

# Multiple-grid model parametrization for seismic tomography with application to the San Jacinto fault zone

Ping Tong<sup>1,2</sup>, Dinghui Yang<sup>3</sup> and Xueyuan Huang<sup>4</sup>

<sup>1</sup>*Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore.*

*E-mail: tongping@ntu.edu.sg*

<sup>2</sup>*Asian School of the Environment, Nanyang Technological University, 637459, Singapore*

<sup>3</sup>*Department of Mathematical Sciences, Tsinghua University, 100875, Beijing, China*

<sup>4</sup>*Earth Observatory of Singapore, Nanyang Technological University, 639798, Singapore*

Accepted 2019 March 20. Received 2018 March 27; in original form 2019 March 15

## SUMMARY

We discuss the performance of the multiple-grid model parametrization in seismic tomographic inversion. Rather than mapping the velocity perturbation  $\Delta c(\mathbf{x}) = c(\mathbf{x}) - c_0(\mathbf{x})$  on only one regular/collocated grid as many previous studies did, we obtain individual  $\Delta c(\mathbf{x})$  models on multiple grids and generate several updated velocity models during one iteration. The average of all the updated velocity models is considered to be the input model of the next iteration. Different grids should partially/fully shift from each other and/or have different grid spacings to form the multiple-grid model parametrization. The efficacy of the multiple-grid model parametrization is demonstrated through the practical example of imaging the *P*-wave velocity structure along the San Jacinto fault, which is one of the most seismically active faults in California. A series of synthetic recovery examples shows that the multiple-grid model parametrization generally has a better or comparable performance in capturing the heterogeneous subsurface structures than a collocated grid. The root mean square values of the traveltimes residuals in the final tomographic models obtained with the multiple-grid model parametrization are smaller than those with collocated grids. Tomographic results reveal strong heterogeneities in the crust along the San Jacinto fault. Significant velocity contrasts are observable across the fault at shallow depths. A low-velocity anomaly dominates the trifurcation area of the San Jacinto fault from the middle crust to the lower crust. Relatively large earthquakes occurred at the boundaries of low-velocity structures but with high-velocity anomalies nearby. All the results suggest that the multiple-grid model parametrization can be a reliable approach in future seismic tomography studies.

**Key words:** Seismic tomography; Dynamics: seismotectonics; Crustal Imaging.

## 1 INTRODUCTION

Seismic tomography is one of the primary tools that image the subsurface structure of the Earth at a variety of scales (e.g. Rawlinson *et al.* 2010; Liu & Gu 2012; Zhao 2015; Tong *et al.* 2017). It has greatly improved our knowledge of the Earth's internal structure and dynamics ever since the pioneer studies of Aki & Lee (1976) and Dziewonski *et al.* (1977). It is well known that the final image of one seismic tomographic investigation is determined by many factors. Driven by the constant demand for higher resolution subsurface images, in the past 40 yr seismic tomography techniques have achieved significant progresses in many aspects including model parametrization and regularization (e.g. Aki & Lee 1976; Thurber 1983; Sambridge & Rawlinson 2005), data usage (e.g. body waves versus surface waves versus ambient noise; Aki & Lee 1976; Shapiro

*et al.* 2005; Zhao *et al.* 2005; Yang & Ritzwoller 2008), approximation of seismic data sensitivity (e.g. geometrical ray theory versus ray-based finite-frequency kernel versus full-wave-equation-based finite-frequency kernel; Thurber 1983; Dahlen *et al.* 2000; Tromp *et al.* 2005; Tong *et al.* 2014a) and many others. These developments enable us to resolve the Earth's internal structure in unprecedented details nowadays.

This study discusses model parametrization, the way of representing subsurface structure. In principle, the dominant factors that affect the choice of a particular model parametrization include forward solvers, inverse algorithms, *a priori* knowledge of Earth's structure and resolution capability of available data set (Kissling *et al.* 2001). To date, different basis functions have been used to express 3-D structure of the Earth's interior (e.g. Rawlinson *et al.* 2010; Zhao 2015). Among them, blocks and grids are

the most basic forms of parametrization. Aki & Lee (1976) used constant slowness (the inverse of velocity) blocks to describe the 3-D crustal structure beneath a seismic array. In this approach, artificial velocity discontinuities are introduced in between block boundaries and velocity anomalies are forced to take the shapes of blocks (e.g. Kissling *et al.* 2001; Zhao 2009). Alternatively, Thurber (1983) represented velocity models on a 3-D grid of variable node spacing. The grid approach together with its later variants have become very popular in local earthquake tomography as well as in teleseismic tomography (e.g. Zhao *et al.* 1992; Thurber & Eberhart-Philips 1999; Zhao 2009). Spectral parametrizations in the wavenumber domain and irregular parametrizations have been widely used and discussed in previous studies as well (e.g. Dziewonski *et al.* 1977; Dziewonski 1984; Sambridge *et al.* 1995; Zhang & Thurber 2005; Bodin & Sambridge 2009; Bodin *et al.* 2012).

In a common seismic velocity tomography study, one needs to determine the perturbation  $\Delta c(\mathbf{x})$  of the true velocity  $c_{\text{true}}(\mathbf{x})$  from the initial/starting velocity  $c_0(\mathbf{x})$ . For the accuracy of ray tracing or numerical simulation of wave propagation, the initial velocity  $c_0(\mathbf{x})$  and its updates are usually discretized on small blocks or fine grids (e.g. Kissling *et al.* 2001; Tape *et al.* 2009; Tong *et al.* 2017). However, mainly due to the constraint caused by the resolving ability of the data, the perturbation field  $\Delta c(\mathbf{x})$  should be constructed on some large blocks or grids with wide spacing (e.g. Aki & Lee 1976; Zhao *et al.* 1992; Kissling *et al.* 2001; Thurber *et al.* 2009). Some recent wave-equation-based seismic tomography studies, where the perturbation fields are directly given by the sensitivity kernels on fine grids, seem to be the exceptions (e.g. Tape *et al.* 2009; Tong *et al.* 2014a). But the dominant wavelength of the waves that are used to construct sensitivity kernels is at least a few times as long as the grid size. In this regard, only the features that are much larger than the size of the fine grid can be recovered. In the current ray-based traveltimes tomography study, we focus on the parametrization of the relative perturbation field  $\Delta c(\mathbf{x})/c(\mathbf{x})$  on relatively coarse grids, as this is frequently used in previous studies (e.g. Thurber 1983; Nakajima & Hasegawa 2007; Lin 2013; Tong *et al.* 2017; Liu *et al.* 2018). The coarse grid that is selected to represent the perturbation field is always called the inversion grid (Kissling *et al.* 2001; Rawlinson *et al.* 2010).

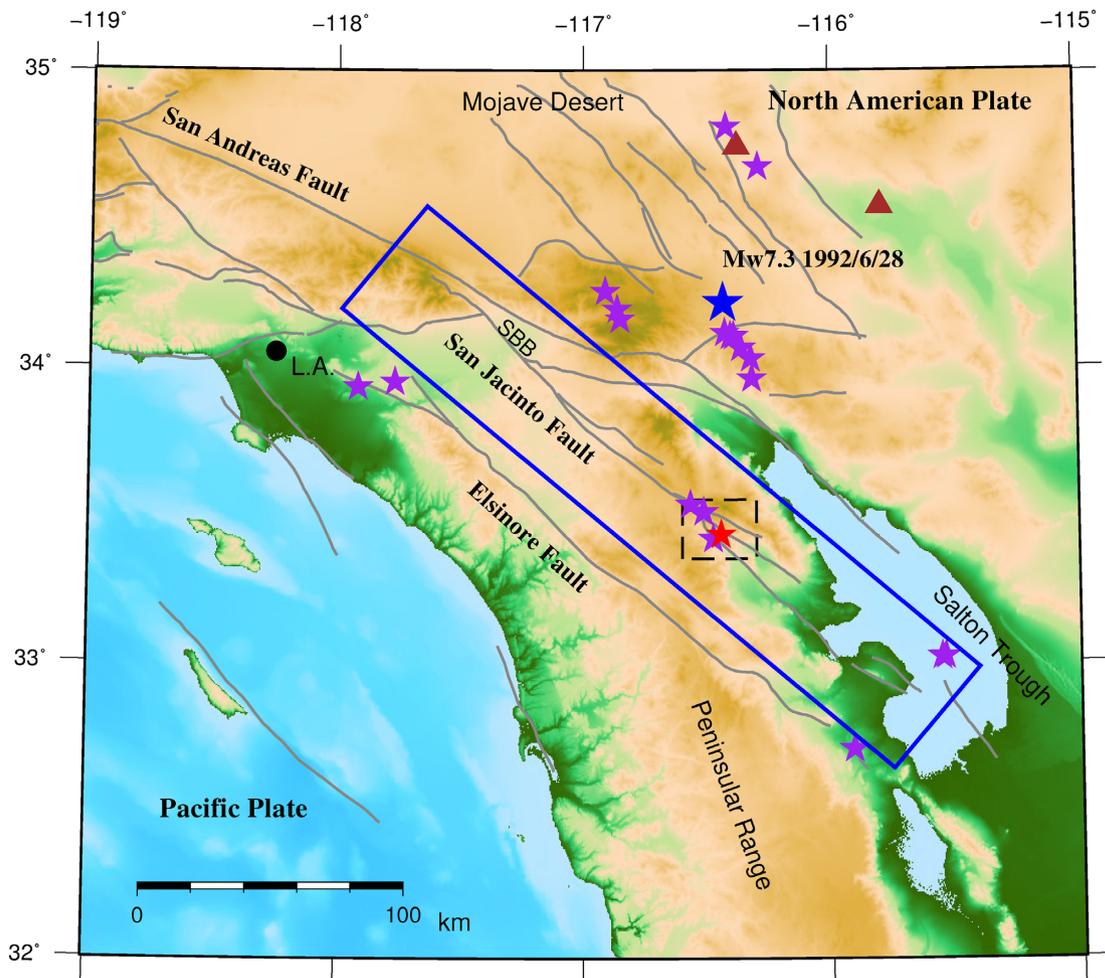
Synthetic resolution tests can reflect the resolving ability of data as well as the efficacy of model parametrization (e.g. Inoue *et al.* 1990; Zhao *et al.* 1992; Kissling *et al.* 2001; Tong *et al.* 2017). In such a test, starting with an initial model  $c_0(\mathbf{x})$ , the synthetic data that is generated with the same source–receiver geometry of the observational experiment are inverted to recover the designed ‘true’ model  $c_{\text{true}}(\mathbf{x})$ . The checkerboard resolution test, with alternating fast and slow anomalies superimposed on the initial model  $c_0(\mathbf{x})$ , is one of the most commonly used approaches (e.g. Zhao 2001; Rawlinson *et al.* 2010; Tong *et al.* 2014c). In many studies with the grid model parametrization (e.g. Nakajima & Hasegawa 2007; Jiang *et al.* 2009; Thurber *et al.* 2009), the designed fast/slow anomalies are usually of size one node (e.g. Zhao *et al.* 1992; Nakajima & Hasegawa 2007; Tong *et al.* 2011) or two nodes in one direction (e.g. Thurber *et al.* 2009; Zhang *et al.* 2009; Hofstetter *et al.* 2012). If the checkerboard pattern can be fully or almost fully recovered, it may suggest that the size of the anomaly can be reduced or remain the same, otherwise the anomaly size should be increased to match the resolution of the data. In the mean time, the grid spacing is changed accordingly. Repeating this process, one can eventually find an optimal grid that is able to resolve anomalies at the resolution scale of the data.

The grid selection is subjective. Even when the optimal grid spacing is determined, where to place the first grid node is still subjective. We know that the property recovered at one grid node is an average within a local volume. The averaging effect may obscure some details such as irregular boundaries between different rocks, little heterogeneities and slow velocity variations (Vesnaver & Bohm 2000). Considering several spatially staggered grids with the same spacing, each of them should have the same opportunity to be selected for the tomographic inversion, and the recovering rate of the checkerboard resolution test on one grid is probably similar to that on another. To reduce the bias caused by the subjective grid selection in seismic tomography, we can invert the same data set on each of the spatially staggered grids and obtain the final result by averaging. This is the so-called staggered-grid model parametrization (e.g. Vesnaver & Bohm 2000; Arato *et al.* 2014). As it roughly operates as a moving average, the staggered-grid model parametrization can reduce ambiguities, smearing and instabilities of the tomographic inversion (Vesnaver & Bohm 2000; Arato *et al.* 2014). A staggered-grid usually contains several regular grids of the same size.

Non-uniqueness is unavoidable in seismic tomographic inversion largely because of poor and uneven data coverage. The multiscale tomography method was introduced to reduce the non-uniqueness in seismic traveltimes tomography (e.g. Zhou 1996, 2003; Mewes *et al.* 2010). By the multiscale method, a series of cells or grids with different sizes are used to represent the velocity model, similar as the staggered-grid approach discussed previously. For instance, Zhou (1996) uses a series of overlapping cells of  $K$  different sizes to represent the mantle, while cells of the same size are non-overlapping. One linear system is formulated on all the overlapping cells and then inverted to generate reliable results. The multiscale approach acts as a regularization technique for the inverse problem. It can accelerate the convergence of the solution (Mewes *et al.* 2010). In comparison with the single-scale seismic tomography method, multiscale traveltimes tomography increases the chance to realize the most suitable cell size with respect to data coverage and anomaly geometry, and generates superior results (Zhou 2003). If the grid approach is adopted, multiscale seismic tomography uses the multi-grid model parametrization. A multigrid usually consists of several regular/collocated grids with different spacings.

Since both the staggered-grid and multigrid model parametrizations have more than one grid involved, we call them as multiple-grid model parametrization in a uniform way. A multiple-grid could be a staggered-grid or a multigrid. In addition to that, several staggered-grids with different spacings can also form a multiple-grid that may inherit the advantages of both the staggered-grid and multigrid approaches. The third type of multiple-grid is named as multiscale-staggered-grid in this study. To evaluate the performance of the multiple-grid model parametrization in real applications, we choose the San Jacinto fault zone in southern California as the test field (Fig. 1).

The San Jacinto fault is a major component of the large San Andreas fault system (Sharp 1967). This right-lateral strike-slip fault is characterized by high seismicity and considered to be the most seismically active fault zone in southern California. Specifically, more than 10 per cent of all earthquakes in southern California since 2000 occurred and clustered around the trifurcation area of the San Jacinto fault zone (Ross *et al.* 2017), including the recent 2016  $M_w$  5.2 Borrego Springs earthquake (Fig. 1). The San Jacinto fault zone has been extensively investigated in previous seismic tomography studies (e.g. Scott *et al.* 1994; Allam & Ben-Zion 2012; Lin 2013). The common features reported by these early studies are the strong velocity contrast and damage zones across the San Jacinto



**Figure 1.** The tectonic setting and surface topography around the San Jacinto fault zone in southern California. The grey curves denote the active faults including the San Jacinto fault, the San Andreas fault and the Elsinore fault. The brown triangles are the volcanoes. The stars represent the locations of relatively large earthquakes ( $M_w > 5.0$ ) since 1992. The 1992  $M_w$  7.3 Lander earthquake is shown as the blue star. The red star is the 2016  $M_w$  5.2 Borrego Springs earthquake. SBB is short for San Bernadino Basin. The dashed box surrounds the trifurcation area. The blue box along the San Jacinto fault indicates the present study area.

fault (Allam & Ben-Zion 2012; Lin 2013). Later results from noise tomography (Zigone *et al.* 2015), wave-equation-based traveltimes seismic tomography (Tong *et al.* 2014c) and wave-equation-based reflection tomography (Huang *et al.* 2016) also show strong velocity contrasts across the San Jacinto fault. In this study, we demonstrate the advantages of the multiple-grid model parametrization by imaging the detailed  $P$ -wave velocity structure of the San Jacinto fault zone.

## 2 METHOD

The ray-based traveltimes seismic tomography method for the velocity structure  $c(\mathbf{x})$  (e.g. Aki & Lee 1976; Thurber 1983; Zhao *et al.* 1992; Tong *et al.* 2017) is conceptually based on the following equation:

$$T_m^{\text{obs}} - T_m^{\text{cal}} = \int_{R_m} \left[ -\frac{1}{c_0(\mathbf{x})} \frac{\Delta c(\mathbf{x})}{c(\mathbf{x})} \right] ds, \quad (1)$$

where  $T_m^{\text{obs}}$  and  $T_m^{\text{cal}}$  are the observed and calculated traveltimes of the  $m$ th recorded seismic phase (such as  $P$  or  $S$  wave).  $R_m$  is the ray path of the wave from earthquake location to the receiver in the initial model  $c_0(\mathbf{x})$ .  $c(\mathbf{x})$  is the true or ‘targeted’ velocity model.

