Geophysical Journal International

Geophys. J. Int. (2023) **00,** 1–22
Advance Access publication 2023 October 23
GJI Seismology

https://doi.org/10.1093/gji/ggad416

Adjoint-state differential arrival time tomography

Ping Tong [1,2,3] Tianjue Li [1,3] Jing Chen [1] and Masaru Nagaso[1]

[1] *Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University,* 637371 *Singapore.*
E-mail: tongping@ntu.edu.sg
[2] *Asian School of the Environment, Nanyang Technological University,* 639798 *Singapore*
[3] *Earth Observatory of Singapore, Nanyang Technological University,* 639798 *Singapore*

Accepted 2023 October 17. Received 2023 October 13; in original form 2023 March 23

SUMMARY

The recently developed adjoint-state traveltime tomography (ATT) method offers an alternative approach to conduct traveltime tomography without the need for ray tracing or waveform modelling. Instead, it utilizes the eikonal equation to depict the minimal traveltime field from an earthquake location to any position in the computational domain. The process of tomographic inversion is formulated as an optimization problem with the goal of minimizing the difference between observed and theoretical first arrival times, which is subsequently solved using the efficient adjoint method. One advantage of differential arrival time data is that it cancels or reduces the influence of common factors, making it more sensitive to a specific subset of model parameters compared to first arrival times. To take advantage of this property, two variants of the ATT method are derived to determine velocity structure and earthquake locations in this study. The first variant, adjoint-state common-source differential arrival time tomography (ATT-CS), uses common-source differential arrival times; while the second variant, adjoint-state common-receiver differential arrival time tomography (ATT-CR), inverts common-receiver differential arrival times. Numerical examples demonstrate that the ATT-CS method is a valuable tool for imaging receiver-side fine-scale velocity structures. Conversely, the ATT-CR method is well suited for resolving source-side velocity structures. Differential arrival times also place constraints on earthquake locations. Compared to common-source differential arrival times, common-receiver differential arrival times are less sensitive to velocity errors and suitable for earthquake location determination. Both common-source and common-receiver differential arrival times are considered first-order differential arrival times. To demonstrate the ease with which the ATT method can be extended to higher-order differential arrival times, we also derive the adjoint-state second-order differential arrival time tomography method. Finally, we discuss how the adjoint-state tomography methods address multipathing.

Key words: Inverse theory; Seismic tomography; Wave propagation.

1 INTRODUCTION

Differential arrival times provide valuable constraints on earthquake locations and seismic structures of the Earth's interior. For instance, the *S-P* differential arrival time and *sP-P* differential arrival time can be used to determine earthquake locations with high precision (e.g. Umino *et al.* 1995; Waldhauser & Ellsworth 2000; Guo & Zhang 2017). In this study, we refer to the differential arrival time of the first *P*- or *S*-waves from two adjacent earthquakes to a common seismic station as common-receiver differential arrival time, and similarly for common-source differential arrival time (Fig. 1). Common-source and common-receiver differential arrival times have been extensively used to reveal velocity heterogeneity of the crust and upper mantle (e.g. Zhang & Thurber 2003; Share *et al.* 2019; Liu *et al.* 2021; Guo *et al.* 2021; Chen *et al.* 2022), such as teleseimic traveltime tomography frequently using common-source differential arrival times (e.g. Liu *et al.* 2018; Boyce *et al.* 2019). Moreover, *S-P* differential arrival times can be inverted to determine the V_P/V_S ratio structure (e.g. Thurber *et al.* 1995; Zenonos *et al.* 2020), *PP-P* differential arrival times are sensitive to the seismic structure of the upper mantle (Obayashi *et al.* 2004), and *SmKS-SKKS* differential arrival times impose constraints on the outermost core structure (Wu & Irving 2020). Additionally, though infrequently used in previous studies, second-order differential arrival times (also known as triple-difference arrival times in previous research) have proven effective in determining seismic velocity structure and accurately locating earthquakes (e.g. Guo & Zhang 2017; Guo *et al.* 2021). Note

© The Author(s) 2023. Published by Oxford University Press on behalf of The Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

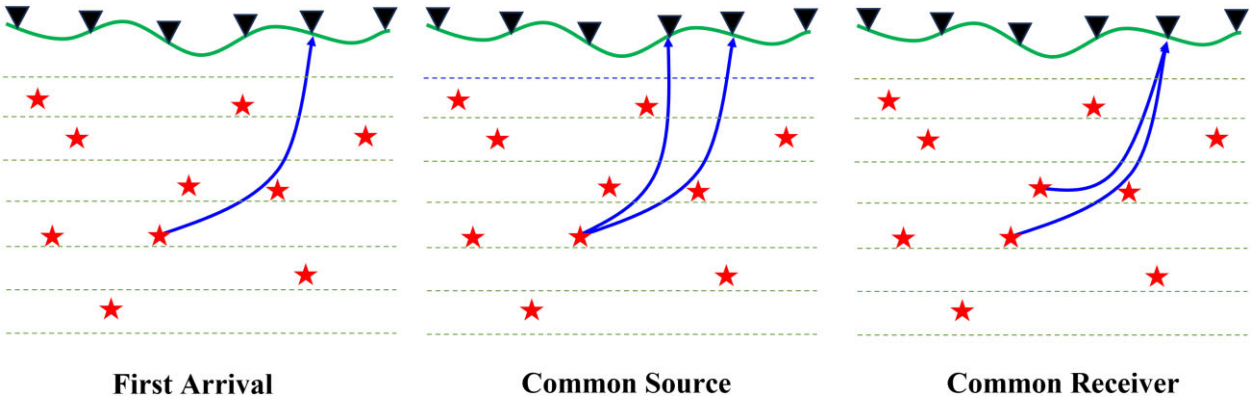

Figure 1. A schematic diagram showing the paths (blue curves) of a first arrival (left-hand panel), common-source differential arrival (middle panel) and common-receiver differential arrival (right-hand panel) used in seismic tomographic inversions. The red stars denote earthquakes. Black inverted triangles are seismic stations deployed on the surface.

that a second-order differential arrival time is the difference between two common-source differential arrival times recorded at the same station pair, or the difference between two common-receiver differential arrival times generated by the same earthquake pair. In this scenario, common-source differential arrival times and common-receiver differential arrival times are first-order differential arrival times. With the increasing number of dense seismic arrays worldwide and the growing use of fibre-optic distributed acoustic sensing in recent years (e.g. Zhan 2020; Cheng *et al.* 2021), we anticipate an exponential growth in the number of measurable differential arrival times from high-quality seismological recordings.

This study focuses on seismic tomographic inversions using common-source differential arrival times and common-receiver differential arrival times to construct subsurface velocity models and determine earthquake locations. As an alternative to ray-based traveltime tomography methods (e.g. Thurber 1983; Zhang & Thurber 2003; Liu & Tong 2021), our approach uses the eikonal equation to simulate the traveltime field of the first *P* or *S* wave, and the inverse problem of traveltime tomography is solved by the efficient adjoint method (e.g. Leung & Qian 2006; Taillandier *et al.* 2009; Tavakoli *et al.* 2017; Tong 2021a). The eikonal equation approach has the advantage of accurately computing the minimal traveltime from the earthquake location to any position of interest (Rawlinson *et al.* 2008). Meanwhile, the adjoint method is a popular method for the efficient computation of the gradient of an objective function in a mathematical optimization problem (e.g. Tromp *et al.* 2005; Fichtner *et al.* 2006; Leung & Qian 2006). The so-called adjoint-state traveltime tomography (ATT) method recently developed by Tong (2021a) is one example that uses the adjoint method to solve the eikonal equation-constrained optimization problem. Unlike some other methods of the same type (e.g. Leung & Qian 2006; Taillandier *et al.* 2009), the ATT method does not require the receivers to be placed on the boundary of the computational or physical domain. Furthermore, it uses the good-performing multiple-grid model parametrization to generate reliable inversion results. Importantly, the computational cost of the ATT method is one order of magnitude lower than that of wave equation-based seismic tomography methods.

In this study, our goal is to develop tomographic inversion schemes that make use of both common-source differential arrival times and common-receiver differential arrival times within the ATT framework. To facilitate the development, we provide a brief review of the ATT method in Section 2, including the implementation of the multiple-grid model parametrization. In Section 3, we derive the sensitivity kernels of the objective function with respect to velocity structure and earthquake parameters for both common-source and common-receiver differential arrival time tomography methods. In Section 4, we provide numerical tests that display various examples of sensitivity kernels and the performance of the ATT method and the two differential arrival time tomography methods in recovering subsurface velocity heterogeneity and earthquake locations. In Section 5, we examine the resolving abilities of the different adjoint-state tomography methods and elucidate how they address the challenge of multipathing. Finally, in Appendix A, we demonstrate the extension of the adjoint-state method to higher-order differential arrival times by deriving the second-order differential arrival time tomography method. In Appendix B, the eikonal equation-based grid search method for earthquake location is derived.

2 REVIEW OF ATT

We consider N earthquakes including natural earthquakes, induced seismicity and artificial explosions that occur at various locations $\mathbf{x}_{s,n}$ (s is for source and $n = 1, 2, \cdots, N$) within the Earth volume Ω, along with M seismic stations $\mathbf{x}_{r,m}$ (r is for receiver and $m = 1, 2, \cdots, M$) situated at or near the Earth's surface $\partial\Omega$. The eikonal equation with a point source condition describes the traveltime $T_n(\mathbf{x})$ of the first P or S wave from the nth earthquake to any position $\mathbf{x}$, which can be expressed as

$$\nabla T_n(\mathbf{x}) \cdot \nabla T_n(\mathbf{x}) = s^2(\mathbf{x}), \qquad T_n(\mathbf{x}_{s,n}) = 0. \tag{1}$$

Here $s(\mathbf{x})$ is the P- or S-wave slowness, the reciprocal of wave speed (velocity). Throughout this discussion, slowness and velocity will be used interchangeably. The arrival time observations $T_n^o(\mathbf{x}_{r,m})$ at these M seismic stations can be inverted to retrieve an optimal slowness model

by minimizing the objective function

$$\chi\left(s(\mathbf{x})\right)=\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{\omega_{n,m}}{2}\left\{T_n(\mathbf{x}_{r,m})-\left[T_n^o(\mathbf{x}_{r,m})-\tau_n\right]\right\}^2. \tag{2}$$

The weight coefficient $\omega_{n,m}$ can be adjusted based on the quality and existence of the associated arrival time $T_n^o(\mathbf{x}_{r,m})$. τ_n is the origin time of the nth earthquake. We distinguish between the theoretical traveltime from the earthquake location $\mathbf{x}_{s,n}$ to the receiver location $\mathbf{x}_{r,m}$, denoted as $T_n(\mathbf{x}_{r,m})$, and the observed traveltime, referred to as $T_n^o(\mathbf{x}_{r,m})-\tau_n$, with $T_n^o(\mathbf{x}_{r,m})$ being the observed arrival time.

An infinitesimal perturbation $\delta s(\mathbf{x})$ in the slowness model $s(\mathbf{x})$ results in an infinitesimal perturbation $\delta\chi(s(\mathbf{x}))$ in the objective function $\chi(s(\mathbf{x}))$. Tong (2021a) derived a mathematical relationship between these two perturbations, given by

$$\delta\chi\left(s(\mathbf{x})\right)=\int_{\Omega}\sum_{n=1}^{N}P_n(\mathbf{x})s^2(\mathbf{x})\frac{\delta s(\mathbf{x})}{s(\mathbf{x})}\mathrm{d}\mathbf{x}, \tag{3}$$

where $P_n(\mathbf{x})$ is the adjoint traveltime field that satisfies

$$\begin{cases}\nabla\cdot\left[P_n(\mathbf{x})\nabla\left(-T_n(\mathbf{x})\right)\right]=\sum_{m=1}^{M}\omega_{n,m}\left\{T_n(\mathbf{x}_{r,m})-\left[T_n^o(\mathbf{x}_{r,m})-\tau_n\right]\right\}\delta(\mathbf{x}-\mathbf{x}_{r,m}), & \mathbf{x}\in\Omega\\ P_n(\mathbf{x})=0, & \mathbf{x}\in\partial\Omega.\end{cases} \tag{4}$$

The misfit kernel $K_s(\mathbf{x})$ that represents the Fréchet derivative of the objective function $\chi(\mathbf{x})$ with respect to the slowness $s(\mathbf{x})$ can be expressed by

$$K_s(\mathbf{x})=\sum_{n=1}^{N}P_n(\mathbf{x})s^2(\mathbf{x})=\sum_{n=1}^{N}K_{s,n}(\mathbf{x})=\sum_{n=1}^{N}\sum_{m=1}^{M}K_{s,n,m}(\mathbf{x}), \tag{5}$$

Here, $K_{s,n}(\mathbf{x})=P_n(\mathbf{x})s^2(\mathbf{x})$ is the event kernel associated with the nth earthquake, and $K_{s,n,m}(\mathbf{x})$ is the individual kernel related to the pair of the nth earthquake and mth seismic station. To compute the objective function value and the misfit kernel, we can use the efficient fast sweeping method (Zhao 2004; Leung & Qian 2006) to solve both the eikonal eq. (1) and the adjoint eq. (4).

The multiple-grid approach, proposed by Tong *et al.* (2019) and Tong (2021a), can be used to discretize the relative slowness perturbation $\delta s(\mathbf{x})/s(\mathbf{x})$ in eq. (3). The approach uses H sets of component grids, each with L_h grid nodes. It defines a series of basis functions $B_{l,h}(\mathbf{x})$ that have a support (non-zero values) locally around the lth node of the hth grid. In this study, the basis functions for trilinear interpolation are chosen due to their simplicity and effectiveness. With any chosen basis functions, slowness $s(\mathbf{x})$ is approximated by $\ln s(\mathbf{x})=\sum_{l=1}^{L_h}C_{l,h}B_{l,h}(\mathbf{x})$, where $C_{l,h}$ are the coefficients. Further, the relative slowness perturbation $\delta s(\mathbf{x})/s(\mathbf{x})$ can be expressed by

$$\frac{\delta s(\mathbf{x})}{s(\mathbf{x})}=\sum_{l=1}^{L_h}\delta C_{l,h}B_{l,h}(\mathbf{x}). \tag{6}$$

The multiple-grid model parametrization takes the average of all the interpolating functions as an estimate for the relative slowness perturbation

$$\frac{\delta s(\mathbf{x})}{s(\mathbf{x})}=\frac{1}{H}\sum_{h=1}^{H}\sum_{l=1}^{L_h}\delta C_{l,h}B_{l,h}(\mathbf{x}). \tag{7}$$

Substituting eq. (7) into eq. (3) yields

$$\frac{\partial\chi\left(s(\mathbf{x})\right)}{\partial C_{l,h}}=\sum_{n=1}^{N}\frac{1}{H}\int_{\Omega}P_n(\mathbf{x})s^2(\mathbf{x})B_{l,h}(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{8}$$

The optimization problem (2) in an infinite-dimensional function space is reduced to a minimization problem in a finite-dimensional vector space. The gradient vector for the minimization problem is given by eq. (8). Optimization algorithms such as the step-size-controlled gradient descent method (Tong 2021a) can be utilized to find an optimal solution for $(C_{1,1},C_{2,1},\cdots,C_{L_1,1},C_{1,2},\cdots,C_{L_H,H})$ and ultimately, an optimal slowness or velocity model.

