Geophysical Journal International

Geophys. J. Int. (2014) **199,** 944–955
GJI Seismology

doi: 10.1093/gji/ggu305

Wavelet-based double-difference seismic tomography with sparsity regularization

Hongjian Fang[1,2] and Haijiang Zhang[1,2]

[1] *Wantai Microseismic Lab of School of Earth and Space Sciences, University of Science and Technology of China, Hefei* 230026*, China.*
E-mail: zhang11@ustc.edu.cn
[2] *Laboratory of Seismology and Physics of Earth's Interior, University of Science and Technology of China, Hefei* 230026*, China*

Accepted 2014 August 1. Received 2014 July 31; in original form 2014 April 9

SUMMARY

We have developed a wavelet-based double-difference (DD) seismic tomography method. Instead of solving for the velocity model itself, the new method inverts for its wavelet coefficients in the wavelet domain. This method takes advantage of the multiscale property of the wavelet representation and solves the model at different scales. A sparsity constraint is applied to the inversion system to make the set of wavelet coefficients of the velocity model sparse. This considers the fact that the background velocity variation is generally smooth and the inversion proceeds in a multiscale way with larger scale features resolved first and finer scale features resolved later, which naturally leads to the sparsity of the wavelet coefficients of the model. The method is both data- and model-adaptive because wavelet coefficients are non-zero in the regions where the model changes abruptly when they are well sampled by ray paths and the model is resolved from coarser to finer scales. An iteratively reweighted least squares procedure is adopted to solve the inversion system with the sparsity regularization. A synthetic test for an idealized fault zone model shows that the new method can better resolve the discontinuous boundaries of the fault zone and the velocity values are also better recovered compared to the original DD tomography method that uses the first-order Tikhonov regularization.

Key words: Wavelet transform; Body waves; Seismic tomography.

1 INTRODUCTION

For seismic tomography, the velocity model is generally parametrized into regular 3-D cells or grid nodes (Aki & Lee 1976; Thurber 1983). However, due to non-uniform station geometry, uneven distribution of seismic sources, missing data and ray bending, the ray coverage is generally highly uneven. Ideally, the distribution of inversion cells or nodes should adaptively match with the resolving power of the data in such a way that the cells/nodes are denser where the ray coverage is better and more sparse in the regions where few rays pass through. However, because of the restriction of the regular cell or grid spacing, it is difficult to parametrize the model to be adaptive with the uneven ray distribution. As a result, the inversion problem becomes relatively ill-conditioned and various regularization methods such as Tikhonov regularization need to be applied to stabilize the inversion (Aster *et al.* 2013). However, these regularization methods to some degree deteriorate the model resolution and small-scale features in the model are generally smeared out even in the region where the ray coverage is good.

To partly account for the mismatch between the regular inversion nodes and uneven ray distribution, Zhou (2003) proposed a multiscale traveltime tomography method that sequentially uses multiple grids with different grid spacing. The coarser grid inversion result serves as the starting model for the finer grid inversion. In this way, different scales of velocity features may be better recovered. However, for the fine grid inversion, strong regularization needs to be applied to make the inversion well conditioned. To better tackle this issue, an adaptive seismic tomography method is preferred in which the cells or nodes are irregularly distributed to match with the uneven ray distribution (Abers & Roecker 1991; Vesnaver 1996; Curtis & Snieder 1997; Sambridge & Gudmundsson 1998; Böhm *et al.* 2000; Spakman & Bijwaard 2001; Zhang & Thurber 2005). As shown by Zhang & Thurber (2005), the adaptive-mesh double-difference (DD) seismic tomography method based on tetrahedral or Voronoi diagrams can better resolve the finer scale structure of the velocity model.

For the adaptive seismic tomography methods, the goal is to match the ray distribution to the density of the inversion cells or nodes. Therefore, they are purely data-adaptive and do not consider whether the model parametrization is appropriate for the model itself. Because of the inherent multiscale nature of the wavelet representation (Daubechies 1992), wavelet-based multiscale tomography method has been proposed to compensate for the mismatch problem of uneven ray distribution and regular cells/nodes (e.g. Chiao & Kuo 2001; Chiao & Liang 2003; Loris *et al.* 2007; Delost *et al.* 2008; Hung *et al.* 2011; Simons *et al.* 2011). In the wavelet-based seismic tomography methods, the wavelet coefficients of the model, instead of the model itself, are the

parameters to be solved for. This provides a natural way to resolve the model at different scales. Compared to purely data-adaptive seismic tomography methods, the wavelet-based methods adapt the inversion cells/nodes implicitly according to the data distribution and the model complexity. If a certain part of the model is smooth, only sparse cells/nodes are required to represent it and thus only wavelet coefficients at large scales are significant and kept in the inversion. On the other hand, if a certain part of the model has small-scale features with good ray coverage, more wavelet coefficients are required to sufficiently represent the model. However, for regions with small-scale features but sparse ray path coverage, wavelet-based tomography methods still cannot resolve fine features.

Previous wavelet-based seismic tomography methods used different wavelet families, different strategies of realizing the wavelet transform, and different regularization tools, and also solved different scales of problems. Chiao & Kuo (2001) and Chiao & Liang (2003) were the first to propose wavelet-based multiresolution parametrization for geophysical inverse problems. In their case, the model components are resolved hierarchically at different scales embedded within the wavelet representation and thus there is no need to invoke additional smoothness regularization in addition to the damping regularization. Following this strategy, Hung *et al.* (2010, 2011) proposed a wavelet-based finite-frequency teleseismic tomography method to image the velocity structure of the central Tibet. The second-generation wavelet transform by the lifting scheme of Sweldens (1996) has been adopted by Chiao & Kuo (2001), Chiao & Liang (2003) and Hung *et al.* (2010, 2011). Delost *et al.* (2008) also used the second-generation wavelet transform for the first-arrival traveltime tomography by designing a bit mask operator to only invert for the wavelet coefficients in specific areas where the resolution is high.

For the wavelet-based seismic tomography methods of Chiao & Kuo (2001), Chiao & Liang (2003) and Hung *et al.* (2010, 2011), the model is solved in the wavelet domain and only damping is applied to regularize the inversion. This is equivalent to minimizing the L_2 norm of the model wavelet coefficients. In comparison, Loris *et al.* (2007) regularized the wavelet-based teleseismic tomography by minimizing the L_1-norm of the model wavelet coefficients to make them sparse. This is based on the fact that a few non-zero coefficients are enough to represent model parameters which is generally smooth with some fine scale features. Simons *et al.* (2011) also proposed using a class of spherical wavelet basis for tomographic inversion of global seismic data. The inversion scheme was formulated to minimize the L_2 norm of the data misfit and the L_1 norm of the model wavelet coefficients. The sparsity-constrained inversion has been shown to be effective to better resolve the model if it is sparse in nature (Aster *et al.* 2013). It allows us to resolve sharp discontinuities and small scale features without the *a priori* smoothing regularization term, which is more subjective and may smear out the fine scale features in the regions with good ray path coverage. For example, Zhu & Bamler (2010) used a compressive sensing approach in synthetic aperture radar tomography and obtained a super-resolution image with highly accurate point localization. In this study, we essentially follow the strategy of Loris *et al.* (2007) to solve the DD tomography problem of Zhang & Thurber (2003) in the wavelet domain but use a different strategy for solving the sparse wavelet coefficients, which uses an iteratively reweighted least-squares (IRLS) method to obtain a sparse solution instead using the soft thresholding method. The code realizing the new method is tested using the same synthetic data set as Zhang & Thurber (2003).

2 METHODOLOGY

For the DD seismic tomography method, in addition to absolute arrival times, differential arrival times from event pairs observed on common stations are also used (Zhang & Thurber 2003):

$$r_k^i = \sum_{l=1}^{3} \frac{\partial T_k^i}{\partial x_l^i} \Delta x_l^i + \Delta\tau^i + \sum_n s_n^{ik} \delta u_n \text{ for absolute data}$$

$$\mathrm{d}r_k^{ij} = r_k^i - r_k^j = \sum_{l=1}^{3} \frac{\partial T_k^i}{\partial x_l^i} \Delta x_l^i + \Delta\tau^i + \sum_n s_n^{ik} \delta u_n - \left(\sum_{l=1}^{3} \frac{\partial T_k^j}{\partial x_l^j} \Delta x_l^j + \Delta\tau^j + \sum_n s_n^{jk} \delta u_n \right) \text{ for differential data.} \quad (1)$$

