range of applications, reducing the need for global simulations and for extensive computational requirements. Stored Green's functions can be accessed when necessary and queried for a chosen set of parameters. The computational costs are therefore reduced to the cost of wavefield injection and extrapolation – embarrassingly parallel processes that read from and write to HDF5 files – as well as to the cost of running the local method.

5.1.2 Method developments

3-D background medium

With the current framework implementing 1-D to 3-D coupling, a 3-D to 3-D method would be a natural extension of the hybrid database approach. AxiSEM3D (Leng et al., 2016) is currently developed towards integration with Instaseis in order to provide a new generation of three-dimensional tomographic databases. Compatible with hybrid injection and extrapolation, databases of 3-D Green's functions will retain the flexible properties of the once-and-for-all approach and render the path effects of the background medium more realistic.

Performance enhancement

The performance of the method can be further enhanced by improved parallelisation of database access and of saving and reading of the interfacing HDF5 files. The hybrid methodology requires information on thousands of points on the boundary of the local mesh, and time series at each point have thousands of samples. Theoretically accessing a database is an embarrassingly parallel process, though file access bottlenecks and a suboptimal load distribution reduce the speed and can be improved for better performance. The same applies to interfacing HDF5 files that collate processed information from the database and make it available to local solvers. Moreover, an efficient implementation of reading and writing HDF5 files in the local solver is necessary for an overall optimal performance of the scheme.

Usability

Based on the Instaseis python package that attaches great importance to a user-friendly interface, the hybrid HDF5 files can be easily created with simple objects and function calls (see

Appendix A for details). The overall usability, however, can be improved with further developments, such as:

- An automated selection of the minimum time for saving wavefields for injection (based on epicentral distance from the source). This reduces the computation time, as zeros preceding the seismic signal are discarded.
- An automated definition of the rotation matrix based on the orientation of the local coordinate system. Currently the user needs to define the rotation matrix manually for the local solver of choice.
- An aided selection of extrapolation points and weights for methods that are not based on the weak form of the wave equation (and which therefore by construction have an optimally chosen set of points and weights to discretise an integral).

Such extensions will help the end user to optimally define the coupling set-up and speed up its design.

5.2 Nuclear explosions case study

5.2.1 Overview

Understanding and quantifying teleseismic contributions of near-source effects is of paramount importance in forensic seismology. Strong scattering in the vicinity of the source, along with complex non-linear interactions, can obscure the signature of the source and interfere with detection and identification of nuclear tests. In an effort to gain more insight into the imprint of the source region on teleseismic signals, we revisit and analyse the digitised dataset of the 1960s-1980s underground explosions from the Degelen test site. The recorded P waves exhibit large variations on each array interpreted as a signature of near-source effects. A visual examination demonstrates a dependence of signal characteristics on the location of the explosion in the mountain range, while depths and yields do not show a consistent effect. The seismograms change progressively along the ridges, and we postulate that this change in the primary P wave and the subsequent coda is related to the mountainous topography. The hypothesis, however, cannot be validated from data alone, as capturing seismogram complexity and topographic features in meaningful measures is not straightforward. The chosen measures prove largely inconclusive – although they suggest that P wave complexity is related to source location – and we turn to deterministic hybrid modelling to test the imprint of Degelen topography at teleseismic distances up to 2 Hz.

By choosing simple model parameters we isolate the effects of topographic scattering alone. We assume an isotropic moment tensor source with a Gaussian moment rate function that ap-

proximates a medium-sized explosion in granite and hence we do not account for the effects of yields (and the corresponding moment rate functions). We also consider only a 1-D anelastic model without any geological heterogeneities in the source region, and we ignore any non-linear contributions to the wavefield. By construction, complex path and receiver effects are not accounted for in the 1-D background model. Despite such simplistic assumptions and uncertainties in source locations we obtain a very good qualitative waveform fit between the observables and synthetics, with amplitudes at all four arrays and numerous features of the P waves reproduced through hybrid modelling. To the best of the author's knowledge, it is the first time that teleseismic 2 Hz waveforms have been compared to synthetics that account for 3-D structures.

5.2.2 Further work

The results suggest that topography is indeed the parameter defining the progressive change of P wave and its coda along the Degelen ridges, as close lying synthetic and real events show distinct similarities. Waveform complexity has not been fully reproduced, however, and further work is needed to refine and improve the synthetics and reliably quantify topographic scattering. The simulated seismograms, for example, have a lower corner frequency than the data. Events in the 1-100 kt range – a range representative of Degelen – have similar spectra and corner frequencies of about 1.3-2 Hz, while the synthetics have been generated to contain most of the energy up to 2 Hz with a minimal computational effort ('dominant frequency' of 2 Hz, i.e. corner frequency of 1 Hz) and do not capture all energy up to 2 Hz. Exact source locations were not considered, and the same approximate source model was used for all simulations regardless of the yield and depth.

Further work at higher resolution is therefore necessary to extend the case study and to carry out a systematic quantitative investigation of the data. Such a comprehensive analysis should consider realistic source models that reflect the effects of yields, and could be used to test for differences in combined teleseismic effects of topography and competing source models. Generating comparable synthetics should also mirror a complete set of real source locations, so that meaningful conclusions can be derived from equivalent probabilistic and deterministic analyses of both datasets, including the relationship between topography and take-off angles and azimuths. A study of a complete synthetic dataset can also encompass three component seismograms and examine S waves, as it has been suggested that topography contributes to their generation. Subsequent steps can also introduce geological complexities of the Degelen mountain range into the modelling, or eventually move beyond anelastic wave propagation to consider local non-linear interactions. Such realistic scenarios could ultimately lead to waveform-based constraints on explosion locations, for example via grid-search methods or more advanced learning algorithms, or even improve nuclear discrimination methods.

Appendices

Appendix A

The modified Instaseis package

The hybrid interface is based on the Instaseis package (van Driel et al., 2015) that was modified to accommodate both wavefield injection and extrapolation. This appendix outlines how to use the hybrid Instaseis package via interfacing HDF5 files. For more information on the usage of Instaseis, refer to the documentation available at `http://instaseis.net/`.

A.1 Instaseis: wavefield injection

A.1.1 Injection requirements

For wavefield injection, Instaseis requires:

- a HDF5 file with coordinates of the mesh points on the boundary of the local solver
- a forward Instaseis database
- an Instaseis-defined source
- a name for the HDF5 file to output background fields
- a specification of the fields to save, as well as of the coordinate system in which the fields should be saved
- a time window for the fields to save (at large distances, one can choose to start the save right before the first arrival in order to reduce the length of the traces; the length of the saved fields defines the subsequent length of the local simulation)

A.1.2 Structure of interfacing HDF5 files

coordinates HDF5 input file

A coordinates HDF5 input file provides Instaseis with all information about the coordinate points of the mesh of the local solver. The group name defines whether the coordinates are given in a local or global coordinate system: i.e. a "local" or a "global" group name. The group needs the following attributes:

- `nb_points` – a total number of points – a float
- `radius_of_box_top` – only necessary for a local coordinate system as it is needed for rotation; in case of a domain at the surface, the radius of the top of the domain is equal to the radius of the Earth – a float
- `rotmat_xyz_loc_to_glob` – only necessary for a local coordinate system; it is the rotation matrix from local Cartesian to global Cartesian coordinates – a 3×3 array

And the following datasets in the relevant coordinate system:

- `coordinates` – an array of size `(nb_of_points_on_boundary, 3)`
- `normals` – an array of size `(nb_of_points_on_boundary, 3)`; only necessary if saving tractions

where the order of the last dimension is (θ, ϕ, r) for global or (x, y, z) for local coordinates.

An example of the file structure for a local coordinate system:

```
GROUP "/" {
  GROUP "local" {
    ATTRIBUTE "nb_points" {
      DATATYPE H5T_STD_I32LE
      DATASPACE SIMPLE { ( 1 ) / ( 1 ) }
    }
    ATTRIBUTE "radius_of_box_top" {
      DATATYPE H5T_IEEE_F64LE
      DATASPACE SIMPLE { ( 1 ) / ( 1 ) }
    }
    ATTRIBUTE "rotmat_xyz_loc_to_glob" {
      DATATYPE H5T_IEEE_F32LE
      DATASPACE SIMPLE { ( 3, 3 ) / ( 3, 3 ) }
    }
```

```
    DATASET "coordinates" {
      DATATYPE H5T_IEEE_F32LE
      DATASPACE SIMPLE { ( 19200, 3 ) / ( 19200, 3 ) }
    }
    DATASET "normals" {
      DATATYPE H5T_IEEE_F32LE
      DATASPACE SIMPLE { ( 19200, 3 ) / ( 19200, 3 ) }
    }
  }
}
```

fields HDF5 output file

The hybrid implementation of Instaseis injection saves the background wavefields into a HDF5 output file. The group name defines whether required fields are output in a local or global coordinate system: i.e. a "local" or a "global" group name. The group needs the following attributes:

- `nb_points` – a total number of points – a float
- `nb_timesteps` – a total number of time steps – a float
- `dt` – the time step between the samples of the wavefields – a float

And at least one of the following datasets in the relevant coordinate system:

- The `displacement`, `velocity` and `traction` are arrays of size (`nb_of_points_on_boundary`, `nb_of_time_steps`, 3), where the order of the last dimension is (θ, ϕ, r) for global or (x, y, z) for local coordinates.
- The `strain` and `stress` are arrays of size (`nb_of_points_on_boundary`, `nb_of_time_steps`, 6), where the order of the last dimension is $(\theta\theta, \phi\phi, rr, r\phi, r\theta, \theta\phi)$ for global or (xx, yy, zz, zy, zx, xy) for local coordinates.

An example of the file structure for a local coordinate system with velocity and stress fields:

```
GROUP "/" {
   GROUP "local" {
      ATTRIBUTE "dt" {
         DATATYPE  H5T_IEEE_F64LE
         DATASPACE  SCALAR
      }
```

```
      ATTRIBUTE "nb_points" {
         DATATYPE  H5T_STD_I64LE
         DATASPACE  SCALAR
      }
      ATTRIBUTE "nb_timesteps" {
         DATATYPE  H5T_STD_I64LE
         DATASPACE  SCALAR
      }
      DATASET "stress" {
         DATATYPE  H5T_IEEE_F32LE
         DATASPACE  SIMPLE { ( 5100, 3000, 6 ) / ( 5100, 3000, 6 ) }
      }
      DATASET "velocity" {
         DATATYPE  H5T_IEEE_F32LE
         DATASPACE  SIMPLE { ( 5100, 3000, 3 ) / ( 5100, 3000, 3 ) }
      }
   }
}
```

A.1.3 Sample code

The example below uses the embarrassingly parallel mpi4py implementation and requires parallel HDF5. A serial implementation is also available. The method `instaseis.hybrid_extraction` generates a HDF5 output file with background fields that need to be imposed on the boundary of the local solver of choice.

```
# import the instaseis package
import instaseis
# import mpi4py to run in parallel
from mpi4py import MPI

# path to the interfacing HDF5 file with the coordinates of the mesh
    points on the boundary of the local solver
path_to_coordinates_hdf5 = "/path/to/coordinatesHDF5/file"

# path and file name of the HDF5 file that will be dumped by
    instaseis
outputfile = "/path/to/outputfileHDF5"
```

```
# path to the forward Instaseis database
fwd_db_path = "/path/to/forward/instaseis/database"

# the time step of the Instaseis dump (as required by the local
    solver)
dt = 0.1

# the seismic source defined via Instaseis
# note: the depth of the source needs to correspond to the depth
    defined in the forward database
source = instaseis.Source(
      latitude=72.9, longitude=1.5, depth_in_m=80000,
      m_rr=4.71E17, m_tt=3.81E15, m_pp=-4.74E17,
      m_rt=3.99E16, m_rp=-8.05E16, m_tp=-1.23E17)

# define dump fields
dumpfields = ("velocity", "stress")

# define dump coordinates
# the injection field is output in the coordinates of the local
    method
dumpcoords = "local"

# define extraction time window
time_window = (300, 600)

# extract wavefields from database and dump the HDF5 output file
instaseis.hybrid_extraction(path_to_coordinates_hdf5,
                            outputfile,
                            fwd_db_path, dt, source,
                            dumpfields=dumpfields,
                            dumpcoords=dumpcoords,
                            time_window=time_window)
```

A.2 Instaseis: wavefield extrapolation

A.2.1 Extrapolation requirements

For wavefield extrapolation, Instaseis requires:

- a HDF5 file with coordinates of the mesh points on the boundary of the local solver
- a HDF5 file with fields on the boundary of the local solver
- a reciprocal Instaseis database
- an Instaseis-defined receiver

A.2.2 Structure of interfacing HDF5 files

coordinates HDF5 input file

Like for the injection case in Section A.1.2, a coordinates HDF5 input file provides Instaseis with all information about the coordinate points of the mesh of the local solver. The group name defines whether the coordinates are given in a local or global coordinate system: i.e. a "local"or a "global" group name. The group needs the following attributes:

- `nb_points` – a total number of points – a float
- `radius_of_box_top` – only necessary for a local coordinate system as it is needed for rotation; in case of a domain at the surface, the radius of the top of the domain is equal to the radius of the Earth – a float
- `rotmat_xyz_loc_to_glob` – only necessary for a local coordinate system; it is the rotation matrix from local Cartesian to the global Cartesian coordinates – a 3×3 array

And the following datasets in the relevant coordinate system:

- `coordinates` – an array of size `(nb_of_points_on_boundary, 3)`
- `weights` – an array of length `(nb_of_points_on_boundary)`
- `normals` – an array of size `(nb_of_points_on_boundary, 3)`

where the order of the last dimension is (θ, ϕ, r) for global or (x, y, z) for local coordinates.

An example of the file structure for a local coordinate system:

```
GROUP "/" {
  GROUP "local" {
    ATTRIBUTE "nb_points" {
      DATATYPE H5T_STD_I32LE
      DATASPACE SIMPLE { ( 1 ) / ( 1 ) }
    }
    ATTRIBUTE "radius_of_box_top" {
      DATATYPE H5T_IEEE_F64LE
      DATASPACE SIMPLE { ( 1 ) / ( 1 ) }
    }
    ATTRIBUTE "rotmat_xyz_loc_to_glob" {
      DATATYPE H5T_IEEE_F32LE
      DATASPACE SIMPLE { ( 3, 3 ) / ( 3, 3 ) }
    }
    DATASET "coordinates" {
      DATATYPE H5T_IEEE_F32LE
      DATASPACE SIMPLE { ( 19200, 3 ) / ( 19200, 3 ) }
    }
    DATASET "weights" {
      DATATYPE H5T_IEEE_F32LE
      DATASPACE SIMPLE { ( 19200 ) / ( 19200 ) }
    }
    DATASET "normals" {
      DATATYPE H5T_IEEE_F32LE
      DATASPACE SIMPLE { ( 19200, 3 ) / ( 19200, 3 ) }
    }
  }
}
```

fields HDF5 input file

A fields HDF5 input file provides Instaseis with all information about the wavefields on the boundary points of the local solver. Again, the group name defines whether the fields are in local or global coordinates (i.e. “local” or “global” group name). The group needs the following attributes:

- `dt` – the time step of the provided fields – a float

And two datasets in the relevant coordinate system. One dataset must be either displacement or velocity, and the other either traction or strain.

- `displacement`, `velocity` and `traction` are arrays of size (`nb_of_points_on_boundary`, `nb_of_time_steps`, 3), where the order of the last dimension is (θ, ϕ, r) for global or (x, y, z) for local coordinates.
- `strain` is an array of size (`nb_of_points_on_boundary`, `nb_of_time_steps`, 6), where the order of the last dimension is $(\theta\theta, \phi\phi, rr, r\phi, r\theta, \theta\phi)$ for global or (xx, yy, zz, zy, zx, xy) for local coordinates.

An example of the file structure for a local coordinate system:

```
GROUP "/" {
  GROUP "local" {
    ATTRIBUTE "dt" {
      DATATYPE H5T_IEEE_F64LE
      DATASPACE SIMPLE { ( 1 ) / ( 1 ) }
    }
    DATASET "displacement" {
      DATATYPE H5T_IEEE_F32LE
      DATASPACE SIMPLE { ( 19200, 8500, 3 ) / ( 19200, 8500, 3 ) }
    }
    DATASET "strain" {
      DATATYPE H5T_IEEE_F32LE
      DATASPACE SIMPLE { ( 19200, 8500, 6 ) / ( 19200, 8500, 6 ) }
    }
  }
}
```

A.2.3 Sample code

The example below uses the embarrassingly parallel mpi4py implementation and requires parallel HDF5. A serial implementation is also available. The method `instaseis.hybrid_repropagation` returns the waveform as an obspy Stream object.

```
# import the instaseis package
import instaseis
# import mpi4py to run in parallel
from mpi4py import MPI
```

```
# path to the interfacing HDF5 file with the coordinates of the mesh
    points on the boundary of the local solver
path_to_coordinates_hdf5 = "/path/to/coordinatesHDF5/file"

# path to the interfacing HDF5 file with the fields from the local
    simulation
path_to_fields_hdf5 = "/path/to/fieldsHDF5/file"

# path to the reciprocal Instaseis database
bwd_db_path = "/path/to/reciprocal/instaseis/database"

# the receiver defined via Instaseis
receiver = instaseis.Receiver(latitude=42.5, longitude=1.5)

# calculate the hybrid seismogram at the specified receiver location
st_hybrid = instaseis.hybrid_repropagation(
                coords_path=path_to_coordinates_hdf5,
                fields_path=path_to_fields_hdf5,
                receiver=receiver,
                bwd_db_path=bwd_db_path,
                components=("Z", "N", "E"))
```

Appendix B

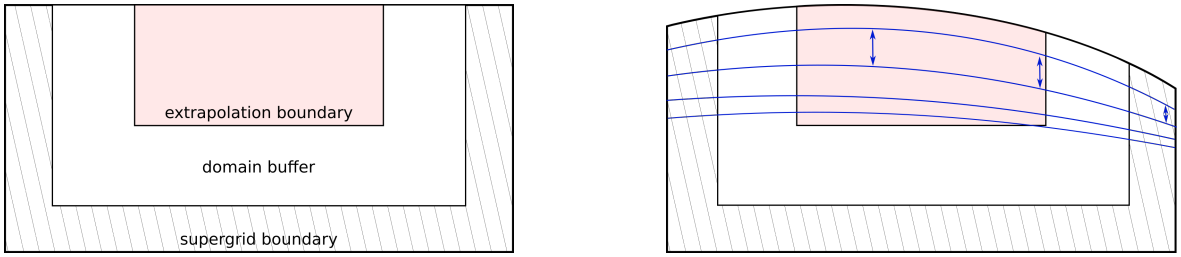
WPP-Instaseis: the spherical Earth

As discussed in Section 2.3.4, the current implementation of the WPP-Instaseis source-side coupling does not respect the curvature of the Earth's surface or of internal discontinuities. For smaller domain sizes, the discrepancy between the flat and the curved surface also remains small and – depending on the simulation period considered – might not generate significant errors, like for the WPP-Instaseis benchmarks in Section 2.4.3. However, larger mismatches between the local and the global domains relative to the simulation period can result in non-negligible effects on waveforms.