Errors in earthquake locations also contribute to the discrepancy between observed and theoretical traveltimes. When earthquake locations are unknown, it becomes crucial to determine seismic velocity and earthquake locations simultaneously. As a result, the objective function can be expanded from eq. (2) to include hypocentres $\mathbf{x}_{s,n}$ and origin times τ_n, given by

$$\chi\left(s(\mathbf{x}),\ \mathbf{x}_{s,1},\tau_1,\cdots,\mathbf{x}_{s,n},\tau_n,\cdots,\mathbf{x}_{s,N},\tau_N\right)=\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{\omega_{n,m}}{2}\left\{T_n(\mathbf{x}_{r,m})-\left[T_n^o(\mathbf{x}_{r,m})-\tau_n\right]\right\}^2. \tag{9}$$

Following the work of Tong (2021a), the full derivative of the objective function can be expressed by

$$\delta\chi\left(s(\mathbf{x}),\ \mathbf{x}_{s,1},\tau_1,\cdots,\mathbf{x}_{s,n},\tau_n,\cdots,\mathbf{x}_{s,N},\tau_N\right)=\int_\Omega\sum_{n=1}^{N}P_n(\mathbf{x})s^2(\mathbf{x})\frac{\delta s(\mathbf{x})}{s(\mathbf{x})}\mathrm{d}\mathbf{x}$$

$$+\sum_{n=1}^{N}\sum_{m=1}^{M}\omega_{n,m}\left\{T_n(\mathbf{x}_{r,m})-\left[T_n^o(\mathbf{x}_{r,m})-\tau_n\right]\right\}\left[\delta\mathbf{x}_{s,n}\cdot\nabla\Gamma_m(\mathbf{x}_{s,n})+\delta\tau_n\right],\tag{10}$$

where $\Gamma_m(\mathbf{x})$ is the traveltime field initiated at the mth receiver location $\mathbf{x}_{r,m}$, and $\Gamma_m(\mathbf{x}_{s,n})=T_n(\mathbf{x}_{r,m})$ holds based on the reciprocity principle (Aki & Richards 2002). One can utilize eq. (10) in conjunction with multiple-grid model parametrization to determine both velocity structure and earthquake locations simultaneously.

3 ADJOINT-STATE DIFFERENTIAL ARRIVAL TIME TOMOGRAPHY

In comparison with first arrival times, differential arrival times have the potential to generate higher-resolution images of specific target regions. Common-source differential arrival times can mitigate the influence of errors in source locations and near-source structural heterogeneity, thereby improving the resolution of receiver-side models (e.g. Yuan *et al.* 2016; Zhang *et al.* 2017). Conversely, common-receiver differential arrival times can effectively constrain source-side structures by reducing the complex effects caused by receiver-side heterogeneity (e.g. Zhang & Thurber 2003; Thurber *et al.* 2006; Allam & Ben-Zion 2012). Moreover, differential arrival times extracted from similar seismograms generally exhibit higher accuracy compared to first arrival time measurements (e.g. Shearer 1997; Zhang & Thurber 2003).

This section develops two methods of differential arrival time tomography: adjoint-state common-source differential arrival time tomography (ATT-CS) and adjoint-state common-receiver differential arrival time tomography (ATT-CR). Additionally, since accurate earthquake locations are essential for constructing reliable seismic velocity models and these two types of data possess unique capabilities in pinpointing earthquake locations, earthquake location methods using common-source differential arrival times and common-receiver differential arrival times have been developed and incorporated into the two differential arrival time tomography methods.

3.1 Velocity estimation using common-source differential arrival times

The first P- or S-wave arrivals of an earthquake occurring at $\mathbf{x}_{s,n}$ ($n=1,2,\cdots,N$) are recorded by two seismic stations $\mathbf{x}_{r,m}$ and $\mathbf{x}_{r,k}$ ($m,\ k=1,2,\cdots,M$). By comparing the observed arrival time difference $T_n^o(\mathbf{x}_{r,m})-T_n^o(\mathbf{x}_{r,k})$, which is equivalent to the observed traveltime difference in this particular context, to the theoretical traveltime difference $T_n(\mathbf{x}_{r,m})-T_n(\mathbf{x}_{r,k})$ computed using the slowness model $s(\mathbf{x})$, an optimal slowness model for the subsurface structure can be determined. The objective function is defined as

$$\chi\left(s(\mathbf{x})\right)=\sum_{m=1}^{M}\sum_{k=1}^{M}\sum_{n=1}^{N}\frac{\omega_n^{m,k}}{2}\left\{\left[T_n(\mathbf{x}_{r,m})-T_n(\mathbf{x}_{r,k})\right]-\left[T_n^o(\mathbf{x}_{r,m})-T_n^o(\mathbf{x}_{r,k})\right]\right\}^2.\tag{11}$$

The weight coefficient $\omega_n^{m,k}$ reflects the existence and quality of the observed arrival time difference $T_n^o(\mathbf{x}_{r,m})-T_n^o(\mathbf{x}_{r,k})$. $\omega_n^{m,k}=\omega_n^{k,m}$ is always assumed.

Directly differentiating eq. (11) with respect to $T_n(\mathbf{x})$, the variation of the objective function due to an infinitesimal perturbation $\delta s(\mathbf{x})$ in the slowness model $s(\mathbf{x})$ can be expressed by

$$\delta\chi\left(s(\mathbf{x})\right)=\sum_{m=1}^{M}\sum_{k=1}^{M}\sum_{n=1}^{N}\omega_n^{m,k}\left\{\left[T_n(\mathbf{x}_{r,m})-T_n(\mathbf{x}_{r,k})\right]-\left[T_n^o(\mathbf{x}_{r,m})-T_n^o(\mathbf{x}_{r,k})\right]\right\}\left[\delta T_n(\mathbf{x}_{r,m})-\delta T_n(\mathbf{x}_{r,k})\right]$$

$$=\sum_{m=1}^{M}\sum_{k=1}^{M}\sum_{n=1}^{N}\int_\Omega 2\omega_n^{m,k}\left\{\left[T_n(\mathbf{x}_{r,m})-T_n(\mathbf{x}_{r,k})\right]-\left[T_n^o(\mathbf{x}_{r,m})-T_n^o(\mathbf{x}_{r,k})\right]\right\}\delta T_n(\mathbf{x})\delta(\mathbf{x}-\mathbf{x}_{r,m})\mathrm{d}\mathbf{x}.\tag{12}$$

According to eq. (1) and under first-order approximation, the traveltime perturbation $\delta T_n(\mathbf{x})$ caused by $\delta s(\mathbf{x})$ satisfies

$$\nabla\left[T_n(\mathbf{x})\right]\cdot\nabla\left[\delta T_n(\mathbf{x})\right]=s(\mathbf{x})\delta s(\mathbf{x}).\tag{13}$$

Multiply an arbitrary test function $P_n(\mathbf{x})$ on both sides of eq. (13) and integrate over the Earth volume Ω

$$\int_\Omega P_n(\mathbf{x})\nabla\left[T_n(\mathbf{x})\right]\cdot\nabla\left[\delta T_n(\mathbf{x})\right]\mathrm{d}\mathbf{x}=\int_\Omega P_n(\mathbf{x})s(\mathbf{x})\delta s(\mathbf{x})\mathrm{d}\mathbf{x}.\tag{14}$$

Using the divergence theorem and assuming that $P_n(\mathbf{x})=0$ on the boundary $\partial\Omega$, we obtain

$$\int_\Omega\delta T_n(\mathbf{x})\nabla\cdot\left[P_n(\mathbf{x})\nabla\left(-T_n(\mathbf{x})\right)\right]\mathrm{d}\mathbf{x}=\int_\Omega P_n(\mathbf{x})s(\mathbf{x})\delta s(\mathbf{x})\mathrm{d}\mathbf{x}.\tag{15}$$

Upon directly comparing eq. (12) and eq. (15), we can make a second assumption regarding $P_n(\mathbf{x})$

$$\nabla \cdot [P_n(\mathbf{x})\nabla(-T_n(\mathbf{x}))] = \sum_{m=1}^{M}\sum_{k=1}^{M} 2\omega_n^{m,k}\left\{\left[T_n(\mathbf{x}_{r,m}) - T_n(\mathbf{x}_{r,k})\right] - \left[T_n^o(\mathbf{x}_{r,m}) - T_n^o(\mathbf{x}_{r,k})\right]\right\}\delta(\mathbf{x}-\mathbf{x}_{r,m}). \tag{16}$$

Following that, we can rewrite eq. (12) as

$$\delta\chi\left(s(\mathbf{x})\right) = \sum_{n=1}^{N}\int_{\Omega} P_n(\mathbf{x})s(\mathbf{x})\delta s(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{17}$$

It is possible to express eq. (17) in a more concise form

$$\delta\chi\left(s(\mathbf{x})\right) = \int_{\Omega} K_s(\mathbf{x})\frac{\delta s(\mathbf{x})}{s(\mathbf{x})}\mathrm{d}\mathbf{x}, \tag{18}$$

where the misfit kernel is

$$K_s(\mathbf{x}) = \sum_{n=1}^{N} K_{s,n}(\mathbf{x}) = \sum_{n=1}^{N} P_n(\mathbf{x})s^2(\mathbf{x}). \tag{19}$$

After computing the misfit kernel (19), we can utilize the multiple-grid model parametrization and the step-size-controlled gradient descent method (Tong 2021a) to find an optimal solution for the objective function (11). This process will ultimately lead to obtaining an optimal velocity model for the subsurface.

It is worth noting that, according to eq. (16), $P_n(\mathbf{x})$ is an adjoint field generated by back-transporting the observational discrepancy at each receiver back to the source location along the direction $\nabla(-T_n(\mathbf{x}))$. We have made two key assumptions regarding $P_n(\mathbf{x})$: first, that $P_n(\mathbf{x}) = 0$ on the boundary $\partial\Omega$, and secondly, that it satisfies a general first-order partial differential eq. (16). These assumptions are both necessary and physically accurate for defining a unique $P_n(\mathbf{x})$, which begins with an arbitrary test function.

3.2 Velocity estimation using common-receiver differential arrival times

First P or S waves emanating from two separate earthquakes $\mathbf{x}_{s,n}$ and $\mathbf{x}_{s,j}$ are recorded by the same seismic station at $\mathbf{x}_{r,m}$. A non-trivial difference between the theoretical traveltime difference $T_n(\mathbf{x}_{r,m}) - T_j(\mathbf{x}_{r,m})$ evaluated in the slowness model $s(\mathbf{x})$ and the observed traveltime difference $\left[T_n^o(\mathbf{x}_{r,m}) - \tau_n\right] - \left[T_j^o(\mathbf{x}_{r,m}) - \tau_j\right]$ indicates the necessity to update $s(\mathbf{x})$ to match the true slowness model. This can be achieved by minimizing the following objective function

$$\chi\left(s(\mathbf{x})\right) = \sum_{n=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\frac{\omega_{n,j}^m}{2}\left\{\left[T_n(\mathbf{x}_{r,m}) - T_j(\mathbf{x}_{r,m})\right] - \left[T_n^o(\mathbf{x}_{r,m}) - T_j^o(\mathbf{x}_{r,m}) - \tau_n + \tau_j\right]\right\}^2. \tag{20}$$

An infinitesimal perturbation of $\delta s(\mathbf{x})$ in the slowness $s(\mathbf{x})$ causes the perturbations in the traveltime fields $T_n(\mathbf{x})$ and $T_j(\mathbf{x})$ and consequently the perturbation in the objective function $\chi(s(\mathbf{x}))$. If we assume $\omega_{n,j}^m = \omega_{j,n}^m$, then the perturbation in the objective function has a relatively simple expression,

$$\delta\chi\left(s(\mathbf{x})\right) = \sum_{n=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\int_{\Omega} 2\omega_{n,j}^m\left\{\left[T_n(\mathbf{x}) - T_j(\mathbf{x})\right] - \left[T_n^o(\mathbf{x}) - T_j^o(\mathbf{x}) - \tau_n + \tau_j\right]\right\}\delta T_n(\mathbf{x})\delta(\mathbf{x}-\mathbf{x}_{r,m})\mathrm{d}\mathbf{x}. \tag{21}$$

Same as the above, we introduce an adjoint function $P_n(\mathbf{x})$ that solves the following boundary value problem

$$\begin{cases} \nabla \cdot [P_n(\mathbf{x})\nabla(-T_n(\mathbf{x}))] = \sum_{j=1}^{N}\sum_{m=1}^{M} 2\omega_{n,j}^m\left\{\left[T_n(\mathbf{x}) - T_j(\mathbf{x})\right] - \left[T_n^o(\mathbf{x}) - T_j^o(\mathbf{x}) - \tau_n + \tau_j\right]\right\}\delta(\mathbf{x}-\mathbf{x}_{r,m}), & \mathbf{x}\in\Omega \\ P_n(\mathbf{x}) = 0, & \mathbf{x}\in\partial\Omega. \end{cases} \tag{22}$$

Then,

$$\delta\chi\left(s(\mathbf{x})\right) = \sum_{n=1}^{N}\int_{\Omega} P_n(\mathbf{x})s(\mathbf{x})\delta s(\mathbf{x})\mathrm{d}\mathbf{x} = \int_{\Omega} K_s(\mathbf{x})\frac{\delta s(\mathbf{x})}{s(\mathbf{x})}\mathrm{d}\mathbf{x}, \tag{23}$$

where the misfit kernel is

$$K_s(\mathbf{x}) = \sum_{n=1}^{N} K_{s,n}(\mathbf{x}) = \sum_{n=1}^{N} P_n(\mathbf{x})s^2(\mathbf{x}). \tag{24}$$

The multiple-grid model parametrization and the step-size-controlled gradient descent method (Tong 2021a) can be adopted to find an optimal solution for the objective function (20).

