

SEISMIC STUDIES OF WESTERN PACIFIC SUBDUCTION ZONES WITH
HIGH-PERFORMANCE COMPUTING AND DEEP LEARNING

By

Ziyi Xi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Computational Mathematics, Science and Engineering—Doctor of Philosophy

2024

ABSTRACT

In this dissertation, I explore seismic phenomena across the Western Pacific subduction zones, employing computational techniques to analyze and interpret complex geophysical data. The study is structured into focused chapters, each contributing to a comprehensive understanding of the seismic velocity structures and the mechanisms driving deep earthquakes in this region.

Chapter 1 provides an overview of subduction zones, crucial regions where one tectonic plate is forced below another, leading to significant geological phenomena such as earthquakes and volcanic activity. Focusing on the Western Pacific, the chapter examines the complex interactions at the subduction zones formed by the convergence of the Pacific and Philippine Sea Plates, especially in the northern regions, and the unique geological setting of the Tonga subduction zone and the Lau back-arc basin in the south. These zones are vital for understanding the mechanics of plate tectonics and the resulting seismic activity. Through a detailed examination of the Earth model and seismicity, the chapter sets the groundwork for the dissertation's main scientific inquiries, aiming to unravel the processes governing seismic phenomena in these tectonically active regions using seismic imaging and machine learning techniques.

Chapter 2 introduces EARA2023, a detailed seismic velocity model of the crust and upper mantle beneath East Asia and the northwestern Pacific. This model, constructed through adjoint full-waveform inversion tomography, is based on waveform data from 142 earthquakes and broadband records from around 2000 stations. EARA2023 offers enhanced images of the subducted oceanic plate, detailing complex slab deformations and providing new perspectives on the origins of notable intraplate volcanoes in East Asia, such as Changbaishan, Datong-Fengzhen, Tengchong, and Hainan.

Chapter 3 shifts the focus to the application of machine learning for detecting deep earthquakes, specifically within the Tonga subduction zone. This part of the research outlines the development of a deep-learning-based workflow centered around PhaseNet-TF, a new phase picker designed for the time-frequency domain analysis. This method proves particularly effective for analyzing data from ocean-bottom seismographs (OBSs), improving the detection and picking of P- and S-wave arrivals.

The chapter discusses the integration of this technology with an improved Gaussian Mixture Model Associator and the application of a semi-supervised learning approach to refine earthquake catalogs, thereby enhancing our understanding of the seismic activity in the Tonga region.

Chapter 4 further advances the discussion by examining the use of converted seismic phases, with an emphasis on PS waves from deep earthquakes in the Tonga subduction zone. The chapter explores the application of PhaseNet-TF to identify PS-converted wave arrivals, alongside direct P and S phases, and details the use of beam-forming and a Markov chain Monte Carlo method for associating detected phases. This approach significantly increases the detection of PS arrivals, contributing to a higher resolution imaging of the subducted slab and offering insights into the structural seismology of the Tonga subduction zone.

Collectively, these chapters present a narrative that not only highlights the utility of computational methods in seismic research but also provides a deeper understanding of the mechanisms underlying seismic activity in the Western Pacific. This dissertation contributes to the field of seismology by demonstrating the potential of combining seismic modeling, machine learning, and the analysis of converted seismic phases to investigate subduction zone dynamics and deep earthquake mechanisms.

Copyright by
ZIYI XI
2024

ACKNOWLEDGEMENTS

Embarking on the journey to pursue my PhD in the United States, leaving the familiarity of my homeland behind, has been one of the most rewarding decisions I've made in my early years. This journey has brought countless invaluable individuals into my life, providing immense support both academically and personally. This chapter of my life, enriched with learning and growth, will forever hold a special place in my heart.

I extend my deepest gratitude to my mentors, Dr. Songqiao "Shawn" Wei and Dr. Min Chen. Their guidance transcends the boundaries of academia; they have been pillars of wisdom, teaching me not only the intricacies of research but also the virtues of being a better individual. Their mentorship is something I hold in high regard.

The memory of my first interaction with Min during a PhD program interview remains vivid. Despite my limited research experience and language barriers at the time, Min's patience and warmth made me feel welcomed, opening the door to this incredible learning opportunity. Min's philosophy of keeping an open mind and the importance of communication have been instrumental in my growth. I cherish the moments when she proudly shared my work with others. Although we were constrained to virtual interactions due to the pandemic and Min's unfortunate departure, her legacy continues to inspire me to forge ahead with resilience.

Shawn has been instrumental in my academic journey, providing me with opportunities to delve into exciting research in the Tonga subduction zone. I would greatly appreciate it that Shawn took the responsibility to be my advisor after Min passed away and financially supported the rest of my PhD study. Our weekly discussions have been a source of significant personal and academic growth. Shawn's ability to grasp concepts quickly and offer insightful advice, coupled with our enjoyable casual conversations, has made our meetings highlights of my week. Shawn's support extended beyond academia, encouraging my industry aspirations and facilitating my transition with flexibility. I am immensely grateful for his guidance and meticulous attention to detail in my thesis work.

My committee members, Dr. Allen McNamara, Dr. Jeff Freymueller, and Dr. Rongrong Wang,

deserve a special thank you for their invaluable advice in Earth science and computational techniques. Their courses have greatly enriched my knowledge base. I also extend my gratitude to Dr. Jeroen Tromp for his willingness to offer guidance as an unofficial committee member.

The camaraderie and support from my lab colleagues have been pivotal in my PhD journey. I am thankful for the wisdom shared by Dongdong Tian, Guoliang Li, Jiaqi Li, Ross Maguire, and Tong Zhou, and the enriching collaborations with Fan Wang, Yaqi Jie, Yurong Zhang, and Zhuoran Zhang. Their friendship has made this journey all the more enjoyable.

I am grateful for the companionship and support of my peers at MSU, including Binbin Huang, Hao Wang, He Lyu, Jiaxing Yang, and Kang Yu, Qi Lyu, Shuyang Qin, as well as friends from the EES department like Guanyuan Shuai, Jiaxing Zhang, Keyi Cheng, Mingda Lyu, and Zechao Zhuo.

My heartfelt thanks go out to my collaborators, Dr. Baoshan Wang, Dr. Gregory C. Beroza, Dr. Thomas P. Ferrand, Dr. Weiqiang Zhu, and Dr. YoungHee Kim, for their pivotal contributions and insights. I also wish to acknowledge my undergraduate advisor, Dr. Daoyuan Sun, for his foundational influence and continued support.

I am deeply appreciative of the financial support from Mr. Raymond Ginther and Mrs. Marie Ginther through the Ginther graduate fellowship, the National Science Foundation, and MSU for their fellowships, which have been crucial in my research endeavors.

Above all, my heartfelt gratitude goes to my parents, whose unwavering support and sacrifices have paved the way for my academic and personal growth. Their strength and love have been my constant source of motivation, and I hope to make them proud.

This acknowledgment is a small token of my gratitude to everyone who has been a part of this journey. Each one of you has left an indelible mark on my life, for which I am eternally grateful.

TABLE OF CONTENTS

CHAPTER 1	INTRODUCTION	1
CHAPTER 2	EARA2023: A NEW RADially ANISOTROPIC SEISMIC VELOCITY MODEL FOR THE CRUST AND UPPER MANTLE BENEATH EAST ASIA AND NORTHWESTERN PACIFIC SUBDUCTION ZONES . . .	3
CHAPTER 3	DEEP LEARNING FOR DEEP EARTHQUAKES: INSIGHTS FROM OBS OBSERVATIONS OF THE TONGA SUBDUCTION ZONE	53
CHAPTER 4	DETECTING CONVERTED SEISMIC PHASES OF TONGA DEEP EARTHQUAKES: IN-DEPTH INSIGHTS FROM DEEP-LEARNING METHODS	81
BIBLIOGRAPHY		105

CHAPTER 1

INTRODUCTION

The intricate tectonic landscape of the Western Pacific, highlighted by the convergence of the Pacific and Philippine Sea Plates along notable trenches such as the Kuril, Japan, Izu-Bonin, Mariana, Yap, and Tonga, presents a dynamic setting for seismic studies. This region's active subduction zones, characterized by frequent earthquakes and arc volcanoes, form a pivotal area for understanding the interactions between tectonic plates and the resultant geological processes [Zhang, 1997, Allégre et al., 1984, Yin and Harrison, 2000]. To the north, the collision between the Indian and Eurasian Plates, forming the Tibetan Plateau and Himalayan Mountains, further contributes to the complexity of this geologically diverse area [Bruce Watson and Brenan, 1987, Tapponnier et al., 1986]. To the south, the Tonga subduction zone, in particular, stands out due to its unique geological setting, hosting the coldest and fastest subducting slab on Earth. This area is crucial for investigating deep-focus earthquakes and understanding slab dehydration processes and the interaction between the slab and surrounding mantle [Green and Houston, 1995, Zhan, 2020]. The Tonga trench provides a natural laboratory for studying deep earthquakes, offering insights into slab geometry and the mechanisms behind intermediate-depth and deep-focus seismic events.

Advances in seismic imaging, notably through the development of spectral-element methods (SEM) and full-waveform inversion (FWI), have significantly enhanced our ability to model the Earth's interior with high fidelity [Komatitsch and Tromp, 2002a, Fichtner et al., 2009]. These methods, leveraging high-quality seismic data and considering all usable segments of seismic waveforms, have led to the development of detailed seismic velocity models such as EARA2023, providing valuable insights into the crustal and upper mantle structures beneath East Asia and the northwestern Pacific.

Parallel to seismic imaging, machine learning, especially deep learning, has revolutionized the analysis of seismic data. Techniques like PhaseNet with the adaptation of the U-Net and EQTransformer with the transformer architectures for seismic phase detection and association have markedly improved the efficiency and accuracy of identifying seismic events [Ross et al., 2018, Zhu and

Beroza, 2019, Mousavi et al., 2020]. These developments have been particularly impactful in detecting and analyzing earthquakes in subduction zones, including the challenging offshore regions covered by ocean-bottom seismographs (OBS) and relatively weak seismic converted phases.

The development of seismic imaging and machine learning has been driven by a confluence of scientific inquiries, algorithmic advancements, and significant leaps in computational capabilities. In recent years, computational power has been greatly enhanced by the advent of high-performance computing (HPC) and the availability of large-scale computing resources, such as those provided by the Texas Advanced Computing Center (TACC) and the Extreme Science and Engineering Discovery Environment (XSEDE), SUMMIT at Oak Ridge National Laboratory, and Frontera at the University of Texas at Austin. These resources have enabled the development of sophisticated seismic models and machine learning algorithms, facilitating the analysis of large seismic datasets and the generation of high-resolution seismic images. The algorithm design and implementation have also been influenced by the computing architectures, such as the Graphics Processing Unit (GPU) and the Tensor Processing Unit (TPU), which have accelerated the training of deep learning models and the computation of seismic wave propagation simulations, such as the GPU-accelerated SPECFEM3D_GLOBE [Komatitsch and Tromp, 2002a].

This dissertation employs advanced computational techniques and machine learning to explore the seismic phenomena characterizing the Western Pacific subduction zones. By integrating sophisticated seismic modeling with innovative data analysis methods, this work aims to deepen our understanding of the complex processes driving seismic activity in this region, with a special focus on the Tonga subduction zone. This research will contribute to the broader scientific community's knowledge of subduction zone dynamics and the mechanisms behind deep earthquakes, ultimately enhancing our ability to predict and mitigate seismic hazards in these tectonically active areas. In the meantime, the development of advanced computational techniques and machine learning algorithms will continue to play a crucial role in advancing our understanding of the Earth's interior and the seismic processes shaping our planet. This dissertation will also serve as a testament to the power of computational science in the geosciences.

CHAPTER 2

EARA2023: A NEW RADIALY ANISOTROPIC SEISMIC VELOCITY MODEL FOR THE CRUST AND UPPER MANTLE BENEATH EAST ASIA AND NORTHWESTERN PACIFIC SUBDUCTION ZONES

An edited version of this chapter is under review at Geophysical Journal International. Ziyi Xi, Min Chen, Songqiao Shawn Wei, Jiaqi Li, Tong Zhou, Baoshan Wang, and YoungHee Kim.

EARA2023: A New Radially Anisotropic Seismic Velocity Model for the Crust and Upper Mantle beneath East Asia and Northwestern Pacific Subduction Zones

2.1 Abstract

We present a new 3-D radially anisotropic seismic velocity model EARA2023 of the crust and mantle beneath East Asia and the northwestern Pacific using adjoint full-waveform inversion tomography. We construct the EARA2023 model by iteratively minimizing the waveform similarity misfit between the synthetic and observed waveforms from 142 earthquakes recorded by about 2,000 broadband stations in East Asia. Compared to previous studies, this new model renders significantly improved images of the subducted oceanic plate in the upper mantle, mantle transition zone, and uppermost lower mantle along the Kuril, Japan, Izu-Bonin, and Ryukyu Trenches. Complex slab deformation and break-offs are observed at different depths. Moreover, our model provides new insights into the origins of intraplate volcanoes in East Asia, including the Changbaishan, Datong-Fengzhen, Tengchong, and Hainan volcanic fields.

2.2 Introduction

East Asia and the northwestern Pacific subduction zones consist of various tectonic structures and host diverse and active geological processes in the crust and upper mantle (Figure 2.1). In the east, the Pacific Plate subducts westward along the Kuril, Japan, Izu-Bonin, Mariana, and Yap Trenches, and the Philippine Sea Plate subducts along the Nankai, Ryukyu, Malina, and Philippine Trenches. These subduction zones, with numerous earthquakes and arc volcanoes, are one of the most geologically active regions in the world. In the middle, the Xing'an-East Mongolia Block, North China Block, and South China Block form the east-central end of the Eurasian continent

[Zhang, 1997]. In the west, the collision between the Indian and Eurasian Plates forms the Tibetan Plateau and Himalayan Mountains [Allégre et al., 1984, Yin and Harrison, 2000]. The complex geological structure and active tectonics in this region provide a natural laboratory for studying inter-plate and intra-plate tectonics. For instance, the collision of the Indian Plate and Eurasian Plate results in the Karakorum, Kunlun, Altyn-Tagh Faults and the Tarim, Junggar, and Qaidam Basins [Bruce Watson and Brennan, 1987]. The resistance of the Yangtze Block in southern China to the subducted Indian Plate causes numerous intra-plate earthquakes in southwestern China [Tapponnier et al., 1986]. Furthermore, the subducted Pacific Plate in the mantle transition zone (MTZ) also has significant influences on the evolution of the East Asian lithosphere [Zhao et al., 2007, 2009, Lei et al., 2013, He et al., 2014, Tang et al., 2014, Guo et al., 2016, 2018].

High-resolution seismic images provide valuable information about the crustal and upper mantle structures and geological processes in this region. Since decades ago, body-wave traveltimes tomography, based on either the high-frequency approximation or 3-D finite-frequency kernels of seismic wave propagation, has been widely used to image East Asia [Aki and Lee, 1976, Dziewonski et al., 1977, Friederich, 2003, Lei and Zhao, 2005, 2016, Huang and Zhao, 2006, Zhao et al., 2007, 2009, Kustowski et al., 2008b, Lei et al., 2009, Tian et al., 2009, Li and Van Der Hilst, 2010, Obrebski et al., 2012]. In addition, ambient noise, surface wave dispersion, and Rayleigh wave ellipticity also have been used to image the subsurface structure beneath East Asia [Yao et al., 2006, 2008, Zheng et al., 2006, Witek et al., 2014, Bao et al., 2015, Li et al., 2016, Yang et al., 2019]. Finite-difference methods of simulating wave propagation provide another tool for studying the East Asian crust and upper mantle [Wang et al., 2013].

The development of the spectral-element method (SEM) provides another way to simulate wave propagation and naturally considers the finite-frequency effects [Komatitsch and Tromp, 2002a,b]. Its combination with the adjoint state method provides an effective way to compute the gradients of 3-D model parameters and thus can be used for tomographic imaging. The full-waveform inversion (FWI) method based on SEM and adjoint state methods images the Earth's interior by iteratively applying SEM to simulate seismic waveforms in 3-D models and then updating the model with

finite-frequency kernels. By considering all usable segments of seismic waveforms, this method uses much more information than other seismic tomography methods, thus potentially enhancing the model resolution. Furthermore, there is no need for a crustal correction, a common practice in conventional tomography methods, because the crust is explicitly included in the simulations. This approach also makes it easy to consider the effects of the Earth’s ellipticity, bathymetry, topography, and internal discontinuities in the full waveform simulation. Several global or regional models have been developed using this method [Tape et al., 2007, Fichtner et al., 2009, 2010, Tape et al., 2009, 2010, Zhu et al., 2012a,b, Lei et al., 2013, French and Romanowicz, 2014, Chen et al., 2015, Bozdağ et al., 2016, Tao et al., 2018, Lei et al., 2020, Zhou et al., 2022, Ma et al., 2022].

In this work, we use SPEC-FEM3D_GLOBE [Komatitsch and Tromp, 2002a,b, Peter et al., 2011] to invert 10 years of high-quality seismic data collected from East Asia and the northwestern Pacific to image the crustal and mantle structure. This paper presents our latest FWI tomography model of this region: EARA2023. First, we will introduce our dataset and methods. Second, we will assess our model with resolution tests and waveform comparisons. Third, we will describe the model and compare our model with previous models. Last, we will discuss slab morphology in northwestern Pacific subduction zones and the origin of four intra-plate volcanoes based on our new model.

2.3 Data and Method

2.3.1 Data selection and processing

We defined our study region in East Asia and the northwestern Pacific as a nearly square region centered at 36°N , 123°E , with a length of 60° on each side, and anti-clockwise rotated by 10° (Figure 2.1). We selected 142 earthquakes with M_w 5.2–7.0 from the global centroid moment tensor (CMT) catalog [Ekström et al., 2012] between 2008 and 2018 in our study region (Figure 2.2). These earthquakes were carefully chosen to provide balanced data coverage. We excluded events less than 2 degrees away from the study region boundary to further reduce the numerical noise from the boundary effect, even though Clayton–Engquist–Stacey absorbing conditions [Clayton and Engquist, 1977, Stacey, 1988] had been used in our simulations. 58 of these 142 earthquakes were

deeper than 150 km for better constraining deep mantle structures, such as the subducted slabs and MTZ. We also excluded events with M_w larger than 7.0 or source half duration time longer than 7 s to avoid complex rupture processes that might complicate the structural inversion. Consequently, our earthquake catalog consisted of 142 events in the Kuril, Japan, Izu-Bonin, Mariana, Nankai, Ryukyu, Manila, and Philippine subduction zones down to about 650 km depth, as well as continental events shallower than 30 km in the Eurasian lithosphere.

We used three-component seismic waveforms recorded by about 2,000 stations from the China Earthquake Administration Array (CEArray) [Zheng et al., 2009], F-net [Okada et al., 2004], NECESSArray [Tang et al., 2014], INDEPTH IV project [Wei et al., 2010, Yue et al., 2012], the central Mongolia seismic experiment [Meltzer et al., 2019], the Korean seismic network and other regional or global seismic networks. More than 90% of the stations had broadband (BH and HH) seismometers, whereas other stations had short-period high-gain (SH) seismometers. After detrending and tapering, we first removed the instrumental response to get displacement waveforms in three components: vertical (Z), east (E), and north (N). We then rotated the two horizontal components (E and N) to the radial (R) and tangential (T) components. In the next step, the R and T components of stations with misorientation were corrected according to the methods described in previous studies for CEArray [Niu and Li, 2011] and the Korean seismic network [Lim et al., 2018]. In the first inversion stage, we applied a second-order zero-phase Butterworth bandpass filter between 10 and 40 s for body waves and 40 to 120 s for surface waves. In the second stage, we tried to achieve higher resolution with higher-frequency signals and filtered body waves at 8–40 s and surface waves at 30–120 s. In the third stage, we kept the same bandpass filter range for body waves as the second stage, and filtered surface waves at 30–120 s. Finally, we down-sampled the filtered data to 10 samples per second.

2.3.2 Configuration of the Spectral-Element-Method Simulation

We use the spectral-element method [Komatitsch and Tromp, 2002a,b] to simulate wave propagation and compute sensitivity kernels based on the adjoint state method. In addition, we use the conjugate gradient method to maximize the waveform similarity and phase matching between

data and synthetics. The 3-D model space extends from the Earth’s surface, as delineated by the Crust1.0 model [Laske et al., 2013], down to the center of the Earth. We construct the initial model by combining two regional models, FWEA18 [Tao et al., 2018] and EARA2014 [Chen et al., 2015], embedded in two global models, S362ANI [Kustowski et al., 2008b] for the mantle and CRUST1.0 for the crust. FWEA18, a radially anisotropic model, contributes to the majority of this initial model. In regions not covered by FWEA18, the initial model relies on EARA2014 instead. This approach leverages the consistency and compatibility of the methodologies behind these models and is driven by the progression in model development, where each new model builds upon the previous one, supplemented by additional data. It is highly likely that the most recent model provides a better estimation of the real Earth. By starting with a model that is as close to reality as possible, we aim to enhance the efficiency and accuracy of our inversion process. Our model construction process involved merging the velocity perturbations of FWEA18 and EARA2014 relative to S362ANI in the mantle and to Crust1.0 in the crust. To facilitate a smooth integration of these models and mitigate numerical instability, we apply Gaussian smoothing to the combined perturbations. The extent of these models are shown in Figure 2.25. We fix the topography of the 410- and 660-km discontinuities to that from S362ANI and the attenuation structure to that from QL6 [Durek and Ekström, 1996] during the inversions.

In the crust and upper mantle, we use five parameters to parameterize our model: density ρ , isotropic bulk sound speed V_C , horizontally propagating S wave speeds V_{SV} (polarized in the vertical direction) and V_{SH} (polarized in the horizontal direction), and a non-dimensional parameter η measuring the anellipticity of P wave phase slowness. Radial anisotropy is introduced to the crustal part of our model because previous studies indicate the presence of radial anisotropy in this region [Witek et al., 2021, Xie et al., 2013, Tao et al., 2018]. Since previous studies [e.g., Panning and Romanowicz, 2006] suggest that the upper part of the lower mantle is nearly isotropic, we parameterize it with isotropic P wave speed V_P , isotropic S wave speed V_S , and density ρ . Because density ρ cannot be well constrained by our data, we scale ρ with isotropic shear wave speed perturbation $\delta \ln V_S$ as $\delta \rho = 0.33 \delta \ln V_S$ [Panning and Romanowicz, 2006]. Here the perturbation reference

model STW105 [Kustowski et al., 2008a] is crucial in the inversion process, as we update model perturbations rather than the absolute values of model parameters.

The horizontal dimension of the spectral element is set as ~ 20 km in the crust and doubled in the mantle, and the vertical dimension is ~ 10 km in the crust and ~ 35 km in the mantle. The shortest period for waveform forward modeling is 8 s. Each spectral element contains 125 Gauss-Lobatto-Legendre interpolation points. In total, the simulation involves about 2.3 million elements. To simulate a 30-minute wave propagation for one event, the SPEC-FEM_GLOBE takes ~ 30 minutes wall time for the forward simulation and ~ 70 minutes for sensitivity kernel calculation using 441 cores on the SKX nodes of the Stampede2 supercomputer at the Texas Advanced Computing Center (TACC). The total volume of seismic data and synthetics in each iteration is ~ 800 GB. To improve I/O efficiency, we develop a Python-based data analysis and inversion workflow based on the Adaptable Seismic Data Format (ASDF) [Krischer et al., 2016] and ObsPy [Beyreuther et al., 2010] to store and process the seismic data.

2.3.3 Waveform Misfit Measurement

We group our measurements into 6 categories: P-SV body waves on the vertical and radial components, SH body waves on the transverse component, Rayleigh waves on the vertical and radial components, and Love waves on the transverse component. We use a conjugate gradient method to iteratively minimize the misfit between recorded waveforms and synthetics. More specifically, we define the misfit function as a normalized zero-lag cross-correlation coefficient (NZCC) because this coefficient reflects the model based on waveform similarity and thus works more effectively compared to other misfit functions that only use traveltime information [Tao et al., 2017, Zhou et al., 2021]. The NZCC for each measurement window is defined as:

$$\text{NZCC}^{c,e,s,w} = \frac{\int s^{c,e,s,w}(t) \cdot u^{c,e,s,w}(t) dt}{\int s^{c,e,s,w}(t) dt \cdot \int u^{c,e,s,w}(t) dt} \quad (2.1)$$

where $s(t)$ is the synthetic and $u(t)$ is the observed waveform in category c and window w for event e recorded by station s . We measure all major body-wave and surface-wave phases, including P, pP, sP, PP, S, sS, SS, ScS, Rayleigh wave, and Love wave.

Compared with other misfit measurements that utilize dispersion or frequency-dependent phase arrival time, the NZCC is a fixed frequency measurement method. Previous studies introduce several waveform inversion strategies to address dispersion, including generalized seismological data functionals [Gee and Jordan, 1992], multi-taper delay time measurements [Tape et al., 2010], instantaneous phase and envelope misfits [Bozdağ et al., 2011, Rickers et al., 2012], exponentiated phase [Yuan et al., 2020], and time-frequency phase misfit for waveform misfit objectives [Fichtner et al., 2009, 2010, 2013, Krischer et al., 2018]. However, these methods are predominantly focused on the first arrival time. The arrival time misfit, dominated by the first arrival’s large amplitude, means that the corresponding adjoint sources do not account for waveform distortions associated with later arrivals. In contrast, NZCC considers the entire waveform window, and thus offers better performance in complex scenarios, as evidenced by both synthetic tests [Tao et al., 2017] and real-data studies [Tao et al., 2018, Zhou et al., 2021, 2022]. To take frequency-dependent arrival times into consideration, we utilize different frequency bands at various stages of our analysis.

Following Tao et al. [2018], we select a measurement window based on a predicted traveltimes because this strategy is more computationally efficient compared to the conventional FLEXWIN method [Maggi et al., 2009]. We define the body-wave measurement windows from 20 s before to 50 s after the predicted phase arrival times based on the AK135 model [Kennett et al., 1995]. If multiple body-wave windows overlap with each other, we merge them into a single one. For surface waves, the measurement window starts 50 s before a predicted phase arrival time, assuming a wave speed of 4.6 km/s, and ends 50 s after a predicted phase arrival time, assuming a wave speed of 3.3 km/s. We only measure surface waves from events shallower than 150 km. In total, we measure ~650,000 body-wave windows and ~150,000 surface-wave windows that satisfy our criteria of signal-to-noise ratio (SNR), cross-correlation coefficient (CC), and time shift. The three criteria are described in Appendix 2.7.1 in detail. It is worth noticing that the absolute amplitude information is not taken into account because the NZCC is normalized. Therefore, our inversion is insensitive to seismic attenuation even though it may slightly distort waveforms. Therefore, we only focus on the Earth’s elastic properties in this study.

We apply categorical and geographical weighting (Table 2.1) to balance the misfits from different categories and regions. The weighted misfit function is

$$\chi = \sum_{c=1}^{N_c} W_c \sum_{e=1}^{N_e} \sum_{s=1}^{N_{e,s}} W_{e,s} \sum_{w=1}^{N_{c,e,s,w}} W_w (1 - \text{NZCC}^{c,e,s,w}) \quad (2.2)$$

where N_c represents the number of categories (i.e., 6) used in our inversion, N_e indicates the total number of events, $N_{e,s}$ is the number of stations for event e , and $N_{c,e,s,w}$ is the number of measurement windows in the corresponding category c for event e and station s . W_c is a categorical weighting term, balancing the misfit contribution from different categories. $W_{e,s}$ is a geographical weighting term, balancing the misfit contribution from different regions and mitigating the bias due to nonuniform station distribution [Ruan et al., 2019]. W_w is a weighting term to control the waveform quality in each measurement window.

The corresponding adjoint source of NZCC is defined as

$$\delta\chi = \sum_{c=1}^{N_c} W_c \sum_{e=1}^{N_e} \sum_{s=1}^{N_{e,s}} W_{e,s} \sum_{w=1}^{N_{c,e,s,w}} W_w \frac{\partial \chi^{c,e,s,w}}{\partial t} \delta t \quad (2.3)$$

where:

$$\frac{\partial \chi^{c,e,s,w}}{\partial t} = \frac{s^{c,e,s,w}(t) - A^{c,e,s,w} u^{c,e,s,w}(t)}{W^{c,e,s,w}} \quad (2.4)$$

$A^{c,e,s,w}$ is defined as

$$A^{c,e,s,w} = \frac{\int s^{c,e,s,w}(t) dt \int u^{c,e,s,w}(t) dt}{\int |u^{c,e,s,w}(t)|^2 dt} \quad (2.5)$$

And the normalization factor $W^{c,e,s,w}$ is defined as

$$W^{c,e,s,w} = \sqrt{\int |s^{c,e,s,w}(t)|^2 dt \int |u^{c,e,s,w}(t)|^2 dt} \quad (2.6)$$

The structural inversion details will be described below in Section 2.3.5.

2.3.4 Source Inversion

Earthquake source parameters also influence seismic waveforms, and thus we have to invert for source and structure iteratively. Here we use the steepest descent method to optimize the source model and the adjoint state method to compute the gradients following Kim et al. [2011]. The misfit

function and the adjoint source of event e for the source inversion are defined in similar ways:

$$\chi_e = \sum_{c=1}^{N_c} W_c \sum_{s=1}^{N_{e,s}} W_{e,s} \sum_{w=1}^{N_{c,e,s,w}} W_w (1 - NZCC^{c,e,s,w}) \quad (2.7)$$

$$\frac{\partial \chi_e}{\partial t} = \sum_{c=1}^{N_c} W_c \sum_{s=1}^{N_{e,s}} W_{e,s} \sum_{w=1}^{N_{c,e,s,w}} W_w \frac{\partial \chi^{c,e,s,w}}{\partial t} \quad (2.8)$$

The subscripts and superscripts here have the same definitions as those in Section 2.3.3. Categorical weighting is defined only for a single event:

$$W_c = \frac{1}{\sum_{s=1}^{N_{e,s}} N_{c,e,s,w}} \quad (2.9)$$

The invertible source parameters include centroid moment tensor M , 3-D centroid location X , and source time function $S(t)$. Readers are referred to Appendix 2.7.2 for technical details.

2.3.5 Inversion Process

The inversion process is divided into three stages, and each stage includes source inversions followed by structure inversions using a preconditioned conjugate gradient method [Polak and Ribiere, 1969]. In the first stage, after 3 iterations of source inversion based on the initial model m_{00} , the source parameters remain unchanged, and we perform 5 iterations of structural inversion, using body waves filtered at 10-40 s and surface waves filtered at 40-120 s. Then we change the dimension of the Gaussian smoother to sensitivity kernels from 50 km horizontally and 25 km vertically to 25 km horizontally and 10 km vertically for another 5 iterations. In the second stage, we start from the output model from the first stage and use higher-frequency body waves filtered at 8-40 s and surface waves filtered at 30-120 s. In this stage, after 3 iterations of source inversion, we perform another 10 iterations of structure inversion. In the third stage, we have also performed the same number of inversions as the second stage, but the surface waves are further filtered at 20-120 s. In sum, we run 9 iterations of source inversion and 30 iterations of structure inversion. Readers are referred to Appendix 2.7.3 for technical details about the structure inversion. We conclude each stage after 10 iterations for efficiency. Although the inversion does not necessarily converge at the end of each stage, the misfit reduction is small with additional iterations (Figure 2.21). More importantly, the subsequent stage can further reduce the data misfit with the frequency band used

in the previous stage. Figure 2.3 shows that the inversions converge after 30 iterations where misfit reduction becomes minimal. The whole inversion process consumes about 4.5 million CPU hours on Stampede2.

2.4 Results and Model Assessment

2.4.1 Average V_P and V_S models

The final structure model is named EARA2023 (East Asia Radial Anisotropic Model 2023). Average V_P and V_S are calculated following Panning and Romanowicz [2006]:

$$V_S^2 = \frac{2V_{SV}^2 + V_{SH}^2}{3} \quad (2.10)$$

$$V_P^2 = \frac{V_{PV}^2 + 4V_{PH}^2}{5} \quad (2.11)$$

The model shows $\delta \ln V$ perturbations as large as 6%. The $\delta \ln V_S$ perturbations are typically larger than the $\delta \ln V_P$ perturbations, but both show similar large-scale features (Figures 2.4 and S1). S-wave radial anisotropy $(V_{SH} - V_{SV})/V_S$ is about 3% in the lithosphere and varies from -3% to 3% in the asthenosphere (Figure 2.5). For the rest of this paper, we focus on the S-wave model because it is better resolved than the P-wave model due to the constraints from surface waves. As discussed by Tao et al. [2018], thin anomalies ~ 60 km above and below the 410- and 660-km discontinuities are likely artifacts due to the discontinuity topography model from S362ANI [Kustowski et al., 2008a] that is fixed during inversions. For the same reason, some high-velocity anomalies, interpreted as subducted slabs, seem to be split into two parts separated by a thin low-velocity layer in between, with the upper part in the MTZ and the lower part in the uppermost lower mantle. These are common artifacts shown in previous models using the same approach of handling discontinuities [Tao et al., 2018, Lei et al., 2020], and should not be over-interpreted. To avoid over-interpretation, we place semi-transparent bands ± 60 km around the 410- and 660-km discontinuities to mask potential artifacts due to the discontinuity topography.

The EARA2023 model reveals several important tectonic structures similar to previous models in this region [Chen et al., 2015, Tao et al., 2018]. Figure 2.4 shows high-velocity anomalies beneath the Kuril, Japan, Izu-Bonin, and Ryukyu subduction zones, as well as the Sichuan Basin, eastern

Tibet Plateau, and Ordos Block in China at lithospheric depths. Compared to previous models, one obvious improvement is the high-resolution structure of the Japan slab beneath the Changbaishan volcano in the MTZ. Figure 2.6 compares cross-sections of multiple tomography models EARA2023, FWEA18 [Tao et al., 2018], EARA2014 [Chen et al., 2015], GLAD_M25 [Lei et al., 2020], SPiRaL [Simmons et al., 2021], and GAP_P4 [Obayashi et al., 2013]. In global models (GLAD_M25, SPiRaL, and GAP_P4), the flat slab appears to be continuous in the MTZ. However, Tang et al. [2014] used regional traveltimes tomography to first image a gap in the flat slab beneath the Changbaishan volcanoes. Over the course of developing FWI models EARA2014, FWEA18, and EARA2023, this gap structure gradually emerges and becomes more complex than previously thought. The EARA2023 model shows that there are at least 2 holes in the flat slab at 125°E and 129°E. Their implications will be discussed in Section 2.5.2.

Here we only highlight a few complex crustal and upper mantle structures beneath the Tibetan Plateau, Sichuan Basin, and North China Craton (Figure 2.7) while keeping thorough interpretations for future studies. The Tibetan Plateau is the world's highest and largest plateau, resulting from the ongoing collision of the Indian and Eurasian Plates. The Sichuan Basin, located to the east of the plateau, is a sedimentary and tectonic basin. We observe high-velocity anomalies in the upper mantle beneath the Tibetan Plateau and Sichuan Basin (Figures 2.7a, 2.7b, and 2.7c). The upper crust of the Sichuan Basin is characterized by high-velocity anomalies (Figure 2.20). West of 100°E, we image a large high-velocity anomaly zone with a maximum $\delta \ln V_S$ perturbation of 4%, potentially indicating a subducted plate reaching the MTZ and uppermost mantle (see discussions in Section 2.5.1). Moving eastwards, another high-velocity anomaly is observed below the Sichuan Basin at about 100 km depth, with a maximum $\delta \ln V_S$ perturbation of 6%, representing the western edge of the South China Block and extending to a depth of around 300 km. We find significant low-velocity anomalies in the crust and uppermost mantle beneath the North China Craton (NCC) (Figures 2.7d and 2.7e), a stable continental region that has experienced significant tectonic deformation and lithosphere thinning [Zhu et al., 2011]. Beneath the Huabei Plain, the eastern part of the NCC, the strong low-velocity anomaly at 80–150 km depths reveals a thinned lithosphere above it.

In contrast, the lithosphere beneath the Ordos Block, indicated by a high-velocity anomaly, extends to at least 200 km depth.

In summary, our study provides a detailed model of the crustal and upper mantle structures beneath different tectonic features in East Asia and the northwestern Pacific. These observations offer valuable insights into the underlying geological processes in this region.

2.4.2 Model Assessment

To assess the model, we first compute the phase time residual (ΔT) between synthetic (Iteration #30) and observed waveforms. Figure 2.8 shows that the EARA2023 model has significantly reduced phase time residuals, particularly for surface waves, compared to the initial model. Additionally, the final model shows improved waveform similarity, as measured by normalized zero-lag cross-correlation coefficient (NZCC) and cross-correlation coefficient (CC). NZCC measures both phase time shift and similarity, while CC only measures phase similarity. These results demonstrate that the final model provides a more accurate representation of the seismic structure in this region.

To further assess the model resolution, we conduct a point spread function (PSF) test [Fichtner and Trampert, 2011b], a widely used method for resolution tests in FWI studies [Chen et al., 2015, Tao et al., 2018, Zhou et al., 2022]. Compared to traditional checkerboard methods, which are computationally expensive for FWI, the PSF test provides a more efficient way to evaluate smoothing, distortion, and parameter trade-offs in the inversion process. The PSF test results are shown in Figure 2.9. For a checkerboard-shaped model perturbation to V_{SV} (Figure 2.9a), we expect the PSF response for V_{SV} to resemble the input in high-resolution regions. The V_{SV} response (Figure 2.9b) shows that the EARA2023 model has good resolutions in central-eastern China, Korea, the Japan Sea, the Yellow Sea, and the East China Sea at 100 km depth. However, due to the event distribution, the V_{SV} PSF response is distorted in Mongolia, the South China Sea, and the Izu-Bonin-Mariana Arc. At greater depths, the V_{SV} PSF response remains similar to that at 100 km but with smaller amplitudes. At depths greater than 900 km, the resolution is low, and the model should not be over-interpreted. Ideally, perturbations to V_{SV} should not affect V_{SH} , and thus the PSF response

for V_{SH} should be 0. However, the results show non-zeros V_{SH} response (Figure 2.9c), suggesting weak trade-offs between V_{SV} and V_{SH} , as they both rely on the same elastic moduli. In other words, a small portion of V_{SV} heterogeneity is artificially mapped into V_{SH} , slightly biasing radial anisotropy. Similarly, Figure 2.9d shows the bulk velocity V_C response being nearly 0 everywhere (Figure 2.9d), suggesting that the inversion can perfectly distinguish V_S and V_C . More technical details of the PSF tests are presented in Appendix 2.7.4.

Waveform comparisons between observed waveforms, synthetic waveforms based on the initial structure and source models, and synthetic waveforms based on the final EARA2023 and source models also demonstrate the improvements. Figure 2.10 shows an example of such comparisons for a 21-km deep event on May 7, 2008 (GCMT ID: 200805071602A) in the Japan subduction zone recorded by a CEArray station (JX.JIJ) in southeastern China. The results show that the final models provide better waveform fitting than the initial models, particularly for surface waves. Figure 2.19 offers another example of fitting triplicated waveforms with the final models. Additionally, Figures 2.22 and 2.23 provide independent validation of our models using recent earthquakes (GCMT ID: 202307011438A and 202306220124A) which are not used for inversions. The waveforms of these earthquakes at several GSN stations are better fitted with our final models.

Since the final models integrate both source and structural inversions, it is crucial to distinguish the improvements attributable to source inversions and structural adjustments. Figure 2.8 compares histograms of ΔT , NZCC, and CC for three different cases: (a) using the initial structural and source models; (b) using the initial structural model with inverted source models after 3 iterations of the source inversion updates in the first stage of inversions; and (c) using the final structural and source models. This approach highlights that while the source inversions mainly adjust phase time shifts ΔT , structural inversion plays a significant role in improving NZCC and CC. Additionally, Figure 2.24 offers a comparative analysis between observed waveforms and synthetics produced for the three different cases mentioned above. This comparison shows the collaborative improvements for the waveform fitting through both the source and structure inversions.

Moreover, we use the phase time residual ΔT as a key indicator of source parameter variations, as

ΔT is mostly influenced by the change of event origin time in the source inversions. For body waves, the source inversions greatly reduce the mean and standard deviation of ΔT . For surface waves, which are more sensitive to the crustal and mantle structures, our analysis leads to insignificant improvement of the mean and standard deviation of ΔT in the source inversions. Therefore, the structure inversions should have contributed most of the improvements for the crustal and mantle structures. Therefore, we are confident with the EARA2023 model in capturing seismic structures beneath East Asia and northwestern Pacific subduction zones.

2.5 Discussion

2.5.1 Slab morphology of Northwestern Pacific Subduction Zones

The northwestern Pacific region hosts a number of subduction zones. The Pacific Plate subducts westwards beneath the Eurasian Plate along the Kuril, Japan Trenches and beneath the Philippine Sea Plate along the Izu-Bonin, Mariana, and Yap Trenches, whereas the Philippine Sea Plate subducts beneath the Eurasian Plate northwestwards along the Nankai, Ryukyu, Malina, and Philippine Trenches. Numerous earthquakes and arc volcanic activities occur in these subduction zones. Figure 2.11 presents vertical cross-sections of the $\delta \ln V_S$ perturbation for the Kuril, Japan, Izu-Bonin, Nankai, and Ryukyu subduction zones. We also create a 3-D slab model in this region by identifying all fast- V_S anomalies of $> 1.5\%$ (Figure 2.12).

The Kuril, Japan, and Izu-Bonin slabs are the subducted Pacific Plate along the Kuril, Japan, and Izu-Bonin Trenches, respectively. The Kuril slab is characterized by a nearly constant dipping angle of $\sim 36^\circ$ and a significant maximum 8% $\delta \ln V_S$ perturbation reaching the 660-km discontinuity (Figure 2.11a). A substantial low-velocity anomaly zone above 200 km depth represents the vast mantle wedge, with a maximum $\delta \ln V_S$ perturbation of 4% . The Japan slab, characterized by high-velocity anomalies, subducts with a shallower dipping angle of $\sim 25^\circ$ in the upper mantle and is deflected and extends horizontally in the mantle transition zone (MTZ) for at least 800 km. There are two holes within the slab at $\sim 125^\circ\text{E}$ and 129°E (Figures 2.11c). Their implications and relationship to Changbaishan volcanoes will be discussed in Section 2.5.2. The Japan slab may extend further to $\sim 115^\circ\text{E}$, with a big slab hole between $\sim 120^\circ\text{E}$ and $\sim 123^\circ\text{E}$. Figure 2.11b shows the transition

between the Kuril slab to the Japan slab, where the subducted Pacific plate starts to flatten in the MTZ. It is worth noting that the Japan slab beneath the East China Sea (Figure 2.11g) is much thicker than that beneath Northeast China (Figure 2.11c), extending into the uppermost lower mantle.

The mantle wedge above the Japan slab shows a low $\delta \ln V_S$ anomaly of as much as -6%, reaching depths between 200 and 300 km. Another low-velocity anomaly is detected beneath the Japan slab in the MTZ, with a peak $\delta \ln V_S$ anomaly of -2%. Because the major part of this anomaly is well above the 660-km discontinuity, we believe it is a real feature and may indicate sub-slab partial melting that is entrained by subduction from the shallow depths [Kawakatsu et al., 2009, Wang et al., 2020].

The Izu-Bonin slab, the Pacific Plate subducted beneath the Philippine Sea Plate along the Izu-Bonin Trench, is characterized by a maximum $\delta \ln V_S$ anomaly greater than 6%, and has a steeper dipping angle of $\sim 40^\circ$ in the upper mantle. The slab appears flat but heavily deformed in the mid-mantle, with pieces of high-velocity anomalies at varying depths in the MTZ and uppermost lower mantle (Figures 2.11d and 2.11e). The slab deformation is particularly clear in the southern part of the Izu-Bonin slab, although we lose the resolution of the Mariana slab to the further south. Our observations of high-velocity anomalies are consistent with a previous study [Zhang et al., 2019a] that interpreted the anomalies as evidence of a single contiguous slab undergoing tearing.

The Philippine Sea Plate is much younger and thus warmer than the Pacific Plate in this region. As the Philippine Sea Plate subducts beneath the Eurasian Plate along the Nankai Trough and Ryukyu Trench, the Nankai and Ryukyu slabs show distinct characteristics. The Nankai slab, one of the warmest slabs globally, subducts with a dipping angle of $\sim 20^\circ$ to 200 km depth, with a maximum $\delta \ln V_S$ perturbation of only 1% (Figure 2.11f). Consequently, the Nankai slab does not appear in our 3-D slab morphology model (Figure 2.13), which shows the 1.5% isosurface of high-velocity anomalies. The Ryukyu slab subducts with a nearly constant dipping angle of $\sim 50^\circ$ and extends to ~ 500 km depth (Figure 2.11g), consistent with previous studies [Li et al., 2008, Tao et al., 2018]. Wu et al. [2016] interpret the thick high-velocity anomaly in the MTZ and uppermost lower mantle beneath the Yellow Sea as the subducted Philippine Sea Plate overlying the Pacific Plate. However,

our model shows a ~300-km-wide gap between the Ryukyu slab beneath the East China Sea and the Pacific Plate beneath the Yellow Sea. Therefore, the thick high-velocity anomaly beneath the Yellow Sea is more likely the Japan slab ponding around the 660-km discontinuity, as suggested by Tao et al. [2018].

Similar to Tao et al. [2018], we also image a NE-SW striking high-velocity anomaly in the MTZ beneath the South China Block. It appears to connect to the Japan slab beneath the Yellow Sea. However, this connection may be an artifact due to low resolution, as there is no evidence suggesting that the subducted Pacific Plate reaches this far to the southwest. Tao et al. [2018] speculate that this anomaly is related to the subduction of a proto-South China Sea, although a plate reconstruction model places this structure further to the east [Wu et al., 2016].

