Geophysical Journal International

Geophys. J. Int. (2022) **228,** 729–743
Advance Access publication 2021 September 04
GJI Seismology

https://doi.org/10.1093/gji/ggab361

Constraining the 410-km discontinuity and slab structure in the Kuril subduction zone with triplication waveforms

Jiaqi Li,[1] Min Chen,[1,2] Jieyuan Ning,[3] Tiezhao Bao,[3] Ross Maguire,[1,4] Megan P. Flanagan[5] and Tong Zhou[6]

[1]*Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824, USA. E-mail: lijiaqi9@msu.edu*
[2]*Department of Earth and Environmental Sciences, Michigan State University, East Lansing, MI 48824, USA*
[3]*School of Earth and Space Sciences, Peking University, Beijing 100871, China*
[4]*Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, NM 87131, USA*
[5]*EditSprings, Boston, MA 02445, USA*
[6]*Earth, Planetary and Space Sciences, University of California Los Angeles, CA 90095, USA*

Accepted 2021 September 2. Received 2021 August 20; in original form 2020 December 27

SUMMARY

The detailed structure near the 410-km discontinuity provides key constraints of the dynamic interactions between the upper mantle and the lower mantle through the mantle transition zone (MTZ) via mass and heat exchange. Meanwhile, the temperature of the subducting slab, which can be derived from its fast wave speed perturbation, is critical for understanding the mantle dynamics in subduction zones where the slab enters the MTZ. Multipathing, i.e. triplicated, body waves that bottom near the MTZ carry rich information of the 410-km discontinuity structure and can be used to constrain the discontinuity depth and radial variations of wave speeds across it. In this study, we systematically analysed the trade-off between model parameters in triplication studies using synthetic examples. Specifically, we illustrated the necessity of using array-normalized amplitude. Two 1-D depth profiles of the wave speed below the Tatar Strait of Russia in the Kuril subduction zone are obtained. We have observed triplications due to both the 410-km discontinuity and the slab upper surface. And, seismic structures for these two interfaces are simultaneously inverted. Our derived 410-km discontinuity depths for the northern and southern regions are at 420±15 and 425±15 km, respectively, with no observable uplift. The slab upper surface is inverted to be located about 50–70 km below the 410-km discontinuity. This location is between the depths of the 1 and 2 per cent *P*-wave speed perturbation contours of a regional 3-D full-waveform inversion (FWI) model, but we found twice the wave speed perturbation amplitude. A wave speed increase of 3.9–4.6 per cent within the slab, compared to 2.0–2.4 per cent from the 3-D FWI model, is necessary to fit the waveforms with the shortest period of 2 s, indicating that high-frequency waves are required to accurately resolve the detailed structures near the MTZ.

Key words: Waveform inversion; Body waves; Interface waves; Subduction zone processes.

1 INTRODUCTION

The 410-km discontinuity defines the top of the mantle transition zone (MTZ). This interface has long been attributed to the mineralogical phase change of olivine to wadsleyite at around 410 km depth, demonstrated by laboratory experimental evidence (Ringwood 1975). The detailed structures near the 410-km discontinuity provide key constraints on the dynamic interactions between the upper and the lower mantle through the MTZ, particularly via mass and heat exchange.

One of the typical interactions involves cold slabs penetrating and elevating the 410-km discontinuity and carrying volatiles into the transition zone (Kawakatsu & Watada 2007). In a pure thermal environment, the topography of the discontinuity can directly reflect the temperature perturbations. That is, the topography can serve as a thermometer in the deep mantle (Vidale & Benz 1992). The 410-km discontinuity transitional thickness (i.e. sharpness) is highly sensitive to the water content (Helffrich & Wood 1996; Van der Meijde *et al.* 2003), which also provides insight into the deep Earth's volatile budget (Thompson 1992).

This paper is dedicated to our co-author Min Chen who unexpectedly passed away before the acceptance of the manuscript.

Although deep-focus earthquakes and cold temperatures in the subducting slab are associated (Isacks & Molnar 1971; Molnar *et al.* 1979), the mechanism for deep-focus earthquakes is still unclear. Interaction between the 410-km discontinuity and the subducting slab can help address this important question. Specifically, the 410-km discontinuity topography inside the subducting slab could be used to infer the existence of a metastable olivine wedge, on the edge of which the transformational faulting from metastable olivine to wadsleyite may cause deep-focus earthquakes (Green II & Burnley 1989; Kirby *et al.* 1991).

To detect and further constrain the upper-mantle discontinuities, secondary seismic phases generated at the interface could be good candidates. The related methods can generally be classified into two categories: one is to use the reflected waves off the interfaces (e.g. Flanagan & Shearer 1998, 1999; Gu & Dziewonski 2002; Schmerr & Garnero 2007; Houser *et al.* 2008; Lawrence & Shearer 2008; Ritsema *et al.* 2009b; Wang *et al.* 2017; Li *et al.* 2019b; Guo & Zhou 2020; Tian *et al.* 2020; Wei *et al.* 2020) and the other is to use the converted wave upon transmissions at the discontinuities (e.g. Vinnik 1977; Collier & Helffrich 1997; Thirot *et al.* 1998; Chevrot *et al.* 1999; Niu *et al.* 2005; Ritsema *et al.* 2009a). Although these secondary phases provide direct constraints on the discontinuities, stacking over hundreds of traces is usually necessary to enhance the visibility of these minor phases (Shearer 2000).

An alternative approach is to use the regional (10–30°) multipathing seismic body waves that bottom near the interface. Unlike the phase conversions and reflections which are too weak to observe on an individual seismogram, these multipathing waves (triplications) are clearly recorded at a single station. Moreover, distinct triplication branches with different moveout slopes can be observed in record sections of dense seismic arrays.

Pioneering work on triplications was done by Niazi & Anderson (1965) and Johnson (1967). After that, a series of studies followed (e.g. Grand & Helmberger 1984; Tajima & Grand 1995; Brudzinski & Chen 2000; Wang *et al.* 2009) to derive regional 1-D upper-mantle seismic structures by waveform matching between observed and synthetic seismograms. However, due to the complexity of the observed waveforms, most of these studies rely on forward modelling approaches.

Besides this trial-and-error modelling, some efforts to automate the inversion procedure have been made by employing the conjugate gradient method (Gao *et al.* 2006). However, for this gradient-based method, it is a challenge to find an appropriate initial model to avoid falling into local minima, especially for the complex triplication data. The exhaustive grid search has also been used, but with a reduced number of model parameters (e.g. Chu *et al.* 2012; Li *et al.* 2017).

With the rapid development of full-waveform inversion (FWI) in earthquake seismology, triplicated waveforms have been recently incorporated into the 3-D FWI with the adjoint method (Tao *et al.* 2018). The adjoint method and its derived sensitivity kernels provide an effective way to minimize the given misfit in 3-D FWI (e.g. Tromp *et al.* 2005; Bozdağ *et al.* 2016; Koroni & Trampert 2021). However, this approach has only been applied to long-period data with the shortest period of ∼8 s (Tao *et al.* 2018) due to the prohibitive computational cost to simulate higher frequency (shorter period) seismic waves, which consequently limits the image resolution. Also, the currently available data coverage may still not be adequate to provide 3-D constraints of the MTZ structure in some regions. In addition, the uncertainty quantification of the inverted models is always challenging, especially for the 3-D case. Therefore, 1-D simulation and inversion, with high-frequency waveform data (up to ∼1 Hz) and fewer parameters, still are complementary and important in characterizing the MTZ discontinuities, especially near the turning points of seismic waves.

The temperature of the subducting slab is critical for understanding the mantle dynamics near the subduction zone. However, the inverted or modelled fast wave speed perturbations within the subducting slab still vary a lot amongst different studies. In the Kuril subduction zone, the global traveltime tomography model (Obayashi *et al.* 2013) shows the wave speed perturbations within the slab are on the order of ∼1 per cent. Although the current regional 3-D FWI model (Tao *et al.* 2018) shows a thinner but more strongly perturbed slab with ∼2–3 per cent wave speed increase, this value is still much smaller than ∼5 per cent, which is derived from the previous traveltime (Ding & Grand 1994) or waveform (Zhan *et al.* 2014) modelling studies. Whether or not and/or to what extent the perturbation of the slab is underestimated in tomographic results deserve in-depth investigation. Wang *et al.* (2014) and Tao *et al.* (2017), through waveform modelling, have shown that the subducting slab near the turning points can influence triplicated waveforms. Therefore, with a carefully selected event and station distribution, triplication can be used to better constrain the slab structures with higher frequency waves. In this paper, we will first introduce the phenomena of triplications. Then, we systematically analyse the trade-off between model parameters, through forward modelling and waveform inversion. We also illustrate the necessity of using array-normalized amplitude. Finally, we present a case study for seismic data sampling below the Tatar Strait of Russia, where triplications due to both the 410-km discontinuity and the slab upper surface are observed. We simultaneously invert the seismic structures of these two interfaces and demonstrate that with high-frequency data (∼2 s), full-waveform inversion can accurately constrain the detailed structures of the MTZ, comparable to the findings from previous modelling studies.

2 MULTIPATHING TRIPLICATED BODY WAVES

Triplications originate when seismic body waves encounter regions where wave speed increases sharply with depth (e.g. at the Moho, the 410-, or 660-km discontinuities, and the slab upper surface). Near such discontinuities or regions with steep wave speed gradients, body waves (both *P* and *S* phases) will propagate along different paths and can be observed on the regional distance seismic stations (10–30°). An example of the ray path geometry and corresponding synthetic seismograms of *P*-wave triplications caused by the 410-km discontinuity is shown in Fig. 1(a). To clearly present the triplicated phases, in this section we use the WKBJ code (Chapman 1978) to separately calculate each of the three branches. The synthetics are computed using the seismic reference model IASP91 (Kennett & Engdahl 1991), with an intermediate-depth earthquake (the same event as analysed in Section 4) at 114 km to minimize the interference of multiple arrivals (e.g. depth phases). The three branches consist of the direct branch (AB), the reflected branch (BC) and the refracted branch (CD), which are illustrated in Figs 1(b)–(d), respectively. We note that here we have not applied normalization to this synthetic case such that the relative amplitude variations between stations are kept. As shown in Fig. 1(a), these triplicated phases provide dense samplings of the 410-km discontinuity. Since the ray paths of the different triplication branches are largely overlapping in the shallow part, the relative traveltimes and amplitudes