In the above equations, r_k^i is the arrival time residual for event i and station k, $\delta\tau^i$ is the origin time perturbation for event i, $\Delta x_l^i (l = 1, 2, 3)$ are the location perturbations in three coordinate directions, δu_n is the slowness perturbation and s_n^{ik} and s_n^{jk} are the weighted ray path segment length with respect to node n for ray paths from event i and j to station k, respectively. In the matrix form, for the absolute data it can be written as:

$$\mathbf{A}\Delta\mathbf{X} + \mathbf{C}\Delta\mathbf{M} = \Delta\mathbf{T}, \quad (2)$$

where $\mathbf{A}$ and $\mathbf{C}$ are the sensitivity matrix to the event location and slowness parameters, respectively, $\Delta\mathbf{X}$ is the vector including event locations and origin times, $\Delta\mathbf{M}$ is the slowness perturbation parameters and $\Delta\mathbf{T}$ is the data misfit vector. For the differential data, we can simply apply a difference operator $\mathbf{Q}$ to eq. (2) as follows (Zhang & Thurber 2006):

$$\mathbf{QA}\Delta\mathbf{X} + \mathbf{QC}\Delta\mathbf{M} = \mathbf{Q}\Delta\mathbf{T}. \quad (3)$$

For a 3-D wavelet transform represented by a matrix $\mathbf{W}$ (Daubechies 1992), we can rewrite eq. (2) as:

$$\mathbf{A}\Delta\mathbf{X} + \mathbf{CW}^{-1}\mathbf{W}\Delta\mathbf{M} = \Delta\mathbf{T}. \quad (4)$$

Eq. (3) can be also rewritten as:

$$\mathbf{QA}\Delta\mathbf{X} + \mathbf{QCW}^{-1}\mathbf{W}\Delta\mathbf{M} = \mathbf{Q}\Delta\mathbf{T}. \quad (5)$$

Then we can combine eqs (4) and (5) to a matrix form:

$$\tilde{\mathbf{G}}\tilde{\mathbf{m}} = \mathbf{d}, \quad (6)$$

where $\tilde{\mathbf{G}} = \begin{bmatrix} \mathbf{A} & \mathbf{CW}^{-1} \\ \mathbf{QA} & \mathbf{QCW}^{-1} \end{bmatrix}$, $\tilde{\mathbf{m}} = \begin{bmatrix} \Delta\mathbf{X} \\ \mathbf{W}\Delta\mathbf{M} \end{bmatrix}$ and $\mathbf{W}^{-1}$ represents the inverse wavelet transform. We can write $\mathbf{W}^{-1} = \mathbf{W}^{\mathbf{T}}$ if we use orthogonal wavelet basis or $\mathbf{W}^{-1} = \tilde{\mathbf{W}}^{\mathbf{T}}$ if we use bi-orthogonal wavelet basis, where $\tilde{\mathbf{W}}$ is the matrix representing the wavelet transform using the dual wavelet basis. Note that we just apply the wavelet transform to the slowness part of the sensitivity matrix and leave out the location and origin time parameters. We use the lifting scheme to achieve the wavelet transform, which decomposes a certain signal into approximation coefficients and detail coefficients at different scales (Sweldens 1996). The approximation coefficients are obtained by representing the signal in the space formed by the scaling functions at different scales, which mostly reflect the smooth component of the signal. In comparison, the detail coefficients are obtained by representing the signal in the space formed by wavelet functions at different scales, which mostly reflect the small-scale features of the signal. The inverse problem is now recast as first seeking the wavelet coefficients of the model and then using the inverse wavelet transform matrix $\mathbf{W}^{-1}$ to get the model on a discrete spatial grid.

Because the linearized equation $\mathbf{Gm} = \mathbf{d}$ for seismic tomography is generally ill conditioned, some regularization methods are required to make the inversion stable. Zeroth-order or higher-order Tikhonov regularization methods can be applied to make the model either not far from an *a priori* model or smooth. For higher-order Tikhonov regularization methods, they may not be appropriate in the case that the model has discontinuities such as faults. For a model that is smooth at large scales but varies rapidly in some regions, its representation is generally sparse in the wavelet domain meaning that only the approximation coefficients and a few non-zero detail coefficients are sufficient to recover the main features of the model (Daubechies 1992). Following Loris *et al.* (2007), instead of using the Tikhonov regularization methods, we apply the sparsity constraint to the model wavelet coefficients. Different from Loris *et al.* (2007), we use the IRLS to solve the inversion system in the wavelet domain to obtain the sparse solution of the model wavelet coefficients. Instead of applying the wavelet transform to the teleseismic tomography problem (Simons *et al.* 2011), we apply it to the local scale DD tomography problem.

The L_0 norm, which is defined as the number of non-zero coefficients, can be used to measure the sparsity of the solution. However, applying the L_0 norm constraint to the inversion model is extremely computational prohibitive since the resulting cost function becomes non-convex (Aster *et al.* 2013). Here we used the L_1 norm instead, which has been shown to also produce a sparse solution (Candès *et al.* 2006; Donoho 2006). The inverse problem is to seek a solution to satisfy $\tilde{\mathbf{G}}\tilde{\mathbf{m}} = \mathbf{d}$ with a minimal $||\tilde{\mathbf{m}}||_1$, leading to a minimization problem as follows,

$$\min||\tilde{\mathbf{G}}\tilde{\mathbf{m}} - \mathbf{d}||_2^2 + \lambda||\tilde{\mathbf{m}}||_1, \tag{7}$$

where λ is the Lagrange multiplier that is used to balance between the data misfit and the sparsity of the solution. It can be determined by the discrepancy principle (Scherzer 1993) if we know the noise level in the data. However, in the real case of seismic tomography, we often do not know the data noise level, thus we generally use the L-curve method or the generalized cross-validation method to find the optimum λ (Aster *et al.* 2013). In comparison, this is a convex problem and we can solve it efficiently.

It is well known that the L_1 norm regularization always leads to a sparse solution and for some very large classes of $\tilde{\mathbf{G}}$, the L_1 minimizer leads to a unique solution when sufficiently sparse solutions exist (Candès *et al.* 2006; Donoho 2006). Eq. (7) can be solved by the IRLS algorithm (Scales *et al.* 1988), in which the L_1-norm functional in (7) is replaced by a weighted L_2-norm functional, and the weights are computed iteratively from the previous model parameters $\tilde{\mathbf{m}}$. For the IRLS algorithm, eq. (7) is recast to solve the following least squares problem (Aster *et al.* 2013):

$$\min \left\| \begin{bmatrix} \tilde{\mathbf{G}} \\ \sqrt{\frac{\lambda}{2}}\sqrt{\dot{\mathbf{W}}} \end{bmatrix} \tilde{\mathbf{m}} - \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix} \right\|^2, \tag{8}$$

with

$$\dot{W}_{i,i} = \begin{cases} 1/|\tilde{m}_i| & |\tilde{m}_i| \geq \xi \\ 1/\xi & |\tilde{m}_i| < \xi \end{cases}, \tag{9}$$

where ξ is a tolerance parameter to avoid the case of division by zero. The above algorithm essentially results in a first-order approximation to the L_1 norm objective function. In the IRLS procedure, because the matrix $\tilde{\mathbf{G}}$ consists of elements corresponding to different physical entities related to source parameters (location and origin time) and wavelet coefficients of the model slowness parameters, the column scaling is applied to mitigate this issue, similar to the original DD tomography method of Zhang & Thurber (2003).

For realizing wavelet-based DD tomography with sparsity regularization using eq. (8), it involves two levels of iterations. The first level of iterations is based on the linearization of non-linear inverse problem by updating the model iteratively for solving (7). For each first level iteration, the second level of iterations is required in the IRLS procedure for solving (8). We first solve (8) by setting $\dot{\mathbf{W}}$ as the identity matrix, then update $\dot{\mathbf{W}}$ and iterate a few times to get a sparse solution. During the second level of iterations, the value of λ is fixed. In our study, we choose ξ to be 10^{-6} considering the machine accuracy for single precision floating-point number.

In practice, at different stages of inversion we use different threshold values to zero out small or nearly zero wavelet coefficients of the sensitivity matrix in the wavelet domain. This procedure keeps the main features of the sensitivity matrix but makes the inversion more stable and data adaptive, as will be shown in the next section. We use relatively larger thresholds for the first few iterations and then smaller thresholds for later iterations. This is equivalent to determining the model at coarser scales at the beginning and then gradually at finer scales. This process naturally results in a model whose wavelet coefficients are mostly zeros except for the region where the model varies discontinuously. For the area with sparse ray coverage, most wavelet coefficients of the model parameters will be zero due to the sparsity constraint, so only large-scale features are recovered. For the region with dense ray coverage, the detail wavelet coefficients could have more non-zero values if the model has small-scale features associated with them. Therefore the wavelet-based seismic tomography method is inherently multiscale and data adaptive.

3 SYNTHETIC TEST

To validate the effectiveness of the wavelet-based DD tomography method, we test it with the same synthetic data set as Zhang & Thurber (2003). The synthetic data set was constructed based on an idealized model of the velocity structure of the San Andreas fault in central California (Kissling *et al.* 1994). To the west of the 'fault' ($X < 0$), the velocity is constant (6 km s^{-1}); to the east, there is a very sharp gradient into a low-velocity zone of 4 km s^{-1} from $X = 1$ to 5 km, a sharp gradient to 5 km s^{-1} at $X = 6$ km, and then a linear increase from 5 to 6 km s^{-1} in the region between 6 and 38 km (Fig. 1a). The events and stations used to construct the synthetic data set are from the actual seismicity and U.S. Geological Survey stations in the Loma Prieta region (Fig. 2). Gaussian random noise with zero mean and a standard deviation of 0.04 s was added to the true arrival times. In addition, a constant noise term was also added to the arrivals at each station from a uniform distribution between −0.3 and 0.3 s. This was to simulate the case that the systematic errors (model errors and pick bias) associated with the arrival times are larger than the random ones. The differential arrival times are constructed by directly subtracting absolute arrival times for event pairs at common stations.

