Geophysical Journal International

Geophys. J. Int. (2015) **201,** 574–594
GJI Seismology

doi: 10.1093/gji/ggv047

Double-difference traveltime tomography with edge-preserving regularization and *a priori* interfaces

Youzuo Lin,[1] Ellen M. Syracuse,[1] Monica Maceira,[1] Haijiang Zhang[2] and Carene Larmat[1]

[1]*Los Alamos National Laboratory, Geophysics Group, MS* D452*, Los Alamos, NM* 87545*, USA. E-mail: ylin@lanl.gov*
[2]*School of Earth and Space Sciences, University of Science and Technology of China, Hefei, Anhui* 230026*, China*

Accepted 2015 January 27. Received 2015 January 26; in original form 2014 September 20

SUMMARY

Conventional traveltime seismic tomography methods with Tikhonov regularization (L_2 norm) typically produce smooth models, but these models may be inappropriate when subsurface structure contains discontinuous features, such as faults or fractures, indicating that tomographic models should contain sharp boundaries. For this reason, we develop a double-difference (DD) traveltime tomography method that uses a modified total-variation regularization scheme incorporated with *a priori* information on interfaces to preserve sharp property contrasts and obtain accurate inversion results. In order to solve the inversion problem, we employ an alternating minimization method to decouple the original DD tomography problem into two separate subproblems: a conventional DD tomography with Tikhonov regularization and a L_2 total-variation inversion. We use the LSQR linear solver to solve the Tikhonov inversion and the split-Bregman iterative method to solve the total-variation inversion. Through our numerical examples, we show that our new DD tomography method yields more accurate results than the conventional DD tomography method at almost the same computational cost.

Key words: Inverse theory; Tomography; Body waves; Seismic tomography; Computational seismology.

1 INTRODUCTION

Double-difference (DD) traveltime seismic tomography is an efficient tool to obtain subsurface velocity models using body-wave arrival time observations. Zhang & Thurber (2003, 2006) extend the DD location method of Waldhauser & Ellsworth (2000) to simultaneously solve for both velocity structure and seismic event locations. Recent developments are focused towards the inversion of different geophysical observations because they sense the subsurface medium differently according to their intrinsic sensitivities to the subsurface structure and composition. Maceira & Ammon (2009) combine surface wave dispersion and gravity observations into one simultaneous joint inversion to increase the resolution of shallow structures. In order to take advantage of complimentary data sets and improve seismic events locations, Zhang *et al.* (2014) develop a joint inversion method, which incorporates body-wave arrival times and surface wave dispersion data.

Due to limited ray path coverage and data noise, tomographic inversions are ill-posed and underdetermined, resulting in infinite models that are able to fit in the data to the desired degree, that is the solution to the inverse problem is not unique. To alleviate the ill-posedness, many numerical methods have been developed, including regularization techniques (e.g. Aster *et al.* 2005), the use of *a priori* information (Tarantola 1984; Delbos *et al.* 2006) and preconditioning methods (Berryman 1989).

Regularization is a commonly used method to stabilize ill-posed inverse problems. The most popular regularization method in geophysical inversions is L_2-norm regularization or Tikhonov regularization (Menke 1989; Tikhonov *et al.* 1995; Tarantola 2005). Also commonly used is the sparsity-promoting regularization, including L_1-norm regularizations such as total-variation (TV) regularization, the Huber function (Farquharson 2008; Lelièvre *et al.* 2009; Loris *et al.* 2010; Loris & Verhoeven 2012), and compressive sensing (Chiao & Kuo 2001; Hung *et al.* 2011). Inversion methods using Tikhonov regularization are usually computationally more efficient than inversions using TV regularization. However, because of the quadratic term used in the Tikhonov regularization, the inversions are biased towards producing smoother models, generating unwanted artefacts and degrading sharp interfaces.

Seismic inversion algorithms using conventional TV regularization promote sharp interfaces (Lelièvre *et al.* 2009; Loris *et al.* 2010). The disadvantage of using TV regularization is its instability and expensive computational costs. The instability arises from the non-differentiability of the TV functional at the origin. The coupling of the TV functional with the data misfit term makes it difficult to solve for gradient-based numerical optimization methods. One option to enable the differentiability of the TV functional is to add a small smoothing parameter to the TV functional (Lelièvre *et al.* 2009; Loris *et al.* 2010). Although an inversion with a TV functional and a smoothing parameter may be solvable, the convergence

is highly sensitive to the selection of the smoothing parameter, increasing the instability of the inversion. Another common approach to bypass the non-differentiability of the TV functional is to approximate the L_1 norm with a quadratic approximation (Nolet 1987). However, inversion methods based on the quadratic approximation, such as iteratively reweighted least squares method (IRLS), can diverge and be unstable due to division by zero or a very small value (Loris *et al.* 2010; Rodriguez 2014). A new solution has recently emerged to the problem with sharp transitions in the model from the use of neighbourhood algorithms allowing irregular distribution of grid nodes. The method has shown success for inversions involving receiver functions, for example, Sambridge (1999), where the size and number of layers are automatically determined by the samples themselves. The neighbourhood approximation is more efficient than classical L_2-norm methods relying on the calculation of the forward modelling for each sample but still suffers from the computation cost required by methods based on L_2-norm.

An alternative approach to reducing the ambiguity and improving the inversion results is to incorporate *a priori* information to guide the inversion, especially in areas of the model poorly resolved by the data set (Tarantola 1984). Ammon & Patton (1992) show the utility of including *a priori* information on spatial variations in geology by computing the difference between minimum roughness (Laplacian) smoothing and physiographic-based smoothing constraints when applied to tomography using a sparse set of phase velocity measurements from the western United States. Lelièvre & Oldenburg (2009a) and Li & Oldenburg (1996, 1998) constrain the smoothness of the inversion with respect to a given reference model. In the continued work of Li & Oldenburg (2000), they further incorporate dip information into inversions. Also from the work of Li & Oldenburg (2003), bound constraint is proved to be effective for the inversion. Chasseriau & Chouteau (2003) estimate a model covariance matrix to normalize the data fitting. Lelièvre & Oldenburg (2009b) include the structural orientation information from geological data and apply this technique to unstructured meshes (Lelièvre & Farquharson 2013). Most of these aforementioned references incorporate the *a priori* information to the original set up of the inversion. Wijns & Kowalczyk (2007) and Barbosa & Silva (2006) develop user interactive inversion approaches, which take user input to direct the inversion towards a geologically reasonable solution. A good summary of incorporating *a priori* information—especially from geological data—into seismic observations can be obtained in Lelièvre (2009).