$\Delta c(\mathbf{x}) = c(\mathbf{x}) - c_0(\mathbf{x})$  measures the difference between the true and initial velocities. If the ray path is connected by a series of segments, and the central point of the  $i$ th segment ( $i = 1, 2, \dots, I_m$ ) is  $\mathbf{x}_i^m$ , then eq. (1) can be discretely expressed as

$$T_m^{\text{obs}} - T_m^{\text{cal}} = \sum_{i=1}^{I_m} \left[ -\frac{1}{c_0(\mathbf{x}_i^m)} \frac{\Delta c(\mathbf{x}_i^m)}{c(\mathbf{x}_i^m)} \right] \Delta r_i^m, \quad (2)$$

where  $\Delta r_i^m$  is the length of the  $i$ th segment. If the relative velocity perturbation  $\Delta c(\mathbf{x}_i^m)/c(\mathbf{x}_i^m)$  at any involved point  $\mathbf{x}_i^m$  is directly treated as an unknown in the inversion, it is probable that a large number of unknowns will be introduced and most of them are located only on one ray path, resulting in an underdetermined system of linear equations. Alternatively,  $1/c_0(\mathbf{x})$  and  $\Delta c(\mathbf{x})/c(\mathbf{x})$  can be approximately expressed in terms of two sets of basis functions  $P_l(\mathbf{x})$  ( $l = 1, 2, \dots, L$ ) and  $Q_n(\mathbf{x})$  ( $n = 1, 2, \dots, N$ ) as

$$\frac{1}{c_0(\mathbf{x})} \approx \sum_{l=1}^L s_l P_l(\mathbf{x}) \quad (3)$$

and

$$\frac{\Delta c(\mathbf{x})}{c(\mathbf{x})} \approx \sum_{n=1}^N u_n Q_n(\mathbf{x}), \quad (4)$$

respectively. Substituting eqs (3) and (4) in eq. (2) gives

$$T_m^{\text{obs}} - T_m^{\text{cal}} = \sum_{n=1}^N \left\{ \sum_{l=1}^L \sum_{i=1}^{I_m} [-P_l(\mathbf{x}_i^m) Q_n(\mathbf{x}_i^m) s_l \Delta r_i^m] \right\} u_n. \quad (5)$$

Given a total of  $M$  observations, we can obtain a linear system  $AX = b$  by defining

$$\begin{aligned} a_{m,n} &= \sum_{l=1}^L \sum_{i=1}^{I_m} [-P_l(\mathbf{x}_i^m) Q_n(\mathbf{x}_i^m) s_l \Delta r_i^m], \\ A &= (a_{m,n})_{M \times N}, \\ X &= (u_1, u_2, \dots, u_N)^T, \\ b &= (T_1^{\text{obs}} - T_1^{\text{cal}}, T_2^{\text{obs}} - T_2^{\text{cal}}, \dots, T_M^{\text{obs}} - T_M^{\text{cal}})^T. \end{aligned} \quad (6)$$

Either because of non-uniqueness or non-existence of  $X$ , the solution of  $AX = b$  is usually obtained by solving a regularized objective function

$$\begin{aligned} \chi(X) &= \frac{1}{2} (AX - b)^T C_d^{-1} (AX - b) + \frac{\lambda^2}{2} X^T C_m^{-1} X \\ &\quad + \frac{\eta^2}{2} X^T D^T D X, \end{aligned} \quad (7)$$

where  $C_d$  and  $C_m$  are the *a priori* data and model covariance matrices,  $D$  is a derivative smoothing operator and  $\lambda$  and  $\eta$  are the damping parameter and smoothing parameter, respectively (e.g. Rawlinson *et al.* 2010; Tong *et al.* 2014a). In a least-squares sense the minimization problem (eq. 7) is equivalent to solving the linear system

$$\begin{pmatrix} C_d^{-1/2} A \\ \lambda C_m^{-1/2} \\ \eta D \end{pmatrix} X = \begin{pmatrix} C_d^{-1/2} b \\ 0 \\ 0 \end{pmatrix}, \quad (8)$$

which can be solved by using the LSQR algorithm or SVD (Paige & Saunders 1982; Rawlinson *et al.* 2010).

There are many options for the basis functions, including polynomials and spherical harmonics (Fichtner *et al.* 2018). In this study we only discuss trilinear interpolation on regularly spaced nodes, where the trilinear basis function  $Q_n(\mathbf{x})$  is defined on a set of  $N$  nodes  $\hat{\mathbf{x}}_n = (\hat{x}_i, \hat{y}_j, \hat{z}_k)$  as

$$Q_n(\mathbf{x}) = q_i^x(x) q_j^y(y) q_k^z(z), \quad (9)$$

where

$$q_i^x(x) = \begin{cases} \frac{x - \hat{x}_{i-1}}{\hat{x}_i - \hat{x}_{i-1}}, & \text{if } \hat{x}_1 \leq \hat{x}_{i-1} \leq x \leq \hat{x}_i; \\ 1 - \frac{x - \hat{x}_i}{\hat{x}_{i+1} - \hat{x}_i}, & \text{if } \hat{x}_i \leq x \leq \hat{x}_{i+1} \leq \hat{x}_{N_x}; \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

$q_j^y(y)$  and  $q_k^z(z)$  are defined in a similar way. We have  $1 \leq i \leq N_x$ ,  $1 \leq j \leq N_y$ ,  $1 \leq k \leq N_z$ ,  $n = i + (j - 1)N_x + (k - 1)N_x N_y$  and  $1 \leq n \leq N = N_x N_y N_z$  in the above expressions (9) and (10). The grid  $G_u$  formed by the  $N$  nodes  $\hat{\mathbf{x}}_n$  is known as the *inversion grid*. Likewise, the velocity  $c_0(\mathbf{x})$  in eq. (3) can be approximated by trilinear interpolation on another grid  $G_0$ .  $G_0$  is usually referred as the *forward grid*. To accurately determine the ray path  $R_m$  and precisely compute  $A = (a_{m,n})_{M \times N}$ ,  $c_0(\mathbf{x})$  is always interpolated on a dense forward grid  $G_0$ . Mainly constrained by the resolving ability of the selected data, the inversion grid  $G_u$  for  $\Delta c(\mathbf{x})/c(\mathbf{x})$  is usually coarser than the forward grid  $G_0$  (Kissling *et al.* 2001). The selection

of both  $G_0$  and  $G_u$  is subjective. The influence of the subjectivity of  $G_0$  on the accuracy of ray tracing and traveltimes computation can be ignored as long as it is dense enough. However, this is not always true for the inversion grid  $G_u$ . A later example in Section 5 will show obvious differences between tomographic results generated on two different inversion grids  $G_u^1$  and  $G_u^2$ . But the two inversion grids have the same horizontal and vertical spacings.

To reduce the influence caused by the subjective choice of a single inversion grid  $G_u$ , we suggest using multiple inversion grids  $G_u^k$  ( $k = 1, 2, \dots, K$ ) in seismic tomographic inversion. Following the ideas of staggered-grid model parametrization and multigrid model parametrization (e.g. Vesnaver & Bohm 2000; Zhou 2003), the  $K$  inversion grids should be spatially staggered and/or have different grid spacings. The relative velocity perturbation  $\Delta c(\mathbf{x})/c(\mathbf{x})$  can be solved on each grid and the solution on the  $k$ th inversion grid is denoted as  $[\Delta c(\mathbf{x})/c(\mathbf{x})]_k$ . Letting the interpolated velocity perturbation on the dense forward grid  $G_0$  be  $[\Delta c(\mathbf{x})]_k$ , we can construct an updated velocity model by using the following relationship:

$$c_0(\mathbf{x}) + [\Delta c(\mathbf{x})]_k = \frac{c_0(\mathbf{x})}{1 - [\Delta c(\mathbf{x})/c(\mathbf{x})]_k}. \quad (11)$$

Considering that each of the  $K$  grids should have the same opportunity to be selected for the tomographic inversion if only one grid is needed, we simply take the average of all the updated velocity models  $c_0(\mathbf{x}) + [\Delta c(\mathbf{x})]_k$  ( $k = 1, 2, \dots, K$ )

$$c_0(\mathbf{x}) + [\Delta \bar{c}(\mathbf{x})]_K = \frac{1}{K} \left( \sum_{k=1}^K \frac{c_0(\mathbf{x})}{1 - [\Delta c(\mathbf{x})/c(\mathbf{x})]_k} \right) \quad (12)$$

as the result of the multiple-grid approach. If we treat the error between the true velocity perturbation  $[\Delta c(\mathbf{x})]_{\text{true}}$  and the velocity perturbation  $[\Delta c(\mathbf{x})]_k$  in eq. (11) or  $[\Delta \bar{c}(\mathbf{x})]_K$  in eq. (12) as a random variable, both  $[\Delta c(\mathbf{x})]_{\text{true}} - [\Delta c(\mathbf{x})]_k$  and  $[\Delta c(\mathbf{x})]_{\text{true}} - [\Delta \bar{c}(\mathbf{x})]_K$  can be assumed to have the same expected value 0, but the variance of  $[\Delta c(\mathbf{x})]_{\text{true}} - [\Delta c(\mathbf{x})]_k$  is  $K$  times as large as the variance of  $[\Delta c(\mathbf{x})]_{\text{true}} - [\Delta \bar{c}(\mathbf{x})]_K$ . This suggests that  $[\Delta \bar{c}(\mathbf{x})]_K$  is a low-risk choice that is unlikely to be far away from the ‘true’ solution.

One can choose to smooth the updated velocity model  $c_0(\mathbf{x}) + [\Delta \bar{c}(\mathbf{x})]_K$  by convolving it with a 3-D Gaussian function

$$G(\mathbf{x}) = G(x, y, z) = \frac{8}{\sqrt{\pi^3} \sigma_x \sigma_y \sigma_z} e^{-\left( \frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} + \frac{z^2}{2\sigma_z^2} \right)}. \quad (13)$$

Here  $\sigma_x$ ,  $\sigma_y$  and  $\sigma_z$  are the smoothing radiuses along the  $x$ ,  $y$  and  $z$  directions, respectively. In this study the smoothing parameter  $\eta$  in eq. (8) is set to be zero but Gaussian smoothing (eq. 13) is applied by selecting an optimal smoothing radius. Because of the nonlinear dependence of the ray path  $L$  on the velocity model  $c_0(\mathbf{x})$ , the velocity model needs to be iteratively improved. The update from  $c_0(\mathbf{x})$  to  $c_0(\mathbf{x}) + [\Delta \bar{c}(\mathbf{x})]_K$  in eq. (12) is only for one iteration. The updated velocity model  $c_0(\mathbf{x}) + [\Delta \bar{c}(\mathbf{x})]_K$  is set as the new starting model  $c_0(\mathbf{x})$  for the next iteration. Throughout iterations, a series of velocity models on the dense forward grid  $G_0$  are generated. Note that the accuracy of the initial model  $c_0(\mathbf{x})$  affects the convergence of inversion and a careful selection of  $c_0(\mathbf{x})$  is necessary before conducting seismic tomographic inversion.

### 3 DATA

In this section we prepare an accurate initial model  $c_0(\mathbf{x})$  and high-quality traveltimes data for the goal of building a detailed  $P$ -wave

velocity model along the San Jacinto fault zone with the multiple-grid model parametrization.

### 3.1 1-D inversion

From the Southern California Earthquake Data Center (SCEDC 2013), we downloaded 97 076 *P*-wave traveltimes of 6613 earthquakes with magnitudes greater than or equal to 2.0 occurred between 1992 January and 2017 October in the study area (Fig. 1). These traveltimes were manually picked from 182 seismic stations deployed in the same region (Fig. 2a). Beginning with an arbitrarily assumed 1-D layered model  $m_{1D}^0$  (Table 1), we iteratively refine it by inverting all the *P*-wave arrivals with epicentral distances less than 90 km. At each iteration, the relative velocity perturbation of each layer is limited within  $\pm 1.0$  per cent by choosing an optimal damping parameter. Since the ray paths of first arrivals in 1-D layered models can be easily traced using Snell's law, this 1-D inversion is very efficient. A stable solution is obtained after about 40 iterations (Table 1). The root mean square value of the traveltime residuals is reduced from 0.867 s in  $m_{1D}^0$  to 0.212 s in  $m_{1D}^{40}$ .

The study area defined by the blue rectangle in Fig. 1 has a length of 270 km and a width of 50 km. In such an area the Moho geometry should be honoured to accurately trace the first arrivals that refract through the mantle. So, the 1-D layered model  $m_{1D}^{40}$  and the undulated Moho discontinuity (Zhu & Kanamori 2000) are used together to build a 3-D *P*-wave velocity model for the study area. The velocity below the Moho is set as  $8.0 \text{ km s}^{-1}$ . The 3-D velocity model is further smoothed across each discontinuity  $z_{\text{dis}}$  based on the following interpolation function when  $z$  is in between the middle of the upper layer and the middle of the lower layer

$$v(z) = \begin{cases} v_1 + \frac{v_2 - v_1}{1.0 + \exp(z_{\text{dis}} - z)}, & \text{if } z < z_{\text{dis}} \\ v_2 + \frac{v_1 - v_2}{1.0 + \exp(z - z_{\text{dis}})}, & \text{if } z \geq z_{\text{dis}} \end{cases} \quad (14)$$

Here  $z$  denotes the depth;  $v_1$  and  $v_2$  are the velocities above and below the discontinuity at the depth of  $z_{\text{dis}}$ . The smoothed 3-D velocity model is used as the initial model  $c_0(\mathbf{x})$  for seismic tomographic inversions.

### 3.2 Data selection

Seismicity in the study area distributes mainly along active faults with a fairly uniform coverage down to the depth of about 15 km (Fig. 2a). To make sure that only high-quality traveltime data are used, we carefully select a subset of the previously downloaded *P*-wave traveltimes for tomographic inversions. The selection criteria are as follows:

- (1) Every selected earthquake in the study area ( $[0, 270\text{km}] \times [0, 50\text{km}]$ ) from the surface to the 30.0 km depth, Fig. 2a) has at least eight *P*-wave picks in the same area.
- (2) To avoid event clustering and reduce data redundancy, the study area is divided into  $1.0 \text{ km} \times 1.0 \text{ km} \times 0.5 \text{ km}$  blocks and from each block only one earthquake with the greatest number of picks is chosen if there is any.
- (3) Only earthquakes occurred deeper than 2.0 km are selected.

A total of 82 105 *P*-wave arrivals from 4010 earthquakes recorded at 182 seismic stations is selected (Fig. 2b). It can be observed that the distribution of the 4010 earthquakes is very similar to that of the 6613 earthquakes (Figs 2a and b), indicating that there is no loss in spatial coverage after data selection.