It would only seem natural to impose the curvature of the Earth's surface as topography that deforms the top layers of the flat Cartesian WPP mesh. Changing the mesh, however, poses a few problems for the current implementation. First of all, it should be clarified that the actual simulation domain in WPP is larger than the coupling surface chosen to repropagate local wavefields in Instaseis. The domain needs to incorporate an absorbing supergrid boundary layer, and is chosen to be even larger as a precaution to avoid potential spurious boundary reflections – the set-up is sketched in Figure B.1a. Given the current implementation of the WPP mesher, imposing a spherical curvature 'stretches' elements under the free surface of the domain relative to the requested node spacing. The extrapolation boundary is on the interior of the WPP domain, resulting in elements of larger sizes (Figure B.1b). Such a topographic modification implies that:

- The mid-point rule that has been implemented for regular grid points to discretise the integral necessary for extrapolation is no longer valid and needs to be adapted accordingly.
- The number of points per wavelength inside of the WPP domain and on the extrapolation surface is decreased, resulting in wavefields of lower accuracy.

Imposing the curvature of the free surface as WPP topography is therefore not straightforward and substantially increases computational costs, as it requires significantly finer meshing and



(a) (b)

Figure B.1: Sketch of the effect of imposing the Earth's curvature as topography in WPP-Instaseis coupling. (a) The WPP coupling set up. The actual simulation domain in WPP is larger than the coupling surface chosen to repropagate local wavefields to receivers in Instaseis (red area). The domain needs to incorporate an absorbing supergrid boundary layer (striped area), and is chosen to be even larger as a precaution to avoid potential spurious boundary reflections (white area). (b) An exaggerated view of stretching of elements in the vertical direction for the curvature imposed as topography. Instaseis extrapolation surface is on the interior of the WPP domain (red area), resulting in elements of larger sizes on that artificial coupling boundary.

includes a thick layer of curvilinear elements. The approach has been tested and rejected for the purpose of this work.

Appendix C

The Degelen test site data

The table on the following two pages summarises all data used for the analysis of source-side effects at the Degelen test site in Kazakhstan (Chapter 3). End-of-adit latitudes and longitudes come from the study by Trabant et al. (2002), whereas GPS locations are as reported by Leith (1998). Depths for pre-1972 events have been published by Bocharov et al. (1989), and new body-wave magnitude (m_b) estimates have recently been published by Peacock et al. (2017). GBA, EKA, YKA and WRA stand for Gauribidanur, Eskdalemuir, Yellowknife and Warramunga, respectively, and the table indicates all recordings available for each array. The data from all four arrays span a total of a 137 events for years 1964-1989. For completeness and ease of cross-referencing, location IDs used in Trabant et al. (2002) are included.

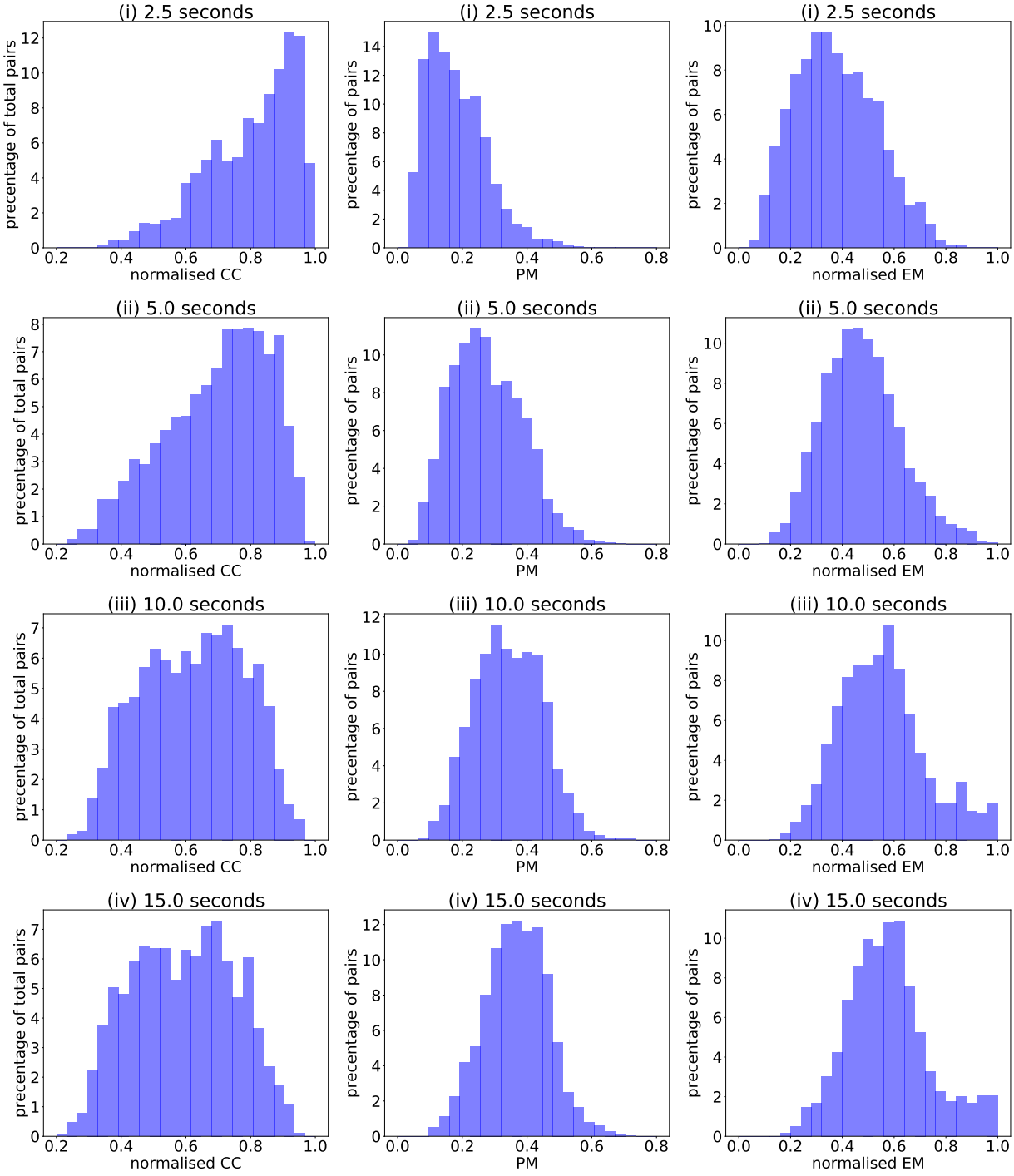
event ID	Trabant et al location ID	GBA	EKA	YKA	WRA	date [yyyy:mm:dd]	time [hh:mm:ss]	mb	GPS latitude	GPS longitude	adit end latitude	adit end longitude	depth [km]
1	n/a		x	x		1964-03-15	08:00:00	5.605	49.8208	78.0986	n/a	n/a	0.22
2	n/a		x	x		1964-05-16	06:01:00	5.594	49.8178	78.0939	n/a	n/a	0.253
3	n/a		x	x		1964-06-06	00:00:00	0	49.7747	77.9881	n/a	n/a	n/a
4	n/a			x		1964-07-19	06:00:01	5.443	49.815	78.0931	n/a	n/a	1.68
5	n/a		x	x		1964-11-16	06:00:00	5.686	49.8139	78.1336	n/a	n/a	0.194
6	134		x	x		1965-02-04	06:00:00	n/a	49.7731	77.9914	49.7669	77.9979	n/a
7	134		x	x		1965-05-11	06:40:00	4.766	49.7731	77.9914	49.7669	77.9979	0.103
8	141		x	x		1965-06-17	03:45:00	5.257	49.8306	78.0578	49.8291	78.0661	0.152
9	135			x		1965-07-29	03:05:00	n/a	49.785	77.9969	49.7784	78.0004	0.126
10	n/a		x	x		1965-09-17	04:00:00	5.223	49.8172	78.1425	n/a	n/a	0.156
11	n/a		x	x		1965-10-08	06:00:00	5.478	49.8306	78.1014	n/a	n/a	0.204
12	n/a		x	x		1965-11-21	04:58:00	5.656	49.8283	78.0597	n/a	n/a	0.278
13	n/a			x		1965-12-24	05:00:00	4.97	49.8117	78.1075	n/a	n/a	0.213
14	n/a		x	x	x	1966-02-13	04:58:00	6.276	49.8142	78.1308	n/a	n/a	0.297
15	3	x	x	x	x	1966-03-20	05:50:00	6.073	49.7517	78.0083	49.7615	78.0226	0.294
16	n/a		x	x	x	1966-04-21	03:58:00	5.402	49.8178	78.0939	n/a	n/a	0.178
17	11	x	x	x	x	1966-05-07	03:58:00	4.748	49.7361	78.1072	49.7437	78.105	0.274
18	148	x	x		x	1966-06-29	06:58:01	5.532	49.8392	78.0708	49.835	78.0726	0.187
19	10	x	x	x	x	1966-07-21	03:58:00	5.395	49.7356	78.1047	49.7376	78.0972	0.17
20	6	x	x	x	x	1966-08-05	03:58:00	5.425	49.7589	78.0486	49.7637	78.042	0.171
21	n/a	x	x	x		1966-08-19	03:53:00	4.641	49.8306	78.1014	n/a	n/a	0.134
22	141			x	x	1966-09-07	03:52:00	4.627	49.8306	78.0578	49.8291	78.0661	0.117
23	4		x	x	x	1966-10-19	03:58:00	5.712	49.7419	78.0256	49.7468	78.0208	0.185
24	5	x		x		1966-12-03	05:02:00	4.552	49.7403	78.0308	49.7473	78.0331	0.153
25	107	x		x		1967-01-30	04:02:00	4.666	49.7731	77.99	49.7705	77.9825	0.131
26	9	x	x	x		1967-02-26	03:58:00	6.08	49.7394	78.0864	49.7464	78.0822	0.241
27	8	x	x	x	x	1967-03-25	05:58:01	5.354	49.7533	78.0536	49.7533	78.0624	0.152
28	11	x	x	x	x	1967-04-20	04:08:00	5.58	49.7361	78.1072	49.7437	78.105	0.225
29	3	x	x	x	x	1967-05-28	04:08:00	5.501	49.7517	78.0083	49.7615	78.0226	0.262
30	118	x	x	x	x	1967-06-29	02:57:00	5.339	49.8222	78.0433	49.8174	78.0476	0.195
31	101	x	x	x	x	1967-07-15	03:27:00	5.399	49.8442	78.1156	49.8366	78.1177	0.161
32	7	x	x	x	x	1967-08-04	06:58:00	5.337	49.7544	78.0503	49.7609	78.0562	0.16
33	137	x	x	x	x	1967-10-17	05:04:00	5.667	49.7844	77.9978	49.7805	78.0031	0.181
34	96	x	x	x	x	1967-10-30	06:04:00	5.444	49.7972	77.9994	49.7942	78.0052	0.173
35	102	x	x	x	x	1967-12-08	06:04:00	5.33	49.8183	78.1708	49.8177	78.1637	0.15
36	n/a	x	x	x	x	1968-01-07	03:47:00	5.005	49.7539	78.0469	n/a	n/a	0.237
37	100	x	x	x		1968-04-24	10:36:00	4.91	49.8417	78.1072	49.8459	78.1035	0.127
38	110	x	x	x	x	1968-06-11	03:06:00	5.262	49.7986	78.15	49.7934	78.1447	0.148
39	113	x	x	x	x	1968-07-12	12:08:00	5.208	49.7556	78.0992	49.7554	78.09	0.172
40	n/a	x			x	1968-08-20	04:06:00	4.788	49.8206	78.0811	n/a	n/a	0.208
41	104	x	x	x	x	1968-09-05	04:06:00	5.472	49.7381	78.0808	49.7423	78.0756	0.162
42	n/a	x	x	x	x	1968-09-29	03:43:00	5.887	49.8042	78.1056	n/a	n/a	0.29
43	99	x			x	1968-10-29	03:54:00	n/a	49.8333	78.0928	49.831	78.0879	n/a
44	111	x		x	x	1968-11-09	02:54:00	4.801	49.7997	78.1444	49.8007	78.1388	0.125
45	103	x	x	x	x	1968-12-18	05:02:00	5.066	49.7408	78.0878	49.7467	78.0916	0.194
46	n/a	x	x	x	x	1969-03-07	08:27:00	5.694	49.8283	78.0597	n/a	n/a	0.214
47	124	x	x	x	x	1969-05-16	04:03:00	5.279	49.7678	78.0817	49.7598	78.0754	0.184
48	125	x	x	x		1969-07-04	02:47:00	5.268	49.7542	78.1092	49.747	78.111	0.219
49	n/a	x	x	x	x	1969-07-23	02:47:00	5.529	49.8197	78.1372	n/a	n/a	0.175
50	98	x	x	x	x	1969-09-11	04:02:00	4.951	49.7778	77.9864	49.7762	77.995	0.19
51	112	x	x	x	x	1969-10-01	04:03:00	5.275	49.7864	78.1081	49.7829	78.097	0.144
52	126	x	x	x	x	1970-01-29	07:03:00	5.624	49.7922	78.1131	49.7957	78.123	0.214
53	115	x	x		x	1970-03-27	05:03:00	4.96	49.7444	78.0033	49.7479	77.9978	0.138
54	120	x	x		x	1970-06-28	01:58:00	5.891	49.7822	78.0533	49.7774	78.0493	0.332
55	26	x	x	x	x	1970-07-24	03:57:00	5.353	49.8136	78.1322	49.8099	78.1286	0.154
56	78	x		x	x	1970-12-17	07:01:00	5.447	49.7519	78.1017	49.7462	78.099	0.193
57	23	x				1971-01-29	05:03:00	n/a	49.8053	78.1686	49.8039	78.1616	n/a
58	121	x	x	x	x	1971-04-25	03:33:00	6.084	49.7803	78.0367	49.7686	78.0331	0.296
59	25	x	x	x	x	1971-05-25	04:03:00	5.097	49.8042	78.145	49.8023	78.1386	0.132
60	16		x	x		1971-11-29	06:03:00	5.49	49.7378	78.0814	49.7441	78.0787	0.203
61	114		x	x		1971-12-30	06:21:00	5.868	49.755	78.0475	49.7607	78.0362	0.249
62	86		x	x		1972-03-10	04:57:00	5.487	49.7519	78.1181	49.7473	78.1199	0.171
63	76	x	x	x	x	1972-03-28	04:22:00	5.199	49.7342	78.0825	49.734	78.0747	0.124
64	145		x	x		1972-07-06	01:03:00	4.363	49.7361	78.1092	49.738	78.1097	0.081
65	123		x	x		1972-08-16	03:17:00	5.13	49.7706	78.0572	49.766	78.0587	0.139
66	11		x	x		1972-12-28	04:27:00	n/a	49.7361	78.1072	49.7437	78.105	0.132
67	22	x	x	x		1973-02-16	05:03:00	5.509	49.8158	78.1067	49.8061	78.1034	n/a
68	n/a		x	x	x	1973-07-10	01:27:00	5.353	49.7911	78.0128	n/a	n/a	n/a
69	89	x	x	x	x	1973-10-26	04:27:00	5.266	49.7522	78.1325	49.7469	78.132	n/a