2.5.2 Intra-plate Volcanoes

Intra-plate volcanoes are intriguing because they cannot be easily explained by the classic theory of plate tectonics. Some of them, e.g., the Hawaiian volcanoes, are related to mantle plumes originating from the lower mantle [e.g., Wolfe et al., 2009, French and Romanowicz, 2015]. Some of them, e.g., Bermuda Island, are rooted in the MTZ [e.g., Mazza et al., 2019]. However, the origins of other intra-plate volcanoes, especially those on continents, e.g., the Yellowstone volcano, are under debate [Nelson and Grand, 2018, Zhou, 2018, Zhou et al., 2018]. Here we focus on four intra-plate volcanic fields, Datong-Fengzhen, Tengchong, Changbaisan, and Hainan in East Asia, and try to explore their origins. These volcanoes are far away from any active subduction zones and thus are unlikely to be related to arc or back-arc volcanism. They are also distant from any known divergent plate boundaries.

The Datong-Fengzhen volcanoes in northern China predominantly consist of Cenozoic basaltic lava flows and pyroclastic deposits [Xu et al., 2005, Smithsonian-Institution, 2023]. Lei [2012] image the upper mantle beneath the North China Craton with teleseismic wave traveltime tomography and suggest that Datong-Fengzhen volcanoes are generated by hot materials upwelling from the lower mantle. In contrast, Kim et al. [2021] image the S-wave velocity structure of East Asia with teleseismic traveltime tomography and suggest that these volcanoes are generated by slab-induced

dehydration of the hydrous mantle transition zone. Moreover, by imaging mantle discontinuities beneath eastern North China with receiver functions, Zuo et al. [2020] find a normal MTZ thickness in this region, ruling out the possibility of an abnormally hotter mantle plume. Instead, they suggest that the mantle return flow induced by lithospheric delamination beneath the southern Great Xing'an Range generates decompression melting to feed these volcanoes. Our results show a low-velocity zone immediately below the Datong-Fengzhen volcanoes, extending to ~ 400 km depth (Figure 2.13a and Figure 2.18a). The origin of the Datong-Fengzhen volcanoes thus appears to be not a mantle plume from greater depths. In addition, the subducted Pacific Plate, characterized by high-velocity anomalies in the MTZ and uppermost lower mantle beneath the Bohai Sea, is hundreds of kilometers away from the low-velocity anomalies beneath Datong-Fengzhen volcanoes. Instead, we observe high-velocity anomalies in the upper mantle and MTZ beneath the southern Great Xing'an Range, possibly indicating lithospheric delamination. Therefore, we propose that the mantle return flow induced by the delaminated lithosphere is the most plausible cause for the Datong-Fengzhen volcanoes, as suggested by Zuo et al. [2020].

The Tengchong volcanic field in southwestern China is characterized by Cenozoic basalts and andesites [Wang et al., 2007, Smithsonian-Institution, 2023]. Previous travelttime tomography images [Wei et al., 2012, Huang et al., 2015, Lei and Zhao, 2016, Lei et al., 2019] show a high-velocity anomaly, interpreted as the subducted Burma microplate beneath Tengchong, suggesting that the volcanism is related to slab dehydration and partial melting in the so-called big mantle wedge (BMW). However, our results show no strong high-velocity anomalies in the MTZ beneath Tengchong, and the Burma slab subducts with a high dip angle to only the 410-km discontinuity northwest of Tengchong (Figure 2.13b). Our observation is consistent with the global tomography model UU-P07 [Amaru, 2007]. Therefore, we conclude that the Tengchong volcanic field is not caused by slab dehydration of the Burma slab. Instead, we suspect the volcanism is related to small-scale mantle convection and upwelling between the subducting Burma slab and the 200-300 km thick lithosphere of the Sichuan Basin.

The Changbaishan volcanoes, located in northeastern China along the border with North Korea,

are primarily composed of trachytic and basaltic lava flows, as well as pyroclastic deposits [Zhang et al., 2018, Smithsonian-Institution, 2023]. These stratovolcanoes have been active since the late Pleistocene epoch and are renowned for their large-scale eruptions. Early seismic traveltime tomography studies of Northeast Asia, such as Huang and Zhao [2006] and Lei and Zhao [2005], image a flattened slab in the MTZ, suggesting the dehydration of the stagnant Pacific slab releases water to the so-called big mantle wedge (BMW) and triggers partial melting in the upper mantle beneath these volcanoes. However, this BMW hypothesis cannot explain the specific location of these volcanoes. In contrast, recent tomography studies empowered by regional permanent and temporal seismic stations [e.g., Tang et al., 2014, Tao et al., 2018] have uncovered a slab gap in the MTZ, suggesting that the sub-slab asthenosphere entrained by subduction escape through the slab gap and upwells to trigger decompression melting to feed the Changbaishan volcanoes. A new receiver-function study Wang et al. [2020] observes hydrous melting atop the MTZ and hydrous slab crust segments in the MTZ, supporting the BMW hypothesis. However, Li et al. [2023] compares seismic waveform modeling with an electrical resistivity model in this region, arguing that seismic anomalies atop the MTZ are likely caused by moduli reduction during phase transformation rather than hydrous melting. Our results show low-velocity anomalies immediately beneath the Changbaishan volcanoes, extending into two slab holes in the MTZ (Figure 2.13c). This observation is consistent with that from Tang et al. [2014] and Tao et al. [2018], suggesting that the Changbaishan volcanoes are fed by decompression melting from upwelling materials through the slab holes. The origin of the upwelling materials remains uncertain. Since there is no surface geological evidence of a deep-rooted mantle plume, Tang et al. [2014] hypothesize that the sub-slab asthenosphere is entrained by the subduction of the Pacific Plate and escapes through the slab holes. Our results show a sub-slab low-velocity anomaly in the MTZ at $\sim 135^{\circ}\text{E}$, possibly indicating the entrained asthenosphere. On the other hand, we cannot rule out another possibility that the upwelling materials come from the lower mantle, as evidenced by the low-velocity anomalies immediately below the slab holes extending down to at least 1000 km depth. To verify the source of the upwelling materials, more tomographic images of the lower mantle in this region are needed.

The Hainan volcanic field in the Leizhou Peninsula and Hainan Island in southern China is covered by late Cenozoic basalts [Zhou and Armstrong, 1982, Smithsonian-Institution, 2023]. An early global finite-frequency tomography model by Montelli et al. [2004] proposed a mantle plume in this region. Later regional travelttime tomography studies with higher resolution [Huang and Zhao, 2006, Lei et al., 2009] showed more details of this Hainan mantle plume. Furthermore, a receiver function study imaged a thinner-than-normal MTZ beneath this region, suggesting that the Hainan mantle plume is about 170-200°C hotter than the surrounding mantle [Wei and Chen, 2016]. In our study, Figure 2.13d shows strong low-velocity anomalies (more than a 2% reduction for V_s) extending from the surface to the lower mantle, suggesting a deep-rooted Hainan mantle plume.

2.6 Conclusions

We present a new 3-D radially anisotropic seismic velocity model EARA2023 of the crust and mantle down to 1,000 km depth beneath East Asia and northwestern Pacific subduction zones using full waveform inversion (FWI) tomography. Adjoint tomography based on a spectral element method is applied to a large dataset of about 640,000 body-wave and 148,000 surface-wave measurement windows. The new model provides high-resolution images of the subducted plates, continental lithosphere, asthenosphere, and mantle transition zone (MTZ) in this region. Compared to the previous FWI models of East Asia (FWEA18 and EARA2014), EARA2023 shows more small-scale velocity anomalies, better delineating the continental lithosphere thickness variations, slab morphology at different depths, and mantle upwelling. In particular, the shear wave velocity structures of EARA2023 provide the following new insights into slab morphology and origins of intraplate volcanoes in this region.

- The Japan slab is flattened in the MTZ west of the Japan Sea. At least two slab holes, ~200 km wide each, are imaged in the subducted Pacific Plate beneath the Changbaishan volcanoes. These holes provide the pathway for mantle upwelling that feeds the intra-plate volcanoes (Figure 2.6, 2.11c, and 2.13c).
- The Izu-Bonin slab is heavily deformed and fragmented, with pieces of subducted Pacific

Plate residing in the MTZ or uppermost lower mantle (Figure 2.11d and 2.11e).

- The Philippine Sea Plate subducts along the Ryukyu Trench and reaches 500 km depth. Although both are in the MTZ, the subducted Philippine Sea Plate beneath the East China Sea is not attached to the subducted Pacific Plate beneath the Yellow Sea (Figure 2.11f).
- Beneath the Datong volcanoes, low-velocity anomalies are imaged from the lithosphere down to ~400 km depth. The mantle upwelling is more likely induced by lithospheric delamination beneath the southern Great Xing'an Range (Figure 2.13a).
- The low-velocity anomalies beneath the Tengchong volcanic field are confined within the top 200 km. There is no obvious connection between the Tengchong volcano and the subducted Burma microplate. The intra-plate volcanism is more likely to be triggered by small-scale asthenospheric convection and upwelling between the subducting Burma slab and the 200-300 km thick lithosphere of the Sichuan Basin (Figure 2.13b).
- We clearly image the Hainan mantle plume, characterized by low-velocity anomalies extending from the surface to the lower mantle (Figure 2.13d).

2.7 Appendix

2.7.1 Data quality control weighting

As shown in Equation (2.2) of the main texts, W_w is the weighting term to control the data quality in each measurement window. Following Tao et al. [2018], W_w is defined as

$$W_w = f_1(\text{CC})f_2(\text{SNR})f_3(\delta t_{\text{CC}}) \quad (2.12)$$

where

$$f_1(\text{CC}) = \begin{cases} 0, & \text{CC} < 0.5 \\ \frac{1}{2} + \frac{1}{2} \cos\left(\pi \frac{0.7 - \text{CC}}{0.7 - 0.5}\right), & 0.5 \leq \text{CC} < 0.7 \\ 1, & \text{CC} \geq 0.7 \end{cases} \quad (2.13)$$

$$f_2(\text{SNR}) = \begin{cases} 0, & \text{SNR} < 10 \\ \frac{1}{2} + \frac{1}{2} \cos\left(\pi \frac{15-\text{SNR}}{15-10}\right), & 10 \leq \text{SNR} < 15 \\ 1, & \text{SNR} \geq 15 \end{cases} \quad (2.14)$$

$$f_3(\delta t_{\text{CC}}) = \begin{cases} 0, & |\delta t_{\text{CC}}| \geq 10 \\ \frac{1}{2} - \frac{1}{2} \cos\left(\pi \frac{10-|\delta t_{\text{CC}}|}{10-8}\right), & 8 \leq |\delta t_{\text{CC}}| < 10 \\ 1, & |\delta t_{\text{CC}}| < 8 \end{cases} \quad (2.15)$$

SNR is the signal-to-noise ratio for each measurement window:

$$\text{SNR} = 10 \log \frac{\frac{1}{t_{\text{signal}}} \int u_{\text{signal}}^2(t) dt}{\frac{1}{t_{\text{noise}}} \int u_{\text{noise}}^2(t) dt} \quad (2.16)$$

where $u_{\text{signal}}(t)$ represents displacement within the measurement window, t_{signal} is the corresponding window length, $u_{\text{noise}}(t)$ is displacement in the noise window defined as from the event origin time to 10 s before the predicted first P wave arrival time based on the AK135 model [Kennett et al., 1995], and t_{noise} is the corresponding window length. CC and δt_{CC} are the cross-correlation coefficient and time shift between the data and synthetics in the measurement window.

2.7.2 Source inversion details

The invertible source parameters include centroid moment tensor M , 3-D centroid location X , and source time function $S(t)$. Assuming a Gaussian source time function, we have

$$S(t) = \frac{1}{\sqrt{\pi\tau}} e^{-\left(\frac{t-t_0}{\tau}\right)^2} \quad (2.17)$$

where τ means the half duration time and t_0 represents the event's origin time. Because the frequency band of our inversion cannot well constrain τ , we only invert for t_0 and use τ determined by the magnitude in the Global CMT solution [Ekström et al., 2012].

Following Kim et al. [2011], we have ten source parameters to invert, including 6 elements of the moment tensor $M_{11}, M_{22}, M_{33}, M_{12}, M_{13}, M_{23}$, 3 elements of the 3-D centroid location x_1, x_2, x_3 , and the centroid origin time t_0 . The gradient to the data misfit χ_e is

$$g = \left\{ \frac{\partial \chi_e}{\partial M_{11}}, \frac{\partial \chi_e}{\partial M_{22}}, \frac{\partial \chi_e}{\partial M_{33}}, \frac{\partial \chi_e}{\partial M_{12}}, \frac{\partial \chi_e}{\partial M_{13}}, \frac{\partial \chi_e}{\partial M_{23}}, \frac{\partial \chi_e}{\partial x_1}, \frac{\partial \chi_e}{\partial x_2}, \frac{\partial \chi_e}{\partial x_3}, \frac{\partial \chi_e}{\partial t_0} \right\} \quad (2.18)$$

By using $\sigma_M = \sqrt{2}((\frac{\partial \chi_e}{\partial M_{11}})^2 + (\frac{\partial \chi_e}{\partial M_{22}})^2 + (\frac{\partial \chi_e}{\partial M_{33}})^2 + (\frac{\partial \chi_e}{\partial M_{12}})^2 + (\frac{\partial \chi_e}{\partial M_{13}})^2 + (\frac{\partial \chi_e}{\partial M_{23}})^2)^{-\frac{1}{2}}$ and $\sigma_X = ((\frac{\partial \chi_e}{\partial x_1})^2 + (\frac{\partial \chi_e}{\partial x_2})^2 + (\frac{\partial \chi_e}{\partial x_3})^2)^{-\frac{1}{2}}$ to non-dimensionalize M and X , we have non-dimensional parameters $\hat{M} = \frac{M}{\sigma_M}$, $\hat{X} = \frac{X}{\sigma_X}$ and gradients $\frac{\partial \chi_e}{\partial \hat{M}} = \sigma_M \frac{\partial \chi_e}{\partial M}$, $\frac{\partial \chi_e}{\partial \hat{X}} = \sigma_X \frac{\partial \chi_e}{\partial X}$. The non-dimensionalization will guarantee different types of parameters contribute equally to the gradient's magnitude.

Each source inversion step contains 3 iterations of the steepest descent update, after which the source parameters will almost not change. To do a line search of the steepest descent update in each iteration, we first calculate the waveform synthetics of the perturbed source model $\{M - \epsilon \frac{\partial \chi_e}{\partial M}, X - \epsilon \frac{\partial \chi_e}{\partial X}, t_0\}$, where $\epsilon = 0.01$. By assuming a linear relationship between the source model and the waveform perturbations, we approximate the waveform synthetics $s_\alpha(x, t)$ for the source model $\{M - \alpha \frac{\partial \chi_e}{\partial M}, X - \alpha \frac{\partial \chi_e}{\partial X}, t'\}$ with arbitrary step length α and origin time t' as

$$s_\alpha(x, t) = s_0(x, t - (t' - t_0)) + \frac{\alpha}{\epsilon}(s_\epsilon(x, t - (t' - t_0)) - s_0(x, t - (t' - t_0))) \quad (2.19)$$

where $s_0(x, t)$ is the synthetics for the starting source model $\{M, X, t_0\}$ of this iteration, $s_\epsilon(x, t)$ is the perturbed synthetics for the source model $\{M - \epsilon \frac{\partial \chi_e}{\partial M}, X - \epsilon \frac{\partial \chi_e}{\partial X}, t_0\}$. $s_0(x, t - (t' - t_0))$ and $s_\epsilon(x, t - (t' - t_0))$ shift the synthetics by $t' - t_0$ to align the waveforms. We use a Bayesian optimization method to efficiently find an optimized step length α_{opt} and the origin time t'_{opt} to minimize the data misfit.

2.7.3 Optimization of the structural model parameters

We use the conjugate gradient descent to update our structure model. In each iteration, the search direction is determined by the gradient of the current iteration g_i and the search direction of the previous iteration d_{i-1} :

$$d_i = -Pg_i + \beta d_{i-1} \quad (2.20)$$

where P is the preconditioner and

$$\beta = \frac{(d_i - d_{i-1}) \cdot Pg_i}{d_{i-1} \cdot Pg_{i-1}} \quad (2.21)$$

In the first iteration, β is set to 0. In the later iterations, we calculate the Powell ratio [Powell and Powell, 1981] defined as:

$$R = \frac{Pg_i \cdot Pg_{i-1}}{Pg_{i-1} \cdot Pg_{i-1}} \quad (2.22)$$

If $R > 0.2$ for all the invertible structure parameters, we reset β to 0 to restart the optimization when a critical point is found, thus speeding up our inversion.

The preconditioner P is chosen to be close to the inverse Hessian matrix [Fichtner and Trampert, 2011a]. In practice, it is difficult to directly calculate the Hessian matrix. Instead, we use the so-called pseudo-Hessian matrix [Bozdağ et al., 2016]:

$$H(x) = \sum_{e=1}^{N_e} \int \frac{\partial^2 s(x, t)}{\partial t^2} \cdot \frac{\partial^2 s^\dagger(x, T-t)}{\partial t^2} dt \quad (2.23)$$

where $s(x, t)$ is the wave field of the forward simulation and $s^\dagger(x, T-t)$ is the backward wave field of the adjoint source. N_e denotes the total number of events.

By setting an appropriate threshold δ of the inverse hessian, we have

$$P(x) = \begin{cases} \frac{1}{\delta} & H(x) < \delta \\ \frac{1}{H(x)} & H(x) \geq \delta \end{cases} \quad (2.24)$$

δ is necessary as P acts as weighting to the gradient. We need δ to suppress a large P . It should be noticed that the choice of δ has a significant impact on the inversion. As P can be regarded as the weighting of the gradient in different regions, a smaller δ will tolerate a larger value of $P(x)$. A region with a larger $P(x)$ usually means a smaller Hessian matrix, i.e., less data coverage. To balance model updating in different regions, we should use appropriate weighting to enhance the gradient for the area with less data coverage [Fichtner and Trampert, 2011a].

For the calculated kernels from the adjoint simulation of each event, we first use Gaussian spheres centered at the sources and receivers to damp out regions with potential singularity [Zhu et al., 2015]. The diameters of the Gaussian spheres are 50 km for the source and 10 km for the receivers. Then we apply the preconditioning after summing up the kernels for all events. In the next step, we apply Gaussian smoothing on the kernel to help stabilize the inversion. In the first 5 iterations of the inversion, the Gaussian smoothing spheres are configured as 50 km horizontally and 25 km vertically. In the subsequent 25 iterations, we choose Gaussian smoothing spheres with 25 km horizontally and 10 km vertically to capture small-scale structures.

We use line search in each iteration to find the optimal step length. Based on the line search direction d_i , which is the negation of the normalized gradient, we forwardly simulate synthetics

for the perturbed model $m_i + \epsilon d_i$, where $\epsilon = 0.01$. Using the linear dependency approximation of waveform to structure [Warner et al., 2013], we estimate the perturbed synthetics for arbitrary model $m_i + \alpha d_i$ as

$$s(x, t; m_i + \alpha d_i) = s(x, t; m_i) + \frac{\alpha}{\epsilon} (s(x, t; m_i + \epsilon d_i) - s(x, t; m_i)) \quad (2.25)$$

where $s(x, t; m_i)$ represents the synthetics for model m_i , and $s(x, t; m_i + \epsilon d_i)$ represents the synthetics for model $m_i + \epsilon d_i$. It is computationally efficient to calculate $s(x, t; m_i + \alpha d_i)$ without performing the SEM simulation. We can simply test different values of α and find the optimized one.

2.7.4 Point spread function tests

Following Fichtner and Trampert [2011b], if we apply a single point perturbation to one of the invertible model parameters, the gradient from the adjoint state method can be regarded as the negation of the Hessian matrix, also called the point spread function (PSF), of the final model. This function can quantify the model's blurring and distortion by showing how it tends to update when only one point is perturbed. For example, when we apply a single point perturbation to V_{SV} , the PSF response for V_{SV} and V_{SH} are $H_{V_{SV}, V_{SV}}$, $H_{V_{SV}, V_{SH}}$, respectively, where H represents the Hessian matrix.

In practice, we apply Gaussian sphere perturbations with a radius of 50 km and a separation of 2 degrees horizontally and 200 km vertically, located at depths of 100 km, 300 km, 500 km, 700 km, and 900 km, to the final model. By using the adjoint state method, the corresponding PSFs are calculated and shown in the main text.

2.8 Figures

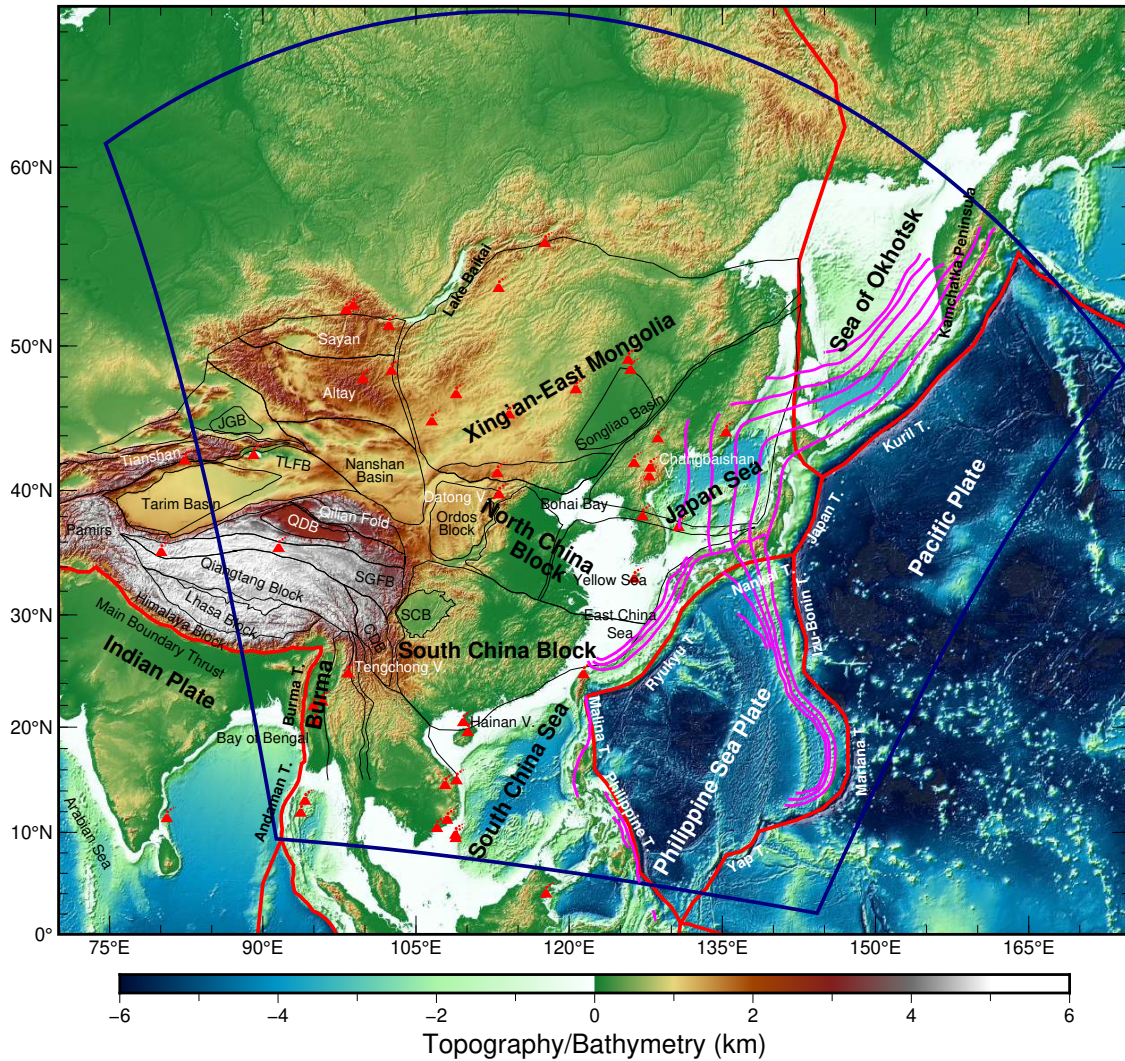


Figure 2.1 Topography/bathymetry map of East Asia and the northwestern Pacific. The simulation region is outlined by navy blue lines. Magenta curves represent slab depth contours with a 100-km interval based on the Slab2 model [Hayes et al., 2018]. The tectonic units are outlined by black curves, and intraplate volcanoes are marked with red triangles. Red thick lines indicate plate boundaries from the NUVEL-1 model [DeMets et al., 1990]. Acronyms of certain tectonic units: CDB - Chuandian Block, JGB - Junggar Basin, QDB - Qaidam Basin, SCB - Sichuan Basin, TLFB - Tulufan Basin, SGFB - Songpan Ganzi Fold Belt.

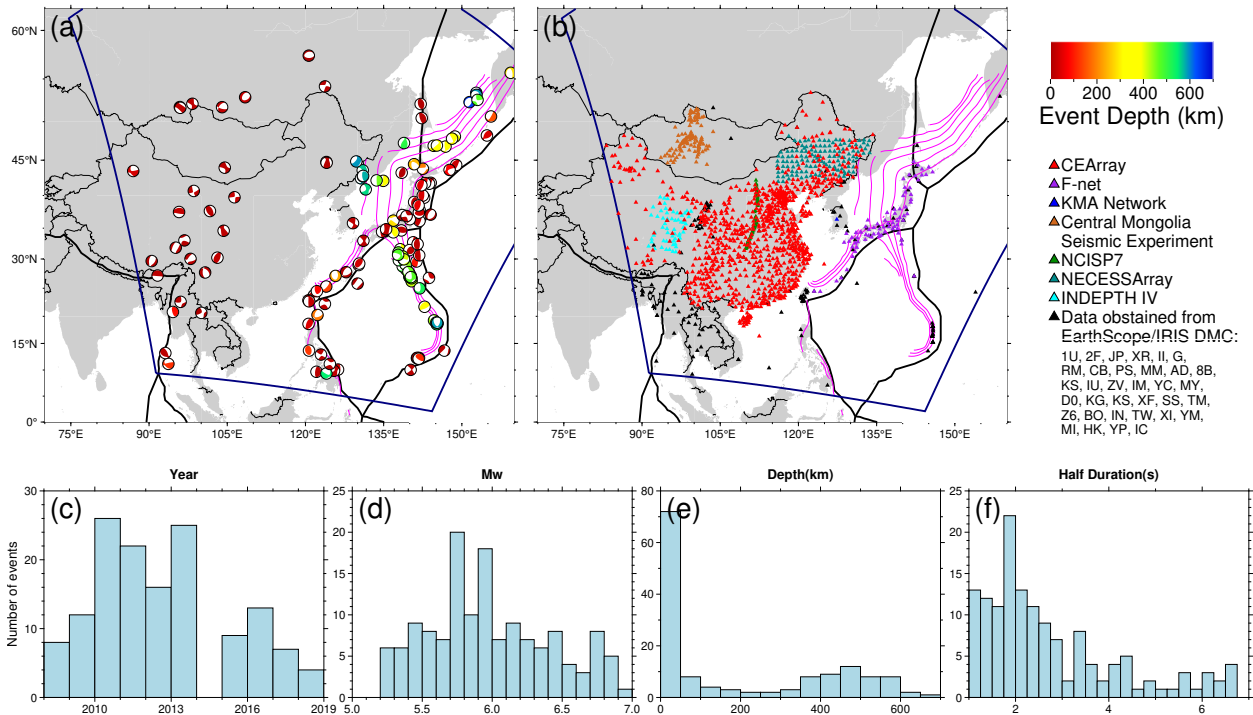


Figure 2.2 Distribution of earthquakes (a) and seismic stations (b) used in this study. Earthquakes in (a) are color-coded by depth. The focal mechanisms are from the Global Centroid Moment Tensor (GCMT) project [Dziewonski and Anderson, 1981, Ekström et al., 2012]. Stations in (b) are color-coded by networks. Other features in (a) and (b) are the same as Fig. 1. (c-f) Histograms of earthquake origin time (c), moment magnitude (d), depth (e), and half duration time of the source time function (f).

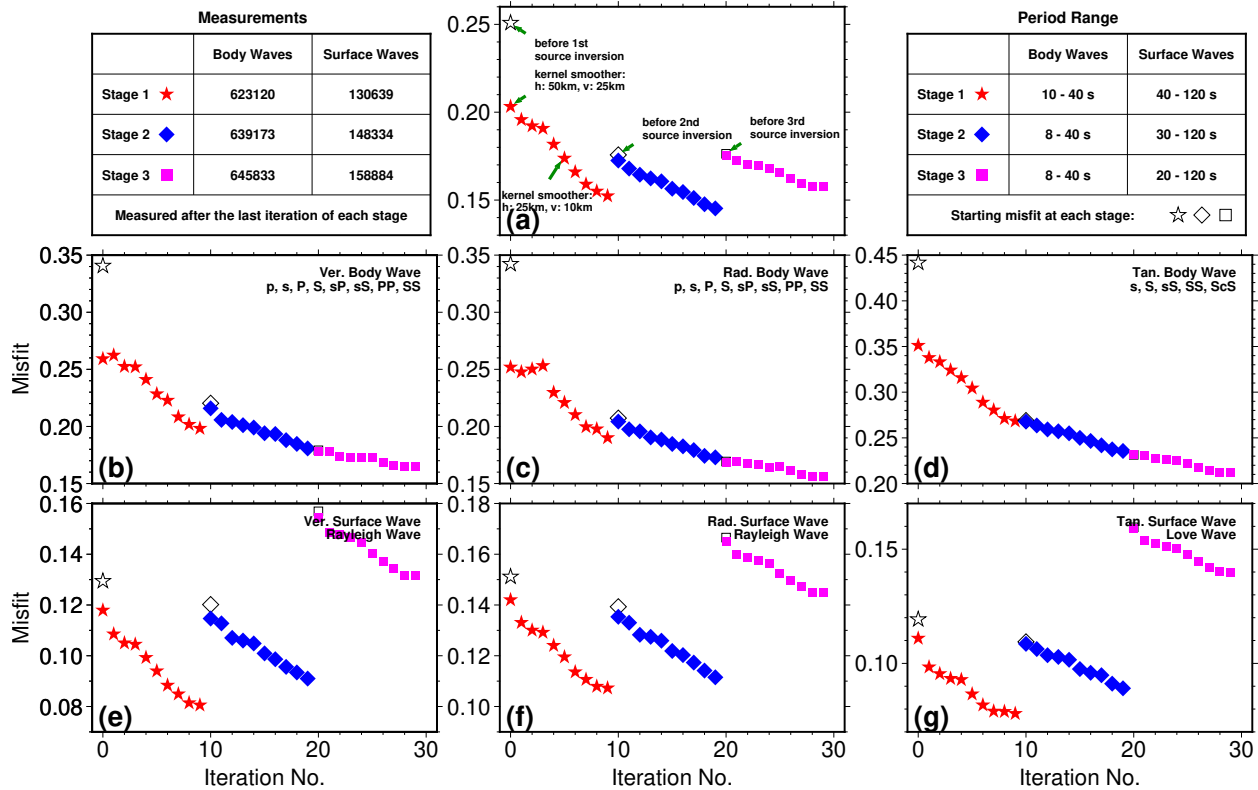


Figure 2.3 Misfit reduction curves in three stages (stage 1 in red stars, stage 2 in blue diamonds, and stage 3 in magenta squares). The open symbols represent the misfit before the source inversion. (a) Overall misfits. (b) Misfits for body waves on the vertical component. (c) Misfits for body waves on the radial component. (d) Misfits for body waves on the transverse component. (e) Misfits for Rayleigh waves on the vertical component. (f) Misfits for Rayleigh waves on the radial component. (g) Misfits for Love waves on the transverse component.

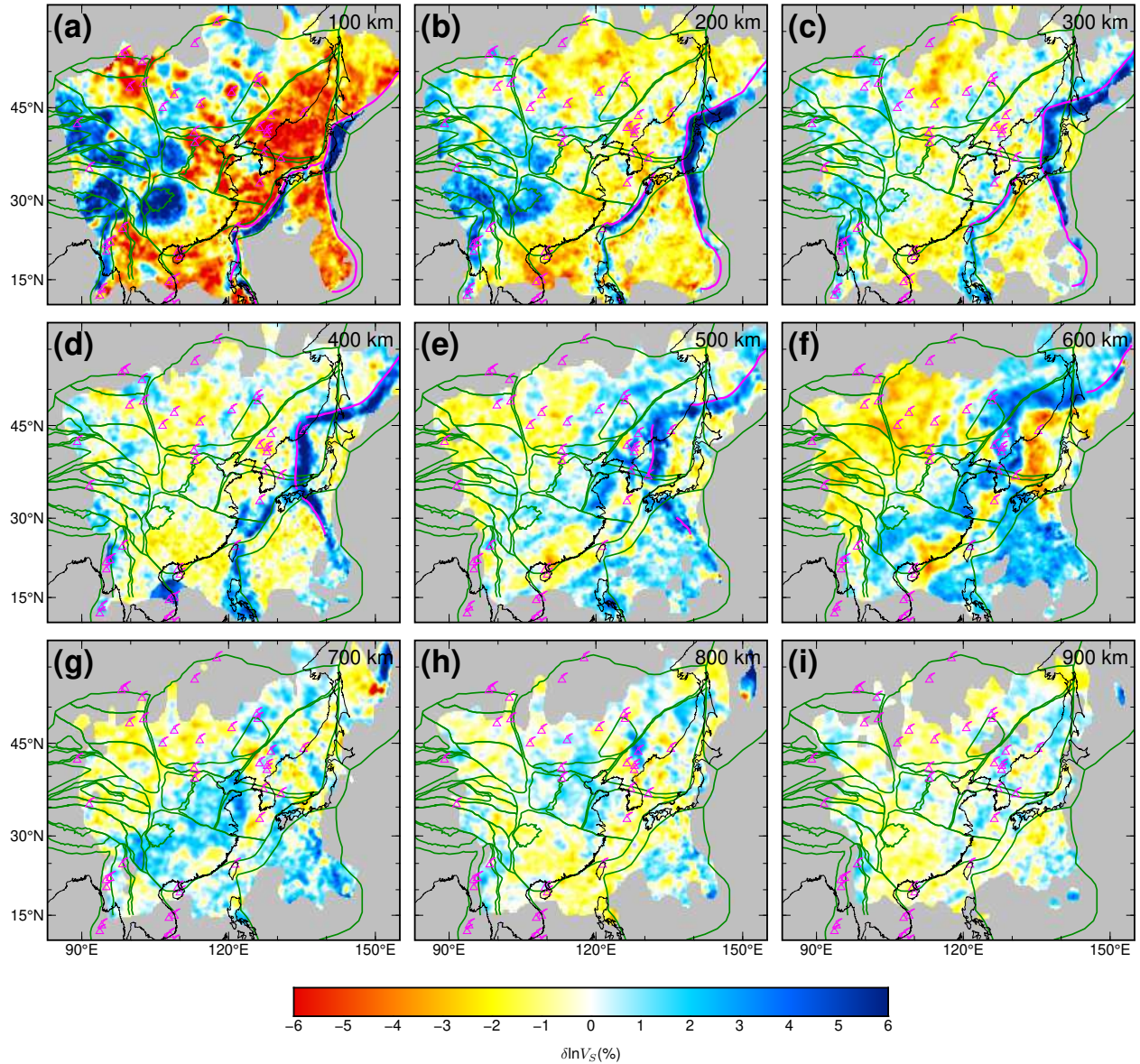


Figure 2.4 Mapviews of $\delta \ln V_S$ perturbation at various depths. The reference model is the 1-D average of the 3-D model. Regions with low resolution based on the point spread function resolution test are masked out. Magenta curves represent the slab depth contour at each depth based on the Slab2 model [Hayes et al., 2018]. Other features are the same as Fig. 2.1.

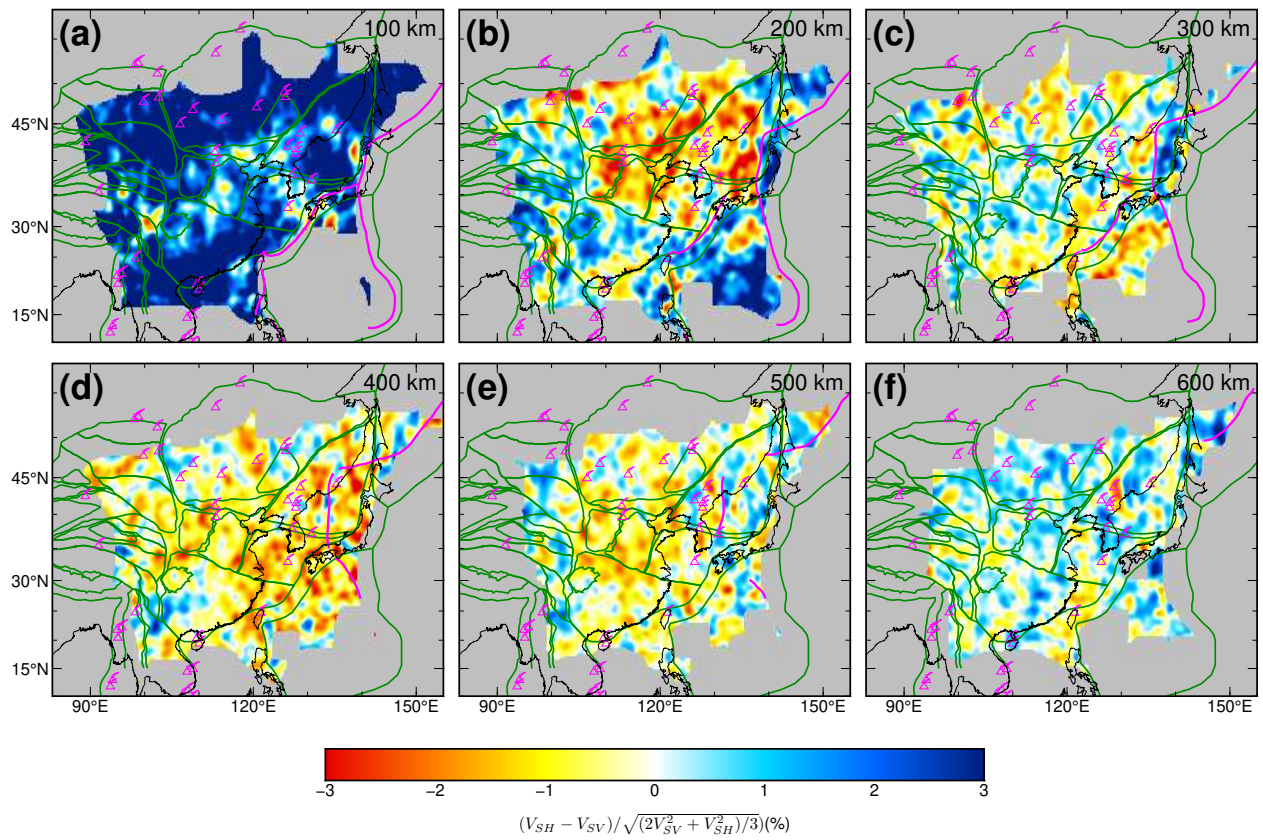


Figure 2.5 Mapviews of radial anisotropy at various depths. Other features are the same as Figure 2.4.

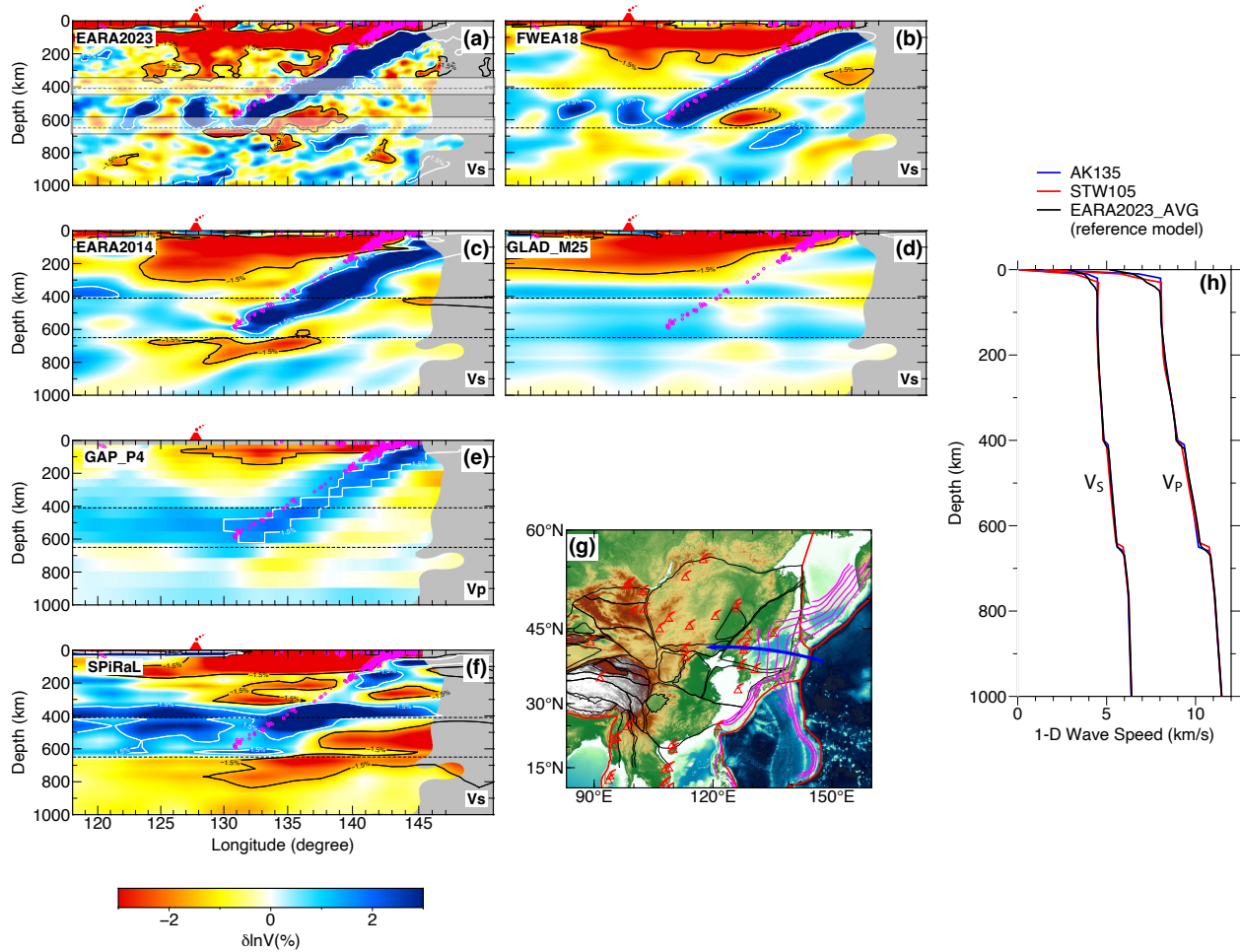


Figure 2.6 (a-f) Comparison of different models along the same cross-section across the Changbaishan volcanoes. The EARA2023, FWEA18 [Tao et al., 2018], EARA2014 [Chen et al., 2015], GLAD_M25 [Lei et al., 2020], and SPiRaL [Simmons et al., 2021] models present the $\delta \ln V_S$ perturbation with respect to EARA2023_AVG (averaged EARA2023), whereas the GAP_P4 model [Obayashi et al., 2013] displays the V_P perturbation with respect to its own reference model. The 410- and 660-km depths are shown as dashed lines. Earthquakes from 1964 to 2019 from the ISC-EHB catalog [Weston et al., 2018] within 100 km of the cross-section are shown as magenta circles. Regions with low resolution based on the point spread function resolution test are masked out. Semi-transparent bands ± 60 km around the 410- and 660-km discontinuities mask potential artifacts due to the discontinuity topography in (a). (g) The location of the cross-section is shown as the blue arrow. Other features are the same as Fig. 2.1. (h) Comparison of 1-D velocity models AK135 [Kennett et al., 1995], STW105 [Kustowski et al., 2008a], and EARA2023_AVG.

