

Comprehensive Study of Seismic Waveform Similarity: Applications to Reliable  
Identification of Repeating Earthquakes and Investigations of Detailed Source  
Process of Induced Seismicity

by

Dawei Gao

B.Eng., Central South University, 2012

M.Sc., University of Victoria, 2016

A Dissertation Submitted in Partial Fulfillment  
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the School of Earth and Ocean Sciences

©Dawei Gao, 2021  
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by  
photocopy or other means, without the permission of the author.

## **Supervisory Committee**

Comprehensive Study of Seismic Waveform Similarity: Applications to Reliable Identification of Repeating Earthquakes and Investigations of Detailed Source Process of Induced Seismicity

by

Dawei Gao

B.Eng., Central South University, 2012

M.Sc., University of Victoria, 2016

### **Supervisory Committee**

Dr. Honn Kao (School of Earth and Ocean Sciences)

**Co-Supervisor**

Dr. Stan Dosso (School of Earth and Ocean Sciences)

**Co-Supervisor**

Dr. Edwin Nissen (School of Earth and Ocean Sciences)

**Departmental Member**

Dr. Tom Gleeson (Department of Civil Engineering, School of Earth and Ocean Sciences)

**Departmental Member**

Dr. Magdalena Bazalova-Carter (Department of Physics and Astronomy)

**Outside Member**

## Abstract

### **Supervisory Committee**

Dr. Honn Kao (School of Earth and Ocean Sciences)

#### **Co-Supervisor**

Dr. Stan Dosso (School of Earth and Ocean Sciences)

#### **Co-Supervisor**

Dr. Edwin Nissen (School of Earth and Ocean Sciences)

#### **Departmental Member**

Dr. Tom Gleeson (Department of Civil Engineering, School of Earth and Ocean Sciences)

#### **Departmental Member**

Dr. Magdalena Bazalova-Carter (Department of Physics and Astronomy)

#### **Outside Member**

This Ph.D. dissertation focuses on a comprehensive study of seismic waveform similarity aiming at two themes: (1) reliable identification of repeating earthquakes (repeaters) and (2) investigation of the detailed source process of induced seismicity through the three-dimensional spatiotemporal evolution of mainly neighbouring earthquakes.

**Theme 1:** Reliable identification of repeaters.

Repeaters, occurring repeatedly on the same fault patch with nearly identical waveforms, are usually identified with the match-filtering (MF) method which essentially measures the degree of waveform similarity between an earthquake pair through the corresponding cross-correlation coefficient (CC). However, the performance of the MF method can be severely affected by the length of the cross-correlation window, the frequency band of the applied digital filter, and the presence of a large-amplitude wave train. To optimize the performance of MF, I first examine the effects of different operational parameters and determine generic rules for selecting the window length and the optimal frequency passband. To minimize the impact of a large-amplitude wave train, I then develop a new method, named the match-filtering with multisegment cross-correlation (MFMC) method. By equally incorporating the contributions from various segments of the waveforms, the new method is much more effective in capturing the minor waveform discrepancy between an event pair due to location difference and hence is more reliable in detecting potential repeaters and discriminating non-repeaters with large inter-event separation. With both synthetic and borehole array waveform data, I further reveal that waveform similarity is controlled by not only the inter-event separation but also many other factors, including station azimuth, epicentral distance, velocity structure, etc. Therefore, in

contrast to the traditional view, the results indicate that waveform similarity alone is insufficient to unambiguously identify true repeaters. For reliable repeater identification, we should rely on a physics-based approach considering both the overlapped source area and magnitude difference. Specifically, I define an event pair to be true repeaters if their inter-event separation is smaller than the rupture radius of the larger event and their magnitude difference is no more than 1. For the precise estimation of inter-event distance in cases of limited data, I develop the differential traveltime double-difference (DTDD) method which relies on the relative S-P differential traveltime. The findings of this study imply that previously identified repeaters and their interpretations/hypotheses potentially can be biased and hence may need a systematic reexamination.

**Theme 2:** Investigation of the detailed source process of induced seismicity.

Earthquakes induced by hydraulic fracturing (HF), especially those with large magnitudes, are often observed to have occurred near/after well completion. The delayed triggering of induced seismicity with respect to injection commencement poses serious challenges for risk mitigation and hazard assessment. By performing waveform cross-correlation and hierarchical clustering analysis, I reveal a high-resolution three-dimensional source migration process with mainshock delayed triggering that is probably controlled by local hydrogeological conditions. The results suggest that poroelastic effects might contribute to induced seismicity but are likely insufficient to activate a non-critically stressed fault of sufficient size. My analysis shows that the rapid pore-pressure build-up from HF can be very localized and capable of producing large, felt earthquakes on non-critically stressed fault segments. I further infer that the number of critically stressed, large intraplate faults should be very limited, and that reactivation of such faults may require sufficient pore-pressure accumulation. The findings of this study may also explain why so few fluid injections are seismogenic.

# Table of Contents

Supervisory Committee .....	ii
Abstract .....	iii
Table of Contents .....	v
List of Tables .....	vii
List of Figures .....	viii
Acknowledgments.....	x
<b>Chapter 1. Introduction.....</b>	<b>1</b>
1.1 What Are Repeating Earthquakes and Why Are They Important? .....	1
1.2 Conventional Ways to Identify Repeating Earthquakes and Their Drawbacks.....	3
1.3 Objectives of this study .....	6
1.4 Structure of This Dissertation.....	8
1.5 Information on Journal Articles Based on the Content of This Dissertation .....	8
1.5.1 Manuscript Corresponding to Chapter 2 .....	8
1.5.1.1 Article Citation.....	9
1.5.1.2 Author's Names and Affiliations .....	9
1.5.1.3 Author and Coauthor Contributions .....	9
1.5.2 Manuscript Corresponding to Chapter 3 .....	9
1.5.2.1 Article Citation.....	9
1.5.2.2 Author's Names and Affiliations .....	10
1.5.2.3 Author and Coauthor Contributions .....	10
1.5.3 Manuscript Corresponding to Chapter 4 .....	10
1.5.3.1 Article Citation.....	10
1.5.3.2 Author's Names and Affiliations .....	10
1.5.3.3 Author and Coauthor Contributions .....	11
<b>Chapter 2. Robust Detection of Potential Repeating Earthquakes .....</b>	<b>12</b>
2.1 Abstract.....	12
2.2 Introduction .....	12
2.3 Factors Affecting CC values.....	14
2.3.1 Window Length.....	15
2.3.2 Frequency Band of the Digital Filter.....	16
2.3.3 Large-amplitude Phase(s).....	19
2.4 Proposed Alternative Form of Cross-Correlation.....	22
2.4.1 Theory .....	22
2.4.2 Effect of Background Noise.....	23
2.5 Verification: Constraining Inter-event Separations Using Synthetic Data .....	26
2.5.1 Horizontal Inter-event Separations.....	29
2.5.2 Vertical Inter-event Separations.....	31
2.5.3 Importance of Station Distribution.....	33
2.6 Demonstration: A True Repeating Event or Not? .....	34
2.6.1 Real Examples from the Blanco Fracture Zone .....	34
2.6.2 Real Examples from the Queen Charlotte Plate Boundary .....	37
2.7 Conclusions .....	39
2.8 Acknowledgements .....	40
2.9 Supporting Information .....	41

<b>Chapter 3. Misconception of Waveform Similarity in the Identification of Repeating Earthquakes .....</b>	<b>57</b>
3.1 Abstract and Plain Language Summary .....	57
3.1.1 Abstract .....	57
3.1.2 Plain Language Summary .....	57
3.2 Introduction .....	58
3.3 Synthetic Experiments.....	59
3.3.1 Synthetic Experiment Setup .....	59
3.3.2 Constraining Inter-event Separation Using Single-channel Data.....	61
3.3.3 Constraining Inter-event Separation Using Single-station (3-channel) Data .....	63
3.3.4 Constraining Inter-event Separation Using Multi-station Data.....	64
3.4 Verification With Real Earthquake Examples.....	64
3.4.1 CC between Non-repeaters.....	65
3.4.2 CC between True Repeaters .....	68
3.5 A Physics-based Definition of Repeaters .....	69
3.6 Discussion.....	72
3.7 Conclusion.....	74
3.8 Acknowledgements .....	75
3.9 Supporting Information .....	76
3.9.1 CC Calculation.....	76
3.9.2 The Differential Traveltime Double-Difference (DTDD) Method.....	77
<b>Chapter 4. Complex 3D Migration and Delayed Triggering of Hydraulic Fracturing-Induced Seismicity .....</b>	<b>98</b>
4.1 Abstract and Plain Language Summary .....	98
4.1.1 Abstract .....	98
4.1.2 Plain Language Summary .....	98
4.2 Introduction .....	99
4.3 Methods.....	102
4.3.1 Waveform Cross-correlation and Hierarchical Clustering Analysis .....	102
4.3.2 Poroelastic Modeling and $\Delta$ CFS Calculation .....	103
4.4 Results .....	104
4.4.1 A High-resolution 3D Pattern of IIE Migration .....	104
4.4.2 Delayed Triggering Due to Pore Pressure Build-up.....	107
4.5 Interpretation and Implications.....	108
4.5.1 Reactivation of A Non-critically Stressed Fault Segment.....	108
4.5.2 Current Stress State of Crustal Faults .....	109
4.5.3 Delayed Triggering of IIE.....	110
4.5.4 Seismogenic vs. Aseismogenic Injection Operations.....	110
4.6 Discussion and Conclusions .....	111
4.7 Acknowledgments .....	112
4.8 Supporting Information .....	112
4.8.1 Technical Details of CC Calculation.....	112
4.8.2 Technical Details of Poroelastic Modeling .....	113
<b>Chapter 5. Conclusions.....</b>	<b>119</b>
<b>Bibliography .....</b>	<b>122</b>

## List of Tables

<b>Table 2.1.</b> Digital filters commonly used in waveform match-filtering.....	17
<b>Table 3.1.</b> A compiled list of different criteria in identifying repeaters.....	91
<b>Table 3.2.</b> A list of digital filters commonly used in identifying repeaters.....	96
<b>Table 3.3.</b> A list of commonly assumed $\Delta\sigma$ in estimating the ERR .....	97
<b>Table 4.1.</b> Solid and fluid properties of each layer used in the model. ....	116
<b>Table 4.2.</b> Injection data. ....	117

## List of Figures

<b>Figure 1.1.</b> Characteristics of repeaters on a creeping fault.....	2
<b>Figure 1.2.</b> Waveform examples of the first two foreshocks of the 1999 Mw 7.6 Izmit earthquake ..	5
<b>Figure 2.1.</b> Effect of band-pass filtering on the result of waveform cross-correlation (CC)..	18
<b>Figure 2.2.</b> Synthetic experiment to test the performance of the match-filtering (MF) method.....	21
<b>Figure 2.3.</b> A flow chart to illustrate the steps of match-filtering with multi-segment cross-correlation (MFMC).....	23
<b>Figure 2.4.</b> A test on the effect of background noise..	24
<b>Figure 2.5.</b> Band-pass filtered (1–10 Hz) and normalized synthetic seismograms of two template events at different depths..	27
<b>Figure 2.6.</b> CC variation due to horizontal inter-event separation with a template starting from the P phase.....	28
<b>Figure 2.7.</b> CC variation due to vertical inter-event separation with a template starting from the P phase. ....	32
<b>Figure 2.8.</b> Normalized waveforms of three events occurred in the Blanco Fracture Zone..	36
<b>Figure 2.9.</b> Normalized waveforms of two event pairs occurred in the Queen Charlotte plate boundary zone.....	38
<b>Figure 2.10.</b> Effect of high-pass (a) and low-pass (b) filtering on the result of waveform cross-correlation.....	41
<b>Figure 2.11.</b> CC variation due to horizontal inter-event separation with a template starting from the S phase.....	42
<b>Figure 2.12.</b> CC variation due to horizontal inter-event separation with a template starting from the P phase.....	43
<b>Figure 2.13.</b> CC variation due to horizontal inter-event separation with a template starting from the S phase.....	44
<b>Figure 2.14.</b> CC variation due to vertical inter-event separation with a template starting from the S phase.....	45
<b>Figure 2.15.</b> CC variation due to vertical inter-event separation with a template starting from the P phase.....	46
<b>Figure 2.16.</b> CC variation due to vertical inter-event separation with a template starting from the S phase.....	47
<b>Figure 2.17.</b> Experiment configuration of the additional tests presented in Section 2.5.3. .	48
<b>Figure 2.18.</b> Experiment results showing the importance of station distribution .....	49
<b>Figure 2.19.</b> CC variation due to horizontal inter-event separation with a template at a depth of 3 km.....	50
<b>Figure 2.20.</b> CC variation due to horizontal inter-event separation with a template at a depth of 10 km.....	51
<b>Figure 2.21.</b> CC variation due to vertical inter-event separation with a template at a depth of 3 km .....	52
<b>Figure 2.22.</b> CC variation due to vertical inter-event separation with a template at a depth of 10 km .....	53

<b>Figure 2.23.</b> Map showing the tectonic setting of the Blanco Fracture Zone (BFZ) .....	54
<b>Figure 2.24.</b> Amplitude spectra of the three BFZ events shown in Figure 2.23 and discussed in Section 2.6.1. ....	55
<b>Figure 2.25.</b> Effect of different filter passbands on the cross-correlation results of BFZ event pairs. ....	56
<b>Figure 3.1.</b> Schematic diagram showing the setup of the synthetic experiments. ....	60
<b>Figure 3.2.</b> Results of our synthetic experiment showing CC variation as a function of horizontal (a and c) and vertical (b and d) inter-event separation.. ....	61
<b>Figure 3.3.</b> CC test results with real earthquake data.....	67
<b>Figure 3.4.</b> A physics-based approach to distinguish repeating and neighbouring events..	71
<b>Figure 3.5.</b> Synthetic waveforms of both template and matched events for a horizontal separation of -2.6 km. ....	79
<b>Figure 3.6.</b> CC variation (obtained with single-channel data) due to inter-event separations .....	80
<b>Figure 3.7.</b> CC variation (obtained with single-station, 3-channel data) due to horizontal inter-event separations. ....	81
<b>Figure 3.8.</b> CC variation (obtained with single-station, 3-channel data) due to vertical inter-event separations. ....	82
<b>Figure 3.9.</b> Effects of filtering on the CC value between two events that have been verified to be non-repeaters (events No. 1 and No. 2, Figure 3.3) with single-channel data.....	83
<b>Figure 3.10.</b> Effects of Twin and filtering on the CC value between two events that have been verified to be non-repeaters (events No. 1 and No. 2, Figure 3.3) with single-station (3-channel) data.....	84
<b>Figure 3.11.</b> An example of how filtering increases waveform similarity at a close station SMNB .....	85
<b>Figure 3.12.</b> An example of how filtering increases waveform similarity at a distant station FROB .....	86
<b>Figure 3.13.</b> Effects of Twin and filtering on the CC value between two events that have been verified to be true repeaters (events No. 1 and No. 3, Figure 3.3) with single-station (3-channel) data.....	87
<b>Figure 3.14.</b> Examples of normalized unfiltered waveforms of two events that have been verified to be true repeaters (events No. 1 and No. 3, Figure 3.3).....	88
<b>Figure 3.15.</b> Zoom-in map showing the epicenters of the 3 events in Figure 3.4.....	89
<b>Figure 3.16.</b> Relationship between earthquake magnitude (M) and ERR. ....	90
<b>Figure 4.1.</b> Comparison of induced seismicity and HF activity.....	101
<b>Figure 4.2.</b> Identification of near-identical earthquakes .....	103
<b>Figure 4.3.</b> East-west cross section of the 2015 Mw 3.9 earthquake sequence .....	105
<b>Figure 4.4.</b> Poroelastic modelling results of (a) Mw 3.2 event and (b) Mw 3.9 event .....	108
<b>Figure 4.5.</b> Histogram of the 49 poorly correlated small earthquakes. ....	115
<b>Figure 4.6.</b> $\Delta P$ contribution in determining the $\Delta CFS$ for the Mw 3.9 event. ....	115

## Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor, **Honn Kao**, for his invaluable guidance, constant support, and kind encouragement. His good nature, wealth of knowledge, and passion for science always motivated me to improve both myself and my work. I also thank him for reviewing and editing drafts and offering comments during the writing of this dissertation and associated scientific journal papers.

I would like to thank my co-supervisor, **Stan Dosso**, for editing the drafts, providing constructive comments and helpful suggestions, and his kind encouragement.

Besides my supervisors, I would like to thank my committee members, **Edwin Nissen**, **Tom Gleeson**, and **Magdalena Bazalova-Carter** for their stimulating advice, kind encouragement, helpful comments, and devoting their time to guiding my research.

I would like to thank all those who have contributed to this dissertation research. **Bei Wang** performed the poroelastic modeling and Coulomb stress change calculation for the 2015 Fox Creek Mw 3.9 earthquake sequence and helped me write that part of the dissertation. **Ryan Visser** collected the injection data used in this study. **Ryan Schultz** and **Rebecca Harrington** provided helpful comments and edits in Chapter 4. **Jiangheng He** helped me resolve the IT issues throughout my research. I also thank **Lupei Zhu** for providing the FK code that is used in generating synthetic seismograms, **Didem Cambaz** for helping collect seismic waveform data, **Jianlong Yuan**, **Fengzhou Tan**, **Ramin Mohammad Hosseini Dokht**, **Douglas A. Dodge**, **Jean-Luc Got**, **Roy Hyndman**, **Toshihiro Igarashi**, **Naoki Uchida**, **David P. Schaff**, **Makoto Naoi**, **Kelin Wang**, **Rachel E. Abercrombie**, **Wen-che Yu**, **Lingsen Meng**, **Hui Huang**, **Jean Schmittbuhl**, **Marco Bohnhoff**, **Christopher Wollin**, **Tomoaki Nishikawa**, **Emily Warren-Smith**, **Luis A. Dominguez**, **Tianhaozhe Sun**, **Taizi Huang**, and **Haipeng Luo** for thoughtful discussions.

I would also like to thank **Shao-Ju Shan** for kind encouragement, and thank my friends, fellow graduate students and colleagues, **Hongyu Yu**, **Ayodeji Kuponiyi**, **Yijie Zhu**, **Jesse Hutchinson**, **Adebayo Oluwaseun Ojo**, **Carlos Herrera**, and **Chet Goerzen** for their help and encouragement.

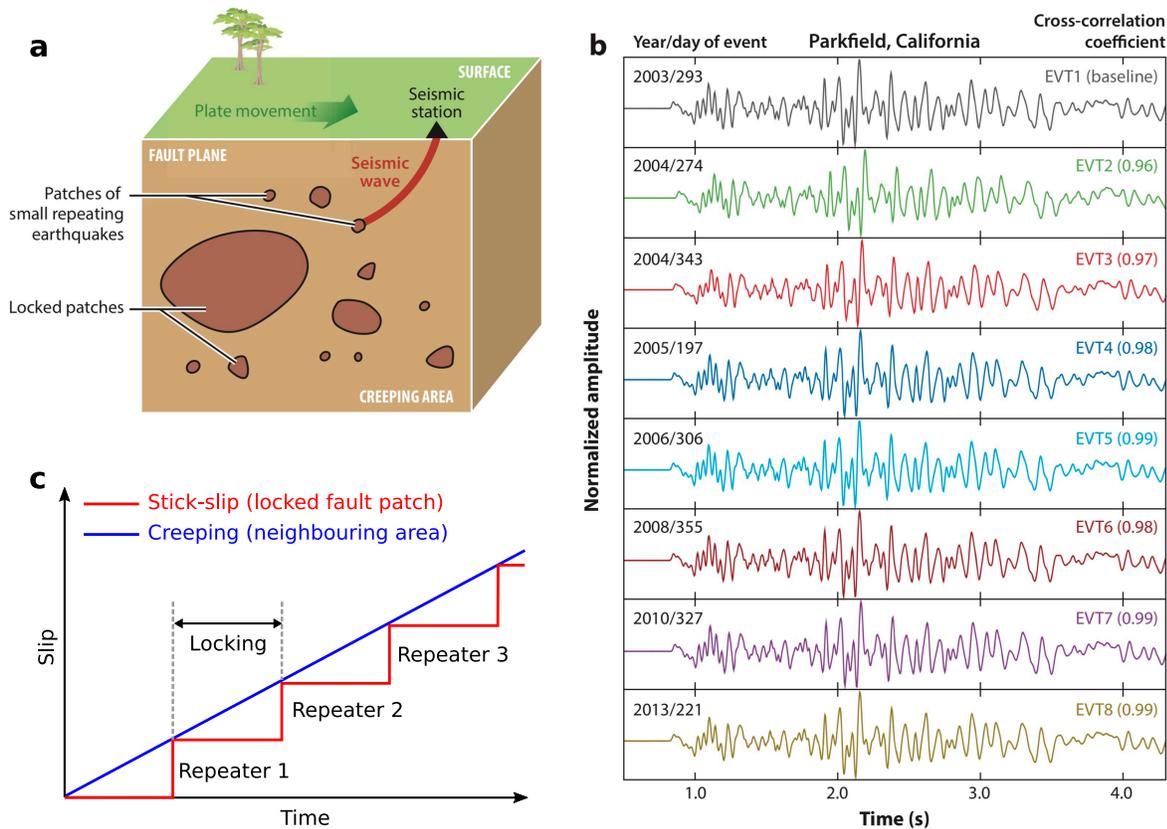
This dissertation is dedicated to my family, **Zhian Yu**, **Danghuai Gao**, **Dapeng Gao**, **Shiqiang Feng**, **Yinbin Feng**, and **Shangqian Gao**, for their love, encouragement, and support.

# Chapter 1. Introduction

## 1.1 What Are Repeating Earthquakes and Why Are They Important?

Repeating earthquakes (repeaters) are events that repeatedly rupture the same fault patch at different times with the same focal mechanisms and nearly identical seismic waveforms (Figure 1). The occurrences of such events are commonly interpreted to be the recurring seismic energy release of a locked seismogenic fault patch loaded by surrounding aseismic slip, prominent on creeping faults (Uchida and Bürgmann, 2019). Figure 1a shows a schematic fault model that includes a number of locked patches embedded on a creeping fault. As the fault overall is creeping (Figure 1.1a), the locked patches are forced to catch up with the neighbouring creeping motion in a stick-slip manner (Figure 1.1c), radiating nearly identical seismic signals (Figure 1.1b) at roughly periodic intervals (Figure 1.1c). In the long term, the fault patch that hosts the repeaters (red line in Figure 1.1c) undergoes almost the same cumulative slip as the neighbouring area (blue line in Figure 1.1c) (Uchida, 2019).

As the repeaters are believed to be caused by fault creeping (Uchida, 2019; Uchida and Bürgmann, 2019), they are important diagnostic indicators of fault behavior (i.e., creeping, partially creeping, or fully locked) which is of great significance in characterizing the potential rupture scenarios of future large earthquakes and the associated seismic hazards (e.g., Bohnhoff et al., 2017; Hayward and Bostock, 2017). Moreover, repeaters are free creepmeters (Uchida, 2019) which can provide quantitative estimations of in situ slip rates at seismogenic depths (e.g., Li et al., 2007, 2011). Quantitatively, the slip rate is the cumulative amount of fault slip, which can be derived from the seismic moment of a repeating cluster, divided by the total duration of the cluster. Repeaters are extremely valuable for better understanding of the deeper part of the strike-slip faults and off-shore near-trench areas of subduction faults where geodetic data are unavailable and/or the geodetic data resolution is very low (Uchida, 2019).



**Figure 1.1.** Characteristics of repeaters on a creeping fault. (a) Schematic fault model of small locked seismic patches loaded by surrounding aseismic slip (Uchida and Bürgmann, 2019). (b) An example of the recorded waveforms (at a common station) of a group of repeaters near Parkfield, California (Kim et al., 2016; Uchida and Bürgmann, 2019). (c) Schematic illustration of the slip history of a locked fault patch and its neighbouring creeping area (modified from Uchida, 2019).

Because of the same origins of repeaters, their seismic waveforms carry a wealth of information about many aspects of geophysics (Uchida and Bürgmann, 2019). Specifically, the subtle waveform change between a group of repeaters is commonly regarded as the manifestation of the change in velocity structure (e.g., Poupinet et al., 1984; Schaff and Beroza, 2004; Sawazaki et al., 2015; Pacheco et al., 2017) or temperature (Wu et al., 2020) along the ray path. If the repeaters are large enough and their seismic energies have traversed the inner core, the differential traveltimes between the BC and DF branches of PKP waves can be used to quantitatively constrain the inner core rotation (e.g., Li and Richards, 2003; Zhang et al., 2005, 2008; Tkalčić et al., 2013). Since repeaters essentially occur at the same location, their apparent

inter-event separations computed from a given catalog can be used to directly evaluate the precision of earthquake locations in that catalog (e.g., Li and Richards, 2003; Meier et al., 2004; Schaff and Richards, 2011; Jiang et al., 2014). Last, but not the least, observations of repeaters before some large earthquakes (e.g., Kato et al., 2012; Kato and Nakagawa, 2014; Meng et al., 2015; Huang and Meng, 2018) are interpreted as the manifestation of an aseismic triggering process. Such knowledge is crucial to accurate prediction, early warning, and mitigation of natural hazards. Given the extensive applications of repeaters, the question of how to reliably identify them is of fundamental importance for many geophysical studies.

## 1.2 Conventional Ways to Identify Repeating Earthquakes and Their Drawbacks

There are three ways of identifying repeating earthquakes. The most straightforward way is based on precise determination of the overlap of the source areas (Waldhauser and Ellsworth, 2002). However, this approach requires the establishment of a very dense array in the source area, which limits its applicability, especially of past events in early times. The second, and the most popular way relies on the match-filtering (MF) method which essentially measures the degree of waveform similarity between an earthquake pair through the corresponding cross-correlation coefficient (CC) (e.g., Nadeau and McEvilly, 1999; Igarashi, 2010). The mathematical definition of CC is given by:

$$CC = \frac{\sum_{i=1}^n [a(i) - \bar{a}][b(i) - \bar{b}]}{\sqrt{\sum_{i=1}^n [a(i) - \bar{a}]^2 \times \sum_{i=1}^n [b(i) - \bar{b}]^2}} \quad (1-1)$$

where  $a$  and  $b$  correspond to the discrete time-series of two events,  $\bar{a}$  and  $\bar{b}$  are the mean values of each time-series, and  $n$  is the total number of samples in the cross-correlation window. A higher CC represents a higher degree of waveform similarity and a value of 1 means two signals are perfectly matched (i.e., identical waveforms). Once the computed CC exceeds a prescribed threshold, a repeater is declared (e.g., Nadeau and McEvilly, 1999; Igarashi et al., 2003; Uchida et al., 2003, 2010; Meier et al., 2004; Schaff and Richards, 2004, 2011; Green and Neuberg, 2006; Petersen, 2007; Yamashita et al., 2012; Kato and Nakagawa, 2014; Meng et al., 2015; Warren-Smith et al., 2018; Chaves et al., 2020; Hatch et al., 2020). The third way is a hybrid method with complementary criteria in addition to CC. The additional criteria vary significantly

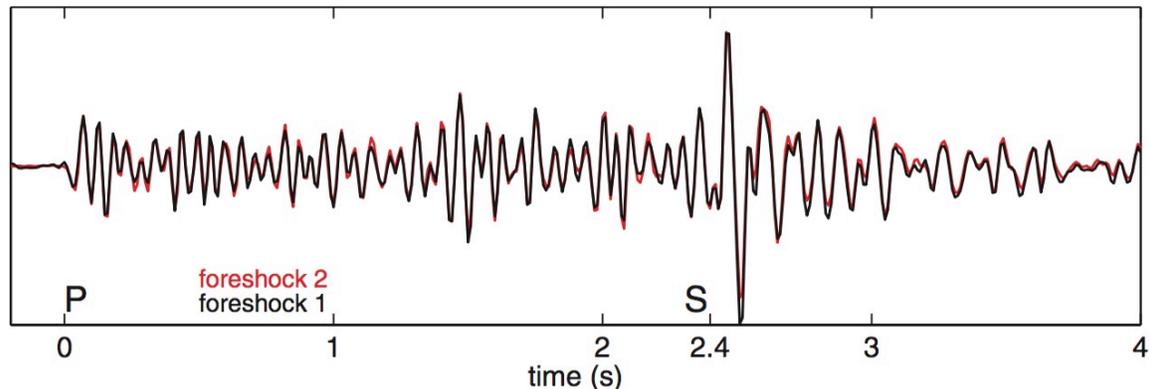
among different studies. Some studies place constraints on the S-P differential times (e.g., Rau et al., 2007; Chen et al., 2008), the origin time difference (e.g., Bohnhoff et al., 2017), and/or magnitude difference (e.g., Huang and Meng, 2018; Nishikawa and Ide, 2018) between an event pair to minimize misidentification, whereas other studies directly verify the repeaters by examining their overlapped source areas (e.g., Cociani et al., 2010; Naoi et al., 2015; Cauchie et al., 2020).

The application of waveform similarity (i.e., CC) in repeater identification is extremely popular in the community mainly because of two reasons. First, the implementation of the MF method is easy, quick, and convenient. Second, and more importantly, there is a greatly reduced data requirement as the MF method can work well even with single-station/channel data (e.g., Li and Richards, 2003; Schaff and Richards, 2004, 2011; Li et al., 2007, 2011; Zhang et al., 2008; Cannata et al., 2013; Yamada et al., 2016; Salvagea and Neuberg, 2016). Although the seismic waveform data constraints no longer exist due to the rapid increase of the number of seismograph stations in recently years, many recent studies simply stick with the MF method (e.g., Kato and Nakagawa, 2014; Warren-Smith et al., 2018; Chaves et al., 2020; Hatch et al., 2020).

However, there are at least two key problems if the repeater identification process is solely based on waveform similarity. The first one is the inconsistent, and sometimes inaccurate, performance of the MF method as CC can be affected by the cross-correlation window length, the bandwidth of seismic signals (e.g., Harris, 1991; Schaff, 2008; Baisch et al., 2008; Uchida, 2019), and especially the presence of a large amplitude wave train (often the shear wave and surface waves) (e.g., Buurman & West, 2010; Calderoni et al., 2015; Li et al., 2017; Myhill et al., 2011; Schmittbuhl et al., 2016). If a large amplitude wave train is present in the waveforms, the computed CC is very likely to lose the resolution in discriminating non-repeaters with very large hypocentre distances.

The second reason, that is probably more troublesome, is the lack of resolution between repeating and neighbouring events based on waveform similarity alone. The fundamental difference between repeaters and neighbouring events is that the former have significantly overlapped source areas while the latter have little or no overlap. Given the proximity of both kinds of events, they can all be characterized by highly similar seismic waveforms.

Consequently, it is usually difficult to differentiate them from each other. If misclassification happens, incorrect interpretations/hypotheses and/or controversy may be incurred.



**Figure 1.2.** Waveform examples of the first two foreshocks of the 1999 Mw 7.6 Izmit earthquake (adopted from Bouchon et al. (2011)). Notice that the noise level is extremely low as indicated by the nearly flat waveforms before the P arrival.

One famous controversial example is the foreshock sequence of the 17 August 1999, Mw 7.6 Izmit, Turkey, earthquake which ruptured the western portion of the strike-slip North Anatolian Fault Zone (Ozalaybey et al., 2002). Few foreshock sequences of large earthquakes have been well recorded by local seismic stations in the past, yet this sequence is apparently an exception as a number of foreshocks were captured by 10 local stations in the epicentral distance range of 10 – 100 km (Ellsworth and Bulut, 2018). Based on the near-identical foreshock waveforms (Figure 1.2), Bouchon et al. (2011) conclude that the Izmit earthquake was triggered by repeated seismic ruptures near the hypocentre of the mainshock (i.e., the so-called preslip model proposed by Ellsworth and Beroza, 1995). This implies an aseismic triggering process and the size of the eventual large earthquake may be predicted (Ellsworth and Beroza, 1995; Gomberg, 2018). With double-difference relocation (Waldhauser and Ellsworth, 2000) and examination of the waveform discrepancies between foreshocks, however, Ellsworth and Bulut (2018) argue that the foreshocks are effectively neighbouring events and the observations can be best explained by the cascade triggering process with small foreshocks migrating from west to east, finally leading to

the mainshock. The cascade triggering process simply means that the occurrence of a large earthquake is a random outcome of interaction between small earthquakes through stress perturbation and cannot be predicted (Ellsworth and Beroza, 1995; Gomberg, 2018). Without additional evidence, this controversy has not been resolved conclusively. Even for such a well-recorded earthquake sequence, two different research groups reach totally opposite conclusions (i.e., repeaters by Bouchon et al. (2011) vs. neighbouring events by Ellsworth and Bulut (2018)) and different models of the mainshock initiation process, highlighting the challenge and significance in reliably discriminating repeaters and neighbouring events. Therefore, improved understanding of waveform similarity and the factors controlling its behaviour will help recognize and differentiate such events and probe the geophysical processes behind them.

### **1.3 Objectives of This Study**

The research reported in this Ph.D. dissertation is designed to conduct a comprehensive study of seismic waveform similarity on two themes: (1) reliable identification of repeating earthquakes, and (2) investigation of the detailed source process of induced seismicity through the spatiotemporal distribution of mainly neighbouring earthquakes.

With respect to the first theme, I first examine the effects of different factors, including the length of the cross-correlation window, the frequency band of the applied digital filter, and the presence of a large-amplitude wave train, on the performance of the MF method. To optimize the MF, I determine generic rules for selecting the window length and the optimal frequency passband. Then, to minimize the influence of a large-amplitude wave train, I develop a new method, named the match-filtering with multisegment cross-correlation (MFMC) method, which equally incorporates the contributions from various segments of the waveforms. Compared with the conventional MF, the new method is much more effective in recognizing the hypocentre location difference between an event pair and hence is more reliable in detecting potential repeating earthquakes and discriminating non-repeaters with large hypocentre distance. It should be noted that (1) the MFMC method can work very well with limited data (e.g., single-station/channel data), and (2) the value of MFMC technique is to effectively decrease the likelihood of the misidentification although it cannot totally eliminate the chance of misidentification. Therefore, the MFMC is a very powerful tool for quick and direct potential

repeater scanning or it can be used as an efficient post-analysis tool to extract potential repeaters from an existing catalog or from the output of the conventional MF. However, the detected events still need to be verified by other means (e.g., Cheng et al., 2007; Li et al., 2007, 2011).

To fully understand the spatial resolution of waveform similarity, I systematically study the relationship between CC and inter-event separation and uncover the overlooked factors seriously affecting CC through a large number of synthetic experiments in which I consider different station azimuths, focal mechanisms, source depths, epicentral distances, the choices of data (one specific channel or all three channels), and orientations of the source separation (horizontal vs. vertical). With near-field, high-quality real earthquake waveforms recorded by a dense local borehole array in Parkfield, California, I illustrate that waveform similarity alone is insufficient to reliably identify true repeaters from the earthquakes with closely spaced source areas. For reliable repeater identification, the final decision should depend on both the overlapped source areas (e.g., Waldhauser and Ellsworth, 2002) and magnitude difference (e.g., Nishikawa and Ide, 2018). In addition, the earthquake magnitude should be comparable with the loading rate and the time interval between earthquakes if such information is available. For the precise estimation of inter-event separation in cases with limited data, I develop the differential traveltime double-difference (DTDD) method which relies on the relative S-P differential traveltime. I validate the DTDD method using earthquakes that occurred in the Fox Creek area, Alberta, Canada. Note, however, that precise estimation of the inter-event separations and source areas for a very large seismic dataset can be very labor-intensive and time-consuming. In such cases, using the newly developed MFMC technique for preliminary repeater scanning can significantly lower the workload of final confirmation. The significance of this work is that the results basically reject the idea of identifying repeaters with 100% confidence using waveform similarity alone, implying that previously identified repeaters and associated interpretations/hypotheses may be unreliable and hence may need a systematic reevaluation. Insights from this study will contribute broadly to not only research based on repeating earthquakes but also other waveform-similarity-based studies.

With respect to the second theme, the key question to answer is the detailed source process of induced seismicity. Fluid injection-induced earthquakes (IIE), especially the mainshocks, are often observed to occur near or after well completion, such as the 2018  $M_L$  4.5 Dawson Creek earthquake in British Columbia (Peña Castro et al., 2020) and 2019  $M_L$  4.18 Red Deer event in

Alberta (Schultz and Wang, 2020). The delayed triggering relative to injection commencement poses serious challenges for seismic hazard mitigation. In this dissertation, I re-examine the source process of the 2015 Mw 3.9 earthquake sequence near Fox Creek, Alberta, Canada. This is the first well-known delayed case with the mainshock occurring about 2 weeks after the hydraulic fracturing (HF) stimulation was completed (Bao and Eaton, 2016). In this particular study, I take the waveform-similarity-based approach to mainly identify neighbouring earthquakes (referred to as near-identical events) occurring on the same/similar fault structures instead of repeaters with overlapped source areas. Then I utilize such events to delineate the corresponding structure of the seismogenic fault, investigate the three-dimensional spatiotemporal evolution of these near-identical events, and uncover the cause of delayed triggering. These findings can deepen our understanding of the current stress state of crustal faults in the source area, and also explain why so few injection operations are seismogenic. Insights from this study not only contribute to seismic hazard mitigation from IIE but also provide economic benefits for HF operations.

## **1.4 Structure of This Dissertation**

Apart from the Introduction (Chapter 1) and Conclusions (Chapter 5), this dissertation contains three separate manuscripts (details are given in Section 1.5). Chapters 2 and 3 address the objective of reliable repeater identification. Chapter 2 is primarily based on a paper published in *Journal of Geophysical Research: Solid Earth* in 2020 and mainly presents a new technique for robustly detecting potential repeating earthquakes and discriminating non-repeating events with large inter-event distance. Chapter 3 represents a submitted manuscript which investigates whether waveform similarity alone is a sufficient proxy for identifying repeating earthquakes. Chapter 4 represents another submitted manuscript focusing on the detailed source process of fluid IIE in the Fox Creek area using neighbouring earthquakes. Chapter 5 summarizes the main findings of this dissertation and provides suggestions for future research.

## **1.5 Information on Journal Articles Based on the Content of This Dissertation**

### **1.5.1 Manuscript Corresponding to Chapter 2**

The main body of Chapter 2 primarily consists of a published journal article (Gao and Kao, 2020) which aims at optimizing the performance of the match-filtering method to more reliably detect potential repeating earthquakes. The article information is given below, while Sections 2.1-2.8 present the article itself. Supporting information accompanied with the published article is presented in Section 2.9.

#### **1.5.1.1 Article Citation**

Gao, D., & Kao, H. (2020). Optimization of the Match-Filtering Method for Robust Repeating Earthquake Detection: The Multisegment Cross-Correlation Approach. *Journal of Geophysical Research: Solid Earth*, 125(7), e2020JB019714.

#### **1.5.1.2 Author's Names and Affiliations**

Dawei Gao<sup>1,2</sup> and Honn Kao<sup>1,2</sup>

<sup>1</sup>School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada

<sup>2</sup>Pacific Geoscience Centre, Geological Survey of Canada, Sidney, British Columbia, Canada

#### **1.5.1.3 Author and Coauthor Contributions**

The author of this dissertation, D.G., generated the synthetic seismograms, carried out the cross-correlation analysis with both synthetic and real earthquake waveforms, and initiated the manuscript. Coauthor H.K. designed the study, initiated the concept of multi-segment cross-correlation, reviewed and edited drafts, and offered comments during the writing of this manuscript.

### **1.5.2 Manuscript Corresponding to Chapter 3**

The main body of Chapter 3 primarily consists of a submitted manuscript (Gao and Kao, 2021, submitted) which examines whether seismic waveform similarity alone is sufficient to reliably identify true repeaters. The article information is given below. Sections (3.1-3.8) present the article itself. Supporting information accompanied with the submitted article is presented in Section 3.9.

#### **1.5.2.1 Article Citation**

Gao, D., & Kao, H. (2021). Misconception of Waveform Similarity in the Identification of Repeating Earthquakes. *Geophysical Research Letters*, under revision.

### **1.5.2.2 Author's Names and Affiliations**

Dawei Gao<sup>1,2</sup> and Honn Kao<sup>1,2</sup>

<sup>1</sup>School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada

<sup>2</sup>Pacific Geoscience Centre, Geological Survey of Canada, Sidney, British Columbia, Canada

### **1.5.2.3 Author and Coauthor Contributions**

The author of this dissertation, D.G., carried out the waveform cross-correlation analysis with both synthetic and real earthquake waveforms, developed the DTDD method, and initiated the manuscript. Coauthor H.K. designed the study, reviewed and edited drafts, and offered comments during the writing of this manuscript.

## **1.5.3 Manuscript Corresponding to Chapter 4**

The main body of Chapter 4 consists of a submitted manuscript (Gao et al., 2021, submitted) which investigates the delayed triggering of large induced events through the three-dimensional spatiotemporal evolution of mainly neighbouring earthquakes. The article information is given below. Subsequent sections (4.1- 4.7) present the article itself. Supporting information accompanied with the article is presented in Section 4.8.

### **1.5.3.1 Article Citation**

Gao, D., Kao, H., Wang, B., Visser, R., Schultz, R., & Harrington, R. (2020). Complex 3D Migration and Delayed Triggering of Hydraulic Fracturing-Induced Seismicity. *Geophysical Research Letters*, submitted.

### **1.5.3.2 Author's Names and Affiliations**

Dawei Gao<sup>1,2</sup>, Honn Kao<sup>1,2</sup>, Bei Wang<sup>1,2</sup>, Ryan Visser<sup>2</sup>, Ryan Schultz<sup>3</sup>, Rebecca Harrington<sup>4</sup>

<sup>1</sup>School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada.

<sup>2</sup>Pacific Geoscience Centre, Geological Survey of Canada, Sydney, British Columbia, Canada.

<sup>3</sup>Department of Geophysics, Stanford University, Stanford, CA, USA.

<sup>4</sup>Ruhr University Bochum, Institute of Geology, Mineralogy, and Geophysics, Bochum, Germany.

### **1.5.3.3 Author and Coauthor Contributions**

The author of this dissertation, D.G. carried out the waveform cross-correlation and hierarchical clustering analysis and initiated the manuscript. Coauthor H.K. designed the study and initiated the interpretation. Coauthor B.W. performed poroelastic modeling and Coulomb stress change ( $\Delta CFS$ ) calculation. Coauthor R.V. contributed to the collection of injection data. All authors contributed to the writing and reviewing of the manuscript.

## Chapter 2. Robust Detection of Potential Repeating Earthquakes

### 2.1 Abstract

Waveform match-filtering (MF), based on cross-correlation between the recorded waveforms of an earthquake pair, is a powerful and widely used tool in seismology. However, its performance can be severely affected by several factors, including the length of the cross-correlation window, the frequency band of the applied digital filter, and the presence of large-amplitude phase(s). To optimize the performance of MF, we first systematically examine the effects of different operational parameters and determine generic rules for selecting the window length and the optimal frequency passband. To minimize the influence of large-amplitude phase(s), we then propose a new approach, namely, MF with multi-segment cross-correlation (MFMC). By equally incorporating the contributions from various segments of the waveforms, this new approach is much more sensitive to small separation between two sources compared to the conventional MF using the entire waveform template. To compare the reliability and effectiveness of both methods in capturing inter-event source separation and identifying potential repeating earthquakes, we systematically conduct experiments with both synthetic data and real observations. The results demonstrate that the conventional MF can detect the existence of an event but sometimes lacks the resolution to tell the large hypocentre distance between the template and detected events, whereas MFMC works in all cases tested. The far-reaching implication from this study is that inferring source separation between an earthquake pair based on the conventional MF method, particularly with data from a single channel/station, may not be reliable.

### 2.2 Introduction

The match-filtering (MF) method uses waveform cross-correlation to determine the similarity between a pair of events. It is a powerful tool in modern seismology to identify repeating earthquakes (e.g., Nadeau et al., 1995; Igarashi et al., 2003; Uchida et al., 2003; Matsuzawa et al., 2004; Schaff and Richards, 2004, 2011; Yamashita et al., 2012; Naoi et al., 2015; Meng et al., 2015; Schmittbuhl et al., 2016; Huang and Meng, 2018) and to detect events that can be

easily missed by conventional phase arrival-based methods (e.g., Gibbons and Ringdal, 2006; Shelly et al., 2007; Peng and Zhao, 2009; Skoumal et al., 2014, 2015, 2019; Zhang and Wen, 2015; Schultz et al., 2014, 2017; Warren-Smith et al., 2017, 2018; Chamberlain and Townend, 2018; Ross et al., 2019). The extensive applications of this technique have led to major observational breakthroughs (e.g., Shelly et al., 2007).

The cross-correlation coefficient (CC), a value characterising the degree of similarity between different waveforms, is often taken as the sole criterion in MF for repeater identification (e.g., Nadeau and McEvelly, 1999; Buurman and West, 2010; Schaff and Richards, 2004, 2011; Ma and Wu, 2013; Hatch et al., 2020) and earthquake detection (e.g., Zhang and Wen, 2015). When dealing with repeating earthquakes with nearly identical waveforms recorded at one or more common stations, it is usually assumed that events with very high CC belong to the same cluster and physically represent repeated ruptures in the vicinity of the same patch of the same fault. The employed CC thresholds typically range from 0.70 to 0.98 (e.g., Nadeau and McEvelly, 1999; Igarashi et al., 2003; Uchida et al., 2003; Matsuzawa et al., 2004; Schaff and Richards, 2004, 2011; Green and Neuberg, 2006; Schultz et al., 2014; Meng et al., 2015; Naoi et al., 2015; Schmittbuhl et al., 2016; Hayward and Bostock, 2017; Yamada et al., 2016; Bohnhoff et al., 2017; Huang and Meng, 2018).

Little attention has been paid to the operational parameters, such as the cross-correlation window length and filter frequency band, that can significantly affect the CC values (e.g., Harris, 1991; Schaff, 2008; Baisch et al., 2008). Because the calculation of CC is most sensitive to the phase(s) with large amplitude(s) within the template window (often the shear wave and surface waves) (e.g., Buurman and West, 2010; Myhill et al., 2011; Calderoni et al., 2015; Schmittbuhl et al., 2016; Li et al., 2017), the contribution of other phases with low amplitudes, such as depth phases (Ma and Atkinson, 2006; Ma, 2010) and coda waves (Snieder and Vrijlandt, 2005; Robinson et al., 2011), which contain additional source location information, can be overwhelmed. Thus, a very high CC due to the match of specific high-amplitude phase(s) does not necessarily represent a small distance between two hypocentres. In other words, the traditional waveform cross-correlation-based MF approach may be an excellent tool to detect events, but how close they are can be very uncertain.

The focus of this study is to understand the limitations of the traditional MF method and to develop a new approach that can more reliably identify potential repeating earthquakes, especially with data from a single station due to sparse network coverage. Even for regions with excellent network coverage and high station density (e.g., Japan), it is a popular practice to use only two stations in searching for repeating earthquakes because of their generally low magnitudes (e.g., Igarashi et al., 2003; Igarashi, 2020; Uchida et al., 2003; Matsuzawa et al., 2004; Nishikawa and Ide, 2018). To optimize the performance of MF, we first investigate how the operational parameters can influence CC values and define generic rules for specifying the window length and the frequency passband of the applied filter. To minimize the effects of large-amplitude phase(s) present in the waveforms, we propose a modification to the existing MF: match-filtering with multi-segment cross-correlation (MFMC). This new approach is very effective in recognizing the hypocentre location difference between a pair of events because it equally considers the contributions from all segments of the waveforms. A slight hypocentre shift causing subtle changes in the waveforms will be reflected in the sudden drop of CC values with the new approach. Finally, we verify the effectiveness of the MFMC method by applying it to both synthetic data and real observations. The results demonstrate that the MFMC approach can unambiguously discriminate event pairs with small separation from those with large separation in either horizontal or vertical directions.