In Zhang & Thurber (2003), the inversion nodes along the X coordinate were set to coincide with the fault boundaries. Here the X–Y nodes used to parametrize the velocity model are different and are much finer compared to the previous DD tomography study. This is to test the ability of the method to recover the structure when there is no *a priori* information available for the model. The current inversion grid nodes are at $X = -35, -15, -10, -5, -3, -1, 0, 1, 2, 3, 4, 5, 6, 10, 20, 35$ km, and at $Y = -60, -40, -20, -10, 0, 10,$

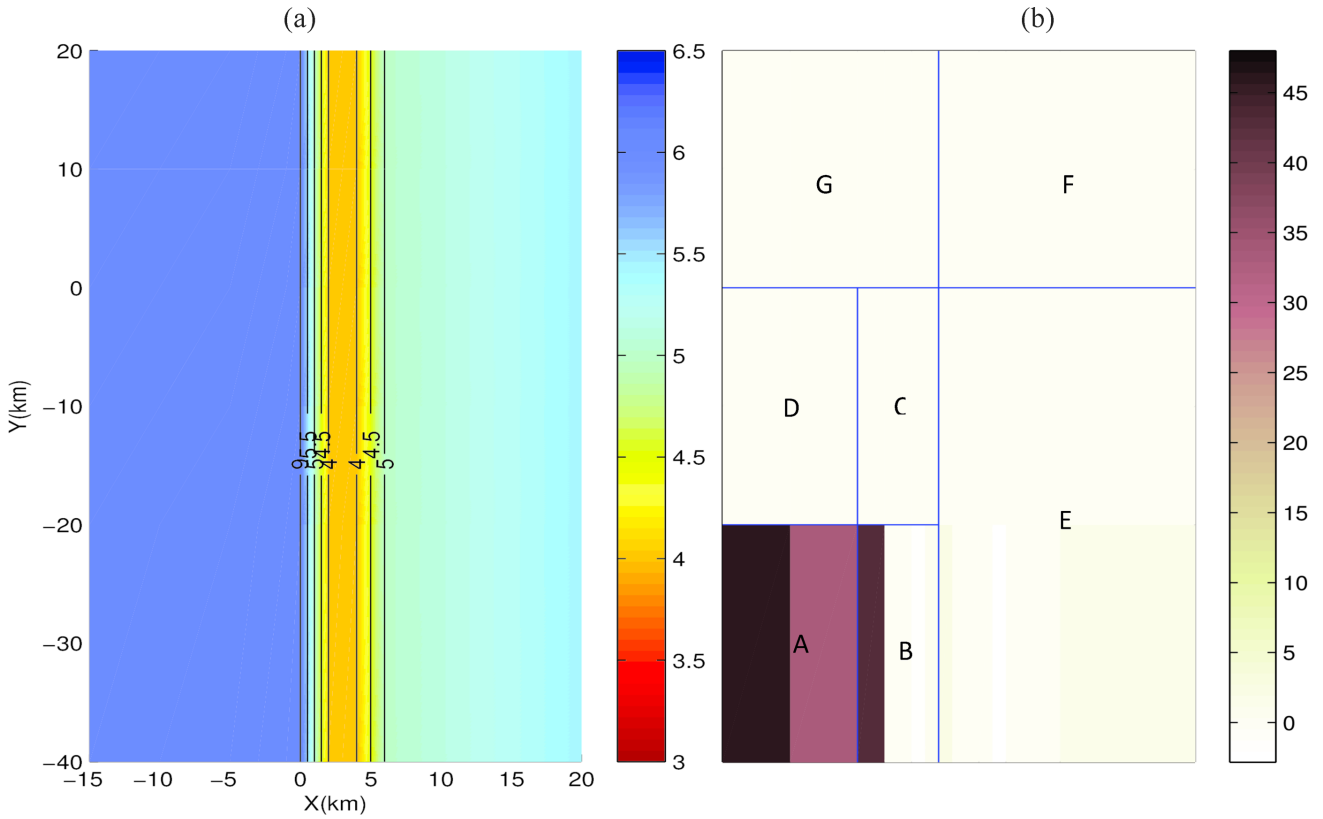

Figure 1. (a) A horizontal slice through the true synthetic velocity model and (b) the approximation and detail coefficients of the model in the Haar wavelet domain. The true velocity model in 3-D is similar to a 'vertical sandwich', which is constant with the depth. The wavelet transform is conducted up to two scales. E, F and G mark detail coefficients at scale 1, respectively. B, C and D mark detail coefficients at scale 2, respectively. A marks the approximation coefficients at scale 2. It can be seen that the model in the wavelet domain is quite sparse due to its step-like features.

20, 35 km. The nodes in depth are positioned at 0, 1, 3, 5, 7, 9, 11 and 16 km.

For this synthetic data set, the 1D wavelet transform using the lifting scheme of Sweldens (1996) is applied sequentially to each dimension of the 3-D model. As a result, the largest approximation coefficients will be located in a small cube around the corner of the 3-D model in the wavelet domain. Due to the difficulty of visualizing the approximation and detail wavelet coefficients in 3-D, we just plot the 2-D wavelet coefficients of the true velocity model in the horizontal plane in Fig. 1(b). Because the model is represented in non-uniform grid nodes, the wavelet transform based on the lifting scheme is appropriate for dealing with the non-uniform model sampling. This is because the lifting scheme only involves the averaging and subtraction calculations for the signal samples at certain scales using the predefined wavelet filters (Sweldens 1996). For seismic traveltime tomography, the sensitivities of traveltime with respect to slowness model parameters are simply weighted ray path lengths for related inversion nodes. In our implementation of $\tilde{\mathbf{G}} = \mathbf{G}\mathbf{W}^{\mathrm{T}}$, we first obtain the sensitivities for the model slowness parameters for each ray path and then apply the wavelet transform to them to construct the new sensitivity matrix. In this way, we can directly build the sensitivity matrix $\tilde{\mathbf{G}}$ in the wavelet domain. We select the Haar and D4 wavelets from the Daubechies wavelet family (Daubechies 1992) due to their orthogonality and compactness. We select a 1-D velocity model increasing from 5.0 to 5.5 km s^{-1} with depth as the initial model, which is very different from the true model.

We first use the standard DD tomography method to see how well the synthetic model can be recovered. 10 iterations of simultaneous inversion are performed. The absolute traveltime root mean square (rms) residual drops from 0.583 s before the inversion to 0.091 s after the inversion, and the differential time rms residual decreases from 0.253 to 0.059 s (Fig. 3a). The horizontal slices of the velocity model at different depths are shown in Fig. 4(a). It can be seen that low velocity zones can be recovered but at various degrees at different depths. The absolute difference between the inverted model from the standard DD tomography and the true velocity model has a mean value of 0.287 km s^{-1}, a median value of 0.182 km s^{-1}, and a standard deviation of 0.086 km s^{-1}.

For the wavelet-based DD tomography, we first use the Haar wavelet and the result converges after five iterations (Fig. 3b). The value of λ in eq. (7) is selected by using the L-curve method. The absolute traveltime RMS residual drops from 0.583 s before the inversion to 0.059 s after the inversion. In comparison, the differential time rms residual decreases from 0.253 to 0.059 s. The horizontal slices of the velocity model resulting from wavelet-based DD tomography using the Haar wavelet are shown in Fig. 4(b). In comparison to the standard DD tomography, the model is better recovered with a more accurate low velocity zone and sharper fault zone boundaries. The absolute difference with the true model is smaller, with the mean value, median value and standard deviation value of 0.154, 0.095 and 0.027 km s^{-1}, respectively. As seen in Fig. 4(b), the low velocity boundaries are much more sharply recovered. We also use the D4 wavelet to test the effect of different wavelet families on inversion. The convergence curves for absolute and differential traveltimes are shown in Fig. 3(c) and the velocity model is shown in Fig. 4(c), respectively. The convergence curves using the Harr and D4 wavelets look very similar (Figs 3b and c). The absolute