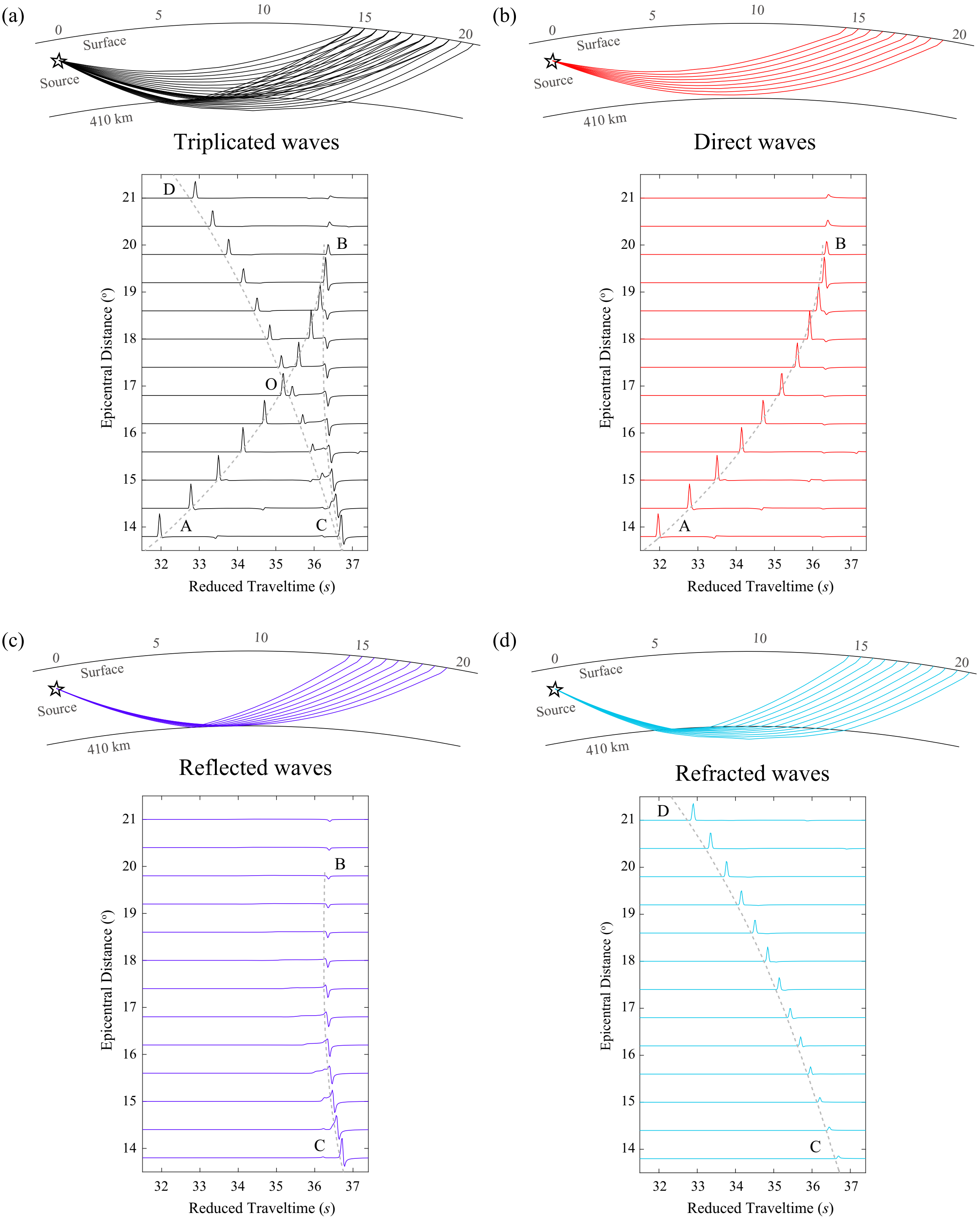

Figure 1. Ray paths and waveforms for *P*-wave triplications. (a) Ray paths and waveforms for all the triplicated *P* waves. In the upper panel, the black star is the earthquake source at 114 km, and black lines show all the ray paths. In the lower panel, the black waveforms are synthetics, and the dashed grey lines are the traveltime curves. AB, BC and CD branches represent the direct waves, reflected waves and refracted waves, respectively. The O point shows the crossover point of the AB and BC branchs. A reducing slowness of 11.5 s $\deg^{-1}$ is used. (b) Ray paths and waveforms for the direct waves AB. (c) Ray paths and waveforms for the reflected waves BC. (d) Ray paths and waveforms for the refracted waves CD.

of triplications can be attributed primarily to the structure near the transition zone.

3 THE TRADE-OFF BETWEEN MODEL PARAMETERS

3.1 Discontinuity depth and wave speed above

The existence of a low wave speed zone above the 410-km discontinuity, indicative of partial melting, can provide evidence for the water content in the mantle transition zone (Bercovici & Karato 2003). Some researchers using converted or transmitted phases have observed the existence of the low wave speed zone above the 410-km discontinuity in some regions (Revenaugh & Sipkin 1994; Schmandt *et al.* 2011; Wei & Shearer 2017). Such anomaly has also been modelled from triplication data (e.g. Song *et al.* 2004; Li *et al.* 2019a; Han *et al.* 2021).

Here, we perform an ideal synthetic case without noise to test the sensitivity of triplications to the low wave speed zone above the interface. For the model set-up, we keep the wave speed above 360 km, the same as the IASP91 model, and decrease the wave speed at 410 km by 0.1 km s^{-1} to represent a low wave speed gradient within 50 km above the 410-km discontinuity (the red line in Fig. 2a).

We calculated both the traveltime curves and waveforms (amplitude normalized by each trace) for this case. We note that for this modelling here and all the others in subsequent parts, we use the QSEIS program (Wang 1999) to calculate the full wavefield, instead of specified phases by the WKBJ program in Fig. 1. QSEIS uses the orthonormal propagator algorithm, a numerically more stable alternative (Wang 1999) to the reflectivity method. In addition, it can directly calculate waveforms starting from the onset time of the triplication phases instead of the origin time of the earthquake, which significantly saves computing time. As shown in the traveltime curves, the low wave speed zone above the discontinuity mainly affects the extension of the OB branch (the red line in Fig. 2b). Specifically, in this case, the direct waves (OB branch) terminate at a larger epicentral distance, thereby with increased OB branch's amplitude compared to the IASP91 synthetics (the shaded grey area in Fig. 2c). This phenomenon has also been observed in previous studies (e.g. Li *et al.* 2017, 2019; Han *et al.* 2021) and has been used to detect the existence of the low wave speed zone.

However, other model candidates also have such equivalent behaviour near cusp B. For example, we show a comparison between this model (the red line in Fig. 2a) and the other equivalent model with a depressed 410-km discontinuity but with a normal wave speed gradient as the IASP91 above the discontinuity (the blue line in Fig. 2a). The traveltime curves (Fig. 2b) show that both of these two models will produce synthetics with an extended OB branch to larger epicentral distances to different extents compared to the IASP91 model. Specifically, the model with a low wave speed layer above the interface extends the OB branch to a relatively larger epicentral distance.

The waveforms of the OB branch from the two models, are quite similar at certain epicentral distances (the shaded grey area in Figs 2c and d), indicating that the trade-off does exist between the discontinuity depth and the wave speeds above the discontinuity. We note that the amplitude of the waveforms shows some discrepancies with the traveltime curves (e.g. Fig. 2b shows that the OB branch terminates at 21° for the model with 15 km depression of the discontinuity, while the OB branch for the same model seems to be extending beyond 21°, in Fig. 2d). This inconsistency is due to two reasons. The first is the difference between the ray theory and the finite frequency effect; the waveform calculated numerically, which takes the finite frequency effect into account, is more reliable and closer to the real physical situation. The second reason is from the normalization by each trace which we will discuss in the next subsection.

This trade-off has also been noted and investigated in previous studies (e.g. Song *et al.* 2004; Wang & Chen 2009). For example, Wang & Chen (2009) analysed similar model pairs for the 660-km discontinuity and rejected the model with a depressed interface based on its different slope for the OC branch in the traveltime curves. According to our test, even if there are some differences for the slope of the OC branch in the traveltime curves (Fig. 2b), the differences in the corresponding waveforms for the OC branch are more subtle (e.g. less than a quarter of the wavelength). The other reason why the waveforms from our two tested models look more identical is that we applied our waveform inversion code to search for the equivalent model to the one with low wave speed above the discontinuity (out of 15 000 models). Song *et al.* (2004) also discussed these two endmember models for the 410-km discontinuity by comparing the waveforms. The model with a depressed interface is ruled out due to its failure to generate the visible waveforms of the OB branch (Song *et al.* 2004). However, the proposed model in our case does generate a clear OB branch whose amplitude is equivalent to the model with a low wave speed zone above the interface. This discrepancy could partly come from the different earthquake sources we choose (different depths and focal mechanisms). The other possibility is that our synthetic model has an extra localized high wave speed anomaly below the interface. Assuming that this anomaly doesn't exist in the MTZ, the CD branch will be delayed. Thus, if viewed in the velocity seismogram (e.g. Song *et al.* 2004), the negative pulses of the delayed CD branch will partly overlap with the OB branch and lower its amplitude.

We also note that this equivalent model we propose might not be consistent with other constraints in certain regions (e.g. the receiver function results in Song *et al.* 2004). However, theoretically, these two models are identical examined by only triplication data. Therefore, triplication data alone cannot well constrain a low wave speed zone due to the trade-off between the interface depth and the wave speed gradient above it, especially when we normalize the amplitude by each trace.

3.2 Array normalization

In most of the previous triplication studies, researchers prefer to normalize the waveforms by each trace. Normalization is needed because of the uncertainties in the source magnitude, fault plane solution, attenuation, and site effects, which make the absolute amplitudes more difficult to constrain. However, when using the normalized amplitude of each trace, information about the relative amplitude variations between stations is lost.

In this paper, we propose to use array normalization rather than trace normalization. In a record section, array normalization means that we normalize all traces relative to one particular reference station. Because all the records are from the same earthquake, the source magnitude uncertainty will not affect the results after array normalization. Furthermore, within the narrow azimuthal range for the particular record section, the effect of uncertainty in the