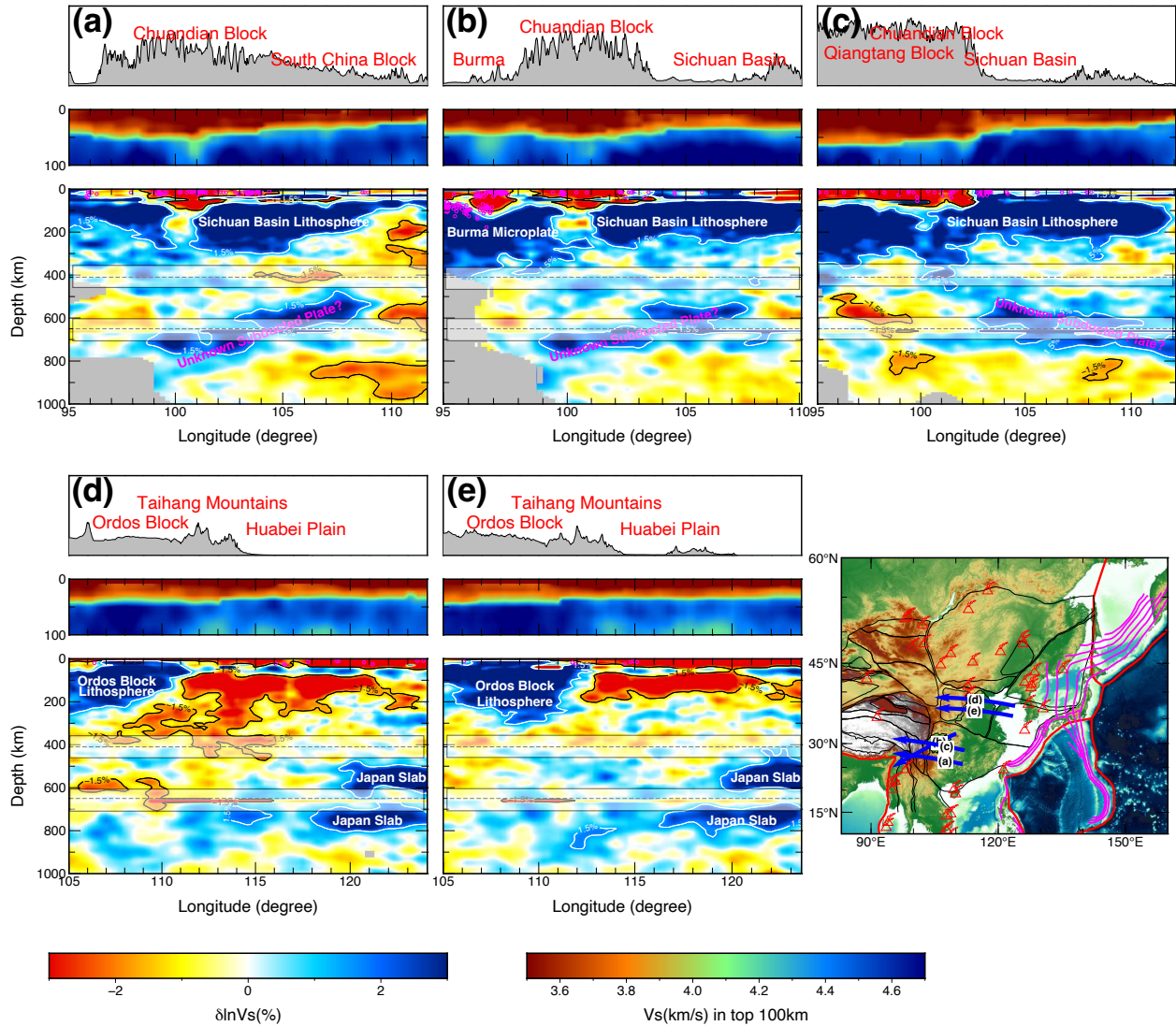


Figure 2.7 Cross-Sections of $\delta \ln V_S$ perturbation beneath Eastern Tibet (a-c) and North China Craton (d-e). The locations of the cross-sections are marked as thick blue arrows on the map. The annotations on the map are similar to the previous figures. Each cross-section includes a topography panel on the top with a 60-times vertical exaggeration, an absolute velocity panel from the surface to 100 km depth, and a panel of $\delta \ln V_S$ perturbation with respect to EARA2023_AVG (the 1-D average of the 3-D model) down to 1000 km depth. Earthquakes from the ISC-EHB catalog [Weston et al., 2018] within 100 km of the cross-section are shown as magenta circles. White and black curves mark the 1.5% contours. Regions with low resolution based on the point spread function resolution test are masked out. Semi-transparent bands ± 60 km around the 410- and 660-km discontinuities mask potential artifacts due to the discontinuity topography.

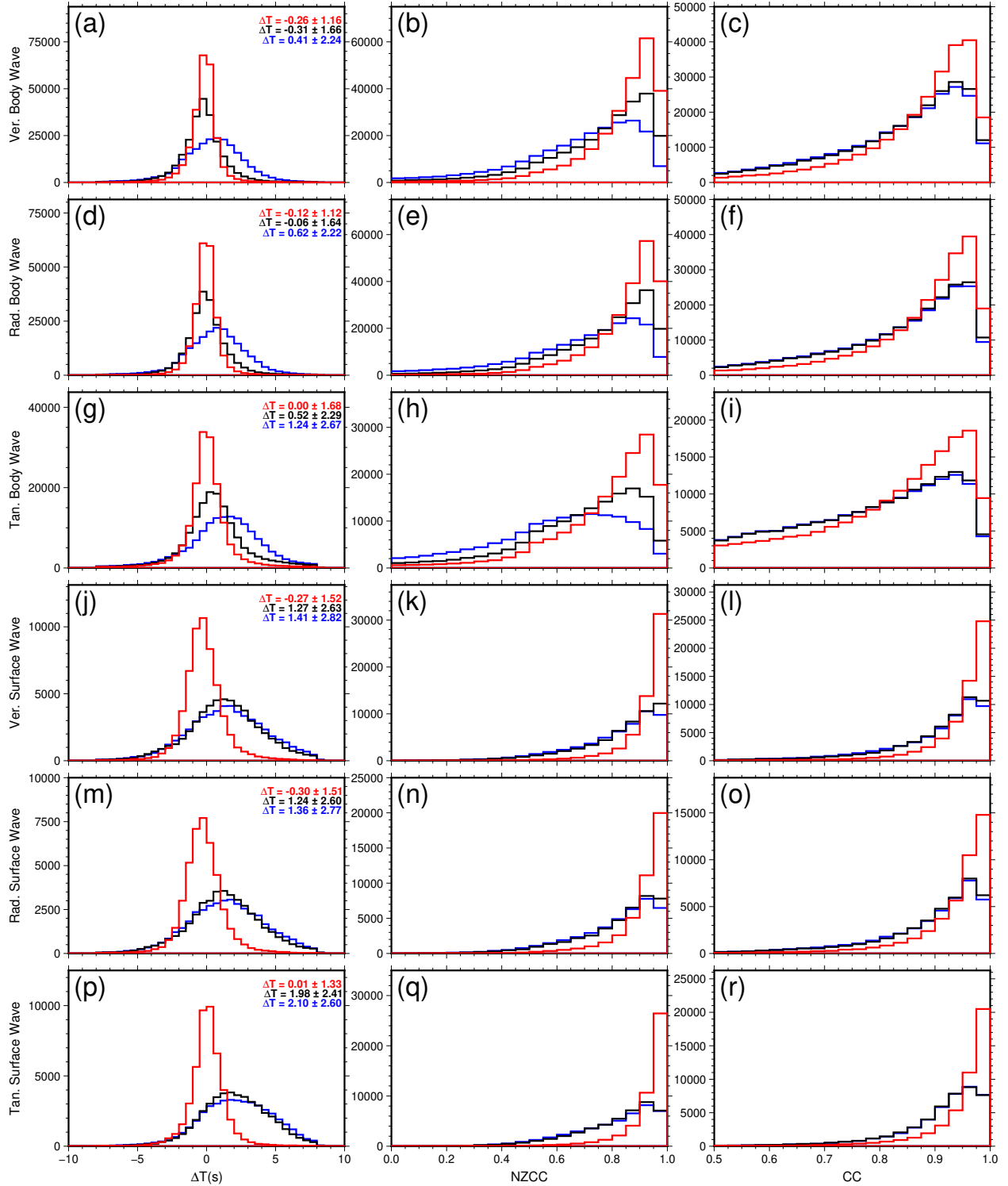


Figure 2.8 Histograms of time shifts, normalized zero-lag cross-correlation coefficients (NZCC), and cross-correlation coefficients (CC) for all 6 measurement categories. Blue histograms represent the initial structural and source models, black histograms represent the initial structural model with inverted source models after 3 iterations in the first stage of source inversions, whereas red ones represent the final structural and source models. For time shifts, also shown the mean and standard deviation.

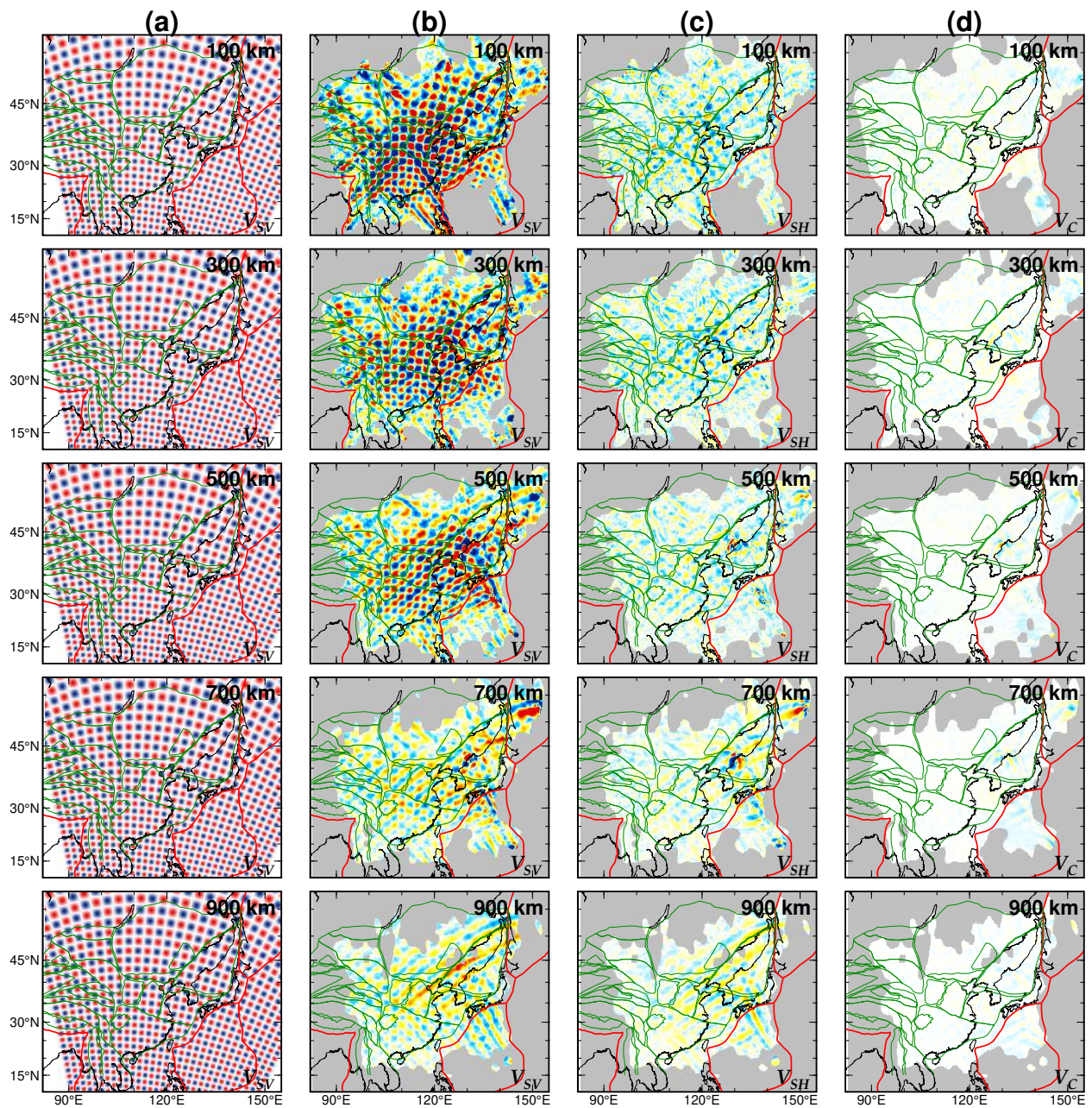


Figure 2.9 Point Spread Function (PSF) test for V_{SV} perturbation at various depths. (a) The input model perturbation applied to V_{SV} for the final model. (b) The PSF response for V_{SV} , which should mimic the input model in an ideal case. (c) The PSF response for V_{SH} , which should be null in an ideal case. (d) The PSF response for bulk velocity V_C , which should be null in an ideal case.

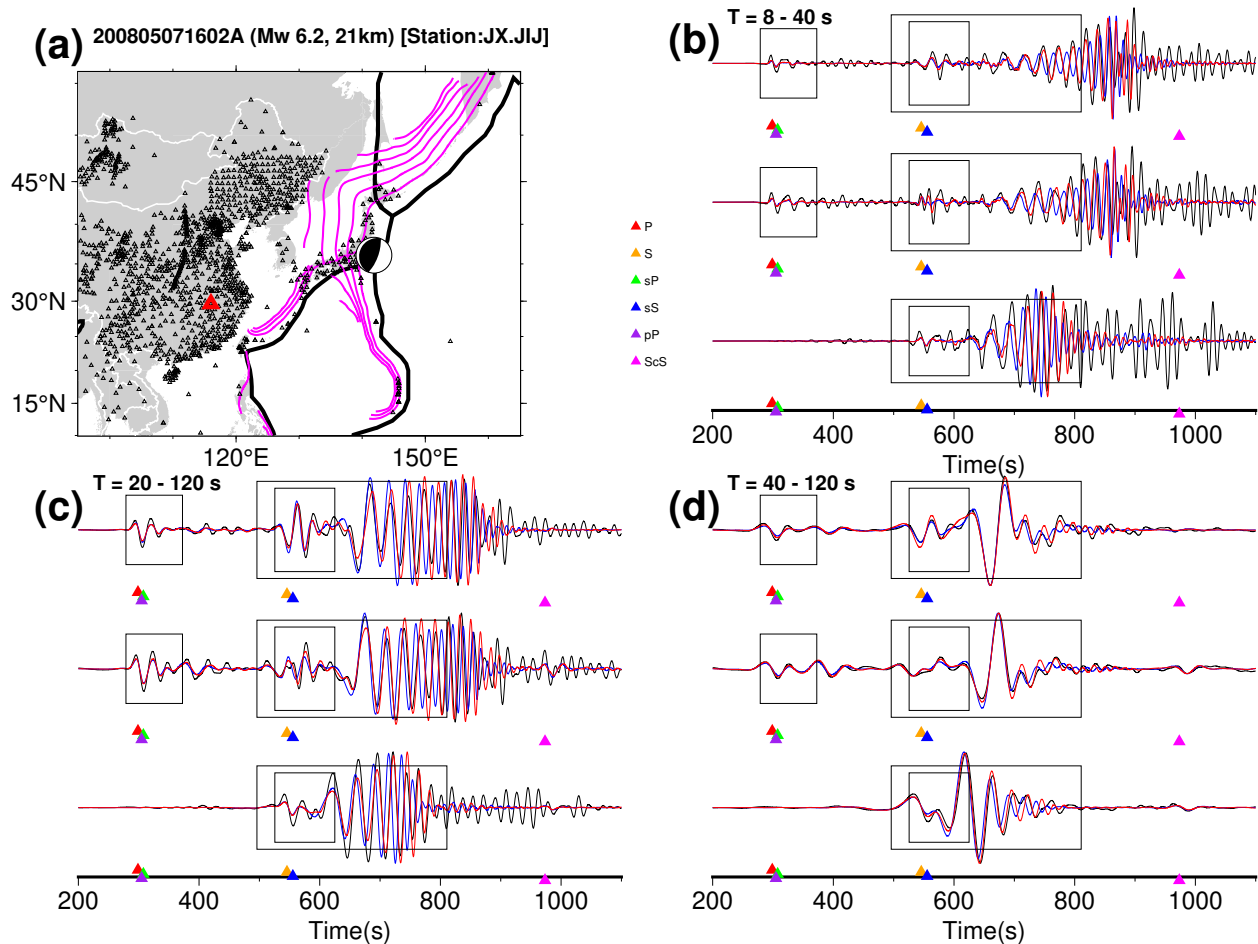


Figure 2.10 Waveform comparisons for a selected earthquake (GCMT ID: 200805071602A) recorded by a station indicated by the red triangle in (a). All other stations used in the inversion are shown as black triangles for reference. Other features are the same as Fig. 2.2b. (b-d) Waveform comparisons between observations (black), synthetics based on the starting model (blue), and synthetics based on the final model (red) at different frequency bands for various seismic phases. Black boxes indicate the measurement windows based on the theoretical phase arrival times predicted by the AK135 model [Kennett et al., 1995].

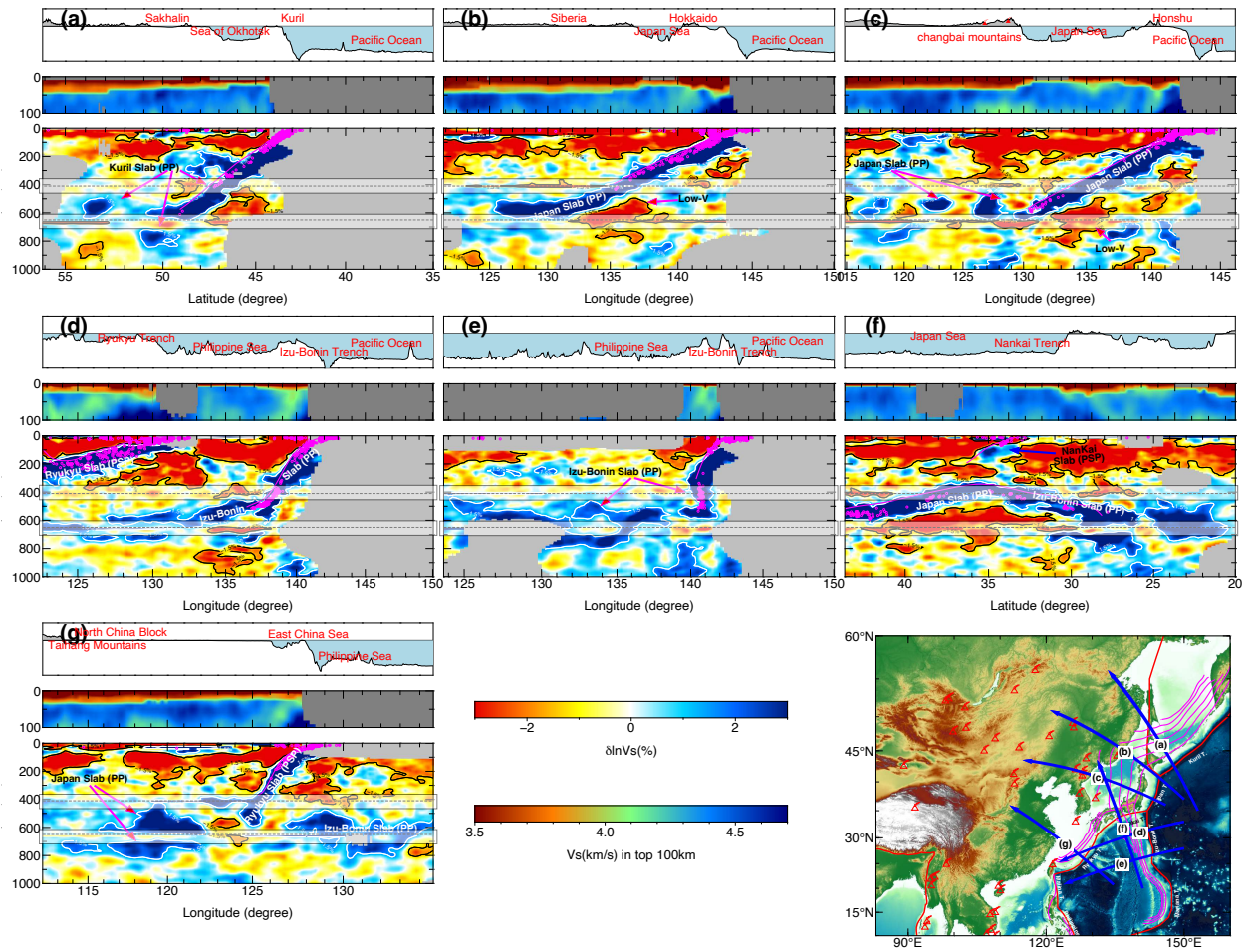


Figure 2.11 Cross-sections of $\delta \ln V_S$ perturbation for northwestern Pacific subduction zones, plotted similar to Fig. 2.7. The reference model is EARA2023_AVG. Acronyms: PP: Pacific Plate, PSP: Philippine Sea Plate.

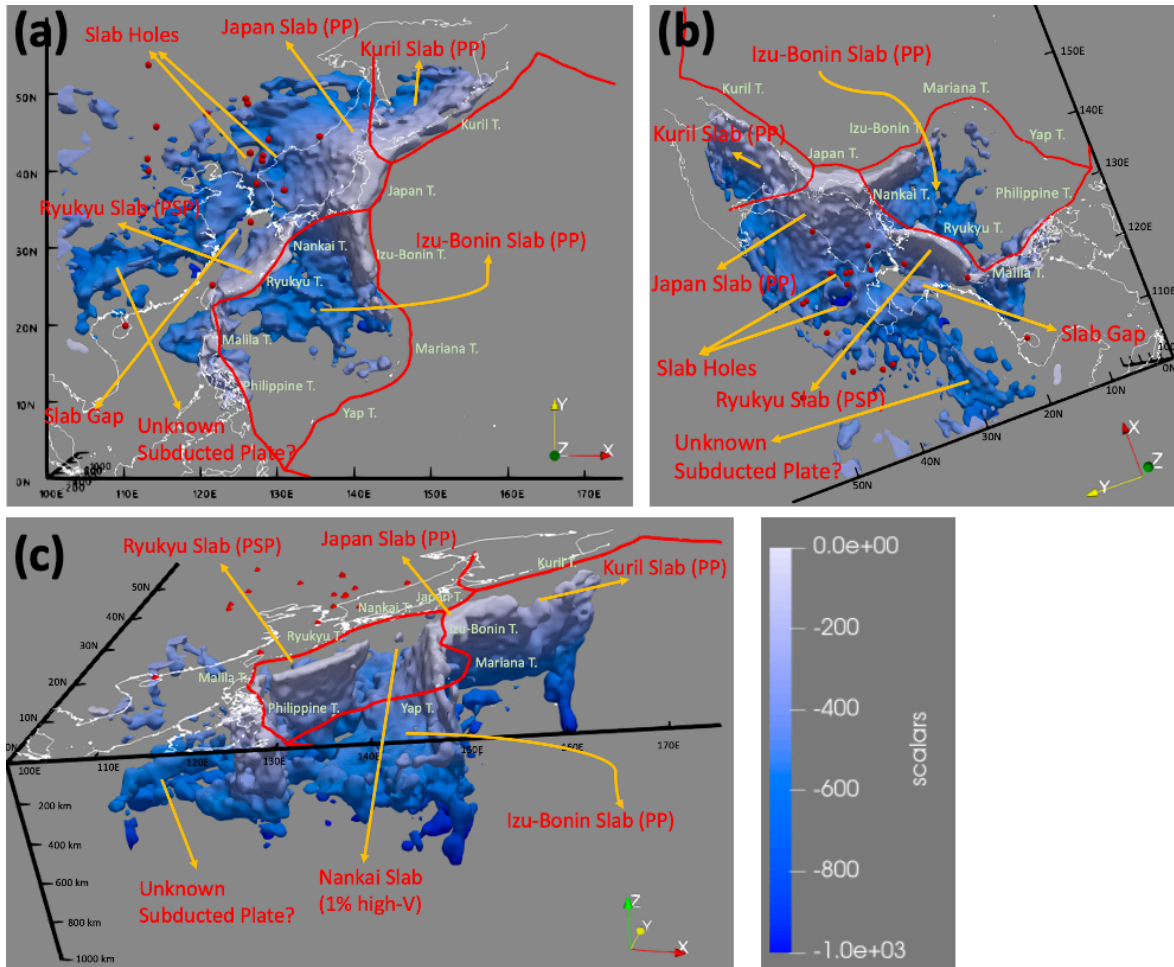


Figure 2.12 3-D images of slab morphology in the northwestern Pacific subduction zones. The slabs are represented by the 1.5% isosurface of high-velocity anomalies and color-coded by depth. Red thick curves indicate the plate boundaries from the NUVEL-1 model [DeMets et al., 1990], white curves show the coastline, and red dots represent intra-plate volcanoes. (a) Mapview. (b) Viewing from the northwest. (c) Viewing from the southeast.

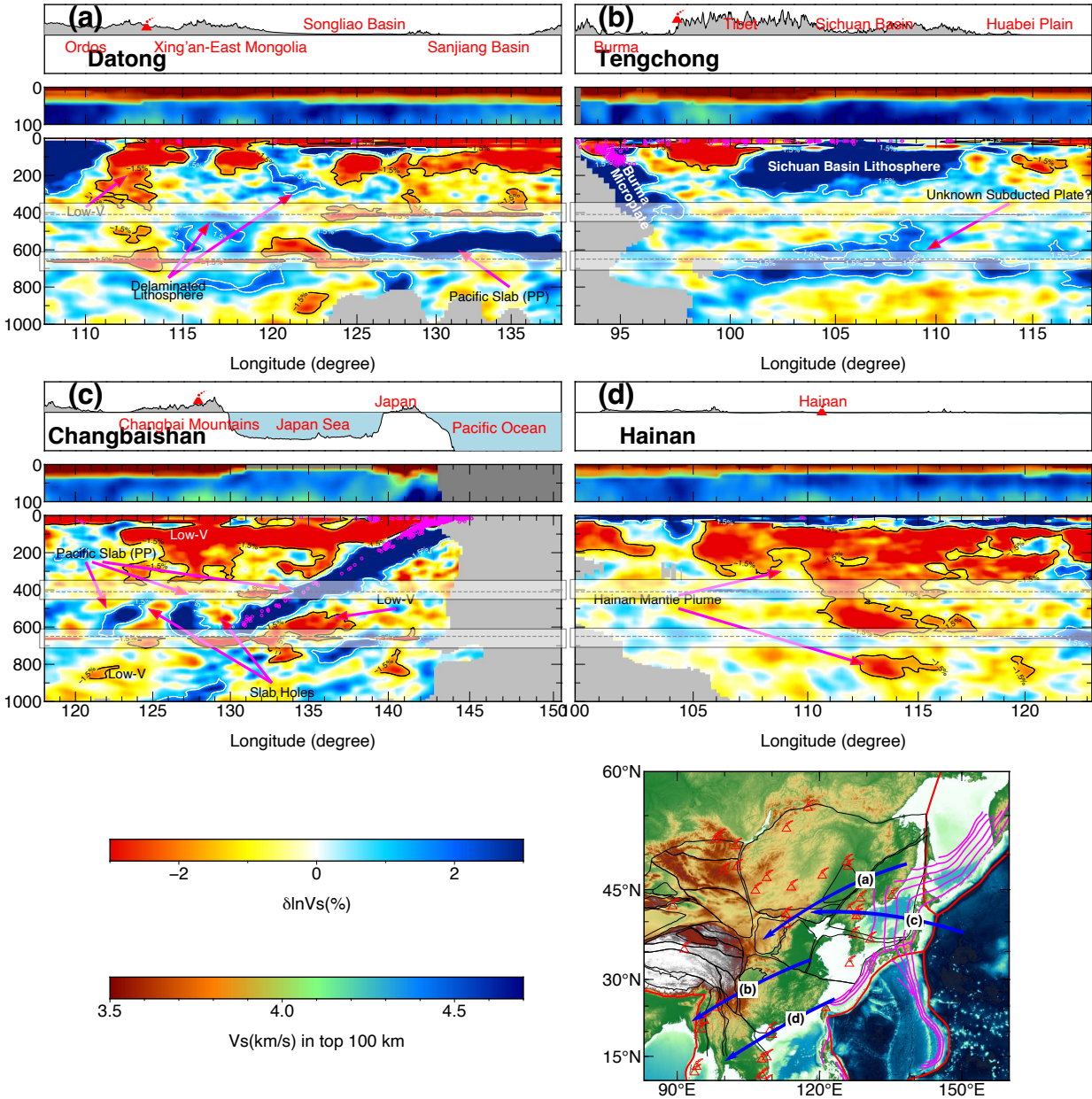


Figure 2.13 Cross-sections of $\delta \ln V_S$ perturbation for four intra-plate volcanic fields in East Asia, plotted similar to Fig. 2.7. The reference model is EARA2023_AVG. (a) Datong-Fengzhen volcanoes. (b) Tengchong volcanic field. (c) Changbaishan volcanoes. (d) Hainan volcanic field.

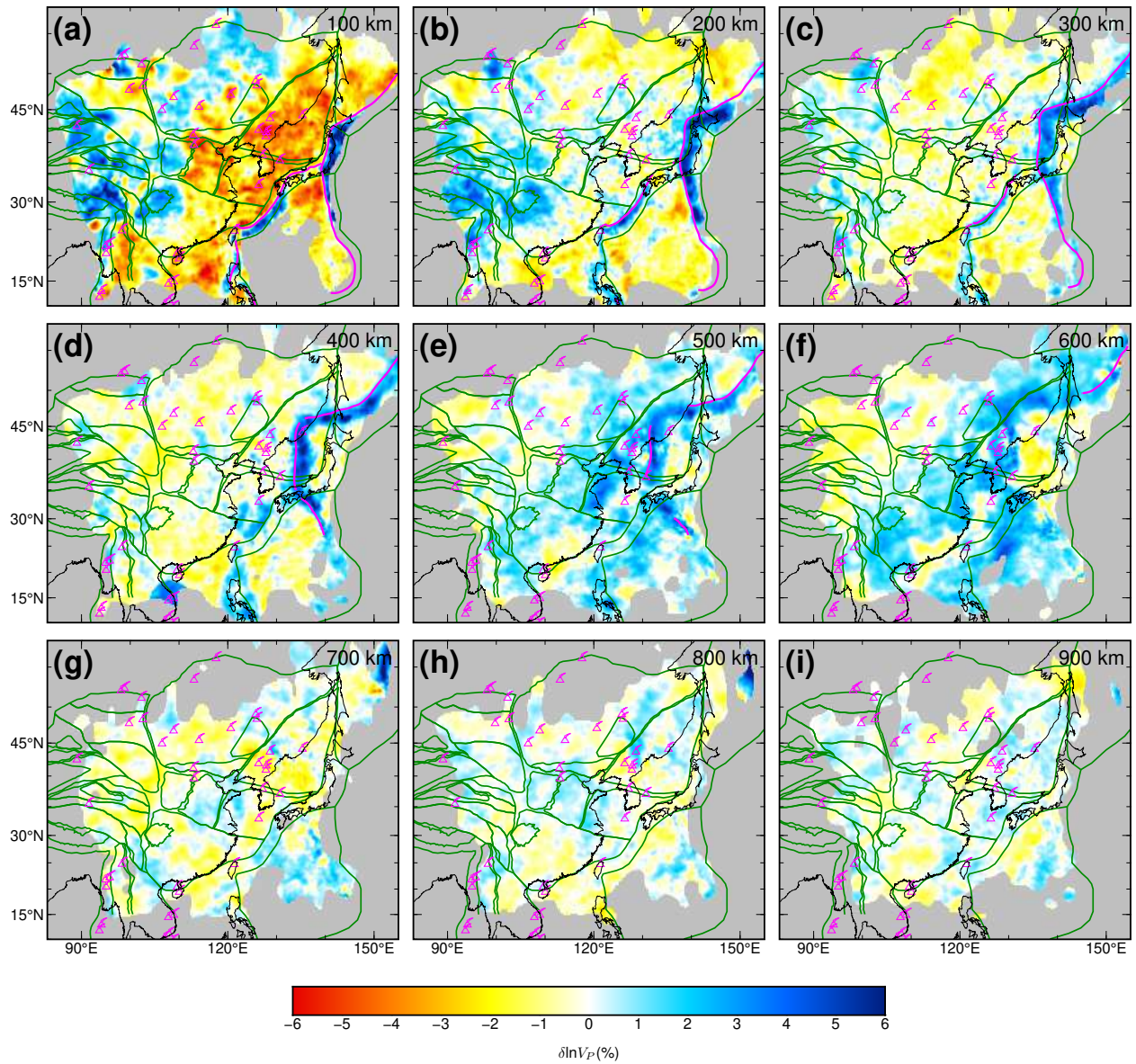


Figure 2.14 Mapviews of $\delta \ln V_P$ perturbation at various depths, similar to Figure 2.4.

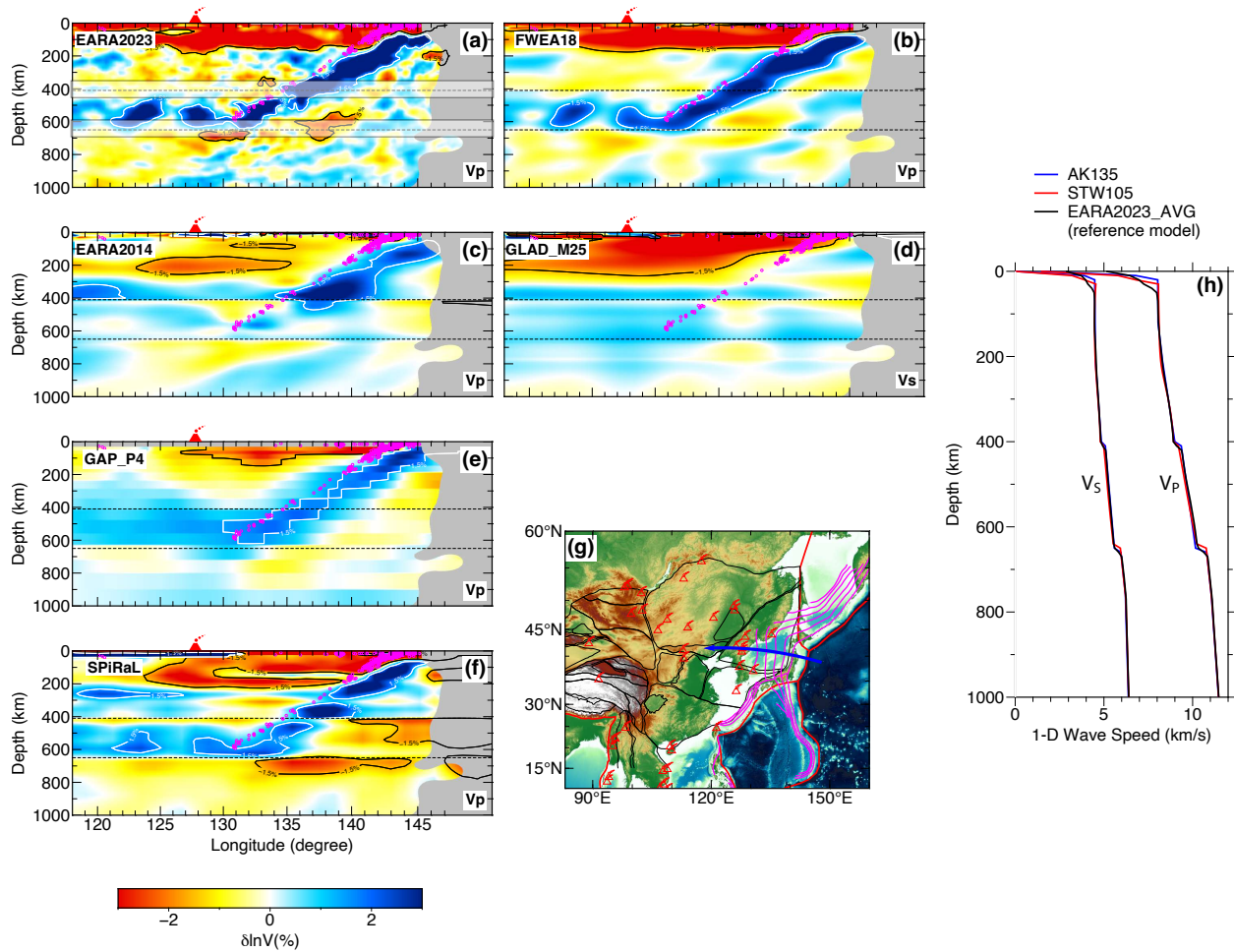


Figure 2.15 Comparison of different models along the same cross-section across the Changbaishan volcanoes. The EARA2023, FWEA18 [Tao et al., 2018], EARA2014 [Chen et al., 2015], GLAD_M25 [Lei et al., 2020], and SPiRaL [Simmons et al., 2021] models present the $\delta \ln V_P$ perturbation with respect to EARA2023_AVG (averaged EARA2023), whereas the GAP_P4 model [Obayashi et al., 2013] displays the $\delta \ln V_P$ perturbation with respect to its own reference model. Annotations are similar to Figure 2.6.

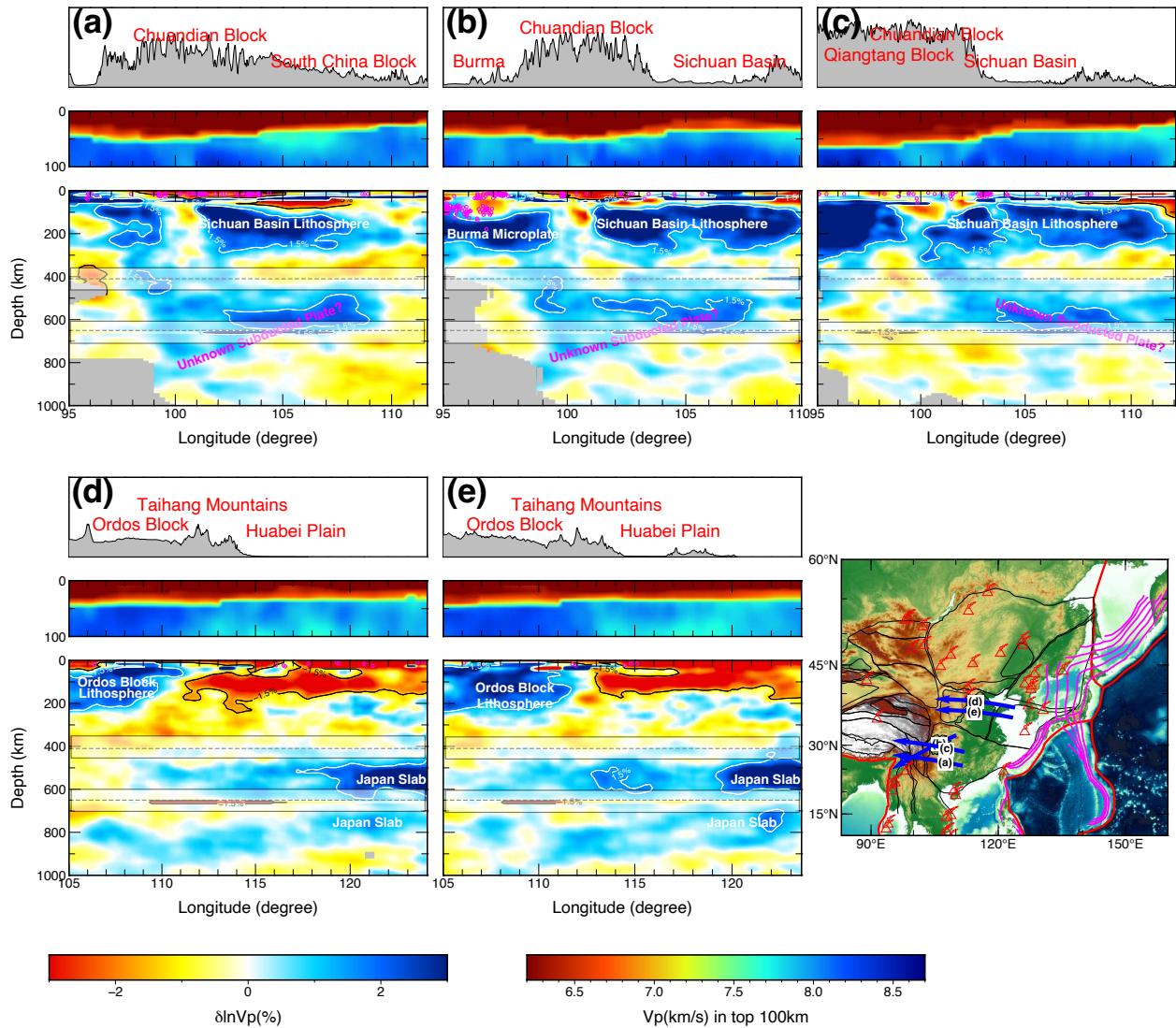


Figure 2.16 Cross-Sections of $\delta \ln V_p$ perturbation beneath Eastern Tibet (a-c) and North China Craton (d-e). Annotations are similar to Figure 2.7.

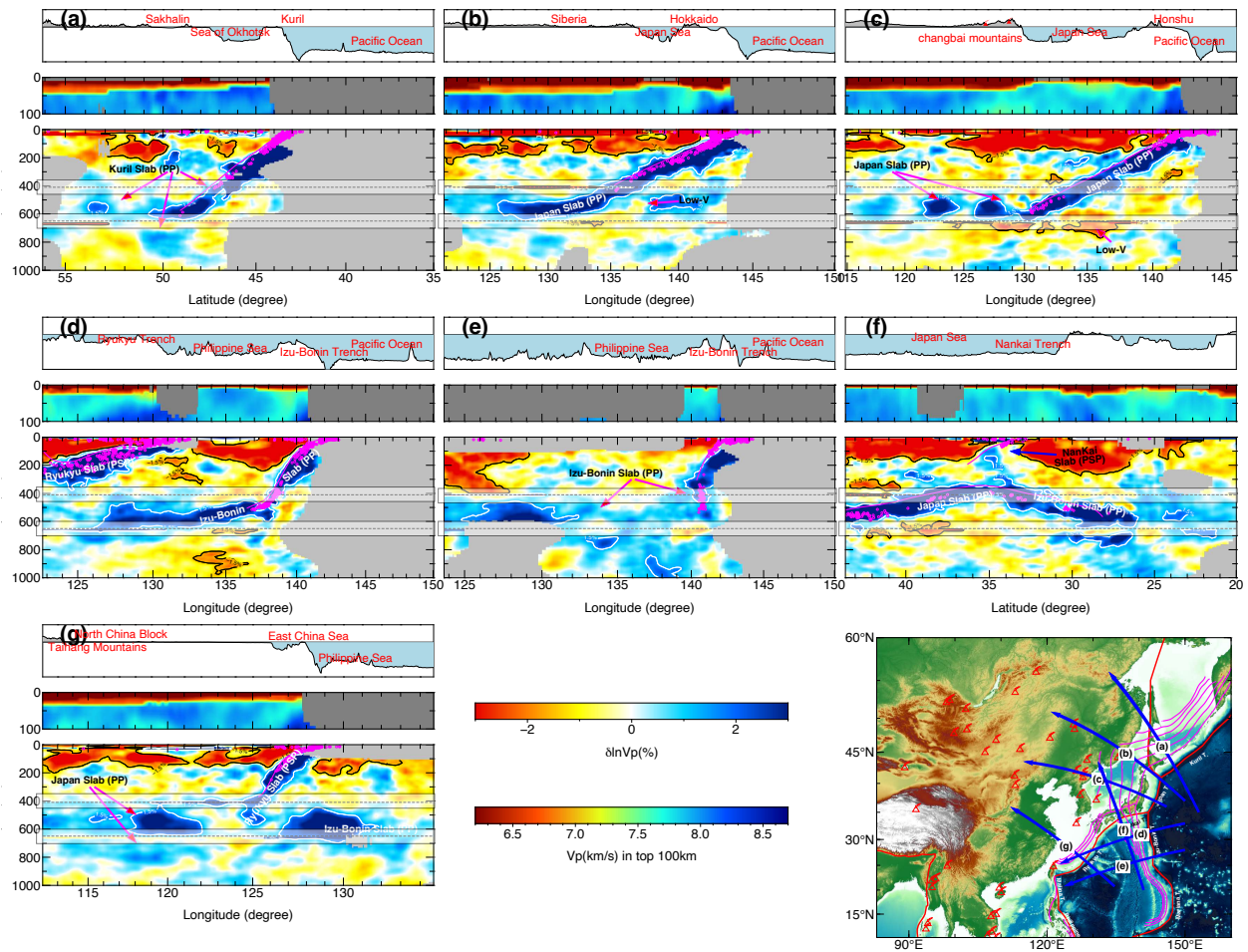


Figure 2.17 Cross-Sections of $\delta \ln V_p$ perturbation for northwestern Pacific subduction zones. Annotations are similar to Figure 2.11.

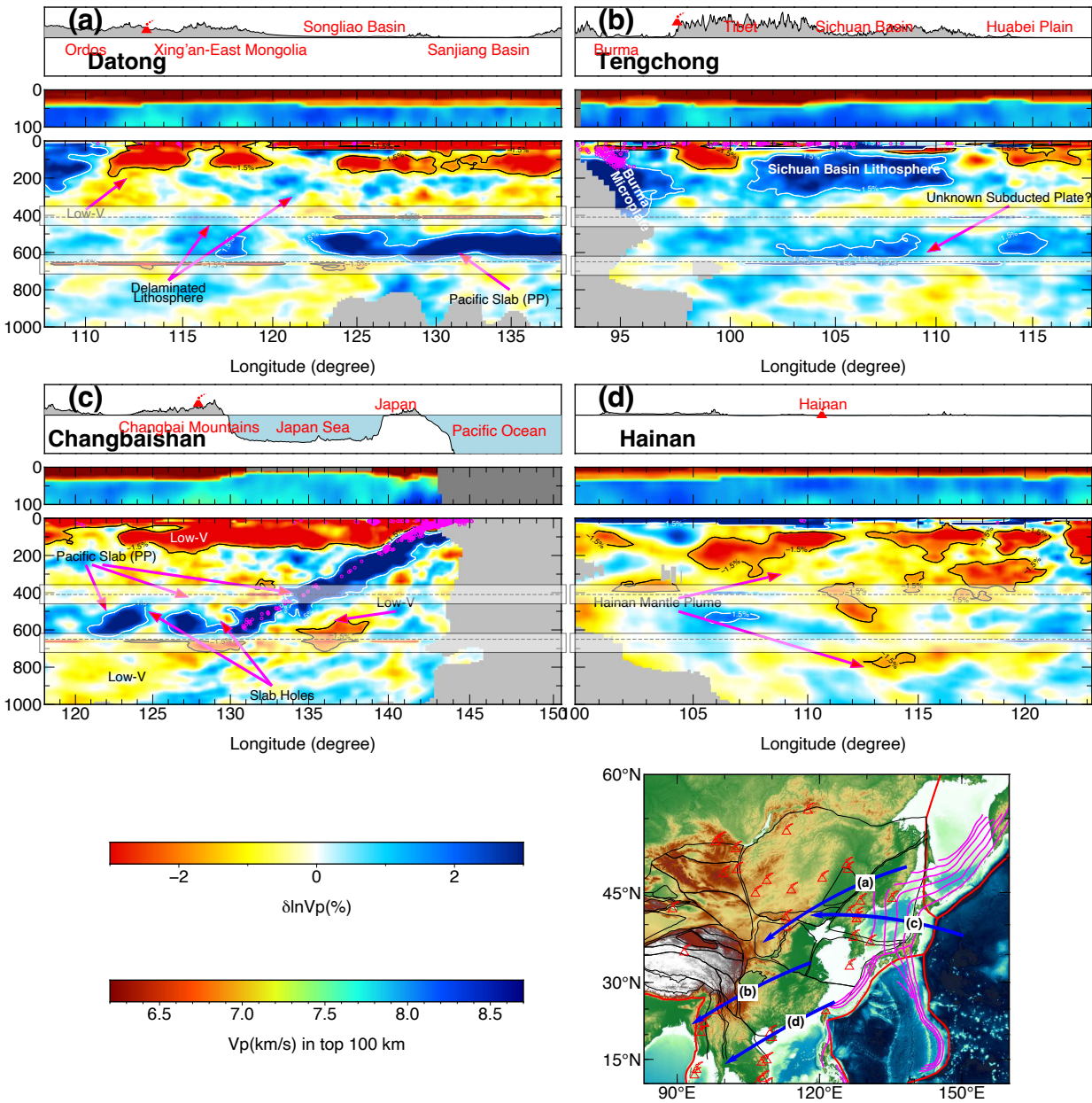


Figure 2.18 Cross-Sections of $\delta \ln V_p$ perturbation for 4 intra-plate volcanic fields in East Asia. Annotations are similar to Figure 2.13.