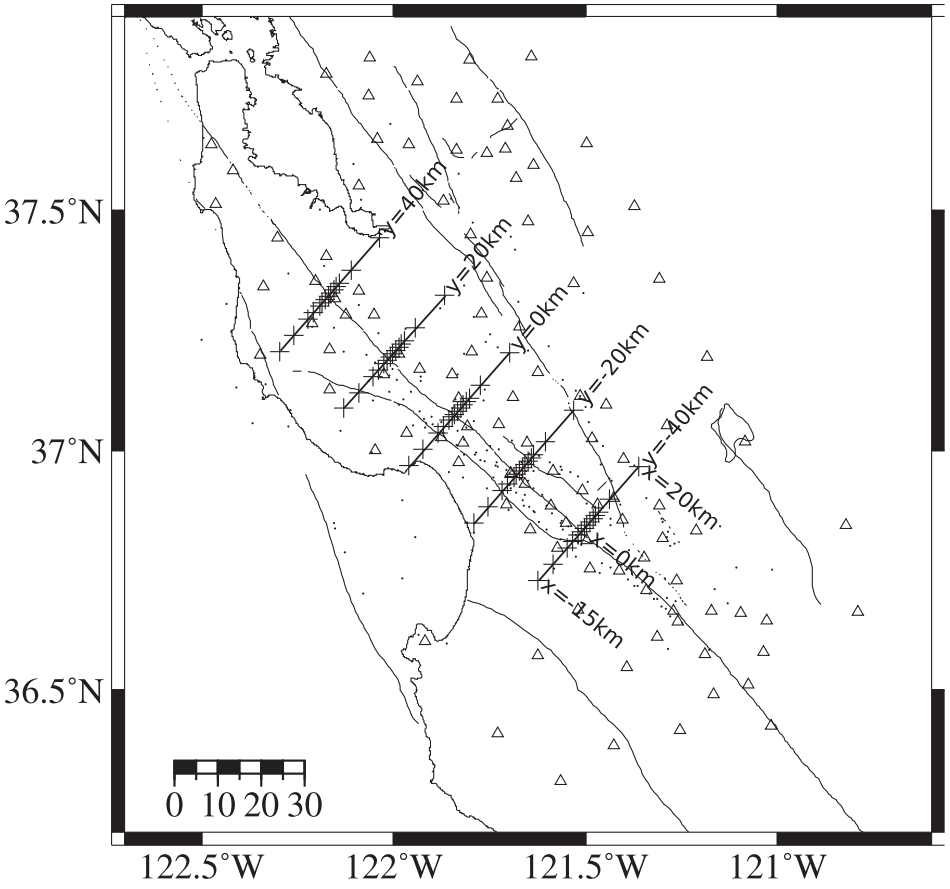

Figure 2. Event location (filled circles) and stations (open triangles) used for the synthetic data set. The tomographic inversion grid nodes are shown as crosses. They are located at $X = -35, -15, -10, -5, -3, -1, 0, 1, 2, 3, 4, 5, 6, 10, 20, 35$ km, at $Y = -60, -40, -20, -10, 0, 10, 20, 35$ km and at $Z = 0, 1, 3, 5, 7, 9, 11, 16$ km, respectively.

difference between the inverted model and the true model has a mean value of 0.191 km s^{-1}, a median value of 0.115 km s^{-1} and the standard deviation value of 0.041 km s^{-1}, which are smaller than the standard DD tomography but larger than the case using the Haar wavelet.

We also test how the inverted model differs when applying the L_2-norm regularization to the model wavelet coefficients, as used by Hung *et al.* (2010, 2011). The convergence curve and the inverted model by using the Haar wavelet are shown in Figs 3(d) and 4(d), respectively. Compared to the case using the L_1-norm regularization (Fig. 4b), the absolute and differential traveltime residuals decrease in a similar manner along with the iteration. The absolute difference between the inverted model and the true model has a mean value of 0.246 km s^{-1}, a median value of 0.178 km s^{-1} and a standard deviation value of 0.053 km s^{-1}, which are about twice of the corresponding values for the model inverted using the L_1-norm regularization but still slightly smaller compared to the model inverted from the standard DD method.

4 DISCUSSIONS

A checkerboard resolution test is also performed to test the effectiveness of the wavelet-based DD tomography method. Compared to the 'vertical sandwich' velocity model (Fig. 1a), the checkerboard model is more complicated because of positive and negative anomalies are added on the velocity model alternatively in three directions (Fig. 5). The recovery degree of the checkerboard pattern is characterized by the resolvability (Zelt 1998), which is used to assess the recovered checkerboard model quantitatively. For each node the resolvability is defined as:

$$R = \frac{\sum_{j=1}^{M} (Dn_{tj} + Dn_{rj})^2}{2\sum_{j=1}^{M} (Dn_{tj}^2 + Dn_{rj}^2)}, \tag{10}$$

with Dn_{tj} and Dn_{rj} the true and recovered velocity anomalies at the jth node inside the model region consisting of M nodes. For this synthetic test, we select two neighbouring nodes in each direction centring the calculated node.

The resolution test shows that the Haar wavelet-based DD tomography method with the L_1-norm regularization performs very well and small-scale checkerboard patterns can be well recovered with an average resolvability of 0.85 (Fig. 5b). The D4 wavelet-based method with the L_1-norm regularization recovers the checkerboard pattern slightly worse with an average resolvability of 0.74 (Fig. 5c), but most small-scale patterns can be recovered. The relatively better performance of the Haar wavelet than the D4 wavelet might simply attribute to the fact that both the synthetic model structure and the checkerboard pattern can be better represented by the Haar wavelet basis. This is because both models have step-like features, similar to the Haar wavelet. In comparison, the Haar wavelet-based DD tomography method with the L_2-norm regularization recovers the checkerboard pattern the best (Fig. 5d), with an

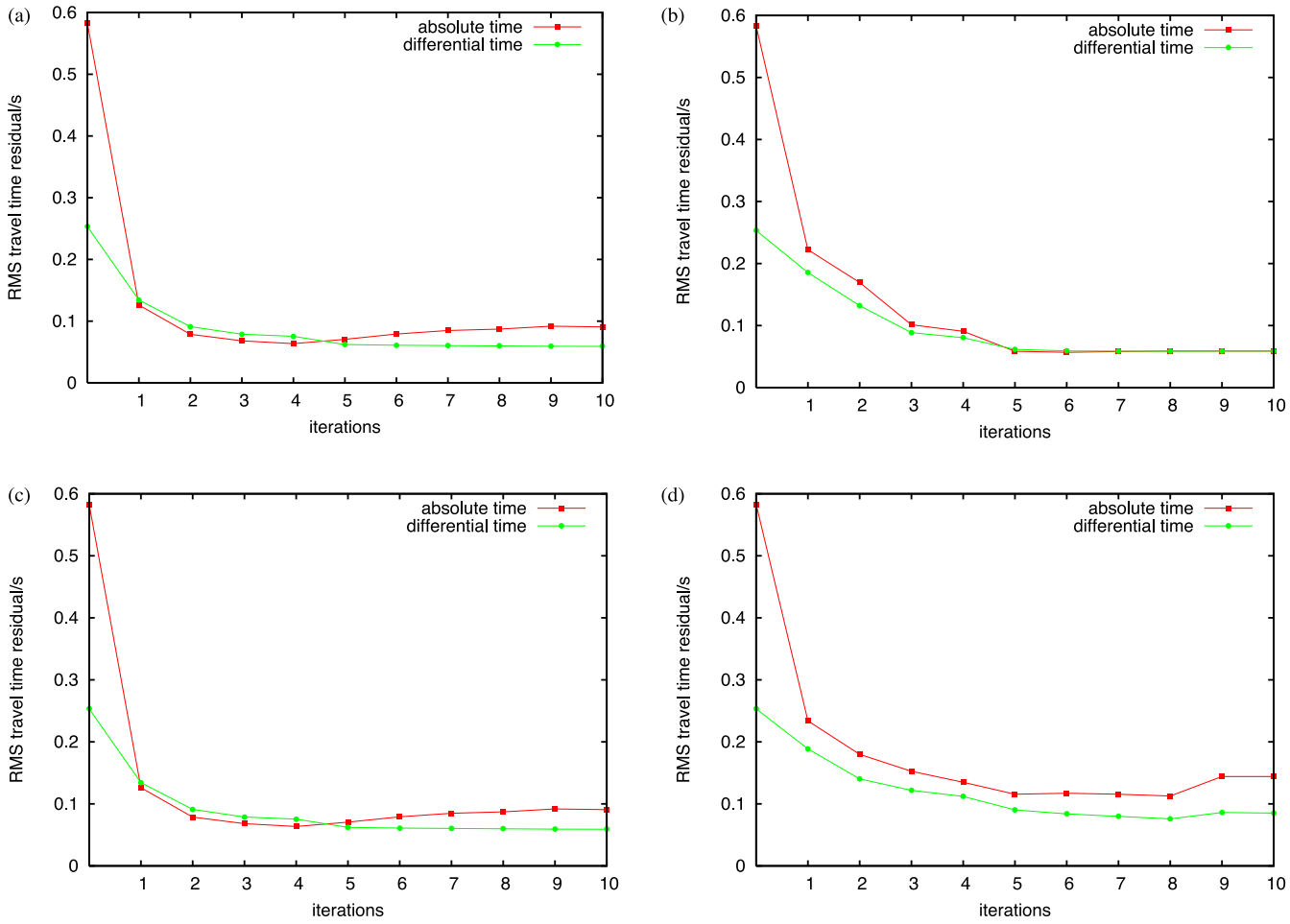

Figure 3. The rms absolute (red line) and differential (green line) traveltime residuals with iterations for (a) the standard DD tomography and the wavelet-based DD tomography with (b) the Haar wavelet by L_1-norm regularization, (c) the D4 wavelet by L_1-norm regularization and (d) the Haar wavelet by L_2-norm regularization.

average resolvability of 0.86, which is slightly better than the recovered model with the L_1-norm regularization. The major difference between the L_1-norm regularization and the L_2-norm regularization is that the former results in a sparse model in the wavelet domain while the latter gives a model with many small wavelet coefficients. In the case that the model has many small-scale features such as the checkerboard model, the L_2-norm regularization may be more appropriate since the model parameter is not sparse in the wavelet domain. For the 'vertical sandwich' model shown in Fig. 1, the model is expected to be sparse in the wavelet domain, thus the L_1-norm regularization shall perform better, as shown in Fig. 4. As expected, the standard DD tomography method performs the worst with an average resolvability of 0.73 (Fig. 5a). It can be seen that many small-scale checkerboard features are smeared out due to the use of the smoothing regularization in the inversion.