3.3 Common-source differential arrival time earthquake location

Consider N earthquakes $(\mathbf{x}_{s,n}, \tau_n)$ $(n = 1, 2, \cdots, N)$ and M seismic stations $\mathbf{x}_{r,m}$ $(m = 1, 2, \cdots, M)$. To relocate the nth earthquake $(\mathbf{x}_{s,n}, \tau_n)$, we can minimize an objective function that utilizes common-source differential arrival time data as follows,

$$\chi\left(\mathbf{x}_{s,n}, \tau_n\right) = \sum_{m=1}^{M}\sum_{k=1}^{M} \frac{\omega_n^{m,k}}{2}\left\{\left[T_n(\mathbf{x}_{r,m}) - T_n(\mathbf{x}_{r,k})\right] - \left[\left(T_n^o(\mathbf{x}_{r,m}) - \tau_n\right) - \left(T_n^o(\mathbf{x}_{r,k}) - \tau_n\right)\right]\right\}^2. \tag{25}$$

To facilitate the derivation of its derivatives, we define a function similar to eq. (25),

$$\bar{\chi}\left(\mathbf{x}_{s,n}, \tau_n\right) = \sum_{m=1}^{M}\sum_{k=1}^{M} \frac{\omega_n^{m,k}}{2}\left\{\left[\Gamma_m(\mathbf{x}_{s,n}) - \Gamma_k(\mathbf{x}_{s,n})\right] - \left[\left(T_n^o(\mathbf{x}_{r,m}) - \tau_n\right) - \left(T_n^o(\mathbf{x}_{r,k}) - \tau_n\right)\right]\right\}^2, \tag{26}$$

where $\Gamma_m(\mathbf{x})$ is the traveltime field initiated at the mth receiver, and similar for $\Gamma_k(\mathbf{x})$. If there is an infinitesimal perturbation $(\delta\mathbf{x}_{s,n}, \delta\tau_n)$ in $(\mathbf{x}_{s,n}, \tau_n)$, then the corresponding perturbation in $\bar{\chi}\left(\mathbf{x}_{s,n}, \tau_n\right)$ is

$$\delta\bar{\chi}\left(\mathbf{x}_{s,n}, \tau_n\right) = \sum_{m=1}^{M}\sum_{k=1}^{M} \omega_n^{m,k}\left\{\left[\Gamma_m(\mathbf{x}_{s,n}) - \Gamma_k(\mathbf{x}_{s,n})\right] - \left[T_n^o(\mathbf{x}_{r,m}) - T_n^o(\mathbf{x}_{r,k})\right]\right\}\left[\nabla\Gamma_m(\mathbf{x}_{s,n}) - \nabla\Gamma_k(\mathbf{x}_{s,n})\right]\cdot\delta\mathbf{x}_{s,n}, \tag{27}$$

The reciprocity principle asserts that the two functions (25) and (26) are always identical (Aki & Richards 2002), and so are their first derivatives as they depend on the same variables $\mathbf{x}_{s,n}$ and τ_n. Hence, exploiting the fact that $\Gamma_m(\mathbf{x}_{s,n}) = T_n(\mathbf{x}_{r,m})$ and $\Gamma_k(\mathbf{x}_{s,n}) = T_n(\mathbf{x}_{r,k})$, we have

$$\delta\chi\left(\mathbf{x}_{s,n}, \tau_n\right) = \sum_{m=1}^{M}\sum_{k=1}^{M} 2\omega_n^{m,k}\left\{\left[\Gamma_m(\mathbf{x}_{s,n}) - \Gamma_k(\mathbf{x}_{s,n})\right] - \left[T_n^o(\mathbf{x}_{r,m}) - T_n^o(\mathbf{x}_{r,k})\right]\right\}\nabla\Gamma_m(\mathbf{x}_{s,n})\cdot\delta\mathbf{x}_{s,n}. \tag{28}$$

Eq. (28) defines an adjoint-state earthquake location method using common-source differential arrival times. Obviously, this earthquake location method has no constraints on earthquake origin times.

3.4 Common-receiver differential arrival time earthquake location

Using common-receiver differential arrival time data to locate earthquakes, in a similar manner, we define

$$\chi\left(\mathbf{x}_{s,n}, \tau_n, \mathbf{x}_{s,j}, \tau_j\right) = \sum_{m=1}^{M} \frac{\omega_{n,j}^{m}}{2}\left\{\left[T_n(\mathbf{x}_{r,m}) - T_j(\mathbf{x}_{r,m})\right] - \left[\left(T_n^o(\mathbf{x}_{r,m}) - \tau_n\right) - \left(T_j^o(\mathbf{x}_{r,m}) - \tau_j\right)\right]\right\}^2, \tag{29}$$

and

$$\tilde{\chi}\left(\mathbf{x}_{s,n}, \tau_n, \mathbf{x}_{s,j}, \tau_j\right) = \sum_{m=1}^{M} \frac{\omega_{n,j}^{m}}{2}\left\{\left[\Gamma_m(\mathbf{x}_{s,n}) - \Gamma_m(\mathbf{x}_{s,j})\right] - \left[\left(T_n^o(\mathbf{x}_{r,m}) - \tau_n\right) - \left(T_j^o(\mathbf{x}_{r,m}) - \tau_j\right)\right]\right\}^2. \tag{30}$$

The first derivatives of $\tilde{\chi}$ can be expressed by

$$\begin{aligned}\delta\tilde{\chi}\left(\mathbf{x}_{s,n}, \tau_n, \mathbf{x}_{s,j}, \tau_j\right) = &\sum_{m=1}^{M} \omega_{n,j}^{m}\left\{\left[\Gamma_m(\mathbf{x}_{s,n}) - \Gamma_m(\mathbf{x}_{s,j})\right] - \left[\left(T_n^o(\mathbf{x}_{r,m}) - \tau_n\right) - \left(T_j^o(\mathbf{x}_{r,m}) - \tau_j\right)\right]\right\}\\ &\times\left[\nabla\Gamma_m(\mathbf{x}_{s,n})\cdot\delta\mathbf{x}_{s,n} - \nabla\Gamma_m(\mathbf{x}_{s,j})\cdot\delta\mathbf{x}_{s,j} + \delta\tau_n - \delta\tau_j\right].\end{aligned} \tag{31}$$

The reciprocity principle also asserts that the two objective functions (eq. 29) and (eq. 30) are identical, and so are their first derivatives. We have

$$\begin{aligned}\delta\chi\left(\mathbf{x}_{s,n}, \tau_n, \mathbf{x}_{s,j}, \tau_j\right) = &\sum_{m=1}^{M} \omega_{n,j}^{m}\left\{\left[\Gamma_m(\mathbf{x}_{s,n}) - \Gamma_m(\mathbf{x}_{s,j})\right] - \left[\left(T_n^o(\mathbf{x}_{r,m}) - \tau_n\right) - \left(T_j^o(\mathbf{x}_{r,m}) - \tau_j\right)\right]\right\}\\ &\times\left[\nabla\Gamma_m(\mathbf{x}_{s,n})\cdot\delta\mathbf{x}_{s,n} - \nabla\Gamma_m(\mathbf{x}_{s,j})\cdot\delta\mathbf{x}_{s,j} + \delta\tau_n - \delta\tau_j\right].\end{aligned} \tag{32}$$

By utilizing the gradient information provided by eq. (32), the accuracy of earthquake locations can be iteratively enhanced.

Notably, the double-difference earthquake location algorithm proposed by Waldhauser & Ellsworth (2000), which is widely used in practice, involves the utilization of both first arrival times and common-receiver differential arrival times to determine hypocentre locations. We can achieve the same results by combining eq. (10) without considering velocity perturbations and eq. (32).

3.5 Differential arrival time tomography

When neither precise earthquake locations nor an accurate velocity model are available, we can utilize common-source differential arrival time data to determine an optimal seismic velocity model and earthquake locations simultaneously, using the adjoint-state common-source

differential arrival time tomography (ATT-CS) method. The objective function of ATT-CS takes the following form

$$\chi\left(s(\mathbf{x}),\ \mathbf{x}_{s,1},\ \tau_1, \cdots,\ \mathbf{x}_{s,n},\ \tau_n,\ \cdots,\ \mathbf{x}_{s,N},\ \tau_N\right)$$
$$= \sum_{m=1}^{M}\sum_{k=1}^{M}\sum_{n=1}^{N}\frac{\omega_n^{m,k}}{2}\left\{\left[T_n(\mathbf{x}_{r,m}) - T_n(\mathbf{x}_{r,k})\right] - \left[T_n^o(\mathbf{x}_{r,m}) - T_n^o(\mathbf{x}_{r,k})\right]\right\}^2. \tag{33}$$

Based on the discussion in Sections 3.1 and 3.3, the first derivatives of the objective function (eq. 33) can be expressed by

$$\delta\chi\left(s(\mathbf{x}),\ \mathbf{x}_{s,1},\ \tau_1, \cdots,\ \mathbf{x}_{s,n},\ \tau_n,\ \cdots,\ \mathbf{x}_{s,N},\ \tau_N\right) = \int_\Omega K_s(\mathbf{x})\frac{\delta s(\mathbf{x})}{s(\mathbf{x})}\mathrm{d}\mathbf{x}$$
$$+\sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{M}2\omega_n^{m,k}\left\{\left[\Gamma_m(\mathbf{x}_{s,n}) - \Gamma_k(\mathbf{x}_{s,n})\right] - \left[T_n^o(\mathbf{x}_{r,m}) - T_n^o(\mathbf{x}_{r,k})\right]\right\}\nabla\Gamma_m(\mathbf{x}_{s,n})\cdot\delta\mathbf{x}_{s,n}, \tag{34}$$

where the misfit kernel $K_s(\mathbf{x})$ is given by eq. (19). As discussed above, by combining the multiple-grid model parametrization and the step-size-controlled gradient descent method, we can effectively search for an optimal slowness model and estimate earthquake hypocentres. Common-source differential arrival time data have no constraints on the origin time of each earthquake τ_n. However, after obtaining an optimal slowness model $s(\mathbf{x})$ and accurate earthquake hypocentres $\mathbf{x}_{s,n}$, if the observed arrival times $T_n^o(\mathbf{x}_{r,j})$ ($j = 1, 2, \cdots, M$) are available, we can equate the derivative of the objective function (eq. 9) with respect to τ_n to zero (extreme value condition) in order to determine the origin time τ_n,

$$\tau_n = \frac{\sum_{m=1}^{M}\omega_{n,m}\left[T_n^o(\mathbf{x}_{r,m}) - T_n(\mathbf{x}_{r,m})\right]}{\sum_{m=1}^{M}\omega_{n,m}}. \tag{35}$$

The adjoint-state common-receiver differential arrival time tomography (ATT-CR) method enables the utilization of common-receiver differential arrival time data to determine an optimal seismic velocity model and accurate earthquake locations. The objective function of ATT-CR is written as

$$\chi\left(s(\mathbf{x}),\ \mathbf{x}_{s,1},\ \tau_1, \cdots,\ \mathbf{x}_{s,n},\ \tau_n,\ \cdots,\ \mathbf{x}_{s,N},\ \tau_N\right)$$
$$= \sum_{n=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\frac{\omega_{n,j}^{m}}{2}\left\{\left[T_n(\mathbf{x}_{r,m}) - T_j(\mathbf{x}_{r,m})\right] - \left[T_n^o(\mathbf{x}_{r,m}) - T_j^o(\mathbf{x}_{r,m}) - \tau_n + \tau_j\right]\right\}^2. \tag{36}$$

Based on the discussion in Sections 3.2 and 3.4, its first derivatives can be computed by

$$\delta\chi\left(s(\mathbf{x}),\ \mathbf{x}_{s,1},\ \tau_1, \cdots,\ \mathbf{x}_{s,n},\ \tau_n,\ \cdots,\ \mathbf{x}_{s,N},\ \tau_N\right) = \int_\Omega K_s(\mathbf{x})\frac{\delta s(\mathbf{x})}{s(\mathbf{x})}\mathrm{d}\mathbf{x}$$
$$+\sum_{n=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}2\omega_{n,j}^{m}\left\{\left[\Gamma_m(\mathbf{x}_{s,n}) - \Gamma_m(\mathbf{x}_{s,j})\right] - \left[\left(T_n^o(\mathbf{x}_{r,m}) - \tau_n\right) - \left(T_j^o(\mathbf{x}_{r,m}) - \tau_j\right)\right]\right\}\left[\nabla\Gamma_m(\mathbf{x}_{s,n})\cdot\delta\mathbf{x}_{s,n} + \delta\tau_n\right]. \tag{37}$$

Here the misfit kernel $K_s(\mathbf{x})$ is given by eq. (24).

If the sole requirement is velocity estimation, it is worth highlighting that source-receiver reciprocity can be effectively utilized to interchange the roles of source and receiver. The primary distinction lies in the computational cost. In a scenario with N sources and M receivers, the common-receiver approach requires N forward simulations (or $2N$ forward simulations if the objective function value is computed first without saving the N traveltime fields), along with N adjoint simulations. Conversely, by using source-receiver reciprocity, the common-source approach requires M forward simulations and M adjoint simulations. The choice between the two approaches depends on which value is greater between M and N.

In addition to first arrival times, common-source differential arrival times, and common-receiver differential arrival times, second-order differential arrival times (called triple-difference arrival times in previous studies) have been proven effective in determining seismic velocity structure and locating earthquakes (e.g., Guo & Zhang 2017; Guo *et al.* 2021). These advancements have been achieved through the utilization of ray-based approaches. The adjoint-state method can also be used to leverage second-order differential arrival time data, sharing similarities with the ATT, ATT-CS and ATT-CR methods. We provide a brief description of the adjoint-state second-order differential arrival time tomography method in Appendix A. A thorough investigation and real data applications will be conducted in a separate study.

4 NUMERICAL EXAMPLES AND DISCUSSION

Accurately and efficiently solving the eikonal eq. (1) and the adjoint eqs (4), (16) and (22) is crucial for the numerical implementation of the ATT, ATT-CS and ATT-CR methods. As frequently discussed in the literature, the fast marching method and the fast sweeping method are the two main types of numerical solvers for the eikonal equation (e.g. Sethian 1996; Zhao 2004; Rawlinson *et al.* 2008). Both methods can obtain accurate numerical approximations to the viscosity solution of the eikonal equation. Mathematically, a viscosity solution satisfies the eikonal

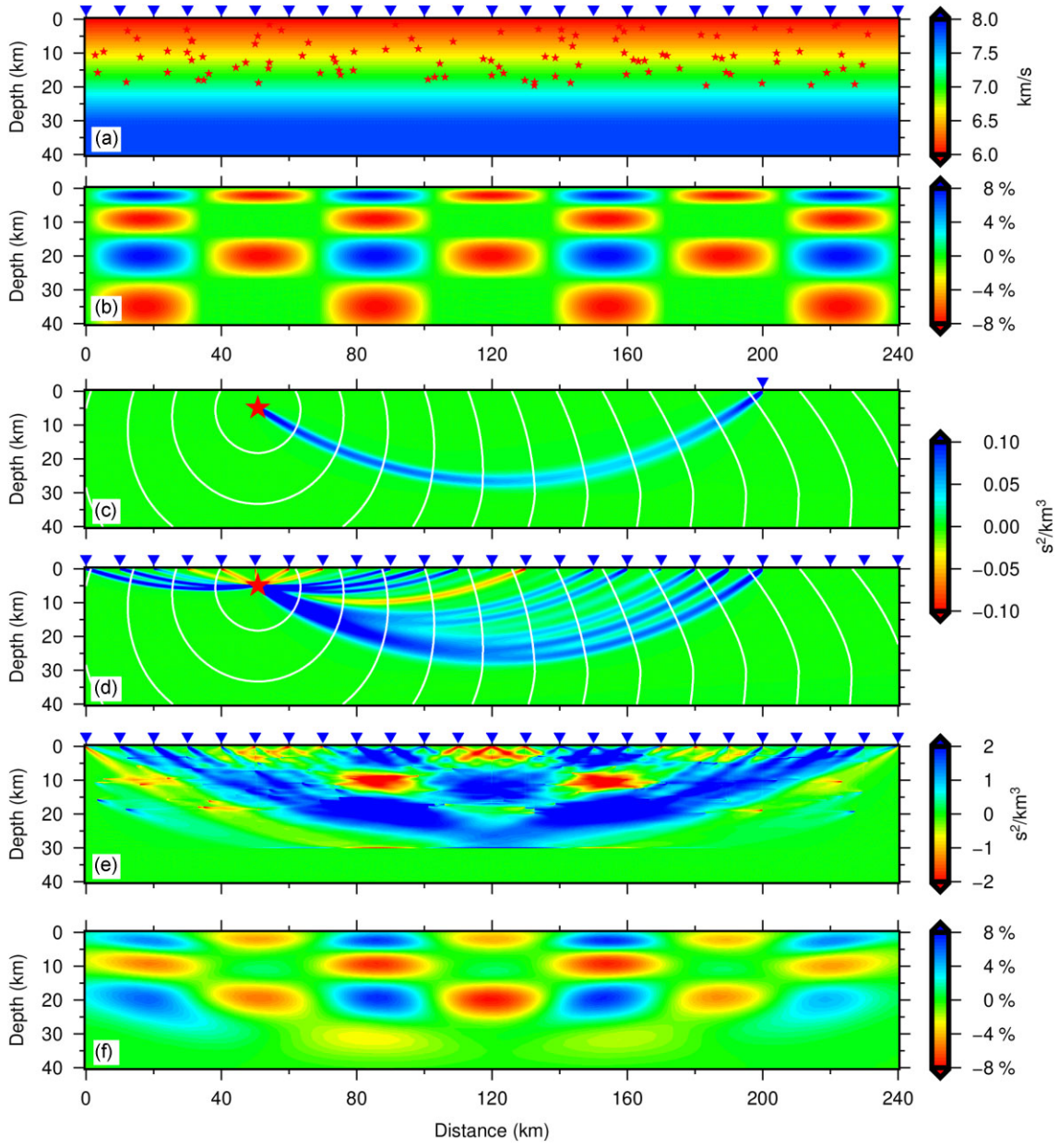

Figure 2. Performance demonstration of the ATT method. (a) A vertically varying velocity model $v(\mathbf{x})$. The red stars are 100 randomly distributed earthquakes. The blue inverted triangles denote 25 seismic stations. (b) The relative velocity perturbation $\Delta v(\mathbf{x})/v(\mathbf{x})$ of the target model from the initial model in (a). (c) An example of an individual kernel. (d) An example of an event kernel corresponding to the same earthquake in (c). (e) The misfit kernel of the objective function. (f) The inverted relative velocity perturbation after 40 iterations. The white curves in (c) and (d) are traveltime isochrones at an interval of 2 s. All the sensitivity kernels in (c)–(e) are with respect to slowness and computed in the initial model $v(\mathbf{x})$. The colour scales on the right hand side are for velocity, relative velocity perturbation and sensitivity kernels, respectively.