70	65		x	x	x	1974-05-16	03:03:00	5.276	49.7397	78.0358	49.7459	78.0374	n/a
71	n/a	x	x			1974-06-25	03:57:00	4.48	49.8306	78.1014	n/a	n/a	n/a
72	80	x	x	x	x	1974-07-10	02:57:00	5.201	49.7772	78.1125	49.7791	78.1027	n/a
73	68	x	x	x		1974-09-13	03:03:00	5.176	49.7697	78.0553	49.7653	78.0549	n/a
74	45	x				1974-12-16	06:41:00	4.969	49.8331	78.0267	49.8329	78.0318	n/a
75	13	x	x	x		1975-03-11	05:43:00	5.454	49.7547	78.1075	49.7462	78.106	n/a
76	56	x	x	x		1975-06-08	03:27:00	5.578	49.7517	78.0058	49.7606	78.0125	n/a
77	27	x	x	x		1975-08-07	03:57:00	5.218	49.8019	78.1306	49.8038	78.1234	n/a
78	18	x	x	x	x	1975-12-13	04:57:00	5.026	49.8133	78.1086	49.8074	78.1091	n/a
79	24	x	x	x	x	1976-01-15	04:47:00	5.227	49.81	78.1714	49.812	78.1576	n/a
80	13	x	x	x		1976-04-21	04:58:00	4.975	49.7547	78.1075	49.7462	78.106	n/a
81	54	x	x	x		1976-05-19	02:57:00	4.761	49.7775	78.0156	49.777	78.0009	n/a
82	72	x	x	x		1976-07-23	02:33:00	4.992	49.7433	78.0517	49.7492	78.0617	n/a
83	41	x				1976-10-30	04:57:00	4.594	49.8314	78.0572	49.8285	78.0516	n/a
84	121	x		x		1976-12-30	03:57:00	5.113	49.7803	78.0367	49.7686	78.0331	n/a
85	122		x			1977-03-29	03:57:00	0	49.7764	78.0175	49.7704	78.0136	n/a
86	18	x	x	x		1977-04-25	04:07:00	5.108	49.8133	78.1086	49.8074	78.1091	n/a
87	64	x	x	x	x	1977-07-30	01:57:00	5.162	49.7506	78.0492	49.75	78.0399	n/a
88	21	x	x	x	x	1977-08-17	04:27:00	5.03	49.8308	78.1139	49.8278	78.1184	n/a
89	27	x	x	x		1977-12-26	04:03:00	4.952	49.8019	78.1306	49.8038	78.1234	n/a
90	117	x		x	x	1978-03-26	03:57:00	5.727	49.7619	77.9825	49.7643	77.9993	n/a
91	88	x	x	x	x	1978-04-22	03:07:00	5.369	49.7517	78.1317	49.7469	78.1251	n/a
92	34	x	x	x	x	1978-05-29	04:57:00	4.717	49.7914	78.0944	49.7986	78.1022	n/a
93	n/a	x	x	x	x	1978-07-28	02:47:00	5.766	49.755	78.145	n/a	n/a	n/a
94	110			x		1978-09-20	05:03:00	4.287	49.7986	78.15	49.7934	78.1447	n/a
95	83	x	x	x		1978-10-15	05:37:00	5.183	49.7367	78.1111	49.7395	78.1127	n/a
96	79	x	x	x	x	1978-10-31	04:17:00	5.281	49.7886	78.1075	49.7875	78.0974	n/a
97	22	x		x	x	1978-12-14	04:43:00	4.76	49.8158	78.1067	49.8061	78.1034	n/a
98	127	x			x	1978-12-20	04:33:00	4.748	49.8108	78.0542	49.8169	78.0542	n/a
99	117	x	x	x	x	1979-05-06	03:17:00	5.255	49.7619	77.9825	49.7643	77.9993	n/a
100	35	x	x	x	x	1979-05-31	05:55:00	5.297	49.8342	78.0803	49.8267	78.0757	n/a
101	64	x		x		1979-09-27	04:13:00	4.459	49.7506	78.0492	49.75	78.0399	n/a
102	30	x	x	x	x	1979-10-18	04:17:00	5.268	49.8242	78.0975	49.82	78.1003	n/a
103	77	x			x	1979-11-30	04:53:01	4.462	49.7831	78.0867	49.7794	78.0953	n/a
104	126		x	x	x	1979-12-21	04:42:00	4.762	49.7922	78.1131	49.7957	78.123	n/a
105	70	x	x	x	x	1980-04-10	04:07:00	5.023	49.7825	78.0572	49.778	78.0547	n/a
106	63	x	x	x	x	1980-05-22	03:57:00	5.567	49.7797	78.0364	49.7739	78.0287	n/a
107	133	x	x	x	x	1980-07-31	03:33:00	5.369	49.7906	78.0908	49.7955	78.0907	n/a
108	142	x	x		x	1980-09-25	06:21:13	4.875	49.7833	78.0806	49.7855	78.0805	n/a
109	74	x	x	x	x	1981-06-30	01:57:15	5.202	49.7675	78.0808	49.7669	78.0744	n/a
110	17		x		x	1981-07-17	02:37:18	5.121	49.8014	78.1314	49.8064	78.1352	n/a
111	71	x	x	x	x	1981-08-14	02:27:15	4.925	49.7522	78.0531	49.7587	78.0565	n/a
112	14	x		x	x	1981-11-20	04:57:05	5.059	49.7367	78.1042	49.7401	78.0965	n/a
113	35	x		x	x	1981-12-22	04:31:05	5.001	49.8342	78.0803	49.8267	78.0757	n/a
114	46	x	x	x		1982-02-19	03:56:13	5.435	49.8233	78.0333	49.8136	78.0319	n/a
115	81	x	x	x	x	1982-06-25	02:03:07	4.637	49.7714	78.1108	49.7749	78.0996	n/a
116	5	x	x	x		1982-08-23	02:43:08	4.531	49.7403	78.0308	49.7473	78.0331	n/a
117	87	x		x	x	1982-09-21	02:57:03	5.201	49.7792	78.1247	49.7839	78.1347	n/a
118	62			x		1982-12-25	04:23:08	4.544	49.7811	78.035	49.7763	78.028	n/a
119	66			x	x	1983-03-30	04:17:10	4.692	49.785	78.0406	49.781	78.0413	n/a
120	73	x		x	x	1983-04-12	03:41:08	4.719	49.7856	78.0847	49.791	78.0807	n/a
121	65		x			1983-06-24	02:56:14	4.544	49.7397	78.0358	49.7459	78.0374	n/a
122	143	x		x	x	1983-09-11	06:33:13	4.559	49.7847	78.0842	49.7854	78.0806	n/a
123	31	x	x	x		1983-12-26	04:29:09	5.527	49.79	78.1092	49.7975	78.1036	n/a
124	75	x	x	x	x	1984-04-15	03:17:11	5.732	49.7606	78.0892	49.7499	78.0825	n/a
125	33	x			x	1984-09-09	02:59:09	4.949	49.8044	78.0875	49.806	78.0997	n/a
126	128	x				1984-11-23	03:55:08	4.41	49.8125	78.0594	49.8176	78.0551	n/a
127	n/a	x				1985-07-25	03:11:09	4.873	49.8192	78.1494	n/a	0	n/a
128	32	x		x		1987-02-26	04:58:24	5.327	49.8342	78.0811	49.8299	78.0835	n/a
129	55	x		x	x	1987-05-06	04:02:08	5.602	49.7758	78.0122	49.7742	77.9986	n/a
130	37	x		x		1987-06-06	02:37:09	5.38	49.8367	78.0617	49.8327	78.0704	n/a
131	57	x		x	x	1987-07-17	01:17:09	5.945	49.7758	78.0197	49.7664	78.0287	n/a
132	55			x	x	1987-12-20	02:55:09	4.514	49.7758	78.0122	49.7742	77.9986	n/a
133	119	x		x	x	1988-04-22	09:30:09	4.724	49.7903	78.1069	49.7942	78.1	n/a
134	12	x	x	x	x	1988-10-18	03:40:09	4.728	49.78	78.0172	49.7799	78.0079	n/a
135	58	x	x	x		1988-11-23	03:57:09	5.277	49.7794	78.0372	49.7726	78.0378	n/a
136	38	x		x	x	1989-02-17	04:01:09	4.973	49.8278	78.0597	49.8235	78.068	n/a
137	60	x				1989-10-04	11:30:00	4.475	49.7483	78.0094	49.7498	78.0117	n/a

Appendix D

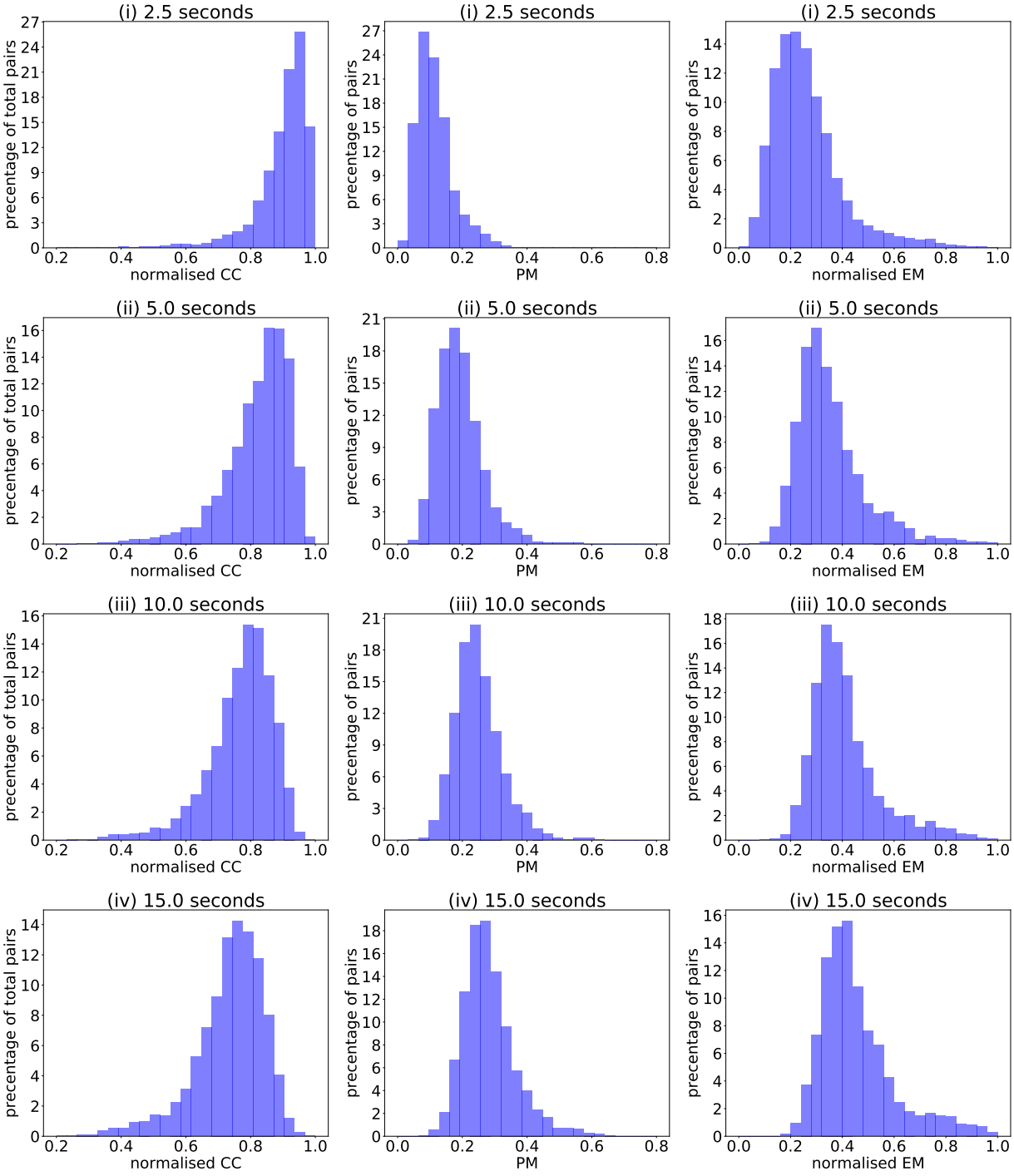
Cross-correlation, envelope and phase misfit coefficients of the Degelen seismograms

Figures D.1 - D.3 show distributions of cross-correlation (CC), phase misfit (PM) and envelope misfit (EM) coefficients for pairs of Degelen seismograms at the GBA, YKA and WRA arrays, complementing Figure 3.11 for EKA in the main text. Four time windows extending to 2.5, 5, 10 and 15 seconds after the onset of the P wave are considered. As for the EKA array in Figure 3.11, the level of similarity between waveforms decreases with an increasing time window, suggesting that the first P arrival is highly consistent. Coda waves, on the other hand, exhibit a larger variation that outweighs the similarity of the first P wave for longer time intervals. However, it is also true that CC coefficients remain high, and PM and EM coefficients remain low for a small number of pairs for all time windows considered, indicating a high level of similarity in P wave coda for some seismogram pairs.



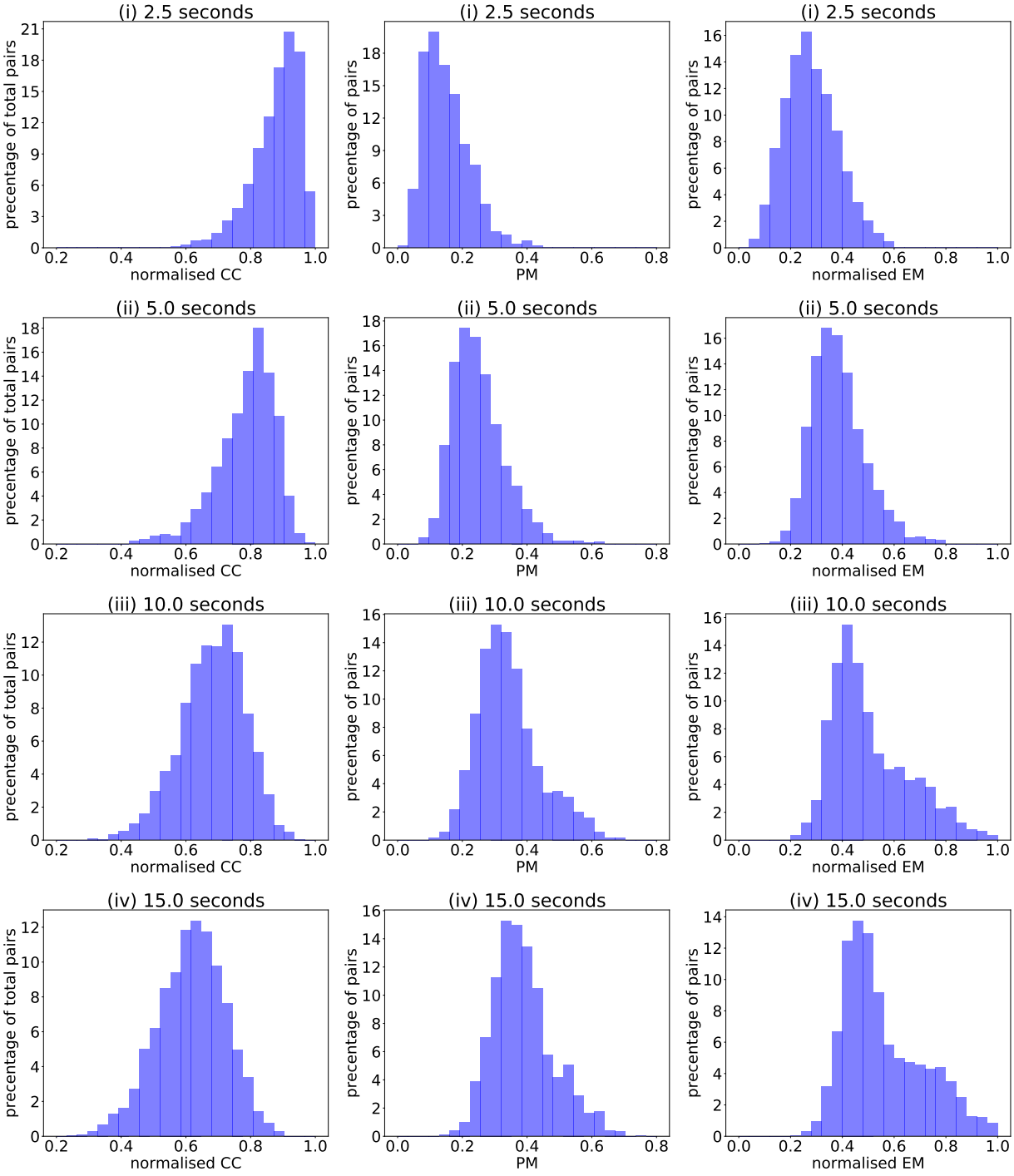
(a) CC (b) PM (c) EM

Figure D.1: Normalised CC, PM, and normalised EM coefficients as a percentage of the total number of pairs for events at GBA for time windows of (i) 2.5, (ii) 5, (iii) 10 and (iv) 15 seconds after the onset of the P wave.



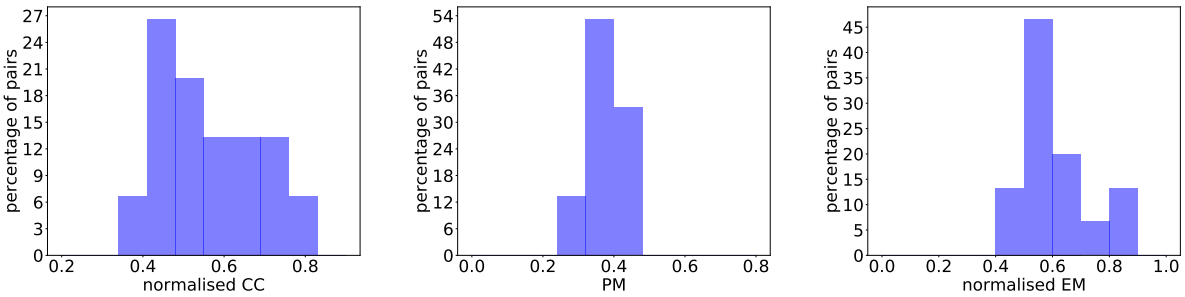
(a) CC (b) PM (c) EM

Figure D.2: Normalised CC, PM, and normalised EM coefficients as a percentage of the total number of pairs for events at YKA for time windows of (i) 2.5, (ii) 5, (iii) 10 and (iv) 15 seconds after the onset of the P wave.

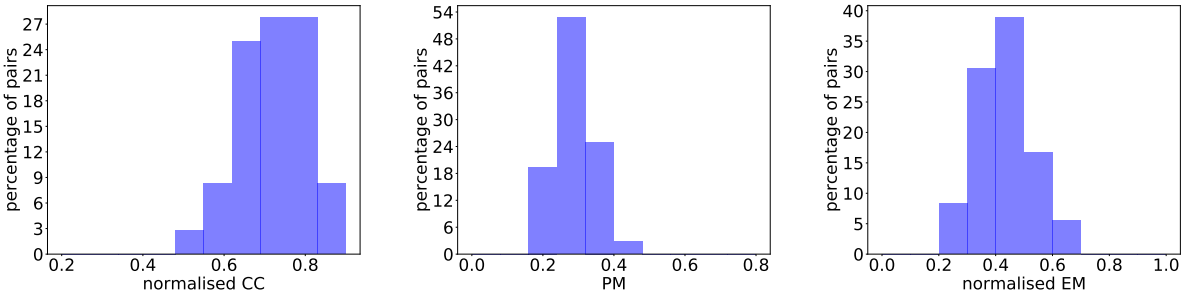


(a) CC (b) PM (c) EM

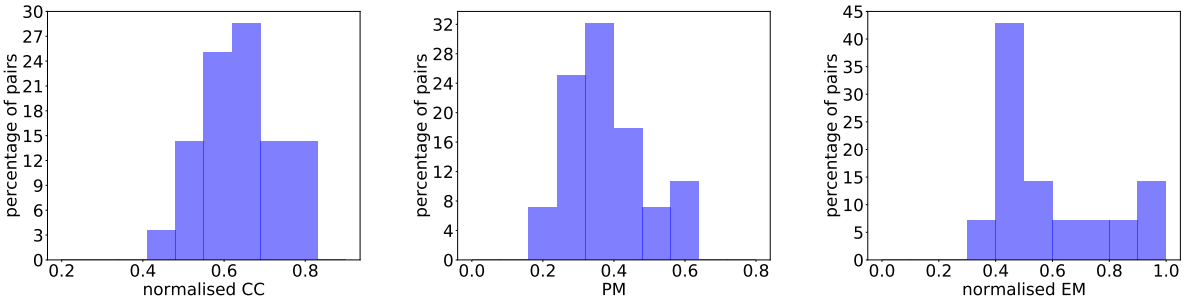
Figure D.3: Normalised CC, PM, and normalised EM coefficients as a percentage of the total number of pairs for events at WRA for time windows of (i) 2.5, (ii) 5, (iii) 10 and (iv) 15 seconds after the onset of the P wave.