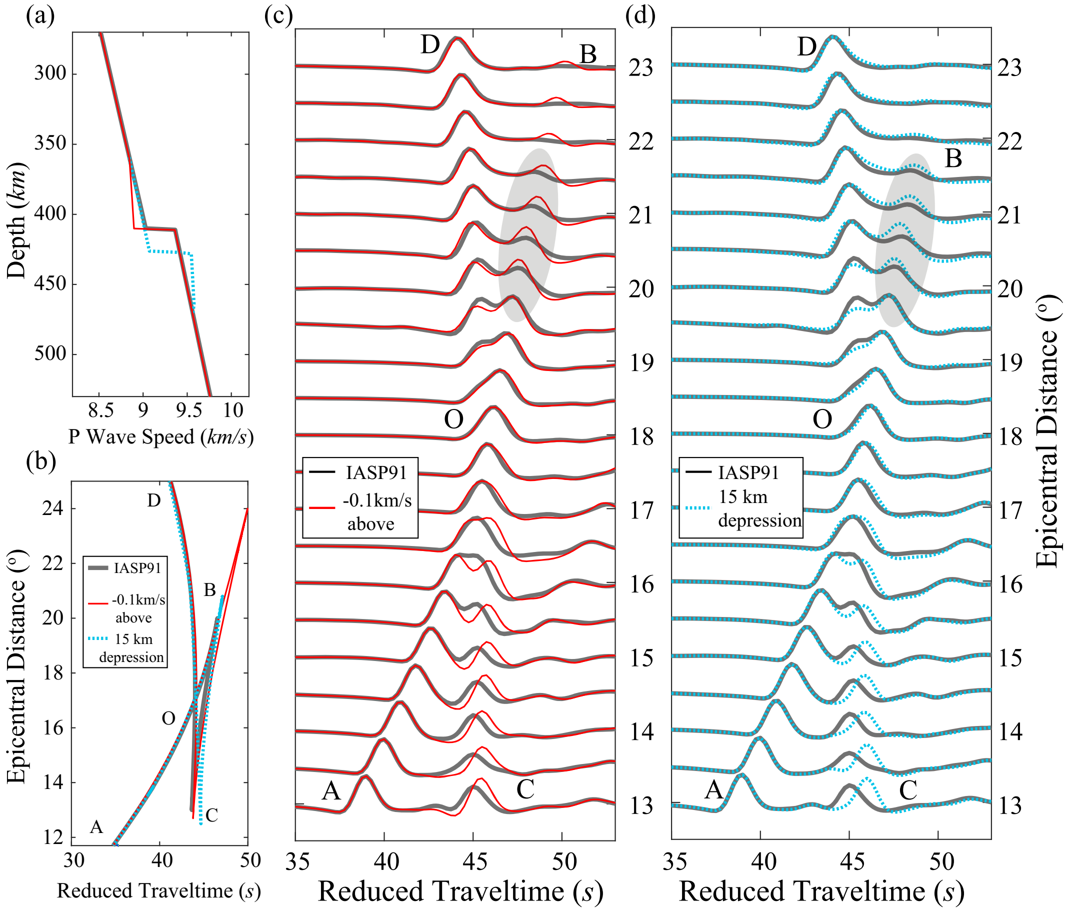

Figure 2. Modelling tests for the trade-off between model parameters. (a) The bold grey line shows the model IASP91, the solid red line indicates the model with a low wave speed zone above the discontinuity and the dashed blue line represents the model with a 15-km depression for the interface. (b) Traveltime curves for the models in (a) with the same line styles. (c) Waveform comparison between model IASP91 (grey) and the solid red model in (a). (d) Waveform comparison between model IASP91 and the dashed blue model in (a). The amplitude is normalized by each trace and waveforms in the grey region are similar.

fault plane solution is also insignificant. When we invert for one discontinuity, the range of epicentre distances is only within about ten degrees, therefore, we expect the attenuation near the discontinuity within this relatively small range should not influence the waveforms dramatically. Nevertheless, suppose we have observed stations with unusual amplitudes either due to attenuation or site effects, we could use trace normalization for these stations or reduce the weights of them in waveform misfit contribution for full-waveform inversion.

We first compare the trace normalization and array normalization for the two models shown before (Fig. 2a). In the array-normalized waveforms (Fig. 3a) where amplitude information between stations is kept, we do observe differences in amplitude between these two models (the shaded grey region in Fig. 3a). More specifically, the amplitude of the OB branch for the blue one is smaller than the red one, although still larger than the IASP91 model. Besides, the amplitude for the OD branch is also different. When we apply the traditional trace normalization (Fig. 3b), since the amplitudes of both the OD and OB branches are magnified for the blue one, there are no obvious differences between the waveforms for these two model types (blue and red). In other words, the larger amplitude of the OB branch comes from the magnification of the trace normalization due to the smaller amplitude of the OD branch. Therefore, relative amplitude information between stations in the array-normalized record section (Fig. 3a) can help to distinguish these two types of models, whose waveforms are almost identical in the traditional trace-normalized record section (Fig. 3b).

We show another comparison between trace normalization and array normalization to illustrate the necessity of applying array normalization. As shown in Figs 4(c) and (d), we present model IASP91 (black) and another designed model (red) with a -0.4 km s^{-1} low wave speed layer only in the shallow part (<160 km). As shown in Fig. 4a, the array-normalized seismogram demonstrates that the different structures in the shallow part will cause an overall time delay (of ~3 s) and affect the amplitude of the direct wave at different epicentral distances (AO). In comparison, the amplitudes of the later phases (CO) remain basically unchanged. However, for trace normalization, because the amplitude for the direct wave (AO) is always the largest within the epicentral distance range before 15°, the amplitude of the direct wave is always unity after trace normalization (Fig. 4b). Therefore, the amplitude of the later phases (CO), which is originally unchanged, seems to have a smaller amplitude relative to the IASP91 synthetics after the trace normalization. We note that the later phases (CO) correspond to the reflected wave at the 410-km discontinuity and the transmitted wave below it. Therefore, the deeper structure is likely to be incorrectly inverted due to trace normalization (Fig. 4d).

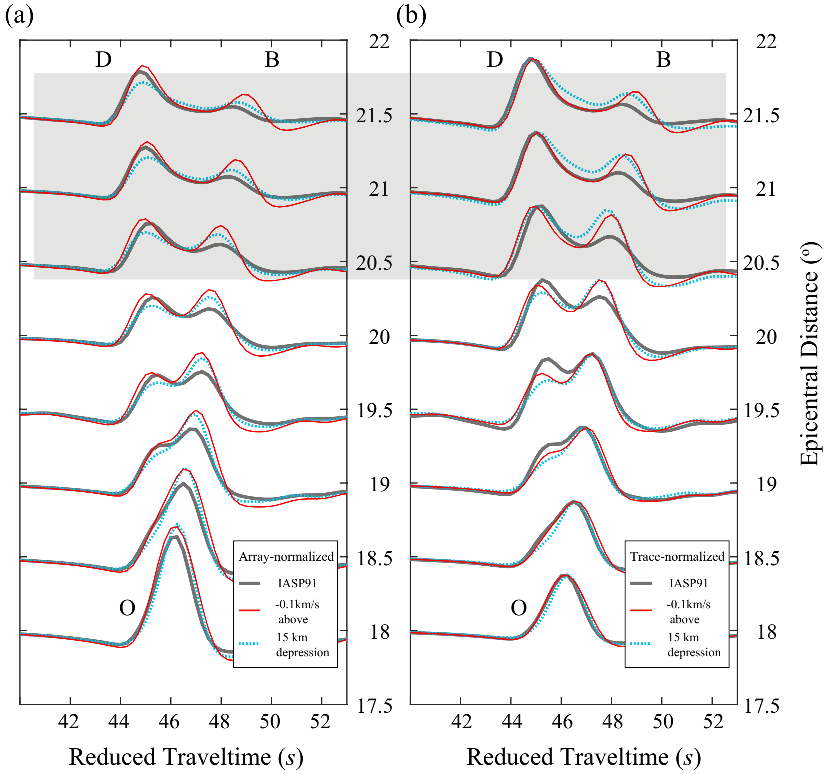

Figure 3. Comparison between array and trace normalization (model trade-off). (a) Array-normalized waveforms. The bold grey waveforms are for model IASP91, the solid red and the dashed blue waveforms represent the corresponding models in Fig. 2(a). Differences between these two deviated models are clearly shown in the shaded grey region. (b) Trace-normalized waveforms. Symbols are the same as (a) and no obvious differences exist between these two models (red and blue).

As such, besides losing the relative waveform amplitude information between stations (increasing trade-off), trace normalization will also lead to the erroneous mapping of the corresponding structure due to the artificially created mismatch in the waveforms. Therefore, we propose and recommend using the array-normalization approach for full-waveform inversion.

3.3 Synthetic inversion test

Using array normalization, the trade-off between the interface's depth and the wave speed gradient above it can be minimized. Here, we will perform a synthetic inversion test to show to what extent this trade-off will be reduced and how much of it remains.

To obtain quantitative error bounds, and to avoid the risk of falling into the local minima faced by the gradient-based inversion method, we adopt the niching genetic algorithm (Koper *et al.* 1999; Li *et al.* 2012, 2021) in the inversion framework of triplicated waveforms. Niching genetic algorithm (NGA) is a non-gradient-based inversion scheme that searches the model space through massive forward modelling and is independent of the initial model. Only the search range of the model space is given as *a priori*. The key difference between NGA and other global search techniques is that NGA promotes model diversity by applying a penalty term for the model similarities. Moreover, because NGA involves numerous samplings in the model space, it can provide a series of acceptable model sets. The mean and variance of these acceptable models can help estimate the uncertainty of the final model.

We design a *P*-wave synthetic test and apply the array normalization. In this test, we set model IASP91 as the 'ground truth', and invert for the 1-D wave speed profiles using its corresponding synthetic displacement waveforms as inputs. We set the maximum epicentral distance to be 21° because within this range the triplicated OB branch is visible. In the inversion set-up, considering the ray path's penetration depths, we only invert the structure between 210 and 570 km depths. Within this depth range, we parametrize the seismic structure with ten unknowns to be inverted for. Three parameters are used to capture the sharp wave speed gradient across the '410-km' discontinuity: two of them are used to describe the wave speed jump across the discontinuity, the third one represents the discontinuity depth perturbation. Seven more parameters are set above and below the interface with a depth interval of ∼40 km to capture more gradual wave speed changes away from the discontinuity. We only invert the wave speed at these points. Between two adjacent anchor points, the wave speeds are linearly interpolated. Beyond this depth range, the wave speeds are the same as the model IASP91. The *P* wave speed at each anchor point is allowed to vary between plus and minus 0.3 km s^{-1}, and the position of the discontinuity varies within plus or minus 20 km, relative to model IASP91 (Fig. 5a). In addition, the effect of attenuation for the *P* wave is considered by applying a constant t^* value of 1 s (Stein & Wysession 2009). We note that the *P* wave speed is the only unknown parameter for each fixed anchor point, and the Poisson's ratio and density are the same as those in the model IASP91. We think this assumption is appropriate for this case, where only the *P*-wave data set is involved. To further constrain the Poisson's ratio,

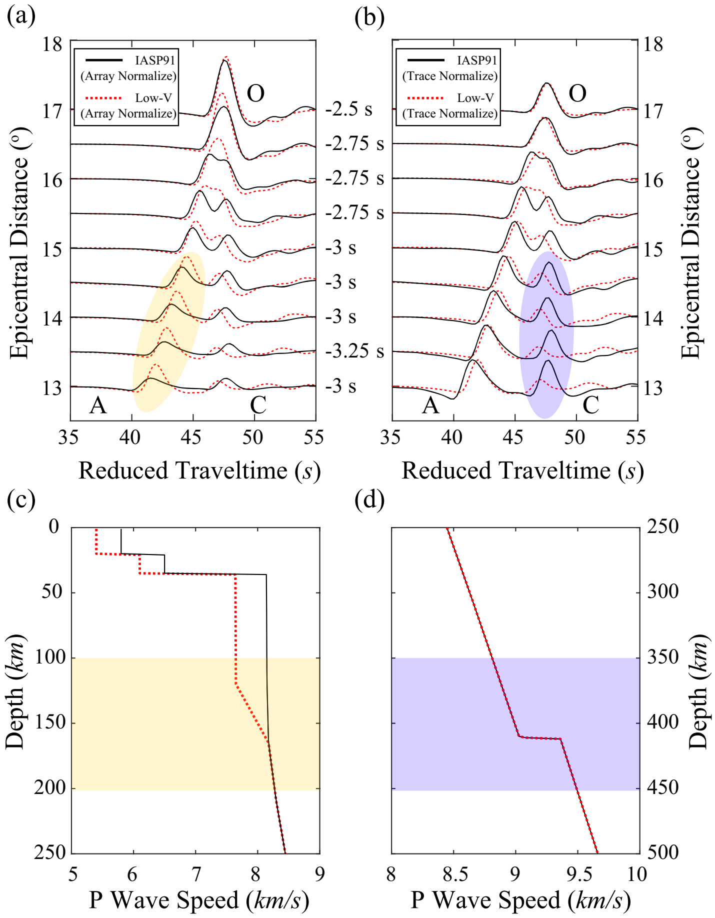

Figure 4. Comparison between array and trace normalization (artificial mismatch). The solid black waveforms are synthetics for model IASP91 and the dashed red waveforms represent the red model in (c). The yellow region shows where the amplitudes are different. The number near the end of each trace denotes the time delay (∼3 s) for each station. (b) Trace-normalized waveforms. The blue region shows where the waveforms are different. (c) The shallow portion of the model. The solid black line is model IASP91, and the dashed red line is the designed model with a -0.4 km s^{-1} zone in the top 160 km. The yellow box roughly shows where the wave speed gradient changes. (d) The deep portion of the model. The blue box roughly shows the structure we tend to modify due to the artificial mismatch of the OC branch.

an additional *S*-wave data set is required. But this joint inversion is currently beyond the scope of this paper.