equation almost everywhere as long as $\nabla T(\mathbf{x})$ is defined (Rouy & Tourin 1992). The fast marching method adopts a fast heap sorting method to locate the minimum traveltime node (e.g. Sethian 1996; Rawlinson *et al.* 2008). In contrast, the fast sweeping method uses Gauss–Seidel iterations with alternating ordering to solve the discretized eikonal equation (e.g. Zhao 2004; Qian *et al.* 2007). Following Leung & Qian (2006) and Tong (2021a), we utilize the fast sweeping method to solve the eikonal eq. (1) and the adjoint eqs (4), (16) and (22).

4.1 Velocity estimation

We evaluate the velocity estimation performance of the ATT, ATT-CS and ATT-CR methods through a uniform restoration test. The restoration test involves using a starting velocity model that varies only in the vertical direction (Fig. 2a) and a 3-D target model. Mathematically, we can express the starting model $v(\mathbf{x})$ and the relative velocity perturbation $\Delta v(\mathbf{x})/v(\mathbf{x})$ of the target model from the starting model as follows:

$$v(\mathbf{x}) = v(x, y, z) = \min(6.0 + 0.06z, 7.80) \tag{38}$$

and

$$\frac{\Delta v(\mathbf{x})}{v(\mathbf{x})} = \begin{cases} 0.1 \sin\left(\frac{7\pi x}{240}\right) \sin\left(\frac{\sqrt{9+8z}-3}{4}\pi\right), & \text{if } \sin\left(\frac{7\pi x}{240}\right) > 0 \text{ or } \sin\left(\frac{\sqrt{9+8z}-3}{4}\pi\right) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{39}$$

25 seismic stations are located on the surface of the vertical x–z (depth) plane at equal intervals of 10 km (Fig. 2a). We use traveltime data from 100 earthquakes occurring between depths of 1 and 20 km in this vertical plane to recover the relative velocity perturbation. Earthquake locations are assumed to be known for simplicity, and their coordinates are randomly generated (Fig. 2a). The velocity perturbation is invariant in the y-direction. For demonstration purposes, only the results in the vertical x–z plane are displayed. However, it should be noted that all of the ATT methods discussed in this study are inherently capable of solving tomographic problems in 2-D and 3-D.

First, we examine the performance of the ATT method using first arrival time data by computing observed traveltimes in the target model (Fig. 2b). Figs 2(c)–(e) show examples of individual kernel, event kernel, and misfit kernel computed in the starting velocity model (Fig. 2a). Every individual kernel connects a source and a receiver pair and is perpendicular to the wave front (Fig. 2c). As indicated by eq. (4), the sign of an individual kernel is determined by the traveltime misfit $T_n(\mathbf{x}) - \left[T_n^o(\mathbf{x}) - \tau_n\right]$ at the receiver location $\mathbf{x}_{r,m}$. The sum of all individual kernels associated with a single earthquake is an event kernel (Fig. 2d), and the misfit kernel (Fig. 2e) is the sum of all event kernels. The misfit kernel can be viewed as an image of the relative velocity perturbation of the target model to the present velocity model. It also suggests a direction to update the present model. However, due to spatial smearing, the misfit kernel may provide incorrect information about the model perturbation at some locations such as ($x = 120$ km, $z = 9$ km). To obtain a reliable result, the model should be iteratively improved until the objective function becomes stable or quantitatively has less than a few percent reduction (e.g. 10 per cent) compared to its performance in the first iteration. We use the step-size-controlled gradient descent method (Tong 2021a), which restricts the relative slowness perturbation within ± 2 per cent at each iteration, roughly equivalent to choosing an optimal damping parameter (Liu *et al.* 2021), to update the velocity model. After a total of 40 iterations, we successfully recover the velocity structure from the surface to approximately 25 km depth (Fig. 2f).

Next, we investigate the effectiveness of the ATT-CS method. The model parametrization and optimization parameters are the same as the previous ATT example. In practical applications, differential arrival times of seismic phases on similar waveforms can be accurately measured using cross-correlation. Similar waveforms are commonly observed at seismic stations in close proximity. To simulate this scenario, only common-source differential arrival times at adjacent stations separated by less than 15 km are used in this test. Figs 3(c)–(e) display examples of individual kernel, event kernel, and misfit kernel of the common-source differential arrival time objective function (eq. 11) with respect to slowness. The individual kernel indicates that common-source differential arrival times are more sensitive to receiver-side structures. Furthermore, a direct comparison between the common-source differential arrival time misfit kernel (Fig. 3e) and the first arrival time misfit kernel (Fig. 2e) also reveals that common-source differential arrival times are more sensitive to velocity heterogeneity from the surface to about 5 km. Fig. 3(f) shows the result after 40 iterations. We observe that the velocity anomalies above 14 km depth are almost fully recovered. In the depth range between 14 and 20 km, the velocity anomalies are only partly recovered, and the performance of the ATT-CS method is inferior to that of the ATT method in imaging structures below 14 km depth. Overall, this test suggests that the ATT-CS method is an efficient tool for receiver-side velocity structure imaging.

Thirdly, we use the ATT-CR method to reconstruct the relative velocity perturbation of the target model to the starting model (Fig. 4b). The synthetic data only include common-receiver differential arrival times of adjacent earthquakes separated by less than 15 km. Again, examples of individual kernel, receiver kernel, and misfit kernel (Figs 4c–e) are computed in the vertically varying starting model (Fig. 4a). Here the receiver kernel is defined as the sum of all the individual kernels associated with a common receiver. The individual kernel indicates that common-receiver differential arrival time measurements are more sensitive to source-side velocity structures, whereas the receiver-side complexities may have a limited influence on the final tomographic result (Fig. 4c). Earthquakes in this test are nearly uniformly distributed in the computational domain from the surface to 20 km depth (Fig. 4a). As a result, velocity anomalies above 25 km depth can be satisfactorily recovered by solely using common-receiver differential arrival times (Fig. 4f), indicating that the ATT-CR method could be a valuable tool for subsurface velocity structure imaging.

Before concluding this section, it is worth discussing how a multiple-grid consisting of several sets of inversion grids is designed to discretize the relative slowness model in the restoration test. As stated in eq. (39), the model of the relative velocity perturbation exhibits a sinusoidal shape along both the horizontal and vertical directions, with a horizontal wavelength of approximately 68.6 km and a vertical wavelength that increases from 6 km at the surface to around 26 km at 20 km depth. To avoid undersampling, every inversion grid must have as many grid nodes as possible. However, the limited resolving ability of the available data requires maintaining a low number of grid nodes. To strike a balance, we suggest reconstructing every wavelength anomaly using approximately five grid nodes. In this restoration test, we set the horizontal spacing of each inversion grid to 16.13 km, with the vertical spacing gradually increasing such that every wavelength anomaly is sampled by about five grid nodes (Figs 5a–e). The spacing of each inversion grid is relatively large compared to the forward grid (which has a spacing of 0.2 km in each direction) on which the eikonal equation and adjoint equations are solved. The positioning of the initial node within an inversion grid is subjective and has the potential to impact the ultimate outcome of the tomographic image, particularly due to the relatively wide spacing between nodes in each inversion grid. To reduce the influence of subjectivity, we use a multiple-grid consisting of five sets of staggered inversion grids (Fig. 5f) to reconstruct the velocity anomaly. This approach yields reliable results (Figs 2f, 3f and 4f).

In practical applications, the sizes (wavelengths) of subsurface velocity anomalies are often unknown. However, we can estimate the resolving ability of the chosen data by conducting recovery tests, such as checkerboard resolution tests. By increasing or decreasing the sizes

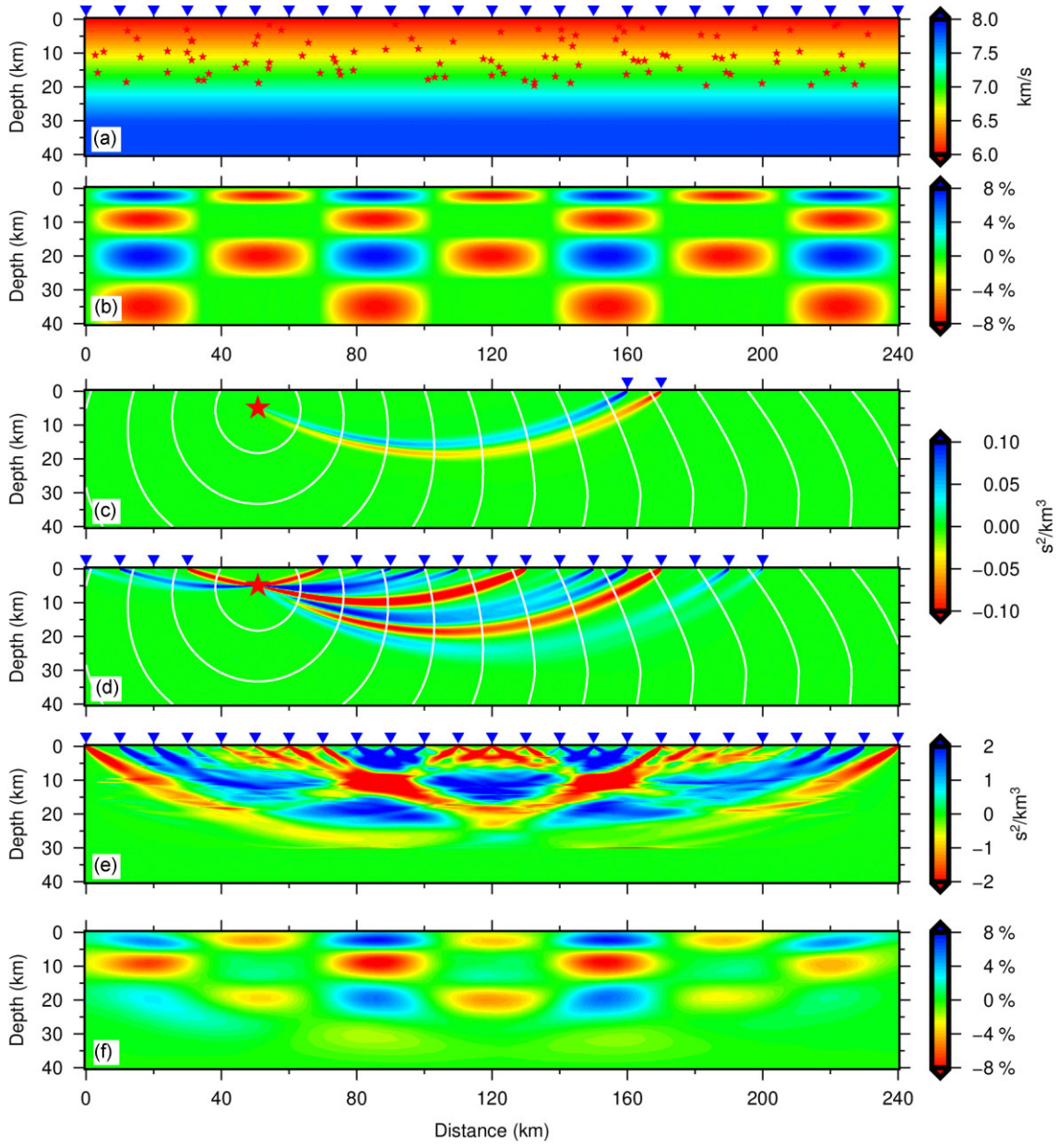

Figure 3. The same as Fig. 2 but for the ATT-CS method. (c) One individual kernel related to one earthquake and two seismic stations. (d) An example of event kernel corresponding to the same earthquake in (c). (e) The misfit kernel of the common-source differential arrival time objective function with respect to slowness computed in the initial model in (a). Differential arrival times are only evaluated at two stations that are separated by less than 15 km. (f) The inverted relative velocity perturbation after 40 iterations.

of anomalies, we can estimate the minimum wavelength that the data can resolve in each direction or around a specific region. Based on the estimated minimum wavelength, we can set the spacing of each inversion grid accordingly to avoid undersampling. This approach ensures that every inversion grid has as many grid nodes as possible while maintaining a low number of nodes, which is necessary due to the limited resolving ability of the available data.

4.2 Resolving abilities of differential arrival time data

The ATT, ATT-CS and ATT-CR methods use different types of arrival time data, as a result, have varying abilities in resolving subsurface velocity structures. A thorough understanding of their resolving abilities is essential for selecting appropriate tomography methods in practical applications. We conduct additional tests to further examine the resolving ability of each type of arrival time data.