In order to make the inversion more stable and converge faster, at each iteration we zero out some small detail coefficients related to the transformed sensitivity matrix $\tilde{\mathbf{G}}$ but keep the approximation coefficients. Similar to the standard DD tomogaphy, we can also calculate the derivative weight sum (DWS) value of the transformed sensitivity matrix for each node (Thurber & Eberhart-Phillips 1999). The DWS values in the space domain can be used to approximately quantify how well the nodes are sampled by the nearby ray paths. A larger DWS value indicates that the velocity at the corresponding grid node is constrained by more ray paths. However, it does not necessarily mean the grid node with a larger DWS value is better resolved since the azimuthal sampling of ray paths is also important. The threshold used to zero out the small wavelet coefficients is based on the mean DWS value in the wavelet domain for all the nodes. We set up a larger threshold value (90 per cent of the mean DWS in the wavelet domain) at the first two iterations and reduce it to 50 per cent in the following three iterations and then 10 per cent in the last three iterations, which naturally leads to a multiscale inversion that first resolves the robust large scale features and then successively adds more and more resolvable fine scale details. Test shows that this thresholding procedure does not change the main feature of the original sensitivity matrix (Fig. 6a). In this test, we zero out half of the wavelet coefficients and then reconstruct the sensitivity matrix by the inverse wavelet transform to see how much the sensitivity matrix has changed. We can see even though half of the wavelet coefficients have been zeroed out, the recovered DWS distribution is similar to the original one, indicating that the sensitivity matrix is just minimally changed (Fig. 6b). By thresholding small wavelet coefficients of the transformed sensitivity matrix, we only need to invert for a limited number of wavelet coefficients of the model. As a result, the inversion becomes more stable.

In comparison, the thresholding we used here has a different meaning from that of Loris *et al.* (2007). Our goal is to apply different threshold values to the sensitivity matrix in the wavelet

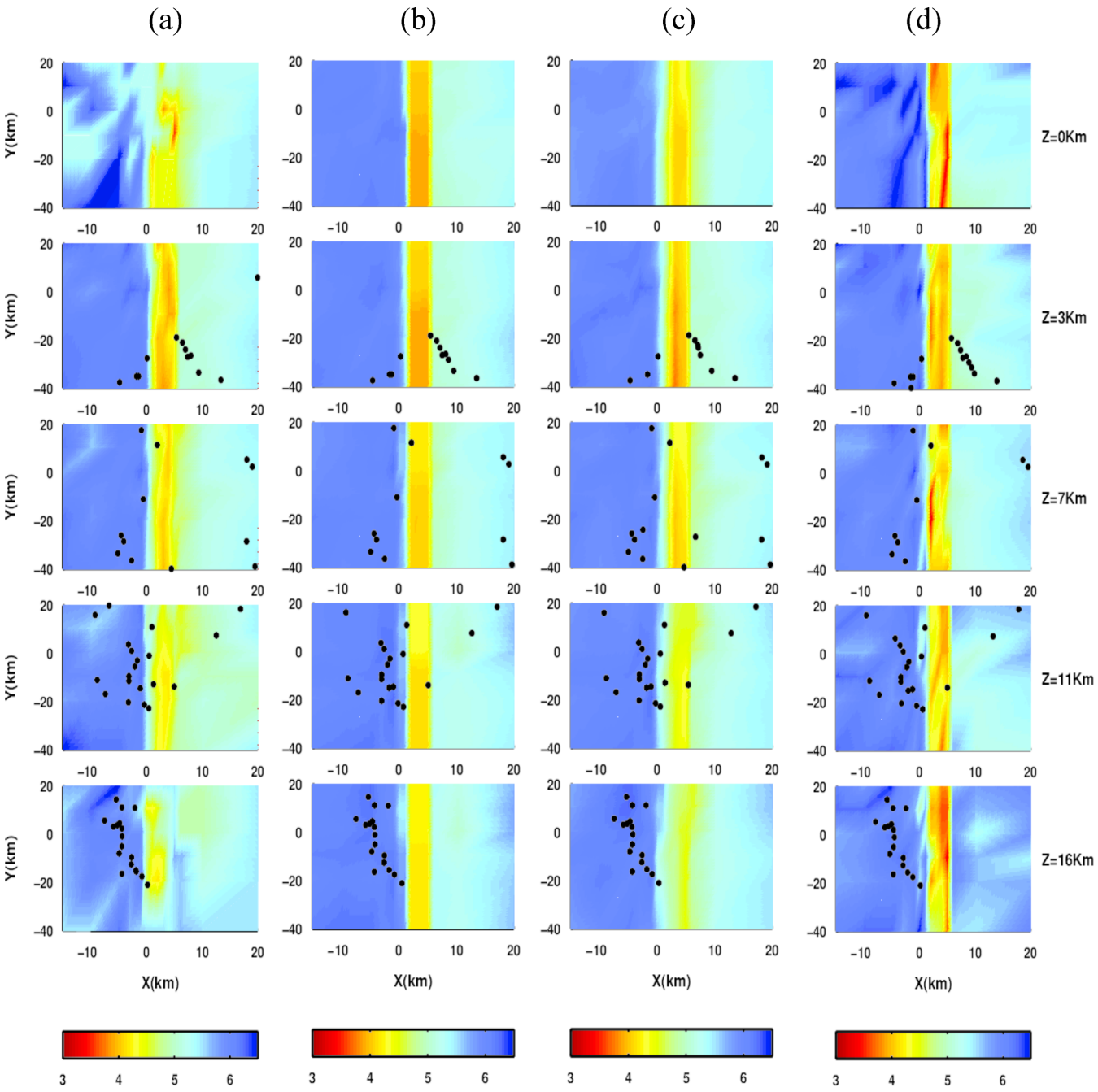

Figure 4. Horizontal slices through the velocity models from (a) the standard DD tomography and the wavelet-based DD tomography with (b) the Haar wavelet by L_1-norm regularization, (c) the D4 wavelet by L_1-norm regularization and (d) the Haar wavelet by L_2-norm regularization.

domain at different inversion stages to achieve a multiscale inversion. In the case of Loris *et al.* (2007), they used the soft-thresholding to the model wavelet coefficients to get a sparse solution during the inversion. In our case, the IRLS procedure is used instead to obtain a sparse solution of model wavelet coefficients without zeroing out small values by the hard or soft thresholding. We also tested using the soft thresholding for removing small values of the sensitivity matrix in the wavelet domain and the inverted model looks similar.

The wavelet coefficients of the sensitivity matrix are partly related to the ray coverage as represented by the average DWS values from all iterations (Fig. 7). Overall, we can see the DWS values at inversion grid nodes are relatively larger when the corresponding model wavelet coefficients are inverted. In comparison, the DWS values are small when the model wavelet coefficients are zeroed out, indicating that only the grid nodes with large DWS values in the space domain are inverted. It is noted that some nodes have relatively large DWS values but are still zeroed out during the inversion (Fig. 7). This is because at first few iterations the threshold value for zeroing out the elements of the transformed sensitivity matrix in the wavelet domain is relatively large. This shows that the wavelet-based tomography method is inherently data adaptive. The combination of zeroing out small wavelet coefficients of the sensitivity matrix and the sparsity constraint to the model wavelet coefficients by the IRLS algorithm leads to very sparse wavelet coefficients for the model. The percentage of non-zero coefficients increases gradually