As for the misfit window, we choose a continuous one from 32 to 52 s (reduced time) which contains the entire triplicated *P* wave train, for the case without noise. Before the calculation of the misfit, we first cross-correlate the synthetic and observed waveforms for the ith station to obtain the time difference Δt_i. This cross-correlation is to mitigate some baseline shifts and emphasize the relative traveltime between the triplicated phases (e.g. Grand & Helmberger 1984; LeFevre & Helmberger 1989; Brudzinski & Chen 2003). After shifting the synthetic trace by Δt_i, we calculate the L2 norm of the differences between the observed and aligned synthetic waveform in the time domain as the misfit function χ_{L2}:

$$\chi_{\mathrm{L2}} = \sum_{i=1}^{N} \int_{t_1}^{t_2} \left| \frac{\mathbf{d}(\mathbf{x}_i, t)}{\max\limits_{t_1 \le t \le t_2} |\mathbf{d}(\mathbf{x}_{\mathrm{ref}}, t)|} - \frac{\mathbf{u}(\mathbf{x}_i, t + \Delta t_i)}{\max\limits_{t_1 \le t \le t_2} |\mathbf{u}(\mathbf{x}_{\mathrm{ref}}, t + \Delta t_{\mathrm{ref}})|} \right|^2 \mathrm{d}t,$$

where $\mathbf{d}(\mathbf{x}_i, t)$ is the displacement data recorded by the *i*th station, $\mathbf{u}(\mathbf{x}_i, t + \Delta t_i)$ is the synthetic data for the *i*th station after a time shift of Δt_i. The start and end time for the misfit window are t_1 and t_2, respectively, and N is the total number of stations used in the inversion. The data and synthetics for the reference station are represented by $\mathbf{d}(\mathbf{x}_{\mathrm{ref}}, t)$ and $\mathbf{u}(\mathbf{x}_{\mathrm{ref}}, t + \Delta t_{\mathrm{ref}})$, respectively, and their maximum absolute values are used for the array normalization. Here we normalize the data and the synthetics separately to minimize the influence from the possible inaccurate magnitude of the event.

After the first 20 generations (100 simulations per generation), the misfit significantly reduces, and after 80 generations the misfit starts to converge (Fig. 5c). From the 100 models in the last generation, we define the acceptable model with a misfit of less than 10 per cent increase than the misfit of the best model. In case, sometimes, the misfit does not readily detect the mismatch, we further examine all the acceptable models by visually comparing the data and synthetics, to determine the final acceptable candidates.

Finally, we have three acceptable model groups (Fig. 5a). The first groups converge to the ground truth model, verifying the effectiveness of our triplication inversion package. The other two model groups (group 2 and group 3) are similar to the pair of models we discussed previously (Fig. 2a). We further use the averaged value of these two groups of models to calculate their corresponding displacement waveforms. Waveforms between these two groups are

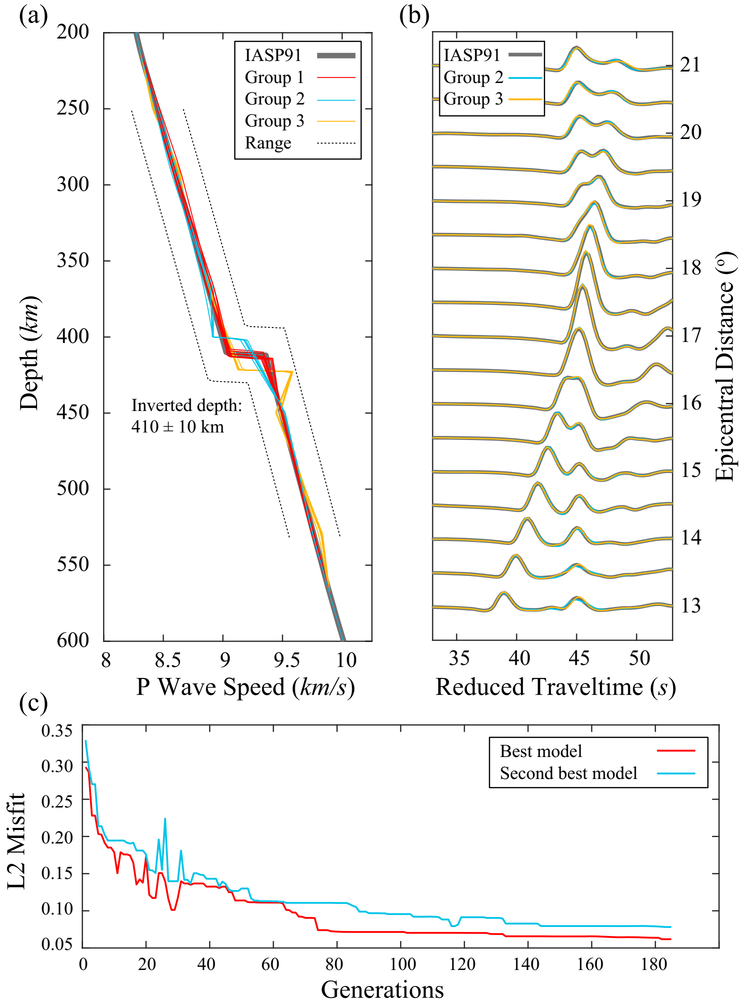

Figure 5. Synthetic tests for Niching Genetic Algorithm. (a) Inverted models. The bold solid grey line shows model IASP91. The red, blue and yellow solid lines indicate different groups of the acceptable models. The dashed black lines represent the model searching range. (b) Waveform fitting. The bold grey waveforms are synthetics for model IASP91, the blue and yellow waveforms are synthetics using the averaged value for model groups 2 and 3, respectively. (c) L2 misfit between data and synthetics. The red and blue lines are the L2 misfit for the best and second-best models, respectively.

almost identical, and both of them are also quite similar to model IASP91 waveforms (Fig. 5b).

This synthetic test shows that even if the array normalization is applied, this trade-off between the interface depth and the wave speed gradient above it cannot be eliminated. The reason is that a depressed interface truly has a similar impact on the amplitude of the OB branch compared with a negative wave speed gradient above the interface (Figs 2c and d). Some differences between these two models in the waveforms are less obvious compared with the traveltime curves (Fig. 2b) due to the finite frequency effect.

Nevertheless, for a given frequency band, we can estimate the depth uncertainty due to this trade-off. These depth limits can be quickly found using our automatic inversion program (Li *et al.* 2021). For this case, given this frequency band (half duration of ∼2 s) and misfit tolerance, the trade-off from the wave speed above the discontinuity will lead to a ∼10 km uncertainty in the depth estimation.

3.4 Discontinuity depth and wave speed in the MTZ

In the western Pacific subduction zone, tomography results (Huang & Zhao 2006; Chen & Pei 2010) indicate a ‘flat slab’ in the MTZ, which increases the wave speeds in the MTZ. We first test the triplication’s sensitivity to the high wave speed perturbations in the MTZ. Here we calculated the traveltime curves when the wave speed below the 410-km discontinuity is increased by 0.1 km s^{-1} (Fig. 6a) relative to model IASP91, using the Taup toolkit (Crotwell *et al.* 1999). Traveltime curves show that the wave speed in the MTZ significantly impacts the CD branch’s traveltime (Fig. 6c). In other words, the increase of the wave speed below the discontinuity will make the transmitted waves (CD) travel faster. Crossover point (O) marks the intersection of the AB and CD branch, where the waveform amplitude reaches its maximum. Therefore it is one of the most obvious signatures of this triplication. In this case, the earlier arrivals of the CD branch will cause the crossover point (O) to appear at a smaller epicentral distance.

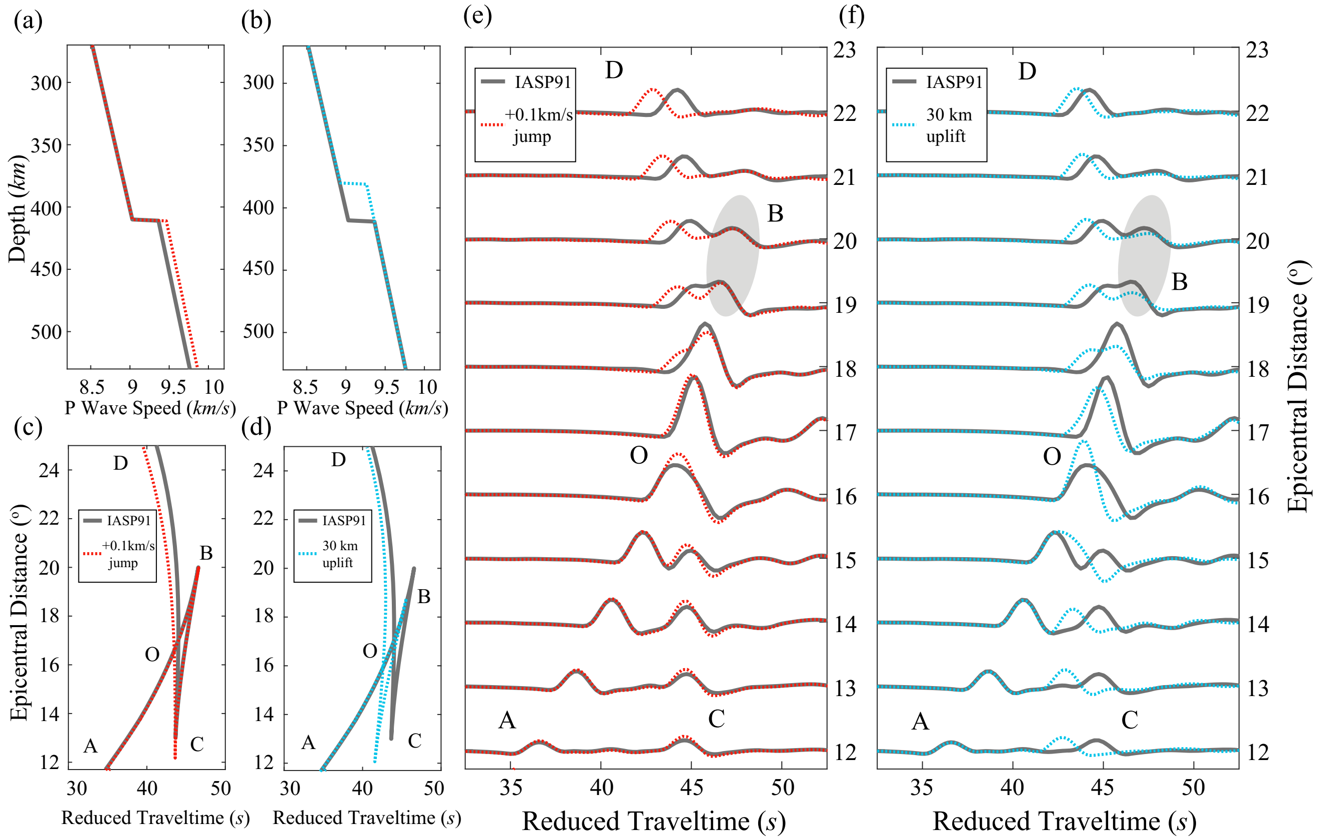

Figure 6. Modelling tests for the discontinuity depth and wave speed jump. (a) The bold grey line shows model IASP91, whereas the dashed blue line is the model with a +0.1 km s^{-1} wave speed jump. (b) Model IASP91 and the model with a 30-km uplift (dashed red). (c) Traveltime curves for models in (a). (d) Traveltime curves for models in (b). (e) Waveform comparison for models in (a). A reducing slowness of 11 s deg^{-1} is applied. (f) Waveform comparison for models in (b). Grey regions in (e) and (f) indicate where the amplitude of the OB branch is different.

