Title
Navigating Microseismicity Characterization with Backprojection, Matched-filter, and Deep Learning

Author
Feng, Tian

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Navigating Microseismicity Characterization with Backprojection, Matched-filter, and Deep Learning

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Geophysics and Space Physics

by

Tian Feng

2022

ABSTRACT OF THE DISSERTATION

Navigating Microseismicity Characterization
with Backprojection, Matched-filter, and Deep Learning

by

Tian Feng
Doctor of Philosophy in Geophysics and Space Physics
University of California, Los Angeles, 2022
Professor Lingsen Meng, Chair

Benefiting from the recent deployments of dense seismic arrays, seismologists have the opportunity to detect earthquakes of small magnitudes ($M < 2$), referring to the occurrence of these events as microseismicity. The characterization of microseismicity is essential in solving many geophysical problems, such as probing fault structures, investigating the process of hydraulic fracturing, and imaging the preseismic, coseismic, and postseismic slip of megathrust earthquakes. Traditionally, earthquakes are identified by picking seismic phases on continuous waveforms using the short-time average to long-time average ratio, kurtosis and skewness, waveform polarization, Akaike information criterion, and discrete wavelet transforms. These single-station methods, however, may fail to detect microseismicity with weak phase arrivals hidden in the noise.

This thesis introduces three new approaches to detecting microseismicity with multi-station data: backprojection, matched-filter, and deep learning. The backprojection approach tracks and back-projects the coherent seismic pulses to the target region to determine the timing and location of any microseismic earthquakes. The matched-filter approach searches for similar patterns of existing template earthquakes in the continuous recordings as suggestive of a new event. We characterize the aftershock sequence of the 2011 M 9 Tohoku earthquake with the backprojection and matched-filter methods. Based on the spatial con-

sistency between aftershock-depleted zones and large coseismic slip, we identify a possible large coseismic slip zone in the near-trench region offshore Fukushima. The deep learning approach is data-driven and it learns to characterize microseismicity with a neural network based on a large number of microseismicity recordings. We trained a Deep Learning Phase-picking model named EdgePhase with a Southern California dataset and applied it to detect the early aftershocks following the 2020 M 7 Samos, Greece earthquake. Compared to a local earthquake catalog, EdgePhase showed 190% more detections with an event distribution that is more conformative to a planar fault interface, suggesting higher fidelity in event locations. In addition to characterizing microseismicity, backprojection can also improve the prediction of earthquake ground motions. We propose a high-frequency distance metric based on backprojection, which outperforms traditional distance metrics in predicting the ground shaking intensity of megathrust earthquakes in Japan and Chile.

The dissertation of Tian Feng is approved.

Paul M. Davis

Gilles F. Peltzer

Peter Bird

Lingsen Meng, Committee Chair

University of California, Los Angeles

2022

To My Family

Table of Contents

List of Figures

List of Tables

Acknowledgments

First of all, I would like to thank fate for giving me this opportunity to pursue a Ph.D. It was during my time in the Ph.D program that I was able to freely think about issues not directly related to my life and career. It is my belief that people only live once, so I desire to leave some signs of my existence behind. During the last six years, I have been engaged in creative research and I have been quite satisfied with my experiences.

I would like to express my sincere appreciation to my advisor, Lingsen Meng. Besides being an excellent advisor, he is also a true and valuable friend. It has been a pleasure to work with Lingsen, as he has always been supportive of my research. He helped me think through questions in depth and explain things with more clarity. He is an advisor who truly cares about a student's future. Furthermore, he connected me with researchers Behzad Hassani and Alan Yong. As well, I am grateful to Paul Davis, who taught me a lot in EPS 200A and we had lots of discussions. I am thankful to Gilles Peltzer for teaching me about remote sensing through EPS150. He is full of knowledge and a joy to learn from. I really appreciate Peter Bird, who patiently gave me several suggestions regarding my research and thesis.

I appreciate both my academic and social interactions with current and former UCLA graduate students and postdocs. Haotian Xu has been a huge help to me in academe and in my life. Zheng Xing and Han Bao are my best gym partners. Aliin Zhang gives me a lot of career advice. Hui Huang and Shuyu Chen are an amazing couple who provide me with much support. I would also apprecaite my friends Yuqing Xie, Tong Zhou, Saeed Mohanna, Liuwei Xu, Xiyuan Bao, and Jie Deng.

My thanks also go to my friends studying seismology at USC and UCSD, Wei Wang, Hao Zhang, Shuye Huang, Yufang Xu, Guoliang Li, and Xiaohua Xu. Furthermore, I am thankful for my friends Yuan Gao, Jiawei Sun, Jiayuan Han, and Qichen Tang.

My final thanks are for my father Shengguo and my mother Xiaoxia. I'm so fortunate to have such wonderful parents. They respect and support every decision I make.

Curriculum Vitae

2012–2016	B.S. in Geophysics, Nanjing University, Nanjing, China
2016–Present	Ph.D. student in Geophysics, University of California, Los Angeles (UCLA).

Publications

Feng, T., & Meng, L. (2018). A High-Frequency Distance Metric in Ground-Motion Prediction Equations Based on Seismic Array Backprojections. Geophysical Research Letters, 45(21), 11-612

Bao, H., Ampuero, J. P., Meng, L., Fielding, E. J., Liang, C., Milliner, C. W., **Feng, T.** & Huang, H. (2019). Early and persistent supershear rupture of the 2018 magnitude 7.5 Palu earthquake. Nature Geoscience, 12(3), 200-205

Hassani, B., Yong, A., Atkinson, G. M., **Feng, T.,** & Meng, L. (2019). Comparison of site dominant frequency from earthquake and microseismic data in California. Bulletin of the Seismological Society of America, 109(3), 1034-1040

Feng, T., Meng, L., & Huang, H. (2020). Detecting Offshore Seismicity: Combining Backprojection Imaging and Matched-filter Detection. Journal of Geophysical Research: Solid Earth, 125(8), e2020JB019599.

Feng, T., Mohanna, S., & Meng, L. EdgePhase: A Deep Learning Model for Multi-station Seismic Phase Picking. (Submitted to Geochemistry, Geophysics, Geosystems)

CHAPTER 1

Introduction

1.1 Importance of Microseismicity

Through the development of dense seismic arrays and new detection techniques, humans are now able to detect earthquakes of magnitudes lower than zero. Although these newly detected microseismic events ($M < 2$) can't be felt by humans, they are still essential in mitigating earthquake hazards, monitoring industry production, and answering scientific questions.

Based on the Gutenberg-Richter law (Gutenberg & Richter, 1944), as earthquake magnitude increases, the number of earthquakes above that magnitude in a given year decreases. This indicates that mega earthquakes are quite rare compared with microseismic ones. Therefore, a modern seismic catalog contains dozens of times more events than an early catalog, due to numerous detectable microseismic events. Furthermore, this results in a reduction of the magnitude of completeness (M_c) of the catalog, leading to the advancement of statistical seismology (e.g., b value, p value, Epidemic Aftershock-Type Sequence model). In Chapter 2, we build a more accurate aftershock catalog for the 2011 M9 Tohoku Earthquake by decreasing M_c to 1.0 on land and 2.0 near the trench as compared to the Japan Meteorological Agency (JMA) catalog. The presence of large b values (≥ 1.2) is found near large aftershocks, possibly indicative of localized pockets of small differential stresses. Near the trench, the p values (0.93–1.11) tend to be higher than those in the inland area (0.64–0.85). This may be due to the larger coseismic slip and hence larger stress drop of the outer-rise normal-faulting events compared to the deeper thrust-faulting events. Microseismicity is also a significant precursor to large earthquakes. A number of large interplate earthquakes have

been preceded by increased seismic activity in the months to days preceding the mainshock (Bouchon, Durand, Marsan, Karabulut, & Schmittbuhl, 2013).

Along with its direct improvements to statistical seismology, microseismicity is also widely used for monitoring anthropogenic earthquakes (e.g., hydrocarbon extraction, geothermal production, CO_2 sequestration, and underground mining; Matzel et al., 2014; T. Chen & Huang, 2020; Warpinski, Du, & Zimmer, 2012; Williams-Stroud et al., 2020). Hydraulic fracturing is an important technique used to enhance the productivity of oil and gas wells. It involves the generation or opening of small fractures by injecting pressurized fluid. According to recent studies, hydraulic fracturing is strongly associated with microseismicity (Ellsworth, 2013; Yin, 2017; Langenbruch & Zoback, 2016). For assessing the performance of hydraulic fracturing, it is essential to track the associated microseismicity in real time. Small to moderate earthquakes may also be induced at nearby faults by the injection activities, posing a seismic hazard to nearby residents. As a result, real-time knowledge of induced microseismicity is crucial for assessing the hazard associated with anthropogenic earthquakes.

Geothermal reservoirs use the thermal energy of the Earth's crust to produce CO_2-free energy and contribute to the development of green economies. Monitoring geothermal reservoirs can also be accomplished through induced microseismicity. The seismic front maps the stress concentration at the tips of aseismic slip, which is primarily caused by an increase in fluid pressure (De Barros, Wynants-Morel, Cappa, & Danré, 2021). Therefore, distance-time plots can be used to monitor the seismogenic state of a reservoir. Furthermore, Holtzman, Paté, Paisley, Waldhauser, and Repetto (2018) find that events with different spectral properties are related to changes in the rate of water injection into the geysers' geothermal reservoir, which indicates that changes in acoustic properties and faulting processes correlate with changes in thermomechanical state.

In addition, microseismicity is used to study rocky slopes (Spillmann et al., 2007; Salvoni & Dight, 2016; Xu et al., 2014) and ice streams (E. Smith, Smith, White, Brisbourne, & Pritchard, 2015; Kufner et al., 2021). It can also help us answer some scientific questions,

such as what is the structure of fault zones, which portion of the fault is ruptured during the duration of a large earthquake, how fluid migrates in the fault system, and what are the magmatic processes before and after a volcanic eruption.

1.1.1 Fault Structure and Rupture Process

Geologists often find it difficult to identify the surface expression of the fault due to sediments, buildings, and vegetation. Surface traces of faults should be interpreted cautiously in regions with complex inherited tectonics (Courboulex et al., 2003). Other than field and satellite Images, the spatial distribution of microseismicity can delineate detailed fault structures and hidden faults (Inbal, Clayton, & Ampuero, 2015; Courboulex et al., 2003; Schaff, Bokelmann, Beroza, Waldhauser, & Ellsworth, 2002; Pananont & Pornsopin, 2020; Wolfson-Schwehr, Boettcher, McGuire, & Collins, 2014). A thorough understanding of fault characteristics (e.g., position, length, strike, dip) is critical for assessing seismic hazards, especially in highly populated areas. An effective means of constraining fault depth is to determine the maximum depth of seismic activity. This helps to understand the thickness of the seismogenic zone, which is important to the study of long-term fault slips (Jiang & Lapusta, 2017). Through examining the focal mechanism of microseismicity, we can also calculate the stress field on the fault planes (Boese, Townend, Smith, & Stern, 2012; Urbancic, Trifu, & Young, 1993; De Matteis et al., 2012).

In addition, aftershocks resulting from a large earthquake may be used to observe the rupture process. When Interferometric Synthetic Aperture Radar (InSAR) and Global Positioning System (GPS) observations are unclear or unavailable, it can be difficult to distinguish the fault plane and auxiliary plane from the mainshock focal mechanism. It is possible to determine the fault plane that participated in the mainshock by examining the distribution of aftershocks with focal mechanisms similar to those of the mainshock. As such, microseismicity is complementary to other observations (e.g., on-site, InSAR) in determining the fault plane.

In addition, microseismicity provides vital information to understand the pre-, co-, and

post-seismic slip processes of large earthquakes (e.g. Frank, Poli, & Perfettini, 2017; L. Meng, Zhang, & Yagi, 2016; Kato et al., 2012; Huang, Meng, et al., 2017; Huang, Xu, Meng, Bürgmann, & Baez, 2017; Uchida & Matsuzawa, 2013; Lengliné, Enescu, Peng, & Shiomi, 2012). According to our findings in Chapter 2, a large coseismic slip zone is possibly present in the near-trench region offshore Fukushima because the aftershock-depleted zones planarly correspond with the zone of large coseismic slip. In general, microseismicity helps to study fault zones and the rupture process associated with large earthquakes.

1.1.2 Monitoring Fluid Migration and Volcanoes

As we learned previously, hydraulic fracturing stages could be monitored through induced microseismicity. In a similar manner, we can also examine the migration pattern of fluids based on natural seismicity, specifically earthquake swarms. Earthquake swarms are sequences of natural seismicity without clear mainshocks. They are frequently associated with areas of high heat flow and are believed to be caused primarily by external processes such as fluid flow or aseismic slip, rather than earthquake triggers connected with stress transfer(Ross, Cochran, Trugman, & Smith, 2020). A study by Ross, Cochran, et al. (2020) of a 4-year-long earthquake swarm at a fault zone in southern California indicates that fluids are naturally injected into the fault zone and diffuse through strike-parallel channels while triggering earthquakes. Initially, a permeability barrier limits up-dip swarm migration but is ultimately circumvented. Observation of the spatiotemporal distribution (migration speed) of microseismicity provides high-resolution constraints on the process by which the swarm initiates, grows, and stops. In addition, Darcy's law can be used to estimate hydrological properties (e.g., diffusivity, permeability) related to fluid migration.

The magma in the volcano chamber also has the property of liquid, and its movement can cause changes in pressure and, consequently, seismicity (McNutt & Roman, 2015; Rubin, Gillard, & Got, 1998; Titos, Bueno, Garcia, & Benitez, 2018). The seismicity of long-period, hybrid, and volcano-tectonic events are useful indicators for tracking the evolution of volcanic processes (Cui, Li, & Huang, 2021; Ripepe et al., 2015). While the magnitudes of these

volcanic earthquakes are typically small, they can illuminate magma reservoirs, monitor magma discharge rates, and predict volcanic eruptions. In some cases, volcano eruptions may be preceded by an upward migration of seismicity associated with magmatic recharge (Lengliné, Duputel, & Ferrazzini, 2016; Cui et al., 2021).

Furthermore, the repeated cycles of uplift and subsidence of the Yellowstone caldera can be explained by the frequent earthquake swarms (Shelly et al., 2013). The occurrence of swarms was the result of the rupture of a zone of confined high-pressure aqueous fluids within a preexisting crustal fault system, causing release of accumulated stresses. As a consequence of the high pressure fluid injection, there may have been hybrid shear and dilatational failures, which are common to hydrothermally affected fault zones that have been exhumed. The above process has likely occurred repeatedly in Yellowstone's history as aqueous fluids migrated into the brittle crust as they were exsolved by magma, and it may have contributed to observed cycles of caldera uplift and subsidence.

This section introduced microseismicity and its applications in research, hazard mitigation, and industrial production. These applications accelerate the development of techniques for characterization of microseismicity. Yet, detecting microseismicity remains challenging in some cases, such as the low signal-to-noise ratio (SNR) of seismograms and sparse station distribution. Next, we will summarize the development history of microseismicity detection methods.

1.2 Conventional Microseismicity Detection Methods

The conventional method of detecting microseismicity consists of phase picking, phase association, earthquake location, and magnitude estimation. The first of these steps, phase picking, is fundamental and has a direct bearing upon the credibility of subsequent research on microseismicity (for example, event identification and source mechanisms). Historically, reliable phase arrivals were selected by seismologists with decades of experience and enormous capabilities of the human brain (Trnkoczy, 2009). Since the late 1970s, automatic phase

picking algorithms have flourished largely due to the rapid development of microcomputers. Thanks to microcomputers and automatic phase-picking algorithms, earthquake study evolved into a new era. It frees seismologists from the heavy labor of phase picking and increases the productivity of seismology research (Saragiotis, Hadjileontiadis, & Panas, 2002). It also eliminates the time-consuming and subjective (human experts pick phases based on their individual experiences) issues associated with manual picks. Nevertheless, developing a phase picking algorithm that can rival human analysts was a challenge, and numerous algorithms and methods have been proposed since the 1970s to the present day. Specifically, we introduce a few widely used methods, such as short-time-average over long-time-average trigger (STA/LTA), polarization analysis, skewness, kurtosis, Akaike information criterion (AIC), and the discrete wavelet transform (DWT).

1.2.1 STA/LTA

The STA/LTA method is the most widely used energy detector technique in detecting microseismic activity (R. V. Allen, 1978; R. Allen, 1982; Ruud & Husebye, 1992; Earle & Shearer, 1994; Saari, 1991; Baer & Kradolfer, 1987). It continuously calculates the average absolute amplitude of seismic trace (acceleration, velocity, or displacement) over two consecutive sliding windows. Short time windows (STA) are sensitive to seismic events, whereas long time windows (LTA) are sensitive to the temporal amplitude of the seismic noise at the site (Trnkoczy, 2009). The baseline (zero) for amplitude is determined by band-pass trigger filters, which should generally accommodate the frequencies of the maximum energy of expected seismic events (Trnkoczy, 2009). In practice, the length of the short and long windows is 30 seconds and one second, respectively. When the STA/LTA ratio exceeds an empirically determined threshold, a tentative event is declared. Assuming that the arrival of a phase is signaled by a change in the frequency content and amplitude in the seismic time series X_i, characteristic functions (CF) are applied on the seismograms to highlight these changes. In previous practices, different CFs were applied to seismic traces before calculating the average of windows, such as the absolute value of the input series $|X_i|$, the square of

the derivative of the input series $|X_i'^2|$, etc. Here is the original version of CF (R. V. Allen, 1978).

$$CF_i = X_i^2 + X_i'^2 + K$$

where X_i' is the derivative of seismic trace and K is a constant. This CF monitors the change of the envelope of the seismic signal and it performs well in the routine detection of earthquakes. As the choice of CF (Akram, Peter, & Eaton, 2018) and parameters in STA/LTA are case-dependent and important for achieving optimal results, numerous versions are in use in various research institutions.

1.2.2 Polarization Analysis

In the polarization analysis method, earthquakes are detected by identifying the highly polarized initial arrivals of seismic phases (Vidale, 1986; Jurkevics, 1988; Magotra, Ahmed, & Chael, 1987, 1989; S. Anderson & Nehorai, 1996; Kurzon, Vernon, Rosenberger, & Ben-Zion, 2014; Ross & Ben-Zion, 2014). The polarization of a 3-component M-sample sequence can be determined based on the covariance matrix S.

$$S_{jk} = \frac{XX^T}{M} = \frac{1}{M}\sum_{i=1}^{M} x_{ij}x_{ik}$$

where x_{ij} is the ith sample of component j. The x has removed the mean value along the channel. The covariance matrix S is a $3 * 3$, semidefinite matrix, whose eigenvalues are real and non-negative.

$$S = \begin{vmatrix} S_{zz} & S_{zn} & S_{ze} \\ S_{nz} & S_{nn} & S_{ne} \\ S_{ez} & S_{en} & S_{ee} \end{vmatrix}.$$

The corresponding eigenvalues ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) and eigenvector matrix (u_1, u_2, u_3) can be calculated through singular value decomposition (Kurzon et al., 2014; Ross & Ben-Zion, 2014). These eigenvalues and eigenvectors provide information on the propagation direction and perpendicular direction of the P wave (The eigenvector u_1 with the largest eigenvalue λ_1

points to the P wave propagation direction). There are two common measures used to detect seismic phases: the first is rectilinearity, which measures the degree of linear polarization, which theoretically equals 1 for a pure body wave.

$$r = 1 - (\frac{\lambda_2 + \lambda_3}{2\lambda_1}) \in [0, 1]$$

The other measure is the apparent vertical incidence angle ϕ.

$$\phi = cos^{-1}|u_{11}|$$

The value of $cos(\phi)$ is close to 1 and 0 for P- and S-phases, respectively. By combining these two measures, Ross and Ben-Zion (2014) construct polarization filters to identify P- and S-phases from seismograms. The P-phase filter has the function of $p = rcos(\phi)$ and the S-phase filter has the function of $s = r(1-cos(\phi))$. As the P- and S-phase filters enhance and depress the signal in the wavefront propagation direction, respectively, they can highlight P- and S-phases (Fig. 1.1).

1.2.3 Skewness and Kurtosis

One way to identify phases is through the use of higher-order statistics in seismograms (Yung & Ikelle, 1997; Saragiotis et al., 2002; Saragiotis, Hadjileontiadis, & Panas, 1999). Two common statistics are introduced here: the skewness and the kurtosis, which are third- and fourth-order statistics, respectively.

$$skewness = \frac{\sum_{i=1}^{M}(X(i) - \hat{m_X})^3}{(M-1)\hat{\sigma_X}^3}$$

$$kurtosis = \frac{\sum_{i=1}^{M}(X(i) - \hat{m_X})^4}{(M-1)\hat{\sigma_X}^4} - 3$$

where $\hat{m_X}$ and $\hat{\sigma_X}$ are the estimates of the mean and standard deviation of the M-sample finite sequence $X(i)$. The skewness measures the symmetry, and the kurtosis measures the heaviness of the tails in the distribution of sequences. The estimates of skewness and kurtosis are zero if the data is Gaussian. Integrating skewness and kurtosis, the phase-picking model

PAI-S/K (Saragiotis et al., 2002) can identify signals with asymmetrical distributions and heavy tails (non-Gaussian signals). By assuming that earthquake signals are non-Gaussian while noise is Gaussian, the transition between Gaussianity and non-Gaussianity coincides with the onset of the seismic event (Fig. 1.2; Saragiotis et al., 2002).

1.2.4 AIC picker

The Akaike information criterion (AIC) can be used to divide a single time series into two portions (early and late) under the assumption that each portion exhibits distinct but stationary characteristics. AIC can be used to detect P and S-phase onsets by assuming that the variance of the time series in the seismogram is different before and after the P and S wave arrivals (Maeda, 1985; Sleeman & Van Eck, 1999). The onset occurs when the AIC has a minimum value (Fig. 1.3).

$$AIC(k) = k * log(var(x[1,k])) + (N-k-1) * log(var(x[k+1,N]))$$

where x is the seismogram of length N, and $k \in [1, N-1]$. The AIC picker was later improved by incorporating the autoregressive process (AR) (Sleeman & Van Eck, 1999; Leonard & Kennett, 1999; Leonard, 2000; H. Zhang, Thurber, & Rowe, 2003). To be more precise, the two stationary time series are fitted with AR processes.

For real-world applications, AIC pickers have several shortcomings since the phase onset is always picked up at the minimum. Even if a window only contains noise, the AIC pickers will define an "onset". Moreover, AIC pickers generate several local minima in low SNR conditions, and the P wave arrival can be located at one of the local minima.

1.2.5 DWT

Discrete wavelet transforms (DWT) have been extensively employed for improving seismic data resolution (Chakraborty & Okaya, 1995), for compressing seismic data (Lervik, Rosten, & Ramstad, 1996), and for picking up seismic phases (H. Zhang et al., 2003). DWT is superior to the Fourier transform in its ability to characterize the signal structure locally

and according to its scale. In particular, DWT displays coarse (low frequency) features on large scales while fine (high frequency) features appear at small scales (H. Zhang et al., 2003). Due to the ability to analyze seismic signals at different resolutions, DWT is widely used in the preceding phase-picking methods. A few examples are Anant and Dowla (1997), which combine DWT and polarization; Saragiotis et al. (1999), which combine DWT with high-order statistics; H. Zhang et al. (2003), which combine DWT and AIC picker; Rodriguez (2011), which combine DWT with STA/LTA; and X. Li, Shang, Wang, Dong, and Weng (2016), which combine DWT with STA/LTA and kurtosis. These hybrid methods improve the accuracy of phase picking at low SNR conditions, and have become increasingly popular since the 2000s.

No matter how complex the mathematical transformations they performed, the conventional phase-picking method can only analyze information from a single station. Some seismic phases, however, are difficult to identify at a single station due to a low SNR. In the manual phase picking process, humans typically rely on the consistency of waveforms from multiple stations in order to determine whether ambiguous cases should be classified as seismic phases. In this instance, a multi-station approach utilizing the waveform consistency between neighboring stations should improve phase picking and microseismicity detection. The following is an introduction to two such approaches based on waveform similarity and machine learning, respectively.

1.3 Models Based on Waveform Similarity

Since the early 2000s, a number of waveform-similarity-based multi-station methods have become popular for microseismicity detection as a result of the development of central processing units (CPUs) and the deployment of dense arrays (e.g., backprojection, matched-filter, autocorrelation, and neighboring-station coherence; Z. Li, Peng, Hollis, Zhu, & McClellan, 2018). The primary advantage of these methods is that they utilize the coherency and similarity of waveforms in dense arrays, which is potentially better suited to detecting smaller events than the traditional single-station methods are, but at the cost of additional compu-

tation resources.

1.3.1 Backprojection

Backprojection (BP) is a recently-developed earthquake rupture imaging technique based on coherent teleseismic P-wave recordings (e.g. Rost & Thomas, 2002; Ishii, Shearer, Houston, & Vidale, 2005; Krüger & Ohrnberger, 2005; L. Meng, Ampuero, Luo, Wu, & Ni, 2012; Kiser & Ishii, 2017). Back-tracking of seismic waves recorded by dense arrays allows BP to determine the timing and location of the energy source that generates the strongest seismic radiations. Tracking these energy sources allows us to visualize the spatiotemporal properties (length, direction, speed, and segmentation) of the rupture process. The development of large-scale dense seismic networks has enabled BP to image the rupture process of large earthquakes in recent decades. Contrary to conventional finite-fault modeling, BP is an imaging procedure which does not require detailed knowledge of Green's functions or the solution of an inverse problem. The fact that BP does not attempt to deterministically fit seismic waves, but instead relies on the coherent phase of seismic array signals, allows BP to be applied to high-frequency (HF) wavefields ($f \geq 1Hz$).

Several recent studies have demonstrated the effectiveness of using BP method to detect missing early aftershocks that are otherwise masked by the mainshock coda wave (Kiser & Ishii, 2013; Fan & Shearer, 2016). Kiser and Ishii (2013), for instance, back-projected the first 25 hours of continuous recordings following the 2011 Mw 9.0 Tohoku earthquake which resulted in 300 new aftershock detections. It is noteworthy that the BP-detected events augment the JMA earthquake catalog by providing evidence of significantly more active seismicity offshore, particularly near the trench. The BP technique determines the origin time, epicenter, and magnitude of a detected microseismic event based on the spatiotemporal distribution of energy, which is why BP is an end-to-end technique. Using a network of seismograms as input, it outputs microseismicity characters directly. In chapter 2, we investigate the ability to detect the offshore microseismic activity near Japan by conducting continuous BP in a more efficient manner. Additionally, we propose an objective threshold to

identify earthquakes, perform spatial calibration to improve epicenter accuracy, and perform temporal calibration to correct the origin time.

The capability of BP to image high-frequency sources is critical in improving the prediction of strong ground motion, since the typical frequency range of intense shaking is between 1 and 10 Hz (e.g., J. Anderson, 2007; Sokolov & Chernov, 1998). The natural frequency of building response falls in this particular frequency band: A rule of thumb is that the building period equals the number of stories divided by 10, so the 1 to 10 Hz band is roughly responsible for the shaking of 1 to 10 story residential and commercial buildings. Our chapter 4 presents a high-frequency distance metric (R_{hf}) based on observations from the BP. Our study of five $Mw > 7.2$ megathrust earthquakes in Japan and Chile finds that the R_{hf} distance metric outperforms the conventional distance metric (R_{rup}) in predicting the ground shaking intensity between 0.5 and 4 Hz. R_{hf} can be used as a complementary measure to the conventional ground motion prediction equation (GMPE) distance metrics, which is a more accurate predictor of ground motion in many cases.

1.3.2 Matched-filter and Autocorrelation

The aforementioned BP method detects microseismicity by identifying high-frequency (HF) radiators, and here we present another detection method named template matching or matched-filter (MF), which is based on the waveform similarity of close events. If the event sources are close together relative to their common distance from the station, then the seismicograms are assumed to be similar, as the focal mechanism, ray path, and site effects are similar. This method searches for a pattern similar to existing template events within continuous recordings that suggests the possibility of a new event. By stacking the multi-channel cross-correlation coefficients (CCs) between the template waveforms and the continuous recordings, the MF method is capable of detecting $\sim 4 - 10$ times more events than those listed in the routine catalogs, and thus reduces the magnitude of completeness by $\sim 0.6 - 1.0$ (e.g. Turin, 1960; Gibbons & Ringdal, 2006; Shelly, Beroza, & Ide, 2007; Peng & Zhao, 2009; Kato et al., 2012; Huang, Meng, et al., 2017).

The drawback of the MF method is that it relies on known earthquake templates. In situations where templates are unavailable or we are interested in exploring new types of events, the autocorrelation method may be an option. The basic concept is similar to MF, but the templates are generated from sliding windows of continuous waveforms rather than known event windows. The numerous templates make scanning months of data almost unaffordable, as the computational cost increases dramatically with the size of the dataset. By scanning through 1-hour seismic data from 6 Hi-Net stations in Japan, Brown, Beroza, and Shelly (2008) has identified both previously known and unknown low frequency earthquakes. According to the computation efficiency, MF is more commonly used than autocorrelation in long term earthquake detection tasks.

In Chapter 2, we presented a combination of the BP and MF approaches (BP-MF), which incorporates those BP-detected events as additional template events into the MF detection in order to retrieve a more complete picture of offshore seismicity, particularly in the shallow parts of the subduction zone. We conducted BP-MF analysis within 600 days of the 2011 M 9.0 Tohoku earthquake. This study finds 44.2% more offshore events than those listed in the JMA catalog. Near the trench, we detect 213% more events.

1.4 Machine Learning Models

1.4.1 Differences between physics-based and machine learning models

Data-driven models learn to deal with a certain task through data or experience, as their name implies. If the learning process is performed by a machine or by a computer, we refer to it as machine learning (ML). As opposed to physics-based models, ML models are based solely on the statistical distribution of data without assuming any physical assumptions. Physics-based models typically utilize simple mathematical functions because physical equations are simple and beautiful, and humans can only effectively adjust dozens of parameters at one time. Conversely, ML models allow for greater flexibility in expression, depending on the task. With a standard scheme of learning, several to billions of parameters can be efficiently

tuned in an ML model.

We use the movement of Mars as an example to highlight the difference between physics-based models and ML models. Newton's law of universal gravitation can be used as a starting point for constructing a physics-based model that predicts Mars' movement. It illustrates the relationship between gravity force and masses, as well as the relationship between force and acceleration. Using a force analysis, it is possible to approximate the movement of Mars given the initial conditions (e.g., position and velocity of the planet). Most of the parameters in the model have physics meanings with constant values, such as the gravitational constant, the mass of Mars, and the mass of the Sun. In practice, it is difficult to make all physical conditions and assumptions valid in real applications (for example, dark matter is not considered in force analysis; Newton's law works for slow speed conditions, however not for high speed conditions).

Physics-based modeling requires an in-depth understanding of physical systems; in contrast, a ML (statistical) model can be used to estimate the movement path of Mars without having to know any laws of physics and physical information. The ML model can be described with a function $\hat{y} = f_\theta(x)$, where x is the time, $\hat{y}$ is the predicted position of Mars, f is a category of functions with parameters θ. These parameters do not have clear physical meanings and are initialized with random values. During the training process, the optimizer decreases the misfit between observations y and predictions $\hat{y}$ by tuning these parameters. Misfit is often referred to as loss function or target function, which indicates how well the model is performing. ML models are assumed to learn the distribution of training data when the loss function reaches a global minimum or maximum. A trained model can be used to predict future data generated by the same or a similar process (the physical system in this example).

A ML model may sometimes appear as a black box, displaying good performance in real-world applications, but no one knows what the physics or rules are. This phenomenon is not unique, as people have used electrical appliances for decades before the discovery of electrons. Our society is currently in an era where the application of machine learning comes

much earlier than its understanding. For instance, the face recognition system was applied to the latest telephones, but the meaning of millions of parameters in the model is still hard to explain.

1.4.2 Categories of Machine Learning Models

Within this section, we will discuss three different methods for categorizing ML models (task, learning approach, and algorithms) and their application to seismology. The ML model can be categorized according to the problems it solves: regression, classification, and clustering. The aforementioned Mars example is a regression task, whose prediction is numerical data. If the prediction is categorical data, it would be classified as a classification task. Another common type is the clustering task in which the dataset is clustered into groups.

Aside from the difference in task types, ML models can also be classified according to their learning characteristics. The aforementioned Mars example is using supervised learning, which is most common in the applications of machine learning. The observation y represents the correct answer (label) for the training sample x, and it guides the model to update its parameters during supervised learning.

In addition to supervised learning, there is unsupervised learning, semi-supervised learning, self-supervised learning, and reinforcement learning. In an unsupervised learning task, the dataset has no labels, but the model learns a pattern to cluster the samples with methods such as K-means, density-based spatial clustering of applications with noise (DBSCAN), and Gaussian Mixture models (GMM). It has been applied to searching for precursors during laboratory seismic cycles (Bolton et al., 2019), clustering rupture directivity modes (Ross, Trugman, Azizzadenesheli, & Anandkumar, 2020), and clustering seismic signals (Seydoux et al., 2020).

The semi-supervised learning is one between the unsupervised and supervised learning, whose dataset is partially labeled. The model is trained with labeled samples and then predicts unlabeled samples. The procedure includes labeling the samples of high prediction confidence with pseudo-labels, and then continue training the model using these pseudo-

labeled samples. The method has been used to detect rockfalls and volcano-tectonic events in continuous waveform data (Köhler, Ohrnberger, & Scherbaum, 2010).

The self-supervised learning approach uses the samples x themselves as labels. The method has been applied to the removal of spatially incoherent noise from seismic field datasets (Birnie, Ravasi, Liu, & Alkhalifah, 2021) and fiber-optic Distributed Acoustic Sensing (DAS) datasets (M. van den Ende et al., 2021).

Finally, reinforcement learning uses intelligent agents to learn how to take action in a given environment so as to maximize the cumulative rewards (Kaelbling, Littman, & Moore, 1996). During training, the model should find a balance between exploration (uncharted territory) and exploitation (current knowledge). It has been widely used in robotics manipulation, games (e.g. AlphaGo; Silver et al., 2016), and self-driving systems (Kiran et al., 2021). Furthermore, it has been used in dynamic threshold selections for single-station earthquake detectors (Draelos et al., 2018), in seismic control systems (online tuning of a mass-drive system; Khalatbarisoltani, Soleymani, & Khodadadi, 2019) and in minimizing seismicity in industrial projects (Papachristos & Stefanou, 2021).

Finally, ML models can also be categorized by the algorithm they employ. Numerous algorithms have been developed over the years, and here we present only some of the most popular algorithms and their applications in seismic analysis. Support vector machine (SVM) has been applied to predict earthquake early warning magnitudes (Reddy & Nair, 2013; Ochoa, Niño, & Vargas, 2018), earthquake casualties (Xing, Zhonglin, & Shaoyu, 2015), earthquake-induced liquefaction potential of soil (J. Zhou, Huang, Wang, & Qiu, 2021), and mantle convection processes (Shahnas, Yuen, & Pysklywec, 2018). The random forest (Breiman, 2001)has been applied to earthquake catalogs declustering (Aden-Antoniów, Frank, & Seydoux, 2022), earthquake early warning (Z. Li, Meier, Hauksson, Zhan, & Andrews, 2018), earthquake-induced landslides (Miao & Wang, 2015), and ground motion prediction (Trugman & Shearer, 2018). The Markov Chain Monte Carlo approach has been applied to seismic tomography (Bodin & Sambridge, 2009; X. Zhang, Curtis, Galetti, & De Ridder, 2018; Piana Agostinetti, Giacomuzzi, & Malinverno, 2015), the estimation of

earthquake source temporal functions (Debski, 2008), and the prediction of tsunamigenic earthquake magnitudes (Shin, Chen, & Kim, 2015).

Aside from these algorithms, neural networks have become particularly popular in the 2010s and have found great success in many fields such as computer vision and natural language processing. As a result of technologies such as graphics processing units (GPUs) and Tensor Processing Units (TPUs), people are able to build neural networks with numerous layers and millions of parameters. These large, complex neural networks are known as deep learning (DL) models. The applications of DL in seismology have flourished since 2017, including earthquake detection (Reynen & Audet, 2017; Perol, Gharbi, & Denolle, 2018; Z. Li, Meier, et al., 2018; Meier et al., 2019; Saad & Chen, 2020; Saad et al., 2021), earthquake location (Mousavi, Zhu, Sheng, & Beroza, 2019; J. D. Smith, Ross, Azizzadenesheli, & Muir, 2022; M. P. van den Ende & Ampuero, 2020; W. Zhu, Tai, Mousavi, Bailis, & Beroza, 2021), earthquake magnitude estimation (M. P. van den Ende & Ampuero, 2020; Mousavi & Beroza, 2020), focal mechanism determination (Kuang, Yuan, & Zhang, 2021), and polarity determination (Tian et al., 2020). This technique has also been used to detect tremors (Nakano, Sugiyama, Hori, Kuwatani, & Tsuboi, 2019) and moonquakes (Civilini, Weber, Jiang, Phillips, & Pan, 2021).

Next, we will discuss the application of DL to the phase-picking task as well as the history of model development. DL models require a large amount of data, and the high volume of microseismic recordings makes the phase-picking process appropriate for DL models.

1.4.3 Artificial, Convolutional, and Recurrent Neural Networks

The "neural network" algorithm mimics the behavior of neuron cells, hence its name. Here, we take a neuron as an analogy of the smallest computing unit in neural networks. A neuron's dendrites receive signals X from its environment or from other neurons. Different dendrites may have different sensitivities, which can be represented by weightings W for each branch. Before transferring the signals received from all dendrites to other neurons, the selected neuron processes the summarized signal $(X * W)$ with a bias B and an "activation

function". This activation function functions like a filter that is only sensitive to certain signals. Rectified Linear Unit (ReLU) is one of the default activation functions in the state-of-the-art neural networks with the format of $ReLU(x) = max(0, x)$ and here we use it as an example. The processed signal $Y = ReLU(X * W + B)$ is then transferred to other neurons, with positive values for exciting instructions and zero values for unexciting instructions. A neural network is composed of layers of these neuron-like computing units.

Early versions of neural networks were known as artificial neural networks (ANNs). These were popular in the 1990s (Van der Baan & Jutten, 2000; Poulton, 2002). The neurons in an ANN network are connected to each other and the number of parameters is approximately proportional to the number of connections. Because of the limited computational resources available in the 1990s, ANN applications on the phase-picking task generally utilize a few layers (usually one hidden layer) and hundreds of parameters. These ANN models use raw waveforms (Dai & MacBeth, 1995, 1997; Y. Zhao & Takano, 1999) or extracted features (such as the ratio between short-term average and long-term average, the ratio between horizontal power and total power, and the short-axis incidence angle of the polarization ellipsoid; J. Wang & Teng, 1995) as input and predict the probability of seismic phases with a moving (sliding) window. Many of these models are underfitted due to the small number of training samples, which ranges from hundreds to thousands. Although its performance is not the best, the ANN has been used in MyShake, a global smartphone seismic network due to its low computational cost (Kong, Allen, Schreier, & Kwon, 2016).

In the 2010s, thanks to the rapid development of GPU technology, deep networks were soon able to be trained, and a wide variety of network variants developed since then, such as the Convolutional Neural Network (CNN; LeCun, Haffner, Bottou, & Bengio, 1999), Recurrent Neural Network (RNN), and Graph Neural Network (GNN). Various neural networks have been designed for the processing of different types of data (e.g. CNN for images, RNN for audio, GNN for social network data). While some literature considers CNN to be a variant of ANN, we specify that the neurons in ANN are fully connected, while the neurons in CNN are partially connected with a fixed kernel. The kernel slides on the neurons in the

previous layer and processes signals that the kernel can perceive as ANN does. The following equation shows the transformation process from signal X^{l-1} at layer $l-1$ to signal X^l at layer l with a 1D kernel K with size of m and activation function δ.

$$X_i^l = \delta(\sum_{a=0}^{m-1} K_a * X_{i+a}^{l-1})$$

The kernel has a similar function as the filter in the seismic data processing. By applying multiple kernels to the same input, we can achieve different transformed signals (such as high-pass and low-pass) and store them in multiple channels. Compared with ANN, CNN uses much fewer parameters with repetitive weights, thus numerous deep CNN models with dozens of layers were applied on the seismic phase picking task (Ross, Meier, Hauksson, & Heaton, 2018; W. Zhu & Beroza, 2019; L. Zhu et al., 2019; Ross, Meier, & Hauksson, 2018; J. Wang, Xiao, Liu, Zhao, & Yao, 2019; Woollam, Rietbrock, Bueno, & De Angelis, 2019). The performance of these models are on par with human analysts, with an accuracy of $\sim 90\%$.

Meanwhile, some researchers incorporate RNN into their models (Mousavi, Ellsworth, Zhu, Chuang, & Beroza, 2020; Mousavi, Zhu, et al., 2019; Y. Zhou, Yue, Kong, & Zhou, 2019; Wiszniowski, Plesiewicz, & Trojanowski, 2014) because it can interpret temporal data and seismograms as sequential data. Generally, models that incorporate RNN outperform models built purely with CNN, because the RNN stores and utilizes seismogram context information. However, RNNs cannot be stacked into very deep models in practice due to the vanishing and exploding gradient problems (Hochreiter, 1998). Specifically, the gradient of the parameters would be either too large or too small to update in the training process. One common solution is to incorporate Long Short-Term Memory (LSTM; Hochreiter & Schmidhuber, 1997; Magana-Zook & Ruppert, 2017) or Gated recurrent units (GRU; Cho, Van Merriënboer, Bahdanau, & Bengio, 2014) in RNNs. It is worth noting that they have been successful in dealing with complex sequential data, such as seismograms, DNA sequences, articles, dialogues, and voices (Liu et al., 2019; Wen et al., 2015; Lehner, Widmer, & Bock, 2015).

1.4.4 Graph Neural Network

The CNN and RNN phase-picking models have produced promising results in some real-world applications, but there is still room for improvement. Early CNN and RNN models either used a single station or a network with a fixed configuration (Zheng, Harris, Li, & Al-Rumaih, 2020; Yang, Hu, Zhang, & Liu, 2021), since the input channels of CNN and RNN should have a fixed size and order. The former models can't benefit from the waveform consistency between multiple stations, like the conventional single-station phase-picking models. The latter models are trained and tested on the same station configuration and region, which limits their generalization capability to other station networks or tectonic settings. A solution for establishing communications between multiple stations is to use the Graph Neural Network (GNN; Gori, Monfardini, & Scarselli, 2005; Scarselli, Gori, Tsoi, Hagenbuchner, & Monfardini, 2008; J. Zhou et al., 2020). GNN is the neural model that captures the dependence of graphs via message passing between the nodes of graphs (J. Zhou et al., 2020). GNN can freely exchange information between multiple stations, which can handle the irregular structures of graphs with a fixed model architecture (Z. Zhang, Cui, & Zhu, 2020). It has been successfully implemented in seismic source characterization (M. P. van den Ende & Ampuero, 2020; McBrearty & Beroza, 2022) and phase association problems (McBrearty & Beroza, 2021).

There are numerous model architectures in the GNN family and they can be categorized into five categories: Graph Recurrent Neural Networks, Graph Convolutional Networks (GCN), Graph Autoencoders, Graph Reinforcement Learning, and Graph Adversarial Methods (Z. Zhang et al., 2020). This thesis only focuses on one specific type of GCN network referred to as graph message passing network (Gilmer, Schoenholz, Riley, Vinyals, & Dahl, 2017), which is suitable for the phase-picking task. It enables exchange of information between neighboring nodes, which enhances the relevant signals shared by adjacent nodes.

In Chapter 3, we build a multi-station phase-picking model named EdgePhase by incorporating a graph message passing network into a state-of-the-art single-station phase-picking model, EQTransformer (Mousavi et al., 2020). The EQTransformer contains the CNN and

RNN modules described above. As compared to EQTransformer, EdgePhase improves both precision and recall by 5% on our training and test sets for Southern California earthquakes. In our study of early aftershocks following the M7.0 earthquake in Samos, Greece in 2020, the EdgePhase model demonstrated a strong regional generalization ability. When compared to a local earthquake catalog, EdgePhase detected 190% more events with a distribution that is closer to a planar fault interface. Through a GNN framework, the seismograms and geographic locations of multiple stations are treated as a graph. The communication between stations through GNN has taken the DL phase-picking model into a new era, where phase detection is not only dependent on seismograms from one station, but also on the waveform consistency in a seismic network.

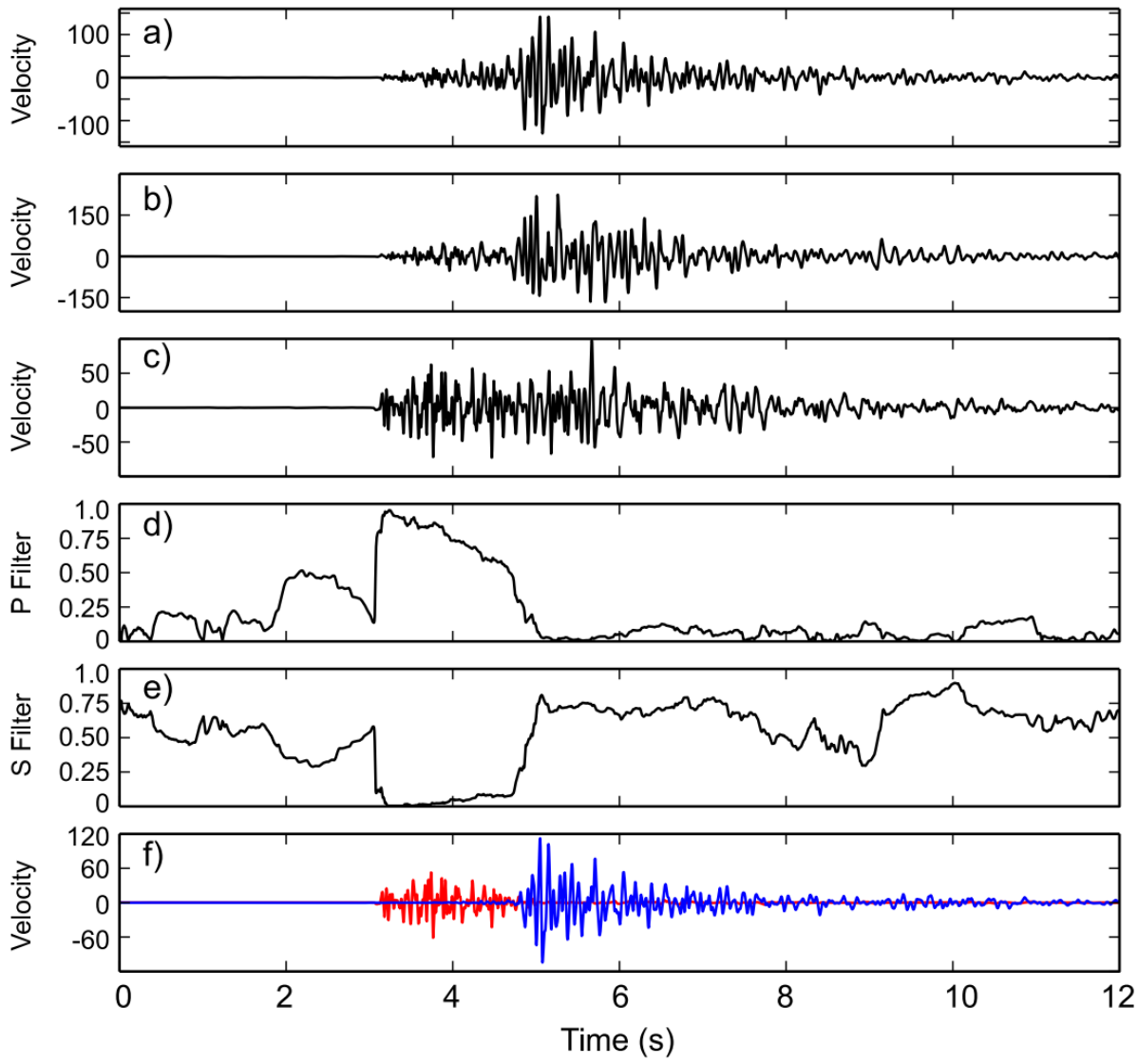

Figure 1.1: An application of polarization analysis to separate a seismogram into P and S phases. Adapted from Ross and Ben-Zion (2014). (a-c) North, east, and vertical component velocity seismograms. (d) A P-polarization filter based on a covariance matrix with high values during the P wave and low values during the S wave. (e) A corresponding polarization filter based on the covariance matrix. (f) A P-polarized vertical trace (red) plotted together with the S-polarized east–west trace (blue).

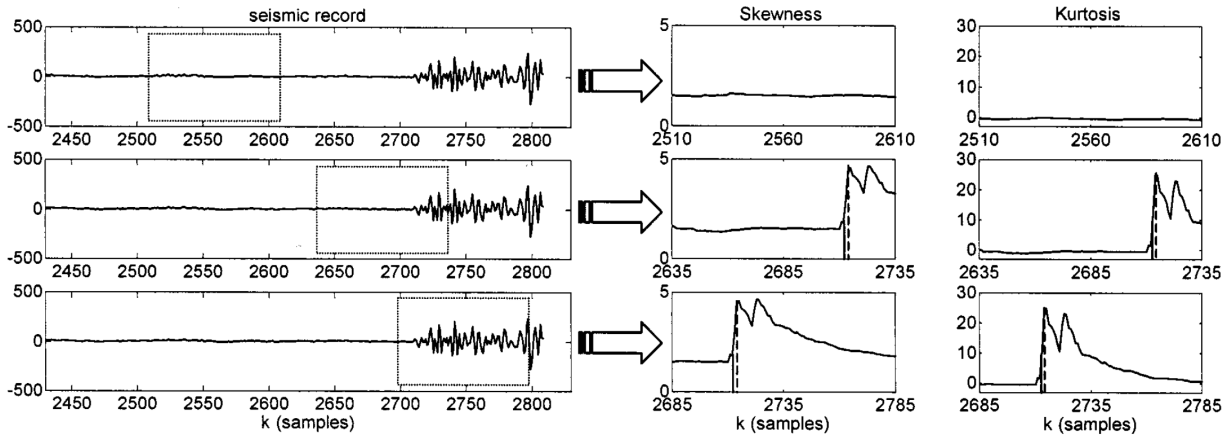

Figure 1.2: Skewness and kurtosis of the waveforms in the sliding windows. Adapted from Saragiotis et al. (2002).

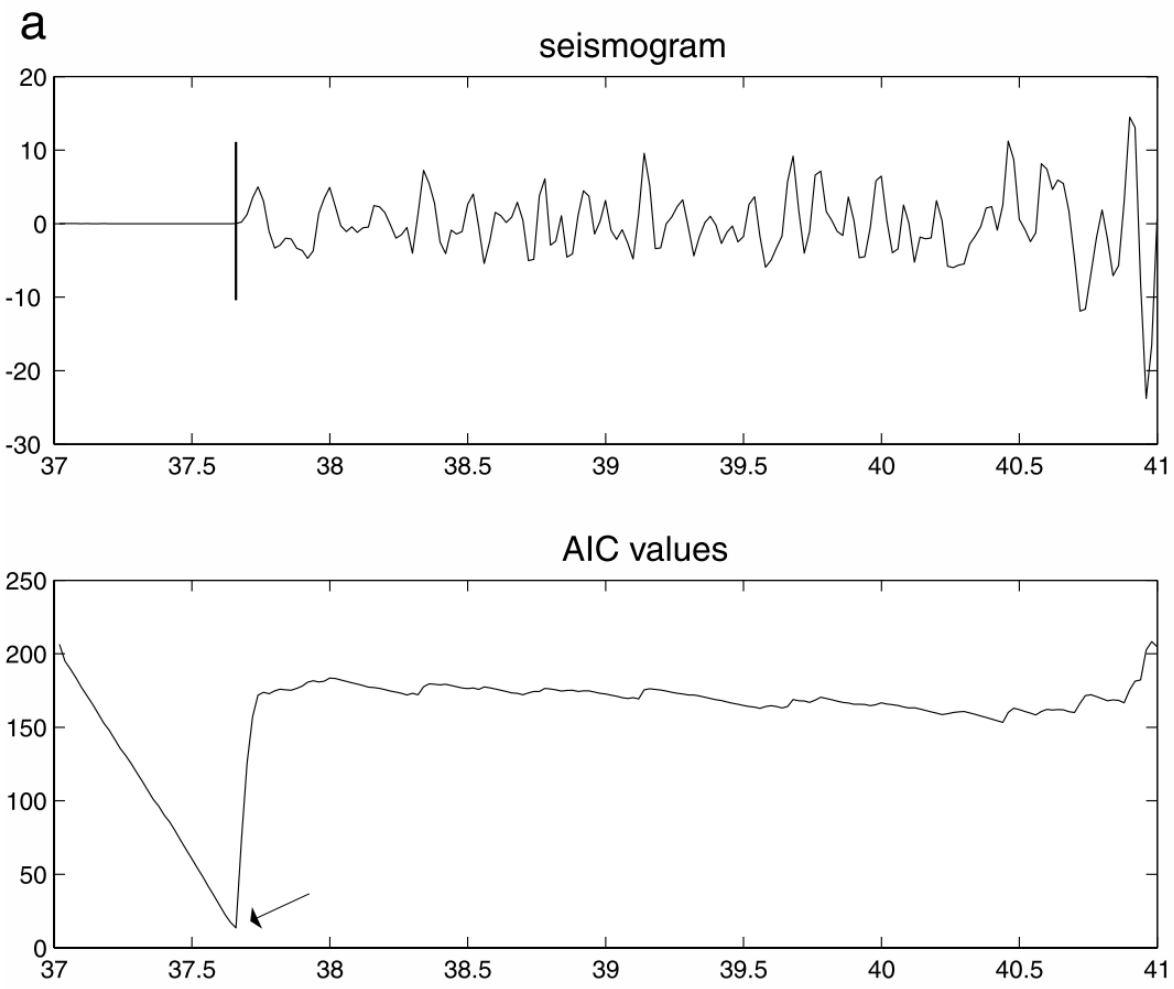

Figure 1.3: Seismograms and the corresponding AIC values. Adapted from H. Zhang et al. (2003).

CHAPTER 2

Detecting Offshore Seismicity: Combining Backprojection Imaging and Matched-filter Detection

Originally published in **Feng, T.**, Meng, L., & Huang, H. (2020). Detecting Offshore Seismicity: Combining Backprojection Imaging and Matched-filter Detection. Journal of Geophysical Research: Solid Earth, 125(8), e2020JB019599. I contributed to the experiments and writing of the paper. Linsen provides instructions on Backprojection and paper revision. Hui provides instructions on the Matched-filter.

Note: the supplementary materials are included in Appendix A.

Abstract

An important manifestation of the background or coseismic deformation surrounding megathrust earthquakes is the offshore microseismicity, which are difficult to be directly detected by land-based seismic instruments. Here, we improve the capability of detecting offshore events by combining two popular techniques: backprojection imaging (BP) and matched-filter detection (MF). The BP method is effective in retrieving offshore seismicity ($Mw > 4.5$) buried in the coda wave of large earthquakes. The capability of the MF method depends on the availability of the template pool, therefore the BP-inferred events can be used as additional templates to expand the MF detections (BP-MF). We performed the BP-MF approach in the period within 600 days after the 2011 Mw 9.0 Tohoku earthquake. We find overall 44.2% more offshore events than those listed in the Japan Meteorological Agency (JMA)

catalog. In the near-trench area, we detect 213% more events. Among the newly detected $Mw > 4$ events, the BP-template contributions are twice more than those matched by the JMA templates. Based on the spatial consistency between aftershock-depleted zones and large coseismic slip, we identify a possible large-coseismic-slip zone in the near-trench region offshore Fukushima. Large b values (≥ 1.2) are found close to large aftershocks, possibly indicating localized pockets of small differential stresses. At several locations close to the trench, p values (0.93 - 1.11) are higher than those in the inland area (0.64 - 0.85). This may be due to the larger coseismic slip and hence larger stress drop of the outer-rise normal-faulting events compared to the deeper thrust-faulting events.

2.1 Introduction

Megathrust earthquakes occur along the plate interface of subduction zones with the large coseismic slip generally distributed in the offshore region. In contrast to the less frequent megathrust earthquakes, small offshore earthquakes occur much more frequently according to the Gutenberg–Richter law (Gutenberg & Richter, 1944). The small offshore earthquakes carry key information in understanding the pre-, co- and post-seismic slip processes of megathrust earthquakes (e.g. Frank et al., 2017; L. Meng et al., 2016; Kato et al., 2012; Huang, Meng, et al., 2017; Huang, Xu, et al., 2017; Uchida & Matsuzawa, 2013; Lengliné et al., 2012). Small offshore earthquakes are also important to monitor the seafloor crustal deformation (Tadokoro et al., 2006; Bungum, Lindholm, & Faleide, 2005), activities of offshore faults (Uchida, Shimamura, Matsuzawa, & Okada, 2015), and gas injection at the offshore platform (Cesca et al., 2014). However, due to the lack of nearby stations, the complete characterization of offshore earthquakes is still challenging.