The restoration test in Section 4.1 demonstrates that all three methods yield satisfactory results. However, if the sizes of velocity anomalies are reduced, we can observe distinct differences in the performance of the three tomography methods. Let us consider a target model that has smaller-size velocity anomalies (Fig. 6b). The relative velocity perturbation of the target model from the starting velocity

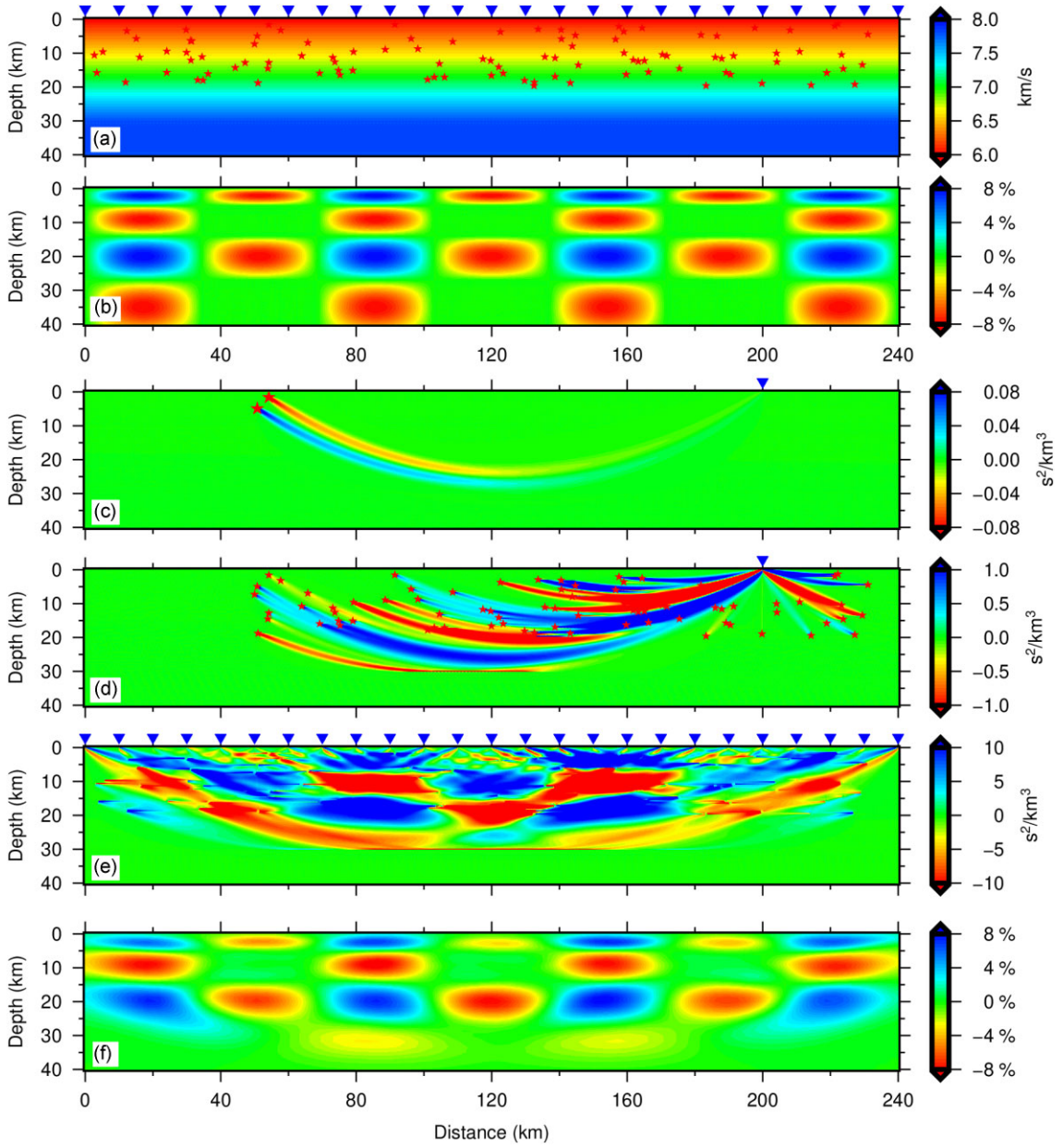

Figure 4. The same as Fig. 2 but for the ATT-CR method. (c) One individual kernel related to two earthquakes and one seismic station. (d) An example of receiver kernel corresponding to the same seismic station in (c). (e) The misfit kernel of the common-receiver differential arrival time objective function with respect to slowness computed in the initial model in (a). Differential arrival times are only evaluated for two earthquakes that are separated by less than 15 km. (f) The inverted relative velocity perturbation after 40 iterations.

model (Fig. 6a) is

$$\frac{\Delta v(\mathbf{x})}{v(\mathbf{x})} = 0.06 \sin\left(\frac{27\pi x}{240}\right) \sin\left(\frac{\sqrt{9+8z}-3}{4}\pi\right), \tag{40}$$

where the wavelength of the velocity perturbation along the x-axis is reduced to 17.78 km. For the sake of simplicity, we assume that there are 47 earthquakes located at a depth of 18 km and are equally spaced at an interval of 5 km (Fig. 6a). It should be noted that this is an unrealistic scenario, but it is not difficult to find 47 earthquakes that occurred around 18 km depth in seismologically active regions such as California. To ensure a fair comparison, all three tomography methods use an identical model parametrization and utilize the same step-size-controlled gradient descent algorithm. Additionally, the iterative process for model updating is terminated after 40 iterations for each method, ensuring consistent stopping criteria across all approaches. The resulting images clearly demonstrate that whereas first arrival time data can only partially recover small-size anomalies (Fig. 6c), common-source differential arrival time data and common-receiver differential arrival time data can fully recover small-size anomalies near the receivers (Fig. 6d) and near the sources (Fig. 6e), respectively. This suggests that differential arrival time data are more effective at imaging smaller-scale structures than first arrival time data.

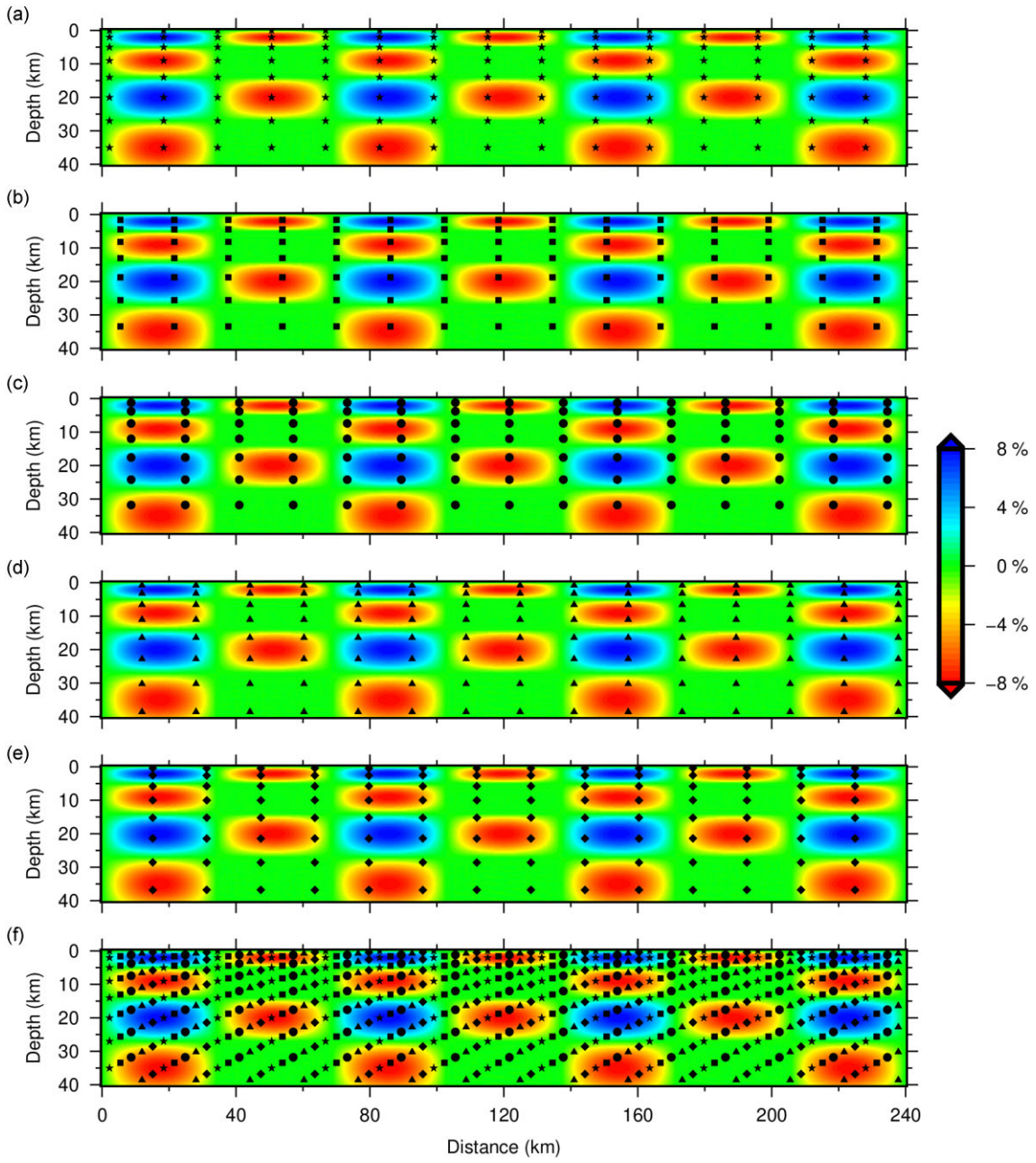

Figure 5. The multiple-grid model parametrization for the restoration test in Section 4.1. Five sets of regular grids placed in a staggered way are used to sample the relative velocity perturbation field. (a)–(e) The spatial distributions of the five regular grids. (f) The multiple-grid formed by the five regular grids. For demonstration purposes, only the grid nodes inside the computational domain are shown. For each regular grid, there is one layer of grid nodes beyond each edge of the computational domain.

It is evident that the two differential arrival time tomography methods have their respective limitations and blind zones (Figs 6d–e). In this particular example, it is possible to fully recover all the velocity anomalies if the final model generated by one differential arrival time tomography method is used as the starting model for the other method. However, in reality, differential arrival times are usually measured when two seismic stations or two earthquakes are in close proximity, resulting in limited illumination of some subregions. One approach to overcome this limitation is to sequentially use the ATT, ATT-CS and ATT-CR methods, where the final model from the previous inversion is used as the starting model for the next inversion. By using this strategy, we can successfully recover all the fine-scale velocity anomalies from the surface to a depth of approximately 25 km (Fig. 6g). Again, the iterative process for model updating is terminated after 40 iterations for each method. An alternative approach is to simultaneously invert all three types of arrival time data, as adopted by Guo *et al.* (2021) and Share *et al.* (2021) through ray-based seismic tomography. The simultaneous inversion of all three types of data using the adjoint-state method will be investigated in a separate study.

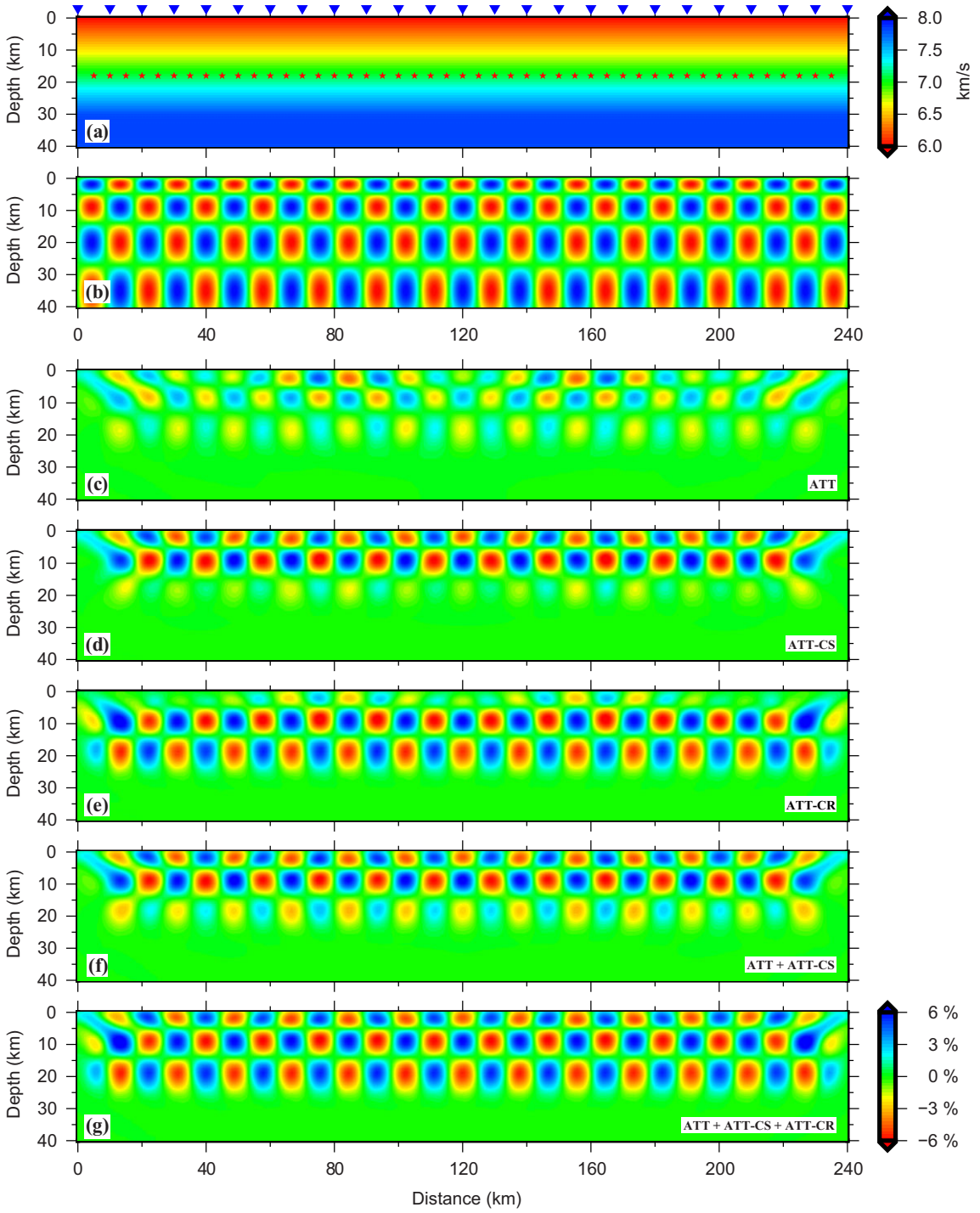

Figure 6. Recovery test of a fine-scale checkerboard structure. (a) The initial vertically varying velocity model with 47 earthquakes located at 18 km depth and 25 seismic stations on the surface. (b) The relative velocity perturbation of the target checkerboard model with respect to the initial velocity model. (c)–(e) The inverted relative velocity perturbation models generated by ATT, ATT-CS and ATT-CR, respectively. (f) The inverted velocity model by sequentially using ATT and ATT-CS. (g) The imaged relative velocity perturbation by ATT-CR with the starting model shown in (f).