In this paper, we develop an efficient DD traveltime tomographic technique with modified TV regularization and *a priori* interfaces. We first define the modified TV regularization by introducing an auxiliary variable and an extra regularization term to conventional TV regularization. With the help of the auxiliary variable, we decouple the TV regularization term from the data misfit term, thereby generating two subproblems: one of a conventional DD traveltime tomography problem using Tikhonov regularization and one of L_2-TV minimization. This significantly simplifies the original DD tomography. To further improve the inversions, we incorporate *a priori* information on interface locations, where available interface locations can be inferred from the geological data such as structural orientations, relative positions of rock units, and changes in geophysical properties across offset faults (Farquharson *et al.* 2008; Lelièvre 2009). The interface locations can be used as a guide to further constrain the inversion.

We develop a computational method for solving our new DD traveltime tomography problem with modified TV regularization and *a priori* interface information. We employ different solvers for the two decoupled subproblems. In particular, we apply the least squares technique with QR decomposition (LSQR; Paige & Saunders 1982a,b) to solve the first subproblem of conventional DD traveltime tomography with Tikhonov regularization. We use the Bregman iterative method (Osher *et al.* 2005; Goldstein & Osher 2009) to solve the second L_2-TV subproblem. There are two main benefits of our computational method. First, we avoid the use of a smoothing parameter in the TV functional, which significantly improves the stability and robustness of our inversion algorithm. Secondly, our algorithm converges to a true TV solution rather than an approximated TV solution, therefore increasing the inversion accuracy. Through an analysis of the computational cost, we show that the extra cost of our new method is trivial compared to the cost of conventional DD traveltime tomography with Tikhonov regularization.

We use body-wave traveltime data generated from a synthetic model to validate the improvement associated with our modified TV regularization and *a priori* interface technique. We demonstrate that our new method recovers velocities much more accurately and better preserves the sharpness of interfaces than those obtained from conventional DD traveltime tomography with Tikhonov regularization. We also provide numerical tests to demonstrate that our new method is a robust inversion algorithm with respect to the accuracy and availability of interface locations.

In the following sections, we first briefly describe the fundamentals of DD traveltime tomography (Section 2). We then introduce the DD traveltime tomography with a modified TV regularization technique incorporated with interface locations as *a priori* information (Section 3). Finally, in the last section, we apply our method to a synthetic data set and give the results (Section 4). Additionally, we also provide the specific computational methods for solving our new method (Section A) and analyse its cost (Section B) as two appendices.

2 THEORY

In this section, we first briefly describe the fundamentals of DD traveltime tomography and then provide the formulation of DD traveltime tomography in both matrix and optimization forms.

2.1 DD traveltime tomography

The DD traveltime tomography method (Zhang & Thurber 2003, 2006) simultaneously inverts for both 3-D seismic velocity models and seismic event locations. Using notation similar to Zhang & Thurber (2006), the DD traveltime tomography can be posed as

$$\begin{cases} r_k^i = (T_k^i)^{\text{obs}} - (T_k^i)^{\text{pred}} \\ r_k^i - r_k^j = (T_k^i - T_k^j)^{\text{obs}} - (T_k^i - T_k^j)^{\text{pred}}, \end{cases} \tag{1}$$

where T_k^i is the arrival time from earthquake i to seismic station k, and T_k^j is the arrival time of a seismic phase from earthquake j recorded at seismic station k. The terms with the superscript 'obs' indicate observations and those with the superscript 'pred' indicate predicted data. r_k^i quantifies the misfit between the observed and predicted arrival times, and $r_k^i - r_k^j$ is the difference between the observed and predicted differential arrival time for earthquakes i and j, the so-called 'double-difference'.

A matrix formulation of eq. (1) is

$$\begin{bmatrix} G_h^T & G_v^T \end{bmatrix} \begin{bmatrix} \delta \mathbf{h} \\ \delta \mathbf{m}_v \end{bmatrix} = \mathbf{d}^T, \tag{2}$$

where G_h^T is the sensitivity matrix for the hypocentre, denoted by $\mathbf{h}$; G_v^T is the sensitivity matrix for the velocity model $\mathbf{m}_v$; and $\mathbf{d}^T$ is the data vector including both absolute and differential data.

To differentiate the *P*-arrivals from the *S*-arrivals, we decompose the model parameters as $\delta\mathbf{m}_v = [\delta\mathbf{m}_p\ \delta\mathbf{m}_s]$. Therefore, eq. (2) can be reformulated as

$$\begin{bmatrix} w_1 G_h^{T_p} & w_1 G_{v_p}^{T_p} & 0 \\ w_2 G_h^{T_s} & 0 & w_2 G_{v_s}^{T_s} \end{bmatrix} \begin{bmatrix} \delta\mathbf{h} \\ \delta\mathbf{m}_p \\ \delta\mathbf{m}_s \end{bmatrix} = \begin{bmatrix} w_1 \mathbf{d}^{T_p} \\ w_2 \mathbf{d}^{T_s} \end{bmatrix}, \tag{3}$$

where $G_h^{T_p}$, $G_h^{T_s}$, $G_{v_p}^{T_p}$, $G_{v_s}^{T_s}$ are the sensitivity matrices of the *P*- and *S*-arrivals with respect to hypocentre parameters ($\mathbf{h}$), compressional-wave velocities v_p and shear wave velocities v_s; $\delta\mathbf{h}$ are perturbations to hypocentre locations and times, $\delta\mathbf{m}_p$ and $\delta\mathbf{m}_s$ are perturbations to v_p and v_s models; w_1 and w_2 are weights for the *P*- and *S*-arrival data, respectively.