Earthquakes are detected by identifying the prominent seismic phases in the continuous waveforms. The phase identification is commonly achieved by examining the short-time average to long-time average ratio (R. V. Allen, 1978; Earle & Shearer, 1994), kurtosis and skewness (Saragiotis et al., 2002), waveform polarization (Vidale, 1986; Ross & Ben-Zion, 2014), Akaike information criteria (Akaike, 1998; Sleeman & Van Eck, 1999; Leonard

& Kennett, 1999; Leonard, 2000) and discrete wavelet transform (Anant & Dowla, 1997; H. Zhang et al., 2003). These methods, however, may fail to detect small events with weak arrivals hidden in the noise or overlapping phases within a short time period.

The matched-filter (MF) method searches for similar patterns of existing template events in the continuous recordings as suggestive of a new event. By stacking the multi-channel cross-correlation coefficients (CCs) between the template waveforms and the continuous recordings, the MF method is capable of detecting $\sim 4-10$ times more events than those listed in the routine catalogs, and thus reduces the magnitude of completeness by ~ 1.0 (e.g. Turin, 1960; Gibbons & Ringdal, 2006; Shelly et al., 2007; Peng & Zhao, 2009; Kato et al., 2012; Huang, Meng, et al., 2017). However, the MF method relies on known template waveforms, which limits its power in detecting offshore events where the template pool is highly incomplete.

The back-projection (BP) method has been widely employed to image the earthquake rupture process based on coherent teleseismic P phases (e.g. Rost & Thomas, 2002; Ishii et al., 2005; Krüger & Ohrnberger, 2005; L. Meng et al., 2012; Kiser & Ishii, 2017). Recent studies have also demonstrated the effectiveness of the BP method in detecting missing early aftershocks that are otherwise masked by the mainshock coda wave (Kiser & Ishii, 2013; Fan & Shearer, 2016). For example, Kiser and Ishii (2013) back-projects the first 25-hour continuous recordings following the 2011 Mw 9.0 Tohoku earthquake, resulting in 300 newly detected aftershocks. These BP-detected events are complementary to the JMA earthquake catalog, revealing significantly more active offshore seismicity, especially near the trench. This study motivates our work to combine the BP and MF approaches (BP-MF), which incorporates the BP-detected events as additional template events into the MF detection, with the aim to retrieve a more complete picture of offshore seismicity, especially in the shallow part of the subduction zone.

In section 2, we introduce the continuous BP method and demonstrate its effectiveness to detect aftershocks of the 2011 Mw 9.0 Tohoku earthquake. Section 3 shows the MF detection based on the combined template pool. The result illustrates the improvement in

mapping the offshore seismicity, and in refining the b and p value statistics. In section 4, we discuss the relationship between the aftershock-depleted area with the large-slip area of the Tohoku mainshock, the advantages, challenges, and the further development of the BP-MF method.

2.2 Backprojection-imaged Aftershocks of the 2011 Mw 9.0 Tohoku Earthquake

2.2.1 Continuous Backprojection

In the previous applications, BP was usually performed in seconds to minutes (e.g., for imaging the rupture properties of large earthquakes). Here, for the purpose of aftershock detections, we conduct continuous BP in a time span of 600 days following the 2011 Mw 9.0 Tohoku earthquake. To efficiently perform continuous BP, we adopt the conventional beamforming (linear stacking) instead of more sophisticated methods such as coherency stacking, MUSIC or compressive sensing (Fletcher, Spudich, & Baker, 2006; L. Meng et al., 2012; Yao, Gerstoft, Shearer, & Mecklenbräuker, 2011) . These techniques are more capable of separating simultaneous earthquake sources, which are essential in imaging rupture details. However, since the detection of aftershocks is less sensitive to the imaging resolution, we choose linear stacking which is more computationally efficient and also provides undistorted amplitude information to constrain the event magnitudes (Kiser & Ishii, 2013).

We obtain the vertical velocity seismograms recorded by 460 broadband stations in North America (Fig. 2.1c) located between 40° and 90° away from the epicenter. We then align the waveforms by the initial P arrivals and filter the data between 0.8 - 2 Hz, the same frequency range as that used in Kiser and Ishii (2013). Considering the waveform coherence in the network, the filtering step is necessary. In the following manuscript, we use the abbreviation 'KI-2013' when referring to their study. The seismogram of each station is normalized by the standard deviation of the 1500-sec window after the initial P arrival of the mainshock. We back-project the 600-day continuous waveforms onto a 7° by 7° source area, with a grid

spacing of 0.1° (Fig. 2.1a). The back-projection is performed in a 20-second long sliding window with a time step of 1 second. The spatial distribution of the beamforming power (Fig. 2.1d) at time t is calculated by:

$$s_i(t) = [\sum_{k=1}^{n} u_k(t + t_{ik} + \delta t_k)]^2 \tag{2.1}$$

where s_i is the beamform power at the ith grid point, u_k is the normalized vertical seismogram at station k, t_{ik} is the theoretical travel time based on a 1D velocity model (IASP91) between the ith grid point and kth station, t is time with respect to a reference event, and n is the number of stations. δt_k is the empirical travel time correction estimated by cross-correlating the 10-second windows around first P arrivals of the mainshock, which accounts for travel time errors due to 3D velocity structures (Ishii et al., 2005; L. Meng et al., 2016).

2.2.2 Aftershock Recognition

The stacking of coherent impulsive energy from earthquakes usually produces a peak in the BP-power time series. For each peak at time t, we compute its maximum stacked energy $p(t)$, spatial background energy $A_s(t)$, temporal background energy $A_t(t)$. We automatically scan through the time series of $p(t)$ and select peaks with the width larger than 3 seconds and the prominence (how much the peak stands out due to its intrinsic height and its location relative to other peaks) larger than 10^{-5}. These peaks are considered as candidate detections (Fig. 2.1e).

$$p(t) = max(s_i(t)) \qquad i \in [1, m] \tag{2.2}$$

$$A_s(t) = \frac{\sum_{i=1}^{m} s_i(t)}{m} \tag{2.3}$$

$$A_t(t) = \frac{\sum_{j=t-\delta t}^{t+\delta t} p(j)}{2\delta t} \tag{2.4}$$

where the total number of grids m is 4,900 and the half-window length δt is 20 seconds, respectively.

The strength of a peak is characterized by the temporal signal-to-noise ratio (SNR) $p(t)/A_t(t)$, and the spatial SNR $p(t)/A_s(t)$, which indicates how concentrated the detected energy pulses appear in the spatial and temporal domain respectively. We expect that earthquakes have higher $p(t)/A_s(t)$ and $p(t)/A_t(t)$ than noise and coda waves. Previously, KI-2013 identify aftershocks as peaks with $p(t)/A_t(t) > 2$. Additional quality control is done by visually inspecting the spatial energy distributions. Here, in order to avoid the subjective decisions in the manual picking process, we establish an automated picking procedure based on both $p(t)/A_s(t)$ and $p(t)/A_t(t)$. To determine the detection threshold, we compare the positions of windows containing earthquakes, background noise, and the coda waves in the $p(t)/A_s(t)$ and $p(t)/A_t(t)$ domain (Fig. 2.2). We randomly select approximately 2 million 20-sec-long windows after the first 24 hours following the mainshock to represent background noise and 3 thousand windows within the first 24 hours following the mainshock to represent the coda waves of the mainshock and its intense early aftershocks. Both the noise and coda windows are taken 60 seconds away from any catalog events to avoid including any earthquake waveforms. Note that the noise or coda windows we used here could, in theory, still contain some earthquakes missed by the JMA catalog. However, due to the temporal sparsity of earthquakes, this effect is negligible.

In Figure 2.2, we find that the majority of large earthquakes ($Mw \geq 4.5$) in the JMA catalog are located in Region I (high $p(t)/A_s(t)$, high $p(t)/A_t(t)$), indicating peaks associated with earthquakes are strong in both the spatial and temporal domain. A small portion of the windows are located in Region IV, but very few are located in Region II and III. The noises are concentrated in Region III (low $p(t)/A_s(t)$, low $p(t)/A_t(t)$), which suggests that most noise windows do not produce strong spatial or temporal peaks. A significant portion of the noise are located in Region IV (low $p(t)/A_s(t)$, high $p(t)/A_t(t)$)). These "noise" windows may carry impulsive energy (e.g. microseisms) but can not be coherently stacked by P-wave travel times in our study area. Most of the Earthquakes smaller than $M4.5$ are also distributed in Region IV overlapping with the noise windows (Fig. 2.2b), indicating that they are too small to produce strong spatial peaks and can not be reliably distinguished from the noise. The

event detection in KI-2013 only requires large $p(t)/A_t(t)$ (Region I and IV), which leads to a high probability of false detections by identifying a large amount of noise windows in Region IV as earthquakes. Most coda waves are also located in Region III, although some strong coda waves produce significant spatial peaks in Region II (high $p(t)/A_s(t)$, low $p(t)/A_t(t)$). These windows might correspond to strong P-wave scatters but since coda waves are not as impulsive as direct arrivals, they seldom form strong peaks in the time domain. In summary, we consider that only the signals within Region I (high $p(t)/A_s(t)$, high $p(t)/A_t(t)$) can be reliably identified as earthquakes.

Next, we determine the boundary of Region I that best separates the earthquakes from noise and coda waves by minimizing the Shannon information entropy (Shannon, 1948). This procedure can be considered as a two-nodes classification tree based on two attributes: $p(t)/A_s(t)$ and $p(t)/A_t(t)$. We measure the information entropy (purity) H, which is the summation of the negative logarithm of the probability mass function for each possible value (Shannon, 1948).

$$H = -\sum_i P_i \cdot log_2(P_i) \tag{2.5}$$

Where P_i is the probability of event i. In our model, the event is either classified as earthquakes or noises, which are denoted by the suffix e and n respectively. The entropy before the classification H_0 is:

$$H_0 = -P_e \cdot log_2(P_e) - P_n \cdot log_2(P_n) \tag{2.6}$$

The entropy after the classification is the weighted summation of Regions I and II, III, IV.

$$\begin{aligned} H_1 \quad = \quad & -(P_{e_{II,III,IV}} \cdot log_2(P_{e_{II,III,IV}}) + P_{n_{II,III,IV}} \cdot log_2(P_{n_{II,III,IV}})) \times \frac{N_{II,III,IV}}{N} \\ & -(P_{e_I} \cdot log_2(P_{e_I}) + P_{n_I} \cdot log_2(P_{n_I})) \times \frac{N_I}{N} \end{aligned} \tag{2.7}$$

The information gain G after segmentation is

$$G = H_0 - H_1 \tag{2.8}$$

The larger the information gain is, the purer the classification will be. The best solutions of $p(t)/A_s(t)$ and $p(t)/A_t(t)$ by grid search are 3.4 and 1.5 respectively (Fig. 2.2c). However, some of the noises in Region I possibly are events missed by the JMA catalog. Since our objective is to detect more missing events by the JMA catalog at a reasonable false-alarm rate, we chose a slightly lower threshold of $p(t)/A_s(t) > 3$ and $p(t)/A_t(t) > 1.2$ rather than the best solution of the purest classification. The temporal peak threshold of $p(t)/A_t(t) > 1.2$ is lower than $p(t)/A_t(t) > 2$ in KI-2013 in order to detect more events buried under the noise. Meanwhile, the spatial threshold $p(t)/A_s(t) > 3$ ensures strong noise is not misidentified as earthquakes. Overall, we consider our detection threshold balances the tradeoff between false positives (0.004) and false negatives (0.208) (Fig. A.1).

The threshold above yields 3,156 detections from continuous BP. We consider a BP detection matched a known event in the JMA catalog if they occurred within 60 s and 50 km from each other. Under such matching criteria, we matched approximately 78% $M \geq 4.5$ events listed in the JMA catalog. Among the 22% (380) "unmatched" events, our BP, in fact, detected 47% of their energy peaks. However, a number of factors prevent us from matching these energy peaks with events in the JMA catalog. First, the strength of the energy peak is smaller than the spatial or temporal threshold (Fig. A.2). We used a conservative detection threshold that keeps a high precision rate (0.999) at the expense of the recall rate (0.792) (Fig. A.1). Second, the energy peak of the event passes the threshold but can not match with the events in the JMA catalog due to large differences in location or origin time (Fig. A.3 and A.4), as a result of the errors of both catalogs. In some rare cases, when two energy peaks occur in less than 30 sec, our automatic matching program only retains the first detection and considers the second energy peak represents the S waves of the first event instead of the P waves of another event (Fig. A.5). For the remaining 53% missing events, the SNR of the waveforms is too low to produce any energy pulses in the spatial and temporal BP domain. The low SNR may be attributed to event-specific factors, such as temporary high

noise levels and low radiated energy in the operating frequency band of BP. We also find 1,090 previously undocumented events (Fig. 2.3a), with a significant portion located in the offshore region. The mean distance separation between matched events is about 21 km. This difference mainly comes from uncertainties of the BP locations (Fig. 2.3b). The uncertainty of the BP locations is up to 30 km based on inspecting the discrepancy of event locations imaged by different arrays (L. Meng et al., 2016), while the location errors listed in the JMA catalog are less than 3 km. In the next section, we developed a procedure to calibrate the epicenter, origin time and magnitude of BP-detected events based on the JMA catalog.

2.2.3 Spatial Calibration

As indicated in Equation 1, BP usually applied a hypocentral alignment to correct travel time errors due to 3-D velocity structures (Ishii et al., 2005). Such correction presumes that all the source locations in the imaging domain share the same travel time error with the epicenter. Apparently, this assumption is valid in the proximity to the epicenter but is less accurate for more distant sources. Additional in-situ travel time errors can be obtained by inspecting the time shifts to align the aftershock waveforms. The travel time errors evaluated at each aftershock location can be either empirically or parametrically interpolated to cover the entire mainshock rupture zone (Ishii, Shearer, Houston, & Vidale, 2007; Kiser & Ishii, 2017; Palo, Tilmann, Krüger, Ehlert, & Lange, 2014; L. Meng et al., 2016; L. Meng, Huang, Xie, Bao, & Dominguez, 2019; Feng & Meng, 2018). However, since the accurate waveform cross-correlation requires a large SNR, the travel time error estimations are limited to $Mw > 5.5$ aftershocks recorded at the teleseismic distance. The travel time based on these sparsely distributed large aftershocks is less effective in capturing the local travel time variation. Here, instead of calibrating the travel time errors, we decide to directly calibrate the BP event locations based on linear spatial interpolation. This strategy takes advantage of more than 2,000 matched aftershocks densely distributed in the $7^\circ \times 7^\circ$ imaging domain. Many of these events are below Mw 5.5 and are too small for cross-correlation, but their locations can be nevertheless imaged with BP and provide better spatial coverage. We presume there

exists a smooth function that maps the BP-inferred locations and to their corresponding JMA hypocenters. Without solving the mapping function explicitly, we linearly interpolate the value of the mapping function at each grid point. We then apply this mapping function to calibrate the locations of 1,090 newly detected BP events undocumented by the JMA catalog.

Figures 2.3 (c-d) shows the distribution of newly detected BP events after the spatial calibration. The mean distance between matched event pairs decreased from 21 km to 11 km (Fig. 2.3d). Before the spatial calibration, the mean location errors of BP events are 4.69 km in the E-W direction and 1.93 km in the N-S direction, respectively (Fig. 2.4a). In other words, BP locations are systematically biased 5 km northeastward. After the calibration, the mean error in the EW and NS directions decreased to 0.4 km and 0.12 km, respectively. The standard deviation of EW and NS errors decreased from 18.95 km and 13.28 km to 12.23 km and 8.08 km, respectively. The smaller spatial errors allow matching 21 additional BP events with the JMA events.

2.2.4 Temporal Calibration and Magnitude Estimation

We also attempt to calibrate the origin time of the BP events with the JMA catalog. The accuracy of the earthquake origin time is essential in extracting template waveforms from the local network recordings. The time errors of the origin time come from two sources: First, the BP event time corresponds to the peak energy time rather than the event onset time. The time delay between BP peak energy time and origin time is approximately half the earthquake duration, ranging from 1 second for Mw 4.5 events to 12 seconds for Mw 6.0 events (Vidale & Houston, 1993). The second error source is the tradeoff between origin time and earthquake location, especially the depth. BP method is not sensitive to the source depth, therefore we perform the BP at the constant mainshock depth. However, the depths of the aftershocks range from 0 to 80 km. Thus the travel time error caused by inaccuracies in depth is not negligible. The origin time error of matched events ranges from 0 to 40 seconds. On average, the BP origin time is 20 seconds ahead of the JMA event (Fig. 2.4b).

Simply subtracting the mean value of time error roughly fills the gap between the origin time and peak energy time. However, this constant time correction does not reduce the variance of the error. Here, we map the origin time error as a function of the source location and apply a temporal calibration in a similar manner to the spatial calibration. As a result, the standard deviation of origin time errors decreased from 10 to 5 seconds (Fig. 2.4b).

Next, we estimate the magnitude of BP events by performing the least squares linear regression between the magnitudes of matched earthquakes and their corresponding logarithmic BP power (Fig. A.6). The empirical relationship between BP energy E_{BP} (maximum value of $p(t)$ in the earthquake duration) and magnitude M is shown below:

$$M = 0.481 \times log(E_{BP}) + 4.645 \tag{2.9}$$

The coefficient of determination is 0.712, which confirms the linear relationship between the BP energy and the earthquake magnitude. The slope between logarithmic moment and moment magnitude is about 0.67 according to the definition. Because the BP technique reveals high-frequency energy burst and large earthquakes tend to have a smaller portion of high-frequency energy, it is not surprising that the slope of Equation 9 is smaller than 0.67. In addition, the 95% prediction limit is around 0.6. After the calibration, we consider that the uncertainty of the epicenter, origin time, and magnitude of BP events after temporal and spatial calibration is around 10 km, 5 sec, and 0.6 respectively. Our method produces similar results to that of KI-2013 in the first 25 hours after the mainshock, although we pick up 10% more events. More detailed comparison between our study and KI-2013 is included in the supplementary materials (Text. A.1, Fig. A.7 and A.8).

2.3 Offshore Seismicity Imaged by MF Detection

2.3.1 Template Quality Control

We combine the JMA earthquake catalog ($M \geq 4.5$) with the BP events ($M \geq 3.8$, Fig. 2.5) into our template event dataset. We collect three-component continuous seismograms

within 600 days after the Tohoku mainshock, which are recorded by 47 Hi-net stations along the coastline (Fig. 2.1b). Template waveforms are extracted from 10 seconds before to 30 seconds after the theoretical S arrival times. We filter the template waveforms with a fourth-order Butterworth filter in a frequency band of 1 - 6 Hz. To ensure the high quality of template waveforms, we first keep the template events with the SNR ratio greater than 5 at least 12 channels. Then we apply the deep-learning-based Generalized Phase Detection (GPD) to identify the template waveforms with high probability ($\geq$ 98%) being S phases (Fig. A.9). The GPD framework is developed based on $\sim$ 4.5 million training seismograms in California and is well generalized to phase detections in other regions (Ross, Meier, Hauksson, & Heaton, 2018).

As a result, we retain 605 out of the 1,080 BP templates and 1,896 out of the 2,201 JMA templates (Fig. 2.5a). Figures A.10 and A.11 show two examples of the BP template waveforms. The theoretical P and S wave arrivals based on the IASP91 model match the observed phase arrivals well at regional recordings, confirming the location accuracy of the BP template events. To investigate how the template dataset is improved by the addition of BP templates, we divide the fault plane into 9 along-dip segments with a spacing of 40 km (Fig. 2.5a). Figure A.12 shows the number of BP and JMA templates and the percentage of BP templates among all templates in every segment. The BP template percentage increases from $\sim$ 20% in the near-offshore region to $\sim$ 30% in the near-trench region, indicating that the BP technique is especially effective in recovering distant offshore earthquakes missed by the JMA catalog. By inspecting the origin times of BP templates, we found that 187 of them (30.9%) occur within the first 24 hours after the Tohoku mainshock. BP templates are rarely found after 500 days since the mainshock, at the rate of 0.06 events per day (Fig. A.13). The BP templates in the last 100 days only take up 1% of the total amount, therefore we stop performing the continuous BP after 600 days since the mainshock.

2.3.2 Matched-filter Detection

We compute the CC between the template and daily continuous waveforms with a 40-s sliding window. The CC series in each channel are averaged after accounting for the travel time differences at different stations. We define positive detections if the mean CC exceeds 9 times the median absolute deviation (MAD) of the background CC on a given day. The magnitude is inferred based on the median value of the maximum amplitude ratios between the detected and template event at all channels (Peng & Zhao, 2009; X. Meng, Yu, Peng, & Hong, 2012). Figure A.14 shows an example of a positive detection that occurred $\sim$ 6 days after the mainshock. The waveforms of the $M4.8$ template and the $M3.7$ detected event show significant similarity across the 47 stations. After removing self and duplicate detections by multiple templates, we detected 98,967 new aftershocks within 600 days after the mainshock, 16,411 of which are matched by the 605 BP template events. Among these 16,411 events, 1,314 (8%) of them are also detected by JMA templates but with smaller CCs, therefore, they are linked to the BP templates. In the density map of MF detections by BP templates, 92.1% detections are offshore events and 34.7% events are within 80 km from the trench (Fig. A.15). In comparison, 55.8% of events are offshore events and 5.7% of events are within 80 km from the trench in the JMA catalog. We then build the BP-MF catalog by combining 223,897 JMA catalog events, 82,556 detections matched by JMA templates, and 16,411 detections matched by BP templates.

The number of events in the BP-MF catalog is 44.2% more than the original JMA catalog. Most of the new detections are within the magnitude range of 0 to 4, with the median near 1.9 (Fig. 2.6a). Among the new detections, $\sim$ 16.6% are detected by the BP template events (24.2% of total templates). Even though BP templates do not detect as many events as JMA templates, they contribute 71.5% of the new detections of earthquakes with magnitudes larger than 4.0. We then focus on the new detections near the trench (east to the 20 km iso-depth of the subducting slab). Most of the new detections are within magnitude 1 to 4, with the median near 2.9 (Fig. 2.6b). Among the new near-trench detections, $\sim$ 31.5% are matched by the BP template events. This percentage is almost doubled compared to the BP-matched

events in the total detections. In conclusion, offshore seismicity detection was significantly improved by incorporating the BP templates in MF detection.

We compare the event density of the JMA catalog and our BP-MF catalog at $0.2° - by - 0.2°$ grids (Fig. 2.7). In the JMA catalog, 94.4% aftershocks are located to the west of the mainshock epicenter and are most densely distributed near the coastline (Fig. 2.7a). The remaining aftershocks (5.6%) are located near the trench with less than 500 events in each grid. In comparison, the BP-MF catalog illuminates additional aftershock-rich areas near the trench (Fig. 2.7b). There are significantly more aftershocks to the east of the mainshock in the BP-MF catalog (39,812) than those in the JMA catalog (13,117). The ratio of seismicity density between the BP-MF and the JMA catalogs reaches 3 near the trench (Fig. A.16). The aftershock activity is intense near the outer trench-slope normal fault zone (gray rectangle in Fig. 2.7b). Most of the events to the east of the trench are determined as non-interplate events in the oceanic plate by ocean bottom seismographs (Obana et al., 2012). These events are likely activated by the Coulomb stress change induced by the mainshock, on the order of 0.15 to 1.5 MPa (Lay, Ammon, Kanamori, Kim, & Xue, 2011; Toda, Lin, & Stein, 2011). The aftershock density reaches 1,500 events per grid near $40.9°N, 145.1°E$ and $38.3°N, 144.7°E$, where vigorous aftershock sequences followed an Mw 6.9 normal-faulting event on 14 March 2012 and an Mw 7.6 event on 11 March 2011. Based on the aftershock density map, we sketch the aftershock-depleted area (blue line in Fig. 2.7b), with a large portion to the north of the mainshock hypocenter (Region A) and a long channel in the south (Region C).

2.3.3 b and p Value Statistics of the BP-MF catalog

Based on our new BP-MF catalog, we also attempt to analyze the statistics of seismicity in the offshore area. We first investigate the frequency-magnitude distribution (FMD) by fitting our observation with the Gutenberg-Richter law (Gutenberg & Richter, 1944):

$$log_{10}(N) = a - b \cdot M \tag{2.10}$$

where N is the cumulative number of earthquakes with magnitudes larger than M, and

a and b are constants. The b value can be used to estimate the stress state, slab age, and fractal dimension of faults (Tormann, Enescu, Woessner, & Wiemer, 2015; Nishikawa & Ide, 2014; Hirata, 1989). To achieve a statistically reliable b value, we remove the events with magnitudes lower than the magnitude of completeness (M_c), the minimum magnitude above which all earthquakes are reliably recorded. We grid search the b value and M_c that best satisfy the log-linear FMD relations. (Wiemer & Wyss, 2000; Woessner & Wiemer, 2005; Chamberlain et al., 2018). The goodness of fit is estimated by the absolute difference (R) of the number of events in each magnitude bin between the observed and synthetic distribution

$$R = 100 - (\frac{\sum_{M_i}^{M_{max}} |B_i - S_i|}{\sum_i B_i} \cdot 100) \tag{2.11}$$

where B_i and S_i observed and predicted the cumulative number of events in each magnitude bin (Wiemer & Wyss, 2000). A larger value of R indicates a better goodness of fit. In the example shown in Figure A.17, R achieves the maximum value when M_c and b value are 3.8 and 1.09, respectively. We then calculate the M_c and b value at $0.2^\circ - by - 0.2^\circ$ grids. We keep the results if the number of events is larger than 100 and the R is larger than 90. Since the BP-MF catalog provides a better fit for the FMD relation than the JMA catalog, we consider that b value and M_c estimation of the BP-MF catalog is more reliable (Fig. A.18).

The spatial distribution of the M_c is shown in Figure 2.8. We find that the M_c generally increases from the coastline to the trench in both the JMA and BP-MF catalog. To highlight the M_c variation in the along-dip direction, we group M_c values into 50 bins according to the distances to the trench and calculate the median M_c in each bin. We find that the median M_c of the BP-MF catalog is approximately 1.0 on land, 2.0 between the coastline to the trench, and 2.5 to the east of the trench (Fig. 2.8b and d). Compared with the original JMA catalog (Fig. 2.8a and c), the BP-MF catalog illuminates M_c in more regions to the east of the trench. On average M_c in the BP-MF catalog is ~ 0.22 smaller than those in the JMA catalog.

The BP-MF catalog fills the gap in the b value map of the JMA catalog, where the seismicity is too sparse for reliable b value estimation (Fig. 2.9a). In the overlapping area

between the two catalogs, the b value distribution is overall consistent but smoother spatially than the JMA counterpart (Fig. 2.9b). There is no systematic variation of b value in the along dip-direction (Fig. 2.9c and d). The median b values in most of the bins are between 0.5 and 1 in both catalogs. We also find that the six large aftershocks ($Mw \geq 6.7$) are spatially correlated with the local high b value area (~ 1.2). This may indicate low local differential stresses due to the stress released by the large aftershocks.

We also investigate the aftershock decay rate following the six large aftershocks. The aftershock decay rate is described by the modified Omori's law (Omori, 1894; Utsu, 1961; Utsu, Ogata, et al., 1995):

$$n(t) = \frac{k}{(c+t)^p} \tag{2.12}$$

where $n(t)$ the aftershock rate at time t after the mainshock, and k, p, c are empirical parameters. The exponent p indicates the aftershock decay rate and is speculated to be controlled by stress changes (Enescu & Ito, 2002). Three of the large aftershocks (a-c) located in the nearshore region are associated with low p values (0.64 - 0.85). In contrast, high p values (0.93 - 1.11) are observed for the other three large aftershocks in the near trench area (Fig. 2.10). This may be due to the higher coseismic slip and hence larger stress drop of large normal-faulting events in the outer-rise compared to the deeper thrust-faulting events (Enescu & Ito, 2002).

2.4 Discussion

The BP-MF method illuminates the trenchward and outer-rise seismicity by doubling the near-trench events in the original JMA catalog. As a result, the BP-MF catalog provides a refined image of the seismicity density, magnitude of completeness, and b and p value distribution, especially in the near trench area. In this section, we discuss the relationship between the aftershock-depleted area with the large-slip area of the mainshock, the advantages, and the further development of the BP-MF method.

2.4.1 Comparison of Aftershock-depleted and Large-slip Areas

Almost a decade after the Tohoku earthquake, there is no consensus on the coseismic slip distribution (Lay, 2018). The uncertainty of the large-slip area remains significant, especially in the very shallow portion of the megathrust (Tajima, Mori, & Kennett, 2013). In addition to direct constraints by teleseismic, geodetic, tsunami observations, the coseismic process of the Tohoku Earthquake can be indirectly inferred by the aftershock distribution. If the stress is significantly released by the mainshock and the residue stress is concentrated at the edge of the large-slip zone, the on-fault thrust aftershocks tend to be temporarily shut off in the large-slip area and instead occur at its fringes (Mendoza & Hartzell, 1988). This assumption is supported by the systematic deficiency of aftershocks in areas of high coseismic slip ($> 15\%$ peak slip) in 101 $Mw > 7.0$ subduction-zone earthquakes (Wetzler, Lay, Brodsky, & Kanamori, 2018). Previous studies confirmed that the interplate events are depleted in the large-slip area of the Tohoku earthquake (Asano et al., 2011; Obana et al., 2013). Kato and Igarashi (2012) delineates the edges of the interplate coseismic rupture based on the sharp seismicity-density contrast in the JMA catalog (pink line in Fig. 2.7b). Here, we attempt to outline a large coseismic-slip zone using our refined BP-MF catalog, which puts further constraints on the rupture extent in the near-trench area (blue line in Fig. 2.7b).

The spatial distribution of interplate events is necessary to properly delineate the large-slip area. Since the BP-MF technique does not provide the focal mechanism and depth information, we are only able to outline the depleted area of aftershocks of all types (Region A and C, the blue line in Fig. 2.7b). It encloses a wide up-dip region (350 km by 150 km) to the south of the Off-Iwate segment and to the north of the Off-Ibaraki segment. The Kato's model is consistent with our aftershock-depleted zone in the Off-Miyagi region (Region A), but shows an additional narrow and southward rupture zone at the 30 - 40 km depth range along the plate interface in the Off-Fukushima region (Region D in Fig. 2.7b). The difference may be caused by two possibilities. First, most aftershocks on the outer edge of our model are likely non-interplate normal and thrust events as suggested by the focal mechanism classification using a waveform similarity approach (Nakamura, Uchida, &

Matsuzawa, 2016). In this case, our model does not include the full extent of the large-slip area in the deep portion and misses the narrow rupture at the Off-Fukushima region in the Kato model. Second, it is equally possible that the Kato model misses offshore interplate events due to the limitation of the JMA catalog and the actual large-slip area is smaller than their model. Regardless of the differences between our model and the Kato model in the deep portion (depth $>$ 20km), our model in the near-trench area provides a new constraint. The Kato model does not provide the boundary of the large-slip area in the shallow portion due to the incomplete JMA catalog at depth $<$ 10 km. In the BP-MF catalog, the sharp density contrast is observed at the 20-km iso-depth contour of the plate boundary, switching from 600 to 100 per $0.2^\circ \times 0.2^\circ$ grid. Such drastic contrast indicates that the very shallow near-trench portion of the Offshore-Fukushima region (Region C, depth $<$ 20m) is a low earthquake-density area. Due to the shut-off effect after the mainshock, this low earthquake-density area is likely a large slip area. Alternatively, Region C might be an aseismic region with low seismic activities both before and after the Tohoku earthquake. The trenchward postseismic movement derived from ocean-floor geodesy seems to suggest the coseismic slip in Region C is at least significantly smaller than that of Region A (Sun & Wang, 2015; Honsho, Kido, Tomita, & Uchida, 2019).

It's worth noting that most of the shallow near-trench areas are depleted of aftershocks except an area between $37^\circ N$ and $38.2^\circ N$ (Region B, black dashed line in Fig. 2.7b). Even though the aftershock density in this area is relatively high, we suspect that most of these aftershocks are non-interplate events because they are spatially disconnected from the deeper interplate seismicity and are consistent with the distribution of previously classified non-interplate events (Asano et al., 2011; Kato & Igarashi, 2012; Obana et al., 2012; Nakamura et al., 2016). In addition, almost no interplate aftershocks occur in the shallow subduction zone (depth $<$ 20km) according to focal mechanisms analysis based on the OBS observations (Obana et al., 2013). These intraplate near-trench aftershocks occurred both within the overriding and subducting plates predominantly controlled by tensional stress normal to the trench axis. We thus consider region B as a portion of the large-slip area which needs to be

validated by further studies.

We then compare our estimated large-slip area with eight published finite fault source models (Wei, Graves, Helmberger, Avouac, & Jiang, 2012; Ide, Baltay, & Beroza, 2011; Shao, Li, Ji, & Maeda, 2011; Ammon, Lay, Kanamori, & Cleveland, 2011; Hayes, 2011, 2017; Yagi & Fukahata, 2011; Yamazaki, Cheung, & Lay, 2018). All the models show consistent high coseismic slip in the Off-Miyagi region (Region A in Figure 2.7b) but differ in whether the mainshock ruptured the near-trench Off-Fukushima region. In the models by Wei et al. (2012), Ide et al. (2011), and Shao et al. (2011), only the shallower portion of Off-Fukushima is ruptured (Fig. 2.11 a-c). In the models by Ammon et al. (2011) and Yamazaki et al. (2018), only the deeper portion of Off-Fukushima is ruptured (Fig. 2.11 d-e). Both portions are ruptured in the model by Ide et al. (2011) (Fig. 2.11f) and neither portions broke in the model by Yagi and Fukahata (2011) and Hayes (2011) (Fig. 2.11 g-h). Overall, our estimated large-slip area is spatially the most consistent with the coseismic slip area ($>$ 15 meters) in Wei et al. (2012) model, indicating significant coseismic slip in the up-dip region of Off-Fukushima. Our estimated large-slip area may put constraints on the post-seismic slip distribution if it is spatially complementary with the high coseismic slip. The afterslip (red line in Fig. 2.11i) of the mainshock Ozawa et al. (2012) is concentrated at the deeper portion of the megathrust, which surrounds our estimated large-slip area. But it is substantially overlapping with Kato's estimation at the Off-Fukushima region. An alternative model shows that the postseismic slip (blue regions in Fig. 2.11i) at the shallow portion of the off-Fukushima region is as large as 0.5 m (Iinuma et al., 2016). However, this result is less certain since the corresponding estimation error is 0.4 m because it heavily relies on only one offshore geodetic site in that region. By tripling the events near the trench, the BP-MF catalog provides a better constraint on the slip at the shallow depth of the megathrust, where the resolution of coseismic and postseismic slip is low.

2.4.2 Advantages of BP-MF

The case study of the aftershock detections following the Tohoku earthquake demonstrates the advantage of the BP-MF method. Our approach is efficient in detecting offshore events, especially those that occurred near the trench, where near-field (<100 km) observations from permanent seismic stations are rarely available. Standard detection methods using regional land-based networks, systematically miss offshore events in the routine catalog. This can be shown in the along-dip distribution of earthquake templates, where the number of JMA templates decreases by 50% from the coast to the trench (Fig. A.12). However, the BP technique, utilizing waveforms recorded at teleseismic distances between 40° and 90°, provides uniformly distributed templates in the study area regardless of their proximity to the landward stations. In fact, the BP-MF approach detects roughly twice more events in the near-trench region and three times more events to the east of the Japan Trench than those of the JMA catalog (Fig. A.16). Such improvement makes the BP-MF method an inexpensive solution for offshore seismicity monitoring compared with deploying ocean-bottom instruments. The recent developments of large-scale seafloor networks such as the S-net project along the Japan trench and the DONET network along the Nankai trough will undoubtedly promote unprecedented progress in the surveillance and research of offshore seismicity and fault structures. However, most of the developing countries close to subductions such as Indonesia, Mexico, and Chile do not have yet permanent seafloor networks. Therefore, it is challenging to detect offshore seismicity with conventional methods relying on close sensors. The BP-MF method can potentially enhance the monitoring of near-trench events in these countries without deploying offshore sensors. In addition, the BP-MF method is easy to implement and computationally efficient. The BP-MF processing is relatively robust with few hyperparameters to tune. Comparing with the autocorrelation-based methods (e.g. Brown et al., 2008) which does not require event templates, the strategy of enriching the template pool in matched-filter detection requires significantly less computing resources. The BP-MF processing can be divided into independent calculations in each sub-region, which is ideal for parallel computing and GPU acceleration.

2.4.3 Further Development of BP-MF

The BP-MF technique can be further improved in several aspects. The BP-MF is capable of capturing moderate to large earthquakes. But it is less effective in detecting small earthquakes ($Mw < 4.5$), which limits the number of high-quality BP templates for the MF detections. The amplitude of the $Mw < 4.5$ events are weak and indistinguishable from that of the ambient noise at teleseismic distances (Fig. 2.2b). In the current BP detection scheme, we only extract two features, the spatial peak ratio $p(t)/A_s(t)$ and the temporal peak ratio $p(t)/A_t(t)$ from the BP energy series in order to separate the seismic events from the background noise. Such task of feature extraction is similar to reducing a video into a two-component 1D time series. The original BP movie, containing the continuous spatial-temporal information of the BP energy distribution, could be represented in more than two components. We can potentially increase the dimensions of the classification input by including additional channels (e.g., BPs using different frequency bands and networks). We can also extract more features from the BP energy using self-defined mathematical functions (e.g., kurtosis, skewness, and maximum absolute deviation), or adopting the automatic methods (e.g., Convolutional Neural Network). With the increased channels and features, it would be interesting to explore whether state-of-art machine learning techniques (e.g., Support Vector Machine, Convolutional Neural Network, Random Forest, and Adaboost) can separate small earthquakes ($Mw < 4.5$) from the background noise.

We can devote more effort to reducing the epicenter and magnitude uncertainties of the BP templates. Currently, the hypocenter of the MF detection is assumed to be identical to the template event and the magnitude is determined according to the amplitude ratio between the MF detections and its corresponding template. The epicenter and magnitude errors of BP templates are around 10 km (after calibration) and 0.6 respectively, larger than the uncertainties listed in the JMA catalog. The new events detected with the BP templates inherit the corresponding location and magnitude errors, which affects the aftershock density, M_c, b and p value analysis. In our future work, we will improve the accuracy of the BP template locations by including arrival-times estimated at the regional network (e.g. Hi-net

and K-net) possibly accounting for the effect of local velocity variations (e.g. Z. Wang & Zhao, 2006). For those events with clear sP depth phases in seismograms, the focal depth of BP templates could be determined (Umino, Hasegawa, & Matsuzawa, 1995; Mishra, Zhao, Umino, & Hasegawa, 2003; Z. Wang & Zhao, 2005; Gamage, Umino, Hasegawa, & Kirby, 2009; D. Zhao, Wang, Umino, & Hasegawa, 2009; D. Zhao, Huang, Umino, Hasegawa, & Kanamori, 2011). Therefore, we can potentially make a distinction between events on the megathrust to those in the outer-rise according to the depth information. In addition, the magnitude estimates are affected by the radiation pattern of the earthquake sources at the teleseismic array, which can be suppressed by averaging the BP estimations using multiple arrays.

We can also aim to improve the quality control of BP-MF detections. Currently, we remove the templates without enough S wave phases detection with the GPD processing and only keep the MF detections above a simple threshold of 9 times MAD. We can potentially combine different existing detection methods (e.g., STA/LTA, polarization analysis, Akaike information criteria picker) through a voting strategy to determine which temples and new detections to keep. This ensembling method can outperform every single classifier if they are independent and have relatively high accuracy. In addition, our approach can benefit from including more datasets. The S-net project constructed by the National Research Institute for Earth Science and Disaster Resilience (NIED) started recording continuous waveforms since 2016. It consists of 150 real-time monitoring observatories connected in line with submarine optical cables from off-Hokkaido to off-Kanto (Mochizuki et al., 2016; Uehira et al., 2018). Performing matched-filter on the S-net recordings with additional BP templates could detect more offshore events that are otherwise buried in the noise. We also expect to obtain more accurate hypocenters, origin times, magnitudes, and event focal mechanisms with the aid of the S-net project.

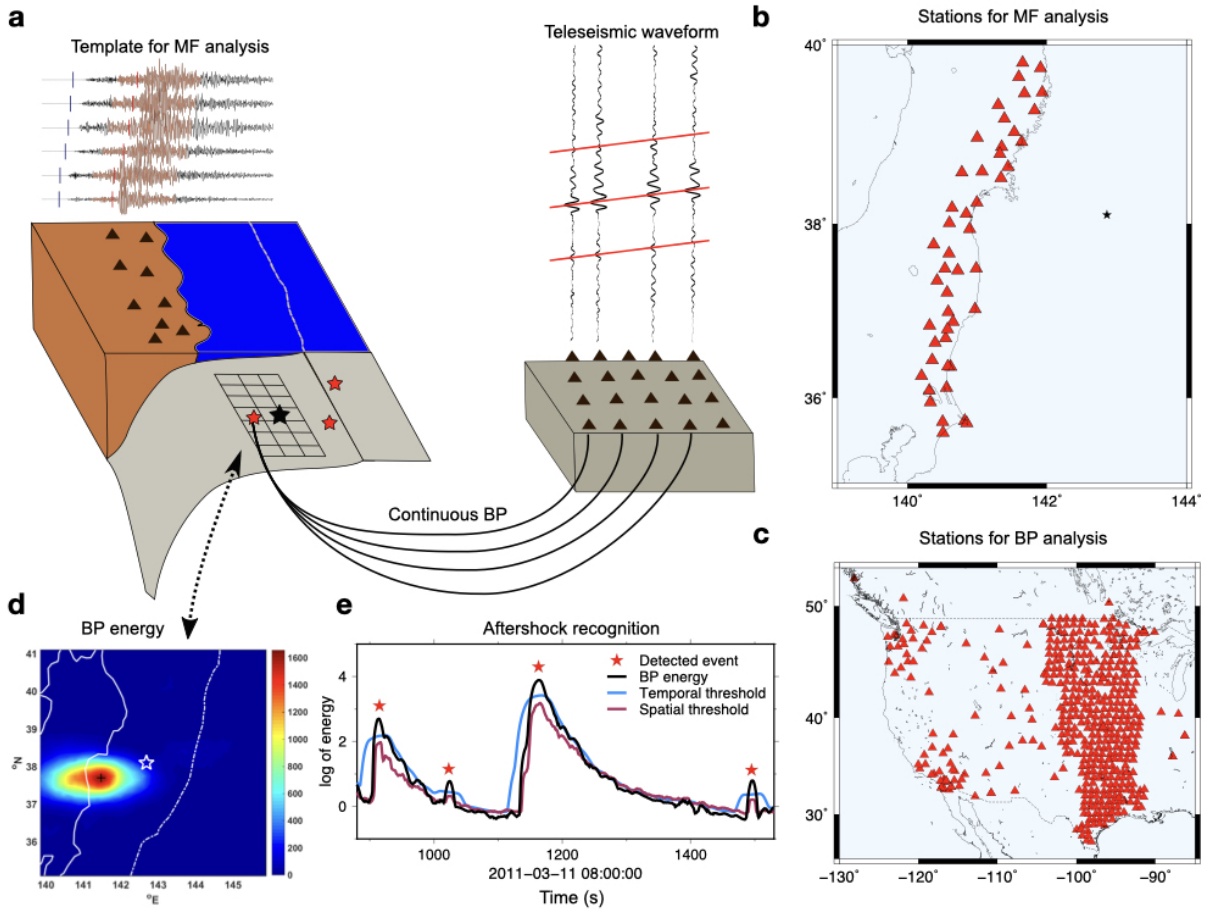

Figure 2.1: (a) Illustration of the back-projection imaging: The rectangular grids indicate the location of testing sources. The true source location (red star) is connected by ray theory to the teleseismic receivers (black triangles). The black curves above the receivers denote the recorded seismograms. In principle, the travel time curve of the true source locations (red lines) brings the seismograms in phase, thus the stack along the moveout reaches the maximum. (b) The locations of 47 selected Hi-net stations for the match-filter analysis (c) The locations of 460 selected stations in North American for BP analysis. (d) The distribution of normalized energy released during a BP-detected event with the epicenter located at the energy peak (blue cross). The white open star is the mainshock epicenter. (e) The BP peak stack energy (black), spatial threshold (red), and temporal threshold (blue) with respect to time from 11 March 2011, 08:00:00 UTC. The red stars are the BP-detected events with the peak energy above the spatial and temporal thresholds.

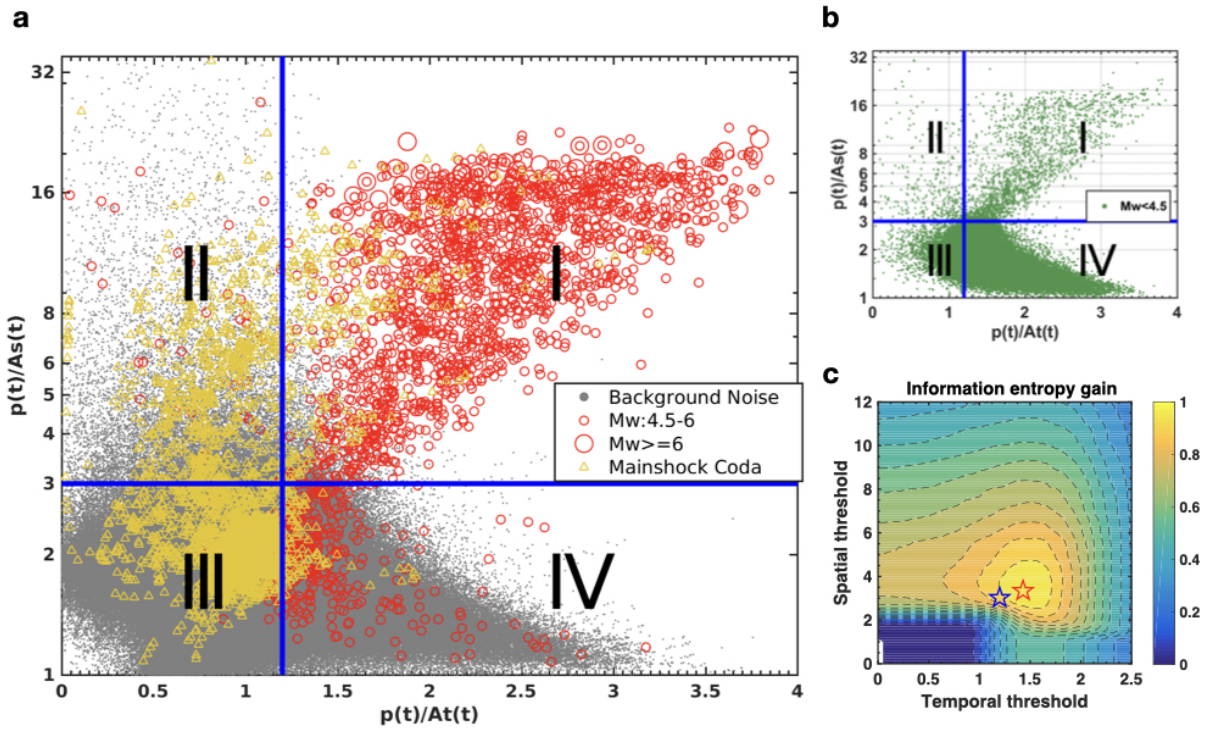

Figure 2.2: (a) The spatial SNR $p(t)/A_s(t)$and temporal SNR $p(t)/A_t(t)$distribution of large earthquakes ($Mw \geq 4.5$) recorded in JMA catalog (open red circles), background noise (gray dots, 1 to 600 days after the mainshock), and coda waves of the mainshock and early aftershocks (yellow triangles, within 24 hours after the mainshock). The blue lines are thresholds to classify earthquakes (upper-right corner) and noises. (b) The spatial SNR $p(t)/A_s(t)$ and temporal SNR $p(t)/A_t(t)$ distribution of small earthquakes ($Mw < 4.5$) recorded in the JMA catalog (green dots). (c) The normalized information entropy gain after the binary classification as a function of the spatial and temporal threshold. The peak gain is represented by a red open star and final selection is represented by a blue open star.

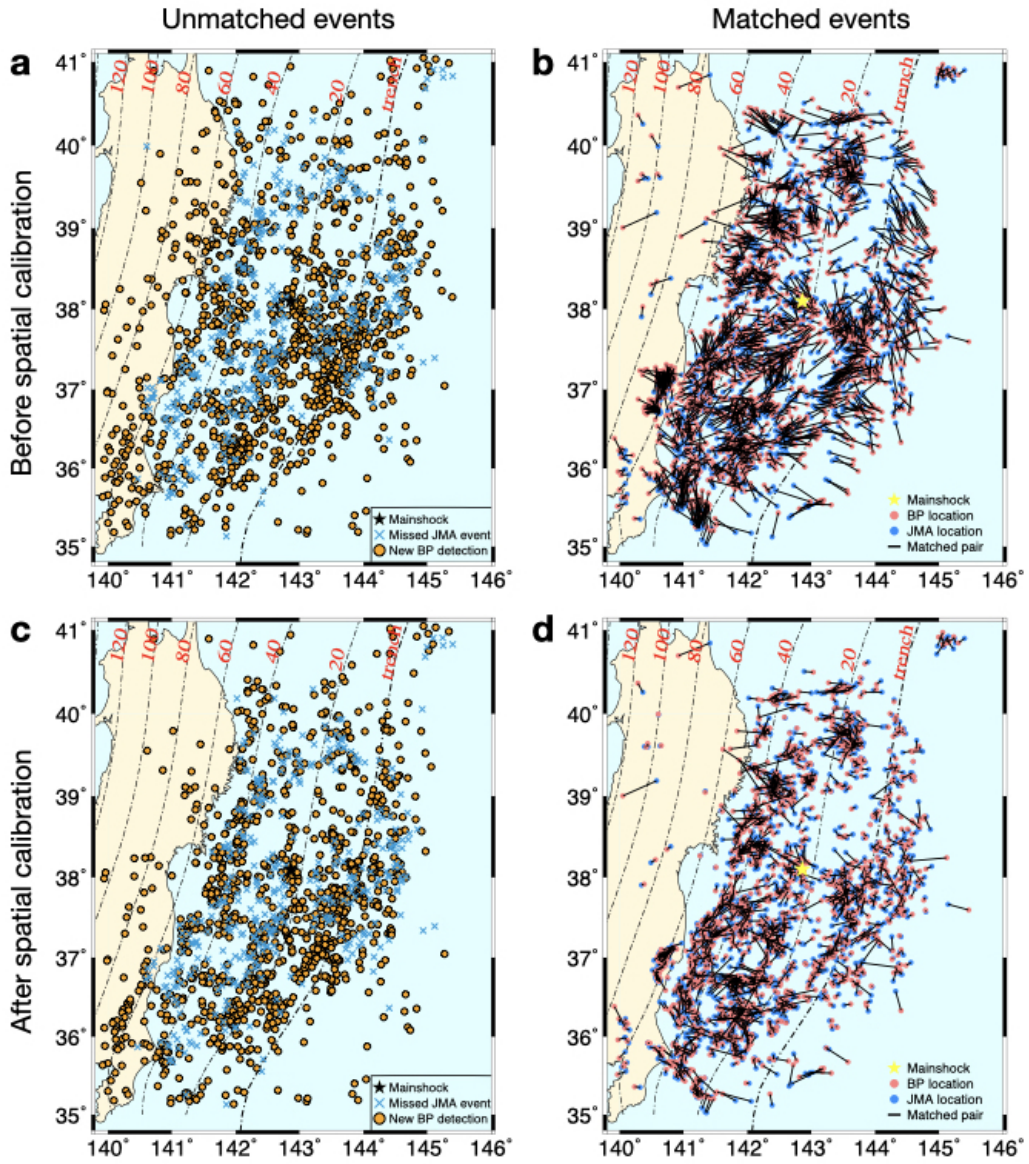

Figure 2.3: Locations of the BP template events before (a-b) and after (c-d) spatial calibration: (a) The locations of 1090 newly detected (orange ball) and 396 missed (blue cross) $Mw \geq 4.5$ events by Backprojection. The black dashed lines represent iso-depth contours of the plate boundary at 20-km intervals (b) The locations of 2066 matched-event pairs, which are recorded by both the BP catalog and JMA catalog. The vectors of event pairs (black arrows) are pointing from JMA locations (blue) to BP locations (red), with an average distance of 21 km. (c) The locations of 1069 newly detected (orange ball) and 380 missed (blue cross) $Mw \geq 4.5$ events by Backprojection. (d) The locations of 2087 matched-event pairs with an average distance of 11 km.

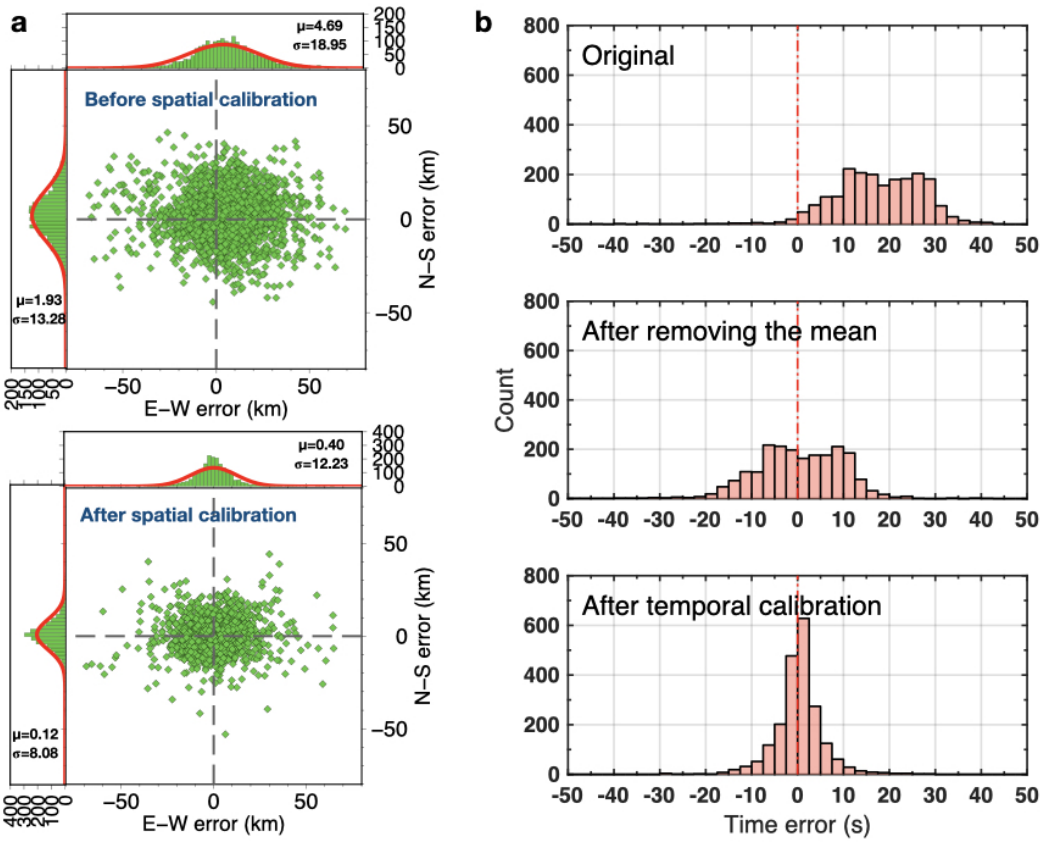

Figure 2.4: (a) Scatter plot of event pair vectors (green dots) before and after calibration. The vectors are pointing from JMA locations to BP locations. Left and Upper panels show the distribution of errors in N-S and E-W directions with the best fitting Gaussian curves (red curves), respectively. (b) Histogram of origin time error of matched event pairs before, after removing the mean, and after temporal calibration.