### 2.3 Factors Affecting CC values

Traditionally, the waveform similarity between an event pair with same window lengths is defined by the normalized cross-correlation coefficient:

$$CC = \frac{\sum_{i=1}^n [a(i) - \bar{a}][b(i) - \bar{b}]}{\sqrt{\sum_{i=1}^n [a(i) - \bar{a}]^2 \times \sum_{i=1}^n [b(i) - \bar{b}]^2}} \quad (2-1)$$

where  $a$  and  $b$  correspond to the discrete time-series of two events,  $\bar{a}$  and  $\bar{b}$  are the mean values of each time-series, and  $n$  is the total number of samples. CC ranges from  $-1$  (reversed shapes) to  $1$  (identical shapes). The CC value of  $0$  represents no correlation (i.e., the signals are orthogonal). For repeater identification, part or all of the wave train of a well-located event is normally utilized as a template event to scan through continuous waveforms. CC calculation is

usually performed for every sample point. Once the computed CC exceeds a certain threshold, a repeating event is declared (e.g., Nadeau and McEvilly, 1999; Igarashi et al., 2003; Uchida et al., 2003, 2010; Meier et al., 2004; Schaff and Richards, 2004, 2011; Green and Neuberg, 2006; Petersen, 2007; Yamashita et al., 2012; Kato and Nakagawa, 2014; Meng et al., 2015; Warren-Smith et al., 2018; Chaves et al., 2020; Hatch et al., 2020). Only a few studies may take additional efforts to confirm the detection (e.g., Cheng et al., 2007; Li et al., 2007, 2011; Naoi et al., 2015).

### 2.3.1 Window Length

It is clear that CC is a function of time window length. However, there is no standard rule of selecting time windows for CC calculation. Different studies favor different window lengths that generally cover high-amplitude phase(s), ranging from seconds to minutes (e.g., Nadeau et al., 1995; Igarashi et al., 2003; Schaff and Richards, 2004, 2011; Skoumal et al., 2014; Zhang and Wen, 2015; Schultz et al., 2017; Huang and Meng, 2018; Ross et al., 2019). Using a shorter time window (normally containing at least one high amplitude phase) is more likely to yield a higher CC value. In the extreme case, for example, the CC will be exactly  $\pm 1$  if the window contains only 1 data point according to Equation 2-1. In contrast, using a longer time window is meant to compare all the phases of interest within that window and tends to result in a relatively lower CC value.

Because the total length of a seismic wave train increases with source-receiver distance, using a time window with a fixed length can only work for specific cases. To properly consider the distance effect, a more general and reasonable choice is to use a dynamic window based on the differential traveltimes between P and S phases (e.g., Baisch et al., 2008), i.e.,

$$T_{win} = k(T_s - T_p) \quad (2-2)$$

where  $T_{win}$  is the window length,  $k$  is a constant, and  $T_s$  and  $T_p$  represent S wave and P wave arrival times, respectively. If the window starts from the P onset,  $k$  is set to be 3. This choice is similar to the preferred value proposed by Baisch et al. (2008). If the window starts from the S onset,  $k$  is usually 2.

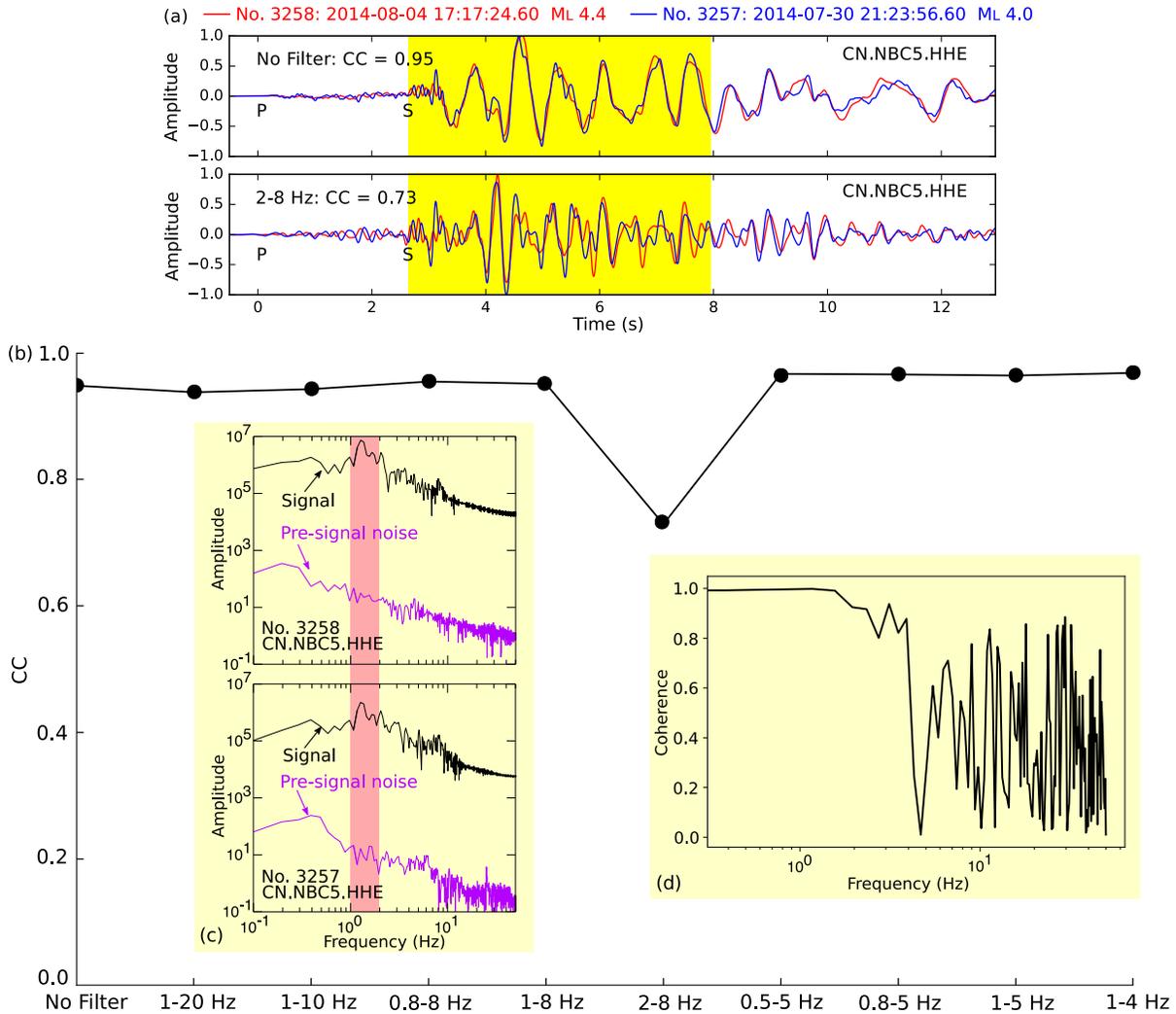
### 2.3.2 Frequency Band of the Digital Filter

A variety of filters ranging from very broad to very narrow frequency bands have been adopted by different studies prior to performing waveform cross-correlation (Table 2.1). Some studies choose the frequency passband based on earthquake magnitude or noise characteristics while the others may subjectively select the filters without explicitly explaining the reason. Overall, no simple rule can be summarized from Table 2.1 for selecting the optimal frequency passband. In this section, we demonstrate that, given the same dataset, different frequency bandwidths used in data processing may yield different CC values. Consequently, bias may occur during the identification of repeaters if we only look at the face value of the computed CC without considering the effect of the chosen filters. Specifically, we apply some commonly used band-pass filters (Table 2.1) as well as high/low-pass filters to a real waveform example.

In Figure 2.1, we present an event pair with similar waveforms occurring in northeastern British Columbia, Canada (Visser et al., 2017). This pair (No. 3258 and 3257) is characterized by high signal-to-noise ratio (SNR) (see waveforms in Figure 2.1a and amplitude spectra in Figure 2.1c). To calculate the CC value of an event pair, we cut the template from the S wave arrival with a length of  $2(T_s - T_p)$  from one event. Then the template is slid from 0.2 s before the S arrival of the other event to 0.2 s after, with a step of one sample. A  $\pm 0.2$  s shift should be adequate to take care of any manual phase-picking error. The maximum value of the cross-correlation function that results from the sliding process is defined as the final CC value of the event pair. It is obvious that utilizing raw or band-pass filtered waveforms yield similar cross-correlation results except for the case of 2–8 Hz band-pass filtering (Figure 2.1b). In this particular case, the improper choice of a 2–8 Hz band-pass filter would remove the dominant frequency band between 1 and 2 Hz in the waveform (Figure 2.1c), thus make the seismic waveforms less similar (Figure 2.1a, bottom panel), as shown by the sudden drop of CC value (Figure 2.1b).

**Table 2.1.** Digital filters commonly used in waveform match-filtering

Filter	References	Magnitude (M)	Epicentral Distance (km)	Frequency Range Dependency
No Filter	Warren-Smith et al. (2018)			
1–20 Hz	Kimura et al. (2006) Cannata et al. (2013)	$2.0 \leq M \leq 4.56$ $0.5 \leq M \leq 4.4$	$\leq \sim 100$ $\leq \sim 150$	Instrument response and noise characteristics
1–10 Hz	Li et al. (2007) Li et al. (2011) Ma and Wu (2013) Ma et al. (2014) Cociani et al. (2010) Schmittbuhl et al. (2016)	$1.1 \leq M \leq 2.8$ $0.9 \leq M \leq 2.8$ $\sim 1.5 \leq M \leq \sim 3.5$ $1.4 \leq M \leq 4.2$ $1.5 \leq M \leq 2.5$ $0.78 \leq M \leq 2.72$	$\leq 150$  $\approx 35$ $\leq \sim 250$ $\leq \sim 35$ $\leq \sim 30$	Noise characteristics
0.8–8 Hz	Schultz et al. (2014)	$\sim 0.5 \leq M \leq \sim 4.0$	$\sim 30$ and $\sim 200$	
1–8 Hz	Matsubara et al. (2005) Huang and Meng (2018) Taira et al. (2014) Dominguez et al. (2016)	$2.0 \leq M$ $2.5 \leq M < 3.0$ $2.0 \leq M \leq 3.5$ $2.5 \leq M \leq 4.5$	$< 300$  $< \sim 150$	Magnitude Magnitude Noise characteristics
2–8 Hz	Yamashita et al. (2012) Igarashi (2020)	$2.0 \leq M$ $2.5 \leq M \leq 2.9$	$< 50$ $\leq 400$	Magnitude
0.5–5 Hz	Green and Neuberg (2006) Schaff and Richards (2011)	$2.0 < M < \sim 6.0$	$\approx 1$ $< \sim 2000$	Noise characteristics
1–4 Hz	Uchida et al. (2003) Matsuzawa et al. (2004) Huang and Meng (2018) Igarashi (2020)	$2.0 \leq M$ $2.0 \leq M$ $3.0 \leq M$ $3.0 \leq M$	   $\leq 400$	Magnitude Magnitude



**Figure 2.1.** Effect of band-pass filtering on the result of waveform cross-correlation (CC). (a) Normalized waveforms of a pair of events (No. 3258 and 3257 taken from Visser et al., 2017) aligned according to S wave arrivals. Top and bottom panels show original and filtered waveforms, respectively. CC value determined from the yellow shaded segment is labelled for each panel. (b) CC values determined after applying some commonly used band-pass filters. (c) Amplitude spectra of each event and its corresponding pre-signal noise. Pink bar outlines the dominant frequency band of seismic energy. (d) The coherence function of this event pair.

We reach the same conclusion when various high-pass filters are applied. Once the dominant seismic energy (in the 1–2 Hz frequency band) is filtered out, the CC value drops significantly (Figure 2.10a in the Supporting Information). With a low-pass filter, however, the CC can be extremely high ( $> 0.99$ ) even though the dominant frequency energy has been removed (corner frequency  $\leq 1$  Hz, Figure 2.10b). This is because the remaining frequency content of the signal becomes so narrow-band that the waveforms of these two events are essentially identical (Figure

2.10c). Another way of verifying the selection of frequency passband is to examine the coherence function (essentially the regular CC function in the frequency domain; Uchida, 2019) between the two signals (Figure 2.1d). Overall the coherence function provides consistent information that the two waveforms match best at lower frequencies. However, it cannot identify which frequency band (1–2 Hz in this case) we have to keep in order to boost the waveform similarity.

Taken together, CC is definitely a function of the frequency band of the applied digital filter. Choosing a proper filter is very important when performing waveform cross-correlation. By checking amplitude spectra of both signal and pre-signal noise, we can design a filter that keeps the dominant seismic energy while reducing background noise more effectively. We note, however, that keeping the dominant energy in the signal does not mean choosing a filter with a very narrow frequency bandwidth, which may lead to meaningless correlation and even false detections (Harris et al., 1991; Schaff, 2008; Carrier and Got, 2014; Dodge and Walter, 2015). Instead, we could consider the time-bandwidth product of the seismogram that characterizes the amount of information contained in the signal to determine the optimal bandwidth. With a larger time-bandwidth product, i.e., longer signal duration and wider frequency bandwidth, the cross-correlation can be of greater statistical significance (Schaff, 2008; Schaff et al., 2018).

We caution that the choice of frequency band should also depend on the source size and rupture process when identifying repeaters (Uchida, 2019). A predominantly low-frequency passband may not have sufficient spatial resolution to identify neighbouring events that do not overlap with each other, whereas a passband focusing on very high frequencies may increase the likelihood of excluding true repeaters that occur in the same source area but with different rupture characteristics (e.g., the rupture nucleation point and/or directivity). Readers interested in further discussion on this aspect are referred to Harris et al. (1991), Schaff (2008), and Uchida (2019) for more details.

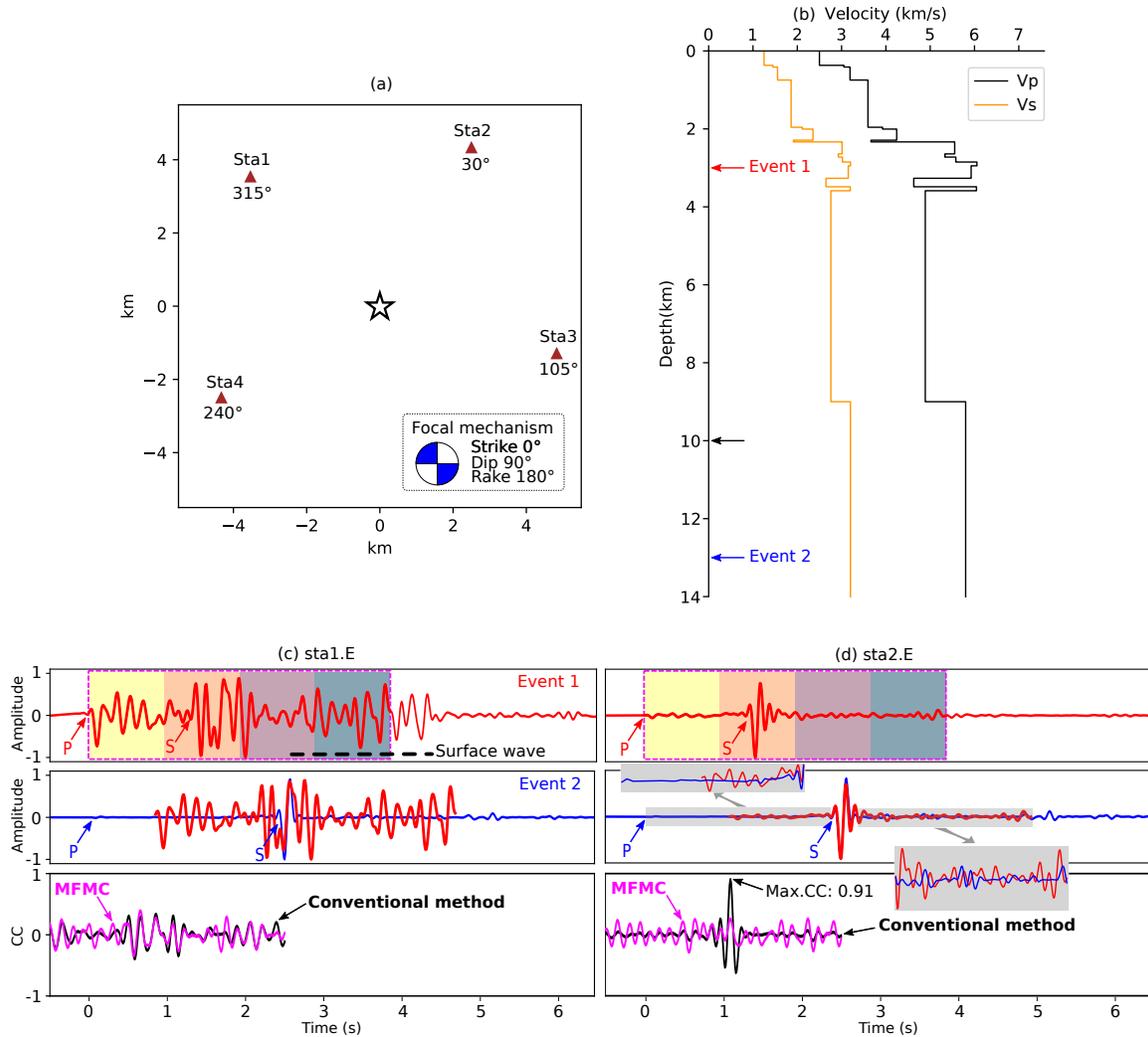
### **2.3.3 Large-amplitude Phase(s)**

As outlined in the Section 2.2, the conventional way of calculating CC is most sensitive to the large-amplitude phase(s) within the window. For demonstration purposes, here we show one synthetic example. The configuration of this experiment is given in Figure 2.2a. We place two

strike-slip events in the centre of an array at depths of 3 km (event 1) and 13 km (event 2), respectively. The four stations are all 5 km away from the epicentre. We calculate the synthetic seismograms using the frequency-wavenumber approach (Zhu and Rivera, 2002). The velocity model used in the calculation (Figure 2.2b) is taken from the ToC2ME experiment (Eaton et al., 2018; Tan et al., 2019). The source time function is assigned as a simple triangular shape with a duration of 0.1 s, suitable for small earthquakes (Harrington and Brodsky, 2009). All synthetic seismograms are filtered in the range of 1–10 Hz, a frequency band typical for local earthquakes (Warren-Smith et al., 2017).

The E channel of event 1 at station 1 (azimuth = 315°) exhibits clear P wave, S wave and surface wave trains (Figure 2.2c, top panel). The amplitudes of the P and S waves are comparable because station 1's location is anti-nodal for P and SV waves and nodal for SH wave. In contrast, the P wave and surface wave of event 2 are obscured (Figure 2.2c, middle panel) because of the much greater focal depth. Here we employ a template window that starts from the P wave onset of event 1 with a length of  $3(T_s - T_p)$  to cross-correlate the corresponding segment of event 2 based on the traditional method (Equation 2-1). The poor CC indicates that these two events are not correlated owing to the very large difference in hypocentral locations (Figure 2.2c, bottom panel).

However, if we take a closer look at the E channel of station 2 (azimuth = 30°), both events show very strong S waves yet very little P and surface waves (Figure 2.2d, top and middle panels). Even though these two events are 10 km apart in depth with very different S-P differential traveltimes and the majority of the waveforms before and after S waves are very different (Figure 2.2d, gray shaded box in the middle panel), the computed CC based on the conventional approach still remains high (>0.9) because of the simple match of the high-amplitude S phase (Figure 2.2d, bottom panel). Therefore, a high CC obtained through the conventional waveform cross-correlation method cannot guarantee the spatial separation to be small, especially in the cases that one (or few) large-amplitude phase dominates the waveform in the calculation window.



**Figure 2.2.** Synthetic experiment to test the performance of the match-filtering (MF) method. (a) Station-source setup. Brown triangles mark the seismic stations with azimuths labelled below the symbols. Open star represents the epicentral location of event 1 and event 2. Insert shows the focal mechanism used in the synthetic waveform calculation. (b) Velocity model from Eaton et al. (2018). Red and blue arrows mark the depths of events 1 and 2, respectively. Red and black arrows also denote the template depths used in the tests in Section 2.5. Top panels in (c) and (d) are band-pass filtered (1–10 Hz) and normalized synthetic seismograms of event 1 at station 1 and station 2, respectively. Dashed fuchsia boxes show the template windows with one-segment used in the conventional MF. Color shaded areas represent the 4 segments employed in the MFMC. Dashed thick black line indicates the surface wave train after the S phase. Middle panels in (c) and (d) show the band-pass filtered (1–10 Hz) and normalized synthetic seismograms of event 2 at station 1 and station 2, respectively. Template from event 1 is superposed at the location of the best match according to the conventional MF. Bottom panels in (c) and (d) display the cross-correlation results computed with the conventional approach and MFMC method, respectively.

## 2.4 Proposed Alternative Form of Cross-Correlation

### 2.4.1 Theory

As shown in Figure 2.2d, the traditional approach is very sensitive to the large-amplitude phase(s) and thus the result may not be reliable in identifying potential repeating earthquakes. To overcome this issue, we introduce the MFMC method. The first step is to divide the template (with  $n$  samples) into  $N_{seg}$  segments of equal lengths ( $L = n/N_{seg}$ ). By shifting the segments together along the continuous waveform, we perform the cross-correlation calculation for each individual segment and its corresponding waveform, i.e.,

$$CC_m = \frac{\sum_{i=(m-1)L+1}^{mL} [a_m(i) - \bar{a}_m][b_m(i) - \bar{b}_m]}{\sqrt{\sum_{i=(m-1)L+1}^{mL} [a_m(i) - \bar{a}_m]^2 \times \sum_{i=(m-1)L+1}^{mL} [b_m(i) - \bar{b}_m]^2}} \quad (2-3)$$

where  $m = 1, 2, \dots, N_{seg}$ . The final  $CC$  at each sample point is defined as:

$$CC = \frac{\sum_{m=1}^{N_{seg}} CC_m}{N_{seg}}. \quad (2-4)$$

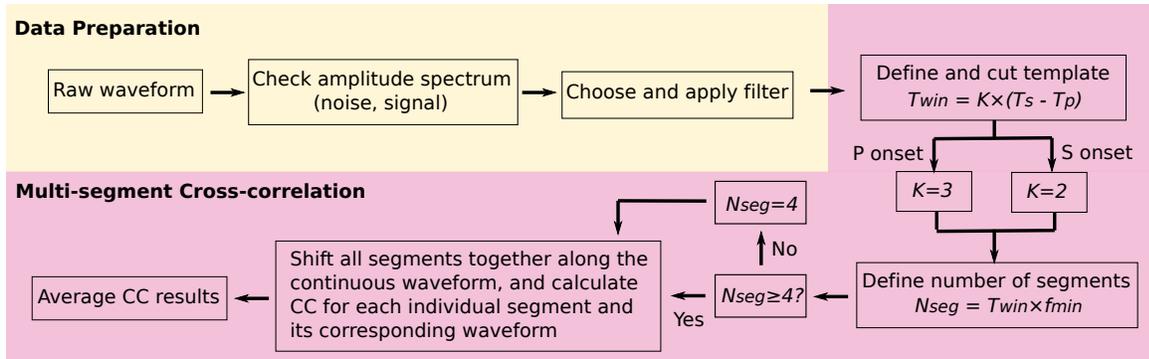
The parameter  $N_{seg}$  essentially acts as a weighting factor for each individual segment (Equation 2-4). If  $N_{seg}$  equals 1, the new method using a template with only one segment is identical to the conventional approach (Equation 2-1). Using a larger  $N_{seg}$  divides the template into more segments and is more sensitive to the waveform changes of another event due to location difference but increases the computing time. A practical way of assigning  $N_{seg}$  is to consider the cycles of the longest period wave ( $1/f_{min}$ ) in the band-pass filtered template, i.e.,

$$N_{seg} = \frac{T_{win}}{\frac{1}{f_{min}}} = T_{win} \times f_{min} \quad (2-5)$$

Based on our experience, a minimum  $N_{seg}$  of 4 is required to achieve the optimal performance of our new method. In Figure 2.3, we summarize the work flow of MFMC.

To demonstrate the strength of our new method, we apply it with 4 segments to the two synthetic events discussed in Section 2.3.3. For the E channel of station 1, both the old and new methods yield very similar results with low  $CC$  values (Figure 2.2c, bottom panel) suggesting that these two events are poorly correlated. For the same channel of station 2, the  $CC$  value obtained with our new method (0.26) is significantly lower than that with the conventional

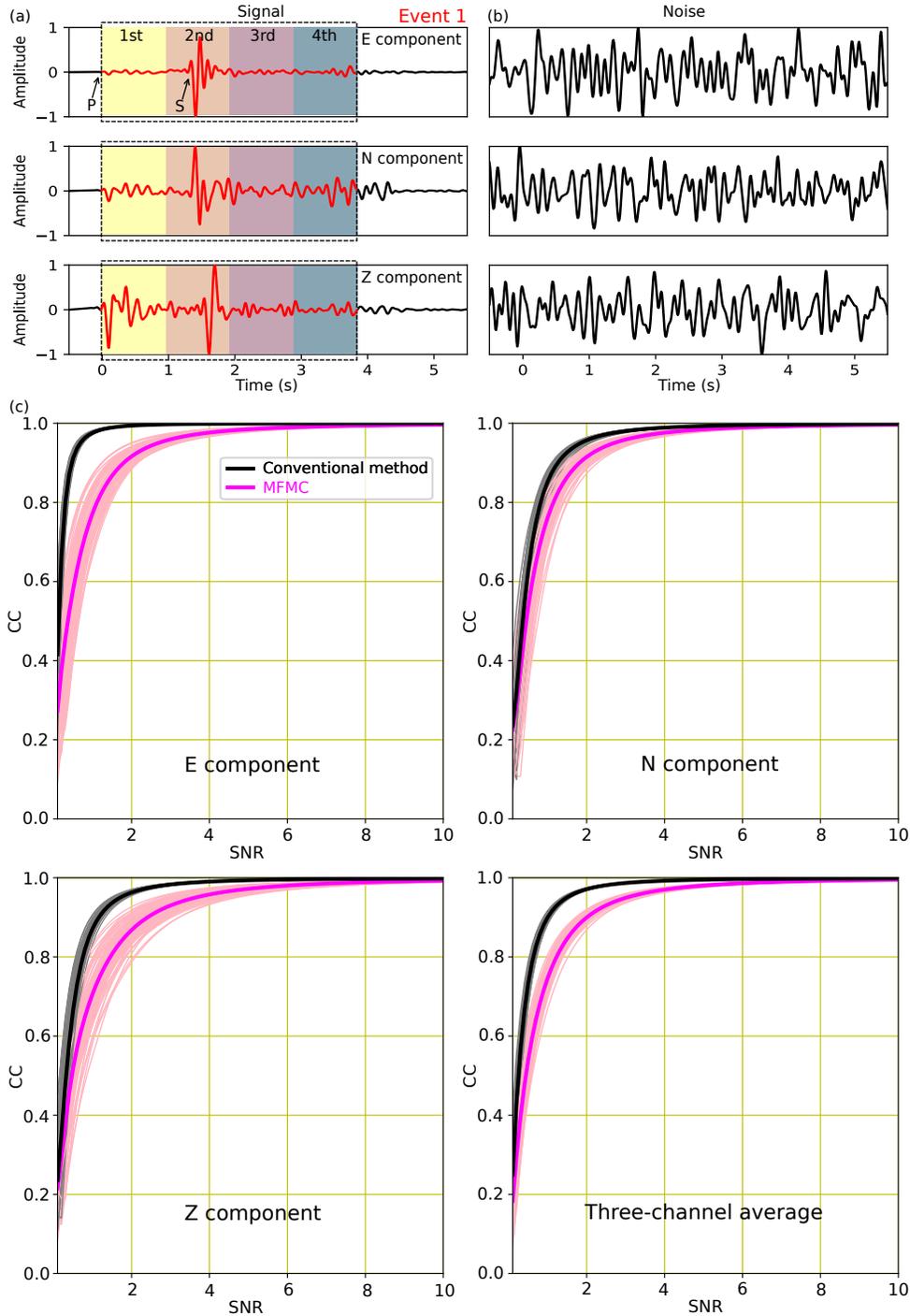
method (0.91), indicating significant waveform differences due to the large spatial separation between these two events. This simple experiment demonstrates that the multi-segment approach is much more effective in recognizing the subtle change in the waveforms due to source location difference.



**Figure 2.3.** A flow chart to illustrate the steps of match-filtering with multi-segment cross-correlation (MFMC).

## 2.4.2 Effect of Background Noise

To test how sensitive our new method is to the level of background noise, we conduct several additional experiments in which we cross-correlate an accurate (simulated) signal and a noisy version of the same signal for various noise levels. Such tests are designed intentionally to imitate the situation of using a template event to detect repeaters occurring at the same fault patch. In practice, usually the template event is carefully selected with sufficient magnitude so that it has an extremely high SNR, although it will never be perfectly accurate (i.e., totally free from noise contamination). In contrast, the detected event may be contaminated by noise due to a slightly lower magnitude.



**Figure 2.4.** A test on the effect of background noise. (a) Normalized waveforms of event 1 at station 2. Symbols are the same as in Figure 2.2d. (b) Examples of normalized Gaussian noise (1 out of 100 realizations) filtered between 1 and 10 Hz. (c) Curves showing the relationship between CC and SNR. Thin gray and light pink lines represent the results of the conventional MF and MFMC with each noise realization, respectively. Thick black and fuchsia lines denote averages over the 100 noise realizations for the conventional and MFMC methods, respectively.

In the experiments, first we take the three channels of event 1 at station 2 as signal (Figure 2.4a) and generate Gaussian noise (Figure 2.4b) with Numpy (Walt et al., 2011; <https://numpy.org>). The noise is filtered in the same frequency range as the signal. We then make noisy seismograms by adding scaled noise to the signal according to a given SNR. The SNR in this paper is defined by the ratio of the mean absolute amplitude of signal to that of noise, i.e.,

$$\text{SNR} = \frac{\sum_{i=1}^n |\text{signal}(i)|/n}{\sum_{i=1}^m |\text{noise}(i)|/m} \quad (2-6)$$

where  $n$  and  $m$  are the numbers of samples in the signal and noise waveforms, respectively. It should be noted that we purposely exclude samples in the segment with the largest amplitude, the second segment in the top panel of Figure 2.4a, for example, to emphasize the comparison between low amplitude phases and noise when calculating the mean absolute amplitude of signal. Waveform cross-correlation is carried out with both the old and new methods between the accurate signal and the noisy seismograms. For each assumed SNR, we repeat the experiment 100 times to account for the randomness of noise.

The results are summarized in Figure 2.4c. With a very high SNR ( $\geq 6$ ), the effect of background noise is negligible for both the conventional MF and MFMC methods. With a low SNR, however, the influence from noise becomes much more pronounced. Overall, the CC value calculated with the multi-segment approach is lower and drops faster than that obtained with the conventional method. This suggests that our new approach is more sensitive to the level of background noise as it intrinsically emphasizes minor details in the waveform.

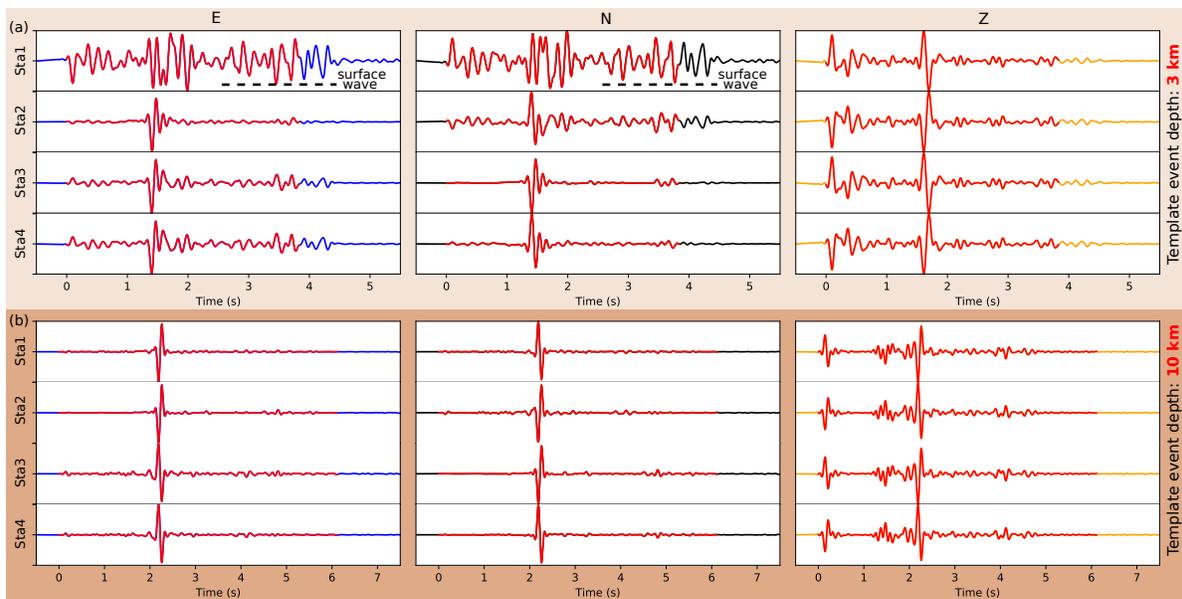
Furthermore, we notice that the CC curves derived with the conventional MF show varied noise sensitivities over different channels while those obtained with MFMC are sensitive to the specific noise added each time (Figure 2.4c). For example, the CC value calculated with the conventional method shows a clear decreasing trend when the SNR is  $\leq 4$  for both N and Z channels. But the same decreasing trend is observed for the E channel when SNR is  $\leq 2$ . In contrast, the CC curves derived with the MFMC in general display a similar pattern over different channels. This indicates that the MFMC results are less dependent on a specific channel used in the calculation. However, the MFMC-derived CC values seem to be relatively more dependent on the specific noise added each time as the variation between individual tests is more apparent.

In summary, there is a trade-off between the ability of reliably detecting potential repeating earthquakes and the tolerance of noise. Both methods can work similarly well when SNR is  $\geq 4$  and suffers from high level of noise when SNR is  $\leq 1$  (Figure 2.4c). As discussed in Section 1.2, even for a well-recorded event pair with nearly no noise contamination (Figure 1.2), it is very difficult to reliably decide whether they are true repeaters or just neighboring events based on the waveform similarity alone. If the noise level is high, the situation can become much worse as the waveform discrepancy can predominantly arise from noise, or hypocentre location difference, or from both. In such cases, it will also be extremely difficult, if not impossible, to precisely estimate other important parameters such as S-P differential traveltimes and source dimension, which are crucial to repeater verification (e.g., Li et al., 2007, 2011). Therefore, for robust detection of potential repeaters, we recommend to use MFMC only in reasonably high SNR cases (such as SNR  $\geq 4$  or even higher) for two reasons. First, the waveform discrepancy between an event pair can be more confidently attributed to hypocentre difference if the noise is negligible. Second, we can verify whether the detected events are true repeaters or not by other means with high confidence. We note that the exact SNR threshold appropriate for the MFMC method may depend on data availability/quality, noise characteristics, and epicentral distance, and more importantly, it may depend on how confident we are to get other important precise estimations such as S-P differential traveltimes and rupture area for final verification.

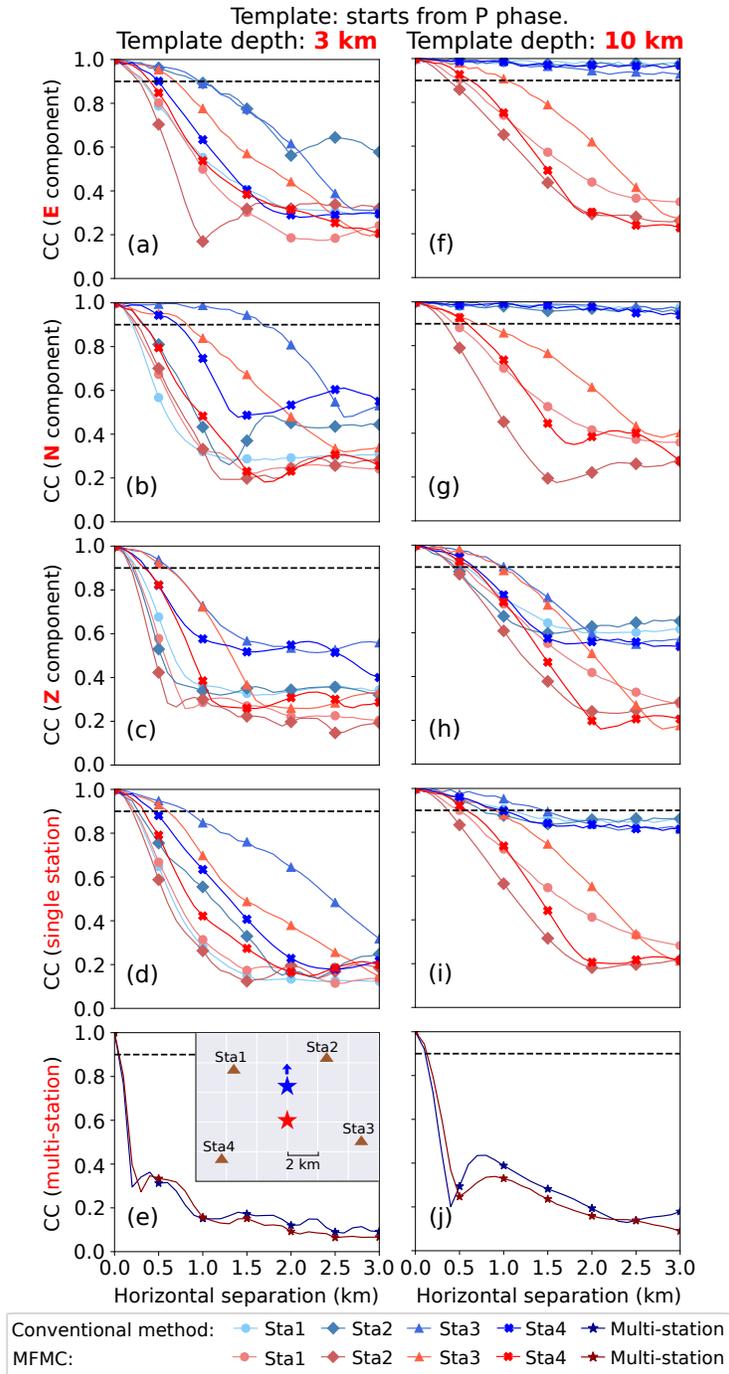
## **2.5 Verification: Constraining Inter-event Separations Using Synthetic Data**

As we have shown that the conventional one-segment cross-correlation approach is not always reliable in differentiating the location difference between two earthquakes, here we apply both methods to synthetic data to systematically compare their sensitivity to inter-event separations. In this experiment, we consider spatial separation in either horizontal or vertical directions. The station geometry is the same as that in Figure 2.2a and one event (the template event) is placed at the centre of the array. Then we incrementally shift the source (the matched event) towards the north or towards the surface by 0.1 km each time. The synthetic seismograms (see examples in Figure 2.5) are generated and processed in the same way as in Section 2.3.3. When implementing our new method, we follow the working procedure illustrated in Figure 2.3.

In our experiment, we considered three different focal mechanisms: strike-slip, normal, and reverse. The results of normal and reverse-faulting essentially reach the same conclusion of those of strike-slip fault (Figures 2.6-2.7 and Figures 2.11-2.22 in the Supporting Information). To keep our discussion focused, only the results from the strike-slip fault are presented and discussed here. The other cases are presented in the Supporting Information (Figures 2.19-2.22) for readers who are interested in more technical details.



**Figure 2.5.** Band-pass filtered (1–10 Hz) and normalized synthetic seismograms of two template events at different depths are shown in (a) and (b). Source-station setup is the same as that in Figure 2.2a. Red lines mark the templates that start from the P phase.



**Figure 2.6.** CC variation due to horizontal inter-event separation with a template starting from the P phase. Detailed configuration of the experiment, including the assumed strike-slip focal mechanism, is shown in Figure 2.2. Plots on the left and right correspond to the results of the template event at depths of 3 and 10 km, respectively. Dashed black line marks the CC value of 0.9 for reference. Insert in (e) schematically shows the arrangement of horizontal separation between two sources. Red star denotes the template event and blue star represents the matched event. Station geometry is the same as that in Figure 2.2a.

### 2.5.1 Horizontal Inter-event Separations

For horizontal inter-event separations, we consider two situations: the template event occurs at shallow depth (3 km) and deep depth (10 km). The most pronounced difference between waveforms of the two template events is the amplitude variation of surface waves. When the source depth is at 3 km, the amplitude of surface waves can be comparable to that of the S waves (Figure 2.5a). In contrast, waveforms of the template event at 10 km are all dominated by the S waves except the vertical channel (Figure 2.5b).

To investigate how the CC value derived from both methods changes in response to horizontal source separations (Figure 2.6e, insert), we tested three scenarios: single channel, single station with three channels (Z, N, E), and multi-station with all channels. For the first scenario, the seismogram of the template event is slid from 0.5 s before the corresponding phase of the matched event to 0.5 s after. We consider two kinds of templates starting from either the P or S onset. The window length is dynamically determined by Equation 2-2. The peak CC value during the sliding is taken as the final CC between the template and matched events. For the second scenario, the CC from all three channels of the same station are averaged and the maximum within the sliding window is taken as the final CC. For the third one, we average the CC values over different stations and channels with a fixed time difference between the template and matched events. In this case, a larger shifting window (from 0.5 s before the earliest P or S arrival to 0.5 s after the latest P or S arrival) is used during the CC calculation.

We first conduct the experiment with a template containing both P and S phases. Figure 2.6a-e presents the results of how the CC value calculated by both methods varies with source separation when the template event is shallow (depth = 3 km). When a single channel is used, both one-segment and multi-segment methods yield very scattered results (Figure 2.6a-c). Overall, the CC obtained by both methods drops with increasing horizontal separation between the template and matched events. Yet slight variation exists. For example, the CC curve derived with the conventional method using the E channel of station 2 reaches the local minima at the separation distance of 2.0 km but increases at larger distances between 2.0 and 2.5 km (Figure 2.6a). This implies that using the CC values to constrain source separation may not be stable when the separation is large. For a given separation distance between two sources, the CC value obtained with our new method in general is lower than that with the conventional method,

suggesting that our new method is relatively more effective in recognizing spatial changes. A particular case should be mentioned that both methods work similarly well for all three channels of station 1 (Figure 2.6a-c). This is because the template waveforms are not dominated by any specific phase as found at other stations (Figure 2.5).

If three channels of the same station are included in the calculation, the scattering effect is reduced (Figure 2.6d). Especially for the new method, stations 1, 2 and 4 all give similar solutions at small separations. If more stations are included, both methods can work extremely well because the template and matched events have different traveltimes to different stations. The computed CC becomes very sensitive to source location difference as a very small hypocentre horizontal shift of only 0.2 km can lead to a dramatic CC drop (below 0.4) no matter which method is applied (Figure 2.6e).

When the template event is at greater depth (10 km), using the conventional method with the E or N channels from the same station is far from sufficient to constrain the hypocentre separation (Figure 2.6f and g). Using all three channels gives the similar conclusion (Figure 2.6i), which is not surprising because of the assigned equal weighting in the final CC calculation. The CC value remains high ( $> 0.9$  for the E, N, and  $> 0.8$  for the average of three channels) even for a separation as large as 3 km (Figure 2.6f, g, and i). The Z channel performs relatively better at small separations ( $< 2$  km) but loses sensitivity at larger separations ( $\geq 2$  km) as the CC curves become almost flat (Figure 2.6h). In comparison, the new method works more effectively in such cases, yet large CC variation with respect to the same change of source location exists when different channels and stations are used. Similar to the previous test, both methods work very well in constraining spatial separations when all stations are included. A point worth noting is that the performance of both methods using an array seems to become slightly poorer when the template event occurs at deeper depth (Figure 2.6e and 6j).

We further conduct experiments with a template that starts from the S phase. Without including the P wave part can make it even harder for the conventional MF to distinguish any source location variation, especially for the Z channel (e.g., Figure 2.11c and 2.11h). In contrast, choosing a template that starts from the P or S phase is not a major concern for the MFMC as the results are very similar (Figures 2.6 and 2.11). The results of these experiments suggest that the MFMC is more stable and less dependent of the choice of template windows. It is also worth

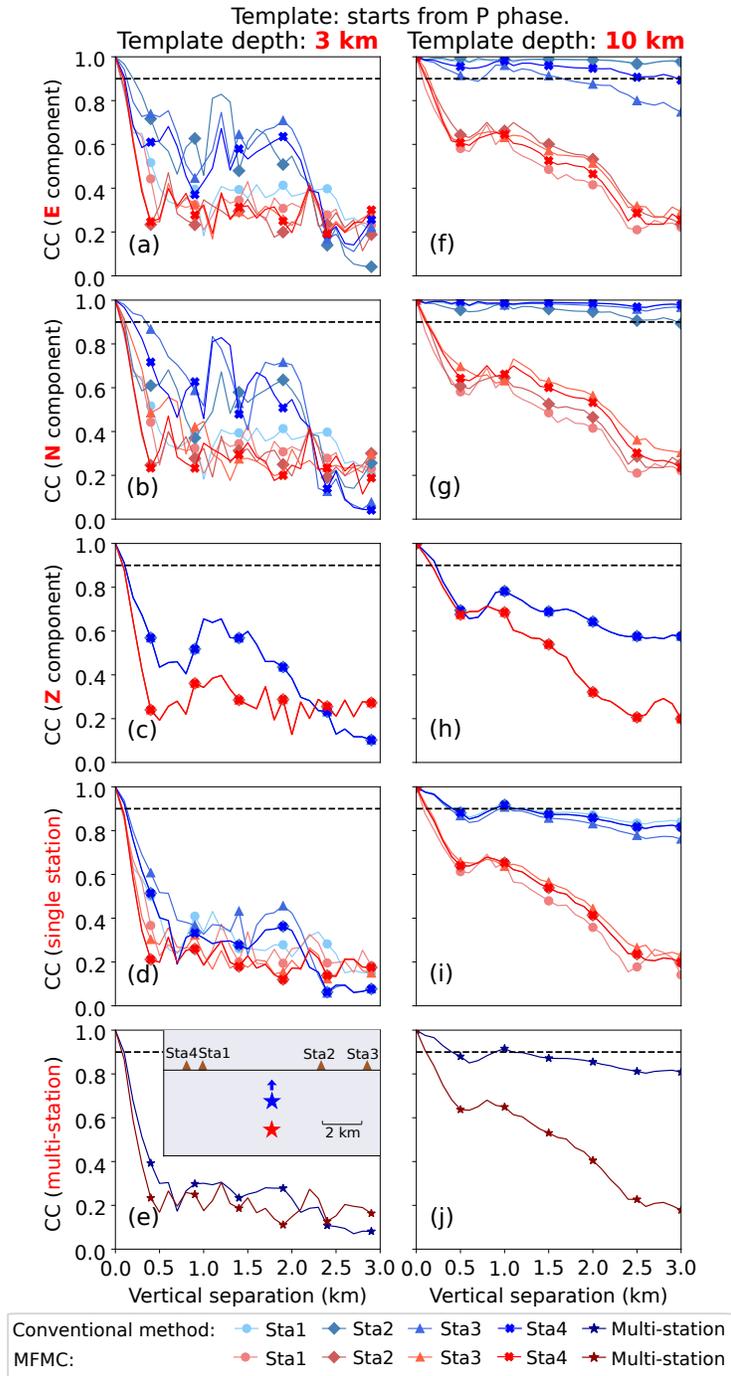
noting that the MFMC with a template starting from the S phase can outperform the conventional MF with a template that contains both P and S phases in discriminating the horizontal inter-event separations (Figures 2.6 and 2.11).