As noted by Zhang & Thurber (2006), DD tomography produces more accurate event locations and velocity models than standard velocity tomography, which only uses absolute arrival times. In order to further improve the inversion accuracy and stability, regularization techniques have been applied to eq. (3). In Zhang & Thurber (2003, 2006), Tikhonov regularization is incorporated with the DD traveltime tomography; the inversion results therein demonstrate the importance of regularization to traveltime tomography.

2.2 DD traveltime tomography with the Tikhonov regularization scheme

Tikhonov regularization is the most commonly used regularization technique for inverse problems, and it has been used widely in traveltime tomography. DD tomography using the Tikhonov regularization scheme can be posed as (Zhang & Thurber 2003; Zhang *et al.* 2014):

$$\begin{bmatrix} w_1 G_h^{T_p} & w_1 G_{v_p}^{T_p} & 0 \\ w_2 G_h^{T_s} & 0 & w_2 G_{v_s}^{T_s} \\ w_h L_h & 0 & 0 \\ 0 & w_p L_{v_p} & 0 \\ 0 & 0 & w_s L_{v_s} \\ \lambda_h I & 0 & 0 \\ 0 & \lambda_p I & 0 \\ 0 & 0 & \lambda_s I \end{bmatrix} \begin{bmatrix} \delta\mathbf{h} \\ \delta\mathbf{m}_p \\ \delta\mathbf{m}_s \end{bmatrix} = \begin{bmatrix} w_1 \mathbf{d}^{T_p} \\ w_2 \mathbf{d}^{T_s} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{4}$$

where L_h, L_{v_p} and L_{v_s} are the first-order smoothing matrices for $\mathbf{h}$, $\mathbf{m}_p$ and $\mathbf{m}_s$ with weights of w_h, w_p and w_s; I is the identity matrix weighted by λ_h, λ_p and λ_s.

By rewriting this in an optimization form, we have an equivalent minimization problem

$$E(\tilde{\mathbf{m}}) = \min_{\tilde{\mathbf{m}}} \left\{ \|\tilde{G}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}\|_2^2 + \|WL\tilde{\mathbf{m}}\|_2^2 + \|\Lambda\tilde{\mathbf{m}}\|_2^2 \right\}, \tag{5}$$

where

$$\tilde{G} = \begin{bmatrix} w_1 G_h^{T_p} & w_1 G_{v_p}^{T_p} & 0 \\ w_2 G_h^{T_s} & 0 & w_2 G_{v_s}^{T_s} \end{bmatrix},$$

$$\tilde{\mathbf{d}} = \begin{bmatrix} w_1 \mathbf{d}^{T_p} \\ w_2 \mathbf{d}^{T_s} \end{bmatrix},$$

$W = \mathrm{diag}(w_h,\ w_p,\ w_s)$, $L = \mathrm{diag}(L_h, L_{v_p}, L_{v_s})$, $\tilde{\mathbf{m}} = (\delta\mathbf{h}^T,\ \delta\mathbf{m}_p^T,\ \delta\mathbf{m}_s^T)^T$ and $\Lambda = \mathrm{diag}(\lambda_h,\ \lambda_p,\ \lambda_s)$.

In eq. (5), the data misfit term $\|\tilde{G}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}\|_2^2$ measures the misfit of the predicted data to the observed data. The second term, $\|WL\tilde{\mathbf{m}}\|_2^2$, quantifies the roughness of the inverted model parameters, and the third term, $\|\Lambda\tilde{\mathbf{m}}\|_2^2$, is the model norm, which penalizes the magnitude of the inverted model parameters. The parameters in W and Λ are the regularization parameters. They balance the weights between the data misfit term of $\|\tilde{G}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}\|_2^2$ and the regularization terms of $\|WL\tilde{\mathbf{m}}\|_2^2$ and $\|\Lambda\tilde{\mathbf{m}}\|_2^2$. Using incorrect values of the regularization parameters might lead to a failure in solving eq. (5). For instance, if both W and Λ are too small, minimizing eq. (5) might yield to a solution that overfits the data, which can generate some unwanted artefacts in the inversion. On the other hand, if the regularization parameters are too large, the solution will be over-regularized and will generate significantly biased inversions towards the specific regularization imposed.

For the simplicity of derivation, we use the optimization form of eq. (5) instead of the matrix form of eq. (4). To numerically solve the linear system in eq. (4) or the minimization problem in eq. (5), LSQR is used, as it is appropriate for solving large-scale sparse ill-posed inverse problems (Paige & Saunders 1982a,b).

The L_2-norm based regularization terms in eq. (5), $\|WL\tilde{\mathbf{m}}\|_2^2$ and $\|\Lambda\tilde{\mathbf{m}}\|_2^2$, are usually referred to as Tikhonov regularization (Tikhonov *et al.* 1995). The DD traveltime tomography with the Tikhonov regularization scheme is effective for velocity anomalies with quadratic interfaces. For piecewise constant velocity models, the DD traveltime tomography with Tikhonov regularization unfortunately may introduce artefacts and thus smooth the sharp interfaces in the velocity models (Loris *et al.* 2010). In order to preserve sharp interfaces, we develop a DD traveltime tomography method using TV regularization schemes.

3 DD TRAVELTIME TOMOGRAPHY USING TOTAL-VARIATION REGULARIZATION TECHNIQUES

In this section, we first provide DD traveltime tomography with a conventional TV regularization scheme. Then we provide our new DD traveltime tomography with a modified total-variation regularization scheme and extend it by incorporating *a priori* interface information.