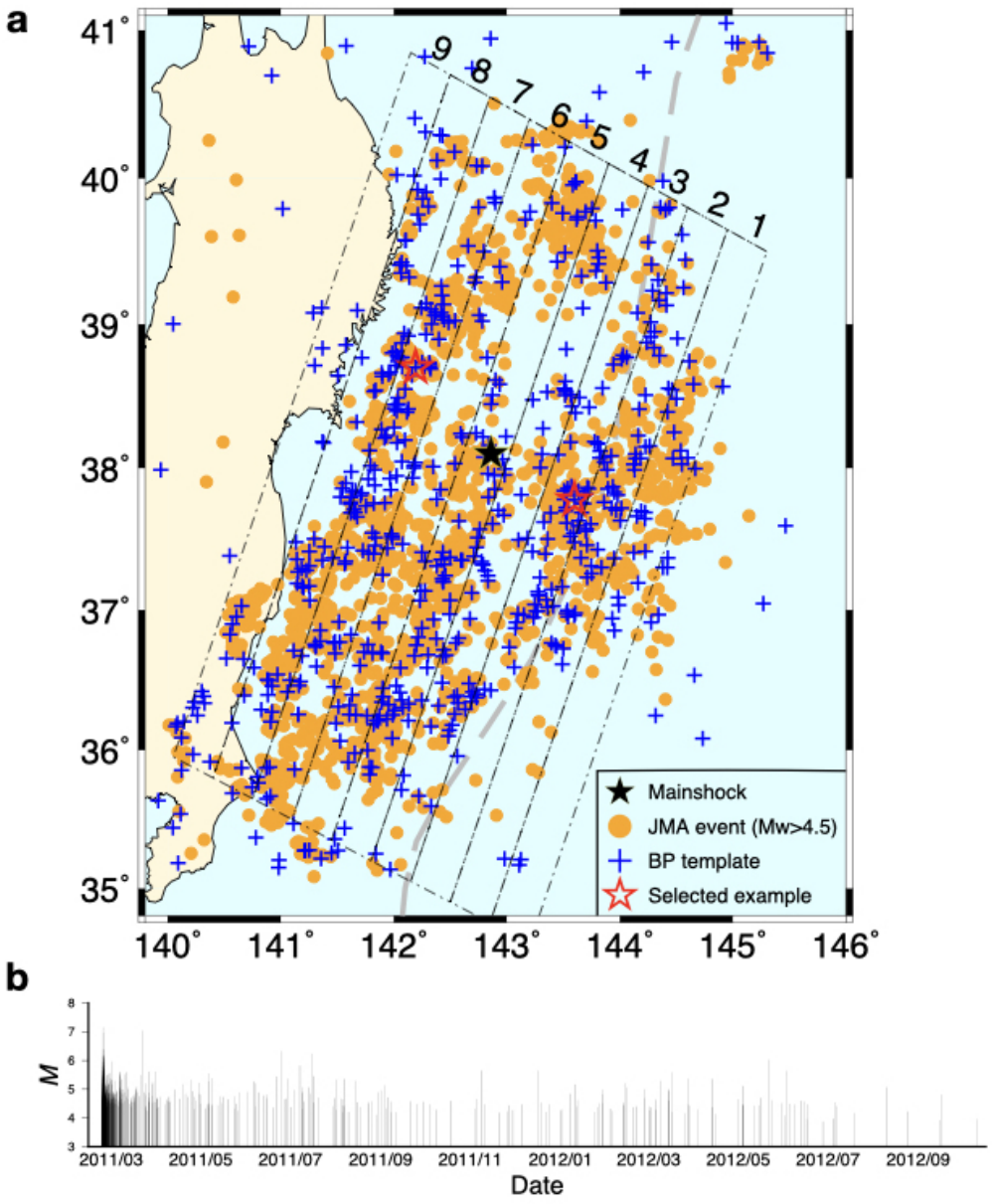

Figure 2.5: (a) The locations of 1,896 large JMA templates ($Mw \geq 4.5$, orange circles) and 605 newly-detected BP templates ($Mw \geq 3.8$, blue crosses). The two red stars are examples of two aftershocks located deeper and shallower than the 2011 Mw 9.0 Tohoku mainshock along the slab interface. The gray dash lines divide the fault plane into 9 segments according to the distance from the trench. (b) Magnitude distribution of BP-detected templates in the period within 600 days after the 2011 Mw 9.0 Tohoku Earthquake.

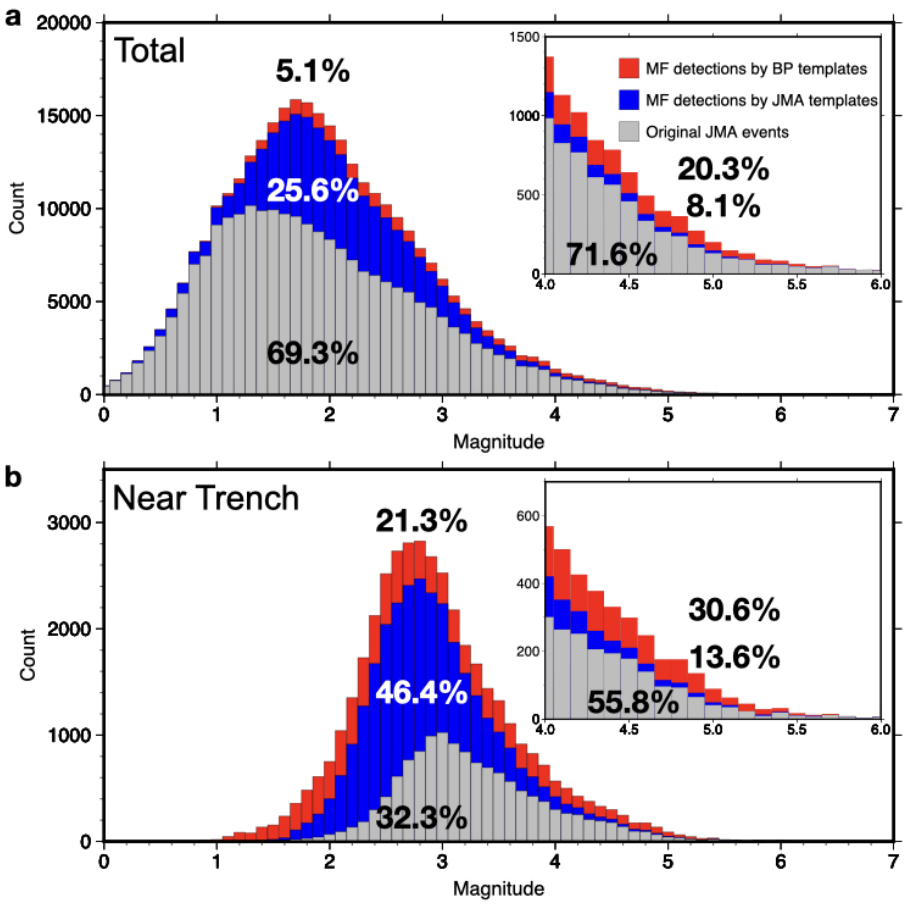

Figure 2.6: (a) Magnitude distribution of original JMA events (gray), new MF detections by JMA templates (blue), and new MF detections by BP templates (red). These three portions of events together constitute the BP-MF catalog. The distribution is zoomed to the range from Mw 4.0 to 6.0 (inset-top-right panel). (b) Magnitude distribution of events near the trench (east to the 20 km iso-depth of the plate interface). The distribution is zoomed to the range from Mw 4.0 to 6.0 (inset-top-right panel).

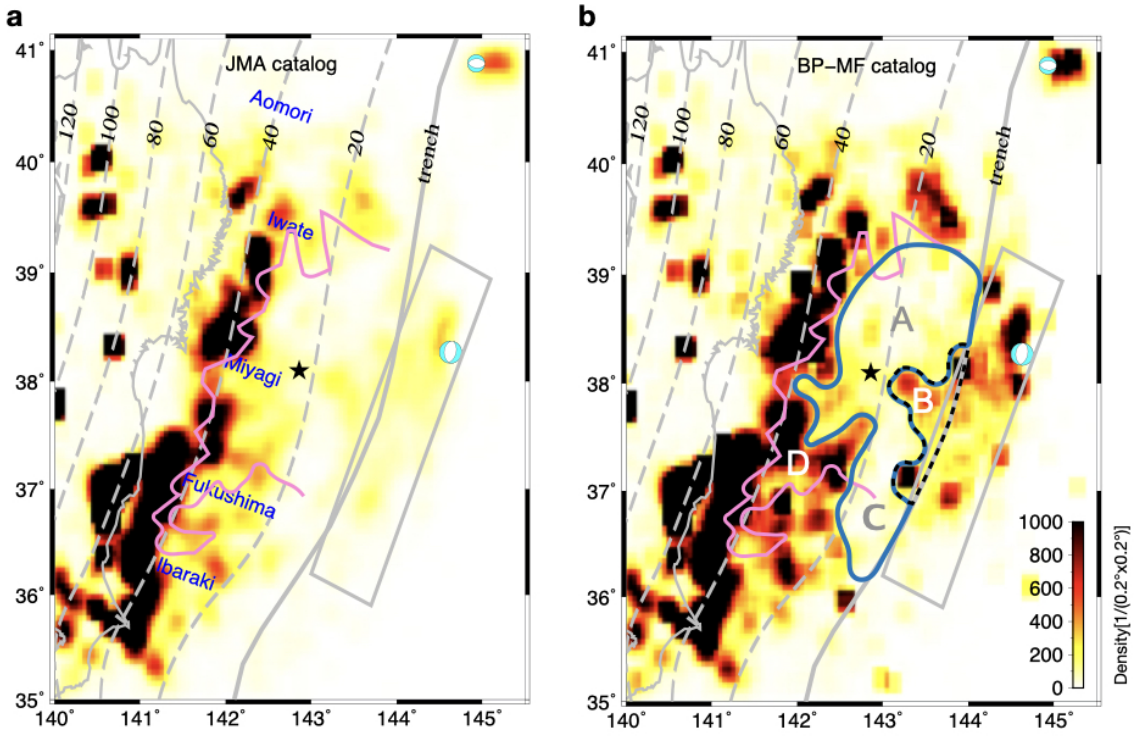

Figure 2.7: Map view of the earthquake density (per 0.2° × 0.2° grid) recorded by the JMA catalog (a) and the BP-MF catalog (b). The black star indicates the epicenter of the Tohoku mainshock. The gray rectangle is the outer trench-slope normal fault zone. The blue line sketches the seismicity gap. The black dash line(Region B) outlines an area of high intraplate-seismicity density. The gray dashed lines represent iso-depth contours of the plate boundary at 20-km intervals. The pink line represents the edges of the interplate coseismic rupture based on the sharp seismicity-density contrast in the JMA catalog (Kato and Igarashi, 2012). Two large aftershocks to the east of the trench are represented with the focal mechanism in cyan (2012/03/14 08:42:8.3 Mw 6.9 and 2011/03/11 06:26:12.6 Mw 7.6).

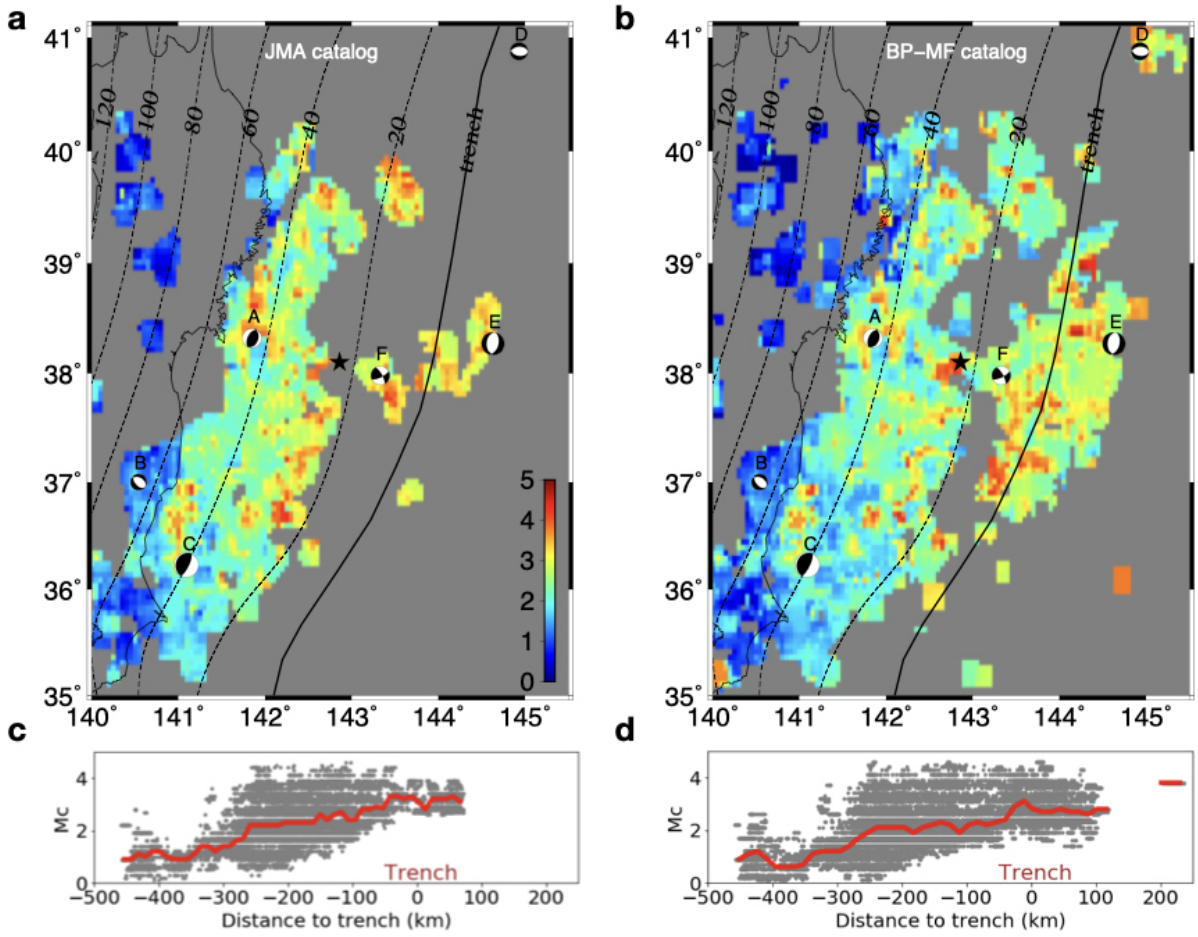

Figure 2.8: Map view of the magnitude of completeness (M_c) distribution based on JMA catalog (a) and new BP-MF catalog (b). The black star indicates the epicenter of the Tohoku mainshock. (A-F) The 6 large aftershocks ($Mw \geq 6.7$) are shown with the focal mechanisms in black (2011/04/07 14:32:50.6 Mw 7.1, 2011/04/11 08:16:19.3 Mw 6.7, 2011/03/11 06:15:58.7 Mw 7.9, 2012/03/14 08:42:8.3 Mw 6.9, 2011/03/11 06:26:12.6 Mw 7.6, and 2011/07/10 00:57:16.3 Mw 7.0). Three of them (A-C) locate nearshore and the other three are near the trench (D-F). (c) and (d) represent the M_c change with distance to the trench (negative: west to the trench, positive: east to the trench), in the JMA and BP-MF catalog respectively. Red dots represent the medians of M_c for respective bins.

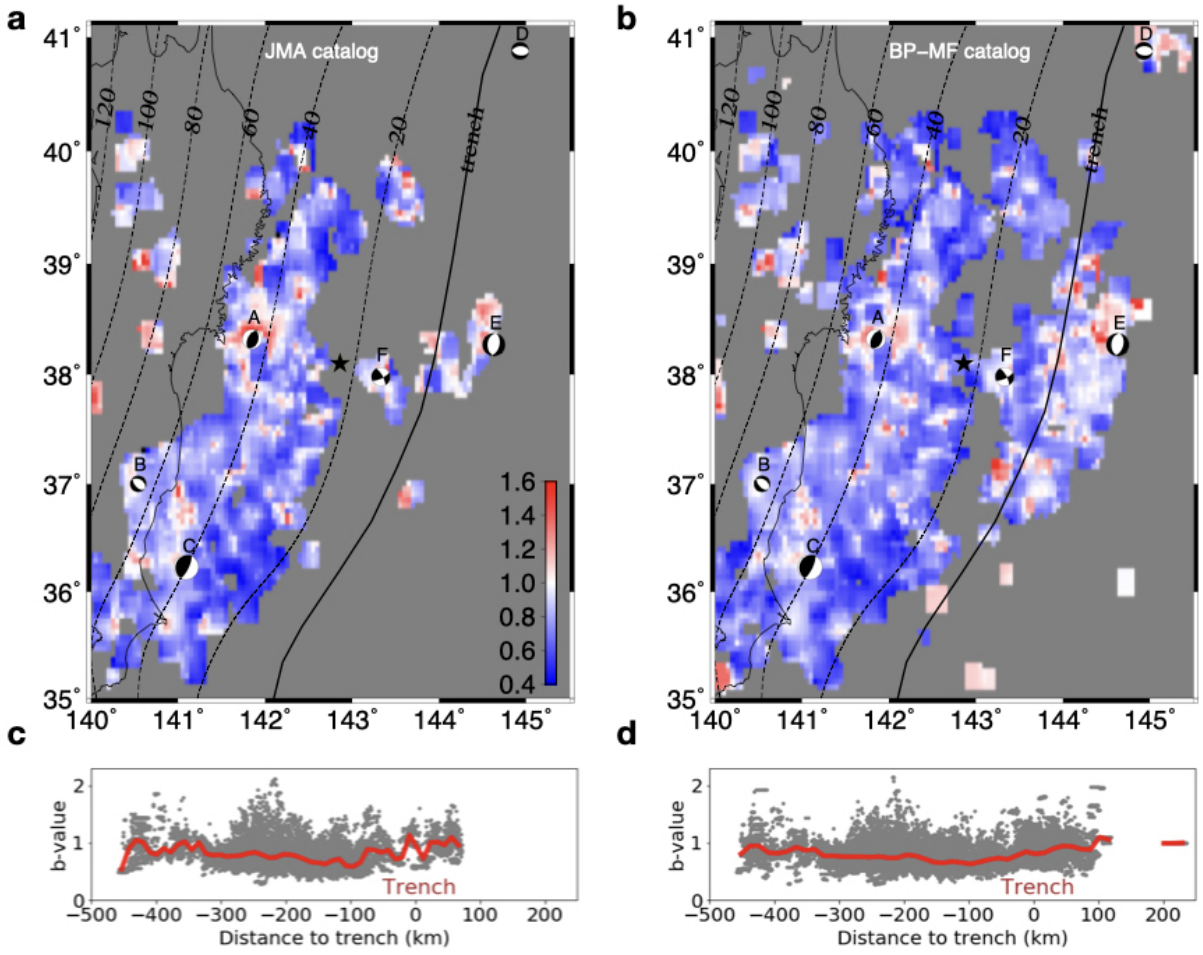

Figure 2.9: Map view of the b value distribution based on the JMA catalog (a) and the new BP-MF catalog (b). The black star indicates the epicenter of the Tohoku mainshock. (A-F) The 6 large aftershocks ($Mw \geq 6.7$) are shown with the focal mechanisms in black (2011/04/07 14:32:50.6 Mw 7.1, 2011/04/11 08:16:19.3 Mw 6.7, 2011/03/11 06:15:58.7 Mw 7.9, 2012/03/14 08:42:8.3 Mw 6.9, 2011/03/11 06:26:12.6 Mw 7.6, and 2011/07/10 00:57:16.3 Mw 7.0). Three of them (A-C) locate nearshore and three near trench (D-F). (c) and (d) represent the b value change with distance to the trench (negative: west to the trench, positive: east to the trench), in the JMA and BP-MF catalog respectively. Red dots represent the medians of b values for respective bins.

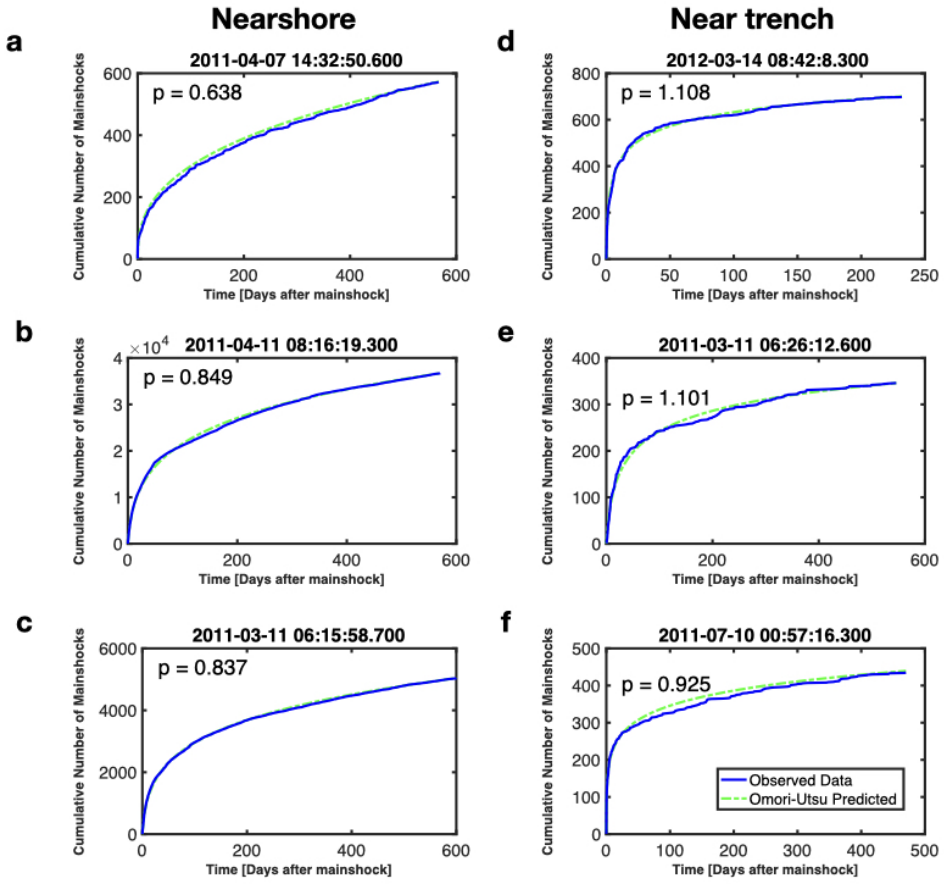

Figure 2.10: The p value of the 6 selected aftershock sequences in Figure 2.8 and 2.9 (2011/04/07 14:32:50.6 Mw 7.1, 2011/04/11 08:16:19.3 Mw 6.7, 2011/03/11 06:15:58.7 Mw 7.9, 2012/03/14 08:42:8.3 Mw 6.9, 2011/03/11 06:26:12.6 Mw 7.6, and 2011/07/10 00:57:16.3 Mw 7.0).

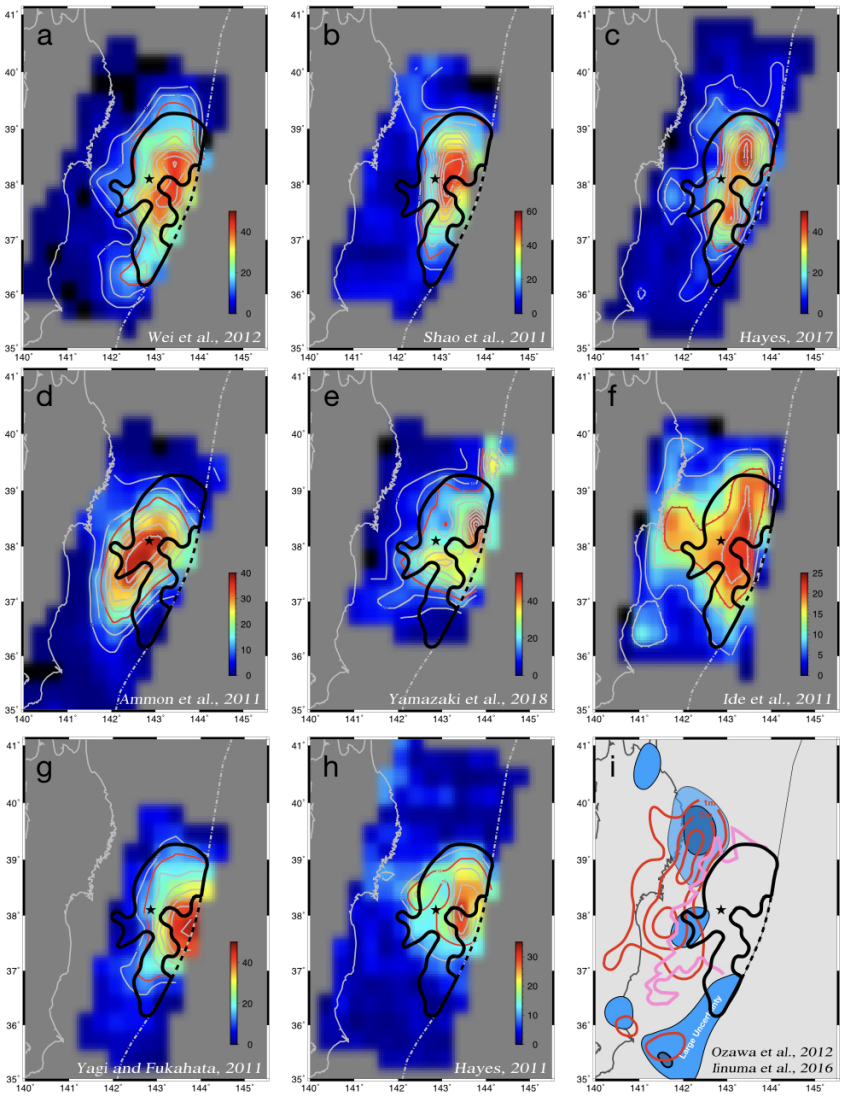

Figure 2.11: (a-h) Comparison of estimated large-slip area (black line) with the large co–seismic slip area in 8 finite fault source models (Wei et al., 2012; Ide et al., 2011; Shao et al., 2011; Ammon et al., 2011; Hayes, 2011, 2017; Yagi & Fukahata, 2011; Yamazaki et al., 2018). The white lines represent iso-slip contours of mainshock slip in 5-m intervals (15-m contour highlighted by red lines). Black stars represent the mainshock epicenter. (i) Comparison of estimated large-slip area (black line) with the large postseismic slip area in two models by Ozawa et al. (2012) (red contour, slip $>$ 1 m) and Iinuma et al. (2016) (blue region, slip $>$ 0.4 m).

CHAPTER 3

EdgePhase: A Deep Learning Model for Multi-station Seismic Phase Picking

Paper under review. **Feng, T.**, Mohanna, S., & Meng, L. EdgePhase: A Deep Learning Model for Multi-station Seismic Phase Picking (Submitted to Geochemistry, Geophysics, Geosystems). I contribute to the experiements and writing of the paper. Lingsen and Saeed provide suggestions on the paper revision.

Note: the supplementary materials are included in Appendix B.

abstract

In this study, we build a multi-station phase-picking model named EdgePhase by integrating an Edge Convolutional module with a state-of-the-art single-station phase-picking model, EQTransformer. The Edge Convolutional module, a variant of Graph Neural Network, exchanges information relevant to seismic phases between neighboring stations. In EdgePhase, seismograms are first encoded into the latent representations, then converted into enhanced representations ("Enhanced" in this context refers to the improved representation ability) by Edge Convolutional module, and finally decoded into the P- and S-phase probabilities. Compared to the standard EQTransformer, EdgePhase increases the precision (fraction of phase identifications that are real) and recall (fraction of phase arrivals that are identified) rate by 5% on our training and test datasets of Southern California earthquakes. To evaluate its performance in regions of different tectonic settings, we applied EdgePhase to detect the early aftershocks following the 2020 M7.0 Samos, Greece earthquake. Compared to a

local earthquake catalog, EdgePhase produced 190% additional detections with an event distribution more conformative to a planar fault interface, suggesting higher fidelity in event locations. This case study indicates that EdgePhase provides a strong regional generalization capability in real-world applications.

3.1 Introduction

The recent expansion of seismic data and computing resources enables flourishing applications of deep learning (DL) in seismology (Valentine & Trampert, 2012; LeCun, Bengio, & Hinton, 2015; Schmidhuber, 2015; Oord et al., 2016; Reynen & Audet, 2017; Perol et al., 2018; Z. Li, Meier, et al., 2018; Meier et al., 2019; Kong, Trugman, et al., 2019; Kong, Inbal, Allen, Lv, & Puder, 2019; Bergen, Johnson, Maarten, & Beroza, 2019; Bianco et al., 2019; Nakano et al., 2019; Mousavi & Beroza, 2019, 2020; Tian et al., 2020; Kuang et al., 2021; Saad et al., 2021; Saad & Chen, 2020; J. D. Smith et al., 2022; T. Wang, Trugman, & Lin, 2021; Civilini et al., 2021; Ross, Yue, Meier, Hauksson, & Heaton, 2019). Many of these studies aim to automatically pick P and S arrivals, especially the signals of microseismicity buried under noises. Dozens of DL models prove to be efficient in detecting phases of local events (epicentral distance smaller than 300 km) (Ross, Meier, Hauksson, & Heaton, 2018; J. Wang et al., 2019; Woollam et al., 2019; W. Zhu & Beroza, 2019; L. Zhu et al., 2019; Ross, Meier, & Hauksson, 2018; Y. Zhou et al., 2019; Dokht, Kao, Visser, & Smith, 2019; Mousavi et al., 2020; Baker, Holt, Pankow, Koper, & Farrell, 2021; Xiao et al., 2021; Liao, Lee, Mu, Chen, & Rau, 2021). Some of these DL models exhibit performance close to manual phase picking with precision and recall rates exceeding 95%. Moreover, they are advantageous in terms of their high sensitivity, computation efficiency, and automatic implementation (Mousavi et al., 2020). Compared to conventional automatic phase picking methods, e.g., short-term average/long-term average and auto-regression with Akaike Information Criterion (R. V. Allen, 1978; Stevenson, 1976; Sleeman & Van Eck, 1999; Akazawa, 2004), detections by DL models show higher accuracy, especially in high-level noise environments (W. Zhu & Beroza, 2019; Mousavi et al., 2020).

Previous models mainly predict phases with seismograms from a single station. However, some phases are difficult to identify at a single station due to a low signal-to-noise ratio (SNR; Xiao et al., 2021). In the manual phase picking process, humans often rely on waveform consistency between multiple stations to decide if an ambiguous case should be marked as a seismic phase. Therefore, a multi-station approach utilizing the waveform consistency between neighboring stations should improve the performance of DL models in phase picking. Several DL models have attempted to utilize waveforms from multiple stations for source characterization tasks (e.g., earthquake detection, phase-picking, earthquake location). These models typically learn the representations of waveforms in a seismic network through feature extraction or mathematical transformations. The representation learning compresses the high-dimensional raw data to low-dimensional vectors, making it easier to discover patterns and anomalies in the downstream tasks. One approach taken by several studies is to jointly convert seismograms of all stations into latent representations (simplified representation of the input data; X. Zhang et al., 2020; Zheng et al., 2020; Yang et al., 2021; Kriegerowski, Petersen, Vasyura-Bathke, & Ohrnberger, 2019). This approach is straightforward and easy to implement with conventional DL modules, such as Convolutional Neural Network (CNN; LeCun et al., 1999) and Recurrent Neural Networks (RNN; Hopfield, 1982). However, models using this approach are trained and tested on the same station configuration and region, which limits their generalization capability to other station networks or tectonic settings.

An alternative approach to utilize multi-station data is to combine the feature embeddings from individual stations with an aggregation module (M. P. van den Ende & Ampuero, 2020; W. Zhu et al., 2021; Xiao et al., 2021). For example, the Siamese Earthquake Transformer (S-EqT) model aggregates the feature embeddings from station pairs with a cross-correlation module (Xiao et al., 2021). S-EqT has strong generalization ability because the aggregation module is designed to accommodate different station configurations. However, since the communication is limited between station pairs, the waveform consistency of more stations is not fully explored.

One solution to establish communications between multiple stations is to utilize the Graph Neural Networks (GNN; Gori et al., 2005; Scarselli et al., 2008; J. Zhou et al., 2020). GNN can freely exchange information between multiple stations, which can handle the irregular structures of graphs with a fixed model architecture (Z. Zhang et al., 2020). It has been successfully implemented in seismic source characterization (M. P. van den Ende & Ampuero, 2020; McBrearty & Beroza, 2022) and phase association problems (McBrearty & Beroza, 2021). In the multi-station phase-picking task, the high dimensionality of the data makes training a GNN model from scratch difficult. One solution to solve this issue is fine-tuning, a process that adopts a pre-trained model for one given task and then tunes or tweaks the model to adapt a second similar task. It accelerates the training process and overcomes the problem of small datasets. For instance, the aforementioned S-EqT model is fine-tuned on EQTransformer, a popular DL model for the single-station phase-picking task (Mousavi et al., 2020). Here, we attempt to integrate the GNN with EQTransformer to improve the detection ability in low SNR conditions. More specifically, we fine-tuned the EQTransformer with a GNN variant named Edge Convolution (Y. Wang et al., 2019), and we named this model EdgePhase. In EdgePhase, the seismograms of each station in the seismic network are first encoded into the latent representations individually by the encoder from EQTransformer. Then the edge convolution module combines the latent representations of the target and a neighboring station into a "message". The "messages" are then used to aggregate the latent representation of the target station into "enhanced representation". Finally, the "enhanced representation" is decoded into the final P- and S-phase probability functions by decoders from EQtransformer.

The rest of this article is organized as follows: In section 2, we introduce the EdgePhase model, which is composed of the EQTransformer and an Edge Convolutional module. In section 3, we describe the Southern California dataset used for fine-tuning. In section 4, we compare the performance of the EdgePhase model with two baseline models. In section 5, we apply the EdgePhase model in a case study of the 2020 M7 Samos earthquake in Greece. In section 6, we discuss the seismotectonics of the Samos Earthquake, advantages and potential

developments of EdgePhase, and its comparison with other multi-station methods.

3.2 Model

3.2.1 EQTransformer

The EdgePhase system is built by integrating the EQTransformer model with the Edge Convolutional module. EQTransformer is one of the most popular DL models for the single-station phase-picking task, and it is also the benchmark model that other models are usually compared with (Mousavi et al., 2020). This state-of-the art model picks P- and S-phases with precision close to manual picks by human analysts. It utilizes plenty of deep-learning modules, like 1D convolutions, bi-directional and uni-directional long-short-term memories (LSTM), residual connections, feed-forward layers, transformer, and self-attentive layers. The transformer it adopts has been widely used in the field of natural language processing and computer vision (Vaswani et al., 2017; Devlin, Chang, Lee, & Toutanova, 2018). To learn more about these DL modules, please refer to Mousavi et al. (2020). The transformer weights the significance of each part of the input data differentially, and highlights the information relevant to the target. In the phase-picking task, the attention of the transformer is focused on durations, P- and S-phases of events in the seismograms. EQTransformer is built with a multiple-task structure that consists of a deep encoder and three independent decoders (Detection/P-phase/S-phase). In our experiments, we keep the encoder and two decoders (P-phase and S-phase) with their pre-trained weights (328.1K parameters in total) based on the STanford Earthquake Dataset (STEAD) (Mousavi, Sheng, Zhu, & Beroza, 2019). The encoder encodes 3-component seismograms (X; dimension: 3*6000) into the latent representations (V; dimension: 32*64) (Fig. 3.1a). The Edge Convolutional module transforms the latent representations into the enhanced representations (V'; dimension: 32*64), by transferring the information (messages) related to P- and S-phases between neighboring stations. This is a light module with only 9.3K trainable parameters, and we will introduce its details in the next section. The two decoders transform latent representations

or enhanced representations into P- and S-phase probabilities respectively.

3.2.2 Edge Convolution

We attempt to improve the phase detection rate in low SNR conditions by incorporating Edge Convolution operations into EQTransformer. Edge Convolution, a variant of graph message passing processing (Gilmer et al., 2017), has achieved great success in classification and segmentation problems (e.g., Y. Wang et al., 2019). Edge Convolution enables exchange of information between neighboring nodes, which enhances the relevant signals shared by adjacent nodes. The graph message passing is composed of three steps: (1) the target node gathers messages or feature embeddings from its neighboring nodes; (2) all messages are aggregated via an aggregate function (e.g., sum, max, mean); (3) the pooled messages are then passed through an update function to renew the embeddings of the target node.

In the phase picking task, we represent the recordings of each earthquake or noise sample with a graph $G = (V, E)$, in which $V = \{v_i | i = 1, ..., N\} \subseteq R^{C*F}$ is the node set with N elements. Each node represents a seismic station, with feature embeddings (channel number is C and embedding length is F) exacted from the seismograms. The $E = \{e_{ij} | v_i, v_j \ are \ connected\}$ is the edge set containing all edges of G. One edge (e_{ij}) represents a path used to send messages (m_{ij}) between two connecting nodes. The message in this task is a vector that indicates the likelihood that a phase arrives. Based on the assumption that the waveforms of neighboring stations are typically more similar than distant ones, we build an edge between two nodes if their represented stations are considered to be close (geographic distance < 100 km). The graph includes a self-loop, meaning each node also points to itself with an edge. An important characteristic of graphs is that they are not defined by the ordering or positioning of the data but only by the relations between data (M. P. van den Ende & Ampuero, 2020). The edge convolution operation first collects all the messages associated with all the edges emanating from each node, and then applies a channel-wise symmetric aggregation operation (here we use maximum operation) to update features for each node. Taking node v_i as an example in Figure 3.1b, we see that it receives

4 messages m_{ij1}, m_{ij2}, m_{ij3}, and m_{ii} from neighbor v_{j1}, v_{j2}, v_{j3}, and itself v_i. Then target node v_i will update its value with $max(m_{ij1}, m_{ij2}, m_{ij3}, m_{ii})$. The number of neighbors may be different in other nodes, but a similar operation can be done for every node in the graph.

In order to incorporate Edge Convolutional module into the existing architecture of Eq-Transformer, the message is designed to be a differentiable function $m_{ij} = h_\theta(v_i, v_j)$, where h_θ is the nonlinear function with a set of learnable parameters θ. We represent h with a neural network composed of 5 one-dimensional convolutional (conv1D) blocks (Fig. 3.1c). Each block has a convolutional layer (kernel size = 3, stride = 1, padding = 1), a Gaussian Error Linear Units activation layer (GELU; Hendrycks & Gimpel, 2016), and a dropout layer (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014) for regularization (dropout rate = 0.1). We concatenate two latent representations along the channel dimension as the input (R^{2C*F}) of the neural network h_θ. Since the input dimension of the EQTransformer decoder is $C * F$, we set the output channel number of the 5 conv1D blocks as $[2C, 2C, C, C, C]$, respectively. Compared to the original version of the Edge Convolution module used in the DGCNN model (Y. Wang et al., 2019), we made the following changes to adapt to the phase-picking task. In the EdgePhase, a CNN instead of a Multilayer perceptron (MLP) is used to construct messages. This is because the station embedding is a large-size two-dimensional vector, using a CNN instead of a MLP can significantly reduce the number of trainable parameters and skip the flattening operation. Traditional Edge Convolution defines neighbors based on the relative distances with k-nearest neighbors (k-NN). However, only absolute distances can guarantee the coherency of waveforms, because the waveform coherence decays dramatically in an exponential fashion with interstation distance and frequency (Luco & Wong, 1986; Zerva & Zervas, 2002; Langston, 2014). Therefore, unlike K-NN, EdgePhase defines neighbors with the fixed threshold of the geographic distance. The DGCNN model is stacked with many layers of Edge Convolutional layers due to the large number of nodes in graphs. Considering that the number of stations in a network is usually around dozens or hundreds, EdgePhase only uses one Edge Convolutional layer.

3.3 Dataset

Our training and testing dataset (SCSN2021) is built with the continuous waveforms and earthquake catalog of the year 2021 recorded by Southern California Seismic Network (SCSN; Hutton, Woessner, & Hauksson, 2010). Here, each sample event (earthquake or noise) is treated as an edgeless graph. Since the number and sequence of stations might vary with different samples, we store the node information of each sample in the dictionary data structure. The edge information is not stored in the dataset, but constructed during the training, validation, and testing process, because edges could change when performing data augmentation (e.g., reordering, resampling stations; W. Zhu, Mousavi, & Berosa, 2020). In addition, the edgeless structure allows for the exploration of different methods and threshold values to establish edges in further research.

In our dataset, the 12,718 earthquake samples with more than 5 labeled phases are selected from 22,619 events that occurred within the year of 2021 in the SCSN catalog. The 15,813 noise samples are randomly selected in the gaps between 5 min before and 40 min after events in the SCSN catalog. Compared to datasets suitable for single-station models (e.g., STEAD with 1.3 million samples), our dataset suffers from the problem of a small training dataset (28.5k samples), because each sample in our dataset is a graph containing all the seismograms of one event from a seismic network, rather than one seismogram at a single station. A small set of training samples makes training deep learning models from scratch difficult, so we decided to fine-tune the EQTransformer model with the SCSN2021 dataset. The pre-processing steps of seismograms are the same as STEAD, including detrending, bandpass filtering the seismograms to 1-45 Hz, and resampling to 100 Hz.

Previous single-station models achieved robust performance in high SNR conditions, so we intentionally make our training dataset noisy to test the performance of multi-station models. STEAD only keeps the seismograms with both P- and S-phase labels, but our dataset includes seismograms with a P-phase label or S-phase label. The seismograms with only one phase label tend to be of low SNR as the other phase is not recognized by the SCSN phase picking system. Compared with samples in STEAD, the waveforms of samples

in the SCSN2021 are more noisy. The median values of the signal-to-noise ratio of the STEAD test set and SCSN2021 test set are 25.0 and 16.3 decibels, respectively (Fig. B.1). The SNR is calculated in the same way as in the STEAD dataset (Mousavi, Sheng, et al., 2019). For the prediction targets (labels), we convert the phase arrival times into a temporal distribution with a triangular shape, just like EQTransformer. If one station has multiple types of channels (e.g., Broadband, High Gain), we consider them as multiple stations with the same geographical location. If some components or segments of waveforms are missed, we pad the gaps with zeros. Finally, we split the datasets into training, development, and test sets, which account for 80%, 10%, and 10% of the total samples respectively.

3.4 Performance

To understand the contribution of the Edge Convolutional module to the overall system of EdgePhase, we compared the EQTransformer model with Edge Convolutional module (EdgePhase) and the standard EQTransformer without Edge Convolution (Baseline-A). For a fair comparison, EdgePhase and Baseline-A are both fine-tuned with the SCSN2021 training set. To study the effect of fine-tuning, we also added Baseline-B to the comparison, which is the standard EQTransformer trained with the STEAD dataset without the fine-tuning. Due to the consistency of region and station configuration between training and test set, the EdgePhase and Baseline-A are expected to achieve better performance than Baseline-B on the SCSN2021 test set. In the fine-tuning process of EdgePhase and Baseline-A, we adopt the binary cross-entropy and Adaptive Moment Estimation (Adam) as our loss function and optimizer, respectively. The training loss was checked every 5,000 training steps. The learning rate for the optimizer started from 10e-4 and decreased by 90% every time the training loss stopped decreasing in the past 10 checks. We update the weights of the whole network in a batch size of 16. Each sample may have a different number of stations, so we make the data generator sampling up to 32 stations per training sample. The EdgePhase and Baseline-A models start to converge at 52k and 40k training steps, respectively.

We define the peak in the probability function above a certain detection threshold as

a phase pick, with the phase arrival time located at the center of the peak. The pick is counted as a true positive (TP) if the time residual of matched labeled phases is smaller than 0.5 sec. If the pick can't match any labeled phases, it is deemed a false positive (FP). If a labeled phase does not match with any picks, then it is considered false negative (FN). Based on TP, FP, and FN values we can calculate three basic performance metrics of machine learning models: precision ($\frac{TP}{TP+FP}$), recall ($\frac{TP}{TP+FN}$), and F1 score ($\frac{2*precision*recall}{precision+recall}$). For the true positives, we evaluate the accuracy of picking time by calculating the mean, standard deviation (Std), and mean absolute error (MAE) of their residuals with the ground truth (Table 1).

To determine the model detection thresholds, we use the F1 score, which is the weighted average of precision and recall. We set the thresholds of the P-phase for EdgePhase, Baseline-A, and Baseline-B at 0.39, 0.64, and 0.05, when they achieve maximum F1 scores at 0.88, 0.84, and 0.66, respectively (Fig. 3.2; Table B.1). Similarly, we set the thresholds of the S-phase for EdgePhase, Baseline-A, and Baseline-B at 0.28, 0.4, and 0.03, when they achieve maximum F1 scores in the test set at 0.86, 0.81, and 0.62, respectively. The Baseline-B model achieves the F1 score of 0.99 and 0.98 for P and S-phases on the STEAD test set, which are higher than on the SCSN2021 test set (0.66 and 0.62). This is due to consistency of region and station distribution between training and test set, and the earthquake waveforms in the SCSN2021 test set being more noisy than STEAD, as we laid out in the previous section. In addition, the Std, and MAE on phase picking time residuals by EdgePhase increase by about 0.01 second, compared to two other models. This is also reasonable as the additional detections made by EdgePhase tend to have blurry phase arrivals, which would contribute to larger picking time residuals.

In previous analysis, EdgePhase outperforms two baseline models with higher precision and recall rate. To study the effect of SNR on the phase picking performance, we compare the number of TP, FN, and FP generated by the three models at low (0 db), medium (10 db), and high (20 db) SNR bands (width: $\pm$1 db). The P-phase TP (matched phases) numbers of EdgePhase, Baseline-A, and Baseline-B models are 183, 109, 5 at low SNR; 1048, 995, 468

at middle SNR; and 988, 959, 915 at high SNR, respectively (Fig. B.2). The P-phase FN (missed phases) numbers of EdgePhase, Baseline-A, and Baseline-B models are 443, 517, 621 at low SNR; 113, 166, 693 at middle SNR; and 39, 69, 112 in high SNR, respectively. The P-phase FP (false alarms) numbers of EdgePhase, Baseline-A, and Baseline-B models are 36, 22, 14 at low SNR; 39, 36, 76 in middle SNR; and 17, 22, 155 in high SNR, respectively. Compared with two baseline models, EdgePhase detects more and misses fewer phases at all three SNR bands. EdgePhase also makes fewer false alarms at high SNR conditions, but more false alarms at low SNR conditions. Compared with two baseline models, EdgePhase is more sensitive to phase arrivals at low SNR conditions by making significantly more phase picks, so it is reasonable to make more false alarms. It demonstrates EdgePhase's superior detection capability at the environment of 0-20 db SNR. Based on Figure B.3, a similar conclusion can also be drawn for S-phases. To sum up, EdgePhase outperforms two baseline models, especially at low to medium noise levels (SNR between 0-20 db).

Next, in Figure 3.3, we show the detection probability functions of the 3 models with an example of an M 1.43 earthquake in the test dataset that occured in 25th April, 2021 (202110425M1). The EdgePhase model outperforms two baseline models by detecting all labeled P- (11) and S-phases (9). The Baseline-A detects 5 P- and 4 S-phases, while Baseline-B detects 4 P- and 7 S-phases with 6 false detections. Some of the missed detections in Baseline-A show a peak in the probability functions, but smaller than the detection threshold. Baseline-B sometimes confuse P-phases with S-phases and there are no corresponding peaks to the missed detections in its probability functions. In general, EdgePhase outperforms Baseline-A, and Baseline-A outperforms Baseline-B. To study the efficiency of the Edge Convolutional module in exchanging information relevant to P- and S-phases, we visualize the input (latent representations) and output (enhanced representations) of this module. We visualize the feature map of the 3 neighboring stations for Event 202110425M1 (Fig. B.2), which are sorted in ascending order based on the mean values of the features. For the enhanced representations, the amplitudes of features are either increased or decreased in the region between the P-phases and S-phases. But there are clear changes of amplitudes at the

arrivals of the P-phases and S-phases in the feature maps, compared with the seismograms before and after the seismic phases. This phenomena is especially clear for the station SLB. The latent representations show fluctuations before the P-phases but no clear change of pattern before and after P-phase arrivals in many channels. The boundaries between phases and background are clearer in the enhanced representations than in the latent representations, which demonstrates that the communication between stations efficiently suppresses the noise before P-phases and after S-phases in the feature spaces.

To conclude, fine-tuning is essential to EQTransformer when applying to new datasets or applications. The Edge Convolutional module, which can convert a single-station model into a multi-station model, makes additional improvements during the fine-tuning process.

3.5 Case Study of the M7.0 Samos Earthquake

To evaluate the performance of EdgePhase in a real-world application, we examined the aftershock region of the M7.0 Samos (Néon Karlovasion) earthquake which occurred 14 km northeast of the Greek island of Samos on 30th October, 2020. It was the deadliest earthquake in 2020, causing more than 100 deaths during the mainshock and tsunami. This devastating earthquake happened due to the north-south extensional stress during slab rollback, where the African plate subducted beneath the Aegean and Anatolian Microplates (Fig. B.5; J. Meng et al., 2021). Preliminary results indicate that the mainshock occurred on the Offshore North Samos (Kaystrios or Samos Basin) Fault, which has not hosted a large earthquake since the mid-eighteenth century (Papadimitriou et al., 2020; Foumelis et al., 2021). The north-dipping candidate plane from the focal mechanism is more likely to be responsible for the 2020 quake, as suggested by the uplift at the western part of Samos Island (footwall), and the over 10cm of subsidence at the northernmost edge of the central part of the island (Papadimitriou et al., 2020; Evelpidou, Karkani, & Kampolis, 2021). However, the distribution of aftershocks in regional catalogs (NKUA, GI-NOA, and AFAD) does not show any preference for the nodal plane (Foumelis et al., 2021; Papadimitriou et al., 2020).

Early aftershocks immediately following the mainshock are usually relevant to the mainshock rupture process (Mendoza & Hartzell, 1988; Wetzler et al., 2018), so we study the distribution of aftershocks in the first month after mainshock. The phase picking model was applied to continuous data from 72 stations around the Aegean Sea (Fig. B.5). These waveforms are not included in the STEAD and SCEC2020 datasets, so they are suitable for testing the generalization ability of EdgePhase. The pre-processing of the raw waveforms included bandpass filtering (1- 45 Hz) and detrending. We then divide the continuous waveforms into the 60-sec sliding windows with a step of 18 sec and feed them into the EdgePhase phase picker. The computational time required to process 24 hours of continuous data from 72 three-component seismometers is around 20 minutes on an economical GeForce-GTX-1070-8GB GPU. As a result of using low detection thresholds (0.18 for P-phases and 0.15 for S-phases), we were able to detect 979,218 P-phases and 1,000,341 S-phases with relatively low SNRs. Then, we applied the rapid earthquake association and location method (REAL), least-squares location method (VELEST), and double-difference earthquake location algorithm (HypoDD), in that order, to perform phase association and event locations (M. Zhang, Ellsworth, & Beroza, 2019; Kissling, Ellsworth, Eberhart-Phillips, & Kradolfer, 1994; Waldhauser & Ellsworth, 2000). The magnitude estimation method and parameter setting of earthquake location methods are included in the supplementary materials (Text. B.1 and Text. B.2).

In total, we located 1,222 events with VELEST and relocated 687 events with HypoDD. Compared to the local catalog (421 events, Fig. B.6) from the NKUA, this study (Fig. B.7) increases the detections by 190% with a distribution more conformative to a linear fault trace (Fig. 3.4a). It demonstrates that the detection ability of the DL phase picking model is on par with that of the manually reviewed catalog. We consider a DL detection to match with a known event in the NKUA catalog if they occurred within 3 sec and 30 km of each other. Under such matching criteria, we matched approximately 53% of the events listed in the NKUA catalog, with the event locations of this study generally being somewhat north of the NKUA catalog (Fig. B.8). Based on matched event pairs, we calibrate the estimated

local magnitude of our detected events using linear regression (Fig. B.9). We then check the 997 new detections found by this study, and find that earthquake signals are quite clear in most traces, especially at close stations (four examples shown in Fig. B.10 - B.13).

3.6 Discussion

3.6.1 Seismotectonics of the Samos Earthquake

Similar to the preliminary report by Papadimitriou et al. (2020), we group aftershocks of the M7 Samos Earthquake into four clusters along the fault strike direction (Fig. 3.4a) with the DBSCAN algorithm (distance $< 4.3\ km$; at least 43 samples in a group). Events in Clusters 2 and 3 started to appear in an earthquake swarm from June 20th to June 30th, 2009. During the period of the swarm, one M_w 5.1 event and more than 80 $M_L > 1.5$ events occurred around Samos Island (Tan et al., 2014). The clustering of events indicates the concentration of stress in these locations as well as the presence of frictional heterogeneities on the fault plane. Furthermore, the slow unlocking process (precursor) of the Offshore North Samos Fault might have existed several years before the 2020 M7 event similar to those observed prior to several megathrust earthquakes (e.g. Huang & Meng, 2018).

The cross-section along the A-A' direction shows that the events (Clusters 1 and 2) west of the mainshock are shallower than the events east of the mainshock (Clusters 3 and 4; 3.4b). Furthermore, the events in the west segments are restricted to a depth of 13-17 km. By examining the cross-section along the B-B' direction, we find that the south-dipping plane of the mainshock by AFAD (dashed lines in 3.4c) seems to fit the lower bound of clusters 1 and 4 well. Cluster 2 and 3 events are close to the mainshock hyocenter, with shallower depths than the mainshock(3.4d).

The results of this case study demonstrate that our fine-tuned DL phase picking method incorporated with traditional earthquake location methods, could automatically monitor earthquakes. Meanwhile, the performance of the multi-station phase-picking model is proven to be comparable to human efforts.

3.6.2 Advantages and Further Development of EdgePhase

In early phase-picking studies, the seismograms/spectrograms were treated as images by CNNs (Ross, Meier, Hauksson, & Heaton, 2018; W. Zhu & Beroza, 2019; L. Zhu et al., 2019; Ross, Meier, & Hauksson, 2018; J. Wang et al., 2019; Woollam et al., 2019), or as audio data by incorporating RNNs (Mousavi et al., 2020; Mousavi, Zhu, et al., 2019; Y. Zhou et al., 2019). In this study, the seismograms and geographical locations of multiple stations are treated as a graph through the GNN framework. The communication between stations through the Edge Convolutional module brings the phase-picking model into a new stage, where the phase detection not only depends on seismograms from a single station, but also waveform consistency in a seismic network. Compared with the single-station model (Baseline-A), the precision and recall of the multi-station model (EdgePhase) increases by 5%, and the detection accuracy in low SNR conditions is improved.

Although the EdgePhase model is fine-tuned with data in Southern California (SCSN2021), it showed strong generalization ability in the case study of the 2020 Samos earthquake sequence in Greece. There are two reasons for the strong generalization ability. First, the encoders and decoders of EQTransformer are pre-trained with a global dataset (STEAD). Second, the information exchange mechanism in EdgePhase only requires that the neighboring stations are within a certain distance range, therefore its application is not restricted to a specific station configuration as is the case of the end-to-end DL models (X. Zhang et al., 2020; Zheng et al., 2020; Yang et al., 2021; Kriegerowski et al., 2019).

In addition to strong generalization ability, the Edge Convolutional module can be easily incorporated into any pre-trained phase-picking models in the encoder-decoder architecture. For example, one potential improvement to PhaseNet (W. Zhu & Beroza, 2019) is to add the Edge Convolutional module right before the first devolutional layer. There is no need to change its U-net architecture and residual connections. We can even pick and choose encoders and decoders from different models, with the Edge Convolutional module connecting them. Reusing the pre-developed encoders, exchange modules, and decoders can accelerate the development of new models through transferring learning or fine-tuning.

There is still potential for improvement of EdgePhase in further work. Possible developments include replacing the Edge Convolutional module with a more general message passing network (Gilmer et al., 2017), trying different functions to build and aggregate messages; fine-tuning with a global dataset (large size), and exploring different algorithms such as k-NN to define neighbors of stations. In the current EdgePhase model, two stations are either classified as neighbor or non-neighbors based on their relative distance. This rough binary classification ignores the differences between close and far neighbors, and an alternative method is to add a distance feature to the edges connecting the stations. The distance feature can potentially serve as weighting coefficients for the waveform embedding of two connected nodes, although the adding edge features may require more computing resources, which scales with the square of the number of nodes.

The EdgePhase is very suitable for solving node-level tasks, in which the model predicts properties for each node in a graph. Besides the phase-picking task, EdgePhase can also be applied to the seismograms denoising problem. The phase-picking and seismogram denoising tasks transform the raw seismograms into phase probability functions and denoised seismograms for each station, respectively. Like the phase-picking task, separating noise from signals also benefits from communication between stations. Based on the assumption that a similar pattern of the earthquake signal is shared at neighboring stations, the Edge Convolutional module can be modified to transfer the feature embeddings relevant to the earthquake signals. One can adopt encoders, decoders, and loss functions in the previous single-station seismograms denoising models (W. Zhu, Mousavi, & Beroza, 2019; Novoselov, Balazs, & Bokelmann, 2020) and fine-tune them as is done in this study.