### 3.3 Earthquake relocation

We relocate all the 4010 earthquakes in the smoothed 3-D velocity model  $c_0(\mathbf{x})$  before conducting tomographic inversion for the velocity structure.  $c_0(\mathbf{x})$  is sampled on a forward grid  $G_0$  with a uniform grid interval of  $\Delta x = \Delta y = \Delta z = 0.4 \text{ km}$ . The ray paths of first arrivals in  $c_0(\mathbf{x})$  are determined by solving eikonal equations with a highly accurate multistencils fast marching method (Hassouna & Farag 2007; Tong *et al.* 2017). Our earthquake relocation algorithm is similar to Geiger's method (Geiger 1910; Tong *et al.* 2016) and is based on the following equation,

$$\begin{aligned} \Delta t &= \frac{\partial t}{\partial \mathbf{x}_s} \cdot \Delta \mathbf{x}_s + \Delta t_0 = -\frac{1}{V_{\mathbf{x}_s}} \frac{ds}{d\mathbf{x}_s} \cdot \Delta \mathbf{x}_s + \Delta t_0 \\ &= -\frac{D_x \cos \alpha}{V_{\mathbf{x}_s}} \cdot \frac{\Delta x_s}{D_x} - \frac{D_y \cos \beta}{V_{\mathbf{x}_s}} \cdot \frac{\Delta y_s}{D_y} \\ &\quad - \frac{D_z \cos \gamma}{V_{\mathbf{x}_s}} \cdot \frac{\Delta z_s}{D_z} + D_t \cdot \frac{\Delta t_0}{D_t}, \end{aligned} \quad (15)$$

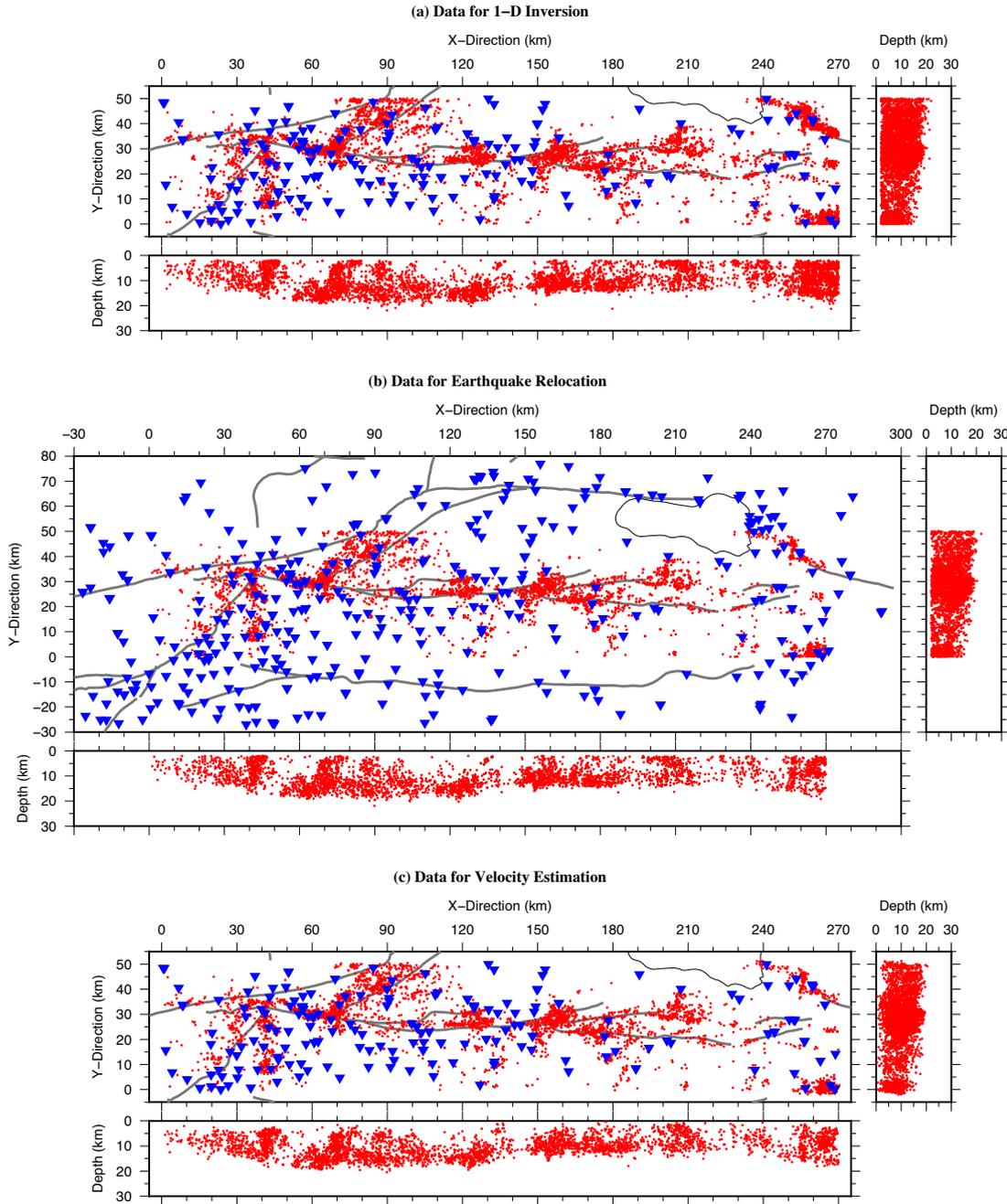
where  $(\cos \alpha, \cos \beta, \cos \gamma)$  is the direction of the ray path at the estimated earthquake location  $\mathbf{x}_s$ ,  $V_{\mathbf{x}_s}$  is the wave speed at  $\mathbf{x}_s$  and  $D_x, D_y, D_z$  and  $D_t$  are the non-dimensionalization scaling parameters for the location and the origin time. Different from Geiger's method, the unknowns in our approach are  $\Delta x_s/D_x, \Delta y_s/D_y, \Delta z_s/D_z$  and  $\Delta t_0/D_t$ . The selection of these scaling parameters is subjective. We choose  $D_x = D_y = D_z = 20 \text{ km}$  and  $D_t = D_x/V_{\mathbf{x}_s}$  in this study.

Some earthquakes are close to the boundaries of the study area, and their selected *P*-wave arrivals have limited azimuth coverage. If only these data are used, we may not be able to get accurate locations of those near-boundary earthquakes. To reduce the influence caused by incomplete data coverage, we relocate the selected earthquakes in an extended area. The earthquakes are distributed in the horizontal domain  $[0, 270 \text{ km}] \times [0, 50 \text{ km}]$ . A total of 131 308 *P*-wave arrivals picked at 395 seismic stations in an extended region  $[-30, 300 \text{ km}] \times [-30, 80 \text{ km}]$  is inverted to relocate all the 4010 earthquakes (Fig. 2b). Every earthquake is surrounded by stations covering almost all azimuth (Fig. 2b) necessary for robust earthquake relocation. After five iterations, the root mean square value of the traveltime residuals is reduced from 0.2458 to 0.2071 s, and the standard deviation is changed from 0.3474 to 0.2922 s. The relocated earthquakes are shown in Fig. 2(c). For those 82 105 *P*-wave arrivals selected for tomographic inversion, with the updated earthquake locations, the root mean square value and standard deviation of the traveltime residuals in  $c_0(\mathbf{x})$  are 0.1905 and 0.2820 s, respectively.

## 4 MULTIPLE-GRID MODEL PARAMETRIZATION AND RESOLUTION TESTS

### 4.1 Single-scale reconstruction test

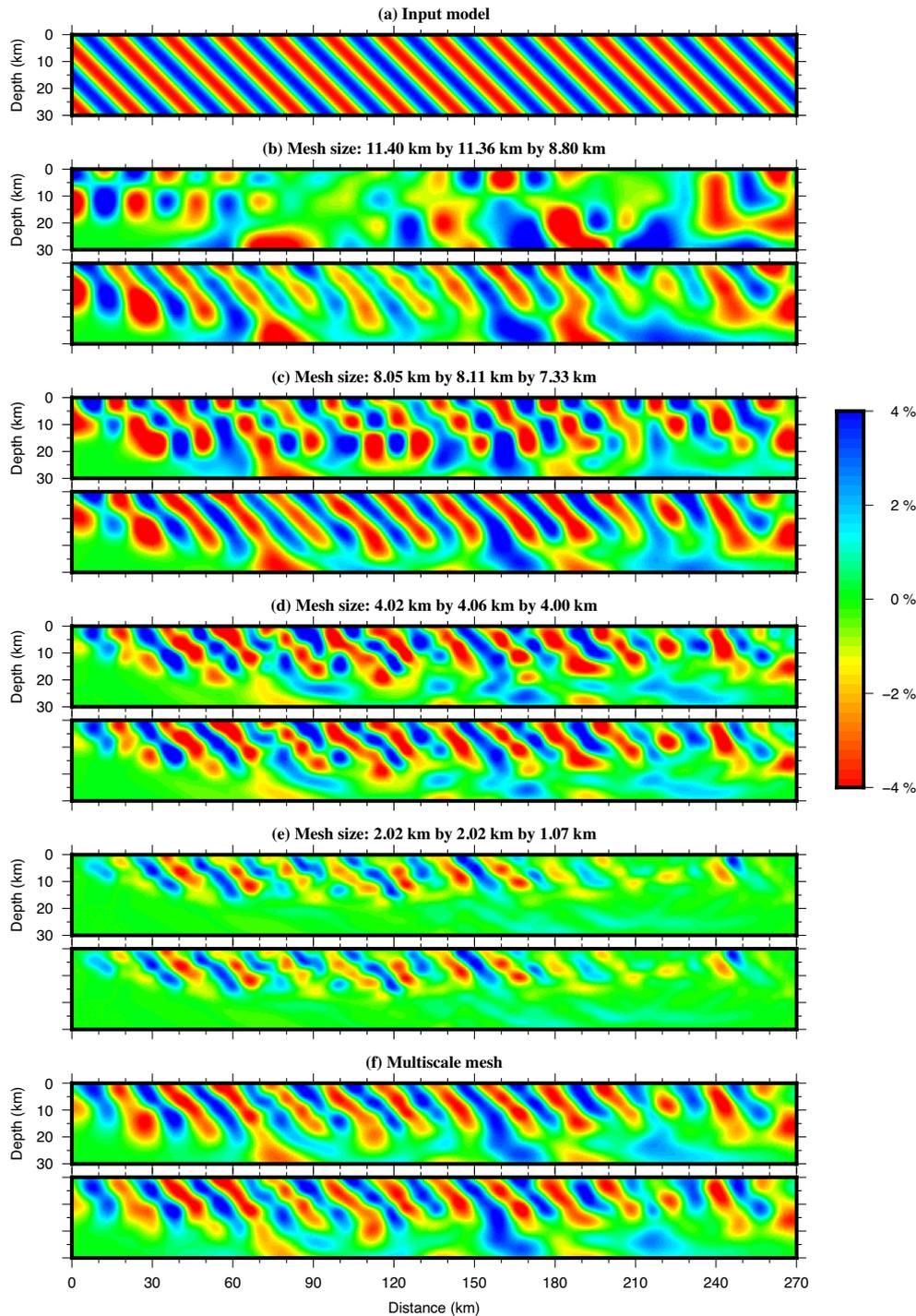
We perform a series of synthetic reconstruction tests to show the efficacy of the multiple-grid model parametrization in seismic tomographic inversions. The arrival times of the 82 105 *P* waves are computed in a velocity model that has a relative perturbation of 4 per cent  $\times \sin(0.1\pi x - 0.1\pi z)$  from the initial model  $c_0(\mathbf{x})$  (Fig. 3a). The velocity perturbation is invariant in the  $y$  direction and has a wavelength of 20 km along the  $x$  and  $z$  directions. Random errors with the mean value of 0.0 and standard deviation of 0.2 s are added to the synthetic data to mimic the measurement errors in real data. Four collocated grids, four staggered-grids, one multigrid



**Figure 2.** The distributions of earthquakes (red dots) and seismic stations (blue inverted triangles). (a) Shows all the 6613 earthquakes with magnitudes greater than or equal to 2.0 downloaded from SCEDC and the 182 seismic stations in the study area. The *P*-wave arrivals of these earthquakes are inverted to generate a 1-D layered model of this region (Table 1). (b) The data of the shown 4010 earthquakes are selected for tomographic inversions. But before that, the selected earthquakes are relocated by using the *P*-wave traveltimes picked from 395 seismic stations in an extended area. (c) Shows the distribution of the 4010 earthquakes after relocation.

**Table 1.** The 1-D layered *P*-wave velocity models. The model  $m_{1D}^{40}$  is obtained by inverting all the 97076 *P*-wave arrivals downloaded from SCEDC.  $m_{1D}^{40}$  is then used to construct the initial model  $c_0(x)$  for earthquake relocation and tomographic inversion.

Depth (km)	<4	[4, 8.5)	[8.5, 13.5)	[13.5, 19)	≥19
$m_{1D}^0$	4.800	5.300	5.800	6.300	6.800
$m_{1D}^{40}$	5.519	6.161	6.377	6.536	6.767

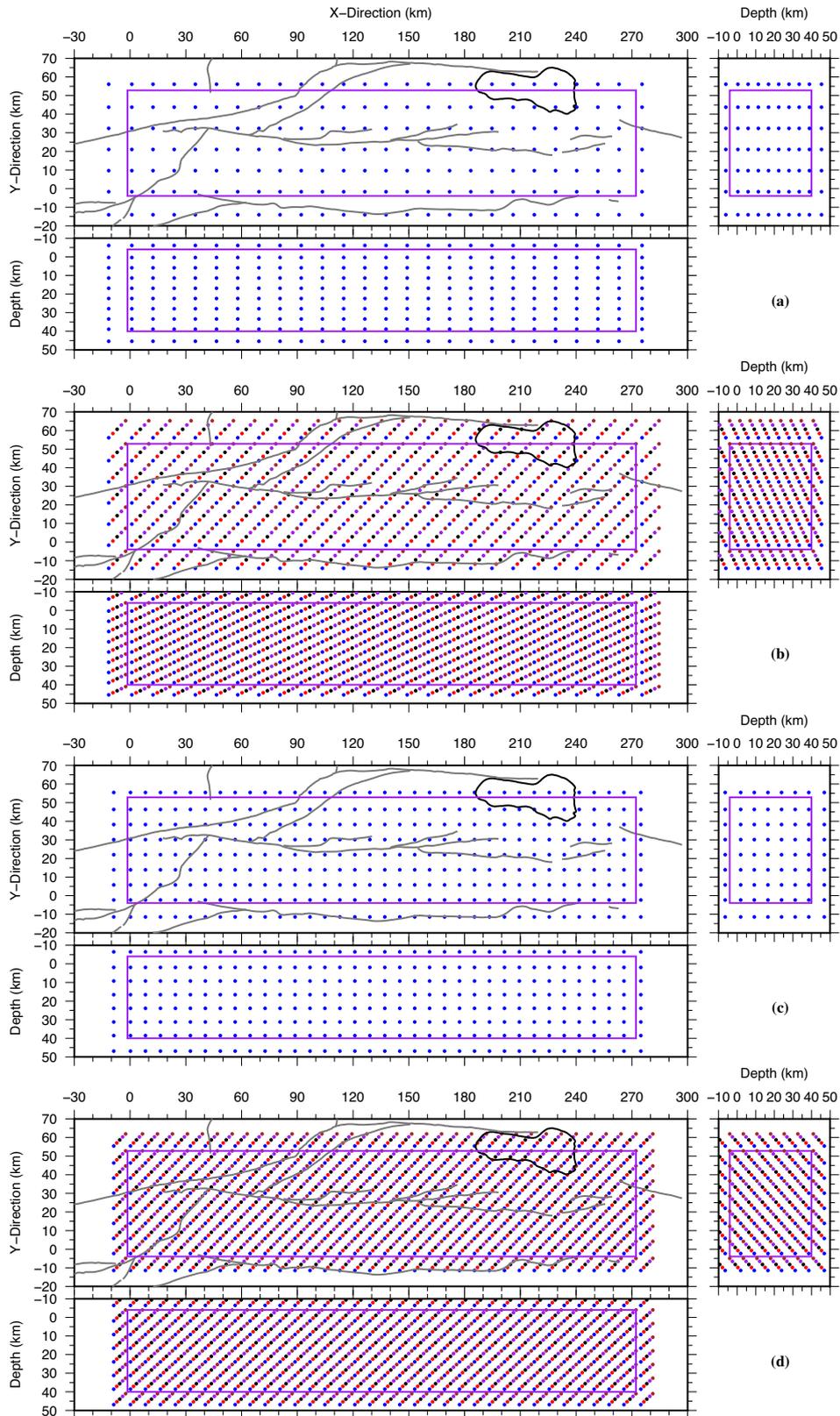


**Figure 3.** (a) The ‘true’ relative velocity perturbation. (b–e) Each shows the vertical views of the relative velocity perturbation field at  $y = 25$  km recovered with a collocated grid (top panel) and a staggered-grid (bottom panel). The spacings of the collocated grid in the  $x$ ,  $y$  and depth directions are labelled above each subfigure. The related staggered-grid consists of five collocated grids with the same spatial spacings. The results shown in (b) are constructed on the grids presented in Figs 4(a) and (b), respectively. Figs 4(c) and (d) show the grids used to generate the images demonstrated in (c). The top panel of (f) is generated on a multigrid formed by the four collocated grids used in (b)–(e). The bottom panel of (f) is constructed on a multiscale-staggered-grid with eight collocated grids at four different scales.