4.3 Earthquake location determination

First arrival times, common-source differential arrival times, and common-receiver differential arrival times contain information on earthquake locations. To evaluate the performance of the associated earthquake location algorithms, we relocate 160 earthquakes assumed to have occurred within a fault zone in a model where velocity increases with depth (Fig. 7a). Initially, the maximum location errors of these 160 earthquakes are 10 km in the direction perpendicular to the fault trace, 2 km along the fault trace, 2.5 km in the depth direction, and 1 s in origin time (Fig. 7a). 21 seismic stations with an average spacing of 4 km are positioned in a direction perpendicular to the fault trace (Fig. 7e). There is a

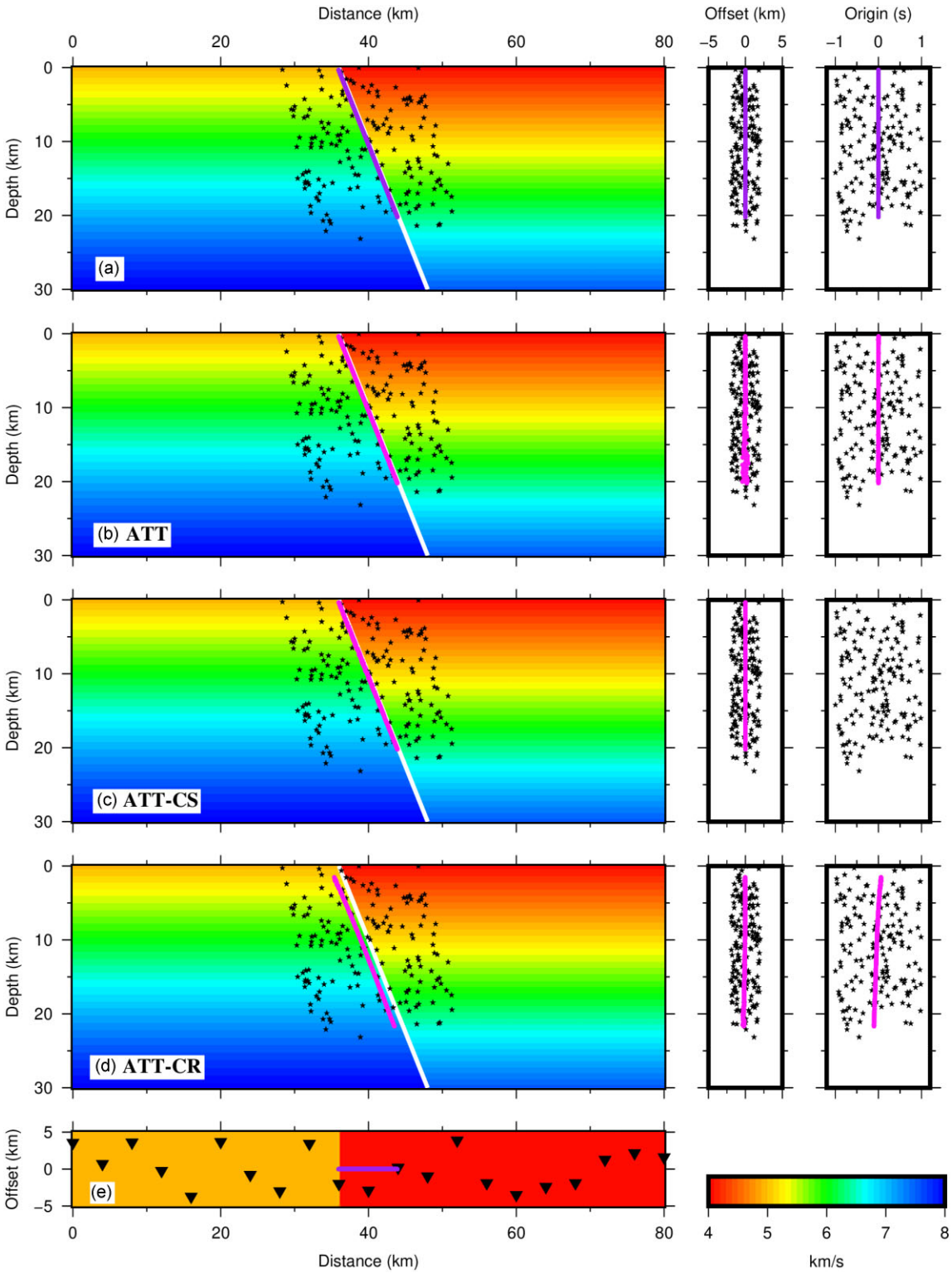

Figure 7. Relocating earthquakes in an accurate velocity model. The black stars and purple circles represent the initial and (a) true or (b–d) inverted earthquake locations. The left panel displays the velocity model in the vertical and fault-perpendicular plane. The white lines denote a fault zone and highlight the strong velocity contrast across the fault. The middle panel displays earthquakes in the vertical and fault-parallel plane. Earthquakes in the vertical and origin time domain are included in the right panel. Earthquake locations are determined using first arrival times (b), common-source differential arrival times (c) and common-receiver differential arrival times (d), respectively. (e) Inverse triangles represent 21 seismic stations on the surface.

maximum deviation of 4 km from this perpendicular direction, meaning that some stations are slightly off the exact perpendicular alignment to the fault trace. We use the arrival time data computed at these seismic stations to locate earthquakes. Every common-source differential arrival time is computed at a station pair separated by less than 15 km, with both source-receiver distances being greater than four times the distance between the station pair. This ensures that the two first arrivals travel along similar paths. Additionally, for every earthquake pair separated by a distance greater than 2 km but less than 15 km, their common-receiver differential arrival times are used, provided that both source-receiver distances are greater than four times the distance between the earthquake pair.

To eliminate the complexity arising from considering other influencing factors, such as velocity inaccuracy, we first determine earthquake locations using the true velocity model (Fig. 7a). As shown in Figs 7(b)–(f), all three types of data accurately determine the locations of the 160 earthquakes in a precise velocity model. This confirms the effectiveness of each earthquake location algorithm. However, it is worth noting a slight leftward shift in the earthquake locations derived from common-receiver differential arrival times, while their relative positions remain highly accurate (Fig. 7d).

The accuracy of the velocity model plays a critical role in earthquake location determination. To examine the influence, we determine the locations of the 160 earthquakes in a simple linearly increasing velocity model (Fig. 8a). It is evident that the inaccuracy of the velocity model introduces systematic errors to earthquake locations (deviations from the fault zone) when utilizing first arrival times (Fig. 8b). The results derived from common-source differential arrival times appear to be even less accurate (Fig. 8c), providing an incorrect representation of the actual earthquake distributions. However, common-receiver differential arrival times have much better performance within such an inaccurate velocity model. This may suggest that common-receiver differential arrival times are less sensitive to velocity errors and offer supplementary constraints on earthquake locations.

4.4 Simultaneous determination of velocity and earthquake locations

When both earthquake locations and velocity structure are not accurately known, it becomes essential to determine them simultaneously. We continue to consider a model with a strong velocity contrast across a fault, along with 160 earthquakes situated within the fault zone. The true model is slightly different from the one depicted in Fig. 7(a), as it incorporates a small-size high-velocity anomaly in the broad low-velocity block on the right-hand side (Fig. 9a). To increase subsurface illumination, we include another 200 randomly distributed earthquakes (Fig. 9a). The initial positions for the 360 earthquakes are determined using the eikonal equation-based grid search method (Appendix B). This method is capable of finding a globally optimal solution. However, conducting the grid search within the initial velocity model may introduce errors into the earthquake locations due to the inaccuracies of the initial model, and make it difficult to escape the local minimum during the subsequent joint inversion of velocity and earthquake locations. To mitigate this, we conduct the grid search in a different velocity model that matches the target model's velocity at the surface and linearly increases to 7.95 km s^{-1} at a depth of 30 km (Fig. 9b).

We use a sequential combination of the ATT, ATT-CS and ATT-CR methods to reconstruct the subsurface velocity structure and determine earthquake locations. The initial velocity model for ATT is a simple vertically increasing velocity model, as illustrated in Fig. 8(a). In all three tomography methods, we utilize a multiple-grid inversion grid, which consists of five sets of regular grids with a uniform horizontal spacing of approximately 3 km, allowing it to capture the small-size high-velocity zone, while accommodating the presence of the large-scale high-velocity anomaly on the left and the general low-velocity body on the right. The criteria for choosing common-source differential arrival times and common-receiver differential arrival times remain consistent with those used in the preceding earthquake location example. Additionally, random errors (with mean 0 s and standard deviation 0.1 s for first arrival times and with mean 0 s and standard deviation 0.05 s for differential arrival times) are introduced to the synthetic data to simulate measurement errors in real-world practice. The earthquake locations and velocity structure are updated iteratively until the objective function values show no significant reduction.

The ATT method successfully captures and reconstructs the large-scale high-velocity anomaly on the left and the low-velocity anomaly on the right. It also accurately identifies the sharp velocity contrast above 15 km depth (Fig. 9c). The accuracies of earthquake locations have been slightly improved, particularly visible for earthquakes inside the small-size high-velocity anomaly. However, the recovery of the small-size high-velocity body is only partial. The utilization of common-source differential arrival times enhances the accuracy of near-surface structures (Fig. 9d). Specifically, the sharp velocity contrast across the fault aligns more closely with the true model when compared to the result of ATT. It is also observable that the accuracies of earthquake locations in the fault-parallel direction have been improved. The use of common-receiver differential arrival times generally improves the accuracies of earthquake locations (Fig. 9e), underscoring the sensitivity of common-receiver differential arrival times to earthquake parameters (hypocentre and origin time). Importantly, the tomographic results refined first by ATT-CS and then by ATT-CR highlight the presence of a small-size high-velocity anomaly in the large-scale low-velocity body. This underlines the advantage of involving differential arrival times in tomographic inversions.

4.5 Multipathing

The adjoint eqs (4), (16) and (22) describe how traveltime misfits or differential traveltime misfits observed at all the receivers are back-projected to the source location. The direction of the back-projection is given by the negative traveltime gradient $-\nabla T(\mathbf{x})$. In a local sense, this gradient is parallel to the tangent of the fastest path between the source and a receiver. Therefore, the adjoint eqs (4), (16) and (22) transport traveltime misfits or differential traveltime misfits along the fastest paths (seismic rays) backwards to the source location. However, even in some mildly heterogeneous media, there may exist multiple fastest paths between a source and a receiver. This phenomenon, known as

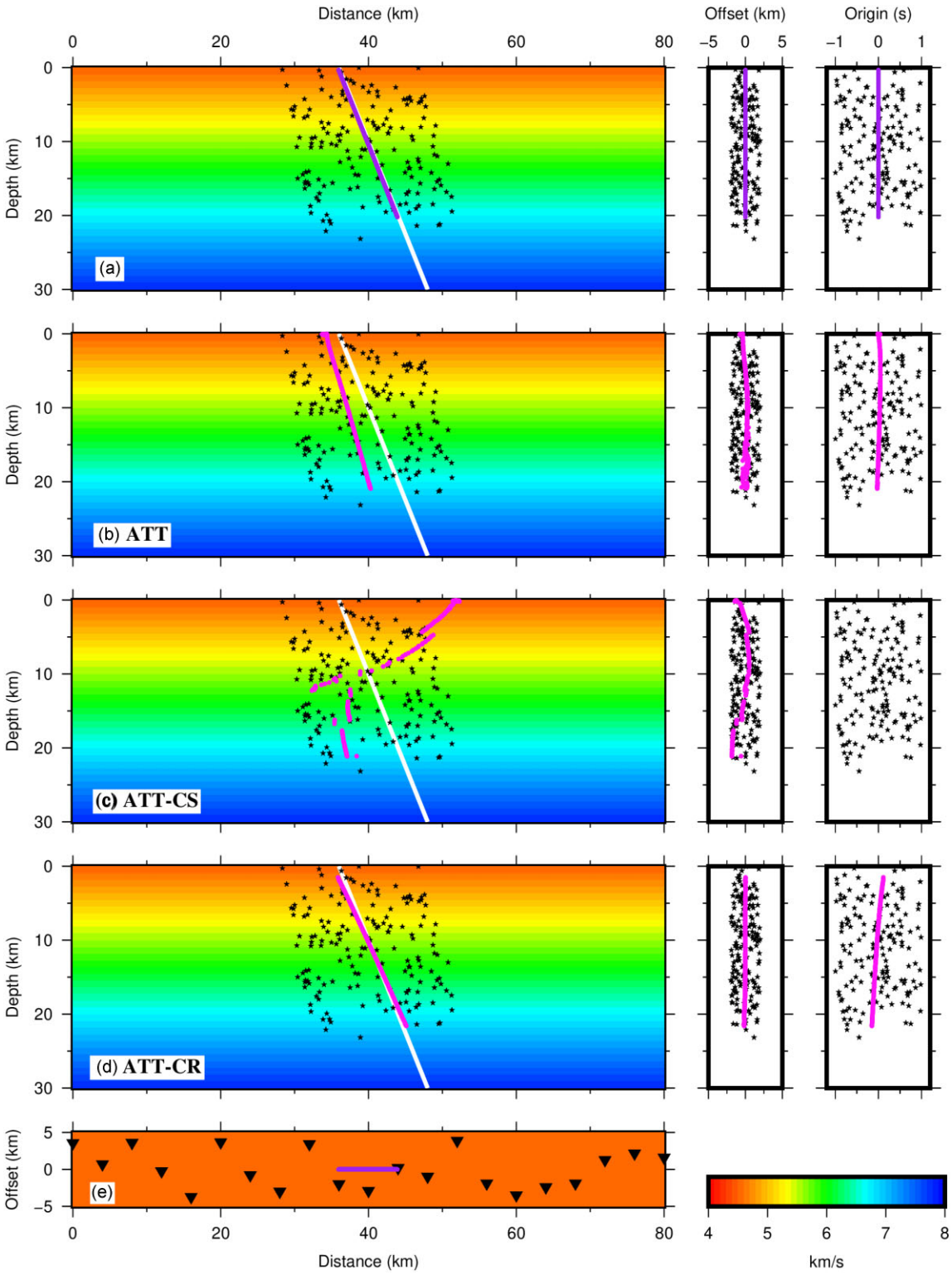

Figure 8. The same as Fig. 7 but earthquake locations are determined in an inaccurate velocity model shown in (a).

multipathing, occurs at singular points where $\nabla T(\mathbf{x})$ is not continuous, resulting in multiple distinct directions of $\nabla T(\mathbf{x})$ in the neighbourhood of each singular point. Ray tracing along the negative direction of $\nabla T(\mathbf{x})$ may become unstable through these singular points and typically can only identify one fastest path. In contrast, our approach evaluates $\nabla T(\mathbf{x})$ at every point in the computational domain, enabling us to identify all the distinct directions in the neighbourhood of a singular point. Consequently, when there are multiple fastest paths between a source and a receiver, solving the adjoint equations directly (without using any ray tracing techniques) allows for the projection of traveltime misfits or differential traveltime misfits onto all possible paths.