Adding considerable Gaussian noise ( $\text{SNR} = 2$ ) to the matched events in the experiment seems only to have pronounced influences on the MFMC-derived CC values at small separations ( $< 1$  km, Figures 2.12 and 2.13). Even with no separation, the CC determined by MFMC can drop from 1.0 to a value slightly larger than 0.8. This is not surprising as we have learned from Section 2.4.2 that MFMC is more sensitive to noise relative to the conventional MF. With the enlarging separation ( $> 1$  km), however, it is the source location shift that matters as the difference in the MFMC-derived CC curves between the cases with and without noise become negligible.

From the tests above, we conclude that our new method is more effective in recognizing horizontal source location changes when data from only a single channel or single station are used. With the help of more stations, however, both methods work similarly well and a slight horizontal hypocentre changes will lead to a significant reduction in CC values.

### **2.5.2 Vertical Inter-event Separations**

For vertical inter-event separations (Figure 2.7e, insert), we also investigate two cases by placing the template event at two different depths. Similar to the finding in the preceding section, our new method is more capable of capturing hypocentre changes than the conventional method (Figures 2.7 and 2.14). For the conventional method, using more channels/stations helps significantly in constraining the vertical hypocentre separations only when the source depth of the template event is shallow (Figures 2.7e and 2.14e). When the template event is deeper, the conventional method fails to recognize the depth difference between the two sources no matter how many channels and stations (up to 12 channels from 4 stations in our experiment) are used in the calculation while the new method is still capable of differentiating source depth changes but with much less sensitivity (Figures 2.7j and 2.14j).



**Figure 2.7.** CC variation due to vertical inter-event separation with a template starting from the P phase. Symbols and layout are the same as that in Figure 2.6.

For the new method, including more channels and/or stations does not appear to be helpful, as the single-channel MFMC can produce results similar to that of using an array (Figures 2.7 and 2.14). It should be noted that, for both methods, the CC curves from Z channels of different

stations are nearly identical (Figure 2.7c and 2.7h). This seems to suggest that, for strike-slip events, the Z channel is independent of station azimuth in constraining the vertical separation of the two sources. Similar to the conclusion obtained in Section 2.5.1, the MFMC with a template starting from the S phase can outperform the conventional MF with a template containing both P and S phases in most cases (Figures 2.7 and 2.14). Including the P phase can improve the depth resolution of MFMC only when the source is deep ( $\geq 10$  km). Also, the effect of adding random noise in the test is generally insignificant (Figures 2.15 and 2.16). It is interesting to note that both the conventional and new methods overall show less CC variation in constraining vertical inter-event separation than horizontal over different channels and stations.

### **2.5.3 Importance of Station Distribution**

We have demonstrated that the conventional multi-station MF lacks the resolution for vertical inter-event separation when the template event is deep (Figures 2.7 and 2.14-2.16). It is worth considering whether it is a worst-case scenario due to the model setup in which the same epicentral distance is used for all stations. Hence, we conduct additional experiments with the identical setup except the epicentral distance is different for each station ranging from 4 to 7 km (Figure 2.17). It should be noted that having two representative cases (one with equal epicentral distances to all stations while the other with different) is sufficient to illustrate the effect of station distribution given the fact that multi-station CC is not widely used (see the compiled Table 3.1 in Chapter 3).

As evident in Figure 2.18, with more variation in the traveltime moveout, conventional multi-station MF is indeed able to distinguish the assumed depth changes more effectively and its performance becomes much closer to that of the multi-station MFMC. For horizontal separations in particular, the resolution is approximately the same for both conventional MF and MFMC independent of station locations (Figure 2.18). Therefore, station distribution can be an important factor that affects the depth resolution of the conventional multi-station MF method, consistent with the finding of Chamberlain and Townend (2018). In contrast, the performance of MFMC is nearly independent of the setup of seismic stations.

## 2.6 Demonstration: A True Repeating Event or Not?

### 2.6.1 Real Examples from the Blanco Fracture Zone

To demonstrate how the conventional MF may misidentify different earthquakes as repeating events due to the lack of resolution in inter-event separation, we apply both the conventional MF and MFMC to a group of real earthquakes that occurred in the Blanco Fracture Zone (BFZ) in northeast Pacific (approximate location shown in Figure 2.23).

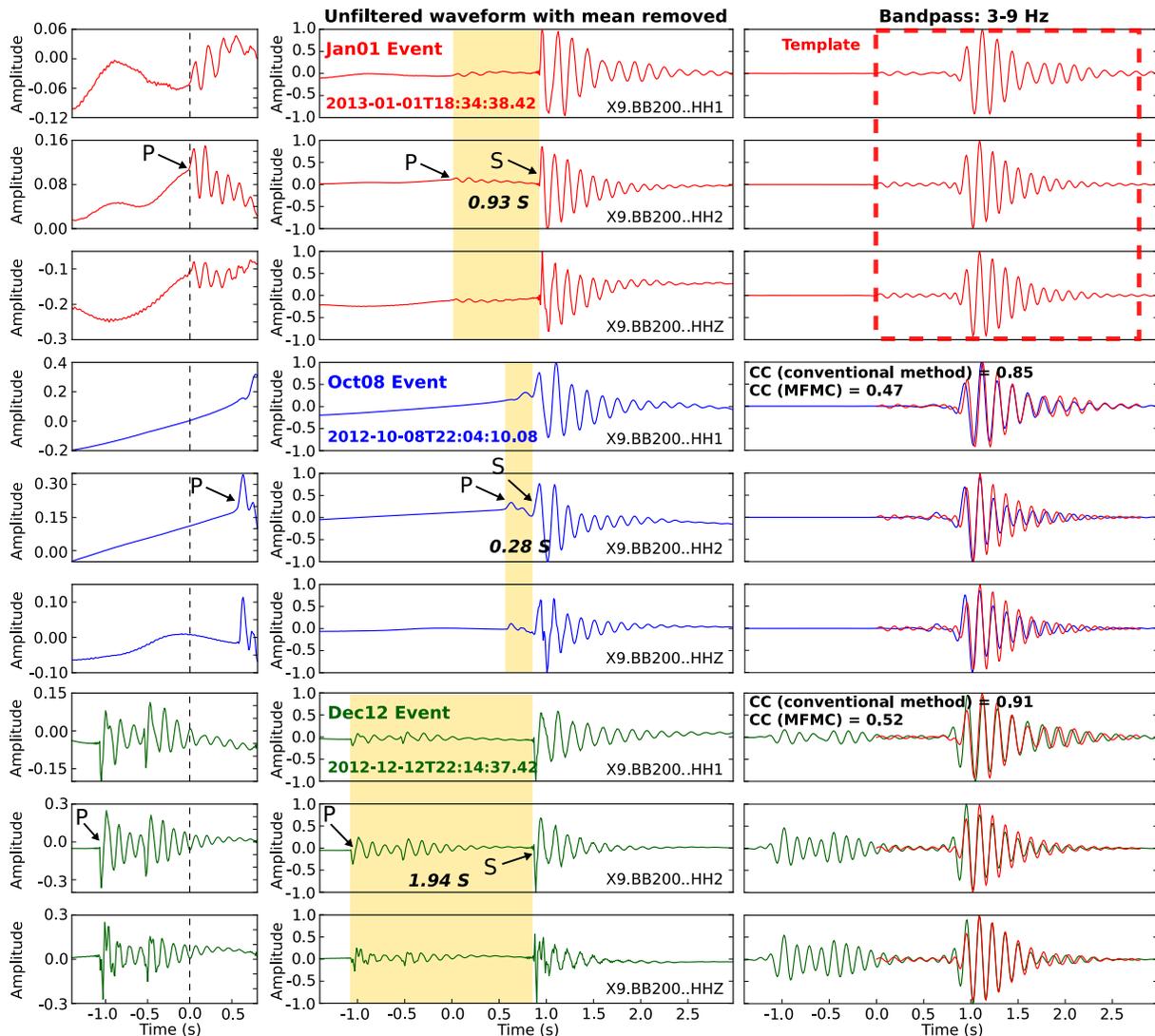
This group consists of three events on 8 Oct 2012, 12 Dec 2012, and 1 Jan 2013. For convenience, we refer them as Oct08, Dec12 and Jan01 events in the following text. These events are so small that only one nearby ocean-bottom seismometer (station X9.BB200, Figure 2.23) records clear signals and hence their precise hypocentres cannot be determined. The waveforms of these events are all dominated by the large S phases and prolonged ringing (middle column, Figure 2.8). Despite the P phases having very small amplitude, they can be clearly recognized from the background noise (left column, Figure 2.8). The S-P differential traveltimes of the three events are 0.93 s (Jan01), 0.28 s (Oct08), and 1.94 s (Dec12). Based on this evidence, there is no doubt that they are not repeating earthquakes. Assuming the average P and S wave velocities of 6.0 and 3.24 km/s, respectively (Kuna et al., 2019) and that these events occurred along the BFZ, we can estimate to first order the inter-event distance in the source-station direction for the event pairs of Jan01–Oct08 and Jan01–Dec12 as 4.58 and 7.11 km, respectively.

Here we take the Jan01 earthquake as the template event. The waveform templates start from the P phase with a length of  $3(T_s - T_p)$  (right column, Figure 2.8). All three events are band-pass filtered between 3 and 9 Hz according to the signal and pre-signal noise spectra (Figure 2.24). Taking the conventional MF approach, the three-channel averaged CC value for the Jan01–Oct08 event pair is as large as 0.85 (mainly due to the good match of the high-amplitude portion), despite the clear 0.65 s difference in the S-P differential traveltime (Figure 2.8). This CC value is higher than the threshold used by many previous studies to define repeating earthquakes (e.g., Buurman and West, 2010; Petersen, 2007; Schaff and Richards, 2004, 2011; Cannata et al., 2013; Ma and Wu, 2013). In a hypothetical scenario, if we incorrectly consider the Oct08 event as a repeating earthquake at the hypocentre of the Jan01 event, the misidentification may inevitably incur a large bias/error in subsequent analyses, such as

estimating fault creep rate (e.g., Nadeau and Johnson, 1998; Uchida et al., 2003, 2006; Yu, 2013; Materna et al., 2018) and/or monitoring subsurface velocity changes (e.g., Poupinet et al., 1984; Schaff and Beroza, 2004; Sawazaki et al., 2015; Li et al., 2017). With the MFMC approach, however, the CC value is lower than 0.5, mainly due to the mismatch of the low-amplitude portions of the waveforms (Figure 2.8), suggesting that they are not repeating earthquakes.

For the Jan01–Dec12 event pair, we notice that the P first motions have opposite polarities (Figure 2.8). This observation provides strong evidence indicative of these two events being non-repeaters, in addition to the large difference in the S-P differential traveltime. Yet the CC calculated from the conventional MF method is very high (0.91) in comparison to 0.52 when the MFMC approach is used.

We also tested three other different filters with much wider frequency passbands for both event pairs and the results are comparable (Figure 2.25). We find that the performance of MFMC is less sensitive to the choice of filter passband. Taking the Jan01-Dec12 event pair for example, the conventional method-derived CC can vary by as much as 0.13 when different filters are applied (e.g., a narrow band-pass filter of 3-9 Hz vs. a wider one of 1-20 Hz; Figure 2.25). In contrast, the MFMC-derived CC is very stable with a difference of only 0.02.



**Figure 2.8.** Normalized waveforms of three events occurred in the Blanco Fracture Zone. Red, blue, and dark green lines correspond to the 3-channel (i.e., HH1, HH2, and HHZ) waveforms of the Jan01, Oct08, and Dec12 events, respectively. The waveforms of both Oct08 and Dec12 events are aligned at the location of the best match with the Jan01 event based on the conventional MF method. The reference time (i.e., time 0) is defined by the P arrival of the Jan01 event. Panels on the left display the zoom-in P waves without filtering. Panels in the middle exhibit the full unfiltered waveforms with mean removed. Color shaded segments indicate the S-P differential traveltimes. Panels on the right show the band-pass filtered waveforms. The templates from Jan01 event are superposed at the location of the best match on the basis of conventional method for the Oct08 and Dec12 events.

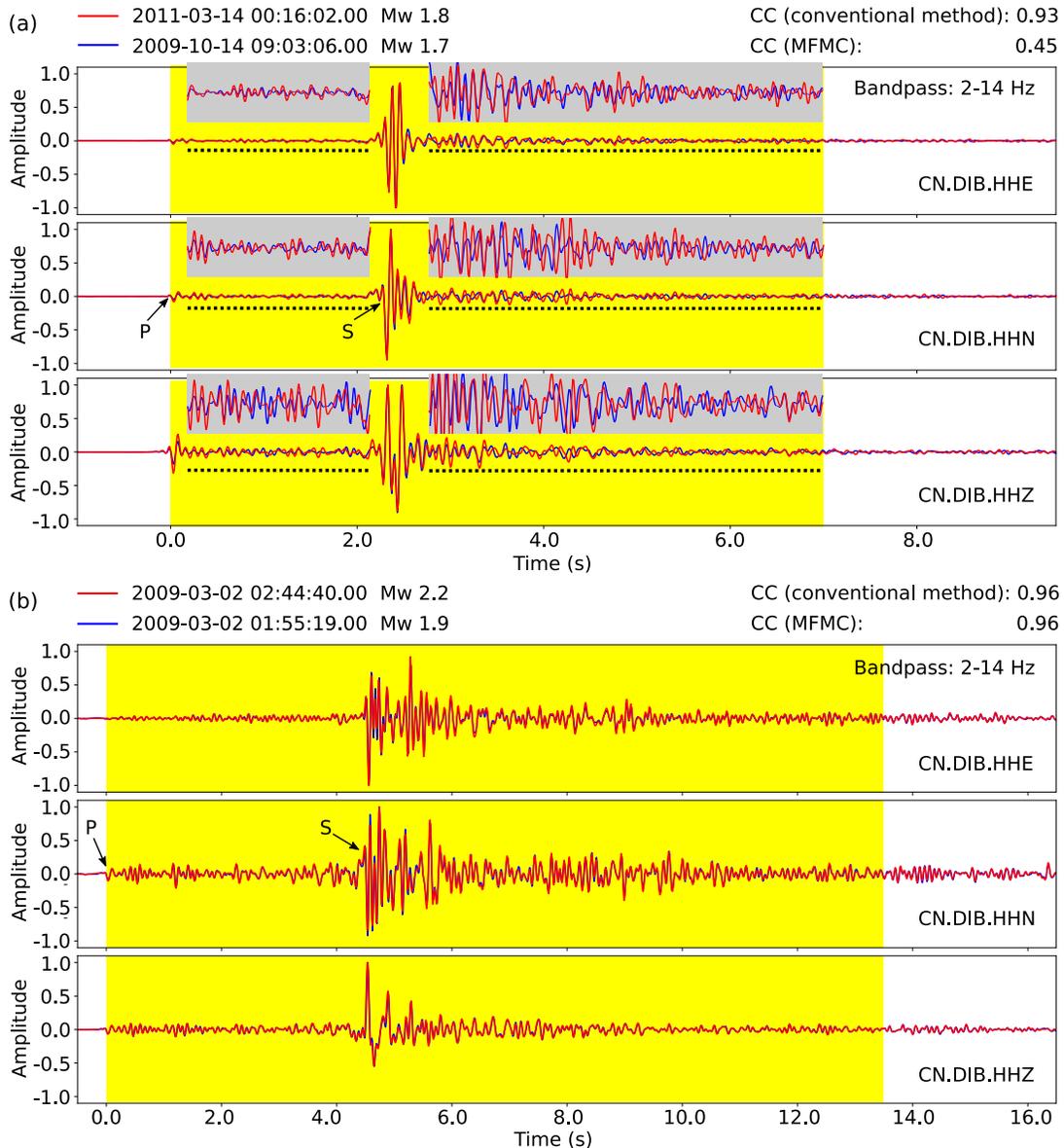
## 2.6.2 Real Examples from the Queen Charlotte Plate Boundary

The BFZ examples above manifest the capability of MFMC in differentiating distant sources. However, for closely spaced neighbouring earthquakes with very similar P and S waveforms and S-P differential traveltimes (e.g., Cheng et al., 2007), it is always a challenge to distinguish if they are just neighbouring events or true repeaters unless a very dense local array is available. Here, we demonstrate a solution to this common but challenging problem by applying the MFMC to two event pairs taken from the repeating earthquake catalog compiled by Hayward and Bostock (2017) for the Queen Charlotte plate boundary mainly on the basis of conventional waveform cross-correlation using data from a single station (Figure 2.9).

We process the waveforms of both pairs with the same filter (2–14 Hz) as that used in Hayward and Bostock (2017). For the first event pair, the conventional method-derived CC is as high as 0.93, in contrast to the low CC of 0.45 calculated with the MFMC. This pair has almost identical P and S waveforms, yet the scattered phases both before and after the S phase are visibly distinct (especially on the vertical channel, gray shaded boxes in Figure 2.9a). Given the extremely low noise level (as indicated by the nearly flat waveforms before the P arrival), the waveform discrepancy between the scattered phases can arise from either changes of scatters along the ray-paths or, more likely, source location difference. Therefore, the first event pair could be neighbouring events rather than repeating earthquakes, in good agreement with the MFMC-derived low CC. For true repeaters, like the second pair, each segment of the waveforms matches almost perfectly when the background noise level is low (Figure 2.9b). In such a case, the CC values derived from both methods are high (0.96).

The demonstrated examples clearly show that the conventional MF method works well in detecting the existence of an event, but lacks the resolution to determine how far the template and matched events are. From this perspective, we point out that the MFMC method should not necessarily be used as a replacement of the conventional MF method. Depending on the purpose of the research, the conventional MF may be the preferred tool if the primary objective is to detect as many events as possible regardless of being repeaters or not. On the other hand, if the objective is to robustly identify potential repeating earthquakes with limited station coverage, then the MFMC method can work much more effectively. We note that the value of MFMC

technique is to effectively decrease the likelihood of the misclassification although it cannot totally eliminate the chance of misclassification.



**Figure 2.9.** Normalized waveforms of two event pairs occurred in the Queen Charlotte plate boundary zone. (a) An example of a misidentified repeating pair. Gray shaded boxes show the amplified waveforms for the segments before and after the S phase (marked by black dotted lines). (b) An example of true repeating pair. For both (a) and (b), yellow shaded segments indicate the windows used for CC calculation.

## 2.7 Conclusions

The main conclusions from this study are summarized as follows:

(1) The value of CC can be affected by many factors, including the window length of the waveform template, the frequency band of applied digital filter, and the existence of large-amplitude phase(s). As a result, a variety of different CC thresholds, ranging from 0.75 to 0.98 depending on different choices of data processing parameters, were used in previous studies to identify repeating earthquakes. It is important to bear in mind that the computed CC should not be taken at face value. The repeaters identified with a given CC threshold need to be verified by other means.

(2) To avoid any ambiguity due to the choice of different operational parameters and to optimize the performance of MF, we propose generic rules for selecting the length of waveform template and the frequency passband of the digital filter used in data processing. Specifically, a dynamic window length dependent on the differential traveltimes between the P and S phases can properly account for the increasing wave train with source-receiver distance. Comparing the spectra of signal and pre-signal noise before performing cross-correlation is of great importance in designing an optimal filter with the highest signal-to-noise ratio. Our results show that the performance of MF is significantly improved when an optimal filter is applied before cross-correlation.

(3) As evident from both synthetic and real earthquake examples, the conventional MF method (i.e., the one-segment cross-correlation approach) is very sensitive to large-amplitude phase(s) in the calculation window. Consequently, the CC value obtained with the conventional approach is not a proper indicator for inter-event spatial separation. A high CC value does not necessarily imply a small hypocentre separation, especially when one (or few) large-amplitude phase(s) dominates the waveform in the calculation window. The situation becomes worse if the template event is deep (e.g.,  $\geq 10$  km). With data from multiple stations, the traditional approach seems to work much better in constraining the horizontal separation but not always the vertical. Significant improvement can be reached if the station distribution is optimized. A key conclusion from these synthetic and real earthquake experiments is that inferring inter-event separation based on the conventional MF method, especially with data from only a single channel or station, may not be reliable.

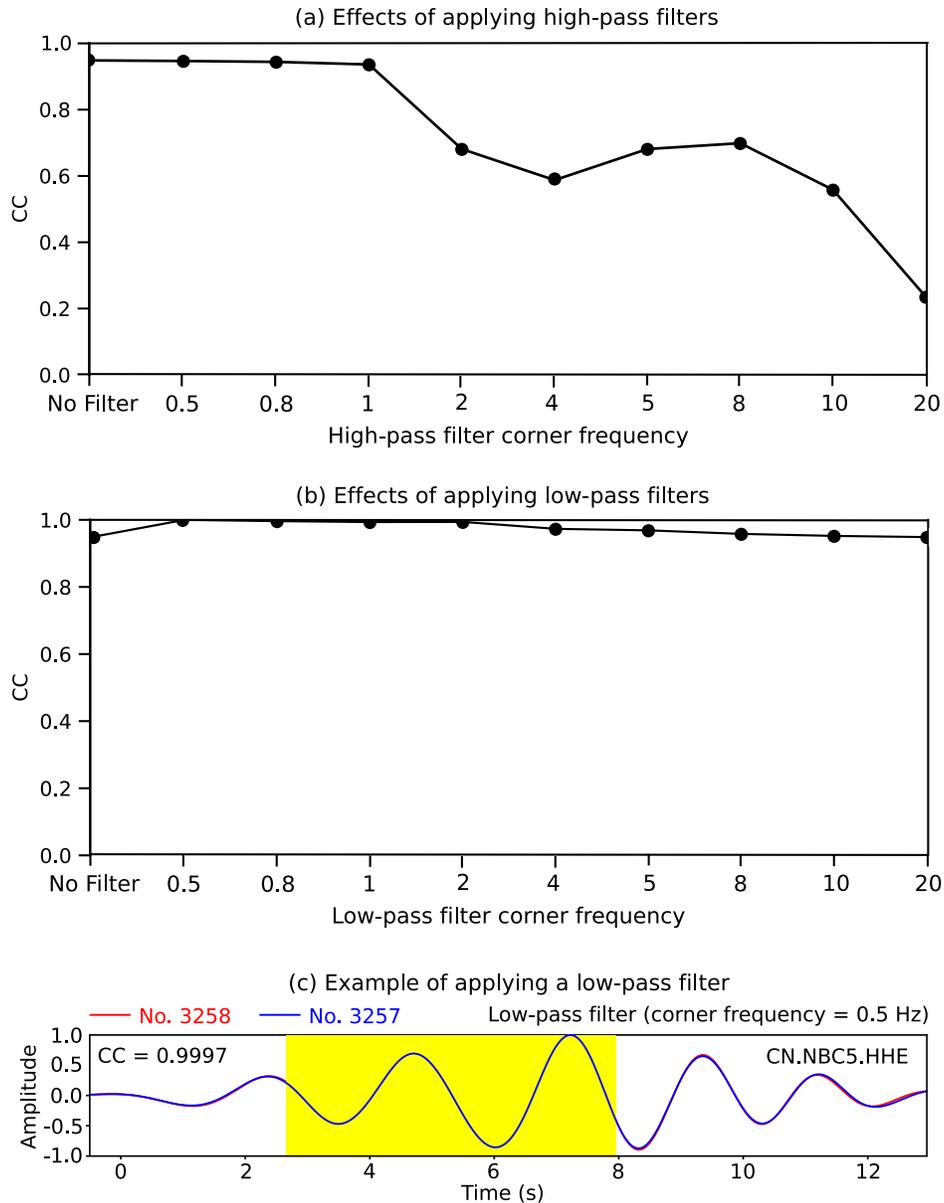
(4) To reliably identify potential repeating earthquakes, we propose the MFMC approach that properly incorporates the contributions from low-amplitude phases in the calculation of CC. The new method is very sensitive to subtle waveform changes caused by minor hypocentre shifts. A slight location difference can lead to an apparent drop in the final CC value. For horizontal inter-event separation, including multiple stations may help reduce ambiguity; for separation in the vertical direction, using a single channel can perform similarly well as using an array. When the template event is deep, the new method becomes slightly less effective in recognizing vertical changes but still outperforms the conventional method.

(5) Finally, the MFMC can be applied directly to continuous waveforms to search for potential repeating earthquakes. It can also be used as a post-analysis tool to quickly extract potential repeaters from an existing catalog or from the output of the conventional MF. It is not a replacement of the conventional MF if the primary objective is to detect as many events as possible regardless of being repeaters or not.

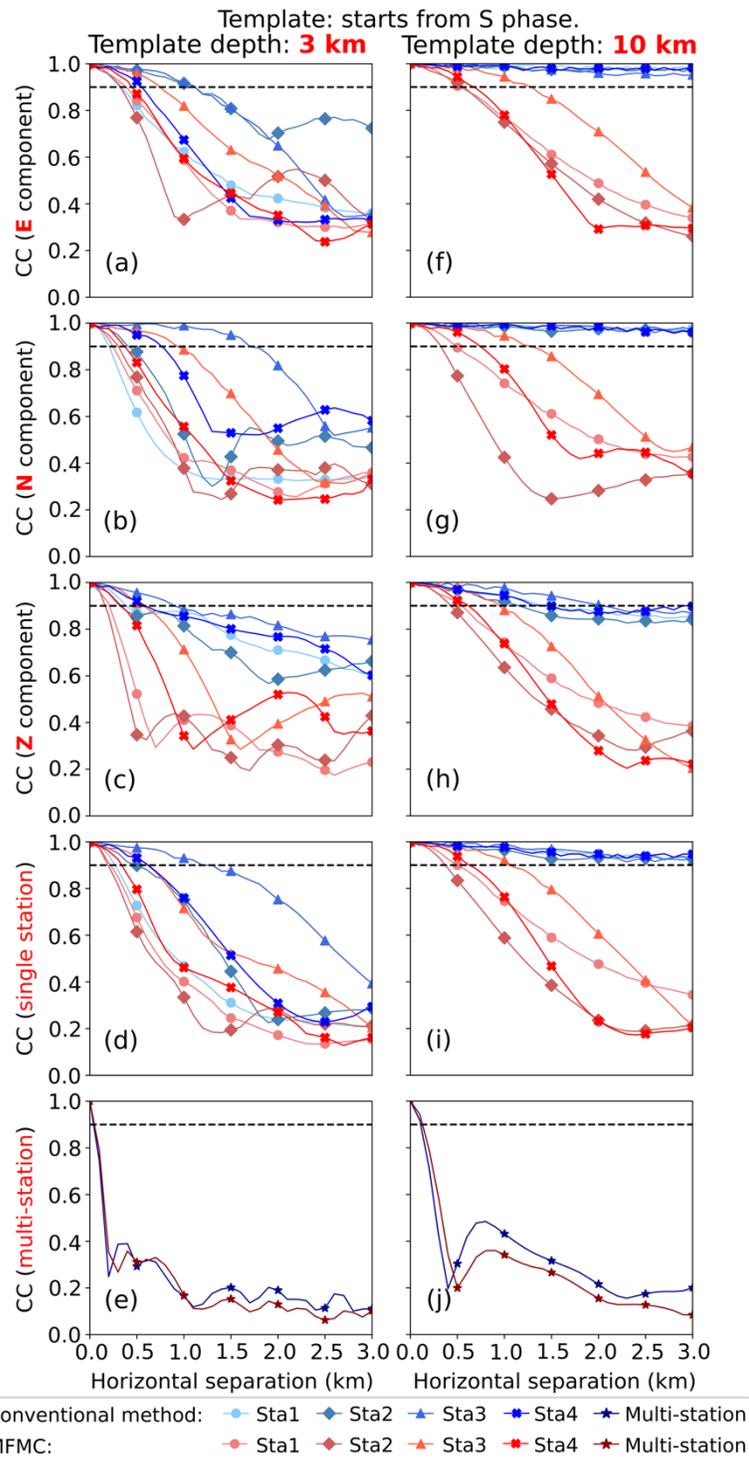
## **2.8 Acknowledgements**

We thank the Associate Editor, Calum J. Chamberlain, and an anonymous reviewer for their constructive comments. We are grateful to Lupei Zhu for providing the FK code that is used in generating synthetic seismograms. Insightful discussions with Jianlong Yuan, Stan Dosso, Ryan Visser, Fengzhou Tan, Ramin Mohammad Hosseini Dokht, Bei Wang, Douglas A. Dodge and Jean-Luc Got are much appreciated. Real waveform data used in this study were downloaded from the Incorporated Research Institutions for Seismology (<http://ds.iris.edu/ds/nodes/dmc/>, last accessed May 2020). Seismic data are processed with Obspy (Beyreuther et al., 2010; <https://github.com/obspy/obspy/>). Figures are made with Matplotlib (Hunter, 2007; <https://matplotlib.org>) and GeoMapApp (Ryan et al., 2009; [www.geomapapp.org](http://www.geomapapp.org)). This study is partially supported by a University of Victoria Fellowship (DG), the Induced Seismicity Research Project of NRCan, and a NSERC Discovery Grant (HK). This paper is NRCan contribution 20200096.

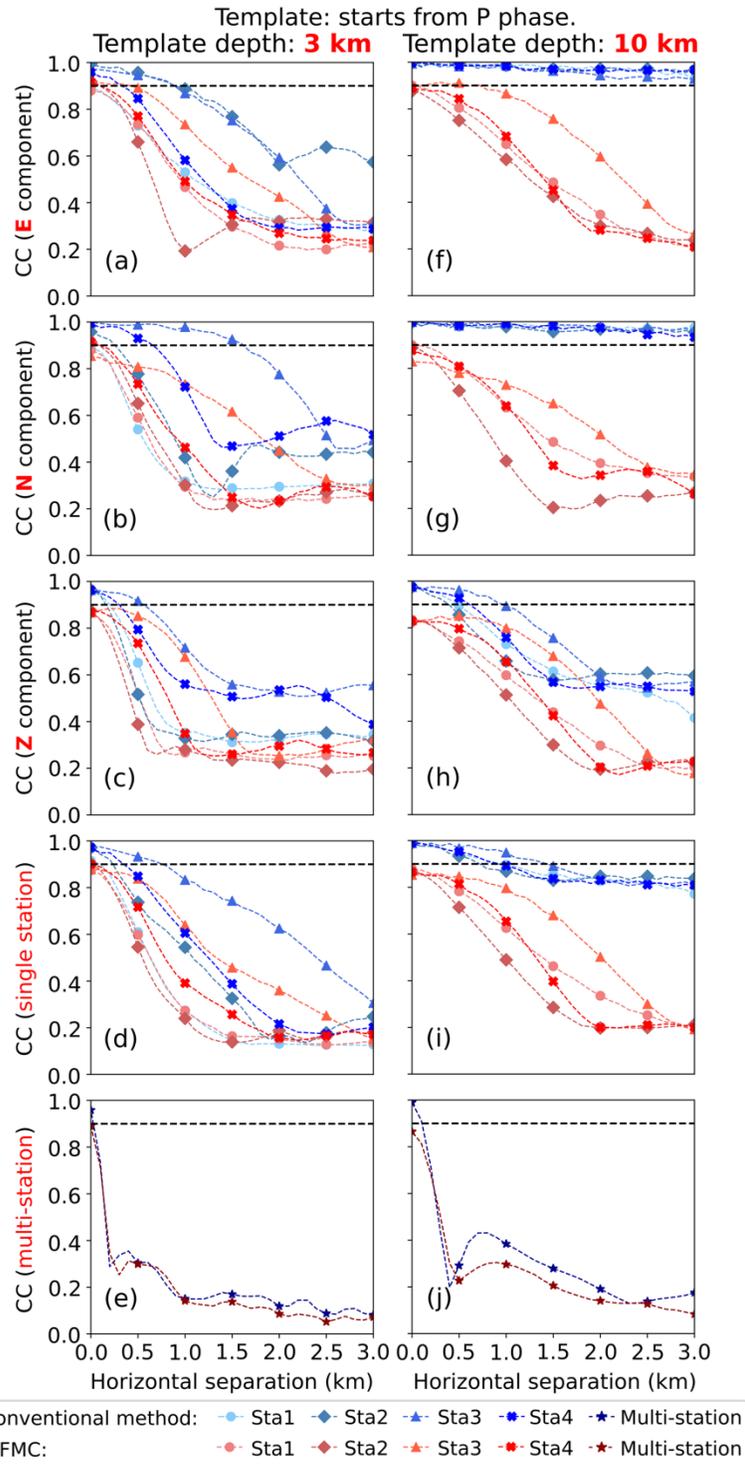
## 2.9 Supporting Information



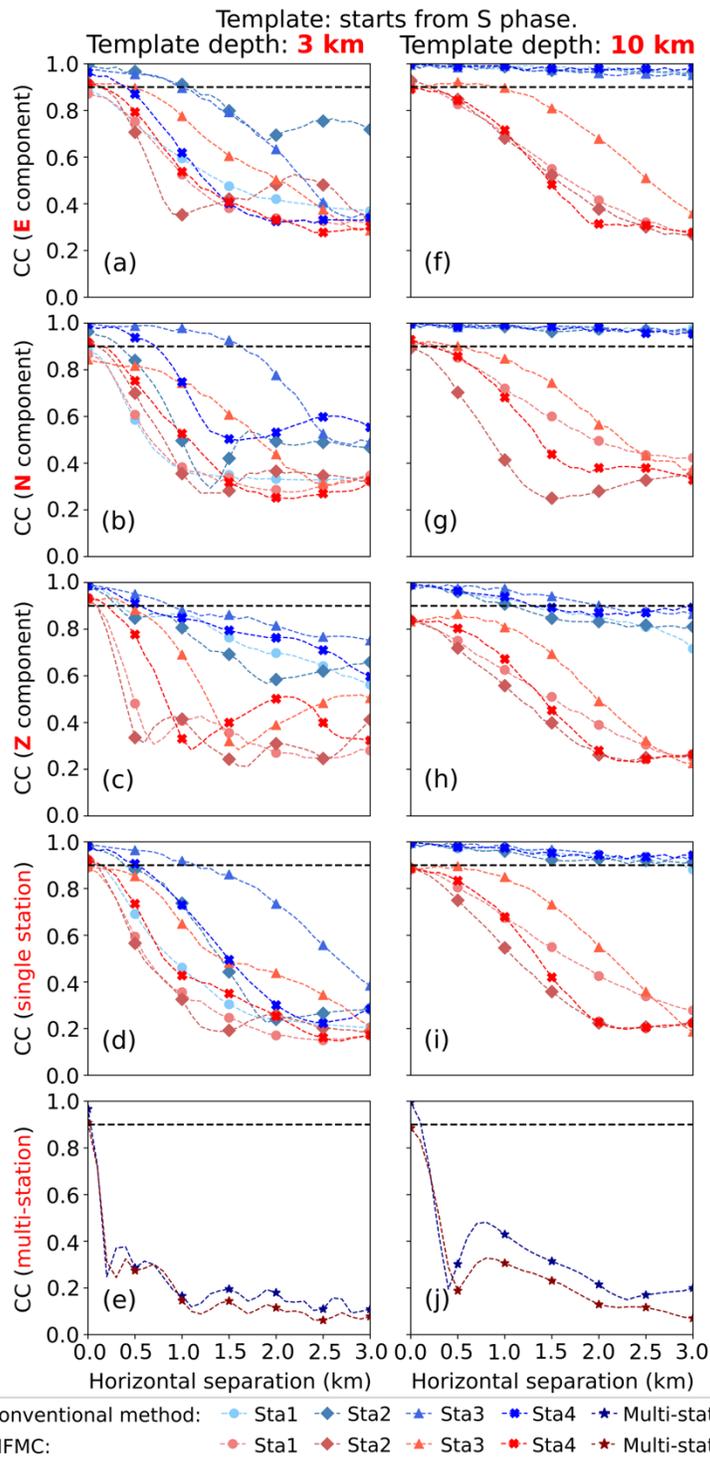
**Figure 2.10.** Effect of high-pass (a) and low-pass (b) filtering on the result of waveform cross-correlation. The corner frequencies of the band-pass filters in Table 2.1 are used as the cut-off frequencies of the high- and low-pass filters. (c) Example of applying a low-pass filter. Symbols are the same as that in Figure 2.1a.



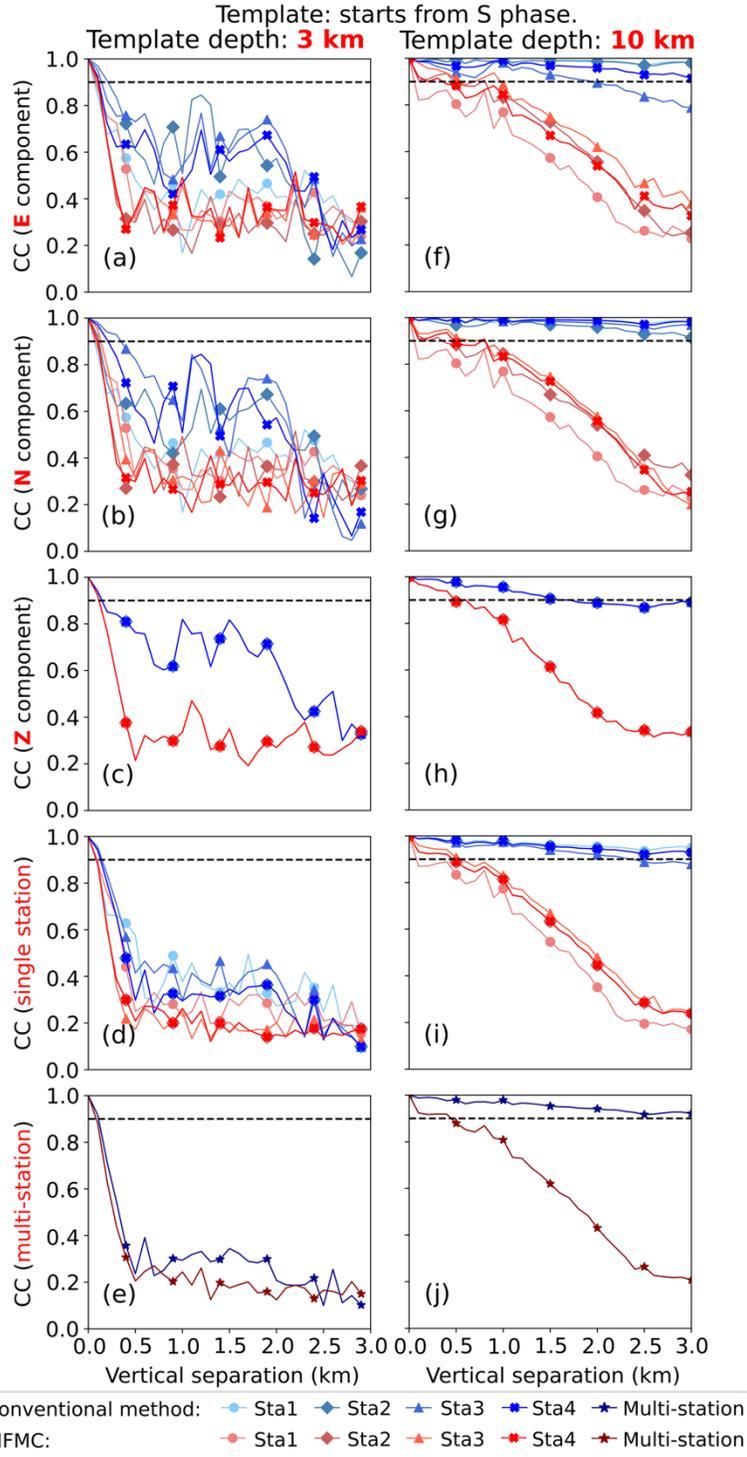
**Figure 2.11.** CC variation due to horizontal inter-event separation with a template starting from the S phase. Detailed configuration of the experiment, including the assumed strike-slip focal mechanism, is shown in Figure 2.2. Symbols and layout are the same as that in Figure 2.6.



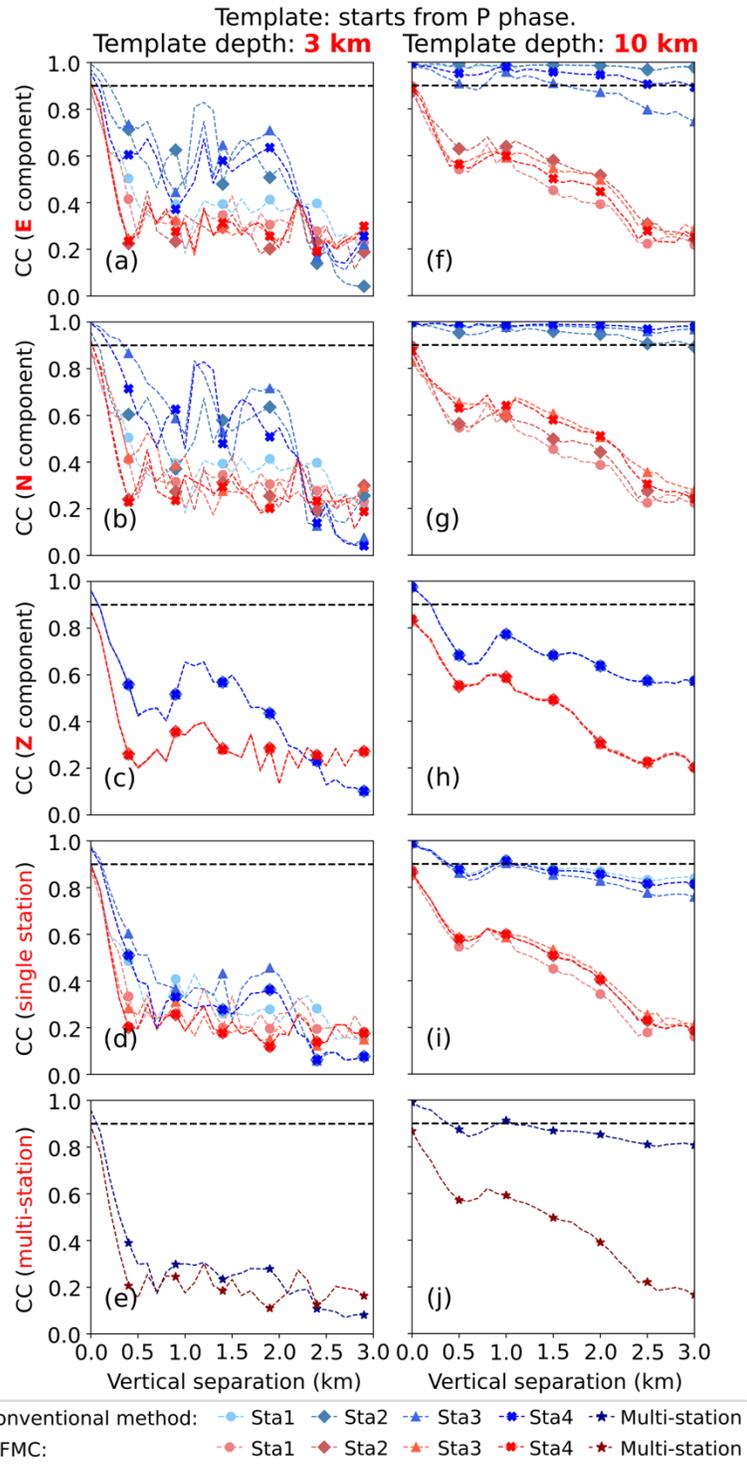
**Figure 2.12.** CC variation due to horizontal inter-event separation with a template starting from the P phase. Experiment setup is the same as that in Figure 2.11. For each assumed separation, background noise equivalent to a SNR of 2 is added in the test with 100 realizations for each channel of each station. Only the average results of 100 realizations is displayed here. Layout is the same as that in Figure 2.6.



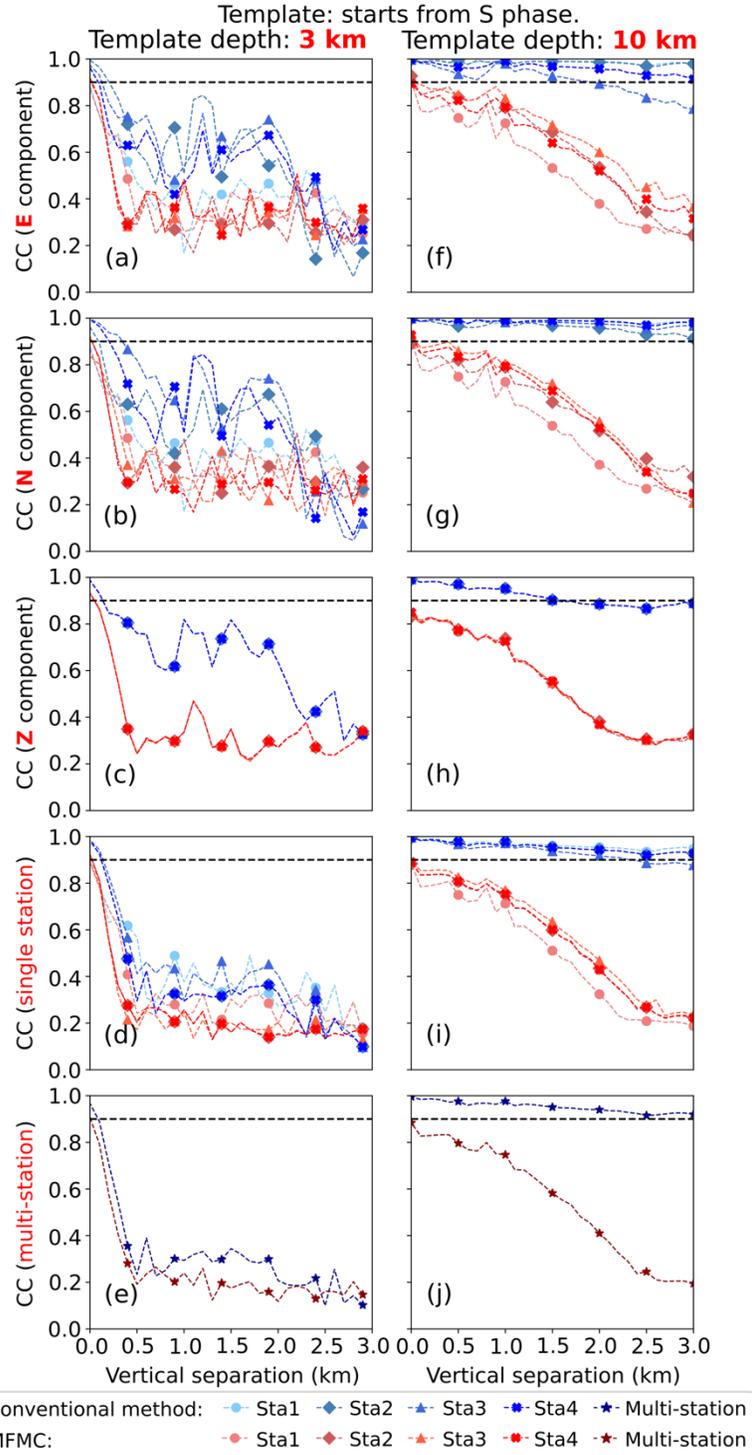
**Figure 2.13.** CC variation due to horizontal inter-event separation with a template starting from the S phase. Experiment setup is the same as that in Figure 2.11. Background noise is added in each test following the exactly same setup as described in Figure 2.12. Layout is the same as that in Figure 2.6.