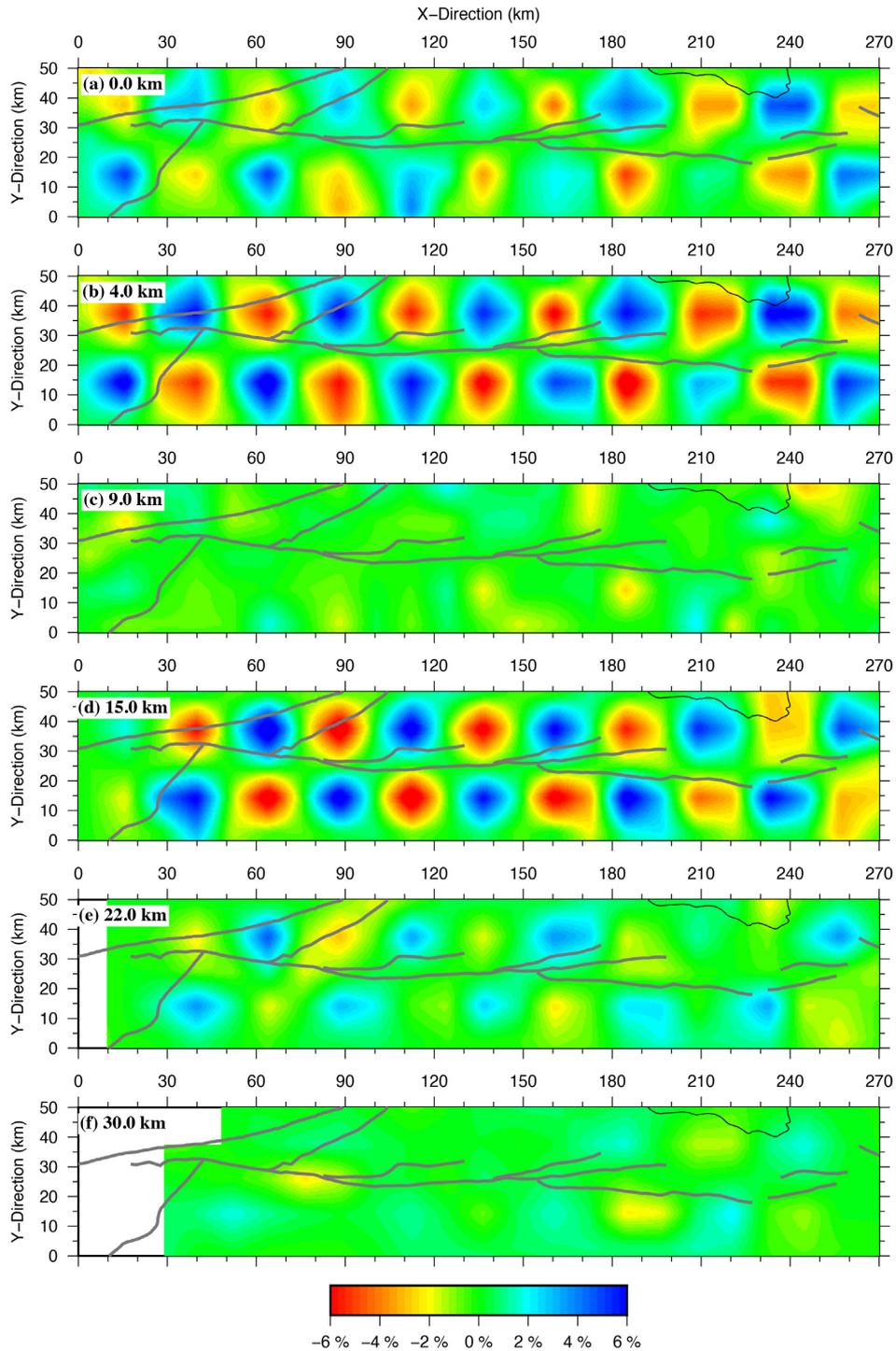
and one multiscale-staggered-grid are used separately to retrieve the single-scale perturbation field.

First, the spacings of the collocated grid are about 11.4 km horizontally and 8.8 km vertically (Fig. 4a), roughly making that one wavelength distance of the sinusoidal perturbation field is sampled

by three grid nodes. Based on Nyquist–Shannon sampling theorem, a satisfactory recovery would be achieved if all the grid nodes lie exactly on the peaks and troughs of the perturbation field. But this rarely happens in real practices. As shown in the top panel of Fig. 3(b), the collocated grid cannot recover the perturbation field



**Figure 4.** (a) A collocated grid with spatial spacings of 11.40, 11.36 and 8.8 km in the  $x$ -,  $y$ - and vertical directions, respectively. (b) A staggered-grid consisting of five collocated grids with the same spatial spacings as the one in (a). Different grids are marked with different colours. (c) A finer collocated grid with spatial spacings of 8.05, 8.11 and 7.33 km. (d) A finer staggered-grid formed by five finer collocated grids.



**Figure 5.** The results of the checkerboard resolution test at different depths. The targeted relative velocity perturbation is defined in eq. (16) with  $a = b = 0.02$ . The wavelength of the anomaly along the  $x$  or  $y$  axis is 50 km. The input velocity perturbations at 0, 9 and 22 km depths are zero.

because of inadequate sampling. A staggered-grid (Fig. 4b) consisting of five collocated grids with the same spacings (11.4 km horizontally and 8.8 km vertically) is then used to recover the perturbation field. Due to the increased spatial sampling, a much better result is obtained (see the bottom panel of Fig. 3b). We reduce the horizontal spacing of the collocated grid to about 8.0 km and its vertical spacing to 7.33 km (Fig. 4c). A new staggered-grid (Fig. 4d) is also formed by superimposing five collocated grids of the same reduced size. As demonstrated in Fig. 3(c), the collocated grid can

roughly reveal the fast and slow anomalies, while the staggered-grid is able to almost fully recover these anomalies. We reduce the grid spacings of the collocated grid and the staggered-grid for another two times (Figs 3d and e). Both the collocated grid and staggered-grid can adequately sample the designed velocity anomaly. The performances of the collocated grid and staggered-grid seem to be the same (Figs 3d and e). More unknowns need to be determined if a finer grid is used in the tomographic inversion. As a result, the data may not have enough constraints on the increased unknowns.

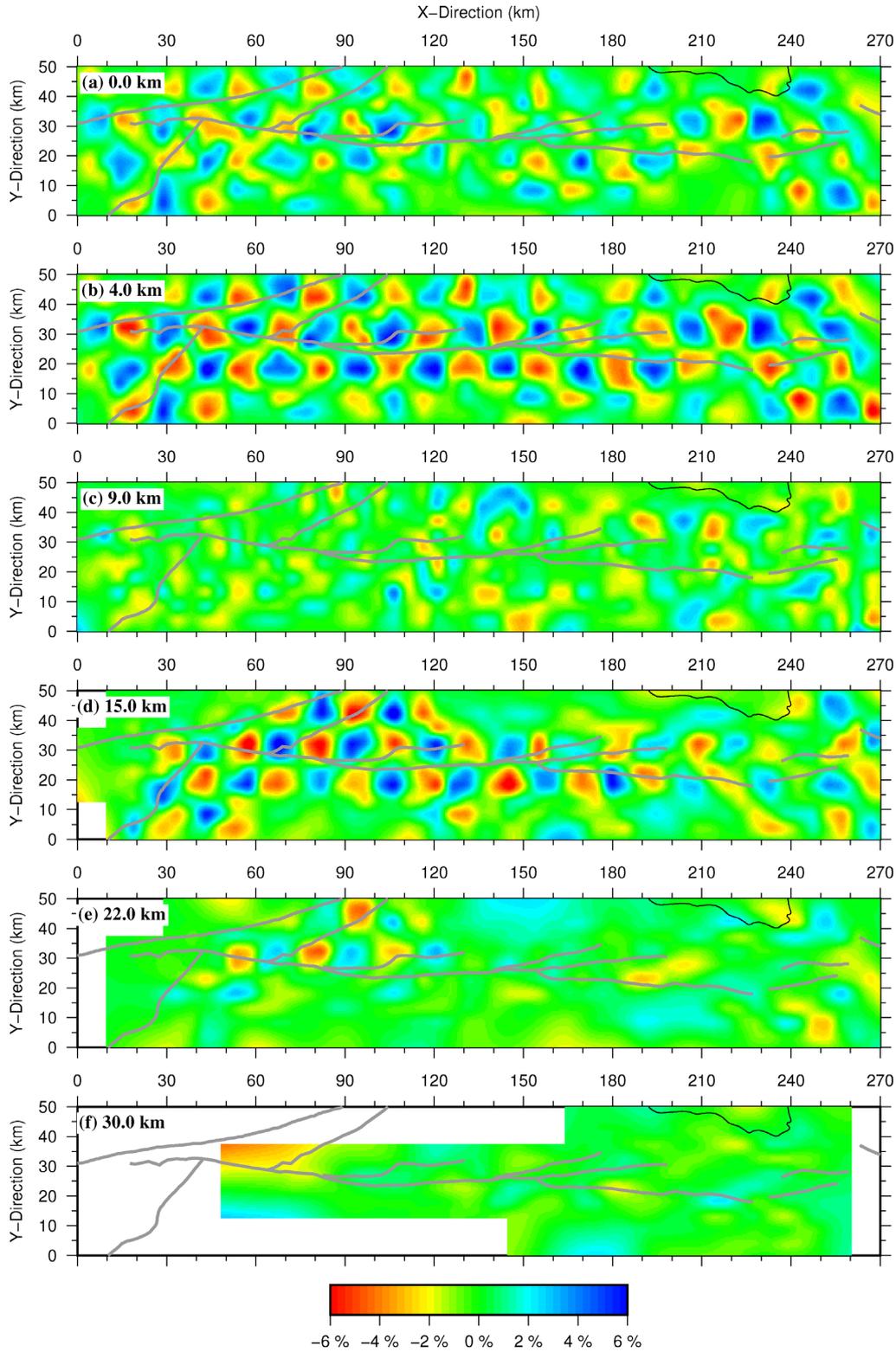


Figure 6. The same as Fig. 5 but  $a = b = 0.04$ . The wavelength of the anomaly along the  $x$  or  $y$  axis is 25 km.

To get optimal results, we carefully selected the regularization parameters (damping parameter  $\lambda$  in eq. 7 and smoothing radius  $\sigma = \sigma_x = \sigma_y = \sigma_z$  in eq. 13) by using the L-curve method (Tong *et al.* 2014b). The details are presented in the Supporting Information. However, the sinusoidal anomaly can only be partially recovered on the finest grids (Fig. 3e). This suggests that optimal grid spacings

should be determined in advance to generate reliable and stable results.

Collocated grids at four different scales have been used in the single-scale reconstruction test. We continue to select four collocated grids at four different scales and superimpose them together to form a multigrid. A multiscale-staggered-grid is also constructed

by combining four staggered-grids at four different scales. Here every staggered-grid consists of two spatially shifted grids of the same size (see Figs 7 and 8 for examples). The similar results generated on the multigrid and the multiscale-staggered-grid show that the velocity anomaly has been largely recovered (Fig. 3f). In comparison with the multigrid approach, the multiscale-staggered-grid can reduce the root mean square value of the traveltimes residuals from 0.20783 s to a slightly smaller value (0.19843 versus 0.19835 s).

In these single-scale reconstruction tests, the best recoveries are achieved on the staggered-grids when the grid spacing of the component grid is about 1/3 to 1/2 of the wavelength of the anomaly (Figs 3b and c). Meanwhile, both the multigrid and the multiscale-staggered-grid approaches can generate reasonably good results and enjoy the freedom in grid spacing selection. These examples suggest that the multiple-grid including the staggered-grid, multigrid and multiscale-staggered-grid is more reliable than or at least as reliable as the collocated grid in recovering subsurface anomalies.

It is common that the velocity perturbation  $\Delta c(\mathbf{x}) = c_{\text{true}}(\mathbf{x}) - c_0(\mathbf{x})$  of the true model from the starting model has multiscale features and some of them with short wavelengths cannot be resolved by the selected data. We have also conducted multiscale reconstruction test to show the advantages of the multiple-grid model parametrization in tomographic inversions. Please see the Supporting Information for the details.

#### 4.2 Checkerboard resolution test

We perform checkerboard resolution tests to assess data resolution of the selected  $P$ -wave arrivals. The checkerboard model  $c_{\text{ckbd}}(\mathbf{x})$  is constructed by perturbing the initial model  $c_0(\mathbf{x})$  with an amount of  $\Delta c(\mathbf{x})$ . The mathematical expression of the checkerboard model used in this study is

$$\begin{aligned} c_{\text{ckbd}}(\mathbf{x}) &= c_0(\mathbf{x}) + \Delta c(\mathbf{x}) \\ &= c_0(\mathbf{x}) \left[ 1 + 0.06 \times \sin(2\pi ax) \sin(2\pi by) \right. \\ &\quad \left. \times \sin \left( 2\pi \frac{\sqrt{49 + 8z - 7}}{8} \right) \right], \end{aligned} \quad (16)$$

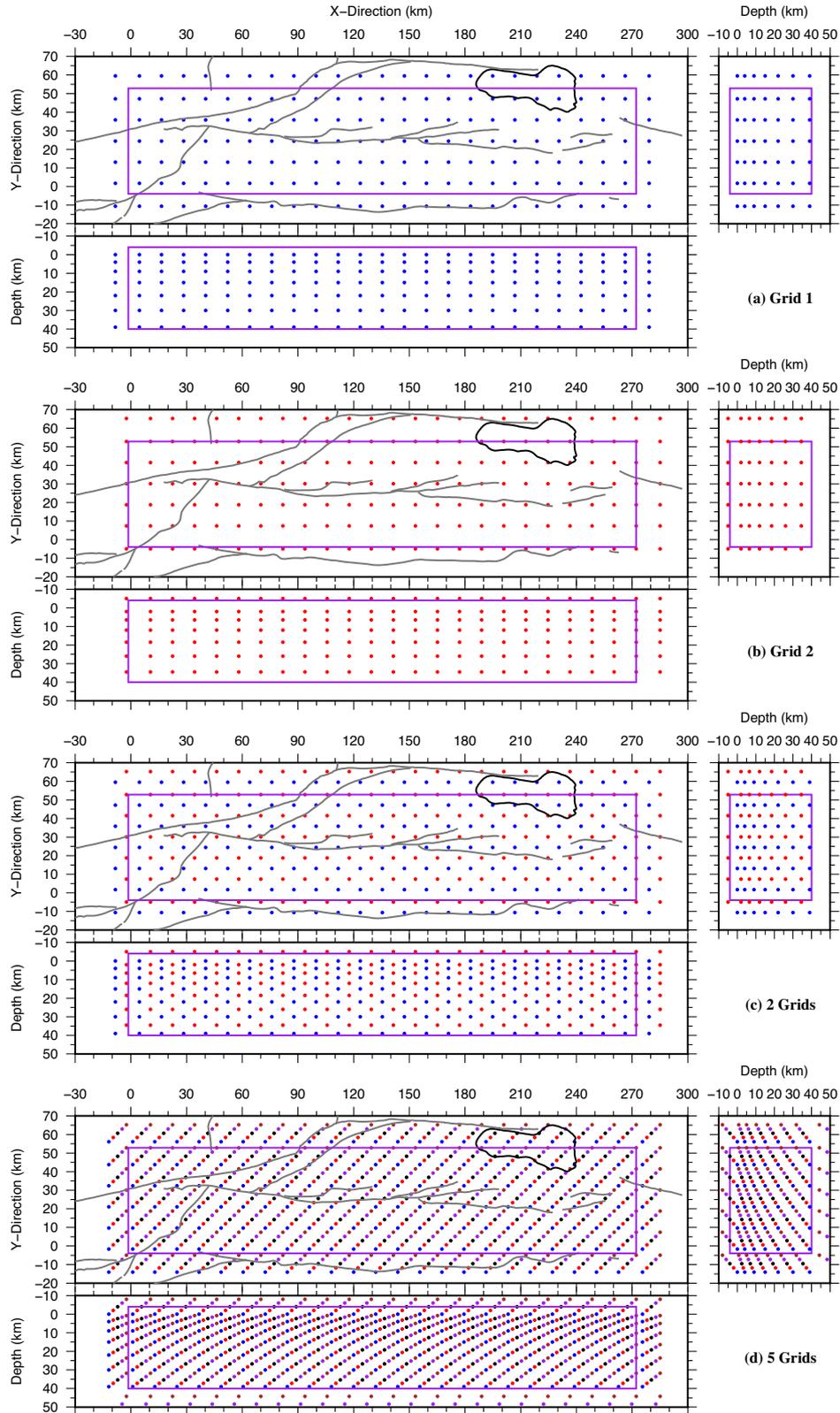
where  $a$  and  $b$  are the wavenumbers of the sinusoidal anomaly along the  $x$  and  $y$  directions, respectively. We choose  $a = b = 0.02$ , indicating that the wavelength of the anomaly along the  $x$  axis (as well as the  $y$  axis) is 50 km. At depths of 0, 4, 9, 15, 22, 30 and 39 km, the third sine function in eq. (16) takes the values of 0, 1, 0,  $-1$ , 0, 1 and 0, respectively. The wavelength of the anomaly in the vertical direction gradually increases from 14 km at 0 km depth to 38 km at 39 km depth. The arrival times of the 82 105  $P$  waves in the checkerboard model  $c_{\text{ckbd}}(\mathbf{x})$  are evaluated by solving eikonal equations. Once again, random errors with the mean value of 0.0 and standard deviation of 0.2 s are added to the synthetic data. A collocated grid with horizontal spacings of 12.05 km along the  $x$  axis and 11.52 km in the  $y$  direction is used to parametrize the relative perturbation field  $\Delta c(\mathbf{x})/c(\mathbf{x})$ . The grid nodes are placed at depths of 0, 4, 9, 15, 22, 30 and 39 km in the vertical direction. In this manner, every wavelength distance of the anomaly is sampled by at least five grid nodes. No undersampling issue exists. We attempt to iteratively recover the checkerboard pattern from the initial model  $c_0(\mathbf{x})$ . In each iteration, we restrict the relative velocity perturbation  $\Delta c(\mathbf{x})/c_{\text{ckbd}}(\mathbf{x})$  within  $\pm 1$  per cent by choosing an optimal damping parameter, and after that the Gaussian smoothing (eq. 13 with a radius  $\sigma = \sigma_x = \sigma_y = \sigma_z = 0.8$  km) is

applied to the inverted  $\Delta c(\mathbf{x})/c_{\text{ckbd}}(\mathbf{x})$ . The final results at different depths after 12 iterations are shown in Fig. 5. The fast, slow and zero velocity perturbations have been largely recovered from the surface to 22 km depth. Leakages of velocity anomaly up to the surface and down to a deeper depth of 22 km are observable at 4 km depth and 15 km depth, respectively. This example suggests that the selected data set can well resolve seismic velocity structure with a horizontal wavelength of 50 km or above from the near surface to about 22 km depth. In the vertical direction, it is able to recover seismic heterogeneities with wavelengths greater than 14 km.