3.1 DD traveltime tomography with a conventional total-variation regularization scheme

The TV regularization term for a 2-D model is defined as (Rudin *et al.* 1992)

$$\|\mathbf{m}\|_{\mathrm{TV}} = \|\nabla\mathbf{m}\|_1 = \sum_{1\le i,j\le n} \sqrt{|(\nabla_x\mathbf{m})_{i,j}|^2 + |(\nabla_z\mathbf{m})_{i,j}|^2}, \tag{6}$$

where $(\nabla_x\mathbf{m})_{i,j} = m_{i+1,j} - m_{i,j}$ and $(\nabla_z\mathbf{m})_{i,j} = m_{i,j+1} - m_{i,j}$ are the spatial derivatives at a spatial gridpoint (i, j) in a Cartesian coordinate system (x, z). The DD traveltime tomography with conventional TV regularization can be therefore posed as

$$E(\tilde{\mathbf{m}}) = \min_{\tilde{\mathbf{m}}} \left\{ \|\tilde{G}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}\|_2^2 + \eta\|\nabla\tilde{\mathbf{m}}\|_1 \right\}, \tag{7}$$

where η is the regularization parameter and the roughness term is a TV regularization term.

Due to the non-differentiability of the TV regularization term at the origin, various numerical methods have been developed to approximate the TV regularization term to enable its differentiability at the origin. One common method is to introduce a small-valued smoothing parameter ϵ such that

$$\|\mathbf{m}\|_{\mathrm{TV},\epsilon} = \sum_{1\le i,j\le n} \sqrt{|(\nabla_x \mathbf{m})_{i,j}|^2 + |(\nabla_z \mathbf{m})_{i,j}|^2 + \epsilon}, \tag{8}$$

where the gradient and the Hessian can now be computed numerically (Vogel 2002). Another popular approach to bypass the non-differentiability of the TV regularization term is to approximate the L_1 norm with a quadratic approximation (Nolet 1987),

$$\|\nabla \mathbf{m}\|_1 \approx \|\Omega D\mathbf{m}\|_2^2, \tag{9}$$

where the operator D is a differential operator and $\Omega = \mathrm{diag}([(D_x\mathbf{m})^2 + (D_y\mathbf{m})^2]^{-0.5})$.

However, a direct numerical method of DD traveltime tomography using the approximated TV regularization term in eq. (8) or eq. (9) can be difficult. The cause is the numerical instability of the approximation of the TV regularization term. Specifically, for the methods based on a smooth approximation of TV regularization such as lagged diffusivity fixed point iteration (Vogel 2002), the convergence of the inversion is highly sensitive to the smoothing parameter ϵ. As pointed out in Loris *et al.* (2010) and Rodriguez (2014), numerical methods based on the quadratic approximation, such as the iteratively reweighted least squares method (IRLS), can diverge and be unstable due to division by zero or a small value. In consideration of these, we add a modified TV regularization scheme to overcome these issues.

3.2 DD traveltime tomography with a modified total-variation regularization scheme

Our new DD traveltime tomography with a modified TV (MTV) regularization scheme is given by

$$E(\tilde{\mathbf{m}}, \tilde{\mathbf{u}}) = \min_{\tilde{\mathbf{m}},\tilde{\mathbf{u}}} \left\{ \|\tilde{G}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}\|_2^2 + \mu\|\tilde{\mathbf{m}} - \tilde{\mathbf{u}}\|_2^2 + \eta\|\nabla\tilde{\mathbf{u}}\|_1 \right\}, \tag{10}$$

where an auxiliary variable $\tilde{\mathbf{u}}$ and an extra Tikhonov regularization term $\|\tilde{\mathbf{m}} - \tilde{\mathbf{u}}\|_2^2$ are added, and μ is a regularization parameter. Equivalently, eq. (10) can be written as (Huang *et al.* 2008; Ye *et al.* 2011)

$$E(\tilde{\mathbf{m}}, \tilde{\mathbf{u}}) = \min_{\tilde{\mathbf{u}}} \left\{ \min_{\tilde{\mathbf{m}}} \left\{ \|\tilde{G}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}\|_2^2 + \mu\|\tilde{\mathbf{m}} - \tilde{\mathbf{u}}\|_2^2 \right\} + \eta\|\nabla\tilde{\mathbf{u}}\|_1 \right\}. \tag{11}$$

The regularization parameters Λ and μ control the trade-off between the data misfit term and the two Tikhonov regularization terms, while the regularization parameter η controls how much of the interface is preserved in the inversion.

We employ an alternating-minimization algorithm to solve the double minimization problem in eq. (11). Using an initial model $\tilde{\mathbf{u}}^{(0)} = \tilde{\mathbf{m}}^{(0)}$, solving eq. (11) leads to the solutions of two minimization subproblems:

$$\tilde{\mathbf{m}}^{(k)} = \underset{\tilde{\mathbf{m}}}{\mathrm{argmin}} \{E_1(\tilde{\mathbf{m}})\} = \underset{\tilde{\mathbf{m}}}{\mathrm{argmin}} \left\{ \|\tilde{G}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}\|_2^2 + \mu\|\tilde{\mathbf{m}} - \tilde{\mathbf{u}}^{(k)}\|_2^2 \right\} \tag{12}$$

$$\tilde{\mathbf{u}}^{(k)} = \underset{\tilde{\mathbf{u}}}{\mathrm{argmin}} \{E_2(\tilde{\mathbf{u}})\} = \underset{\tilde{\mathbf{u}}}{\mathrm{argmin}} \left\{ \|\tilde{\mathbf{m}}^{(k)} - \tilde{\mathbf{u}}\|_2^2 + \eta\|\nabla\tilde{\mathbf{u}}\|_1 \right\}, \tag{13}$$

for iteration step $k = 1, 2, \ldots$. The two minimization problems in eqs (12) and (13) have distinct physical meanings. The subproblem in eq. (12) is similar to the problem in eq. (5). We solve for model parameters of $\tilde{\mathbf{m}}^{(k)}$ by minimizing the data misfit term with the constraints of a Tikhonov regularization term, where a reference model $\tilde{\mathbf{u}}^{(k)}$ is introduced as *a priori* information. Eq. (13) is a standard L_2-TV minimization problem (Wohlberg & Rodríguez 2007; Micchelli *et al.* 2013), which seeks a reference model $\tilde{\mathbf{u}}^{(k)}$ with a sharpened interface with the help of the TV regularization term.