Similar behaviour of the traveltime curves occurs when the depth of the interface is shallower, and near the subducting slab, the 410-km discontinuity can be elevated due to the positive Clapeyron slope (e.g. Bina & Helffrich 1994; Flanagan & Shearer 1998). Assuming a situation where the 410-km discontinuity has a 30-km uplift (Fig. 6b), the CD branch arrives earlier, and consequently, the crossover point (O) occurs at a smaller distance (Fig. 6d). This is because, in this situation, this elevated interface is equivalent to a high wave speed anomaly between 380 km and 410 km depth.

It is critical to have stations with smaller epicentral distances (before the crossover point) to distinguish the model with uplifted discontinuity and the one with high wave speed within MTZ. One difference between these two models (Figs 6a and b) is that when the 410-km discontinuity is uplifted, the earlier arrival of the CD branch can be seen at smaller epicentral distances where this branch just emerges (cusp C in Figs 6d and f). While for the other case where a high wave speed exits in the MTZ, the advance of the CD branch is not obvious until at epicentral distances greater than the crossover distance (O in Figs 6c and e).

However, stations close to the epicentre near the subduction zone are often scarce which makes it difficult to distinguish between these two models. Nevertheless, the additional detailed waveform differences may help differentiate these two models. For example, in the case of a more considerable wave speed jump beneath the discontinuity, the amplitude near cusp B remains unchanged (the shaded grey area in Fig. 6e). Meanwhile, for the model with an uplifted 410-km discontinuity, the amplitude near cusp B is smaller (the shaded grey area in Fig. 6f). Therefore, even if the traveltime differences between the OB and OD branches are almost identical for these two situations (Figs 6c and d), we can make an unambiguous distinction between them based on the waveform details at certain epicentral distances (Figs 6e and f).

4 APPLICATION TO THE KURIL SUBDUCTION ZONE

We focus on an intermediate depth (114 km) event that occurred in the Kuril subduction zone on 2009 October 10 with moment magnitude of ∼5.9 (Fig. 7b). We choose this intermediate-depth event to avoid the interference of the depth phases in the triplicated waveforms. The observed waveforms are selected from a subset of the broad-band CEArray stations (Zheng *et al.* 2010) and the NECESSArray (the NorthEast China Extended SeiSmic Array) in northeast China. We choose the *P*-wave data to achieve a better resolution of the MTZ structure because the *P* wave is typically observed at a higher frequency than the *S* wave due to its smaller attenuation. Therefore, even though the wave speed of the *P* wave is larger than that of the *S* wave, the *P* wave still has a smaller Fresnel zone. After removing the instrument response, we have applied a first-order, zero-phase shift Butterworth filter with a frequency band of 0.05–1 Hz to the data. We choose this relatively broad frequency band to avoid waveform distortion due to narrow-band filtering. We further divide the stations into two sublinear arrays according to their azimuthal angles (282.5° for the northern region along with

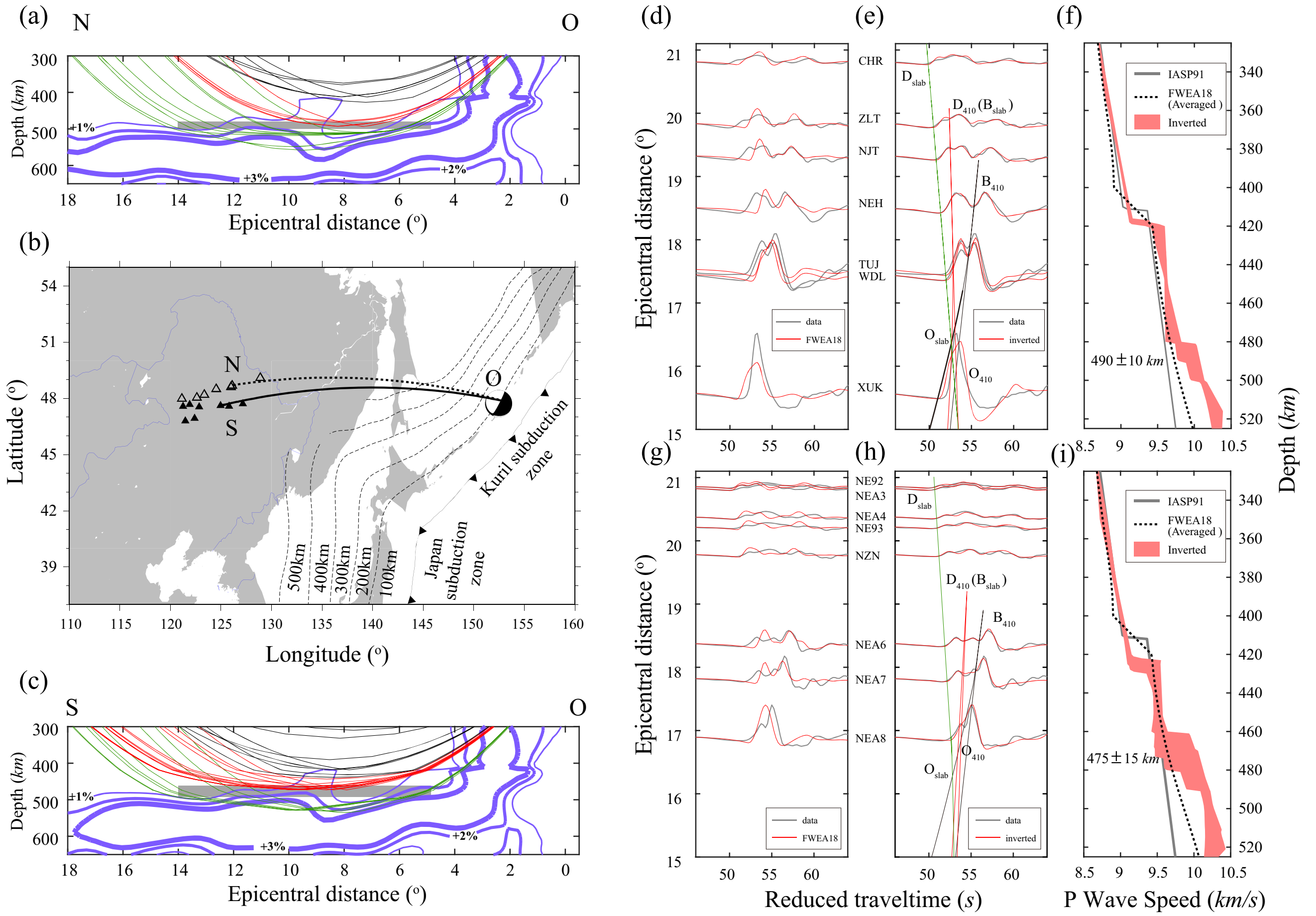

Figure 7. Research region and inversion results. (a and c) Cross-sections ON and OS as shown in (b). The blue lines are wave speed contours of model FWEA18. The black, red and green lines are the ray paths corresponding to the traveltime curves in (e) and (h). The shaded grey regions indicate the locations of the inverted slab upper surfaces with uncertainties. (b) Research region. Beach ball represents the earthquake. Hollow and solid triangles denote the stations for the northern and southern regions, respectively. ON and OS show the location of the cross-sections. Dashed lines are the depth contours of the slab. (d and g) Displacement waveform (*P* wave in the *Z* component) comparison between aligned data (grey) and FWEA18 synthetics (red, 2-D forward modelling) in the northern and southern regions, respectively. A reducing slowness of 10.5 s deg^{-1} is applied. (e and h) Waveform comparison between aligned data (grey) and 1-D inverted synthetics (red) for the best-fitting model in the northern and southern regions, respectively. (f and i) *P* wave speed inversion models in the northern and southern regions, respectively. The shaded red region indicates all the acceptable models, whereas the solid grey line shows model IASP91. The dashed black line represents the averaged (from 5° to 14°) value for model FWEA18.

ON and 279.5° for the southern region along with OS) and the observed distinct waveform patterns (Figs 7e and h). Within each sublinear array, the azimuth range is relatively narrow (∼1°), and one model should explain all the waveforms in this particular record section.

Compared with the synthetic waveforms shown before (e.g. Fig. 5b), the observed data are more complex because there are two triplications in each record section. Specifically, there is a third phase (along the red line in Figs 7e and h) between the first (the green line) and the last phase (the black line). This extra phase requires, in the inverted models, another discontinuity (high-wave speed gradient) below the 410-km discontinuity. By parametrizing two interfaces in the inversion, we obtain acceptable models indicated by the shaded red region in Figs 7f and i. We choose one of them to generate the synthetic waveforms, which show good agreement with the observed waveforms in terms of both the relative waveform timing and amplitudes on each trace and the relative amplitude variations between stations (Figs 7e and h).

Certain *a priori* information is assumed in the inversion. First, given the fact that with this triplication data alone we cannot exclusively verify the presence of a low wave speed zone above the 410-km discontinuity, therefore we force the wave speed gradient above the 410-km discontinuity to be no less than the value in model IASP91. As such, we can reduce the model unknowns and focus more on inverting the parameters pertaining to the secondary discontinuity. But we know that given the half duration of 2 s, a negative wave speed gradient above the discontinuity can introduce a topographic uplift of ∼10 km for the 410-km discontinuity in an equivalently accepted model (Fig. 5a). Second, we set all the interfaces as a sharp discontinuity to further reduce the number of model unknowns in the inversion. We will systematically test the sharpness of the discontinuities in the discussion.

Results show that the average depth for the first discontinuity is 420 km and 425 km for the northern and southern regions, respectively (Figs 7f and i). The depth uncertainty is estimated to be ∼±5 km from all the acceptable models given the frequency band (2–20 s). We note that this uncertainty is based on our *a* *priori* assumptions. If we fully consider the trade-off between different model parameters, another ± 10-km uncertainty should be taken into account. Therefore, the overall estimated depth should be 420±15 and 425±15 km for the northern and southern regions, respectively. Turning points, that is the most sensitive regions of the triplicated wave paths, are below the Tatar Strait of Russia. Our inverted 410-km discontinuity depths at 420±15 and 425±15 km are consistent with the insignificantly uplifted 410-km discontinuity in this region derived from ScS reverberations (Wang *et al.* 2017). Furthermore, our result has improved resolution due to the smaller Fresnel zone of the *P* wave at a higher frequency (∼ 0.5 Hz).