3.6.3 Comparison to Other Multi-station DL Models

In this section, we compare the EdgePhase model with three other multi-station DL models that combine the feature embeddings from individual stations with an aggregation module (M. P. van den Ende & Ampuero, 2020; W. Zhu et al., 2021; Xiao et al., 2021). M. P. van den Ende and Ampuero (2020) aggregate the feature embeddings of all stations with a node-wise

maximum function (we will use EA2020 as the abbreviation for their model). For each element of the latent representation, the node-wise maximum function selects the largest value among embeddings of all stations, therefore the aggregation result is not affected by the number or the order of stations. The EQNet by (W. Zhu et al., 2021) adopted a different approach to aggregate feature embeddings related to P- and S-phases through a shift-and-stack operation similar to the backprojection process. The shift operation is performed on P- and S-phase embeddings, which can be regarded as a preprocessing step. The stacking operation is to node-wise sum up the shifted embeddings, which works as an aggregation function. The S-EqT model (Xiao et al., 2021) retrieves phase picks in target seismograms (low SNR) by referencing confident phase picks in the template seismograms (usually has high SNR). Specifically, it first extracts feature embeddings from seismograms with the encoder of EQTransformer. Then it performs enhancing and cross-correlation operations on the feature embeddings of the template and target seismograms. The cross-correlation operation works as an aggregation module to gather information about similarity between template and target stations.

These models attempt to aggregate station embeddings with the knowledge of graphs and earthquake physics (e.g., phase velocity of the P and S arrivals, waveform similarity), which correspond to GNN modules and physics-based modules in models, respectively. An elementary GNN module transforms nodes, edges, and the global context of graphs, while a physics-based module is modified from some widely used multi-station geophysical technologies or models (e.g., waveform cross-correlation, backprojection, and focal mechanism inversion). Compared to the GNN modules, physics-based modules benefit from additional physical constraints but are only applicable to some specific tasks. The S-EqT and EQNet models include physics-based modules, while EdgePhase and EA2020 only utilize GNN modules.

In addition to the presence of physics-based modules, these four models show differences in methods to aggregate feature embeddings. The EA2020 and EQNet models aggregate feature embeddings of all nodes with a function that satisfies the law of permutation invari-

ant, such as maximum (EA2020) and summation (EQNet). Permutation invariant means that the values of these functions do not change when exchanging the order of the components/elements of input. The aggregation modules of EQNet and EA2020 treat every station equally and transform the feature embeddings of the whole seismic network into a matrix of a fixed size, which does not change proportionally with the number of stations. This design is suitable for solving graph-level tasks, which predict properties for the entire graph. Here, we take the source characteristic task as an example: the amount of parameters involved in predicting magnitude and location only depends on the source (graph) itself, not the number of stations (nodes). On the other hand, S-EqT and EdgePhase use the message passing mechanism, which applies the aggregation module N times for a N-station network. In each aggregation operation, there is a non-repeating primary station and several reference stations, and the fixed-size output represents the feature embeddings for the primary station. By scaling the number of aggregation operations with the number of stations, the size of all stations' embeddings is proportional to the number of stations, therefore message passing networks are suitable for solving node-level tasks.

Next, we will compare S-EqT and EdgePhase in more detail, because they are both built with a message passing mechanism and for the same task (multi-station phase-picking). In the S-EqT model, a primary station pairs another referencing station with reliable phase picks as its neighbor. In the EdgePhase model, more neighbors are connected with the primary station based on the geographic distance, which means more data is used to predict phases for each station. The geographic locations of the template and searching stations are not considered in the S-EqT model. However, the distances between stations is still useful information for phase picking, since nearby stations have closer phase arrival times on the seismograms for the same event. In addition, S-EqT constructs a message between the primary and referencing station through the cross-correlation between two station embeddings. In the EdgePhase, messages are learnt by a neural network, which takes the embeddings of sending and receiving stations as input. The neural network, rather than simply calculating the similarity between two embeddings, is more expressive than the cross-correlation func-

tion. Therefore, a neural network could extract more relevant information that nodes intend to communicate with each other through optimizing the weights to minimize loss function. In the training process, EdgePhase makes two additional improvements compared with S-EqT. First, the parameters of the EqTransformer model are fixed during the training process of S-EqT, but are trainable during the fine-tuning process of EdgePhase. Making EqTransformer parameters trainable increases the degrees of freedom of the model, but achieves better performance after fine-tuning. Second, S-EqT is optimized with Stochastic gradient descent by setting batch size to 1 for convenience, while EdgePhase is optimized with mini-batch gradient descent (batch size = 16). Stochastic gradient descent updates parameters once per sample, so it usually takes a longer time to converge to the minimum value of the loss function and involves numerous random and less efficient update steps compared with mini-batch gradient descent.

To conclude, although these models have a lot of differences in model structures, they all analyze multi-station waveforms in a GNN framework. As we gain a better understanding of constructing graphic networks for geophysical problems, the next generations of GNN models will have a stronger expression and generalization ability. Finally, we expect more developments about how to efficiently exchange and aggregate information among multiple stations, how to design GNN models for different geophysical tasks, and how to build a dataset suitable for these models.

3.7 Conclusions

By integrating an Edge Convolutional module with EQTransformer, we construct a multi-station phase-picking model named EdgePhase. EdgePhase takes into account the waveform consistency between neighboring stations and improves the phase detection capability at low to medium noise levels (SNR between 0-20 db). We compared EdgePhase with the fine-tuned EQTransformer (Baseline-A) and the original EQTransformer (Baseline-B) using the Southern California dataset (SCSN2021). The best F1 scores for EdgePhase, Baseline-A, and Baseline-B are 0.88, 0.84, 0.66 in P-phase detections; 0.86, 0.81, 0.62 in S-phase detections,

respectively. It demonstrates that fine-tuning the EQTransformer with the Edge Convolutional module makes significant improvements to phase picking tasks. The EdgePhase also shows a strong regional generalization ability, as it was trained with the Southern California dataset but performs well in an independent dataset of the 2020 Samos earthquake sequence. Compared with the NKUA catalog, EdgePhase detects 190% extra events, with the fidelity of the event locations being higher. With its comparable performance to human seismologists, EdgePhase can automatically detect and monitor regional seismic activities in real-time. Furthermore, the Edge Convolutional module could potentially be applied to other pre-trained phase picking models (e.g., PhaseNet) and other node-level tasks (e.g., multi-station seismograms denoising).

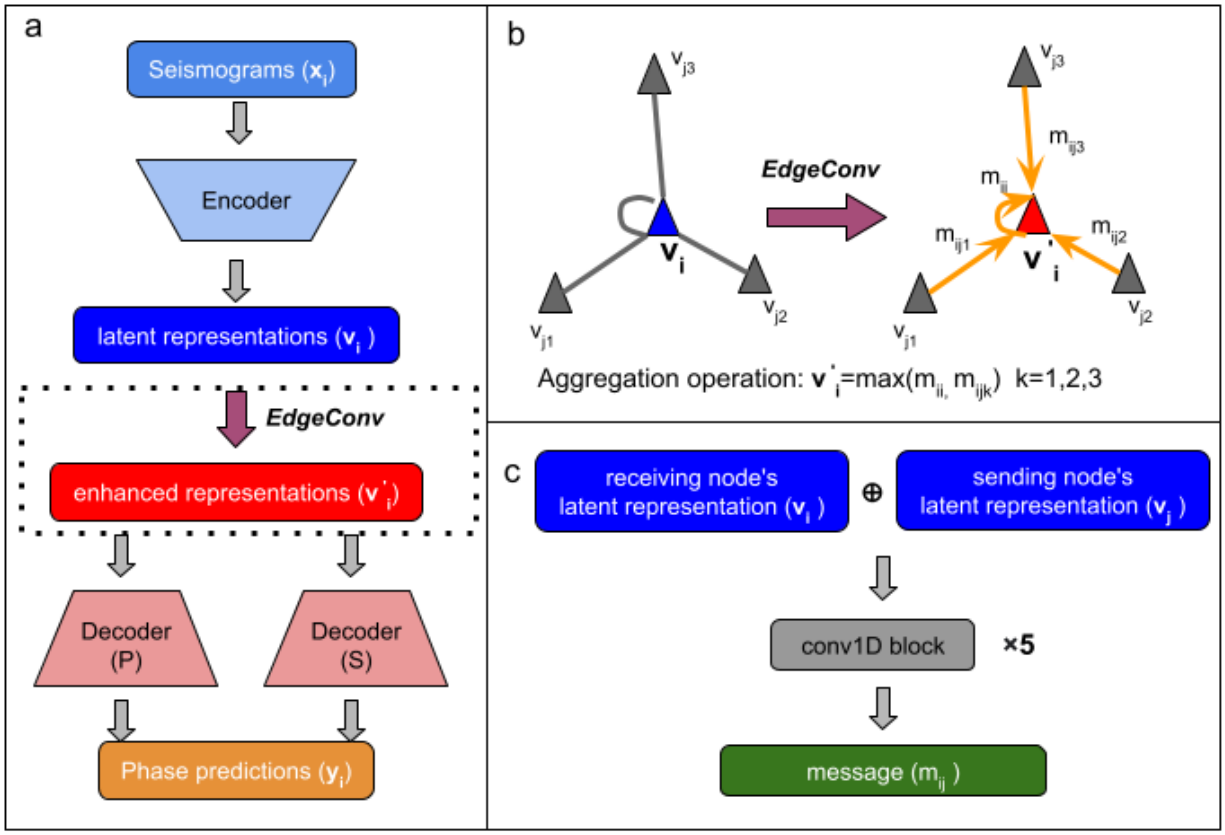

Figure 3.1: Network architecture. (a) shows the components of EQTransfromer and inserted Edge Convolution module. (b) shows the process of collecting messages (m_{ij}) from neighboring nodes, and aggregation to enhanced representations. (c) shows the neural network in constructing messages (m_{ij}) between two nodes. The operation $\oplus$ represents concatenation of two latent representations along the channel dimension.

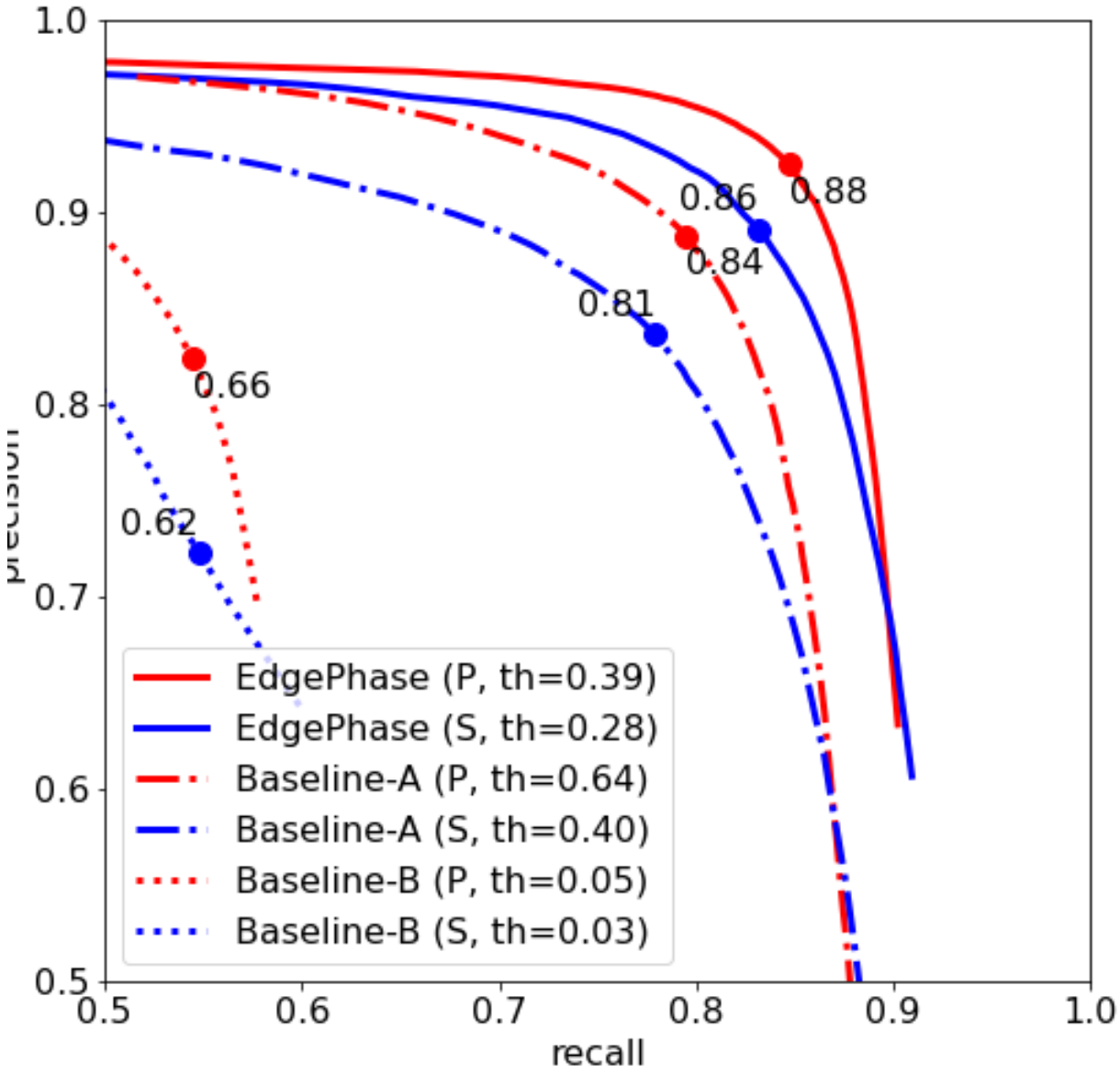

Figure 3.2: The precision-recall curve of EdgePhase, Baseline-A, and Baseline-B models on the SCSN2021 test set. The precision and recall of P-phases (blue) and S-phases (red) are calculated separately. The optimal threshold (circle) of each model is selected according to the maximum F1 score. The maximum F1 score in P-phases for EdgePhase, Baseline-A, and Baseline-B models are 0.88, 0.84, and 0.66, respectively. For S-phases, they are 0.86, 0.81, and 0.62, respectively.

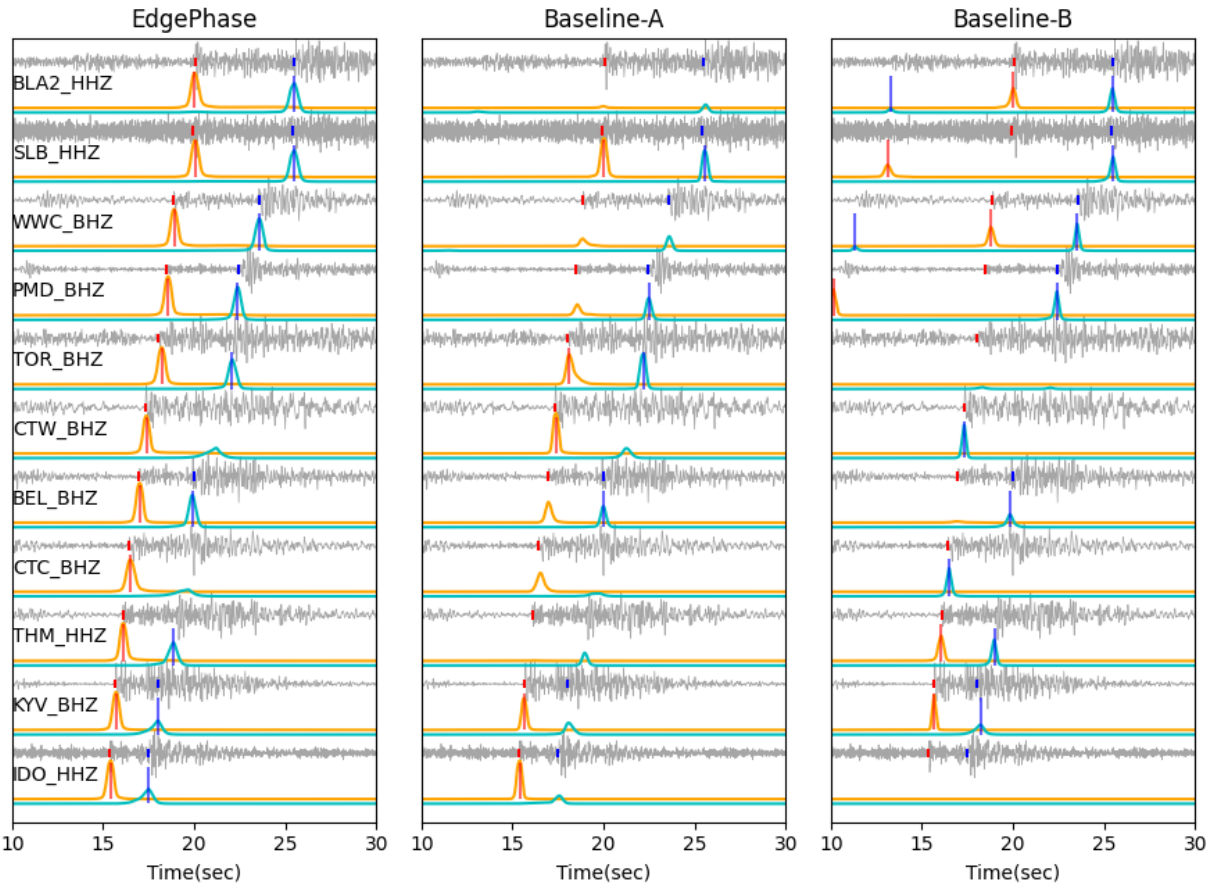

Figure 3.3: Detection probability functions of EdgePhase (left), Baseline-A (middle), and Baseline-B (right) models in an M 1.43 earthquake from the SCSN2021 test set. The earthquake occured at 01:45:56 UTC on 25th April, 2021 (202110425M1). The red and blue dots on the waveforms (gray curves) represent the ground truth P- and S-phases. Here, we missed the S-phase ground truth labels for CTC and CTW stations. The orange and cyan curves represent the predicted probability of P- and S-phases, with red and blue vertical lines on them if the peak value is larger than the best threshold determined in Table S1. Here, we only visualize the waveforms of a portion of vertical channels, since the total channel number (219) is too large to visualize all in one figure.

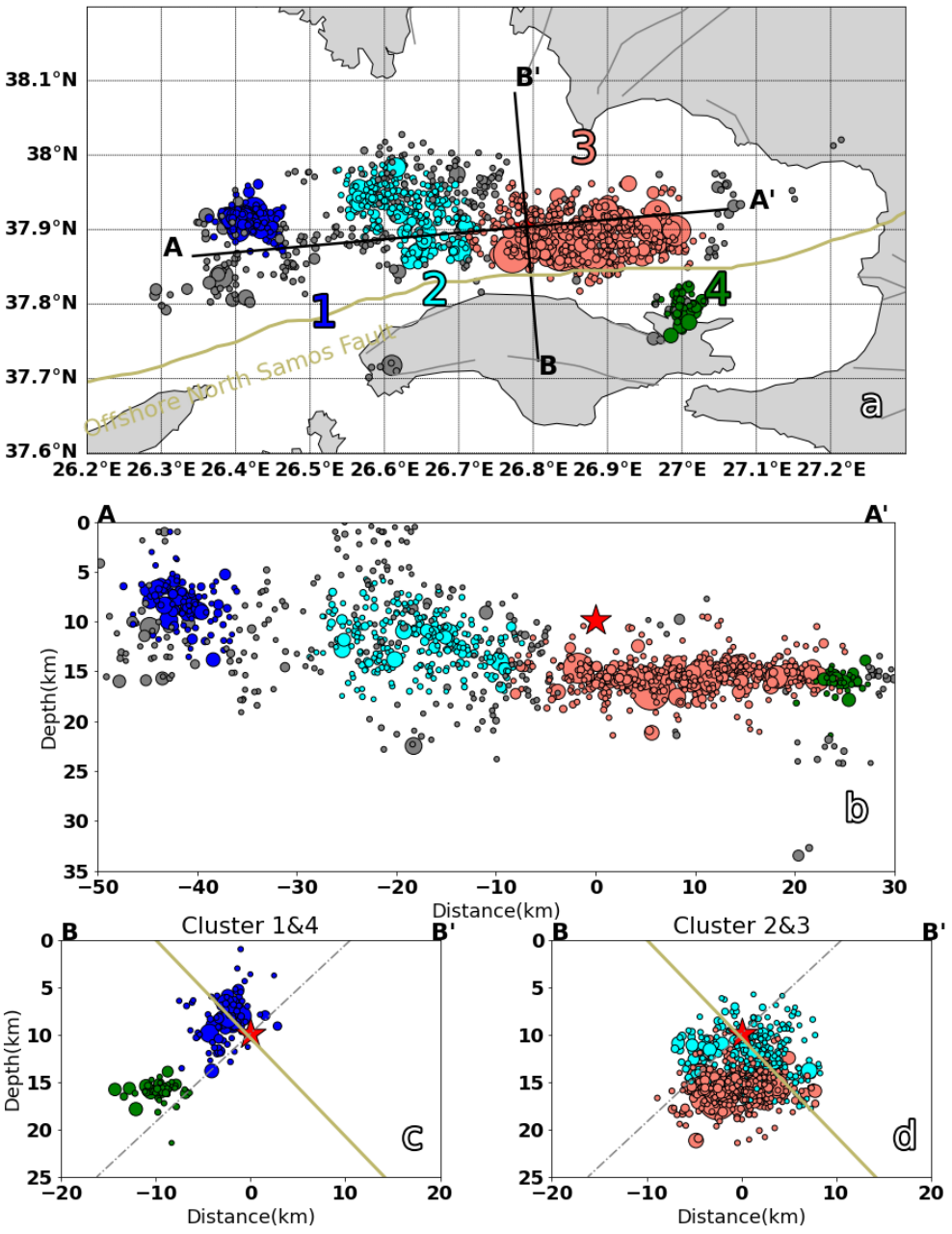

Figure 3.4: Map view and vertical cross-sections of seismicity from this study. (a) shows clustered events (colored circles) detected by this study. (b-d) represent the vertical cross–sections, respectively. We project events from Cluster 1 and 4 onto the B-B' plane in panel (c), and events from 2 and 3 onto the B-B' plane in panel (d). The yellow line and gray dashed line indicate the two nodal planes of the mainshock's focal mechanism provided by AFAD.

CHAPTER 4

A High-Frequency Distance Metric in Ground-Motion Prediction Equations Based on Seismic Array Backprojections

Originally published in **Feng, T.**, & Meng, L. (2018). A High-Frequency Distance Metric in Ground-Motion Prediction Equations Based on Seismic Array Backprojections. Geophysical Research Letters, 45(21), 11-612. I contribute to the experiements and writing of the paper, focusing on the analysis of the ground motion. Lingsen provides instructions on Backprojection and paper revision.

Note: the supplementary materials are included in Appendix A.

Abstract

Typical ground-motion prediction equations (GMPEs) measure source-to-site distances relative to the closest point on the rupture plane (R_{rup}). However, for megathrust earthquakes ($M_w > 8$), the oversimplification of the earthquake source characteristics in distance metrics results in significant bias. Recent studies suggest that the high-frequency (HF) and low-frequency (LF) energy tend to emanate from different portions of the megathrusts. This phenomenon motivates an alternative distance metric based on the array backprojection imaging technique that effectively captures regions releasing HF energy. Herein, we define an HF distance metric (R_{hf}) as the distance from the site to the high-frequency radiation zone. We study five $M_w > 7.2$ megathrust earthquakes in Japan and Chile and find that R_{hf} outperforms R_{rup} in predicting the ground shaking intensity between 0.5 and 4 Hz. We

consider R_{hf} as a complementary measure to conventional GMPE distance metrics and a more accurate ground-motion predictor in many cases.

4.1 Introduction

Recent advances in earthquake source imaging of major subduction earthquakes highlight the frequency-dependent and depth-varying seismic radiation along the plate interfaces. Low-frequency (LF) energy mainly emanates from the shallower portion of the megathrusts, while dominant high-frequency (HF) energy often radiates from the deeper portion of the megathrust (Lay et al., 2012; Yao, Shearer, & Gerstoft, 2013). In the 2011 Mw 9.0 Tohoku-Oki earthquake, the along-dip separation between fault segments that radiate HF and LF seismic waves is as large as about 100 km (L. Meng, Inbal, & Ampuero, 2011; Yao et al., 2011). These frequency-dependent source behaviors have significant impacts on the spatial pattern of the near-field ground shaking and therefore should be considered in ground-motion predictions, which are crucial to both earthquake source sciences and civil engineering applications (J. Anderson, 2007; Bozorgnia et al., 2014). The standard approach to predict the expected seismic shaking intensity is the ground-motion prediction equations (GMPEs). A GMPE provides estimates of ground-motion intensity measures based on regression analysis of the source, path, and site effects (N. A. Abrahamson, Silva, & Kamai, 2014; Atkinson & Boore, 2007; Boore, Joyner, & Fumal, 1993, 1997; Boore, Stewart, Seyhan, & Atkinson, 2014; Gregor et al., 2014). GMPEs are widely used in probabilistic seismic hazard analysis (PSHA) due to their simplicity, as it does not require local 3-D velocity models, previously observed strong motion records, or additional input such as empirical Green's functions.

The source-to-site distance is an important parameter of the GMPEs that represents the path effect of ground motion during an earthquake (Figure C.1 in the supporting information). The classical GMPEs use point-source measures relative to the epicentral and hypocentral distances (R_{epi} and R_{hypo}, respectively). Considering the finite dimensions of large earthquakes, point-source distance metrics were replaced by extended-source distance metrics. Two widely used metrics are the rupture distance R_{rup} (Douglas, 2003), represent-

ing the closest distance (orthogonal) to the fault plane, and the Joyner-Boore distance R_{JB} (Joyner & Boore, 1981), the shortest distance to the surface projection of the fault plane. A variant of R_{JB}, the equivalent hypocentral distance EHD, measures the shortest distance to the area of principle moment release on the fault (Ohno, Ohta, Ikeura, & Takemura, 1993; Si & Midorikawa, 2000). Many current GMPEs based on these frequency-independent distance metrics rely on numerous modifications of distance to accurately model near-field ground motions. Because these modifications (e.g., additional distance definitions and saturation terms) often result in complicate functional forms without seismological basis, Thompson and Baltay (2018) advocate for the use of mean rupture distance R_p, which is frequency dependent and in a simpler form. Frequency-independent distance metrics perform reasonably well in predicting the shaking intensity of small earthquakes, but the frequency dependence of source distances can become more significant for large subduction-zone earthquakes (Lay et al., 2012; L. Meng et al., 2011; Yao et al., 2013). In the case of the Tohoku earthquake, GMPEs using frequency-independent distance metrics produce substantial biases in the estimated level of shaking. Stewart et al. (2013) compared the performances of four different GMPEs (N. Abrahamson, Gregor, & Addo, 2016; Atkinson & Boore, 2007; Si & Midorikawa, 2000; J. X. Zhao et al., 2006), all of which underestimate the near-field ($< 100\ km$) ground-motion intensity of the Tohoku earthquake at short periods and overestimate shaking at long periods. To address this bias, a statistical event term is applied after residual analysis is performed, comparing GMPEs to observed ground-motion intensity (Stewart et al., 2013). This term is positive at short periods and zero or negative at long periods. The physical basis of this term is unknown, and it is not clear how the event term can be generalized to other megathrust earthquakes.

Here based on the observations that the short- and long-period seismic radiations are generated from differ- ent portions of the megathrust, we attributed this event term to the frequency-dependent source-to-site distances. In the following sections, we propose a new frequency-dependent distance metric R_{hf} based on backprojection (BP) imaging and compare it with R_{rup} in five large megathrust events: the 2011 Mw 9.0 Tohoku earthquake,

the 2003 Mw 8.0 Tokachi earthquake, the 2005 Mw 7.2 off-Miyagi earthquake, the 2014 Mw 8.1 Iquique earthquake, and the 2015 Mw 8.3 Illapel earthquake. In all cases, R_{hf} outperforms R_{rup} in the HF band (0.5–2 Hz) by showing little or no bias in ground-motion estimations.

4.2 High-Frequency Distance Metric

4.2.1 Backprojection Imaging

The frequency-dependent radiation patterns, particularly at HFs, can be effectively captured with the BP approach. BP is an earthquake-rupture imaging technique utilizing the coherent teleseismic P wavefield based on seismic array processing (Ishii et al., 2005; Kiser & Ishii, 2017; Krüger & Ohrnberger, 2005; L. Meng et al., 2011). Back-tracking of seismic waves recorded by dense arrays allows BP to determine the spatiotemporal properties of the rupture (length, direction, speed, and segmentation; Figure C.2). Over recent decades, development of large-scale dense seismic networks has enabled BP imaging of the rupture process of major large earthquakes.

The advantages of BP are threefold. First, in contrast to conventional finite-fault modeling, BP is an imaging procedure that does not involve the solution of an inverse problem. Thus, it does not suffer from being ill-posed or underdetermined and does not rely on restrictive parameterizations of the rupture kinematics and fault geometry or additional smoothing (Kiser & Ishii, 2017). Because fewer assumptions and choices are involved, BP results tend to be more robust and are more mutually consistent between different research groups than finite fault models (Kiser & Ishii, 2017). Second, BP algorithms (such as beamforming) are simple and do not require heavy computations, such as large-scale matrix inversions or wave-propagation simulations. Therefore, BP is easily automated (e.g., IRIS BP products) and well suited for implementation in real time for the purpose of rapid seismic hazard assessment or earthquake/tsunami early warning (An & Meng, 2016; L. Meng, Allen, & Ampuero, 2014). Third, since BP does not attempt to deterministically fit seismic waves and relies only

on the coherent phase of seismic array signals, it does not require the detailed knowledge of the Green's functions. This benefit allows BP to be applied to HF wavefield ($f \geq 1\ Hz$), which is generally thought to be too noisy and uncertain to be modeled deterministically.

In comparison, classic source inversions requiring the deterministic Green's function are limited to relatively low frequencies ($f < 0.1\ Hz$), for which the crustal structure is reasonably approximated. The complex rupture patterns that emerge in dynamic ruptures require finer details of the source process that are apparent in high frequencies ($f > 1\ Hz$), which are hard to retrieve through using conventional source inversions. Furthermore, the ability of BP to image HF sources is crucial to improve the prediction of strong ground motion, since the typical frequency range of intense shaking is between 1 and 10 Hz (e.g., J. Anderson, 2007; Sokolov & Chernov, 1998). The natural frequency of building response falls in this particular frequency band: A rule of thumb is that the building period equals the number of stories divided by 10, so the 1 to 10 Hz band is roughly responsible for the shaking of 1 to 10 story residential and commercial buildings.

4.2.2 High-Frequency Radiator and High-Frequency Distance

To systematically evaluate the performances of the frequency-dependent distance metrics in ground-motion predictions, we studied five large earthquakes in Japan and Chile (Table C.1 in the supporting information). These earthquakes are chosen because they occurred after 2003, when strong motion data are available for systematic strong motion analysis, and because their GCMT solutions indicate typical thrust mechanism (Figure C.3). In our study, we adopt the hypocenters and magnitudes determined by Japan Meteorological Agency and Centro Sismólogico Nacional (CSN). We perform BP analysis of the vertical velocity seismograms recorded by broadband station in North America for the five earthquakes at four representative frequencies (0.5, 1, 2, and 4 Hz). The data are available through the IRIS data management center. We chose stations at teleseismic distances between 40° and 85° because of simplicity and the high coherency of the direct P arrivals. We adopt the Multiple Signal Classification BP method with the reference window strategy that provides superior

resolution and less artifact than conventional beamforming (L. Meng et al., 2011, 2012). We first filter the seismograms aligned by initial P arrivals and then determine the HF sources (large energy bursts) as the peak locations of short-period energy released in each time frame with an interval of 1 s. We name the locations of these energy bursts as HF radiators (HFRs) since 0.5–4 Hz is relatively high in the ground-motion spectrum of earthquakes.

The spatial distribution of HFRs at four frequencies show clear along-dip patterns, with higher frequency HFRs closer to the mainland (Figure 4.1a). Assuming the coseismic rupture occurring on the megathrust, we estimate the depths of HFRs by projecting HFRs onto the predefined plate interface (SLAB 1.0, Hayes, Wald, & Johnson, 2012). We then define the HF distance (R_{hf}) as the shortest distance from a given site to all HFRs during the earthquake, analogous to the distance metric R_{rup}. Since the locations of HFRs can be evaluated with BP performed at different frequency bands, the distance metric R_{hf} (f) is also frequency-dependent. To avoid the contamination of the coda wave, only the HFRs within the earthquake duration are used to determine R_{hf}. The event duration is determined according to the average SCARDEC moment-release functions (Vallée & Douet, 2016), and we consider the earthquake over when the moment release function falls below 1% of the peak moment rate.

4.2.3 Effective Period and Uncertainty of R_{hf}

It is worth noting that the effective period of R_{hf} metric is limited from 0.1 to 10 s. In this study, we choose 0.25 to 2 s, where teleseismic BP performs best. When the period is longer than 10 s, the BP result is less reliable because of the swimming artifact, a systematic drift resulting from signal nonstationarity (L. Meng et al., 2012). The swimming artifact produces apparent energy bursts migrating in the direction of the array as time lapses because of the trade-off between source origin time and distances (Koper, Hutko, Lay, Ammon, & Kanamori, 2011). The swimming artifact is most prominent in LF beamforming BP and can lead to apparent frequency-dependent rupture behavior (L. Meng et al., 2012). On the other hand, when period is shorter than 0.1 s, the waveforms are too noisy and incoherent

to achieve the reliable BP results. To extend the valid frequency range of GMPEs, we could combine R_{hf} with other distance metrics (e.g., to use R_{hypo} when studying longer period ground motion).

Recent studies show that the spatial uncertainty of BP mainly comes from the travel time error due to 3-D path effects (Y. Chen, Meng, Zhang, & Wen, 2018; Fan & Shearer, 2017; Kiser & Ishii, 2013; L. Meng et al., 2016). This uncertainty is generated by applying a uniform the hypocenter time correction across the entire rupture area without considering the structural heterogeneity of the Earth's interior. The uncertainty can be inferred by comparing the BP-imaged aftershocks with their catalog locations. Fan and Shearer (2017) analyze the BPs of 46 M5–M7 earthquakes in Japan and determine that the median location error is about 25 km. Similarly, Y. Chen et al. (2018) study two deep earthquakes ($M > 5$) near the 2015 Mw 7.9 Bonin earthquake and estimate the spatial error around 15 km. Besides, (L. Meng et al., 2018) perform the slowness-enhanced BP based on the aftershocks of 2015 Mw 8.3 Illapel earthquake and reduce the spatial error from 24 to 8 km. Here we consider that HFR location error and therefore R_{hf} error are on the order of 20 km, which is reasonably small comparing to the dimension of megathrust events ($M > 7$) and the gaps between HF and LF sources.

4.2.4 Comparison Between R_{hf} and R_{rup}

Traditionally, when calculating R_{rup}, the preferred finite fault model is often trimmed by removing low-slip ($< 10\%$ of the peak slip) subfaults. Then, R_{rup} is measured as the closest distance to the edge of trimmed fault area ($> 10\%$ of peak slip, blue pattern in Figure 4.1a). This procedure allows R_{rup} to represent the source-receiver distance from the principle slip area and is a routine processing in the PEER NGA-West2 project (Ancheta et al., 2012; Stewart et al., 2012). Because different data sources (e.g., teleseismic, strong motion, GPS, InSAR, and tsunami) are used alone or jointly in the finite fault model inversions, the edge of trimmed fault areas could be different under the same criteria. For the Iquique and Illapel earthquakes, the edge is spatially correlated with higher frequency HFRs (Figure 4.1a).

Thus, we can deduce that frequency-independent R_{rup} is close to R_{hf} at high frequency while smaller than R_{hf} at low frequency. This difference between R_{rup} and R_{hf} increases from HF to LF. However, for the Tokachi, off-Miyagi, and Tohoku earthquake, the edge is closer to lower frequency HFRs. Accordingly, R_{hf} is close to R_{rup} at low frequency and decreases with the frequency.

4.3 Ground-Motion Predictions

4.3.1 Strong-Motion Data

We collect the strong motion data (horizontal accelerograms) recorded on the Japanese K-Net (Kyoshin network) and Chilean CSN stations, and then compare them with the GMPE predictions. Owing to the densely distributed K-Net stations, over 300 recordings of the ground-motion observations are available to study GMPEs for each Japan event. For the two Chile events, 11 and 48 recordings are used for analysis, respectively. We use 5% damping pseudo spectra acceleration (PSa), the maximum acceleration response of equivalent single degree of freedom systems, as our ground-motion intensity measure, extracted directly from the seismograms. PSa is commonly used in GMPEs, designed to approximate first-order building response. We use the Newmark Method for linear systems to compute the PSa in our study (Chopra, 1995).

4.3.2 Ground-Motion Prediction Equations

As an example of the utility of representing the HF source generation, we replace the distance metric in existing GMPE with R_{hf} and compare the performances of the GMPE with different distance metrics. We adopted the BC2016 GMPE designed for subduction earthquakes with magnitudes up to Mw 9.0 (N. Abrahamson et al., 2016). Even though some other GMPEs use subduction zone data, they are not suitable for large earthquakes in the subduction zone; that is, they do not consider the fore-arc/back-arc effect, which Stewart et al. (2013) show is essential for the 2011 Tohoku earthquake. The original form of the

GMPE uses rupture distance R_{rup}, the shortest distance from the site to the fault plane. The original form of BC2016 for the plate-interface earthquakes is

$$ln(PSa) = \delta C_1 + F_M(M) + F_D(R_{rup}, M) + F_{FABA}(R_{rup}) + F_S(Vs30, PGA_{1000}) \quad (4.1)$$

where PSa is 5% pseudo spectra acceleration, M is moment magnitude, $Vs30$ is the average shear velocity of the top 30 m, and PGA_{1000} is the reference median PGA for $Vs30 = 1000\, m/s$. F_M, F_D, F_{FABA}, and F_S represent the magnitude scaling, distance scaling, fore-arc/back-arc scaling, and site attenuation functions. The event term C_1, suggested by Stewart et al. (2013), is estimated from the residual analysis of the 2010 Mw 8.8 Maule and 2011 Mw 9.0 Tohoku earthquakes. In our analysis, we replace R_{rup} with R_{hf}. To focus on the effect of the distance metric, we also drop the event and site amplification functions. The GMPE is simplified as

$$ln(PSa) = F_M(M) + F_D(R_{rup}, M) + F_{FABA}(R_{rup}) \quad (4.2)$$

To compare the accuracy of different distance metrics in GMPEs, we compute the residual R_i between the observed and model-predicted PSa.

$$R_i = ln(PSa_i) - ln(\hat{PSa_i}) \quad (4.3)$$

where PSa_i is the PSa measured from recording i and $\hat{PSa_i}$ is the corresponding value predicted by the GMPE. This residual is later used for the correction of site attenuation, specifically in reference to the scaling of ground motions with Vs30 (Stewart et al., 2013). We then utilize the root-mean-square error (RMSE) to represent the average logarithmic residuals between predictions with the data:

$$RMSE = \sqrt{\sum_{i=1}^{n} R_i^2 / n} \quad (4.4)$$

Note that n is the total number of recordings.

4.4 Results

4.4.1 The 2011 Mw 9.0 Tohoku Earthquake

First, we select six representative stations in different regions across Japan (Figures 4.3 and C.3) and study the PSa predictions (Figure 4.2a). In all six stations, we find that R_{rup} underestimates the ground-motion intensity at the short periods ($T < 1$ s). On the other hand, the prediction based on R_{hf} fits the observation very well in a broad spectral range ($0.1 < T < 10$ s). Note that R_{hf} still underestimates the shaking at MYG0081, possibly due to the local site effect (Figure 4.2a).

Next, we compare performances of two distance metrics R_{rup} and R_{hf} at all K-Net stations, using BC 2016 (Figure 4.2b), shown for two periods of $T = 2s$ and $T = 0.5$ s. For each case, we combine the fore-arc and back-arc predictions to obtain a single RMSE. We find that the R_{hf}'s RMSE for both periods remains a small value of about 0.74; that is, the predicted PSa is on average 2.15 times either larger or smaller than the observations. On the other hand, while the RMSE of R_{rup} at $T = 2$ s is reasonably small at 0.745, it increases dramatically to 1.38 at $T = 0.25$ s, which renders R_{rup} a poor distance metric for HF. Our results indicate that R_{hf} is a more robust and accurate metric for this event. However, even though R_{hf} performs significantly better than R_{rup} in the near-field ($R < 100$ km), there is still bias in R_{hf}, overestimating at $T = 0.25$ s, while underestimating at $T = 2$ s (Figure 4.2b). We then add two more comparisons in the periods of $T = 1$ s and $T = 0.5$ s (Figures C.4 and C.5). It indicates that while RMSE of R_{hf} remains stable in both frequency bands, the RMSE of R_{rup} increases with frequency (Figure C.4). The RMSEs of R_{hf} and R_{rup} are both reasonably small at $T = 2$ s. We consider that the prediction by R_{rup} is invalid for $T < 0.5$ s for the Tohoku earthquake.

The advantage of R_{hf} can be also verified with the site attenuation analysis (Text C.1 in the supporting information and Figures C.6 and C.7). We focus on the rate $c(T)$ at which site attenuation term is changing with respect to logarithmic value of Vs30. $c(T)$, in other words, is the slope between residual $R(T)$ and $ln(Vs30)$ (Figure C.7). In the Tohoku

earthquake, the variation of $c(T)$ (Stewart et al., 2013) with the period T is much larger than the previous study in active crustal regions (ACRs; Choi & Stewart, 2005) when using R_{rup}. However, when using R_{hf}, $c(T)$ is much more consistent with that inferred from ACR data (Figure C.7). This agreement between the Tohoku earthquake case and ACR study indicates that R_{hf} is a better approximation of the distance term. A more detailed discussion of the methodology can be found in the supporting information (Barani, De Ferrari, Ferretti, & Spallarossa, 2010; Borcherdt, 1994; Borcherdt & Glassmoyer, 1989; Choi & Stewart, 2005; Stewart et al., 2013).

4.4.2 Analysis of All Five Megathrust Events

To further understand the performance of R_{hf} in other earthquakes, we analyzed four additional megathrust earthquakes in Japan and Chile in the same manner as the Tohoku earthquake (Table C.1): the 2003 Mw 8.0 Tokachi earthquake (Koketsu, Hatayama, Furumura, Ikegami, & Akiyama, 2005; Nozu & Irikura, 2008), the 2005 Mw 7.2 off-Miyagi earthquake (Shao & Ji, 2013), 2014 Mw 8.1 Iquique earthquake (Wei, 2014), and the 2015 Mw 8.3 Illapel earthquake (Figures C.8–C.15; Okuwaki, Yagi, Aránguiz, González, & González, 2017). Combining the results of all five earthquakes, we compare the RMSE of R_{rup} and R_{hf} as a function of frequency (Figure 4.1b). For the three Japanese earthquakes (left), the mean RMSE of R_{hf} remains around 0.8 for all frequency bands, while that of R_{rup} increases rapidly with frequency, from 0.8 ($T = 2$ s) to 1.2 ($T = 0.25$ s). This can be explained by our previous analysis (see section 4.2.4): R_{rup} is close to R_{hf} at low frequency and the difference increases with frequency. For the two Chilean earthquakes (right), the mean RMSE of R_{rup} shows the opposite trend, decreasing rapidly with the frequency, from 1.3 ($T = 2$ s) to 0.7 ($T = 0.25$ s) with the corresponding explanation that R_{rup} is close to R_{hf} at high frequency and the difference decreases with frequency. Our results suggest that frequency-independent R_{rup} only works well in one single frequency, and this frequency depends on the fault trimming process and the earthquake itself. Thus, R_{hf} outperforms R_{rup} despite that the trend of the errors is different for Japanese and Chilean earthquakes within the overall HF band

(0.5–2 Hz).

4.5 Discussion

Here we demonstrate that modeling the locations of HF generating areas from BP gives a new distance metric, R_{hf}, that more accurately predicts strong ground motion at distance as compared to the traditionally used metric of R_{rup}. To understand the role of HFRs in generating HF ground motions, we compare HFRs at the period of 0.25 s with the strong motion generation areas (SMGAs) and the coseismic slip of the 2011 Tohoku Event (Figure 4.3) and Tokachi Event (Figure C.16). The SMGA is defined as the source areas that are responsible for the large amplitude of HF near-field strong motions (Miyake, Iwata, & Irikura, 2003). SMGAs are typically interpreted as the source of large slip velocities or large stress release ($\sim$ 10 MPa). The locations of the SMGAs are estimated based on the onset times of the wave packets at near-field stations (Kurahashi & Irikura, 2011). Another terminology, asperity, in the context of kinematic source models, is a region on the fault surface that produces large coseismic slip relative to the average slip. These asperities often locate at shallower depths in the dip-slip faults of crustal earthquakes (Somerville et al., 1999). SMGAs are often associated with the asperities in the corresponding source models by assuming that large slip velocity is often proportional to large static slip (Miyake et al., 2003). However, in the case of the Tohoku earthquake, the identified SMGAs are more consistent with the HFRs inferred from teleseismic BP rather than the large shallow asperities (Stewart et al., 2013; Ye, Lay, & Kanamori, 2013). This agreement between SMGAs and HFRs confirms the common origin of HF bursts ($\sim$ 10 Hz) observed at teleseismic and local distances. In comparison, the region with large slip velocity is not necessarily the region of large slip (Figure 4.3).

One possible scenario to produce large slip velocity with small slip is the short rise times. Average slip velocities equal the total slip divided by the rise time. When the rise time is small, the slip velocity of subfaults can be large even with small final slip. This hypothesis is supported by the observation that the HFRs are colocated with downdip subfaults of short rise times ($< 10\ s$) and moderate localized slip ($\sim 5m$) in Tohoku earthquake (Wei et al.,

2012). The generation of these HFRs can be due to the presence of frictional heterogeneities at the seismogenic zone (Madariaga, 1983; L. Meng et al., 2011). The conditionally stable background is dotted with plenty of patchy, small-scale regions, which produce coherent short-period radiations when broken (Lay et al., 2012).

In future work, we will continue to test and improve the performances of R_{hf} in GMPEs and develop new attenuation relationship based on R_{hf}. The GMPEs using R_{hf} can be used to guide the interpolation of sparse data when producing the USGS ShakeMap. When a finite fault model is unavailable immediately after an earthquake, a quick, robust prediction of ground-motion intensity using R_{hf} can be developed based on near real-time automatic BP. One can determine the HFRs directly from IRIS Data Management Center's automated BP results for global $M > 6.5$ earthquakes, based on coherent P wave energy from regional arrays (North America, Europe, and Australia) and Global Seismographis Network (Trabant et al., 2012). The duration and magnitude of an earthquake can also be estimated from BP based on an empirical equation (D. Wang et al., 2017). Thus, we have all the necessary information to estimate ground-motion intensity. An important aspect of the GMPE is to estimate the ground motion intensity for the future earthquake scenarios, which requires the assumptions of source locations. For PSHA purposes, R_{hf} cannot be simply estimated using BP as it is only available after the earthquake. These deep HFRs are the closest to the landward strong motion stations and therefore dominate the determinations of R_{hf}. The mechanism of repeating earthquakes and HFRs are both interpreted as the repeated failures of small patches driven by either slow-slip or dynamic ruptures. Thus, these small patches in brittle-ductile transition could be the potential locations to generate coseismic HF energy in megathrust earthquakes. The spatial consistency between HFRs during the 2011 Tohoku earthquake (Figure 4.3) and the 2003 Tokachi earthquake (Figure C.16) and preseismic repeating earthquakes is an evidence for such hypothesis (Igarashi, Matsuzawa, Umino, & Hasegawa, 2001; Lay et al., 2012; L. Meng et al., 2011; Uchida, Iinuma, Nadeau, Bürgmann, & Hino, 2016). Next, we will explore which repeating earthquakes could be used as proxies for HFRs of potential future ruptures for PSHA.

4.6 Conclusion

Recent studies exploring the frequency-dependent and depth-varying seismic radiation at plate interfaces from several major megathrust earthquakes motivate us to consider frequency-dependent source effect for GMPEs. The corresponding HF distance metric R_{hf} is based on HFRs, which are quickly and robustly located by BP imaging technique without any prior knowledge of rupture kinematics. We demonstrate that R_{hf} performs better in GMPEs over the traditional frequency-independent metric R_{rup} in several recent large megathrust earthquakes based on the analysis of the 2011 Tohoku earthquake, the 2003 Tokachi earthquake, the 2005 off-Miyagi earthquake, the 2014 Iquique earthquake, and the 2015 Illapel earthquake. R_{rup} shows significant bias in the prediction of 5% pseudo spectral acceleration at high frequencies (0.5–4 Hz), which is close to the natural frequency of low-rise buildings. However, the distance metric R_{hf} based on BP shows little or no bias in ground-motion estimations. R_{hf} therefore potentially has the ability to improve the accuracy of GMPE predictions. GMPEs utilizing on R_{hf} can potentially provide real-time ground-motion intensity estimations for large earthquakes based on automated BP.

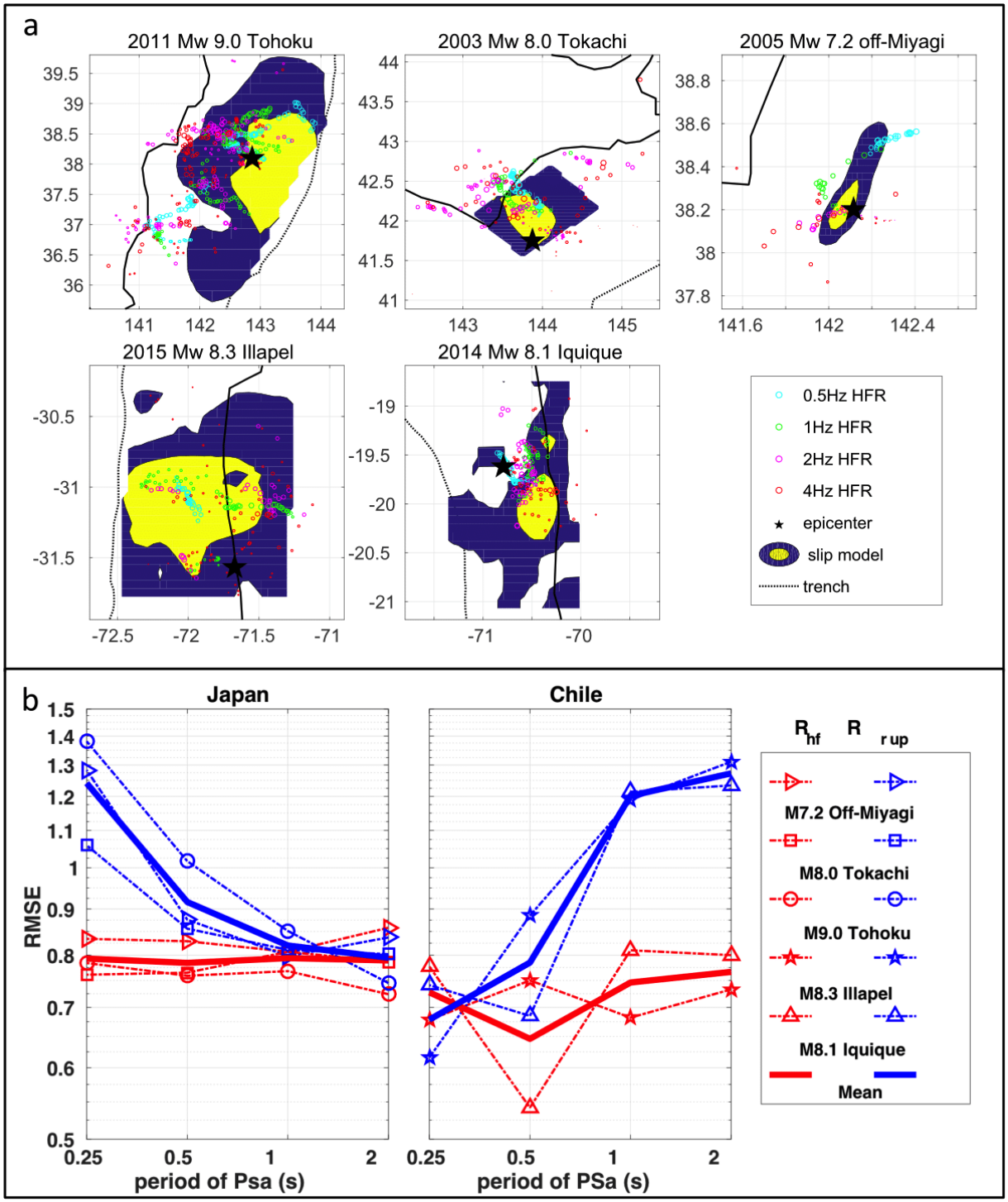

Figure 4.1: (Caption next page.)

(a) Backprojection results of five large earthquakes in Japan and Chile: the 2011 Mw 9.0 Tohoku earthquake, the 2003 Mw 8.0 Tokachi earthquake, the 2005 Mw 7.2 off-Miyagi earthquake, the 2014 Mw 8.1 Iquique earthquake, and the 2015 Mw 8.3 Illapel earthquake. The black star marks the hypocenters of the five earthquakes. The color open circles represent high-frequency radiators in four frequency bands. The yellow and blue patches represent areas with coseimic slip larger than 50% and 10% of peak slip, respectively. (b) Comparison of the performance between R_{rup} and R_{hf} in five earthquakes. The blue dashed lines represent R_{rup}, while the red lines represent R_{hf}. (left) The R_{rup}'s mean root-mean- square error (RMSE) value of three Japanese events (solid blue line) decreases with period, while R_{hf}'s mean RMSE (solid red line) changes little. (right) The R_{rup}'s mean RMSE value of three Chilean events increases with period while R_{hf}'s mean RMSE remains small.

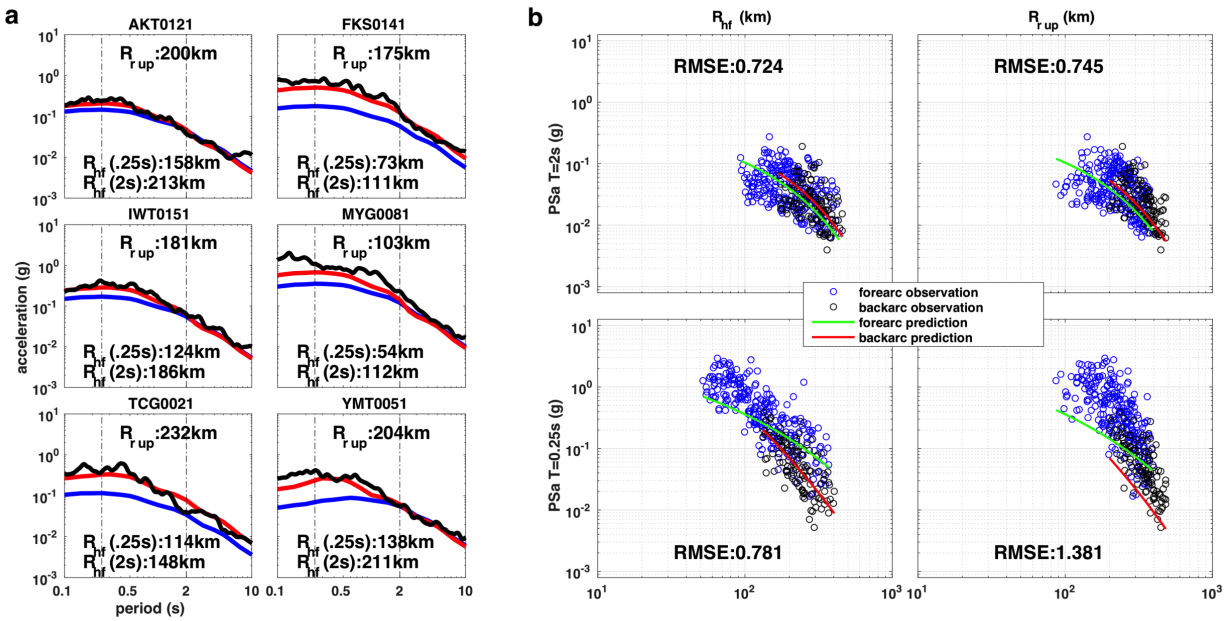

Figure 4.2: (a) Comparisons of pseudo spectra acceleration (PSa; g) predicted with distance metrics R_{hf} and R_{rup} of 2011 Tohoku earthquake (Mw 9.0) at six selected K- Net stations. The black, blue, and red lines denote the observation, the prediction based on R_{rup}, and the prediction based on R_{hf}, respectively. (b) Comparisons of PSa in different frequencies using R_{hf} and R_{rup} for the 2011 Tohoku earthquake. The x axis marks distance in km, and the y axis denotes PSa in gravitational acceleration g. The green and red traces denote the fore-arc and back-arc PSa values predicted with GMPE BC2016. The blue and black open circles are the observed PSa values at fore-arc and back-arc K-Net stations, respectively.