There are two major benefits of using the MTV regularization scheme in eq. (10). The first benefit is that we reduce the complexity of the minimization problem in eq. (7). The separation of the TV regularization term from the data misfit term enables the solution of two simpler minimization problems posed in eqs (12) and (13). The second benefit is that this technique is much more robust than solving the inversion scheme using eq. (7) with approximations in eqs (8) or (9) (Huang *et al.* 2008; Ye *et al.* 2011). This is based on the following reasons. Conventional optimization methods for solving the minimization problem in eq. (7) rely on the gradient of approximated TV functional (Nolet 1987; Vogel 2002). Numerical instability may arise during the iteration as pointed out in Rodriguez (2014) and Loris *et al.* (2010). Notice, however, that solving for $\mathbf{u}^{(k)}$ in eq. (13) is a standard L_2-TV minimization. Efficient gradient-free optimization methods exist for L_2-TV minimization problems provided with a unitary system matrix, which means the system matrix satisfies

$$\tilde{G}^T\tilde{G} = \tilde{G}\tilde{G}^T = I, \tag{14}$$

where I is the identity matrix and $(\,\cdot\,)^T$ is the matrix transpose operator. By using the properties of a unitary matrix, solving for a L_2-TV minimization relies on a shrinkage operation, which is defined as (Elad 2010)

$$S(y, v) = \mathrm{sgn}(y) \cdot \max\{|y| - v, 0\}, \tag{15}$$

where v is the threshold value and $\mathrm{sgn}(y)$ takes the sign of the value y. This can be rather efficient and robust. The system matrix of our problem in eq. (13) is an identity matrix, which is unitary, and therefore it can be solved efficiently and robustly. This is generally not true for the DD traveltime tomography with conventional TV regularization in eq. (7).

3.3 DD traveltime tomography using *a priori* interface information

The minimization problem in eq. (10) seeks a subsurface model that best fits the geophysical data. In some applications, both geophysical and geological data can be obtained for a region, and the geological data can be incorporated as constraints into the inversion. Subsurface structural information or the location of known interfaces based on discontinuities in geophysical properties can be utilized. For instance, we may know that a particular rock type lies above another type, or we may know the location of a fault. Such information is available in independent data sets such as gravity, surface geological maps, and even topography. Lelièvre & Farquharson (2013) and Lelièvre & Oldenburg (2009b) incorporate subsurface structural orientation information into inversions with the help of a rotated gradient operator, which improves the resulting inversion. In this work, we focus on *a priori* interface information and impose it as a constraint for our inversion problem. We therefore modify the MTV regularization scheme given in eq. (10) to incorporate interface information by adding a weighting parameter w, such

that

$$E(\tilde{\mathbf{m}}, \tilde{\mathbf{u}}) = \min_{\tilde{\mathbf{m}},\tilde{\mathbf{u}}} \left\{ \|\tilde{G}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}\|_2^2 + \mu\|\tilde{\mathbf{m}} - \tilde{\mathbf{u}}\|_2^2 + \eta\|w\,\nabla\tilde{\mathbf{u}}\|_1 \right\}, \quad (16)$$

where the weighting parameter w controls the amount of regularization among adjacent spatial gridpoints. We set the weighting value to be:

$$w_{i,j} = \begin{cases} 0 & \text{if point (i, j) is on the interface} \\ 1 & \text{if point (i, j) is off the interface} \end{cases}. \quad (17)$$

By assigning a zero weight to the node adjacent to an interface, that node is not penalized by the regularization. The weighting parameter w therefore varies according to the interface location.

Analogous to eqs (11)–(13), we incorporate interface information in eq. (16) as

$$E(\tilde{\mathbf{m}}, \tilde{\mathbf{u}}) = \min_{\tilde{\mathbf{u}}} \left\{ \min_{\tilde{\mathbf{m}}} \left\{ \|\tilde{G}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}\|_2^2 + \mu\|\tilde{\mathbf{m}} - \tilde{\mathbf{u}}\|_2^2 \right\} + \eta\|w\,\nabla\tilde{\mathbf{u}}\|_1 \right\}, \quad (18)$$

which can be solved by an alternating minimization method:

$$\tilde{\mathbf{m}}^{(k)} = \operatorname*{argmin}_{\tilde{\mathbf{m}}} \{E_1(\tilde{\mathbf{m}})\} = \operatorname*{argmin}_{\tilde{\mathbf{m}}} \left\{ \|\tilde{G}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}\|_2^2 + \mu\|\tilde{\mathbf{m}} - \tilde{\mathbf{u}}^{(k)}\|_2^2 \right\} \quad (19)$$

$$\tilde{\mathbf{u}}^{(k)} = \operatorname*{argmin}_{\tilde{\mathbf{u}}} \{E_2(\tilde{\mathbf{u}})\} = \operatorname*{argmin}_{\tilde{\mathbf{u}}} \left\{ \|\tilde{\mathbf{m}}^{(k)} - \tilde{\mathbf{u}}\|_2^2 + \eta\,\|w\,\nabla\tilde{\mathbf{u}}\|_1 \right\}. \quad (20)$$

The interleaving solving of the two subproblems in eqs (19) and (20) leads to an inversion that both improves the minimization of the data misfit and enhances the sharpness of velocity interfaces.

In the appendices, we provide the detailed computational methods for solving these two minimization subproblems and compare the computational costs of solving eqs (19) and (20) in comparison to that of conventional DD traveltime tomography.

3.4 Selection of the regularization parameters

Regularization parameters play an important role in obtaining accurate inversion results. In our DD traveltime tomography technique with MTV regularization and *a priori* interfaces, we have two regularization parameters that need to be appropriately selected: μ in eq. (19) and η in eq. (20).

The regularization parameter μ is used as part of the Tikhonov regularization. Many methods have been developed to suitably estimate its value (Hansen 1998; Vogel 2002), and we use the L-curve method (Hansen & O'Leary 1993) due to its simplicity and efficiency. The L-curve is a log–log plot of the norm of a regularized solution versus the norm of the corresponding residual norm. It is a graphical tool for displaying the trade-off between the size of a solution and its fit to the input data for a set of regularization parameters. The plot often results in the shape of letter L. The best regularization parameter lies at the corner of the L shape.