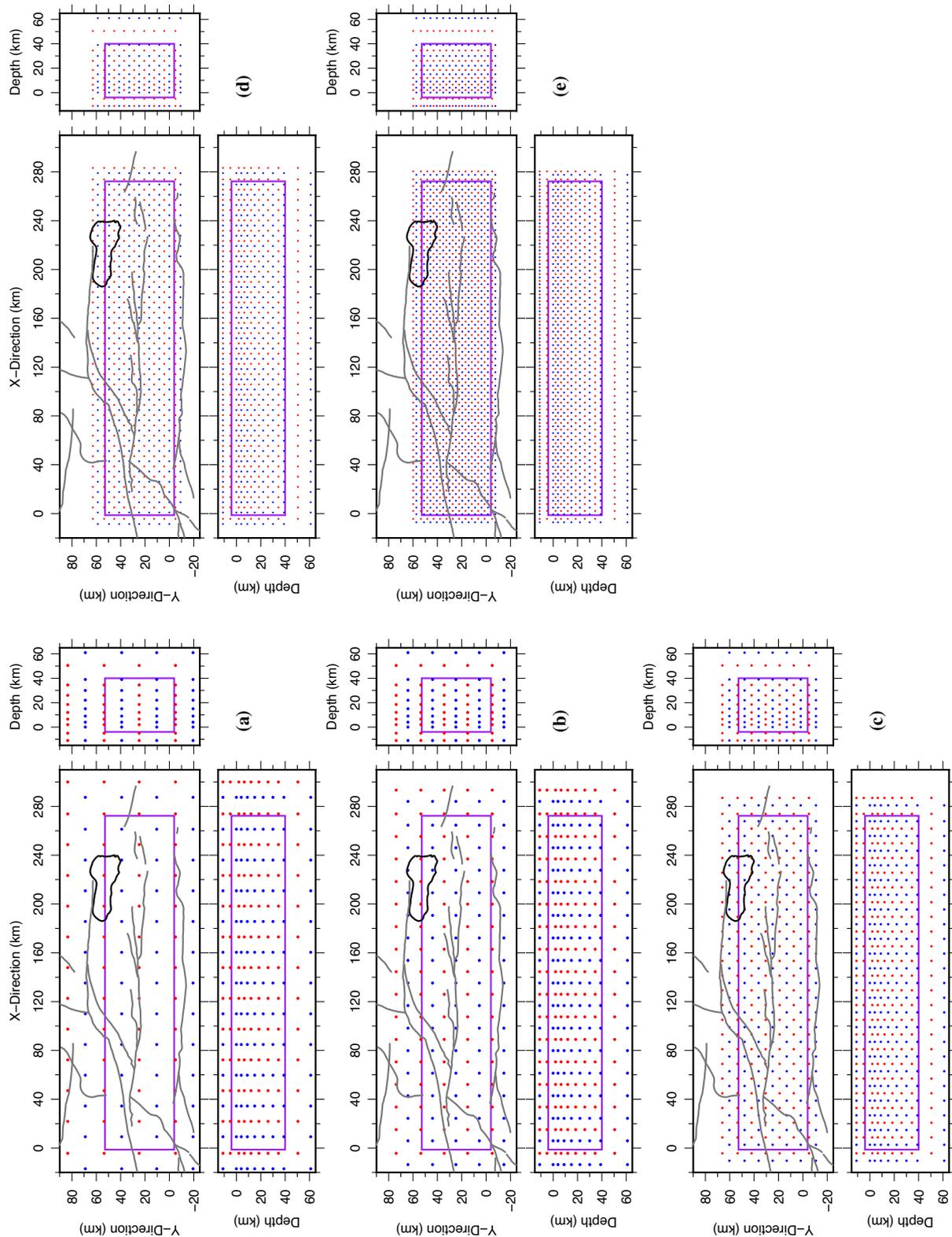
To check if the data can resolve small-scale anomalies, we further choose  $a = b = 0.04$ . The wavelength of the anomaly along the horizontal axis is reduced to 25 km. Synthetic data with random errors are inverted on a finer collocated grid with spacings of 5.13 km in the  $x$  direction and 5.24 km in the  $y$  direction. Since the wavelength of the perturbation field remains the same in the vertical direction, the grid nodes are placed at the same depths as the previous case of  $a = b = 0.02$ . Along the San Jacinto fault the checkerboard pattern has been recovered from 4 km depth to 15 km depth (Fig. 6). Artefacts are observable at depths of 0, 9 and 22 km where the input velocity perturbation is zero (Fig. 6). Generally speaking, the selected data set is able to resolve heterogeneous structures with horizontal wavelengths greater than 25 km along the San Jacinto fault. We have also investigated the case of  $a = b = 0.08$ . An even finer collocated grid with an average spacing of 2.6 km is used to recover the anomaly with a wavelength of 12.5 km along the horizontal axes. The inverted results show that artefacts are persistent from the surface to the lower crust, and the checkerboard pattern is not identifiable (Fig. S6, Supporting Information). Based on all the checkerboard resolution tests, we can roughly say that the chosen  $P$ -wave data can resolve structures with horizontal wavelengths greater than 25 km and vertical wavelengths greater than 14 km along the San Jacinto fault. This information is useful for choosing suitable inversion grid in tomographic inversions with the real data.

## 5 TOMOGRAPHIC RESULTS

In this section we show the tomographic results of the San Jacinto fault zone generated by using five different model parametrizations: Collocated Grid 1, Collocated Grid 2, Staggered-grid A, Staggered-grid B and a multiscale-staggered-grid (Figs 7 and 8). For all the grids, the grid spacing selection is based on the checkerboard resolution test discussed in the previous section. Fig. 7(a) shows a regular or collocated grid with spatial spacings of 11.9 km in the  $x$  direction and 11.4 km along the  $y$  direction. Vertically, the grid nodes are located at  $-11, 0, 4$  and  $9, 15, 22, 30, 39$  and 61 km depth. We call this grid as Grid 1. With an equal opportunity to be selected for the tomographic inversion, another grid with the same lateral spacings is shown in Fig. 7(b). The second grid is named as Grid 2. Its nodes are placed at  $-5, 2, 6.5, 12, 18.5, 26, 34.5$  and 50.5 km depth. Grids 1 and 2 are spatially staggered. We proceed to construct a staggered-grid by superimposing Grid 2 onto Grid 1 and call it as Staggered-grid A (Fig. 7c). Different from the staggered-grid finite-difference method for solving differential equations, a staggered-grid used in tomographic inversion can have more than two sets of collocated grids. So, we construct another staggered-grid formed by five sets of collocated grids to show the performance of the staggered-grid model parametrization with different configurations (Fig. 7d). The spatial spacings of the five grids are almost the same as those of Grids 1 and 2. This second staggered-grid is



**Figure 7.** (a) The collocated Grid 1. (b) The collocated Grid 2. Grid 2 has a spatial shift from Grid 1. (c) Staggered-grid A formed by combining Grids 1 and 2. (d) Staggered-grid B consisting of five collocated grids with the same spatial spacings as Grids 1 and 2.

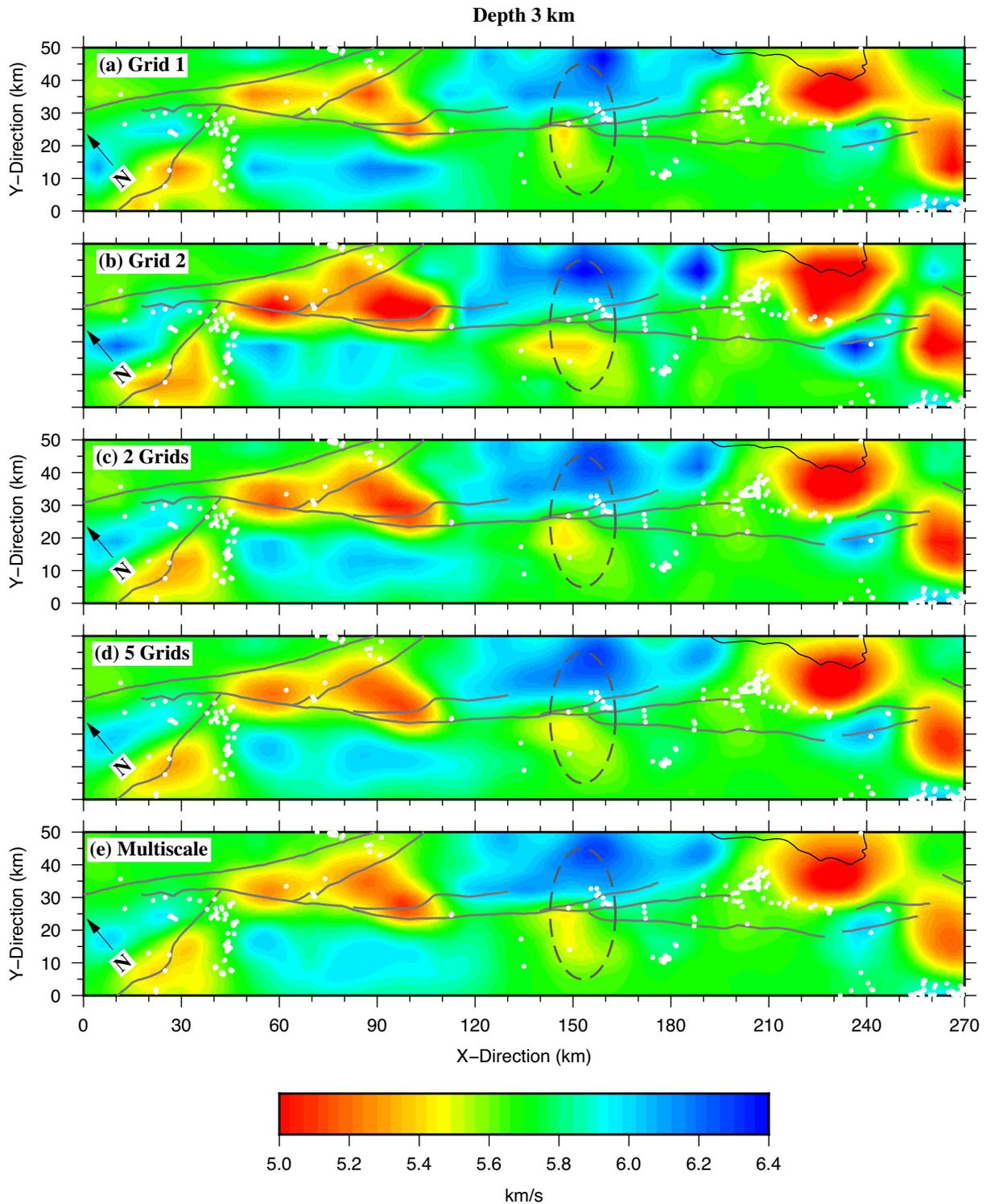


**Figure 8.** Staggered-grids at five different scales. Each staggered-grid consists of two collocated grids marked by red and blue colours. The average grid intervals in the horizontal plane are (a) 26.5 km, (b) 18.6 km, (c) 11.6 km, (d) 8.2 km and (e) 5.3 km, respectively.

called Staggered-grid B. A multiscale-staggered-grid that composes of five staggered-grids or 10 collocated grids (Fig. 8) is also used in the tomographic inversions. The lateral grid spacings of the 10 collocated grids vary among 26.5, 18.6, 11.6, 8.2 and 5.3 km.

### 5.1 Plan views

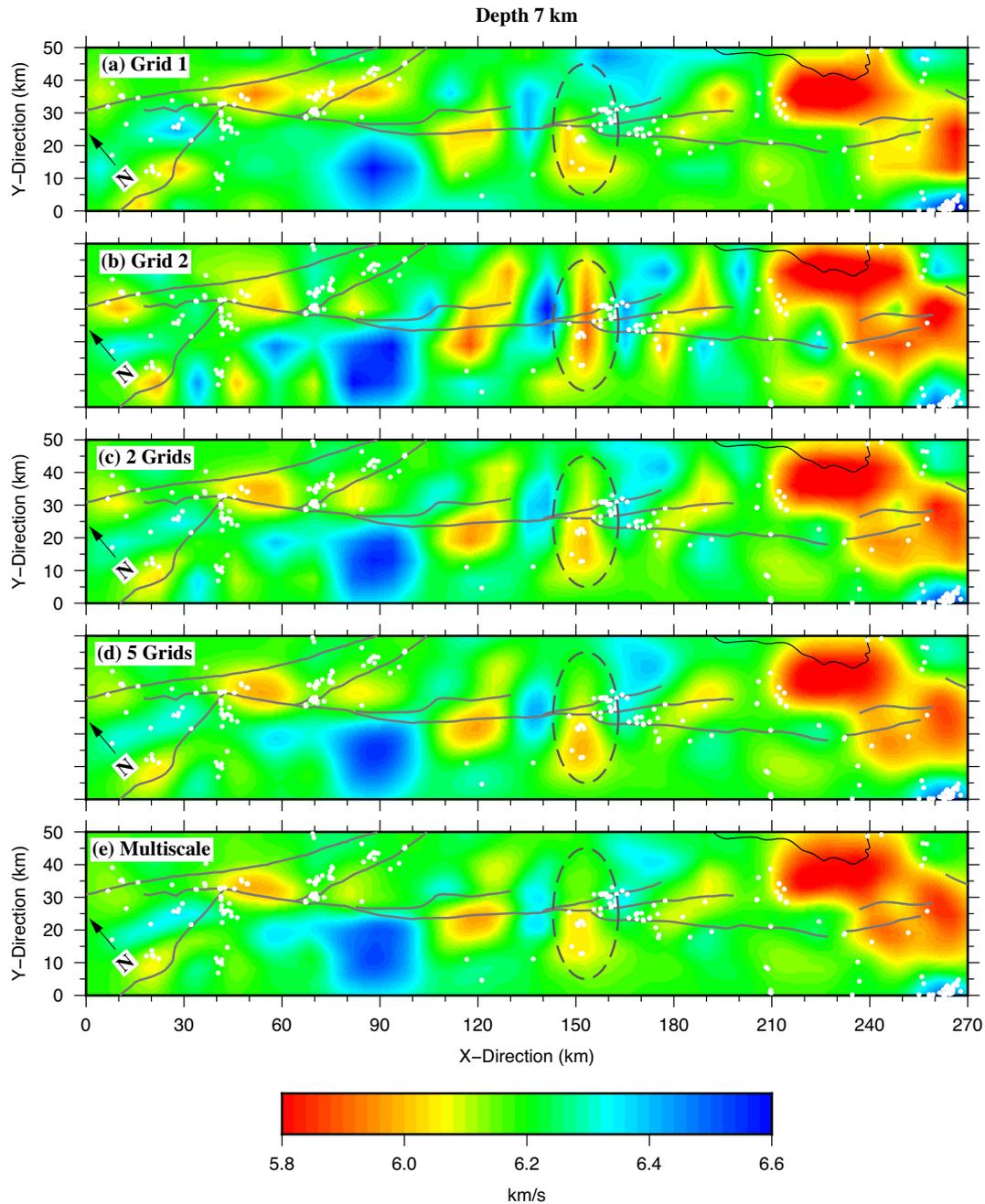
Figs 9–12 show plan views of the recovered *P*-wave velocity structure at four representative depths (3, 7, 11 and 15 km) along the San Jacinto fault. More plan views at other depths (5, 9, 13 and 17 km)



**Figure 9.** Plan views of the inverted velocity models with five different model parametrizations at the depth of 3 km. *N* points to the north direction. The red colour indicates low velocity and blue colour shows high-velocity structures.

are shown in Supporting Information Figs S7–S10. Generally speaking, the distributions and sizes of high- and low-velocity anomalies revealed with the five different model parametrizations are similar. The velocity perturbations at shallow depths are up to and beyond  $\pm 10$  per cent, which indicate strong lateral heterogeneities near the surface (Fig. 9). The velocity variations are relatively smaller at deeper depths with a few percent of perturbation (Figs 11 and 12).