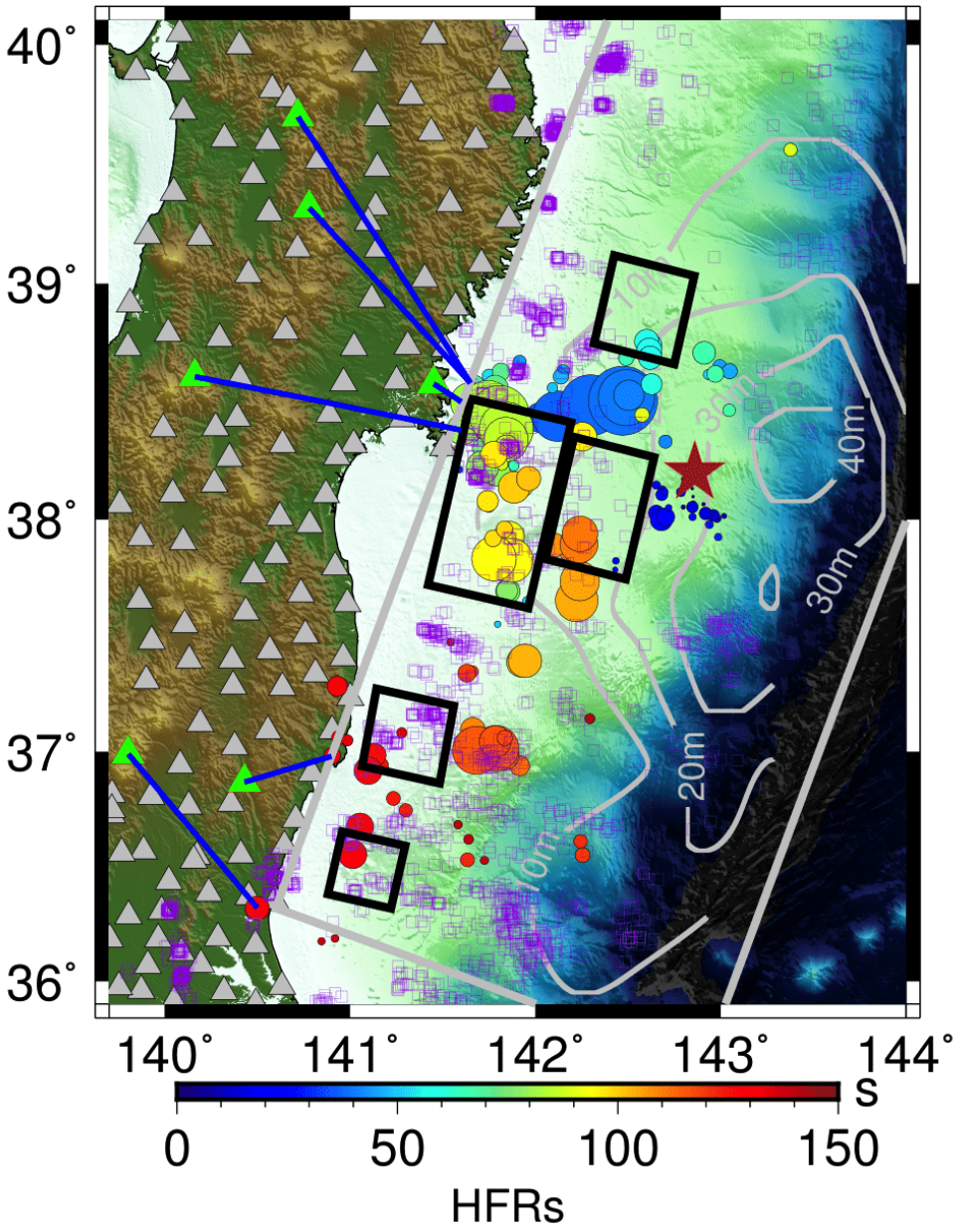

Figure 4.3: Finite fault model (Wei et al., 2012) of 2011 Mw 9.0 Tohoku earthquake (gray rectangle), strong motion generation areas from Kurahashi and Irikura (2011) (black rectangles), the 4-Hz high-frequency radiator locations (color spheres), and deep preseismic repeating earthquakes (Uchida et al., 2016)(purple open squares). The contours indicate coseismic slip (m), and the red star is the JMA epicenter of the main shock. The blue lines denote the surface projection of R_{hf} at six selected stations (green triangles).

CHAPTER 5

Discussion and Conclusion

My Ph.D. research aims at characterizing microseismicity through the use of machine-learning and waveform-similarity-based approaches. In our case study of the aftershock sequence of 2011 M9 Tohoku, Japan earthquake, we propose a waveform-similarity-based model (BP-MF) which integrates BP and MF. The BP-MF method is highly effective in detecting offshore microseismicity. After spatial and temporal calibration, the BP-MF aftershock catalogue contains 44.2% more offshore events than the JMA catalog. We inferred the large co-seismic slip area from the aftershock distribution, which is spatially complementary with post-seismic slip. The large coseismic slip area inferred by our model is consistent with several finite fault models that indicate large coseismic slip in the near-trench region offshore Fukushima. In contrast to deploying ocean bottom instruments, the BP-MF method makes use of waveforms recorded in teleseismic distances, resulting in an inexpensive solution for offshore seismicity monitoring.

In the case study of the aftershock sequence of the 2020 M7 Somas, Greece earthquake, we build a multiple-station phase-picking model (EdgePhase) by integrating GNN with EQ-Transformer. Compared to a local earthquake catalog, EdgePhase showed 190% additional detections with event distribution more conformative to a planar fault interface, suggesting a higher degree of fidelity in event locations. This case study shows that EdgePhase, when combined with traditional earthquake location methods, can automatically monitor earthquakes. Meanwhile, the performance of the multi-station phase-picking model is demonstrated to be comparable to that of humans.

In addition to studying microseismicity, my thesis investigates ground motion during

megathrust earthquakes. We propose a new frequency-dependent distance metric based on the BP imaging technique. The method improves the accuracy achievable in rapid post-seismic mapping of probable strong ground motions, especially where strong-motion stations are sparse. We study five large subduction zone earthquakes in Japan and Chile and R_{hf} outperforms traditional distance metrics. We consider R_{hf} as a complementary measure to conventional GMPE distance metrics and a more accurate ground-motion estimator in seismic hazard analysis.

The British statistician George Edward Pelham Box FRS said, "All models are wrong, but some are useful". While all models are inaccurate due to their simplification of reality, simplifying the realities of life can be quite useful, since it helps us explain, predict and better understand the universe and all its various components. Many generations of iterations are necessary to create a good model. The models presented in this thesis are not entirely new, but rather are improved upon existing models. In future studies, I believe there is still considerable room for improvement in the study of microseismicity. I will highlight these improvements in the following three aspects: data, model, and task.

My thesis only makes use of the conventional seismic data, but new unconventional data has recently been used in the study of microseismicity and oceanic seismicity. Data collected from distributed acoustic sensing (DAS; H. F. Wang et al., 2018; Lindsey et al., 2017; Martin et al., 2018; Marra et al., 2018) shows high spatial and temporal resolution, and thus is suitable for detecting microseismicity. It is necessary to compress DAS data prior to applying it to make predictions, as the volume of data from DAS is very large. Self-supervised learning and semi-supervised learning are effective approaches for extracting latent representations from DAS data. In addition, A study of the state of polarization (SOP) of a transcontinental submarine cable (Zhan et al., 2021) detected multiple moderate-to-large oceanic earthquakes along the cable. Apart from this, we can also utilize the data provided by citizen such as felt reports (Atkinson & Wald, 2007), social media (Sakaki, Okazaki, & Matsuo, 2010; Earle, Bowden, & Guy, 2011), web traffic (Bossu et al., 2008), acceleration data recorded by smartphones (Kong et al., 2016), or Internet of things (IoT) devices.

In addition to new data, we can also build models based on new concepts. As we discussed in Chapter 3, combining physics-based models with ML models is an important trend in characterizing microseismicity. Physics-based models offer the benefit of understanding the physics of systems; machine-learning models are able to take advantage of large volumes of data and high levels of complexity. In Chapter 3, we compared EdgePhase with two other DL models that included physics-based modules (similar to BP and MF). This is a good strategy for developing a new generation of microseismicity characterizing methods, but the research is still in an early stage. There are some challenges in integrating physics-based module into ML architecture. For example, how can physical parameters be incorporated and made applicable to multiple types of instruments, regions, and events? A model of this kind would show superior performance if the generalization problem can be solved efficiently. Another trend is to integrate several types of data when developing a model. In a microseismicity task, these data could consist of seismic data, GPS data, and DAS data. In a broader sense, these data types could encompass geology data, geophysics data, and geochemistry data.

Within this thesis, we only examine the task of detecting microseismicity and seismic phases with supervised learning, primarily because the labels in this task are most complete and reliable. Once we have accumulated enough experiences and techniques, we will be able to address more challenging questions with partially labeled data, such as predicting focal mechanisms of microseismicity. There have been some studies that have attempted to solve this problem, but the experiments were conducted in a very limited number of instances (Kuang et al., 2021; Uchide, 2020). For more general applications, we need to use more innovative and specific ML models, and the traditional supervised learning approach is unlikely to be suitable, perhaps semi-supervised or self-supervised learning is appropriate. There is also the possibility of applying machine learning to earthquake nucleation and triggering. Our hypothesis depends on establishing a comprehensive catalog, which is what we are attempting to do in this thesis. Utilizing unsupervised learning algorithms, we were able to cluster events easily for the subsequent analysis. Furthermore, GNNs can also be used to predict the triggering relationships between events. In addition, we can use a ML model

to identify the migration of earthquakes as well as to trace repeating earthquakes. Through these studies, we might gain a deeper understanding of cascade and slow slip triggering models.

If we can make progress in the three aspects outlined above, we may be able to create an artificial intelligent model that behaves like a human seismologist. Seismicity can be analyzed by the model automatically based on a variety of sources of seismic data. As compared to human experts, it has the advantages of being objective, robust, and fast. Besides microseismology on Earth, the system may also be effective in detecting tremors, slow earthquakes (Rouet-Leduc, Hulbert, McBrearty, & Johnson, 2020; Hulbert et al., 2019), moonquakes, and marsquakes through transfer learning (Civilini et al., 2021).

Apart from studying microsemicity, machine learning has many potential applications for solving geophysics problems. Listed here are some noteworthy applications:

(1) In Chapter 2, we determine microseismicity from the BP energy series by looking at spatial peak ratios and temporal peak ratios. It would be advantageous to detect and locate events directly from the original BP movies (animated maps of energy radiation). For this task, I recommend using object detection models used in computer vision, such as R-CNN (Ren, He, Girshick, & Sun, 2015) and YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016). By taking advantage of different frequencies and networks in the conduct of BP, we can augment the BP movies, in much the same way we use multiple cameras when detecting objects.

Additionally, we perform spatial and temporal calibration to improve the hypocenters and origin times of newly detected events by BP with linear interpolation. Using the historical events recorded in a conventional catalog as ground truth, this calibration task is very suitable for supervise learning. Linear interpolation may capture the majority of errors, but cannot quantify the velocity heterogeneity that caused them. With the aid of machine learning, we can build a more general model to estimate the slowness term, and inversely, to help BP perform better by considering 3-D velocity structures. Han, one of my colleagues, discovered that one slowness correction may not be applicable to a large geographical area

due to the structural heterogeneity. By using the K-means algorithm, he clustered the region into several groups and fitted slowness separately. Both clustering techniques and supervised learning can be used in this task.

Based on the aformationed study of detection and calibration, it is possible to train a BP imaging model in conjunction with ML algorithms. Using continuous waveforms and a local catalog as training data, a ML model is supervised to learn to perform BP by considering three-dimensional velocity structures. An objective function may be defined as the spatial and temporal difference between the BP energy peaks and catalog events. With trained BP model, the travel time errors should be theoretically corrected, and BP imaging results should be more accurate than current versions without the need for calibration. The model can be constructed in the same way as the EQNet (W. Zhu et al., 2021) we discussed in Chapter 3. EQNet's shift-and-stack module actually functions as BP. EQNet only shifts features with constant P and S velocities. Here, I recommend replacing shift-and-stack module with a learnable model, whose parameters can be used to characterize 3D velocity structures. Furthermore, it would be great if this could be applied to more than one network (Du, 2021), making BP results more robust on multiple networks.

(2) As we illustrated in Chapter 3, EdgePhase model (message passing neural network) is very suitable for signal denoising problems when there are multiple sensors present. Denoising is not limited to seismic data, as it can also be applied to GPS and infrasound data. Machine learning has great potential in denoising geodetic data as they require complex preprocessing procedures to highlight transient signals. One of the difficulties is that we lack knowledge of the ground truth (clean) signal in many instances, so we need to rely on synthetic data for model training. Accordingly, supervised learning may not be the best solution, so it might be useful to adopt unsupervised/self-supervised approaches, such as generative adversarial networks (GANs) or variational autoencoders (VAEs).

(3) Another trend is to employ ML models to approximate or simplify complex function or processes in physics. It has been used to accelerate calculations on large-scale viscoelastic calculations (DeVries, Thompson, & Meade, 2017). It could also help with waveform sim-

ulation problems. The traditional simulation method may be very time consuming when studying the properties of waveforms at high frequencies. It is possible to use a ML model as a fast computing approximator and therefore help save a great deal of resources when we require a quick resolution or real-time monitoring.

In addition to the aforementioned tasks, ML can be applied to many other topics. As an example, to examine the structure of the Earth, as well as other planets, especially when we have seismic data from the InSight Mission on Mars. Several studies have applied machine learning to the study of tomographs (Mauricio, Jennings, Adler, & Dahlke, 2018; Valentine & Woodhouse, 2010; Paitz, Gokhberg, & Fichtner, 2018; Bianco & Gerstoft, 2018). Dicitonary learning may be an interesting direction to explore. Also, we can make use of machine learning to automate some tasks, such as categorizing volcanic ash particles (Shoji, Noguchi, Otsuki, & Hino, 2018). In a similar way, we can apply machine learning to identify images of rocks, faults, landslides, craters and other objects based on techniques used in computer vision.

Overall, there are a large number of applications to be considered. At present, we are at a very early stage of applying ML tools to the study of geophysical problems. Machine learning, especially deep learning, is developing at an astounding rate, with new models emerging each month. Hopefully, previously unsolvable problems will soon be resolved.

APPENDIX A

Supplementary Materials of Chapter 2

A.1 Text S1 Comparison between this study and KI-2013

. Since KI-2013 (Kiser & Ishii, 2013) only performs the BP up to 25 hours after the mainshock, we compare our BP catalog with it in the same time span. KI-2013 and our method detected 600 and 661 events respectively, 437 of them were matched (Fig. A.7). We consider two events are matched in two catalogs if they occur within 60 sec and 50 km from each other.

We also match the BP detections with large events ($Mw > 4.5$) in the JMA catalog (431). KI-2013 detected 266 of them and our catalog matched 49 more events. Among these matched events, 243 of them are detected by both BP methods (Portion A in Fig. A.7). Compared to the KI-2013 catalog, the event location determined by our method is overall more landward (Fig. A.8a). This difference might come from the selection of mainshock hypocenter and settings of hyperparameters. Also, we perform the spatial calibration, which systematically corrects the BP location according to event hypocenters from the JMA catalog. In addition, the origin time determined by KI-2013 is overall 10 sec later than our method which has the same median origin time with the JMA catalog (Fig. A.8c). KI-2013 did not perform the temporal calibrations.

However, both methods are not able to match 93 events in the JMA catalog ($Mw > 4.5$). Some events are in fact detected by the BP techniques, but the spatial or temporal differences are larger than the matching thresholds. The remaining events are missed due to the low signal-to-noise ratio of the teleseismic waveforms. Please check Section 2.2 for more details.

334 and 346 events in KI-2013 and our BP catalogs are not matched with large events ($Mw > 4.5$) in the JMA catalog. The majority of them are newly detected events, and some of them are events recorded in the JMA catalog but with smaller magnitude ($Mw < 4.5$). We then used the same criteria to match the remaining events in two BP catalogs, finding 194 matched pairs (Portion B in Fig. A.7). The spatial and temporal differences between these two catalogs shows the same pattern: the hypocenter determined by KI-2013 is more trenchward and origin time is 11 sec later than our method (Fig. A.8b,d). Compared with events matched with the JMA catalog ($Mw > 4.5$, Portion A), these unmatched events (Portion B) in both BP studies are more trenchward. This pattern indicates the advantages of BP in detecting near-trench events.

Overall, our method detected more events and matched more events with the JMA catalog than KI-2013. Two BP catalogs have about 70% overlapping events. The different performance may come from the following aspects: First, KI-2013 manually identifies events from the energy peaks, while our method sets up an objective threshold to identify earthquakes. Second, we add additional quality control by inspecting the spatial energy distribution. Third, KI-2013 used a coherency function approach while we used conventional beamforming.

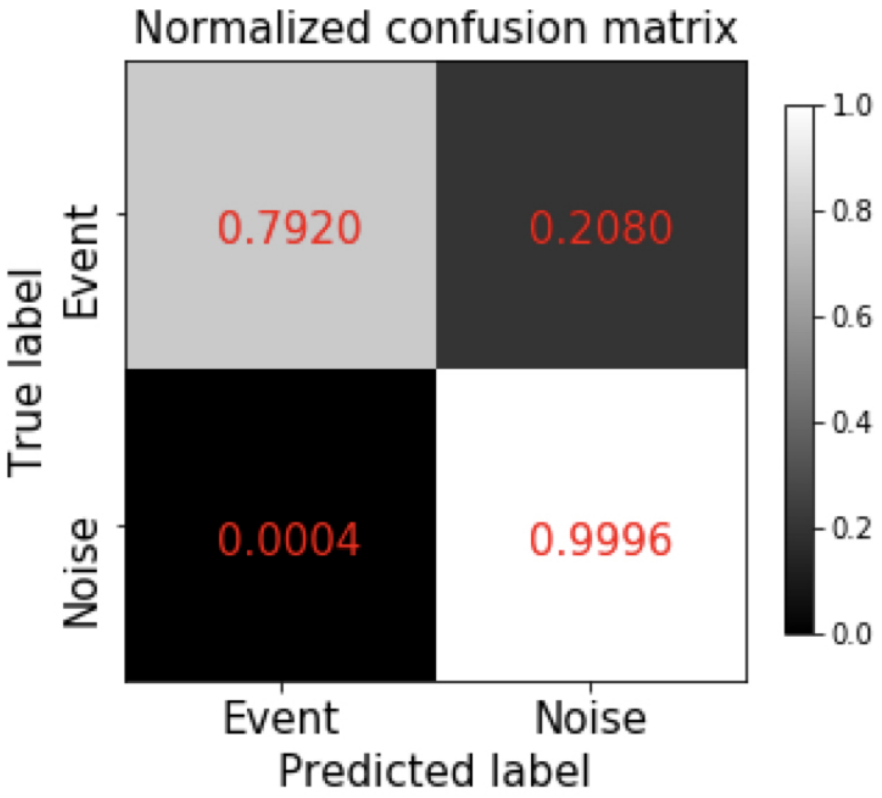

Figure A.1: Normalized confusion matrix for binary classification (event and noise). The true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) are 0.7920, 0.0004, 0.2080, and 0.9996 respectively.

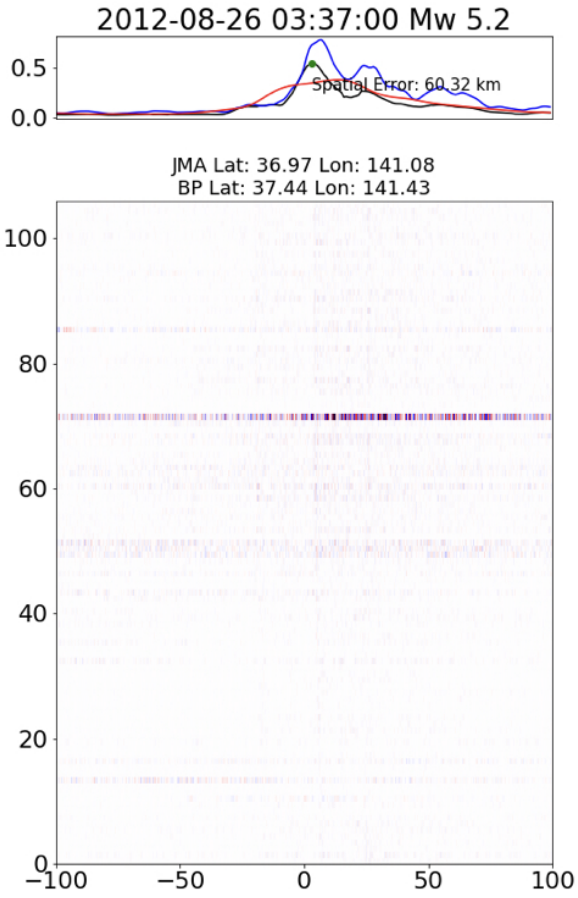

Figure A.2: Upper panel: The BP peak stack energy (black), temporal threshold (red), spatial threshold (blue), and the waveforms (lower panel) of an Mw 5.2 event on 2012-08-26 03:37:00. This event is missed by BP detection because the energy peak is smaller than the spatial threshold.

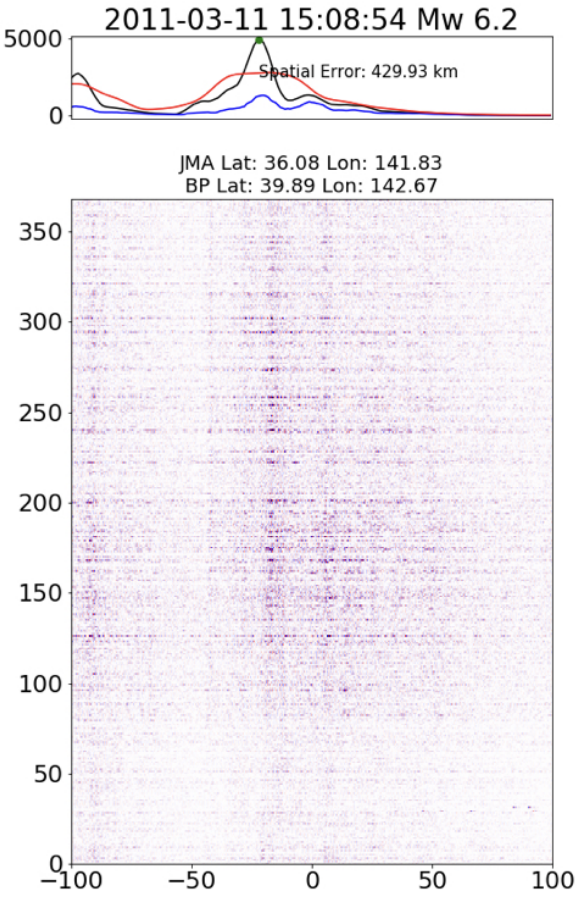

Figure A.3: Upper panel: The BP peak stack energy (black), temporal threshold (red), spatial threshold (blue), and the waveforms (lower panel) of an Mw 6.2 event on 2011-03-11 15:08:54. This detection is not matched with the JMA catalog because the hypocentral difference is larger than the threshold of 50 km.

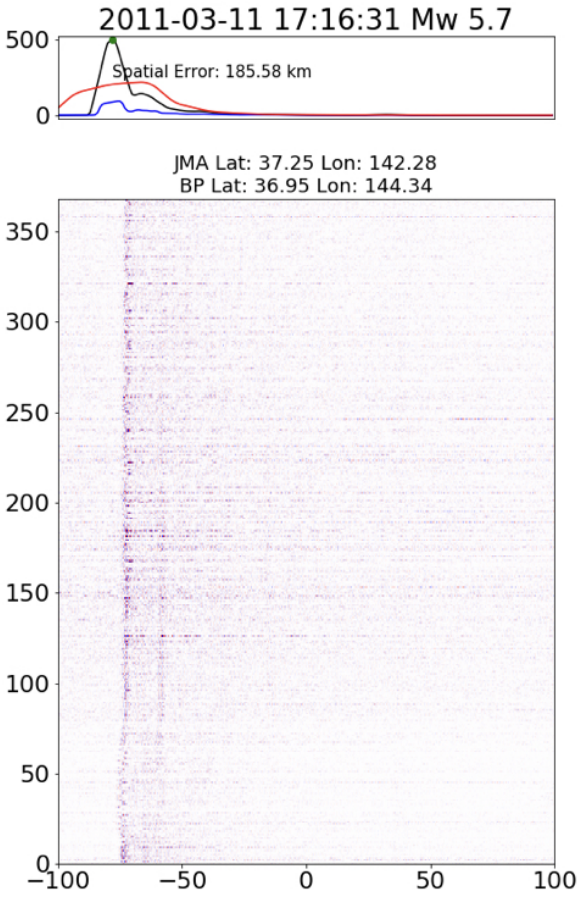

Figure A.4: Upper panel: The BP peak stack energy (black), temporal threshold (red), spatial threshold (blue), and the waveforms (lower panel) of an Mw 5.7 event on 2011-03-11 17:16:31. This detection is not matched with the JMA catalog because the origin time difference is larger than the threshold of 60 sec.

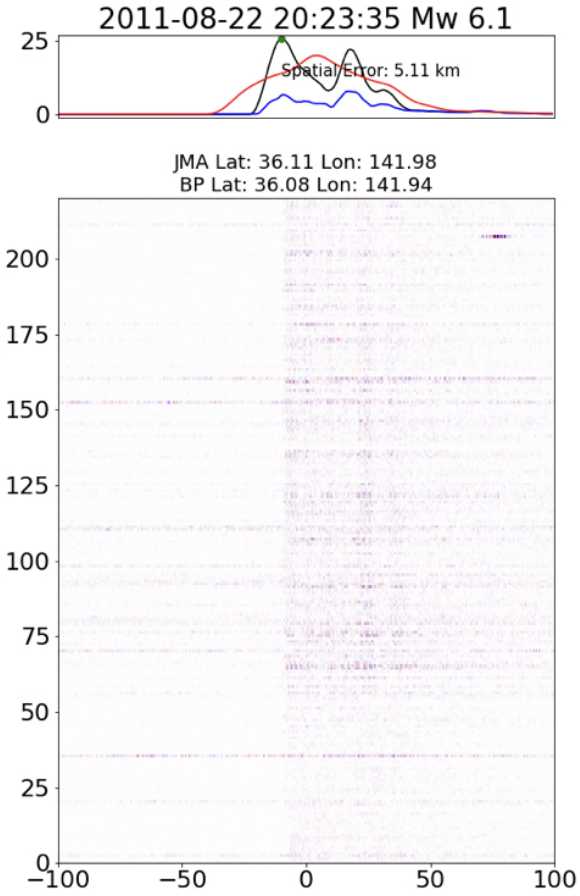

Figure A.5: Upper panel: The BP peak stack energy (black), temporal threshold (red), spatial threshold (blue), and the waveforms (lower panel) of an Mw 6.1 event on 2011-08-22 20:23:35 (Lat = 36.1072, Lon=141.9840). In the JMA catalog, an Mw 4.4 event occurs 24 sec earlier than this Mw 6.1 event (Lat = 36.0703, Lon = 141.9413). Our BP program matched the first energy peak with the Mw 4.4 event and considered the second energy peak represents the S wave.

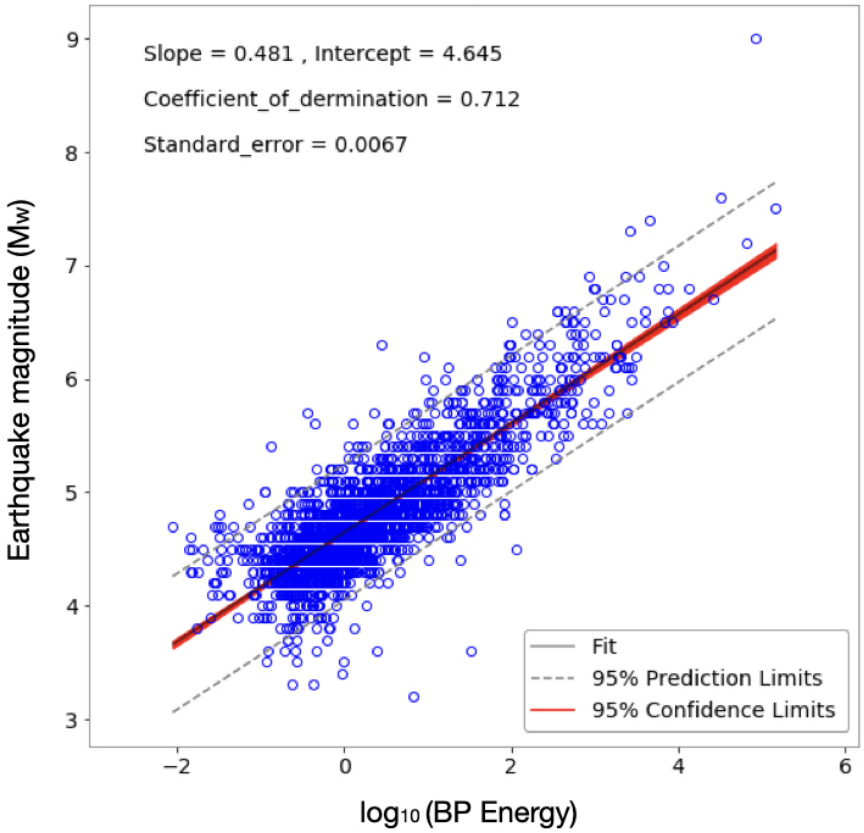

Figure A.6: The relationship between the BP energy and the earthquake magnitude in the JMA catalog. The best fit line (gray line) with 95% confidence limits (red zone) is presented together with 95% prediction limits (gray dashed lines).

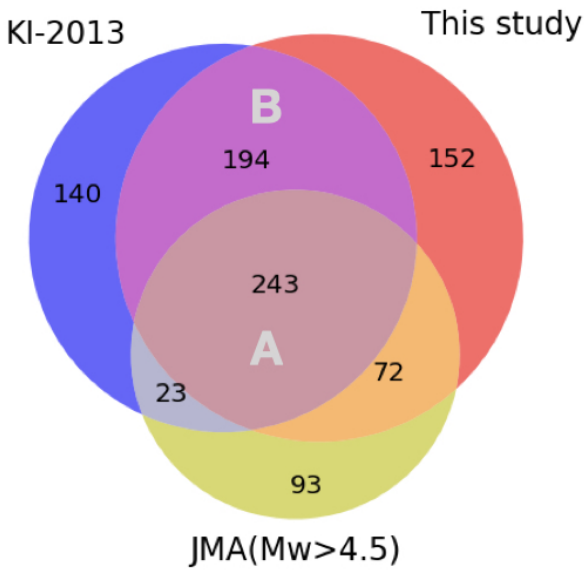

Figure A.7: Comparison between KI-2013 BP catalog (600 events), BP catalog of this study (661 events), and the JMA catalog ($Mw > 4.5$, 431 events) in the first 25 hours after the 2011 Mw 9.0 Tohoku mainshock. The portion A is the overlap of three catalogs, while the Portion B are matched events in two BP catalogs but not present in the JMA catalog ($Mw > 4.5$).

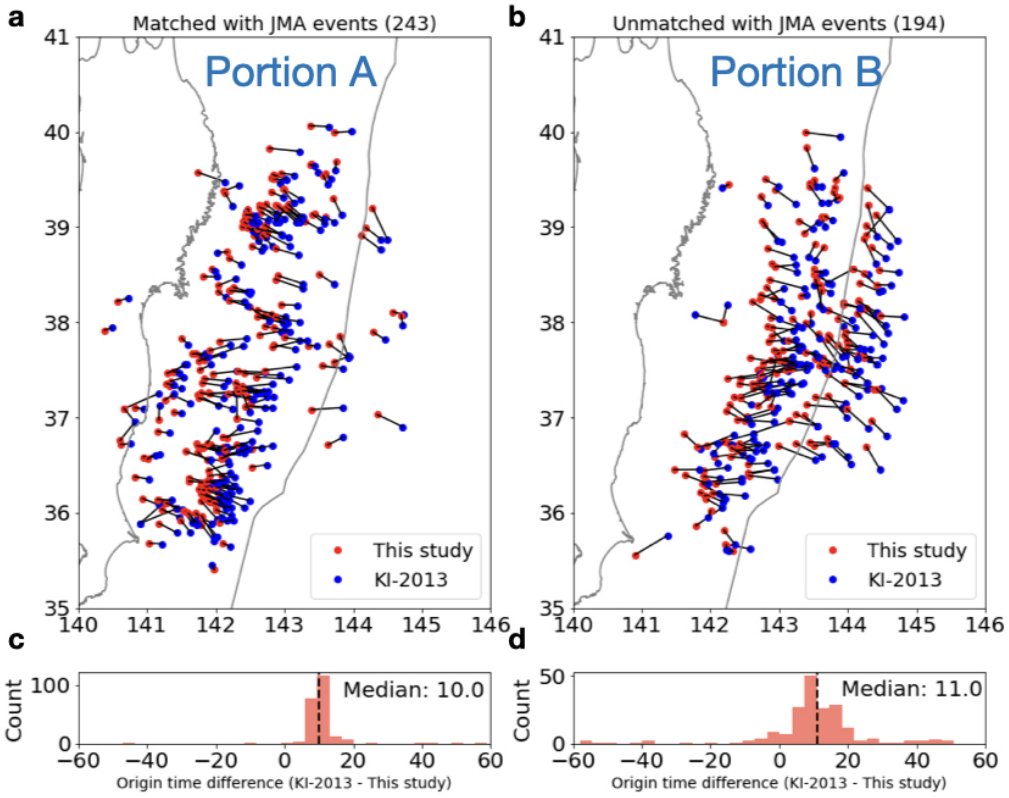

Figure A.8: The spatial (a-b) and temporal (c-d) differences of matched events between KI-2013 and this study. The left panels (a, c) are events of Portion A in Fig. A.7. The right panels (b, d) are events of Portion B in Fig. A.7.

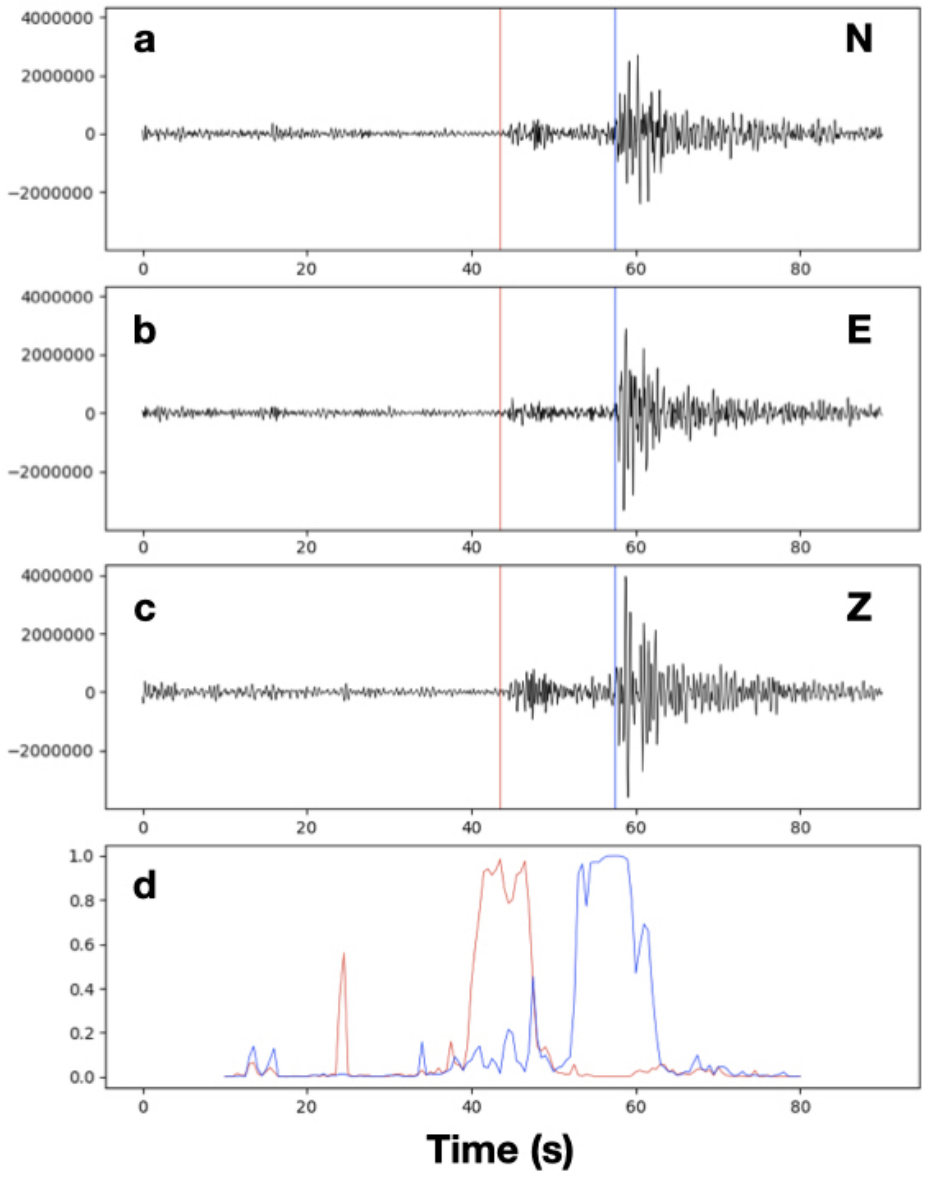

Figure A.9: An example of the GPD algorithm based on deep learning (Ross, Meier, Hauksson, & Heaton, 2018). It takes the 3-component seismic waveforms (a-c) as input and outputs the probability of S and P phases as a function of time (d). The red and blue lines indicate the offset of the estimated P and S phases (a-c).

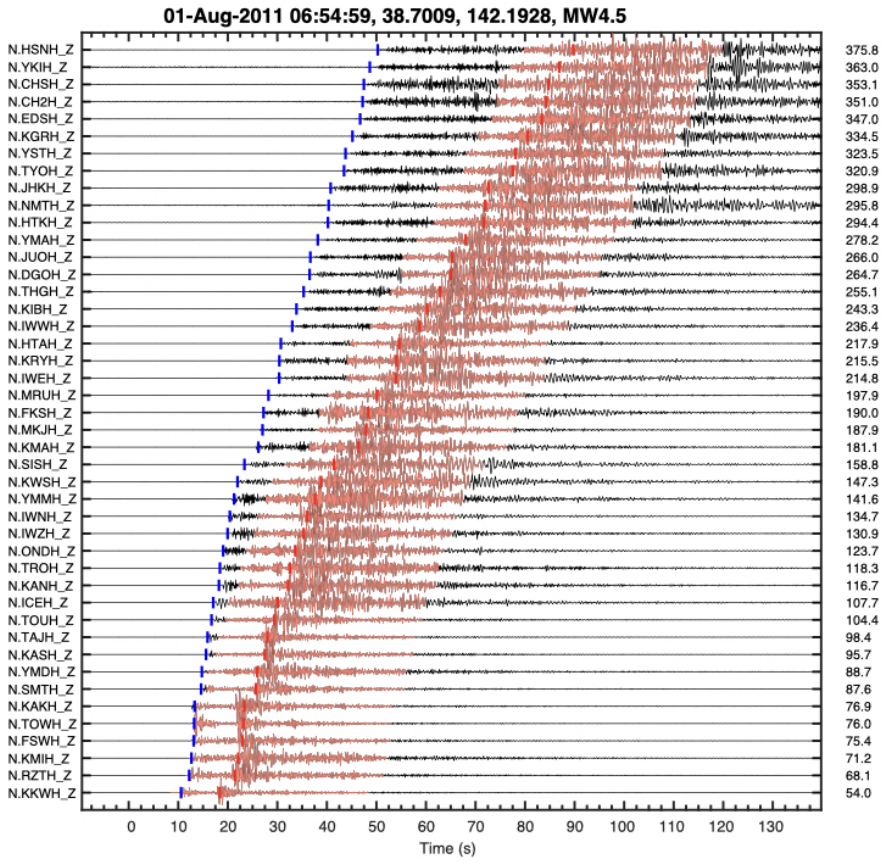

Figure A.10: Waveforms of an Mw 4.5 near-coast event recorded by 47 Hi-net stations. Recorded P and S wave trains (salmon color) are close to theoretical P (blue) and S (red) arrivals predicted by the IASP91 model.

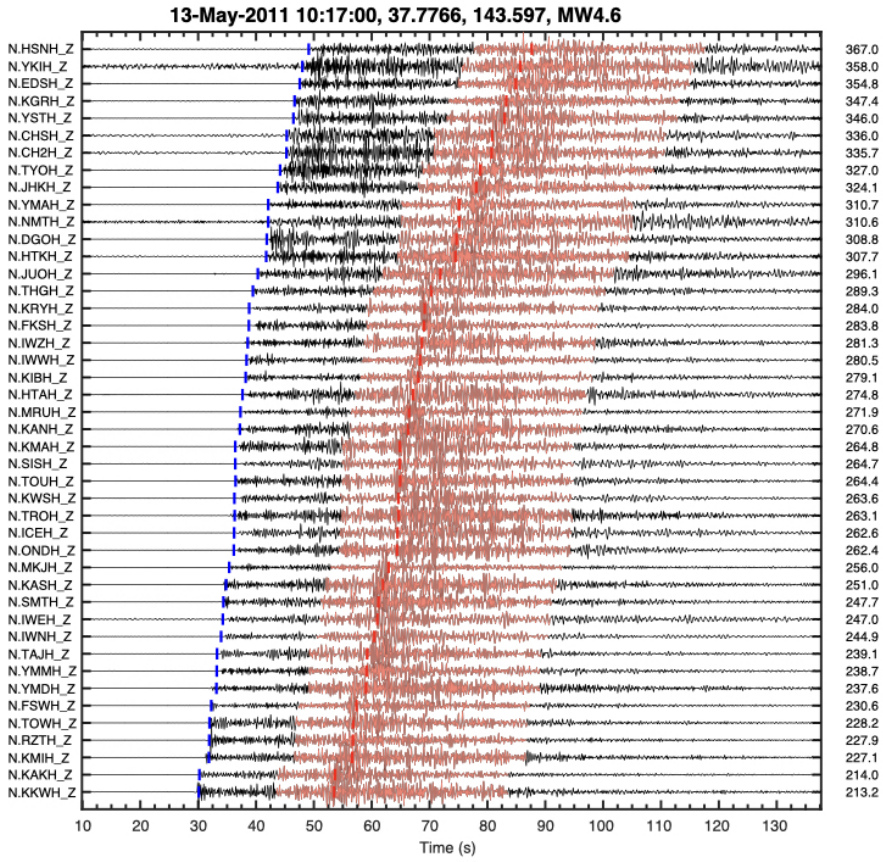

Figure A.11: Waveforms of an Mw 4.6 near-trench event recorded by 47 Hi-net stations. Recorded P and S wave trains (salmon color) are close to theoretical P (blue) and S (red) arrivals predicted by the IASP91 model.

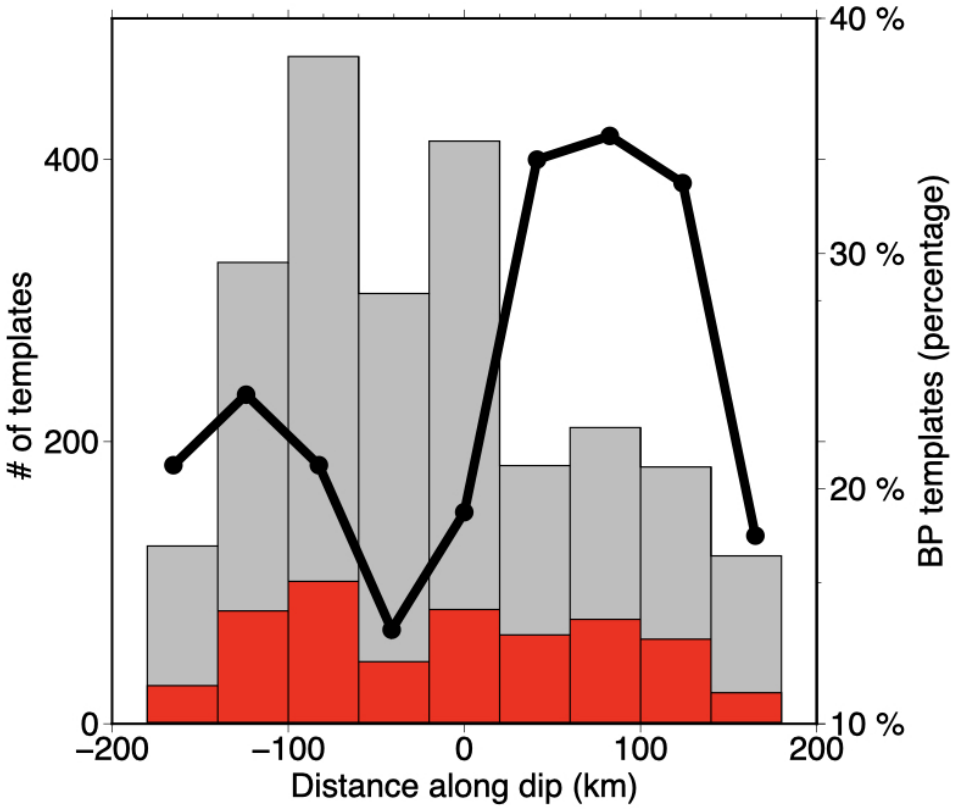

Figure A.12: The relationship between the number of BP (red columns) and JMA (gray columns) templates and distance along the dip. The percentage of BP templates in total templates are also presented (black line).

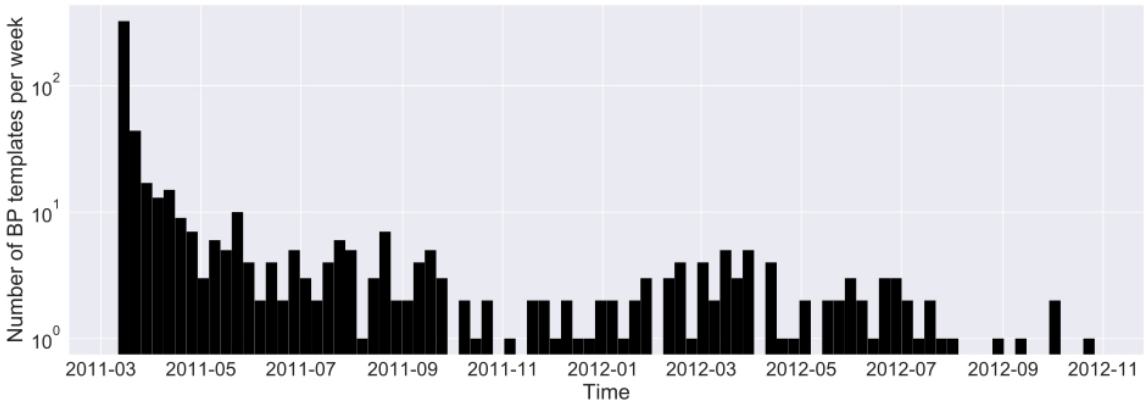

Figure A.13: The occurrence of BP templates in the period within 600 days after the 2011 Mw 9.0 Tohoku earthquake.

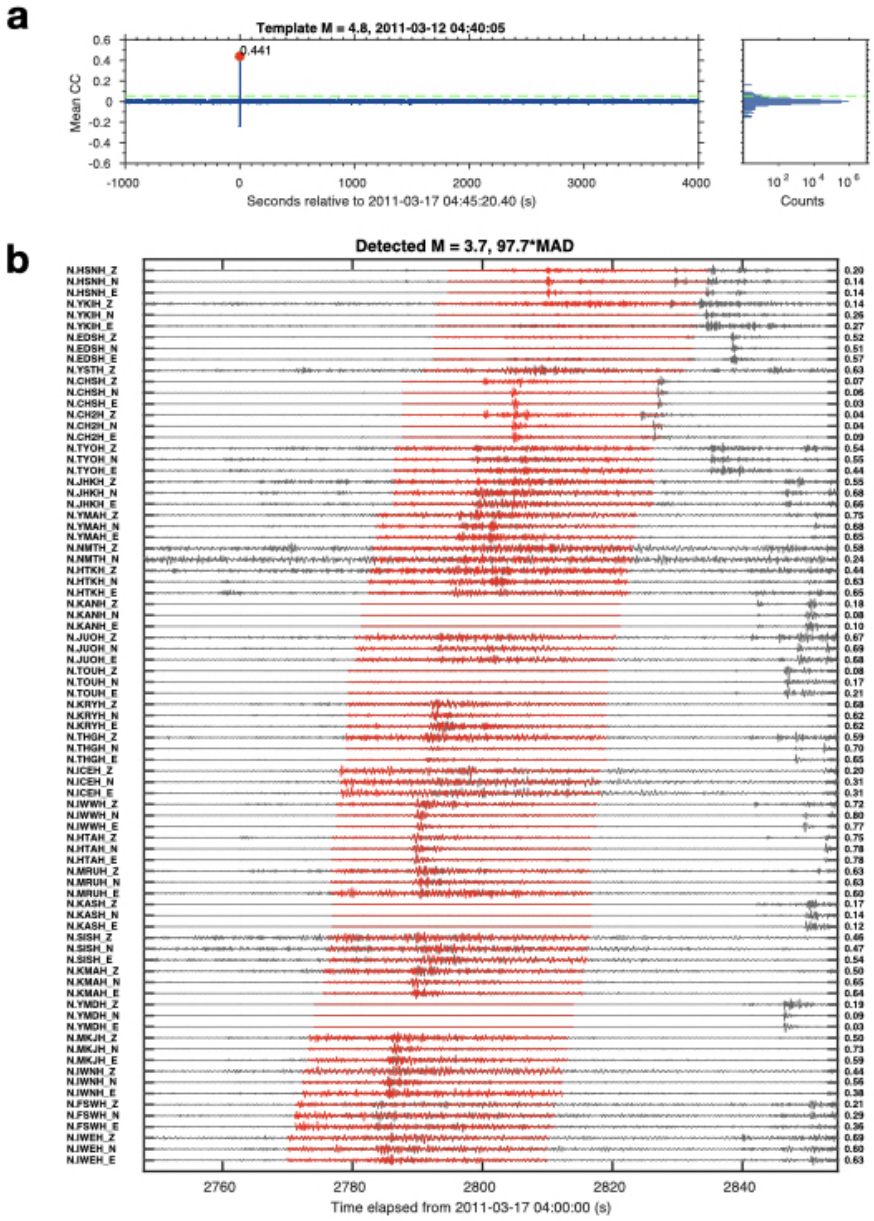

Figure A.14: An example of the matched-filter detection with 47 stations. (a) The left panel shows the stacked CC trace around the detected event 2011-03-17, 04:45:20.40, which is not listed in the JMA catalog. The right panel shows the histogram of the mean CC trace of the day. The green dashed line denotes the detection threshold (9 times the MAD). (b) shows the template waveforms (red) superimposed at the position of matched continuous waveforms (gray). The channel name and the corresponding CC values are labeled on the left and right sides, respectively.

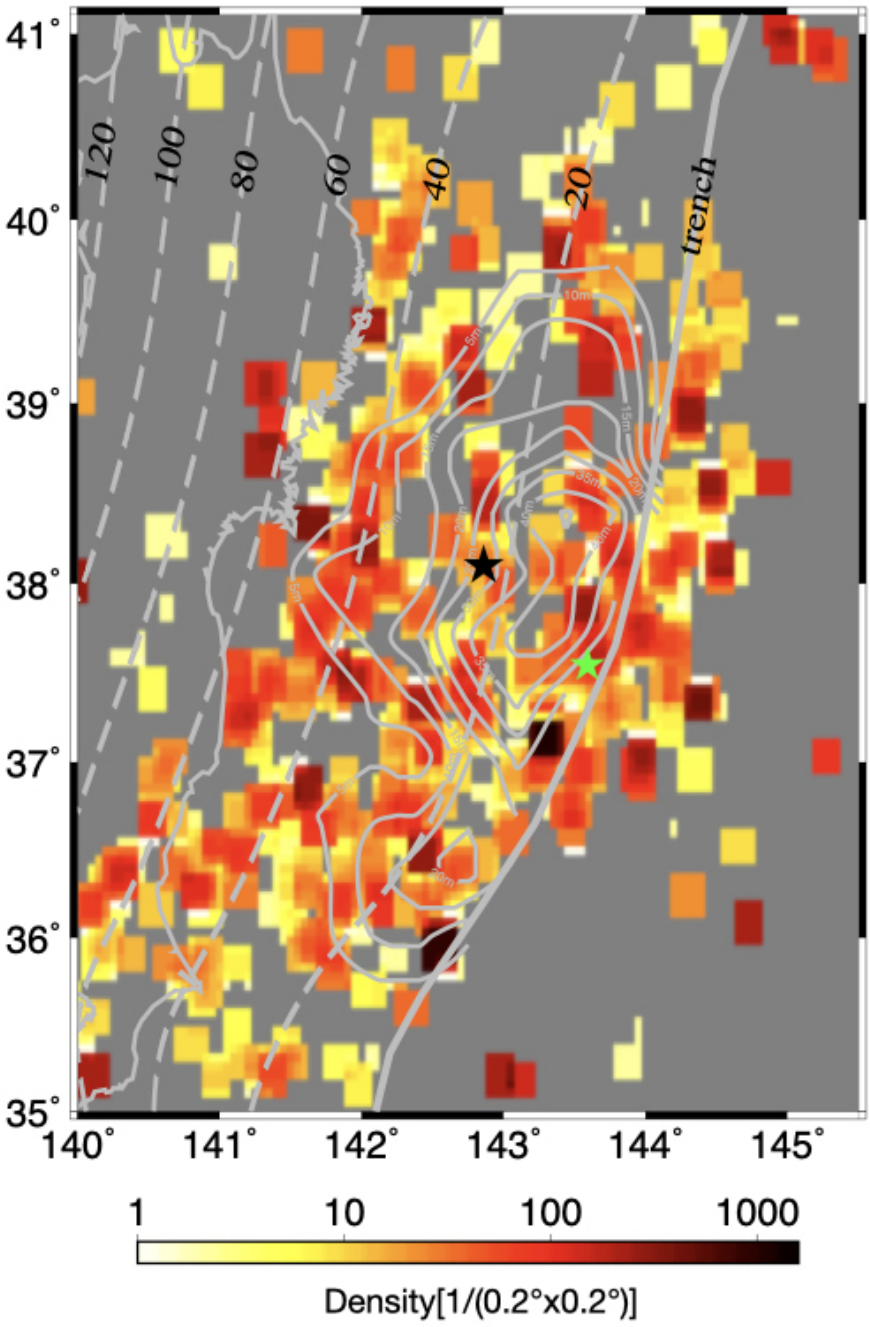

Figure A.15: The density of newly detected events (per $0.2° \times 0.2°$ grid) within 600 days after the mainshock by the MF detection using BP templates. The green star in the outer rise denotes the location of the template event shown in Fig. A.14, while the black star shows the epicenter of the M 9.0 mainshock. The gray dashed lines represent iso-depth contours of the plate boundary at 20-km intervals.

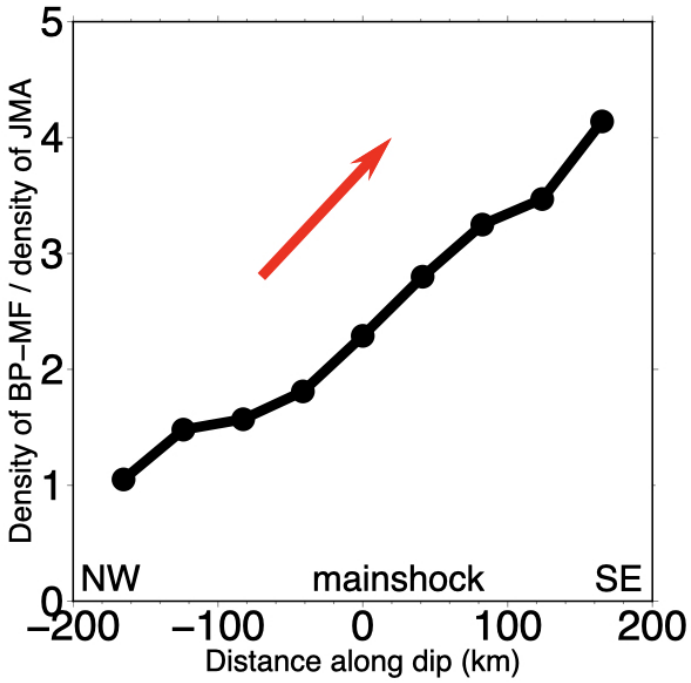

Figure A.16: The ratio of seismicity density between the BP-MF catalog and JMA catalog as a function of along-dip distance.