For the secondary discontinuity, it is located at 490±10 and 475±15 km for the northern and southern regions, respectively (Figs 7f and i). The second discontinuity is located at depths between +1 and +2 per cent wave speed contours (Figs 7a and c) of the regional FWI model FWEA18 (Tao *et al.* 2018). We also note that the trade-off between the interface depth and the wave speed gradient above it can be observed from the acceptable models. In this region, the upper slab surface is located ∼50–70 km below the 410-km discontinuity (Figs 7f and i). The cooling effect from this relatively distant slab should be weak, which also explains the insignificant uplift of the 410-km discontinuity.

5 DISCUSSION

5.1 Comparison with the 3-D regional FWI model

Tomographic velocity model FWEA18 (Tao *et al.* 2018) is the currently highest resolution FWI model in this region. Because FWEA18 is inverted with body waves with the shortest period of 8 s, thus it is important to examine how well this model can predict the body waves of a higher frequency of 2 s and if further refinement of the model is necessary. Due to the prohibitive computational cost of 3-D simulations for 2-s waves, 2-D simulations are performed for the profiles close to the selected stations which are sublinear. The 2-D models along with profiles ON and OS (Figs 7a and c) are extracted from 3-D model FWEA18 and implemented in the 2-D finite-difference (FD) package (Li *et al.* 2014) to generate the 2-D synthetics (Figs 7d and g). This 2-D package has incorporated several corrections, e.g. out-of-plane spreading, point-source excitation, and Earth-flattening, to better account for the 3-D wavefield spreading (Li *et al.* 2014). We note that this 2-D package is only used to compute seismograms for comparison here, not for the inversion.

Waveform fitting comparison indicates that although model FWEA18 predicts the observed 2-s waveforms much better (Figs 7d and g) than the model IASP91 (Supporting Information Figs S1a and b), its synthetics still can't fit the data completely. The mismatch mainly comes from the differential traveltime between the direct wave and the refracted wave, which is smaller in the FWEA18 synthetics (Figs 7d and g) than in the data (e.g. the black and green lines in Figs 7e and h). This mismatch suggests that although the high *P* wave speed perturbation within the slab of model FWEA18 is stronger, twice the perturbation within the slab of model GAP-P4 (Obayashi *et al.* 2013), the perturbation is still not large enough to advance the refracted phases relative to the direct waves. The underestimation of wave speed perturbation within the slab model FWEA18 is likely caused by the relatively long-period data (>8 s) used in the model inversion. Through the 1-D full-waveform inversion in this study, we can find acceptable models with adequate wave speed jump across the slab upper surface. More significantly, these models can predict not only the differential traveltime between the direct and the refracted waves but also an additional moveout in the record sections (indicated by the red line in Figs 6e and h). Therefore, 1-D full-waveform inversion based on a higher frequency (∼2 s) waveforms is indispensable to reveal precise structures such as the wave speed jump across the slab upper surface, which will be discussed in the following subsections.

For all the stations in the northern region, both the data and synthetics are normalized according to station WDL so that the relative amplitudes between stations are preserved. We also observe a large amplitude difference between the data and the FWEA18 synthetics for station XUK, but a much smaller amplitude difference between data and the synthetics of the model inverted in this study. This large amplitude discrepancy between data and FWEA18 synthetics can be explained by the misfit function used in the inversion, which is based on normalized-zero-lag cross-correlation (Tao *et al.* 2017) and is insensitive to the amplitude differences. This suggests that full-waveform inversion with preserved relative waveform amplitude is necessary to recover more realistic models that can predict the waveform data across the array.

5.2 Discontinuity sharpness

The discontinuity sharpness is one of the critical parameters to distinguish models with mineralogical phase changes and chemical layering (Benz & Vidale 1993). In addition, the sharpness of the 410-km discontinuity is sensitive to the water content (Helffrich & Wood 1996) and is critical to understand the deep water cycle of the Earth (Thompson 1992).

In our inversions, to reduce the model unknowns, we set all the discontinuities as sharp interfaces. However, it is important to discuss if the 2-s waveforms in this study can resolve the sharpness across the 410-km discontinuity and the slab upper surface. The systematic waveform modelling tests with different 410-km discontinuity thickness (0, 20 and 40 km as shown in Supporting Information Fig. S2a) show that the discontinuity thickness has the biggest impact on the extent of the BC branch, with thicker discontinuity corresponding to a smaller extent on the traveltime curve (Supporting Information Fig. S2b). In these tests, we keep the wave speeds the same above and below the 410-km transitional zone as model IASP91, but only vary the thickness centred at the depth of 410 km.

However, different from the impact of the discontinuity thickness on the traveltime curves calculated based on ray theory, the corresponding 3-s waveforms of the BC branch exhibit no sensitivity to the discontinuity thickness even up to 40 km (Supporting Information Fig. S2c), which is likely due to the wave-front healing, a finite frequency effect (Nolet & Dahlen 2000).

We perform a set of forward modelling tests to further investigate the sensitivity of waveforms with different dominant frequency periods (3, 2 and 1 s) to the discontinuity thickness up to 40 km (Supporting Information Figs S2c–e). As the frequency increases (e.g. from 2 to 1 s), the waveform differences become more apparent between the sharper and gradual discontinuities, especially for the pre-critical reflections at a smaller epicentral distance (Supporting Information Figs S2d and e). A similar frequency-dependent sensitivity to the discontinuity thickness has also been observed in previous triplication studies (Melbourne & Helmberger 1998; Zhang *et al.* 2019). Given the fact that 2-s waveforms cannot discern a model with a sharp jump across the 410-km discontinuity from the model with a certain thickness, therefore, we set the discontinuity as a sharp interface in the inversion. Nevertheless, the inverted sharp interface's depth should reflect the centre of alternative gradual interfaces with a certain thickness.

To investigate the sharpness of the slab upper surface, we apply waveform inversion to the real data. In the southern region, we fix the midpoint of the interface at 480 km and set its thickness to be 20, 40, 60 and 80 km (Supporting Information Fig. S3a). For each interface thickness, the full-waveform inversion is performed to search for the acceptable models. Up to 60 km for the slab upper interface thickness, the models inverted can predict the second triplication on the observed waveform (annotated by the black arrow in Supporting Information Fig. S3b). However, when the thickness reaches 80 km, the wave speed gradient is too small to generate the second triplication. Therefore, triplication waveforms of 2 s alone, recorded from one event, are unable to discern the slab interface thickness between 0 and 60 km, but they certainly require the thickness to be less than 80 km. There is also a challenge to use shorter period (higher frequency) waveforms from lower magnitude earthquakes, due to the low signal-to-noise ratio. Nevertheless, combining triplication data with converted or underside reflected phases could better constrain the discontinuity's sharpness in the future. Additionally, for the accepted model with a slab interface thickness of 60 km, the positive wave speed gradient within the inverted slab is twice the average value of model FWEA18, which will be further discussed in the next section.

5.3 Wave speed jumps across the slab upper surface

Although model FWEA18 shows a 2-D structure, in the cross-section roughly parallel to the strike direction of the slab, its structure near the turning points of the wave paths varies little laterally and can be treated as a 1-D layered model (the grey region in Figs 7a and c) with averaged velocity from the epicentral distance of 5° to 14° (dashed black lines in Figs 7f and i). To obtain the wave speed jump across the slab interface, we choose the wave speed values at points 30 km (on the order of one wavelength of the waveform) above and below the inverted interface for both the inverted models and model FWEA18. This helps avoid the complication due to the trade-off between the interface depth and the wave speed in its vicinity. We also isolate the wave speed jump due to the cold slab from that due to pressure and temperature increase in the ambient mantle, by removing a value of 0.2 km s^{-1}, for both the inverted models and model FWEA18.

The inverted wave speed jump across the slab upper surface is ∼3.9 per cent and ∼ 4.6 per cent for the northern and southern regions, respectively. These wave speed jumps are about twice the values from the averaged FWEA18 model, that is ∼2.0 and ∼2.4 per cent for the northern and southern regions, respectively. The 2.0–2.4 per cent wave speed perturbations for the slab can neither fit the relative timing nor produce the extra triplicated phase observed in the data (Figs 7d and g), therefore can be treated as the minimum limit value for the subducting slab in this region. As for the 3.9–4.6 per cent wave speed jump from our inverted results, they are robustly constrained and not affected much by the slab interface thickness, for example a sharp interface or a 60-km thick interface (Supporting Information Fig. S3a). Nevertheless, our inverted wave speed jump of 3.9–4.6 per cent can be viewed as the maximum limit value because these values might be overestimated when the slab perturbations, in the source region, are not accounted for. Therefore, the wave speed jump across the slab upper surface, over a depth range of 60 km, should be between 2.4 and 4.6 per cent. This relatively larger wave speed perturbation of the slab is consistent with the value in the slab core (∼5 per cent) discovered in the same Kuril subduction zone by modelling the teleseismic waveforms recorded in the downdip direction (Zhan *et al.* 2014).

5.4 Future improvement

Because triplications are most sensitive to the turning points of the wave paths, we only inverted for the structures in the deeper region where there are enough path crossings. However, the effect of shallow structures should be considered. Although triplication can minimize the influence from the shallower structure due to the similar wave paths in the shallower part (e.g. Li *et al.* 2021), when the unconstrained structure is close to the depth range to be inverted, the triplicated waveforms can still be influenced (Fig. 4a). To overcome this, the shallow part can be pre-constrained with *a priori* information from independent studies (e.g. Chu & Helmberger 2014). In this study, for the region above 300 km, we use the value from the averaged model FWEA18 to minimize the influence from the shallower structure. In the results, the inverted wave speeds are consistent with the averaged values of model FWEA18 even down to 360 km (Figs 7f and i), which confirms that choosing the pre-constrained model down to the depth of 300 km is reasonable.

Strong lateral heterogeneities can also influence the FWI results based on 1-D assumptions. Previous studies indicate that 2-D and 3-D slab structures near the wave path turning points can affect the triplicated waveforms (e.g. Wang *et al.* 2014 and Tao *et al.* 2017). To minimize such influence from lateral heterogeneities, we purposefully choose the event–station configuration with wave paths roughly parallel to the slab's depth contours where the slab structure can be approximated as 1-D near the turning points (Figs 7a and c). However, near the earthquake source, the high wave speed slab is roughly parallel to the ray paths. This accumulated effect of the source-side wave speed perturbation along the ray paths cannot be neglected (Li *et al.* 2016). Otherwise, the slab wave speed perturbation near the wave path turning points can be overestimated and our inverted wave speed jump of 3.9–4.6 per cent can only be viewed as the upper bound. To overcome this, the current 1-D simulation tool QSEIS can be replaced by 2-D simulation tools in the future, which can take into account the lateral heterogeneities either in the shallower part or near the source side with *a priori* information, for example tomography results from other studies. To reduce the computational costs for our non-gradient-based inversion approach, a GPU version of the 2-D FD method (Li *et al.* 2014) can be utilized. The current non-gradient-based framework still works with the fast speed of the GPU-based 2-D FD simulations, although it is beyond the scope of this study.