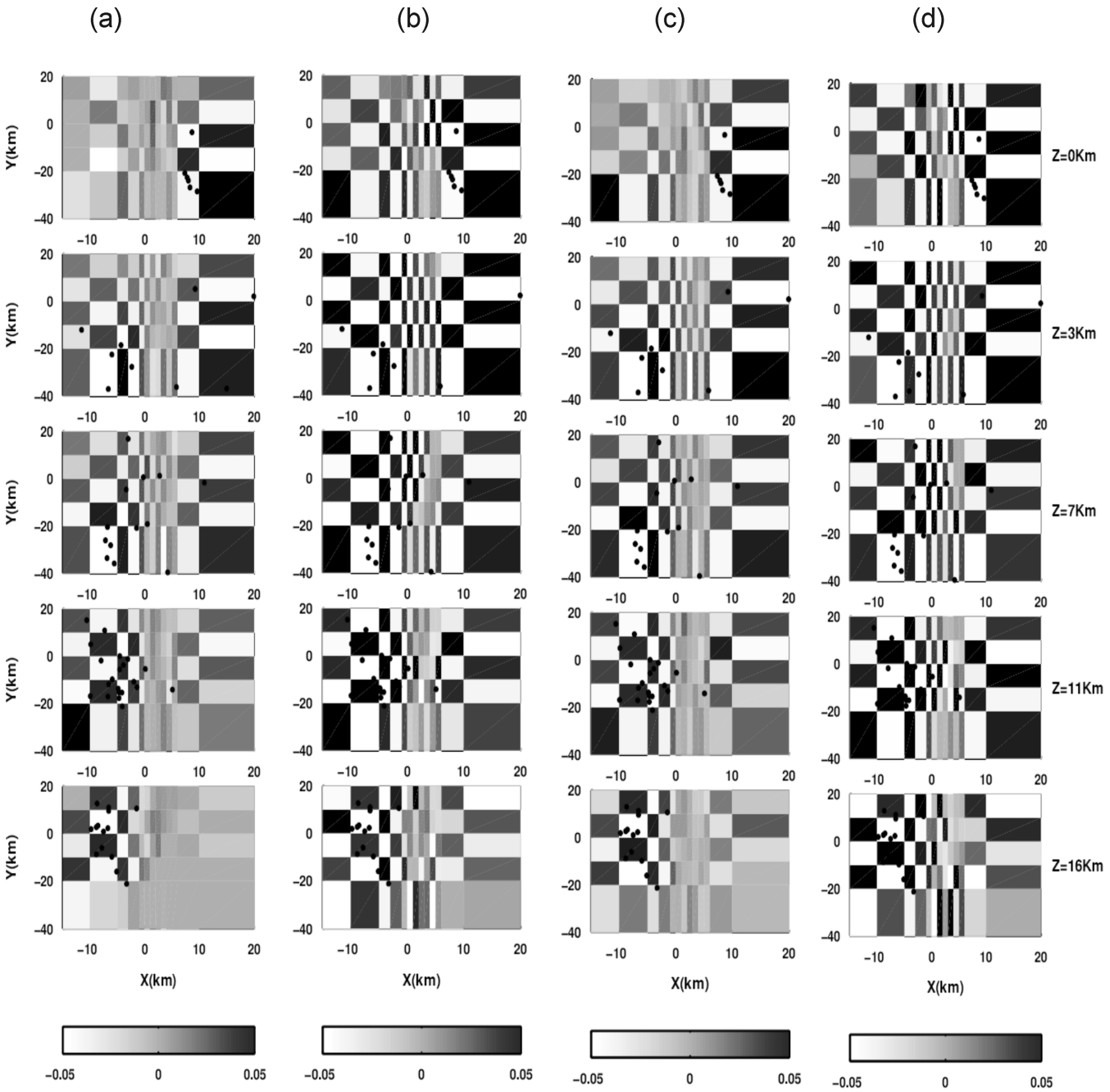

Figure 5. The checkerboard resolution tests for (a) the standard DD tomography and the wavelet-based DD tomography using (b) the Haar wavelet by the L_1-norm regularization, (c) the D4 wavelet by the L_1-norm regularization and (d) the Haar wavelet by the L_2-norm regularization.

from about 2 to 9 per cent along with the iterations, indicating that more and more fine scale features are resolved.

When applying the wavelet transform to the sensitivity matrix, the transformed $\tilde{\mathbf{G}}$ matrix in the wavelet domain has different characteristics. Let us take a ray path in two dimensions as an example. Fig. 8(a) shows one row of the sensitivity matrix that represents the ray path segments in each cell. When we transform it into the wavelet domain, it resembles a 'fat ray' (Figs 8b and c). It is known that by using the fat ray, the tomographic inversion is more stable (Husen & Kissling 2001). Actually, for the synthetic test, the condition number of the sensitivity matrix is 75.7 in the wavelet domain using the Haar wavelet, smaller than 121.9 of the original sensitivity matrix with the same damping applied. This shows that the sensitivity matrix in the wavelet domain is less ill-conditioned and the inversion is more stable.

To further test the ability of the wavelet-based DD tomography method, we only choose the approximation coefficients of the transformed sensitivity matrix to do the inversion (Fig. 9). It is surprising to see after only the first iteration the main features of the 'vertical-sandwich' model can be resolved by just using the approximation coefficients. In this case, less than 2 per cent of the elements of the transformed sensitivity matrix are non-zero. Because all of the detail coefficients are set to be zero, the norm of the recovered sensitivity matrix is about 35 per cent of that of the original sensitivity matrix, which means we can never recover the true sensitivity matrix. However, with more iterations, the amplitude of the velocity model

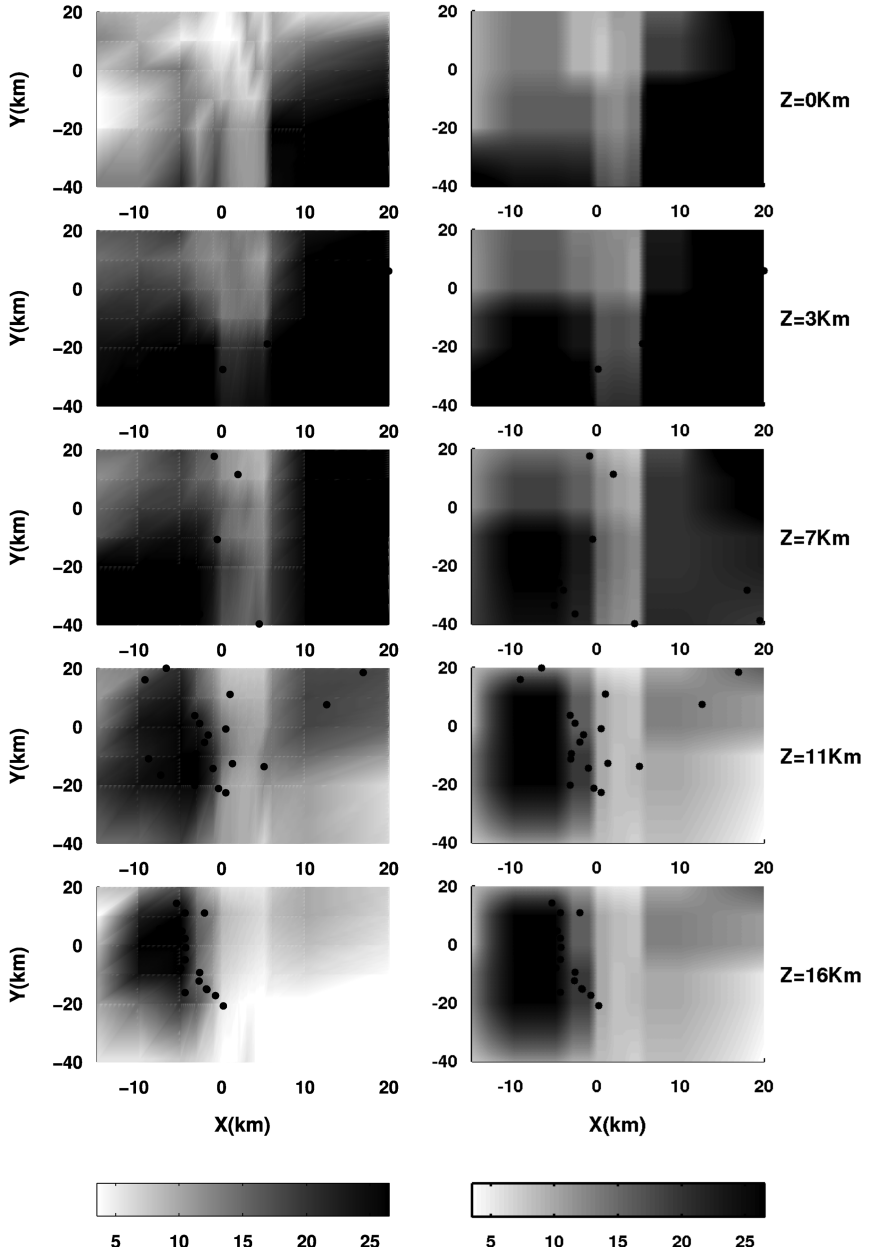

Figure 6. The DWS distribution in the space domain for (a) the standard DD tomography and (b) Haar wavelet-based DD tomography. In (b), the DWS values are calculated from the inverse wavelet transform of the sensitivity matrix in the wavelet domain whose small values are zeroed out by a hard thresholding.

is better recovered. For the final model, its absolute difference has a mean value of 0.303 km s^{-1}, a median value of 0.189 km s^{-1} and the standard deviation value of 0.087 km s^{-1}. This test shows that we can recover the large-scale features of the model by mainly using the approximation coefficients at the first few iterations and then gradually resolve the small scale features by adding the detail coefficients. In this way, the inversion may more likely avoid getting stuck in the local minimum and find a global minimum to better fit the data.