Specifically, three dominant features present in the *P*-wave velocity model. (1) Strong velocity contrasts and a reversal in the velocity contrast polarity across the San Jacinto fault are observable. In details, the east side of the northern section of the fault (from about 40 to 110 km along the *x* axis) has lower velocities than the west side from the surface to the lower crust. However, the east side of the central section of the fault (in between 110 and 200 km) has

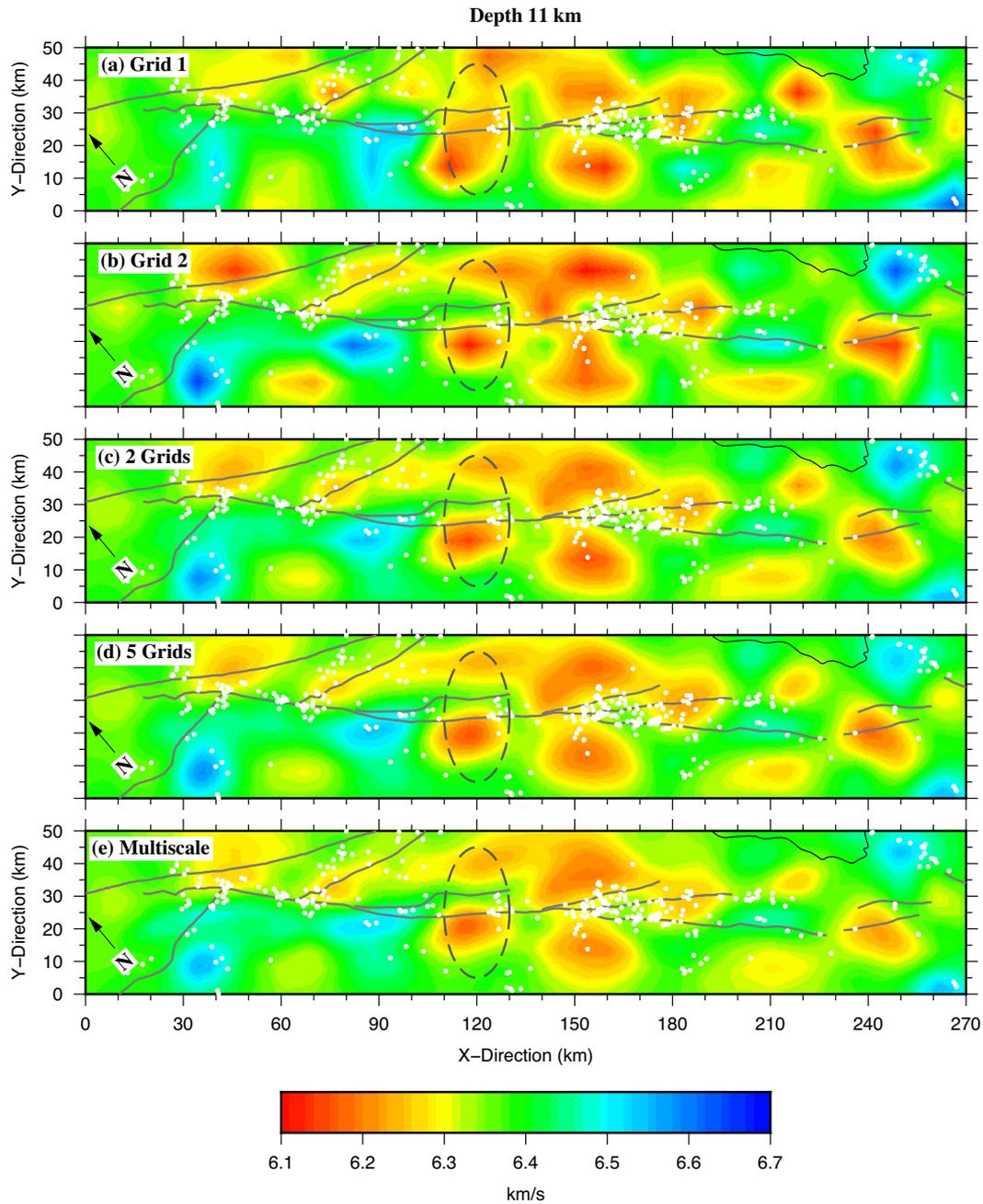


**Figure 10.** The same as Fig. 9 but at the depth of 7.0 km.

higher velocities than the west side from the surface to about 5 km. (2) Above a pronouncing high-velocity anomaly, a significant low-velocity anomaly exists beneath the Salton Trough roughly down to the depth of 11 km. (3) A wide low-velocity zone is observable along the San Jacinto fault at about 11 km depth. This low-velocity structure continues to the lower crust with a decreasing size as the depth increases. Below the depth of 13 km, the low-velocity anomaly mainly concentrates around the central section of the San Jacinto fault zone (around the trifurcation area), and the northwest and southeast ends are occupied by high  $P$ -wave velocity structures. This low-velocity anomaly was also discovered by a reflection tomography study using both  $P$  and  $PmP$  phases and was interpreted as a possible manifestation of the upwelling of partial melt from the mantle (Huang *et al.* 2016). The relatively large earthquakes with

magnitudes greater than 5.0 including the 2016  $M_w$  5.2 Borrego Springs earthquake occurred mainly in the trifurcation area at about 13 km depth. We can observe that these earthquakes were located at the boundaries of the low-velocity region and surrounded by high velocities (Fig. S9). The 1989  $M_w$  6.9 Loma Prieta earthquake and 2014  $M_w$  6.0 South Napa earthquake in northern California were also found in transition zones between high and low velocities (Tong *et al.* 2017). It is probable that the interactions between high- and low-velocity anomalies triggered these relatively large earthquakes.

The present study area almost fully coincides with but is a little smaller than that of Allam & Ben-Zion (2012). Our inversion results show a high level of consistency with theirs. The low-velocity zone atop a high-velocity anomaly beneath the Salton Trough reflects crustal thinning in a region of extension (Allam & Ben-Zion 2012).

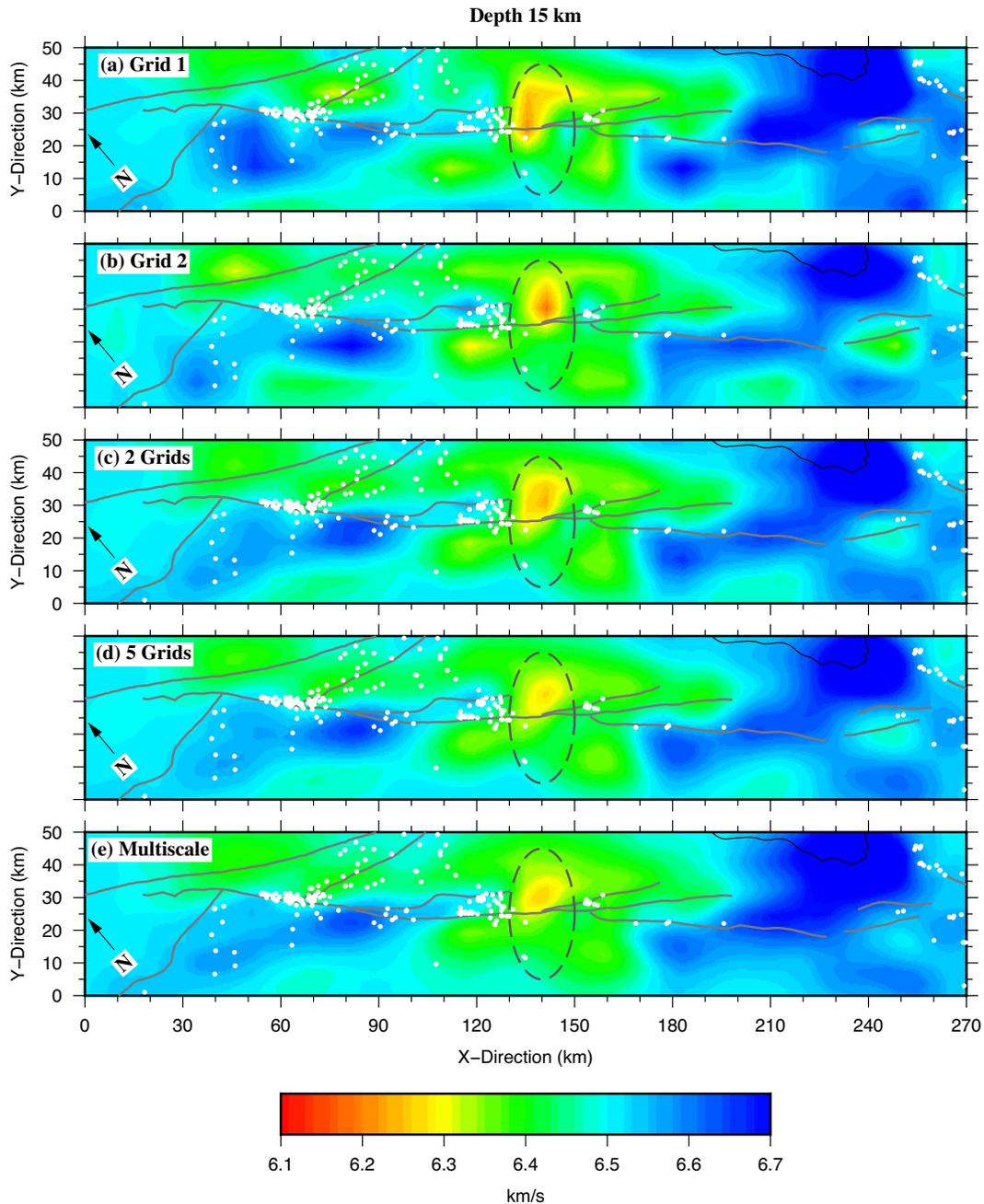


**Figure 11.** The same as Fig. 9 but at the depth of 11.0 km.

However, the thickness of the low-velocity zone is estimated to be 9–11 km in this study, different from 7 km of Allam & Ben-Zion (2012). Reversed velocity contrasts along the San Jacinto fault at shallow depths and a wide low-velocity zone around 11 km depth are revealed by both studies. The main difference is that the  $P$ -wave velocity variation at each representative depth in Allam & Ben-Zion (2012) is larger than that of this study. Scott *et al.* (1994) constructed a  $P$ -wave velocity model of the San Jacinto fault zone near Anza (around the trifurcation area). Their model shows overall faster velocities in the east side of the fault from 0 to 6 km, which can also be observed in our model (Figs 9 and 10). Lin (2013) determined a 3-D velocity model near the Salton Trough and the San Jacinto fault zone and reported a change from low velocity to high velocity at 10 km depth beneath the Salton Trough, consistent

with the findings of this study. This feature is considered to be related to the late Cenozoic spreading events (Lin 2013).

A closer observation can easily detect the differences in location, shape, size and amplitude among the tomographic results generated by using different model parametrizations (Figs 9–12). For example, at each representative depth and around the trifurcation area, the low-velocity anomalies highlighted by dashed ellipses have different shapes and amplitudes (Figs 9–12). At 3 km depth, the low-velocity structure to the east of the northern section of the San Jacinto fault revealed by Grid 1 (Fig. 9a) seems to be less significant than those revealed by the other four approaches (Figs 9b–e). We subtract the velocity given by the mutiscale-staggered-grid approach and the ones generated on the other four grids. Fig. 13 shows the differences at depths of 3, 7, 11 and 15 km. Near the



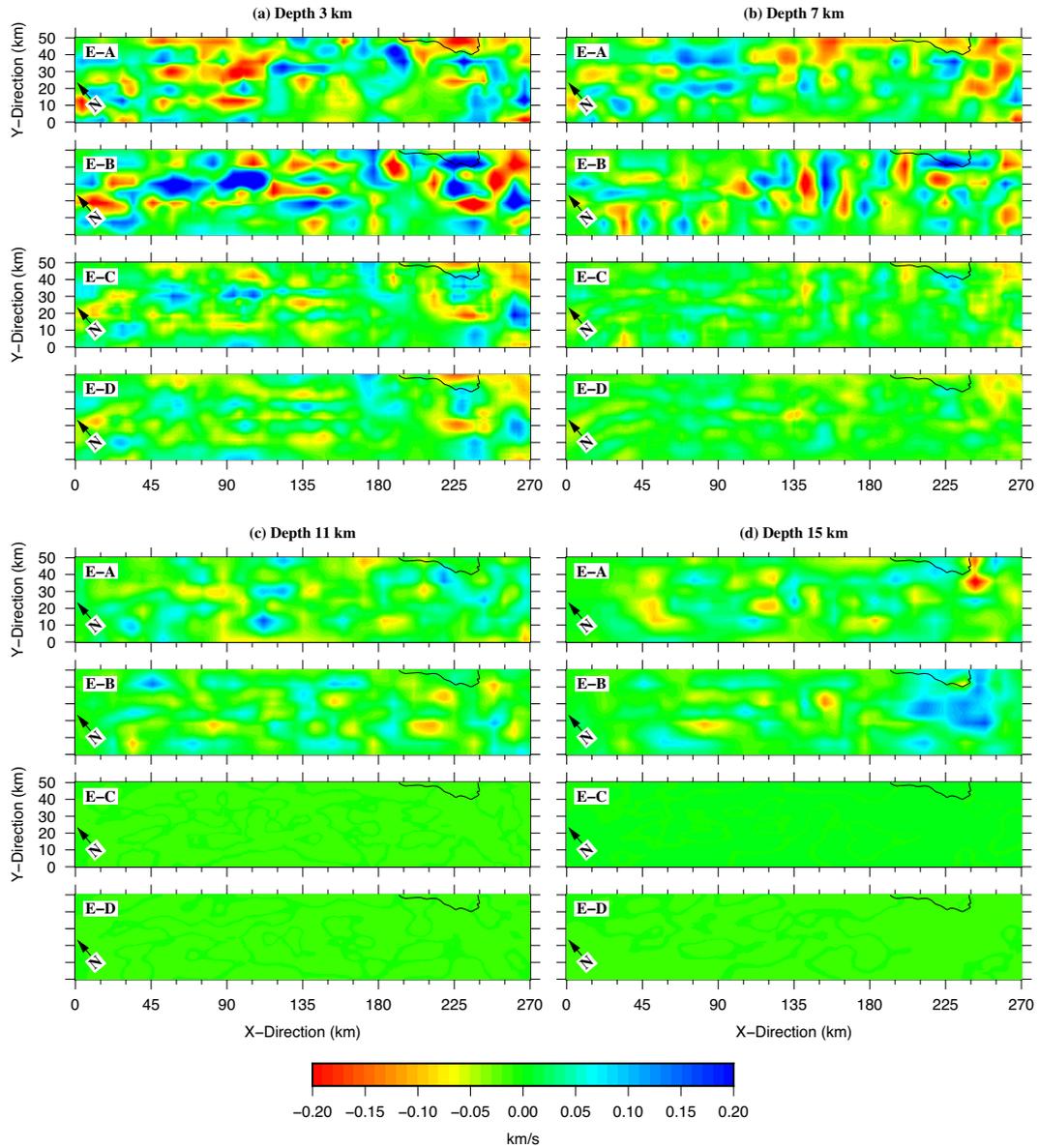
**Figure 12.** The same as Fig. 9 but at the depth of 15.0 km.

surface the dominant differences exceed  $0.2 \text{ km s}^{-1}$  (Fig. 13a) but get reduced as the depth increases. These minor differences may not affect the macroscale observations we previously made, but indicate the necessity of carefully selecting model parametrization in seismic inversions to gain reliable information about local velocity anomalies. The tomographic results generated on Staggered-grid A, Staggered-grid B and the multiscale-staggered-grid show the highest level of similarity, in contrast to the obvious differences between the results of Grids 1 and 2. The increased spatial sampling of the staggered-grids and the multiscale-staggered-grid is probably the main factor that prevents local structural anomalies migrating to other places or being distorted. This suggests that the multiple-grid model parametrization has the advantage of generating stable tomographic results in seismic inversions. Besides that, staggered-grids

and the multiscale-staggered-grid have better performances in reducing the root mean square value of the traveltimes residuals. As shown in Table 2, the multiscale-staggered-grid gives the smallest root mean square value after 20 iterations, and Staggered-grid A gives the second smallest.