We note that this work is a case study showing the effectiveness of the high-frequency (∼2 s) triplications in resolving MTZ discontinuities and the slab upper surface. More events and stations are needed to better constrain the model because the full-waveform inversion results in this study strongly depend on the details of high-quality waveforms. For instance, in the northern region, if we consider another model with a low wave speed zone between the 410-km discontinuity and the slab upper surface (Supporting Information Fig. S4c), the waveform fitting is similar but slightly different (Figs S4a and d) from the model without this zone (Supporting Information Fig. S4b). Due to the limited number of stations, the details of the waveform cannot be verified as consistent and robust waveform features or subjected to noise contamination. To definitively discriminate these two models, more high-quality waveforms with better spatial coverage are required, not only for

1-D full-waveform inversion but also for obtaining a 3-D model of discontinuity structure in this region (e.g. Stahler *et al.* 2012; Takeuchi *et al.* 2014).

6 CONCLUSIONS

Triplicated body waves effectively sample the structure near the transition zone and carry rich information of the discontinuities regarding their depths and wave speed gradients. The 1-D non-gradient-based waveform inversion of triplication waveforms is a useful and efficient tool in accurately mapping the MTZ structural details to the first order from the high-frequency waveforms.

We systematically analysed the trade-off between the depth of the discontinuity and the low wave speed gradient above it, discussed the necessity of using array-normalized amplitude, and applied the non-gradient-based waveform inversion to obtain 1-D structures below the Tatar Strait of Russia.

We have observed triplications due to both the 410-km discontinuity and the slab upper surface, the seismic structures of which are simultaneously inverted. Our derived 410-km discontinuity depths for the northern and southern regions are at 420±15 and 425±15 km, respectively, with no observable uplift. The average depth of the slab upper surface is inverted to be located about 50–70 km below the 410-km discontinuity. This location is between the 1 and 2 per cent P-wave speed perturbation contours of the regional tomography results (Tao *et al.* 2018), but we found twice the amplitude of the wave speed perturbation. A strong wave speed jump between 2.4 and 4.6 per cent (potentially over a depth range of 60 km) is necessary to both fit the differential traveltime between main phases and generate an extra triplication phase observed in the data. Our inverted wave speed jump across the slab upper surface is consistent with the strong wave speed perturbation of ∼5 per cent in the cold slab core (Zhan *et al.* 2014) as well as the results from the residual sphere method (Ding & Grand 1994) in the same region. Our study also indicates that full-waveform inversion at a relatively higher frequency band (∼2 s) is required to resolve the detailed and precise structures near the MTZ. Due to the prohibitive computational cost with 3-D full-waveform inversion, the method used in this study provide an incremental yet effective approach to probe the MTZ structure perturbed by the subducting slabs.

ACKNOWLEDGEMENTS

Seismic records used in this study came from the CEArray and the NECESSArray, and we thank the team members for their deployments. We thank editor Ebru Bozdağ, assistant editor Fern Storey, reviewer Maria Koroni and another anonymous reviewer for their constructive suggestions. We thank Chunquan Yu for helping with the 2-D simulation and Wang-Ping Chen, Kai Tao, Shawn S. Wei, Mingda Lv, Xiaobo He and Ziyi Xi for valuable discussion. We acknowledge the course 'English Composition for Geophysical Research' by Li Zhao of Peking University for help in improving this manuscript. We thank the IRIS Data Management Center for the access to waveforms used in the focal depth inversion. We thank the Institute for Cyber-Enabled Research (ICER) at Michigan State University, the Extreme Science and Engineering Discovery Environment (XSEDE supported by NSF grant ACI-1053575), and the High-performance Computing Platform of Peking University for providing the high-performance computing resources. The map-view figure is produced using the GMT software of Wessel & Smith (1998). This research was supported by NSF grant 1802247 and the startup fund of Min Chen at Michigan State University.

This section is in memory of our co-author Dr Min Chen, a seismologist who passed away on 2021 July 18. Min was our loved colleague, advisor, teacher, and friend. She also worked tirelessly to improve diversity, equity and inclusion in the geoscience community. Min was always passionate about life and always ready to help others all over the world—from Asia, America to Africa. She indeed passed on her kind heart and smile to every friend. Min left us too soon. The world lost a growing mind in seismology, a thoughtful advisor and a good mother (most importantly for her two daughters, Vivian and Mia). It is very hard to say goodbye to Min. With deepest condolences, we hope Min will rest in peace forever.

DATA AVAILABILITY

The data underlying this paper will be shared on reasonable request to the corresponding author.

CONFLICT OF INTEREST

The authors acknowledge that there are no conflicts of interest recorded.

REFERENCES

Green, H.W., II & Burnley, P.C., 1989. A new self-organizing mechanism for deep-focus earthquakes, *Nature,* **341**(6244), 733–737, doi.org/10.1038/341733a0.

Benz, H.M. & Vidale, J.E., 1993. Sharpness of upper-mantle discontinuities determined from high-frequency reflections, *Nature,* **365**(6442), 147–150.

Bercovici, D. & Karato, S.I., 2003. Whole-mantle convection and the transition-zone water filter, *Nature,* **425**(6953), 39–44.

Bina, C.R. & Helffrich, G., 1994. Phase transition Clapeyron slopes and transition zone seismic discontinuity topography, *J. geophys. Res.,* **99**(B8), 15 853–15 860.

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N. & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophys. J. Int.,* **207**(3), 1739–1766.

Brudzinski, M.R. & Chen, W.P., 2000. Variations in P wave speeds and outboard earthquakes: evidence for a petrologic anomaly in the mantle transition zone, *J. geophys. Res.,* **105**(B9), 21 661–21 682.

Brudzinski, M.R. & Chen, W.P., 2003. A petrologic anomaly accompanying outboard earthquakes beneath Fiji-Tonga: corresponding evidence from broadband *P* and *S* waveforms, *J. geophys. Res.,* **108**(B6), doi:10.1029/2002JB002012.

Chapman, C.H., 1978. A new method for computing synthetic seismograms, *Geophys. J. Int.,* **54**(3), 481–518.

Chen, Y.J. & Pei, S., 2010. Tomographic structure of East Asia: II. Stagnant slab above 660 km discontinuity and its geodynamic implications, *Earthq. Sci.,* **23**(6), 613–626.

Chevrot, S., Vinnik, L. & Montagner, J.P., 1999. Global-scale analysis of the mantle Pds phases, *J. geophys. Res.,* **104**(B9), 20 203–20 219.

Chu, R. & Helmberger, D., 2014. Lithospheric waveguide beneath the Midwestern United States; massive low-velocity zone in the lower crust, *Geochem. Geophys. Geosyst.,* **15**(4), 1348–1362.

Chu, R., Schmandt, B. & Helmberger, D.V., 2012. Upper mantle *P* velocity structure beneath the Midwestern United States derived from triplicated waveforms, *Geochem. Geophys. Geosyst.,* **13**(2), doi:10.1029/2011GC003818.

Collier, J.D. & Helffrich, G.R., 1997. Topography of the "410" and "660" km seismic discontinuities in the Izu-Bonin subduction zone, *Geophys. Res. Lett.,* **24**(12), 1535–1538.

Crotwell, H.P., Owens, T.J. & Ritsema, J., 1999. The TauP Toolkit: flexible seismic travel-time and ray-path utilities, *Seismol. Res. Lett.,* **70**(2), 154–160.

Ding, X.Y. & Grand, S.P., 1994. Seismic structure of the deep Kurile subduction zone, *J. geophys. Res.,* **99**(B12), 23 767–23 786.

Flanagan, M.P. & Shearer, P.M., 1998. Global mapping of topography on transition zone velocity discontinuities by stacking SS precursors, *J. geophys. Res.,* **103**(B2), 2673–2692.

Flanagan, M.P. & Shearer, P.M., 1999. A map of topography on the 410-km discontinuity from PP precursors, *Geophys. Res. Lett.,* **26**(5), 549–552.

Gao, W., Matzel, E. & Grand, S.P., 2006. Upper mantle seismic structure beneath eastern Mexico determined from P and S waveform inversion and its implications, *J. geophys. Res.,* **111**(B8), doi:10.1029/2006JB004304.

Grand, S.P. & Helmberger, D.V., 1984. Upper mantle shear structure of North America, *Geophys. J. Int.,* **76**(2), 399–438.

Gu, Y.J. & Dziewonski, A.M., 2002. Global variability of transition zone thickness, *J. geophys. Res.,* **107**(B7), ESE 2–1-ESE 2-17.

Guo, Z. & Zhou, Y., 2020. Finite-frequency imaging of the global 410- and 660-km discontinuities using SS precursors, *Geophys. J. Int.,* **220**(3), 1978–1994.

Han, G., Li, J., Guo, G., Mooney, W.D., Karato, S.I. & Yuen, D.A., 2021. Pervasive low-velocity layer atop the 410-km discontinuity beneath the northwest Pacific subduction zone: implications for rheology and geodynamics, *Earth planet. Sci. Lett.,* **554,** 116642.

Helffrich, G.R. & Wood, B.J., 1996. 410 km discontinuity sharpness and the form of the olivine α-β phase diagram: resolution of apparent seismic contradictions, *Geophys. J. Int.,* **126**(2), F7–F12.

Houser, C., Masters, G., Flanagan, M. & Shearer, P., 2008. Determination and analysis of long-wavelength transition zone structure using SS precursors, *Geophys. J. Int.,* **174**(1), 178–194.

Huang, J. & Zhao, D., 2006. High-resolution mantle tomography of China and surrounding regions, *J. geophys. Res.,* **111**(B9), doi:10.1029/2005JB004066.

Isacks, B. & Molnar, P., 1971. Distribution of stresses in the descending lithosphere from a global survey of focal-mechanism solutions of mantle earthquakes, *Rev. Geophys.,* **9**(1), 103–174.

Johnson, L.R., 1967. Array measurements of P velocities in the upper mantle, *J. geophys. Res.,* **72**(24), 6309–6325.

Kawakatsu, H. & Watada, S., 2007. Seismic evidence for deep-water transportation in the mantle, *Science,* **316**(5830), 1468–1471.

Kennett, B.L.N. & Engdahl, E.R., 1991. Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.,* **105**(2), 429–465.

Kirby, S.H., Durham, W.B. & Stern, L.A., 1991. Mantle phase changes and deep-earthquake faulting in subducting lithosphere, *Science,* **252**(5003), 216–225.

Koper, K.D., Wysession, M.E. & Wiens, D.A., 1999. Multimodal function optimization with a niching genetic algorithm: a seismological example, *Bull. seism. Soc. Am.,* **89**(4), 978–988.

Koroni, M. & Trampert, J., 2021. Imaging global mantle discontinuities: a test using full-waveforms and adjoint kernels, *Geophys. J. Int.,* **226,** 1498–1516,

Lawrence, J.F. & Shearer, P.M., 2008. Imaging mantle transition zone thickness with SdS-SS finite-frequency sensitivity kernels, *Geophys. J. Int.,* **174**(1), 143–158.

LeFevre, L.V. & Helmberger, D.V., 1989. Upper mantle P velocity structure of the Canadian shield, *J. geophys. Res.,* **94**(B12), 17 749–17 765.