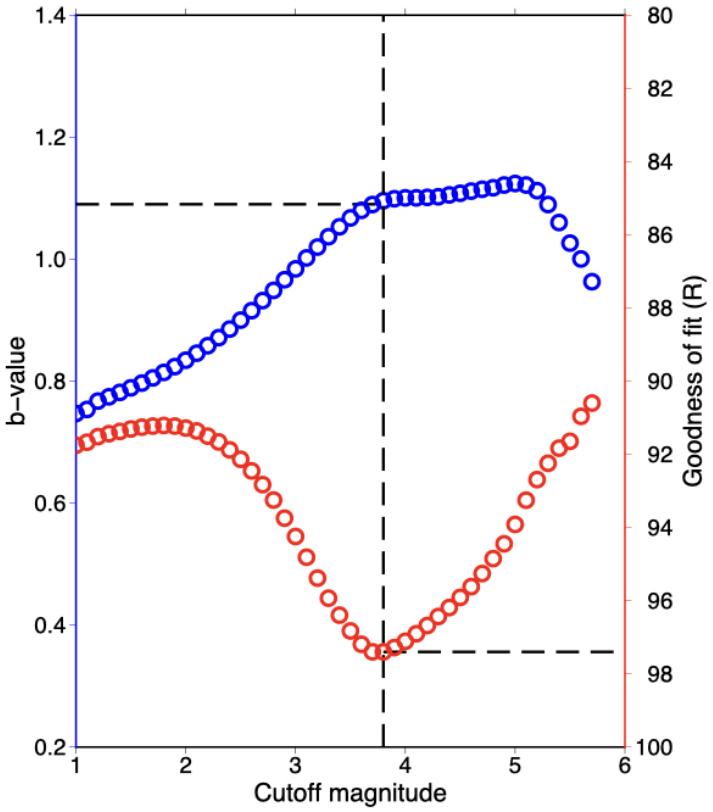

Figure A.17: An example showing the grid search process of b value (blue circles) and magnitude of completeness (M_c). The goodness of fit (R, red circles) achieves the maximum value when M_c and b value equal 3.8 and 1.09, respectively.

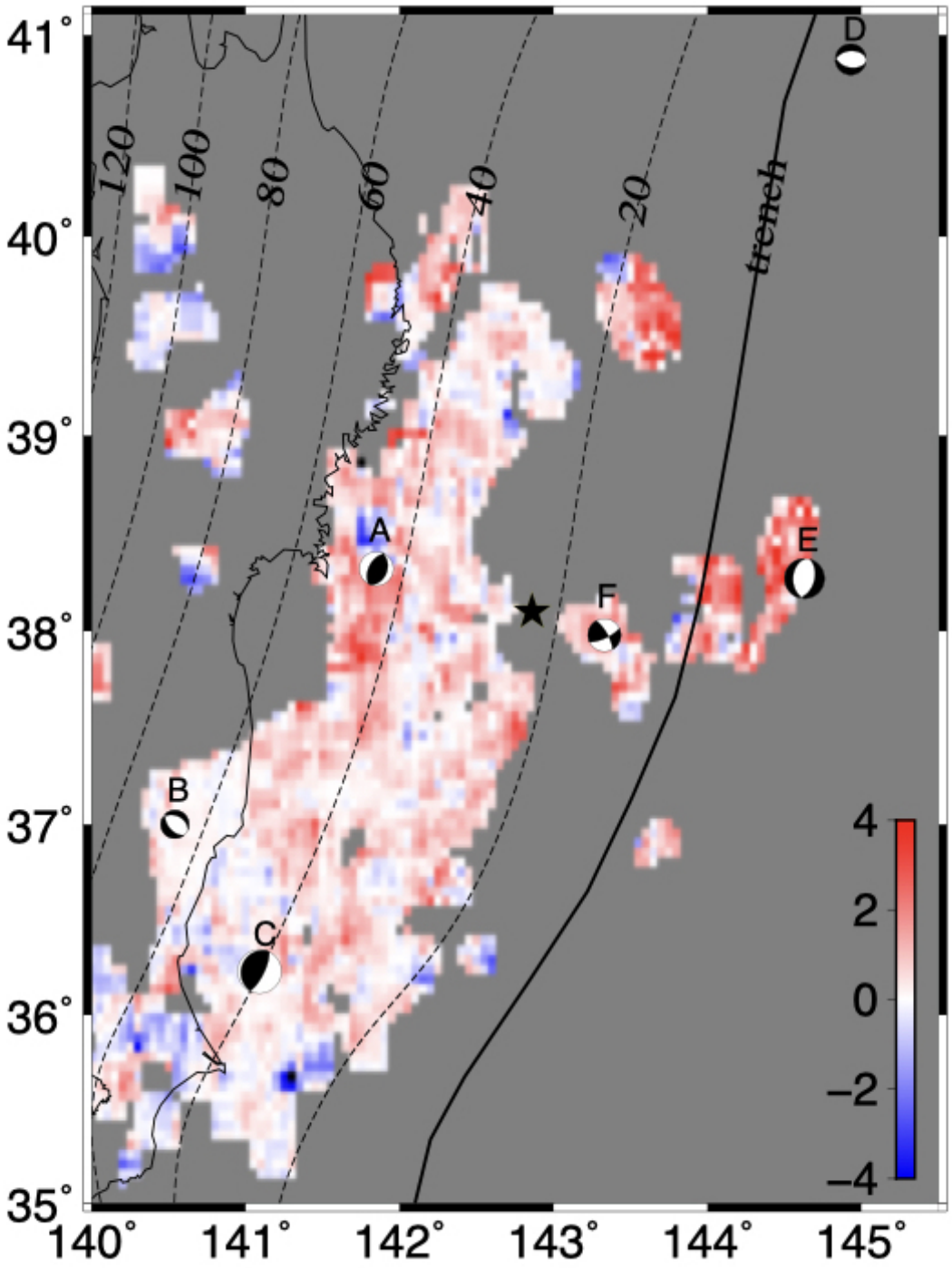

Figure A.18: The difference of goodness of fit (R) between BP-MF and JMA catalog in fitting the frequency-magnitude distribution (BP-MF minus JMA). The Positive value indicates that the BP-MF catalog fits Gutenberg-Richter law better than the JMA catalog.

APPENDIX B

Supplementary Materials of Chapter 3

B.1 Text S1 Earthquake magnitude calculatation

To calculate earthquake magnitude, we first simulate the displacement waveform on a Wood-Anderson seismograph (in meters), trim to the event window (5 sec before the arrival of theoretical P-phase arrival and 45 sec after), and calculate the zero-to-peak amplitude *amp* according to the larger absolute maximum of the two horizontal components (Hutton & Boore, 1987; Tan, 2013). Afterwards, we estimate the local magnitude M_L with epicentral distance d_{epi} and zero-to-peak amplitude based on an empirical function for each station. The median magnitude is used for the final estimation.

$$M_L = log_{10}(amp \cdot 1000\frac{mm}{m}) + 0.018 \cdot d_{epi} + 2.17 \qquad d_{epi} < 60km \tag{B.1}$$

$$M_L = log_{10}(amp \cdot 1000\frac{mm}{m}) + 0.0038 \cdot d_{epi} + 3.02 \qquad d_{epi} \geq 60km \tag{B.2}$$

B.2 Text S2 The parameters setting for REAL, VELEST, and HypoDD

In the REAL algorithm, the search grid size is set to 0.03 degrees in latitude and longitude direction and 5 km in the depth direction. The earthquake detection threshold is either at least 4 P-phase picks or at least 8 P- and S-phases in total. Under these standards, we detected and roughly located 27,007 events.

In the VELEST algorithm, the maximal epicentral distance of 400 km is used to filter the seismic phases. The observations at stations with a distance greater than 400 km are ignored. The weighting factor for S-phase relative to P-phase is 0.75, and the Vp/Vs ratio is set to 1.73. The local velocity model has been adapted from Ozer, Gok, and Polat (2018). The inversion process is composed of 9 forward and backward simulation iterations. Finally, it deletes the events with a station gap greater than 180 degrees or inversion residual greater than 0.5 seconds. As a result, we deleted 24,242 events and kept 2,765 events.

In the HypoDD algorithm, we limit the research area to the longitude between 26.0 E and 27.4 E and the latitude between 37.5 N and 38.3 N. We set the maximum distance between cluster centroid and station as 400 km; maximum hypocentral separation as 100 km; the maximum number of neighbors per event as 20; and the minimum number of links required to define a neighbor as 1. For the event clustering, we set the minimum number of observation pairs to be 4. According to these standards, we relocated 687 events out of 1,222 events.

Table B.1: Performance of EdgePhase, Baseline-A, and Baseline-B models on the SCSN2021 test set. The best performance is highlighted in bold.

	Phase	Precision	Recall	F1	Mean(s)	Std(s)	MAE(s)	Threshold
Baseline-B	P-Phase	0.82	0.55	0.66	-0.02	**0.10**	**0.06**	0.05
	S-Phase	0.77	0.53	0.62	**-0.01**	0.15	**0.10**	0.03
Baseline-A	P-Phase	0.89	0.80	0.84	**0.00**	**0.10**	**0.06**	0.64
	S-Phase	0.84	0.78	0.81	0.03	**0.13**	**0.10**	0.40
EdgePhase	P-Phase	**0.92**	**0.85**	**0.88**	0.03	0.11	0.07	0.39
	S-Phase	**0.89**	**0.83**	**0.86**	-0.04	0.14	0.11	0.28

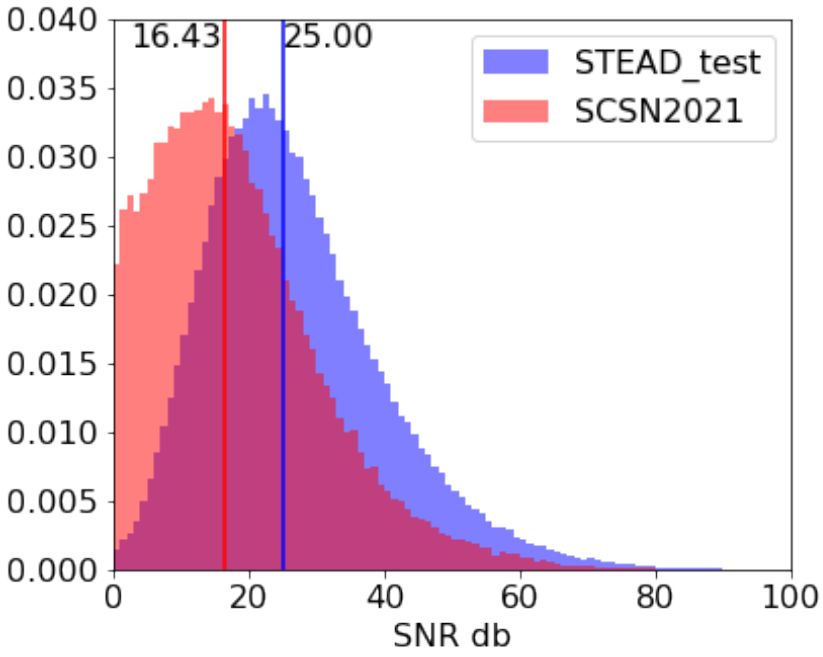

Figure B.1: The signal-to-noise ratio (db) density distribution of STEAD test set and SCSN2021 test set. The median value of SNR for STEAD (red vertical line) and SCSN2021 (blue vertical line) test sets are 16.43 and 25.00 db.

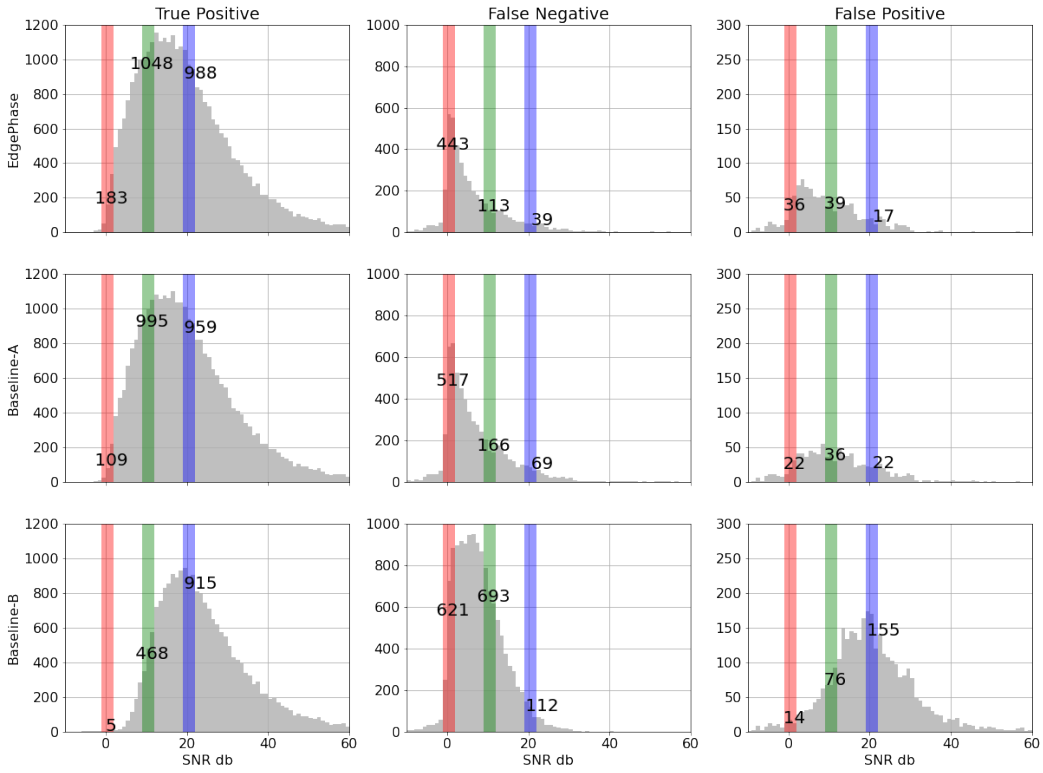

Figure B.2: The SNR (db) density distribution of True Positives (TP, matched detections), False Negatives (FN, missed phases), and False Positives (FP, false detections) counts for P-phases on the SCSN2021 test set generated by EdgePhase, Baseline-A, and Baseline-B. The selected low (0 db), medium (10 db), and high (20 db) SNR bands (width ± 1 db) are represented by the red, green, blue color bands.

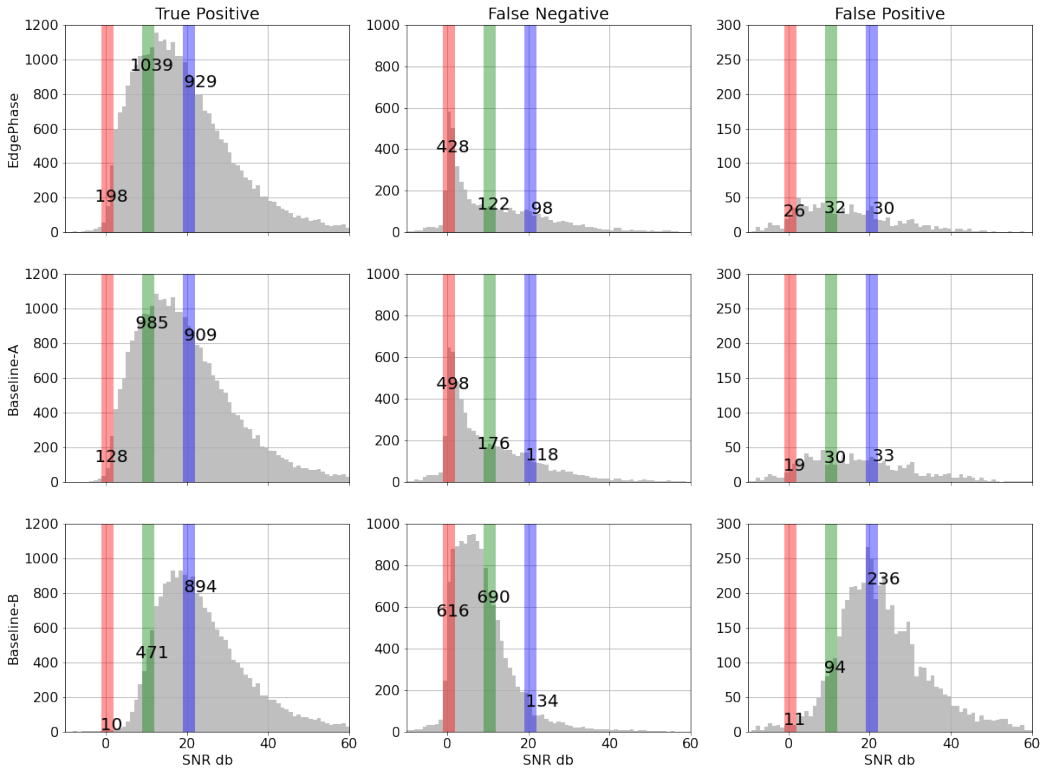

Figure B.3: The SNR (db) density distribution of True Positives (TP, matched detections), False Negatives (FN, missed phases), and False Positives (FP, false detections) counts for S-phases on the SCSN2021 test set generated by EdgePhase, Baseline-A, and Baseline-B. The selected low (0 db), medium (10 db), and high (20 db) SNR bands (width ± 1 db) are represented by the red, green, blue color bands.

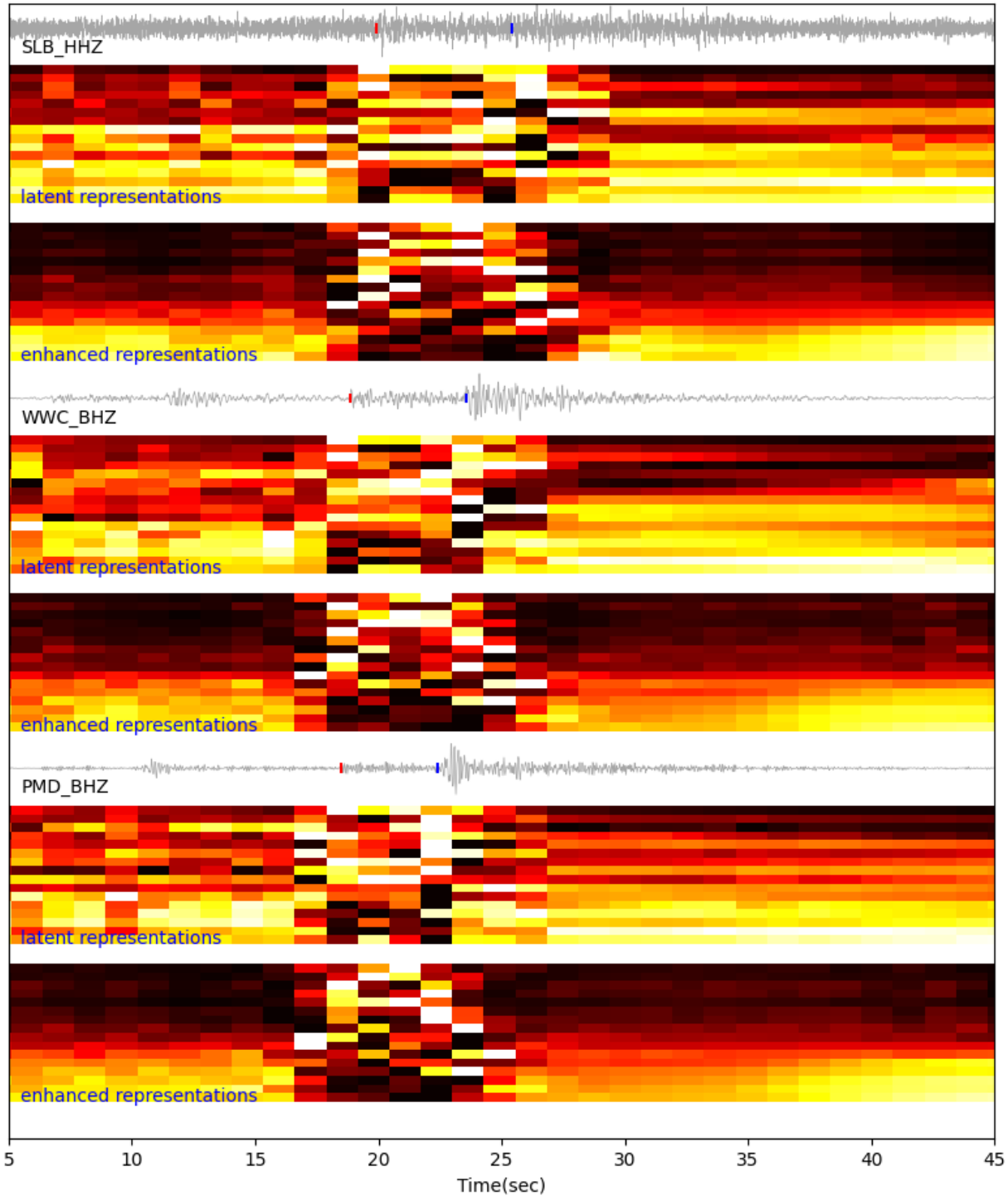

Figure B.4: (Caption next page.)

Feature map of latent representations and enhanced representations in EdgePhase. The earthquake example is the same as Figure 3.2. We only visualize the vertical seismogram of 3 neighboring stations (SLB, WWC, and PMD), since the total channel number (219) is too large to visualize all in one figure. The red and blue dots on the waveforms (gray curves) represent the ground truth P- and S-phases. We process the feature map with min-max normalization, more specifically, the minimum value of that feature gets transformed into a 0 (black), the maximum value gets transformed into a 1 (white), and every other value gets transformed into a decimal between 0 and 1. Finally, we sorted the feature map in ascending order based on the mean value of the feature.

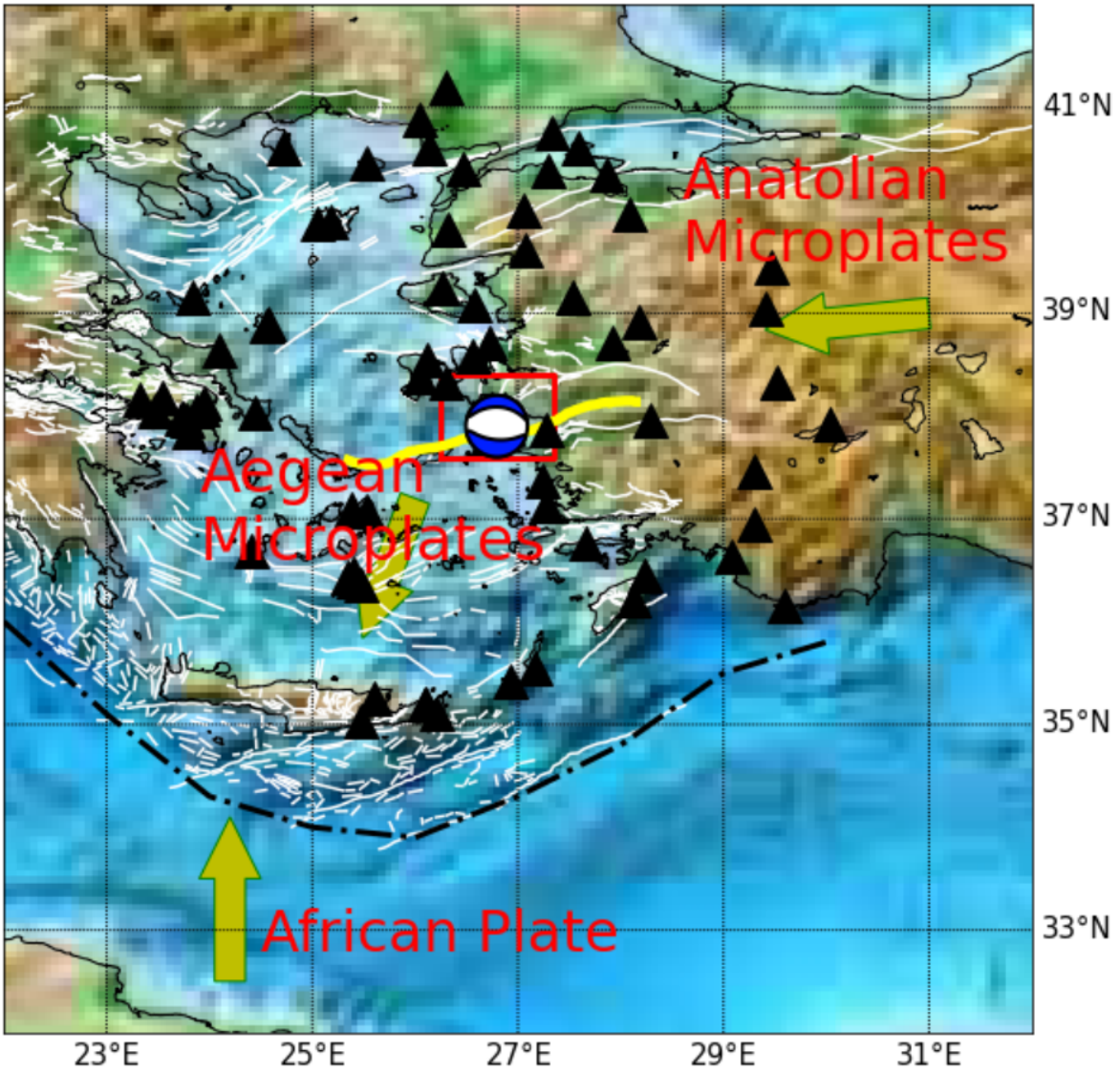

Figure B.5: Tectonics in the Aegean Sea with movement directions of African Plate, Aegean Microplates, and Anatolian Microplates. The focal mechanism (red stars) and faults (white curves) are from AFAD and the New Seismotectonic Atlas of Greece (Kassaras et al., 2020), respectively. The study region is highlighted by a red box. We select 72 stations around the Aegean Sea from 6 networks (HA, HC, HI, HL, HT, KO) and 2 different types of channels (HH: High gain and HN: accelerometer) with a 100 Hz sampling rate. The distribution of stations shows good azimuthal coverage of the 2020 Mw 7.0 Samos earthquake (red star).

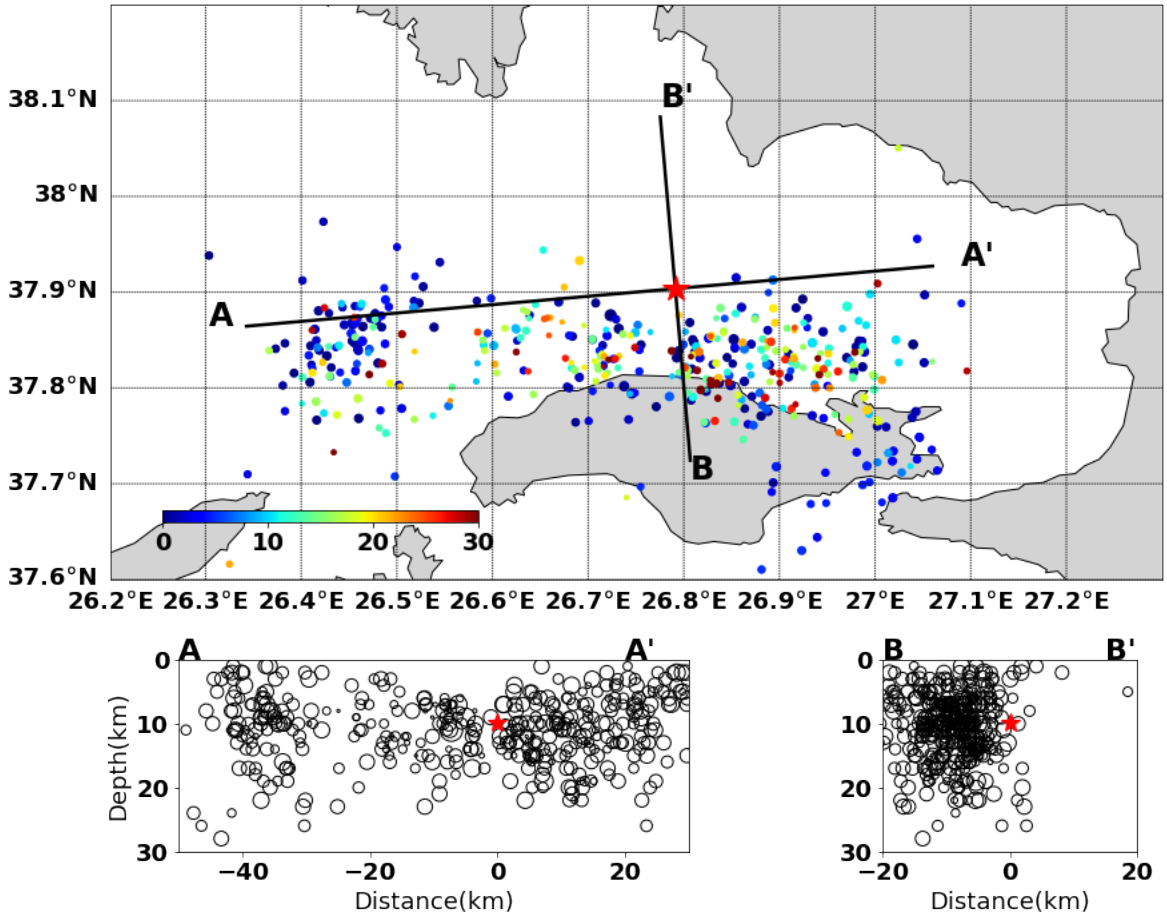

Figure B.6: Map view and vertical cross-sections of seismicity from NKUA. Empty circles are the events from the earthquake catalog from the Seismological Laboratory of National and Kapodistrian University of Athens (SL-NKUA), 421 in total. The mainshock hypocenter is highlighted by red stars. The color of circles in the top panel indicates the days after the mainshock.

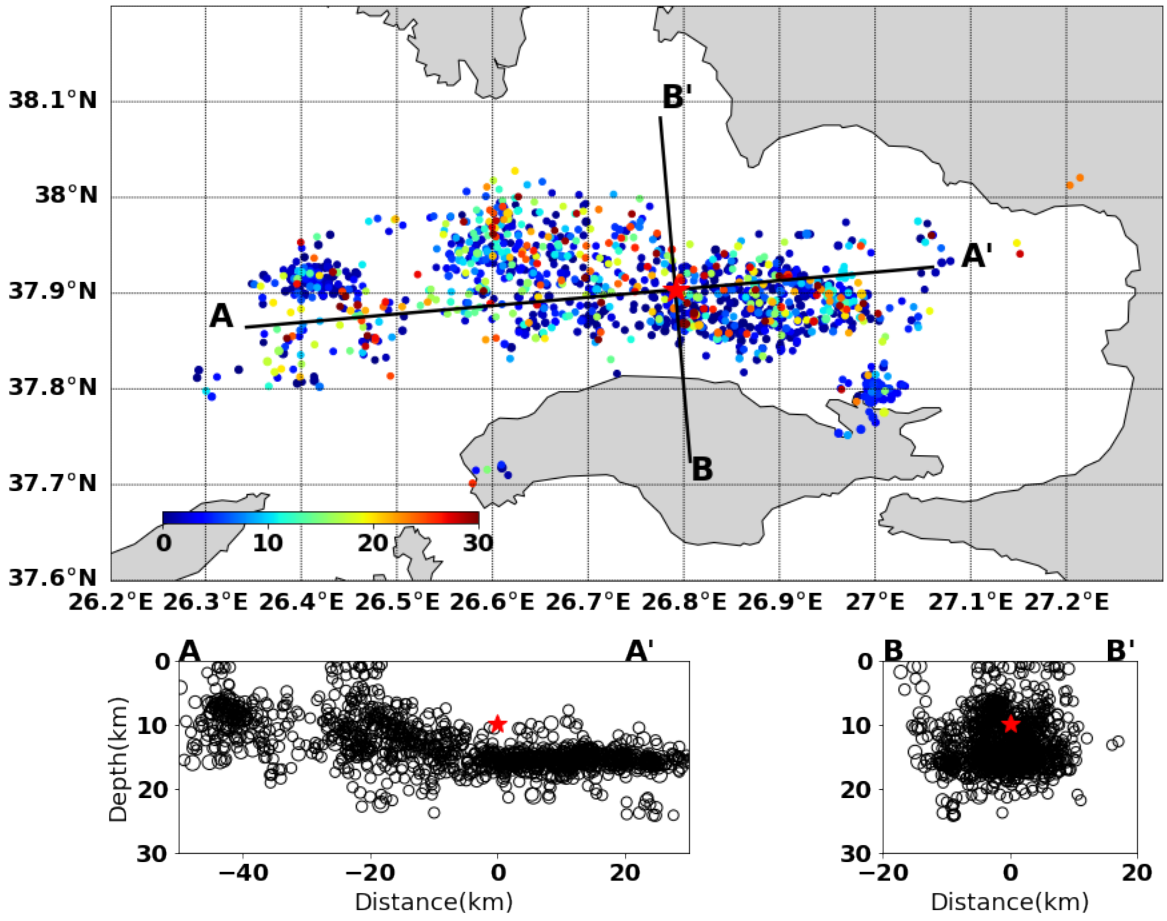

Figure B.7: Map view and vertical cross-sections of seismicity from this study. Empty circles are the events detected with EdgePhase, relocated by VELEST and HypoDD methods. The total number of events is 1222, 687 of them are relocated by HypoDD. The mainshock hypocenter is highlighted by red stars. The color of circles in the top panel indicates the days after the mainshock.

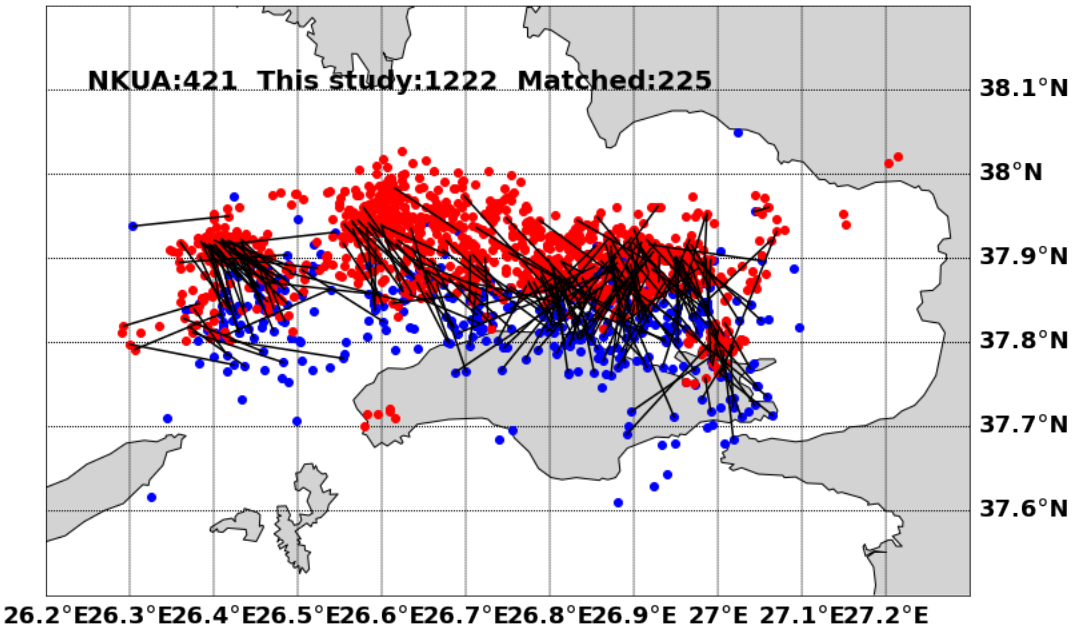

Figure B.8: Map view of matched events. Two events from the NKUA catalog (blue) and the catalog of this study (red) are matched (black lines) if the temporal and spatial difference is smaller than 3 sec and 30 km.

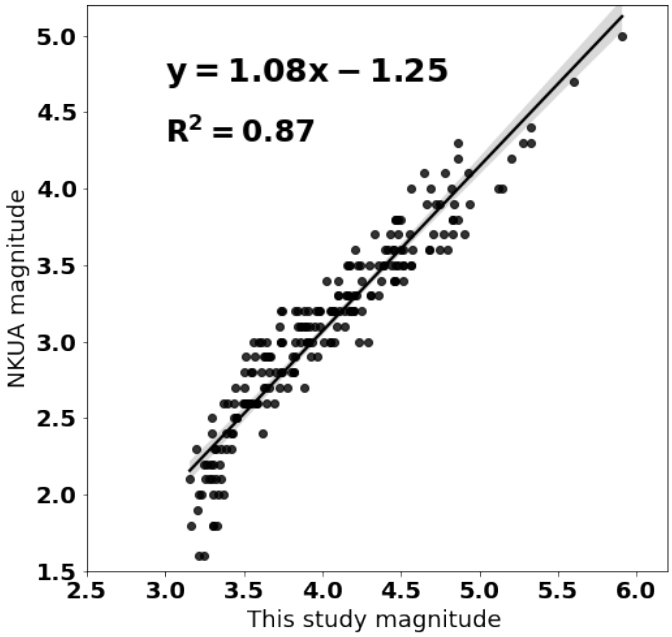

Figure B.9: Linear regression of magnitude between matched event pairs. The coefficient of determination (R^2) is 0.87.

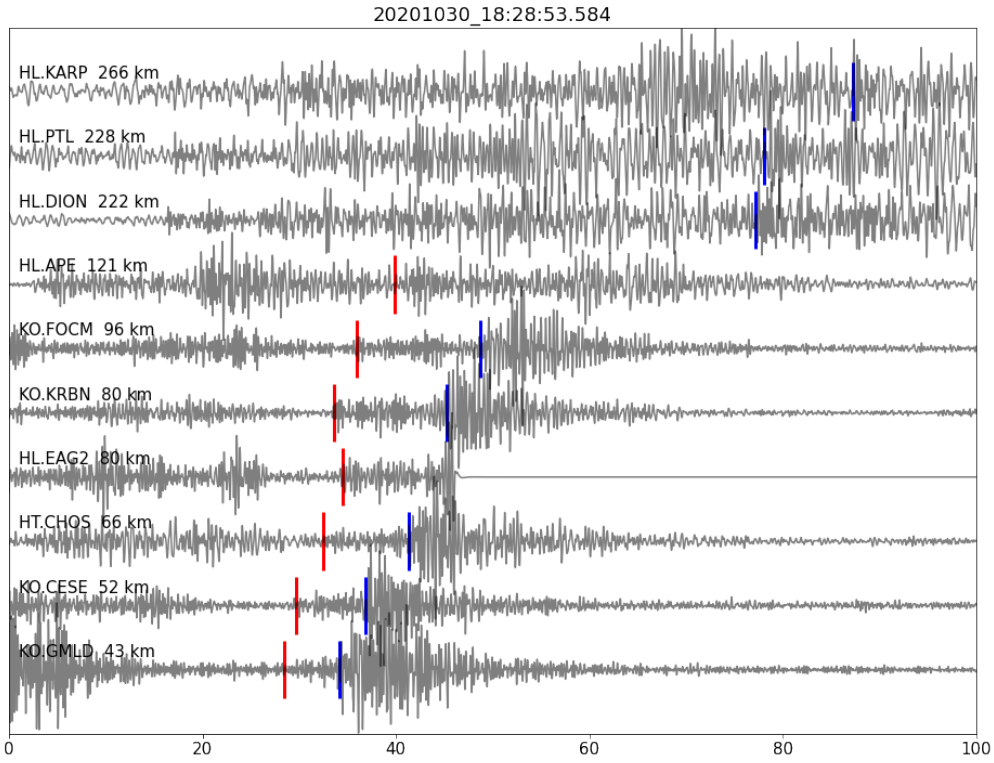

Figure B.10: An example of a newly detected M 3.1 event by EdgePhase. The relocated hypocenter is 37.840 N, 26.434 E with a depth of 21 km. The detected 7 P- and 8 S- phases are represented by red and blue vertical bars.

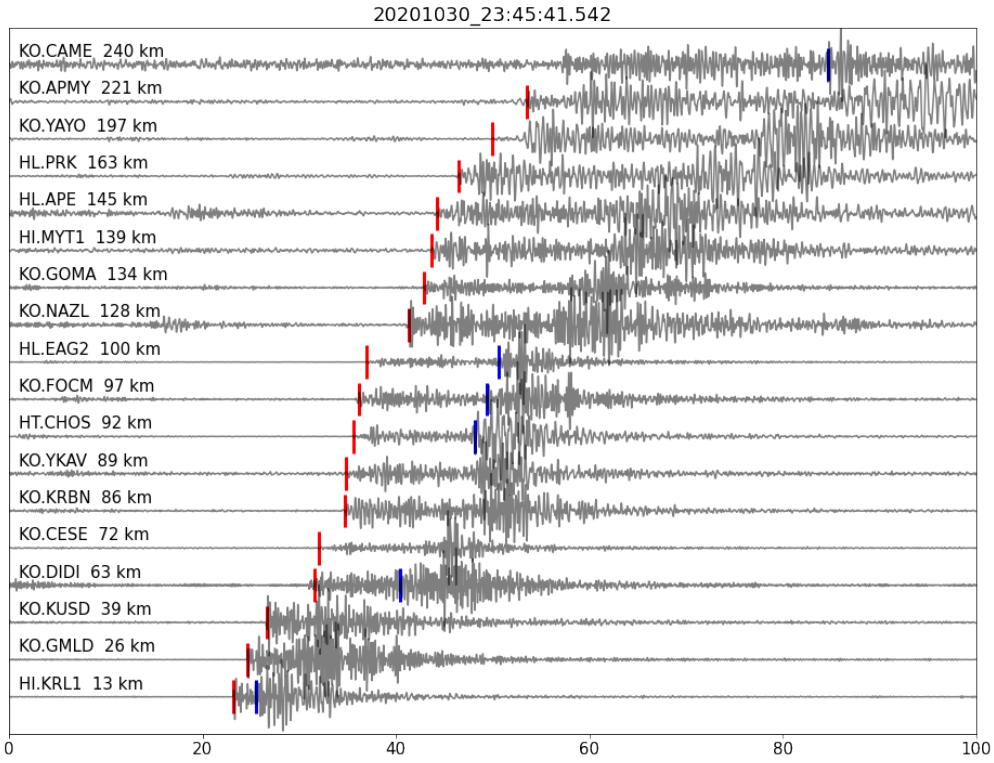

Figure B.11: An example of a newly detected M 3.6 event by EdgePhase. The relocated hypocenter is 37.902 N, 26.837 E with a depth of 14 km. The detected 17 P- and 6 S- phases are represented by red and blue vertical bars.

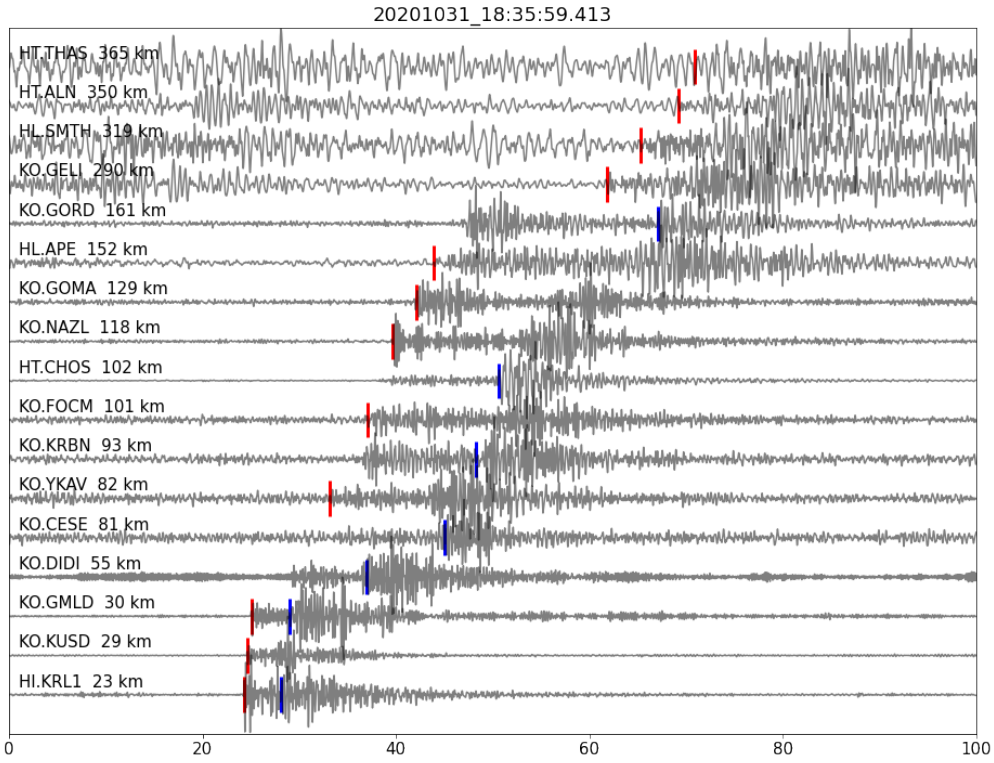

Figure B.12: An example of a newly detected M 2.9 event by EdgePhase. The relocated hypocenter is 37.880 N, 26.979 E with a depth of 15 km. The detected 12 P- and 7 S- phases are represented by red and blue vertical bars.

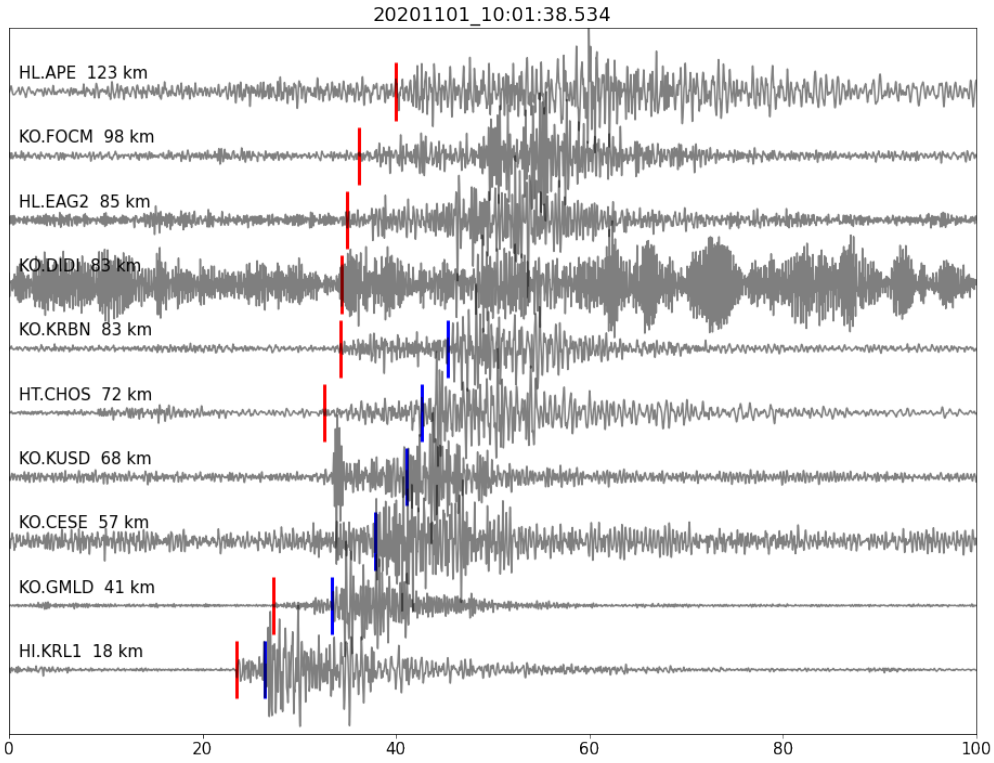

Figure B.13: An example of a newly detected M 3.0 event by EdgePhase. The relocated hypocenter is 37.815 N, 26.979 E with a depth of 14 km. The detected 8 P- and 6 S- phases are represented by red and blue vertical bars.

APPENDIX C

Supplementary Materials of Chapter 4

C.1 Text S1. Site Attenuation Analysis

Site attenuation term associated with local ground response or basin response can be evaluated with the amplification factors $F(T)$, which is defined as the ratio of an observed ground-motion intensity to the intensity measurement of a reference site (often the rock type) at similar source distances, (Stewart et al., 2013):

$$F(T) = PSa_{obs}(T)/PSa_{ref}(T) \tag{C.1}$$

Where $PSa_{ref}(T)$ is the ground-motion intensity measure for the reference site and $PSa_{obs}(T)$ is the intensity of the observe site at the period T. The site amplification $F(T)$ can be compared with the residual $R(T)$ between observation and GMPE predictions according to the following two arguments. The observed intensity at the reference station $PSa_{ref}(T)$ is equivalent with the GMPE prediction $\hat{PSa_{ref}}(T)$ plus a site attenuation term $Fs_{ref}(T)$).

$$PSa_{ref}(T) = \hat{PSa_{ref}}(T) + Fs_{ref}(T) \tag{C.2}$$

Since the observation station and reference station are at similar distances, the GMPE predications at the two stations are equal: $\hat{PSa_{ref}}(T) = \hat{PSa_{obs}}(T)$. Also, by definition the site attenuation term is zero at the reference site: $Fs_{ref}(T) = 0$. Thus, we can use the estimation at the observation site $\hat{PSa_{obs}}(T)$ by GMPE to represent the value of observation $PSa_{ref}(T)$ at the reference site.

$$PSa_{ref}(T) = \hat{PSa_{ref}}(T) + Fs_{ref}(T) = \hat{PSa_{obs}}(T) \tag{C.3}$$

Taking the natural logarithm of both sides of Equation C.1, we have:

$$ln(F(T)) = ln(PSa_{obs}(T)) - ln(PSa_{ref}(T)) = ln(PSa_{obs}(T)) - ln(\hat{PSa_{obs}}(T)) = R(T) \tag{C.4}$$

Equation C.4 indicates that the GMPE performance can be evaluated by comparing the observed log amplification factor $ln(F(T))$ with the residual $R(T)$ of the GMPE prediction. The site amplification factor $F(T)$ depends on both period T and the average sub-surface shear velocity $Vs30$. It is often factorized as a linear function of $ln(Vs30)$ with the coefficient a and c depending on the period T. (Fig. S4, Borcherdt, 1994; Borcherdt & Glassmoyer, 1989; Barani et al., 2010):

$$R(T) = ln(F(T)) = a(T) + c(T) * ln(Vs30) \tag{C.5}$$

Since $a(T)$ also depends on the choice of the reference station, we focus on the slope $c(T)$ which can be compared across different datasets. If a GMPE adequately considers the source and path effects, the residual $R(T)$ should only account for the site effect. Thus the $c(T)$ in the residual scaling approach should match the parameters in previous studies. Stewart et al. (2013) find that in the Tohoku earthquake the variation of $c(T)$ with the period T is much larger than the previous study in active crustal regions (ACR). They attribute the discrepancies to a region-specific variation of the site response. However, in our work, $c(T)$ is much more consistent with that inferred from ACR data (Fig. C.5, Choi & Stewart, 2005). Our results indicate that the effect of the frequency-dependent distance metric is mapped into the residual calculation in Stewart et al. (2013). When using frequency-dependent distance R_{hf}, the term $c(T)$ is closer to ACR data than that calculated with R_{rup}. This agreement between the observed and inferred site attenuation term indicates that R_{hf} is a better approximation of the distance term. We consider that most of source and path effects are excluded from the term $c(T)$ when using R_{hf} as the distance metric. R_{hf} also properly accounts for the frequency dependent source effect and is thus a better distance metric.

Table C.1: Basic information of three megathrust events in the Japan Subduction zone.

Event	Time	Duration(s)	Mw	Location	Depth(km)	Fault model
Tohoku	2011/03/11	150	9.0	38.10N 142.86E	24	(Wei et al., 2012)
Tokachi	2003/09/26	70	8.0	41.78N 144.07E	42	(Koketsu et al., 2005)
off-Miyagi	2005/08/16	25	7.2	38.15N 142.28E	42	(Shao & Ji, 2013)
Illapel	2015/09/16	100	8.3	31.55S 71.86W	11	(Okuwaki et al., 2017)
Iquique	2014/04/01	80	8.1	19.57S 70.91W	17	(Wei, 2014)

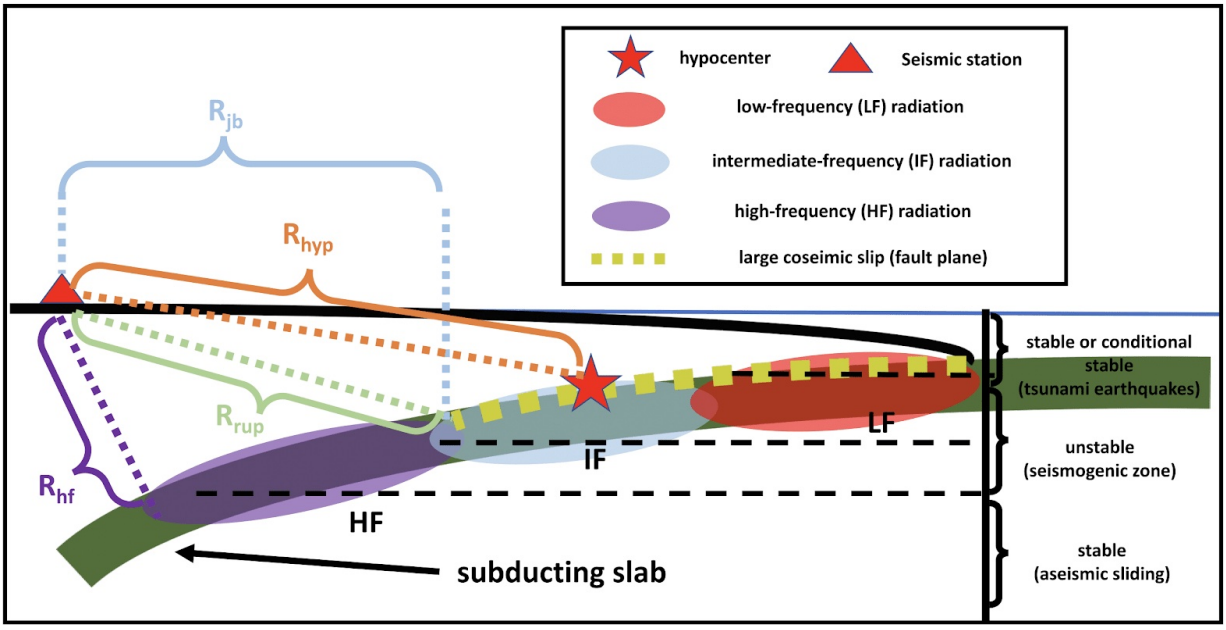

Figure C.1: Schematic diagram of distance metrics with several frequency-dependent radiation domains from Yao et al. (2013); Lay et al. (2012). Red, blue and purple portions represent low-frequency, intermediate-frequency and high-frequency radiation zones along the slab interface accordingly.

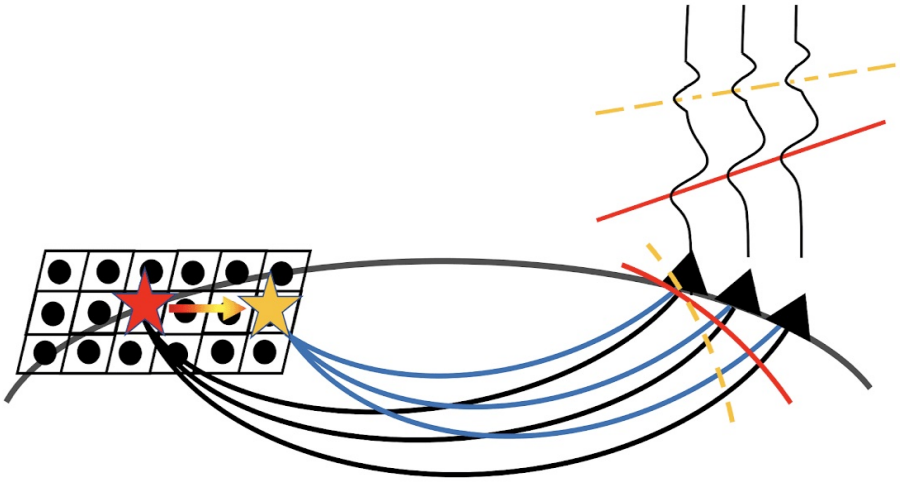

Figure C.2: Sketch shows the integration of Back-Projection. The black dots in the center of the rectangular grids indicate the locations of testing sources in Back-Projection. The red star represents the hypocenter. The moveout of recorded seismograms is shown with the red line. In principle, the moveout of the actual source locations brings the seismograms in phase; thus, the stack along the moveout reaches the maximum. When the earthquake rupture to a new location, represented by the yellow star, the moveout of seismograms move from the red line to the yellow dashed line.