For the wavelet-based DD tomography method, another question is how to choose the appropriate wavelet function. The major properties of the wavelet function are its orthogonality, support, and the number of vanishing moments (Daubechies 1992). For our purpose, as seen in eqs (4) and (5), the wavelet basis function needs to be orthogonal or bi-orthogonal so that $\mathbf{W}^{-1} = \mathbf{W}^{T}$ or $\mathbf{W}^{-1} = \tilde{\mathbf{W}}^{\mathbf{T}}$. The Haar wavelet has the smallest support of only 2, which is appropriate when the number of grid nodes is small and the model structure is relatively simple. Depending on the model itself, to detect the singularities in the *j*th derivative, the wavelet function needs to be sufficiently regular with at least *j* vanishing moments (Daubechies 1992). For the Haar wavelet, the number of vanishing moments is one and can detect the first-order singularities. Therefore it is appropriate for the synthetic model because there are only eight inversion nodes in the both *Y* and *Z* directions and the model is constant in the *Z* direction and has discontinuities in the *X* direction. For the D4 wavelet, it has the support of 4 and the vanishing moment of 2. Compared to the Haar wavelet, the D4 wavelet has a slightly higher computational overhead. As shown in Fig. 4, the model inverted using the Haar wavelet in all three directions is better than that inverted using the D4 wavelet. The fault zone boundaries are

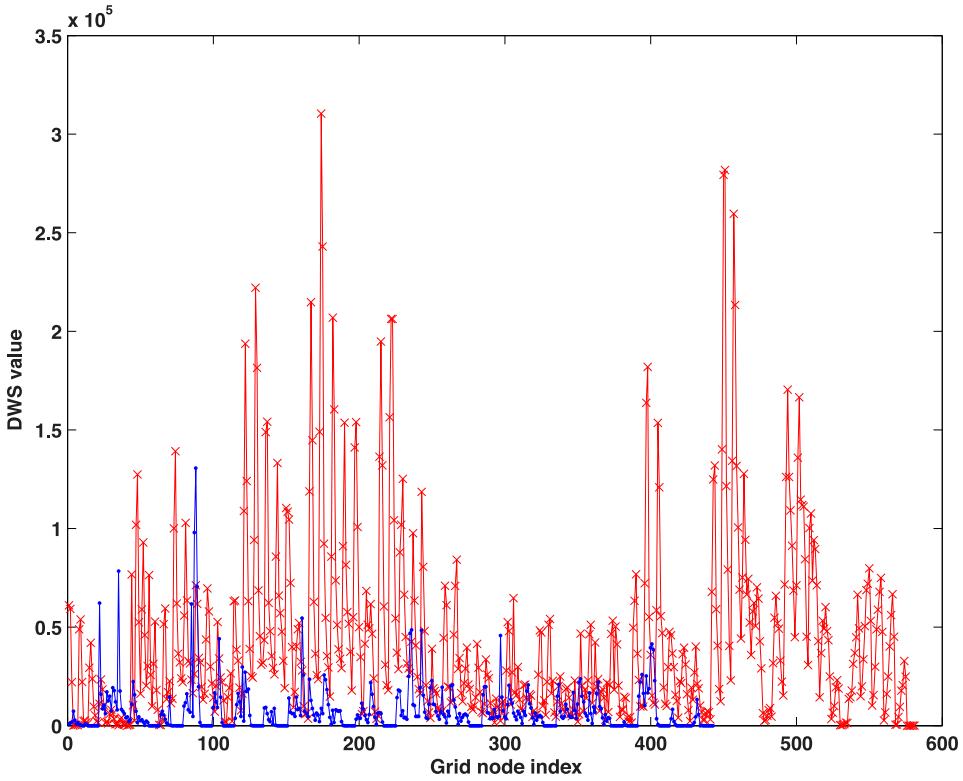

Figure 7. The average DWS values from all iterations of inversion for nodes corresponding to inverted (red) and zeroed out (blue) wavelet coefficients. The horizontal axis refers to the inversion node index in the space domain. The node index associated with wavelet coefficient is converted from the wavelet domain to the space domain.

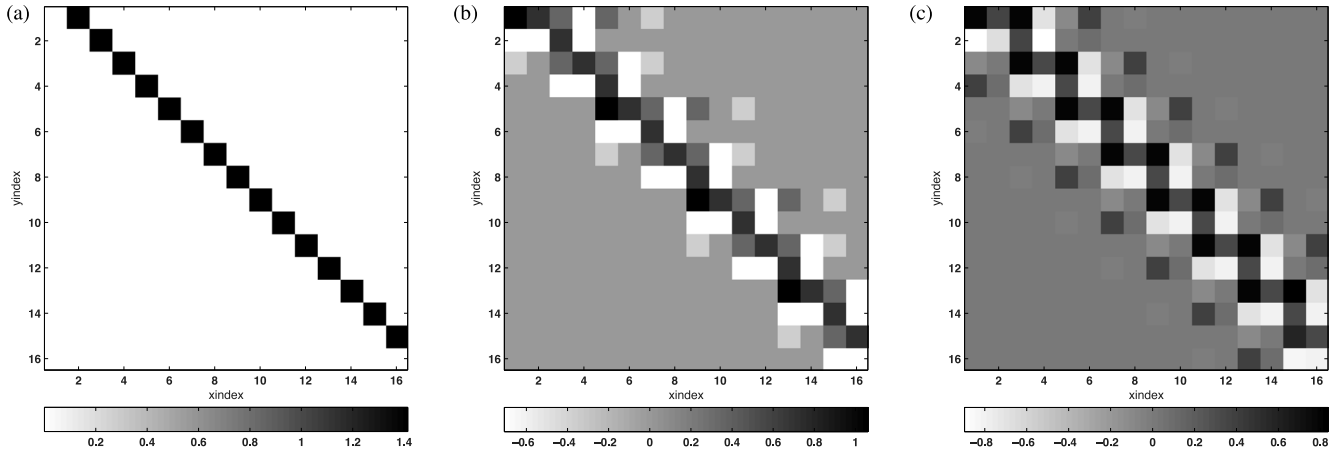

Figure 8. (a) A row of the sensitivity matrix in 2-D. The representation of the row of the sensitivity matrix in the wavelet domain with (b) the Haar wavelet and (c) the D4 wavelet.

somewhat smeared out and the low velocity anomalies are less accurate when using the D4 wavelet. This is because the Haar wavelet has the smallest support but is regular enough to characterize the synthetic model. For the wavelet function with a longer support, the 'fat ray' area in the wavelet domain is larger (Fig. 8), indicating that more wavelet coefficients are needed to resolve the model parameters for individual grid nodes. If some related wavelet coefficients cannot be resolved due to the poor ray sampling, the model will be smeared out because only larger scale features are recovered from the inverse wavelet transform. For most of the applications, the D4 wavelet may be more appropriate because it has a good balance between the support and the number of vanishing moments, and sometimes the model structures are too complicated for the Haar wavelet to represent.

5 CONCLUSIONS

To reduce the effect of the mismatch between the regular inversion nodes and ray distribution, we have developed a wavelet-based DD seismic tomography method with the sparsity constraint by applying the L_1-norm regularization to the model wavelet coefficients. The new method takes advantage of the multiscale property of the