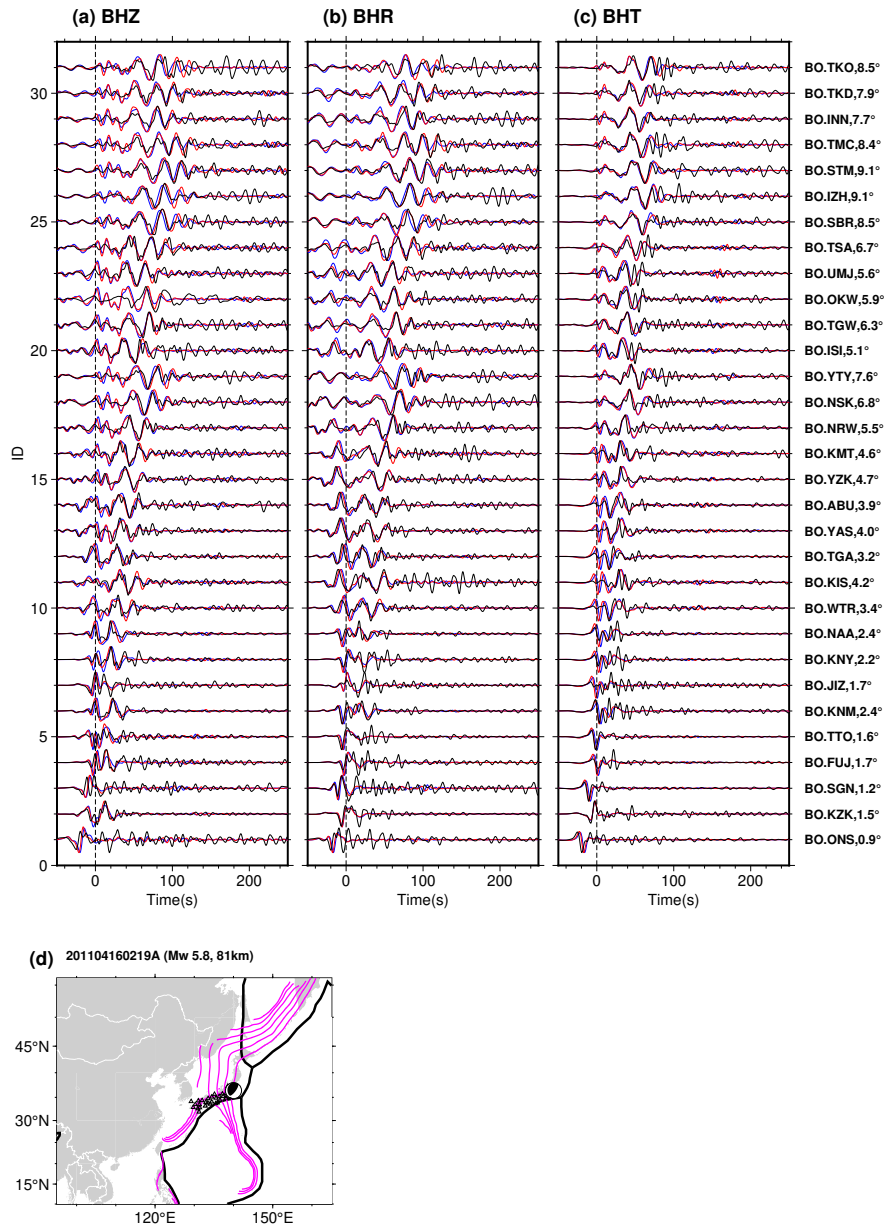


Figure 2.19 Waveform comparisons and S-wave triplication for Earthquake GCMT ID: 201104160219A. (a-c) Waveform comparisons across stations (black triangles in panel (d)) for the vertical (BHZ), radial (BHR), and tangential (BHT) components, featuring observed data (black), starting model synthetics (blue), and final model synthetics (red). Each station's waveforms are normalized and aligned with theoretical S wave arrival times based on the AK135 model [Kennett et al., 1995], indicated by black dashed lines. Station names and great circle distances to the event are displayed adjacent to (c). Waveforms are ranked by the station's longitude.

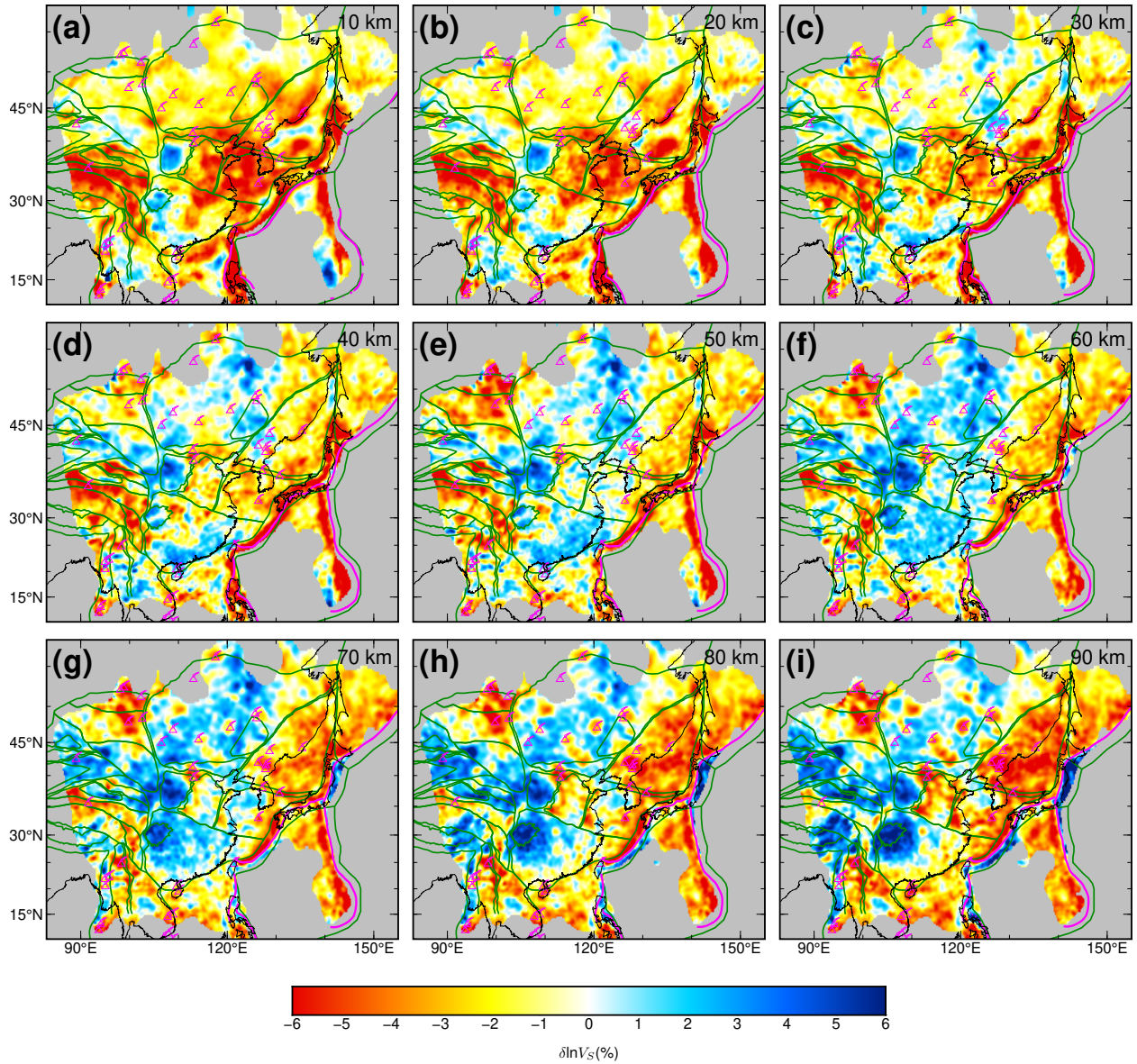


Figure 2.20 Maps of $\delta \ln V_s$ perturbations above 100 km depth. The reference model is the 1-D average of the 3-D model. Other features are the same as Figure 2.4.

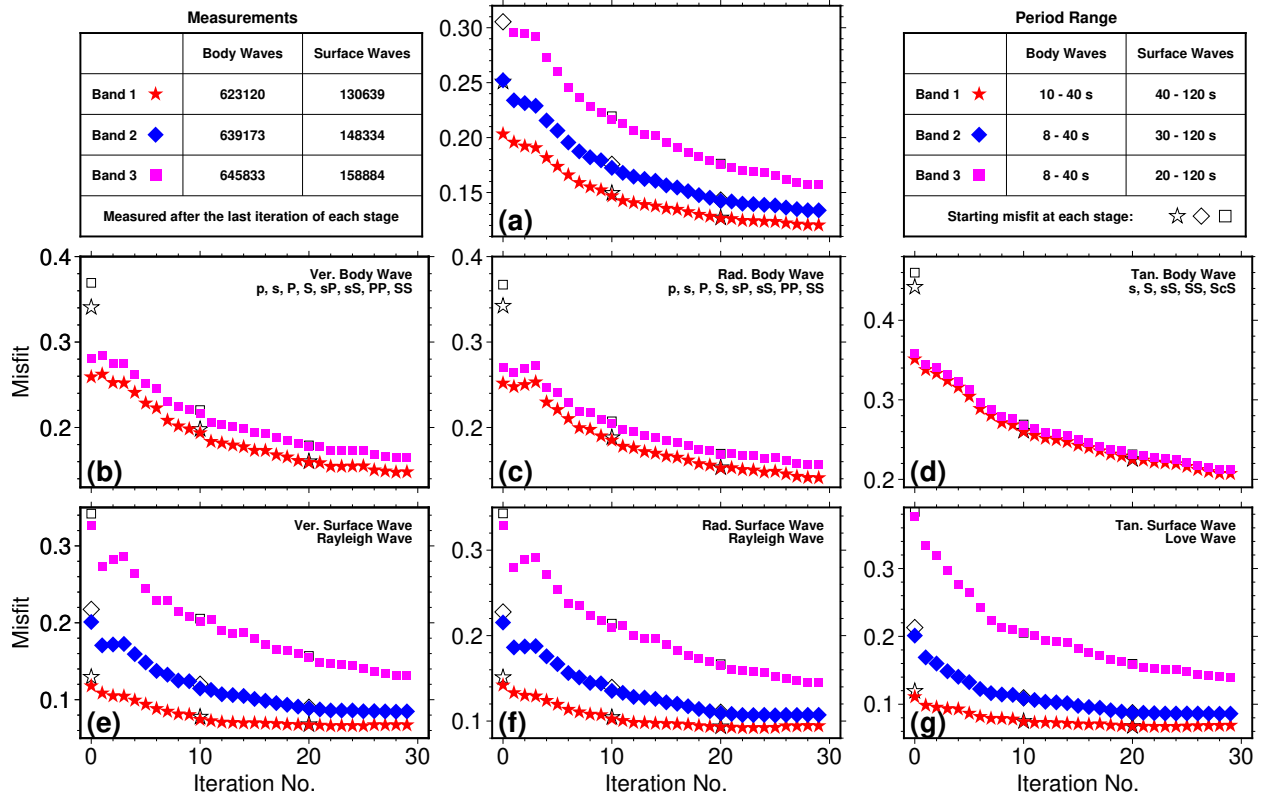


Figure 2.21 Misfit reduction curves with respect to frequency bands in three stages: Stage 1 (red stars), Stage 2 (blue diamonds), and Stage 3 (magenta squares), similar to Figure 2.3. Open symbols represent the misfit prior to source inversion. Unlike Figure 2.3, this figure additionally displays the misfit for each stage, encompassing more than just their respective iterations.

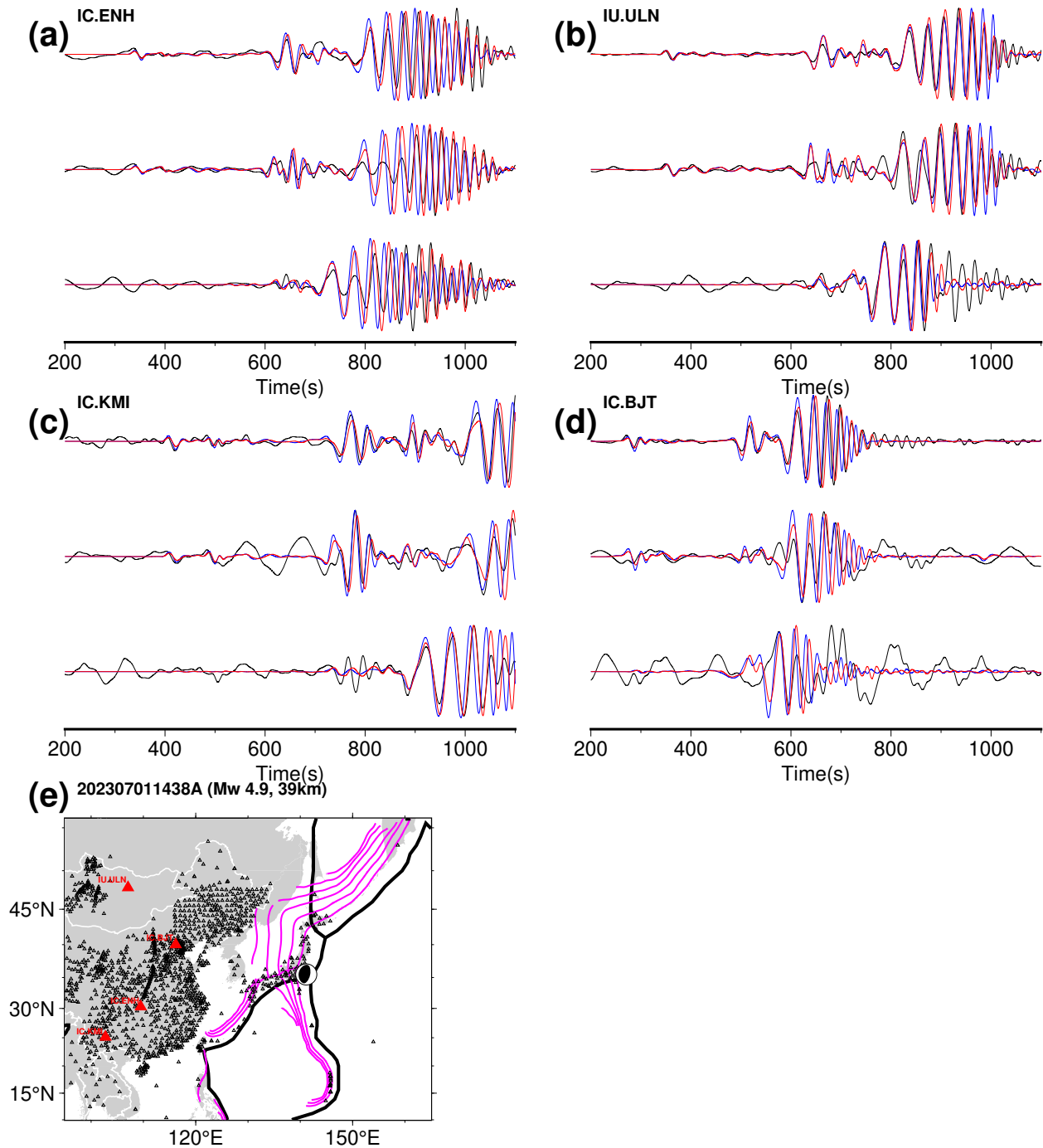


Figure 2.22 Waveform comparisons for an earthquake (GCMT ID: 202307011438A) recorded by selected stations marked by the red triangle in (e). The waveforms are filtered between 20 to 120 s. Other features follow the layout of Figure 2.10. (a-d) Waveforms at different stations, comparing observed data (black) with synthetics from the starting model and the GCMT source model (blue), and synthetics from the final model and the GCMT source (as we have not inverted the source for this event) (red).

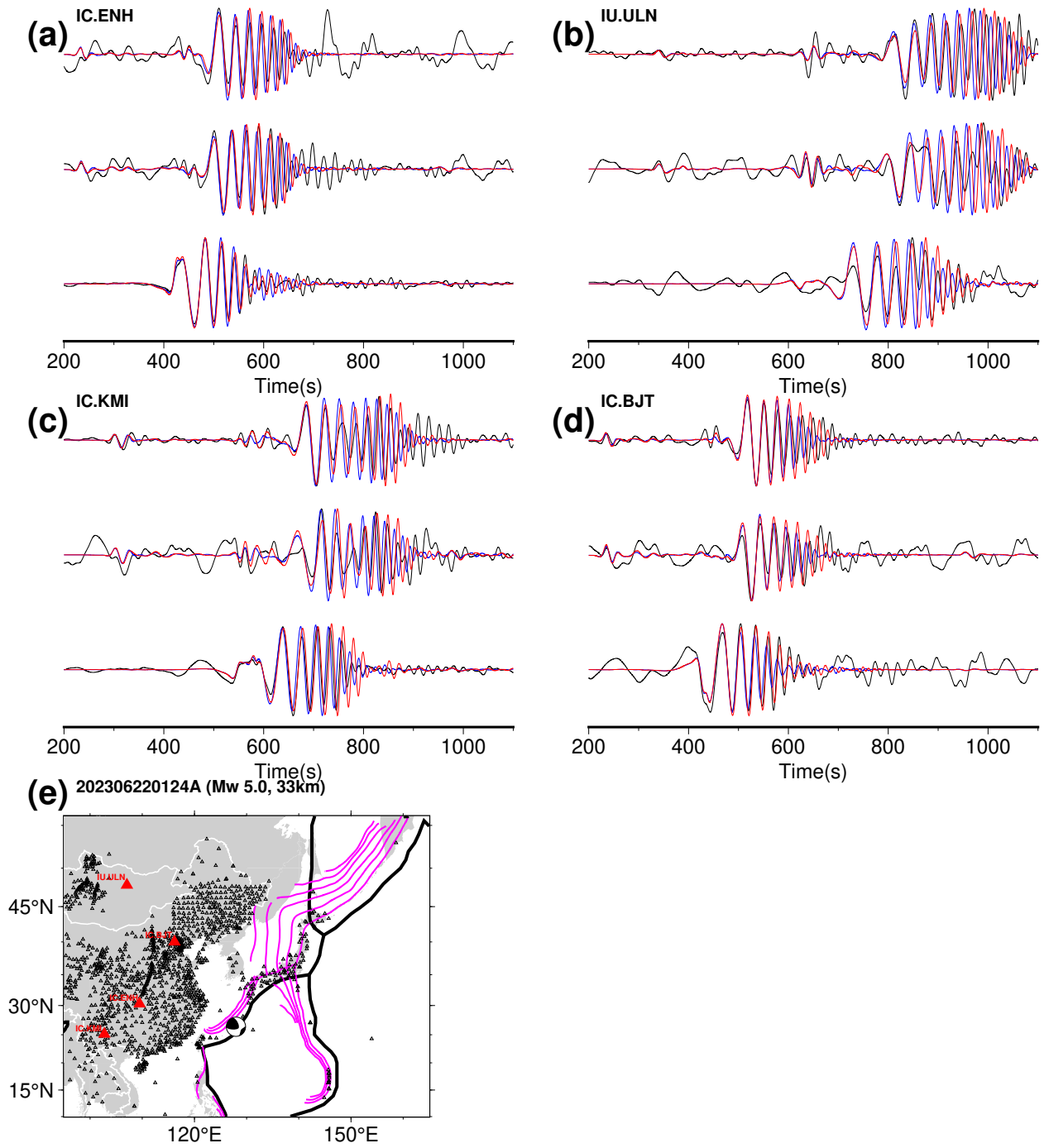


Figure 2.23 Similar to Figure 2.22 but for a different earthquake (GCMT ID: 202306220124A).

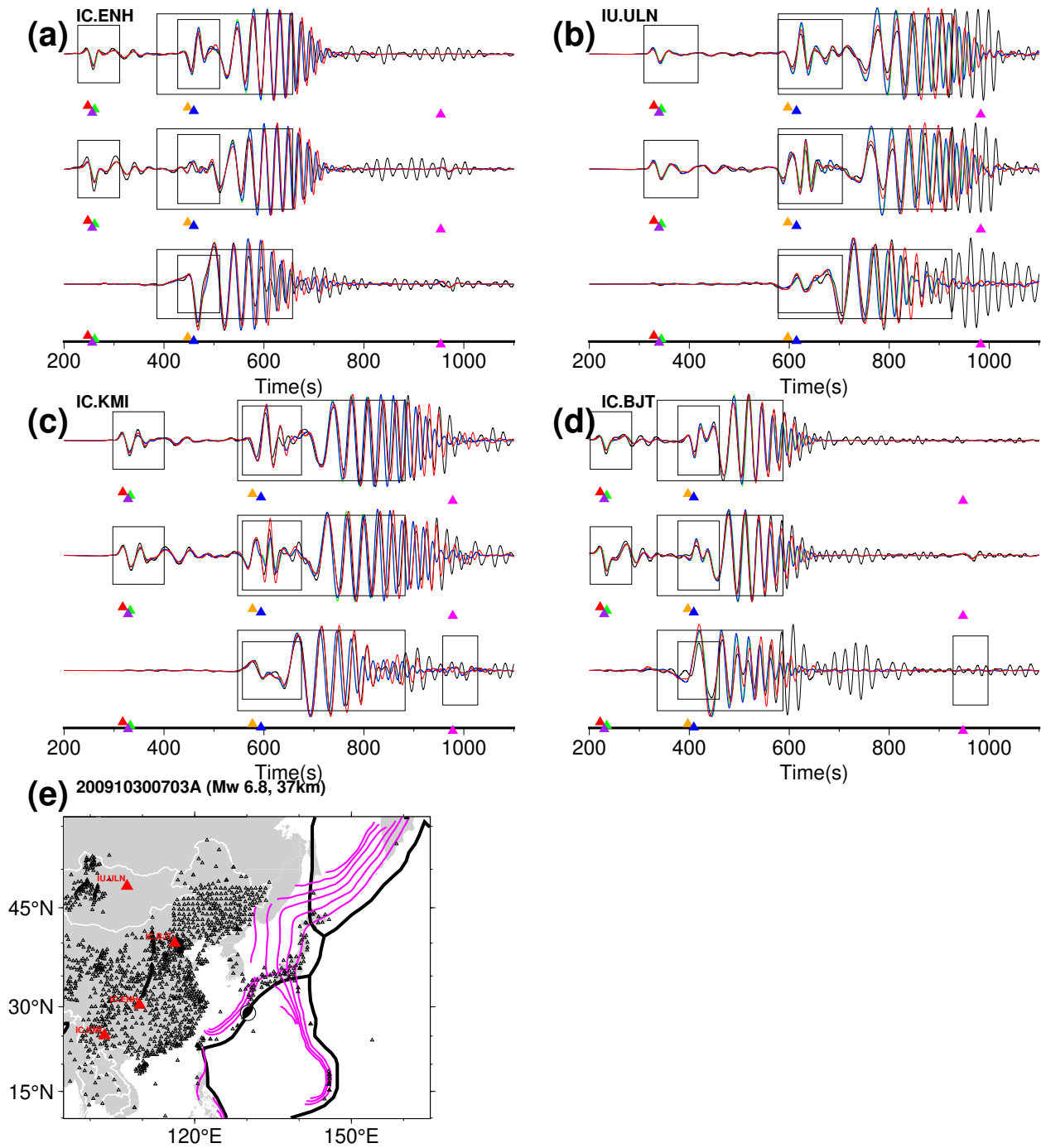


Figure 2.24 Similar to Figure 2.22, this figure relates to a different earthquake (GCMT ID: 200805071602A). As this event is included in the inversion process, measurement windows are marked with black boxes for clarity. Additionally, synthetics from the initial model and the inverted source model in stage 1 are displayed in green, closely aligning with those from the starting model and the GCMT source model, illustrated in blue.

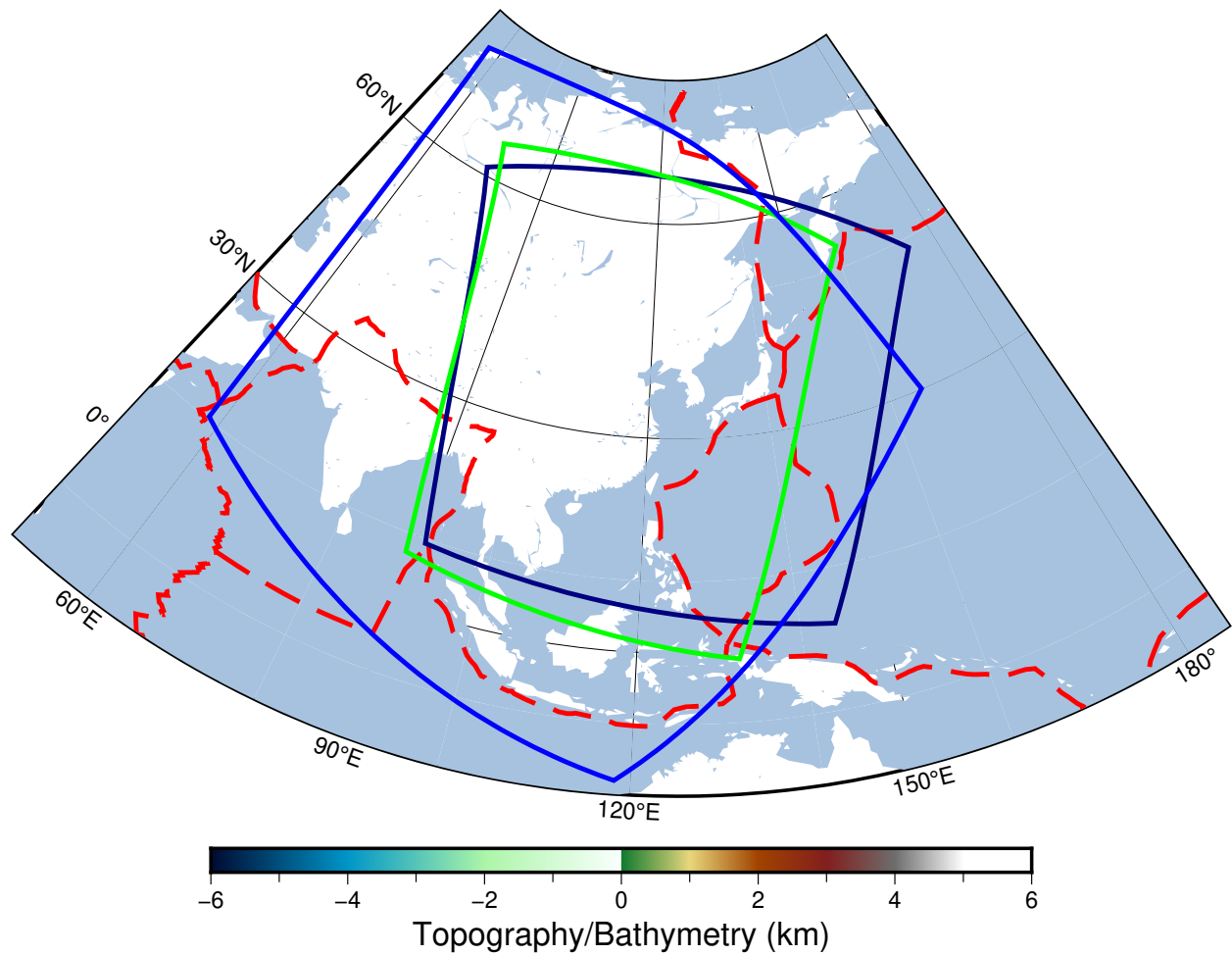


Figure 2.25 Study region comparison for EARA2023 (navy blue), FWEA18 (green), and EARA2014 (blue), with plate boundaries from the NUVEL-1 model [DeMets et al., 1990] delineated by red dashed lines. Tectonic units within China are outlined in black.

Table 2.1 Summary of weighting terms in the misfit function

Weighting term	Formula	Applied on	Comment
categorical weighting W_c	$\frac{1}{\sum_{e=1}^{N_e} \sum_{s=1}^{N_{e,s}} N_{c,e,s,w}}$	Each category	e : event; s : station; c : category; w : window; N_e : number of events; $N_{e,s}$: number of stations for event e ; $N_{c,e,s,w}$: number of measurement windows in category c for event e and station s
geographical weighting $W_{e,s}$	$\frac{1}{\sum_{j=1}^{N_{e,s}} e^{-\left(\frac{\Delta_{sj}}{\Delta_0}\right)^2}}$	Each station	s : current station; j : other stations but include s itself; e : event; $N_{e,s}$: number of stations for event e ; Δ_{sj} : Epicenter distance between station s and station j ; Δ_0 : reference distance, with detail in note ^a
quality weighting W_w	$f_1(\text{CC})f_2(\text{SNR})f_3(\delta t_{\text{CC}})$	Each measurement window	CC: cross-correlation coefficient; SNR: signal to noise ratio; δt_{CC} : time shift obtained from cross-correlation; f_1 , f_2 , and f_3 are defined in the same manner as equation (3) to equation (6) in Tao et al. [2018]

^aFor the geographical weighting, we select Δ_0 as $\frac{1}{\sum_{j=1}^{N_{e,s}} e^{-\left(\frac{\Delta_{sj}}{\Delta_0}\right)^2}} = \frac{1}{3} \max_x \frac{1}{\sum_{j=1}^{N_{e,s}} e^{-\left(\frac{\Delta_{sj}}{x}\right)^2}}$. This definition is consistent with Ruan et al. [2019].

CHAPTER 3

DEEP LEARNING FOR DEEP EARTHQUAKES: INSIGHTS FROM OBS OBSERVATIONS OF THE TONGA SUBDUCTION ZONE

An edited version of this chapter is under review at Geophysical Journal International. Ziyi Xi, Songqiao Shawn Wei, Weiqiang Zhu, Gregory C. Beroza, Yaqi Jie, and Nooshin Saloor. Deep Learning for Deep Earthquakes: Insights from OBS Observations of the Tonga Subduction Zone

3.1 Abstract

Applications of machine learning in seismology have greatly improved our capability of detecting earthquakes in large seismic data archives. Most of these efforts have been focused on continental shallow earthquakes, but here we introduce an integrated deep-learning-based workflow to detect deep earthquakes recorded by a temporary array of ocean-bottom seismographs (OBSs) and land-based stations in the Tonga subduction zone. We develop a new phase picker, PhaseNet-TF, to detect and pick P - and S -wave arrivals in the time-frequency domain. The frequency-domain information is critical for analyzing OBS data, particularly the horizontal components because they are contaminated by signals of ocean-bottom currents and other noise sources in certain frequency bands. PhaseNet-TF shows a much better performance in picking S waves compared to its predecessor PhaseNet. The predicted phases are associated using an improved Gaussian Mixture Model Associator GAMMA-1D and then relocated with a double-difference package *teletomoDD*. We further enhance the model performance with a semi-supervised learning approach by iteratively refining labeled data and retraining PhaseNet-TF. This approach effectively suppresses false picks and significantly improves the detection of small earthquakes. The new catalog of Tonga deep earthquakes contains more than 10 times more events compared to the reference catalog that was analyzed manually. This deep-learning-enhanced catalog reveals Tonga seismicity in unprecedented detail, and better defines the lateral extent of the double-seismic zone at intermediate depths and the location of 4 large deep-focus earthquakes relative to background seismicity. It also offers new potential for deciphering deep earthquake mechanisms, refining tomographic models, and understanding subduction processes.

3.2 Introduction

Detecting and locating earthquakes in subduction zones plays a pivotal role in advancing the understanding of subduction processes and earthquake physics. In particular, earthquakes deeper than 50 km provide critical information on slab geometry, slab mineral dehydration and transformation, and the interaction between the slab and surrounding mantle [Green and Houston, 1995, Zhan, 2020]. Initial studies, such as Ruff and Kanamori [1980], established meaningful connections between seismicity and physical attributes of subduction zones, such as the lateral extent and penetration depth of the Wadati-Benioff zone, the age of the subducting lithosphere, convergence rates, and back-arc spreading. In recent years, comprehensive global slab models like Slab1.0 [Hayes et al., 2012] and Slab2 [Hayes et al., 2018] have leveraged high-accuracy regional seismicity catalogs to refine slab geometry. Precise earthquake distributions also help reveal the underlying mechanisms of intermediate-depth ($\sim 70\text{--}300$ km) and deep-focus (300–700 km) earthquakes [Wiens et al., 1993, Brudzinski et al., 2007, Kita et al., 2010, Wei et al., 2017, Chen et al., 2019b, Florez and Prieto, 2019]. However, most global studies suffer from limited local station coverage, especially offshore, and most regional studies with temporary deployments lack sufficient duration, which limits high-precision earthquake locations in large numbers.

Recent advances in the applications of deep learning methods in seismology greatly increase the information content that can be extracted from seismic datasets by detecting many more earthquakes [Mousavi and Beroza, 2022, 2023]. Initial efforts, such as Gentili and Michelini [2006], Ross et al. [2018], used simple neural networks for detecting seismic phase arrivals, a foundational step in earthquake localization. Subsequent developments incorporated biomedical image segmentation algorithms, notably the U-Net architecture [Ronneberger et al., 2015], to create highly effective deep-neural network (DNN) phase pickers like PhaseNet [Zhu and Beroza, 2019]. The Transformer architecture [Vaswani et al., 2017] has further inspired new models, such as the EQTransformer [Mousavi et al., 2020], which leverages attention mechanisms to enhance phase detectability. For seismic phase association that links seismic arrivals to preliminary event origins, significant improvements have been achieved through Gaussian mixture models [Zhu et al., 2022] and graph

neural networks (GNN) [McBrearty and Beroza, 2023]. These machine-learning-based techniques outperform traditional methods in both phase-picking [Baer and Kradolfer, 1987, Sleeman and van Eck, 1999] and phase association [Zhang et al., 2019b].

Most machine-learning studies have been focused on continental, shallow earthquakes. Limited attention has been given to deep earthquakes in subduction zones. For instance, PhaseNet and GaMMA were developed using seismic data from Northern California, where most earthquakes occur at depths shallower than 20 km. EQTransformer was trained with the STanford EArthquake Dataset (STEAD) [Mousavi et al., 2019], which, despite its global scope, contains earthquakes predominantly shallower than 100 km. Generalized Seismic Phase Detection [Ross et al., 2018] is developed with vast hand-labelled data archives of the Southern California Seismic Network, which is also dominated by continental earthquakes. Studies utilizing these methods, [e.g., Chai et al., 2020, Liu et al., 2020, Park et al., 2020, Ross et al., 2020, Tan et al., 2021, Wilding et al., 2023, Liu et al., 2023, Gong et al., 2023] similarly concentrate on continental earthquakes.

Since many subduction zones are covered by oceans, offshore seismic data is critical for investigating subduction zone earthquakes. However, data from ocean-bottom seismographs (OBSs) is generally noisier than that from land-based stations because of ocean-bottom currents, seismometer tilting, instrument coupling, etc. Recent efforts, such as PickBlue, utilize existing packages such as PhaseNet and EQTransformer to process OBS data but show a lower performance compared to continental data [Bornstein et al., 2023]. Therefore, a new phase picker is needed for processing OBS data.

The Tonga subduction zone hosts abundant intermediate-depth earthquakes and produces the majority of deep-focus earthquakes, and thus serves as a unique natural laboratory for studying deep earthquakes. However, studying Tonga's earthquakes has been challenging. Global catalogs, such as the ISC EHB catalog, mainly rely on a handful of land-based stations on the islands of Tonga and Fiji. These catalogs provide foundational information on Tonga earthquakes, but the hypocentre precisions, particularly in the vertical direction, are limited due to the lack of local stations. Since 1993, a few temporary seismic deployments, including broadband OBSs, greatly

improved the data coverage and earthquake hypocentre precisions in this region, leading to groundbreaking discoveries [Wiens et al., 1993, 1994, Wei et al., 2017]. However, the short duration of these deployments has limited the number of analyzed earthquakes and motivates a systematic effort to mine these datasets for more small earthquakes.

Innovative approaches are needed to tackle earthquake detection and location in the Tonga subduction zone. First, the phase picker must be capable of handling noise in OBS data that is sometimes enriched in specific frequency ranges. To solve this problem, we introduce a new phase picker PhaseNet-TF to detect seismic arrivals in the time-frequency domain. Second, the phase associator should be more efficient than the conventional back-projection-based methods and account for the change in seismic velocity with respect to depth. We develop a new associator GaMMA-1D to associate arrivals output from PhaseNet-TF. Finally, given the limited data, optimizing the use of existing data is crucial. Here we build a new semi-supervised learning-based workflow to analyze seismic data from a 1-year temporary deployment.

3.3 Data

In this study, we analyzed seismic data recorded by a temporary seismic array deployed in the Tonga subduction zone from November 2009 to December 2010 (Figure 3.1). This array included 49 OBSs (network ID YL) with either Guralp CMG3T_120sec (100 Hz sample rate) or LDEO OBS Sensor Mk2 seismometers (40 Hz sample rate), 17 island-based stations (network ID Z1) with Guralp CMG40T, Streckeisen STS-2, or Nanometrics Trillium 120 seismometers (40 Hz sample rate), and one GSN station MSVF (network ID II) with a Geotech KS-54000 Borehole seismometer (20 Hz sample rate).

We compiled a reference catalog of local earthquakes, consisting of 1,163 events, 42,256 *P*-wave arrivals, and 14,852 *S*-wave arrivals (Table 3.1) that were manually picked with the Antelope software [Wei et al., 2017]. This catalog is hereinafter called the manually picked reference catalog or the reference catalog. We created a reference dataset by windowing three-component waveforms 5 minutes before and 5 minutes after each *P*-wave arrival. This window is sufficiently long to include the corresponding *S* arrival for the same event. We subsequently removed the instrumental

response to obtain displacement waveforms in three components: vertical (Z), east (E or 1), and north (N or 2). We further resample the data to 40 Hz. No additional preprocessing, such as filtering and component rotation, was applied.

We also created a continuous dataset by partitioning the three-component continuous waveforms recorded by this array into 10-day segments and subsequently removing the instrumental response. Since not all stations had complete data from November 2009 to December 2010, we replaced the missing data with zeros. Incomplete components were also accepted, with missing channels filled with zeros. Given the large volume of the data, exceeding 2TB in the miniSEED format, we used the mseedindex package and an ObsPy wrapper [Beyreuther et al., 2010] to construct a miniSEED database. This facilitated efficient data analysis and improved machine learning I/O performance.

3.4 Methods

3.4.1 Phase arrival-time picking by PhaseNet-TF

We develop a new phase picker PhaseNet-TF, based on its predecessor PhaseNet [Zhu and Beroza, 2019], to leverage the benefits of the time-frequency domain, which excels in capturing both temporal and spectral features of seismic data. PhaseNet-TF adapts the architecture of DeepLabv3+ [Chen et al., 2018] to accommodate data in the time-frequency domain (Figure 3.2). DeepLabv3+ is a state-of-the-art semantic image segmentation model that incorporates an encoder-decoder structure to refine object boundaries in segmentation tasks. As part of the renowned DeepLab model series, it offers top-tier performance in a wide array of applications, ranging from autonomous driving to medical image analysis, and outperforms earlier models such as U-Net [Ronneberger et al., 2015], which was used in PhaseNet. A seismic spectrogram acts as an image that represents the time-frequency distribution of phase signal and noise. A spectrogram is generated by applying the short-time Fourier transform (STFT) to three-component time-domain waveforms, and thus consists of 6 components: i.e., the real and imaginary parts of the three-component waveform spectra. Using the spectrogram as an input, DeepLabv3+ produces a pixel-level classification image that matches the dimension of the spectrogram. This output highlights the relative positions of signal and noise and is subsequently processed by a multilayer perceptron to estimate phase and noise

probabilities in the time domain.

The manually picked reference dataset is divided into training, validation, and test datasets using stratified sampling, with a distribution ratio of 90:5:5. The ratio for the training dataset is relatively high as we have a limited amount of data. This approach ensures an equitable representation of both P and S waves, especially given the fact that the number of S wave picks is about 1/5 of that of P wave picks. The input waveform window is 120 seconds long, with the P wave arrival initially centered at the 10-second timestamp. We augment the training dataset in 3 ways. First, we randomly shift the waveform windows to prevent the model from overfitting to specific phase arrival positions. The 10-minute long waveform in our dataset is sufficient for cutting and shifting the 120-second window. Second, we randomly stack two signal-bearing windows or one signal-bearing and one noise-only window. The ratio for stacking is fine-tuned as a hyperparameter, allowing the model to adapt to more complex real-world scenarios and preventing it from mistakenly learning that a 120-second window always contains only two phases (P and S). Third, we stabilize the input data through normalization by subtracting the mean and dividing by the standard deviation of the 120-second window. When applying random stacking, each window is also normalized before stacking, and then the entire stacked waveform is normalized. For the validation and testing sets, we do not shift or stack windows and only stabilize the input data through normalization.

We formulate the training labels to represent the probability of phase arrival times and use the Kullback-Leibler (KL) divergence for the loss function. The KL divergence differs from the cross-entropy loss used in PhaseNet only by a constant value, so they are equivalent for optimization. We define the probability at time t as follows:

$$y_{\text{true}}(t) = e^{-\frac{(t-t_0)^2}{2\delta^2}}, \quad \text{where } |t - t_0| \leq 3\delta \quad (3.1)$$

Here, t_0 is the phase arrival time, and δ is the width of the label (20 points, or 0.5 s in our case). This label definition smooths the phase arrival time and allows for the quantification of classification uncertainty through the shape of the model predictions. The KL divergence, measuring the

similarity between two probability distributions, is defined as:

$$L(y_{\text{true}}, y_{\text{predict}}) = y_{\text{true}} * \log \frac{y_{\text{true}}}{y_{\text{predict}}} \quad (3.2)$$

where y_{predict} is the model prediction. A lower KL divergence value indicates a higher similarity between y_{true} and y_{predict} . The KL divergence is evaluated across three output channels: probabilities of P , S , and noise. The sum of these probabilities is fixed as 1 at each time stamp.

We use the open-source PyTorch library, including AdamW [Loshchilov and Hutter, 2019] for optimization and MultiStepLR for learning rate scheduling. AdamW is widely used in computer vision tasks and deviates from the traditional Adam optimizer by decoupling the weight decay from the gradient update. MultiStepLR adjusts the learning rate at specific epochs, decreasing it by a fixed rate of 0.6 in our case. Our model is trained for 400 epochs, starting with a learning rate of 0.0004, which decays at epochs 15, 30, 45, and 60. We also add an L2 regularization term with a weight decay of 0.001 to the loss function to mitigate overfitting. Early stopping is implemented to prevent overfitting and save computational time. Training is halted if the validation loss does not improve for 30 consecutive epochs. For our reference dataset, the training took about 5 hours on 16 NVIDIA Tesla V100 GPU cards at the MSU HPCC.

We apply the trained PhaseNet-TF model to the continuous dataset for phase detection. The output is continuous probability distributions for P waves, S waves, and noise. We first partition the continuous waveforms into 120-second segments with a 60-second overlap between consecutive segments. Each segment is normalized in the same way used for model training. Then we applied the model to the entire continuous dataset, which took 16 hours on 4 NVIDIA Tesla V100 GPUs to process. The output is 120-second segments of probability distributions for P wave, S wave, and noise. We combine these 120-second segments into a single continuous time series by taking the final output probability as the maximum value from the overlapping predictions. Peak probabilities larger than 0.5 are counted as positive picks.

3.4.2 Phase association by GaMMA-1D

Associating phase picks to specific earthquakes is necessary for locating events and eliminating unreliable picks. We use GaMMA-1D, a Bayesian Gaussian Mixture Model Associator with a 1D

velocity structure, which is an improved version of GaMMA [Zhu et al., 2022]. While GaMMA-1D retains the Gaussian mixture model framework of its predecessor GaMMA for phase association, it improves calculating phase arrival times by using a fast-sweeping method to solve the Eikonal equation based on a 1D velocity model AK135 [Kennett et al., 1995]. In contrast to GaMMA which used a uniform half-space for arrival time predictions, Gamma-1D uses a 1D velocity model, which is critical for the large depth range of earthquakes in Tonga. Events associated with less than 10 picks are discarded. Figure 3.3 shows the association results for a densely packed sequence of phase picks.