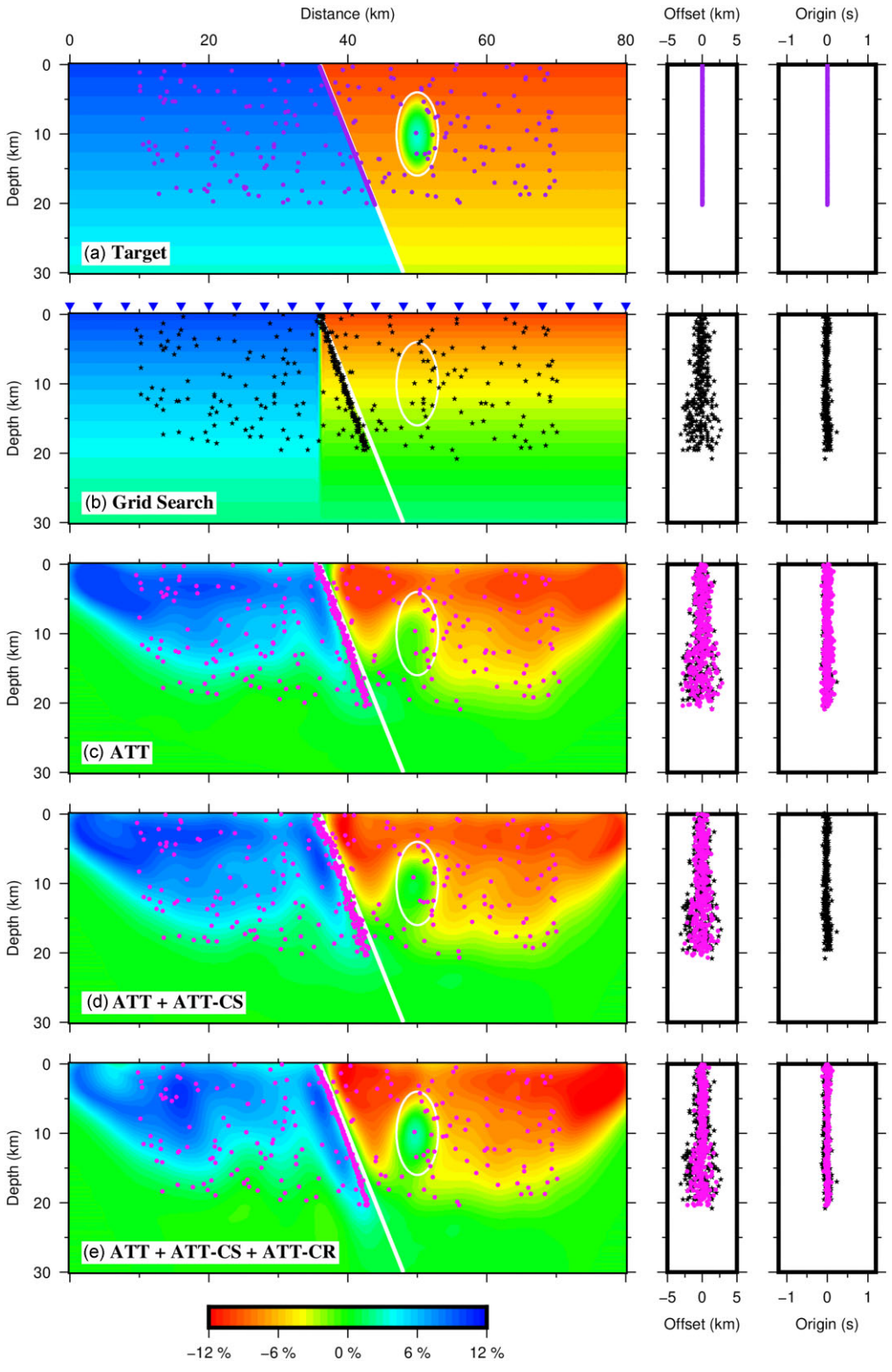

Figure 9. Simultaneous determination of velocity structure and earthquake locations. All the velocity models are presented in terms of relative velocity perturbation compared to the initial model (Fig. 8a). (a) Earthquakes at their true locations and the target velocity model. The white ellipse outlines a high-velocity anomaly embedded in the general low-velocity block on the right. (b) Black stars represent earthquakes at their initial locations. The initial earthquake locations are determined using the eikonal equation-based grid search method (see Appendix B) within a velocity model that matches the target model's velocity at the surface and linearly increases to 7.95 km s^{-1} at a depth of 30 km. (c) Inversion results using first arrival times. (d) Subsequent refinement of the velocity model and earthquake locations using common-source differential arrival times, building upon the outcome of ATT. (e) Further updated velocity model and earthquake locations by common-receiver differential arrival times, building upon the outcome of ATT and ATT-CS. All the others are the same as Fig. 7.

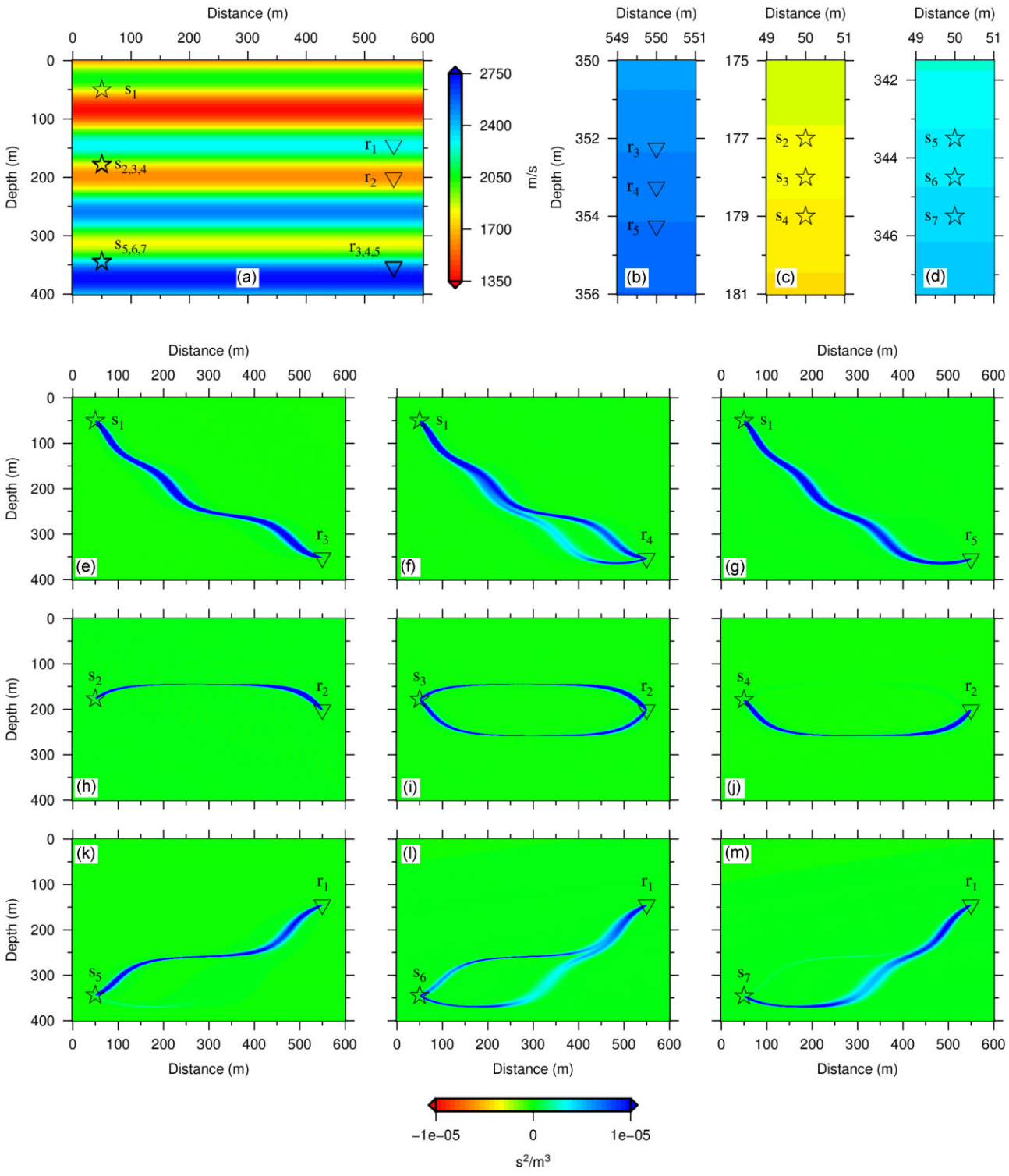

Figure 10. (a) A vertically varying velocity model in a cross-borehole setting. The stars in the left well are seven sources, and the inverted triangles in the right well denote five receivers. (b)–(d) Zoom-in views of receivers 3–5, sources 2–4 and sources 5–7, respectively. (e)–(g) Individual kernels from source 1 to receiver 3, 4 and 5, respectively. (h)–(j) Individual kernels from sources 2, 3, 4 to receiver 2, respectively. (k)–(m) Individual kernels from sources 5, 6, 7 to receiver 1, respectively. All the individual kernels are computed in the velocity model of (a) by assuming 1 s traveltime misfit at the respective receiver locations with the ATT method. The colour scale for the individual kernels is placed at the bottom.

Figs 10(a)–(d) depict a cross-borehole setting with seven sources located in one well and five receivers placed in the other well. The reference velocity model between the two wells only varies in the vertical direction. Individual traveltime sensitivity kernels are computed by assuming 1 s traveltime misfit at the respective receiver locations (Figs 10e–m). Multipathing phenomena can be observed between some sources and receivers (Figs 10f, i and l). As discussed earlier, traveltime misfits are projected along the multipathing routes. In general,

multipathing indicates the complexity of seismic wave propagation and difficulty of reliable seismic imaging in heterogeneous media. To generate reliable velocity models for geologically complex regions, advanced inversion techniques such as the adjoint-state tomography methods developed in this study are necessary.

5 CONCLUSIONS

We have successfully developed two adjoint-state differential arrival time tomography methods by extending the ATT method to invert common-source differential arrival times and common-receiver differential arrival times. The only difference among these tomography methods is the source term of the adjoint equation, which characterizes the observational misfits at receiver locations. The three methods, namely the ATT, ATT-CS and ATT-CR methods, use the same multiple-grid model parametrization, which is needed when updating the velocity model, and the same optimization algorithm to iteratively update velocity models and/or earthquake locations. Synthetic tests have verified the validity of these three tomography methods, indicating their readiness for real data inversions. The main features that distinguish ATT, ATT-CS and ATT-CS from other ray-based traveltime tomography approaches, including Thurber (1983), Waldhauser & Ellsworth (2000), Zhang & Thurber (2003), Guo & Zhang (2017) and Guo *et al.* (2021), can be summarized in three aspects:

(1) Ray tracing independence: ATT, ATT-CS and ATT-CS do not rely on traditional ray tracing methods.
(2) Efficient sensitivity kernel computation: These methods use the efficient adjoint-state technique to calculate sensitivity kernels of the objective function with respect to model parameters.
(3) Incorporation of multiple-grid model parametrization.

It is also worth noting that the ATT method has already been extended to image both velocity heterogeneity and azimuthal anisotropy by using a more sophisticated elliptical eikonal equation to model first-arrival traveltime fields (Tong 2021b; Chen *et al.* 2023).

Common-source differential arrival times are primarily sensitive to receiver-side velocity structures. On the other hand, common-receiver differential arrival times provide additional constraints on source-side velocity structures. Both adjoint-state differential arrival time tomography methods demonstrate improved capability in recovering small-scale velocity structures compared to first-arrival tomography. When both earthquake locations and velocity structure are not accurately known, it becomes crucial to determine them simultaneously. Numerical examples further support the sensitivity of common-receiver differential arrival times to earthquake locations, a characteristic that has already been used by the double-difference earthquake location method proposed by Waldhauser & Ellsworth (2000). To fully utilize all three types of arrival time data, it is recommended to perform a joint or sequential inversion of first arrival times, common-source differential arrival times, and common-receiver differential arrival times, in order to obtain reliable subsurface velocity models and earthquake locations, as has been done by Guo *et al.* (2021) and Share *et al.* (2021) using ray-based seismic tomography.

The present study has shown that the three seismic tomography methods, ATT, ATT-CS and ATT-CR, have the ability to accurately identify subsurface influence zones on arrival time misfits or differential arrival time misfits at receiver locations, even in the presence of multipathing, by numerically solving the adjoint equations with the fast sweeping method. We conclude that the ATT method using first arrival times and the two adjoint-state differential arrival time tomography methods are promising techniques that have the potential to become standard tools in future seismic tomographic inversion studies.

ACKNOWLEDGMENTS

We are grateful for Carl Tape, Ebru Bozdag, Tianze Liu and one anonymous reviewer for their valuable and insightful comments, which have greatly contributed to the improvement of this paper. This work was supported by the Ministry of Education, Singapore, under its MOE AcRF Tier 1 grant (RG86/22). PT and TL were also supported by the National Research Foundation Singapore and the Ministry of Education, Singapore, under the Research Centres of Excellence initiative (04MNS001953A620). All the figures are made with the Generic Mapping Tool (GMT; Wessel & Smith 1991). The codes that support the findings of this study are available from the corresponding author, PT, upon reasonable request.

DATA AVAILABILITY

No new data were generated or analysed in support of this research.

REFERENCES

Aki, K. & Richards, P.G., 2002. *Quantitative Seismology: Theory and Methods,* 2nd edn, University Science Books.

Allam, A. & Ben-Zion, Y., 2012. Seismic velocity structures in the southern California plate-boundary environment from double-difference tomography, *Geophys. J. Int.,* **190,** 1181–1196.

Boyce, A., Bastow, I.D., Golos, E.M., Rondenay, S., Burdick, S. & Van der Hilst, R.D., 2019. Variable modification of continental lithosphere during the proterozoic grenville orogeny: evidence from teleseismic P-wave tomography, *Earth planet. Sci. Lett.,* **525,** doi:10.1016/j.epsl.2019.115763.

Chen, J., Chen, G., Wu, H., Yao, J. & Tong, P., 2022. Adjoint tomography of Northeast Japan revealed by common-source double-difference traveltime data, *Seismol. Res. Lett.,* **93,** 1835–1851.

Chen, J., Chen, G., Nagaso, M. & Tong, P., 2023. Adjoint-state traveltime tomography for azimuthally anisotropic media in spherical coordinates, *Geophys. J. Int.,* **234**(1), 712–736.

Cheng, F., Chi, B., Lindsey, N.J., Dawe, T.C. & Ajo-Franklin, J.B., 2021. Utilizing distributed acoustic sensing and ocean bottom fiber optic cables for submarine structural characterization, *Sci. Rep.,* **11,** doi:10.1038/s41598-021-84845-y.

Fichtner, A., Bunge, H.P. & Igel, H., 2006. The adjoint method in seismology – I. Theory, *Phys. Earth planet. Inter.,* **157,** 86–104.

Guo, H. & Zhang, H., 2017. Development of double-pair double difference earthquake location algorithm for improving earthquake locations, *Geophys. J. Int.,* **208,** 333–348.

Guo, H., McGuire, J.J. & Zhang, H., 2021. Correlation of porosity variations and rheological transitions on the southern Cascadia megathrust, *Nat. Geosci.,* **14,** 341–348.

Leung, S. & Qian, J., 2006. An adjoint state method for three-dimensional transmission traveltime tomography using first-arrivals, *Commun. Math. Sci.,* **4,** 249–266.

Liu, S., Suardi, I., Yang, D., Wei, S. & Tong, P., 2018. Teleseismic traveltime tomography of northern Sumatra, *Geophys. Res. Lett.,* **45**(24), 13 231–13 239.

Liu, S., Suardi, I., Xu, X., Yang, S. & Tong, P., 2021. The geometry of the subducted slab beneath Sumatra revealed by regional and teleseismic traveltime tomography, *J. geophys. Res.,* **126**(1), e2020JB020169. doi: 10.1029/2020JB020169.

Liu, Y. & Tong, P., 2021. Eikonal equation-based P-wave seismic azimuthal anisotropy tomography of the crustal structure beneath northern California, *Geophys. J. Int.,* **226,** 287–301.

Obayashi, M., Suetsugu, D. & Fukao, Y., 2004. PP-P differential traveltime measurement with crustal correction, *Geophys. J. Int.,* **157,** 1152–1162.

Qian, J., Zhang, Y. & Zhao, H., 2007. A fast sweeping method for static convex Hamilton-Jacobi equations, *J. Sci. Comput.,* **31,** 237–271.