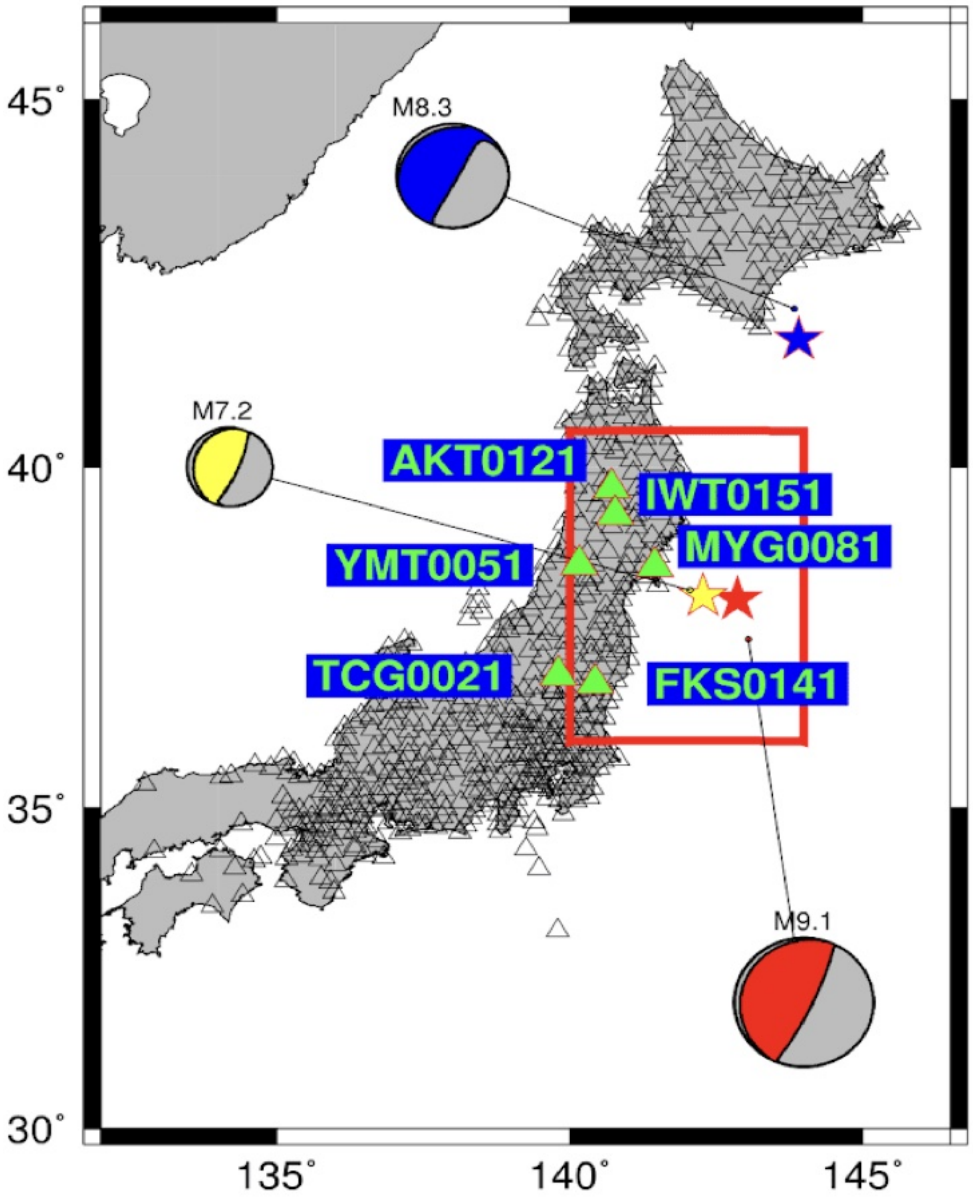

Figure C.3: K-Net stations (black open triangles) used in GMPE analysis of 2011 Mw 9.0 Tohoku-Oki earthquake as well as GCMT (Global Centroid Moment Tensor) resolution of three large megathrust events in the Japan subduction zone. The red rectangle indicates the area in Figure 4.3. Green triangles represent the six stations selected for the analysis.

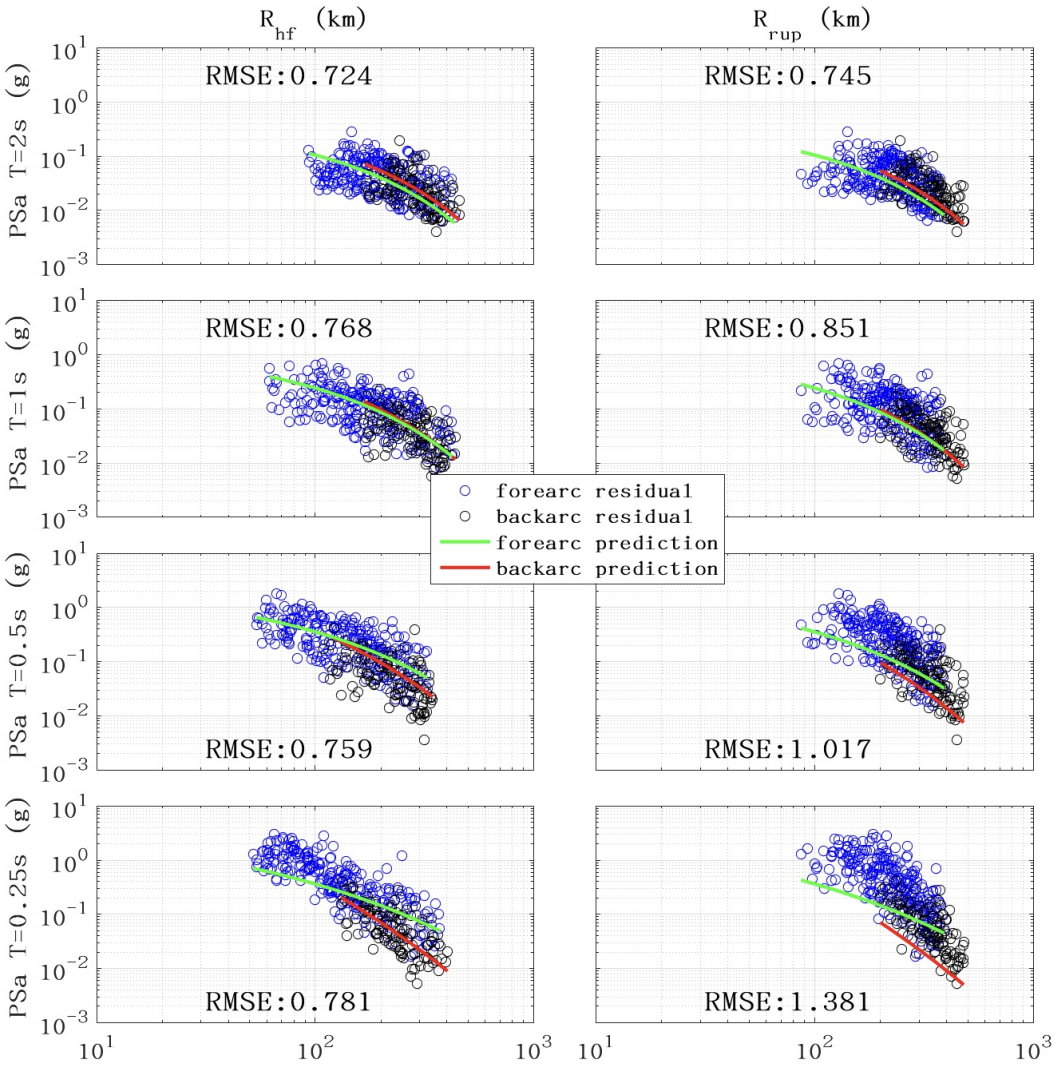

Figure C.4: Comparisons of PSa in different frequencies between R_{hf} metric and R_{rup} metric of 2011 Tohoku earthquake (Mw 9.0).

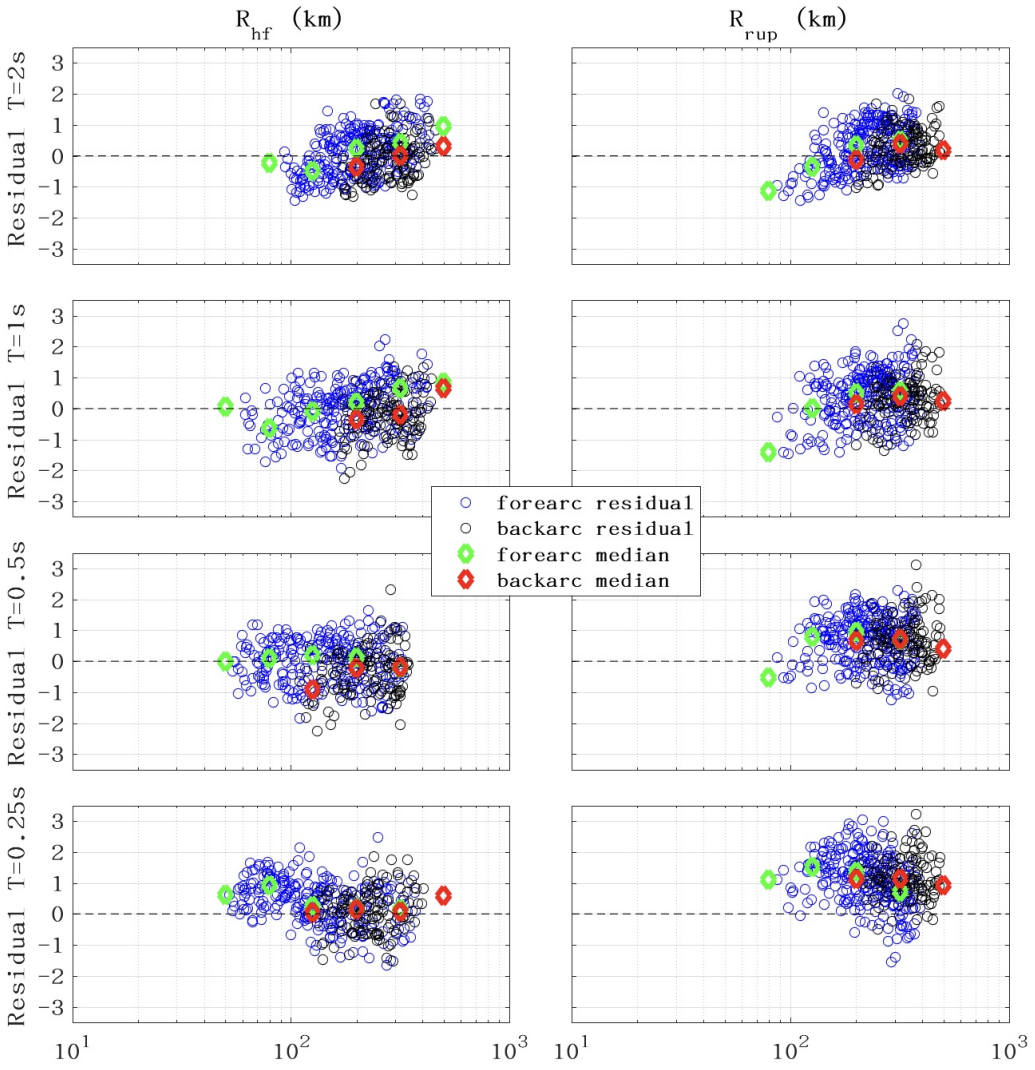

Figure C.5: Comparisons of logarithmic residual in different frequencies between R_{hf} metric and R_{rup} metric of 2011 Tohoku earthquake (Mw 9.0).

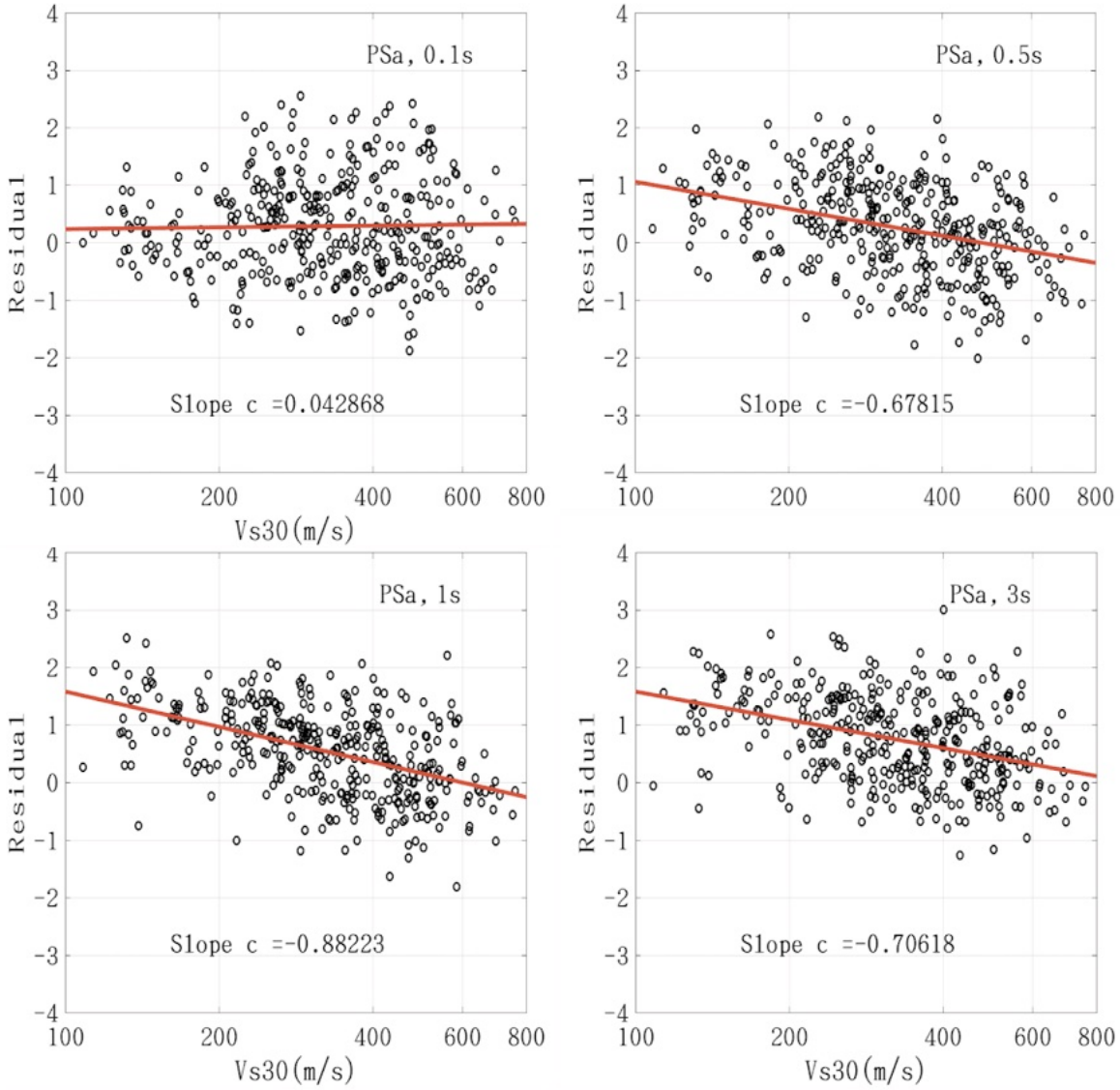

Figure C.6: Scaling relationship between residuals from Tohoku-Oki recordings and logistic Vs30 values at the period of 0.1 s, 0.5 s, 1 s, and 3 s. The red lines are the linear regression fit of residuals (black open dots).

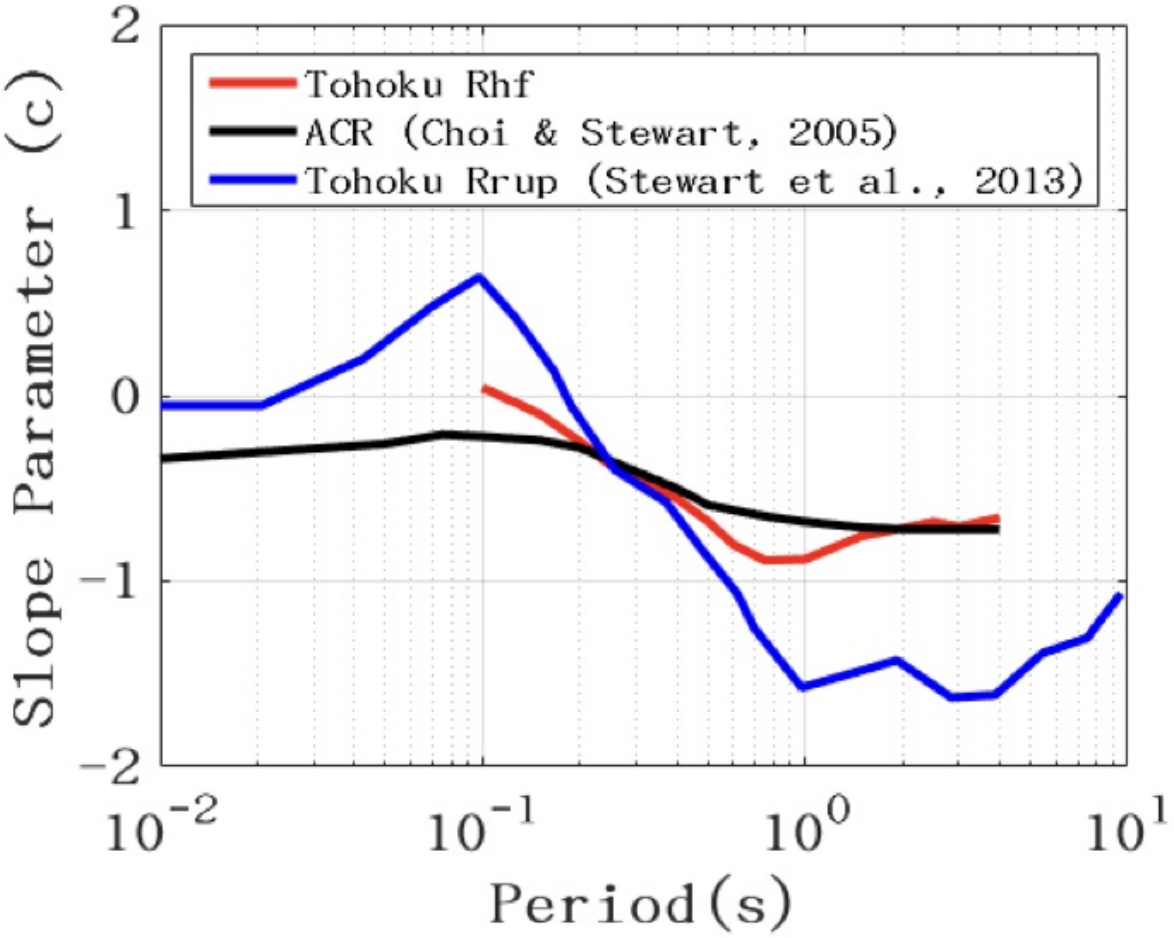

Figure C.7: Vs30-scaling of residuals from Tohoku-Oki recordings (c parameter) along with comparable values for data from active crustal regions (Choi & Stewart, 2005).

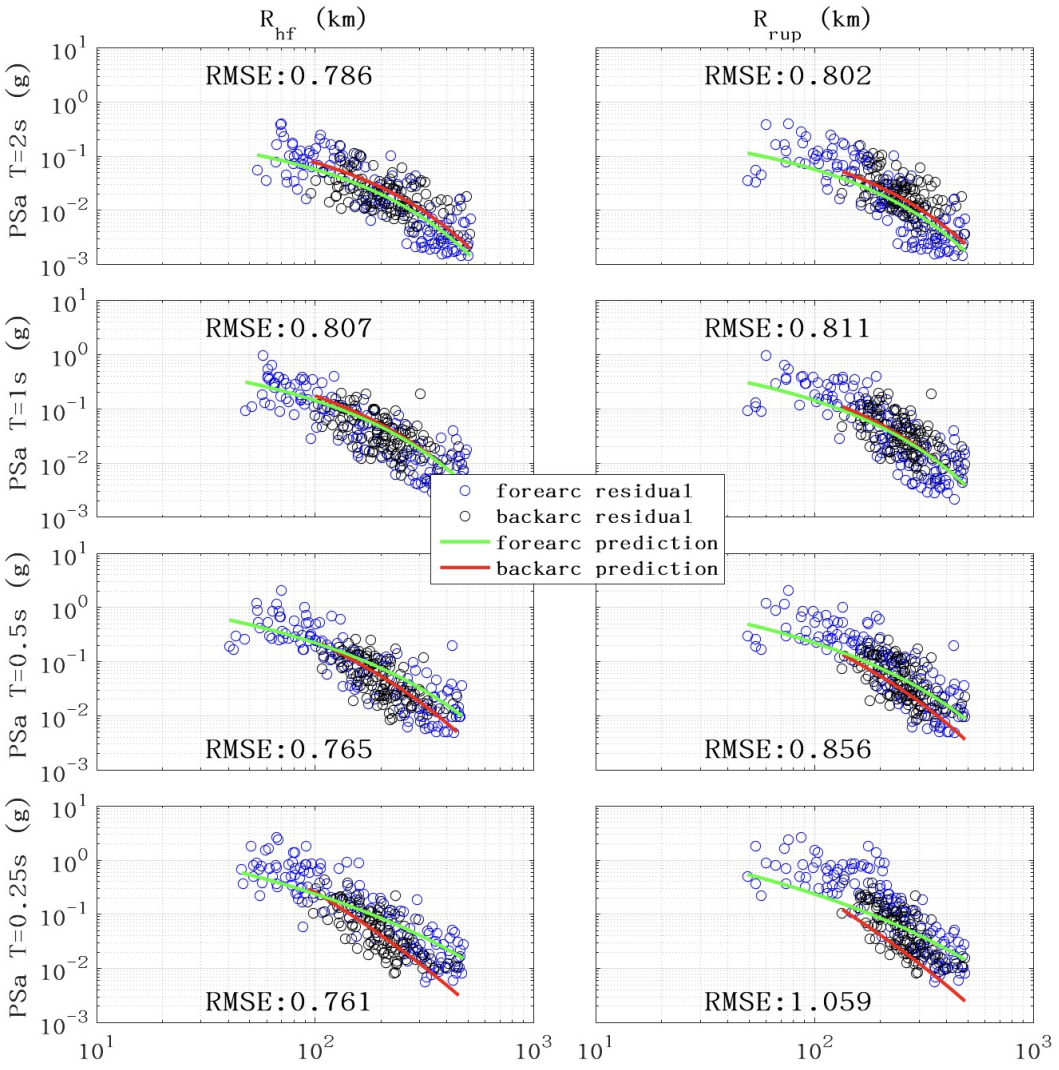

Figure C.8: Comparisons of PSa in different frequencies between R_{hf} metric and R_{rup} metric of 2003 Tokachi earthquake (Mw 8.0).

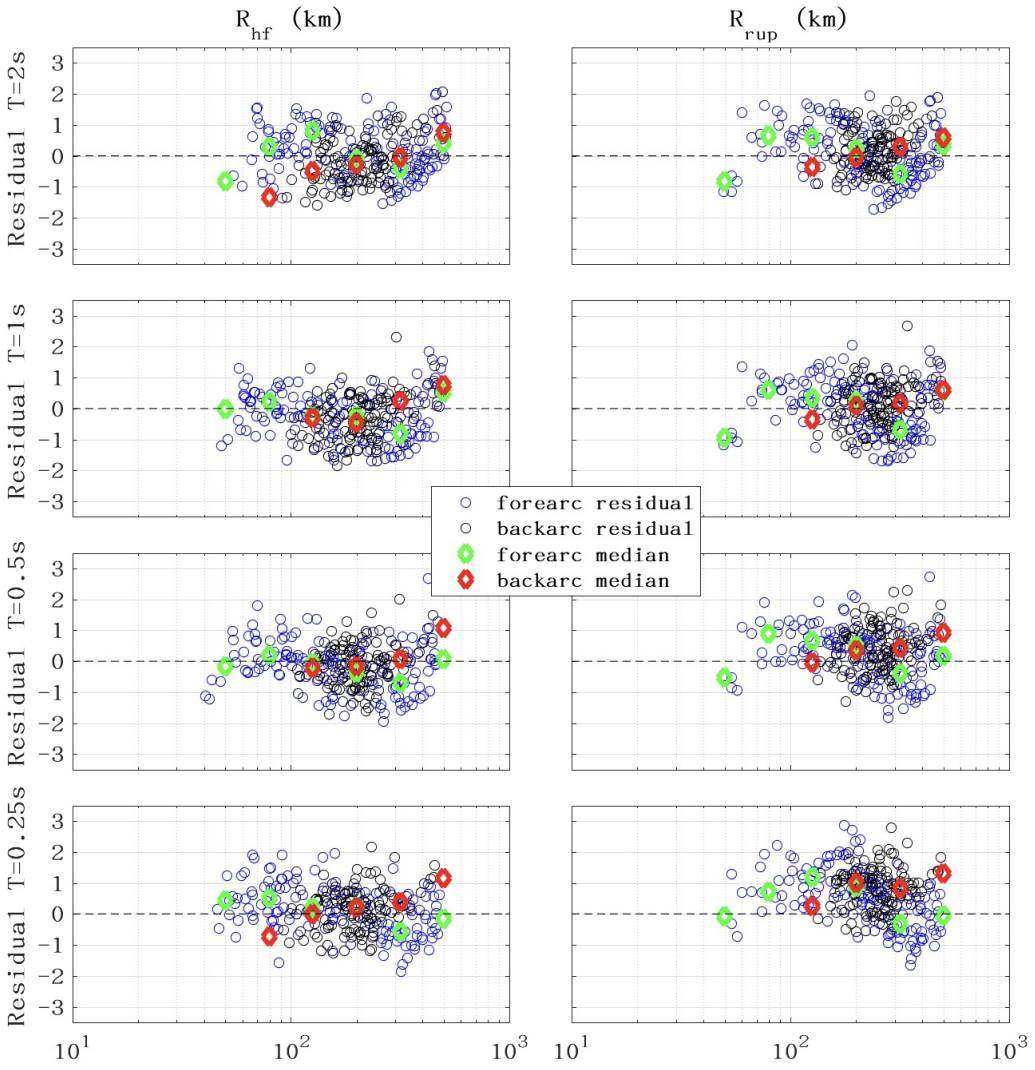

Figure C.9: Comparisons of logarithmic residual in different frequencies between R_{hf} metric and R_{rup} metric of 2003 Tokachi earthquake (Mw 8.0).

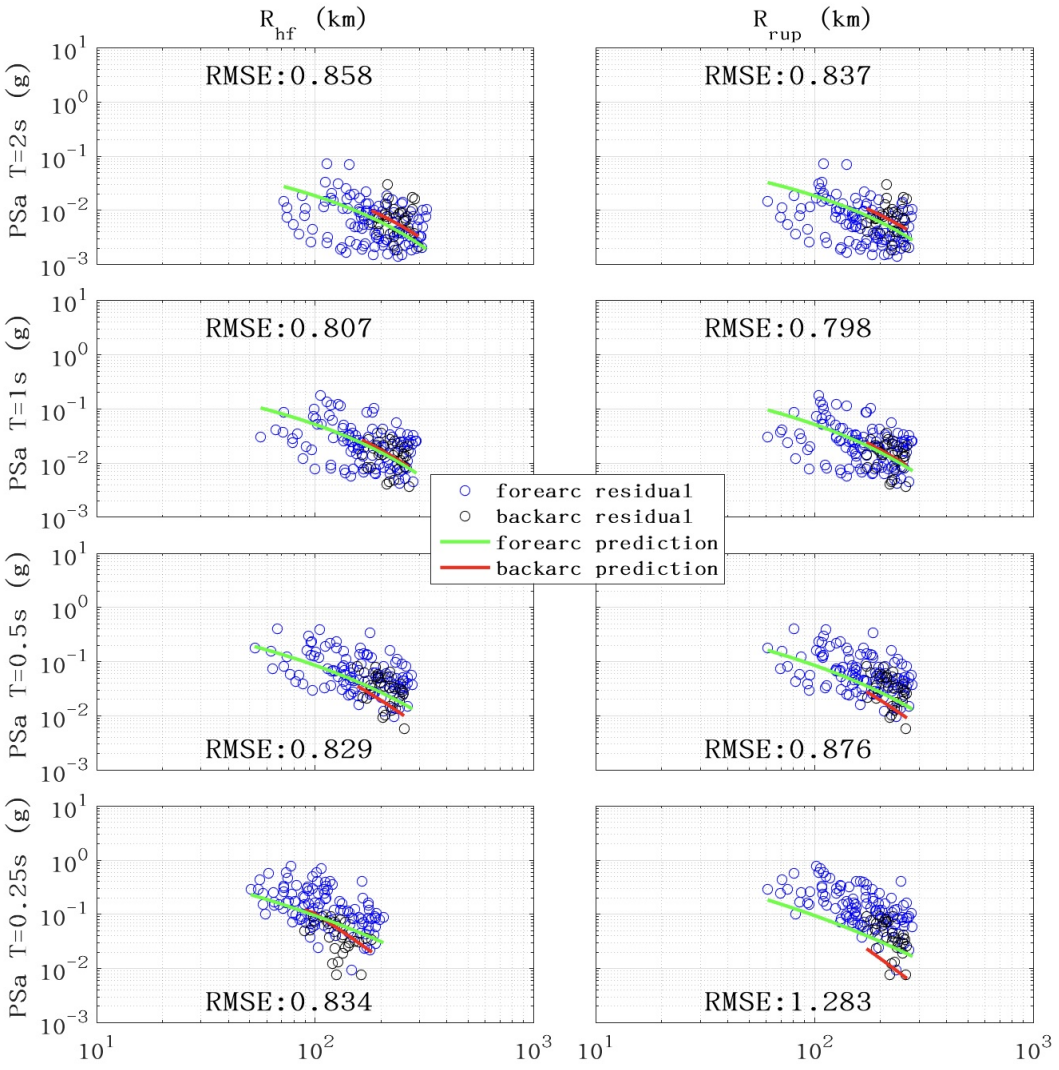

Figure C.10: Comparisons of PSa in different frequencies between R_{hf} metric and R_{rup} metric of 2005 off-Miyagi earthquake (Mw 7.2).

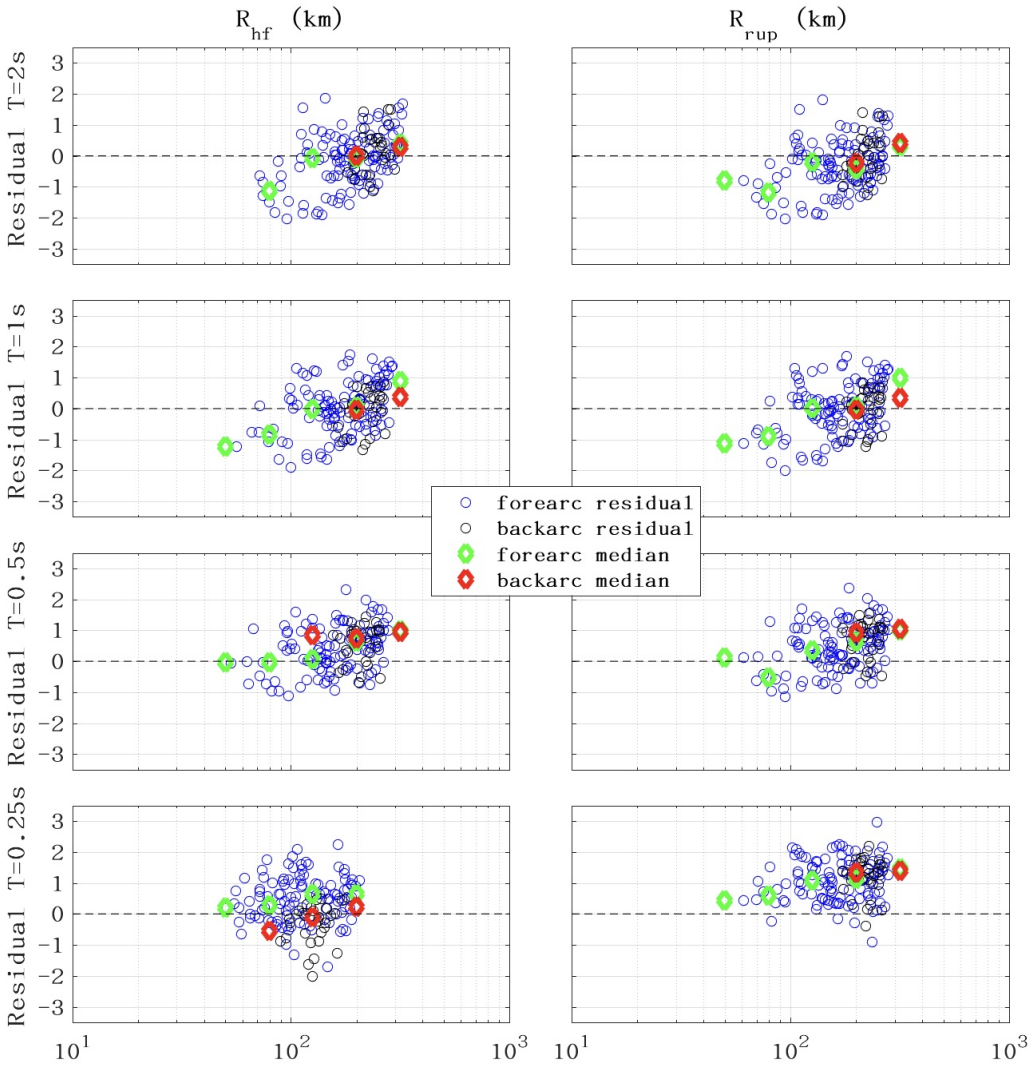

Figure C.11: Comparisons of logarithmic residual in different frequencies between R_{hf} metric and R_{rup} metric of 2005 off-Miyagi earthquake (Mw 7.2).

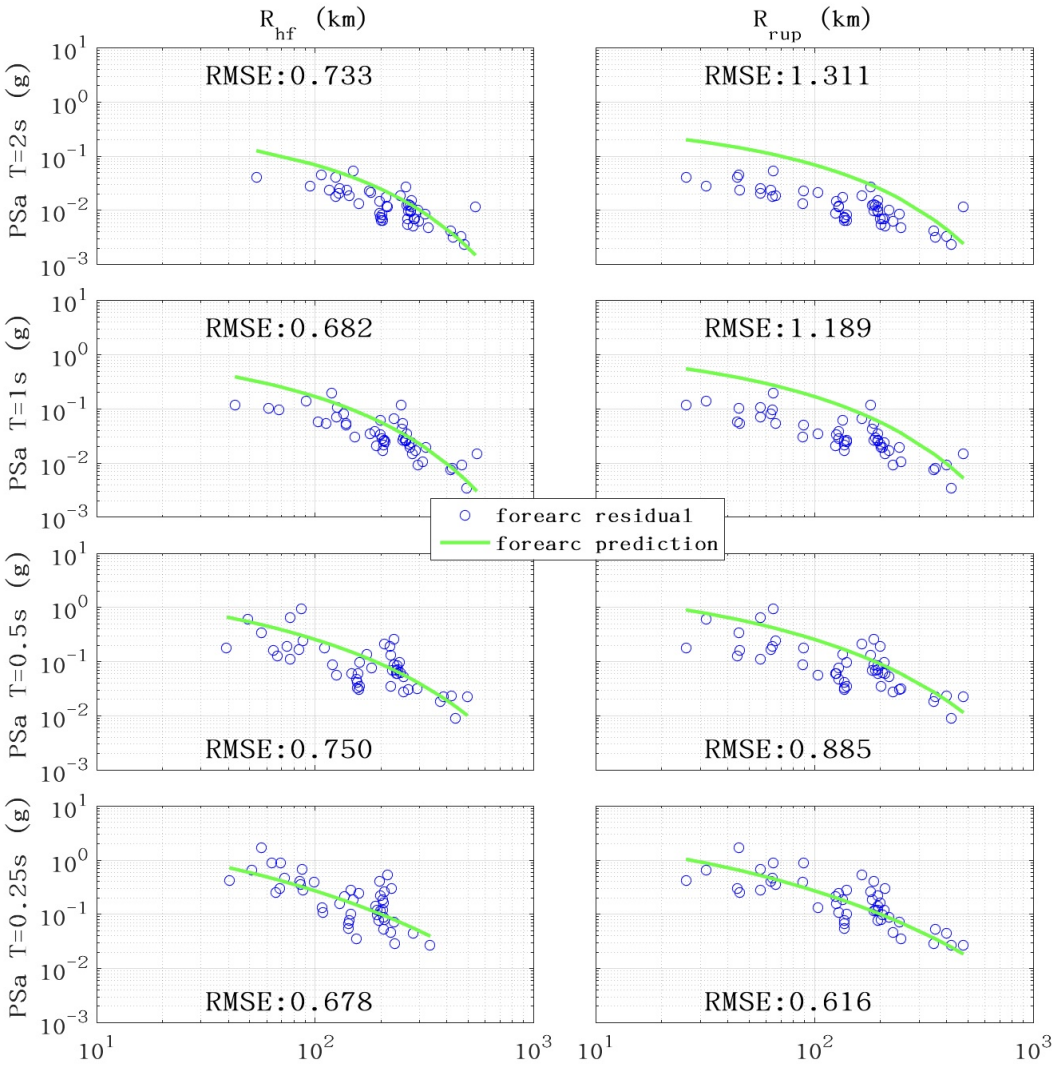

Figure C.12: Comparisons of PSa in different frequencies between R_{hf} metric and R_{rup} metric of 2015 Illapel earthquake (Mw 8.3).

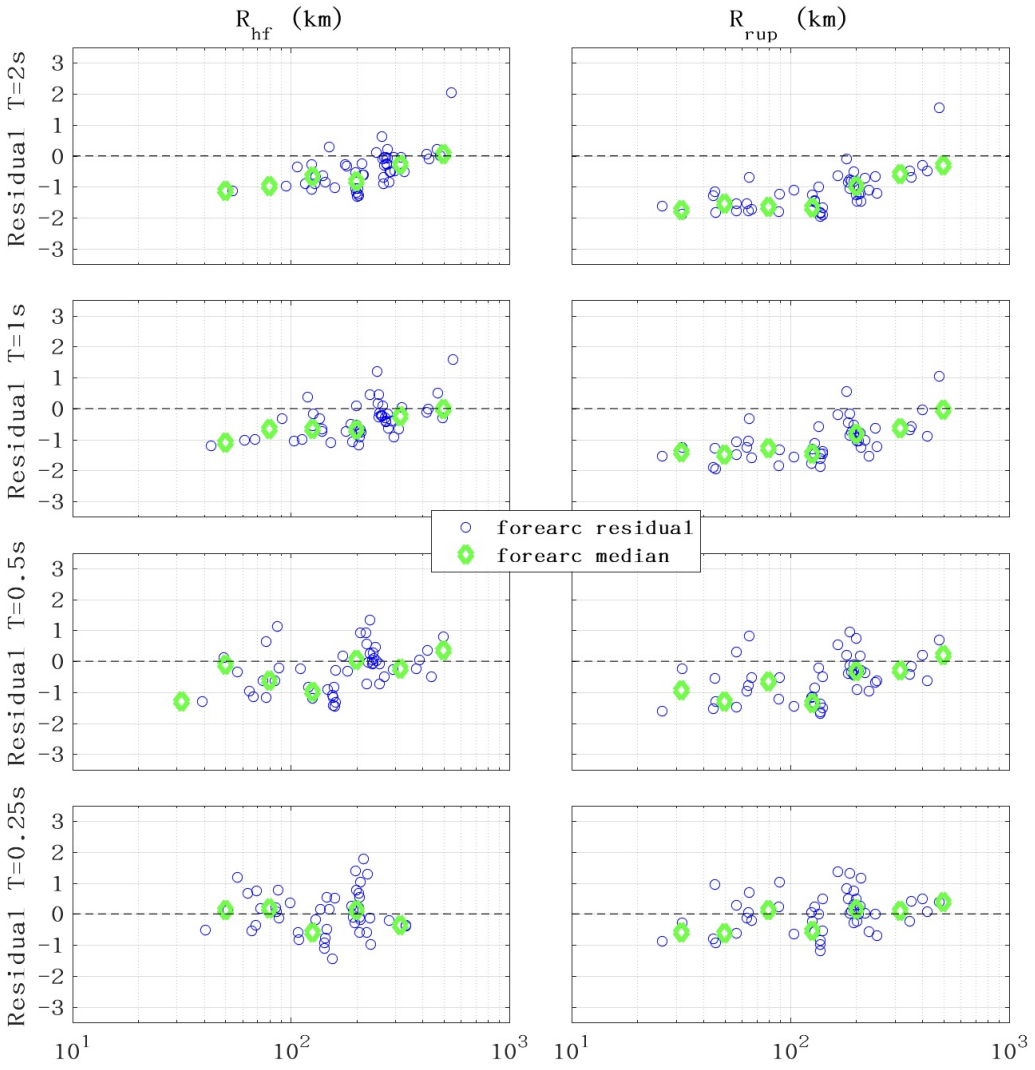

Figure C.13: Comparisons of logarithmic residual in different frequencies between R_{hf} metric and R_{rup} metric of 2015 Illapel earthquake (Mw 8.3).

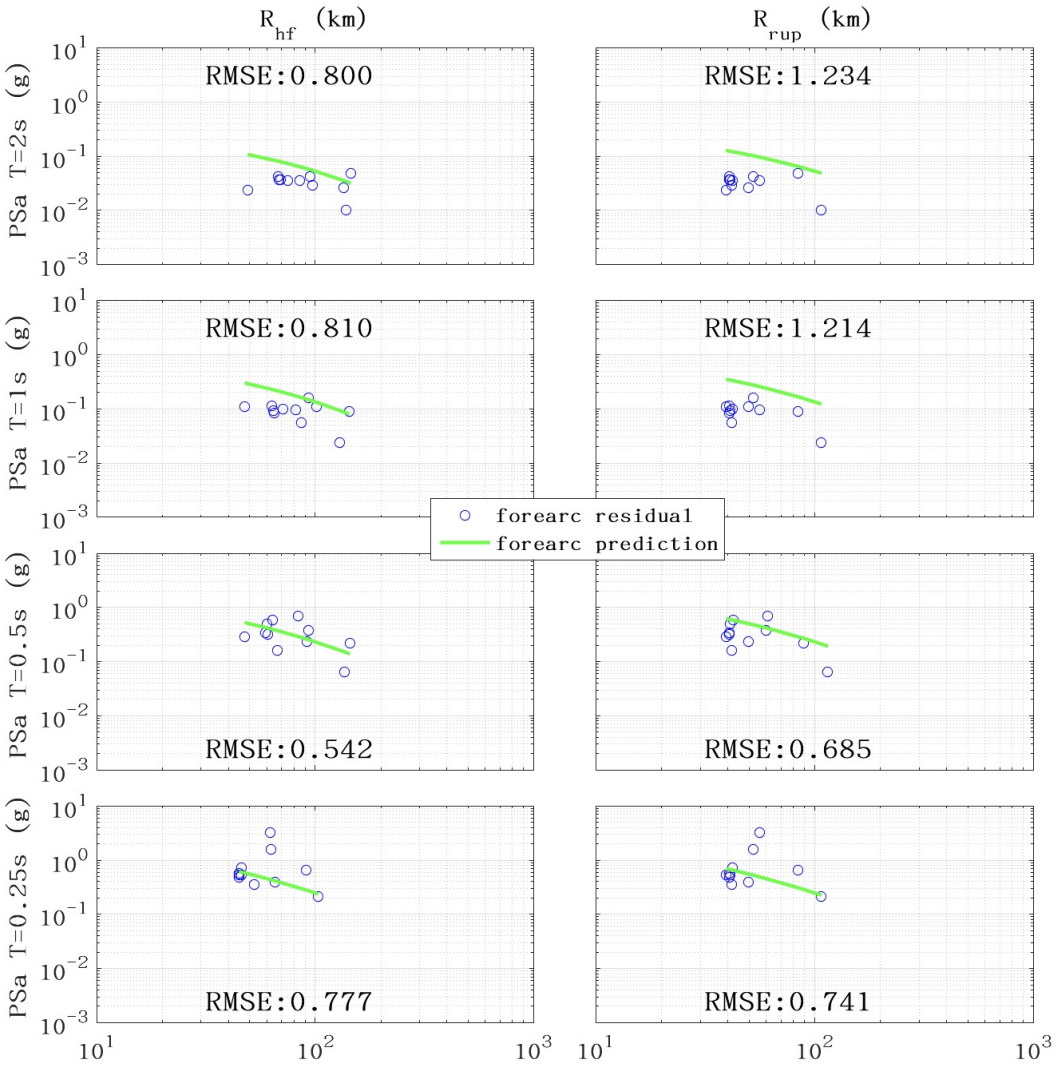

Figure C.14: Comparisons of PSa in different frequencies between R_{hf} metric and R_{rup} metric of 2014 Iquique earthquake (Mw 8.1).

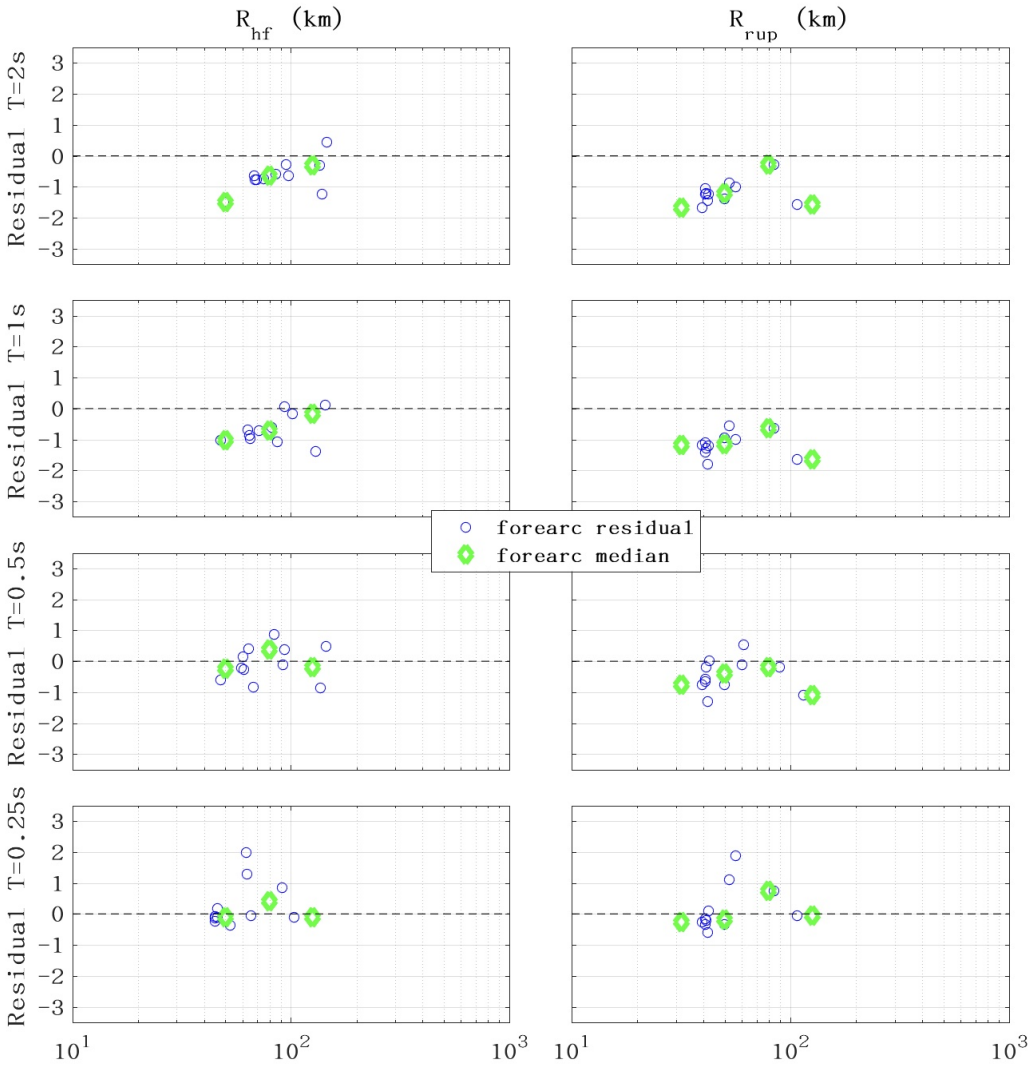

Figure C.15: Comparisons of logarithmic residual in different frequencies between R_{hf} metric and R_{rup} metric of 2014 Iquique earthquake (Mw 8.1).

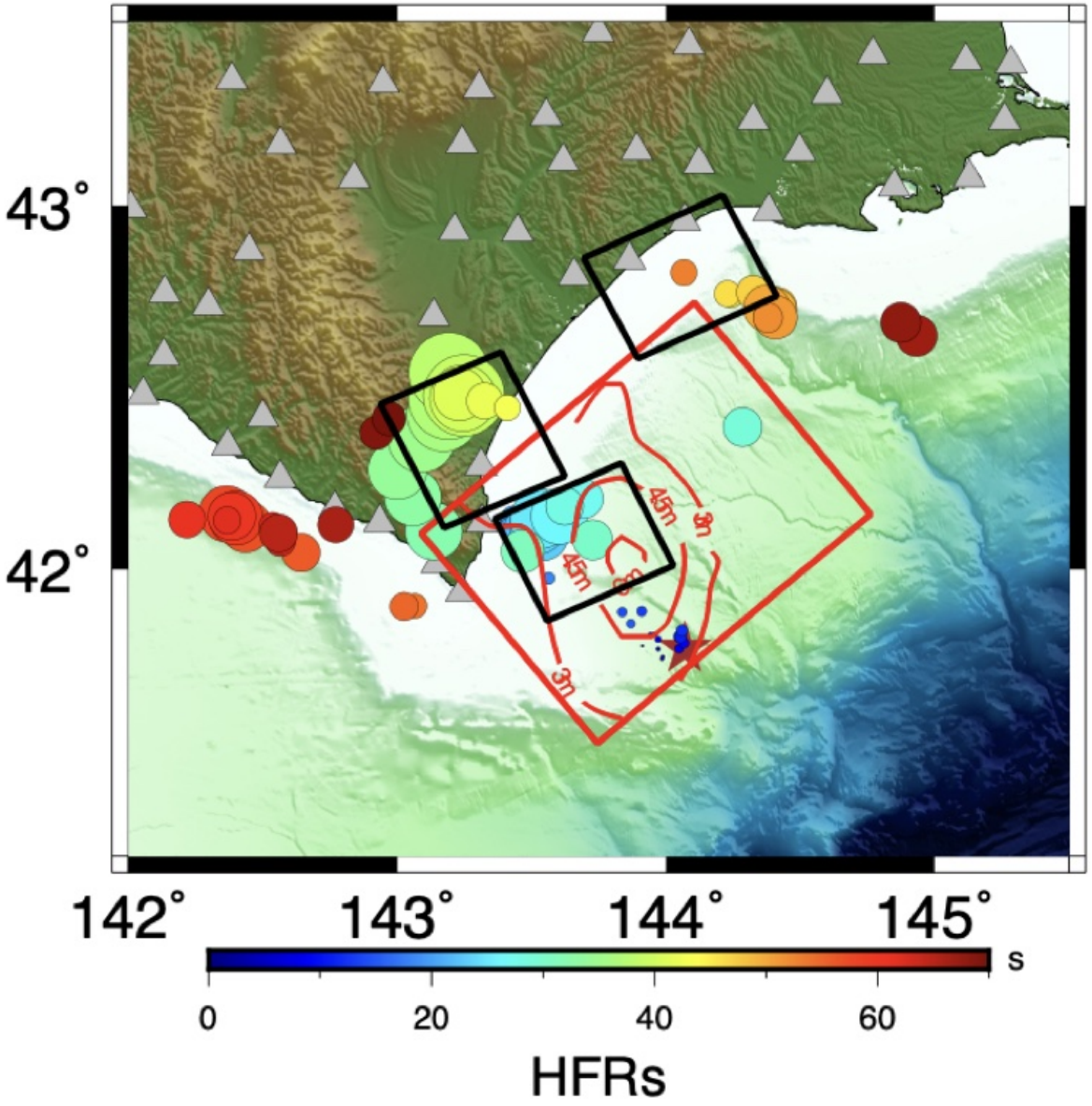

Figure C.16: Finite fault model of 2003 Mw 8.0 Tokachi earthquake (Koketsu et al., 2005) (red rectangle), strong motion generation areas from (Nozu & Irikura, 2008) (black rectangles), the 2 Hz high-frequency radiator locations (color spheres), and deep pre-seismic repeating earthquakes (Uchida et al., 2016) (purple open squares). Contours indicate coseismic slip (m) and the red star is the JMA epicenter of the main shock.

Bibliography

Abrahamson, N., Gregor, N., & Addo, K. (2016). Bc hydro ground motion prediction equations for subduction earthquakes. *Earthquake Spectra*, *32*(1), 23–44.

Abrahamson, N. A., Silva, W. J., & Kamai, R. (2014). Summary of the ask14 ground motion relation for active crustal regions. *Earthquake Spectra*, *30*(3), 1025–1055.

Aden-Antoniów, F., Frank, W., & Seydoux, L. (2022). An adaptable random forest model for the declustering of earthquake catalogs. *Journal of Geophysical Research: Solid Earth*, *127*(2), e2021JB023254.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199–213). Springer.

Akazawa, T. (2004). A technique for automatic detection of onset time of p-and s-phases in strong motion records. In *Proc. of the 13th world conf. on earthquake engineering.*

Akram, J., Peter, D., & Eaton, D. (2018). A k-mean characteristic function to improve sta/lta detection. *Proceedings of the Geoconvention, Calgary, AB, Canada*, 7–11.

Allen, R. (1982). Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, *72*(6B), S225–S242.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, *68*(5), 1521–1532.

Ammon, C. J., Lay, T., Kanamori, H., & Cleveland, M. (2011). A rupture model of the 2011 off the pacific coast of tohoku earthquake. *Earth, Planets and Space*, *63*(7), 33.

An, C., & Meng, L. (2016). Application of array backprojection to tsunami prediction and early warning. *Geophysical Research Letters*, *43*(8), 3677–3685.

Anant, K. S., & Dowla, F. U. (1997). Wavelet transform methods for phase identification in three-component seismograms. *Bulletin of the Seismological Society of America*, *87*(6), 1598–1612.

Ancheta, T., Bozorgnia, Y., Darragh, R., Silva, W., Chiou, B., Stewart, J., ... others (2012). Peer nga-west2 database: A database of ground motions recorded in shallow crustal earthquakes in active tectonic regions. In *Proceedings, 15th world conference*

on earthquake engineering.

Anderson, J. (2007). Physical processes that control strong ground motion. *Earthquake seismology*, *4*, 513–565.

Anderson, S., & Nehorai, A. (1996). Analysis of a polarized seismic wave model. *IEEE Transactions on signal processing*, *44*(2), 379–386.

Asano, Y., Saito, T., Ito, Y., Shiomi, K., Hirose, H., Matsumoto, T., ... Sekiguchi, S. (2011). Spatial distribution and focal mechanisms of aftershocks of the 2011 off the pacific coast of tohoku earthquake. *Earth, planets and space*, *63*(7), 669–673.

Atkinson, G. M., & Boore, D. M. (2007). *Boore-atkinson nga ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters.* Pacific Earthquake Engineering Research Center.

Atkinson, G. M., & Wald, D. J. (2007). "did you feel it?" intensity data: A surprisingly good measure of earthquake ground motion. *Seismological Research Letters*, *78*(3), 362–368.

Baer, M., & Kradolfer, U. (1987). An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, *77*(4), 1437–1445.

Baker, B., Holt, M. M., Pankow, K. L., Koper, K. D., & Farrell, J. (2021). Monitoring the 2020 magna, utah, earthquake sequence with nodal seismometers and machine learning. *Seismological Society of America*, *92*(2A), 787–801.

Barani, S., De Ferrari, R., Ferretti, G., & Spallarossa, D. (2010). Calibration of soil amplification factors for real-time ground-motion scenarios in italy. In *Proceedings of the 15th international conference on recent advances in geotechnical earthquake engineering and soil dynamics.*

Bergen, K. J., Johnson, P. A., Maarten, V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid earth geoscience. *Science*, *363*(6433).

Bianco, M. J., & Gerstoft, P. (2018). Travel time tomography with adaptive dictionaries. *IEEE Transactions on Computational Imaging*, *4*(4), 499–511.

Bianco, M. J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M. A., Gannot, S., ... Li, W. (2019). Machine learning in acoustics: a review. *arXiv e-prints*.

Birnie, C., Ravasi, M., Liu, S., & Alkhalifah, T. (2021). The potential of self-supervised networks for random noise suppression in seismic data. *Artificial Intelligence in Geosciences*, *2*, 47–59.

Bodin, T., & Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, *178*(3), 1411–1436.

Boese, C., Townend, J., Smith, E., & Stern, T. (2012). Microseismicity and stress in the vicinity of the alpine fault, central southern alps, new zealand. *Journal of Geophysical Research: Solid Earth*, *117*(B2).

Bolton, D. C., Shokouhi, P., Rouet-Leduc, B., Hulbert, C., Rivière, J., Marone, C., & Johnson, P. A. (2019). Characterizing acoustic signals and searching for precursors during the laboratory seismic cycle using unsupervised machine learning. *Seismological Research Letters*, *90*(3), 1088–1098.