3.4.3 Earthquake relocation by teletomoDD

We use teletomoDD [Pesicek et al., 2010], a package for double-difference seismic tomography and relocation, to relocate all events associated with GaMMA-1D in the previous step. The 3D seismic velocity model is fixed during inversions and is adopted from the TX2019slab model [Lu et al., 2019]. We apply a bootstrap resampling technique to estimate relocation uncertainties and filter out events with large uncertainties. We create 1,000 subsets of the data by randomly excluding 30% of the stations from each subset. After relocating events in these 1,000 subsets, we compute the mean and standard deviation of the hypocentre and origin time of each event. We eliminate events with a standard deviation in longitude and latitude greater than 0.1 degrees, in depth of greater than 10 km, or in origin time of greater than 1 second. This approach effectively removes unreliable picks from PhaseNet-TF and GaMMA-1D as well as events that are poorly constrained. The relocation output catalog contains the hypocentres and origin times of the remaining events and the corresponding *P*- and *S*-wave arrival times.

3.4.4 Semi-supervised-learning-based workflow

Since there are only 1,163 manually picked events out of presumably tens of thousands of earthquakes in the Tonga subduction, the reference catalog and dataset may limit phase detection capability. Therefore, we utilize a semi-supervised learning strategy to iteratively refine labeled picks and retrain PhaseNet-TF (Figure 3.4). This approach integrates a limited labeled dataset with a larger pool of unlabeled data for model training. In Iteration #1, we train PhaseNet-TF with the

original labeled dataset, i.e., the manually picked reference dataset. This model is then applied to the continuous dataset, generating new phase picks that may include false detections. The subsequent steps of phase association by GAMMA-1D and event relocation by teletomoDD filter out unreliable picks and events with large uncertainties. Compared to the reference catalog, the output catalog thus contains a larger number of reliable picks. Similar to the reference dataset, we create a new labeled dataset by windowing three-component waveforms 5 minutes before and 5 minutes after each P wave arrival from the output catalog. This newly labeled dataset is divided into the training, validation, and test datasets at a ratio of 90:5:5 to train the PhaseNet-TF model in the next iteration. As more picks predicted by deep learning are added to the training dataset, one can expect more picks and events to be detected, at the cost of increasing arrival time residuals compared to the reference dataset. We continue this iterative workflow for several iterations until the number of events reaches a plateau and the arrival time residuals do not increase dramatically.

3.5 Results

3.5.1 PhaseNet-TF model assessment

We assess the PhaseNet-TF model in each semi-supervised-learning iteration using the test dataset, which is 5% of the labeled dataset in the corresponding iteration (Table 3.2). The evaluation metrics include precision, recall, F1 score, and arrival-time residuals compared to the labeled catalog. This labeled catalog is the manually picked reference catalog for Iteration #1 and is the output catalog from the previous iteration for Iteration #2 and #3. Predicted picks with arrival-time residuals smaller than 1 second are considered true positives, whereas predicted picks with larger arrival-time residuals are false positives. Labeled picks that are not predicted by PhaseNet-TF are considered false negatives. Precision, recall, and F1 score are defined as

$$\text{Precision: } P = \frac{T_P}{T_P + F_P} \quad (3.3)$$

$$\text{Recall: } R = \frac{T_P}{T_P + F_N} \quad (3.4)$$

$$\text{F1: } F1 = \frac{2PR}{P + R} \quad (3.5)$$

where T_P , F_P , and F_N are the numbers of true positives, false positives, and false negatives, respectively.

In Iteration #1, the PhaseNet-TF output is evaluated against the manually picked reference catalog. For the P wave, the model exhibited a precision of 0.99, a recall of 0.99, and an F1 score of 0.99. For the S wave, the corresponding values are 0.97, 0.99, and 0.98, respectively. Figure 3.5 shows examples of seismograms, spectrograms, and prediction probabilities from iteration #1 at 2 OBSs and 2 land-based stations. In the following iterations, the evaluation metrics remain at the same high level, validating the semi-supervised learning approach and the robustness of PhaseNet-TF in accurately identifying phase arrivals.

3.5.2 Phase association and earthquake relocation assessments

Phase association and earthquake relocation serve as critical filters for eliminating unreliable predicted picks and poorly constrained events. When manual picks are associated with GaMMA-1D, some of them are missed in the association catalog and considered false negative picks. The picks associated with GaMMA-1D are subsequently used for event relocation by teletomoDD with bootstrap resampling. Events with large uncertainties and their corresponding picks are discarded during the relocation process. We assess these processes (Table 3.3) using the reference catalog that contains manual picks associated with the Antelope software [Wei et al., 2017]. When comparing the reference catalog and the output catalog by GaMMA-1D or teletomoDD, events with origin-time residuals smaller than 15 seconds are considered true positive events, whereas events with larger origin-time residuals are false positives. If some picks that were associated with a single event by Antelope are associated with multiple events by GaMMA-1D, the new events are also counted as false positives. Because there are no picks added during this processing, the recall for P - or S -wave arrivals reflects the picks eliminated during association and relocation, and a high recall value is desired. In contrast, new events may be added during the association process, lowering precision, whereas existing events may be discarded during the association and relocation processes, lowering recall. Thus, the F1 score for events that balance precision and recall serves as a better indicator of the filtering performance.

As shown in Table 3.3, the recall for *P*-wave arrivals is 0.97 after association and 0.95 after relocation, suggesting that GaMMA-1D and teletomoDD are highly effective in retaining manually picked *P*-wave arrivals. However, the recall for *S*-wave arrivals is 0.92 after association and 0.86 after relocation. These numbers indicate that about 8% of the manually picked *S*-wave arrivals are not successfully associated with GaMMA-1D, which impacts the subsequent relocation performance. This could be attributed to either the limitations of GaMMA-1D in associating *S*-wave arrivals or inaccurate manual *S*-wave picks. Nonetheless, the overall performance of the association and relocation filtering processes remains promising.

3.5.3 Phase detection and event relocation on continuous data

In Iteration #1, the PhaseNet-TF model is trained by the manually picked reference dataset. When applying this model to the continuous data, PhaseNet-TF detects 294,050 *P*-wave arrivals and 112,547 *S*-wave arrivals, which is substantially more than the number of picks in the reference catalog. Figure 3.6 demonstrates the performance on one hour of continuous data. These arrivals are associated with GaMMA-1D in a preliminary catalog. In this step, about 10% of *P*- and 30% of *S*-wave arrivals are discarded, and the associated catalog consists of 13,111 events with 265,439 *P*-wave, and 79,380 *S*-wave arrivals. These events generally align with the reference catalog but are more scattered (Figures 3.7a, 3.7b, 3.8a, and 3.8b). Many events in the mantle wedge are not reliable as they have a large azimuthal coverage gap. The subsequent relocation and error estimation filter out most of these outlier events, leaving a new catalog of 9,427 events with 217,254 *P*-wave, and 63,590 *S*-wave arrivals (Figures 3.7c and 3.8c, Table 3.1). When comparing this catalog with the reference catalog, the recall for *P*-waves, *S*-waves, and events are 0.94, 0.83, and 0.96, respectively, and the standard deviations of arrival-time residuals are 0.14 and 0.15 seconds for *P* and *S* waves, respectively (Figures 3.9a and 3.9b), suggesting that our workflow can effectively detect seismic arrivals and earthquakes that were manually picked. More importantly, the new catalog contains dramatically more *P*- and *S*-wave arrivals and events (Table 3.1).

Leveraging this new catalog, we assemble a new labeled dataset enriched with phase arrivals detected in Iteration #1 of the semi-supervised learning workflow. This new dataset (120 GB) is

substantially larger than our initial reference dataset (22 GB). The increased dataset size requires additional computational resources, extending the training time from 5 to 24 hours while utilizing the same number of GPUs. In Iteration #2, this new PhaseNet-TF model is applied to the continuous dataset again, resulting in significantly more arrivals and events (Figures. 3.7d and 3.8d). The standard deviation of arrival-time residuals for P waves remains 0.15 seconds (Figure 3.9c), but that for S waves increases from 0.15 to 0.23 seconds (Figure 3.9d). In Iteration #3 which uses the output catalog from Iteration #2 for training, the numbers of arrivals and events and arrival-time residuals remain stable (Figs. 3.7e and 3.8e). We thus cease the semi-supervised learning workflow after Iteration #3, anticipating diminishing returns in further iterations.

Our final catalog from Iteration #3 contains 13,406 relocated events with 372,774 P -wave arrivals and 78,853 S -wave arrivals. Compared with the manually picked reference catalog, our final compilation boasts a factor of 11 times more events, 8 times more P -wave phases, and 5 times more S -wave phases. Figures 3.7 and 3.8 show that our final catalog offers enhanced delineation of both the slab geometry and double seismic zone, demonstrating its superiority over the reference catalog.

3.6 Discussion

3.6.1 Comparison with previous packages

In this study, PhaseNet-TF detects seismic arrivals in the time-frequency domain, different from most other deep-learning phase pickers that work in the time domain. Using a manually picked reference catalog and dataset, we conduct a quantitative and fair comparison between PhaseNet and PhaseNet-TF. We first test the original PhaseNet model that was trained by the Northern California data [Zhu and Beroza, 2019]. We also retrain the PhaseNet architecture with our training dataset from Tonga. The PhaseNet-TF model in Iteration #1 is used for comparison. All models are evaluated on the same testing dataset to ensure a fair comparison.

Table 3.2 highlights the superiority of PhaseNet-TF over PhaseNet in detecting seismic arrivals, particularly for OBS data. PhaseNet with its original weights shows poor performance for both P and S waves. This is an unsurprising outcome given that it was trained with a dataset from a different

tectonic setting and on land-based vs. OBS instruments. When retrained with the Tonga dataset, PhaseNet’s performance is similar to PhaseNet-TF for detecting *P*-wave arrivals but displays a lower performance for *S* waves. Furthermore, the standard deviation of *S*-waves arrival-time residuals for PhaseNet-TF (Iteration #1) are significantly smaller than those for PhaseNet models (Figure 3.9b and 3.9h). These differences indicate that including the time-frequency domain and the new architecture enhances the model’s capability to detect *S* waves. This is because the horizontal components that record *S* waves are much noisier for OBS data compared to land-based stations, due to seismometer tilting and ocean-bottom currents [Webb and Crawford, 1999, Wei et al., 2015].

For associating seismic arrivals from Tonga deep earthquakes, GaMMA-1D exhibits higher performance compared to GaMMA, which was designed for California [Zhu et al., 2022]. That is because GaMMA-1D uses a 1D velocity model AK135, whereas GaMMA assumes a uniform velocity model. We test both GaMMA and GaMMA-1D to associate all manual picks and compare the output catalogs against the reference catalog. When comparing the origin-time and depth residuals, GaMMA-1D consistently achieves superior accuracy to its predecessor GaMMA (Figure 3.10). Table 3.3 lists the evaluation metrics for GaMMA-1D and GaMMA. Compared to GaMMA, GaMMA-1D achieves a similar performance for associating P- and S-wave picks to certain events. However, the low precision of event association (0.76) suggests that GaMMA tends to break a single event into multiple events. This problem will impact the next step of earthquake relocation, resulting in more events with poorer constraints and/or misassociated phases.

Our workflow is readily adaptable to a cloud computing setting through modifications to the machine learning models employed in Quakeflow [Zhu et al., 2023b]. Quakeflow is a cloud-based earthquake monitoring system designed for detecting seismic activity and analyzing source characteristics from continuous waveform data. It currently utilizes PhaseNet for phase picking, GaMMA for phase association, and HypoDD for event relocation. Given the compatibility in model inputs and outputs, these can be smoothly swapped with PhaseNet-TF, GaMMA-1D, and teletomoDD. A future version of Quakeflow potentially provides an efficient and effective solution for enhancing real-time earthquake surveillance and for in-depth analysis of historical seismic data, particularly

for deep earthquakes in subduction zones and OBS data.

3.6.2 Tonga deep earthquakes revealed by the new catalog

With only 1 year of data, our results show unprecedented detail in the Wadati-Benioff zone. Figure 3.11 compares the new catalog against the manually picked reference catalog [Wei et al., 2017] and the ISC EHB Bulletin from 1964 to 2020 [International-Seismological-Centre, 2023]. The latter uses the EHB algorithm [Engdahl et al., 1998] to minimize hypocentre errors, particularly in the vertical direction, and is arguably the most precise global catalog. The general pattern of earthquake distribution remains similar across all catalogs.

Double seismic zones (DSZs), in which intermediate-depths earthquakes occur along two planes parallel to the dip of the slab, are observed in many subduction zones and are attributed to metamorphic reactions that release fluids or volatiles in the slab [e.g., Hacker et al., 2003, Yamasaki and Seno, 2003, Brudzinski et al., 2007]. The Tonga DSZ was initially discovered by Kawakatsu [1985] using focal mechanisms constrained by global data. Using the 2009-2010 temporary deployment, Wei et al. [2017] confirmed the existence of the Tonga DSZ and revealed more details. Our new catalog shows a much clearer DSZ along cross-section B-B' that extends to about 300 km depth. Although hinted by Wei et al. [2017], the new results explicitly suggest that the lower plane of the DSZ is confined between the latitudes of 19°S and 21°S (the cyan box in Figure 3.11a), diminishing to the north and south. A DSZ with a limited extent in Tonga is in agreement with similar observations in Japan [Igarashi et al., 2001] and Alaska [Wei et al., 2021], suggesting that deeper parts of the slab mantle are not uniformly hydrated. We do not observe a deeper DSZ at 350–460 km depths that was interpreted as the edges of a metastable olivine wedge in the slab [Wiens et al., 1993]. This could be because there were not enough earthquakes occurring at these depths during the 1-year deployment to delineate that feature.

The Tonga subduction zone hosts the majority of deep-focus earthquakes in the world, including 4 notable large events: Mw 7.6 on 9 March 1994, Mw 7.3 on 9 November 2009, Mw 8.2 on 19 August 2018, and Mw 7.9 on 6 September 2018. Our results show that the 1994 Mw 7.6 and 2018 Mw 8.2 events occurred at the bottom of a highly seismogenic region (Figure 3.11d), consistent

with the previous suggestion that these events initiated rupturing in the cold slab [McGuire et al., 1997, Fan et al., 2019]. In contrast, the 2018 Mw 7.9 event occurred in a previously aseismic region (Figure 3.11g), possibly rupturing through a dissipative process at the edge of a warm fossil slab [Fan et al., 2019, Jia et al., 2020]. The 2009 Mw 7.3 event occurred at the western end of a seismicity band corresponding to a fossil slab subducted at the now inactive Vitiaz Trench [Cai and Wiens, 2016].

3.7 Conclusions

Our integrated workflow has proven highly effective at detecting and locating deep earthquakes in the Tonga subduction zone. We use PhaseNet-TF to detect *P*- and *S*-wave arrivals in the time-frequency domain, showing superior performance over traditional ways of picking arrivals in the time domain. Detecting arrivals in the time-frequency domain is critical for analyzing OBS data, particularly the horizontal components with much higher noise levels compared with land-based stations. We use GaMMA-1D to associate arrivals, teletomoDD to relocate events, and a bootstrap resampling approach to estimate uncertainties. This workflow effectively removes artificial arrivals and events with poor constraints. Furthermore, through semi-supervised learning, the PhaseNet-TF model improves iteratively, leading to a more comprehensive and accurate earthquake catalog compared to the initial manual picks.

This research opens new avenues for in-depth studies in subduction zones, particularly those with limited local data coverage like Tonga. The new catalog with more events and arrivals potentially benefits future work of high-resolution tomography imaging and earthquake similarity analyses, helping us better understand deep earthquake mechanisms and subduction processes. While our study focuses on Tonga, the methods and workflow are readily adaptable for application in other subduction zones, offering a versatile tool for both real-time monitoring and historical data analyses.

3.8 Figures

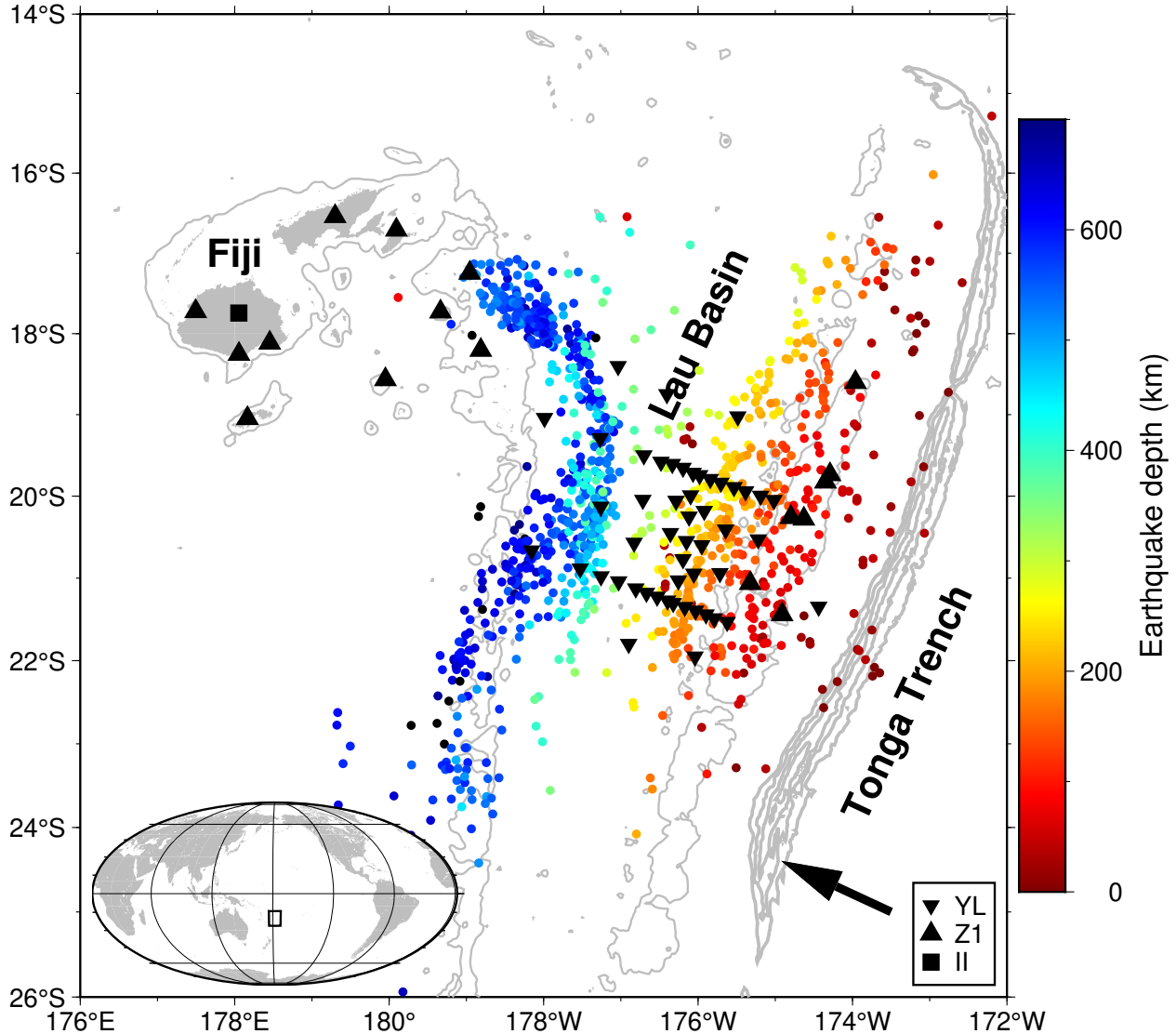


Figure 3.1 Map of the Tonga subduction zone and adjacent regions. Earthquakes in the manually picked reference catalog are shown as dots color-coded by depth. Black triangles, inverted triangles, and squares represent land-based stations, ocean-bottom seismographs, and a GSN station, respectively, deployed from November 2009 to December 2010. Land areas are shaded in grey. The bathymetry contours of 1 km highlight features such as the Tonga Ridge, Lau Ridge, and Fiji Plateau. Additional bathymetry contours at 7, 8, 9, and 10 km delineate the Tonga Trench. The black arrow indicates the Pacific Plate's motion relative to Tonga. Inset provides a global context for the study region.

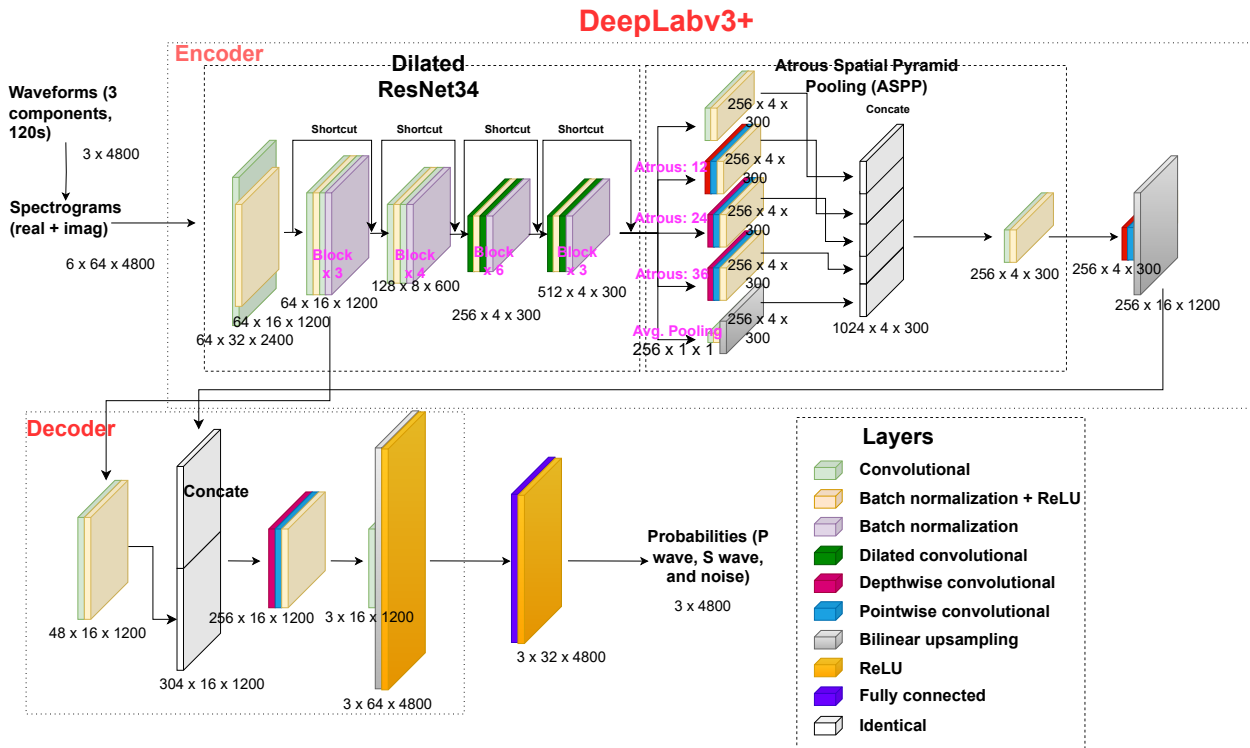


Figure 3.2 PhaseNet-TF architecture. The input consists of 120-second three-component seismograms with a sample rate of 40 Hz, so the input has a dimension of 3×4800 . The output includes three probability time series, corresponding to *P*-wave arrival, *S*-wave arrival, and noise, with the same length as the input. Dimensions for each layer are denoted adjacent to the layer, with the format of "number of channels \times layer height \times layer width". The input seismograms are first transformed into spectrograms before being processed through the DeepLabv3+ network. DeepLabv3+ includes an encoder and a decoder and is equipped with skip layers. The decoder features a dilated ResNet34 and Atrous Spatial Pyramid Pooling (ASPP). The output of DeepLabv3+ is subsequently refined through a multilayer perceptron.

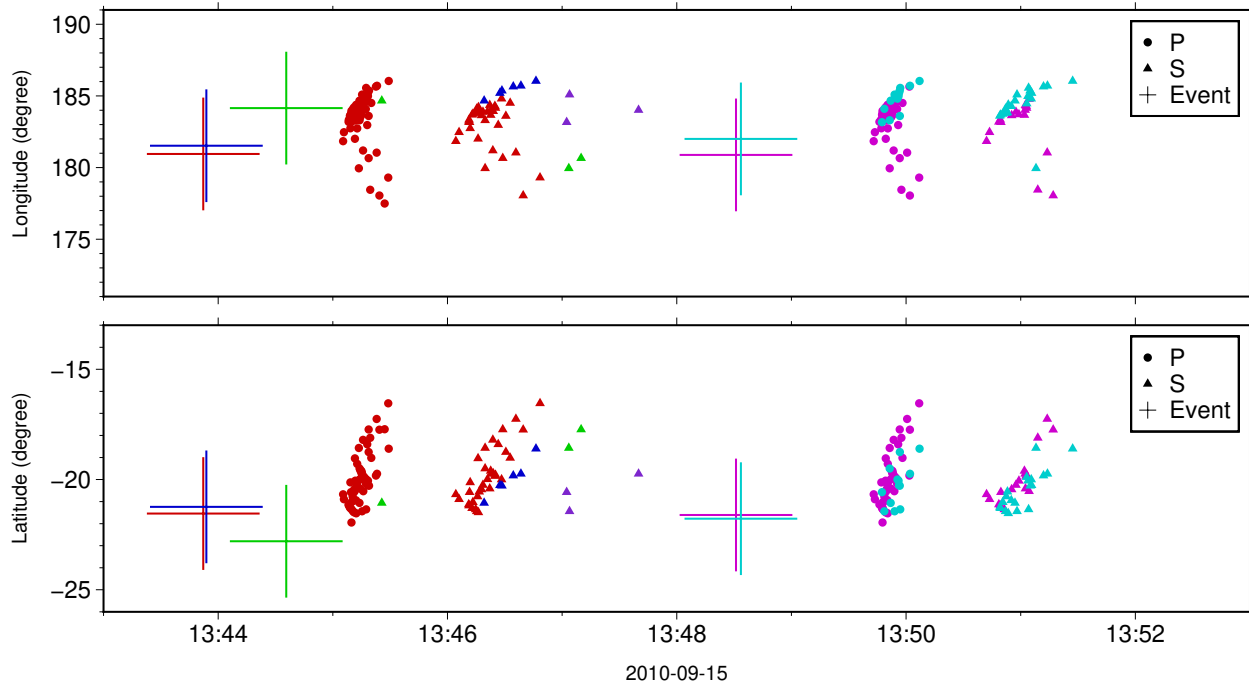


Figure 3.3 Examples of GaMMA-1D association from UTC 13:43 to 13:53 on September 15, 2010. The top panel shows associated arrivals with respect to longitude, whereas the bottom panel shows them against latitude. In both panels, individual colors denote distinct associated events. Dots and triangles indicate *P*- and *S*-wave arrivals, respectively, while crosses mark the origin time and locations of the associated events. Events associated with less than 10 phase arrivals (e.g., the green-colored event and arrivals) will be discarded in the following step.

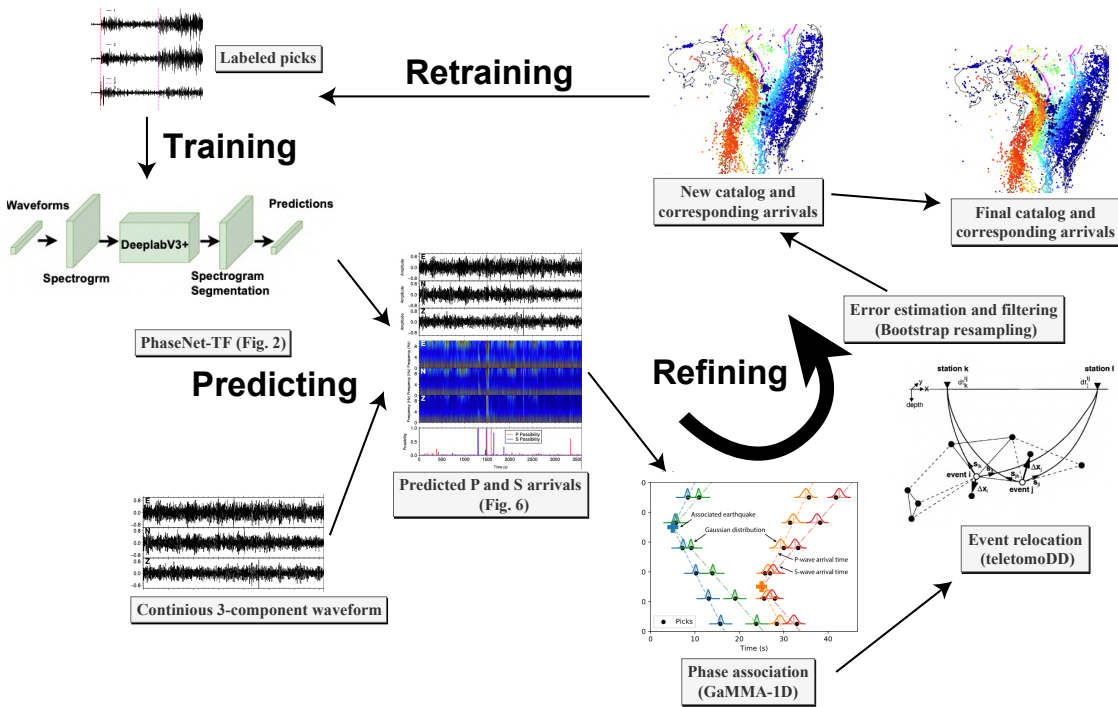


Figure 3.4 Schematic of our semi-supervised learning workflow. The process begins with training the PhaseNet-TF model using manually picked arrivals. The trained model is then applied to continuous seismic data to obtain *P*- and *S*-wave arrival probabilities. The following phase association by GaMMA-1D, event relocation by teletomoDD, and error estimation/filtering with a bootstrap resampling approach together produce a refined event catalog and associated phase arrivals. This updated catalog serves as a new labeled dataset for training and validating PhaseNet-TF in the next iteration. The entire workflow is iteratively repeated until no significant improvements are observed, culminating in the final event catalog and corresponding phase arrivals.

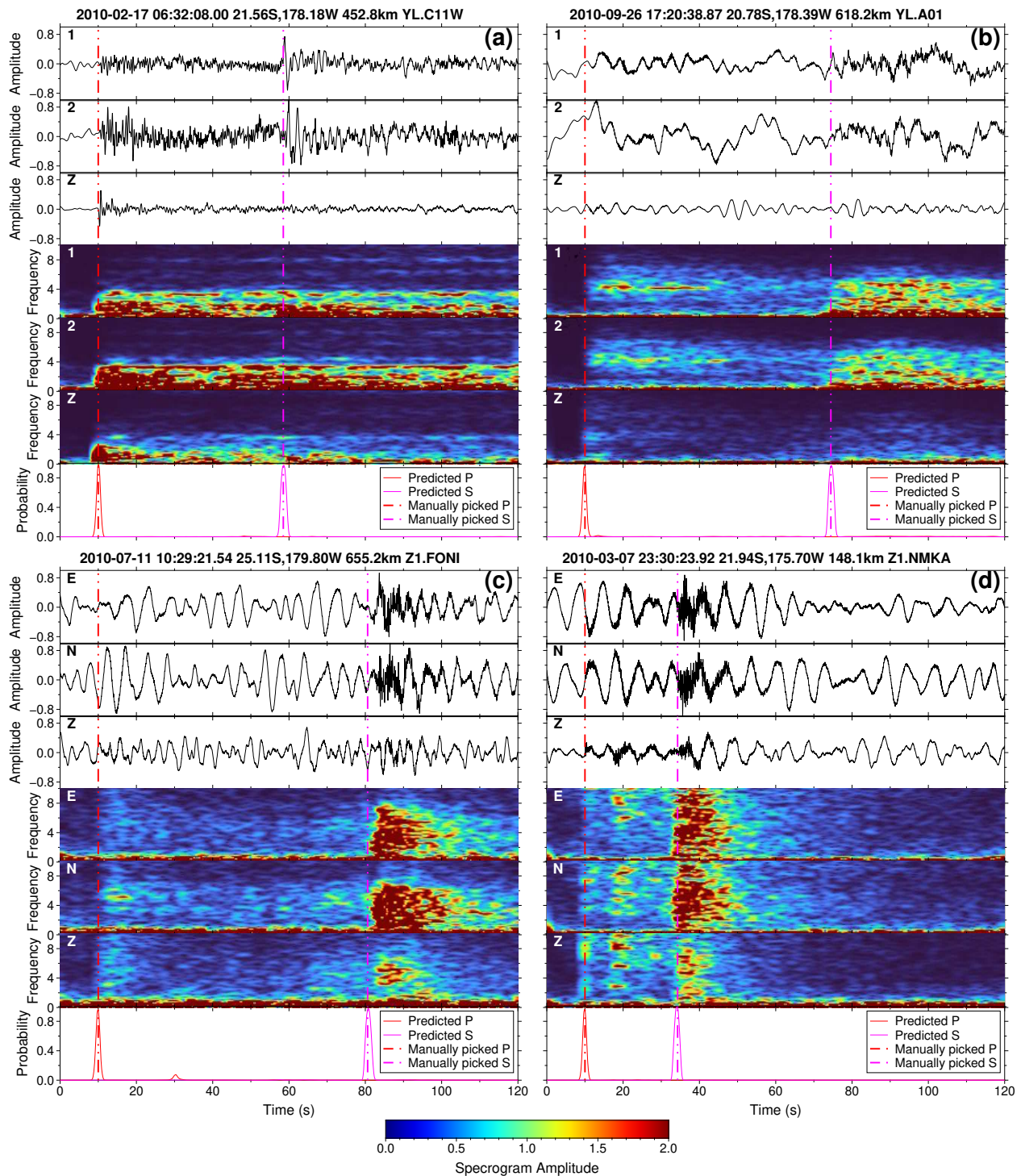


Figure 3.5 Examples of PhaseNet-TF prediction on the test dataset. (a) and (b) show deep earthquakes recorded at OBSs, whereas (c) and (d) show events recorded at land-based stations. In each panel, the title includes the origin time, hypocenter, and station name. The top 3 sections display the three-component waveforms in the time domain, followed by 3 sections of spectrograms (power value) of the three components. The bottom section shows the probabilities of predicted P - and S -wave arrivals. Manually picked P - and S -wave arrivals are indicated by dashed lines.

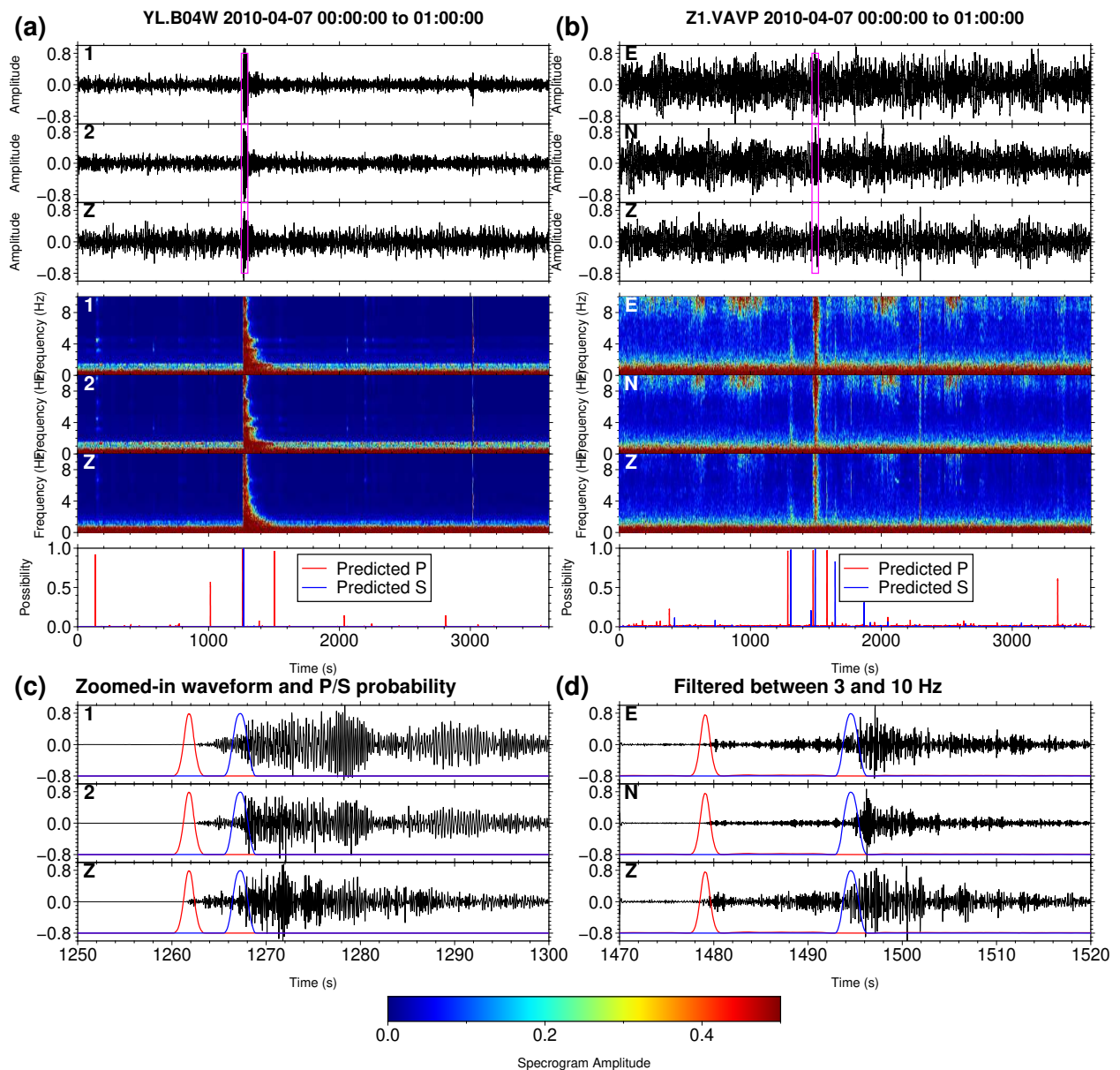


Figure 3.6 Examples of PhaseNet-TF prediction on continuous data from UTC 00:00 to 01:00 on 7 April 2010 on an OBS B04W (a) and a land-based station VAVP (b). (a) and (b) are similar to Figure 3.5, except that there are no manual picks. (c) Zoomed-in waveform filtered at 3-10 Hz and P/S probability within the time window indicated by the magenta box in (a). (d) Zoomed-in waveform filtered at 3-10 Hz and P/S probability within the time window indicated by the magenta box in (b). The P and S arrivals are confirmed to be true by the following steps of phase association and event relocation.

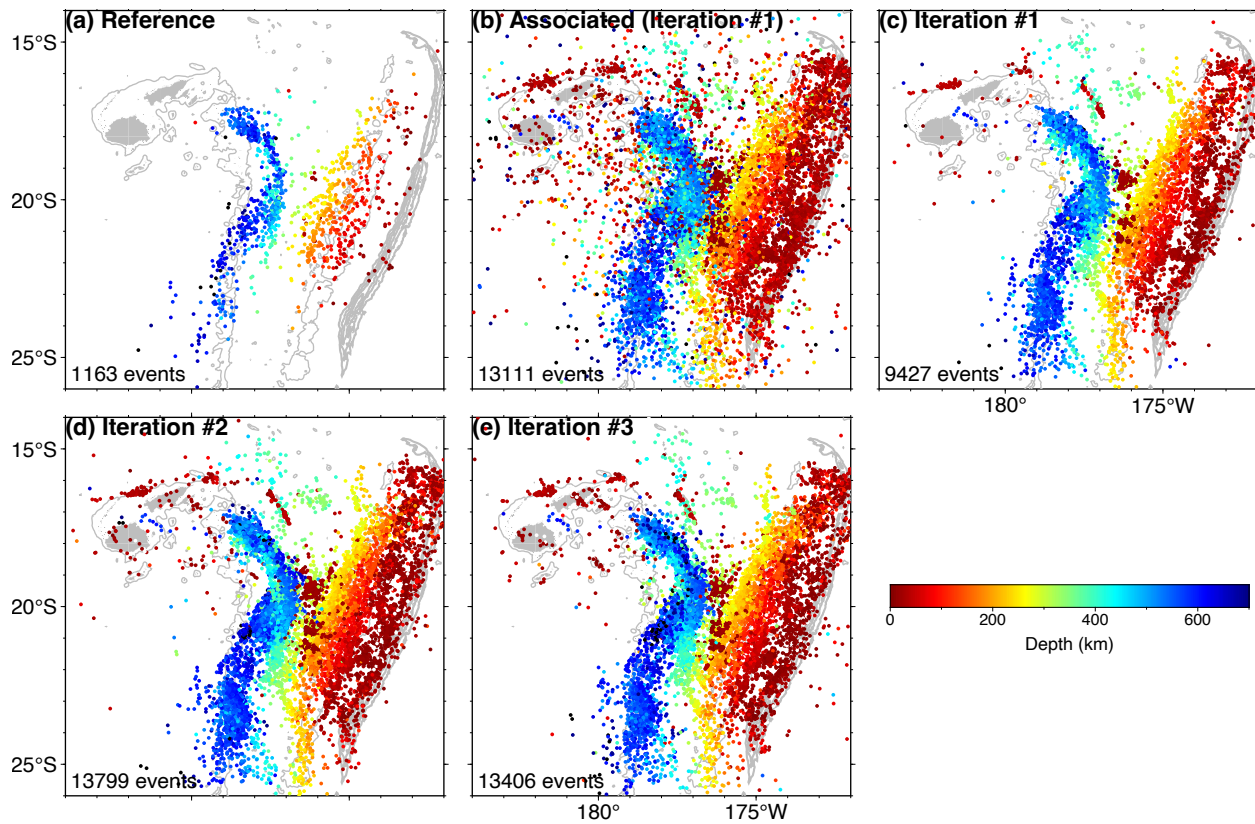


Figure 3.7 Maps of event distributions at various stages of the analysis. (a) The manually picked reference catalog by Wei et al. [2017]. (b) Events predicted by PhaseNet-TF and associated by GaMMA-1D in Iteration #1. (c) Relocated events serve as the output catalog of Iteration #1. (d) Output catalogue of Iteration #2. (e) Output catalog of Iteration #3, considered the final catalog. All panels are plotted similarly to Figure 3.1. The number of events for each step is shown in the bottom-left corner of each panel.

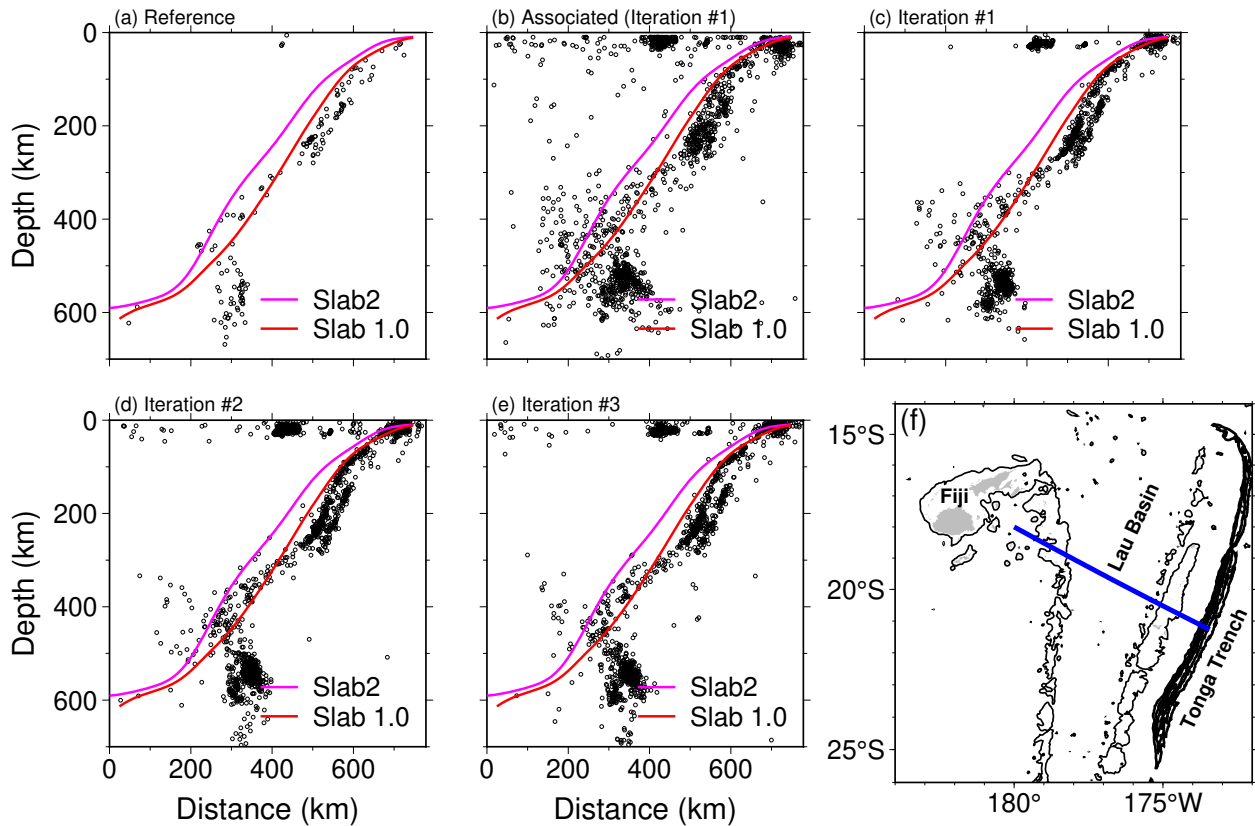


Figure 3.8 Cross-sections of event distributions at various stages of the analysis. (a) The manually picked reference catalog by Wei et al. [2017]. (b) Events predicted by PhaseNet-TF and associated by GaMMA-1D in Iteration #1. (c) Relocated events serve as the output catalog of Iteration #1. (d) Output catalogue of Iteration #2. (e) Output catalog of Iteration #3, considered the final catalog. (f) The location of the cross-section (blue line). In each of (a) – (e), black circles indicate earthquakes within 70 km away from the cross-section. Red and magenta curves show the slab upper interface according to the Slab1.0 [Hayes et al., 2012] and Slab2 [Hayes et al., 2018] models, respectively. Neither of these models is accurate in the Tonga subduction zone.