Rawlinson, N., Hauser, J. & Sambridge, M., 2008. Seismic ray tracing and wavefront tracking in laterally heterogeneous media, *Adv. Geophys.,* **49,** 203–273.

Rouy, E. & Tourin, A., 1992. A viscosity solutions approach to shape-from-shading, *SIAM J. Numer. Anal.,* **29,** 867–884.

Sambridge, M.S. & Kennett, B. L.N., 1986. A novel method of hypocentre location, *J. geophys. Int.,* **87**(2), 679–697.

Sethian, J.A., 1996. A fast marching level set method for monotonically advancing fronts, *Proc. Natl. Acad. Sci. U.S.A.,* **93,** 1591–1595.

Share, P.E., Guo, H., Thurber, C., Zhang, H. & Ben-Zion, Y., 2019. Seismic imaging of the Southern California plate boundary around the south-central Transverse Ranges using double-difference tomography, *Pure Appl. Geophys.,* **176,** 1117–1143.

Share, P.-E., Castro, R., Vidal-Villegas, J., Mendoza, L. & Ben-Zion, Y., 2021. High-resolution seismic imaging of the plate boundary in northern Baja California and Southern California using double-pair double-difference tomography, *Earth planet. Sci. Lett.,* **568,** doi:10.1016/j.epsl.2021.117004.

Shearer, P.M., 1997. Improving local earthquake locations using the L1 norm and waveform cross correlation: application to the Whittier Narrows, California, afterschock sequence, *J. geophys. Res.,* **102,** 8269–8283.

Taillandier, C., Noble, M., Chauris, H. & Calandra, H., 2009. First-arrival traveltime tomography based on the adjoint-state method, *Geophysics,* **74,** WCB57–WCB66.

Tavakoli, F., Operto, S., Ribodetti, A. & Virieux, J., 2017. Slope tomography based on eikonal solvers and the adjoint-state method, *Geophys. J. Int.,* **209,** 1629–1647.

Thurber, C., Zhang, H., Waldhauser, F., Hardebeck, J., Michael, A. & Eberhart-Phillips, D., 2006. Three-dimensional compressional wavespeed model, earthquake relocations, and focal mechanisms for the Parkfield, California, region, *Bull. seism. Soc. Am.,* **96**(4B), S38–S49.

Thurber, C.H., 1983. Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California, *J. geophys. Res.,* **88,** 8226–8236.

Thurber, C.H., Atre, S. & Eberhart-Phillips, D., 1995. Three-dimensional Vp and Vp/Vs structure at Loma Prieta, California, from local earthquake tomography, *Geophys. Res. Letts.,* **22,** 3079–3082.

Tong, P., 2021a. Adjoint-state traveltime tomography: eikonal equation-based methods and application to the Anza area in southern California, *J. geophys. Res.,* **126,** e2021JB021818. doi: 10.1029/2021JB021818.

Tong, P., 2021b. Adjoint-state traveltime tomography for azimuthally anisotropic media and insight into the crustal structure of central California near Parkfield, *J. geophys. Res.,* **126,** e2021JB022365. doi: 10.1029/2021JB022365.

Tong, P., Yang, D. & Huang, X., 2019. Multiple-grid model parametrization for seismic tomography with application to the San Jacinto fault zone, *Geophys. J. Int.,* **218,** 200–223.

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

Umino, N., Hasegawa, A. & Matsuzawa, T., 1995. sP depth phase at small epicentral distances and estimated subducting plate boundary, *Geophys. J. Int.,* **120,** 356–366.

Waldhauser, F. & Ellsworth, W.L., 2000. A double-difference earthquake location algorithm: method and application to the northern Hayward fault, California, *Bull. seism. Soc. Am.,* **90**(6), 1353–1368.

Wessel, P. & Smith, W. H.F., 1991. Free software helps map and display data, *EOS, Trans. Am. geophys. Un.,* **72**(1), 441–448.

Wu, W. & Irving, J. C.E., 2020. Array-based iterative measurements of SmKS travel times and their constraints on outermost core structure, *J. geophys. Res.,* **125**(3), e2019JB018162.

Yuan, Y.O., Simons, F.J. & Tromp, J., 2016. Double-difference adjoint seismic tomography, *Geophys. J. Int.,* **206,** 1599–1618.

Zenonos, A., Siena, L.D., Widiyantoro, S. & Rawlinson, N., 2020. Direct inversion of S-P differential arrival times for Vp/Vs ratio in SE Asia, *J. geophys. Res.,* **125,** e2019JB019152.

Zhan, Z., 2020. Distributed acoustic sensing turns fiber-optic cables into sensitive seismic antennas, *Seismol. Res. Lett.,* **91,** 1–15.

Zhang, H. & Thurber, C.H., 2003. Double-difference tomography: the method and its application to the Hayward fault, California, *Bull. seism. Soc. Am.,* **93**(5), 1875–1889.

Zhang, H., Nadeau, R.M. & Guo, H., 2017. Imaging the nonvolcanic tremor zone beneath the San Andreas fault at Cholame, California using station-pair double-difference tomography, *Earth planet. Sci. Lett.,* **460,** 76–85.

Zhao, H., 2004. A fast sweeping method for Eikonal equations, *Math. Comput.,* **74**(250), 603–627.

APPENDIX A: SECOND-ORDER DIFFERENTIAL ARRIVAL TIME TOMOGRAPHY

A second-order differential arrival time measures the difference between two common-source differential arrival times corresponding to two earthquakes ($\mathbf{x}_{s,n}$, $\mathbf{x}_{s,j}$) and the same pair of seismic stations ($\mathbf{x}_{r,m}$, $\mathbf{x}_{r,k}$), that is the difference between $T_n^o(\mathbf{x}_{r,m}) - T_n^o(\mathbf{x}_{r,k})$ and $T_j^o(\mathbf{x}_{r,m}) - T_j^o(\mathbf{x}_{r,k})$. It is equivalent to the difference between two common-receiver differential arrival times corresponding to two seismic stations and the same pair of earthquakes, that is the difference between $\left[T_n^o(\mathbf{x}_{r,m}) - \tau_n\right] - \left[T_j^o(\mathbf{x}_{r,m}) - \tau_j\right]$ and $\left[T_n^o(\mathbf{x}_{r,k}) - \tau_n\right] - \left[T_j^o(\mathbf{x}_{r,k}) - \tau_j\right]$. Accordingly, the objective function of second-order differential arrival time tomography can be defined as

$$\chi\left(s(\mathbf{x}),\ \mathbf{x}_{s,1},\ \tau_1, \cdots,\ \mathbf{x}_{s,n},\ \tau_n, \cdots,\ \mathbf{x}_{s,N},\ \tau_N\right) = \sum_{m=1}^{M}\sum_{k=1}^{M}\sum_{n=1}^{N}\sum_{j=1}^{N}\frac{\omega_{n,j}^{m,k}}{2}\left(\Delta t_{n,j}^{m,k}\right)^2, \tag{A1}$$

where

$$\Delta t_{n,j}^{m,k} = \left\{\left[T_n(\mathbf{x}_{r,m}) - T_n(\mathbf{x}_{r,k})\right] - \left[T_j(\mathbf{x}_{r,m}) - T_j(\mathbf{x}_{r,k})\right]\right\} - \left\{\left[T_n^o(\mathbf{x}_{r,m}) - T_n^o(\mathbf{x}_{r,k})\right] - \left[T_j^o(\mathbf{x}_{r,m}) - T_j^o(\mathbf{x}_{r,k})\right]\right\}, \tag{A2}$$

and $\omega_{n,j}^{m,k}$ is the weight associated with the second-order differential arrival time.

The variation of the objective function due to an infinitesimal perturbation $\delta s(\mathbf{x})$ in the slowness model $s(\mathbf{x})$ can be expressed by

$$\delta\chi = \sum_{m=1}^{M}\sum_{k=1}^{M}\sum_{n=1}^{N}\sum_{j=1}^{N} 4\omega_{n,j}^{m,k}\Delta t_{n,j}^{m,k}\int_{\Omega}\delta T_n(\mathbf{x})\delta(\mathbf{x}-\mathbf{x}_{r,m})\mathrm{d}\mathbf{x}, \tag{A3}$$

where $\delta T_n(\mathbf{x})$ represents the perturbation in the traveltime field. Similar to the discussion in Sections 3.1 and 3.2, introducing a test function $P_n(\mathbf{x})$ that satisfies the boundary value problem (adjoint equation),

$$\begin{cases}\nabla\cdot\left[P_n(\mathbf{x})\nabla\left(-T_n(\mathbf{x})\right)\right] = \sum_{m=1}^{M}\sum_{k=1}^{M}\sum_{j=1}^{N} 4\omega_{n,j}^{m,k}\Delta t_{n,j}^{m,k}\delta(\mathbf{x}-\mathbf{x}_{r,m}), & \mathbf{x}\in\Omega\\ P_n(\mathbf{x}) = 0, & \mathbf{x}\in\partial\Omega,\end{cases} \tag{A4}$$

we can establish the relationship between $\delta\chi$ and the causing factor $\delta s(\mathbf{x})/s(\mathbf{x})$ as

$$\delta\chi = \sum_{n=1}^{N}\int_{\Omega} P_n(\mathbf{x})s(\mathbf{x})\delta s(\mathbf{x})\mathrm{d}\mathbf{x} = \int_{\Omega} K_s(\mathbf{x})\frac{\delta s(\mathbf{x})}{s(\mathbf{x})}\mathrm{d}\mathbf{x}, \tag{A5}$$

where the misfit kernel is

$$K_s(\mathbf{x}) = \sum_{n=1}^{N} K_{s,n}(\mathbf{x}) = \sum_{n=1}^{N} P_n(\mathbf{x})s^2(\mathbf{x}). \tag{A6}$$

To locate earthquakes, we examine the variation of the objective function due to an infinitesimal perturbation $(\delta\mathbf{x}_{s,n},\ \delta\tau_n)$ in the earthquake location $(\mathbf{x}_{s,n},\ \tau_n)$,

$$\begin{aligned}\delta\chi &= \sum_{m=1}^{M}\sum_{k=1}^{M}\sum_{n=1}^{N}\sum_{j=1}^{N}\omega_{n,j}^{m,k}\Delta t_{n,j}^{m,k}\left\{\left[\delta T_n(\mathbf{x}_{r,m}) - \delta T_n(\mathbf{x}_{r,k})\right] - \left[\delta T_j(\mathbf{x}_{r,m}) - \delta T_j(\mathbf{x}_{r,k})\right]\right\}\\ &= \sum_{m=1}^{M}\sum_{k=1}^{M}\sum_{n=1}^{N}\sum_{j=1}^{N} 4\omega_{n,j}^{m,k}\Delta t_{n,j}^{m,k}\delta T_n(\mathbf{x}_{r,m}).\end{aligned} \tag{A7}$$

We assume that $\Gamma_m(\mathbf{x})$ is the traveltime field initiated at the receiver location $\mathbf{x}_{r,m}$. Based on the reciprocity principle (Aki & Richards 2002), we have $\delta T_n(\mathbf{x}_{r,m}) = \nabla\Gamma_m(\mathbf{x}_{s,n})\cdot\delta\mathbf{x}_{s,n}$. Therefore, the full derivative of the objective function (eq. A1) is

$$\begin{aligned}&\delta\chi\left(s(\mathbf{x}),\ \mathbf{x}_{s,1},\ \tau_1,\cdots,\ \mathbf{x}_{s,n},\ \tau_n,\ \cdots,\ \mathbf{x}_{s,N},\ \tau_N\right)\\ &= \int_{\Omega} K_s(\mathbf{x})\frac{\delta s(\mathbf{x})}{s(\mathbf{x})}\mathrm{d}\mathbf{x} + \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{M}\sum_{j=1}^{N} 4\omega_{n,j}^{m,k}\Delta t_{n,j}^{m,k}\nabla\Gamma_m(\mathbf{x}_{s,n})\cdot\delta\mathbf{x}_{s,n},\end{aligned} \tag{A8}$$

where $K_s(\mathbf{x})$ is given by eq. (A6). The subsequent steps are similar to those discussed in the ATT, ATT-CS and ATT-CR methods.

APPENDIX B: EIKONAL EQUATION-BASED GRID SEARCH FOR EARTHQUAKE LOCATION

Grid search is a popular earthquake location algorithm known for its simplicity and capacity to determine globally optimal hypocentre parameters (Sambridge & Kennett 1986). Here, we present an eikonal equation-based earthquake location method, which follows a grid-search approach for earthquake localization. A set of N earthquakes is denoted as $(\mathbf{x}_{s,n},\ \tau_n)$ $(n = 1, 2, \cdots, N)$, and M seismic stations as $\mathbf{x}_{r,m}$ $(m = 1, 2, \cdots, M)$. $T_n(\mathbf{x})$ represents the traveltime field generated by the nth earthquake. To accurately relocate the nth earthquake $(\mathbf{x}_{s,n},\ \tau_n)$, we define a misfit function that uses the first arrival times $T_n^o(\mathbf{x}_{r,m})$ from this earthquake to all seismic stations,

$$G(\mathbf{x},\ \tau) = \sum_{m=1}^{M}\frac{\omega_{n,m}}{2}\left\{\Gamma_m(\mathbf{x}) - \left[T_n^o(\mathbf{x}_{r,m}) - \tau\right]\right\}^2, \tag{B1}$$

where $\Gamma_m(\mathbf{x})$ represents the traveltime field initiated at the mth receiver. Following the reciprocity principle assumption (Aki & Richards 2002), $\Gamma_m(\mathbf{x}_{s,n}) = T_n(\mathbf{x}_{r,m})$, and thus, the minimizer of $G(\mathbf{x},\ \tau)$ can be regarded as the actual location of the nth earthquake $(\mathbf{x}_{s,n},\ \tau_n)$.

At any local minimum ($\mathbf{x}^*$,, τ^*), the derivative of $G(\mathbf{x}, \tau)$ with respect to τ is zero. Therefore, by equating $\partial G(\mathbf{x}, \tau)/\partial\tau$ to zero, we obtain a general function for τ,

$$\tau(\mathbf{x}) = \frac{\sum_{m=1}^{M} \omega_{n,m} \left[T_n^o(\mathbf{x}_{r,m}) - \Gamma_m(\mathbf{x})\right]}{\sum_{m=1}^{M} \omega_{n,m}}. \tag{B2}$$

Substituting $\tau(\mathbf{x})$ into eq. (B1), we arrive at a function dependent solely on $\mathbf{x}$,

$$F(\mathbf{x}) = G\left(\mathbf{x}, \tau(\mathbf{x})\right) = \sum_{m=1}^{M} \frac{\omega_{n,m}}{2} \left\{\Gamma_m(\mathbf{x}) - \left[T_n^o(\mathbf{x}_{r,m}) - \tau(\mathbf{x})\right]\right\}^2, \tag{B3}$$

For a given grid with L nodes, M eikonal equations are solved to obtain $\Gamma_m(\mathbf{x})$ at every node, and hence the value of $F(\mathbf{x})$. If $F(\mathbf{x})$ reaches its minimum at the grid node $\mathbf{x}^*$, then $\mathbf{x}^*$ is considered the earthquake location, along with the origin time $\tau(\mathbf{x}^*)$.

© The Author(s) 2023. Published by Oxford University Press on behalf of The Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.