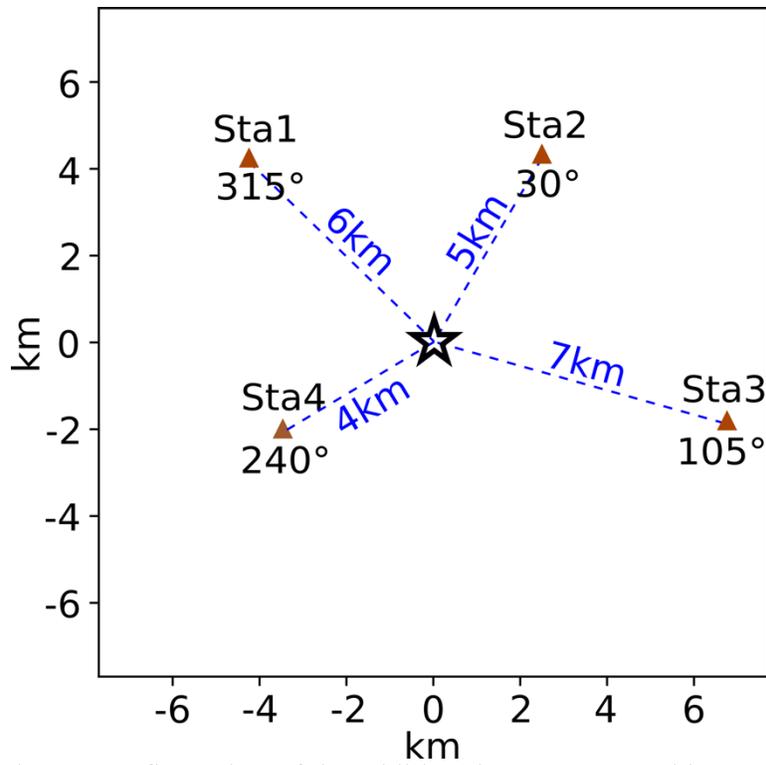
**Figure 2.14.** CC variation due to vertical inter-event separation with a template starting from the S phase. Experiment setup is the same as that in Figure 2.11. Symbols and layout are the same as that in Figure 2.6.



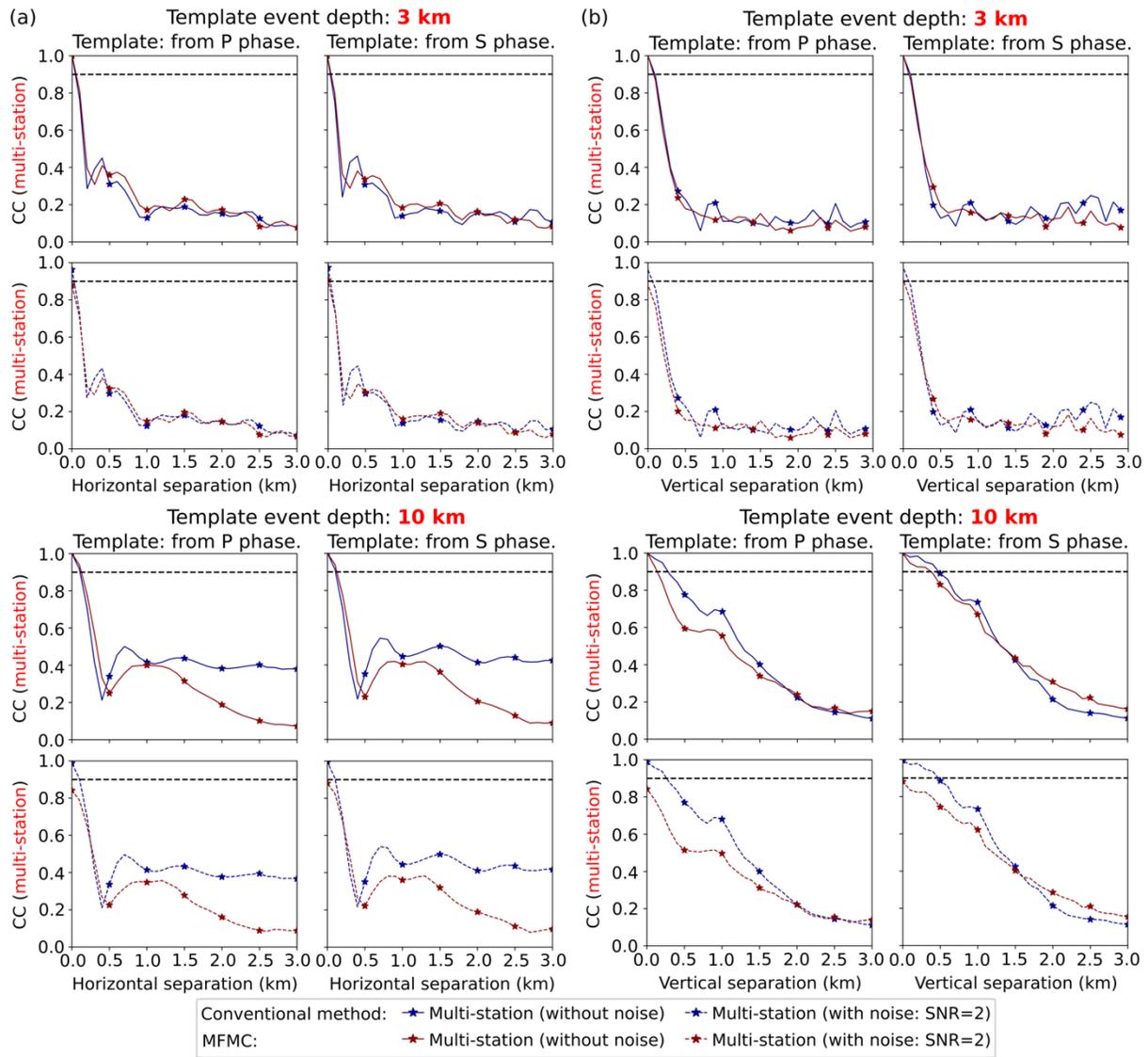
**Figure 2.15.** CC variation due to vertical inter-event separation with a template starting from the P phase. Experiment setup is the same as that in Figure 2.11. Background noise is added in each test following the exactly same setup as described in Figure 2.12. Layout is the same as that in Figure 2.6.



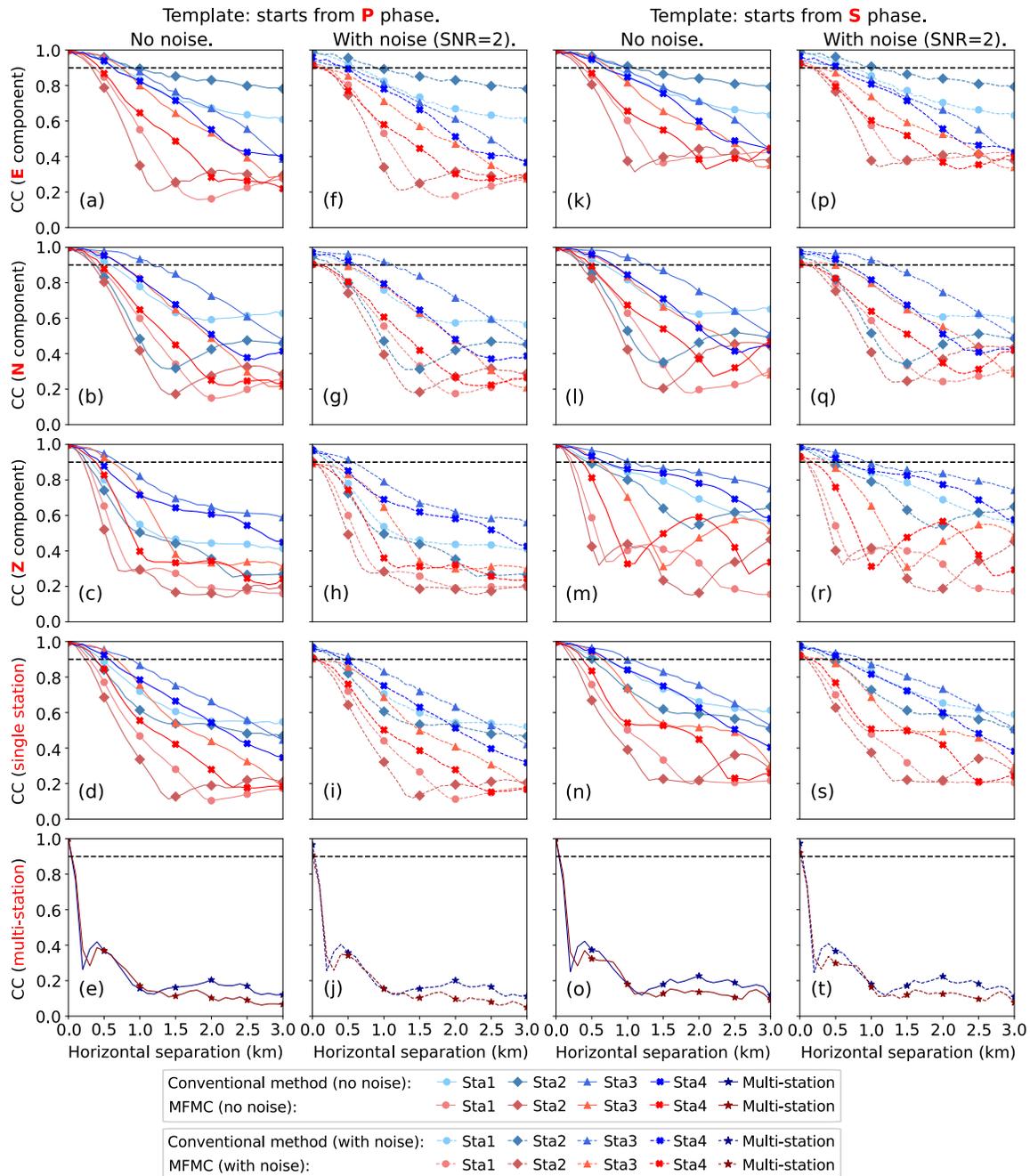
**Figure 2.16.** CC variation due to vertical inter-event separation with a template starting from the S phase. Experiment setup is the same as that in Figure 2.11. Background noise is added in each test following the exactly same setup as described in Figure 2.12. Layout is the same as that in Figure 2.6.



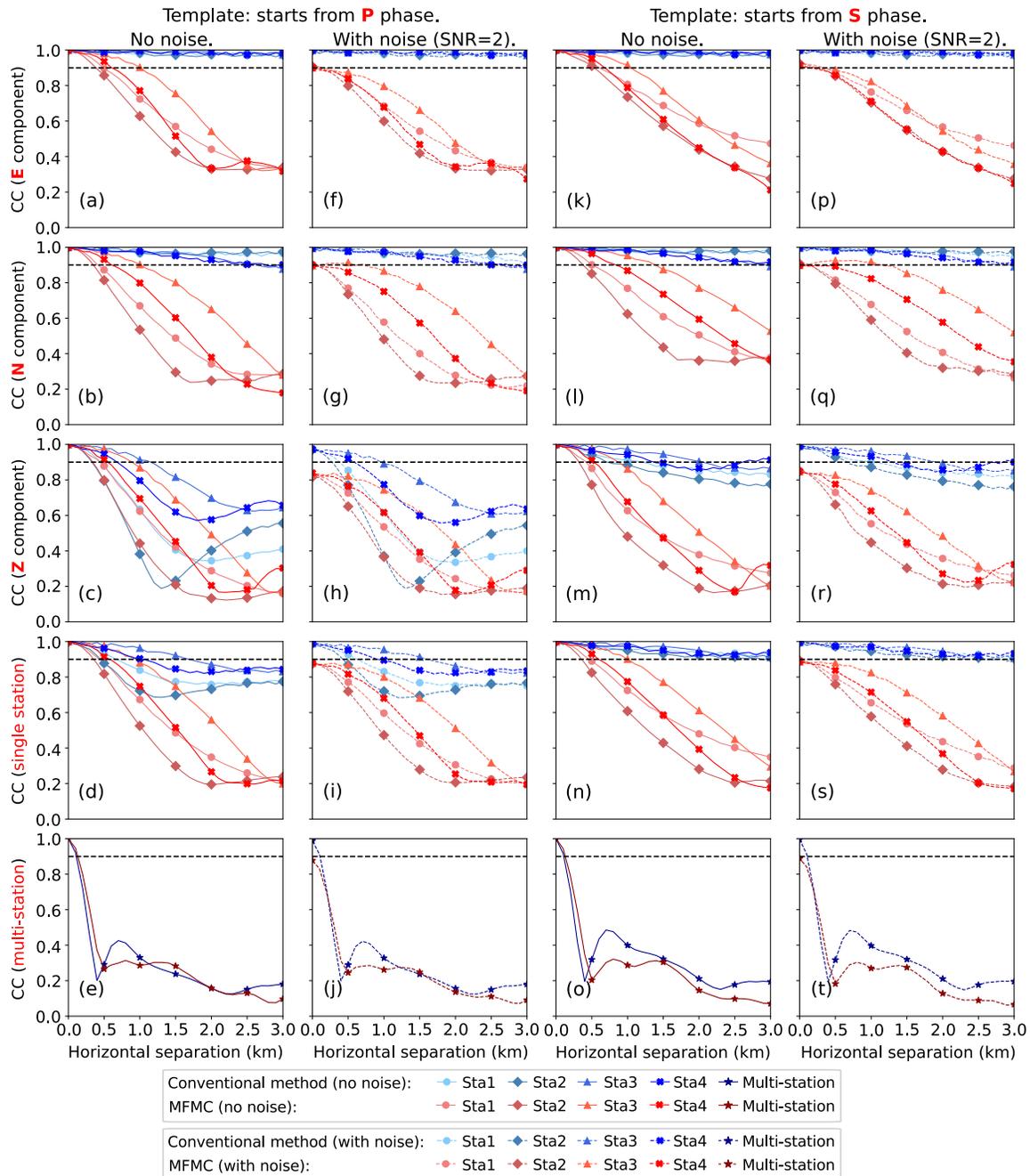
**Figure 2.17.** Experiment configuration of the additional tests presented in Section 2.5.3. Symbols are the same as that in Figure 2.2a.



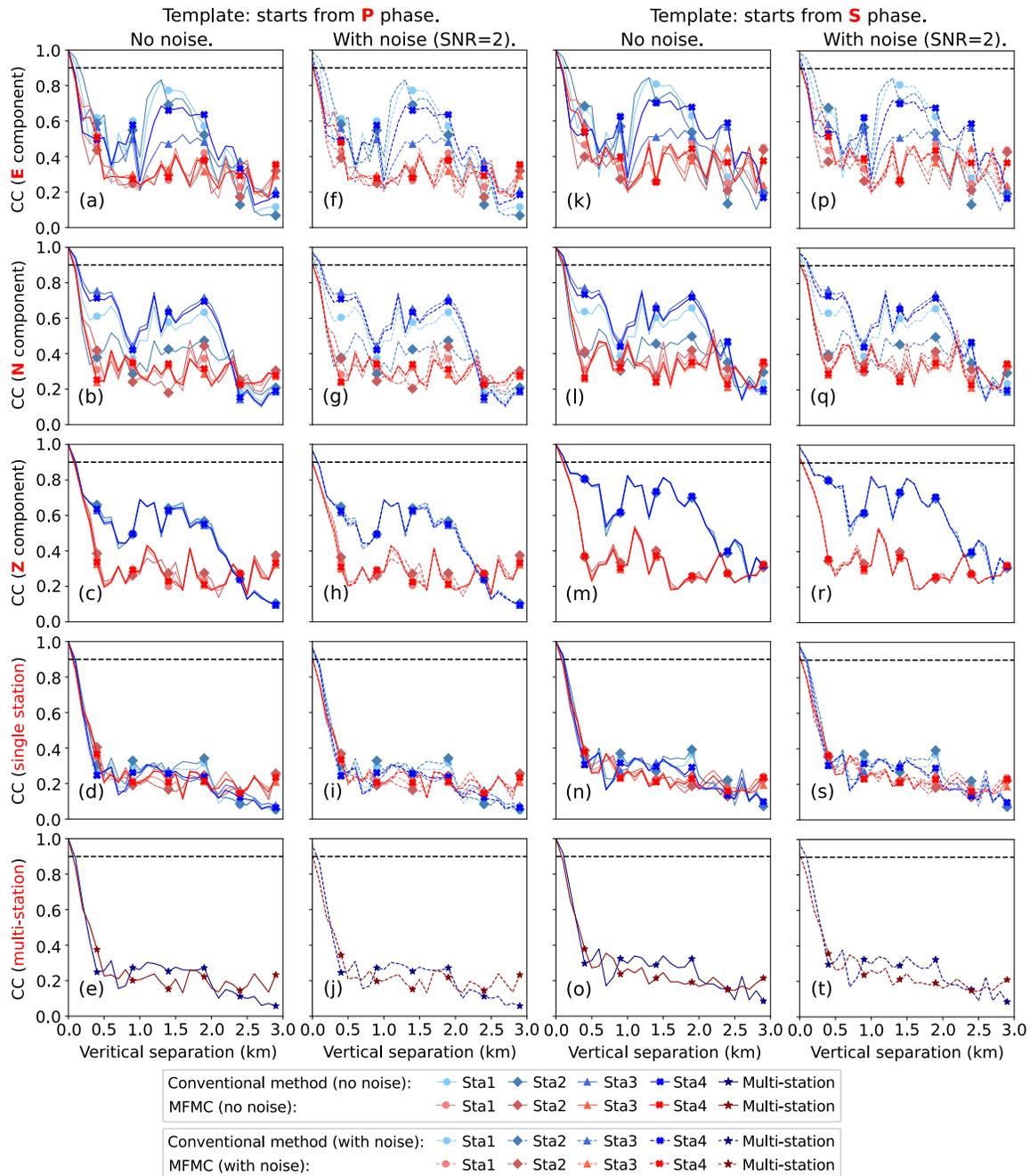
**Figure 2.18.** Experiment results showing the importance of station distribution. (a) CC variation due to horizontal inter-event separation. (b) CC variation due to vertical inter-event separation. Similar to previous tests, background noise is added following the exact same setup as described in Figure 2.12.



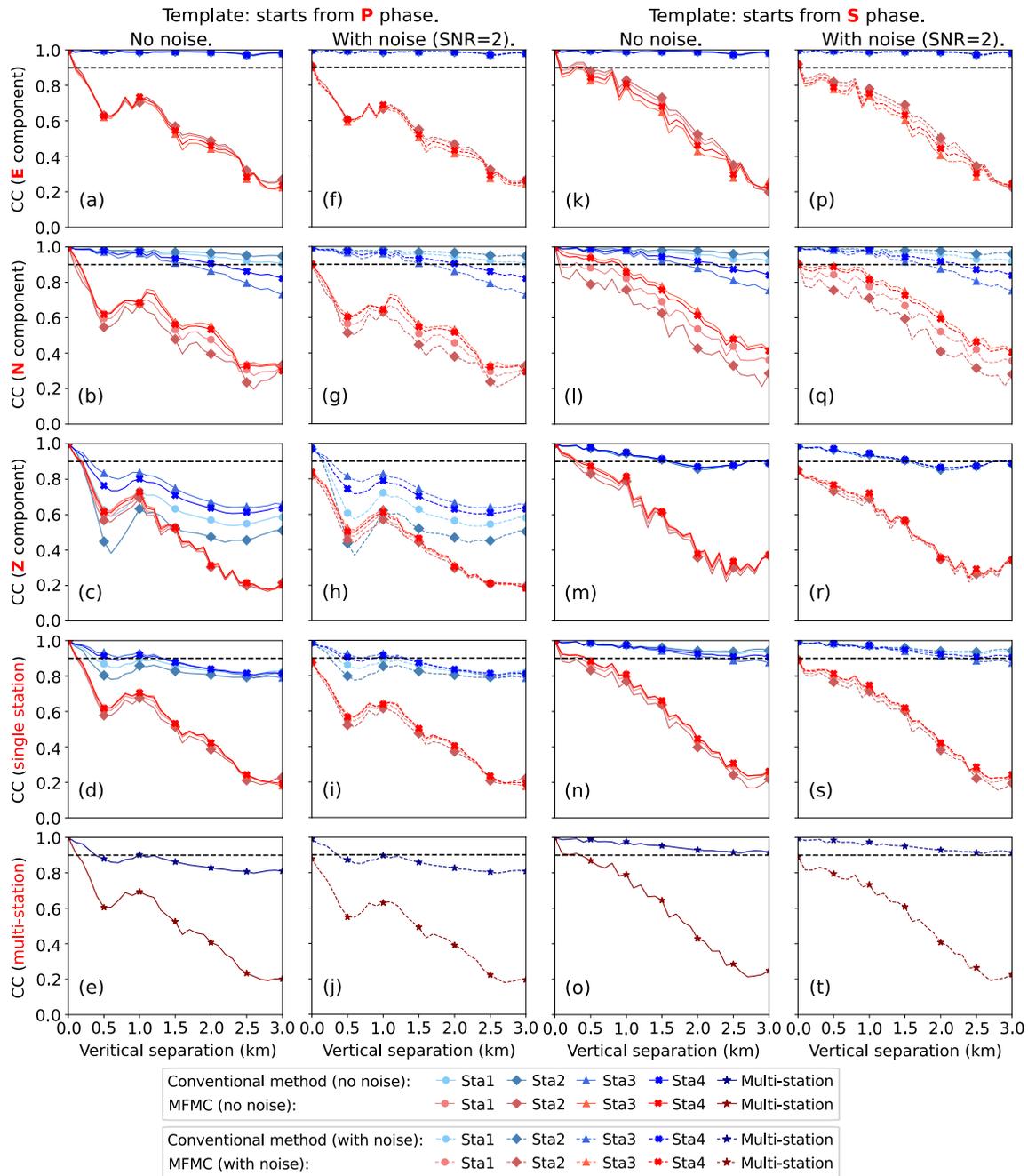
**Figure 2.19.** CC variation due to horizontal inter-event separation with a template at a depth of 3 km. In this experiment, a pure normal-faulting mechanism is assumed. Notice that the CC results of normal and reverse-faulting are the same, hence the results of reversed fault are not displayed. Background noise is added in each test following the exactly same setup as described in Figure 2.12.



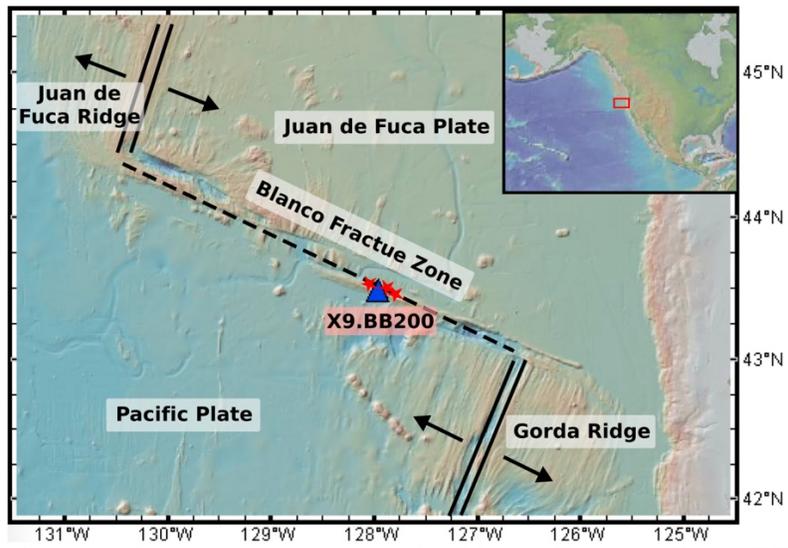
**Figure 2.20.** CC variation due to horizontal inter-event separation with a template at a depth of 10 km. Experiment setup is the same as that in Figure 2.19. Background noise is added in each test following the exactly same setup as described in Figure 2.12.



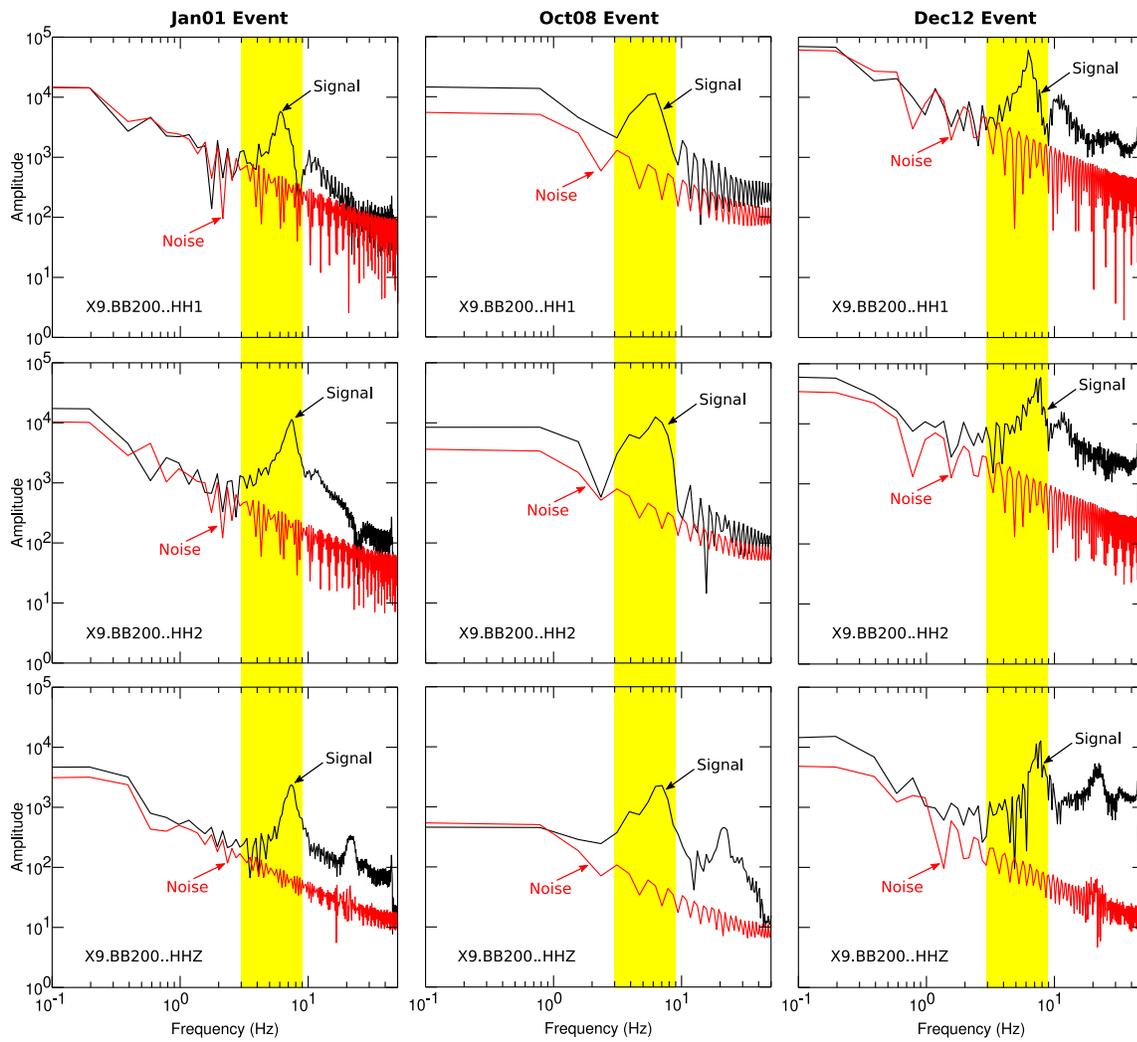
**Figure 2.21.** CC variation due to vertical inter-event separation with a template at a depth of 3 km. Experiment setup is the same as that in Figure 2.19. Background noise is added in each test following the exactly same setup as described in Figure 2.12.



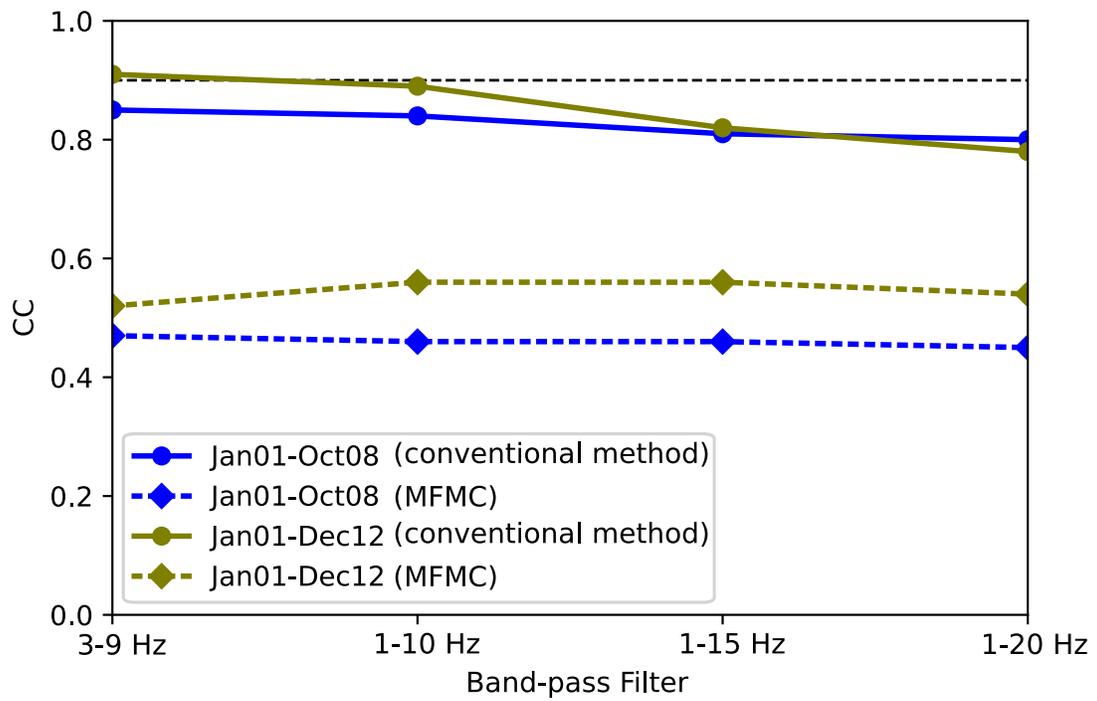
**Figure 2.22.** CC variation due to vertical inter-event separation with a template at a depth of 10 km. Experiment setup is the same as that in Figure 2.19. Background noise is added in each test following the exactly same setup as described in Figure 2.12.



**Figure 2.23.** Map showing the tectonic setting of the Blanco Fracture Zone (BFZ). Blue triangle marks the location of an ocean-bottom seismometer (OBS) station X9.BB200. Red stars represent the approximate locations of three small earthquakes discussed in Section 2.6.1.



**Figure 2.24.** Amplitude spectra of the three BFZ events shown in Figure 2.23 and discussed in Section 2.6.1. Yellow bars indicate the frequency band of dominant seismic energy.



**Figure 2.25.** Effect of different filter passbands on the cross-correlation results of BFZ event pairs. Dashed black line marks the CC threshold of 0.9 for reference.

## **Chapter 3. Misconception of Waveform Similarity in the Identification of Repeating Earthquakes**

### **3.1 Abstract and Plain Language Summary**

#### **3.1.1 Abstract**

Identification of repeating earthquakes (repeaters) usually depends on waveform similarity expressed as the corresponding cross-correlation coefficient (CC) being above a prescribed threshold, typically ranging between 0.70 and 0.98. However, the robustness and effectiveness of such a strategy have never been fully examined. In this study, we examine whether CC is a valid proxy for repeater identification through both synthetic and real earthquake experiments. We reveal that CC is controlled by not only the inter-event distance but also many other factors, including station azimuth, epicentral distance, velocity structure, etc. Consequently, CC lacks resolution in identifying true repeaters. For reliable repeater identification, we should consider both the inter-event overlap and magnitude difference. Specifically, we define an event pair to be true repeaters if their inter-event separation is smaller than the rupture radius of the larger event and their magnitude difference is  $\leq 1$ . Our results imply that a systematic reevaluation of previously identified repeaters and associated interpretations/hypotheses may be important and necessary.

#### **3.1.2 Plain Language Summary**

Repeating earthquakes (repeaters) are events that occur repeatedly on the same fault patch with nearly identical waveforms. They provide important insights into a variety of geophysical subjects such as fault behavior, subsurface structural change, inner core rotation, and nucleation processes of earthquakes and landslides. The identification of repeaters is usually solely based on waveform similarity, but the criteria can vary significantly from one case to another. With both synthetic and real data, we find that waveform similarity is controlled by many factors, in addition to inter-event distance. Therefore, a higher degree of waveform similarity does not necessarily imply a smaller hypocentre separation, and vice versa. Our results suggest that waveform similarity alone is insufficient to reliably identify true repeaters. To more reliably identify repeaters, we should rely on the overlap of the source areas and magnitude difference.

Quantitatively, we define a repeating pair as two events with an inter-event distance smaller than the rupture radius of the larger event and a difference in magnitude of no more than 1. Our results imply that previously-identified repeaters and associated interpretations/hypotheses may benefit from a systematic reexamination.

### **3.2 Introduction**

Repeating earthquakes (repeaters) are events that recurrently rupture the same fault patch with the same focal mechanisms, often characterized by nearly identical waveforms (Uchida and Bürgmann, 2019). These events are of great importance in many aspects of geophysics, such as monitoring subtle temporal changes of crustal properties (e.g., Poupinet et al., 1984; Schaff and Beroza, 2004; Sawazaki et al., 2015; Pacheco et al., 2017) and oceanic temperature (Wu et al., 2020), estimating fault creep (e.g., Nadeau and Johnson, 1998; Uchida et al., 2003, 2006; Matsubara et al., 2005; Yu, 2013; Materna et al., 2018), investigating inner core rotation (e.g., Li and Richards, 2003; Zhang et al., 2005, 2008; Tkalčić et al., 2013), evaluating the precision of earthquake locations (e.g., Li and Richards, 2003; Meier et al., 2004; Schaff and Richards, 2011; Jiang et al., 2014), and providing insights into the nucleation process of earthquakes (Kato et al., 2012; Kato and Nakagawa, 2014; Meng et al., 2015; Huang and Meng, 2018) and landslides (Yamada et al., 2016).

There are three ways of identifying repeaters. The most straightforward approach is to consider the overlap of the source areas (e.g., Waldhauser and Ellsworth, 2002). However, this physics-based approach requires a dense near-source array to precisely estimate the inter-event distance and source dimensions. The second, and perhaps the most popular way relies on waveform similarity between an event pair (64% of the cases compiled in Table 3.1 in Section 3.9 Supporting Information). The degree of similarity can be expressed by the value of cross-correlation coefficient (CC) in the time domain or, equivalently, the coherence in the frequency domain. The employed CC threshold is somewhat arbitrary, ranging from as low as 0.70 for regions with sparse network coverage to as high as 0.98 for areas with denser instrumentation (Table 3.1). The third way is a hybrid approach with complementary criteria in addition to CC (36% of cases compiled in Table 3.1). The additional criteria differ significantly among various studies. Some studies examine the time interval between the event pair, the difference between

the two event's S-P differential times, and/or the magnitude difference to reduce misidentification, while others directly verify the repeaters by relocating their hypocentres (Table 3.1).

With the increasing computing power, detecting repeaters through waveform similarity has become a routine process in seismology (e.g., Tepp, 2018; Chamberlain et al., 2018, 2020). However, a growing body of literature suggests that similar waveforms may only imply source proximity (Cheng et al., 2007; Ellsworth and Bulut, 2018) and/or similar focal mechanisms (Kilb and Rubin, 2002), but not necessarily repeating ruptures. One of the most striking examples is in central Japan where Cheng et al. (2007) analyze high-quality data from an extremely dense seismic array of 56 stations in a small area of  $\sim 20$  km by 20 km and found no true repeaters among 807 very similar events ( $CC \geq 0.80$ ). Their results highlight the serious reliability issue of using only waveform similarity in repeater identification.

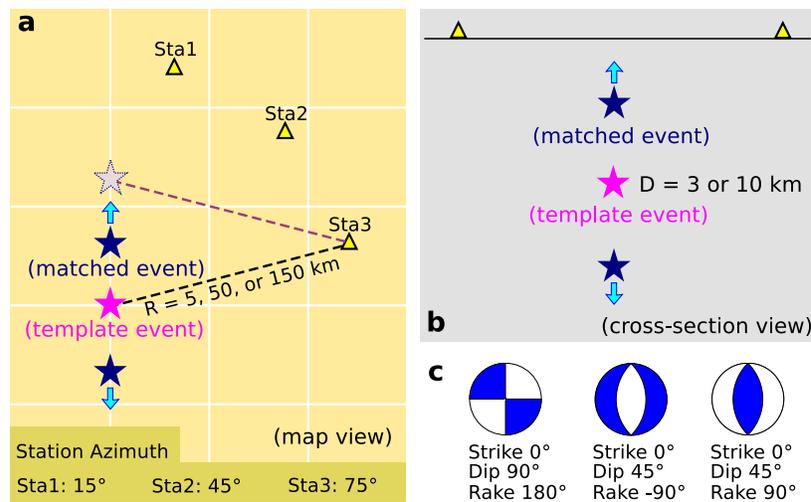
The focus of this study, therefore, is to systematically investigate the deficiency of waveform similarity in repeater identification. We first examine how the CC varies with inter-event separation and uncover the overlooked factors through a large number of synthetic experiments. We then illustrate that waveform similarity indeed lacks the resolution to determine whether two events are true repeaters or not using a dense local borehole array data in Parkfield, California. To more reliably identify repeating earthquakes, a physics-based approach that considers not only the source area overlap but also the magnitude difference should be used. We illustrate the effectiveness of this approach using events that occurred in the Fox Creek area, Alberta, Canada, where earthquake source parameters are well constrained by local stations.

### **3.3 Synthetic Experiments**

#### **3.3.1 Synthetic Experiment Setup**

Figure 3.1a illustrates the configuration of our synthetic experiments. We place one event (the template event) at the centre of an array. Then we incrementally shift the other event (the matched event with the same focal mechanism) by 0.1 km each time in either a north-south (Figure 3.1a) or vertical direction (Figure 3.1b). Due to the symmetrical setup of our experiment, it is sufficient to focus our investigation on stations in the first quadrant. For the demonstration

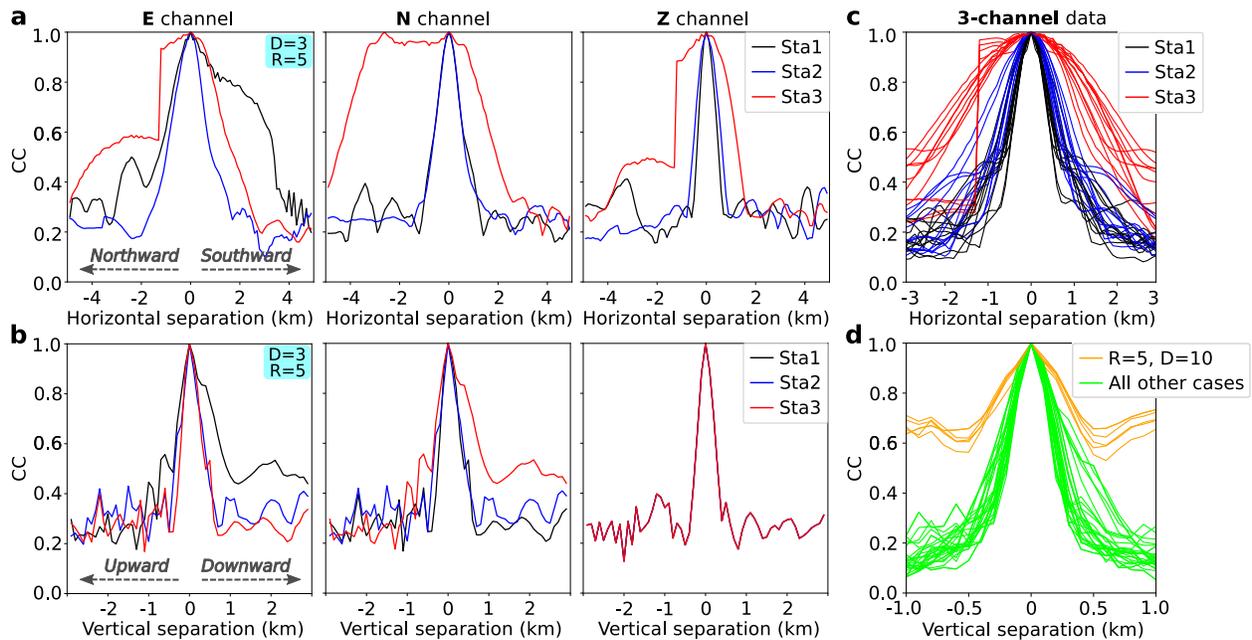
purpose, we show three stations with azimuths of  $15^\circ$ ,  $45^\circ$ , and  $75^\circ$  in Figure 3.1. To explore the effects from various parameters, we conduct our tests with two different focal depths ( $D = 3$  and  $10$  km), three representative focal mechanisms (strike-slip, normal, and reverse faulting), three different epicentral distances ( $R = 5, 50,$  and  $150$  km, Figure 3.1), and two different data choices (single-channel and single-station (3-channel) data). In total, 72 scenarios that correspond to either horizontal ( $\pm 5$  km) or vertical ( $\pm 3$  km) spatial separation are studied in detail. The synthetic seismograms are generated and processed exactly the same way as an earlier study (Gao and Kao, 2020). The CC is calculated with the newly developed match filtering with multi-segment cross-correlation (MFMC) technique (Gao and Kao, 2020). The technical details about the MFMC calculation are presented in Section 3.9.1.



**Figure 3.1.** Schematic diagram showing the setup of the synthetic experiments. (a) For horizontal inter-event separation, two sources are placed along a line trending N–S separated by a short distance. The template event (fuchsia star) is fixed in the middle while the matched event (navy blue star) moves away from the template event in both directions. Stations (triangles) are placed at three different epicentral distances ( $R=5, 50,$  or  $150$  km). The gray star marks the location of the matched event such that one of the stations (Sta3 in this case) is of equal distance to both the template and matched events. (b) For vertical inter-event separation, the template event is placed at two different depths ( $D=3$  or  $10$  km) with the matched event moving up or down. (c) Three different types of focal mechanisms are used in the calculation of synthetic seismograms.

### 3.3.2 Constraining Inter-event Separation Using Single-channel Data

Single-channel CC has been used in numerous previous studies to infer the existence of repeaters (Table 3.1), thus we first examine how the CC varies with source separation using single-channel (i.e., E, N, or Z) data. In Figure 3.2a and 3.2b, we present the results of a representative case, namely, a strike-slip earthquake (template event) at the depth of 3 km with a station 5 km away from the epicentre.



**Figure 3.2.** Results of our synthetic experiment showing CC variation as a function of horizontal (a and c) and vertical (b and d) inter-event separation. The setup of sources and receivers is depicted in Figure 3.1. (a) and (b) correspond to a representative case with single-channel data, whereas (c) and (d) compile all test results with single-station (3-channel) data. Individual test results are presented in Figures 3.7 and 3.8 (in Section 3.9 Supporting Information). For (a) and (c), positive and negative horizontal separations indicate the matched event is shifted to the south and north, respectively; for (b) and (d), positive and negative vertical separations indicate the matched event is shifted down or up, respectively.

For horizontal inter-event separation, our results indicate that single-channel waveforms can have very different sensitivities (Figure 3.2a). In general, the CC value decreases nonlinearly with increasing hypocentral separation. It quickly drops from 1 when the two sources are perfectly co-located to  $<0.5$  when the pair is  $\sim 1$  km apart. Beyond that, the CC curves appear to fluctuate between 0.2 and 0.4 without a clear monotonic trend. This implies that using the CC

value to constrain the difference between two nearby hypocentres may not be ideal once the separation is on the order of kilometers, similar to the results of Schaff (2010) and Han et al. (2014) although they use the traditional cross-correlation method.

Another important point in Figure 3.2a is that the CC value may be strongly affected by the combined effect of focal mechanism and relative position between the source and station. This effect is best illustrated by Station 3 as the inter-event distance increases. For all 3 channels, the CC value decreases when the matched event shifts northward from 0 to  $-1.3$  km. Once past the  $-1.3$  km mark, the CC value has a sudden drop on both E and Z channels but continues to increase on the N channel. This unexpected result happens when Station 3 is located very close to one of the assumed nodal planes (Figure 3.1a). As the matched event shifts northward, Station 3's relative position moves across the nodal plane and therefore causes polarity reversal on the Z and E channels. When the inter-event separation reaches  $-2.6$  km, Station 3 is nearly of equal distance to both the template and matched events (Figure 3.1a), leading to identical waveforms on the N channel but reversed shapes on the other two channels (Figure 3.5 in Section 3.9, Supporting Information). Consequently, the final (maximum) CC values would be 1 for the N channel (taken when the two P phases coincide) and  $\sim 0.5$ – $0.6$  for the Z and E channels (taken when the two P phases are offset by half a cycle), even though the two events are 2.6 km apart. We have tested other types of focal mechanisms (pure normal or thrust-faulting) and the profound effect remains (Figure 3.6).

Unlike the cases of horizontal separation, the CC curves obtained with different channels and stations overall show similar trends when the two sources are vertically separated (Figure 3.2b), hinting that using the CC value to constrain the vertical inter-event separation is probably independent of data channel and station azimuth. Especially for the vertical channel, stations with different azimuths can have identical sensitivities to the inter-event separation when the focal mechanism is pure strike-slip (Figure 3.2b, right panel). Notice that the CC curves derived from the E and N channels of Station 1 are identical to those from the N and E channels of Station 3, respectively (Figure 3.2b), due to the symmetrical station location on the focal sphere (Figure 3.1a). Results of these tests once again suggest that a smaller CC does not necessarily represent a larger separation once the vertical separation exceeds a certain threshold ( $\sim 0.5$  km). We also find that results from different focal mechanisms are comparable (Figure 3.6). Last but not the least, the CC value generally drops much faster with increasing vertical source separation

(Figure 3.2a vs. 3.2b) as a result of more minor discrepancies between waveforms. In other words, the CC seems to be much more sensitive to vertical source shift than horizontal, consistent with the results of Schaff (2010) and Han et al. (2014).

The simple tests above demonstrate that, in addition to inter-event distance, CC can be strongly dependent on the specific channel used, combined effect of focal mechanism and relative position between the source and station, and source separation direction (horizontal vs. vertical). Results from other single-channel CC cases (i.e., different combinations of focal depth, epicentral distance, and focal mechanism) are overall comparable.

### **3.3.3 Constraining Inter-event Separation Using Single-station (3-channel) Data**

If data from all three channels are included, we find that the CC sensitivity to source separation increases dramatically for the cases of horizontal separation (e.g., Figures 3.2a vs. 3.7a) but insignificantly for those of vertical separation (e.g., Figures 3.2b vs. 3.8a). For a given horizontal separation, Stations 1 and 3 tend to have the lowest and highest CC values, respectively (Figures 3.2c and 3.7), strongly suggesting that station azimuth is an important factor that cannot be overlooked. In contrast, the influences of focal depth, epicentral distance, and source focal mechanisms seem to be limited (Figure 3.7). Our results indicate that a station approximately in line with the template and matched events can be more effective in delineating the inter-event separation (e.g., Station 1 in our case, Figure 3.1a).