(a) GBA

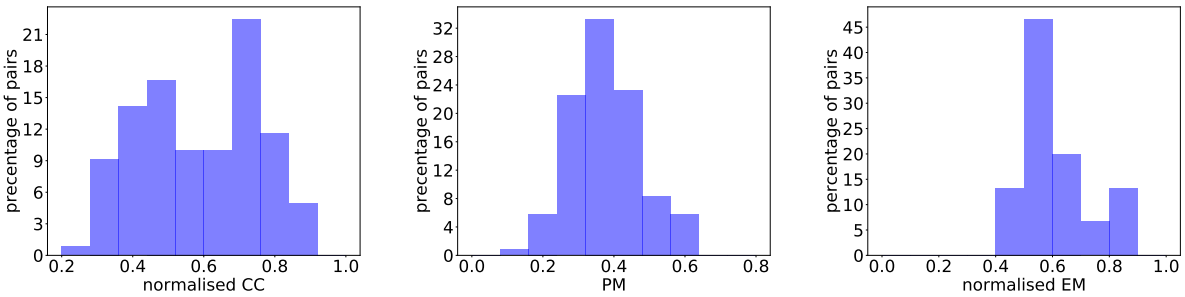


(b) YKA

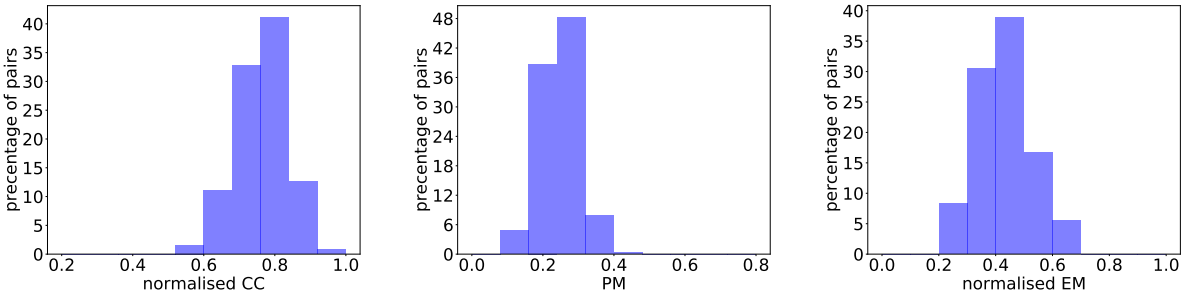


(c) WRA

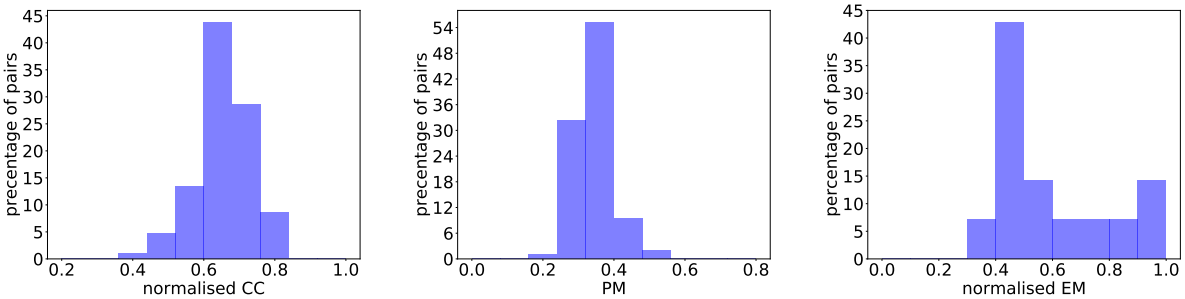
Figure D.4: Normalised CC, PM, and normalised EM coefficients as a percentage of the total number of pairs for events with depths between 180 and 200 metres at: (a) GBA, (b) YKA, and (c) WRA. Comparing distributions of the chosen subset to distributions of the total dataset for each array (Figures D.1 - D.3), we conclude that depth is not the primary factor explaining similarities and differences in the Degelen dataset.



(a) GBA



(b) YKA



(c) WRA

Figure D.5: Normalised CC, PM, and normalised EM coefficients as a percentage of the total number of pairs for events with body-wave magnitudes m_b between 5.3 and 5.5 at: (a) GBA, (b) YKA, and (c) WRA. Comparing distributions of the chosen subset to distributions of the total dataset for each array (Figures D.1 - D.3), we conclude that m_b is not the primary factor explaining similarities and differences in the Degelen dataset.

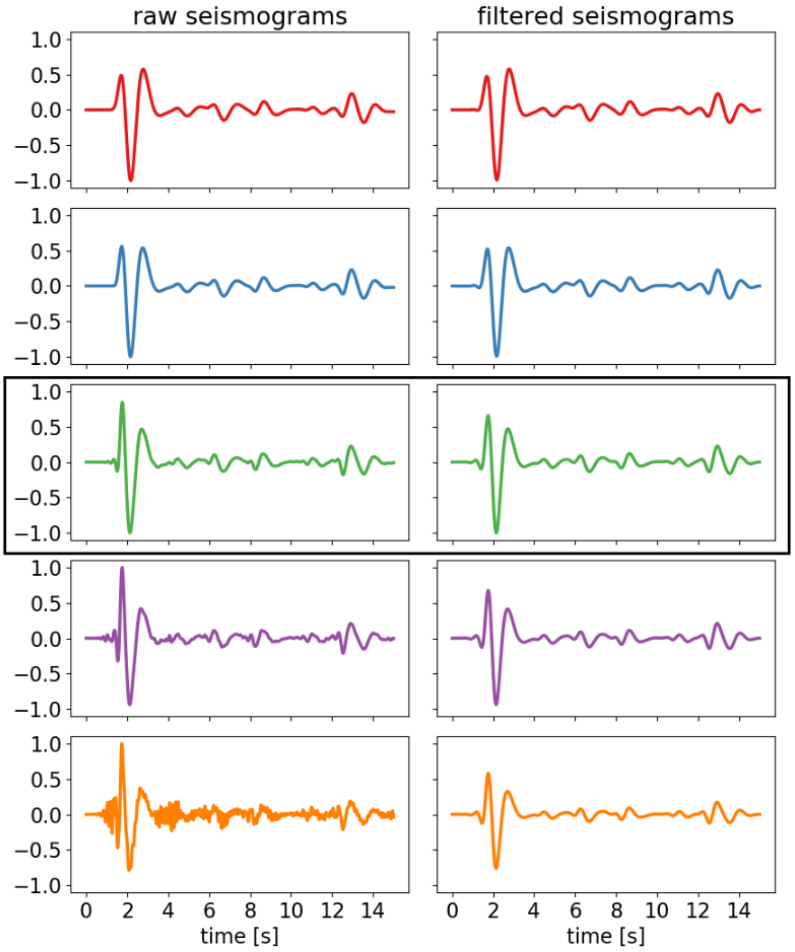
Appendix E

The resolution limit of the mesh: choosing a moment rate function

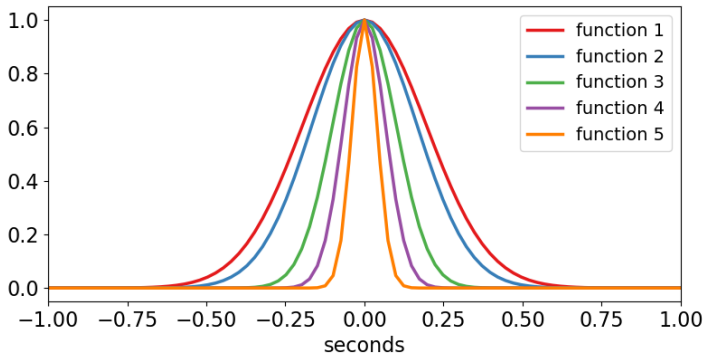
As discussed in Section 4.2.3, the Gaussian moment rate function of 0.25 s half-width is at the resolution limit of the global AxiSEM mesh. Increasing the half-width of the Gaussian below the resolution limit would reduce spectral amplitudes in the 1-2 Hz range, which in turn weakens sensitivity to small-scale topographic features and compromises on waveform detail. We therefore choose the half-width at the resolution limit for all hybrid simulations of the Degelen topography in Chapter 4, and we acknowledge that some of the energy in the 2-4 Hz range will not be well resolved.

In Figure E.1 we test the effects of convolving Gaussians of various half-widths into the 2 Hz Green's functions database in order to examine the resolution issues of the chosen moment rate. Figure E.1a displays raw (left column) and filtered (right column) seismograms that have been obtained by convolving Gaussians of progressively narrower half widths. As the half-width gets smaller, noise becomes more apparent. Most of the noise – with the exception of the negative arrival before the P wave onset – can be filtered out, although the final amplitudes of the signal are significantly affected for the Gaussian with the shortest half-width.

The moment rate function with 0.25 s half-width chosen for the simulations is indicated in green in Figure E.1b (function 3). The corresponding raw seismogram has a noisy onset and very low amplitude noise later in the signal. The low-amplitude noise is filtered out, but the negative arrival prior to the P wave onset remains in the final seismogram. The amplitudes are not severely affected and we choose to accept the instability in the onset in favour of retaining higher spectral amplitudes in the well-resolved 1-2 Hz range.



(a)



(b)

Figure E.1: Verification of effects of numerical noise for Gaussian functions above the resolution of the global AxiSEM mesh. The left column in (a) shows raw seismograms, while the right column shows the same seismograms filtered with a 2 Hz Butterworth zero-phase filter. The colours of the waveforms in (a) correspond to the Gaussian functions used in the respectrive convolutions and shown in (b). The Gaussian chosen for hybrid simulations of the Degelen topography is indicated in green (function 3), and corresponding seismograms in (a) are highlighted.

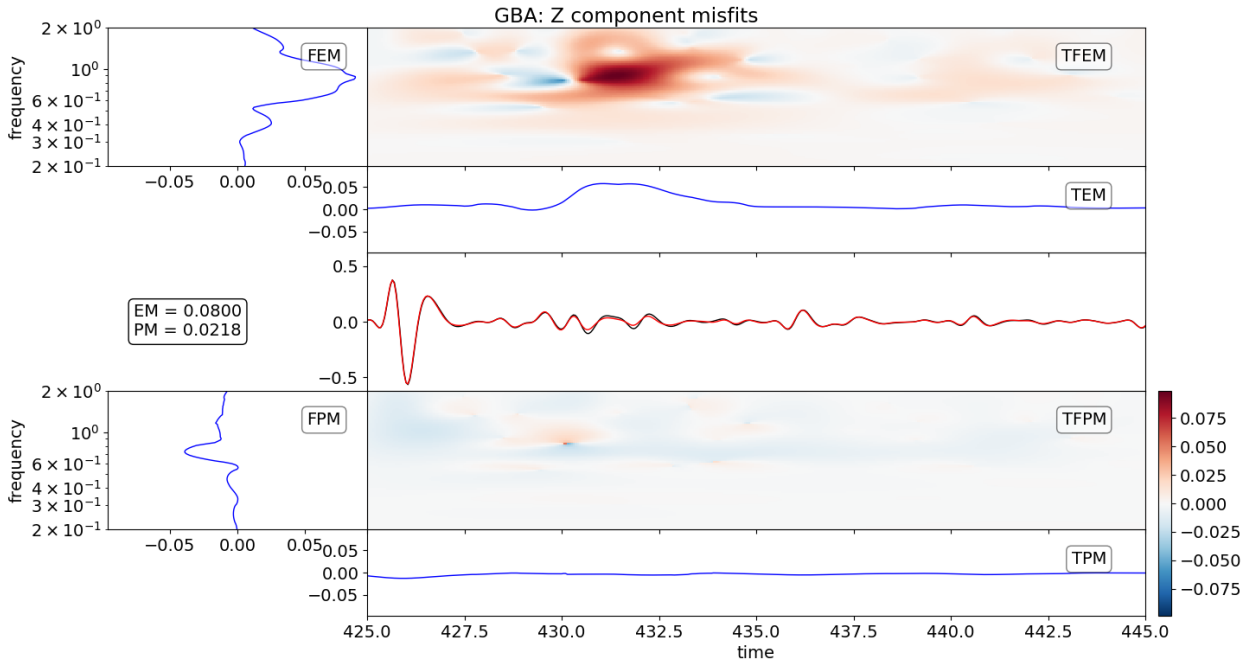
Appendix F

Degelen simulations: 1-D benchmarks

Figures F.1 and F.2 in this appendix show 1-D benchmarks of the Degelen simulations for the GBA, EKA and WRA array locations, and complement those presented Section 4.3.1 for the YKA array:

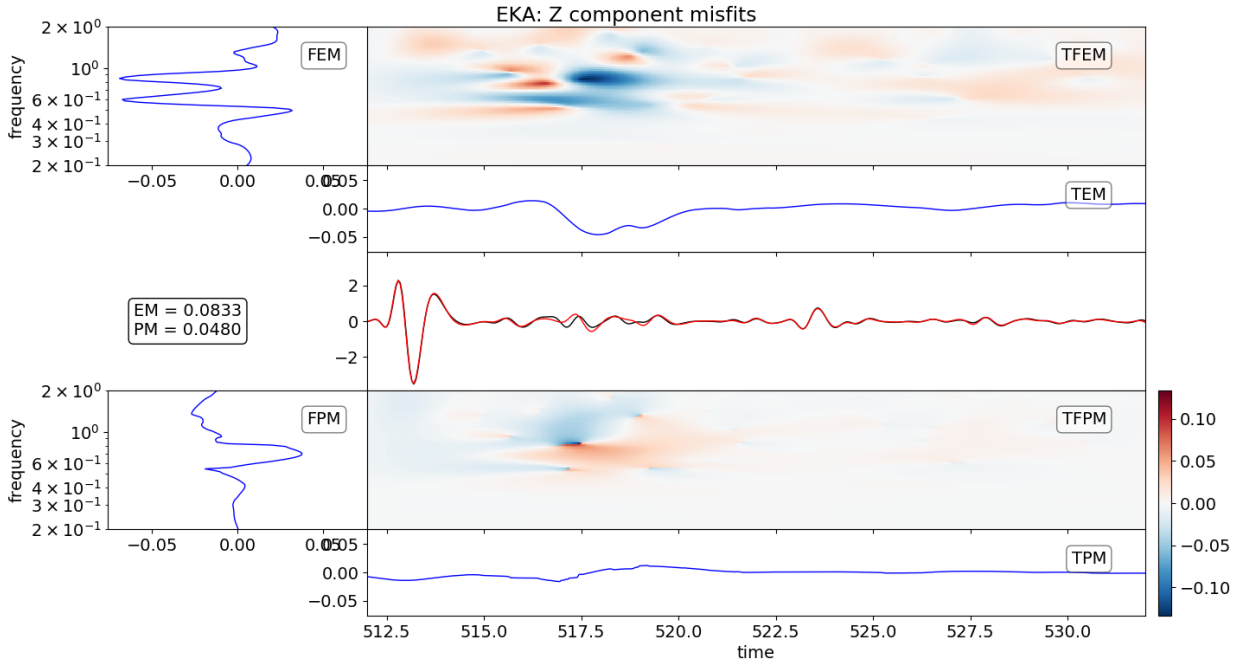
- Figure F.1 is an extension of Figure 4.7a, showing benchmarks for $30 \times 30 \times 10$ km WPP domain with 100 m spacing in WPP runs, and a coarser 200 m spacing for Instaseis repropagation. This is the configuration used in the final simulations.
- Figure F.2 is an extension of Figure 4.7b, where we can see that benchmarks improve for a smaller domain ($10 \times 10 \times 10$ km), as the discrepancy between the top surface of the WPP domain and the global spherical domain is smaller.

Figure F.3 shows the effect of grid refinement on the quality of the $10 \times 10 \times 10$ km benchmarks. In Figure F.3a, the WPP spacing remains at 100 m, but the repropagation grid is refined to 100 m. In Figure F.3b, both the WPP spacing and the repropagation grid are refined to 50 m.

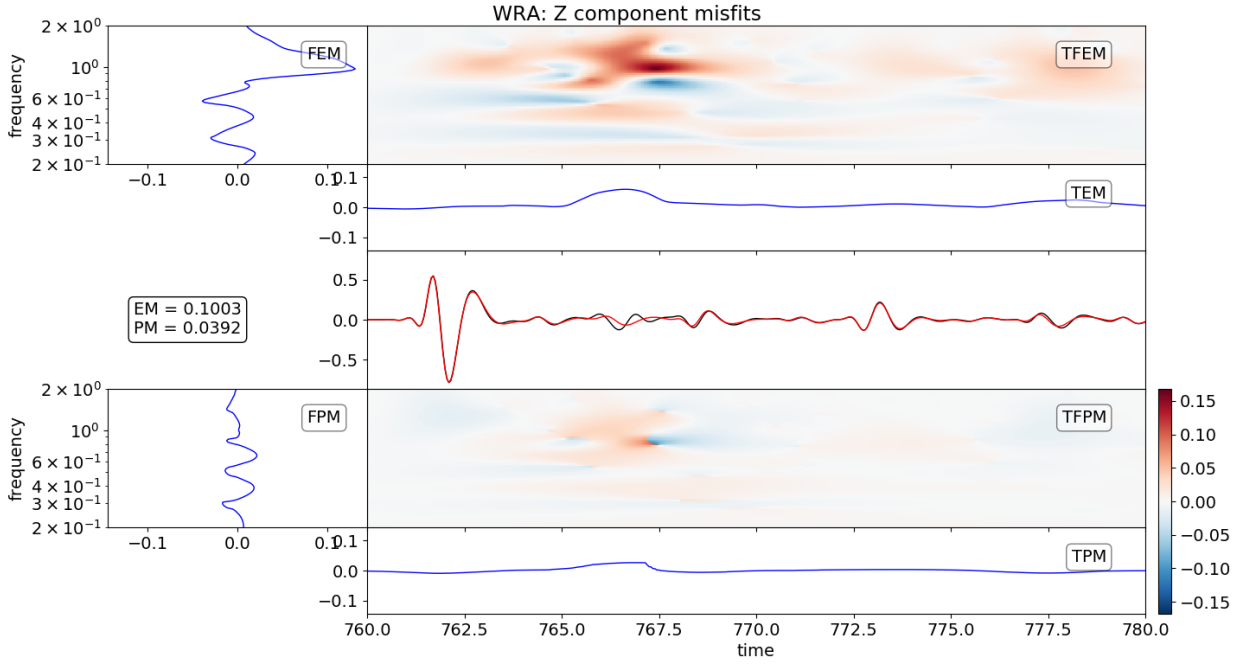


(a) GBA

Figure F.1: 1-D benchmark of the hybrid Z-component seismogram for $30 \times 30 \times 10$ km WPP domain (black line) against the AxiSEM 1-D seismogram (red line) at: (a) GBA, (b) EKA, and (c) WRA. The WPP domain is discretised with 100 m spacing in x, y and z, while a coarser grid of 200 m in all three dimensions is used for the Instaseis repropagation. The plot details the frequency-envelope misfit (FEM), the time-envelope misfit (TEM), the time-frequency-envelope misfit (TFEM), the frequency-phase misfit (FPM), the time-phase misfit (TPM), and the time-frequency-phase misfit (TFPM). Misfits result mostly from the mismatch between the local and global models due to the lack of the spherical surface in the Cartesian WPP domain.