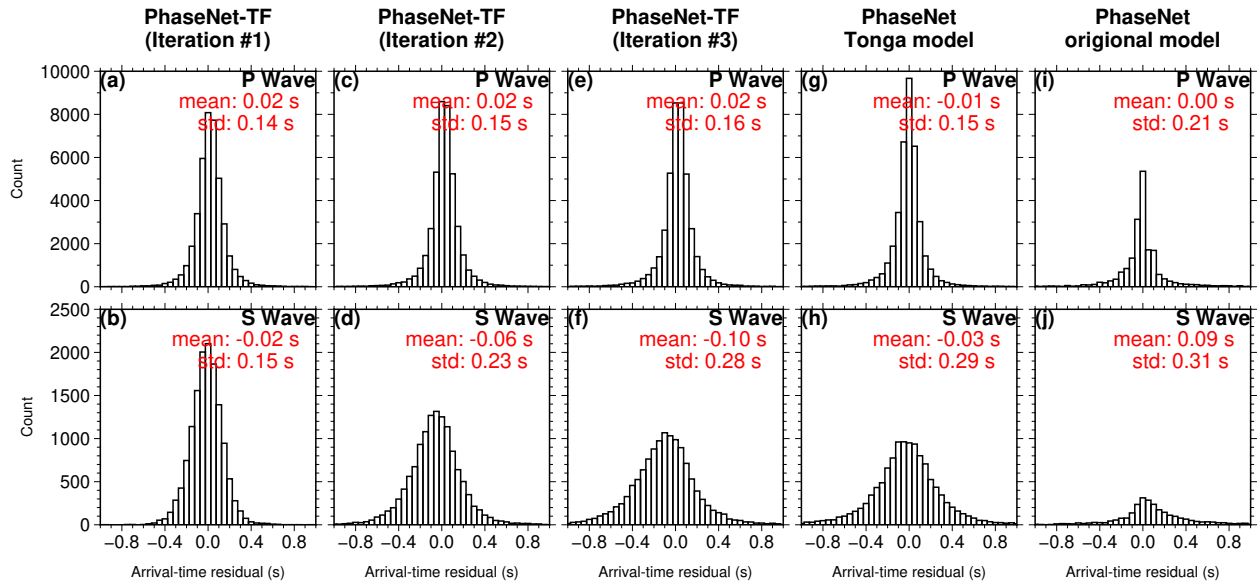


Figure 3.9 Arrival-time residuals of P (top) and S (bottom) waves across different models and iterations. When each model is applied to the continuous dataset, the predicted arrivals are compared to manual picks when they exist, and the arrival-time differences contribute to the histogram. (a-f) Results of PhaseNet-TF from three consecutive iterations. (g-h) Outcomes of PhaseNet retrained with the Tonga dataset. (i-j) Results of PhaseNet with its original model weights trained with data from Northern California.

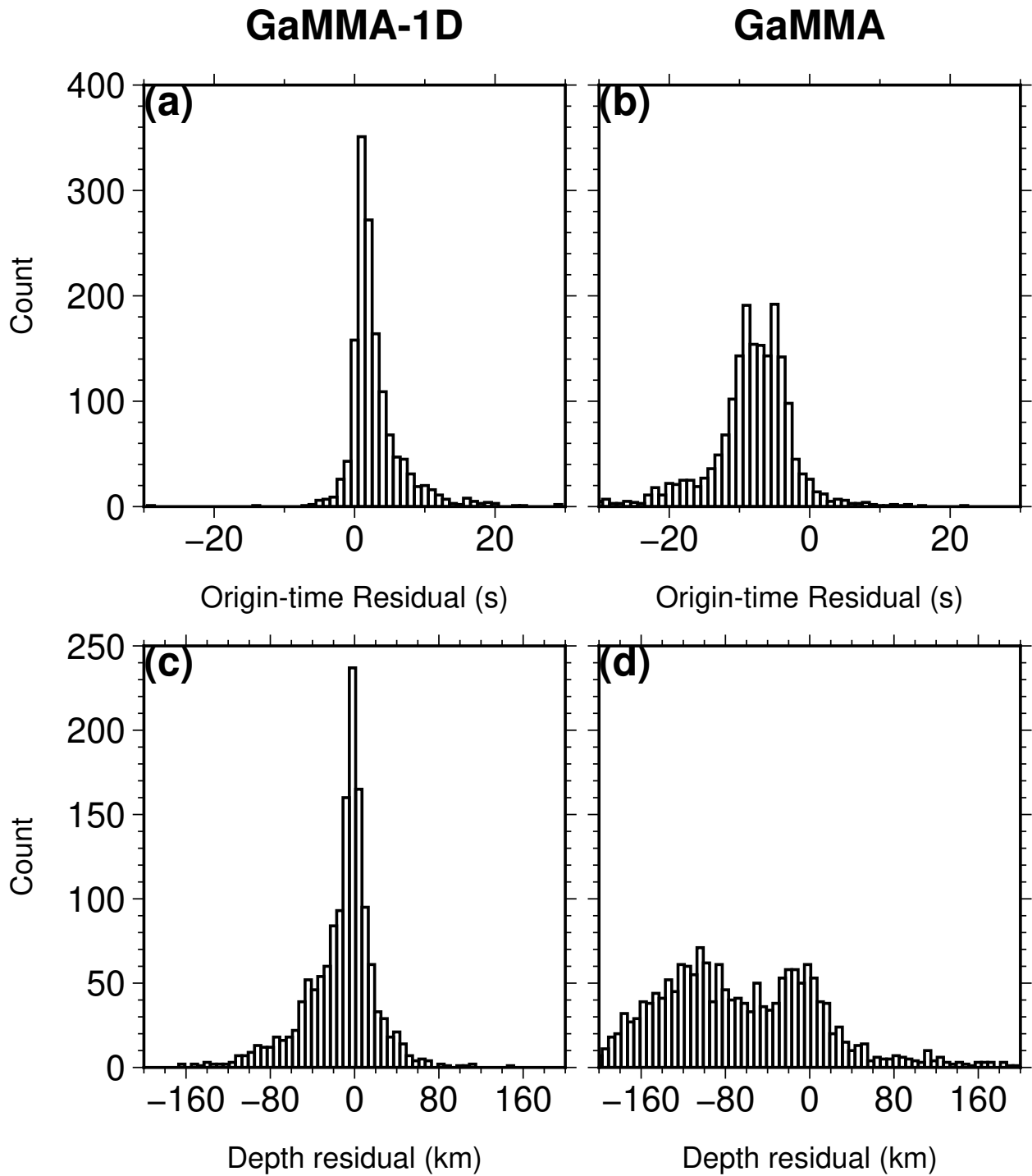


Figure 3.10 Origin-time (top) and depth (bottom) residuals for GaMMA-1D and GaMMA. When manual picks are associated by GaMMA-1D or GaMMA, the event origin times and depths are compared against the reference catalog that was associated by the Antelope software [Wei et al., 2017].

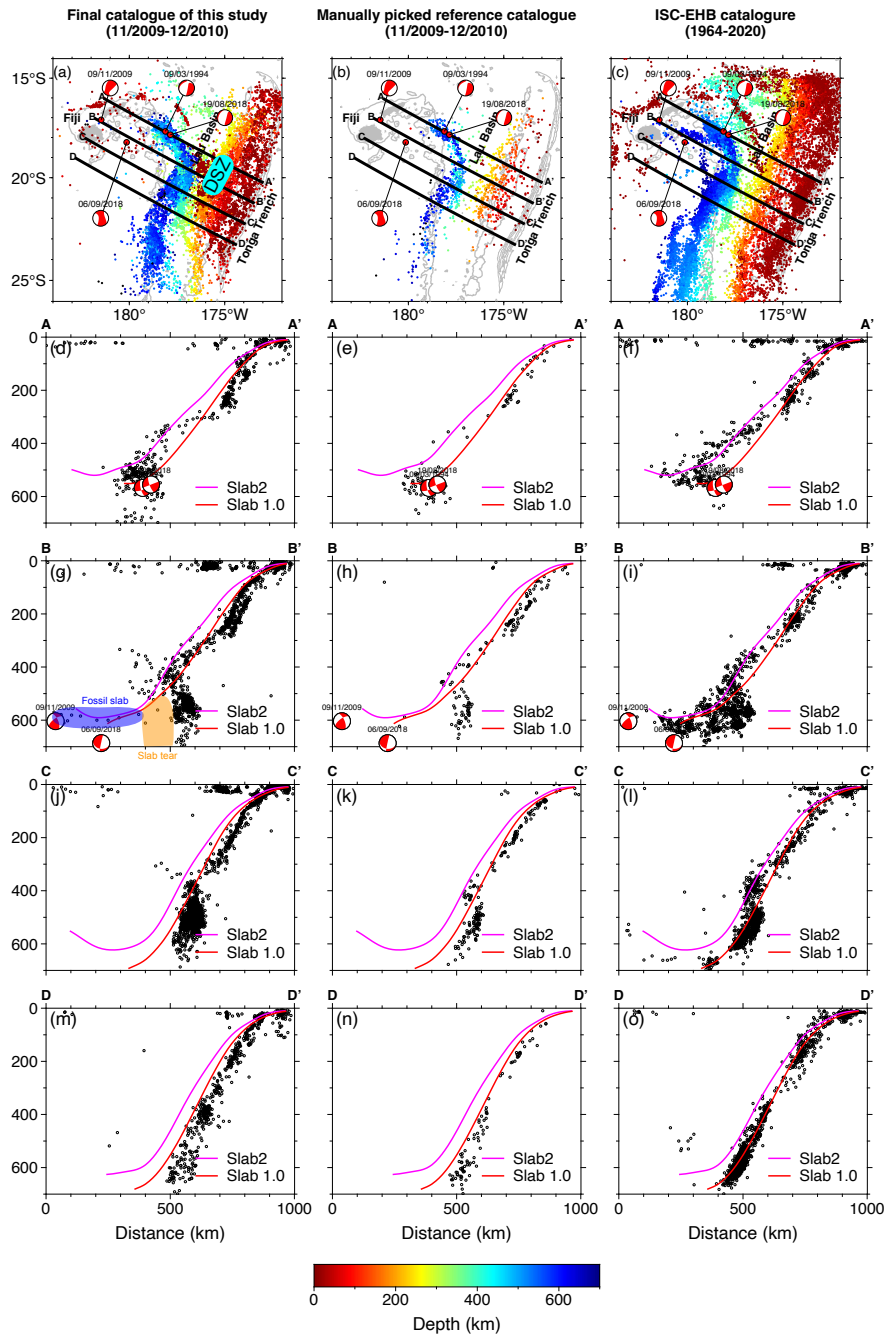


Figure 3.11 Event distributions across 3 distinct catalogs. (Left column) Our final catalog in this study. (Middle column) Reference catalog by Wei et al. [2017]. (Right column) ISC-EHB catalogue [International-Seismological-Centre, 2023]. In each column, the top panel shows the map view, similar to Figure 3.7, and the following panels show 4 cross-sections, similar to Figure 3.8. Beachballs show the focal mechanisms of 4 notable large deep earthquakes (1994/3/9 Mw 7.6, 2009/11/9 Mw 7.3, 2018/8/19 Mw 8.2, and 2018/9/6 Mw 7.9) from the Global CMT catalog [Dziewonski et al., 1981, Ekström et al., 2012].

3.9 Tables

Table 3.1 Numbers of picks and events in the manually picked reference catalogue and output catalogues. Recalls are evaluated against the reference catalogue. The picks with arrival-time residuals < 1 s compared to the reference catalogue are counted as true positive picks, whereas the events with origin-time residuals < 15 s are considered true positive events. The output catalogue of Iteration #3 is considered the final catalogue. The low recall values of S -wave arrivals result from the filtering processes of phase association and relocation, as many S -wave arrivals are discarded.

	P -wave arrival		S -wave arrival		Event	
	Recall	Number	Recall	Number	Recall	Number
Reference	N/A	42,256	N/A	14,852	N/A	1,163
Iteration #1 output	0.94	217,254	0.83	63,590	0.96	9,427
Iteration #2 output	0.94	343,247	0.80	79,593	0.96	13,799
Iteration #3 output (Final)	0.89	372,774	0.75	78,853	0.91	13,406

Table 3.2 Evaluation metrics of phase pickers on the test dataset for different architectures and models. Picks with arrival-time residuals < 1 s compared to the manually picked reference catalogue are counted as true positive picks. The metrics in the table are evaluated on the test dataset partitioned from the labelled dataset, which is the manually picked reference catalogue for PhaseNet-TF Iteration #1 and two PhaseNet models, and the output catalogue from the previous iteration for PhaseNet-TF Iterations #2 and #3.

	P -wave arrival			S -wave arrival		
	Precision	Recall	F1	Precision	Recall	F1
PhaseNet-TF (Iteration #1)	0.99	0.99	0.99	0.97	0.99	0.98
PhaseNet-TF (Iteration #2)	0.99	0.99	0.99	0.96	0.99	0.98
PhaseNet-TF (Iteration #3)	0.99	0.98	0.99	0.96	0.99	0.98
PhaseNet retrained by the Tonga dataset	0.97	0.97	0.97	0.85	0.94	0.90
PhaseNet with original model weights (trained with Northern California data)	0.89	0.66	0.76	0.48	0.28	0.36

Table 3.3 Evaluation of association and relocation filtering on the manually picked reference catalogue (Iteration #1). Events with origin time residuals < 15 s are counted as true positive events, and phase arrival-time residuals < 1 s are counted as true positive picks.

	<i>P</i> -wave arrival	<i>S</i> -wave arrival	Event		
	Recall	Recall	Precision	Recall	F1
Association by GaMMA-1D	0.97	0.92	0.97	0.98	0.97
Relocation using GaMMA-1D output and filtering (epicentre uncertainty $< 0.1^\circ$, depth uncertainty < 10 km, and origin-time uncertainty < 1 s)	0.95	0.86	0.98	0.95	0.96
Association by GaMMA	0.97	0.91	0.76	0.96	0.85

CHAPTER 4

DETECTING CONVERTED SEISMIC PHASES OF TONGA DEEP EARTHQUAKES: IN-DEPTH INSIGHTS FROM DEEP-LEARNING METHODS

An edited version of this chapter is in preparation for submission to Geophysics Journal International.

4.1 Abstract

Converted seismic phases have been instrumental in advancing structural seismology. Recently, the collection of seismic data in the Tonga subduction zone offers a good opportunity to explore the Tonga slab. Among this data are numerous P-to-S (PS) waves from deep earthquakes converted at the slab top interface, which are critical to understanding the geometry and internal structure of the subducted slab. However, identifying PS phases in a large seismic dataset is challenging, primarily due to the inconsistent distribution of energy in the frequency domain, varying picking criteria by human analysts, and non-unique interpretations of the signals. To address these challenges, we adopt a deep-learning-based package, PhaseNet-TF, specifically to detect PS-wave arrivals in a dataset collected from an amphibious seismic array in the Tonga-Fiji area. We further use filters of locations, travel-time, and energy and beamforming analyses to quality control the predicted arrivals. With a semi-supervised learning strategy, we iteratively refine the arrival labels and retrain PhaseNet-TF. This approach has significantly increased the number of PS arrival measurements, which will be used to image the Tonga slab in detail. Our study demonstrates that deep learning offers a promising avenue for improving the detection of converted phases in structural seismology.

4.2 Introduction

Detecting converted seismic waves is essential for understanding the complexities of slab structures and the dynamics of subduction processes. The analysis of P-slab-S and S-slab-P waves is particularly crucial, as these waves provide insights distinct from those offered by direct P and S waves, which primarily map the crust and mantle's structure. Converted waves are sensitive to variations in seismic wave speed within the slab and at its interface, offering critical information for subduction studies. Research in northeastern Japan, the eastern Aleutian, and the Nazca sub-

duction zone has emphasized the role of PS waves in providing detailed information on the slab's upper interface and identifying low-velocity zones within the slab [Matsuzawa et al., 1986, 1990, Zhao et al., 1997a, Shiina et al., 2013, 2014, 2017, Helffrich and Abers, 1997, Bock et al., 2000]. Converted SP waves have been studied predominantly in northeastern Japan, influenced by the origin of these waves in the lower plane of the double seismic zone [Ohmi and Hori, 2000]. Very few such data are available from other subduction zones.

Recent advancements in applying deep learning techniques for seismic wave detection have significantly enhanced the ability to identify seismic events without compromising accuracy. The evolution from early applications of simple neural networks for phase detection [Gentili and Michelini, 2006, Ross et al., 2018] to the adoption of U-Net [Ronneberger et al., 2015]-inspired deep neural networks like PhaseNet [Zhu and Beroza, 2019] and transformer [Vaswani et al., 2017]-inspired architectures such as EQTransformer [Mousavi et al., 2020] demonstrates deep learning's reliability and efficiency in seismic wave detection. The development of PhaseNet-TF [Xi et al., 2023], which integrates time-frequency domain information and semantic image segmentation techniques to identify noisy seismic signals from ocean bottom seismometers, has motivated a focus on detecting weak signals alongside the traditionally more emphasized strong seismic signals.

However, the detection of converted seismic phases with current deep learning methods faces challenges, including limited labeled datasets and the presence of weak and noisy signals. One significant challenge is that PS waves are not confined to fixed frequency bands; their behavior is greatly influenced by source parameters and wave propagation effects, such as attenuation and scattering. As a result, PS wave signals often mix with noise, complicating their identification for phase picking. The concept of applying time-frequency domain analysis, akin to methods used in models like PhaseNet-TF, emerges as a promising approach for detecting PS waves. Additionally, leveraging array analysis techniques, such as beam-forming [Gong and McGuire, 2021], could also help refine phase-picking results derived from deep learning models.

The Tonga subduction zone offers a unique opportunity to study converted seismic waves. Being the coldest and fastest subducting slab on Earth, it serves as an exceptional case for examining the

mechanisms behind intermediate-depth earthquakes. These earthquakes are widely believed to be associated with slab dehydration, although the precise mechanisms remain a subject of debate. Given the cold and rapid subduction of the Tonga slab, it's highly plausible that hydrous minerals carried by the subduction could reach greater depths, potentially extending into the mantle transition zone (MTZ). This scenario could significantly revise our understanding of Earth's deep water cycle and our estimates of the global water budget. Previously, the Tonga slab was not extensively imaged due to limited data availability. The Lau Basin Ocean Bottom Seismometer (OBS) Experiment, which operated from 2009 to 2010, has provided valuable data, complemented by years of manual phase picking by human experts and a recent catalog developed using deep learning methods [Xi et al., 2023]. However, the availability of converted phases through manual picking has been notably lacking.

Detected converted seismic waves in the Tonga subduction zone can accurately pinpoint the slab interface location and provide direct evidence of the low-velocity zone (LVZ) which can support the presence of a hydrous slab crust and uppermost mantle. Since the lower plane of the Tonga double seismic zone is less active compared to the upper plane, which experiences compressional stress across the slab [Isacks and Molnar, 1971], PS waves are particularly relevant for studies in this area. Previous estimates of the slab's upper interface in the Tonga subduction zone, such as Slab1.0 [Hayes et al., 2012] and Slab2 [Hayes et al., 2018], were primarily derived from seismicity data. Since intermediate-depth earthquakes typically occur within the slab crust, these estimates often suggest a deeper interface than its actual depth. Inverting PS waves could more precisely constrain converting points on the slab interface, with potential errors less than 1 km [Matsuzawa et al., 1986, Zhao et al., 1997a]. Previous tomographic imaging of the Tonga slab using direct waves [van der Hilst, 1995, Zhao et al., 1997b, Conder and Wiens, 2006, Wei et al., 2015, 2016, Adam et al., 2016] lacked resolution on the slab crust. Incorporating PS waves promises improvements in imaging this structure.

To accurately and efficiently detect PS waves in the Tonga subduction zone, new methods must be developed. First, the phase picker should have the capability to identify weak PS wave signals

with high precision and recall, even in the presence of noise commonly found in OBS data. Second, implementing a robust quality control process is essential to minimize the risk of false phase detection. This process should ideally incorporate techniques from seismic array analysis, such as beamforming, to enhance detection accuracy. Lastly, considering the scarcity of labeled data, it is imperative to employ a semi-supervised learning approach that leverages both labeled and unlabeled datasets. This approach has already shown efficiency in detecting direct phases in the Tonga subduction zone [Xi et al., 2023] and in picking seismic arrival times on distributed acoustic sensing (DAS) data [Zhu et al., 2023a].

4.3 Data

In this study, we analyzed seismic data recorded by an amphibious array in Tonga and Fiji from November 2009 to December 2010 (Figure 4.1b). This array included 49 OBSs (network ID YL), 17 island-based stations (network ID Z1), and one Global Seismograph Network (GSN) station MSVF (network ID II). 1,163 local earthquakes with 42,256 P-wave and 14,852 S-wave arrivals were manually picked by Wei et al. [2017]. In addition, 5,127 PS-wave arrivals, usually 10 s after the corresponding P-wave arrivals, were manually picked in the time-frequency domain. These manually picked events and arrivals are hereinafter called the reference dataset. In our previous study [Xi et al., 2023], we developed a semi-supervised deep-learning workflow to detect and locate 13,406 earthquakes with 372,774 P-wave and 78,853 S-wave arrivals from this dataset. In this study, we further extend the deep-learning workflow to include the manually picked PS-wave arrivals and focus on detecting PS waves from events that were previously located.

Similar to Xi et al. [2023] we created the training dataset by windowing three-component waveforms 5 minutes before and 5 minutes after each P-wave arrival. We also subsequently removed the instrumental response to get displacement waveforms of the three components and further resampled the data to 40 HZ with no additional processing. We also created a 1-year continuous dataset by segmenting three-component waveforms into 10-day segments and subsequently removing the instrumental response. All 3 channels were included, and any missing data from November 2009 to December 2010 were filled with zeros to ensure the continuity of the dataset.

4.4 Methods

4.4.1 Phase picking by PhaseNet-TF

In this study, we use PhaseNet-TF, a deep-learning-based phase picker developed by Xi et al. [2023], to detect PS waves in the time-frequency domain. PhaseNet-TF is modeled after the architecture of DeepLabv3+ [Chen et al., 2018], designed to handle data within the time-frequency domain efficiently. The performance of PhaseNet-TF is comparable to that of PhaseNet [Zhu and Beroza, 2019] for detecting P waves and showing superior efficacy in identifying S waves with OBS data [Xi et al., 2023]. Originally, PhaseNet-TF processed three-component waveform inputs to predict the probability of phase arrivals for P and S waves, along with noise probability. Here we introduce an additional output channel dedicated to the probability of PS wave arrivals, while the rest of the model architecture remains unchanged. This modification allows PhaseNet-TF to maintain its accuracy in predicting P and S wave arrivals, as established in the previous study while adding the capability of accurately predicting PS wave arrivals.

In refining PhaseNet-TF, we retain the original model’s loss function, optimizer, and learning rate scheduler but modify the weighting strategy for loss functions to account for PS waves. Given the scarcity of labeled PS wave arrivals compared to P- and S-wave arrivals, addressing the imbalance in the dataset is crucial. We implement a strategy of differential weighting for our loss functions, specifically using the Kullback-Leibler (KL) divergence for distinct categories:

$$L = \frac{\sum_i W_i \sum_x y_{true,i}(x) * \log \frac{y_{true,i}(x)}{y_{predict,i}(x)}}{\sum_i W_i} \quad (4.1)$$

where x denotes the summation performed across all points, and i represents one of the four categories utilized in constructing the loss function: P-wave arrival probability, S-wave arrival probability, PS-wave arrival probability, and noise probability. The sum of these probabilities at any given timestamp equals one. The weights for these categories W_i are set as 1:1:3:3 for noise, P, S, and PS probabilities, respectively. $y_{predict,i}(x)$ is the model prediction for the i -th category, and $y_{true,i}(x)$ corresponds to the reference.

PhaseNet-TF is trained by the manually picked reference dataset with 42,256 P-wave, 14,852

S-wave, and 5,127 PS-wave arrivals. This trained PhaseNet-TF model is then applied to the continuous dataset to predict PS waves. Following Xi et al. [2023], we partition the continuous waveforms into 120-second segments with a 60-second overlap between segments. The output segments of PS-wave probabilities are combined into a single continuous time series by selecting the maximum value from overlapping predictions. Since PS signals are typically weaker than P and S signals, peak probabilities exceeding 0.3 are considered positive picks for PS waves, lower than the threshold of 0.5 for P and S waves. We call these PS-wave predications X arrivals before confirming they are the PS waves converted at the slab top interface. We leverage the pre-established earthquake catalog from Xi et al. [2023] and associate each X arrival with a P-wave arrival immediately preceding it, and thus with the corresponding earthquake in the earthquake catalog.

These X arrivals inevitably include false PS arrivals because (1) the probability threshold of 0.3 is low and (2) the DeepLabv3+ architecture is mainly focused on signal processing and does not carry much seismological knowledge. Consequently, the predictions may represent some seismic signals other than the PS waves converted at the slab-top interface. As described in the following subsections 4.4.2 and 4.4.3, we apply several quality control processes based on seismological knowledge to filter X arrivals and only keep P-slab-S arrivals. These processes will eliminate most false positive PS detections but cannot address false negative predictions. Since the scientific motivation of this study is to obtain reliable PS arrivals for seismic imaging, this approach will lead to a robust, although not exhaustive, dataset of PS waves.

4.4.2 Quality control (QC): source depth, station location, travel time, and energy

We first examine the properties of individual X arrivals. According to Fig. 4.1a, the PS wave source should be intra-slab intermediately-depth or deep-focus earthquakes. Intra-slab earthquakes shallower than 50 km may also generate P-slab-S waves converted at the megathrust. However, it is difficult to distinguish P-slab-S waves from P-moho-S waves converted at the overriding plate Moho from these shallow intra-slab earthquakes. In addition, these P-slab-S arrivals, if successfully picked, can help constrain the megathrust at <50 km depth, which is usually well constrained by active-source or surface-wave studies, thus are less critical than the P-slab-S arrivals from deep

earthquakes. Therefore, we discard all X predictions from earthquakes shallower than 50 km. Given the ray geometry, we also require the station to be between the earthquake and the Tonga Trench, discarding all X predictions recorded at Fijian stations.

Travel-time is an effective measurement to filter X arrivals [Matsuzawa et al., 1986]. Given the earthquake catalog from Xi et al. [2023], we first require each X arrival to be between the P- and S-wave arrivals of the same event. If an event has no S-wave arrival detected, we estimate the S-wave arrival based on a linear regression of all S-P differential travel-times versus hypocentral distances (Fig. 4.3). X predictions that miss this requirement are discarded. Second, we exclude any predicted X arrivals occurring within 5 seconds after the corresponding P-wave arrivals. This step helps us eliminate P-moho-S waves converted at the overriding plate Moho (Fig. 4.1a) because the crustal thickness in this region is only about 6–36 km from the Lau back-arc basin to the Tonga fore-arc [Wei et al., 2016, Chen et al., 2019a]. This step can also eliminate S-slab-P waves converted at the slab top interface, which usually arrive less than 3 seconds after the corresponding P waves.

In addition, we analyze the energy of X arrivals on 3 components to exclude any signals that arrive as P waves. Given the shear wave nature of PS waves, the PS-wave arrivals should have the strongest energy in the radial direction. Although the PS-wave energy may leak to the vertical component due to the varying incident angle and complex site effects, we still expect higher energy on horizontal components. Therefore, we discard X arrivals with the energy on the vertical component higher than that on the horizontal components.

4.4.3 Quality control (QC): Source-side beamforming analysis

While straightforward quality control measures based on location, travel-time, and energy information successfully eliminate a significant number of falsely detected PS arrivals, they fall short in filtering out phases that share similar arrival times and receiver-side polarity with the PS wave, such as the SPS wave, along with other falsely detected signals not corresponding to any meaningful phases. In addressing this limitation, we use a source-side beam-forming method to investigate the nature of X arrivals near earthquakes. Specifically, we estimate the wave speed, azimuthal angle, and take-off angle of X arrivals derived from source-side beamforming analyses and compare them

with the theoretical values of PS waves.

In contrast to previous 3D source-side beamforming studies [Spudich and Bostwick, 1987, Nakata and Shelly, 2018, Gong and McGuire, 2021], we do not stack waveforms for several reasons: (1) Unlike station-side beamforming, the waveforms in source-side beamforming are heavily influenced by the event source mechanism, and directly stacking waveforms may not effectively amplify the signal. (2) The X signals are noisy in the time domain, even if we filter the waveforms within the optimal frequency band. Instead, we stack delta functions that represent the X arrivals because only the travel-time information is needed for our purpose.

We follow Gong and McGuire [2021] to perform the source-side beamforming analysis. In applying this method to real data, waveform segments are aligned with the direct P wave, respectively. The synthetic delta function-shaped pulse at the phase arrival time predicted by the PhaseNet-TF model is then applied to represent the PS wave arrivals. The slowness vector, expressed in spherical coordinates, depends on the takeoff angle, azimuth angle, and near-source velocity, and can be obtained through a grid search to maximize the PS wave signal stacking energy. For a single experiment of the source-side beamforming analysis, we have a group of X arrivals as the input and the slowness vector as the output. Meanwhile, the deduced slowness vector should match the tomographic data, the earthquake-station geometry, and ray tracing results, from where we have the theoretical or expected slowness vector. Our goal is to discard those unreliable X arrivals and obtain trustable PS arrivals as the output of this stage.

The synthetic delta function-shaped pulse is defined as:

$$D(t) = e^{-\frac{(t-t_0)^2}{2(\frac{w}{6})^2}} \quad (4.2)$$

where $D(t)$ represents the displacement at time t , t_0 denotes the PS arrival time, and w is a parameter controlling the pulse's width, set to 1 second in our study.

We use a Markov chain Monte Carlo (MCMC) approach to evaluate the trustworthiness of individual X arrivals based on the collective assessment from the source-side beamforming analysis. The trustworthiness of the i -th X arrival is considered a random variable R_i with its expectation $E(R_i) = t_i$ which serves as a measure to determine the retention or discard of this X arrival. If

we denote D_j as the beamforming results of the j -th subset, the Bayesian framework allows us to express the probability of the trustworthiness of all X arrivals as

$$P(R_1, R_2, \dots, R_m | D_1, D_2, \dots, D_n) = \frac{P(D_1, D_2, \dots, D_n | R_1, R_2, \dots, R_m) P(R_1, R_2, \dots, R_m)}{P(D_1, D_2, \dots, D_n)} \quad (4.3)$$

To obtain $P(R_1, R_2, \dots, R_m | D_1, D_2, \dots, D_n)$, we aim to derive the best estimation of the distribution for R_1, R_2, \dots, R_m based on the current information provided by D_1, D_2, \dots, D_n , known as the posterior distribution. Here, m denotes the total number of X arrivals, while n indicates the total number of beamforming experiments conducted. $P(D_1, D_2, \dots, D_n | R_1, R_2, \dots, R_m)$ represents the likelihood. The prior distribution $P(R_1, R_2, \dots, R_m)$ reflects our initial assumptions about the distribution of R_1, R_2, \dots, R_m before beamforming. We adopt a set of independent and identically distributed (iid) random variables, modeled by a Beta distribution characterized by parameters $\alpha = 5$ and $\beta = 1$. This prior distribution aims to balance our initial belief about the trustworthiness of X arrivals with the adjustability based on the beamforming outcomes.

Acknowledging the conditional independence of D_1, D_2, \dots, D_n given R_1, R_2, \dots, R_m , the likelihood function can be represented as a product of individual likelihoods for each beamforming result:

$$P(D_1, D_2, \dots, D_n | R_1, R_2, \dots, R_m) = \prod_{j \leq n} P(D_j | R_1, R_2, \dots, R_m) \quad (4.4)$$

For each beamforming result D_j , its likelihood given the trustworthiness of X arrivals is

$$\begin{aligned} P(D_j | R_1, R_2, \dots, R_m) &= P(D_j | \text{all } R_i, i \in I_j) \\ &= \frac{1}{\text{Length}(I_j)} \sum_{i \in I_j} R_i L(V_j; V_{\text{theoretical}}, \sigma_{V, \text{good}}) L(\theta_j; \theta_{\text{theoretical}}, \sigma_{\theta, \text{good}}) \\ &\quad + (1 - R_i) L(V_j; V_{\text{theoretical}}, \sigma_{V, \text{bad}}) L(\theta_j; \theta_{\text{theoretical}}, \sigma_{\theta, \text{bad}}) \end{aligned} \quad (4.5)$$

where I_j denotes the set of all X arrivals utilized in obtaining D_j . The likelihood functions $L(V_j; V_{\text{theoretical}}, \sigma_{V, \text{good}})$ and $L(V_j; V_{\text{theoretical}}, \sigma_{V, \text{bad}})$ reflect the probability of observing the beamforming wave speed V_j under the assumptions of "good" and "bad" PS wave pickings, respectively, modeled by normal distributions with theoretical wave speed $V_{\text{theoretical}}$ from the IASP91

Earth model [Kennett and Engdahl, 1991] and standard deviations $\sigma_{V,good} = 3$ km/s and $\sigma_{V,bad} = 5$ km/s to account for uncertainties in beamforming. Similarly, $L(\theta_j; \theta_{theoretical}, \sigma_{\theta,good})$ and $L(\theta_j; \theta_{theoretical}, \sigma_{\theta,bad})$ assess take-off angles with $\sigma_{\theta,good} = 15$ degrees for "good" picks and $\sigma_{\theta,bad} = 45$ degrees for "bad" picks. The azimuthal angle is not considered here for computational efficiency, as its observed values consistently align well with theoretical predictions through our experiments.

We use the No-U-Turn sampling algorithm [Hoffman and Gelman, 2014] to estimate the posterior distribution $P(R_1, R_2, \dots, R_m | D_1, D_2, \dots, D_n)$. The marginal posterior distribution $P(R_i | D_1, D_2, \dots, D_n)$ for any $i \leq m$ can be directly computed through the MCMC sampling. This distribution effectively represents the trustworthiness of any given X arrival post-analysis. By integrating this distribution, we can calculate the expectation value t_i for the trustworthiness of the i -th X arrival:

$$t_i = E(R_i) = \int_0^1 P(R_i(x) | D_1, D_2, \dots, D_n) dx \quad (4.6)$$

This expectation t_i serves as a quantitative measure of trustworthiness for each PS arrival, allowing us to make informed decisions about which arrivals to retain for further analysis based on their calculated trustworthiness levels. We discard all X arrivals with $t_i < 0.5$.

The accuracy of estimating t_i , the trustworthiness of X arrivals, relies on the results of the beam-forming analysis, D_j . To obtain these crucial data points, we conduct extensive source-side beam-forming analyses on subsets of X arrivals. Given the wide distribution of earthquakes in our study, we divide the X arrivals into subsets for beamforming. We first construct a grid with step lengths of 0.25° in longitude and latitude, and 25 km in depth, organizing the study region into boxes that are 1° wide in longitude, 1° wide in latitude, and 100 km wide in depth. Each box is designed to contain at least 20 X arrivals from the same station. This approach excludes X arrivals from events that are far from the others. Within each box, we subsample the events to generate more beamforming observations. We repeat this process 1000 times with a subsampling rate of 70%. In total, we obtain more than 1 million beamforming results for uniquely different subsets of X arrivals.

4.4.4 Semi-supervised-learning-based workflow

Integrating Semi-Supervised Learning into our workflow significantly enhances the detection capabilities of PhaseNet-TF, employing a strategy that iteratively refines labeled PS arrival picks and retrains the model. This semi-supervised learning approach can effectively leverage a blend of limited labeled phases with an extensive pool of unlabeled data. As illustrated in the diagram of Figure 4.6, in the initial iteration, given the scarcity of manually picked PS arrivals, we bypass the quality control steps related to timing and energy, operating under the assumption that most manually picked phase arrivals are reliable. This iteration of the model is then applied to the continuous dataset, which is constrained by the catalog from a prior study [Xi et al., 2023], to generate new phase picks, including potential false detections.

The introduction of subsequent steps, such as quality control through energy and timing analysis, along with PS wave association via beamforming and the MCMC method, assists in filtering out less reliable picks. These refined PS arrivals from the first iteration serve as labels for training PhaseNet-TF in the second iteration. With each iteration, as the training dataset is augmented with more accurate picks, the model’s proficiency in detecting PS arrivals is expected to improve. This iterative process is repeated until the addition of significant new PS arrivals ceases, ensuring a progressively refined and accurate detection mechanism.

Both deep learning and beamforming require a large amount of PS arrivals. However, the manually picked reference catalog has only 5,127 PS-wave arrivals from 1,163 local earthquakes. Similar to Xi et al. [2023], we use a semi-supervised learning strategy to iteratively increase the number of PS arrivals, refine them with QC filters, and retrain PhaseNet-TF (Figure 4.6). This approach can effectively leverage a blend of limited labeled phases with an extensive pool of unlabelled data. Before Iteration #1, we only apply the QC filters of source location, station location, travel-time, and energy to clean the reference dataset (Section 4.4.2), and do not use the QC filter of beamforming analysis (Section 4.4.3), given the scarcity of manually picked arrivals. In Iteration #1, we use the cleaned reference dataset to train PhaseNet-TF and apply the trained model to the continuous dataset to predict new X arrivals. The predictions are subsequently filtered by all QC measures. In

the following iterations, we use the cleaned X picks from the previous iteration to retrain PhaseNet-TF and predict more X arrivals, which are cleaned by QC filters explained in Sections 4.4.2 and 4.4.3. This iterative process is repeated until the addition of significant new PS arrivals ceases, ensuring a progressively refined and accurate detection mechanism.

4.5 Results

4.5.1 PhaseNet-TF Model assessment

The performance of PhaseNet-TF in detecting PS arrivals is rigorously evaluated using precision, recall, and F1 score metrics on a test dataset, which is randomly selected as 5% of the reference catalog for the initial iteration and 5% of the refined catalog for subsequent iterations. Table 4.1 summarizes these results, showing that PhaseNet-TF consistently achieves high accuracy in predicting P and S arrivals, with metrics above 95% across all iterations, and maintains over 95% precision and recall in detecting PS arrivals. Figure 4.7 also shows an example of newly detected PS arrival at the second supervised learning stage. This level of precision is comparable to manual picks by human analysts, indicating the model's robustness.

However, an essential aspect of the semi-supervised learning process involves the exclusion of some PS arrivals from the second iteration onwards. This is not due to inaccuracies in their detection or direct dismissal by the MCMC process but originates from the requirement for a minimum of 20 PS arrivals to form a box for beamforming analysis, which is a prerequisite not always met. Such exclusions, while potentially omitting valid PS arrivals, represent a necessary compromise to adhere to the structural requirements of our workflow. This is also suggested in Figure 4.8, where the prediction and the reference residuals of the PS wave keep almost the same at different stages of the semi-supervised framework, as well as a low standard deviation, but with a smaller number of PS arrivals preserved in catalogs of stages 2 and beyond.

4.5.2 catalog refined with beamforming

In refining our PS arrival catalog, we began by training the PhaseNet-TF model on 5,127 manually selected PS arrivals from an existing reference catalog, which enabled the model to predict approximately 46,210 PS arrivals from a dataset spanning an entire year. Utilizing constraints from

the previous study (Xi et al. [2023]) as described in the data section, we focused the initial large set of predictions down to 18,290 PS arrivals. As shown in 4.3(b), through subsequent quality control measures that emphasized precise timing and energy distribution analysis, this number was further refined to 15,939 PS arrivals. Only around 12,000 PS arrivals can perform beam-forming through the boxing constraint. The further refinement process incorporated beam-forming analysis and MCMC simulations as critical components, enabling a detailed evaluation of each PS arrival prediction. Figure 4.4 shows the PS arrivals are clearly clustered into two groups, i.e., trustable or not, through this approach. This stage was instrumental in selecting approximately 8,979 PS arrivals, enriching the catalog with both previously recognized arrivals and 6,200 new detections. Building on this enhanced dataset, the PhaseNet-TF model underwent additional training iterations. With the detection of new PS arrivals plateauing after the third iteration, we concluded the semi-supervised learning process. Figure 4.9 compares the station-counted PS arrivals between the reference and the subsequent stages' catalog, which showcases this improvement.

In total, for the reference catalogue, it takes around 5 hours and 400 CPU cores on MSU ICER for the beam-forming calculation, and around 1 hour and 4 Nvidia V100S GPUs for the 4-chains MCMC simulation. For the PS arrivals predicted by the PhaseNet-TF model (first semi-supervised iteration, see the next section), it takes around 36 hours and 400 CPUS for the beam-forming calculation, and around 4 hours and 4 Nvidia V100S GPUs for the 4 chains MCMC simulation. Figure 4.5 shows a comparison before and after applying the abovementioned method to the PS arrivals from the reference catalog, which shows its effectiveness in detecting the potentially falsely detected PS arrivals.

This methodical and iterative approach to catalog refinement not only expanded the number of detected PS arrivals but also significantly improved the accuracy and reliability of the catalog. With ongoing developments in Ocean Bottom Seismometer (OBS) deployments, such as the contributions from the SaLOON Project, our refined methodology is set to further advance PS arrival detection in future seismic research, ensuring a more comprehensive and precise understanding of seismic activities.

4.6 Discussion and Conclusions

In our study, we have developed a comprehensive workflow for the detection and analysis of converted seismic waves within the Tonga subduction zone. This workflow represents a significant advancement in seismic research, particularly in the context of improving the accuracy and reliability of PS wave arrival detection.

To be specific, we find:

- Our study introduces a robust workflow that combines the enhanced PhaseNet-TF model with a semi-supervised learning approach. This integration is crucial for addressing the complex challenge of detecting weak and noisy signals in seismic data, thereby enhancing the precision of phase picking.
- A novel aspect of our workflow is the incorporation of Markov chain Monte Carlo simulations and source-side beamforming as quality control measures. These techniques allow us to refine our detection process significantly, ensuring that our analysis accurately distinguishes between true seismic events and background noise.
- By applying our workflow, we have substantially extended the existing catalog of PS arrivals in the Tonga subduction zone. This expanded catalog is not only a valuable resource for ongoing research but also holds great potential for future studies focused on seismic tomography and the dynamics of earthquakes within this subduction zone.

4.7 Figures

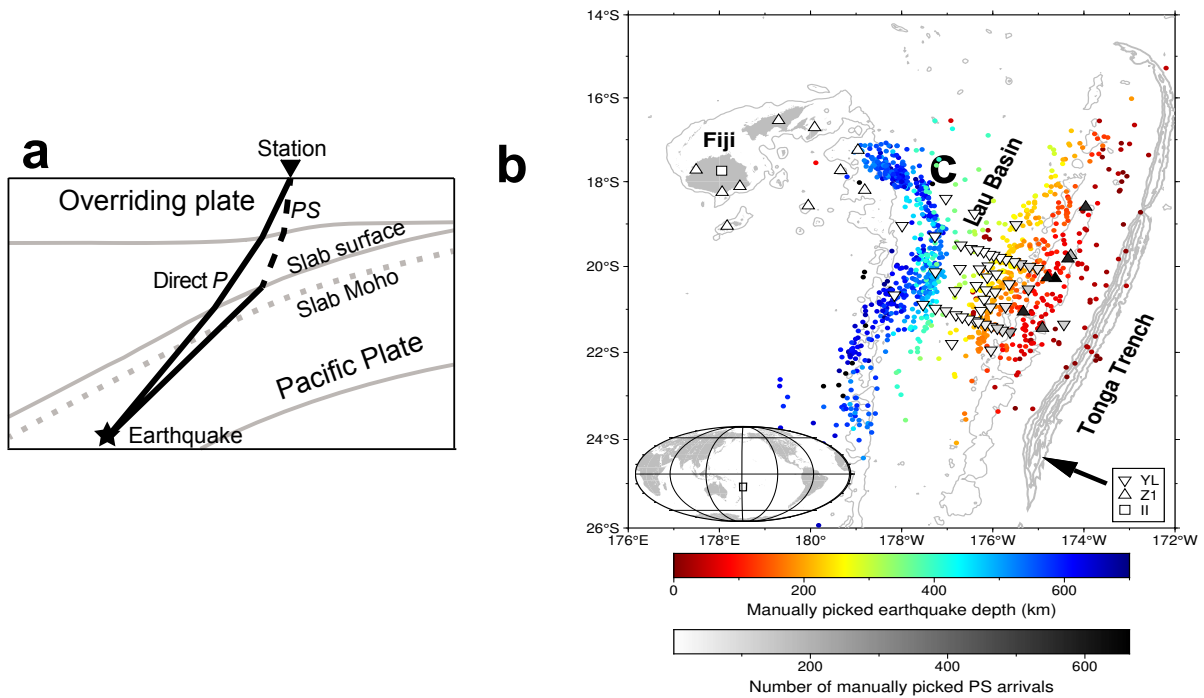


Figure 4.1 (a) Diagrams showing direct P waves and PS waves from an earthquake near the slab surface. The figures illustrate the paths of seismic waves as they travel through the Earth's layers. (b) Map showing the Tonga subduction zone and surrounding areas. Triangles represent land-based seismic stations, inverted triangles denote ocean-bottom seismographs, and the square marks a GSN (Global Seismograph Network) station. These stations were operational from November 2009 to December 2010. The color coding of the stations indicates the number of PS waves recorded at each station for the reference catalog. The distribution of earthquakes from the manually picked reference catalog is depicted with dots, color-coded according to their depth, showcasing the spatial and depth variation of seismic activity in the region. Land regions are marked in grey. The bathymetry, with 1 km contours, highlights geographical features such as the Tonga Ridge, Lau Ridge, and Fiji Plateau, while deeper contours at 7, 8, 9, and 10 km outline the Tonga Trench. The black arrow shows the direction of the Pacific Plate's movement relative to the Tonga area. An inset map provides a global perspective of the study area.