Overall, the computed CC is very sensitive to vertical inter-event separation with the only exception when the source is deep and the station is close (e.g.,  $D=10$  km and  $R=5$  km, Figures 3.2d and 3.8). For a close station ( $R=5$  km) and a shallow source ( $D=3$  km), even a very small (0.2 km) vertical separation can lead to a dramatic drop of CC to  $<0.8$  (Figure 3.8a and 3.8c), but the sensitivity gets worse when the source is deeper (Figure 3.8b and 3.8d). This is mainly a velocity structure effect caused by smaller seismic velocity variation at deep depths. In other words, the CC sensitivity would become higher when the corresponding velocity structure (and therefore the observed waveforms) are more complicated. An important observation to point out is that the CC is very sensitive to vertical inter-event separations when the epicentral distance is large (e.g.,  $R = 50$  or  $150$  km), regardless of the focal depth (Figure 3.8e-l). This is opposite to what is expected for earthquake depth determination as seismic phase arrival times at distant

stations usually have less depth constraint. It turns out that waveforms at distant stations can have better developed depth phases (i.e., seismic phases reflected from either the free surface or Moho). Consequently, a subtle change of source depth may lead to a significant waveform difference and therefore an apparent CC drop.

Therefore, our experiments in this section further demonstrate that CC can be affected by the number of channels used, station azimuth, velocity structure, and epicentral distance.

### **3.3.4 Constraining Inter-event Separation Using Multi-station Data**

For areas with excellent network coverage, it is common to use a minimum of two stations (usually only the vertical channel) for identifying repeaters (Table 3.1). The majority of prior work (Table 3.1) calculate CC separately for each station. This approach essentially uses more stations with different azimuths and/or epicentral distances but may not necessarily improve the sensitivity if all available stations happen to be the ones with lower sensitivities (Figure 3.2c and 3.2d). An alternative way is to calculate the CC simultaneously across the network (Yao et al., 2017; Chaves et al., 2020) which includes the constraint of traveltimes moveout. In such a case, the computed CC can be extremely sensitive to hypocentre difference (Gao and Kao, 2020). We refrain from investigating the multi-station scenario for two reasons. First, the multi-station CC is not widely adopted by previous studies (2 out of 58 cases, Table 3.1 in Section 3.9). Second, and perhaps more importantly, the CC sensitivity is known to be strongly affected by network geometry (Chamberlain and Townend, 2018; Gao and Kao, 2020), and thus no general/common rules can be inferred objectively.

In summary, our synthetic experiments reveal that CC is a very complex function of many aforementioned factors. A higher CC value does not necessarily represent a smaller inter-event separation, and vice versa. Therefore, in contrast to the conventional wisdom, our synthetic results indicate that CC is not a robust indicator of two events being true repeaters or not.

## **3.4 Verification With Real Earthquake Examples**

Using real earthquake waveforms to examine the spatial resolution of waveform similarity is a challenging task since we often do not know the ground truth. As summarized in Table 3.1 (in

Section 3.9 Supporting Information), the majority of earlier studies use data only from one or two stations/channels to identify repeaters. With very limited data, we cannot confidently verify whether the cataloged repeaters are true repeaters or not. For the cases where the data are relatively richer, either the repeater catalog is not explicitly documented in the publication or the seismic waveform data are not ideal (publicly unavailable or contaminated by a high level of noise).

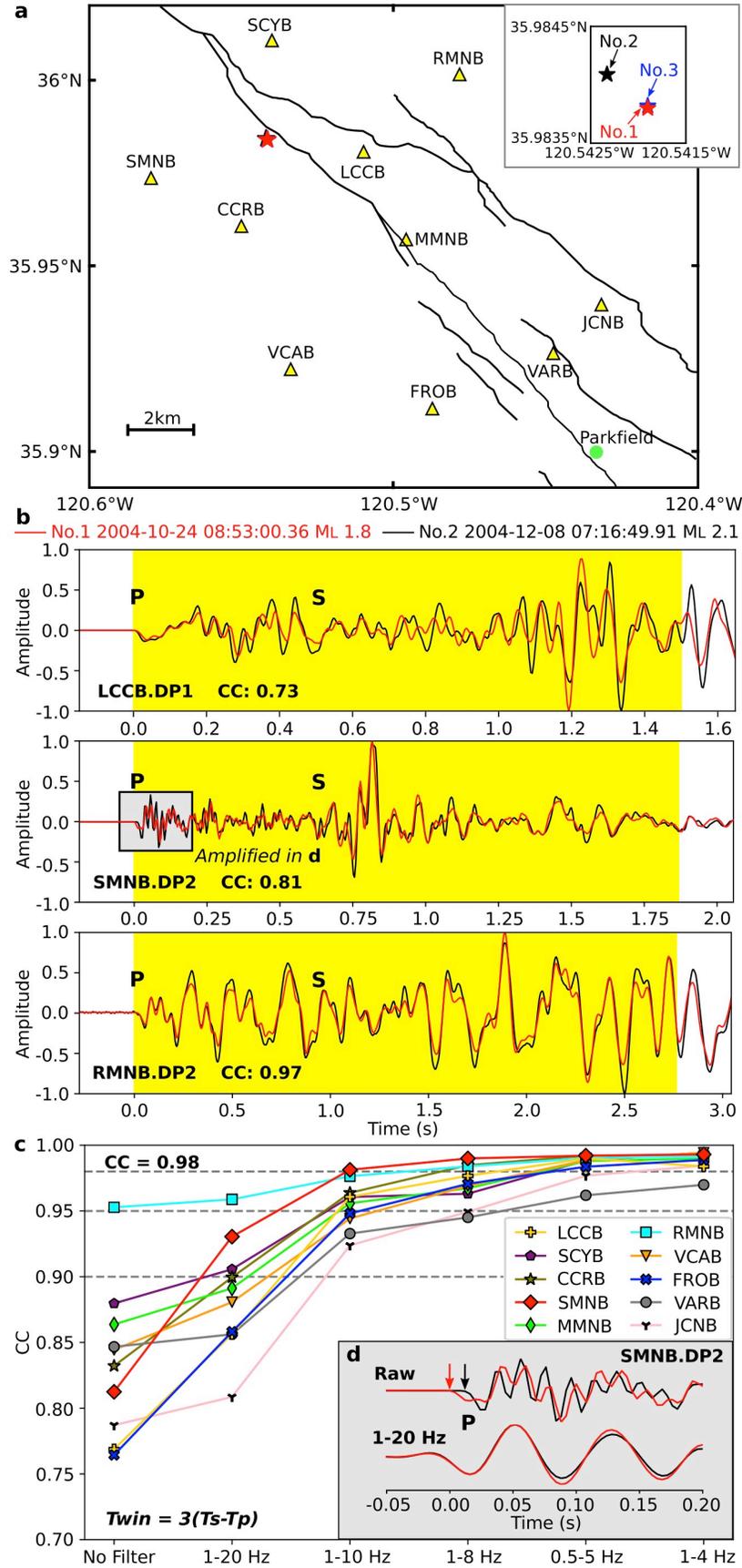
Therefore, we limit our search to the Parkfield area, California, where both a repeater catalog and high-quality waveform data are publicly available (e.g., Dreger et al., 2007; Abercrombie, 2014). The High-Resolution Seismic Network (HRSN, Figure 3.3a) is a dense local array of borehole seismometers deployed in this area and operated by the Berkeley Seismological Laboratory. The HRSN waveform data generally have exceptionally high signal-to-noise ratio (SNR) and hence are ideal for the purpose of this study to verify whether waveform similarity is a good proxy of repeater identification. Here we select three events (No. 1-3, Figure 3.3a) from two well studied repeating earthquake clusters in Parkfield (e.g., Dreger et al., 2007; Chen et al., 2010; Zoback et al., 2011; Abercrombie, 2014; Kim et al., 2016). Among them, events No. 1 and 3 belong to the same cluster with similar source areas while event No. 2 occurred on a neighbouring fault patch. All the selected events have well-determined source parameters (including location and stress drop, Dreger et al., 2007; Abercrombie, 2014) and nearly noise-free waveform recordings from by a large number of stations ( $\geq 9$ ) with different station azimuths and epicentral distances, providing a valuable opportunity to thoroughly examine the spatial resolution of waveform similarity with high confidence.

### **3.4.1 CC between Non-repeaters**

We first calculate the CC between non-repeaters, i.e., No. 1 and 2. We only use data from stations nearly free from noise contamination, as indicated by the flat waveforms before the P wave arrival (one example is shown in Figure 3.3b). The most striking result of our analysis is that the CC derived from unfiltered 3-channel waveforms indeed differ significantly among different stations, ranging from 0.76 to just above 0.95 (Figure 3.3c). Such a wide CC range is consistent with the inference from our synthetic tests that the CC can be severely affected by station azimuth and/or source-receiver position/path even under nearly noise-free circumstances.

Additionally, the CC may be further affected by local structures of velocity discontinuities as this region is featured by complex fault zones (Figure 3.3a). Our study reveals that the waveforms of non-repeaters can show apparent difference (top panel in Figure 3.3b), slight difference (middle panel in Figure 3.3b) or little difference (bottom panel in Figure 3.3b) at different stations. For any given station, the CC values of different channels can be either similar or different (Figure 3.9). Together, waveform similarity indeed lacks the resolution to decide whether two events are repeaters or not.

Because nearly all prior works practically identify repeaters through filtered waveforms for the purpose of mitigating the noise impact, we next examine the effects of commonly used band-pass filters (Table 3.2 in Section 3.9 Supporting Information). Our results indicate that the CC obtained from different stations all show a clear increasing trend when the passband becomes narrower (Figures 3.3c and 3.9), consistent with the findings from synthetic experiments of an earlier study (Baisch et al., 2008). Especially for the very narrow but very popular 1-4 Hz band-pass filter used by many previous studies (Table 3.2), 9 out of 10 stations have  $CC > 0.98$  (Figure 3.3c), which is the highest CC threshold used in the literature in selecting repeaters (Table 3.1). This simple experiment highlights the often-overlooked fact that filtering could remove the important frequency content in the signal that distinguishes the physical separation of the two events, in addition to reducing the unwanted noise. For example, even a very wide band-pass filter (1-20 Hz) would remove the very high frequency signal with poor similarity and thus lead to very similar waveforms as shown in Figure 3.3d. What makes it worse is that filtering would change both the shape and width of the P wave and therefore make the subtle difference in the S-P differential traveltime (0.012s in Figure 3.3d) unresolvable, effectively throwing away the most critical information on the relative distance between the two sources. The results (from nearly noise-free waveform experiments) here strongly imply that band-pass filtering of waveforms may lead to misidentification of repeaters if the selection criterion is solely based on waveform similarity. We also tested the effect of template window length ( $T_{win}$ ) associated with different filters (Table 3.2) in calculating CC (Section 3.9.1) and the results are comparable (Figure 3.10). Two examples of how filtering increases the waveform similarity at close and distant stations, respectively, are presented in Figures 3.11 and 3.12 for reference.



**Figure 3.3.** CC test results with real earthquake data. (a) Map showing the distribution of earthquake epicentres (colored stars) and HRSN seismograph stations (yellow triangles). Black lines denote the surface traces of the San Andreas Fault system. The town of Parkfield is shown as a green dot. Insert shows the zoom-in locations of events No. 1-3. (b) Examples of normalized unfiltered waveforms of two events that have been verified to be non-repeaters (No. 1 and No. 2), aligned at the S wave arrival. The highlighted segment indicates the window of dynamic length (see Section 3.9.2) used for CC calculation. The gray box in the middle panel outlines the waveform segment amplified in (d). (c) Effects of filtering on the CC values between events No. 1 and No. 2 determined with individual single-station (3-channel) data and dynamic window lengths. (d) An example of waveform change due to filtering. Red and black arrows mark the P wave onset of events No.1 and No.2, respectively. The slight time difference (0.012 s) between the two arrows is overlooked after band-pass filtering between 1 and 20 Hz.

### 3.4.2 CC between True Repeaters

Interestingly, we also find that, for true repeaters like events No. 1 and 3, the CC value obtained from different stations still differ significantly from each other (Figure 3.13). The unfiltered waveforms can be nearly identical at one station (Figure 3.14a) but also can be of minor difference at another station even with nearly no noise (zoom-in box in Figure 3.14b). Without noise contamination, the waveform difference between true repeaters may arise from the variability of the rupture process (such as the slight difference in earthquake initiation point) (Uchida, 2019) and/or seismic velocity change (e.g., Poupinet et al., 1984; Sawazaki et al., 2015; Pacheco et al., 2017). With the band-pass filters applied, the waveform discrepancy overall becomes much smaller as indicated by the increasing CC values (Figure 3.13). This is similar to, but less dramatic than the case of non-repeaters.

Taken together, non-repeaters indeed can have very similar waveforms (bottom panel in Figure 3.3b) while the waveforms of true repeaters may display minor difference (Figure 3.14b), further strengthening the argument made from the large number of synthetic experiments. In contrast to the traditional view, our observations suggest that waveform similarity is not a good proxy for repeater identification, especially with band-pass filters applied.

### 3.5 A Physics-based Definition of Repeaters

Since repeaters physically represent the re-rupturing of the same fault patch, the most fundamental concern of two events being repeaters or not is whether their ruptures significantly overlap with each other. Implicitly, the magnitudes (or equivalently the rupture sizes) of the two events should be comparable. Therefore, we define two events to be repeaters if the hypocentre of the smaller event fall within the rupture area of the bigger event and the magnitude difference is no more than 1, similar to the criteria used by some earlier studies (e.g., Zhao and Peng, 2009). One advantage of this definition is that it does not involve the source dimension of the smaller event which may be difficult to precisely estimate. It is worth noting that there will always be a cut-off threshold no matter what criteria are employed. In our case, we only consider events with significant source-area overlap and comparable magnitudes to be repeaters.

There are different ways in seismology to characterize the rupture area of an earthquake source (Stein and Wysession, 2009). For small earthquakes, the most common, and perhaps the easiest, method is to estimate the equivalent rupture radius (ERR), which is defined as the radius of a circle whose area is the same as the source rupture area. It is worth noting that the ERR has been used as an additional requirement for repeater identification in some earlier studies (Table 3.1 in Section 3.9 Supporting Information) and very often used to estimate fault dimension and/or slip rate of repeaters (e.g., Li et al., 2007, 2011; Ma et al., 2014; Hatakeyama et al., 2017; Mesimeri and Karakostas, 2018). Although the true source area of a small earthquake may not be exactly circular, the circular-fault assumption should be a sufficient first-order approximation (e.g., Mesimeri and Karakostas, 2018; Cheng et al., 2007; Hatakeyama et al., 2017; Igarashi, 2020; Kelly et al., 2013; Lengline' and Marsan, 2009; Li et al., 2007, 2011; Ma et al., 2014; Naoi et al., 2015; Schaff and Richards, 2011; Shirzaei et al., 2013; Uchida et al., 2007; Uchida, 2019; Uchida and Matsuzawa, 2013; Yamaguchi et al., 2018).

The ERR can be directly derived from the corner frequency ( $f_c$ ) of the P or S wave spectrum (Brune, 1970):

$$\text{ERR} = \frac{kv}{f_c} \quad (3-1)$$

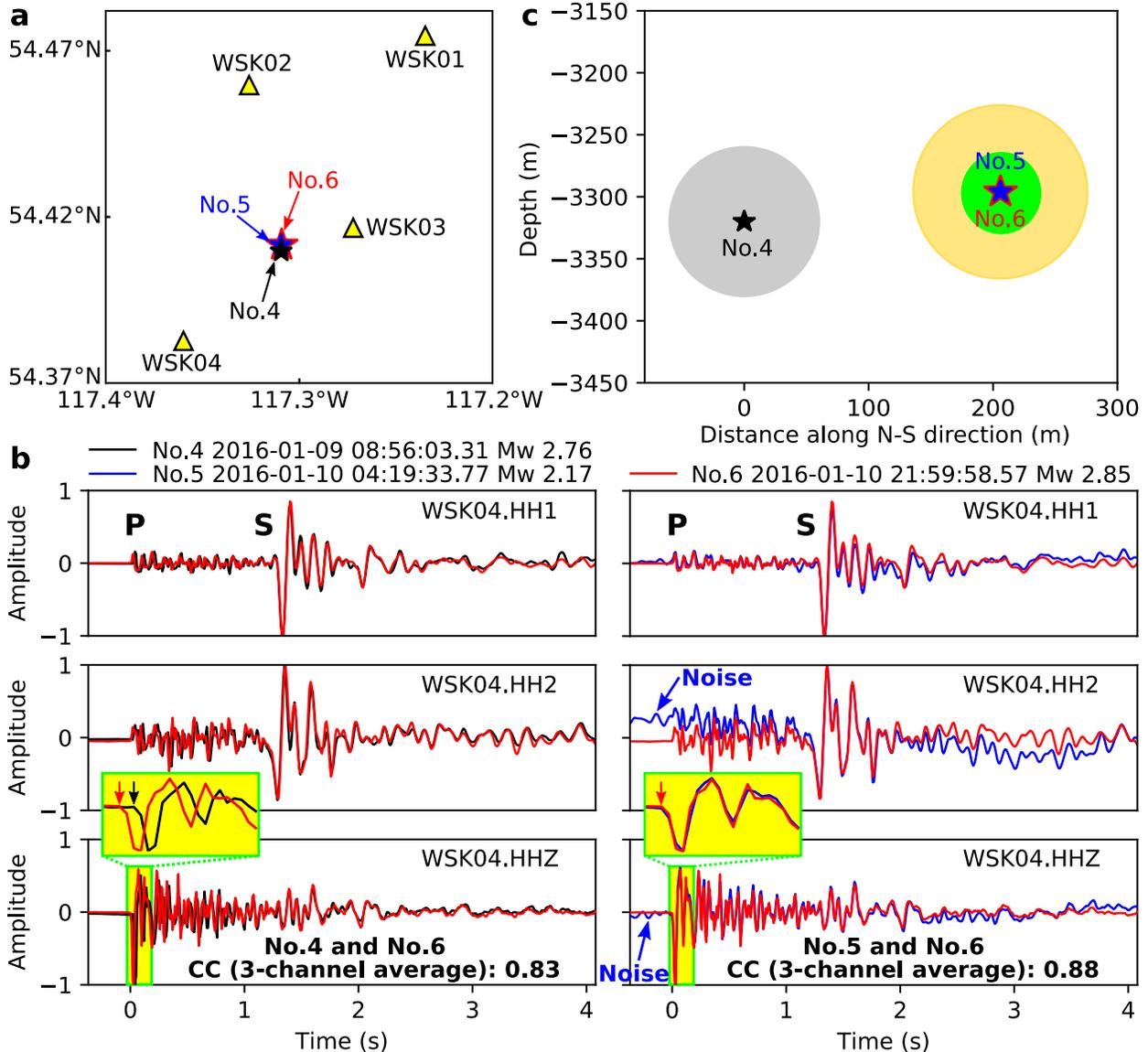
where  $k$  is a constant depending on the phase types (P or S) and  $v$  is phase (P or S) velocity. Alternatively, and more popularly, the ERR can be estimated from the event's stress drop ( $\Delta\sigma$ ) via the well-established circular dislocation model (Eshelby, 1957):

$$\text{ERR} = \sqrt[3]{\frac{7M_0}{16\Delta\sigma}} \quad (3-2)$$

where  $M_0$  is the event's scalar seismic moment. For repeater related studies (e.g., Cheng et al., 2007; Li et al., 2007, 2011; Lengliné and Marsan, 2009; Schaff and Richards, 2011; Ma et al., 2014; Hatakeyama et al., 2017; Mesimeri and Karakostas, 2018; Uchida, 2019),  $M_0$  is commonly converted from the event's moment magnitude ( $M_w$ ) as  $M_0 = 10^{(1.5M_w+9.1)}$  (Hanks and Kanamori, 1979). The value of  $\Delta\sigma$  can either be reasonably assumed (Table 3.3 in Section 3.9 Supporting Information) or directly derived (e.g., Abercrombie, 2014; Ellsworth and Bulut, 2018).

Using the ERR as a proxy for the size of the rupture area, we can rephrase the definition of a repeater as: (1) the inter-event distance is less than the ERR of the larger event, and (2) the magnitude difference between the two events is  $< 1$ . We note that ERR is the most popular, but not the only parameter suitable for this purpose. More precisely determined rupture distribution and/or geometry should be used to compare with the inter-event distance for repeater identification whenever possible.

For the estimation of inter-event distance, however, it is always a challenge to get a precise measurement unless a very dense local array is available (Cheng et al., 2007). In case of limited data, we propose a variant of the double difference method (HypoDD; Waldhauser and Ellsworth, 2000) that minimizes the residual between observed and predicted relative S-P differential traveltimes through three-dimensional (3D) grid search to precisely estimate the inter-event distance. We explain the details of our method, named the differential traveltimes double-difference (DTDD) method, in the Supporting Information (Section 3.9.2).



**Figure 3.4.** A physics-based approach to distinguish repeating and neighbouring events. (a) Map showing the distribution of earthquake epicentres (colored stars) and seismograph stations (yellow triangles). A zoom-in map of the source area is presented in Figure 3.15. (b) Normalized unfiltered waveforms, aligned at the S wave arrival. Black, blue, and red arrows in the zoom-in windows mark the P wave onset of events No. 4, 5 and 6, respectively. The slight time difference (0.02 s) between the red and black arrows in the left zoom-in window represents the subtle difference in S-P time between events No. 4 and 6. In the right zoom-in window, the blue and red arrows are overlapped with each other due to the identical S-P time of events No. 5 and 6. (c) North-south cross section showing the relative event locations. Gray, lime, and gold circles are the ERRs of events No. 4, 5 and 6, respectively.

Figure 3.4 presents an example of using the DTDD method to determine the precise relative position of three events with similar waveforms and small magnitude difference ( $< 1$ ) recorded at 4 nearby stations in Fox Creek, Alberta, Canada. Among them, events No. 4 and 6 are found to be  $\sim 200$  m apart, consistent with the subtle difference in S-P time (left zoom-in window in Figure 3.4b). In comparison, events No. 5 and 6 are effectively co-located (Figures 3.4c and 3.15) as indicated by the identical S-P time at all four stations (right yellow zoom-in window in Figure 3.4b).

To decide whether these three events are repeaters or not, we need to further consider the ERRs. Using the method proposed in Wang et al. (2020), we estimate the  $M_0$  of events No. 4, 5, and 6 to be  $1.76 \times 10^{13}$ ,  $2.26 \times 10^{12}$ , and  $2.24 \times 10^{13}$  N·m, respectively. The  $\Delta\sigma$  of events No. 4 and 6 are 35 and 30 MPa, respectively. Since the waveforms of event No.5 is contaminated by a high level of noise (Figure 3.4b), its stress drop is assumed to be the average of events No. 4 and 6 (i.e., 32.5 MPa). Based on Equation 3-2, the corresponding ERRs of these three events are 60 (No. 4), 31 (No. 5), and 70 (No. 6) m. Therefore, we conclude that events No. 5 and No. 6 are true repeating events (i.e.,  $ERR > \text{inter-event distance}$  and  $\text{magnitude difference} < 1$ ), whereas events No. 4 and No. 6 are at most neighbouring events ( $ERR < \text{inter-event distance}$ ). Notice that our repeater identification criteria do not involve the source dimension of the smaller event. Hence the stress drop assumption of event No. 5 (the smallest event) will not affect our results.

We finally note that both event pairs have high CC values (0.83 between events No. 4 and 6, 0.88 between events No. 5 and 6; Figure 3.4b). Consequently, the determination of these event pairs being true repeaters or not can be arbitrary depending on the choice of the CC threshold (e.g., the choice of 0.80, 0.85, or 0.90 will lead to completely different outcomes). Our results clearly indicate that the CC can drop below 0.9 even for true repeating events when the waveforms of the smaller event are contaminated with noise (event No. 5; Figure 3.4b). It further underscores the challenge in identifying repeaters based solely on waveform similarity with the presence of noise.

### **3.6 Discussion**

ERR and inter-event distance are the two key parameters in our proposed framework of identifying repeaters. The estimation of ERR is usually based on Equation 3-2 with a reasonably

assumed  $\Delta\sigma$  (Table 3.3 in Section 3.9 Supporting Information).). While most previous studies consider  $\Delta\sigma$  to be no more than 10 MPa (Table 3.3), detailed analysis of dense local borehole array data in the Parkfield area suggests that  $\Delta\sigma$  of small ( $M \sim 2.1$ ) events can be as high as 65 MPa (Abercrombie, 2014). Since a smaller  $\Delta\sigma$  value will yield a larger ERR, underestimation of  $\Delta\sigma$  is likely to misclassify neighbouring events as repeaters, and vice versa. Figure 3.16 gives the quantitative relationship between earthquake magnitude and ERR with different assumed values of  $\Delta\sigma$ . Taking an  $M 2$  event for example, the ERR can be 82, 57, 38, or 20 m if the corresponding  $\Delta\sigma$  is 1, 3, 10, 65 MPa, respectively. Among the four assumed  $\Delta\sigma$ , 1 and 10 MPa represent the lower and upper bound of  $\Delta\sigma$  of typical tectonic earthquakes (e.g., Stein and Wysession, 2009), respectively. Most repeater related studies assume  $\Delta\sigma$  to be 3 MPa (Table 3.3), whereas a value of 65 MPa is considered extreme for a small ( $M 2$ ) event (Abercrombie, 2014). If the inter-event separation between the  $M 2$  pair is 50 m, then this event pair can be classified as repeaters (ERR = 82 or 57 m) or neighbouring events (ERR = 38 or 20 m) depending on the assumed  $\Delta\sigma$ . Such an ambiguity highlights the significant risk and uncertainty due to a poorly constrained (or wrongly assumed)  $\Delta\sigma$  value. To uniquely determine the ERR, we therefore recommend to directly calculate the source parameters (i.e.,  $f_c$ ,  $\Delta\sigma$ , and  $M_0$ ) whenever possible (e.g., Dreger et al., 2007; Abercrombie, 2014; Ellsworth and Bulut, 2018; Wang et al., 2020). In case of very limited data, a quantitative ERR uncertainty analysis is necessary before declaring an event pair to be repeaters/non-repeaters.

The DTDD method that we develop to estimate the inter-event distance with limited waveform data relies on precise measurement of the S-P differential traveltime (Section 3.9.2). For a typical crustal model (i.e.,  $V_p = 6.5$  km/s and  $V_p/V_s = 1.73$ ), a 0.01s difference in the S-P time corresponds to a hypocentral difference of  $\sim 90$  m (Hayward and Bostock, 2017). Consequently, we need to pay attention to two potential issues. First, the application of digital filtering, such as the band-pass filters used in previous studies, may change the width of the P wave, leading to a slight shift between the P and S phases (one example is shown in Figure 3.3d). The small bias (0.012 s in the case shown in Figure 3.3d) is equivalent to a mislocation of  $\sim 100$  m that is sufficient to cause misinterpretation for events with small source dimension. Therefore, we prefer to use unfiltered broadband waveforms in the measurement of S-P differential traveltime to avoid any possible bias from waveform filtering.

The second issue is the resolution limit defined by the sampling rate of waveform data. It can be particularly problematic if the original sampling rate is less than 100 Hz (i.e.,  $\geq 0.01$ s between samples) so the hypocentral uncertainty becomes comparable to the source dimension of small events. A straightforward solution is to pre-process waveforms with interpolation to achieve sub-sample precision in measuring the S-P times (Li et al., 2007, 2011). Similarly, the grid size used in the DTDD source-searching process should be much smaller than the source dimension of the targeted events to achieve optimal spatial resolution.

We note that, in the extreme case of limited data from only one or two stations, the DTDD solution can be highly non-unique. Consequently, a priori constraints must be introduced to quantitatively estimate the inter-event distance. One commonly adopted remedy is to require the two events to occur on a given fault plane (e.g., Li et al., 2007). Another commonly adopted constraint comes from the ambient tectonic loading rate, i.e., the recurrence interval between two repeaters should be proportional to the size of the second event (e.g., Li et al., 2007, 2011; Bohnhoff et al., 2017). If the two events occur very closely to each other in time, they are more likely to be neighbouring events because the fault patch ruptured during the first event has not healed yet.

In summary, robust identification of repeaters is very challenging. The significance of this study is to recognize that waveform similarity lacks the resolution in discriminating true repeaters and neighbouring events. Although the physics-based approach is very attractive, precise estimation of the inter-event distance and ERRs for a very large seismic dataset can be very labor-intensive and time-consuming. In such cases, using waveform-similarity based method such as MFMC (Gao and Kao, 2020) for preliminary repeater scanning can significantly reduce the workload of final verification. We note that choosing a proper filter by checking amplitude spectra of both signal and noise is important for the optimal performance of MFMC. However, using unfiltered waveform data is necessary to maintain the correct S-P time for repeater confirmation.

### **3.7 Conclusion**

In this study we reveal that CC can be severely affected by many factors, including station azimuth, epicentral distance, velocity structure, focal mechanism, network geometry, orientation

of the source separation (horizontal vs. vertical), the choice of data (one specific channel or all three channels), and the spectral passband used in data processing. Additionally, the level of background noise and the length of template waveforms may further influence the CC's sensitivity to inter-event distance (Gao and Kao, 2020). Therefore, it is very difficult to reliably identify repeaters solely based on a given CC value, implying that a systematic recheck of previously identified repeaters and associated interpretations/hypotheses may be important and necessary.

To more reliably identify true repeaters, we should consider both overlapped source area and earthquake magnitude. Specifically, we define an event pair to be repeaters if their inter-event separation is no larger than the ERR of the larger event and their magnitude difference is smaller than 1. For the precise estimation of inter-event distance in case of limited data, we develop the DTDD method which relies on the relative S-P differential traveltime. Finally, we illustrate the effectiveness of the DTDD method using earthquakes that occurred in the Fox Creek area, Alberta, Canada. The findings of this work have far-reaching impact on not only repeating earthquake research but also other waveform-similarity-based studies.

### **3.8 Acknowledgements**

We are grateful to Lupei Zhu for providing the FK code that is used in generating synthetic seismograms. We specially thank Bei Wang for helping calculate the source parameters of the Fox Creek events, and thank Didem Cambaz and Edwin Nissen for the help offered in this study. Insightful discussions with Stan Dosso, Toshihiro Igarashi, Naoki Uchida, David P. Schaff, Makoto Naoi, Jianlong Yuan, Kelin Wang, Ryan Visser, Rachel E. Abercrombie, Fengzhou Tan, Ramin Mohammad Hosseini Dokht, Wen-che Yu, Lingsen Meng, Hui Huang, Jean Schmittbuhl, Marco Bohnhoff, Christopher Wollin, Tomoaki Nishikawa, Emily Warren-Smith, Luis A. Dominguez, Tianhaozhe Sun, Taizi Huang, and Haipeng Luo are much appreciated. The HRSN and Fox Creek event waveform data used in this study were downloaded from the Northern California Earthquake Data Centre (NCEDC) (<http://ncedc.org/hrsn/>) and Incorporated Research Institutions for Seismology (IRIS) (<http://ds.iris.edu/ds/nodes/dmc/>), respectively. Seismic data are processed with Obspy (Beyreuther et al., 2010; <https://github.com/obspy/obspy/>). Figures are made with Matplotlib (Hunter, 2007; <https://matplotlib.org>) and Inkscape (<https://inkscape.org>).

This study is partially supported by a University of Victoria Fellowship (DG), the Induced Seismicity Research Project of NRCan (HK), Geoscience BC (HK), and a NSERC Discovery Grant (HK).

## 3.9 Supporting Information

### 3.9.1 CC Calculation

To calculate the CC value of an earthquake pair, we utilize the recently developed match-filtering with multi-segment cross-correlation (MFMC) technique (Gao and Kao, 2020) instead of the classical cross-correlation method which can be severely biased by the existence of large-amplitude phases such as S wave and surface waves (e.g., Kraft and Deichmann, 2014; Myhill et al., 2011; Li et al., 2017; Gao and Kao, 2020). Compared with the conventional one-segment approach, the MFMC technique splits the template into a number of consecutive segments during the cross-correlation process. Such a procedure is designed to mitigate the impact of the large-amplitude phases and essentially gives more weights to important low-amplitude phases such as depth phases (Ma and Atkinson, 2006; Ma, 2010) and coda waves (Snieder and Vrijlandt, 2005; Robinson et al., 2011) that carry additional source location information. Thus the MFMC technique is more reliable in capturing the waveform discrepancy and differentiating the source location difference between an event pair (Gao and Kao, 2020).

In our MFMC CC calculation, we first cut the waveform from the template event starting at the onset of the P wave with a length of  $T_{win}$ . For synthetic experiments, we use a dynamic template window length of  $T_{win} = 3(T_s - T_p)$ , where  $T_s$  and  $T_p$  are the S- and P-phase arrival times, respectively. Using a dynamic template window length based on the differential traveltime between P and S phases is necessary to properly account for the increasing wave train with epicentral distance ( $R = 5, 50, \text{ or } 150 \text{ km}$  in our experiment, Figure 3.1a) (Baisch et al., 2008; Gao and Kao, 2020). For real earthquake waveform tests in Parkfield area, we tested both dynamic (i.e.,  $3(T_s - T_p)$ ) and fixed  $T_{win}$  (1.5 s, 3.0 s, 4.5 s, and 6.0 s) for each station given the small study area (Figure 3a). It should be noted that a fixed  $T_{win}$  of 6.0 s, equivalent to  $12(T_s - T_p)$  of the closest station LCCB (Figure 3.3a), is sufficiently long to cover much of the low-amplitude coda waves. Then the template waveform is divided into  $N_{seg}$  segments of equal

length. For unfiltered waveforms,  $N_{seg}$  is assigned as 4; for filtered waveforms, the value of  $N_{seg}$  is determined by the cycles of the longest period wave ( $1/f_{min}$ ) in the band-pass filtered waveform (i.e.,  $N_{seg} = T_{win} \times f_{min}$ ) and the minimum value of  $N_{seg}$  is set to be 4 as suggested by an earlier study (Gao and Kao, 2020). Finally, all the segments are shifted together one sample point at a time from 0.5 s before the P arrival of the targeted event to 0.5 s after. The cross-correlation calculation is performed individually for each segment, and the CC value at each sample point is given by the average of all segments. The maximum CC value of all time steps is defined as the final CC value of the earthquake pair. If data from three channels of the same station are used, the CC value of all time steps from all three channels are averaged, and the maximum within the sliding window is taken as the final CC.

### 3.9.2 The Differential Traveltime Double-Difference (DTDD) Method

To precisely estimate the relative location between an event pair, we develop the differential traveltime double-difference (DTDD) method which minimizes the residual between observed and predicted relative S-P differential traveltime through three-dimensional (3D) grid search. The relative S-P differential traveltime,  $\Delta SP_{ij}^k$ , between events  $i$  and  $j$  at station  $k$  is given by:

$$\Delta SP_{ij}^k = SP_i^k - SP_j^k \quad (3-3)$$

where  $SP_i^k$  and  $SP_j^k$  are the S-P differential traveltimes of events  $i$  and  $j$  at station  $k$ , respectively. If we take the event  $i$  as the reference, the relative location of event  $j$  is obtained by searching all the possible locations around the reference event in a 3D space. The best location is defined as the grid node which yields the minimum sum of the residuals,  $R$ , between observed and predicted relative S-P differential traveltimes:

$$R = \sum_k |(\Delta SP_{ij}^k)^{obs} - (\Delta SP_{ij}^k)^{cal}| \quad (3-4)$$

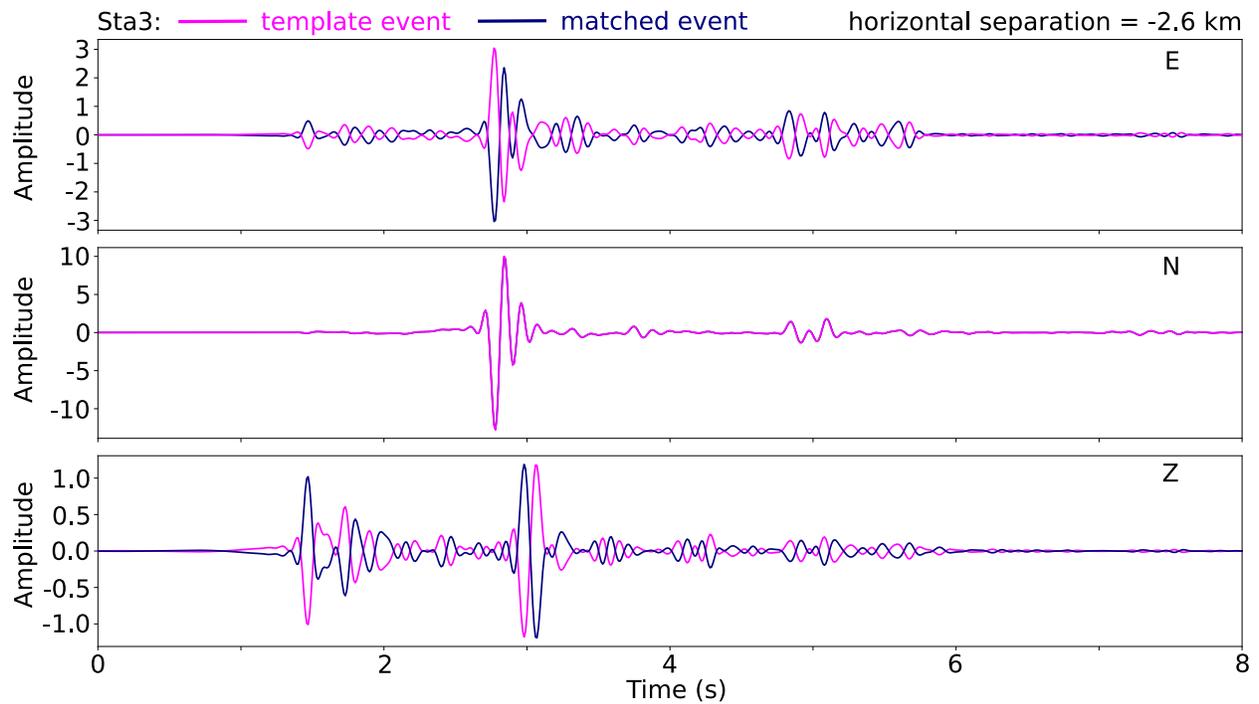
where  $(\Delta SP_{ij}^k)^{obs}$  and  $(\Delta SP_{ij}^k)^{cal}$  are the observed and theoretical relative S-P differential traveltimes, respectively. Combining (3-3) and (3-4) yields:

$$R = \sum_k |(SP_i^k - SP_j^k)^{obs} - (SP_i^k - SP_j^k)^{cal}| \quad (3-5)$$

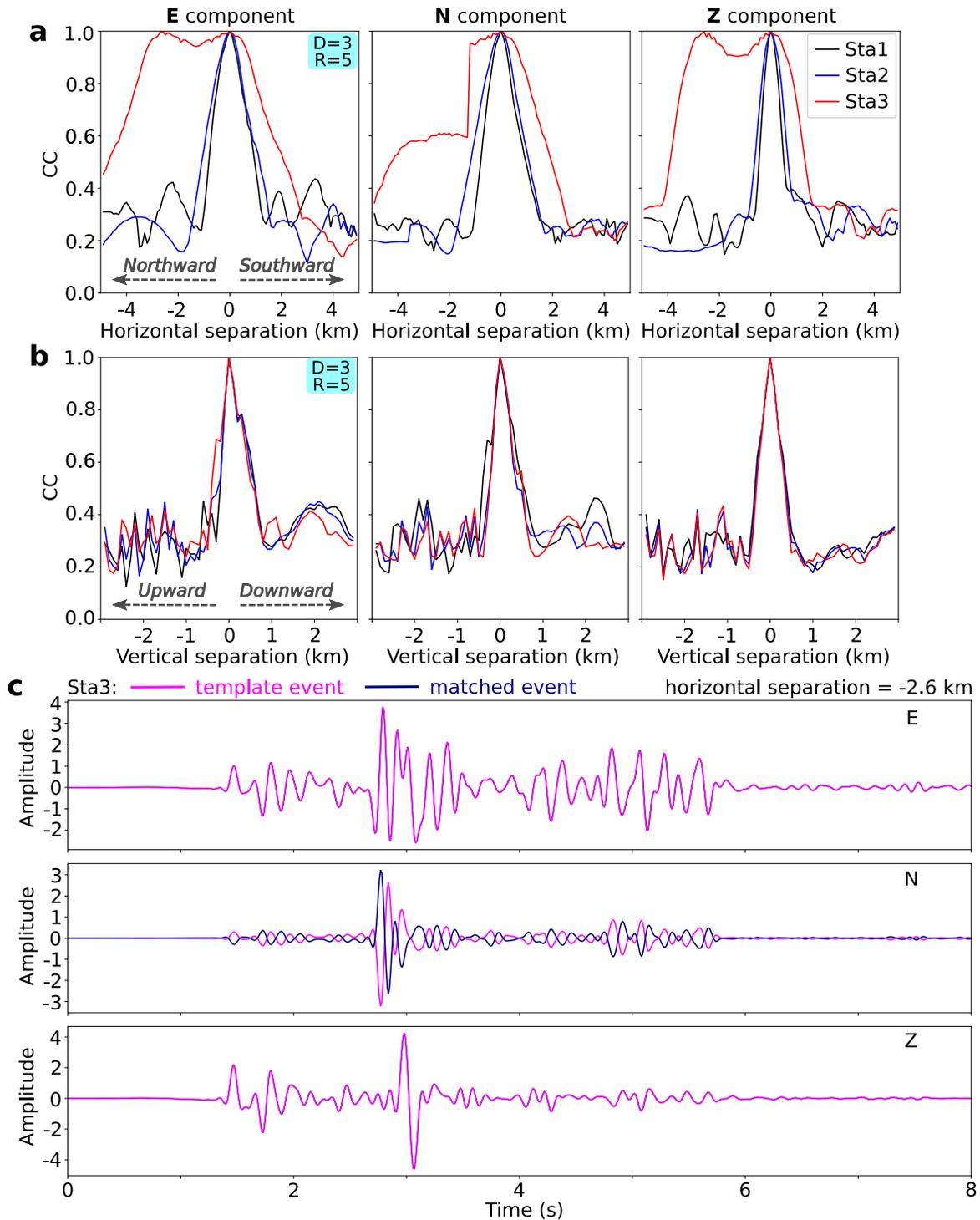
which is the double-difference of the S-P differential traveltime. It should be noted that, in the measure of misfit (Equations 3-4 and 3-5), we use the L1 normal instead of the widely used L2 norm (least squares) because L1 normal is more robust and less sensitive to outliers (Shearer, 1997; Shearer, 2009; Trugman and Shearer, 2017).

When implementing the 3D grid search, we employ a two-step strategy to minimize the calculation, i.e., using a large grid size for initial locating and a very small one for the final solution. Taking the Fox Creek events (No. 4-6) for example, we take the largest event No. 6, which is the best recorded event with the highest SNR, as the reference location. Such a choice is meant to get more reliable relative locations. Then we define the search volume as the cubic space of  $1 \text{ km} \times 1 \text{ km} \times 1 \text{ km}$  centred at the reference event. The 1-D velocity model for calculating theoretical relative S-P differential traveltimes is taken from Bao and Eaton (2016), suitable for Fox Creek area. The preliminary solution is determined with a grid search at the spacing of 10 m. Once the preliminary location is identified, we conduct a finer search at the interval of 1 m within the  $20 \text{ m} \times 20 \text{ m} \times 20 \text{ m}$  cubic space centred at the preliminary solution to obtain the final solution. The effectiveness of our approach is evident from the located results shown in Figure 3.15. The hypocentres of both reference (No. 6) and located (No. 4 and No. 5) events generally fall on a line trending N-S, consistent with the inference of a local N-S trending strike-slip fault system (Schultz et al., 2017; Wang et al., 2017).

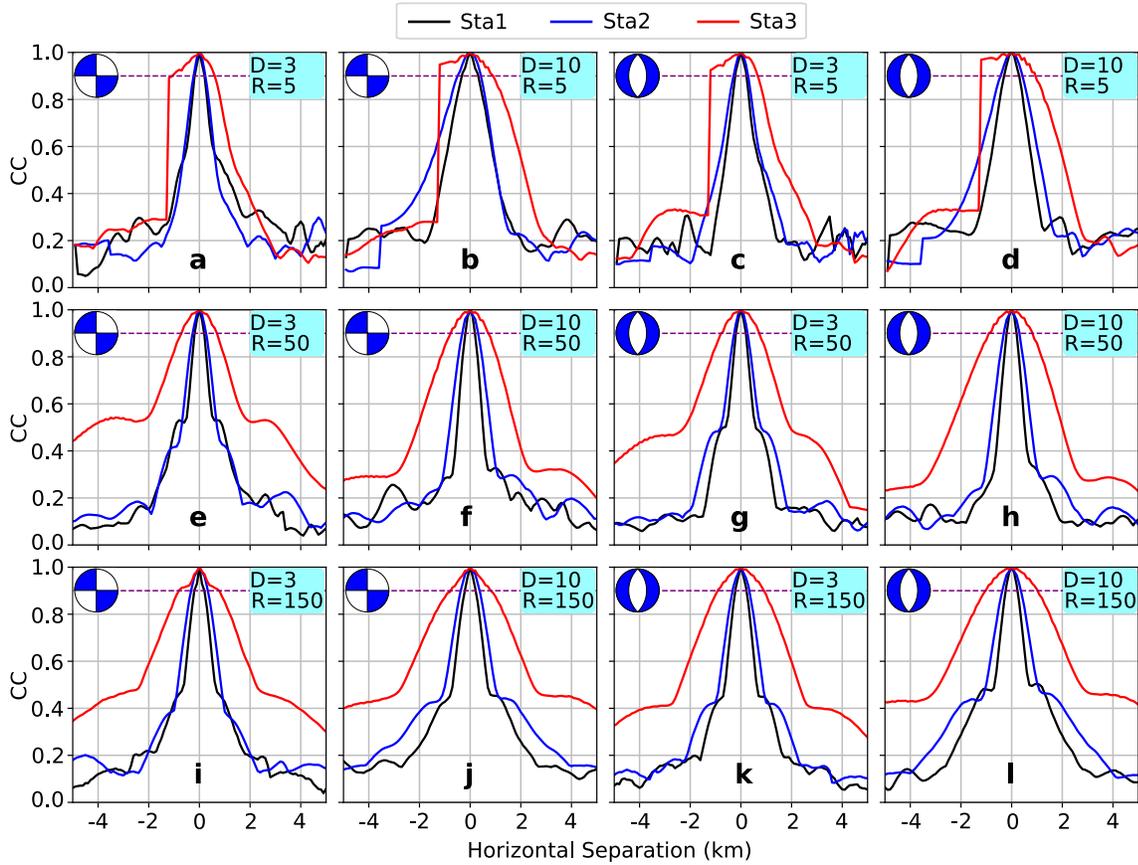
Compared with the conventional grid search approach which scans all possible locations and origin times to minimize the time residuals between observed and theoretical traveltimes of certain phases (e.g., P and/or S phases; Shearer, 2009), one obvious advantage of the DTDD method is that the relative position is determined without the events' origin times. Another added benefit is that it can cancel any path effects due to unmodeled velocity heterogeneities and hence lead to very precise relative location. We note that precise relative location based on the relative S-P differential traveltime can also be solved by conventional inversion schemes (e.g., Cheng et al., 2007). However, the conventional inversion approach may suffer from being trapped to local minima, whereas the forward grid-search approach adopted by the DTDD method can always find a solution sufficiently close to the global minimum given our extremely small grid size.