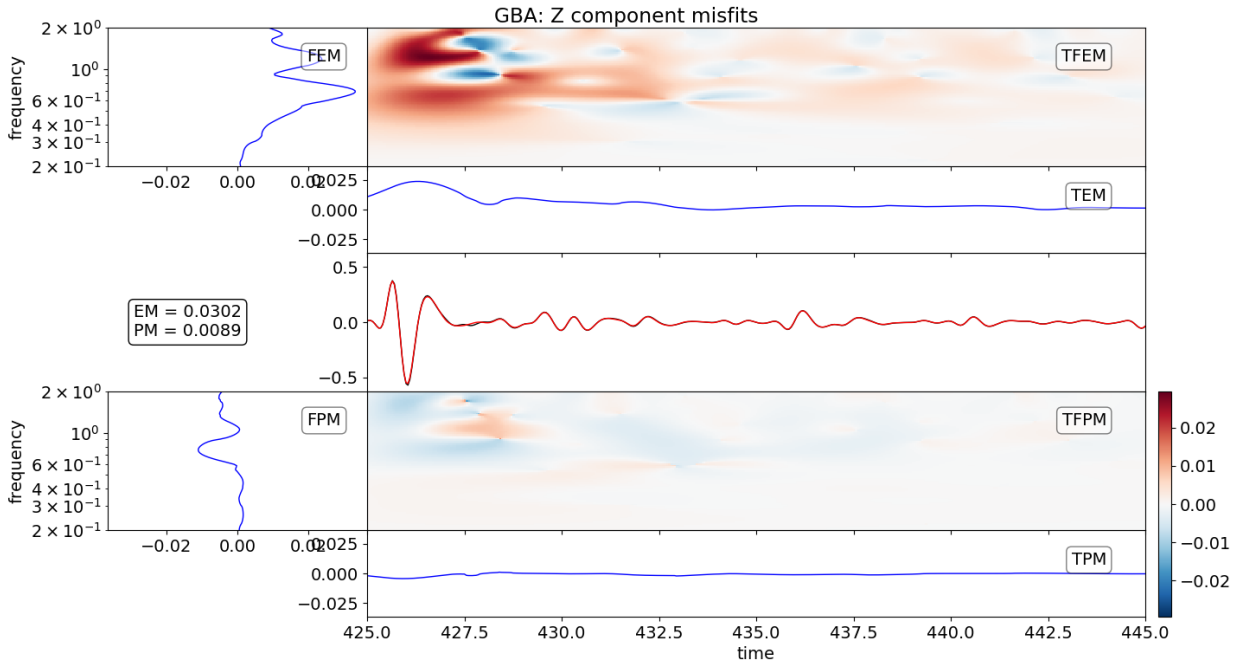


(b) EKA



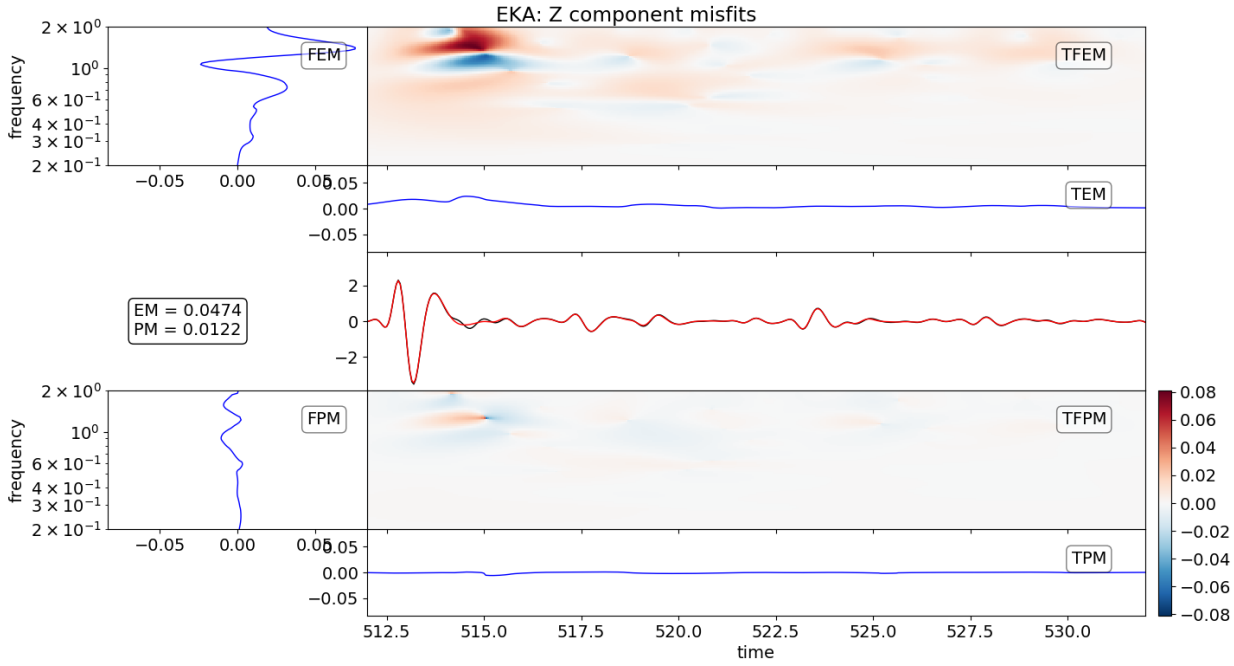
(c) WRA

Figure F.1: (continued).

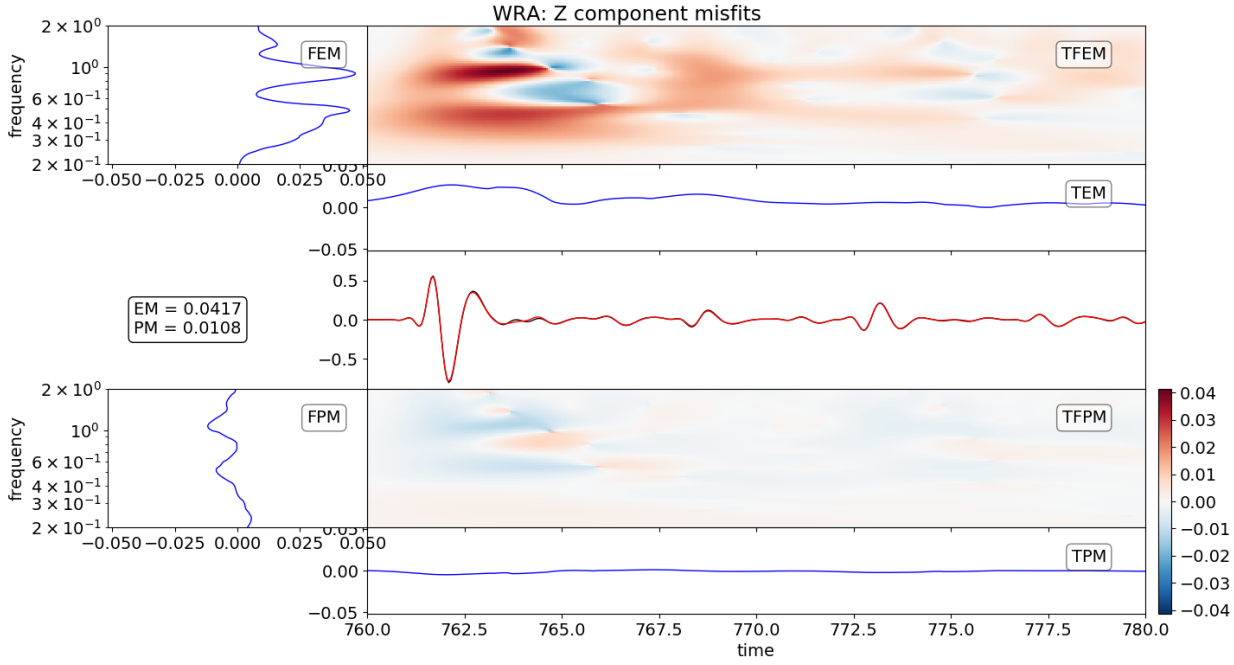


(a) GBA

Figure F.2: 1-D benchmark of the hybrid Z-component seismogram for $10 \times 10 \times 10$ km WPP domain (black line) against the AxiSEM 1-D seismogram (red line) at: (a) GBA, (b) EKA, and (c) WRA. The WPP domain is discretised with 100 m spacing in x, y and z, while a coarser grid of 200 m in all three dimensions is used for the Instaseis repropagation. The plot details the frequency-envelope misfit (FEM), the time-envelope misfit (TEM), the time-frequency-envelope misfit (TFEM), the frequency-phase misfit (FPM), the time-phase misfit (TPM), and the time-frequency-phase misfit (TFPM). Misfits are significantly reduced relative to the $30 \times 30 \times 10$ domain, as a smaller area reduces the mismatch between the local and global models due to the lack of the spherical surface in the Cartesian WPP domain.

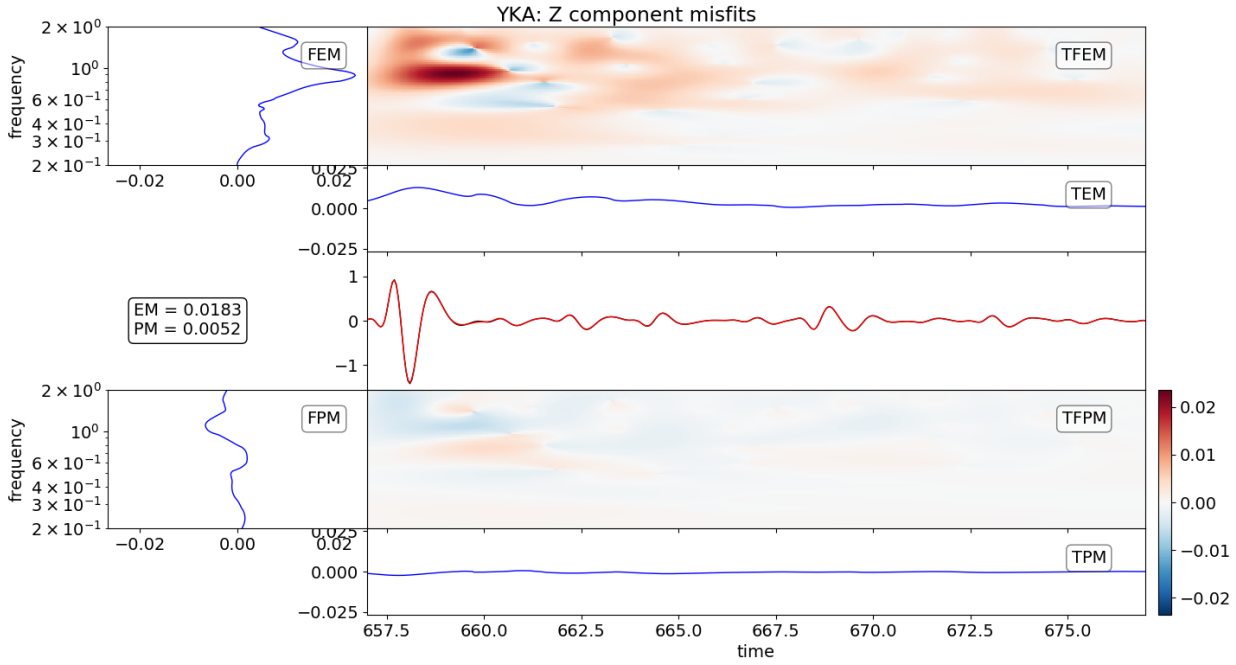


(b) EKA



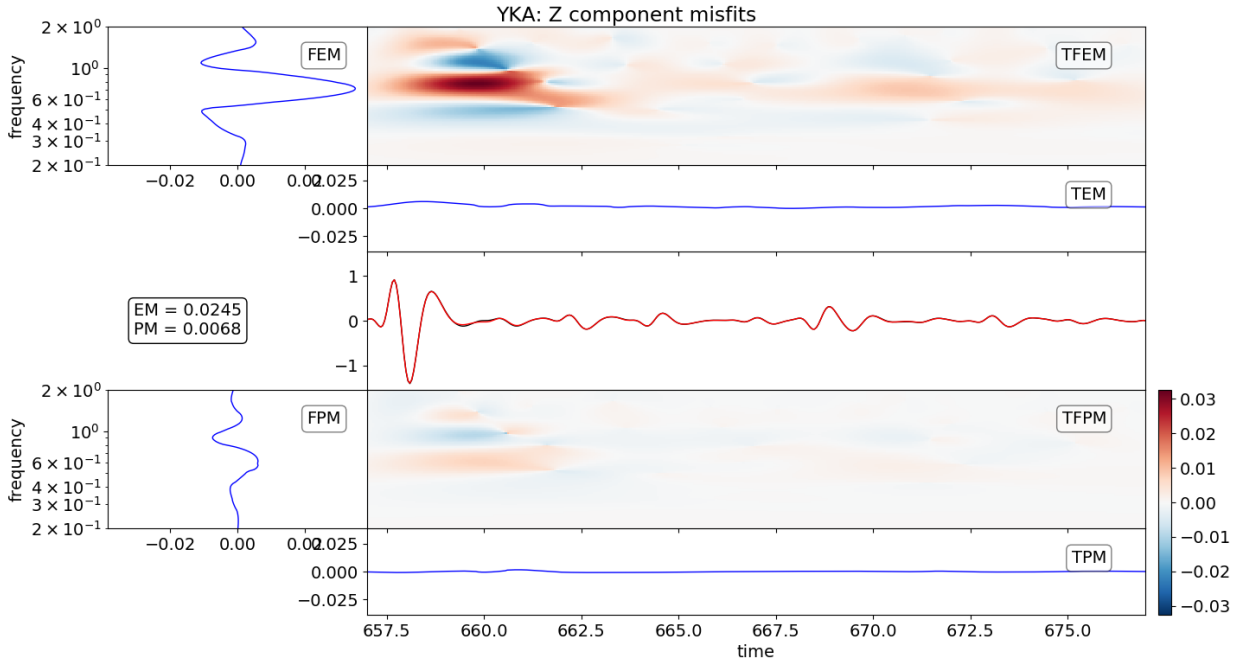
(c) WRA

Figure F.2: (continued).



(a)

Figure F.3: Example of the effect of grid refinement on the quality of 1-D benchmarks. The hybrid Z-component seismogram for $10 \times 10 \times 10$ km WPP domain (black line) is plotted against the AxiSEM 1-D seismogram (red line). (a) The WPP domain is discretised with 100 m spacing in x, y and z, and the same spacing of 100 m is used for the Instaseis repropagation.



(b)

Figure F.3: (continued). (b) The WPP domain is discretised with 50 m spacing in x, y and z – a configuration recommended for best accuracy in WPP, resulting in 15 points per wavelength for the chosen moment rate function – and the same spacing of 50 m is used for Instaseis repropagation. Notice that the numerical error does not decrease relative to (a).

Appendix G

Overview maps of synthetic results

Figures G.1-G.3 complement the overview of synthetics results for YKA (Figure 4.13, Chapter 4) with corresponding overviews for GBA, EKA and WRA. The figures illustrate the extent to which the Degelen topography affects teleseismic signals. First arrivals are either amplified or de-amplified, while coda waves change with location on the mountain ridges: distinct phases appear and disappear, some arrivals shift, and some change amplitudes.

All events have been simulated with an isotropic moment tensor source of 10^{16} Nm buried 100 m under the topography. The distances between locations in each group vary from 100 m in group E to nearly 1000 m in group D. Exact locations of events, as well as directions towards the arrays, are shown in each overview figure.