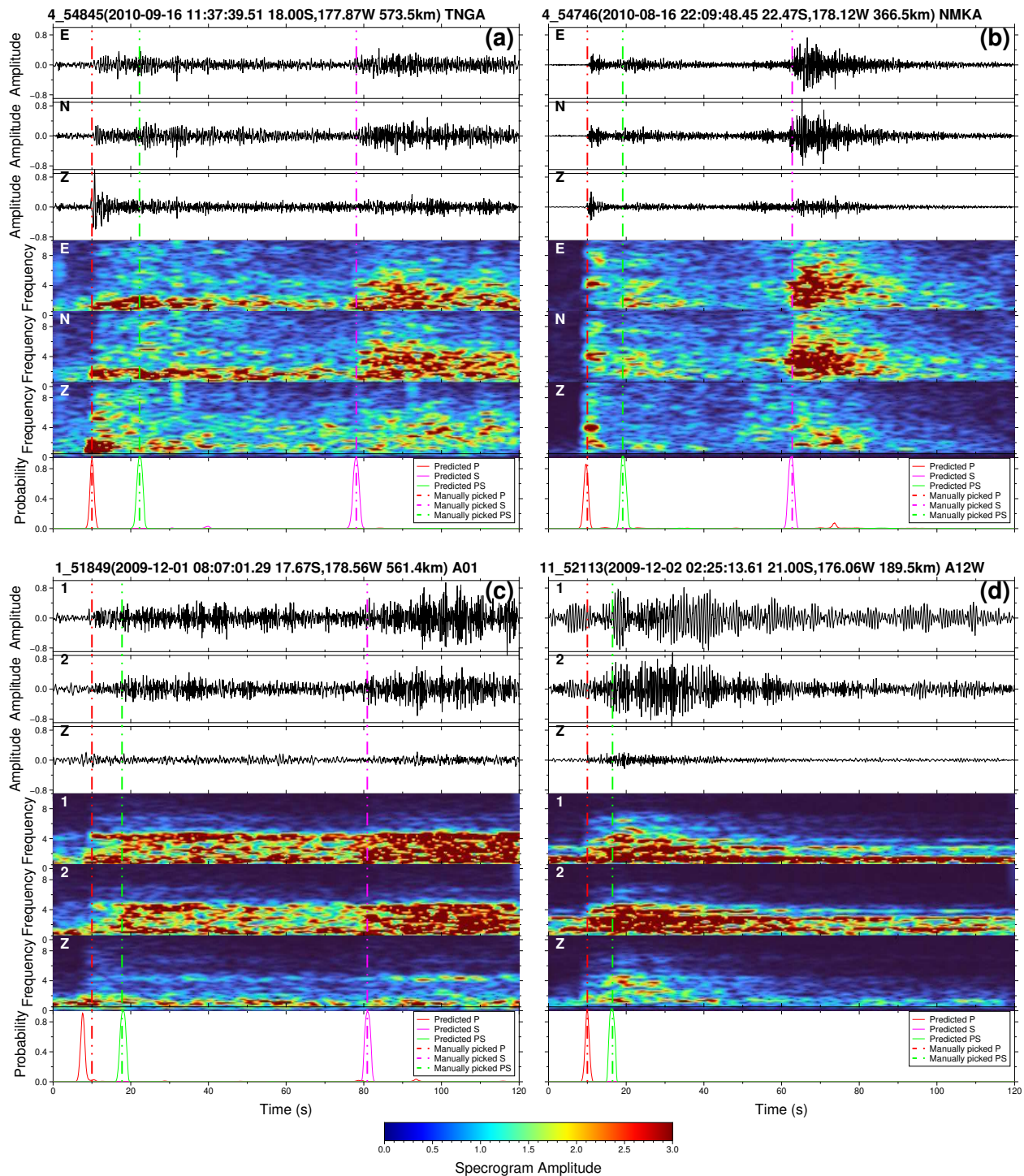


Figure 4.2 PhaseNet-TF model predictions on selected examples from the test dataset, derived from the reference dataset. Panels (a) and (b) illustrate recordings of deep earthquakes at land-based fore-arc stations, while panels (c) and (d) present recordings from back-arc stations equipped with Ocean Bottom Seismometers (OBS). Each panel is labeled with the earthquake's origin time, hypocenter, and the respective station name. The initial three sections in each panel depict waveform data across three components (vertical, north-south, and east-west) in the time domain. These are followed by three corresponding sections showing spectrograms, which represent the power values of each waveform component. The final section in each panel visualizes the probabilities of detected P, S, and PS wave arrivals as predicted by the model. Manually identified arrivals for P, S, and PS waves are marked with dashed lines for comparison.

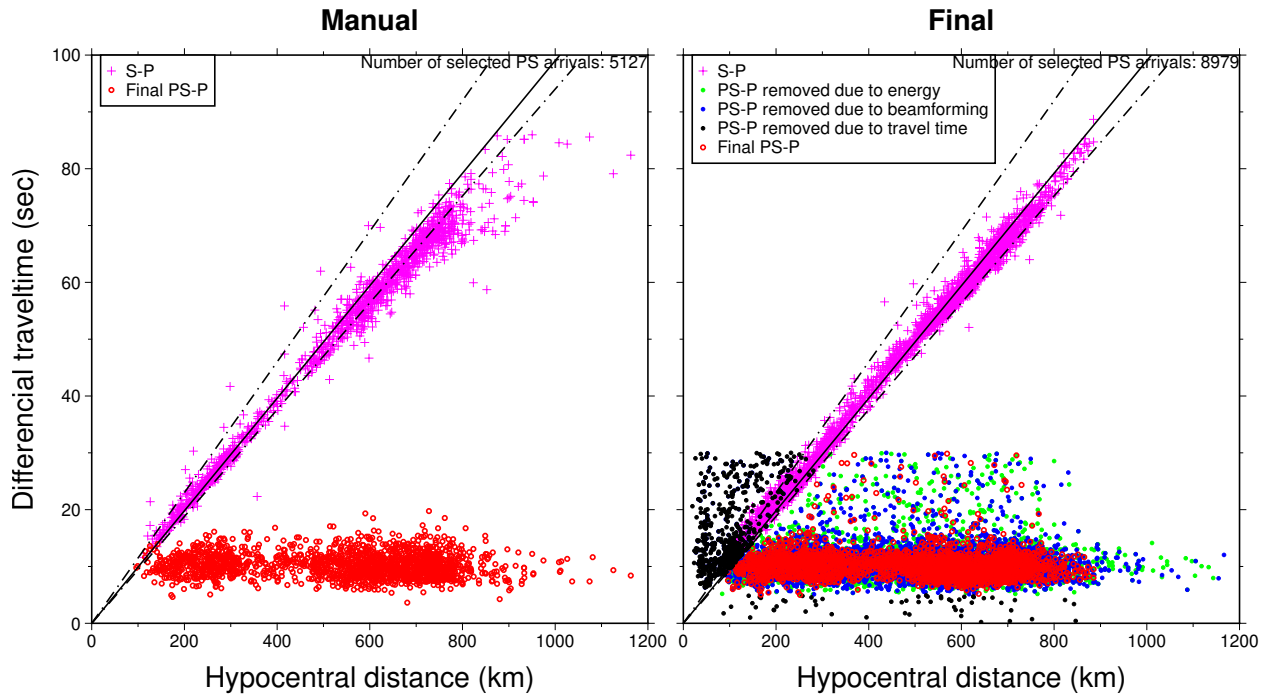


Figure 4.3 Comparison of PS-P and S-P arrival time differences versus hypocentral distance between the reference and final catalogs. The final catalog was developed by applying energy, timing, and beamforming-based quality control measures to PhaseNet-TF predictions of PS arrivals, represented as colored dots. The solid black line illustrates the least squares fit of the PS-P time difference curve of the final catalog. The two dashed black lines indicate the 5% and 95% boundaries, marking the regions where the majority of PS arrivals from the final catalog fall within these limits.

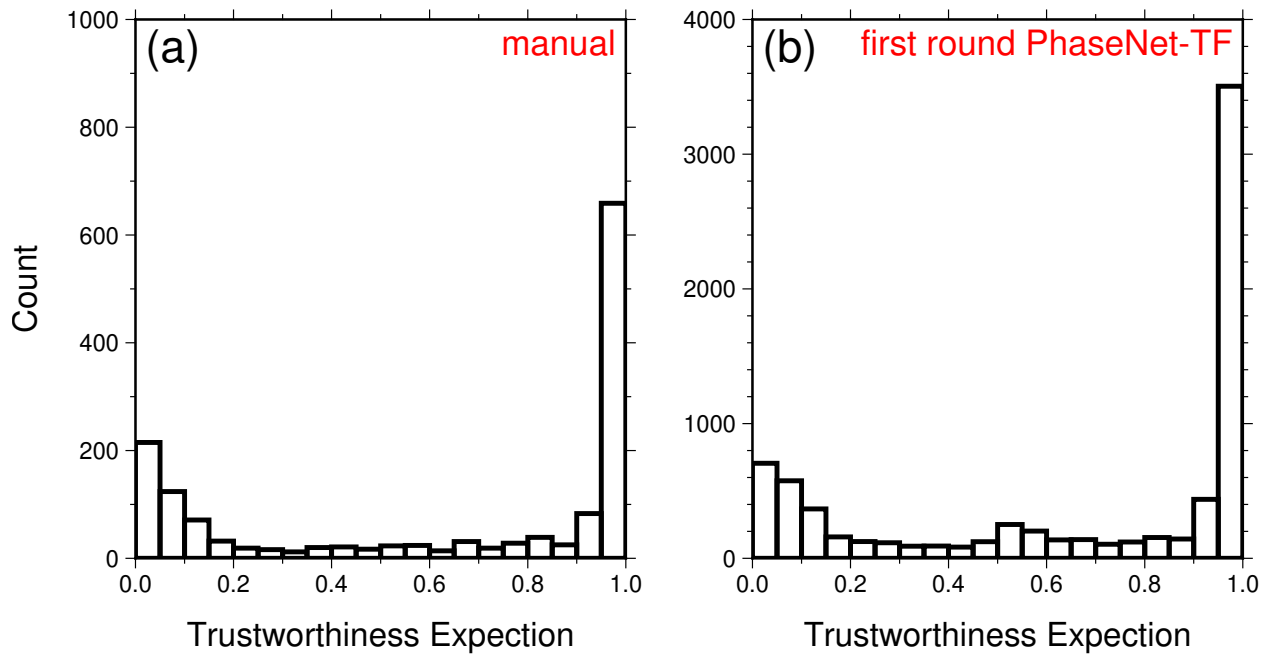


Figure 4.4 Distribution of trustworthiness values derived from beam-forming MCMC simulations, where higher values indicate greater reliability of the corresponding PS arrival. Panel (a) displays the evaluation of this method using the reference dataset, while panel (b) shows the outcomes after the first iteration of our semi-supervised learning framework.

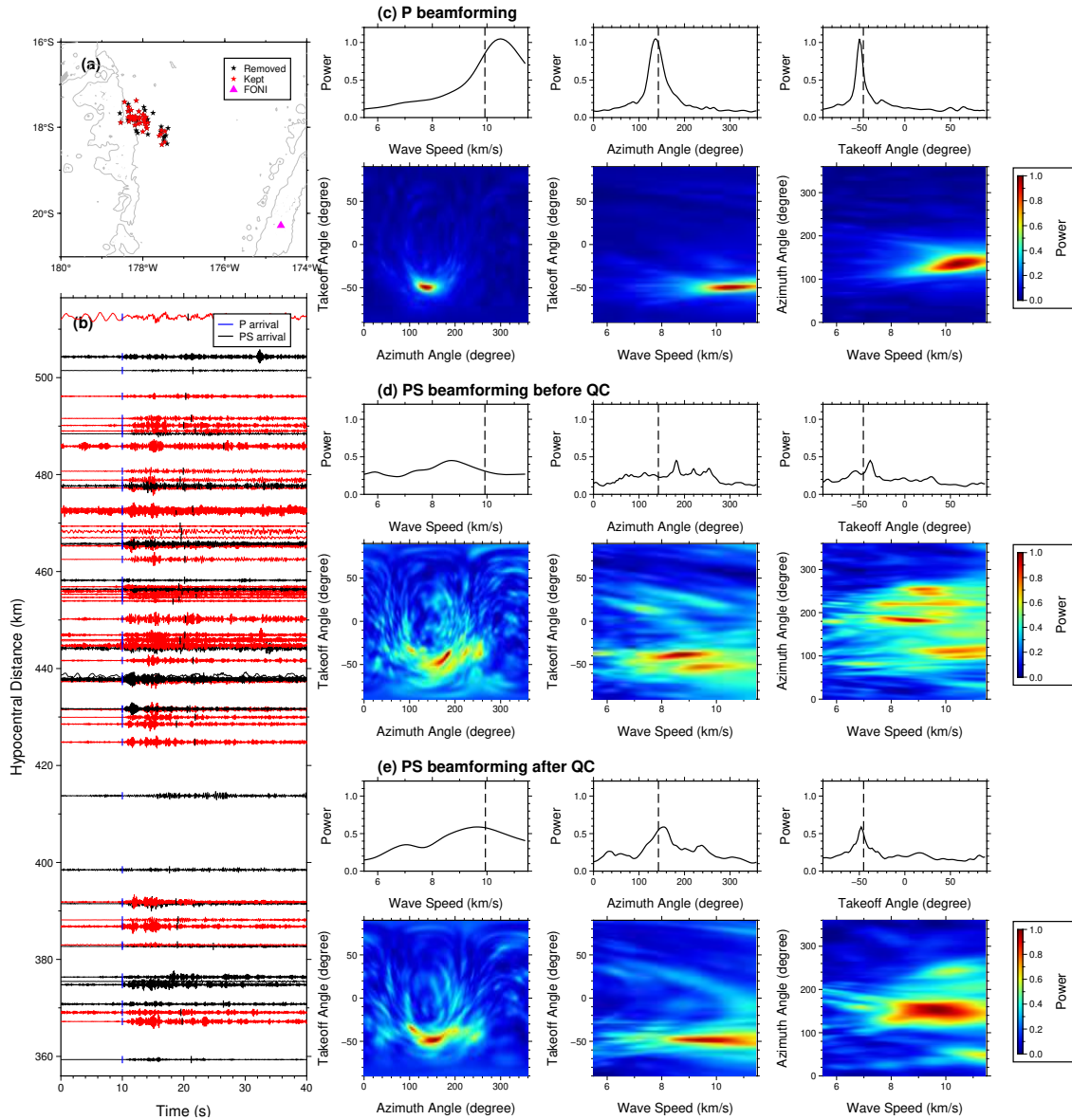


Figure 4.5 Comparison of beamforming results before and after applying MCMC-based quality control on the reference dataset. The top-left map displays station FONI and associated earthquake events, represented as either black or red dots. Red dots indicate events whose PS arrivals are retained post-MCMC analysis, while black dots represent discarded arrivals. The bottom-left plot shows waveform data for PS arrivals, aligned at the P arrival time, with red indicating arrivals kept after beamforming analysis and black indicating those discarded. This data has been filtered around the center frequency of manually identified PS waves, with a bandwidth of 4 Hz. On the right, beamforming results are presented in a series of plots: the top plot shows results for all P arrivals; the middle plot displays results for all PS arrivals before MCMC filtering; the bottom plot illustrates results for PS arrivals post-MCMC filtering. Following the method described by Gong and McGuire [2021], we optimize the power spectrum, and the plots depict this spectrum as a function of various parameters: wave speed, azimuthal angle, and takeoff angle, as well as their combinations, such as takeoff angle and azimuthal angle, takeoff angle and wave speed, and azimuthal angle and wave speed.

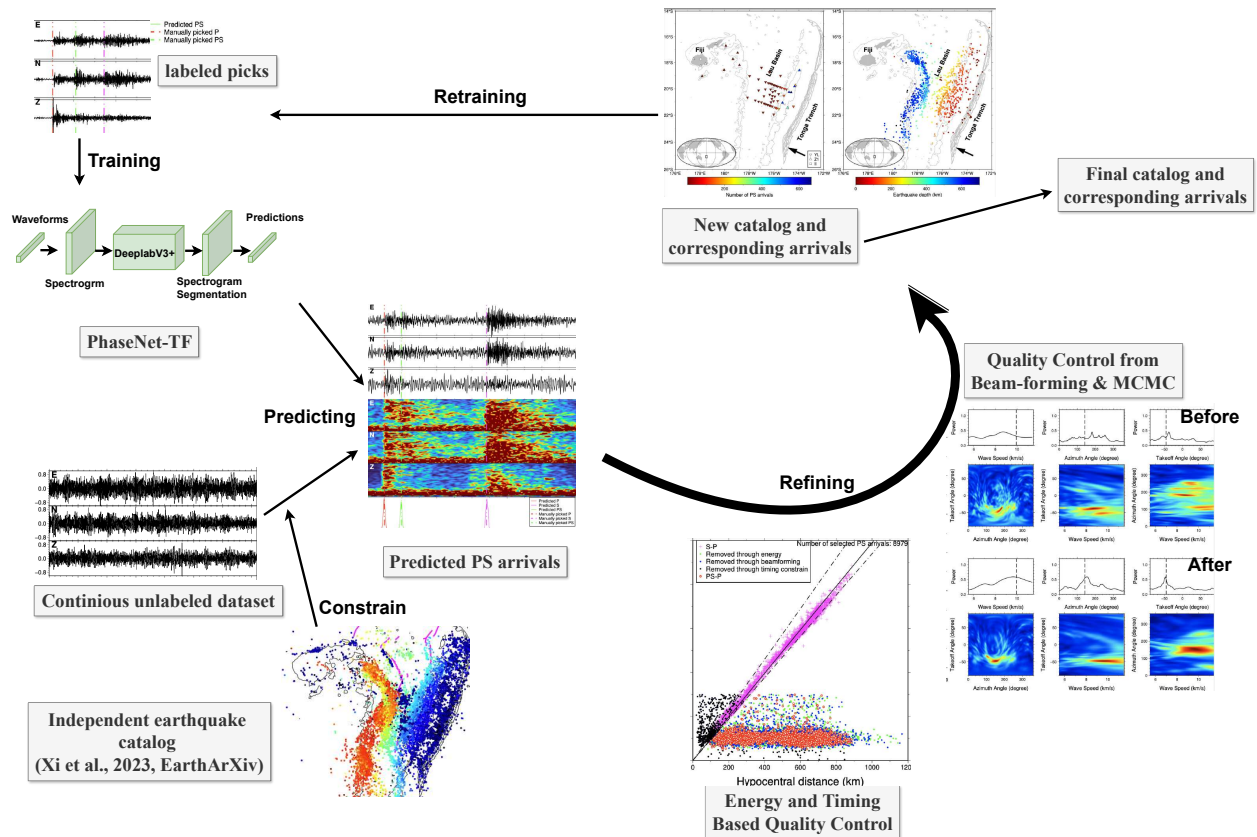


Figure 4.6 Schematic illustration of our semi-supervised learning workflow. The process initiates with the training of the PhaseNet-TF model on a dataset of manually picked seismic arrivals. Subsequently, this trained model is deployed on continuous seismic records to predict probabilities for P, S, and PS wave arrivals. These predictions, specifically for PS arrivals, are then cross-referenced with a finalized catalog from previous studies, such as the one by Xi et al. (2023), to ensure consistency and reliability. Quality control measures, including analysis of timing discrepancies, the ratio of horizontal to vertical energy, and beamforming-based MCMC simulations, are employed to refine this set of PS arrivals, resulting in an updated catalog. This refined catalog is then utilized as a newly labeled dataset for further training and validation of the PhaseNet-TF model in subsequent iterations. The workflow is iteratively executed until the enhancements in the catalog become marginal, culminating in a comprehensive and accurate final catalog of seismic events and their associated phase arrivals.

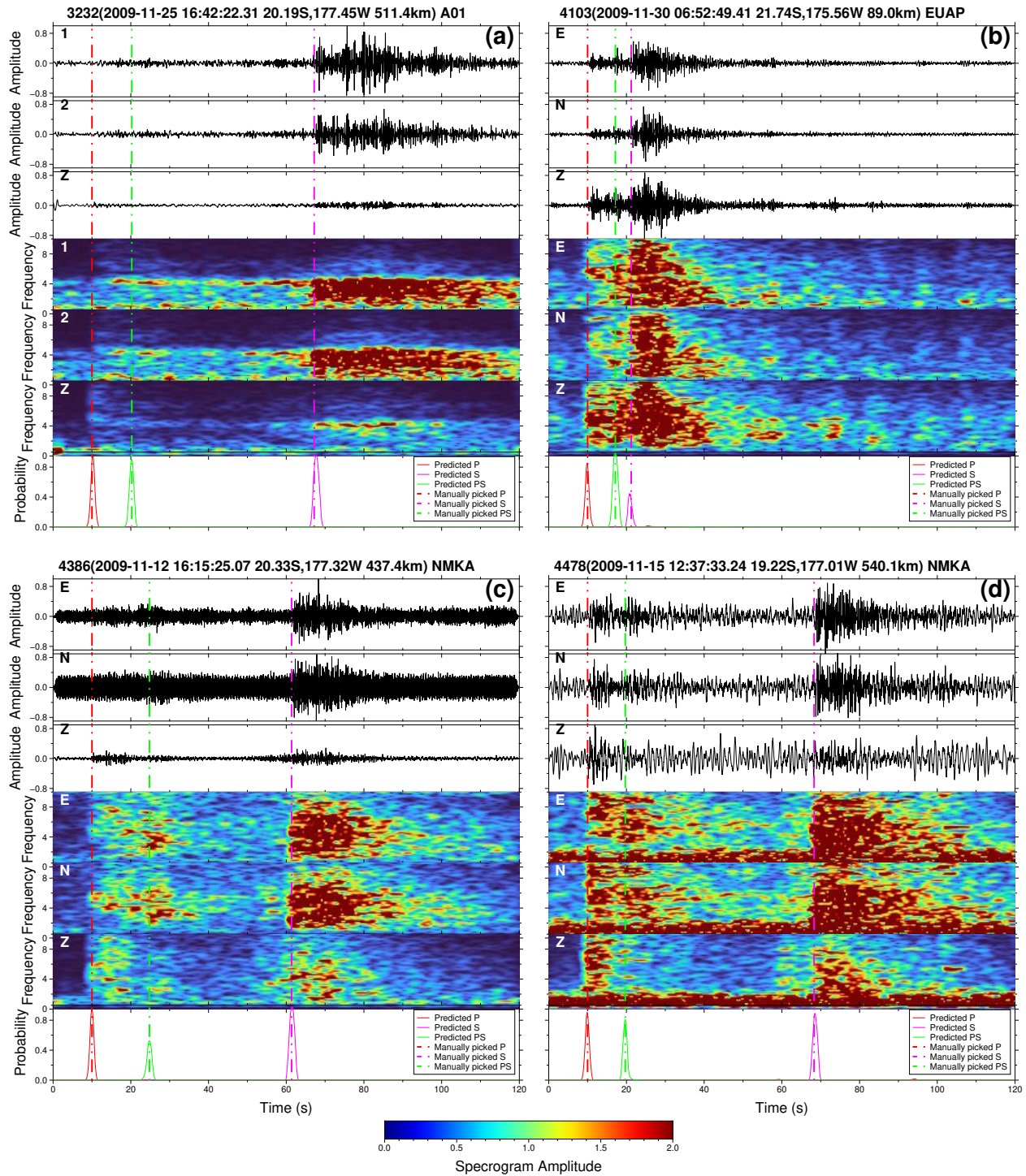


Figure 4.7 Examples of PhaseNet-TF predictions from the test dataset, which was sampled from the training dataset of the second-stage semi-supervised learning. Annotations and layout follow those described in figure 4.2.

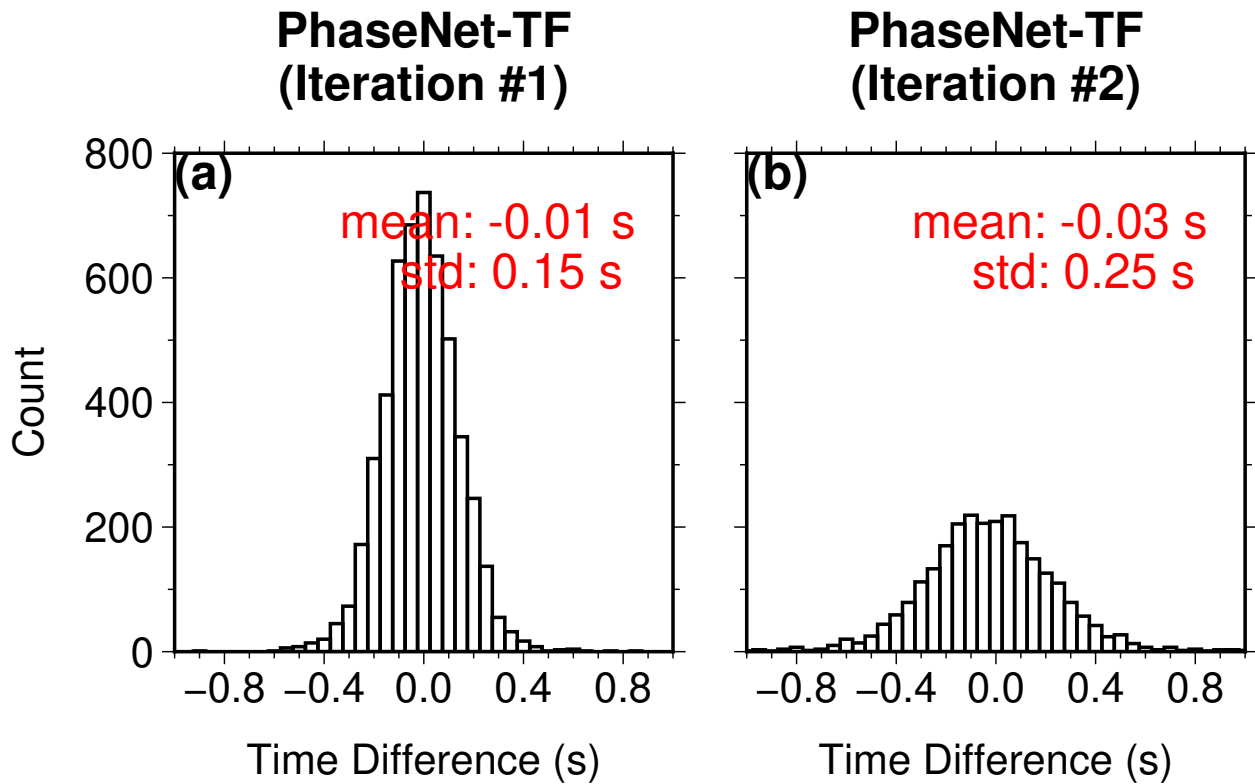


Figure 4.8 Arrival-time residuals for PS waves throughout various iterations of semi-supervised learning. Each histogram represents the differences between predicted arrival times, as determined by the PhaseNet-TF model at each iteration, and manual picks, for instances where both are available. Panel (a) showcases the residuals from the first iteration of the PhaseNet-TF model application, while panel (b) displays the outcomes from the second iteration.

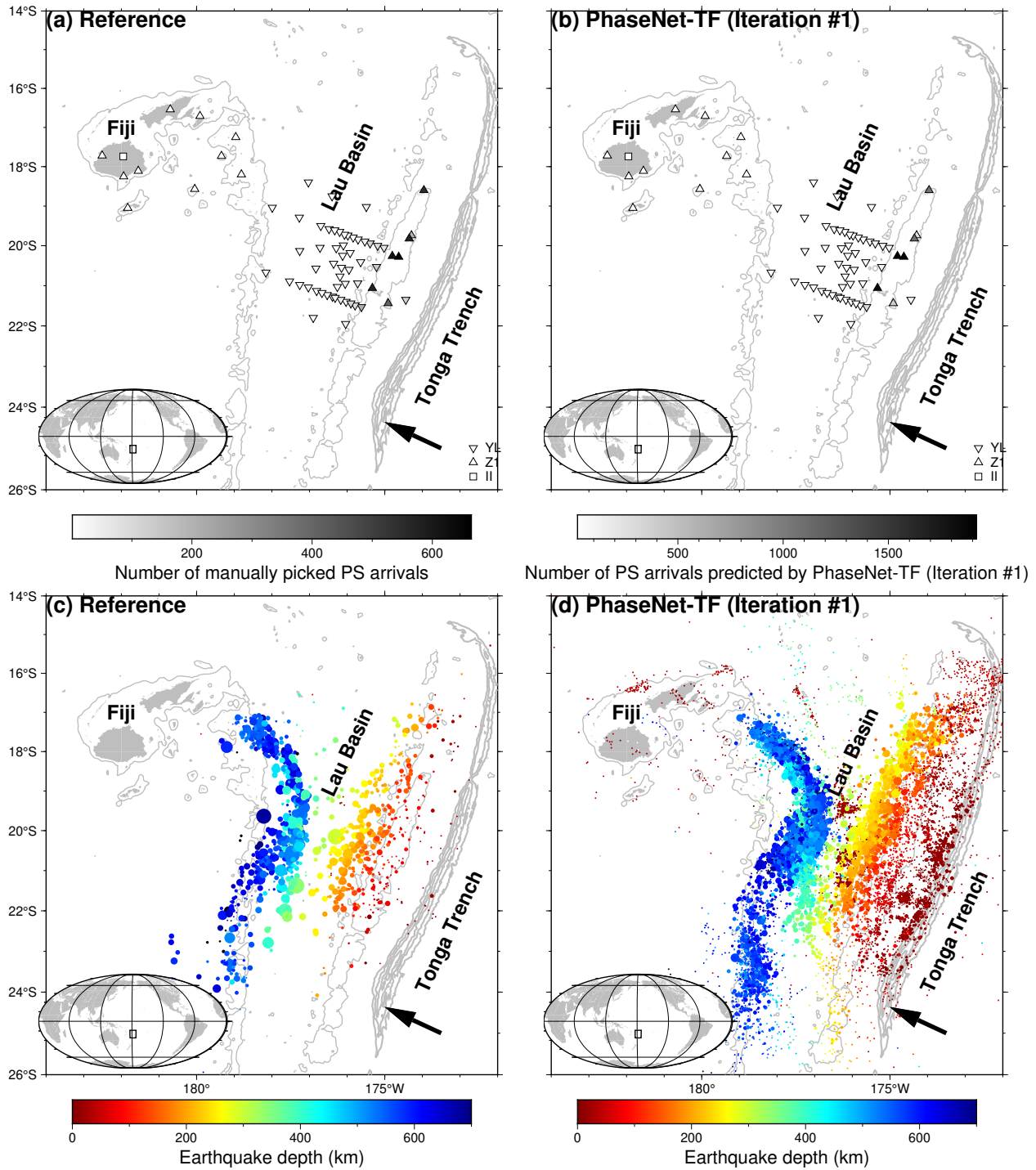


Figure 4.9 Comparison of the reference and final catalogs, with stations color-coded according to the number of PS arrivals recorded at each location (a-b), and events color-coded according to the depth (c-d). In (c-d), the sizes of events are scaled based on the number of PS arrivals. The map includes features similar to those described in Figure 4.1, such as station symbols, land shading, and bathymetry contours.

4.8 Tables

Table 4.1 Evaluation metrics for phase pickers applied to the test dataset across different stages of semi-supervised learning models. Picks are considered true positives if their arrival-time residuals are less than 1 second when compared to the corresponding labels. The table presents metrics evaluated on a test dataset, which is partitioned from the labelled dataset. For PhaseNet-TF Iteration #1, this labelled dataset is the manually picked reference catalogue. For subsequent iterations, such as PhaseNet-TF Iteration #2, the labelled dataset is derived from the output catalogue of the preceding iteration.

	<i>P</i> -wave arrival			<i>S</i> -wave arrival			<i>PS</i> -wave arrival		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
PhaseNet-TF (Iteration #1)	0.99	0.99	0.99	0.95	0.99	0.97	0.98	0.999	0.99
PhaseNet-TF (Iteration #2)	0.98	0.97	0.97	0.94	0.99	0.97	0.93	0.99	0.97

BIBLIOGRAPHY

- John Adam, Simon Turner, and Tracy Rushmer. The genesis of silicic arc magmas in shallow crustal cold zones. *Lithos*, 264:472–494, 2016. doi: 10.1016/j.lithos.2016.07.036. Publisher: Elsevier.
- Keiiti Aki and W. H. K. Lee. Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model. *Journal of Geophysical Research*, 81(23):4381–4399, August 1976. ISSN 01480227. doi: 10.1029/JB081i023p04381.
- C. J. Allégre, V. Courtillot, P. Tapponnier, A. Hirn, M. Mattauer, C. Coulon, J. J. Jaeger, J. Achache, U. Schärer, J. Marcoux, J. P. Burg, J. Girardeau, R. Armijo, C. Gariépy, C. Göpel, Li Tindong, Xiao Xuchang, Chang Chenfa, Li Guangqin, Lin Baoyu, Teng Jiwen, Wang Naiwen, Chen Guoming, Han Tonglin, Wang Xibin, Den Wanming, Sheng Huaibin, Cao Yougong, Zhou Ji, Qiu Hongrong, Bao Peisheng, Wang Songchan, Wang Bixiang, Zhou Yaoxiu, and Ronghua Xu. Structure and evolution of the Himalaya–Tibet orogenic belt. *Nature*, 307(5946):17–22, January 1984. ISSN 0028-0836, 1476-4687. doi: 10.1038/307017a0.
- M. L. Amaru. *Global travel time tomography with 3-D reference models*, volume 274. Utrecht University, 2007.
- M. Baer and U. Kradolfer. An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, 77(4):1437–1445, August 1987. ISSN 0037-1106. doi: 10.1785/BSSA0770041437.
- Xuewei Bao, Xiaodong Song, and Jiangtao Li. High-resolution lithospheric structure beneath Mainland China from ambient noise and earthquake surface-wave tomography. *Earth and Planetary Science Letters*, 417:132–141, May 2015. ISSN 0012821X. doi: 10.1016/j.epsl.2015.02.024.
- M. Beyreuther, R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann. ObsPy: A Python Toolbox for Seismology. *Seismological Research Letters*, 81(3):530–533, May 2010. ISSN 0895-0695. doi: 10.1785/gssrl.81.3.530.
- Günter Bock, Bernd Schurr, and Günter Asch. High-resolution image of the oceanic moho in the subducting Nazca Plate from *P-S* converted waves. *Geophysical Research Letters*, 27(23):3929–3932, December 2000. ISSN 0094-8276, 1944-8007. doi: 10.1029/2000GL011881.
- Thomas Bornstein, Dietrich Lange, Jannes Münchmeyer, Jack Woollam, Andreas Rietbrock, Grace Barcheck, Ingo Grevemeyer, and Frederik Tilmann. PickBlue: Seismic phase picking for ocean bottom seismometers with deep learning, April 2023. arXiv:2304.06635 [physics].
- Ebru Bozdağ, Jeannot Trampert, and Jeroen Tromp. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. *Geophysical Journal International*, 185(2):845–870, May 2011. ISSN 0956-540X. doi: 10.1111/j.1365-246X.2011.04970.x.

- Ebru Bozdağ, Daniel Peter, Matthieu Lefebvre, Dimitri Komatitsch, Jeroen Tromp, Judith Hill, Norbert Podhorszki, and David Pugmire. Global adjoint tomography: first-generation model. *Geophysical Journal International*, 207(3):1739–1766, December 2016. ISSN 0956-540X, 1365-246X. doi: 10.1093/gji/ggw356.
- E. Bruce Watson and James M. Brennan. Fluids in the lithosphere, 1. Experimentally-determined wetting characteristics of CO₂H₂O fluids and their implications for fluid transport, host-rock physical properties, and fluid inclusion formation. *Earth and Planetary Science Letters*, 85(4): 497–515, October 1987. ISSN 0012821X. doi: 10.1016/0012-821X(87)90144-0.
- Michael R. Brudzinski, Clifford H. Thurber, Bradley R. Hacker, and E. Robert Engdahl. Global Prevalence of Double Benioff Zones. *Science*, 316(5830):1472–1474, June 2007. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1139204.
- Chen Cai and Douglas A. Wiens. Dynamic triggering of deep earthquakes within a fossil slab. *Geophysical Research Letters*, 43(18):9492–9499, September 2016. ISSN 0094-8276, 1944-8007. doi: 10.1002/2016GL070347.
- Chengping Chai, Monica Maceira, Hector J. Santos-Villalobos, Singanallur V. Venkatakrisnan, Martin Schoenball, Weiqiang Zhu, Gregory C. Beroza, Clifford Thurber, and EGS Collab Team. Using a Deep Neural Network and Transfer Learning to Bridge Scales for Seismic Phase Picking. *Geophysical Research Letters*, 47(16):e2020GL088651, 2020. ISSN 1944-8007. doi: 10.1029/2020GL088651.
- Jie Chen, Yongshun John Chen, Douglas A. Wiens, S. Shawn Wei, Yang Zha, Jordi Julià, and Chen Cai. Crustal and lithospheric structure of inactive volcanic arc terrains in Fiji. *Tectonophysics*, 750:394–403, 2019a. Publisher: Elsevier.
- Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, August 2018. arXiv:1802.02611 [cs].
- Min Chen, Fenglin Niu, Qinya Liu, Jeroen Tromp, and Xiufen Zheng. Multiparameter adjoint tomography of the crust and upper mantle beneath East Asia: 1. Model construction and comparisons. *Journal of Geophysical Research: Solid Earth*, 120(3):1762–1786, March 2015. ISSN 21699313. doi: 10.1002/2014JB011638.
- Min Chen, Vlad Constantin Manea, Fenglin Niu, S. Shawn Wei, and Eric Kiser. Genesis of Intermediate-Depth and Deep Intraslab Earthquakes beneath Japan Constrained by Seismic Tomography, Seismicity, and Thermal Modeling. *Geophysical Research Letters*, 46(4):2025–2036, 2019b. ISSN 1944-8007. doi: 10.1029/2018GL080025.
- Robert Clayton and Björn Engquist. Absorbing boundary conditions for acoustic and elastic wave equations. *Bulletin of the Seismological Society of America*, 67(6):1529–1540, December 1977. ISSN 1943-3573, 0037-1106. doi: 10.1785/BSSA0670061529.

- James A. Conder and Douglas A. Wiens. Seismic structure beneath the Tonga arc and Lau back-arc basin determined from joint Vp, Vp/Vs tomography. *Geochemistry, Geophysics, Geosystems*, 7(3):2005GC001113, March 2006. ISSN 1525-2027, 1525-2027. doi: 10.1029/2005GC001113.
- C. DeMets, R. G. Gordon, D. F. Argus, and S. Stein. Current plate motions. *Geophysical Journal International*, 101(2):425–478, May 1990. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.1990.tb06579.x.
- Joseph J. Durek and Göran Ekström. A radial model of anelasticity consistent with long-period surface-wave attenuation. *Bulletin of the Seismological Society of America*, 86(1A):144–158, February 1996. ISSN 0037-1106. doi: 10.1785/BSSA08601A0144.
- A. M. Dziewonski, T.-A. Chou, and J. H. Woodhouse. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research: Solid Earth*, 86(B4):2825–2852, April 1981. ISSN 0148-0227. doi: 10.1029/JB086iB04p02825.
- Adam M. Dziewonski and Don L. Anderson. Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, 25(4):297–356, June 1981. ISSN 00319201. doi: 10.1016/0031-9201(81)90046-7.
- Adam M. Dziewonski, Bradford H. Hager, and Richard J. O’Connell. Large-scale heterogeneities in the lower mantle. *Journal of Geophysical Research*, 82(2):239–255, January 1977. ISSN 01480227. doi: 10.1029/JB082i002p00239.
- G. Ekström, M. Nettles, and A.M. Dziewoński. The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200-201: 1–9, June 2012. ISSN 00319201. doi: 10.1016/j.pepi.2012.04.002.
- E. Robert Engdahl, Rob van der Hilst, and Raymond Buland. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bulletin of the Seismological Society of America*, 88(3):722–743, 1998. Publisher: The Seismological Society of America.
- Wenyuan Fan, S. Shawn Wei, Dongdong Tian, Jeffrey J. McGuire, and Douglas A. Wiens. Complex and Diverse Rupture Processes of the 2018 M_w 8.2 and M_w 7.9 Tonga-Fiji Deep Earthquakes. *Geophysical Research Letters*, 46(5):2434–2448, March 2019. ISSN 0094-8276, 1944-8007. doi: 10.1029/2018GL080997.
- Andreas Fichtner and Jeannot Trampert. Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International*, 185(2):775–798, May 2011a. ISSN 0956540X. doi: 10.1111/j.1365-246X.2011.04966.x.
- Andreas Fichtner and Jeannot Trampert. Resolution analysis in full waveform inversion. *Geophysical Journal International*, 187(3):1604–1624, December 2011b. ISSN 0956540X. doi:

10.1111/j.1365-246X.2011.05218.x.

Andreas Fichtner, Brian L. N. Kennett, Heiner Igel, and Hans-Peter Bunge. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, 179(3):1703–1725, December 2009. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.2009.04368.x.

Andreas Fichtner, Brian L.N. Kennett, Heiner Igel, and Hans-Peter Bunge. Full waveform tomography for radially anisotropic structure: New insights into present and past states of the Australasian upper mantle. *Earth and Planetary Science Letters*, 290(3-4):270–280, February 2010. ISSN 0012821X. doi: 10.1016/j.epsl.2009.12.003.

Andreas Fichtner, Jeannot Trampert, Paul Cupillard, Erdinc Saygin, Tuncay Taymaz, Yann Capdeville, and Antonio Villaseñor. Multiscale full waveform inversion. *Geophysical Journal International*, 194(1):534–556, July 2013. ISSN 0956-540X. doi: 10.1093/gji/ggt118.

M. A. Florez and G. A. Prieto. Controlling Factors of Seismicity and Geometry in Double Seismic Zones. *Geophysical Research Letters*, 46(8):4174–4181, April 2019. ISSN 0094-8276, 1944-8007. doi: 10.1029/2018GL081168.

S. W. French and B. A. Romanowicz. Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography. *Geophysical Journal International*, 199(3):1303–1327, December 2014. ISSN 1365-246X, 0956-540X. doi: 10.1093/gji/ggu334.

Scott W. French and Barbara Romanowicz. Broad plumes rooted at the base of the Earth’s mantle beneath major hotspots. *Nature*, 525(7567):95–99, September 2015. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature14876.

Wolfgang Friederich. The S -velocity structure of the East Asian mantle from inversion of shear and surface waveforms. *Geophysical Journal International*, 153(1):88–102, April 2003. ISSN 0956540X, 1365246X. doi: 10.1046/j.1365-246X.2003.01869.x.

Lind S. Gee and Thomas H. Jordan. Generalized seismological data functionals. *Geophysical Journal International*, 111(2):363–390, November 1992. ISSN 0956-540X. doi: 10.1111/j.1365-246X.1992.tb00584.x.

S. Gentili and A. Michelini. Automatic picking of P and S phases using a neural tree. *Journal of Seismology*, 10(1):39–63, January 2006. ISSN 1573-157X. doi: 10.1007/s10950-006-2296-6.

Jianhua Gong and Jeffrey J. McGuire. Constraints on the Geometry of the Subducted Gorda Plate From Converted Phases Generated by Local Earthquakes. *Journal of Geophysical Research: Solid Earth*, 126(2):e2020JB019962, February 2021. ISSN 2169-9313, 2169-9356. doi: 10.1029/2020JB019962.

Jianhua Gong, Wenyuan Fan, and Ross Parnell-Turner. Machine Learning-Based New Earthquake

- Catalog Illuminates On-Fault and Off-Fault Seismicity Patterns at the Discovery Transform Fault, East Pacific Rise. *Geochemistry, Geophysics, Geosystems*, 24(9):e2023GC011043, 2023. ISSN 1525-2027. doi: 10.1029/2023GC011043.
- Harry W. Green and Heidi Houston. The Mechanics of Deep Earthquakes. *Annual Review of Earth and Planetary Sciences*, 23(1):169–213, May 1995. ISSN 0084-6597, 1545-4495. doi: 10.1146/annurev.ea.23.050195.001125.
- Zhen Guo, Y. John Chen, Jieyuan Ning, Yingjie Yang, Juan Carlos Afonso, and Youcai Tang. Seismic evidence of on-going sublithosphere upper mantle convection for intra-plate volcanism in Northeast China. *Earth and Planetary Science Letters*, 433:31–43, January 2016. ISSN 0012821X. doi: 10.1016/j.epsl.2015.09.035.
- Zhen Guo, Kai Wang, Yingjie Yang, Youcai Tang, Y. John Chen, and Shu-Huei Hung. The Origin and Mantle Dynamics of Quaternary Intraplate Volcanism in Northeast China From Joint Inversion of Surface Wave and Body Wave. *Journal of Geophysical Research: Solid Earth*, 123(3): 2410–2425, March 2018. ISSN 21699313. doi: 10.1002/2017JB014948.
- Bradley R. Hacker, Simon M. Peacock, Geoffrey A. Abers, and Stephen D. Holloway. Subduction factory 2. Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? *Journal of Geophysical Research: Solid Earth*, 108(B1):2001JB001129, January 2003. ISSN 0148-0227. doi: 10.1029/2001JB001129.
- Gavin P. Hayes, David J. Wald, and Rebecca L. Johnson. Slab1.0: A three-dimensional model of global subduction zone geometries. *Journal of Geophysical Research: Solid Earth*, 117(B1), 2012. ISSN 2156-2202. doi: 10.1029/2011JB008524.
- Gavin P. Hayes, Ginevra L. Moore, Daniel E. Portner, Mike Hearne, Hanna Flamme, Maria Furtney, and Gregory M. Smoczyk. Slab2, a comprehensive subduction zone geometry model. *Science*, 362(6410):58–61, October 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aat4723.
- Chuansong He, Shuwen Dong, Xuanhua Chen, M. Santosh, and Shuyin Niu. Seismic evidence for plume-induced rifting in the Songliao Basin of Northeast China. *Tectonophysics*, 627:171–181, July 2014. ISSN 00401951. doi: 10.1016/j.tecto.2013.07.015.
- George Helffrich and Geoffrey A. Abers. Slab low-velocity layer in the eastern Aleutian subduction zone. *Geophysical Journal International*, 130(3):640–648, 1997. doi: 10.1111/j.1365-246X.1997.tb01858.x. Publisher: Blackwell Publishing Ltd Oxford, UK.
- Matthew D. Hoffman and Andrew Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.
- Jinli Huang and Dapeng Zhao. High-resolution mantle tomography of China and surrounding regions. *Journal of Geophysical Research*, 111(B9):B09305, 2006. ISSN 0148-0227. doi: 10.1029/2005JB004066.