The regularization parameter η is part of the L_2-TV regularization problem. Lin *et al.* (2010) showed that the unbiased predictive risk estimator (UPRE) method can robustly and accurately estimate η for the L_2-TV regularization problem. The UPRE method is based on a statistical estimator of the mean squared norm of the predictive error. The best regularization parameter therefore yields the minimum predictive error. We apply this method here.

4 NUMERICAL RESULTS

We use synthetic seismic data calculated for a known velocity model to validate the improvement in results obtained using our new DD traveltime tomography algorithm. We first provide a test to compare the numerical results using Tikhonov and MTV regularization with/without the *a priori* interface information. To further validate the robustness of our DD traveltime tomography algorithm on the *a priori* interface information, we provide numerical tests constrained by incorrect or limited interface information. We also assess the robustness of our method given a sparse data coverage through a numerical test.

In our synthetic velocity model, a pair of vertical north–south velocity interfaces are added to an otherwise 1-D background velocity model discretized in 3-D, in which velocities in the central part of the model are reduced by 0.25 km s^{-1} from the background model at all depths, and velocities outside of this central region are increased by 0.25 km s^{-1} from the background model, as shown in Fig. 1. As a result, there is a sharp gradient on the eastern and western edge of this central region. In order to avoid the overly regular data sampling, which may occur with an entirely synthetic data distribution, we use the events and stations from actual seismicity at Kilauea Volcano, Hawaii, recorded by the Hawaii Volcano Observatory in 2007 to construct the synthetic data set (Syracuse *et al.* 2010). The actual ray path coverage for 2665 earthquakes and 116 stations are used, as shown in Fig. 1. Gaussian noise is added to the synthetic arrival times. The differential arrival times are obtained by directly subtracting absolute arrival times from pairs of neighbouring events recorded at common stations.

The velocity model is synthetic, although it represents some realistic situations. The sharp velocity perturbation can represent faults or other low-velocity zones. An example of this is the Hayward fault in California, USA (Zhang & Thurber 2003). We discretize the velocity model using 13 grid nodes in the north-south direction (19.15°N, 19.25–19.45°N with 0.02° spacing, 19.55°N) and 22 gridpoints in the west–east direction (155.50°W, 155.40°W, 155.32–154.98°W with 0.02° spacing, 154.90°W, 154.80°W), based on the distribution of the earthquakes and stations. The grid nodes are between 1.0 and 15.0 km in depth, with 2 km spacing. As the depth increases, the velocity also increases. For the ease of the illustration, we present the shallowest six slices in our numerical results.

4.1 Comparison of numerical results using the Tikhonov and MTV regularization schemes

We first compare the results of our DD traveltime tomography using MTV regularization and *a priori* interface information (TomoDDEMTV) with inversions using Tikhonov regularization (TomoDDTK) and using MTV regularization (TomoDDMTV).

To implement the TomoDDTK method, we use LSQR to solve for the linear systems in eq. (4). We implement the TomoDDMTV method by using Algorithm 2 provided in the Appendix B, and setting the weighting parameter $w = 1.0$. We implement TomoDDEMTV by using Algorithm 2 incorporated with the interface information. The regularization parameters are selected according to the methods described in Section 3.4. Six iterations of simultaneous inversions are performed for each method.

Fig. 2 shows the inverted *P*-wave velocity models resulting from each method. *S*-wave velocities are also solved for in the inversion, but are not discussed here. Fig. 3 shows the differences between each recovered model and the synthetic model. We also illustrate the ray path coverage at each depth, as indicated by the derivative