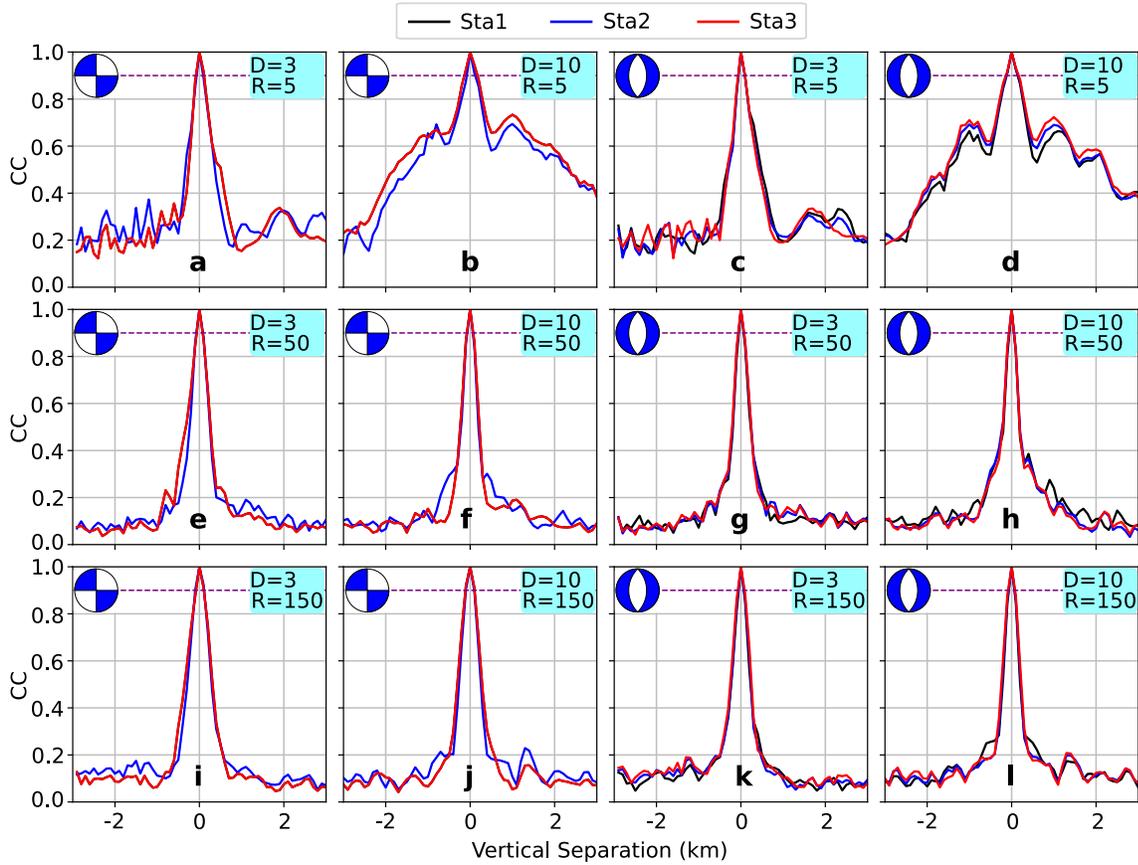
**Figure 3.5.** Synthetic waveforms of both template and matched events for a horizontal separation of -2.6 km. The template event is a strike-slip earthquake and is placed at the depth of 3 km. The receiver has an epicentral distance of 5 km with respect to the template event.



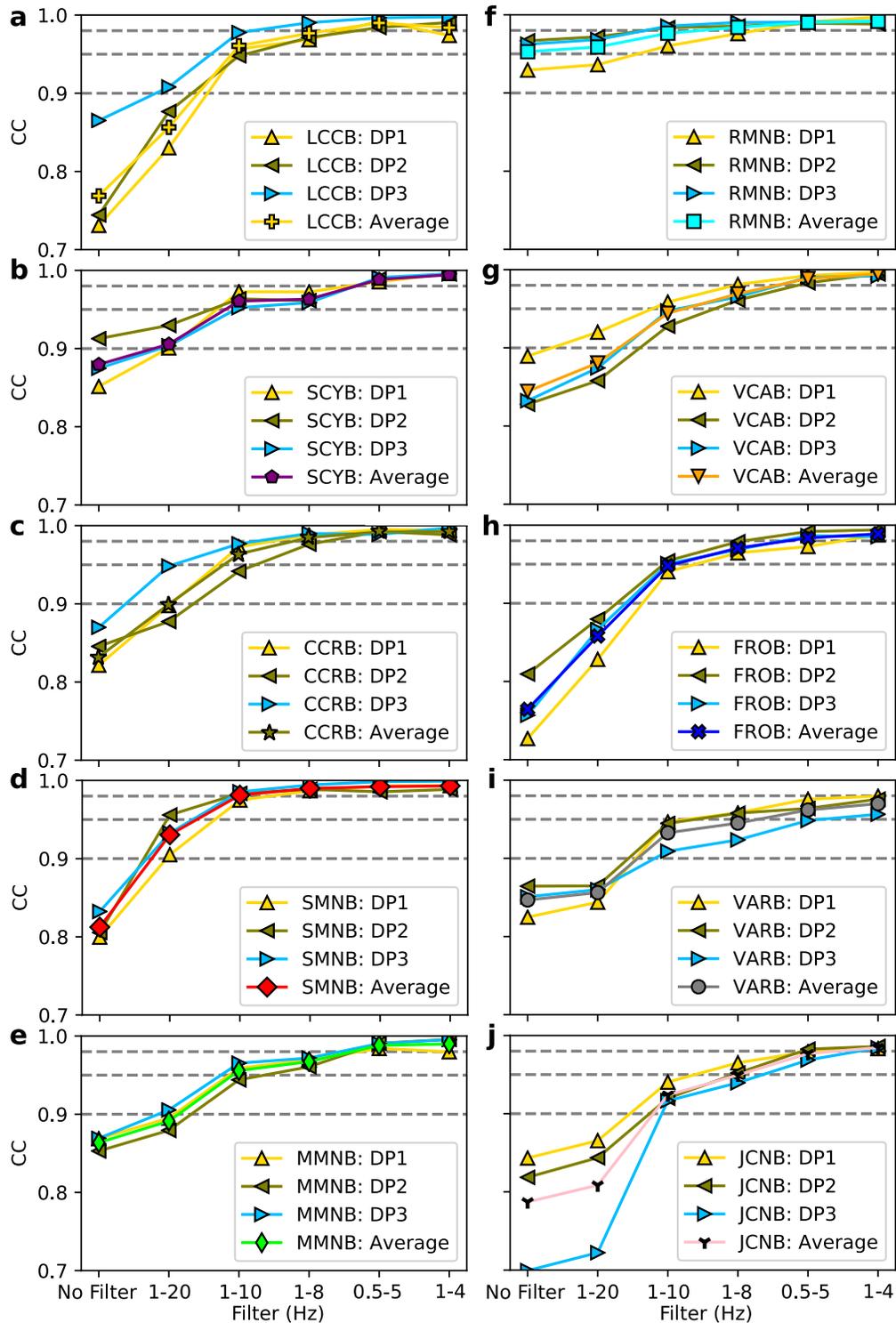
**Figure 3.6.** CC variation (obtained with single-channel data) due to inter-event separations. The template event is a normal-faulting earthquake at the depth of 3 km. The receiver has an epicentral distance of 5 km with respect to the template event. (a) CC variation due to horizontal separations. (b) CC variation due to vertical separations. (c) Synthetic waveforms of both template and matched events for a horizontal separation of  $-2.6$  km.



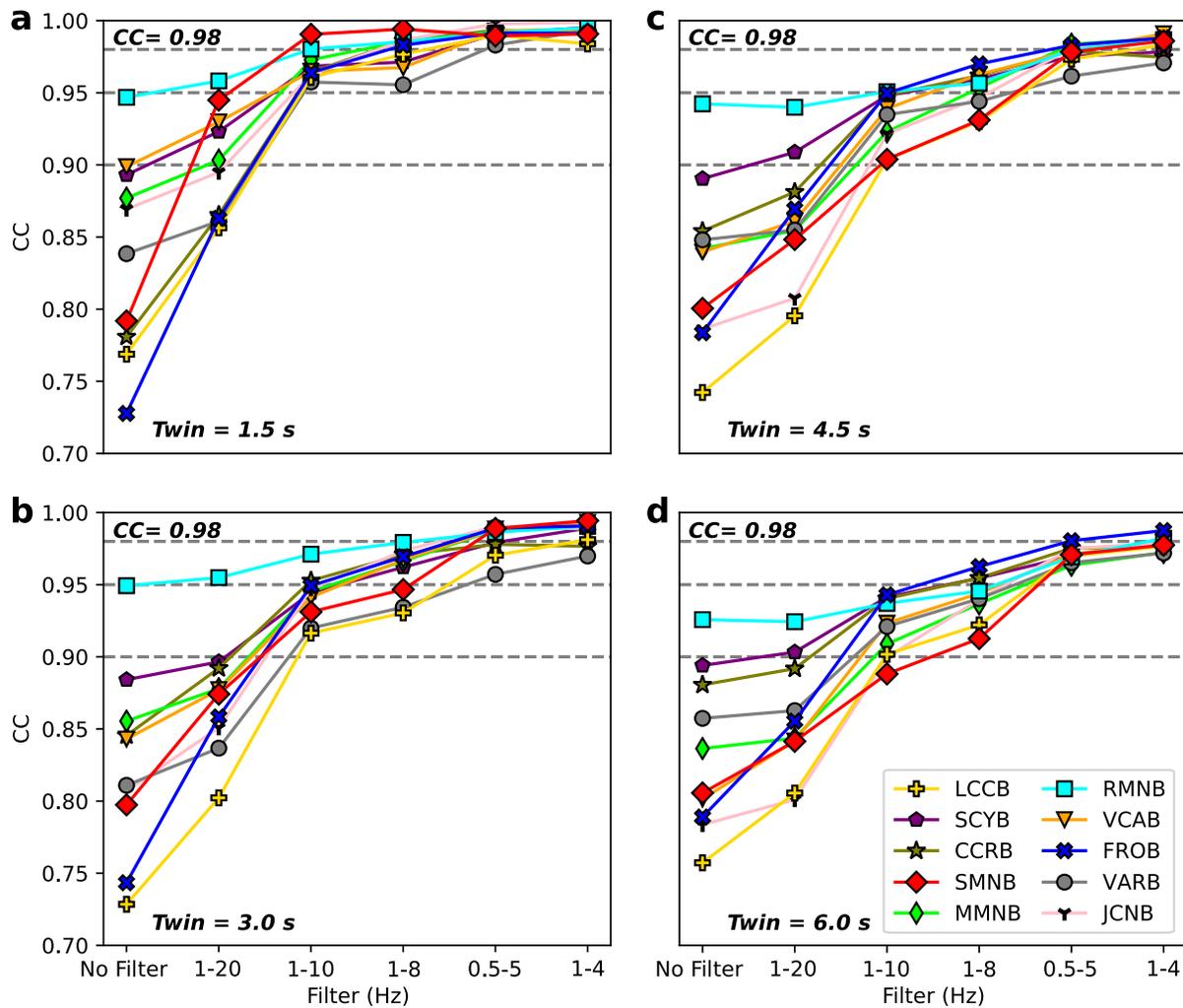
**Figure 3.7.** CC variation (obtained with single-station, 3-channel data) due to horizontal inter-event separations. Dashed purple line marks the CC value of 0.90 for reference. Note that the results of a reverse fault are the same as those of a normal fault and hence are not displayed.



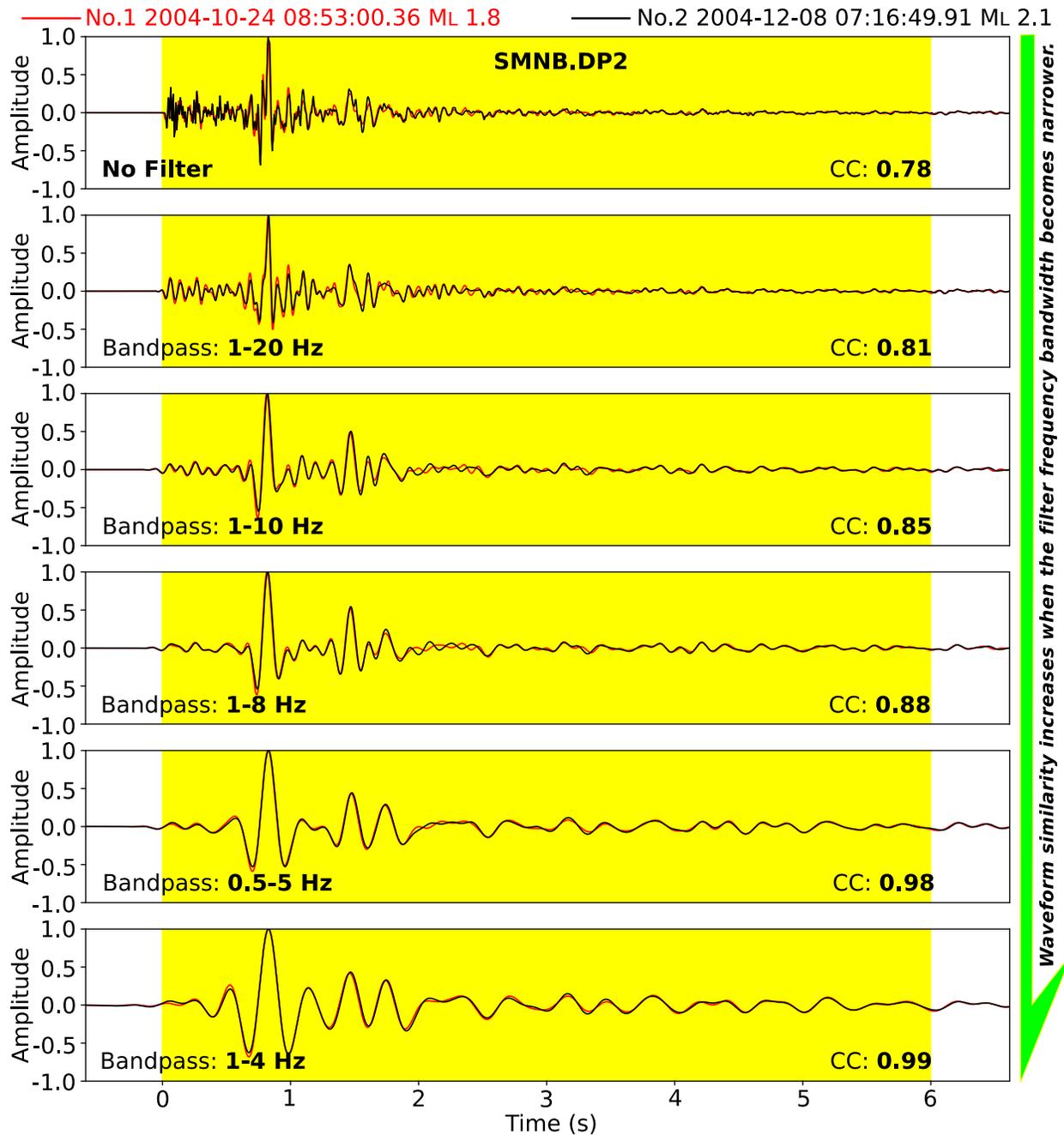
**Figure 3.8.** CC variation (obtained with single-station, 3-channel data) due to vertical inter-event separations. Layout is the same as Figure 3.7. Note that the results of a reverse fault are the same as those of a normal fault and hence are not displayed.



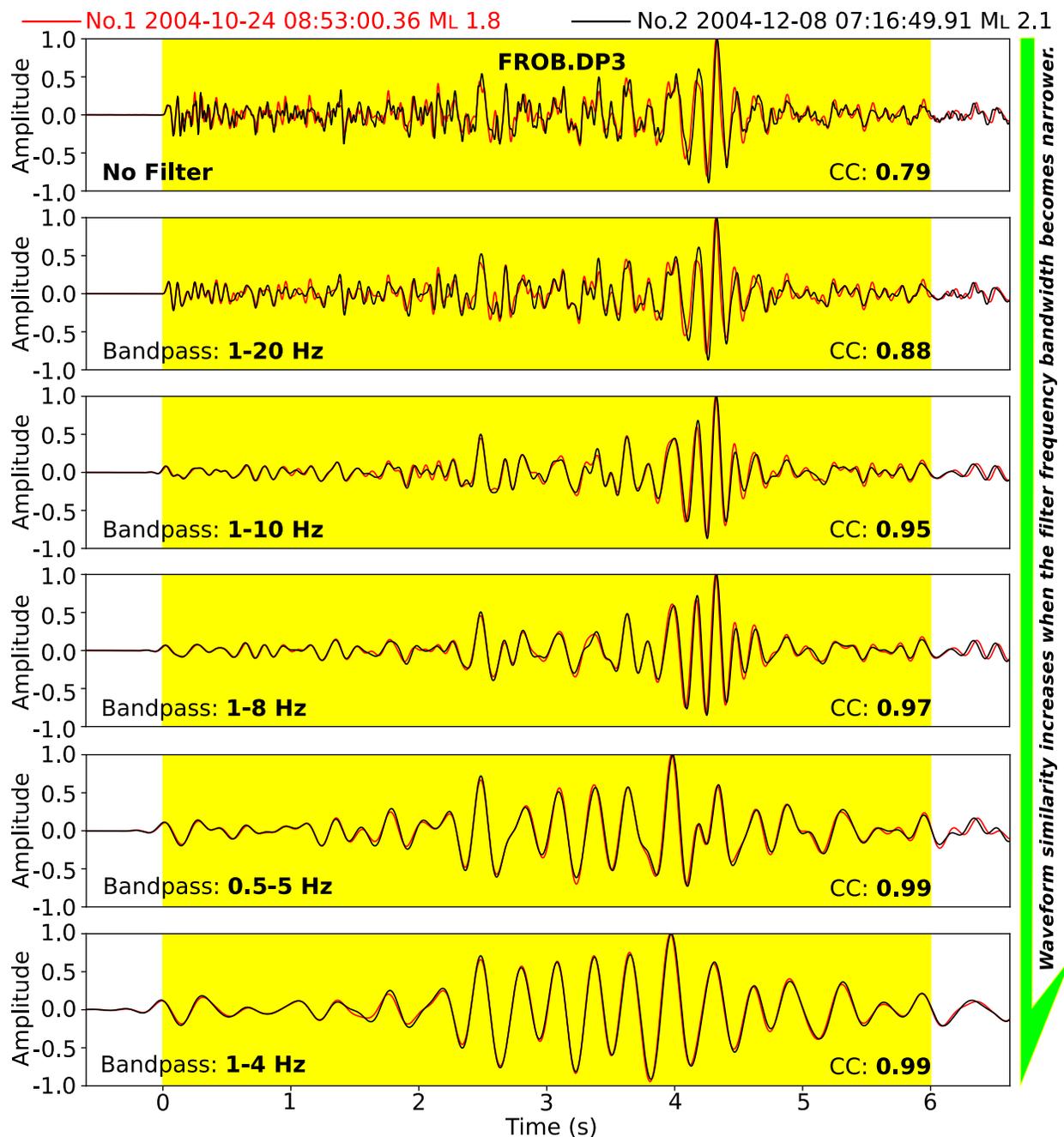
**Figure 3.9.** Effects of filtering on the CC value between two events that have been verified to be non-repeaters (events No. 1 and No. 2, Figure 3.3) with single-channel data. All CC values are determined with the dynamic  $T_{win}$  (see Section 3.9.2 for more details). For each panel, the station and channel names (e.g., LCCB: DP1) are given in the legend box.



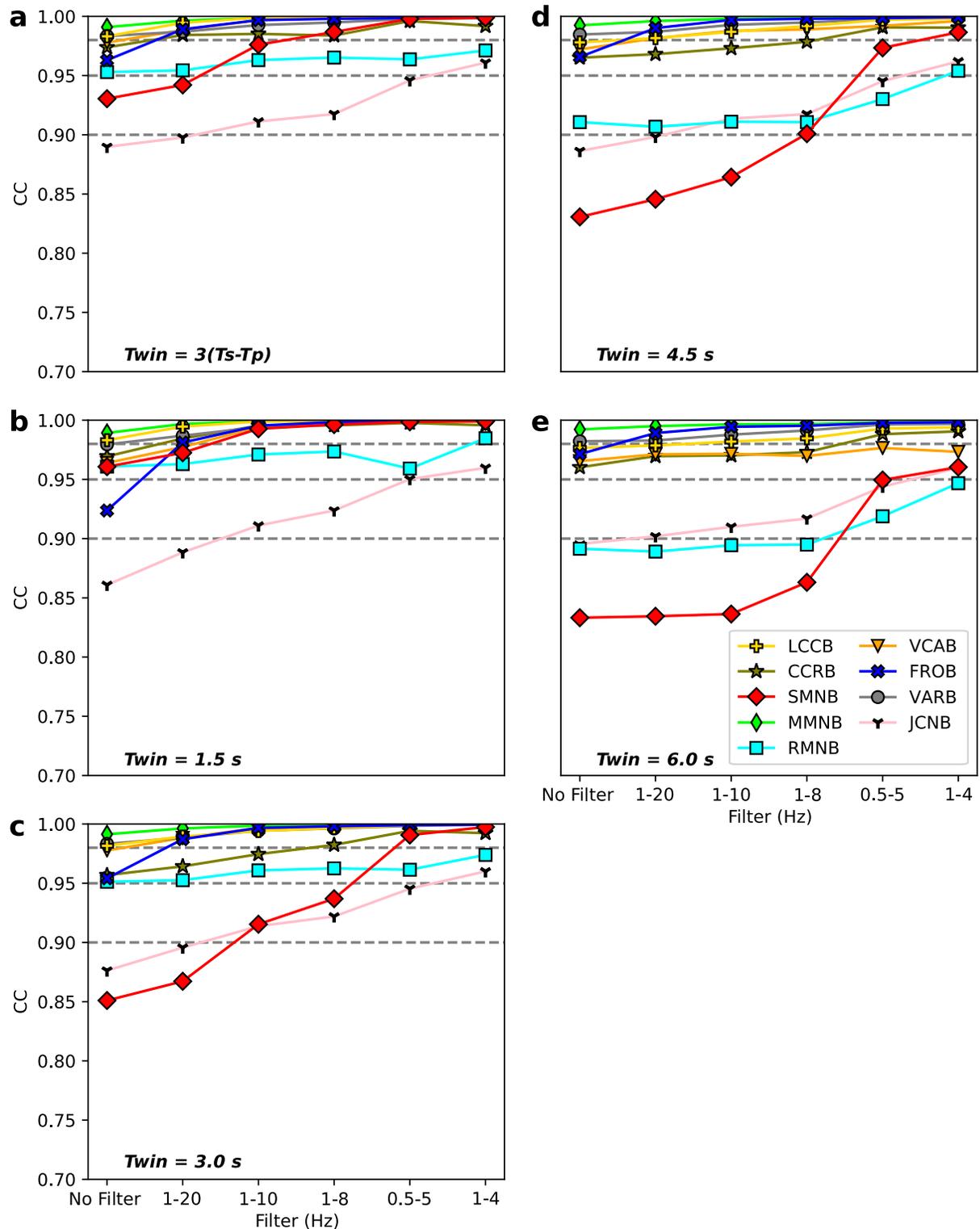
**Figure 3.10.** Effects of  $T_{win}$  and filtering on the CC value between two events that have been verified to be non-repeaters (events No. 1 and No. 2, Figure 3.3) with single-station (3-channel) data. (a)-(d) correspond to the  $T_{win}$  of 1.5, 3.0, 4.5 and 6.0 s, respectively.



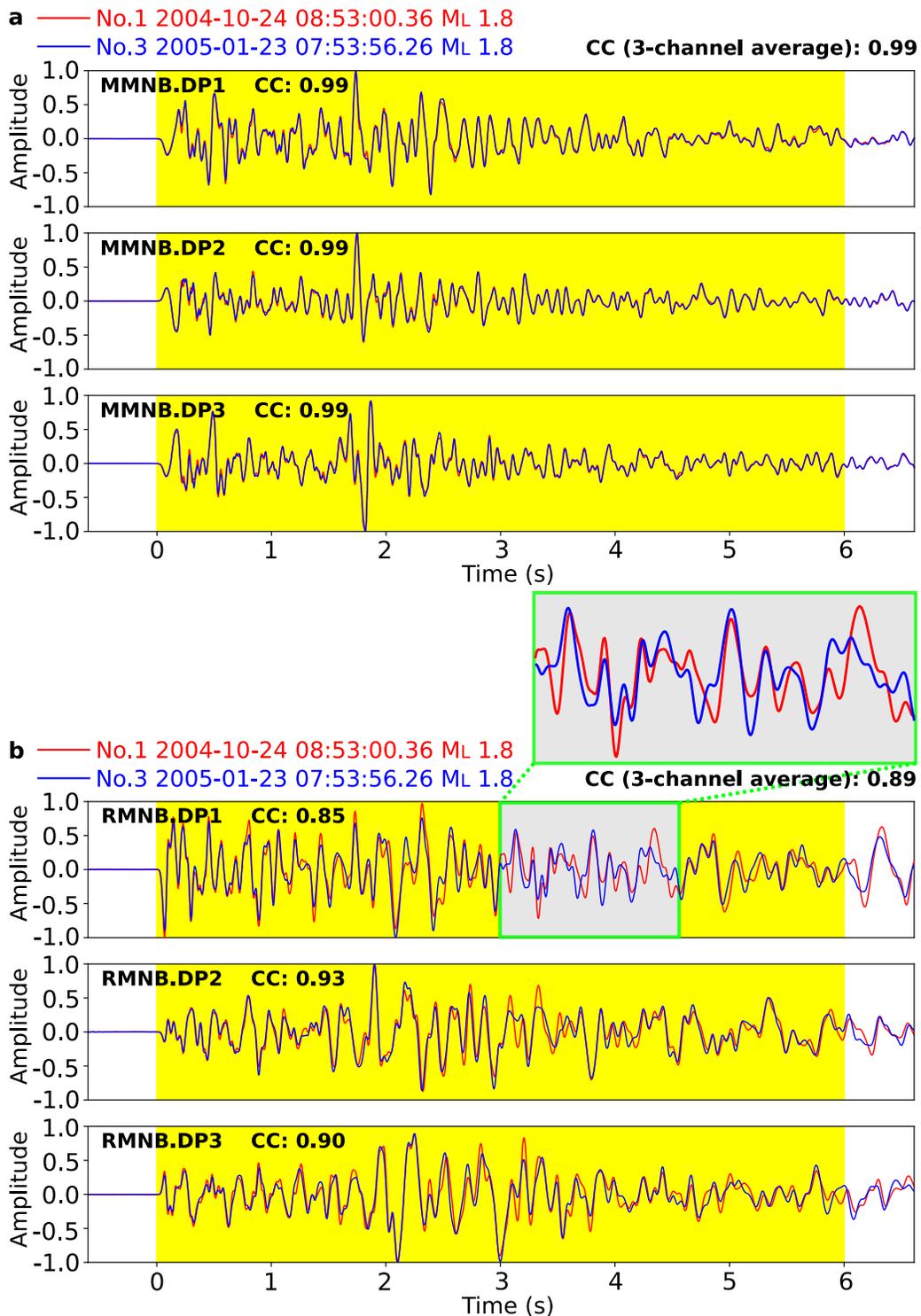
**Figure 3.11.** An example of how filtering increases waveform similarity at a close station SMNB. The highlighted segment ( $T_{win} = 6.0$  s) is used to calculate the CC value which is labelled at the lower-right corner. Note that a  $T_{win}$  of 6.0 s, equivalent to  $10(T_s - T_p)$  at station SMNB, is sufficiently long to cover much of the low-amplitude coda waves.



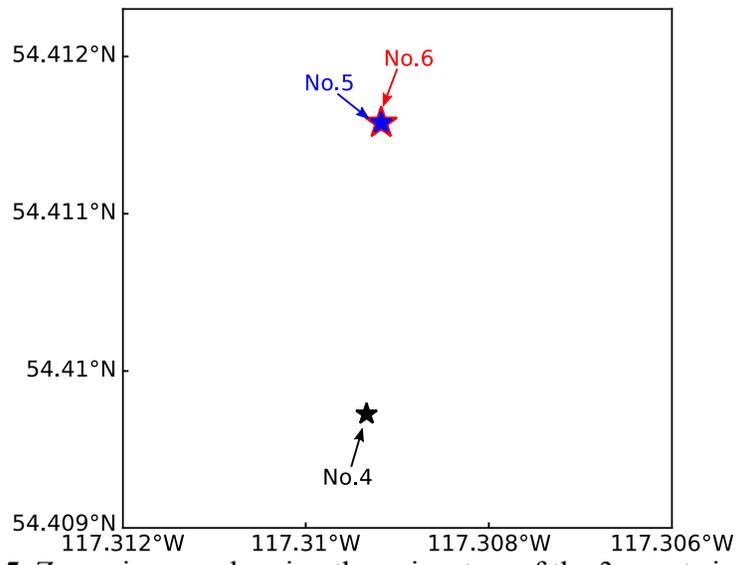
**Figure 3.12.** An example of how filtering increases waveform similarity at a distant station FROB. Layout is the same as Figure 3.11.



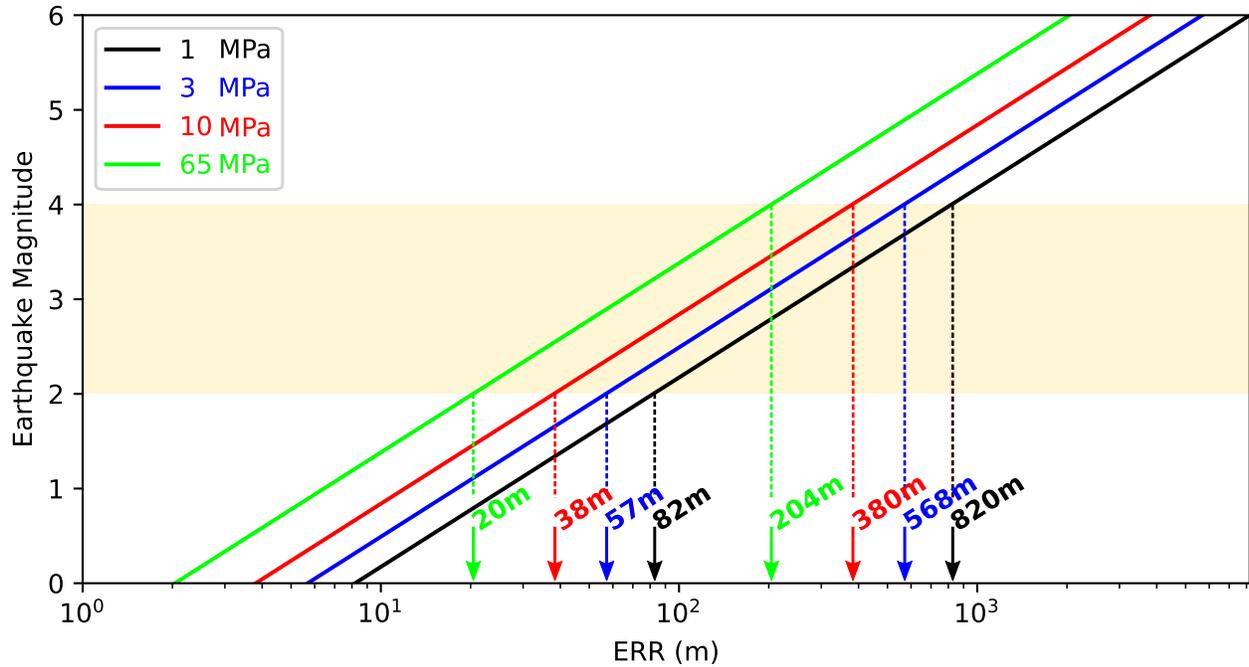
**Figure 3.13.** Effects of  $T_{win}$  and filtering on the CC value between two events that have been verified to be true repeaters (events No. 1 and No. 3, Figure 3.3) with single-station (3-channel) data. Notice that station SCYB is not used in this case because of heavy noise contamination. (a) CC determined by  $T_{win}$  that is dynamically adjusted for each station. (b)-(d) correspond to the  $T_{win}$  of 1.5, 3.0, 4.5, and 6.0 s, respectively.



**Figure 3.14.** Examples of normalized unfiltered waveforms of two events that have been verified to be true repeaters (events No. 1 and No. 3, Figure 3.3). (a) Nearly identical waveforms at station MMNB. (b) Waveforms with minor difference at station RMNB. The highlighted segment ( $T_{win} = 6.0$  s) is used to calculate the CC value.



**Figure 3.15.** Zoom-in map showing the epicentres of the 3 events in Figure 3.4.



**Figure 3.16.** Relationship between earthquake magnitude ( $M$ ) and ERR derived from Equation 3-2 and moment ( $M_0$ )-magnitude ( $M$ ) relationship established by Hanks and Kanamori (1979). Color shaded area represents the typical repeater magnitude range (Table 2.1).

**Table 3.1.** A compiled list of different criteria in identifying repeaters

References	Region	CC Threshold	Additional Criteria/Analysis	Min. Num. of Stations <sup>a,b</sup>	Note
Green and Neuberg (2006)	Soufrière Hills Volcano, Montserrat	0.70		1 station (vertical channel)	
Salvagea and Neuberg (2016)	Soufrière Hills Volcano, Montserrat	0.70		1 station (–)	
Yamada et al. (2016)	Hokkaido, Japan	0.70		1 station (three channels)	
Thelen et al. (2010)	Bezymianny Volcano, Russia	0.70/0.80		2 stations (–)	CC = 0.70 for daily detection; CC = 0.80 for merging results from different days
Thelen et al. (2011)	Mount St. Helens (MSH), Washington; Bezymianny Volcano (BV), Russia	0.70/0.80		2 stations for MSH (–); 2 stations for BV (vertical channel)	CC = 0.70 for daily detection; CC = 0.80 for merging results from different days
Rau et al. (2007)	Longitudinal valley fault, Taiwan	0.70/0.85	composite selection criteria (both CC and S-P differential time)	multi-station (vertical channel)	
Chen et al. (2008)	Chihshang fault, Taiwan	0.70/0.85	composite selection criteria (both CC and S-P differential time)	multi-station (vertical channel)	
Buurman et al. (2013)	Redoubt Volcano, south-central Alaska	0.75		1 station (single-channel)	
Schultz et al. (2014)	Alberta, Canada	0.75		1 station (–)	
Cauchie et al. (2020)	Soultz-sous-Forêts, France	0.75	overlap of rupture areas (inter-event distance < the	1 station (vertical channel)	

			sum of event pair radii)		
De Angelis and Henton (2011)	Soufrière Hills Volcano, Montserrat	0.75		3 stations (–)	
Zhang et al. (2005)	South Sandwich Islands region	0.79			
Schaff and Richards (2004)	China	0.80		1 station (BHZ channel)	
Schaff and Richards (2011)	in and near China	0.80		1 station (BHZ channel)	
Buurman and West (2010)	Augustine Volcano, Alaska	0.80		1 station (vertical channel)	
Petersen (2007)	Shishaldin Volcano, Alaska	0.80		1 station (vertical channel)	
Cannata et al. (2013)	Mt. Etna volcano, Italy	0.80		1 station (vertical channel)	
Li and Richards (2003)	South Sandwich Islands region	0.80		1 station (vertical channel)	visually check waveform similarity
Ma and Wu (2013)	Longmen Shan Fault Zone, China	0.80		1 station (three channels)	
Ma et al. (2014)	Longmen Shan Fault Zone, China	0.80		1 station (three channels)	
Nadeau and McEvelly (2004)	San Andreas Fault	0.80	cross-coherence, visual inspection, relocation, arrival time analysis		CC = 0.80 is used for preliminary scanning
Yu (2013)	Tonga-Kermadec-Vanuatu	0.80	entirely overlapping source areas and similar seismic moment	5 stations (vertical channel)	
Li et al. (2007)	Tangshan fault, China	0.85	average recurrence	1 station (–)	

			interval >100 days; internal consistency of travel time picking; relocation		
Zhang et al. (2008)	South Sandwich Islands, Aleutian Islands, Kuril Islands, Tonga–Fiji–Solomon Islands, Bucaramanga earthquake nest	0.90		1 station (vertical channel)	visually check waveform similarity
Li et al. (2011)	Longmen Shan fault zone, China	0.90	average recurrence interval >100 days; internal consistency of travel time picking; relative distance < rupture sizes	1 station (–)	
Cociani et al. (2010)	Gulf of Corinth, Greece	0.90	inter-event overlap confirmed by relocation and source dimension estimation	1 station (–)	
Hayward and Bostock (2017)	Queen Charlotte plate boundary, Canada	0.90	inter-event overlap suggested by coda wave interferometry	1 station (three channels)	
Bohnhoff et al. (2017)	Marmara, Turkey	0.90	recurrence time > 30 days; hypocentres $\leq 5$ km epicentral distance; magnitude difference $\leq \pm 0.2$	2 stations (vertical channel)	

Naoi et al. (2015)	Cooke 4 mine, South Africa	0.90	inter-event separation < half of the rupture radius of the larger event	2 stations (single channel)	
Yamaguchi et al. (2018)	Cooke 4 mine, South Africa	0.90	inter-event separation < half of the rupture radius of the larger event	3 stations (single channel)	
Schmittbuhl et al. (2016)	Main Marmara Fault, Turkey	0.90	confirm overlapping with waveform stretching and spectral analysis	1 station (vertical and horizontal channel)	
Zhao and Peng (2009)	Calaveras fault, California	0.90	magnitude difference < 1; 50% overlapping of the rupture area	3 stations (vertical channel)	CC threshold is the median value at $\geq 3$ stations
Meier et al. (2004)	Hellenic subduction zone	0.90		3 stations (-)	
Yao et al. (2017)	Nicoya Peninsula, Costa Rica	0.90	overlap of source area	9 channels	Network-averaged CC
Huang et al. (2020)	Ridgecrest, California	0.90	magnitude difference < 0.5; 50% overlapping of the rupture area; horizontal location error < $0.3 \times$ source radius	6 stations (vertical channel)	
Obana et al. (2003)	western Nankai Trough, Japan	0.93/0.95		2 stations (vertical channel)	
Shirzaei et al. (2013)	Hayward fault, California	0.95	coherency > 0.95	1 station (vertical channel)	
Peng and Ben-Zion (2005)	Karadere-Düzce branch of the North Anatolian Fault, Turkey	0.95			

Igarashi et al. (2003)	northeastern Japan	0.95		2 stations (vertical channel)	
Uchida et al. (2003)	northeastern Japan	0.95		2 stations (vertical channel)	
Matsuzawa et al. (2004)	east off northern Honshu, Japan	0.95		2 stations (vertical channel)	
Kimura et al. (2006)	Kanto, Japan	0.95		2 stations (-)	
Igarashi (2010)	Japan	0.95		2 stations (vertical channel)	
Yamashita et al. (2012)	southwestern Japan	0.95		2 stations (vertical channel)	
Kato and Igarashi (2012)	Tohoku, Japan	0.95		2 stations (-)	
Kato et al. (2012)	Tohoku, Japan	0.95		2 stations (-)	
Meng et al. (2015)	Northern Chile	0.95		2 stations (vertical channel)	
Huang and Meng (2018)	central Chile	0.95	magnitude difference $\leq 0.5$ ; stations span a distance $> 50$ km	2 stations (vertical channel)	
Taira et al. (2014)	San Juan Bautista, San Andreas fault	0.95	phase coherency $\geq 0.95$	2 stations (vertical channel)	
Nishikawa and Ide (2018)	Ibaraki-Oki, Japan	0.95	magnitude difference $\leq 0.5$	2 stations (three channels)	
Igarashi (2020)	Japan	0.95	composite selection criteria (both CC and S-P differential time)	2 stations (vertical channel)	
Matsubara et al. (2005)	northern Japan	0.95		3 stations (vertical channel)	
Kato and Nakagawa (2014)	Chile	0.95		4 stations (vertical channel)	

Warren-Smith et al. (2018)	New Zealand	0.95		4 stations (-)	
Chaves et al. (2020)	Nicoya Peninsula, Costa Rica	0.95		Multi-station (vertical channel)	Network-averaged CC
Nadeau et al. (1995)	Parkfield, California	0.98			
Nadeau and McEvelly (1999)	Parkfield, California	0.98			
Hatch et al. (2020)	Virginia, Nevada	0.98		4 stations (-)	

<sup>a</sup> The minimum number of stations required for a pair of events to be classified as repeaters with the CC value exceeding the threshold value. In this column, the employed channel(s) is given in the brackets if it is explicitly documented in the reference or confirmed by the author, otherwise a dash symbol “-” is denoted.

<sup>b</sup> For the cases using data from two or more stations, only Yao et al. (2017) and Chaves et al. (2020) calculate the CC value simultaneously across multiple stations, and then take the average. All others compute the CC value at individual stations separately and claim an event pair to be repeaters if a certain number of stations have the CC values exceeding a given CC threshold.

**Table 3.2.** A list of digital filters commonly used in identifying repeaters

Filter	References
No Filter	Warren-Smith et al. (2018)
1–20 Hz	Cannata et al. (2013); Kimura et al. (2006)
1–10 Hz	Li et al. (2007, 2011); Ma and Wu (2013); Ma et al. (2014); Cociani et al. (2010); Schmittbuhl et al. (2016)
1–8 Hz	Matsubara et al. (2005); Meng et al. (2015); Huang and Meng (2018) ; Taira et al. (2014)
0.5–5 Hz	Green and Neuberg (2006); Schaff and Richards (2004, 2011)
1–4 Hz	Igarashi et al. (2003); Igarashi (2010, 2020); Uchida et al. (2003); Matsuzawa et al. (2004); Kato et al. (2012); Meng et al. (2015); Huang and Meng (2018)

**Table 3.3.** A list of commonly assumed  $\Delta\sigma$  in estimating the ERR

Stress drop	References
3 MPa	Li et al. (2007); Lengliné and Marsan (2009); Schaff and Richards (2011); Shirzaei et al. (2013); Ma et al. (2014); Mesimeri and Karakostas (2018); Huang and Meng (2018); Huang et al. (2020); Igarashi (2020)
5 MPa	Li et al. (2011)
10 MPa	Igarashi et al. (2003); Matsuzawa et al. (2004); Uchida and Matsuzawa (2013); Hatakeyama et al. (2017)

## **Chapter 4. Complex 3D Migration and Delayed Triggering of Hydraulic Fracturing-Induced Seismicity**

### **4.1 Abstract and Plain Language Summary**

#### **4.1.1 Abstract**

Earthquakes resulting from hydraulic fracturing (HF) can have delayed triggering relative to injection commencement over a varied range of time scales, with many cases exhibiting the largest events near/after well completion. This poses serious challenges for risk mitigation and hazard assessment. Here, we document a high-resolution, three-dimensional source migration process with delayed mainshock triggering that is probably controlled by local hydrogeological conditions. Our results reveal that poroelastic effects might contribute to induced seismicity, but are likely insufficient to activate a non-critically stressed fault of sufficient size. The rapid pore-pressure build-up from HF can be very localized and capable of producing large, felt earthquakes on non-critically stressed fault segments. We interpret the delayed triggering as a manifestation of pore-pressure build-up along pre-existing faults needed to facilitate seismic failure. Our findings can deepen our understanding of the current stress state of crustal faults and also explain why so few injection operations are seismogenic.

#### **4.1.2 Plain Language Summary**

Fluid injection-induced earthquakes (IIE), especially the mainshocks, are often observed to occur near or after well completion. Such delayed triggering relative to injection commencement poses serious challenges for both regulators and the energy industry to establish an effective mitigation strategy for the potential seismic risk. In this study, we reveal a high-resolution, complex three-dimensional pattern of IIE migration in Fox Creek, Alberta, Canada. The observed first-outward-then-inward IIE sequence highlights the significance of hydrogeological networks in facilitating fluid pressure migration and the associated seismic failure. The detailed spatiotemporal distribution of IIE suggests that the effect of pore-pressure build-up from hydraulic fracturing (HF) can be very localized. The delayed triggering is probably a combined

result from the fluid pressure migration and the current stress state of the hosting fault system away from the HF wells. The findings from this study also provide plausible explanations on why only a very limited number of fluid injections are seismogenic.

## 4.2 Introduction

Fluid injection-induced earthquakes (IIE), especially relatively large ones, are often observed to have delayed triggering relative to injection commencement. For long-term wastewater disposal (WD), the delay time can be as long as decades (Keranen et al., 2013). For relatively short-term hydraulic fracturing (HF) operations, the delay time varies from days to weeks. In many cases, the largest events occur near or after well completion (Schultz et al., 2015a, 2015b, 2017, 2020; Schultz and Wang, 2020; Lei et al., 2017; Igonin et al., 2021; Wang et al., 2020; Peña-Castro et al., 2020) which severely challenges the designing of effective risk mitigation strategy. Understanding the controlling factor(s) of delayed triggering of induced seismicity is of paramount importance. However, the underlying physics is surprisingly far from clear due to limited observations and/or incomplete injection databases.