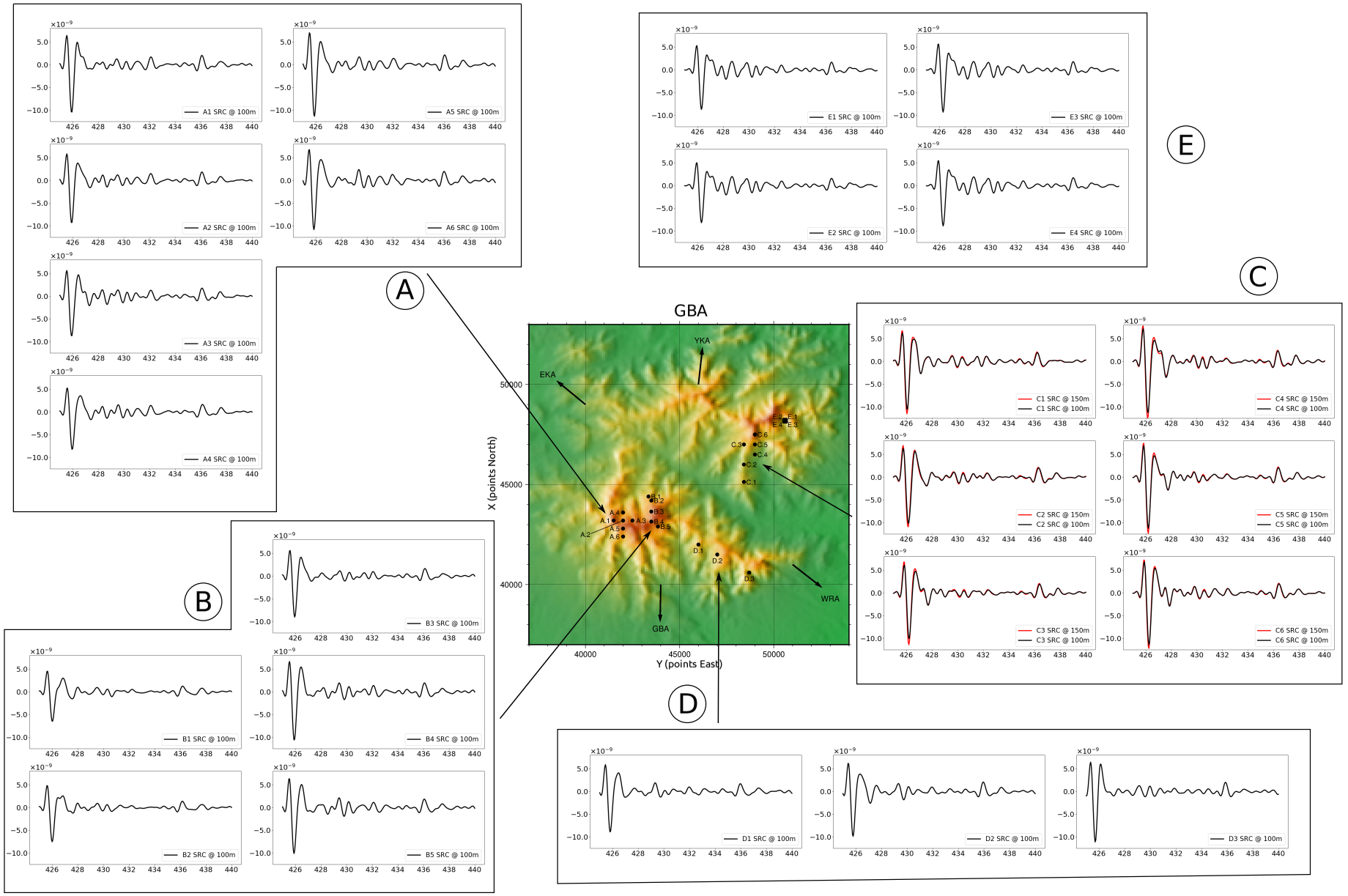


Figure G.1: Overview of synthetic waveforms from the Degelen mountain at the GBA array. First arrivals are either amplified or reduced, and coda waves change with location – new arrivals appear and disappear, some shift, and some change amplitudes. The direction from Degelen to GBA is indicated with an arrow.

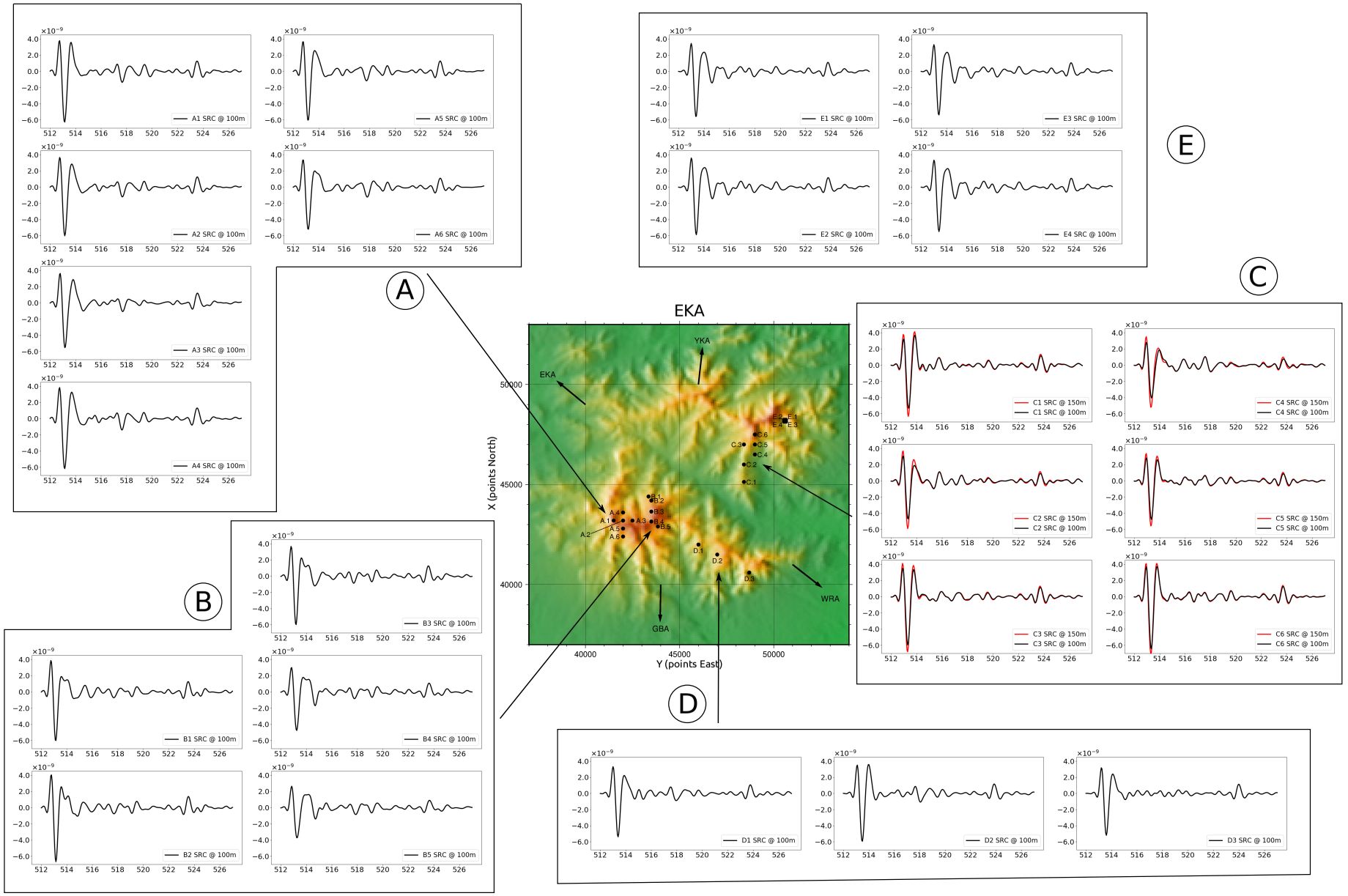


Figure G.2: Overview of synthetic waveforms from the Degelen mountain at the EKA array. First arrivals are either amplified or reduced, and coda waves change with location – new arrivals appear and disappear, some shift, and some change amplitudes. The direction from Degelen to EKA is indicated with an arrow.

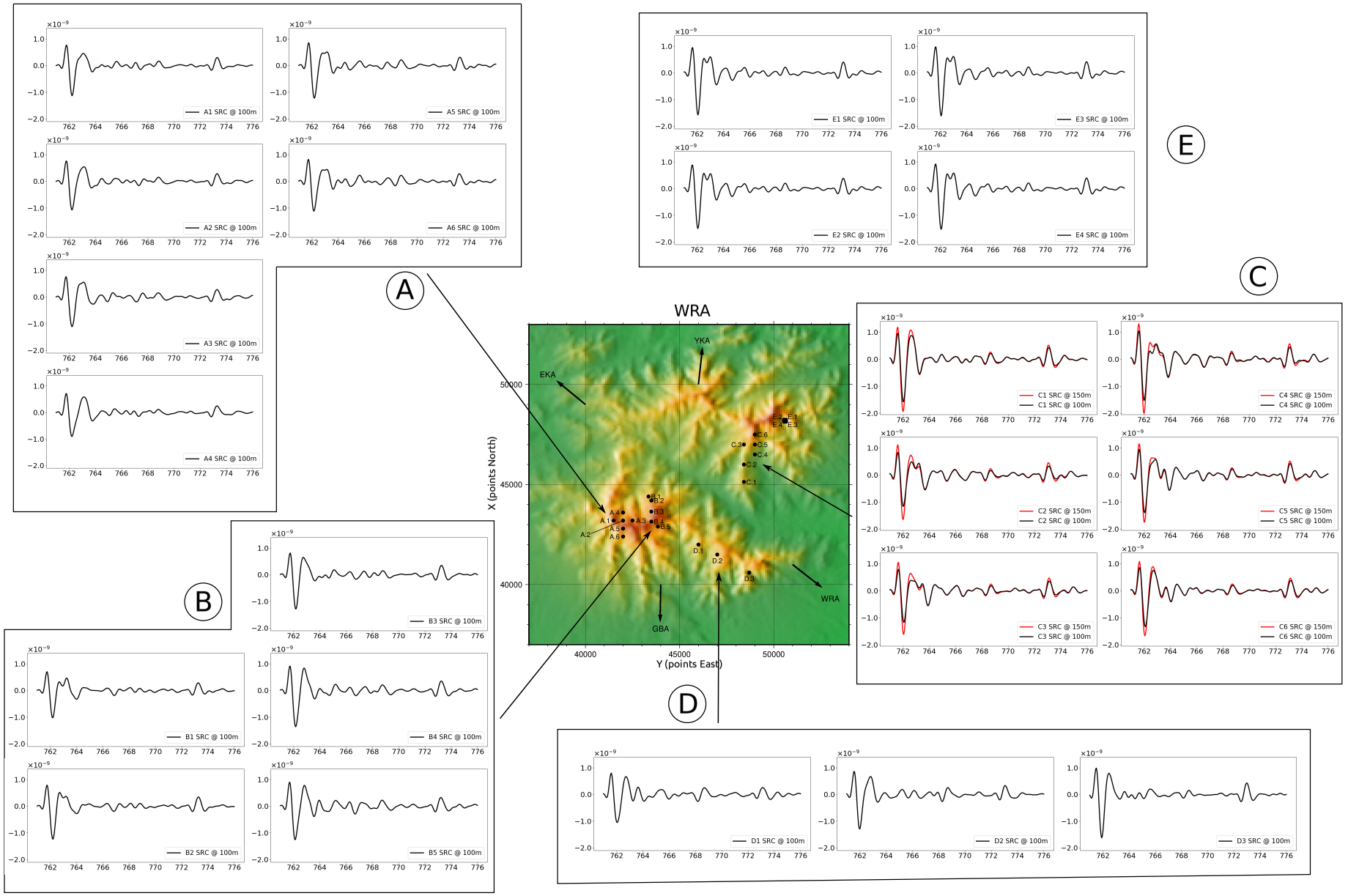


Figure G.3: Overview of synthetic waveforms from the Degelen mountain at the WRA array. First arrivals are either amplified or reduced, and coda waves change with location – new arrivals appear and disappear, some shift, and some change amplitudes. The direction from Degelen to WRA is indicated with an arrow.

Appendix H

Local source-topography interactions: location B3

Figures H.1 and H.2 show local wavefield interactions at location B2 in the Degelen mountain complex and complement similar Figures 4.17 and 4.18 for location C4 presented in Section 4.3.3. Both cross-sections intersect the source location $(13.6, 13.5)$ km: Figure H.1 is the cross section for y fixed at 13.5 km, and Figure H.2 at x fixed at 13.6 km.

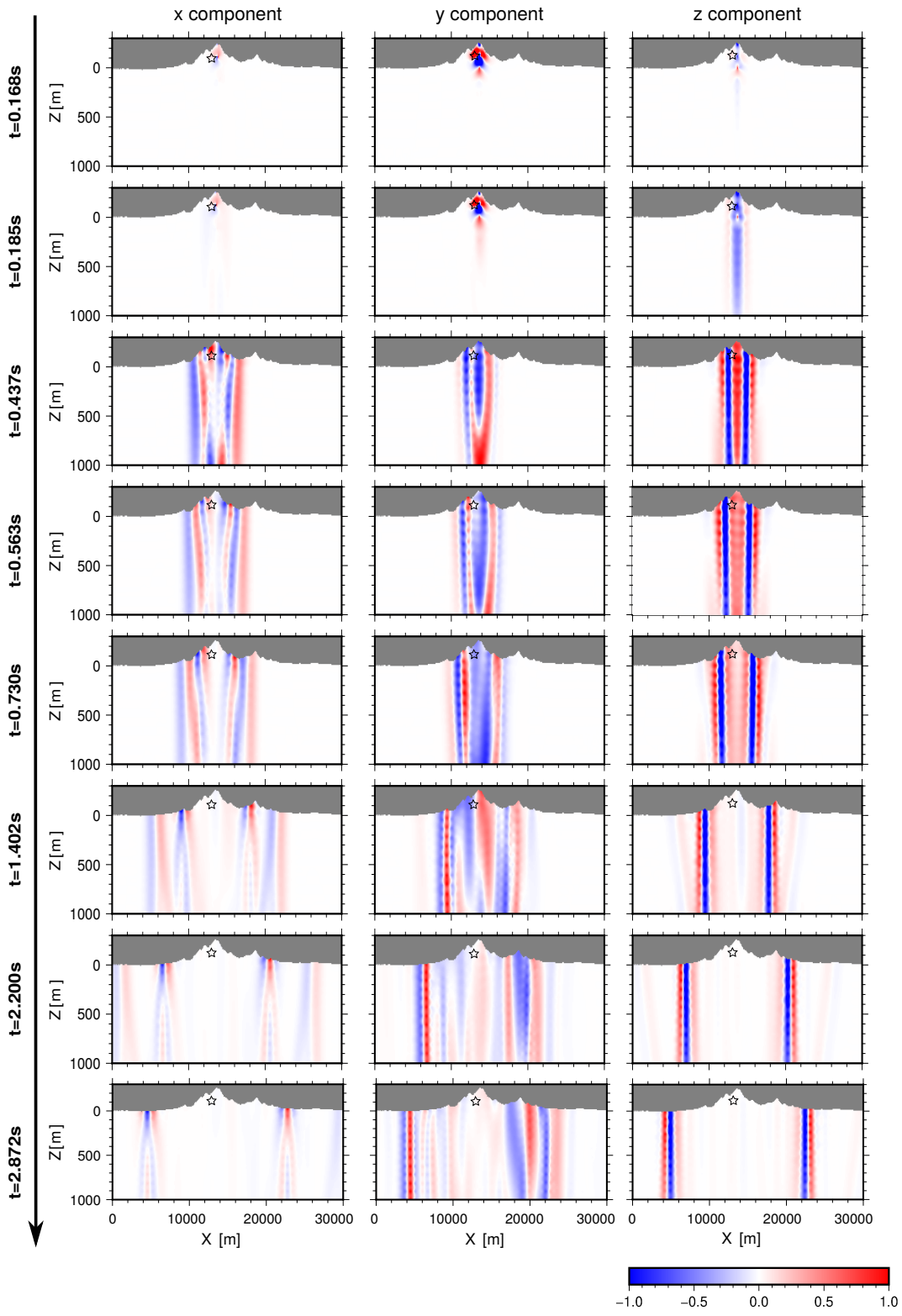


Figure H.1: Cross sections of displacement for $y = 13.5$ km for a source located at B3 at $(13.6, 13.5)$ km. All plots have been normalised by the maximum displacement amplitude at the given time step. Due to the domain size, the wavefield is horizontally exaggerated. The visible ringing is a result of a coarse WPP mesh of 8.5 elements per wavelength. As discussed in Section 4.3.1, refining the WPP mesh showed only a marginal improvement teleseismically.

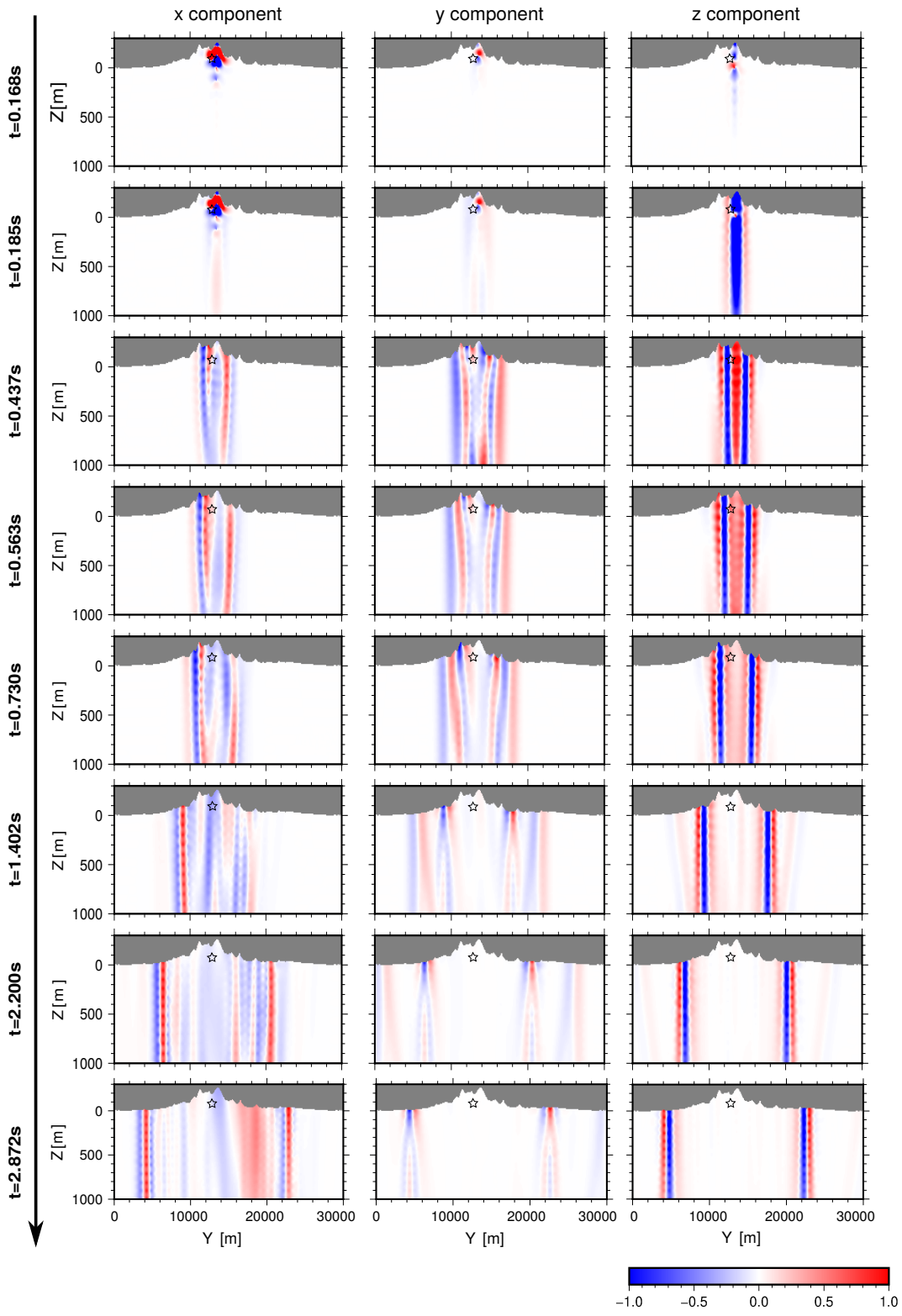


Figure H.2: Cross sections of displacement for $x = 13.6$ km for a source located at B3 at $(13.6, 13.5)$ km. All plots have been normalised by the maximum displacement amplitude at the given time step. Due to the domain size, the wavefield is horizontally exaggerated. The visible ringing is a result of a coarse WPP mesh of 8.5 elements per wavelength. As discussed in Section 4.3.1, refining the WPP mesh showed only a marginal improvement teleseismically.