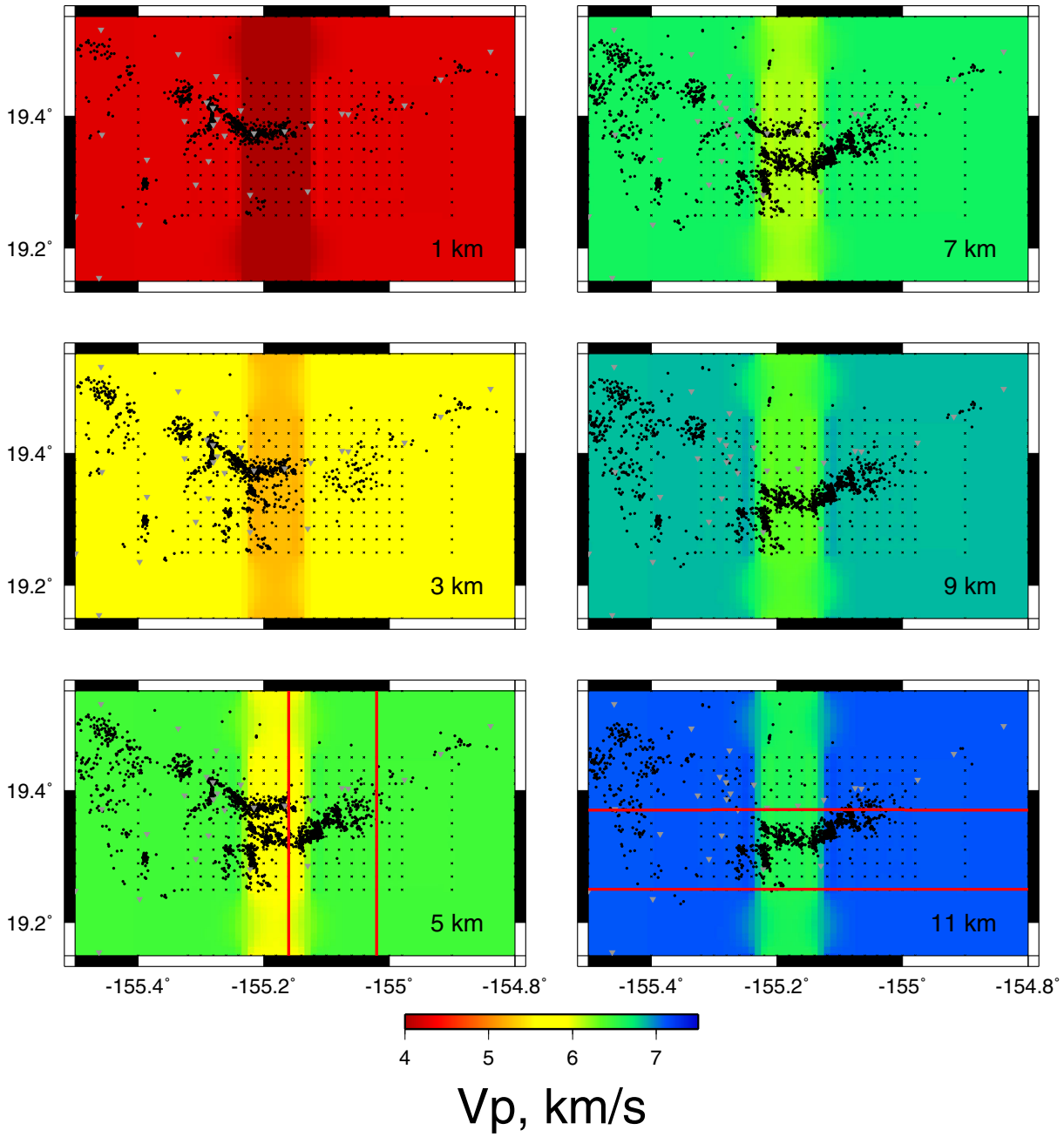

Figure 1. Map views of the true synthetic velocity model. Black circles indicate the location of the 2665 earthquakes considered in our inversion tests. Grey inverted triangles indicate the locations of the 116 seismic stations considered. The red lines indicate the locations of profiles discussed in the text.

weighted sum (DWS) in both Figs 2 and 3 (Thurber & Eberhart-Phillips 1999).

These comparisons demonstrate that there are some short-wavelength artefacts in the TomoDDTK results, and the synthetic velocity interfaces are poorly recovered in this inversion. The TomoDDMTV inversion reduces the inversion noise and artefacts in comparison to the TomoDDTK results, although the velocity interfaces are still not very well recovered. The TomoDDEMTV results show that this method eliminates most of the artefacts in the TomoDDTK and TomoDDMTV results, and the velocity interfaces are well recovered. By including the interface information in the inversion, the recovered model has been significantly improved in some areas of poor ray coverage, such as the southeastern area of the model (Fig. 2). In the absence of ray coverage where an interface is imposed as *a priori* information, the TomoDDEMTV inversion favours extending the velocities recovered at neighbouring well-sampled portions of the velocity interface to poorly sampled parts of the model. To illustrate the results of each method in Fig. 4 we compare profiles of the velocity through each recovered model at locations shown in Fig. 1. The locations of the earthquake events are indicated in Fig. 1. The TomoDDEMTV method yields the most accurately recovered velocities of the three tested methods; it produces the fewest oscillations. To quantify the accuracy of these methods, we calculate the inversion error (in m s^{-1}) to the true model in Fig. 1 by

$$\text{Inversion error} = \frac{\sqrt{\sum\sum_{(i,j)\in\text{nodes}} \left((\mathbf{m}_{(i,j)})^{\text{pred}} - (\mathbf{m}_{(i,j)})^{\text{true}}\right)^2}}{\text{Number of nodes}} \times 1000, \quad (21)$$