Li, D., Helmberger, D., Clayton, R.W. & Sun, D., 2014. Global synthetic seismograms using a 2-D finite-difference method, *Geophys. J. Int.,* **197**(2), 1166–1183.

Li, G., Bai, L., Zhou, Y., Wang, X. & Cui, Q., 2017. Velocity structure of the mantle transition zone beneath the southeastern margin of the Tibetan Plateau, *Tectonophysics,* **721,** 349–360.

Li, G., Li, Y.E., Zhang, H., Bai, L., Ding, L., Li, W., Cui, Q. & Zhou, Y., 2019a. Detection of a thick and weak low-velocity layer atop the mantle transition zone beneath the Northeastern South China Sea from triplicated *P*-wave waveform modeling, *Bull. seism. Soc. Am.,* **109**(4), 1181–1193.

Li, J., Chen, M., Koper, K.D., Zhou, T., Xi, Z., Li, S. & Li, G., 2021. FastTrip: a fast MPI-accelerated 1D triplication waveform inversion package for constraining mantle transition zone discontinuities, *Seismol. Res. Lett.,* **92**(4), 2647–2656.

Li, J., Wang, S., Cai, C. & Ning, J., 2016. A computational scheme for quantitatively removing the effects of lateral velocity variation on 1-D triplicated wave velocity inversion, *Acta Sci. Nat. Univ. Pekin.,* **52**(3), 420–426.

Li, L., Chen, Y.W., Zheng, Y., Hu, H. & Wu, J., 2019b. Seismic evidence for plume-slab interaction by high-resolution imaging of the 410-km discontinuity under Tonga, *Geophys. Res. Lett.,* **46**(23), 13 687–13 694.

Li, S., Wang, Y., Liang, Z., He, S. & Zeng, W., 2012. Crustal structure in southeastern Gansu from regional seismic waveform inversion, *Chin. J. Geophys.,* **55**(2), 206–218.

Melbourne, T. & Helmberger, D., 1998. Fine structure of the 410-km discontinuity, *J. geophys. Res.,* **103**(B5), 10 091–10 102.

Molnar, P., Freedman, D. & Shih, J.S., 1979. Lengths of intermediate and deep seismic zones and temperatures in downgoing slabs of lithosphere, *Geophys. J. Int.,* **56**(1), 41–54.

Niazi, M. & Anderson, D.L., 1965. Upper mantle structure of western North America from apparent velocities of P waves, *J. geophys. Res.,* **70**(18), 4633–4640.

Niu, F., Levander, A., Ham, S. & Obayashi, M., 2005. Mapping the subducting Pacific slab beneath southwest Japan with Hi-net receiver functions, *Earth planet. Sci. Lett.,* **239**(1-2), 9–17.

Nolet, G. & Dahlen, F.A., 2000. Wave front healing and the evolution of seismic delay times, *J. geophys. Res.,* **105**(B8), 19043–19054.

Obayashi, M.,Yoshimitsu, J. Nolet, G. Fukao, Y. Shiobara, H. Sugioka, H. Miyamachi , H. & Gao, Y. 2013. Geophysical Research Letters *5652,*, 00948276

Revenaugh, J. & Sipkin, S.A., 1994. Seismic evidence for silicate melt atop the 410-km mantle discontinuity, *Nature,* **369**(6480), 474–476.

Ringwood, A.E., 1975. *Composition and Petrology of the Earth's Mantle,* MacGraw-Hill, pp. 618.

Ritsema, J., Cupillard, P., Tauzin, B., Xu, W., Stixrude, L. & Lithgow-Bertelloni, C., 2009a. Joint mineral physics and seismic wave traveltime analysis of upper mantle temperature, *Geology,* **37**(4), 363–366.

Ritsema, J., Xu, W., Stixrude, L. & Lithgow-Bertelloni, C., 2009b. Estimates of the transition zone temperature in a mechanically mixed upper mantle, *Earth planet. Sci. Lett.,* **277**(1-2), 244–252.

Schmandt, B., Dueker, K.G., Hansen, S.M., Jasbinsek, J.J. & Zhang, Z., 2011. A sporadic low-velocity layer atop the western US mantle transition zone and short-wavelength variations in transition zone discontinuities, *Geochem. Geophys. Geosyst.,* **12**(8), doi:10.1029/2011GC003668.

Schmerr, N. & Garnero, E.J., 2007. Upper mantle discontinuity topography from thermal and chemical heterogeneity, *Science,* **318**(5850), 623–626.

Shearer, P.M., 2000. Upper mantle seismic discontinuities, *Geophys. Monogr. Ser.,* **117,** 115–132.

Song, T.R.A., Helmberger, D.V. & Grand, S.P., 2004. Low-velocity zone atop the 410-km seismic discontinuity in the northwestern United States, *Nature,* **427**(6974), 530–533.

Stähler, S.C., Sigloch, K. & Nissen-Meyer, T., 2012. Triplicated *P*-wave measurements for waveform tomography of the mantle transition zone, *Solid Earth,* **3**(2), 339–354.

Stein, S. & Wysession, M., 2009. *An Introduction to Seismology, Earthquakes, and Earth Structure*. John Wiley & Sons.

Tajima, F. & Grand, S.P., 1995. Evidence of high velocity anomalies in the transition zone associated with southern Kurile subduction zone, *Geophys. Res. Lett.,* **22**(23), 3139–3142.

Takeuchi, N. *et al.*, 2014. Upper mantle tomography in the northwestern Pacific region using triplicated P waves, *J. geophys. Res.,* **119**(10), 7667–7685.

Tao, K., Grand, S.P. & Niu, F., 2017. Full-waveform inversion of triplicated data using a normalized-correlation-coefficient-based misfit function, *Geophys. J. Int.,* **210**(3), 1517–1524.

Tao, K., Grand, S.P. & Niu, F., 2018. Seismic structure of the upper mantle beneath eastern Asia from full waveform seismic tomography, *Geochem. Geophys. Geosyst.,* **19**(8), 2732–2763.

Thirot, J.L., Montagner, J.P. & Vinnik, L., 1998. Upper-mantle seismic discontinuities in a subduction zone (Japan) investigated from P to S converted waves, *Phys. Earth planet. Inter.,* **108**(1), 61–80.

Thompson, A.B., 1992. Water in the Earth's upper mantle, *Nature,* **358**(6384), 295–302.

Tian, D., Lv, M., Wei, S.S., Dorfman, S.M. & Shearer, P.M., 2020. Global variations of Earth's 520- and 560-km discontinuities, *Earth planet. Sci. Lett.,* **552,** 116600.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160**(1), 195–216.

Van der Meijde, M., Marone, F., Giardini, D. & Van der Lee, S., 2003. Seismic evidence for water deep in Earth's upper mantle, *Science,* **300**(5625), 1556–1558.

Vidale, J.E. & Benz, H.M., 1992. Upper-mantle seismic discontinuities and the thermal structure of subduction zones, *Nature,* **356**(6371), 678–683.

Vinnik, L.P., 1977. Detection of waves converted from P to SV in the mantle, *Phys. Earth planet. Inter.,* **15**(1), 39–45.

Wang, R., 1999. A simple orthonormalization method for stable and efficient computation of Green's functions, *Bull. seism. Soc. Am.,* **89**(3), 733–741.

Wang, T. & Chen, L., 2009. Distinct velocity variations around the base of the upper mantle beneath northeast Asia, *Phys. Earth planet. Inter.,* **172**(3-4), 241–256.

Wang, T., Revenaugh, J. & Song, X., 2014. Two-dimensional/three-dimensional waveform modeling of subducting slab and transition zone beneath Northeast Asia, *J. geophys. Res.,* **119**(6), 4766–4786.

Wang, X., Li, J. & Chen, Q.F., 2017. Topography of the 410 km and 660 km discontinuities beneath the Japan Sea and adjacent regions by analysis of multiple-ScS waves, *J. geophys. Res.,* **122**(2), 1264–1283.

Wang, Y., Wen, L. & Weidner, D., 2009. Array triplication data constraining seismic structure and composition in the mantle, *Surv. Geophys.,* **30**(4), 355–376.

Wei, S.S. & Shearer, P.M., 2017. A sporadic low-velocity layer atop the 410 km discontinuity beneath the Pacific Ocean, *J. geophys. Res.,* **122**(7), 5144–5159.

Wei, S.S., Shearer, P.M., Lithgow-Bertelloni, C., Stixrude, L. & Tian, D., 2020. Oceanic plateau of the Hawaiian mantle plume head subducted to the uppermost lower mantle, *Science,* **370**(6519), 983–987.

Wessel, P. & Smith, W.H., 1998. New, improved version of Generic Mapping Tools released, *EOS, Trans. Am. geophys. Un.,* **79**(47), 579.

Zhan, Z., Helmberger, D.V. & Li, D., 2014. Imaging subducted slab structure beneath the Sea of Okhotsk with teleseismic waveforms, *Phys. Earth planet. Inter.,* **232,** 30–35.

Zhang, M., Sun, D., Wang, Y. & Wu, Z., 2019. Fine structure of the 660-km discontinuity beneath southeastern China, *Geophys. Res. Lett.,* **46**(13), 7304–7314.

Zheng, X.F., Yao, Z.X., Liang, J.H. & Zheng, J., 2010. The role played and opportunities provided by IGP DMC of China National Seismic Network in Wenchuan earthquake disaster relief and researches, *Bull. seism. Soc. Am.,* **100**(5B), 2866–2872.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1: Displacement waveform comparison between aligned data (bold grey) and IASP91 synthetics (red) in the northern region (a) and southern region (b).

Figure S2: Modelling tests for the 410-km discontinuity sharpness. (a) Models used in this test. The black line is the model IAPS91, while the blue and red lines are models in which the 410-km discontinuity is replaced by a gradual transition with thicknesses of 20 km and 40 km, respectively. (b) The traveltime curves for models in (a). (c) Synthetic waveforms for models in (a). Although there are significant differences in the traveltime curves as shown in (b), the waveforms are almost the same with this period of 3 s. (d) Waveforms comparison with a period of 2 s. (e) Waveforms comparison with a period of 1 s. We note that for all these cases, a $t^* \sim 1$ s is convolved.

Figure S3: Inversion tests for slab upper surface thickness in the southern region. (a) Models. The shaded red region marks the accepted models with interface thickness from 20 to 60 km. The bold blue line shows the unaccepted model with a thickness of 80 km. The bold grey line indicates the model IASP91. The dashed black line is the averaged value for model FWEA18 (from 5° to 14°). (b) Waveforms comparison between data (bold grey), and the best-fitting model (red) and the model with a gradual interface of 80-km thickness (blue).

Figure S4: Inversion tests for the low wave speed zone in the northern region. (a) Waveforms comparison between data (bold grey), and the best-fitting model (red) without a low wave speed zone. (b) The shaded red region marks the accepted models without a low wave speed zone. The solid grey line indicates the model IASP91. The dashed black line is the averaged value for model FWEA18 (from 5° to 14°). (c) The shaded red region marks the accepted models with a low wave speed zone between the 410-km discontinuity and the slab upper surface. Other symbols are the same as (b). (d) Waveforms comparison between data (bold grey), and the best-fitting model (red) with a low wave speed zone. The grey circle indicates the waveform differences compared with (a).