Appendix I

Data v.s. synthetics

Figures in this appendix complement Figure 4.21 in Section 4.4.2 for GBA and WRA. Unfortunately, event 90 was not recorded on EKA, while both events 51 and 96 had a low signal-to-noise ratio. An equivalent plot comparing synthetics and observations was therefore not possible for EKA. The included GBA and WRA plots, however, support Figure 4.21 in the main text and show that events in close proximity result in seismograms of high similarity, while slightly larger distances on the order of 500 m can result in larger differences both in synthetic and observed signals.

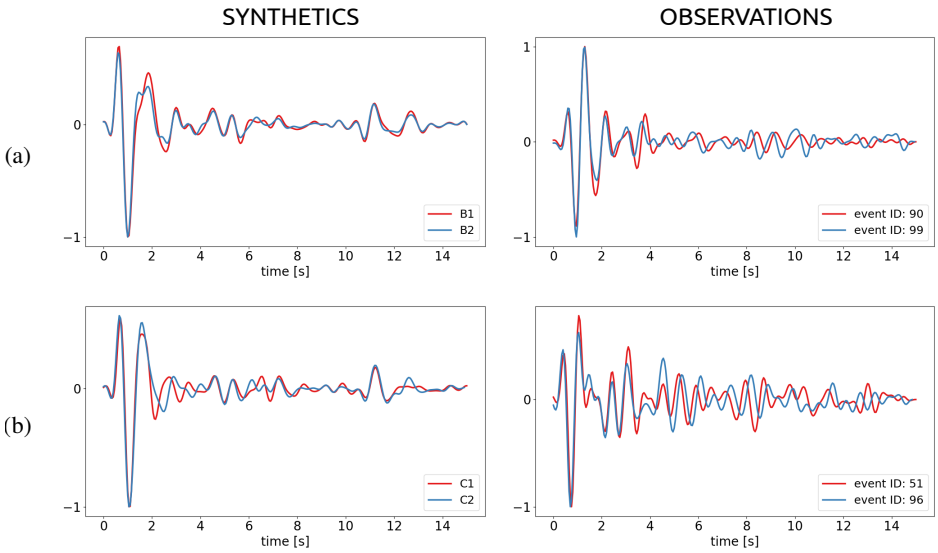


Figure I.1: Illustration of dependence of waveform similarity on source proximity in observed and synthetic datasets at GBA. Event IDs of the observed Degelen events correspond to those listed in Appendix C. Both synthetic and real locations are shown in Figure 4.20.
(a) When events are within 200 m, the first arrival and the coda (after 1.5-2 s after the onset) are nearly identical for event pairs both in synthetics (B1&B2) and in real data (90&99).
(b) When events are located 500 m apart, differences in coda start to appear in both datasets (synthetics C1&C2, observations 51&96), although an overall high level of coherence is preserved.

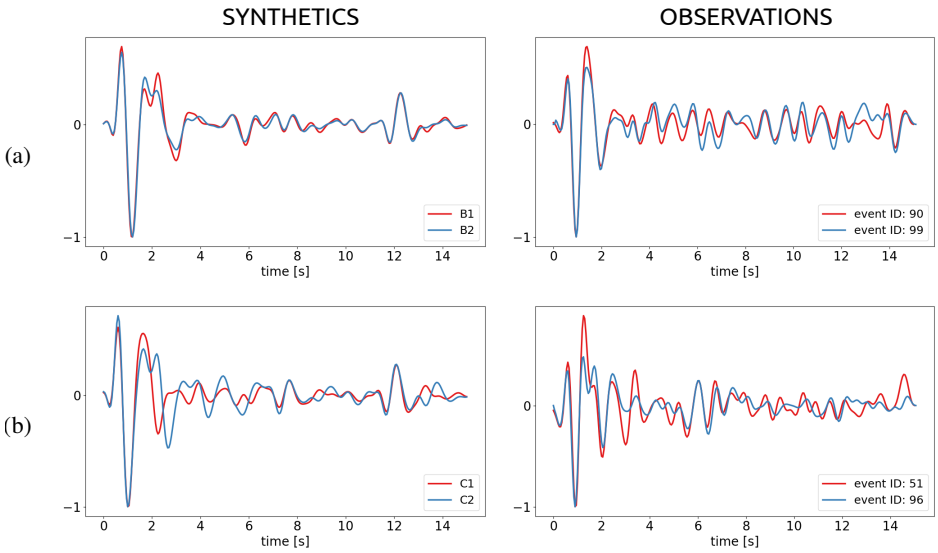


Figure I.2: Illustration of dependence of waveform similarity on source proximity in observed and synthetic datasets at WRA. Event IDs of the observed Degelen events correspond to those listed in Appendix C. Both synthetic and real locations are shown in Figure 4.20.
(a) When events are within 200 m, the first arrival and the coda (after 1.5-2 s after the onset) are nearly identical for event pairs both in synthetics (B1&B2) and in real data (90&99).
(b) When events are located 500 m apart, differences in coda start to appear in both datasets (synthetics C1&C2, observations 51&96), although an overall high level of coherence is preserved.

References

Adushkin, V. and W. Leith (2001). *The Containment of Soviet Underground Nuclear Explosions*. Open File Report 01-312. U.S. Department of the Interior Geological Survey.

Aki, K. and P. G. Richards (2002). *Quantitative Seismology*. Second Edition. Sausalito, CA: University Science Books.

Al-Attar, D. and J. H. Woodhouse (2008). "Calculation of Seismic Displacement Fields in Self-Gravitating Earth Models—Applications of Minors Vectors and Symplectic Structure". In: *Geophysical Journal International* 175.3, pp. 1176–1208.

Alterman, Z. and F. Karal (1968). "Propagation of Elastic Waves in Layered Media by Finite Difference Methods". In: *Bulletin of the Seismological Society of America* 58.1, pp. 367–398.

Auer, L., L. Boschi, T. W. Becker, T. Nissen-Meyer, and D. Giardini (2014). "Savani: A Variable Resolution Whole-mantle Model of Anisotropic Shear Velocity Variations Based on Multiple Data Sets". In: *Journal of Geophysical Research: Solid Earth* 119.4, pp. 3006–3034.

Avants, M. (2014). "Effects Of Near-Source Heterogeneity On Wave Fields Emanating From Crustal Sources Observed At Regional And Teleseismic Distances". PhD thesis. Santa Cruz: University of California.

Bache, T. C., P. D. Marshall, and L. B. Bache (1985). "Q for Teleseismic P Waves from Central Asia". In: *Journal of Geophysical Research* 98.B5, pp. 3575–3587.

Becker, T. W. and L. Boschi (2002). "A Comparison of Tomographic and Geodynamic Mantle Models". In: *Geochemistry, Geophysics, Geosystems* 3.1.

Beller, S, V Monteiller, S Operto, G Nolet, A Paul, and L Zhao (2018). "Lithospheric Architecture of the South-Western Alps Revealed by Multiparameter Teleseismic Full-Waveform Inversion". In: *Geophysical Journal International* 212.2, pp. 1369–1388.

Belyashova, N. N., V. I. Shacilov, N. N. Mikhailova, I. I. Komarov, Z. I. Sinyova, A. V. Belyashov, and M. N. Malakhova (2001). "On the Use of Calibration Explosions at the Former Semipalatinsk Test Site for Compiling a Travel-Time Model of the Crust and Upper Mantle". In: *Pure and Applied Geophysics* 158.1-2, pp. 193–209.

Ben-Zion, Y. and J.-P. Ampuero (2009). "Seismic Radiation from Regions Sustaining Material Damage". In: *Geophysical Journal International* 178.3, pp. 1351–1356.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, and J. Wasserman (2010). "ObsPy: A Python Toolbox for Seismology". In: *Seismological Research Letters* 81.3, pp. 530–533.

Bielak, J. (2003). "Domain Reduction Method for Three-Dimensional Earthquake Modeling in Localized Regions, Part I: Theory". In: *Bulletin of the Seismological Society of America* 93.2, pp. 817–824.

Bielak, J. and P. Christiano (1984). "On the Effective Seismic Input for Non-Linear Soil-Structure Interaction Systems". In: *Earthquake Engineering & Structural Dynamics* 12.1, pp. 107–119.

Bocharov, V. S., S. A. Zelentsov, and V. N. Mikhajlov (1989). "Characteristics of the 96 underground nuclear explosions on the Semipalatinsk test site". In: *Atomnaya Ehnergiya* 67.3, pp. 210–214.

Boschi, L., I. Molinari, and M. Reinwald (2018). "A Simple Method for Earthquake Location by Surface-Wave Time Reversal". In: *Geophysical Journal International* 215.1, pp. 1–21.

Böttger, G., C. Denecke, R. Görner, M. Viehweg, H. Krause, and L. Schneider (1998). *Evaluation of the Suitability of the Sites and Facilities at Degelen and Azgir in the Republic of Kazakhstan as a Final Repository for Radioactive Waste*. Tech. rep. EUR 17633. Luxembourg: Office for Official Publications of the European Communities.

Bouchon, M. (1976). "Teleseismic Body Wave Radiation from a Seismic Source in a Layered Medium". In: *Geophysical Journal International* 47.3, pp. 515–530.

Bowers, D. (1996). "On the Probability of Mistaking an Earthquake for an Explosion Using the Simplicity of P". In: *Bulletin of the Seismological Society of America* 86.6, pp. 1925–1934.

Bowers, D. and N. D. Selby (2009). "Forensic Seismology and the Comprehensive Nuclear-Test-Ban Treaty". In: *Annual Review of Earth and Planetary Sciences* 37, pp. 209–236.

Broggini, F., M. Vasmel, J. O. A. Robertsson, and D.-J. van Manen (2017). "Immersive Boundary Conditions: Theory, Implementation, and Examples". In: *GEOPHYSICS* 82.3, T97–T110.

Burgos, G., Y. Capdeville, and L. Guillot (2016). "Homogenized Moment Tensor and the Effect of Near-Field Heterogeneities on Nonisotropic Radiation in Nuclear Explosion". In: *Journal of Geophysical Research: Solid Earth* 121.6, pp. 4366–4389.

Capdeville, Y., A. To, and B. Romanowicz (2003a). "Coupling Spectral Elements and Modes in a Spherical Earth: An Extension to the 'Sandwich' Case". In: *Geophysical Journal International* 154.1, pp. 44–57.

Capdeville, Y., E. Chaljub, and J. P. Montagner (2003b). "Coupling the Spectral Element Method with a Modal Solution for Elastic Wave Propagation in Global Earth Models". In: *Geophysical Journal International* 152.1, pp. 34–67.

Cerjan, C., D. Kosloff, R. Kosloff, and M. Reshef (1985). "A Nonreflecting Boundary Condition for Discrete Acoustic and Elastic Wave Equations". In: *GEOPHYSICS* 50.4, pp. 705–708.

Chaljub, E., P. Moczo, S. Tsuno, P.-Y. Bard, J. Kristek, M. Kaser, M. Stupazzini, and M. Kristekova (2010). "Quantitative Comparison of Four Numerical Predictions of 3D Ground

Motion in the Grenoble Valley, France". In: *Bulletin of the Seismological Society of America* 100.4, pp. 1427–1455.

Chaljub, E., D. Komatitsch, J.-P. Vilotte, Y. Capdeville, B. Valette, and G. Festa (2007). "Spectral-Element Analysis in Seismology". In: *Advances in Geophysics* 48, pp. 365–419.

Chevrot, S., N. Favier, and D. Komatitsch (2004). "Shear Wave Splitting in Three-Dimensional Anisotropic Media". In: *Geophysical Journal International* 159.2, pp. 711–720.

Clouzet, P., Y. Masson, and B. Romanowicz (2018). "Box Tomography: First Application to the Imaging of Upper-Mantle Shear Velocity and Radial Anisotropy Structure beneath the North American Continent". In: *Geophysical Journal International* 213.3, pp. 1849–1875.

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999). "The TauP Toolkit: Flexible Seismic Travel-Time and Ray-Path Utilities". In: *Seismological Research Letters* 70.2, pp. 154–160.

Cupillard, P., E. Delavaud, G. Burgos, G. Festa, J.-P. Vilotte, Y. Capdeville, and J.-P. Montagner (2012). "RegSEM: A Versatile Code Based on the Spectral Element Method to Compute Seismic Wave Propagation at the Regional Scale". In: *Geophysical Journal International* 188.3, pp. 1203–1220.

Dahlen, F. A. and J. Tromp (1999). *Theoretical Global Seismology*. Princeton, NJ: Princeton University Press.

Douglas, A. (1967). *P-Signal Complexity and Source Radiation Patterns*. VESIAC Report 7885-1-x. University of Michigan.

Douglas, A., P. D Marshall, and D. J. Corbishley (1971). "Absorption and the Complexity of P Signals". In: *Nature Physical Science* 233, pp. 50–51.

Douglas, A., P. D Marshall, P. G. Gibbs, J. B. Young, and C. Blamey (1973). "P Signal Complexity Re-Examined". In: *Geophysical Journal of the Royal Astronomical Society* 33.2, pp. 195–221.

Douglas, A., J. A. Hudson, and B. J. Barley (1981). *Complexity of Short Period P Seismograms: What Does Scattering Contribute?* Technical Report AWRE O 3/81. Aldermaston, Berkshire: Atomic Weapons Research Establishment.

Douglas, A. (1966). *A Special Purpose Least Squares Program*. Technical Report AWRE O-54/66. London: Her Majesty's Stationary Office.

— (2013). *Forensic Seismology and Nuclear Test Bans*. Cambridge, UK: Cambridge University Press.

Douglas, C. C. and J. Mandel (1992). "An Abstract Theory for the Domain Reduction Method". In: *Computing* 48.1, pp. 73–96.

Dumbser, M. and M. Käser (2006). "An Arbitrary High-Order Discontinuous Galerkin Method for Elastic Waves on Unstructured Meshes - II. The Three-Dimensional Isotropic Case". In: *Geophysical Journal International* 167.1, pp. 319–336.

Dziewonski, A. M and B. Romanowicz (2007). "Deep Earth Seismology: An Introduction and Overview". In: *Treatise on Geophysics*. Second Edition. Vol. 1. Elsevier.

Dziewonski, A. M. and D. L. Anderson (1981). "Preliminary Reference Earth Model". In: *Physics of the Earth and Planetary Interiors* 25.4, pp. 297–356.

Dziewonski, A. M., V. Lekic, and B. A. Romanowicz (2010). "Mantle Anchor Structure: An Argument for Bottom up Tectonics". In: *Earth and Planetary Science Letters* 299.1, pp. 69–79.

Earthquake Hazards Program, U. S.G. S. (2017). "Finite Fault Database (ANSS ComCat)". In: *U.S. Geological Survey, Earthquake Hazards Program*.

Fink, M., C. Prada, F. Wu, and D. Cassereau (1989). "Self Focusing in Inhomogeneous Media with Time Reversal Acoustic Mirrors". In: *IEEE Ultrasonics Symposium*. Montreal, Que., Canada: IEEE, pp. 681–686.

Fokkema, J. T. and P. M. van den Berg (1993). *Seismic Applications of Acoustic Reciprocity*. Amsterdam: Elsevier Science Publishers.

Frankel, A. and W. Leith (1992). "Evaluation of Topographic Effects on P and S-Waves of Explosions at the Northern Novaya Zemlya Test Site Using 3-D Numerical Simulations". In: *Geophysical Research Letters* 19.18, pp. 1887–1890.

Friederich, W. and J. Dalkolmo (1995). "Complete Synthetic Seismograms for a Spherically Symmetric Earth by a Numerical Computation of the Green's Function in the Frequency Domain". In: *Geophysical Journal International* 122.2, pp. 537–550.

Gaebler, P., L. Ceranna, N. Nooshiri, A. Barth, S. Cesca, M. Frei, I. Grünberg, G. Hartmann, K. Koch, C. Pilger, J. O. Ross, and T. Dahm (2019). "A Multi-Technology Analysis of the 2017 North Korean Nuclear Test". In: *Solid Earth* 10.1, pp. 59–78.

Gaffet, S. and B. Massinon (1997). "Far-Field Seismogram Perturbations Induced by Topographic Heterogeneity". In: *Geophysical Research Letters* 24.24, pp. 3313–3316.

Galis, M., P. Moczo, and J. Kristek (2008). "A 3-D Hybrid Finite-Difference-Finite-Element Viscoelastic Modelling of Seismic Wave Motion". In: *Geophysical Journal International* 175.1, pp. 153–184.

Harrell, E. and D. Hoffman (2013). *Plutonium Mountain. Inside the 17-Year Mission to Secure a Dangerous Legacy of Soviet Nuclear Testing*. Report by Project on Managing the Atom. Cambridge, MA: Belfer Center for Science and International Affairs, John F. Kennedy School of Government, Harvard University.

Harris, R. A. et al. (2018). "A Suite of Exercises for Verifying Dynamic Earthquake Rupture Codes". In: *Seismological Research Letters* 89.3, pp. 1146–1162.

Helffrich, G. R. and B Wood (2001). "The Earth's Mantle". In: *Nature* 412, pp. 501–7.

Igel, H., T. Nissen-Meyer, and G. Jahnke (2002). "Wave Propagation in 3D Spherical Sections: Effects of Subduction Zones". In: *Physics of the Earth and Planetary Interiors* 132.1-3, pp. 219–234.