## 5.2 Vertical views

To further explore the final results generated by using five different model parametrizations, the relative  $P$ -wave velocity perturbation from the averaged 1-D velocity profile of the smoothed 3-D initial model  $c_0(\mathbf{x})$  (Fig. 14b) is displayed in a series of vertical cross-sectional views from the surface to the uppermost mantle (Figs 15 and 16). The locations of these cross-sections are shown in Fig. 14(a). We can observe again that the overall



**Figure 13.** Velocity differences at four representative depths. The difference is obtained by subtracting the final model given by the multiscale-staggered-grid and the one obtained on Grid 1 (E-A), Grid 2 (E-B), Staggered-Grid A (E-C) or Staggered-Grid B (E-D). Here E stands for subplot (e) in Figs 9, 10, 11 and 12, similar for A, B, C and D.

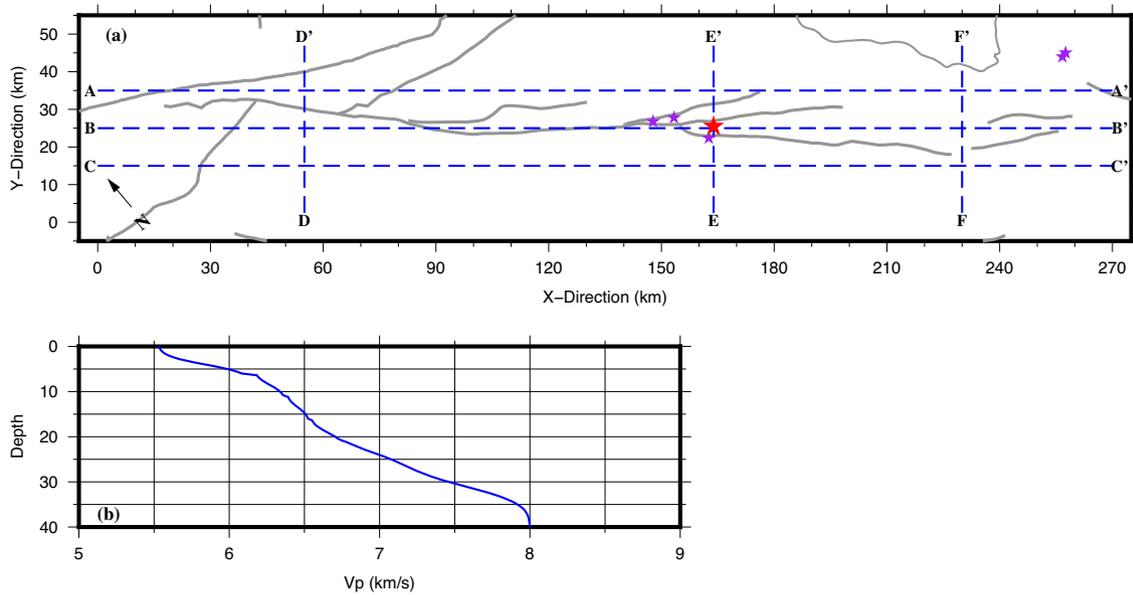
**Table 2.** The root mean square value of the traveltimes residuals in iterative velocity models with different model parametrizations for the tomographic inversions.

	$m_0$	$m_5$	$m_{10}$	$m_{15}$	$m_{20}$
Grid 1	0.1905 s	0.1541 s	0.1411 s	0.1351 s	0.1303 s
Grid 2	0.1905 s	0.1605 s	0.1451 s	0.1361 s	0.1318 s
Staggered-grid A	0.1905 s	0.1565 s	0.1411 s	0.1330 s	0.1275 s
Staggered-grid B	0.1905 s	0.1566 s	0.1424 s	0.1341 s	0.1293 s
Multiscale-staggered-grid	0.1905 s	0.1561 s	0.1418 s	0.1327 s	0.1269 s

patterns of the velocity structure computed with the collocated grids, staggered grids and multiscale-staggered-grid are similar, although the sizes and amplitudes of anomalies show some obvious differences.

Profile BB' nearly aligns with the San Jacinto fault, while Profile AA' and CC' are on the northeast and southwest sides of the fault, respectively. Near the surface (less than 5 km), Profile AA' and

CC' show opposite velocity perturbations from  $x = 40$  km to  $x = 250$  km, clearly demonstrating the strong velocity contrasts across the San Jacinto fault and the reversal in contrast polarity at different segments of the fault. A low-velocity zone sitting on the top of a high-velocity structure is revealed around the southeast ends of the three profiles AA'-CC'. The trifurcation area (roughly from  $x = 120$  km to  $x = 180$  km) generally shows low-velocity structures



**Figure 14.** (a) The map locations of the cross-sections AA', BB', CC', DD', EE' and FF' shown in Figs 15 and 16. (b) The averaged 1-D velocity profile of the 3-D smoothed initial model  $c_0(\mathbf{x})$  in the depth direction.

in the middle and lower crust. The seismicity along Profile BB' mainly occurred in relatively high-velocity structures which can be interpreted as a brittle seismogenic layer (Fig. 16). Similar observations along the San Jacinto fault were reported in Allam & Ben-Zion (2012) and Lin (2013). Profiles DD', EE' and FF' are perpendicular to the fault. The strong velocity contrasts across the San Jacinto fault at shallow depths are clearly shown along the three profiles (Fig. 16). Profiles DD' and FF' have higher velocities on the west side and lower velocities on the east side, contrary to that of Profile EE'. The reversal in contrast polarity indicates the complex structure of the San Jacinto fault. The near-surface low velocity along Profile DD' correlates well with the San Bernadino basin. In comparison with Profiles DD' and FF', Profile EE' passing through the trifurcation area has a wider transition zone from the low velocity on the west side to the high velocity on the east side. We can also find the large-size low-velocity zone beneath the trifurcation area along Profile EE'. Profile FF' exhibits a high-velocity structure below the depth of 10 km. The 2016  $M_w$  5.2 Borrego Springs earthquake is at the intersection of Profile BB' and EE' (Fig. 14). Some other earthquakes with magnitudes greater than 5.0 are also around the meeting section of the two profiles. The tomographic results in Profiles BB' and EE' may suggest that the occurrence of relatively large earthquakes is closely related to the underlying low-velocity zone and the top high velocities.

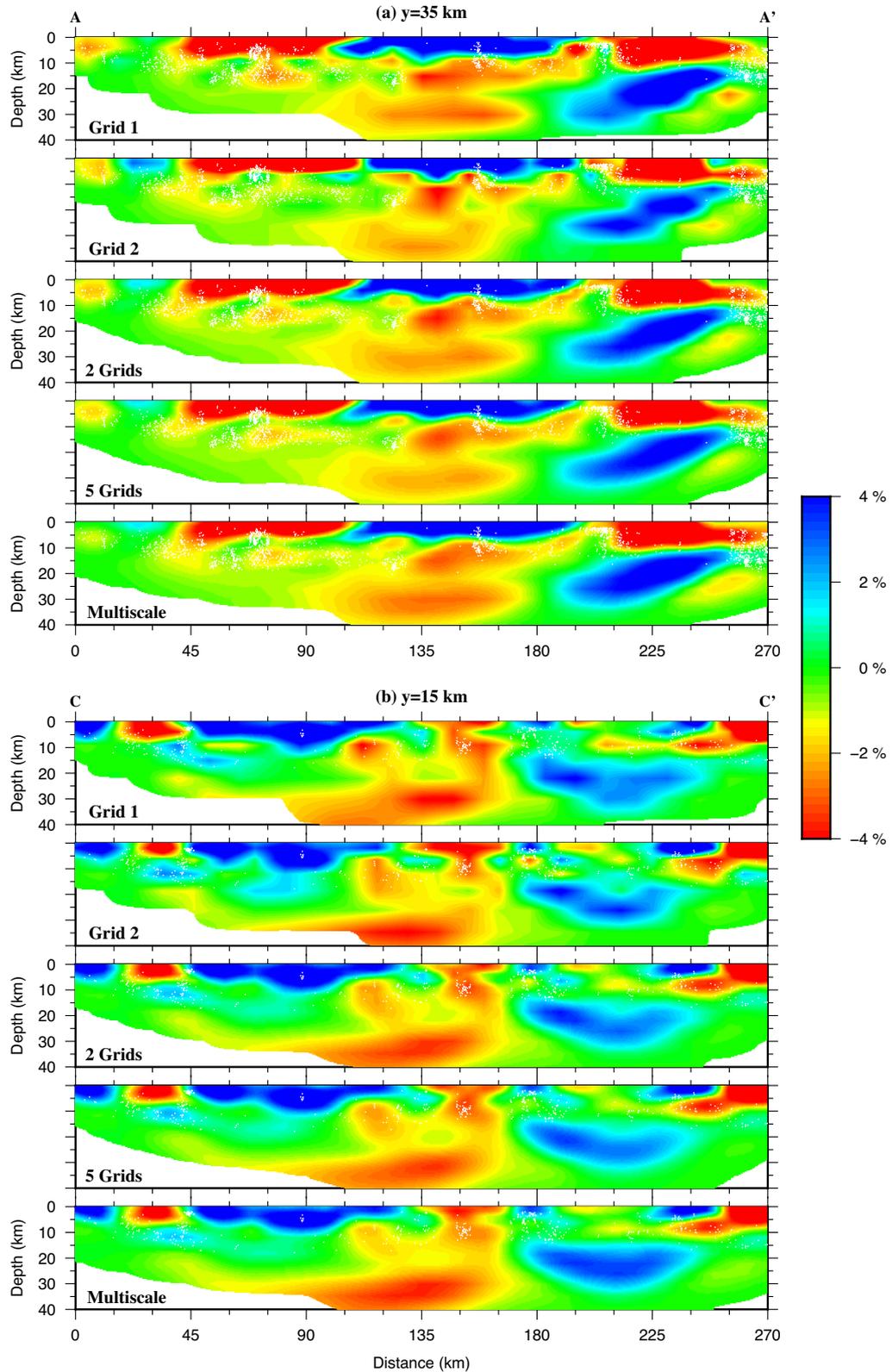
The velocity anomalies along the six vertical profiles revealed by Grids 1 and 2 show obvious differences in location, shape, size and magnitude (Figs 15 and 16). But the results of the multiple-grids including the staggered-grids and the multiscale-staggered-grids are almost the same. For example, the near-surface high-velocity zone along Profile AA' has a length of 90 km from  $x = 100$  km to  $x = 190$  km in the top panel of Fig. 15(a). But its length is reduced to about 80 km from  $x = 115$  km to  $x = 195$  km in the second panel with Grid 2. The multiple-grid approaches uniformly suggest that its location is from  $x = 105$  km to  $x = 200$  km (Fig. 15a). A significant low-velocity anomaly along Profile BB' from  $x = 50$  to  $x = 110$  km is recovered by Grid 2 and the multiple-grids (Figs 16a). But we cannot see it in the result of Grid 1 (Fig. 16a). The near-surface transition from low to high velocities along Profile EE'

occurs on the interval [15, 20] km in the first panel of Fig. 16(c) but happens in between 20 and 25 km in the second panel with Grid 2. The discrepancies between the results of Grids 1 and 2 may get reduced if their grid spacings are reduced. But a finer grid may go beyond the resolution scale of the data. However, the increased spatial sampling of the multiple-grid is helpful in accurately recovering the heterogeneous subsurface structures.

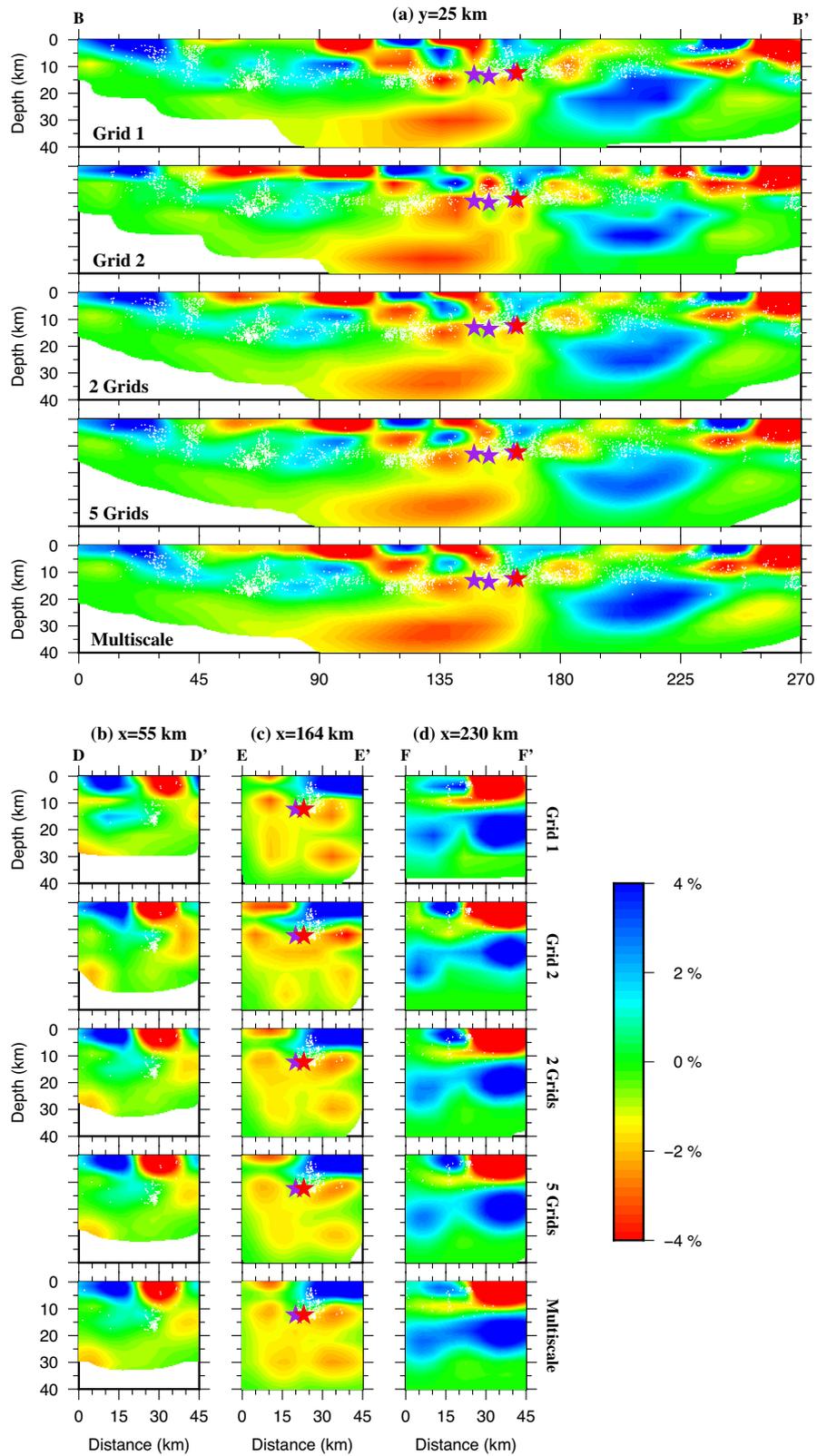
## 6 DISCUSSION AND CONCLUSIONS

Model parametrization greatly influences the final output of the tomographic results. Having an advantage of easy implementation, the multiple-grid model parametrization is proposed to improve the performance of the conventional grid approach in tomographic inversions. The key idea of the multiple-grid approach is that subsurface models are simultaneously constructed on several non-overlapping collocated grids and their average is taken as the output velocity model. Three different types of multiple-grid including the staggered-grid, multigrid and multiscale-staggered-grid have been discussed in this study. The performances of the staggered-grid and multigrid model parametrizations were reported in previous studies (Vesnaver & Bohm 2000; Zhou 2003; Arato *et al.* 2014). Their ability of detecting small velocity change, increasing resolution and minimizing the inversion ambiguities and instability is confirmed again in this study.

Different from reducing the intervals of a collocated grid to have a higher sampling rate and building the subsurface model at a higher but sometimes unrealizable resolution scale, the staggered-grid approach combines the information collected from spatially shifted collocated grids to generate more reliable subsurface images. There are at least two advantages of using staggered-grid model parametrization in seismic tomography studies. First, it is easier to capture dominant anomalies with the increased number of grid nodes. For example, in Figs 3(b) and (c) the sinusoidal anomaly can be efficiently sampled by a staggered-grid that consists of five collocated grids. But this seems to be impossible for one collocated



**Figure 15.** The cross-section views of the relative velocity perturbation field  $\Delta c(x)/c(x)$  given by Grid 1 (Fig. 7a), Grid 2 (Fig. 7b), Staggered-grid A with two collocated grids (Fig. 7c), Staggered-grid B with five collocated grids (Fig. 7d) and (e) the multiscale-staggered-grid (Fig. 8), respectively. Here  $\Delta c(x)$  is the difference between the final model  $c(x)$  and the 1-D velocity profile in Fig. 14(b). The locations of Profile AA' ( $y = 35$  km) and CC' ( $y = 15$  km) are shown in Fig. 14(a). The white dots are earthquakes within a distance of 5.0 km from each profile. The red and blue mean low- and high-velocity anomalies, respectively.