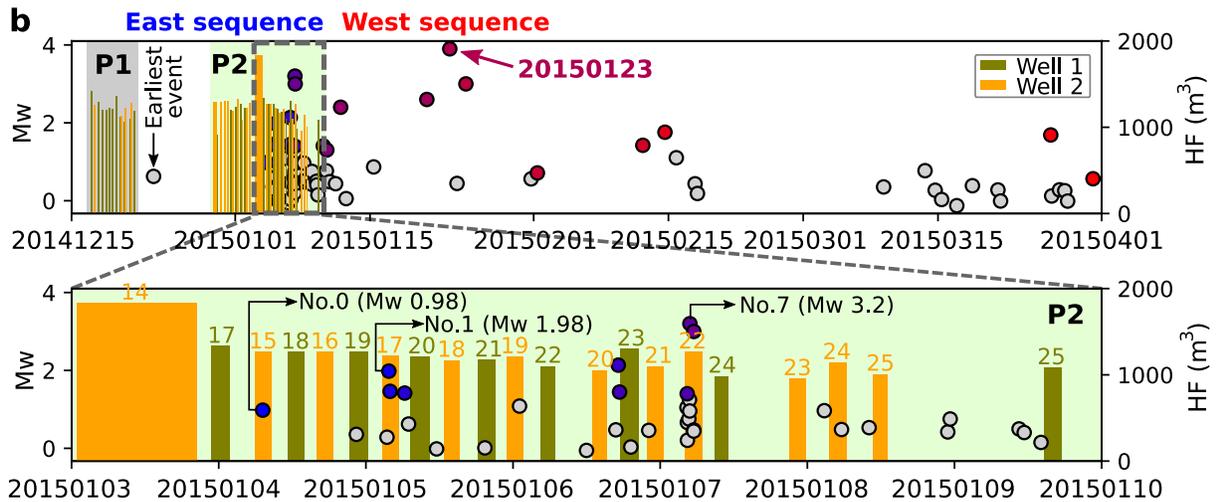
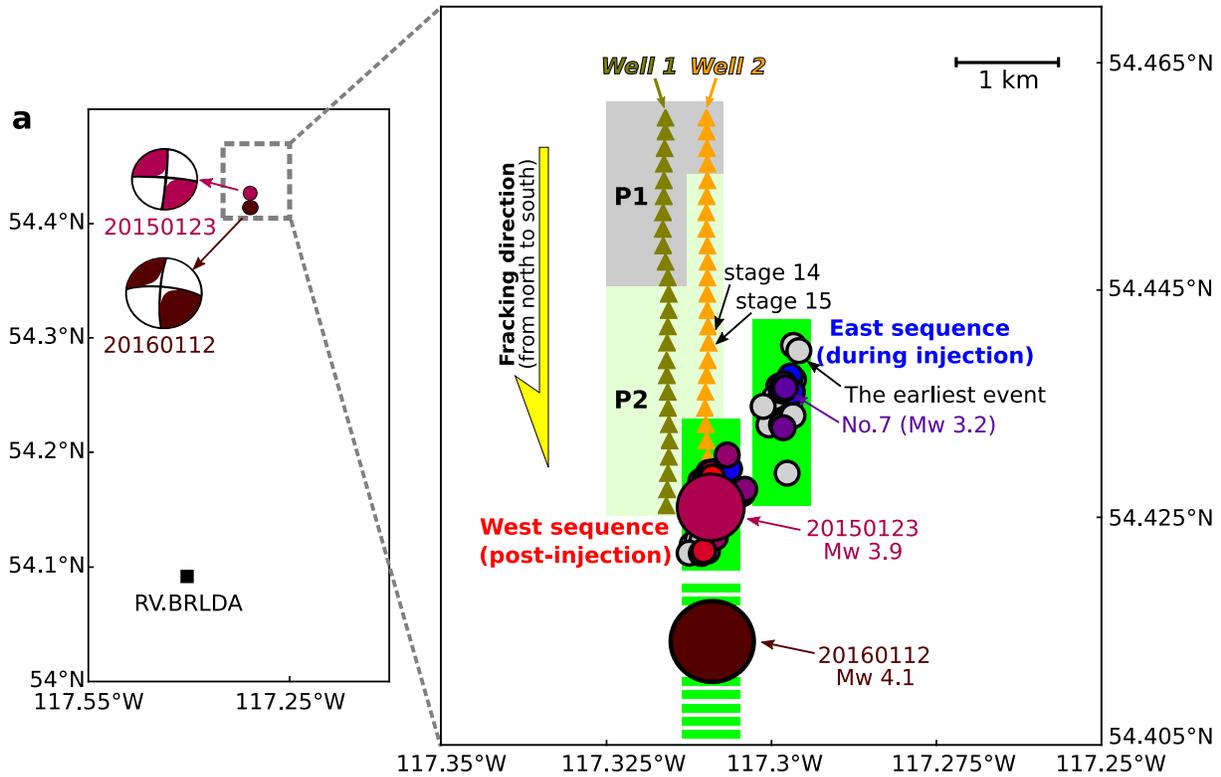
The 2015 Mw 3.9 earthquake sequence near Fox Creek, Alberta, Canada is the first well-known delayed HF-induced case with a ~2-week gap between the stimulation completion and the mainshock. The local seismograph array data contributed by the industry enables precise determination of earthquake hypocentres in comparison to other induced seismicity studies which often rely on regional stations (Bao and Eaton, 2016). During the post-stimulation process, only ~7% of the injected fluids, in contrast to a typical value of ~50% in western Canada, were recovered, unambiguously indicating that a tremendous amount of fluid has leaked off into nearby fault zones (Bao and Eaton, 2016). Given the robust earthquake locations, comprehensive stimulation database, and large volume of fluid loss, the 2015 Fox Creek sequence provides a unique opportunity to infer the corresponding three-dimensional (3D) fluid migration process and the spatiotemporal interactions between the hosting structures and injected fluid at an unprecedented resolution.

According to Bao and Eaton (2016), the Coulomb stress change ( $\Delta CFS$ ) due to fracture opening and pore-pressure diffusion are responsible for the earlier events that occurred during

the HF stimulation (referred to as the east sequence) and the delayed post-stimulation events (west sequence, including the Mw 3.9 strike-slip mainshock), respectively (Figure 4.1a). However, this model has at least two serious issues. First, it is inconsistent with the observed chronological sequence of stimulation and seismicity. There are two periods of stage stimulation from north to south with a  $\sim$ 1-week gap (Figure 4.1). The earliest event (i.e., the east sequence) actually occurred about 2 days after the last stage of the first stimulation period (P1 in Figure 4.1b). This is contradictory to their assumed elastic stress triggering mechanism, which should be instantaneous. Instead, the 2-day delay suggests that pressure migration might have begun during or shortly after P1. Moreover, the west sequence seems to initiate at greater depth relative to the injection well with a clear upward trend of propagation (Bao and Eaton, 2016). Hence it is very unlikely that the west sequence was caused by fluids from the wellbore directly above (Figure 4.1a).

Second, the initial model results in an overestimation of static  $\Delta$ CFS in triggering the earlier events (east sequence). The sudden increase of seismicity of the east sequence (including an Mw 3.2 earthquake) happened halfway through the second stimulation period (P2 in Figure 4.1b), when the treatment approached the vertical fault hosting the seismicity sequence (stages 14 and 15 of well 2 in Figure 4.1a). Thus, the actual  $\Delta$ CFS in triggering these events is significantly overestimated by simply summing the effects of all HF stages. Furthermore, the extremely large injected volume ( $\sim$ 50% more) and long duration ( $\sim$ 5.75 times longer) of stage 14 compared with other stages suggest the likely start time of serious fluid leakage (Figure 4.1b) (Peña-Castro et al., 2020). Consequently, it is inappropriate to calculate the net  $\Delta$ CFS by assuming that the total fracture (opening) volume equals the total volume of injected fluid (Bao and Eaton, 2016).

Here we revisit the 2015 Fox Creek sequence with tight constraints from local geological structures and injection parameters. We first employ waveform cross-correlation and hierarchical clustering analysis to identify near-identical events with highly similar waveforms. The distribution of these events is used to delineate the geometry of corresponding fault structures. We then analyse the spatiotemporal evolution of these on-fault near-identical events. By taking advantage of the complete stimulation database, we further conduct poroelastic modeling to investigate the delayed triggering process. Our results reveal a high-resolution, complex 3D pattern of IIE migration that is probably controlled by local fault architecture and its hydrogeological properties. Finally we discuss the broad implications of this study.



**Figure 4.1.** Comparison of induced seismicity and HF activity. (a), Location of the 2015 Mw 3.9 earthquake sequence in Fox Creek, Alberta, Canada. Beach balls of the 2015 Mw 3.9 and 2016 Mw 4.1 events are taken from prior work (Schultz et al., 2017). The black square in the left panel shows the seismic station used in this study. In the right panel, solid lime bars denote the two fault strands of the 2015 sequence; dashed lime bar marks the hosted fault of the 2016 Mw 4.1 event; triangles represent HF stages. (b), Injection history associated with the occurrence of induced earthquakes. The height of each colored bar represents the total volume of fluid injected at each stage while the width depicts the stage duration. In the bottom panel, stage ID is labelled above each treatment. In both (a) and (b), colored circles are the near-identical events (Figure 4.2); gray circles represent the uncorrelated small events; gray and pale green shaded areas represent P1 and P2 injection periods, respectively.

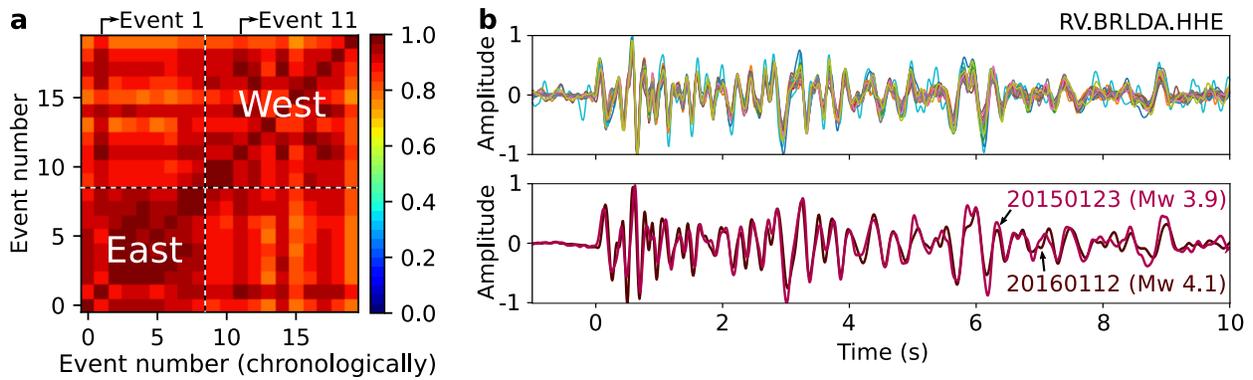
## 4.3 Methods

### 4.3.1 Waveform Cross-correlation and Hierarchical Clustering Analysis

Near-identical waveforms between events are commonly interpreted as indication of a similar source location and focal mechanism (Schultz et al., 2014). Here we directly adopt the high accuracy earthquake catalog (69 events in total) reported in the literature (Bao and Eaton, 2016) and perform pair-wise waveform cross-correlation and clustering analysis (Schultz et al., 2014, 2015, 2017; Hayward and Bostock, 2017) to identify near-identical events. The cataloged events were mainly determined by local seismograph array data contributed by the industry in addition to regional seismic stations and were relocated with hypoDD (Bao and Eaton, 2016).

The waveform similarity can be quantitatively characterized by cross-correlation coefficients (CC). Since data availability of the private seismograph array used by prior work (Bao and Eaton, 2016) is restricted, we choose to calculate the CC values between event pairs with seismograms from the station BRLDA (Figure 4.1a) that have a generally high signal-to-noise ratio (Schultz et al., 2015, 2017) and are publicly accessible from Incorporated Research Institutions for Seismology (<http://ds.iris.edu/ds/nodes/dmc/>, last accessed July 2020). Notice that the purpose of this study is to identify earthquakes that occurred on similar fault structures instead of repeating events, therefore we employ the classical one-segment cross-correlation method to calculate the CC values. The technical details of CC calculation are presented in Section 4.8.1.

The aforementioned pair-wise cross-correlation yields a  $[69 \times 69]$  similarity matrix. We obtain the near-identical events by implementing a hierarchical clustering algorithm based on the unweighted pair-group method using the average approach (UPGMA), available as a SciPy package (Jones et al., 2001, <https://docs.scipy.org>). Compared with the “chain-like” methods (e.g., Igarashi et al., 2003), the UPGMA method yields more robust results in grouping earthquakes (Hayward and Bostock, 2017). Here we define a cluster as a group of events in which the CC of all pairs are higher than 0.75 (Figure 4.2a). Such a CC threshold, the same as the value used in other IIE related clustering studies (e.g., Schultz et al., 2014; Cauchie et al., 2020), is determined by visually inspecting the waveforms in the corresponding cluster. Eventually, we obtain 1 cluster with 20 near-identical events. The high similarity of the event waveforms, including the coda train, justifies our choice of the threshold value (Figure 4.2b).



**Figure 4.2.** Identification of near-identical earthquakes. (a) Cross-correlation matrix of 20 correlated events (i.e., near-identical earthquakes). (b) Normalized waveforms of the 20 near-identical earthquakes (top panel) and of the two large events (bottom panel).

### 4.3.2 Poroelastic Modeling and $\Delta$ CFS Calculation

To investigate the predominant triggering mechanism, we conduct poroelastic modeling that takes into account the interaction between pore pressure change ( $\Delta P$ ) and rock matrix deformation. We use the COMSOL Multiphysics® software (version 5.3a) to model the evolution of pore pressure and poroelastic stress surrounding the two HF horizontal wells. COMSOL Multiphysics® software employs the finite-element algorithm to simulate the fluid-solid coupling in a realistic scenario, thus we can estimate the pore pressure and poroelastic stress simultaneously. In this study, we apply the solid mechanics module and Darcy’s fluid flow module to simulate the coupling process. The technical details of poroelastic modeling are given in Section 4.8.2.

The  $\Delta$ CFS has been commonly used to study the earthquake triggering process (e.g., Stein, 1999; Deng et al., 2016). After we obtain the stress tensor and pore pressure change from the COMSOL model, then we use the following equation to calculate the  $\Delta$ CFS resolved on the specific fault plane (Xu et al., 2010):

$$\begin{aligned}
\Delta\text{CFS} = & \sin \lambda \left[ -\frac{1}{2} \sin^2 \phi \sin(2\tilde{\delta}) \sigma^{11} + \frac{1}{2} \sin(2\phi) \sin(2\tilde{\delta}) \sigma^{12} + \sin \phi \cos(2\tilde{\delta}) \sigma^{13} - \right. \\
& \left. \frac{1}{2} \cos^2 \phi \sin(2\tilde{\delta}) \sigma^{22} - \cos \phi \sin(2\tilde{\delta}) \sigma^{23} + \frac{1}{2} \sin(2\tilde{\delta}) \sigma^{33} \right] + \\
& \cos \lambda \left[ -\frac{1}{2} \sin(2\phi) \sin \tilde{\delta} \sigma^{11} + \cos(2\phi) \sin \tilde{\delta} \sigma^{12} + \cos \phi \cos \tilde{\delta} \sigma^{13} + \right. \\
& \left. \frac{1}{2} \sin(2\phi) \sin \tilde{\delta} \sigma^{22} + \sin \phi \cos \tilde{\delta} \sigma^{23} \right] + \mu \left[ \sin^2 \phi \sin^2 \tilde{\delta} \sigma^{11} - \sin(2\phi) \sin^2 \tilde{\delta} \sigma^{12} - \right. \\
& \left. \sin \phi \sin(2\tilde{\delta}) \sigma^{13} + \cos^2 \phi \sin^2 \tilde{\delta} \sigma^{22} + \cos \phi \sin(2\tilde{\delta}) \sigma^{23} + \cos^2 \phi \sigma^{33} + \Delta P \right] \quad (4-1)
\end{aligned}$$

where  $\mu = 0.6$  is the friction coefficient,  $\phi$ ,  $\tilde{\delta}$ , and  $\lambda$  are the strike, dip, and rake of the receiver fault, respectively,  $\sigma^{ij}$  is the stress tensor, where  $i, j = 1, 2, 3$  are the 3D components in the Cartesian coordinate system, and  $\Delta P$  is the pore pressure change. Based on the Coulomb failure criteria, seismic slip is promoted for a positive  $\Delta\text{CFS}$ , and vice versa (King et al., 1994).

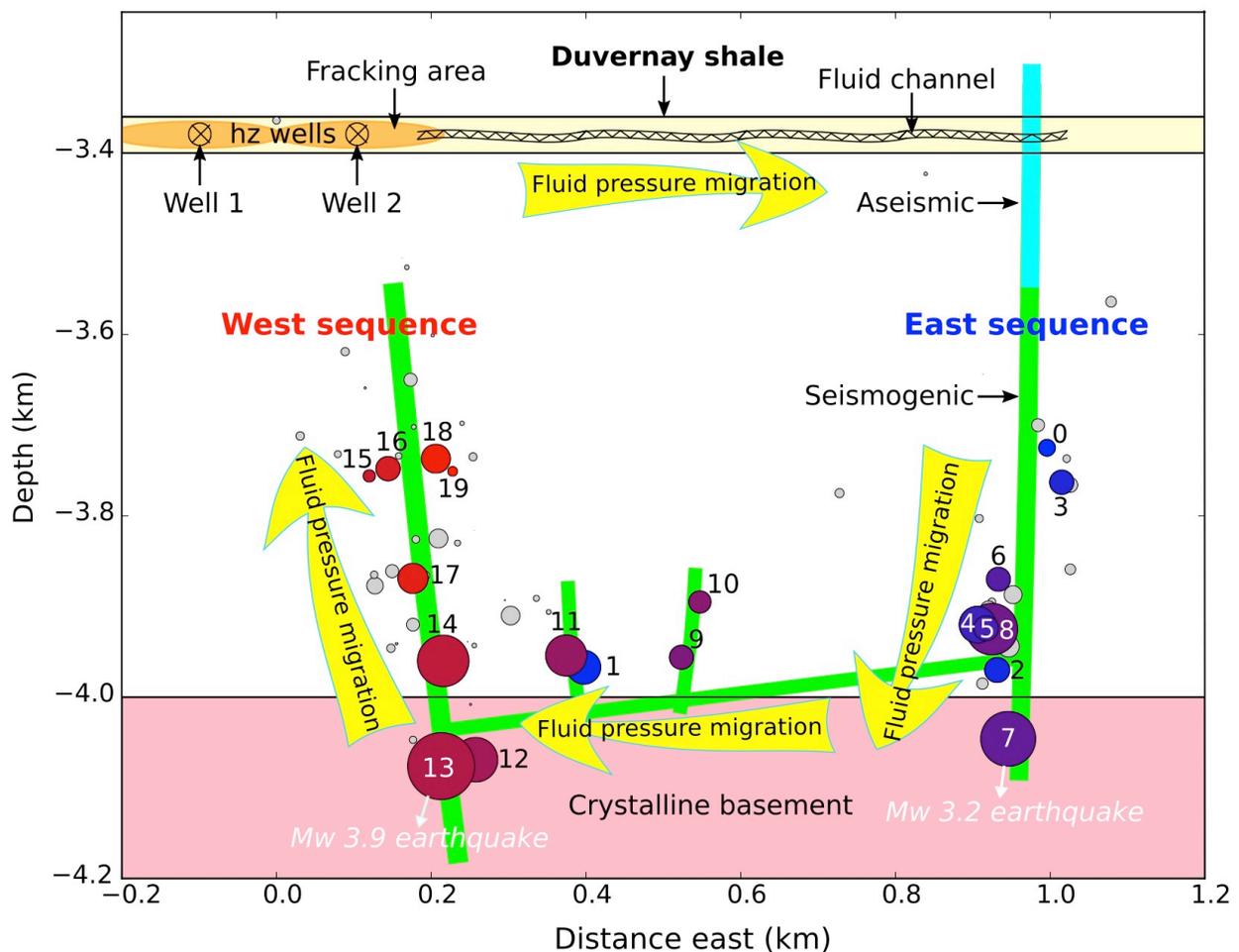
## 4.4 Results

### 4.4.1 A High-resolution 3D Pattern of IIE Migration

Based on the results of waveform analysis (Section 4.3.1), 20 out of 69 events are found with high CC ( $>0.75$ ) and near-identical waveforms, implying that they have ruptured on similar structures with similar focal mechanisms. Overall, the similarity matrix of these near-identical events shows two high CC patches (Figure 4.2a) – one corresponds to the earlier events in the east sequence and the other to the later events in the west sequence (Figure 4.3). Such a two-patch pattern is consistent with the two main near-vertical fault structures (Figure 4.3) inferred from earthquake focal mechanisms (Schultz et al., 2017). According to the “flower structure” model, these two near-vertical faults may merge together in the basement (Wang et al., 2017). The remaining 49 poorly correlated events are generally small (overall  $M_w \leq 1$ , Figure 4.5) and likely to have occurred on the nearby tiny fractures with possibly different orientations and/or focal mechanisms.

It is worth noting that the event magnitudes generally increase with focal depth for both the east and west sequences (Figure 4.3). The overall pattern of relative location among hypocentres should be very robust as they are determined by the high-resolution double-difference method (Waldhauser and Ellsworth, 2000) with data from a local seismic array established by the private

industry (Bao and Eaton, 2016). Most of the largest events appear to have occurred near/in the crystalline basement, possibly due to the varied degrees of fault maturity at different depths (Kozłowska et al., 2018). In comparison, there is no significant event immediately above or below the HF-targeted Duvernay shale formation (Figure 4.3). It appears that the aseismic region can extend up to 200 m surrounding the horizontal wells (Guglielmi et al., 2015; Eyre et al., 2019a).



**Figure 4.3.** East-west cross section of the 2015 Mw 3.9 earthquake sequence. 20 correlated events are marked by numbered and color-coded circles, with the smaller numbers and cooler colors corresponding to earlier events. Gray circles represent the 49 poorly correlated events. Circle sizes are scaled according to earthquake magnitudes. Crossed circles mark two HF horizontal wells.

The spatiotemporal evolution of the near-identical on-fault events (colored circles in Figure 4.3) clearly shows how the seismicity migrates in a 3D way: first in the east from shallow to deep, then shifting to the west, finally from deep to shallow. The seismicity migration, along with the huge fluid loss (Bao and Eaton, 2016), inherently implies the migration of the leaked fluid along pre-existing geological faults. Although the east sequence falls out of the target fracturing region (which is usually within a few hundred meters of the well), it is highly likely that a direct fluid connection exists between the injection well and triggered seismicity through permeable pathways. Such an inference is supported by many other cases documented in the literature (e.g., Wolhart et al., 2005; Davies et al., 2013; Galloway et al., 2018; Igonin et al., 2021) where the maximum fluid communication distance can be as far as ~1 km (Wilson et al., 2018; Igonin et al., 2021; Fu and Dehghanpour, 2020). The uppermost part of the east sequence fault seems to be aseismic, possibly due to the close proximity to the injection area (Guglielmi et al., 2015; De Barros et al., 2016) and/or high clay and organic content in the shale formation that favors stable sliding (Kohli and Zoback, 2013; Eyre et al., 2019a). Upon fluid injection, the fault permeability in the vicinity of fluid channel (i.e., preexisting permeable fracture corridors) may increase dramatically during the aseismic period (Guglielmi et al., 2015) which, in turn, facilitates rapid downward fluid pressure migration, eventually leading to seismic failures towards the basement. The fluid then migrates from east to west through faults in the basement as evident from the timing and location of the induced seismicity. Finally the fluid pressure may migrate vertically (Birdsell et al., 2015) along the west sequence fault hinted by the seismicity pattern (Haagenson and Rajaram, 2020). A lack of typical Omori-type aftershock sequences after the Mw 3.9 event on the west sequence fault (Bao and Eaton, 2016) provides another piece of evidence of the involvement of an external force (fluid pressure) (Lei et al., 2017; 2019) and thus explains the fluid's origin (from the east) and upward earthquake migration on the west.

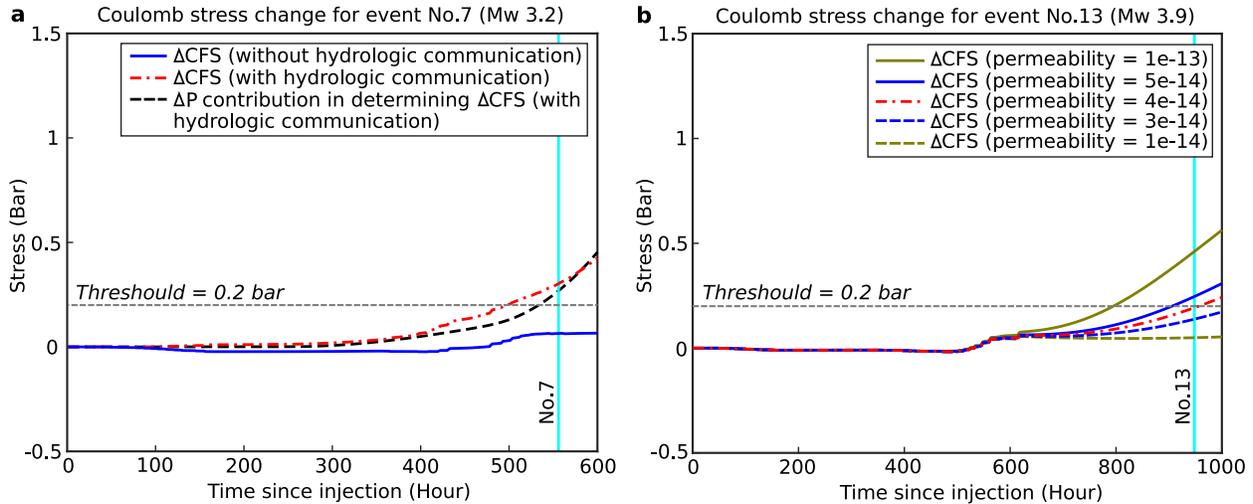
In summary, in contrast to the conventional wisdom that the geomechanical effects due to fluid injection migrate outward from the injection site, our results reveal a high-resolution, complex 3D pattern of IIE migration that can go both outward and inward as controlled by local fault architecture and its hydrogeological properties. The pore pressure build-up due to rapid fluid pressure migration has caused the Mw 3.2 earthquake on the east sequence fault and Mw 3.9 event on the west (Figure 4.3). This first-outward-then-inward sequence highlights the

significance of hydrologic networks in facilitating fluid pressure migration and the associated seismic failure. However, event No. 1 (Mw 1.98) appears to be an exception. It occurred very early (soon after the start of stage 17 of well 2 in P2, Figure 4.1b), not on the east sequence fault but on the west. The hypocentre is close to event No. 11 as evident by both high CC values (Figure 4.2a) and precise hypocentre locations (Figure 4.3). Given the timing and location, event No. 1 may have been caused by poroelastic effects rather than the pore pressure perturbation.

#### **4.4.2 Delayed Triggering Due to Pore Pressure Build-up**

We verify the hypothesis of the pore pressure build-up being the predominant triggering mechanism through poroelastic modelling (Section 4.3.2). In the model, we consider two scenarios for the east sequence fault: one where the near-vertical east sequence fault intersects the inferred horizontal fluid channel, and the other where it does not (Figure 4.3). Our model results indicate that the  $\Delta\text{CFS}$  due to poroelastic effects alone (i.e., without hydrologic communication) is only  $\sim 0.06$  bar (Figure 4.4a). Such a small change is likely insufficient to trigger the Mw 3.2 event on the east sequence fault as it is significantly below the triggering threshold (0.2 bar) adopted by previous studies (e.g., Fischer et al., 2008; Wang et al., 2021). Instead, allowing fluid pressure migration to the seismogenic east sequence fault can explain the observations very well. Figure 4.4a clearly shows that the  $\Delta P$  dominates the  $\Delta\text{CFS}$  in elevating stress to sufficient levels to cause the Mw 3.2 event. We also tested a range of physically reasonable permeability values (Cappa, 2009; Farrel and Taylor, 2014) for the inferred near-horizontal basement fault that facilitates rapid fluid pressure migration from the east sequence fault towards the west. A minimum permeability of  $4 \times 10^{-14} \text{ m}^2$ , about 4 orders higher than that of the low-permeability country rock ( $10^{-18} \text{ m}^2$ , Table 4.1), is found to be required to cause seismic failures on the west sequence fault for the observed time scale (Figure 4.4b). Such a high permeability value is consistent with the laboratory results of well-developed fault damage zones ( $10^{-16} - 10^{-14} \text{ m}^2$ ) that lead to rapid fluid flow (Evans et al., 1997). Thus, we conclude that the pore pressure build-up associated with fluid pressure migration is probably the key mechanism triggering the 2015 Fox Creek earthquake sequence, and that the complex 3D spatiotemporal pattern of hypocentres is dictated by the local hydrogeological setting. Our results also demonstrate that local hydrological pathways, fault structures, and a complete stimulation

database (e.g., accurate stage timing and volume) must all be properly incorporated in the modelling to avoid incorrect outcomes and misinterpretation (Bao and Eaton, 2016).



**Figure 4.4.** Poroelastic modelling results for the Mw 3.2 event (a) and Mw 3.9 event (b). Vertical cyan lines mark the origin times of the two earthquakes. Note for (b), the results of using a permeability lower than  $1 \times 10^{-14} \text{ m}^2$  are nearly identical to that of  $1 \times 10^{-14} \text{ m}^2$  and hence are not displayed for simplicity. In such low permeability cases, the  $\Delta P$  contribution in determining the  $\Delta CFS$  for the Mw 3.9 event is negligible as the fluid pressure can not reach the west sequence fault for the observed time scale. The corresponding  $\Delta P$  contributions of using different permeabilities in (b) are given in Figure 4.6.

## 4.5 Interpretation and Implications

### 4.5.1 Reactivation of A Non-critically Stressed Fault Segment

Previous studies have suggested that the hosting fault must be critically stressed for relatively large ( $M > 2$ ) IIE to occur (Atkinson et al., 2020). However, our observations suggest that the east sequence fault was probably not critically stressed before stimulation, as no event was triggered by poroelastic effects when the stage stimulation started. Instead, the largest event in the east sequence (event No. 7) occurred  $\sim 3$  days after event No. 0 (Figure 1b). The  $\sim 3$ -day delay time may suggest that stage stimulation can dramatically alter the stress state from non-critical to critical over an extremely short period (on the order of days), in contrast to the tectonic loading cycle (on the order of tens/hundreds of years).

Another hint of reactivating a non-critically stressed fault by HF comes from the west sequence fault that hosts the 2015 Mw 3.9 event. About one year later, another comparable-sized event (Mw 4.1) was also induced by HF slightly to the south (Figure 4.1a; Wang et al., 2017; Eyre et al., 2019a, b). These two events share near-identical focal mechanisms (Figure 4.1a) and waveforms (Figure 4.2b), have adjacent locations (epicentres less than 1.5 km apart, and similar depths within ~1 km; Schultz et al., 2017, Eyre et al., 2019b), and both occurred after the completion of HF operations with potentially significant fluid leakage (Bao and Eaton, 2016; Eyre et al., 2019b). Thus, the two large events are most likely to have occurred on two adjacent segments of the same N-S striking fault. Having two nearby ruptures of limited size instead of rupturing the whole west sequence fault at once suggests that the hosting fault is well below the critical state. This inference is also supported by a recent slip tendency analysis (Shen et al., 2019). Our observations indicate that the effect of HF stimulation can be very localized for a non-critically stressed fault given the relatively small injected volume. Therefore, it may only elevate the stress state of a limited segment of the hosting fault to facilitate seismic failure.

#### **4.5.2 Current Stress State of Crustal Faults**

Our observations show that both the east and west sequence faults were unlikely to be critically stressed before stimulation, as no large earthquakes occurred at the very beginning of stimulation and/or were caused by poroelastic effects. Furthermore, the west sequence fault hosted two large earthquakes of comparable size on neighbouring segments instead of rupturing the whole fault at once. Considering the facts that (i) most injection operations are not seismogenic (Atkinson et al., 2016; Schultz et al., 2017, 2020; Rubinstein and Mahani, 2015; Weingarten et al., 2015), (ii) events triggered by poroelastic effects are usually of small magnitudes (Deng et al., 2016; Kozłowska et al., 2018; Yu et al., 2019), (iii) the elevation of pore pressure is widely considered to be the primary cause of relatively large IIE (Lei et al., 2019; Peña-Castro et al., 2020; Wang et al., 2021; Schultz and Wang, 2020), and (iv) for the majority of HF-induced IIE cases, the largest events often occur near or after well completion (Schultz et al., 2015a, 2015b, 2017; Schultz and Wang, 2020; Lei et al., 2017; Igonin et al., 2021; Wang et al., 2020; Peña-Castro et al., 2020), we infer that the number of critically stressed, large

intraplate faults should be very limited, and that reactivation of such faults may require sufficient pore-pressure accumulation.

#### **4.5.3 Delayed Triggering of IIE**

Taking advantage of the high-resolution distribution of hypocentres and complete HF stimulation database, our study reveals that a complex 3D source migration process with the delay of large earthquakes is probably controlled by the local hydrogeological setting. Numerical modelling demonstrates that poroelastic effects alone (i.e., without direct hydrological connection) are likely insufficient to activate the east sequence fault. Instead, the delayed occurrence of two relatively large events (i.e., Mw 3.2 and Mw 3.9) on the time scale of days to weeks can be well-explained by the pore-pressure build-up along the complex local fault system involving an initially outward path at the shallow depth and a later inward one at a greater depth. Although the actual fluid channel and fault architecture could be even more complicated than what we have assumed (Figure 4.3), our model succeeds in explaining the IIE migration process to the first order.

Therefore, the complexity of the hydrologic network determines whether and how fast the fluid can reach the fault; and the current stress state of the hosting fault determines how long it takes for pore-pressure build-up to facilitate seismic failure. This might explain why no large IIE thus far occur at the onset of HF stimulation. Instead, they tend to occur near the end of, or even after the stage stimulation with a wide range of time delays (Schultz et al., 2015a, 2015b, 2017; Schultz and Wang, 2020; Lei et al., 2017; Igonin et al., 2021; Wang et al., 2020; Peña-Castro et al., 2020).

#### **4.5.4 Seismogenic vs. Aseismogenic Injection Operations**

Direct fluid communication should be geologically rare (Galloway et al., 2018). Whether earthquakes can be triggered by an injection operation depends on: (i) the probability of connecting the injection to a pre-existing seismogenic fault, and (ii) whether the amount of injected fluid is sufficient to bring the fault to critical state. Even if direct fluid communication exists, the largest magnitude of triggered events will depend on both the dimension of the pre-

existing fault and the cumulative volume of injected fluid (Schultz et al., 2018). Meeting all these conditions may be statistically demanding, and thus may explain why the majority of seismogenic wells do not produce large felt IIE. This essentially agrees with the Gutenberg-Richter law that smaller earthquakes occur much more frequently than the larger ones.

#### **4.6 Discussion and Conclusions**

Waveform similarity has been a powerful seismological tool recently to study earthquake source characteristics (Schultz et al., 2014, 2020). While there are increasing evidences that waveform CC alone cannot reliably distinguish repeating earthquakes from neighboring events (e.g., Ellsworth and Bulut, 2018), nearly identical waveforms are useful in identifying nearby earthquakes with similar focal mechanisms. In fact, using single-station CC values to identify earthquakes with similar origins has been a common practice in previous studies, especially for areas with limited station availability (e.g., Li and Richards, 2003; Schaff and Richards, 2004; Li et al., 2011; Buurman et al., 2013; Schultz et al., 2014, 2015, 2017; Yamada et al., 2016; Hayward and Bostock, 2017; Cauchie et al., 2020; Gao and Kao, 2020). We have tried different CC threshold values in our hierarchical clustering analysis, and the results are all similar. Although our cross-correlation and clustering analysis are based on single-station data, the overall match of the similarity matrix of the near-identical events (Figure 4.2a) and their hypocentre locations (Figure 4.3) demonstrate the effectiveness of our approach.

We take a more conservative approach in the investigation of the predominant triggering mechanism of IIE by assuming a triggering threshold of  $\Delta CFS = 0.2$  bar (e.g., Fischer et al., 2008; Wang et al., 2021). Some studies have considered a lower value of 0.1 bar to define the triggering threshold (e.g., Stein, 1999; King et al., 1994). Regardless which triggering threshold (0.1 or 0.2 bar) is used, the  $\Delta CFS$  due to poroelastic effects alone is much smaller (0.06 bar, Figure 4.4a) and hence is probably insufficient to trigger the Mw 3.2 event on the east sequence fault. We speculate that the poroelastic effects are at most a contributor in triggering the Mw 3.2 mainshock, whereas rapid pore-pressure build-up through permeable pathways may play a more important role.

To summarize, our study reveals that (i) poroelastic effects of HF stimulation might contribute to the occurrence of IIE, but are probably insufficient to activate a non-critically

stressed fault segment of sufficient size, (ii) the effect of HF can be very localized and non-critically stressed fault segments can produce large felt IIE with rapid pore-pressure build-up, and (iii) the spatiotemporal distribution of IIE can exhibit a very complicated 3D pattern depending on the specific local hydrogeological setting. Therefore, mapping pre-existing geological faults and avoiding direct hydrologic connection to them may be of paramount importance in mitigating short-term seismic hazard from IIE. Precise and accurate assessment of the state of stress of local fault systems is probably the key step in the strategy of maximizing the economic benefit of HF operations and minimizing the potential impact to the safety of local communities and infrastructure.

## **4.7 Acknowledgements**

Insightful discussions with Fengzhou Tan, Ramin Mohammad Hosseini Dokht, Yajing Liu, and Stan Dosso are much appreciated. Waveform data used in this study were downloaded from the Incorporated Research Institutions for Seismology (<http://ds.iris.edu/ds/nodes/dmc/>, last accessed July 2020). Seismic data are processed with Obspy (Beyreuther et al., 2010). Figures are made with Matplotlib (Hunter, 2007) and Inkscape (<https://inkscape.org>). This study is partially supported by a University of Victoria Fellowship (DG), the Induced Seismicity Research Project of NRCAN (HK), Geoscience BC (HK), and a NSERC Discovery Grant (HK). This paper is NRCAN contribution 2021xxxx.

## **4.8 Supporting Information**

### **4.8.1 Technical Details of CC Calculation**

When performing the waveform cross-correlation, the seismic waveforms are band-pass filtered from 1 to 10 Hz (Schmittbuhl et al., 2016; Warren-Smith et al., 2017, 2018). The cross-correlation window is set to be 10 s, starting from 1 s before to 9 s after the theoretical predicted S-arrival (Schultz et al., 2017) based on the ak135 velocity model (Kennett et al., 1995). The choice of a 10-s window length is meant to capture the strongest and cleanest arrival and sufficient coda waves with a lower level of noise contamination. We do not choose a window starting from the P phase because the P waves are very small compared with the S phases, thus

the result can be easily contaminated by noise (Schultz et al., 2017). The correlation is performed by sliding the waveform of one event from 4 s before the predicted S arrival of the other event to 4 s after, in one-sample increments. A  $\pm 4$  s shift should be adequate to account for any predicted phase onset error due to an imperfect velocity model. The maximum value of the CC results during the sliding is defined as the final CC value of the event pair.

#### 4.8.2 Technical Details of Poroelastic Modeling

By assuming that the medium is homogeneous and isotropic, the evolution of pore pressure can be calculated by solving the coupled diffusion equations, as listed below (equivalent forms of the equations can be found in the literature, e.g., Wang and Kumpel, 2003),

$$\rho S \frac{\partial p}{\partial t} - \nabla \cdot \left( \rho \frac{\kappa}{\mu_d} \nabla p \right) = Q_m(\mathbf{x}, t) - \rho \alpha \frac{\partial \varepsilon_{vol}}{\partial t} \quad (4-2)$$

$$S = \chi_f \theta + \chi_p (1 - \theta) \quad (4-3)$$

$$\mathbf{q} = - \frac{\kappa}{\mu_d} \nabla p \quad (4-4)$$

where  $\rho$  is the pore fluid density,  $S$  is the linearized storage parameter,  $p$  is the fluid pore pressure,  $\kappa$  is the permeability of the medium,  $\mu_d$  is its dynamic viscosity,  $Q_m$  is the volumetric flow rate for a fluid source,  $\alpha$  is the Biot-Willis coefficient,  $\varepsilon_{vol}$  is the volumetric strain of the porous matrix,  $\chi_f$  is the compressibility of the fluid,  $\chi_p$  is the compressibility of the rock,  $\theta$  is the porosity, and  $\mathbf{q}$  is the velocity variable which gives a volume flow rate per unit area of porous material. The governing equations for the poroelastic model are then given by:

$$-\nabla \cdot \boldsymbol{\sigma} = \mathbf{F}_v \quad (4-5)$$

$$\sigma_{ij} = \frac{2Gv}{(1-2v)} \varepsilon_{kk} \delta_{ij} + 2G\varepsilon_{ij} - \alpha p \delta_{ij} \quad (4-6)$$

$$\varepsilon_{ij} = \frac{1}{2} ((\nabla \mathbf{u})^T + \nabla \mathbf{u}) \quad (4-7)$$

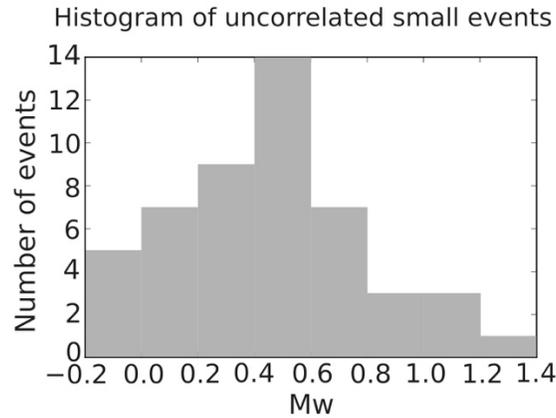
where  $\boldsymbol{\sigma}$  is the stress tensor,  $\mathbf{F}_v$  is the volume force vector (i.e.,  $\mathbf{F}_v = (\rho\theta + \rho_b)\mathbf{g}$ , where  $\mathbf{g}$  is the acceleration of gravity, and  $\rho_b$  is the bulk density),  $\delta_{ij}$  is the Kronecker delta (equal to 1 when  $i = j$ , and to 0 when  $i \neq j$ ),  $G$  is Young's modulus,  $v$  is the Poisson's ratio, and  $\mathbf{u}$  is the displacement vector.

We build a 3D model of  $5 \text{ km} \times 10 \text{ km} \times 5 \text{ km}$  in the x, y and z directions, respectively, and split the model into four simplified layers (Table 4.1). From top to bottom, the four layers correspond to the upper sedimentary section, the Duvernay shale formation in which the HF horizontal wells are located, the lower sedimentary section, and the crystalline basement (Bao and Eaton, 2016). The solid and hydrogeological properties of each layer are listed in Table 4.1. Within the model, we set the so-called roller condition as the side solid boundaries, i.e., no vertical movement is permitted for the solid material on the boundary. We then set the bottom and top solid boundaries as fixed and free surfaces, respectively. Next, we set the fluid boundaries to have no flow. In addition, at the top, we add a standard atmospheric pressure, and set the pore pressure at the top surface to 0. Finally, we set the original fluid condition to be hydrostatic equilibrium.

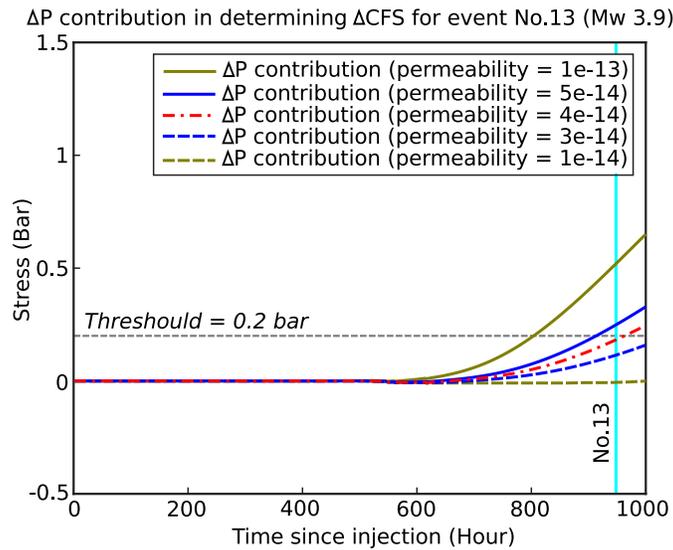
In our model, besides the stimulation points, we assume that the HF operations have created a fracture zone surrounding the horizontal wells (note that the fracture zone is confined in the Duvernay shale layer), leading to an increased permeability compared to the unfractured shale formation. The width of the fractures centred at the stimulation points is set to be 200 m. We assume the permeability of the fractures to be  $5 \times 10^{-15} \text{ m}^2$ , the same as that of the fluid channel but three orders higher than the low-permeability unfractured shale formation ( $1 \times 10^{-18} \text{ m}^2$ , Table 4.1). As mentioned in the main text, there are two inferred fault systems, i.e., the east sequence fault and the west sequence fault. In the model, we create two near-vertical faults on the basis of the Mw 3.9 and Mw 3.2 mainshock and their aftershock locations. We also assume that there is a near-horizontal basement fault connecting the two vertical fault systems (Figure 4.3). We set the permeability along the fault surface for the two vertical faults to be three orders of magnitude larger than the confining rock (Table 4.1), as the fault damage zone could enhance the permeability (Yehya et al., 2018). For the near-horizontal fault, it is worth noting that we have tested multiple permeability values, ranging from the same as the surrounding rock ( $1 \times 10^{-18} \text{ m}^2$ ) to five orders larger than the surrounding rock ( $1 \times 10^{-13} \text{ m}^2$ , Figure 4.4b). This range of permeability includes not only the scenario of a high-permeable horizontal fault, but also one where the horizontal fault does not exist.

To simulate the multi-stage fluid injection process, we assume that fluid is injected at a single point of each stage, and the consecutive stages migrate along the horizontal well bore. Each

stage's fluid injection rate is the ratio between stage injection volume and duration time (calculated from Table 4.2). The outcomes (stress tensor and pore pressure change) from the poroelastic modeling are then used to calculate the  $\Delta\text{CFS}$  as discussed in the main text.



**Figure 4.5.** Histogram of the 49 poorly correlated small earthquakes.



**Figure 4.6.**  $\Delta\text{P}$  contribution in determining the  $\Delta\text{CFS}$  for the Mw 3.9 event.

**Table 4.1.** Solid and fluid properties of each layer used in the model. Note that for the fracture zone, the permeability is  $5 \times 10^{-15} \text{ m}^2$  and the solid properties are the same as that of the confining Duvernay shale layer. For the two vertical faults, their solid properties are the same as the horizontal layers, and the permeability along the fault is  $5 \times 10^{-15} \text{ m}^2$ .

	Layer 1	Layer 2 (HF layer)	Layer 3	Layer 4
Depth	0-3.3 km	3.3 km-3.4 km	3.4 km-4.1 km	4.1 km-5 km
Biot-Willis	0.7	0.7	0.7	0.7
P-wave velocity	5000 m/s	6100 m/s	6300 m/s	6900 m/s
S-wave velocity	2800 m/s	3520 m/s	3630 m/s	3983 m/s
Bulk Density ( $\rho_b$ )	2500 kg/m <sup>3</sup>	2600 kg/m <sup>3</sup>	2750 kg/m <sup>3</sup>	2900 kg/m <sup>3</sup>
Permeability	$7.5 \times 10^{-16} \text{ m}^2$	$1 \times 10^{-18} \text{ m}^2$	$5 \times 10^{-18} \text{ m}^2$	$1 \times 10^{-18} \text{ m}^2$
Porosity ( $\theta$ )	0.1	0.05	0.05	0.08
Fluid density ( $\rho$ )	1000 kg/m <sup>3</sup>	1000 kg/m <sup>3</sup>	1000 kg/m <sup>3</sup>	1000 kg/m <sup>3</sup>
Fluid compressibility ( $\chi_f$ )	$4.5 \times 10^{-10} \text{ Pa}^{-1}$			
Fluid dynamic viscosity ( $\mu_d$ )	$0.79 \times 10^{-3} \text{ Pa*s}$			

**Table 4.2.** Injection data.

Well	Stage	Stage-start	Stage-end	Total Fluid (m <sup>3</sup> )
1	1	17-12-2014--00:38	17-12-2014--04:02	1417
1	2	17-12-2014--18:08	17-12-2014--21:30	1288
1	3	18-12-2014--04:17	18-12-2014--07:10	1196
1	4	18-12-2014--13:53	18-12-2014--16:31	1196
1	5	18-12-2014--21:37	19-12-2014--00:11	1220
1	6	19-12-2014--04:50	19-12-2014--07:25	1217
1	7	19-12-2014--14:30	19-12-2014--18:47	1367
1	8	19-12-2014--23:44	20-12-2014--02:16	1123
1	9	20-12-2014--08:46	20-12-2014--11:01	1065
1	10	21-12-2014--00:33	21-12-2014--02:50	1101
1	11	21-12-2014--10:23	21-12-2014--13:36	1189
1	12	30-12-2014--03:43	30-12-2014--06:00	907.4
1	13	31-12-2014--11:45	31-12-2014--14:23	1199
1	14	31-12-2014--23:28	01-01-2015--02:26	1234
1	15	01-01-2015--11:35	01-01-2015--14:20	1266
1	16	01-01-2015--22:40	02-01-2015--01:34	1226
1	17	03-01-2015--22:43	04-01-2015--01:45	1333
1	18	04-01-2015--11:17	04-01-2015--13:56	1265
1	19	04-01-2015--21:24	05-01-2015--00:21	1268
1	20	05-01-2015--07:21	05-01-2015--10:26	1210.7
1	21	05-01-2015--18:17	05-01-2015--21:10	1174
1	22	06-01-2015--04:29	06-01-2015--06:55	1102
1	23	06-01-2015--17:30	06-01-2015--20:33	1301
1	24	07-01-2015--08:54	07-01-2015--11:05	987
1	25	09-01-2015--14:35	09-01-2015--17:26	1084
2	1	17-12-2014--06:52	17-12-2014--10:44	1253
2	2	20-12-2014--03:56	20-12-2014--07:00	1128
2	3	20-12-2014--13:06	20-12-2014--15:55	1219
2	4	21-12-2014--04:44	21-12-2014--07:21	1282
2	5	29-12-2014--17:52	30-12-2014--01:40	1294

2	6	30-12-2014--11:45	30-12-2014--14:36	1290
2	7	30-12-2014--20:26	30-12-2014--23:11	1305
2	8	31-12-2014--04:33	31-12-2014--07:20	1309
2	9	31-12-2014--17:58	31-12-2014--20:53	1192
2	10	01-01-2015--06:03	01-01-2015--08:49	1324
2	11	01-01-2015--16:46	01-01-2015--19:02	1087
2	12	02-01-2015--03:29	02-01-2015--06:18	1219
2	13	02-01-2015--10:36	02-01-2015--13:19	1278
2	14	03-01-2015--00:58	03-01-2015--20:28	1834
2	15	04-01-2015--06:01	04-01-2015--08:39	1267
2	16	04-01-2015--16:04	04-01-2015--18:42	1266
2	17	05-01-2015--02:36	05-01-2015--05:20	1218
2	18	05-01-2015--12:44	05-01-2015--15:17	1171
2	19	05-01-2015--23:03	06-01-2015--01:45	1212
2	20	06-01-2015--13:02	06-01-2015--15:22	1056
2	21	06-01-2015--21:58	07-01-2015--00:35	1101
2	22	07-01-2015--04:03	07-01-2015--06:49	1267
2	23	07-01-2015--21:09	07-01-2015--23:45	957
2	24	08-01-2015--03:35	08-01-2015--06:26	1144
2	25	08-01-2015--10:42	08-01-2015--13:04	1010

## Chapter 5. Conclusions

This Ph.D. dissertation focuses on a comprehensive study of seismic waveform similarity, aiming at two themes: reliable identification of repeating earthquakes, and investigation of the detailed source process of induced seismicity through the 3D spatiotemporal evolution of mainly neighbouring events. The main contributions of this dissertation are summarized as follows.