Imperatori, W. and P. M. Mai (2013). "Broad-Band near-Field Ground Motion Simulations in 3-Dimensional Scattering Media". In: *Geophysical Journal International* 192.2, pp. 725–744.

Ishii, M., P. M. Shearer, H. Houston, and J. E. Vidale (2005). "Extent, Duration and Speed of the 2004 Sumatra–Andaman Earthquake Imaged by the Hi-Net Array". In: *Nature* 435.7044, pp. 933–936.

Kawai, K., N. Takeuchi, and R. J. Geller (2006). "Complete Synthetic Seismograms up to 2 Hz for Transversely Isotropic Spherically Symmetric Media". In: *Geophysical Journal International* 164.2, pp. 411–424.

Kennett, B. L. N. and E. R. Engdahl (1991). "Traveltimes for Global Earthquake Location and Phase Identification". In: *Geophysical Journal International* 105.2, pp. 429–465.

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995). "Constraints on Seismic Velocities in the Earth from Traveltimes". In: *Geophysical Journal International* 122.1, pp. 108–124.

Komatitsch, D., Q. Liu, J. Tromp, P. Süss, C. Stidham, and J. H. Shaw (2004). "Simulations of Ground Motion in the Los Angeles Basin Based upon the Spectral-Element Method". In: *Bull. Seismol. Soc. Am.* 94.1, pp. 187–206.

Komatitsch, D. and J. Tromp (2002a). "Spectral-Element Simulations of Global Seismic Wave Propagation-I. Validation". In: *Geophysical Journal International* 149.2, pp. 390–412.

— (2002b). "Spectral-Element Simulations of Global Seismic Wave Propagation-II. Three-Dimensional Models, Oceans, Rotation and Self-Gravitation". In: *Geophysical Journal International* 150.1, pp. 303–318.

Kosloff, R. and D. Kosloff (1986). "Absorbing Boundaries for Wave Propagation Problems". In: *Journal of Computational Physics* 63.2, pp. 363–376.

Kristeková, M., J. Kristek, and P. Moczo (2009). "Time-Frequency Misfit and Goodness-of-Fit Criteria for Quantitative Comparison of Time Signals". In: *Geophysical Journal International* 178.2, pp. 813–825.

Kumagai, H., T. Saito, G. O'Brien, and T. Yamashina (2011). "Characterization of Scattered Seismic Wavefields Simulated in Heterogeneous Media with Topography". In: *Journal of Geophysical Research: Solid Earth* 116, B03308.

Lay, T. (1987). "Analysis of Near-Source Contributions to Early P-Wave Coda for Underground Explosions. 3. Inversion for Isotropic Statterers". In: *Bulletin of the Seismological Society of America* 77.5, pp. 1767–1783.

Lay, T. and T. Zhang (1992). "Near-Source Contributions to Teleseismic P Waves and P-Wave Coda for Underground Explosions". In: *Bulletin of the Seismological Society of America* 82.1, pp. 383–405.

Lee, S.-J., D. Komatitsch, B.-S. Huang, and J. Tromp (2009). "Effects of Topography on Seismic-Wave Propagation: An Example from Northern Taiwan". In: *Bulletin of the Seismological Society of America* 99.1, pp. 314–325.

Leith, W. (1998). *Degelen Nuclear Test and Tunnel Data*. Technical Memorandum DNA001-95-C-1079. Defense Threat Reduction Agency, US Geological Survey.

Leng, K., T. Nissen-Meyer, and M. van Driel (2016). "Efficient Global Wave Propagation Adapted to 3-D Structural Complexity: A Pseudospectral/Spectral-Element Approach". In: *Geophysical Journal International* 207.3, pp. 1700–1721.

Leng, K., T. Nissen-Meyer, M. van Driel, K. Hosseini, and D. Al-Attar (2019). "AxiSEM3D: Broad-Band Seismic Wavefields in 3-D Global Earth Models with Undulating Discontinuities". In: *Geophysical Journal International* 217.3, pp. 2125–2146.

Lilwall, R. C. and J. Farthing (1990). *Joint Epicentre Determination of Soviet Underground Nuclear Explosions 1973-89 at the Semipalatinsk Test Site*. Technical Report AWERE O 12/90. Aldermaston, Berkshire: Atomic Weapons Research Establishment.

Lynnes, C. S. and T. Lay (1988). "Observations of Teleseismic P Wave Coda for Underground Explosions". In: *Pure and Applied Geophysics* 128.1-2, pp. 231–249.

— (1989). "Inversion of P Coda for Isotropic Scatterers at the Yucca Flat Test Site". In: *Bulletin of the Seismological Society of America* 79.3, pp. 790–804.

Ma, S., R. J. Archuleta, and M. T. Page (2007). "Effects of Large-Scale Surface Topography on Ground Motions, as Demonstrated by a Study of the San Gabriel Mountains, Los Angeles, California". In: *Bulletin of the Seismological Society of America* 97.6, pp. 2066–2079.

Magnoni, F., E. Casarotti, A. Michelini, A. Piersanti, D. Komatitsch, D. Peter, and J. Tromp (2014). "Spectral-Element Simulations of Seismic Waves Generated by the 2009 L'Aquila Earthquake". In: *Bulletin of the Seismological Society of America* 104.1, pp. 73–94.

Marshall, P. D., T. C. Bache, and R. C. Lilwall (1984). *Body Wave Magnitudes and Locations of Soviet Underground Explosions at the Semipalatinsk Test Site*. Technical Report AWERE O 16/84. OCLC: 35427126. Aldermaston, Berkshire: Atomic Weapons Research Establishment.

Marshall, P. D., D. Porter, and P. Peachell (1992). *Analysis of Seismograms from Nuclear Explosions of Known Yield at Degelen Mountain and Konystan in East Kazakhstan, USSR*. Technical Report AWERE O 2/92. Aldermaston, Berkshire: Atomic Weapons Research Establishment.

Masson, Y. and B. Romanowicz (2017a). "Box Tomography: Localized Imaging of Remote Targets Buried in an Unknown Medium, a Step Forward for Understanding Key Structures in the Deep Earth". In: *Geophysical Journal International* 211.1, pp. 141–163.

— (2017b). "Fast Computation of Synthetic Seismograms within a Medium Containing Remote Localized Perturbations: A Numerical Solution to the Scattering Problem". In: *Geophysical Journal International* 208.2, pp. 674–692.

Masson, Y., P. Cupillard, Y. Capdeville, and B. Romanowicz (2014). "On the Numerical Implementation of Time-Reversal Mirrors for Tomographic Imaging". In: *Geophysical Journal International* 196.3, pp. 1580–1599.

McLaughlin, K. L. and R.-S. Jih (1988). "Scattering from Near-Source Topography: Teleseismic Observations and Numerical Simulations". In: *Bulletin of the Seismological Society of America* 78.4, pp. 1399–1414.

Moczo, P., E. Bystrick, J. Kristek, J. M. Carcione, and M. Bouchon (1997). "Hybrid Modeling of P-SV Seismic Motion at Inhomogeneous Viscoelastic Topographic Structures". In: *Bulletin of the Seismological Society of America* 87.5, pp. 1305–1323.

Moczo, P., J. Kristek, and M. Galis (2014). *The Finite-Difference Modelling of Earthquake Motions: Waves and Ruptures*. First Edition. Cambridge, UK: Cambridge University Press.

Monteiller, V., S. Chevrot, D. Komatitsch, and N. Fuji (2013). "A Hybrid Method to Compute Short-Period Synthetic Seismograms of Teleseismic Body Waves in a 3-D Regional Model". In: *Geophysical Journal International* 192.1, pp. 230–247.

Mueller, R. A. and J. R. Murphy (1971). "Seismic Characteristics of Underground Nuclear Detonations, Part I: Seismic Spectrum Scaling". In: *Bulletin of the Seismological Society of America* 61.6, pp. 1675–1692.

Murphy, J. R. (1977). *Seismic Coupling and Magnitude/Yield Relations for Underground Nuclear Detonations in Salt, Granite, Tuff/Rhyolite and Shale Emplacement Media*. Technical Report CSC-TR-77-0004. Computer Science Corporation.

Myers, S. C., J. L. Wagoner, and S. C. Larsen (2007). "Numerical Experiments Investigating the Source of Explosion S-Waves". In: *Proceedings of the 29th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*. Denver, CO: Los Alamos National Laboratory, pp. 175–184.

Nielsen, L., H. Thybo, A. Levander, and L. N. Solodilov (2003). "Origin of Upper-Mantle Seismic Scattering – Evidence from Russian Peaceful Nuclear Explosion Data". In: *Geophysical Journal International* 154.1, pp. 196–204.

Nilsson, S., N. A. Petersson, B. Sjogreen, and H.-O. Kreiss (2007). "Stable Difference Approximations for the Elastic Wave Equation in Second Order Formulation". In: *SIAM Journal on Numerical Analysis* 45, pp. 1902–1936.

Nissen-Meyer, T., M. van Driel, S. C. Stähler, K. Hosseini, S. Hempel, L. Auer, A. Colombi, and A. Fournier (2014). "AxiSEM: Broadband 3-D Seismic Wavefields in Axisymmetric Media". In: *Solid Earth* 5, pp. 425–445.

Nissen-Meyer, T., F. A. Dahlen, and A. Fournier (2007). "Spherical-Earth Fréchet Sensitivity Kernels". In: *Geophysical Journal International* 168.3, pp. 1051–1066.

Opršal, I., C. Matyska, and K. Irikura (2009). "The Source-Box Wave Propagation Hybrid Methods: General Formulation and Implementation". In: *Geophysical Journal International* 176.2, pp. 555–564.

Opršal, I. and J. Zahradnik (2002). "Three-Dimensional Finite Difference Method and Hybrid Modeling of Earthquake Ground Motion". In: *Journal of Geophysical Research* 107.B8, ESE2:1–16.

Peacock, S., A. Douglas, and D. Bowers (2017). "Joint Maximum-Likelihood Magnitudes of Presumed Underground Nuclear Test Explosions". In: *Geophysical Journal International* 210.2, pp. 621–644.

Petersson, N. A. and B. Sjogreen (2013). *User's Guide to WPP Version 2.2*. Livermore, CA: Lawrence Livermore National Laboratory.

Pitarka, A., D. V. Helmberger, and S. Ni (2007). "Analysis and Simulation of Three-Dimensional Scattering Due to Heterogeous Crustal Structure and Surface Topography on Regional Phases; Magnitude and Discrimination". In: *Proceedings of the 29th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*. Denver, CO: Los Alamos National Laboratory, pp. 205–213.

Pitarka, A., R. J. Mellors, W. R. Walter, S. Ezzedine, O. Vorobiev, T. Antoun, J. L. Wagoner, E. M. Matzel, S. R. Ford, A. J. Rodgers, L. Glenn, and M. Pasyanos (2015). "Analysis of Ground Motion from An Underground Chemical Explosion". In: *Bulletin of the Seismological Society of America* 105.5, pp. 2390–2410.

Ripperger, J., H. Igel, and J. Wasserman (2003). "Seismic Wave Simulation in the Presence of Real Volcano Topography". In: *Journal of Volcanology and Geothermal Research* 128.1-3, pp. 31–44.

Robertsson, J. O. A. and C. H. Chapman (2000). "An Efficient Method for Calculating Finite-difference Seismograms after Model Alterations". In: *GEOPHYSICS* 65.3, pp. 907–918.

Rodean, H. C. (1981). "Inelastic Processes in Seismic Wave Generation by Underground Explosions". In: *Identification of Seismic Sources – Earthquake or Underground Explosions*. Dordrecht, Holland: D. Reidel Publishing Company, pp. 97–189.

Rodgers, A. J., N. A. Petersson, and B. Sjogreen (2010). "Simulation of Topographic Effects on Seismic Waves from Shallow Explosions near the North Korean Nuclear Test Site with Emphasis on Shear Wave Generation". In: *Journal of Geophysical Research* 115, B11309.

Roecker, S., B. Baker, and J. McLaughlin (2010). "A Finite-Difference Algorithm for Full Waveform Teleseismic Tomography". In: *Geophysical Journal International* 181.2, pp. 1017–1040.

Rost, S. and C. Thomas (2002). "Array Seismology: Methods and Applications". In: *Reviews of Geophysics* 40.3, 2:1–27.

Sharrock, D. S., I. G. Main, and A. Douglas (1995). "A Two-Layer Attenuation Model for the Upper Mantle at Short Periods". In: *Geophysical Research Letters* 22.19, pp. 2561–2564.

Shearer, P. M. (2009). *Introduction to Seismology*. Second Edition. Cambridge: Cambridge University Press.

Simpson, D. W. and J. R. Cleary (1977). "P-Signal Complexity and Upper Mantle Structure". In: *Geophysical Journal International* 49.3, pp. 747–756.

Stacey, R. (1988). "Improved Transparent Boundary Formulations for the Elastic-Wave Equation". In: *Bulletin of the Seismological Society of America* 78.6, pp. 2089–2097.

Stead, R. J. and D. V. Helmberger (1988). "Numerical-Analytical Interfacing in Two Dimensions with Applications to Modeling NTS Seismograms". In: 128, p. 37.

Steinberger, B. and T. H. Torsvik (2012). "A Geodynamic Model of Plumes from the Margins of Large Low Shear Velocity Provinces". In: *Geochemistry, Geophysics, Geosystems* 13.1.

Stevens, J. L. and H. Xu (2008). "Wave Propagation from Complex 3D Sources Using the Representation Theorem". In: *Proceedings of the 2008 Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*. Portmouth, VA: Los Alamos National Laboratory, pp. 683–691.

Stuart, A. and K. Ord (1994). *Kendall's Advanced Theory of Statistics*. Sixth Edition. Vol. 1. Wiley-Blackwell.

Takemura, S., T. Furumura, and T. Maeda (2015). "Scattering of High-Frequency Seismic Waves Caused by Irregular Surface Topography and Small-Scale Velocity Inhomogeneity". In: *Geophysical Journal International* 201.1, pp. 459–474.

Taylor, S. R. and D. N. Anderson (2009). "Rediscovering Signal Complexity as a Teleseismic Discriminant". In: *Pure and Applied Geophysics* 166.3, pp. 325–337.

Team, T. S. (2019). *SeisSol Documentation*.

Tong, P., C.-w. Chen, D. Komatitsch, P. Basini, and Q. Liu (2014). "High-Resolution Seismic Array Imaging Based on an SEM-FK Hybrid Method". In: *Geophysical Journal International* 197.1, pp. 369–395.

Trabant, C., C. Thurber, and W. Leith (2002). "Ground Truth Seismic Events and Location Capability at Degelen Mountain, Kazakhstan". In: *Physics of the Earth and Planetary Interiors* 131, pp. 155–171.

van Driel, M., L. Krischer, S. C. Stähler, K. Hosseini, and T. Nissen-Meyer (2015). "Instaseis: Instant Global Seismograms Based on a Broadband Waveform Database". In: *Solid Earth* 6.2, pp. 701–717.

van Driel, M. (2015). "Efficient Methods in Global Seismic Wave Propagation". PhD thesis. ETH Zurich.

van Driel, M. and T. Nissen-Meyer (2014a). "Optimized Viscoelastic Wave Propagation for Weakly Dissipative Media". In: *Geophysical Journal International* 199.2, pp. 1078–1093.

— (2014b). "Seismic Wave Propagation in Fully Anisotropic Axisymmetric Media". In: *Geophysical Journal International* 199.2, pp. 880–893.

van Manen, D.-J., J. O. A. Robertsson, and A. Curtis (2007). "Exact Wave Field Simulation for Finite-Volume Scattering Problems". In: *The Journal of the Acoustical Society of America* 122.4, EL115–EL121.

Vidale, J. E. (1987). "Waveform Effects of a High Velocity, Subdacted Slab". In: *Geophysical Research Letters* 14.5, pp. 542–545.

Wang, N., Y. Shen, A. Flinders, and W. Zhang (2016). "Accurate Source Location from Waves Scattered by Surface Topography". In: *Journal of Geophysical Research: Solid Earth* 121.6, pp. 4538–4552.

Wen, L. and D. V. Helmberger (1998). "A Two-Dimensional P-SV Hybrid Method and Its Application to Modeling Localised Structures near the Core-Mantle Boundary". In: *Journal of Geophysical Research* 103.B8, pp. 17901–17918.

Willemsen, B., A. Malcolm, and W. Lewis (2016). "A Numerically Exact Local Solver Applied to Salt Boundary Inversion in Seismic Full-Waveform Inversion". In: *Geophysical Journal International* 204.3, pp. 1703–1720.

Wollherr, S., A.-A. Gabriel, and C. Uphoff (2018). "Off-Fault Plasticity in Three-Dimensional Dynamic Rupture Simulations Using a Modal Discontinuous Galerkin Method on Unstructured Meshes: Implementation, Verification and Application". In: *Geophysical Journal International* 214.3, pp. 1556–1584.

Wu, W., S. Ni, Z. Zhan, and S. Wei (2018). "An SEM-DSM Three-Dimensional Hybrid Method for Modelling Teleseismic Waves with Complicated Source-Side Structures". In: *Geophysical Journal International* 215.1, pp. 133–154.

Xie, X.-B. and T. Lay (1994). "The Excitation of Lg Waves by Explosions: A Finite-Difference Investigation". In: *Bulletin of the Seismological Society of America* 84.2, pp. 324–342.

Xu, H., A. Rodgers, I. Lomov, N. Petersson, B. Sjogreen, and O. Vorobiev (2013). "Simulation of Explosion Ground Motions Using a Hydrodynamic-to-Elastic Coupling Approach in Three Dimensions". In: *The Bulletin of the Seismological Society of America* 103, pp. 1629–1639.

Xu, Y., K. D. Koper, O. Sufri, L. Zhu, and A. R. Hutko (2009). "Rupture Imaging of the Mw 7.9 12 May 2008 Wenchuan Earthquake from Back Projection of Teleseismic P Waves". In: *Geochemistry, Geophysics, Geosystems* 10.4.

Yuan, S., N. Fuji, S. Singh, and D. Borisov (2017). "Localized Time-Lapse Elastic Waveform Inversion Using Wavefield Injection and Extrapolation: 2-D Parametric Studies". In: *Geophysical Journal International* 209.3, pp. 1699–1717.

Zhao, L., L. Wen, L. Chen, and T. Zheng (2008). "A Two-Dimensional Hybrid Method for Modeling Seismic Wave Propagation in Anisotropic Media". In: *Journal of Geophysical Research* 113, B12307.