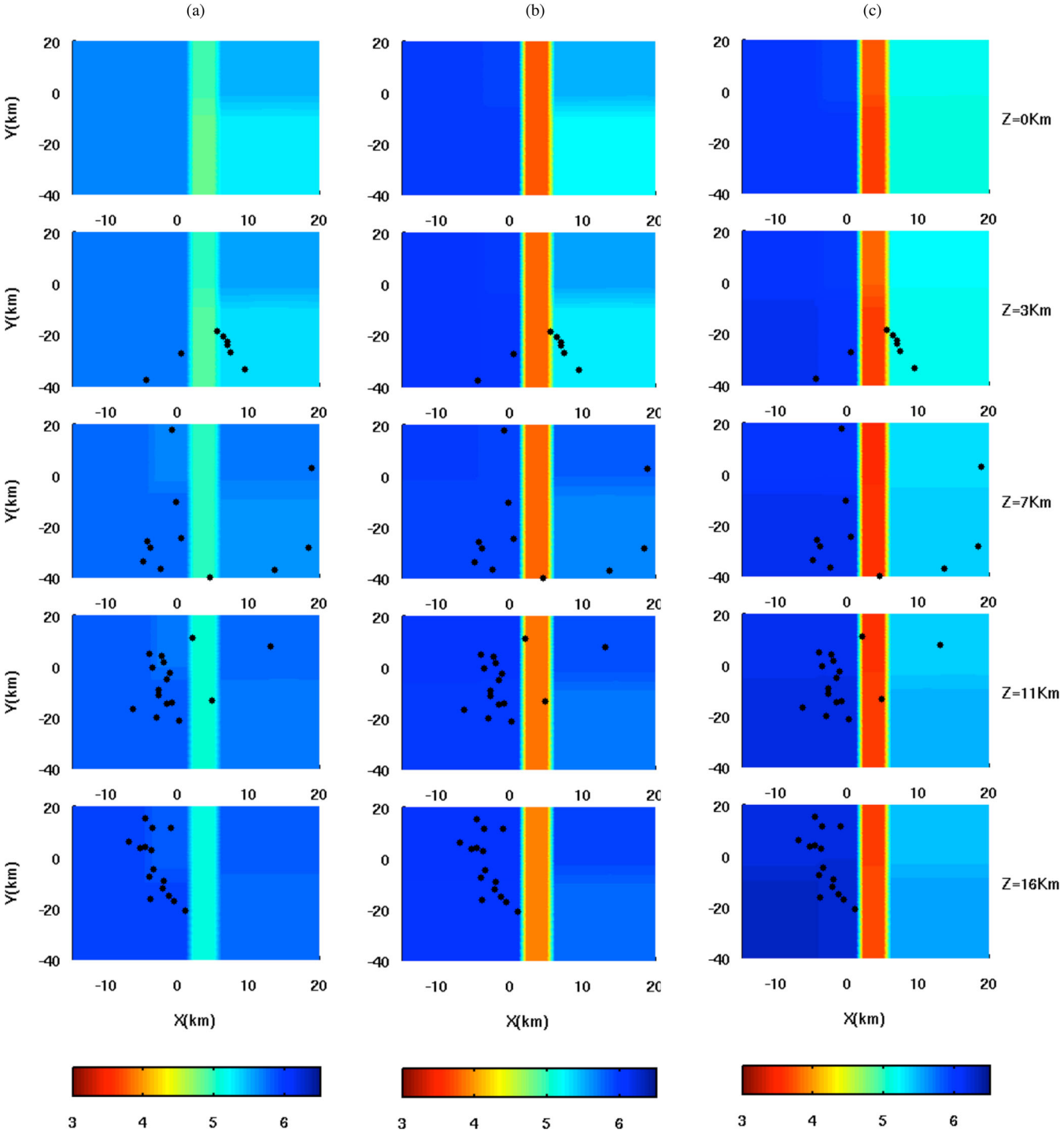

Figure 9. The tomographic result using just the approximation coefficients after (a) the first iteration, (b) the forth iteration and (c) the last iteration with the Haar wavelet by sparsity regularization.

wavelet representation and solves the model in the wavelet domain. Different scales of the model are implicitly recovered from the inverted sparse wavelet coefficients of the model at different scales. The new method is both data- and model-adaptive, providing a natural and easy way to resolve the model features from larger to smaller scales. Synthetic test shows the superior performance of the new method compared to the standard DD tomography of Zhang & Thurber (2003) on the 'vertical sandwich' velocity model. Both smooth and sharp features are better resolved without introducing blurring or other significant artefacts. The new method has the potential to be applied in various regions with highly uneven data distribution to get the fine scale features of the model where the data coverage is relatively good.

ACKNOWLEDGEMENTS

We are very grateful for the comments from Clifford Thurber and the two anonymous reviewers that help improve the paper. This research

is partly supported by the Natural Science Foundation of China under grant No. 41274055, the Fundamental Research Funds for the Central Universities under the grant number WK2080000053, and by China's State Administration of Foreign Experts Affairs International Partnership Program for Creative Research Teams.

REFERENCES

Abers, G.A. & Roecker, S.W., 1991. Deep structure of an arc-continent collision: earthquake relocation and inversion for upper mantle P and S wave velocities beneath Papua New Guinea, *J. geophys. Res.,* **96**(B4), 6379–6401.

Aki, K. & Lee, W.H.K., 1976. Determination of 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**(23), 4381–4399.

Aster, R.C., Borchers, B. & Thurber, C.H., 2013. *Parameter Estimation and Inverse Problems,* Academic Press.

Böhm, G., Galuppo, P. & Vesnaver, A., 2000. 3D adaptive tomography using Delaunay triangles and Voronoi polygons, *Geophys. Prospect.,* **48**(4), 723–744.

Candès, E.J., Romberg, J.K. & Tao, T., 2006. Stable signal recovery from incomplete and inaccurate measurements, *Commun. Pure Appl. Math.,* **59**(8), 1207–1223.

Chiao, L.Y. & Kuo, B.Y., 2001. Multiscale seismic tomography, *Geophys. J. Int.,* **145**(2), 517–527.

Chiao, L.Y. & Liang, W.T., 2003. Multiresolution parameterization for geophysical inverse problems, *Geophysics,* **68**(1), 199–209.

Curtis, A. & Snieder, R., 1997. Reconditioning inverse problems using the genetic algorithm and revised parameterization, *Geophysics,* **62**(5), 1524–1532.

Daubechies, I., 1992. *Ten Lectures on Wavelets,* Vol. 61, pp. 198–202, Society for Industrial and Applied Mathematics.

Delost, M., Virieux, J. & Operto, S., 2008. First arrival traveltime tomography using second generation wavelets, *Geophys. Prospect.,* **56**(4), 505–526.

Donoho, D.L., 2006. For most large underdetermined systems of linear equations the minimal L1 norm solution is also the sparsest solution, *Commun. Pure Appl. Math.,* **59**(6), 797–829.

Hung, S.H., Chen, W.P., Chiao, L.Y. & Tseng, T.L., 2010. First multi-scale, finite-frequency tomography illuminates 3-D anatomy of the Tibetan Plateau, *Geophys. Res. Lett.,* **37**(6), L06304, doi:10.1029/2009gl041875.

Hung, S.H., Chen, W.P. & Chiao, L.Y., 2011. A data-adaptive, multiscale approach of finite-frequency, traveltime tomography with special reference to P and S wave data from central Tibet, *J. geophys. Res.: Solid Earth (1978–2012),* **116**(B6), doi:10.1029/2010JB008190.

Husen, S. & Kissling, E., 2001. Local earthquake tomography between rays and waves: fat ray tomography, *Phys. Earth planet. Inter.,* **123**(2), 127–147.

Kissling, E., Ellsworth, W.L., Eberhart-Phillips, D. & Kradolfer, U., 1994. Initial reference models in local earthquake tomography, *J. geophys. Res.: Solid Earth (1978–2012),* **99**(B10), 19 635–19 646.

Loris, I., Nolet, G., Daubechies, I. & Dahlen, F.A., 2007. Tomographic inversion using $\ell 1$-norm regularization of wavelet coefficients, *Geophys. J. Int.,* **170**(1), 359–370.

Sambridge, M. & Guđmundsson, Ó., 1998. Tomographic systems of equations with irregular cells, *J. geophys. Res.,* **103**(B1), 773–781.

Scales, J.A., Gersztenkorn, A. & Treitel, S., 1988. Fast Lp solution of large, sparse, linear systems: application to seismic travel time tomography, *J. Comput. Phys.,* **75**(2), 314–333.

Scherzer, O., 1993. The use of Morozov's discrepancy principle for Tikhonov regularization for solving nonlinear ill-posed problems, *Computing,* **51**(1), 45–60.

Simons, F.J., Loris, I., Nolet, G., Daubechies, I.C., Voronin, S., Judd, J.S. & Vonesch, C., 2011. Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity, *Geophys. J. Int.,* **187**(2), 969–988.

Spakman, W. & Bijwaard, H., 2001. Optimization of cell parameterizations for tomographic inverse problems, *Pure appl. Geophys.*, **158,** 1401–1423.

Sweldens, W., 1996. The lifting scheme: a custom-design construction of biorthogonal wavelets, *Appl. Comput. Harmonic Anal.,* **3**(2), 186–200.

Thurber, C.H., 1983. Earthquake locations and three-dimensional crustal structure in the Coyote Lake Area, central California, *J. geophys. Res.,* **88**(B10), 8226–8236.

Thurber, C. & Eberhart-Phillips, D., 1999. Local earthquake tomography with flexible gridding, *Comput. Geosci.,* **25**(7), 809–818.

Vesnaver, A.L., 1996. Irregular grids in seismic tomography and minimum-time ray tracing, *Geophys. J. Int.,* **126**(1), 147–165.

Zelt, C.A., 1998. Lateral velocity resolution from three-dimensional seismic refraction data, *Geophys. J. Int.,* **135**(3), 1101–1112.

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. & Thurber, C. 2005. Adaptive mesh seismic tomography based on tetrahedral and Voronoi diagrams: application to Parkfield, California, *J. geophys. Res.,* **110**(B4), B04303, doi:10.1029/2004JB003186.

Zhang, H. & Thurber, C., 2006. Development and applications of double-difference seismic tomography, *Pure appl. Geophys.,* **163**(2–3), 373–403.

Zhou, H.W., 2003. Multiscale traveltime tomography, *Geophysics,* **68**(5), 1639–1649.

Zhu, X.X. & Bamler, R., 2010. Tomographic SAR inversion by L1 norm regularization—the compressive sensing approach, *IEEE Trans. Geosci. Remote Sens.,* **48**(10), 3839–3846.