**Figure 16.** The same as Fig. 15 but along Profile BB', DD', EE' and FF'. The stars represent the locations of relatively large earthquakes ( $M_w > 5.0$ ) since 1992. The red star is the 2016  $M_w$  5.2 Borrego Springs earthquake.

grid if we have no information about the distribution of the underground anomaly (Figs 3b and c). Second, the rising artefacts on one collocated grid can be expelled or suppressed by other collocated grids from the same staggered-grid. Synthetic tests of this study have shown the efficacy of the staggered-grid model parametrization (Fig. 3, Supporting Information Figs S3–S5). The real data example in the San Jacinto fault zone also shows that the results generated on the staggered-grids and the multiscale-staggered-grid are more consistent. In addition to that, the staggered-grid approach can reduce the root mean square value of the traveltimes residuals to smaller values than those given by the conventional grid method (Table 2).

The multigrid method is efficient in recovering multiscale subsurface features and able to reduce the non-uniqueness in seismic tomography (Zhou 2003). The artefacts introduced by finer grids can be partially smoothed or suppressed by coarser grids. The finer grids can also make a remedy for the sampling loss of coarser grids. Reliable results can be obtained on a multigrid without knowing any prior information about the sizes of subsurface anomalies (Fig. 3f). The newly presented multiscale-staggered-grid approach is a simple combination of the staggered-grid and multigrid model parametrizations. In comparison with the multigrid method, it has an increased spatial sampling. Different from the staggered-grid method, its component grids have multiscale spatial spacings. We have obtained reliable results with the multiscale-staggered-grid model parametrization (Fig. 3f).

The San Jacinto fault zone is one of the most seismically active regions in southern California and has been intensively investigated by tomographic studies. Given the well-known subsurface structure of this region, it is an ideal place to show the efficacy of the multiple-grid approach in seismic tomography study. Five tomographic models along the San Jacinto fault zone (Figs 9–12 & 15 and 16) are revealed by using two collocated grids, two staggered-grids and one multiscale-staggered-grid. The general patterns of the recovered subsurface structures with the five different model parametrizations are similar to each other and also similar to the results of previous studies (e.g. Allam & Ben-Zion 2012; Lin 2013). Summarily, we can observe strong heterogeneities near the surface, strong velocity contrasts across the San Jacinto fault, reversal in contrast polarity along the fault and a low-velocity zone atop a high-velocity structure near the Salton Trough which reflects the thinning of the crust. The trifurcation area is the host of several relatively large ( $M_w > 5.0$ ) earthquakes in the past 30 yr. Observing from the tomographic results, we believe that the occurrence of these large earthquakes is partially due to the interactions between the hypocentral low-velocity structure and the nearby high-velocity anomalies. A dominant low-velocity zone exists in the middle and lower crust of the trifurcation area. This low-velocity structure may be related to the upwelling of partial melt from the mantle (Huang *et al.* 2016). The differences among the five tomographic models are also obvious (Fig. 13). We have found that the location, shape, size and magnitude of each high- or low-velocity anomaly in one model are different or slightly different from those of the same anomaly in any of the other four models. This phenomenon is mainly caused by the different spatial distributions of the inversion grids. Increasing the spatial sampling through the multiple-grid approach is likely to give more reliable results. This is partially supported by the observation that the models generated on the two staggered-grids and the multiscale-staggered-grid have higher consistency than the ones constructed on the collocated grids (Figs 9–13 & 15 and 16).

This study in the San Jacinto fault zone has shown the efficacy of the multiple-grid model parametrization in practical seismic tomography studies. The multiple-grid approach can be used to suppress artefacts and generate reliable tomographic results. Admittedly, the number of constructing collocated grids of a multiple-grid is chosen subjectively. That how many collocated grids we should choose to form a multiple-grid to maximize its performance remains a question. But we usually use five spatially shifted collocated grids to form a multiple-grid in our tomographic inversions and reliable results can be obtained. Given the promising results of this study, the multiple-grid approach is worthy of further and deeper investigations. We will implement the multiple-grid model parametrization in ray-based teleseismic tomography as well as full waveform inversion in our future studies.

## ACKNOWLEDGEMENTS

We thank the Southern California Earthquake Data Center (SCEDC) for providing the high-quality seismic data used in this study. We are grateful to Dr. Carl Tape and two anonymous reviewers for their comments and suggestions that have greatly improved the original manuscript. This work was supported by the MOE AcRF Tier 1 grant (grant no. M4011899.110). PT and XH were also supported by the National Research Foundation and the Ministry of Education, Singapore under the Research Centres of Excellence initiative. DY was supported by National Natural Science Foundation of China (grant no. 41230210). All the figures are made with the Generic Mapping Tool (GMT; Wessel & Smith 1991).

## REFERENCES

- Aki, K. & Lee, W., 1976. Determination of the three-dimensional velocity anomalies under a seismic array using first  $P$  arrival times from local earthquakes 1. A homogeneous initial model, *J. geophys. Res.*, **81**, 4381–4399.
- 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.
- Arato, A., Godio, A. & Sambuelli, L., 2014. Staggered grid inversion of cross hole 2-D resistivity tomography, *J. Appl. Geophys.*, **107**, 60–70.
- Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, **178**, 1411–1436.
- Bodin, T., Sambridge, M., Rawlinson, N. & Arroucau, P., 2012. Transdimensional tomography with unknown data noise, *Geophys. J. Int.*, **189**, 1536–1556.
- Dahlen, F., Nolet, G. & Hung, S., 2000. Fréchet kernels for finite-frequency traveltimes - I. Theory, *Geophys. J. Int.*, **141**, 157–174.
- Dziewonski, A., 1984. Mapping the lower mantle: determination of lateral heterogeneity in  $P$  velocity up to degree and order 6, *J. geophys. Res.*, **89**, 5929–5952.
- Dziewonski, A.M., Hager, B.H. & O'Connell, R.J., 1977. Large-scale heterogeneities in the lower mantle, *J. geophys. Res.*, **82**, 239–255.
- Fichtner, A. *et al.*, 2018. The collaborative seismic Earth model: generation 1, *Geophys. Res. Lett.*, **45**, 4007–4016.
- Geiger, L., 1910. Herdbestimmung bei erdbeben ans den ankunftszeiten, *Königlichen Gesellschaft der Wissenschaften Zu Göttingen*, **4**, 331.
- Hassouna, M.S. & Farag, A.A., 2007. Multistencils fast marching methods: a highly accurate solution to the eikonal equation on Cartesian domains, *IEEE Trans. Pattern Anal. Mach. Intell.*, **29**, 1563–1574.
- Hofstetter, R., Dorbath, C. & Calo, M., 2012. Crustal structure of the dead sea basin from local earthquake tomography, *Geophys. J. Int.*, **189**, 554–568.
- Huang, X., Yang, D., Tong, P., Badal, J. & Liu, Q., 2016. Wave equation-based reflection tomography of the 1992 Landers earthquake area, *Geophys. Res. Lett.*, **43**, 1884–1892.

- Inoue, H., Fukao, Y., Tanabe, K. & Ogata, Y., 1990. Whole mantle  $P$ -wave travel time tomography, *Phys. Earth planet. Inter.*, **59**, 294–328.
- Jiang, G., Zhao, D. & Zhang, G., 2009. Seismic tomography of the Pacific slab edge under Kamchatka, *Tectonophysics*, **465**, 190–203.
- Kissling, E., Husen, S. & Haslinger, F., 2001. Model parametrization in seismic tomography: a choice of consequence for the solution quality, *Phys. Earth planet. Inter.*, **123**, 89–101.
- Lin, G., 2013. Three-dimensional seismic velocity structure and precise earthquake relocations in the Salton Trough, Southern California, *Bull. seism. Soc. Am.*, **103**, 2694–2708.
- Liu, Q. & Gu, Y.J., 2012. Seismic imaging: from classical to adjoint tomography, *Tectonophysics*, **566–567**, 31–66.
- Liu, S., Suardi, I., Yang, D., Wei, S. & Tong, P., 2018. Teleseismic traveltimes tomography of northern Sumatra, *Geophys. Res. Lett.*, **45**, GL078610, doi:10.1029/2018GL078610.
- Mewes, A., Kullessa, B., McKinley, J.D. & Binley, A.M., 2010. Anisotropic seismic inversion using a multigrid Monte Carlo approach, *Geophys. J. Int.*, **183**, 267–276.
- Nakajima, J. & Hasegawa, A., 2007. Tomographic evidence for the mantle upwelling beneath southwestern Japan and its implications for arc magmatism, *Earth planet. Sci. Lett.*, **254**, 90–105.
- Paige, C. & Saunders, M., 1982. LSQR: An algorithm for sparse linear equations and sparse least-squares, *Trans. Math. Softw.*, **8**, 43–71.
- Rawlinson, N., Pozgay, S. & Fishwick, S., 2010. Seismic tomography: a window into deep Earth, *Phys. Earth planet. Inter.*, **178**, 101–135.
- Ross, Z.E., Hauksson, E. & Ben-Zion, Y., 2017. Abundant off-fault seismicity and orthogonal structures in the San Jacinto fault zone, *Sci. Adv.*, **3**, e1601946, doi:10.1126/sciadv.1601946
- Sambridge, M., Braun, J. & McQueen, H., 1995. Geophysical parametrization and interpolation of irregular data using natural neighbours, *Geophys. J. Int.*, **122**, 837–857.
- Sambridge, M. & Rawlinson, N., 2005. Seismic tomography with irregular meshes, in *Seismic Earth: Array Analysis of Broadband Seismograms—Geophysical Monograph*, pp. 49–65, eds Lavender, A. & Nolet, G., Am. Geophys. Un.
- SCEDC, 2013. Southern California Earthquake Center. Caltech.Dataset, doi:10.7909/C3WD3xH1.
- Scott, J., Masters, T.G. & Vernon, F.L., 1994. 3-D velocity structure of the San Jacinto fault zone near Anza, California—I.  $P$ -waves, *Geophys. J. Int.*, **119**, 611–626.
- Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, **307**(5715), 1615–1618.
- Sharp, R.V., 1967. San Jacinto fault zone in the peninsular ranges of southern California, *Bull. geol. Soc. Am.*, **78**, 705–730.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science*, **325**, 988–992.
- Thurber, C. & Eberhart-Philips, D., 1999. Local earthquake tomography with flexible gridding, *Comput. Geosci.*, **25**, 809–818.
- Thurber, C., Zhang, H., Brocher, T. & Langenheim, V., 2009. Regional three-dimensional seismic velocity model of the crust and uppermost mantle of northern California, *J. geophys. Res.*, **114**, B01304, doi:10.1029/2008JB005766.
- Thurber, C.H., 1983. Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California, *J. geophys. Res.*, **88**, 8226–8236.
- Tong, P., Chen, C.-W., Komatitsch, D., Basini, P. & Liu, Q., 2014a. High-resolution seismic array imaging based on an SEM-FK hybrid method, *Geophys. J. Int.*, **197**(1), 369–395.
- Tong, P., Yang, D., Li, D. & Liu, Q., 2017. Time-evolving seismic tomography: the method and its application to the 1989 Loma Prieta and 2014 South Napa earthquake area, California, *Geophys. Res. Lett.*, **44**, 3165–3175.
- Tong, P., Yang, D., Liu, Q., Yang, X. & Harris, J., 2016. Acoustic wave-equation-based earthquake location, *Geophys. J. Int.*, **205**, 464–478.
- Tong, P., Zhao, D. & Yang, D., 2011. Tomography of the 1995 Kobe earthquake area: comparison of finite-frequency and ray approaches, *Geophys. J. Int.*, **187**, 278–302.
- Tong, P., Zhao, D., Yang, D., Yang, X., Chen, J. & Liu, Q., 2014b. Wave-equation-based travel-time seismic tomography—Part 1: Method, *Solid Earth*, **5**, 1151–1168.
- Tong, P., Zhao, D., Yang, D., Yang, X., Chen, J. & Liu, Q., 2014c. Wave-equation-based travel-time seismic tomography—Part 2: Application to the 1992 Landers earthquake area, *Solid Earth*, **5**, 1169–1188.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**, 195–216.
- Vesnaver, A. & Bohm, G., 2000. Staggered or adapted grids for seismic tomography?, *Leading Edge*, **19**, 944–950.
- Wessel, P. & Smith, W.H.F., 1991. Free software helps map and display data, *EOS, Trans. Am. geophys. Un.*, **72**(1), 441–448.
- Yang, Y. & Ritzwoller, M.H., 2008. Teleseismic surface wave tomography in the western U.S. using the Transportable Array component of USArray, *Geophys. Res. Lett.*, **35**(4), L04308, doi:10.1029/2007GL032278.
- Zhang, H., Sarkar, S., Toksoz, M.N., Kuleli, H.S. & Al-kindy, F., 2009. Passive seismic tomography using induced seismicity at a petroleum field in Oman, *Geophysics*, **74**, WCB57–WCB69.
- Zhang, H. & Thurber, C., 2005. Adaptive-mesh seismic tomography based on tetrahedral and voronoi diagrams: application to Parkfield, California, *J. geophys. Res.*, **110**, B04303, doi:10.1029/2004JB003186.
- Zhao, D., 2001. New advances of seismic tomography and its applications to subduction zones and earthquake fault zones: a review, *Island Arc*, **10**, 68–84.
- Zhao, D., 2009. Multiscale seismic tomography and mantle dynamics, *Gondwana Res.*, **15**, 297–323.
- Zhao, D., 2015. *Multiscale Seismic Tomography*, Springer.
- Zhao, D., Hasegawa, A. & Horiuchi, S., 1992. Tomographic imaging of  $P$  and  $S$  wave velocity structure beneath northeastern Japan, *J. geophys. Res.*, **97**, 19 909–19 928.
- Zhao, D., Todo, S. & Lei, J., 2005. Local earthquake reflection tomography of the Landers aftershock area, *Earth planet. Sci. Lett.*, **235**, 623–631.
- Zhou, H., 2003. Multiscale traveltimes tomography, *Geophysics*, **68**, 1639–1649.
- Zhou, H.-W., 1996. A high-resolution  $P$  wave model for the top 1200 km of the mantle, *J. geophys. Res.*, **101**, 27 791–27 810.
- Zhu, L. & Kanamori, H., 2000. Moho depth variation in southern California from teleseismic receiver functions, *J. geophys. Res.*, **105**, 2969–2980.
- Zigone, D., Ben-Zion, Y., Campillo, M. & Roux, P., 2015. Seismic tomography of the southern California plate boundary region from noise-based Rayleigh and Love waves, *Pure appl. Geophys.*, **172**, 1007–1032.

## SUPPORTING INFORMATION

Supplementary data are available at ([0:italic](https://doi.org/10.1017/GJI.000000))GJI/([0:italic](https://doi.org/10.1017/GJI.000000)) online.

**Figure S1.** (a) The trade-off analysis between the data variance and model variance with different damping parameters.  $\lambda = 8$  is the optimal damping parameter that lies at the corner of the L-shaped curve. (b) The smoothness of the obtained velocity model and the root mean squared value of the traveltimes residuals with different smoothing radiuses.  $\sigma = 1.6$  is an optimal choice.

**Figure S2.** (a)–(g) Each shows the vertical view of the relative velocity perturbation field at  $y = 25$  km recovered on a collocated grid with different smoothing radiuses  $\sigma$ . The grid size is 2.02 km by 2.02 km by 1.07 km. The smoothing radius is shown on top of each. (h) The ‘true’ relative velocity perturbation.

**Figure S3.** The results of the multiscale reconstruction test at the depths of 0.0 km (I) and 4.0 km (II). The input models at the two depths are shown in (a) and (e). The results of the collocated grid are presented in (b) and (f). (c) and (g) show the inverted results on a staggered-grid consisting of five collocated grids. (d) and (h) are the images generated on a multiscale-staggered-grid.

**Figure S4.** The same as Fig. S3 but at the 9.0 km (III) and 15.0 km (IV) depths.

**Figure S5.** The same as Fig. S3 but at the 22.0 km (V) and 30.0 km (VI) depths.

**Figure S6.** The same as Fig. 5 in the main text but  $a = b = 0.08$ . The wavelength of the anomaly along the  $x$  or  $y$  axis is 12.5 km.

**Figure S7.** Plan views of the inverted velocity models with five different model parametrizations at the depth of 5 km.  $N$  points to the north direction. The red colour indicates low velocity and blue colour shows high-velocity structures.

**Figure S8.** The same as Fig. S7 but at the depth of 9 km. The stars are earthquakes with magnitudes greater than 5.0 since 1992.

**Figure S9.** The same as Fig. S8 but at the depth of 13 km.

**Figure S10.** The same as Fig. S7 but at the depth of 17 km.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.