- Zhouchuan Huang, Pan Wang, Mingjie Xu, Liangshu Wang, Zhifeng Ding, Yan Wu, Mijian Xu, Ning Mi, Dayong Yu, and Hua Li. Mantle structure and dynamics beneath SE Tibet revealed by new seismic images. *Earth and Planetary Science Letters*, 411:100–111, February 2015. ISSN 0012821X. doi: 10.1016/j.epsl.2014.11.040.
- Toshihiro Igarashi, Toru Matsuzawa, Norihito Umino, and Akira Hasegawa. 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, February 2001. ISSN 0148-0227. doi: 10.1029/2000JB900386.
- International-Seismological-Centre. ISC-EHB dataset. 2023.
- Bryan Isacks and Peter Molnar. Distribution of stresses in the descending lithosphere from a global survey of focal-mechanism solutions of mantle earthquakes. *Reviews of Geophysics*, 9(1):103–174, February 1971. ISSN 8755-1209, 1944-9208. doi: 10.1029/RG009i001p00103.
- Zhe Jia, Zhichao Shen, Zhongwen Zhan, Chenyu Li, Zhigang Peng, and Michael Gurnis. The 2018 Fiji Mw 8.2 and 7.9 deep earthquakes: One doublet in two slabs. *Earth and Planetary Science Letters*, 531:115997, 2020. Publisher: Elsevier.
- Hitoshi Kawakatsu. Double seismic zone in Tonga. *Nature*, 316(6023):53–55, 1985. Publisher: Nature Publishing Group UK London.
- Hitoshi Kawakatsu, Prakash Kumar, Yasuko Takei, Masanao Shinohara, Toshihiko Kanazawa, Ei-ichiro Araki, and Kiyoshi Suyehiro. Seismic Evidence for Sharp Lithosphere-Asthenosphere Boundaries of Oceanic Plates. *Science*, 324(5926):499–502, April 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1169499.
- B. L. N. Kennett and E. R. Engdahl. Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, 105(2):429–465, May 1991. ISSN 0956-540X. doi: 10.1111/j.1365-246X.1991.tb06724.x.
- B. L. N. Kennett, E. R. Engdahl, and R. Buland. Constraints on seismic velocities in the Earth from traveltimes. *Geophysical Journal International*, 122(1):108–124, July 1995. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.1995.tb03540.x.
- Yong-Woo Kim, Sung-Joon Chang, Michael Witek, Jieyuan Ning, and Jingchong Wen. S-Velocity Mantle Structure of East Asia From Teleseismic Traveltime Tomography: Inferred Mechanisms for the Cenozoic Intraplate Volcanoes. *Journal of Geophysical Research: Solid Earth*, 126(3), March 2021. ISSN 2169-9313, 2169-9356. doi: 10.1029/2020JB020345.
- YoungHee Kim, Qinya Liu, and Jeroen Tromp. Adjoint centroid-moment tensor inversions. *Geophysical Journal International*, 186(1):264–278, July 2011. ISSN 0956540X. doi: 10.1111/j.1365-246X.2011.05027.x.

- Saeko Kita, Tomomi Okada, Akira Hasegawa, Junichi Nakajima, and Toru Matsuzawa. Anomalous deepening of a seismic belt in the upper-plane of the double seismic zone in the Pacific slab beneath the Hokkaido corner: Possible evidence for thermal shielding caused by subducted fore-arc crust materials. *Earth and Planetary Science Letters*, 290(3-4):415–426, 2010. Publisher: Elsevier.
- Dimitri Komatitsch and Jeroen Tromp. Spectral-element simulations of global seismic wave propagation-I. Validation. *Geophysical Journal International*, 149(2):390–412, May 2002a. ISSN 0956540X, 1365246X. doi: 10.1046/j.1365-246X.2002.01653.x.
- Dimitri Komatitsch and Jeroen Tromp. Spectral-element simulations of global seismic wave propagation-II. Three-dimensional models, oceans, rotation and self-gravitation. *Geophysical Journal International*, 150(1):303–318, July 2002b. ISSN 0956540X, 1365246X. doi: 10.1046/j.1365-246X.2002.01716.x.
- Lion Krischer, James Smith, Wenjie Lei, Matthieu Lefebvre, Youyi Ruan, Elliott Sales De Andrade, Norbert Podhorszki, Ebru Bozdağ, and Jeroen Tromp. An Adaptable Seismic Data Format. *Geophysical Journal International*, 207(2):1003–1011, November 2016. ISSN 0956-540X, 1365-246X. doi: 10.1093/gji/ggw319.
- Lion Krischer, Andreas Fichtner, Christian Boehm, and Heiner Igel. Automated Large-Scale Full Seismic Waveform Inversion for North America and the North Atlantic. *Journal of Geophysical Research: Solid Earth*, 123(7):5902–5928, 2018. ISSN 2169-9356. doi: 10.1029/2017JB015289.
- B. Kustowski, G. Ekström, and A. M. Dziewoński. Anisotropic shear-wave velocity structure of the Earth’s mantle: A global model. *Journal of Geophysical Research*, 113(B6):B06306, June 2008a. ISSN 0148-0227. doi: 10.1029/2007JB005169.
- B. Kustowski, G. Ekström, and A. M. Dziewoński. The shear-wave velocity structure in the upper mantle beneath Eurasia. *Geophysical Journal International*, 174(3):978–992, September 2008b. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.2008.03865.x.
- Gabi Laske, Guy Masters, Zhitu Ma, and Mike Pasyanos. Update on CRUST1. 0—A 1-degree global model of Earth’s crust. In *Geophys. res. abstr*, volume 15, page 2658, 2013.
- Jianshe Lei. Upper-mantle tomography and dynamics beneath the North China Craton. *Journal of Geophysical Research: Solid Earth*, 117(B6), June 2012. ISSN 01480227. doi: 10.1029/2012JB009212.
- Jianshe Lei and Dapeng Zhao. P-wave tomography and origin of the Changbai intraplate volcano in Northeast Asia. *Tectonophysics*, 397(3-4):281–295, March 2005. ISSN 00401951. doi: 10.1016/j.tecto.2004.12.009.
- Jianshe Lei and Dapeng Zhao. Teleseismic P-wave tomography and mantle dynamics beneath

- Eastern Tibet. *Geochemistry, Geophysics, Geosystems*, 17(5):1861–1884, May 2016. ISSN 1525-2027, 1525-2027. doi: 10.1002/2016GC006262.
- Jianshe Lei, Dapeng Zhao, Bernhard Steinberger, Bateer Wu, Fanluan Shen, and Zhixiong Li. New seismic constraints on the upper mantle structure of the Hainan plume. *Physics of the Earth and Planetary Interiors*, 173(1-2):33–50, March 2009. ISSN 00319201. doi: 10.1016/j.pepi.2008.10.013.
- Jianshe Lei, Furen Xie, Qicheng Fan, and M. Santosh. Seismic imaging of the deep structure under the Chinese volcanoes: An overview. *Physics of the Earth and Planetary Interiors*, 224:104–123, November 2013. ISSN 00319201. doi: 10.1016/j.pepi.2013.08.008.
- Jianshe Lei, Dapeng Zhao, Xiwei Xu, Yi-Gang Xu, and Mofei Du. Is there a big mantle wedge under eastern Tibet? *Physics of the Earth and Planetary Interiors*, 292:100–113, July 2019. ISSN 00319201. doi: 10.1016/j.pepi.2019.04.005.
- Wenjia Lei, Youyi Ruan, Ebru Bozdağ, Daniel Peter, Matthieu Lefebvre, Dimitri Komatitsch, Jeroen Tromp, Judith Hill, Norbert Podhorszki, and David Pugmire. Global adjoint tomography—model GLAD-M25. *Geophysical Journal International*, 223(1):1–21, October 2020. ISSN 0956-540X, 1365-246X. doi: 10.1093/gji/ggaa253.
- Chang Li and Robert D. Van Der Hilst. Structure of the upper mantle and transition zone beneath Southeast Asia from traveltimes tomography. *Journal of Geophysical Research*, 115(B7):B07308, July 2010. ISSN 0148-0227. doi: 10.1029/2009JB006882.
- Chang Li, Robert D. Van Der Hilst, E. Robert Engdahl, and Scott Burdick. A new global model for *P* wave speed variations in Earth’s mantle. *Geochemistry, Geophysics, Geosystems*, 9(5), May 2008. ISSN 15252027. doi: 10.1029/2007GC001806.
- Guoliang Li, Haichao Chen, Fenglin Niu, Zhen Guo, Yingjie Yang, and Jun Xie. Measurement of Rayleigh wave ellipticity and its application to the joint inversion of high-resolution *S* wave velocity structure beneath northeast China. *Journal of Geophysical Research: Solid Earth*, 121(2):864–880, February 2016. ISSN 2169-9313, 2169-9356. doi: 10.1002/2015JB012459.
- Shaohua Li, Jiaqi Li, Thomas P. Ferrand, Tong Zhou, Mingda Lv, Ziyi Xi, Ross Maguire, Guangjie Han, Juan Li, Xiyuan Bao, Yiran Jiang, and Tiezhao Bao. Deep Geophysical Anomalies Beneath the Changbaishan Volcano. *Journal of Geophysical Research: Solid Earth*, 128(4):e2022JB025671, April 2023. ISSN 2169-9313, 2169-9356. doi: 10.1029/2022JB025671.
- Hobin Lim, YoungHee Kim, Teh-Ru Alex Song, and Xuzhang Shen. Measurement of seismometer orientation using the tangential P-wave receiver function based on harmonic decomposition. *Geophysical Journal International*, 212(3):1747–1765, March 2018. ISSN 0956-540X, 1365-246X. doi: 10.1093/gji/ggx515.
- Min Liu, Miao Zhang, Weiqiang Zhu, William L. Ellsworth, and Hongyi Li. Rapid Characterization

- of the July 2019 Ridgecrest, California, Earthquake Sequence From Raw Seismic Data Using Machine-Learning Phase Picker. *Geophysical Research Letters*, 47(4):e2019GL086189, 2020. ISSN 1944-8007. doi: 10.1029/2019GL086189.
- Min Liu, Lu Li, Miao Zhang, Xinglin Lei, Mladen R. Nedimović, Alexandre P. Plourde, Rumeng Guo, Weitao Wang, and Hongyi Li. Complexity of initiation and evolution of the 2013 Yunlong earthquake swarm. *Earth and Planetary Science Letters*, 612:118168, June 2023. ISSN 0012-821X. doi: 10.1016/j.epsl.2023.118168.
- Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization, January 2019. arXiv:1711.05101 [cs, math].
- Chang Lu, Stephen P. Grand, Hongyu Lai, and Edward J. Garnero. TX2019slab: A New P and S Tomography Model Incorporating Subducting Slabs. *Journal of Geophysical Research: Solid Earth*, 124(11):11549–11567, 2019. ISSN 2169-9356. doi: 10.1029/2019JB017448.
- Jincheng Ma, Hans-Peter Bunge, Solvi Thrastarson, Andreas Fichtner, Dirk-Philip van Herwaarden, You Tian, Sung-Joon Chang, and Tingting Liu. Seismic full-waveform inversion of the crust-mantle structure beneath China and adjacent regions. *Journal of Geophysical Research: Solid Earth*, 127(9), 2022.
- Alessia Maggi, Carl Tape, Min Chen, Daniel Chao, and Jeroen Tromp. An automated time-window selection algorithm for seismic tomography. *Geophysical Journal International*, 178(1):257–281, July 2009. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.2009.04099.x.
- Toru Matsuzawa, Norihito Umino, Akira Hasegawa, and Akio Takagi. Upper mantle velocity structure estimated from PS-converted wave beneath the north-eastern Japan Arc. *Geophysical Journal International*, 86(3):767–787, 1986. doi: 10.1111/j.1365-246X.1986.tb00659.x. Publisher: Blackwell Publishing Ltd Oxford, UK.
- Toru Matsuzawa, Toshio Kono, Akira Hasegawa, and Akio Takagi. Subducting plate boundary beneath the northeastern Japan arc estimated from SP converted waves. *Tectonophysics*, 181(1-4):123–133, 1990. doi: 10.1016/0040-1951(90)90012-W. Publisher: Elsevier.
- Sarah E. Mazza, Esteban Gazel, Michael Bizimis, Robert Moucha, Paul Béguelin, Elizabeth A. Johnson, Ryan J. McAleer, and Alexander V. Sobolev. Sampling the volatile-rich transition zone beneath Bermuda. *Nature*, 569(7756):398–403, May 2019. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-019-1183-6.
- Ian W. McBrearty and Gregory C. Beroza. Earthquake phase association with graph neural networks. *Bulletin of the Seismological Society of America*, 113(2):524–547, 2023. Publisher: Seismological Society of America.
- Jeffrey J. McGuire, Douglas A. Wiens, Patrick J. Shore, and Michael G. Bevis. The March 9, 1994 (M_w 7.6), deep Tonga earthquake: Rupture outside the seismically active slab. *Journal of*

- Geophysical Research: Solid Earth*, 102(B7):15163–15182, July 1997. ISSN 0148-0227. doi: 10.1029/96JB03185.
- Anne Meltzer, Joshua C. Stachnik, Demberel Sodnomsambuu, Ulziibat Munkhuu, Baasanbat Tsagaan, Mungunsuren Dashdondog, and Raymond Russo. The Central Mongolia Seismic Experiment: Multiple Applications of Temporary Broadband Seismic Arrays. *Seismological Research Letters*, 90(3):1364–1376, May 2019. ISSN 0895-0695, 1938-2057. doi: 10.1785/0220180360.
- Raffaella Montelli, Guust Nolet, F. A. Dahlen, Guy Masters, E. Robert Engdahl, and Shu-Huei Hung. Finite-Frequency Tomography Reveals a Variety of Plumes in the Mantle. *Science*, 303(5656):338–343, January 2004. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1092485.
- S. Mostafa Mousavi and Gregory C. Beroza. Deep-learning seismology. *Science*, 377(6607): eabm4470, August 2022. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abm4470.
- S. Mostafa Mousavi and Gregory C. Beroza. Machine Learning in Earthquake Seismology. *Annual Review of Earth and Planetary Sciences*, 51(1):105–129, May 2023. ISSN 0084-6597, 1545-4495. doi: 10.1146/annurev-earth-071822-100323.
- S. Mostafa Mousavi, Yixiao Sheng, Weiqiang Zhu, and Gregory C. Beroza. STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. *IEEE Access*, 7:179464–179476, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2947848. Conference Name: IEEE Access.
- S. Mostafa Mousavi, William L. Ellsworth, Weiqiang Zhu, Lindsay Y. Chuang, and Gregory C. Beroza. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1):3952, August 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-17591-w. Number: 1 Publisher: Nature Publishing Group.
- Nori Nakata and David R. Shelly. Imaging a Crustal Low-Velocity Layer Using Reflected Seismic Waves From the 2014 Earthquake Swarm at Long Valley Caldera, California: The Magmatic System Roof? *Geophysical Research Letters*, 45(8):3481–3488, April 2018. ISSN 0094-8276, 1944-8007. doi: 10.1029/2018GL077260.
- Peter L. Nelson and Stephen P. Grand. Lower-mantle plume beneath the Yellowstone hotspot revealed by core waves. *Nature Geoscience*, 11(4):280–284, April 2018. ISSN 1752-0894, 1752-0908. doi: 10.1038/s41561-018-0075-y.
- Fenglin Niu and Juan Li. Component azimuths of the CEArray stations estimated from P-wave particle motion. *Earthquake Science*, 24(1):3–13, February 2011. ISSN 1674-4519, 1867-8777. doi: 10.1007/s11589-011-0764-8.
- Masayuki Obayashi, Junko Yoshimitsu, Guust Nolet, Yoshio Fukao, Hajime Shiobara, Hiroko Sugiooka, Hiroki Miyamachi, and Yuan Gao. Finite frequency whole mantle P wave tomography: Im-

- provement of subducted slab images. *Geophysical Research Letters*, 40(21):5652–5657, November 2013. ISSN 00948276. doi: 10.1002/2013GL057401.
- Mathias Obrebski, Richard M. Allen, Fengxue Zhang, Jiatie Pan, Qingju Wu, and Shu-Huei Hung. Shear wave tomography of China using joint inversion of body and surface wave constraints. *Journal of Geophysical Research: Solid Earth*, 117(B1), January 2012. ISSN 01480227. doi: 10.1029/2011JB008349.
- Shiro Ohmi and Sadaki Hori. Seismic wave conversion near the upper boundary of the Pacific plate beneath the Kanto district, Japan. *Geophysical Journal International*, 141(1):136–148, April 2000. ISSN 0956-540X. doi: 10.1046/j.1365-246X.2000.00086.x.
- Yoshimitsu Okada, Keiji Kasahara, Sadaki Hori, Kazushige Obara, Shoji Sekiguchi, Hiroyuki Fujiwara, and Akira Yamamoto. Recent progress of seismic observation networks in Japan —Hi-net, F-net, K-NET and KiK-net—. *Earth, Planets and Space*, 56(8):xv–xxviii, August 2004. ISSN 1880-5981. doi: 10.1186/BF03353076.
- Mark Panning and Barbara Romanowicz. A three-dimensional radially anisotropic model of shear velocity in the whole mantle. *Geophysical Journal International*, 167(1):361–379, October 2006. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.2006.03100.x.
- Yongsoo Park, S. Mostafa Mousavi, Weiqiang Zhu, William L. Ellsworth, and Gregory C. Beroza. Machine-Learning-Based Analysis of the Guy-Greenbrier, Arkansas Earthquakes: A Tale of Two Sequences. *Geophysical Research Letters*, 47(6):e2020GL087032, 2020. ISSN 1944-8007. doi: 10.1029/2020GL087032.
- J. D. Pesicek, C. H. Thurber, H. Zhang, H. R. DeShon, E. R. Engdahl, and S. Widiyantoro. Teleseismic double-difference relocation of earthquakes along the Sumatra-Andaman subduction zone using a 3-D model. *Journal of Geophysical Research*, 115(B10):B10303, October 2010. ISSN 0148-0227. doi: 10.1029/2010JB007443.
- Daniel Peter, Dimitri Komatitsch, Yang Luo, Roland Martin, Nicolas Le Goff, Emanuele Casarotti, Pieyre Le Loher, Federica Magnoni, Qinya Liu, Céline Blitz, Tarje Nissen-Meyer, Piero Basini, and Jeroen Tromp. Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes. *Geophysical Journal International*, 186(2):721–739, August 2011. ISSN 0956540X. doi: 10.1111/j.1365-246X.2011.05044.x.
- E. Polak and G. Ribiere. Note sur la convergence de méthodes de directions conjuguées. *Revue française d’informatique et de recherche opérationnelle. Série rouge*, 3(16):35–43, 1969. ISSN 0373-8000. doi: 10.1051/m2an/196903R100351.
- M. J. D. Powell and Professor of Applied Numerical Analysis M. J. D. Powell. *Approximation Theory and Methods*. Cambridge University Press, March 1981. ISBN 978-0-521-29514-7. Google-Books-ID: ODZ1OYR3w4cC.

- Florian Rickers, Andreas Fichtner, and Jeannot Trampert. Imaging mantle plumes with instantaneous phase measurements of diffracted waves. *Geophysical Journal International*, 190(1): 650–664, August 2012. ISSN 0956-540X. doi: 10.1111/j.1365-246X.2012.05515.x.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Lecture Notes in Computer Science, pages 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4. doi: 10.1007/978-3-319-24574-4_28.
- Zachary E. Ross, Men-Andrin Meier, Egill Hauksson, and Thomas H. Heaton. Generalized Seismic Phase Detection with Deep Learning. *Bulletin of the Seismological Society of America*, 108(5A): 2894–2901, August 2018. ISSN 0037-1106. doi: 10.1785/0120180080.
- Zachary E. Ross, Elizabeth S. Cochran, Daniel T. Trugman, and Jonathan D. Smith. 3D fault architecture controls the dynamism of earthquake swarms. *Science*, 368(6497):1357–1361, June 2020. doi: 10.1126/science.abb0779. Publisher: American Association for the Advancement of Science.
- Youyi Ruan, Wenjie Lei, Ryan Modrak, Rıdvan Örsvuran, Ebru Bozdağ, and Jeroen Tromp. Balancing unevenly distributed data in seismic tomography: a global adjoint tomography example. *Geophysical Journal International*, 219(2):1225–1236, November 2019. ISSN 0956-540X, 1365-246X. doi: 10.1093/gji/ggz356.
- Larry Ruff and Hiroo Kanamori. Seismicity and the subduction process. *Physics of the Earth and Planetary Interiors*, 23(3):240–252, October 1980. ISSN 0031-9201. doi: 10.1016/0031-9201(80)90117-X.
- Takahiro Shiina, Junichi Nakajima, and Toru Matsuzawa. Seismic evidence for high pore pressures in the oceanic crust: Implications for fluid-related embrittlement. *Geophysical Research Letters*, 40(10):2006–2010, May 2013. ISSN 0094-8276, 1944-8007. doi: 10.1002/grl.50468.
- Takahiro Shiina, Junichi Nakajima, Genti Toyokuni, and Toru Matsuzawa. Guided wave observations and evidence for the low-velocity subducting crust beneath Hokkaido, northern Japan. *Earth, Planets and Space*, 66(1):69, December 2014. ISSN 1880-5981. doi: 10.1186/1880-5981-66-69.
- Takahiro Shiina, Junichi Nakajima, Toru Matsuzawa, Genti Toyokuni, and Saeko Kita. Depth variations in seismic velocity in the subducting crust: Evidence for fluid-related embrittlement for intermediate-depth earthquakes. *Geophysical Research Letters*, 44(2):810–817, January 2017. ISSN 0094-8276, 1944-8007. doi: 10.1002/2016GL071798.
- N A Simmons, S C Myers, C Morency, A Chiang, and D R Knapp. SPiRaL: a multiresolution global tomography model of seismic wave speeds and radial anisotropy variations in the crust and mantle. *Geophysical Journal International*, 227(2):1366–1391, November 2021. ISSN

0956-540X. doi: 10.1093/gji/ggab277.

Reinoud Sleeman and Torild van Eck. 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):265–275, June 1999. ISSN 0031-9201. doi: 10.1016/S0031-9201(99)00007-2.

Smithsonian-Institution. Global Volcanism Program, 2023. [Database] Volcanoes of the World (v. 5.0.4; 17 Apr 2023), April 2023. Distributed by Smithsonian Institution, compiled by Venzke, E.

Paul Spudich and Todd Bostwick. Studies of the seismic coda using an earthquake cluster as a deeply buried seismograph array. *Journal of Geophysical Research: Solid Earth*, 92(B10): 10526–10546, September 1987. ISSN 0148-0227. doi: 10.1029/JB092iB10p10526.

Richard Stacey. Improved transparent boundary formulations for the elastic-wave equation. *Bulletin of the Seismological Society of America*, 78(6):2089–2097, December 1988. ISSN 1943-3573, 0037-1106. doi: 10.1785/BSSA0780062089.

Yen Joe Tan, Felix Waldhauser, William L. Ellsworth, Miao Zhang, Weiqiang Zhu, Maddalena Michele, Lauro Chiaraluce, Gregory C. Beroza, and Margarita Segou. Machine-Learning-Based High-Resolution Earthquake Catalog Reveals How Complex Fault Structures Were Activated during the 2016–2017 Central Italy Sequence. *The Seismic Record*, 1(1):11–19, May 2021. ISSN 2694-4006. doi: 10.1785/0320210001.

Youcai Tang, Masayuki Obayashi, Fenglin Niu, Stephen P. Grand, Yongshun John Chen, Hitoshi Kawakatsu, Satoru Tanaka, Jieyuan Ning, and James F. Ni. Changbaishan volcanism in northeast China linked to subduction-induced mantle upwelling. *Nature Geoscience*, 7(6):470–475, June 2014. ISSN 1752-0894, 1752-0908. doi: 10.1038/ngeo2166.

Kai Tao, Stephen P. Grand, and Fenglin Niu. Full-waveform inversion of triplicated data using a normalized-correlation-coefficient-based misfit function. *Geophysical Journal International*, 210(3):1517–1524, September 2017. ISSN 0956-540X, 1365-246X. doi: 10.1093/gji/ggx249.

Kai Tao, Stephen P. Grand, and Fenglin Niu. Seismic Structure of the Upper Mantle Beneath Eastern Asia From Full Waveform Seismic Tomography. *Geochemistry, Geophysics, Geosystems*, 19(8): 2732–2763, August 2018. ISSN 1525-2027, 1525-2027. doi: 10.1029/2018GC007460.

Carl Tape, Qinya Liu, and Jeroen Tromp. Finite-frequency tomography using adjoint methods—Methodology and examples using membrane surface waves. *Geophysical Journal International*, 168(3):1105–1129, March 2007. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.2006.03191.x.

Carl Tape, Qinya Liu, Alessia Maggi, and Jeroen Tromp. Adjoint Tomography of the Southern California Crust. *Science*, 325(5943):988–992, August 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1175298.

- Carl Tape, Qinya Liu, Alessia Maggi, and Jeroen Tromp. Seismic tomography of the southern California crust based on spectral-element and adjoint methods. *Geophysical Journal International*, 180(1):433–462, January 2010. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.2009.04429.x.
- P. Tapponnier, G. Peltzer, and R. Armijo. On the mechanics of the collision between India and Asia. *Geological Society, London, Special Publications*, 19(1):113–157, January 1986. ISSN 0305-8719, 2041-4927. doi: 10.1144/GSL.SP.1986.019.01.07.
- You Tian, Dapeng Zhao, Ruomei Sun, and Jiwen Teng. Seismic imaging of the crust and upper mantle beneath the North China Craton. *Physics of the Earth and Planetary Interiors*, 172(3-4): 169–182, February 2009. ISSN 00319201. doi: 10.1016/j.pepi.2008.09.002.
- Ron van der Hilst. Complex morphology of subducted lithosphere in the mantle beneath the Tonga trench. *Nature*, 374(6518):154–157, 1995. doi: 10.1038/374154a0. Publisher: Nature Publishing Group UK London.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- Suyun Wang, Fenglin Niu, and Guomin Zhang. Velocity structure of the uppermost mantle beneath East Asia from *P_n* tomography and its dynamic implications. *Journal of Geophysical Research: Solid Earth*, 118(1):290–301, January 2013. ISSN 21699313. doi: 10.1002/jgrb.50085.
- Xin Wang, Qi-Fu Chen, Fenglin Niu, Shengji Wei, Jieyuan Ning, Juan Li, Weijun Wang, Johannes Buchen, and Lijun Liu. Distinct slab interfaces imaged within the mantle transition zone. *Nature Geoscience*, 13(12):822–827, December 2020. ISSN 1752-0894, 1752-0908. doi: 10.1038/s41561-020-00653-5.
- Yu Wang, Xuemin Zhang, Chaosong Jiang, Haiquan Wei, and Jinglin Wan. Tectonic controls on the late Miocene–Holocene volcanic eruptions of the Tengchong volcanic field along the southeastern margin of the Tibetan plateau. *Journal of Asian Earth Sciences*, 30(2):375–389, April 2007. ISSN 13679120. doi: 10.1016/j.jseaes.2006.11.005.
- Michael Warner, Andrew Ratcliffe, Tenice Nangoo, Joanna Morgan, Adrian Umpleby, Nikhil Shah, Vetle Vinje, Ivan Štekl, Lluís Guasch, Caroline Win, Graham Conroy, and Alexandre Bertrand. Anisotropic 3D full-waveform inversion. *GEOPHYSICS*, 78(2):R59–R80, March 2013. ISSN 0016-8033, 1942-2156. doi: 10.1190/geo2012-0338.1.
- Spahr C. Webb and Wayne C. Crawford. Long-period seafloor seismology and deformation under ocean waves. *Bulletin of the Seismological Society of America*, 89(6):1535–1542, 1999. Publisher: The Seismological Society of America.
- S. Shawn Wei and Y. John Chen. Seismic evidence of the Hainan mantle plume by receiver function

- analysis in southern China. *Geophysical Research Letters*, 43(17):8978–8985, September 2016. ISSN 00948276. doi: 10.1002/2016GL069513.
- S. Shawn Wei, Douglas A. Wiens, Yang Zha, Terry Plank, Spahr C. Webb, Donna K. Blackman, Robert A. Dunn, and James A. Conder. Seismic evidence of effects of water on melt transport in the Lau back-arc mantle. *Nature*, 518(7539):395–398, 2015. Publisher: Nature Publishing Group UK London.
- S. Shawn Wei, Yang Zha, Weisen Shen, Douglas A. Wiens, James A. Conder, and Spahr C. Webb. Upper mantle structure of the Tonga-Lau-Fiji region from Rayleigh wave tomography. *Geochemistry, Geophysics, Geosystems*, 17(11):4705–4724, November 2016. ISSN 1525-2027, 1525-2027. doi: 10.1002/2016GC006656.
- S. Shawn Wei, Douglas A. Wiens, Peter E. van Keken, and Chen Cai. Slab temperature controls on the Tonga double seismic zone and slab mantle dehydration. *Science Advances*, 3(1):e1601755, January 2017. doi: 10.1126/sciadv.1601755. Publisher: American Association for the Advancement of Science.
- S. Shawn Wei, Philipp Ruprecht, Sydney L. Gable, Ellyn G. Huggins, Natalia Ruppert, Lei Gao, and Haijiang Zhang. Along-strike variations in intermediate-depth seismicity and arc magmatism along the Alaska Peninsula. *Earth and Planetary Science Letters*, 563:116878, 2021. Publisher: Elsevier.
- Songqiao Wei, Y. John Chen, Eric Sandvol, Shiyong Zhou, Han Yue, Ge Jin, Thomas M. Hearn, Mingming Jiang, Haiyang Wang, Wenyan Fan, Zheng Liu, Zengxi Ge, Yanbin Wang, Yongge Feng, and James Ni. Regional earthquakes in northern Tibetan Plateau: Implications for lithospheric strength in Tibet. *Geophysical Research Letters*, 37(19):n/a–n/a, October 2010. ISSN 00948276. doi: 10.1029/2010GL044800.
- Wei Wei, Jiandong Xu, Dapeng Zhao, and Yaolin Shi. East Asia mantle tomography: New insight into plate subduction and intraplate volcanism. *Journal of Asian Earth Sciences*, 60:88–103, October 2012. ISSN 13679120. doi: 10.1016/j.jseaes.2012.08.001.
- J Weston, E R Engdahl, J Harris, D Di Giacomo, and D A Storchak. ISC-EHB: reconstruction of a robust earthquake data set. *Geophysical Journal International*, 214(1):474–484, July 2018. ISSN 0956-540X, 1365-246X. doi: 10.1093/gji/ggy155.
- Douglas A. Wiens, Jeffrey J. McGuire, and Patrick J. Shore. Evidence for transformational faulting from a deep double seismic zone in Tonga. *Nature*, 364(6440):790–793, 1993. Publisher: Nature Publishing Group UK London.
- Douglas A. Wiens, Jeffrey J. McGuire, Patrick J. Shore, Michael G. Bevis, Kitone Draunidalo, Gajendra Prasad, and Saimon P. Helu. A deep earthquake aftershock sequence and implications for the rupture mechanism of deep earthquakes. *Nature*, 372(6506):540–543, 1994. Publisher: Nature Publishing Group UK London.

- John D. Wilding, Weiqiang Zhu, Zachary E. Ross, and Jennifer M. Jackson. The magmatic web beneath Hawai'i. *Science*, 379(6631):462–468, February 2023. doi: 10.1126/science.ade5755. Publisher: American Association for the Advancement of Science.
- M. Witek, S.-J. Chang, D. Y. Lim, S. Ning, and J. Ning. Radial Anisotropy in East Asia From Multimode Surface Wave Tomography. *Journal of Geophysical Research: Solid Earth*, 126(7): e2020JB021201, 2021. ISSN 2169-9356. doi: 10.1029/2020JB021201.
- Michael Witek, Suzan van der Lee, and Tae-Seob Kang. Rayleigh wave group velocity distributions for East Asia using ambient seismic noise. *Geophysical Research Letters*, 41(22):8045–8052, November 2014. ISSN 00948276. doi: 10.1002/2014GL062016.
- Cecily J. Wolfe, Sean C. Solomon, Gabi Laske, John A. Collins, Robert S. Detrick, John A. Orcutt, David Bercovici, and Erik H. Hauri. Mantle Shear-Wave Velocity Structure Beneath the Hawaiian Hot Spot. *Science*, 326(5958):1388–1390, December 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1180165.
- Jonny Wu, John Suppe, Renqi Lu, and Ravi Kanda. Philippine Sea and East Asian plate tectonics since 52 Ma constrained by new subducted slab reconstruction methods: PHILIPPINE SEA PLATE TECTONICS-SLABS. *Journal of Geophysical Research: Solid Earth*, 121(6):4670–4741, June 2016. ISSN 21699313. doi: 10.1002/2016JB012923.
- Ziyi Xi, Songqiao Shawn Wei, Weiqiang Zhu, Greg Beroza, Yaqi Jie, and Nooshin Saloor. Deep Learning for Deep Earthquakes: Insights from OBS Observations of the Tonga Subduction Zone. 2023. doi: 10.31223/X5C105. Publisher: EarthArXiv.
- Jiayi Xie, Michael H. Ritzwoller, Weisen Shen, Yingjie Yang, Yong Zheng, and Longquan Zhou. Crustal radial anisotropy across Eastern Tibet and the Western Yangtze Craton. *Journal of Geophysical Research: Solid Earth*, 118(8):4226–4252, 2013. ISSN 2169-9356. doi: 10.1002/jgrb.50296.
- Yi-Gang Xu, Jin-Long Ma, Frederick A. Frey, Mark D. Feigenson, and Jiang-Feng Liu. Role of lithosphere–asthenosphere interaction in the genesis of Quaternary alkali and tholeiitic basalts from Datong, western North China Craton. *Chemical Geology*, 224(4):247–271, December 2005. ISSN 00092541. doi: 10.1016/j.chemgeo.2005.08.004.
- Tadashi Yamasaki and Tetsuzo Seno. Double seismic zone and dehydration embrittlement of the subducting slab. *Journal of Geophysical Research: Solid Earth*, 108(B4):2002JB001918, April 2003. ISSN 0148-0227. doi: 10.1029/2002JB001918.
- Yu Yang, Jianshe Lei, Yinshuang Ai, Guangwei Zhang, Changqing Sun, Enbo Fan, Long Li, Qi Mi, Mingwen Lu, Jing He, Jian Wang, Mofei Du, Bing Zhang, Fanfan Tian, Chen Ma, and Zemin Liu. Crustal structure beneath Northeast China from ambient noise tomography. *Physics of the Earth and Planetary Interiors*, 293:106257, August 2019. ISSN 00319201. doi: 10.1016/j.pepi.2019.04.008.

- Huajian Yao, Robert D. Van Der Hilst, and Maarten V. De Hoop. Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis - I. Phase velocity maps. *Geophysical Journal International*, 166(2):732–744, August 2006. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.2006.03028.x.
- Huajian Yao, Caroline Beghein, and Robert D. Van Der Hilst. Surface wave array tomography in SE Tibet from ambient seismic noise and two-station analysis - II. Crustal and upper-mantle structure. *Geophysical Journal International*, 173(1):205–219, April 2008. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.2007.03696.x.
- An Yin and T. Mark Harrison. Geologic Evolution of the Himalayan-Tibetan Orogen. *Annual Review of Earth and Planetary Sciences*, 28(1):211–280, May 2000. ISSN 0084-6597, 1545-4495. doi: 10.1146/annurev.earth.28.1.211.
- Yanhua O Yuan, Ebru Bozdağ, Caio Ciardelli, Fuchun Gao, and F J Simons. The exponentiated phase measurement, and objective-function hybridization for adjoint waveform tomography. *Geophysical Journal International*, 221(2):1145–1164, May 2020. ISSN 0956-540X. doi: 10.1093/gji/ggaa063.
- Han Yue, Y. John Chen, Eric Sandvol, James Ni, Thomas Hearn, Shiyong Zhou, Yongge Feng, Zengxi Ge, Andrea Trujillo, Yanbin Wang, Ge Jin, Mingming Jiang, Youcai Tang, Xiaofeng Liang, Songqiao Wei, Haiyang Wang, Wenyuan Fan, and Zheng Liu. Lithospheric and upper mantle structure of the northeastern Tibetan Plateau. *Journal of Geophysical Research: Solid Earth*, 117(B5), May 2012. ISSN 01480227. doi: 10.1029/2011JB008545.
- Zhongwen Zhan. Mechanisms and Implications of Deep Earthquakes. *Annual Review of Earth and Planetary Sciences*, 48(1):147–174, May 2020. ISSN 0084-6597, 1545-4495. doi: 10.1146/annurev-earth-053018-060314.
- Haijiang Zhang, Fan Wang, Robert Myhill, and Hao Guo. Slab morphology and deformation beneath Izu-Bonin. *Nature Communications*, 10(1):1310, March 2019a. ISSN 2041-1723. doi: 10.1038/s41467-019-09279-7.
- K.-J Zhang. North and South China collision along the eastern and southern North China margins. *Tectonophysics*, 270(1-2):145–156, February 1997. ISSN 00401951. doi: 10.1016/S0040-1951(96)00208-9.
- Maoliang Zhang, Zhengfu Guo, Jiaqi Liu, Guoming Liu, Lihong Zhang, Ming Lei, Wenbin Zhao, Lin Ma, Vincenzo Sepe, and Guido Ventura. The intraplate Changbaishan volcanic field (China/North Korea): A review on eruptive history, magma genesis, geodynamic significance, recent dynamics and potential hazards. *Earth-Science Reviews*, 187:19–52, December 2018. ISSN 00128252. doi: 10.1016/j.earscirev.2018.07.011.
- Miao Zhang, William L. Ellsworth, and Gregory C. Beroza. Rapid Earthquake Association and Location. *Seismological Research Letters*, 90(6):2276–2284, September 2019b. ISSN 0895-

0695. doi: 10.1785/0220190052.

Dapeng Zhao, Toru Matsuzawa, and Akira Hasegawa. Morphology of the subducting slab boundary in the northeastern Japan arc. *Physics of the Earth and Planetary Interiors*, 102(1):89–104, June 1997a. ISSN 0031-9201. doi: 10.1016/S0031-9201(96)03258-X.

Dapeng Zhao, Yingbiao Xu, Douglas A. Wiens, LeRoy Dorman, John Hildebrand, and Spahr Webb. Depth Extent of the Lau Back-Arc Spreading Center and Its Relation to Subduction Processes. *Science*, 278(5336):254–257, October 1997b. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.278.5336.254.

Dapeng Zhao, Shigenori Maruyama, and Soichi Omori. Mantle dynamics of Western Pacific and East Asia: Insight from seismic tomography and mineral physics. *Gondwana Research*, 11(1-2): 120–131, January 2007. ISSN 1342937X. doi: 10.1016/j.gr.2006.06.006.

Dapeng Zhao, You Tian, Jianshe Lei, Lucy Liu, and Sihua Zheng. Seismic image and origin of the Changbai intraplate volcano in East Asia: Role of big mantle wedge above the stagnant Pacific slab. *Physics of the Earth and Planetary Interiors*, 173(3-4):197–206, April 2009. ISSN 00319201. doi: 10.1016/j.pepi.2008.11.009.

Tianyu Zheng, Ling Chen, Liang Zhao, Weiwei Xu, and Rixiang Zhu. Crust–mantle structure difference across the gravity gradient zone in North China Craton: Seismic image of the thinned continental crust. *Physics of the Earth and Planetary Interiors*, 159(1-2):43–58, November 2006. ISSN 00319201. doi: 10.1016/j.pepi.2006.05.004.

Xiu-Fen Zheng, Biao Ouyang, Dong-Ning Zhang, Zhi-Xiang Yao, Jian-Hong Liang, and Jie Zheng. Construction of the Technical System of Data Backup Centre for China Seismograph Network and Its Support to Wenchuan Earthquake Researches. *Chinese Journal of Geophysics*, 52(3): 657–662, May 2009. ISSN 08989591. doi: 10.1002/cjg2.1387.

Quan Zhou, Lijun Liu, and Jiashun Hu. Western US volcanism due to intruding oceanic mantle driven by ancient Farallon slabs. *Nature Geoscience*, 11(1):70–76, January 2018. ISSN 1752-0894, 1752-0908. doi: 10.1038/s41561-017-0035-y.

Tong Zhou, Ziyi Xi, Min Chen, and Jiaqi Li. Assessment of seismic tomographic models of the contiguous United States using intermediate-period 3-D wavefield simulation. *Geophysical Journal International*, 228(2):1392–1409, November 2021. ISSN 0956-540X, 1365-246X. doi: 10.1093/gji/ggab406.

Tong Zhou, Jiaqi Li, Ziyi Xi, Guoliang Li, and Min Chen. CUSRA2021: A Radially Anisotropic Model of the Contiguous US and Surrounding Regions by Full-Waveform Inversion. *Journal of Geophysical Research: Solid Earth*, 127(8), August 2022. ISSN 2169-9313, 2169-9356. doi: 10.1029/2021JB023893.

X Zhou and R Armstrong. Cenozoic volcanic rocks of eastern China — secular and geographic

- trends in chemistry and strontium isotopic composition. *Earth and Planetary Science Letters*, 58(3):301–329, May 1982. ISSN 0012821X. doi: 10.1016/0012-821X(82)90083-8.
- Ying Zhou. Anomalous mantle transition zone beneath the Yellowstone hotspot track. *Nature Geoscience*, 11(6):449–453, June 2018. ISSN 1752-0894, 1752-0908. doi: 10.1038/s41561-018-0126-4.
- Hejun Zhu, Ebru Bozdağ, Daniel Peter, and Jeroen Tromp. Seismic wavespeed images across the Iapetus and Tornquist suture zones. *Geophysical Research Letters*, 39(18), September 2012a. ISSN 00948276. doi: 10.1029/2012GL053053.
- Hejun Zhu, Ebru Bozdağ, Daniel Peter, and Jeroen Tromp. Structure of the European upper mantle revealed by adjoint tomography. *Nature Geoscience*, 5(7):493–498, July 2012b. ISSN 1752-0894, 1752-0908. doi: 10.1038/ngeo1501.
- Hejun Zhu, Ebru Bozdağ, and Jeroen Tromp. Seismic structure of the European upper mantle based on adjoint tomography. *Geophysical Journal International*, 201(1):18–52, April 2015. ISSN 1365-246X, 0956-540X. doi: 10.1093/gji/ggu492.
- RiXiang Zhu, Ling Chen, FuYuan Wu, and JunLai Liu. Timing, scale and mechanism of the destruction of the North China Craton. *Science China Earth Sciences*, 54(6):789–797, June 2011. ISSN 1674-7313, 1869-1897. doi: 10.1007/s11430-011-4203-4.
- Weiqiang Zhu and Gregory C Beroza. PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273, January 2019. ISSN 0956-540X. doi: 10.1093/gji/ggy423.
- Weiqiang Zhu, Ian W. McBrearty, S. Mostafa Mousavi, William L. Ellsworth, and Gregory C. Beroza. Earthquake Phase Association Using a Bayesian Gaussian Mixture Model. *Journal of Geophysical Research: Solid Earth*, 127(5):e2021JB023249, 2022. ISSN 2169-9356. doi: 10.1029/2021JB023249.
- Weiqiang Zhu, Ettore Biondi, Jiaxuan Li, Jiuxun Yin, Zachary E. Ross, and Zhongwen Zhan. Seismic Arrival-time Picking on Distributed Acoustic Sensing Data using Semi-supervised Learning, March 2023a. arXiv:2302.08747 [physics].
- Weiqiang Zhu, Alvin Brian Hou, Robert Yang, Avoy Datta, S Mostafa Mousavi, William L Ellsworth, and Gregory C Beroza. QuakeFlow: a scalable machine-learning-based earthquake monitoring workflow with cloud computing. *Geophysical Journal International*, 232(1):684–693, January 2023b. ISSN 0956-540X. doi: 10.1093/gji/ggac355.
- Jiahui Zuo, Liwei Wang, and Fenglin Niu. Multiple source downwellings beneath eastern North China revealed by 3-D CCP migration of receiver function data. *Journal of Asian Earth Sciences*, 192:104266, May 2020. ISSN 13679120. doi: 10.1016/j.jseas.2020.104266.