Boore, D. M., Joyner, W. B., & Fumal, T. E. (1993). Estimation of response spectra and peak accelerations from western north american earthquakes: an interim report.

Boore, D. M., Joyner, W. B., & Fumal, T. E. (1997). Equations for estimating horizontal response spectra and peak acceleration from western north american earthquakes: A summary of recent work. *Seismological research letters*, *68*(1), 128–153.

Boore, D. M., Stewart, J. P., Seyhan, E., & Atkinson, G. M. (2014). Nga-west2 equations for predicting pga, pgv, and 5% damped psa for shallow crustal earthquakes. *Earthquake Spectra*, *30*(3), 1057–1085.

Borcherdt, R. D. (1994). Estimates of site-dependent response spectra for design (methodology and justification). *Earthquake spectra*, *10*(4), 617–653.

Borcherdt, R. D., & Glassmoyer, G. (1989). Influences of local geology on strong and weak ground motions recorded in the san francisco bay region and their implications for site-specific building-code provisions. *The Lorna Prieta, California, Earthquake of October*, *17*, 77–108.

Bossu, R., Mazet-Roux, G., Douet, V., Rives, S., Marin, S., & Aupetit, M. (2008). Internet users as seismic sensors for improved earthquake response. *Eos, Transactions American Geophysical Union*, *89*(25), 225–226.

Bouchon, M., Durand, V., Marsan, D., Karabulut, H., & Schmittbuhl, J. (2013). The long precursory phase of most large interplate earthquakes. *Nature geoscience*, *6*(4), 299–302.

Bozorgnia, Y., Abrahamson, N. A., Atik, L. A., Ancheta, T. D., Atkinson, G. M., Baker, J. W., ... others (2014). Nga-west2 research project. *Earthquake Spectra*, *30*(3), 973–987.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Brown, J. R., Beroza, G. C., & Shelly, D. R. (2008). An autocorrelation method to detect low frequency earthquakes within tremor. *Geophysical Research Letters*, *35*(16).

Bungum, H., Lindholm, C., & Faleide, J. (2005). Postglacial seismicity offshore mid-norway with emphasis on spatio-temporal–magnitudal variations. *Marine and Petroleum Geology*, *22*(1-2), 137–148.

Cesca, S., Grigoli, F., Heimann, S., González, A., Buforn, E., Maghsoudi, S., ... Dahm, T. (2014). The 2013 september–october seismic sequence offshore spain: a case of seismicity triggered by gas injection? *Geophysical Journal International*, *198*(2), 941–953.

Chakraborty, A., & Okaya, D. (1995). Frequency-time decomposition of seismic data using wavelet-based methods. *Geophysics*, *60*(6), 1906–1916.

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.

Chen, T., & Huang, L. (2020). Optimal design of microseismic monitoring network: Synthetic study for the kimberlina co2 storage demonstration site. *International Journal of Greenhouse Gas Control*, *95*, 102981.

Chen, Y., Meng, L., Zhang, A., & Wen, L. (2018). Source complexity of the 2015 mw 7.9 bonin earthquake. *Geochemistry, Geophysics, Geosystems*, *19*(7), 2109–2120.

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

Choi, Y., & Stewart, J. P. (2005). Nonlinear site amplification as function of 30 m shear wave velocity. *Earthquake spectra*, *21*(1), 1–30.

Chopra, A. K. (1995). Dynamics of structures: Theory and applications to earthquake engineering, prentice hall. *Inc., Upper Saddle River, NJ*.

Civilini, F., Weber, R., Jiang, Z., Phillips, D., & Pan, W. D. (2021). Detecting moonquakes using convolutional neural networks, a non-local training set, and transfer learning. *Geophysical Journal International*, *225*(3), 2120–2134.

Courboulex, F., Larroque, C., Deschamps, A., Gélis, C., Charreau, J., & Stéphan, J.-F. (2003). An unknown active fault revealed by microseismicity in the south-east of france. *Geophysical research letters*, *30*(15).

Cui, X., Li, Z., & Huang, H. (2021). Subdivision of seismicity beneath the summit region of kilauea volcano: Implications for the preparation process of the 2018 eruption. *Geophysical Research Letters*, *48*(20), e2021GL094698.

Dai, H., & MacBeth, C. (1995). Automatic picking of seismic arrivals in local earthquake data using an artificial neural network. *Geophysical journal international*, *120*(3), 758–774.

Dai, H., & MacBeth, C. (1997). The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings. *Journal of Geophysical Research: Solid Earth*, *102*(B7), 15105–15113.

De Barros, L., Wynants-Morel, N., Cappa, F., & Danré, P. (2021). Migration of fluid-induced seismicity reveals the seismogenic state of faults. *Journal of Geophysical Research: Solid Earth*, *126*(11), e2021JB022767.

Debski, W. (2008). Estimating the earthquake source time function by markov chain monte carlo sampling. *Pure and Applied Geophysics*, *165*(7), 1263–1287.

De Matteis, R., Matrullo, E., Rivera, L., Stabile, T. A., Pasquale, G., & Zollo, A. (2012). Fault delineation and regional stress direction from the analysis of background microseismicity in the southern apennines, italy. *Bulletin of the Seismological Society of America*, *102*(4), 1899–1907.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidi-

rectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

DeVries, P. M., Thompson, T. B., & Meade, B. J. (2017). Enabling large-scale viscoelastic calculations via neural network acceleration. *Geophysical Research Letters*, *44*(6), 2662–2669.

Dokht, R. M., Kao, H., Visser, R., & Smith, B. (2019). Seismic event and phase detection using time–frequency representation and convolutional neural networks. *Seismological Research Letters*, *90*(2A), 481–490.

Douglas, J. (2003). Earthquake ground motion estimation using strong-motion records: a review of equations for the estimation of peak ground acceleration and response spectral ordinates. *Earth-Science Reviews*, *61*(1-2), 43–104.

Draelos, T. J., Peterson, M. G., Knox, H. A., Lawry, B. J., Phillips-Alonge, K. E., Ziegler, A. E., ... Faust, A. (2018). Dynamic tuning of seismic signal detector trigger levels for local networks. *Bulletin of the Seismological Society of America*, *108*(3A), 1346–1354.

Du, H. (2021). Estimating rupture front of large earthquakes using a novel multi-array back-projection method. *Frontiers in Earth Science*, *9*, 482.

Earle, P. S., Bowden, D., & Guy, M. (2011). Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics*, *54*(6), 708–715.

Earle, P. S., & Shearer, P. M. (1994). Characterization of global seismograms using an automatic-picking algorithm. *Bulletin of the Seismological Society of America*, *84*(2), 366–376.

Ellsworth, W. L. (2013). Injection-induced earthquakes. *science*, *341*(6142), 1225942.

Enescu, B., & Ito, K. (2002). Spatial analysis of the frequency-magnitude distribution and decay rate of aftershock activity of the 2000 western tottori earthquake. *Earth, planets and space*, *54*(8), 847–859.

Evelpidou, N., Karkani, A., & Kampolis, I. (2021). Relative sea level changes and morphotectonic implications triggered by the samos earthquake of 30th october 2020. *Journal of Marine Science and Engineering*, *9*(1), 40.

Fan, W., & Shearer, P. M. (2016). Local near instantaneously dynamically triggered aftershocks of large earthquakes. *Science*, *353*(6304), 1133–1136.

Fan, W., & Shearer, P. M. (2017). Investigation of backprojection uncertainties with m6 earthquakes. *Journal of Geophysical Research: Solid Earth*, *122*(10), 7966–7986.

Feng, T., & Meng, L. (2018). A high-frequency distance metric in ground-motion prediction equations based on seismic array backprojections. *Geophysical Research Letters*, *45*(21), 11–612.

Fletcher, J. B., Spudich, P., & Baker, L. M. (2006). Rupture propagation of the 2004 parkfield, california, earthquake from observations at the upsar. *Bulletin of the Seismological Society of America*, *96*(4B), S129–S142.

Foumelis, M., Papazachos, C., Papadimitriou, E., Karakostas, V., Ampatzidis, D., Moschopoulos, G., ... others (2021). On rapid multidisciplinary response aspects for samos 2020 m7. 0 earthquake. *Acta Geophysica*, 1–24.

Frank, W. B., Poli, P., & Perfettini, H. (2017). Mapping the rheology of the central chile subduction zone with aftershocks. *Geophysical Research Letters*, *44*(11), 5374–5382.

Gamage, S. S., Umino, N., Hasegawa, A., & Kirby, S. H. (2009). Offshore double-planed shallow seismic zone in the ne japan forearc region revealed by sp depth phases recorded by regional networks. *Geophysical Journal International*, *178*(1), 195–214.

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.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. In *International conference on machine learning* (pp. 1263–1272).

Gori, M., Monfardini, G., & Scarselli, F. (2005). A new model for learning in graph domains. In *Proceedings. 2005 ieee international joint conference on neural networks* (Vol. 2, pp. 729–734).

Gregor, N., Abrahamson, N. A., Atkinson, G. M., Boore, D. M., Bozorgnia, Y., Campbell, K. W., ... others (2014). Comparison of nga-west2 gmpes. *Earthquake Spectra*, *30*(3), 1179–1197.

Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in california. *Bulletin of*

the Seismological Society of America, *34*(4), 185–188.

Hayes, G. P. (2011). Rapid source characterization of the 2011 m w 9.0 off the pacific coast of tohoku earthquake. *Earth, planets and space*, *63*(7), 4.

Hayes, G. P. (2017). The finite, kinematic rupture properties of great-sized earthquakes since 1990. *Earth and Planetary Science Letters*, *468*, 94–100.

Hayes, G. P., Wald, D. J., & Johnson, R. L. (2012). Slab1. 0: A three-dimensional model of global subduction zone geometries. *Journal of Geophysical Research: Solid Earth*, *117*(B1).

Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*.

Hirata, T. (1989). A correlation between the b value and the fractal dimension of earthquakes. *Journal of Geophysical Research: Solid Earth*, *94*(B6), 7507–7514.

Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *6*(02), 107–116.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Holtzman, B. K., Paté, A., Paisley, J., Waldhauser, F., & Repetto, D. (2018). Machine learning reveals cyclic changes in seismic source spectra in geysers geothermal field. *Science advances*, *4*(5), eaao2929.

Honsho, C., Kido, M., Tomita, F., & Uchida, N. (2019). Offshore postseismic deformation of the 2011 tohoku earthquake revisited: Application of an improved gps-acoustic positioning method considering horizontal gradient of sound speed structure. *Journal of Geophysical Research: Solid Earth*, *124*(6), 5990–6009.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554–2558.

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., Plasencia, M., Wang, Y., Wang, L., & Xu, M. (2017). Matched-filter detection of the missing pre-mainshock events and aftershocks in the 2015 gorkha, nepal earthquake sequence. *Tectonophysics*, *714*, 71–81.

Huang, H., Xu, W., Meng, L., Bürgmann, R., & Baez, J. C. (2017). Early aftershocks and afterslip surrounding the 2015 mw 8.4 illapel rupture. *Earth and Planetary Science Letters*, *457*, 282–291.

Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Rivière, J., Bolton, D. C., & Marone, C. (2019). Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience*, *12*(1), 69–74.

Hutton, L., & Boore, D. M. (1987). The ml scale in southern california. *Bulletin of the Seismological Society of America*, *77*(6), 2074–2094.

Hutton, L., Woessner, J., & Hauksson, E. (2010). Seventy-seven years (1932–2009) of earthquake monitoring in southern california. *Bull. Seismol. Soc. Am*, *100*(2), 423–446.

Ide, S., Baltay, A., & Beroza, G. C. (2011). Shallow dynamic overshoot and energetic deep rupture in the 2011 mw 9.0 tohoku-oki earthquake. *Science*, *332*(6036), 1426–1429.

Igarashi, T., Matsuzawa, T., Umino, N., & Hasegawa, A. (2001). Spatial distribution of focal mechanisms for interplate and intraplate earthquakes associated with the subducting pacific plate beneath the northeastern japan arc: A triple-planed deep seismic zone. *Journal of Geophysical Research: Solid Earth*, *106*(B2), 2177–2191.

Iinuma, T., Hino, R., Uchida, N., Nakamura, W., Kido, M., Osada, Y., & Miura, S. (2016). Seafloor observations indicate spatial separation of coseismic and postseismic slips in the 2011 tohoku earthquake. *Nature communications*, *7*(1), 1–9.

Inbal, A., Clayton, R. W., & Ampuero, J.-P. (2015). Imaging widespread seismicity at midlower crustal depths beneath long beach, ca, with a dense seismic array: Evidence for a depth-dependent earthquake size distribution. *Geophysical Research Letters*, *42*(15), 6314–6323.

Ishii, M., Shearer, P. M., Houston, H., & Vidale, J. E. (2005). Extent, duration and speed of the 2004 sumatra–andaman earthquake imaged by the hi-net array. *Nature*, *435*(7044),

933–936.

Ishii, M., Shearer, P. M., Houston, H., & Vidale, J. E. (2007). Teleseismic p wave imaging of the 26 december 2004 sumatra-andaman and 28 march 2005 sumatra earthquake ruptures using the hi-net array. *Journal of Geophysical Research: Solid Earth*, *112*(B11).

Jiang, J., & Lapusta, N. (2017). Connecting depth limits of interseismic locking, microseismicity, and large earthquakes in models of long-term fault slip. *Journal of Geophysical Research: Solid Earth*, *122*(8), 6491–6523.

Joyner, W. B., & Boore, D. M. (1981). Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 imperial valley, california, earthquake. *Bulletin of the seismological Society of America*, *71*(6), 2011–2038.

Jurkevics, A. (1988). Polarization analysis of three-component array data. *Bulletin of the seismological society of America*, *78*(5), 1725–1743.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, *4*, 237–285.

Kassaras, I., Kapetanidis, V., Ganas, A., Tzanis, A., Kosma, C., Karakonstantis, A., ... Papadimitriou, P. (2020). The new seismotectonic atlas of greece (v1. 0) and its implementation. *Geosciences*, *10*(11), 447.

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., 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.

Khalatbarisoltani, A., Soleymani, M., & Khodadadi, M. (2019). Online control of an active seismic system via reinforcement learning. *Structural Control and Health Monitoring*, *26*(3), e2298.

Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., Al Sallab, A. A., Yogamani, S., & Pérez, P. (2021). Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*.

Kiser, E., & Ishii, M. (2013). Hidden aftershocks of the 2011 mw 9.0 tohoku, japan earthquake imaged with the backprojection method. *Journal of Geophysical Research: Solid Earth*, *118*(10), 5564–5576.

Kiser, E., & Ishii, M. (2017). Back-projection imaging of earthquakes. *Annual Review of Earth and Planetary Sciences*, *45*, 271–299.

Kissling, E., Ellsworth, W., Eberhart-Phillips, D., & Kradolfer, U. (1994). Initial reference models in local earthquake tomography. *Journal of Geophysical Research: Solid Earth*, *99*(B10), 19635–19646.

Köhler, A., Ohrnberger, M., & Scherbaum, F. (2010). Unsupervised pattern recognition in continuous seismic wavefield records using self-organizing maps. *Geophysical Journal International*, *182*(3), 1619–1630.

Koketsu, K., Hatayama, K., Furumura, T., Ikegami, Y., & Akiyama, S. (2005). Damaging long-period ground motions from the 2003 mw 8.3 tokachi-oki, japan earthquake. *Seismological Research Letters*, *76*(1), 67–73.

Kong, Q., Allen, R. M., Schreier, L., & Kwon, Y.-W. (2016). Myshake: A smartphone seismic network for earthquake early warning and beyond. *Science advances*, *2*(2), e1501055.

Kong, Q., Inbal, A., Allen, R. M., Lv, Q., & Puder, A. (2019). Machine learning aspects of the myshake global smartphone seismic network. *Seismological Research Letters*, *90*(2A), 546–552.

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, *90*(1), 3–14.

Koper, K. D., Hutko, A. R., Lay, T., Ammon, C. J., & Kanamori, H. (2011). Frequency-dependent rupture process of the 2011 mw 9.0 tohoku earthquake: Comparison of short-period p wave backprojection images and broadband seismic rupture models. *Earth, planets and space*, *63*(7), 599–602.

Kriegerowski, M., Petersen, G. M., Vasyura-Bathke, H., & Ohrnberger, M. (2019). A deep convolutional neural network for localization of clustered earthquakes based on

multistation full waveforms. *Seismological Research Letters*, *90*(2A), 510–516.

Krüger, F., & Ohrnberger, M. (2005). Tracking the rupture of the m w= 9.3 sumatra earthquake over 1,150 km at teleseismic distance. *Nature*, *435*(7044), 937–939.

Kuang, W., Yuan, C., & Zhang, J. (2021). Real-time determination of earthquake focal mechanism via deep learning. *Nature communications*, *12*(1), 1–8.

Kufner, S.-K., Brisbourne, A. M., Smith, A. M., Hudson, T. S., Murray, T., Schlegel, R., ... Lee, I. (2021). Not all icequakes are created equal: Basal icequakes suggest diverse bed deformation mechanisms at rutford ice stream, west antarctica. *Journal of Geophysical Research: Earth Surface*, *126*(3), e2020JF006001.

Kurahashi, S., & Irikura, K. (2011). Source model for generating strong ground motions during the 2011 off the pacific coast of tohoku earthquake. *Earth, planets and space*, *63*(7), 571–576.

Kurzon, I., Vernon, F., Rosenberger, A., & Ben-Zion, Y. (2014). Real-time automatic detectors of p and s waves using singular value decomposition. *Bulletin of the Seismological Society of America*, *104*(4), 1696–1708.

Langenbruch, C., & Zoback, M. D. (2016). How will induced seismicity in oklahoma respond to decreased saltwater injection rates? *Science advances*, *2*(11), e1601542.

Langston, C. A. (2014). Coherence of teleseismic p and s waves across the transportable array. *Bulletin of the Seismological Society of America*, *104*(5), 2253–2265.

Lay, T. (2018). A review of the rupture characteristics of the 2011 tohoku-oki mw 9.1 earthquake. *Tectonophysics*, *733*, 4–36.

Lay, T., Ammon, C. J., Kanamori, H., Kim, M. J., & Xue, L. (2011). Outer trench-slope faulting and the 2011 m w 9.0 off the pacific coast of tohoku earthquake. *Earth, planets and space*, *63*(7), 37.

Lay, T., Kanamori, H., Ammon, C. J., Koper, K. D., Hutko, A. R., Ye, L., ... Rushing, T. M. (2012). Depth-varying rupture properties of subduction zone megathrust faults. *Journal of Geophysical Research: Solid Earth*, *117*(B4).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. (1999). Object recognition with gradient-

based learning. In *Shape, contour and grouping in computer vision* (pp. 319–345). Springer.

Lehner, B., Widmer, G., & Bock, S. (2015). A low-latency, real-time-capable singing voice detection method with lstm recurrent neural networks. In *2015 23rd european signal processing conference (eusipco)* (pp. 21–25).

Lengliné, O., Duputel, Z., & Ferrazzini, V. (2016). Uncovering the hidden signature of a magmatic recharge at piton de la fournaise volcano using small earthquakes. *Geophysical Research Letters*, *43*(9), 4255–4262.

Lengliné, O., Enescu, B., Peng, Z., & Shiomi, K. (2012). Decay and expansion of the early aftershock activity following the 2011, mw9. 0 tohoku earthquake. *Geophysical Research Letters*, *39*(18).

Leonard, M. (2000). Comparison of manual and automatic onset time picking. *Bulletin of the Seismological Society of America*, *90*(6), 1384–1390.

Leonard, M., & Kennett, B. (1999). Multi-component autoregressive techniques for the analysis of seismograms. *Physics of the Earth and Planetary Interiors*, *113*(1-4), 247–263.

Lervik, J. M., Rosten, T., & Ramstad, T. A. (1996). Subband seismic data compression: optimization and evaluation. In *1996 ieee digital signal processing workshop proceedings* (pp. 65–68).

Li, X., Shang, X., Wang, Z., Dong, L., & Weng, L. (2016). Identifying p-phase arrivals with noise: An improved kurtosis method based on dwt and sta/lta. *Journal of Applied Geophysics*, *133*, 50–61.

Li, Z., Meier, M.-A., Hauksson, E., Zhan, Z., & Andrews, J. (2018). Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical Research Letters*, *45*(10), 4773–4779.

Li, Z., Peng, Z., Hollis, D., Zhu, L., & McClellan, J. (2018). High-resolution seismic event detection using local similarity for large-n arrays. *Scientific reports*, *8*(1), 1–10.

Liao, W.-Y., Lee, E.-J., Mu, D., Chen, P., & Rau, R.-J. (2021). Arru phase picker: Attention recurrent-residual u-net for picking seismic p-and s-phase arrivals. *Seismological Society*

of America, *92*(4), 2410–2428.

Lindsey, N. J., Martin, E. R., Dreger, D. S., Freifeld, B., Cole, S., James, S. R., ... Ajo-Franklin, J. B. (2017). Fiber-optic network observations of earthquake wavefields. *Geophysical Research Letters*, *44*(23), 11–792.

Liu, Q., Fang, L., Yu, G., Wang, D., Xiao, C.-L., & Wang, K. (2019). Detection of dna base modifications by deep recurrent neural network on oxford nanopore sequencing data. *Nature communications*, *10*(1), 1–11.

Luco, J., & Wong, H. (1986). Response of a rigid foundation to a spatially random ground motion. *Earthquake Engineering & Structural Dynamics*, *14*(6), 891–908.

Madariaga, R. (1983). High frequency radiation from dynamic earthquake. *Ann. Geophys*, *1*, 17.

Maeda, N. (1985). A method for reading and checking phase times in autoprocessing system of seismic wave data. *Zisin*, *38*, 365–379.

Magana-Zook, S. A., & Ruppert, S. D. (2017). Explosion monitoring with machine learning: A lstm approach to seismic event discrimination. In *Agu fall meeting abstracts* (Vol. 2017, pp. S43A–0834).

Magotra, N., Ahmed, N., & Chael, E. (1987). Seismic event detection and source location using single-station (three-component) data. *Bulletin of the Seismological Society of America*, *77*(3), 958–971.

Magotra, N., Ahmed, N., & Chael, E. (1989). Single-station seismic event detection and location. *IEEE Transactions on Geoscience and Remote Sensing*, *27*(1), 15–23.

Marra, G., Clivati, C., Luckett, R., Tampellini, A., Kronjäger, J., Wright, L., ... others (2018). Ultrastable laser interferometry for earthquake detection with terrestrial and submarine cables. *Science*, *361*(6401), 486–490.

Martin, E. R., Huot, F., Ma, Y., Cieplicki, R., Cole, S., Karrenbach, M., & Biondi, B. L. (2018). A seismic shift in scalable acquisition demands new processing: Fiber-optic seismic signal retrieval in urban areas with unsupervised learning for coherent noise removal. *IEEE Signal Processing Magazine*, *35*(2), 31–40.

Matzel, E., White, J., Templeton, D., Pyle, M., Morency, C., & Trainor-Guitton, W. (2014).

Microseismic techniques for avoiding induced seismicity during fluid injection. *Energy Procedia*, *63*, 4297–4304.

Mauricio, A.-P., Jennings, J., Adler, A., & Dahlke, T. (2018). Deep-learning tomography. *The Leading Edge*, *37*(1), 58–66.

McBrearty, I. W., & Beroza, G. C. (2021). Earthquake phase association with graph neural networks. In *Agu fall meeting 2021.*

McBrearty, I. W., & Beroza, G. C. (2022). Earthquake location and magnitude estimation with graph neural networks. *arXiv preprint arXiv:2203.05144*.

McNutt, S. R., & Roman, D. C. (2015). Volcanic seismicity. In *The encyclopedia of volcanoes* (pp. 1011–1034). Elsevier.

Meier, M.-A., Ross, Z. E., Ramachandran, A., Balakrishna, A., Nair, S., Kundzicz, P., ... Yue, Y. (2019). Reliable real-time seismic signal/noise discrimination with machine learning. *Journal of Geophysical Research: Solid Earth*, *124*(1), 788–800.

Mendoza, C., & Hartzell, S. H. (1988). Aftershock patterns and main shock faulting. *Bulletin of the Seismological Society of America*, *78*(4), 1438–1449.

Meng, J., Sinoplu, O., Zhou, Z., Tokay, B., Kusky, T., Bozkurt, E., & Wang, L. (2021). Greece and turkey shaken by african tectonic retreat. *Scientific Reports*, *11*(1), 1–10.

Meng, L., Allen, R., & Ampuero, J.-P. (2014). Application of seismic array processing to earthquake early warning. *Bulletin of the Seismological Society of America*, *104*(5), 2553–2561.

Meng, L., Ampuero, J.-P., Luo, Y., Wu, W., & Ni, S. (2012). Mitigating artifacts in back-projection source imaging with implications for frequency-dependent properties of the tohoku-oki earthquake. *Earth, planets and space*, *64*(12), 1101–1109.

Meng, L., Bao, H., Huang, H., Zhang, A., Bloore, A., & Liu, Z. (2018). Double pincer movement: Encircling rupture splitting during the 2015 mw 8.3 illapel earthquake. *Earth and Planetary Science Letters*, *495*, 164–173.

Meng, L., Huang, H., Xie, Y., Bao, H., & Dominguez, L. A. (2019). Nucleation and kinematic rupture of the 2017 mw 8.2 tehuantepec earthquake. *Geophysical Research Letters*, *46*(7), 3745–3754.

Meng, L., Inbal, A., & Ampuero, J.-P. (2011). A window into the complexity of the dynamic rupture of the 2011 mw 9 tohoku-oki earthquake. *Geophysical Research Letters*, *38*(7).

Meng, L., Zhang, A., & Yagi, Y. (2016). Improving back projection imaging with a novel physics-based aftershock calibration approach: A case study of the 2015 gorkha earthquake. *Geophysical Research Letters*, *43*(2), 628–636.

Meng, X., Yu, X., Peng, Z., & Hong, B. (2012). Detecting earthquakes around salton sea following the 2010 mw7. 2 el mayor-cucapah earthquake using gpu parallel computing. *Procedia Computer Science*, *9*, 937–946.

Miao, T., & Wang, M. (2015). Susceptibility analysis of earthquake-induced landslide using random forest method. In *International conference on computer information systems and industrial applications* (pp. 771–775).

Mishra, O., Zhao, D., Umino, N., & Hasegawa, A. (2003). Tomography of northeast japan forearc and its implications for interplate seismic coupling. *Geophysical Research Letters*, *30*(16).

Miyake, H., Iwata, T., & Irikura, K. (2003). Source characterization for broadband ground-motion simulation: Kinematic heterogeneous source model and strong motion generation area. *Bulletin of the Seismological Society of America*, *93*(6), 2531–2545.

Mochizuki, M., Kanazawa, T., Uehira, K., Shimbo, T., Shiomi, K., Kunugi, T., ... others (2016). S-net project: Construction of large scale seafloor observatory network for tsunamis and earthquakes in japan. In *Agu fall meeting abstracts.*

Mousavi, S. M., & Beroza, G. C. (2019). Bayesian-deep-learning estimation of earthquake location from single-station observations. *arXiv preprint arXiv:1912.01144*.

Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, *47*(1), e2019GL085976.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, *11*(1), 1–12.

Mousavi, S. M., Sheng, Y., Zhu, W., & Beroza, G. C. (2019). Stanford earthquake dataset (stead): A global data set of seismic signals for ai. *IEEE Access*, *7*, 179464–179476.

Mousavi, S. M., Zhu, W., Sheng, Y., & Beroza, G. C. (2019). Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection. *Scientific reports*, *9*(1), 1–14.

Nakamura, W., Uchida, N., & Matsuzawa, T. (2016). Spatial distribution of the faulting types of small earthquakes around the 2011 tohoku-oki earthquake: A comprehensive search using template events. *Journal of Geophysical Research: Solid Earth*, *121*(4), 2591–2607.

Nakano, M., Sugiyama, D., Hori, T., Kuwatani, T., & Tsuboi, S. (2019). Discrimination of seismic signals from earthquakes and tectonic tremor by applying a convolutional neural network to running spectral images. *Seismological Research Letters*, *90*(2A), 530–538.

Nishikawa, T., & Ide, S. (2014). Earthquake size distribution in subduction zones linked to slab buoyancy. *Nature Geoscience*, *7*(12), 904–908.

Novoselov, A., Balazs, P., & Bokelmann, G. (2020). Sedenoss: Separating and denoising seismic signals with dual-path recurrent neural network architecture. *Journal of Geophysical Research: Solid Earth*, e2021JB023183.

Nozu, A., & Irikura, K. (2008). Strong-motion generation areas of a great subduction-zone earthquake: waveform inversion with empirical green's functions for the 2003 tokachi-oki earthquake. *Bulletin of the Seismological Society of America*, *98*(1), 180–197.

Obana, K., Fujie, G., Takahashi, T., Yamamoto, Y., Nakamura, Y., Kodaira, S., ... Shinohara, M. (2012). Normal-faulting earthquakes beneath the outer slope of the japan trench after the 2011 tohoku earthquake: Implications for the stress regime in the incoming pacific plate. *Geophysical Research Letters*, *39*(7).

Obana, K., Kodaira, S., Shinohara, M., Hino, R., Uehira, K., Shiobara, H., ... others (2013). Aftershocks near the updip end of the 2011 tohoku-oki earthquake. *Earth and Planetary Science Letters*, *382*, 111–116.

Ochoa, L. H., Niño, L. F., & Vargas, C. A. (2018). Fast magnitude determination using a single seismological station record implementing machine learning techniques. *Geodesy and Geodynamics*, *9*(1), 34–41.

Ohno, S., Ohta, T., Ikeura, T., & Takemura, M. (1993). Revision of attenuation formula considering the effect of fault size to evaluate strong motion spectra in near field. *Tectonophysics*, *218*(1-3), 69–81.

Okuwaki, R., Yagi, Y., Aránguiz, R., González, J., & González, G. (2017). Rupture process during the 2015 illapel, chile earthquake: Zigzag-along-dip rupture episodes. In *The chile-2015 (illapel) earthquake and tsunami* (pp. 23–32). Springer.

Omori, F. (1894). *On the after-shocks of earthquakes* (Vol. 7). The University.

Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.

Ozawa, S., Nishimura, T., Munekane, H., Suito, H., Kobayashi, T., Tobita, M., & Imakiire, T. (2012). Preceding, coseismic, and postseismic slips of the 2011 tohoku earthquake, japan. *Journal of Geophysical Research: Solid Earth*, *117*(B7).

Ozer, C., Gok, E., & Polat, O. (2018). Three-dimensional seismic velocity structure of the aegean region of turkey from local earthquake tomography. *Annals of Geophysics*, *61*(1), SE111–SE111.

Paitz, P., Gokhberg, A., & Fichtner, A. (2018). A neural network for noise correlation classification. *Geophysical Journal International*, *212*(2), 1468–1474.

Palo, M., Tilmann, F., Krüger, F., Ehlert, L., & Lange, D. (2014). High-frequency seismic radiation from maule earthquake (m w 8.8, 2010 february 27) inferred from high-resolution backprojection analysis. *Geophysical Journal International*, *199*(2), 1058–1077.

Pananont, P., & Pornsopin, P. (2020). A possible hidden fault in chiang mai, northern thailand from microseismicity. In *3rd asia pacific meeting on near surface geoscience & engineering* (Vol. 2020, pp. 1–4).

Papachristos, E., & Stefanou, I. (2021). Controlling earthquake-like instabilities using artificial intelligence. *arXiv preprint arXiv:2104.13180*.

Papadimitriou, P., Kapetanidis, V., Karakonstantis, A., Spingos, I., Kassaras, I., Sakkas, V., ... others (2020). Preliminary report on the mw= 6.9 samos earthquake of 30

october 2020. *Department of Geophysics and Geothermy, National and Kapodistrian University of Athens: Athens, Greece*.

Peng, Z., & Zhao, P. (2009). Migration of early aftershocks following the 2004 parkfield earthquake. *Nature Geoscience*, *2*(12), 877–881.

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2), e1700578.

Piana Agostinetti, N., Giacomuzzi, G., & Malinverno, A. (2015). Local three-dimensional earthquake tomography by trans-dimensional monte carlo sampling. *Geophysical Journal International*, *201*(3), 1598–1617.

Poulton, M. M. (2002). Neural networks as an intelligence amplification tool: A review of applications. *Geophysics*, *67*(3), 979–993.

Reddy, R., & Nair, R. R. (2013). The efficacy of support vector machines (svm) in robust determination of earthquake early warning magnitudes in central japan. *Journal of Earth System Science*, *122*(5), 1423–1434.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 779–788).

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.

Reynen, A., & Audet, P. (2017). Supervised machine learning on a network scale: Application to seismic event classification and detection. *Geophysical Journal International*, *210*(3), 1394–1409.

Ripepe, M., Donne, D. D., Genco, R., Maggio, G., Pistolesi, M., Marchetti, E., ... Poggi, P. (2015). Volcano seismicity and ground deformation unveil the gravity-driven magma discharge dynamics of a volcanic eruption. *Nature communications*, *6*(1), 1–6.

Rodriguez, I. V. (2011). Automatic time-picking of microseismic data combining sta/lta and the stationary discrete wavelet transform. In *Cspg cseg cwls convention, convention abstracts.*

Ross, Z. E., & Ben-Zion, Y. (2014). Automatic picking of direct p, s seismic phases and fault zone head waves. *Geophysical Journal International*, *199*(1), 368–381.

Ross, Z. E., Cochran, E. S., Trugman, D. T., & Smith, J. D. (2020). 3d fault architecture controls the dynamism of earthquake swarms. *Science*, *368*(6497), 1357–1361.

Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, *123*(6), 5120–5129.

Ross, Z. E., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, *108*(5A), 2894–2901.

Ross, Z. E., Trugman, D. T., Azizzadenesheli, K., & Anandkumar, A. (2020). Directivity modes of earthquake populations with unsupervised learning. *Journal of Geophysical Research: Solid Earth*, *125*(2), e2019JB018299.

Ross, Z. E., Yue, Y., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2019). Phaselink: A deep learning approach to seismic phase association. *Journal of Geophysical Research: Solid Earth*, *124*(1), 856–869.

Rost, S., & Thomas, C. (2002). Array seismology: Methods and applications. *Reviews of geophysics*, *40*(3), 2–1.

Rouet-Leduc, B., Hulbert, C., McBrearty, I. W., & Johnson, P. A. (2020). Probing slow earthquakes with deep learning. *Geophysical research letters*, *47*(4), e2019GL085870.

Rubin, A. M., Gillard, D., & Got, J.-L. (1998). A reinterpretation of seismicity associated with the january 1983 dike intrusion at kilauea volcano, hawaii. *Journal of Geophysical Research: Solid Earth*, *103*(B5), 10003–10015.

Ruud, B., & Husebye, E. (1992). A new three-component detector and automatic single-station bulletin production. *Bulletin of the seismological society of America*, *82*(1), 221–237.

Saad, O. M., & Chen, Y. (2020). Earthquake detection and p-wave arrival time picking using capsule neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(7), 6234–6243.

Saad, O. M., Huang, G., Chen, Y., Savvaidis, A., Fomel, S., Pham, N., & Chen, Y. (2021). Scalodeep: a highly generalized deep learning framework for real-time earthquake detection. *Journal of Geophysical Research: Solid Earth*, *126*(4), e2020JB021473.

Saari, J. (1991). Automated phase picker and source location algorithm for local distances using a single three-component seismic station. *Tectonophysics*, *189*(1-4), 307–315.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on world wide web* (pp. 851–860).

Salvoni, M., & Dight, P. M. (2016). Rock damage assessment in a large unstable slope from microseismic monitoring-mmg century mine (queensland, australia) case study. *Engineering Geology*, *210*, 45–56.

Saragiotis, C. D., Hadjileontiadis, L. J., & Panas, S. M. (1999). A higher-order statistics-based phase identification of three-component seismograms in a redundant wavelet transform domain. In *Proceedings of the ieee signal processing workshop on higher-order statistics. spw-hos'99* (pp. 396–399).

Saragiotis, C. D., Hadjileontiadis, L. J., & Panas, S. M. (2002). Pai-s/k: A robust automatic seismic p phase arrival identification scheme. *IEEE Transactions on Geoscience and Remote Sensing*, *40*(6), 1395–1404.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks*, *20*(1), 61–80.

Schaff, D. P., Bokelmann, G. H., Beroza, G. C., Waldhauser, F., & Ellsworth, W. L. (2002). High-resolution image of calaveras fault seismicity. *Journal of Geophysical Research: Solid Earth*, *107*(B9), ESE–5.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85–117.

Seydoux, L., Balestriero, R., Poli, P., Hoop, M. d., Campillo, M., & Baraniuk, R. (2020). Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. *Nature communications*, *11*(1), 1–12.

Shahnas, M., Yuen, D., & Pysklywec, R. (2018). Inverse problems in geodynamics using

machine learning algorithms. *Journal of Geophysical Research: Solid Earth*, *123*(1), 296–310.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, *27*(3), 379–423.

Shao, G., & Ji, C. (2013). Preliminary result of the aug 16, 2005 mw 7.19 honshu earthquake. *UCSB http://www. geol. ucsb. edu/faculty/ji/big_earthquakes/2005/08/smooth/honshu. html (accessed 22 August 2013)*.

Shao, G., Li, X., Ji, C., & Maeda, T. (2011). Focal mechanism and slip history of the 2011 m w 9.1 off the pacific coast of tohoku earthquake, constrained with teleseismic body and surface waves. *Earth, planets and space*, *63*(7), 559–564.

Shelly, D. R., Beroza, G. C., & Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, *446*(7133), 305–307.

Shelly, D. R., Hill, D. P., Massin, F., Farrell, J., Smith, R. B., & Taira, T. (2013). A fluid-driven earthquake swarm on the margin of the yellowstone caldera. *Journal of Geophysical Research: Solid Earth*, *118*(9), 4872–4886.

Shin, J. Y., Chen, S., & Kim, T.-W. (2015). Application of bayesian markov chain monte carlo method with mixed gumbel distribution to estimate extreme magnitude of tsunamigenic earthquake. *KSCE Journal of Civil Engineering*, *19*(2), 366–375.

Shoji, D., Noguchi, R., Otsuki, S., & Hino, H. (2018). Classification of volcanic ash particles using a convolutional neural network and probability. *Scientific reports*, *8*(1), 1–12.

Si, H., & Midorikawa, S. (2000). New attenuation relations for peak ground acceleration and velocity considering effects of fault type and site condition. In *Proceedings of 12th world conference on earthquake engineering.*

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... others (2016). Mastering the game of go with deep neural networks and tree search. *nature*, *529*(7587), 484–489.

Sleeman, R., & Van Eck, T. (1999). Robust automatic p-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Physics of the earth*

and planetary interiors, *113*(1-4), 265–275.

Smith, E., Smith, A., White, R., Brisbourne, A., & Pritchard, H. (2015). Mapping the ice-bed interface characteristics of rutford ice stream, west antarctica, using microseismicity. *Journal of Geophysical Research: Earth Surface*, *120*(9), 1881–1894.

Smith, J. D., Ross, Z. E., Azizzadenesheli, K., & Muir, J. B. (2022). Hyposvi: Hypocentre inversion with stein variational inference and physics informed neural networks. *Geophysical Journal International*, *228*(1), 698–710.

Sokolov, V. Y., & Chernov, Y. K. (1998). On the correlation of seismic intensity with fourier amplitude spectra. *Earthquake spectra*, *14*(4), 679–694.

Somerville, P., Irikura, K., Graves, R., Sawada, S., Wald, D., Abrahamson, N., ... Kowada, A. (1999). Characterizing crustal earthquake slip models for the prediction of strong ground motion. *Seismological Research Letters*, *70*(1), 59–80.

Spillmann, T., Maurer, H., Green, A. G., Heincke, B., Willenberg, H., & Husen, S. (2007). Microseismic investigation of an unstable mountain slope in the swiss alps. *Journal of Geophysical Research: Solid Earth*, *112*(B7).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929–1958.

Stevenson, P. R. (1976). Microearthquakes at flathead lake, montana: A study using automatic earthquake processing. *Bulletin of the Seismological Society of America*, *66*(1), 61–80.

Stewart, J. P., Lanzo, G., Pagliaroli, A., Scasserra, G., Di Capua, G., Peppoloni, S., ... Gregor, N. (2012). Ground motion recordings from the mw 6.3 2009 l'aquila earthquake in italy and their engineering implications. *Earthquake Spectra*, *28*(1), 317–345.

Stewart, J. P., Midorikawa, S., Graves, R. W., Khodaverdi, K., Kishida, T., Miura, H., ... Campbell, K. W. (2013). Implications of the mw9. 0 tohoku-oki earthquake for ground motion scaling with source, path, and site parameters. *Earthquake Spectra*, *29*(1_suppl), 1–21.

Sun, T., & Wang, K. (2015). Viscoelastic relaxation following subduction earthquakes and

its effects on afterslip determination. *Journal of Geophysical Research: Solid Earth*, *120*(2), 1329–1344.

Tadokoro, K., Ando, M., Ikuta, R., Okuda, T., Besana, G. M., Sugimoto, S., & Kuno, M. (2006). Observation of coseismic seafloor crustal deformation due to m7 class offshore earthquakes. *Geophysical research letters*, *33*(23).

Tajima, F., Mori, J., & Kennett, B. L. (2013). A review of the 2011 tohoku-oki earthquake (mw 9.0): Large-scale rupture across heterogeneous plate coupling. *Tectonophysics*, *586*, 15–34.

Tan, O. (2013). The dense micro-earthquake activity at the boundary between the anatolian and south aegean microplates. *Journal of Geodynamics*, *65*, 199–217.

Tan, O., Papadimitriou, E. E., Pabucçu, Z., Karakostas, V., Yörük, A., & Leptokaropoulos, K. (2014). A detailed analysis of microseismicity in samos and kusadasi (eastern aegean sea) areas. *Acta Geophysica*, *62*(6), 1283–1309.

Thompson, E. M., & Baltay, A. S. (2018). The case for mean rupture distance in ground-motion estimationthe case for mean rupture distance in ground-motion estimation. *Bulletin of the Seismological Society of America*, *108*(5A), 2462–2477.

Tian, X., Zhang, W., Zhang, X., Zhang, J., Zhang, Q., Wang, X., & Guo, Q. (2020). Comparison of single-trace and multiple-trace polarity determination for surface microseismic data using deep learning. *Seismological Research Letters*, *91*(3), 1794–1803.

Titos, M., Bueno, A., Garcia, L., & Benitez, C. (2018). A deep neural networks approach to automatic recognition systems for volcano-seismic events. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(5), 1533–1544.

Toda, S., Lin, J., & Stein, R. S. (2011). Using the 2011 m w 9.0 off the pacific coast of tohoku earthquake to test the coulomb stress triggering hypothesis and to calculate faults brought closer to failure. *Earth, planets and space*, *63*(7), 39.

Tormann, T., Enescu, B., Woessner, J., & Wiemer, S. (2015). Randomness of megathrust earthquakes implied by rapid stress recovery after the japan earthquake. *Nature Geoscience*, *8*(2), 152–158.

Trabant, C., Hutko, A. R., Bahavar, M., Karstens, R., Ahern, T., & Aster, R. (2012).

Data products at the iris dmc: Stepping stones for research and other applications. *Seismological Research Letters*, *83*(5), 846–854.

Trnkoczy, A. (2009). Understanding and parameter setting of sta/lta trigger algorithm. In *New manual of seismological observatory practice (nmsop)* (pp. 1–20). Deutsches GeoForschungsZentrum GFZ.

Trugman, D. T., & Shearer, P. M. (2018). Strong correlation between stress drop and peak ground acceleration for recent m 1–4 earthquakes in the san francisco bay area. *Bulletin of the Seismological Society of America*, *108*(2), 929–945.

Turin, G. (1960). An introduction to matched filters. *IRE transactions on Information theory*, *6*(3), 311–329.

Uchida, N., Iinuma, T., Nadeau, R. M., Bürgmann, R., & Hino, R. (2016). Periodic slow slip triggers megathrust zone earthquakes in northeastern japan. *Science*, *351*(6272), 488–492.

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., Shimamura, K., Matsuzawa, T., & Okada, T. (2015). Postseismic response of repeating earthquakes around the 2011 tohoku-oki earthquake: Moment increases due to the fast loading rate. *Journal of Geophysical Research: Solid Earth*, *120*(1), 259–274.

Uchide, T. (2020). Focal mechanisms of small earthquakes beneath the japanese islands based on first-motion polarities picked using deep learning. *Geophysical Journal International*.

Uehira, K., Mochizuki, M., Kanazawa, T., Shiomi, K., Kunugi, T., Aoi, S., ... others (2018). S-net project: Construction of large-scale seismic and tsunami observation system on seafloor along the japan trench. In *Egu general assembly conference abstracts* (Vol. 20, p. 12000).

Umino, N., Hasegawa, A., & Matsuzawa, T. (1995). sp depth phase at small epicentral distances and estimated subducting plate boundary. *Geophysical Journal International*, *120*(2), 356–366.

Urbancic, T. I., Trifu, C.-I., & Young, R. P. (1993). Microseismicity derived fault-planes and their relationship to focal mechanism, stress inversion, and geologic data. *Geophysical Research Letters*, *20*(22), 2475–2478.

Utsu, T. (1961). A statistical study on the occurrence of aftershocks. *Geophys. Mag.*, *30*, 521–605.

Utsu, T., Ogata, Y., et al. (1995). The centenary of the omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, *43*(1), 1–33.

Valentine, A. P., & Trampert, J. (2012). Data space reduction, quality assessment and searching of seismograms: autoencoder networks for waveform data. *Geophysical Journal International*, *189*(2), 1183–1202.

Valentine, A. P., & Woodhouse, J. H. (2010). Approaches to automated data selection for global seismic tomography. *Geophysical Journal International*, *182*(2), 1001–1012.

Vallée, M., & Douet, V. (2016). A new database of source time functions (stfs) extracted from the scardec method. *Physics of the Earth and Planetary Interiors*, *257*, 149–157.

van den Ende, M., Lior, I., Ampuero, J.-P., Sladen, A., Ferrari, A., & Richard, C. (2021). A self-supervised deep learning approach for blind denoising and waveform coherence enhancement in distributed acoustic sensing data. *IEEE Transactions on Neural Networks and Learning Systems*.

van den Ende, M. P., & Ampuero, J.-P. (2020). Automated seismic source characterization using deep graph neural networks. *Geophysical Research Letters*, *47*(17), e2020GL088690.

Van der Baan, M., & Jutten, C. (2000). Neural networks in geophysical applications. *Geophysics*, *65*(4), 1032–1047.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Vidale, J. E. (1986). Complex polarization analysis of particle motion. *Bulletin of the Seismological society of America*, *76*(5), 1393–1405.

Vidale, J. E., & Houston, H. (1993). The depth dependence of earthquake duration and

implications for rupture mechanisms. *Nature*, *365*(6441), 45–47.

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.

Wang, D., Kawakatsu, H., Zhuang, J., Mori, J., Maeda, T., Tsuruoka, H., & Zhao, X. (2017). Automated determination of magnitude and source length of large earthquakes using backprojection and p wave amplitudes. *Geophysical Research Letters*, *44*(11), 5447–5456.

Wang, H. F., Zeng, X., Miller, D. E., Fratta, D., Feigl, K. L., Thurber, C. H., & Mellors, R. J. (2018). Ground motion response to an ml 4.3 earthquake using co-located distributed acoustic sensing and seismometer arrays. *Geophysical Journal International*, *213*(3), 2020–2036.

Wang, J., & Teng, T.-L. (1995). Artificial neural network-based seismic detector. *Bulletin of the Seismological Society of America*, *85*(1), 308–319.

Wang, J., Xiao, Z., Liu, C., Zhao, D., & Yao, Z. (2019). Deep learning for picking seismic arrival times. *Journal of Geophysical Research: Solid Earth*, *124*(7), 6612–6624.

Wang, T., Trugman, D., & Lin, Y. (2021). Seismogen: Seismic waveform synthesis using gan with application to seismic data augmentation. *Journal of Geophysical Research: Solid Earth*, *126*(4), e2020JB020077.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2019). Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, *38*(5), 1–12.

Wang, Z., & Zhao, D. (2005). Seismic imaging of the entire arc of tohoku and hokkaido in japan using p-wave, s-wave and sp depth-phase data. *Physics of the Earth and Planetary Interiors*, *152*(3), 144–162.

Wang, Z., & Zhao, D. (2006). Suboceanic earthquake location and seismic structure in the kanto district, central japan. *Earth and Planetary Science Letters*, *241*(3-4), 789–803.

Warpinski, N. R., Du, J., & Zimmer, U. (2012). Measurements of hydraulic-fracture-induced seismicity in gas shales. *SPE Production & Operations*, *27*(03), 240–252.

Wei, S. (2014). *Apr./0.1/2014 (mw 8.1), iquique, chile. source models of large earthquakes.*

Wei, S., Graves, R., Helmberger, D., Avouac, J.-P., & Jiang, J. (2012). Sources of shaking and flooding during the tohoku-oki earthquake: A mixture of rupture styles. *Earth and Planetary Science Letters*, *333*, 91–100.

Wen, T.-H., Gasic, M., Mrksic, N., Su, P.-H., Vandyke, D., & Young, S. (2015). Semantically conditioned lstm-based natural language generation for spoken dialogue systems. *arXiv preprint arXiv:1508.01745*.

Wetzler, N., Lay, T., Brodsky, E. E., & Kanamori, H. (2018). Systematic deficiency of aftershocks in areas of high coseismic slip for large subduction zone earthquakes. *Science advances*, *4*(2), eaao3225.

Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from alaska, the western united states, and japan. *Bulletin of the Seismological Society of America*, *90*(4), 859–869.

Williams-Stroud, S., Bauer, R., Leetaru, H., Oye, V., Stanek, F., Greenberg, S., & Langet, N. (2020). Analysis of microseismicity and reactivated fault size to assess the potential for felt events by co2 injection in the illinois basin. *Bulletin of the Seismological Society of America*, *110*(5), 2188–2204.

Wiszniowski, J., Plesiewicz, B. M., & Trojanowski, J. (2014). Application of real time recurrent neural network for detection of small natural earthquakes in poland. *Acta Geophysica*, *62*(3), 469–485.

Woessner, J., & Wiemer, S. (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America*, *95*(2), 684–698.

Wolfson-Schwehr, M., Boettcher, M. S., McGuire, J. J., & Collins, J. A. (2014). The relationship between seismicity and fault structure on the d iscovery transform fault, e ast p acific r ise. *Geochemistry, Geophysics, Geosystems*, *15*(9), 3698–3712.

Woollam, J., Rietbrock, A., Bueno, A., & De Angelis, S. (2019). Convolutional neural network for seismic phase classification, performance demonstration over a local seismic network. *Seismological Research Letters*, *90*(2A), 491–502.

Xiao, Z., Wang, J., Liu, C., Li, J., Zhao, L., & Yao, Z. (2021). Siamese earthquake transformer: A pair-input deep-learning model for earthquake detection and phase picking on a seismic array. *Journal of Geophysical Research: Solid Earth*, *126*(5), e2020JB021444.

Xing, H., Zhonglin, Z., & Shaoyu, W. (2015). The prediction model of earthquake casuailty based on robust wavelet v-svm. *Natural Hazards*, *77*(2), 717–732.

Xu, N., Dai, F., Liang, Z., Zhou, Z., Sha, C., & Tang, C. (2014). The dynamic evaluation of rock slope stability considering the effects of microseismic damage. *Rock Mechanics and Rock Engineering*, *47*(2), 621–642.

Yagi, Y., & Fukahata, Y. (2011). Rupture process of the 2011 tohoku-oki earthquake and absolute elastic strain release. *Geophysical Research Letters*, *38*(19).

Yamazaki, Y., Cheung, K. F., & Lay, T. (2018). A self-consistent fault slip model for the 2011 tohoku earthquake and tsunami. *Journal of Geophysical Research: Solid Earth*, *123*(2), 1435–1458.

Yang, S., Hu, J., Zhang, H., & Liu, G. (2021). Simultaneous earthquake detection on multiple stations via a convolutional neural network. *Seismological Research Letters*, *92*(1), 246–260.

Yao, H., Gerstoft, P., Shearer, P. M., & Mecklenbräuker, C. (2011). Compressive sensing of the tohoku-oki mw 9.0 earthquake: Frequency-dependent rupture modes. *Geophysical Research Letters*, *38*(20).

Yao, H., Shearer, P. M., & Gerstoft, P. (2013). Compressive sensing of frequency-dependent seismic radiation from subduction zone megathrust ruptures. *Proceedings of the National Academy of Sciences*, *110*(12), 4512–4517.

Ye, L., Lay, T., & Kanamori, H. (2013). Ground shaking and seismic source spectra for large earthquakes around the megathrust fault offshore of northeastern honshu, japan. *Bulletin of the Seismological Society of America*, *103*(2B), 1221–1241.

Yin, C. (2017). Fault detection based on microseismic events. *Applied Geophysics*, *14*(3), 363–371.

Yung, S., & Ikelle, L. T. (1997). An example of seismic time picking by third-order bicoher-

ence. *Geophysics*, *62*(6), 1947–1952.

Zerva, A., & Zervas, V. (2002). Spatial variation of seismic ground motions: an overview. *Appl. Mech. Rev.*, *55*(3), 271–297.

Zhan, Z., Cantono, M., Kamalov, V., Mecozzi, A., Müller, R., Yin, S., & Castellanos, J. C. (2021). Optical polarization–based seismic and water wave sensing on transoceanic cables. *Science*, *371*(6532), 931–936.

Zhang, H., Thurber, C., & Rowe, C. (2003). Automatic p-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings. *Bulletin of the Seismological Society of America*, *93*(5), 1904–1912.

Zhang, M., Ellsworth, W. L., & Beroza, G. C. (2019). Rapid earthquake association and location. *Seismological Research Letters*, *90*(6), 2276–2284.

Zhang, X., Curtis, A., Galetti, E., & De Ridder, S. (2018). 3-d monte carlo surface wave tomography. *Geophysical Journal International*, *215*(3), 1644–1658.

Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., & Li, W. (2020). Locating induced earthquakes with a network of seismic stations in oklahoma via a deep learning method. *Scientific reports*, *10*(1), 1–12.

Zhang, Z., Cui, P., & Zhu, W. (2020). Deep learning on graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*.

Zhao, D., Huang, Z., Umino, N., Hasegawa, A., & Kanamori, H. (2011). Structural heterogeneity in the megathrust zone and mechanism of the 2011 tohoku-oki earthquake (mw 9.0). *Geophysical Research Letters*, *38*(17).

Zhao, D., Wang, Z., Umino, N., & Hasegawa, A. (2009). Mapping the mantle wedge and interplate thrust zone of the northeast japan arc. *Tectonophysics*, *467*(1-4), 89–106.

Zhao, J. X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., ... others (2006). Attenuation relations of strong ground motion in japan using site classification based on predominant period. *Bulletin of the Seismological Society of America*, *96*(3), 898–913.

Zhao, Y., & Takano, K. (1999). An artificial neural network approach for broadband seismic phase picking. *Bulletin of the Seismological Society of America*, *89*(3), 670–680.

Zheng, J., Harris, J. M., Li, D., & Al-Rumaih, B. (2020). Sc-psnet: A deep neural network

for automatic p-and s-phase detection and arrival-time picker using 1c recordingsdeep learning for phase auto-pickers. *Geophysics*, *85*(4), U87–U98.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57–81.

Zhou, J., Huang, S., Wang, M., & Qiu, Y. (2021). Performance evaluation of hybrid ga–svm and gwo–svm models to predict earthquake-induced liquefaction potential of soil: a multi-dataset investigation. *Engineering with Computers*, 1–19.

Zhou, Y., Yue, H., Kong, Q., & Zhou, S. (2019). Hybrid event detection and phase-picking algorithm using convolutional and recurrent neural networks. *Seismological Research Letters*, *90*(3), 1079–1087.

Zhu, L., Peng, Z., McClellan, J., Li, C., Yao, D., Li, Z., & Fang, L. (2019). Deep learning for seismic phase detection and picking in the aftershock zone of 2008 mw7. 9 wenchuan earthquake. *Physics of the Earth and Planetary Interiors*, *293*, 106261.

Zhu, W., & Beroza, G. C. (2019). Phasenet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–273.

Zhu, W., Mousavi, S. M., & Berosa, G. (2020). Seismic signal augmentation to improve generalization of deep neural networks. *Advances in Geophysics*, *61*, 151–177.

Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 9476–9488.

Zhu, W., Tai, K. S., Mousavi, S. M., Bailis, P., & Beroza, G. C. (2021). An end-to-end earthquake detection method for joint phase picking and association using deep learning. *arXiv preprint arXiv:2109.09911*.