Conventionally, repeating earthquakes are mainly identified with the MF method through a threshold value of CC. To optimize the performance of the MF method, I examine the effects of different operational parameters in determining CC and propose generic rules for selecting the length of waveform template and the frequency passband of the digital filter. With both synthetic and real earthquake data, I show that the presence of a large-amplitude wave train (e.g., S wave and surface waves) may bias the CC calculation and hence can lead to misidentification of repeating earthquakes (Chapter 2). To mitigate the impact of a large-amplitude wave train, I develop the MFMC method which essentially gives more weights to important low-amplitude phases such as depth phases and coda waves that carry additional source location information. The new method is more effective in recognizing the subtle waveform discrepancy between an event pair due to location difference and can be applied directly to continuous seismic waveforms to search for potential repeating earthquakes or it can be used as a post-analysis tool to quickly remove the misidentified repeating events from an existing repeater catalog or from the output of the conventional MF method.

Based on a large number of synthetic experiments and test results with high-quality borehole array waveform data, I find that waveform similarity is controlled by not only the inter-event distance but also many other factors, including station azimuth, epicentral distance, velocity structure, etc. (Chapter 3). The consequence is that a higher degree of waveform similarity does not necessarily represent a smaller hypocentre separation, and vice versa. Therefore, I contest the traditional view and conclude that it is difficult to reliably identify repeaters solely based on a given CC threshold. The waveform similarity can be an easy, quick, and convenient indicator in recognizing potential repeaters, but the final decision should depend on a physics-based approach considering both the inter-event overlap and magnitude difference. Quantitatively, we

define an event pair to be true repeaters if their inter-event distance is smaller than the rupture radius of the larger event and their magnitude difference is no more than 1. For the precise estimation of inter-event distance in cases with limited data, I develop the DTDD method which is based on the relative S-P differential traveltimes. I verify the effectiveness of the DTDD method and validate the physics-based approach in identifying repeaters using earthquakes occurred in the Fox Creek area, Alberta, Canada. It should be noted that precise estimation of the inter-event separation and rupture radius for a very large seismic dataset can be very labor-intensive and time-consuming. In such cases, using the MFMC technique developed in Chapter 2 for preliminary repeater scanning can significantly lower the workload of final verification.

The findings of Chapter 3 are of great significance. First, waveform similarity has long been used as the sole criterion for repeater identification in numerous earlier studies. The results in Chapter 3 simply imply that (1) the classical approach may not work as reliably as anticipated, and (2) previously identified repeaters potentially can be incorrect and their interpretations/hypotheses may be misleading. Hence a systematic reexamination of previous research results may be necessary. Second, waveform similarity also plays an important role in other studies such as coda wave interferometry. How the controlling factors of waveform similarity uncovered in Chapter 3 affect the performance and/or the results of other waveform-similarity-based methods should also be systematically investigated.

With an improved understanding of waveform similarity from the results of Chapters 2 and 3, I mainly identify neighbouring events with very similar waveforms to investigate the source characteristics of IIE in Chapter 4. By applying waveform cross-correlation and hierarchical clustering analysis to the 2015 Mw 3.9 earthquake sequence near Fox Creek, Alberta, I reveal a high-resolution three-dimensional source migration process that is probably controlled by local hydrogeological conditions. The mainshock occurred ~2-week after the HF stimulation was completed, a phenomenon known as delayed triggering. My analysis shows that poroelastic effects might be responsible for the occurrence of small earthquakes but it is likely insufficient to activate a non-critically stressed fault of sufficient size. The results also indicate that rapid pore-pressure build-up due to HF operations, despite the small injected volume, can be very localized and capable of producing large, felt earthquakes on non-critically stressed fault segments. I interpret the commonly observed delayed occurrence of large induced events as a manifestation of pore-pressure build-up along pre-existing geological faults needed for seismic failure. I further

infer that the number of critically stressed, large intraplate faults should be very limited, and that reactivation of such faults may require sufficient pore-pressure accumulation. The findings of this study may also explain why only a very small number of fluid injections are seismogenic.

For mitigating seismic hazard from IIE, future efforts can be made in three directions:

(1) More accurate mapping of pre-existing geological faults before injection.

As pore-pressure build-up along pre-existing geological faults is inferred to be the predominant triggering mechanism of large IIE, more accurate characterization of the subsurface fault structure before injection can help avoid direct hydrologic connection with the injection well and therefore significantly reduce the likelihood of fault reactivation. Additionally, accurate estimation of the fault dimension contributes directly to better seismic hazard assessment.

(2) More accurate assessment of the state of stress of local fault systems before and during injection.

The current stress state of the hosting fault determines how long it takes for pore-pressure build-up to facilitate seismic failure. Therefore, more accurate assessment of the state of stress of local fault zones is extremely important for both regulators and operators to establish an effective risk mitigation strategy while simultaneously maintaining the economic benefit of developing unconventional resources. I recommend that this assessment should focus on the fault systems at distances no more than ~1 km away from the injection well, which is the reported upper limit of fluid communication distance (e.g, Wilson et al., 2018; Igonin et al., 2019; Fu and Dehghanpour, 2020).

(3) Real-time monitoring of the fluid leakage during injection.

Fluid leakage will not only affect the efficiency of HF operations but also may trigger earthquakes on pre-existing geological faults by means of rapid pore-pressure accumulation through permeable pathways. Therefore, monitoring the potential fluid leakage in real-time may be the key step in the strategy of maximizing the fracking efficiency and mitigating short-term seismic hazard from IIE.

## Bibliography

- Abercrombie, R. E. (2014). Stress drops of repeating earthquakes on the San Andreas fault at Parkfield. *Geophysical Research Letters*, 41(24), 8784-8791.
- Atkinson, G. M., Eaton, D. W., & Igonin, N. (2020). Developments in understanding seismicity triggered by hydraulic fracturing. *Nature Reviews Earth & Environment*, 1-14.
- Atkinson, G. M., Eaton, D. W., Ghofrani, H., Walker, D., Cheadle, B., Schultz, R., ... & Kao, H. (2016). Hydraulic fracturing and seismicity in the Western Canada Sedimentary Basin. *Seismological research letters*, 87(3), 631-647.
- Baisch, S., Ceranna, L., & Harjes, H. P. (2008). Earthquake cluster: What can we learn from waveform similarity?. *Bulletin of the Seismological Society of America*, 98(6), 2806-2814.
- Bao, X., & Eaton, D. W. (2016). Fault activation by hydraulic fracturing in western Canada. *Science*, 354(6318), 1406-1409.
- Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A Python toolbox for seismology. *Seismological Research Letters*, 81(3), 530-533.
- Bohnhoff, M., Wollin, C., Domigall, D., Küperkoch, L., Martínez-Garzón, P., Kwiątek, G., ... & Malin, P. E. (2017). Repeating Marmara Sea earthquakes: indication for fault creep. *Geophysical Journal International*, 210(1), 332-339.
- Bouchon, M., Karabulut, H., Aktar, M., Özalaybey, S., Schmittbuhl, J., & Bouin, M. P. (2011). Extended nucleation of the 1999 Mw 7.6 Izmit earthquake. *science*, 331(6019), 877-880.
- Brune, J. N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of geophysical research*, 75(26), 4997-5009.
- Buurman, H., & West, M. E. (2010). Seismic precursors to volcanic explosions during the 2006 eruption of Augustine Volcano. In J. A. Power, M. L. Coombs, & J. T. Freymueller, (Eds.), *The 2006 eruption of Augustine Volcano, Alaska, Professional Paper (Vol. 1769, chap. 2, pp. 41–57)*. Washington, DC: U.S. Geological Survey.
- Buurman, H., West, M. E., & Thompson, G. (2013). The seismicity of the 2009 Redoubt eruption. *Journal of Volcanology and Geothermal Research*, 259, 16-30.
- Calderoni, G., Rovelli, A., & Di Giovambattista, R. (2015). Transient anomaly in fault zone-trapped waves during the preparatory phase of the 6 April 2009, Mw 6.3 L'Aquila earthquake. *Geophysical Research Letters*, 42(6), 1750-1757.
- Cannata, A., Alparone, S., & Ursino, A. (2013). Repeating volcano-tectonic earthquakes at Mt. Etna volcano (Sicily, Italy) during 1999–2009. *Gondwana research*, 24(3-4), 1223-1236.
- Cappa, F. (2009). Modelling fluid transfer and slip in a fault zone when integrating heterogeneous hydromechanical characteristics in its internal structure. *Geophysical Journal International*, 178(3), 1357-1362.
- Carrier, A., & Got, J. L. (2014). A maximum a posteriori probability time-delay estimation for seismic signals. *Geophysical Journal International*, 198(3), 1543-1553.

- Cauchie, L., Lengliné, O., & Schmittbuhl, J. (2020). Seismic asperity size evolution during fluid injection: case study of the 1993 Soultz-sous-Forêts injection. *Geophysical Journal International*, 221(2), 968-980.
- Chamberlain, C. J., & Townend, J. (2018). Detecting Real Earthquakes Using Artificial Earthquakes: On the Use of Synthetic Waveforms in Matched-Filter Earthquake Detection. *Geophysical Research Letters*, 45(21), 11-641.
- Chamberlain, C. J., Hopp, C. J., Boese, C. M., Warren-Smith, E., Chambers, D., Chu, S. X., ... & Townend, J. (2018). EQcorrscan: Repeating and near-repeating earthquake detection and analysis in python. *Seismological Research Letters*, 89(1), 173-181.
- Chamberlain, C. J., Townend, J., & Gerstenberger, M. C. (2020). RT-EQcorrscan: Near-Real-Time Matched-Filtering for Rapid Development of Dense Earthquake Catalogs. *Seismological Society of America*, 91(6), 3574-3584.
- Chaves, E. J., Schwartz, S. Y., & Abercrombie, R. E. (2020). Repeating earthquakes record fault weakening and healing in areas of megathrust postseismic slip. *Science advances*, 6(32), eaaz9317.
- Chen, K. H., Nadeau, R. M., & Rau, R. J. (2008). Characteristic repeating earthquakes in an arc-continent collision boundary zone: The Chihshang fault of eastern Taiwan. *Earth and Planetary Science Letters*, 276(3-4), 262-272.
- Cheng, X., Niu, F., Silver, P. G., Horiuchi, S., Takai, K., Iio, Y., & Ito, H. (2007). Similar microearthquakes observed in western Nagano, Japan, and implications for rupture mechanics. *Journal of Geophysical Research: Solid Earth*, 112(B4).
- Cociani, L., Bean, C. J., Lyon-Caen, H., Pacchiani, F., & Deschamps, A. (2010). Coseismic velocity variations caused by static stress changes associated with the 2001 Mw= 4.3 Agios Ioanis earthquake in the Gulf of Corinth, Greece. *Journal of Geophysical Research: Solid Earth*, 115(B7).
- Davies, R., Foulger, G., Bindley, A., & Styles, P. (2013). Induced seismicity and hydraulic fracturing for the recovery of hydrocarbons. *Marine and petroleum geology*, 45, 171-185.
- De Angelis, S., & Henton, S. M. (2011). On the feasibility of magma fracture within volcanic conduits: constraints from earthquake data and empirical modelling of magma viscosity. *Geophysical Research Letters*, 38(19).
- De Barros, L., Daniel, G., Guglielmi, Y., Rivet, D., Caron, H., Payre, X., ... & Gourlay, M. (2016). Fault structure, stress, or pressure control of the seismicity in shale? Insights from a controlled experiment of fluid-induced fault reactivation. *Journal of Geophysical Research: Solid Earth*, 121(6), 4506-4522.
- Deng, K., Liu, Y., & Harrington, R. M. (2016). Poroelastic stress triggering of the December 2013 Crooked Lake, Alberta, induced seismicity sequence. *Geophysical Research Letters*, 43(16), 8482-8491.
- Dodge, D. A., & Walter, W. R. (2015). Initial global seismic cross-correlation results: Implications for empirical signal detectors. *Bulletin of the Seismological Society of America*, 105(1), 240-256.

- Eaton, D. W., Igonin, N., Poulin, A., Weir, R., Zhang, H., Pellegrino, S., & Rodriguez, G. (2018). Induced seismicity characterization during hydraulic-fracture monitoring with a shallow-wellbore geophone array and broadband sensors. *Seismological Research Letters*, 89(5), 1641-1651.
- Ellsworth, W. L., & Beroza, G. C. (1995). Seismic evidence for an earthquake nucleation phase. *Science*, 268(5212), 851-855.
- Ellsworth, W. L., & Bulut, F. (2018). Nucleation of the 1999 Izmit earthquake by a triggered cascade of foreshocks. *Nature Geoscience*, 11(7), 531-535.
- Evans, J. P., Forster, C. B., & Goddard, J. V. (1997). Permeability of fault-related rocks, and implications for hydraulic structure of fault zones. *Journal of structural Geology*, 19(11), 1393-1404.
- Eyre, T. S., Eaton, D. W., Garagash, D. I., Zecevic, M., Venieri, M., Weir, R., & Lawton, D. C. (2019a). The role of aseismic slip in hydraulic fracturing-induced seismicity. *Science advances*, 5(8), eaav7172.
- Eyre, T. S., Eaton, D. W., Zecevic, M., D'Amico, D., & Kolos, D. (2019b). Microseismicity reveals fault activation before M w 4.1 hydraulic-fracturing induced earthquake. *Geophysical Journal International*, 218(1), 534-546.
- Farrell, N. J. C., Healy, D., & Taylor, C. W. (2014). Anisotropy of permeability in faulted porous sandstones. *Journal of Structural Geology*, 63, 50-67.
- Fischer, A. D., Peng, Z., & Sammis, C. G. (2008). Dynamic triggering of high-frequency bursts by strong motions during the 2004 Parkfield earthquake sequence. *Geophysical Research Letters*, 35(12).
- Fu, Y., & Dehghanpour, H. (2020). How far can hydraulic fractures go? A comparative analysis of water flowback, tracer, and microseismic data from the Horn River Basin. *Marine and Petroleum Geology*, 115, 104259.
- Galloway, E., Hauck, T., Corlett, H., Panã, D., & Schultz, R. (2018). Faults and associated karst collapse suggest conduits for fluid flow that influence hydraulic fracturing-induced seismicity. *Proceedings of the National Academy of Sciences*, 115(43), E10003-E10012.
- Gao, D., & Kao, H. (2020). Optimization of the Match-Filtering Method for Robust Repeating Earthquake Detection: The Multisegment Cross-Correlation Approach. *Journal of Geophysical Research: Solid Earth*, 125(7), e2020JB019714.
- Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, 165(1), 149-166.
- Gomberg, J. (2018). Unsettled earthquake nucleation. *Nature Geoscience*, 11(7), 463-464.
- Green, D. N., & Neuberg, J. (2006). Waveform classification of volcanic low-frequency earthquake swarms and its implication at Soufrière Hills Volcano, Montserrat. *Journal of Volcanology and Geothermal Research*, 153(1-2), 51-63.
- Guglielmi, Y., Cappa, F., Avouac, J. P., Henry, P., & Elsworth, D. (2015). Seismicity triggered by fluid injection-induced aseismic slip. *Science*, 348(6240), 1224-1226.

- Harrington, R. M., & Brodsky, E. E. (2009). Source duration scales with magnitude differently for earthquakes on the San Andreas Fault and on secondary faults in Parkfield, California. *Bulletin of the Seismological Society of America*, 99(4), 2323-2334.
- Harris, D. B. (1991). A waveform correlation method for identifying quarry explosions. *Bulletin of the Seismological Society of America*, 81(6), 2395-2418.
- Hatakeyama, N., Uchida, N., Matsuzawa, T., & Nakamura, W. (2017). Emergence and disappearance of interplate repeating earthquakes following the 2011 M9.0 Tohoku-oki earthquake: Slip behavior transition between seismic and aseismic depending on the loading rate. *Journal of Geophysical Research: Solid Earth*, 122(7), 5160-5180.
- Hatch, R. L., Abercrombie, R. E., Ruhl, C. J., & Smith, K. D. (2020). Evidence of Aseismic and Fluid-Driven Processes in a Small Complex Seismic Swarm Near Virginia City, Nevada. *Geophysical Research Letters*, 47(4), e2019GL085477.
- Hayward, T. W., & Bostock, M. G. (2017). Slip behavior of the queen Charlotte plate boundary before and after the 2012, MW 7.8 Haida Gwaii earthquake: evidence from repeating earthquakes. *Journal of Geophysical Research: Solid Earth*, 122(11), 8990-9011.
- Hotovec-Ellis, A. J., & Jeffries, C. (2016). Near real-time detection, clustering, and analysis of repeating earthquakes: Application to Mount St. Helens and Redoubt volcanoes. In *Seismological Society of America Annual Meeting*.
- Huang, H., & Meng, L. (2018). Slow unlocking processes preceding the 2015 Mw 8.4 Illapel, Chile, earthquake. *Geophysical Research Letters*, 45(9), 3914-3922.
- Huang, H., Meng, L., Bürgmann, R., Wang, W., & Wang, K. (2020). Spatio-temporal foreshock evolution of the 2019 M 6.4 and M 7.1 Ridgecrest, California earthquakes. *Earth and Planetary Science Letters*, 551, 116582.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(3), 90-95.
- Igarashi, T. (2010). Spatial changes of inter-plate coupling inferred from sequences of small repeating earthquakes in Japan. *Geophysical research letters*, 37(20).
- Igarashi, T. (2020). Catalog of small repeating earthquakes for the Japanese Islands. *Earth, Planets and Space*, 72, 1-8.
- Igarashi, T., Matsuzawa, T., & Hasegawa, A. (2003). Repeating earthquakes and interplate aseismic slip in the northeastern Japan subduction zone. *Journal of Geophysical Research: Solid Earth*, 108(B5).
- Igonin, N., Verdon, J. P., Kendall, M., & Eaton, D. (2019, June). The role of parallel fracture networks for induced seismicity in the Duvernay Formation. In *81st EAGE Conference and Exhibition 2019 Workshop Programme (Vol. 2019, No. 1, pp. 1-5)*. European Association of Geoscientists & Engineers.
- Igonin, N., Verdon, J. P., Kendall, J. M., & Eaton, D. W. (2021). Large-scale fracture systems are permeable pathways for fault activation during hydraulic fracturing. *Journal of Geophysical Research: Solid Earth*, 126(3), e2020JB020311.

- Jiang, C., Wu, Z., Li, Y., & Ma, T. (2014). "Repeating Events" as Estimator of Location Precision: The China National Seismograph Network. *Pure and Applied Geophysics*, 171(3-5), 413-423.
- Jones, E., Oliphant, T., & Peterson, P. (2001). *SciPy: Open source scientific tools for Python*.
- Kato, A., & Igarashi, T. (2012). Regional extent of the large coseismic slip zone of the 2011 Mw 9.0 Tohoku-Oki earthquake delineated by on-fault aftershocks. *Geophysical Research Letters*, 39(15).
- Kato, A., & Nakagawa, S. (2014). Multiple slow-slip events during a foreshock sequence of the 2014 Iquique, Chile Mw 8.1 earthquake. *Geophysical Research Letters*, 41(15), 5420-5427.
- Kato, A., Obara, K., Igarashi, T., Tsuruoka, H., Nakagawa, S., & Hirata, N. (2012). Propagation of slow slip leading up to the 2011 Mw 9.0 Tohoku-Oki earthquake. *Science*, 335(6069), 705-708.
- Kennett, B. L., Engdahl, E. R., & Buland, R. (1995). Constraints on seismic velocities in the Earth from traveltimes. *Geophysical Journal International*, 122(1), 108-124.
- Keranen, K. M., Savage, H. M., Abers, G. A., & Cochran, E. S. (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence. *Geology*, 41(6), 699-702.
- Kilb, D., & Rubin, A. M. (2002). Implications of diverse fault orientations imaged in relocated aftershocks of the Mount Lewis, ML 5.7, California, earthquake. *Journal of Geophysical Research: Solid Earth*, 107(B11), ESE-5.
- Kim, A., Dreger, D. S., Taira, T. A., & Nadeau, R. M. (2016). Changes in repeating earthquake slip behavior following the 2004 Parkfield main shock from waveform empirical Green's functions finite-source inversion. *Journal of Geophysical Research: Solid Earth*, 121(3), 1910-1926.
- Kimura, H., Kasahara, K., Igarashi, T., & Hirata, N. (2006). Repeating earthquake activities associated with the Philippine Sea plate subduction in the Kanto district, central Japan: A new plate configuration revealed by interplate aseismic slips. *Tectonophysics*, 417(1-2), 101-118.
- King, G. C., Stein, R. S., & Lin, J. (1994). Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America*, 84(3), 935-953.
- Kohli, A. H., & Zoback, M. D. (2013). Frictional properties of shale reservoir rocks. *Journal of geophysical research: solid earth*, 118(9), 5109-5125.
- Kozłowska, M., Brudzinski, M. R., Friberg, P., Skoumal, R. J., Baxter, N. D., & Currie, B. S. (2018). Maturity of nearby faults influences seismic hazard from hydraulic fracturing. *Proceedings of the National Academy of Sciences*, 115(8), E1720-E1729.
- Kraft, T., & Deichmann, N. (2014). High-precision relocation and focal mechanism of the injection-induced seismicity at the Basel EGS. *Geothermics*, 52, 59-73.
- Kuna, V. M., Nábělek, J. L., & Braunmiller, J. (2019). Mode of slip and crust–mantle interaction at oceanic transform faults. *Nature Geoscience*, 12(2), 138-142.

- Lei, X., Huang, D., Su, J., Jiang, G., Wang, X., Wang, H., ... & Fu, H. (2017). Fault reactivation and earthquakes with magnitudes of up to Mw4.7 induced by shale-gas hydraulic fracturing in Sichuan Basin, China. *Scientific reports*, 7(1), 1-12.
- Lei, X., Wang, Z., & Su, J. (2019). The December 2018 ML 5.7 and January 2019 ML 5.3 earthquakes in South Sichuan basin induced by shale gas hydraulic fracturing. *Seismological Research Letters*, 90(3), 1099-1110.
- Lenclivé, O., & Marsan, D. (2009). Inferring the coseismic and postseismic stress changes caused by the 2004 Mw= 6 Parkfield earthquake from variations of recurrence times of microearthquakes. *Journal of Geophysical Research: Solid Earth*, 114(B10).
- Li, A., & Richards, P. G. (2003). Using earthquake doublets to study inner core rotation and seismicity catalog precision. *Geochemistry, Geophysics, Geosystems*, 4(9).
- Li, L., Chen, Q. F., Cheng, X., & Niu, F. (2007). Spatial clustering and repeating of seismic events observed along the 1976 Tangshan fault, north China. *Geophysical research letters*, 34(23).
- Li, L., Chen, Q. F., Niu, F., & Su, J. (2011). Deep slip rates along the Longmen Shan fault zone estimated from repeating microearthquakes. *Journal of Geophysical Research: Solid Earth*, 116(B9).
- Li, L., Niu, F., Chen, Q. F., Su, J., & He, J. (2017). Post-seismic velocity changes along the 2008 M 7.9 Wenchuan earthquake rupture zone revealed by S coda of repeating events. *Geophysical Journal International*, 208(2), 1237-1249.
- Lockner, D. A., & Beeler, N. M. (1999). Premonitory slip and tidal triggering of earthquakes. *Journal of Geophysical Research: Solid Earth*, 104(B9), 20133-20151.
- Ma, S. (2010). Focal depth determination for moderate and small earthquakes by modeling regional depth phases sPg, sPmP, and sPn. *Bulletin of the Seismological Society of America*, 100(3), 1073-1088.
- Ma, S., & Atkinson, G. M. (2006). Focal depths for small to moderate earthquakes ( $m_N \geq 2.8$ ) in Western Quebec, Southern Ontario, and Northern New York. *Bulletin of the Seismological Society of America*, 96(2), 609-623.
- Ma, X. J., & Wu, Z. L. (2013). 'Negative repeating doublets' in an aftershock sequence. *Earth, Planets and Space*, 65(8), 923-927.
- Ma, X., Wu, Z., & Jiang, C. (2014). 'Repeating earthquakes' associated with the WFSD-1 drilling site. *Tectonophysics*, 619, 44-50.
- Materna, K., Taira, T. A., & Bürgmann, R. (2018). Aseismic transform fault slip at the Mendocino triple junction from characteristically repeating earthquakes. *Geophysical Research Letters*, 45(2), 699-707.
- Matsubara, M., Yagi, Y., & Obara, K. (2005). Plate boundary slip associated with the 2003 Off-Tokachi earthquake based on small repeating earthquake data. *Geophysical Research Letters*, 32(8).

- Matsuzawa, T., Uchida, N., Igarashi, T., Okada, T., & Hasegawa, A. (2004). Repeating earthquakes and quasi-static slip on the plate boundary east off northern Honshu, Japan. *Earth, planets and space*, 56(8), 803-811.
- Meier, T., Rische, M., Endrun, B., Vafidis, A., & Harjes, H. P. (2004). Seismicity of the Hellenic subduction zone in the area of western and central Crete observed by temporary local seismic networks. *Tectonophysics*, 383(3-4), 149-169.
- Meng, L., Huang, H., Bürgmann, R., Ampuero, J. P., & Strader, A. (2015). Dual megathrust slip behaviors of the 2014 Iquique earthquake sequence. *Earth and Planetary Science Letters*, 411, 177-187.
- Mesimeri, M., & Karakostas, V. (2018). Repeating earthquakes in western Corinth Gulf (Greece): implications for aseismic slip near locked faults. *Geophysical Journal International*, 215(1), 659-676.
- Myhill, R., McKenzie, D., & Priestley, K. (2011). The distribution of earthquake multiplets beneath the southwest Pacific. *Earth and Planetary Science Letters*, 301(1-2), 87-97.
- Nadeau, R. M., & Johnson, L. R. (1998). Seismological studies at Parkfield VI: Moment release rates and estimates of source parameters for small repeating earthquakes. *Bulletin of the Seismological Society of America*, 88(3), 790-814.
- Nadeau, R. M., & McEvilly, T. V. (1999). Fault slip rates at depth from recurrence intervals of repeating microearthquakes. *Science*, 285(5428), 718-721.
- Nadeau, R. M., & McEvilly, T. V. (2004). Periodic pulsing of characteristic microearthquakes on the San Andreas fault. *Science*, 303(5655), 220-222.
- Nadeau, R. M., Foxall, W., & McEvilly, T. V. (1995). Clustering and periodic recurrence of microearthquakes on the San Andreas fault at Parkfield, California. *Science*, 267(5197), 503-507.
- Naoi, M., Nakatani, M., Igarashi, T., Otsuki, K., Yabe, Y., Kgarume, T., ... & Moriya, H. (2015). Unexpectedly frequent occurrence of very small repeating earthquakes ( $-5.1 \leq MW \leq -3.6$ ) in a South African gold mine: implications for monitoring intraplate faults. *Journal of Geophysical Research: Solid Earth*, 120(12), 8478-8493.
- Nishikawa, T., & Ide, S. (2018). Recurring slow slip events and earthquake nucleation in the source region of the M 7 Ibaraki-Oki earthquakes revealed by earthquake swarm and foreshock activity. *Journal of Geophysical Research: Solid Earth*, 123(9), 7950-7968.
- Obana, K., Kodaira, S., Kaneda, Y., Mochizuki, K., Shinohara, M., & Suyehiro, K. (2003). Microseismicity at the seaward updip limit of the western Nankai Trough seismogenic zone. *Journal of Geophysical Research: Solid Earth*, 108(B10).
- Ozalaybey, S., Ergin, M., Aktar, M., Tapirdamaz, C., Biçmen, F., & Yörük, A. (2002). The 1999 Izmit earthquake sequence in Turkey: seismological and tectonic aspects. *Bulletin of the Seismological Society of America*, 92(1), 376-386.
- Pacheco, K., Nishimura, T., & Nakahara, H. (2017). Seismic velocity changes of P and S waves associated with the 2011 Tohoku-Oki earthquake (M w 9.0) as inferred from analyses of repeating earthquakes. *Geophysical Journal International*, 209(1), 517-533.

- Peña Castro, A. F., Roth, M. P., Verdecchia, A., Onwuemeka, J., Liu, Y., Harrington, R. M., ... & Kao, H. (2020). Stress Chatter via Fluid Flow and Fault Slip in a Hydraulic Fracturing-Induced Earthquake Sequence in the Montney Formation, British Columbia. *Geophysical Research Letters*, 47(14), e2020GL087254.
- Peng, Z., & Ben-Zion, Y. (2005). Spatiotemporal variations of crustal anisotropy from similar events in aftershocks of the 1999 M 7.4 Izmit and M 7.1 Düzce, Turkey, earthquake sequences. *Geophysical Journal International*, 160(3), 1027-1043.
- Peng, Z., & Zhao, P. (2009). Migration of early aftershocks following the 2004 Parkfield earthquake. *Nature Geoscience*, 2(12), 877.
- Petersen, T. (2007). Swarms of repeating long-period earthquakes at Shishaldin Volcano, Alaska, 2001–2004. *Journal of Volcanology and Geothermal Research*, 166(3-4), 177-192.
- Poupinet, G., Ellsworth, W. L., & Frechet, J. (1984). Monitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras Fault, California. *Journal of Geophysical Research: Solid Earth*, 89(B7), 5719-5731.
- Rau, R. J., Chen, K. H., & Ching, K. E. (2007). Repeating earthquakes and seismic potential along the northern Longitudinal Valley fault of eastern Taiwan. *Geophysical research letters*, 34(24).
- Robinson, D. J., Sambridge, M., & Snieder, R. (2011). A probabilistic approach for estimating the separation between a pair of earthquakes directly from their coda waves. *Journal of Geophysical Research: Solid Earth*, 116(B4).
- Ross, Z. E., Trugman, D. T., Hauksson, E., & Shearer, P. M. (2019). Searching for hidden earthquakes in Southern California. *Science*, 364(6442), 767-771.
- Rubinstein, J. L., & Mahani, A. B. (2015). Myths and facts on wastewater injection, hydraulic fracturing, enhanced oil recovery, and induced seismicity. *Seismological Research Letters*, 86(4), 1060-1067.
- Ryan, W. B., Carbotte, S. M., Coplan, J. O., O'Hara, S., Melkonian, A., Arko, R., ... & Bonczkowski, J. (2009). Global multi-resolution topography synthesis. *Geochemistry, Geophysics, Geosystems*, 10(3).
- Sáez, M., Ruiz, S., Ide, S., & Sugioka, H. (2019). Shallow nonvolcanic tremor activity and potential repeating earthquakes in the Chile Triple Junction: seismic evidence of the subduction of the active Nazca–Antarctic spreading center. *Seismological Research Letters*, 90(5), 1740-1747.
- Salvage, R. O., & Neuberg, J. W. (2016). Using a cross correlation technique to refine the accuracy of the Failure Forecast Method: Application to Soufrière Hills volcano, Montserrat. *Journal of Volcanology and Geothermal Research*, 324, 118-133.
- Sawazaki, K., Kimura, H., Shiomi, K., Uchida, N., Takagi, R., & Snieder, R. (2015). Depth-dependence of seismic velocity change associated with the 2011 Tohoku earthquake, Japan, revealed from repeating earthquake analysis and finite-difference wave propagation simulation. *Geophysical Journal International*, 201(2), 741-763.

- Schaff, D. (2010). Improvements to detection capability by cross-correlating for similar events: a case study of the 1999 Xiuyan, China, sequence and synthetic sensitivity tests. *Geophysical Journal International*, 180(2), 829-846.
- Schaff, D. P. (2008). Semiempirical statistics of correlation-detector performance. *Bulletin of the Seismological Society of America*, 98(3), 1495-1507.
- Schaff, D. P., & Beroza, G. C. (2004). Coseismic and postseismic velocity changes measured by repeating earthquakes. *Journal of Geophysical Research: Solid Earth*, 109(B10).
- Schaff, D. P., & Richards, P. G. (2004). Repeating seismic events in China. *Science*, 303(5661), 1176-1178.
- Schaff, D. P., & Richards, P. G. (2011). On finding and using repeating seismic events in and near China. *Journal of Geophysical Research: Solid Earth*, 116(B3).
- Schaff, D. P., Kim, W. Y., Richards, P. G., Jo, E., & Ryoo, Y. (2018). Using waveform cross correlation for detection, location, and identification of aftershocks of the 2017 nuclear explosion at the North Korea test site. *Seismological Research Letters*, 89(6), 2113-2119.
- Schmittbuhl, J., Karabulut, H., Lengliné, O., & Bouchon, M. (2016). Long-lasting seismic repeaters in the Central basin of the Main Marmara fault. *Geophysical Research Letters*, 43(18), 9527-9534.
- Schultz, R., & Wang, R. (2020). Newly emerging cases of hydraulic fracturing induced seismicity in the Duvernay East Shale Basin. *Tectonophysics*, 228393.
- Schultz, R., Atkinson, G., Eaton, D. W., Gu, Y. J., & Kao, H. (2018). Hydraulic fracturing volume is associated with induced earthquake productivity in the Duvernay play. *Science*, 359(6373), 304-308.
- Schultz, R., Mei, S., Pană, D., Stern, V., Gu, Y. J., Kim, A., & Eaton, D. (2015b). The Cardston earthquake swarm and hydraulic fracturing of the Exshaw Formation (Alberta Bakken play). *Bulletin of the Seismological Society of America*, 105(6), 2871-2884.
- Schultz, R., Skoumal, R. J., Brudzinski, M. R., Eaton, D., Baptie, B., & Ellsworth, W. (2020). Hydraulic fracturing-induced seismicity. *Reviews of Geophysics*, 58(3), e2019RG000695.
- Schultz, R., Stern, V., & Gu, Y. J. (2014). An investigation of seismicity clustered near the Cordel Field, west central Alberta, and its relation to a nearby disposal well. *Journal of Geophysical Research: Solid Earth*, 119(4), 3410-3423.
- Schultz, R., Stern, V., Novakovic, M., Atkinson, G., & Gu, Y. J. (2015). Hydraulic fracturing and the Crooked Lake Sequences: Insights gleaned from regional seismic networks. *Geophysical Research Letters*, 42(8), 2750-2758.
- Schultz, R., Stern, V., Novakovic, M., Atkinson, G., & Gu, Y. J. (2015a). Hydraulic fracturing and the Crooked Lake sequences: Insights gleaned from regional seismic networks. *Geophysical Research Letters*, 42(8), 2750-2758.
- Schultz, R., Wang, R., Gu, Y. J., Haug, K., & Atkinson, G. (2017). A seismological overview of the induced earthquakes in the Duvernay play near Fox Creek, Alberta. *Journal of Geophysical Research: Solid Earth*, 122(1), 492-505.

- Shearer, P. M. (1997). Improving local earthquake locations using the L1 norm and waveform cross correlation: Application to the Whittier Narrows, California, aftershock sequence. *Journal of Geophysical Research: Solid Earth*, 102(B4), 8269-8283.
- Shearer, P. M. (2009). *Introduction to seismology*. Cambridge university press.
- Shelly, D. R., Beroza, G. C., & Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, 446(7133), 305.
- Shen, L. W., Schmitt, D. R., & Schultz, R. (2019). Frictional stabilities on induced earthquake fault planes at Fox Creek, Alberta: a pore fluid pressure dilemma. *Geophysical Research Letters*, 46(15), 8753-8762.
- Shirzaei, M., Bürgmann, R., & Taira, T. A. (2013). Implications of recent asperity failures and aseismic creep for time-dependent earthquake hazard on the Hayward fault. *Earth and Planetary Science Letters*, 371, 59-66.
- Skoumal, R. J., Brudzinski, M. R., & Currie, B. S. (2015). Distinguishing induced seismicity from natural seismicity in Ohio: Demonstrating the utility of waveform template matching. *Journal of Geophysical Research: Solid Earth*, 120(9), 6284-6296.
- Skoumal, R. J., Brudzinski, M. R., Currie, B. S., & Levy, J. (2014). Optimizing multi-station earthquake template matching through re-examination of the Youngstown, Ohio, sequence. *Earth and Planetary Science Letters*, 405, 274-280.
- Skoumal, R. J., Brudzinski, M. R., Currie, B. S., & Ries, R. (2019). Temporal patterns of induced seismicity in Oklahoma revealed from multi-station template matching. *Journal of Seismology*, 1-15.
- Snieder, R., & Vrijlandt, M. (2005). Constraining the source separation with coda wave interferometry: Theory and application to earthquake doublets in the Hayward fault, California. *Journal of Geophysical Research: Solid Earth*, 110(B4).
- Stein, R. S. (1999). The role of stress transfer in earthquake occurrence. *Nature*, 402(6762), 605-609.
- Stein, S., & Wysession, M. (2009). *An introduction to seismology, earthquakes, and earth structure*. John Wiley & Sons.
- Taira, T. A., Bürgmann, R., Nadeau, R. M., & Dreger, D. S. (2014). Variability of fault slip behavior along the San Andreas Fault in the San Juan Bautista Region. *Journal of Geophysical Research: Solid Earth*, 119(12), 8827-8844.
- Tan, F., Kao, H., Nissen, E., & Eaton, D. (2019). Seismicity-Scanning Based on Navigated Automatic Phase-Picking. *Journal of Geophysical Research: Solid Earth*.
- Tepp, G. (2018). A repeating event sequence alarm for monitoring volcanoes. *Seismological Research Letters*, 89(5), 1863-1876.
- Thelen, W., Malone, S., & West, M. (2011). Multiplets: Their behavior and utility at dacitic and andesitic volcanic centers. *Journal of Geophysical Research: Solid Earth*, 116(B8).
- Thelen, W., West, M., & Senyukov, S. (2010). Seismic characterization of the fall 2007 eruptive sequence at Bezymianny Volcano, Russia. *Journal of Volcanology and Geothermal Research*, 194(4), 201-213.

- Tkalčić, H., Young, M., Bodin, T., Ngo, S., & Sambridge, M. (2013). The shuffling rotation of the Earth's inner core revealed by earthquake doublets. *Nature Geoscience*, 6(6), 497-502.
- Trugman, D. T., & Shearer, P. M. (2017). GrowClust: A hierarchical clustering algorithm for relative earthquake relocation, with application to the Spanish Springs and Sheldon, Nevada, earthquake sequences. *Seismological Research Letters*, 88(2A), 379-391.
- Uchida, N. (2019). Detection of repeating earthquakes and their application in characterizing slow fault slip. *Progress in Earth and Planetary Science*, 6(1), 40.
- Uchida, N., & Bürgmann, R. (2019). Repeating earthquakes. *Annual Review of Earth and Planetary Sciences*, 47, 305-332.
- Uchida, N., & Matsuzawa, T. (2013). Pre-and postseismic slow slip surrounding the 2011 Tohoku-oki earthquake rupture. *Earth and Planetary Science Letters*, 374, 81-91.
- Uchida, N., Matsuzawa, T., Hasegawa, A., & Igarashi, T. (2003). Interplate quasi-static slip off Sanriku, NE Japan, estimated from repeating earthquakes. *Geophysical Research Letters*, 30(15).
- Uchida, N., Matsuzawa, T., Hirahara, S., & Hasegawa, A. (2006). Small repeating earthquakes and interplate creep around the 2005 Miyagi-oki earthquake (M= 7.2). *Earth, planets and space*, 58(12), 1577-1580.
- Visser, R., Smith, B., Kao, H., Hutchinson, J., & McKay, J. E. (2017). A comprehensive earthquake catalogue for northeastern British Columbia and western Alberta, 2014-2016. Geological Survey of Canada.
- Waldhauser, F., & Ellsworth, W. L. (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California. *Bulletin of the Seismological Society of America*, 90(6), 1353-1368.
- Waldhauser, F., & Ellsworth, W. L. (2002). Fault structure and mechanics of the Hayward fault, California, from double-difference earthquake locations. *Journal of Geophysical Research: Solid Earth*, 107(B3), ESE-3.
- Walt, S. V. D., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2), 22-30.
- Wang, B., Harrington, R. M., Liu, Y., Kao, H., & Yu, H. (2020). A study on the largest hydraulic-fracturing-induced earthquake in Canada: Observations and static stress-drop estimation. *Bulletin of the Seismological Society of America*, 110(5), 2283-2294.
- Wang, B., Verdecchia, A., Kao, H., Harrington, R. M., Liu, Y., & Yu, H. (2021). A Study on the Largest Hydraulic Fracturing Induced Earthquake in Canada: Numerical Modeling and Triggering Mechanism. *Bulletin of the Seismological Society of America*.
- Wang, R., & Kümpel, H. J. (2003). Poroelasticity: Efficient modeling of strongly coupled, slow deformation processes in multilayered half-space. *Geophysics*, 68(2), 705-717.
- Wang, R., Gu, Y. J., Schultz, R., Zhang, M., & Kim, A. (2017). Source characteristics and geological implications of the January 2016 induced earthquake swarm near Crooked Lake, Alberta. *Geophysical Journal International*, 210(2), 979-988.

- Warren-Smith, E., Fry, B., Kaneko, Y., & Chamberlain, C. J. (2018). Foreshocks and delayed triggering of the 2016 MW7. 1 Te Araroa earthquake and dynamic reinvigoration of its aftershock sequence by the MW7. 8 Kaikōura earthquake, New Zealand. *Earth and Planetary Science Letters*, 482, 265-276.
- Warren-Smith, E., Chamberlain, C. J., Lamb, S., & Townend, J. (2017). High-precision analysis of an aftershock sequence using matched-filter detection: The 4 May 2015 ML 6 Wanaka earthquake, Southern Alps, New Zealand. *Seismological Research Letters*, 88(4), 1065-1077.
- Weingarten, M., Ge, S., Godt, J. W., Bekins, B. A., & Rubinstein, J. L. (2015). High-rate injection is associated with the increase in US mid-continent seismicity. *Science*, 348(6241), 1336-1340.
- Wilson, M. P., Worrall, F., Davies, R. J., & Almond, S. (2018). Fracking: How far from faults?. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, 4(2), 193-199.
- Wolhart, S. L., Harting, T. A., Dahlem, J. E., Young, T., Mayerhofer, M. J., & Lolon, E. P. (2006, January). Hydraulic fracture diagnostics used to optimize development in the Jonah field. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.
- Wu, W., Zhan, Z., Peng, S., Ni, S., & Callies, J. (2020). Seismic ocean thermometry. *Science*, 369(6510), 1510-1515.
- Xu, C., Wang, J., Li, Z., & Drummond, J. (2010). Applying the Coulomb failure function with an optimally oriented plane to the 2008 Mw 7.9 Wenchuan earthquake triggering. *Tectonophysics*, 491(1-4), 119-126.
- Yamada, M., Mori, J., & Matsushi, Y. (2016). Possible stick-slip behavior before the Rausu landslide inferred from repeating seismic events. *Geophysical Research Letters*, 43(17), 9038-9044.
- Yamaguchi, J., Naoi, M., Nakatani, M., Moriya, H., Igarashi, T., Murakami, O., ... & Ogasawara, H. (2018). Emergence and disappearance of very small repeating earthquakes on a geological fault in a gold mine in South Africa. *Tectonophysics*, 747, 318-326.
- Yamashita, Y., Shimizu, H., & Goto, K. (2012). Small repeating earthquake activity, interplate quasi-static slip, and interplate coupling in the Hyuga-nada, southwestern Japan subduction zone. *Geophysical Research Letters*, 39(8).
- Yao, D., Walter, J. I., Meng, X., Hobbs, T. E., Peng, Z., Newman, A. V., ... & Protti, M. (2017). Detailed spatiotemporal evolution of microseismicity and repeating earthquakes following the 2012 Mw 7.6 Nicoya earthquake. *Journal of Geophysical Research: Solid Earth*, 122(1), 524-542.
- Yehya, A., Yang, Z., & Rice, J. R. (2018). Effect of fault architecture and permeability evolution on response to fluid injection. *Journal of Geophysical Research: Solid Earth*, 123(11), 9982-9997.
- Yu, H., Harrington, R. M., Liu, Y., & Wang, B. (2019). Induced seismicity driven by fluid diffusion revealed by a near-field hydraulic stimulation monitoring array in the Montney Basin, British Columbia. *Journal of Geophysical Research: Solid Earth*, 124(5), 4694-4709.
- Yu, W. C. (2013). Shallow-focus repeating earthquakes in the Tonga–Kermadec–Vanuatu subduction zones. *Bulletin of the Seismological Society of America*, 103(1), 463-486.

- Zhang, J., Richards, P. G., & Schaff, D. P. (2008). Wide-scale detection of earthquake waveform doublets and further evidence for inner core super-rotation. *Geophysical Journal International*, 174(3), 993-1006.
- Zhang, J., Song, X., Li, Y., Richards, P. G., Sun, X., & Waldhauser, F. (2005). Inner core differential motion confirmed by earthquake waveform doublets. *Science*, 309(5739), 1357-1360.
- Zhang, M., & Wen, L. (2015). An effective method for small event detection: match and locate (M&L). *Geophysical Journal International*, 200(3), 1523-1537.
- Zhao, P., & Peng, Z. (2009). Depth extent of damage zones around the central Calaveras fault from waveform analysis of repeating earthquakes. *Geophysical Journal International*, 179(3), 1817-1830.
- Zhu, L., & Rivera, L. A. (2002). A note on the dynamic and static displacements from a point source in multilayered media. *Geophysical Journal International*, 148(3), 619-627.