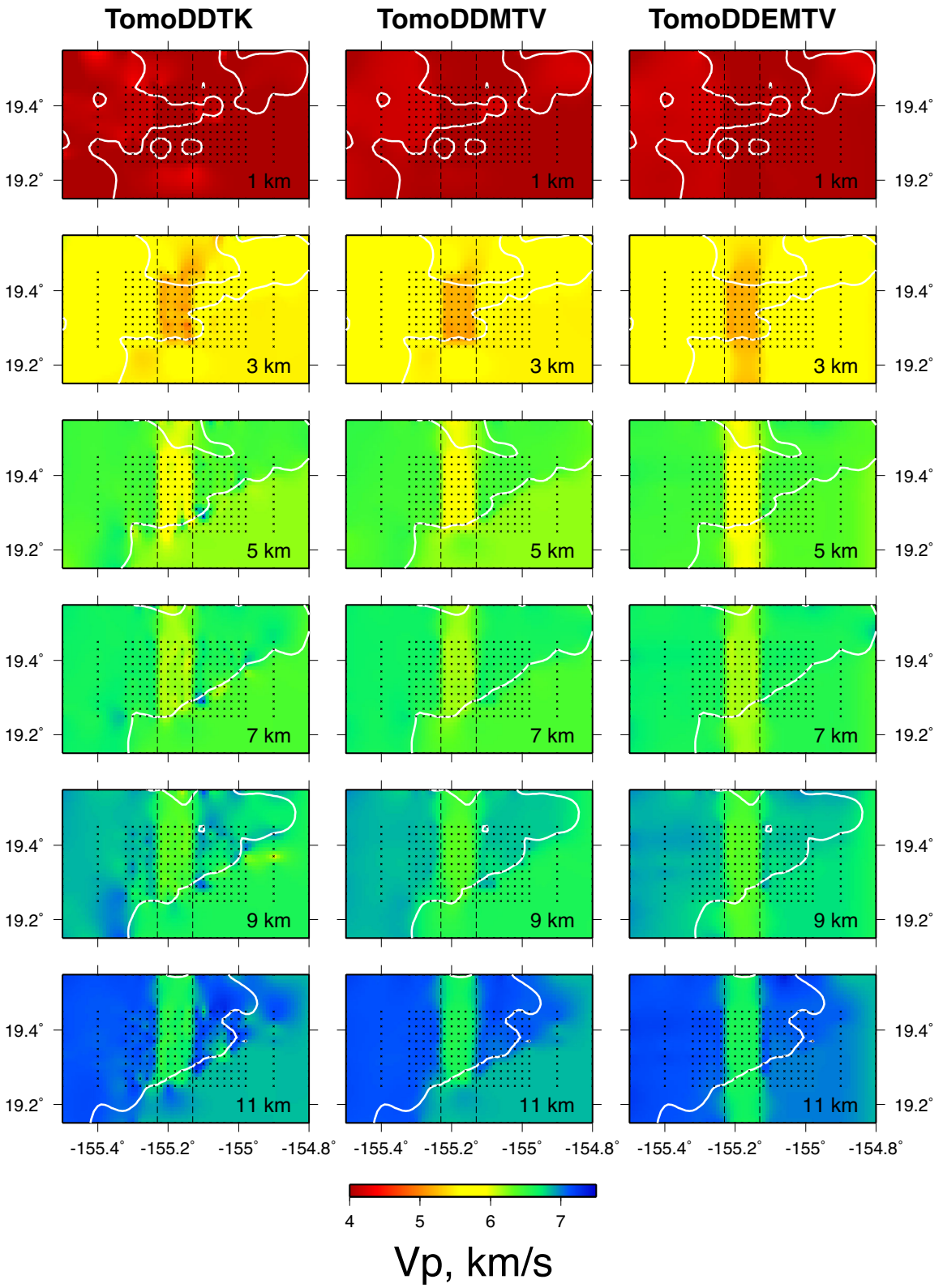

Figure 2. Inversion results produced using Tikhonov regularization (left-hand column), MTV regularization (middle column), and MTV regularization with *a priori* interfaces (right-hand column). The dashed black lines indicate the true locations of the interfaces. The white lines show the area that is well sampled based on DWS values. The MTV regularization with *a priori* interfaces yields a more accurate result than those obtained using Tikhonov or MTV regularization methods.

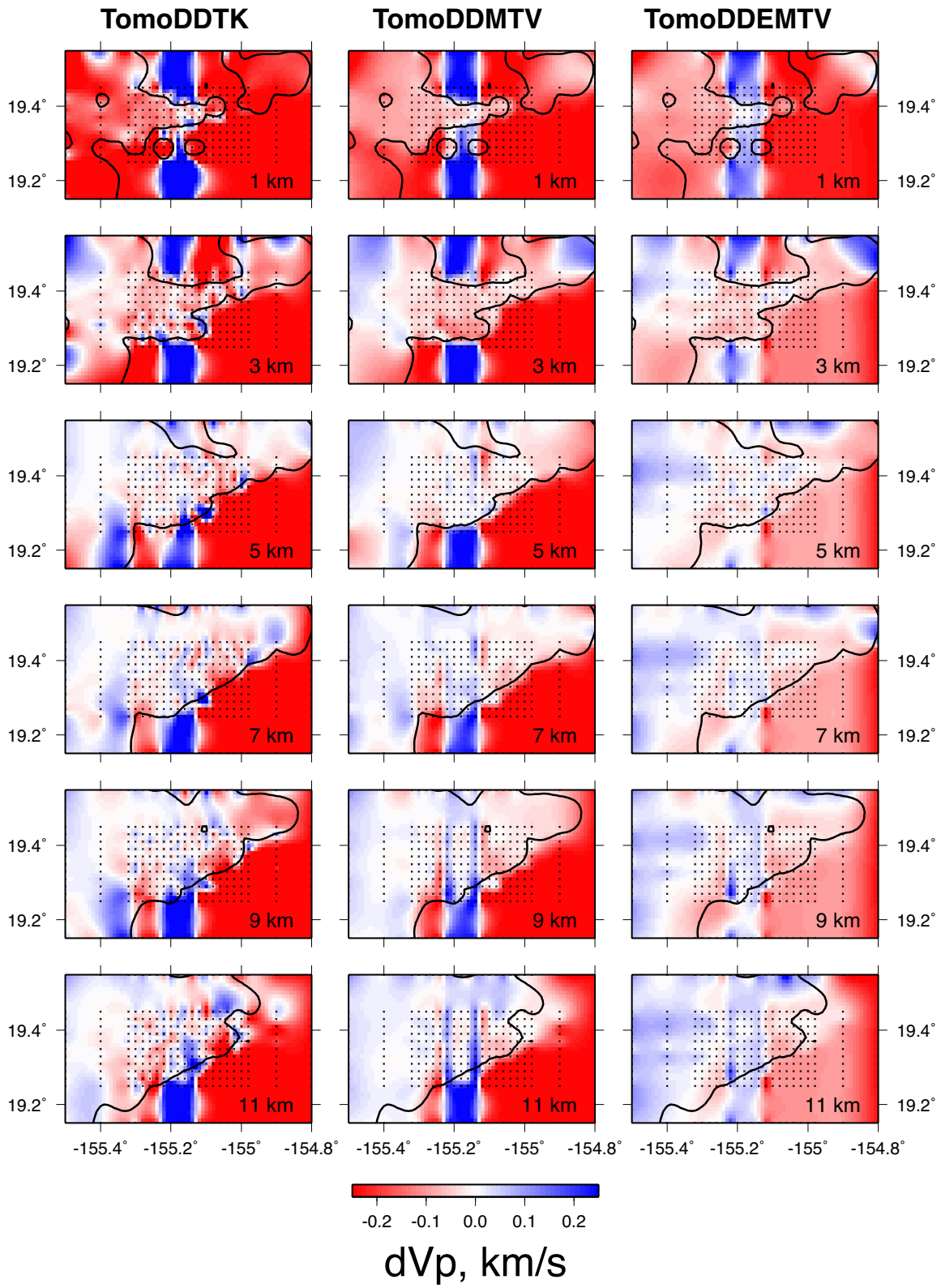

Figure 3. Perturbations from the synthetic model for inversions using Tikhonov regularization (left-hand column), MTV regularization (middle column), and MTV regularization with *a priori* interfaces (right-hand column). Blue indicates that the recovered model is faster than the synthetic model, red indicates that it is slower than the synthetic model. The bold black lines show the area that is well sampled based on DWS values. The MTV regularization with *a priori* interfaces yields the smallest differences to the synthetic model.

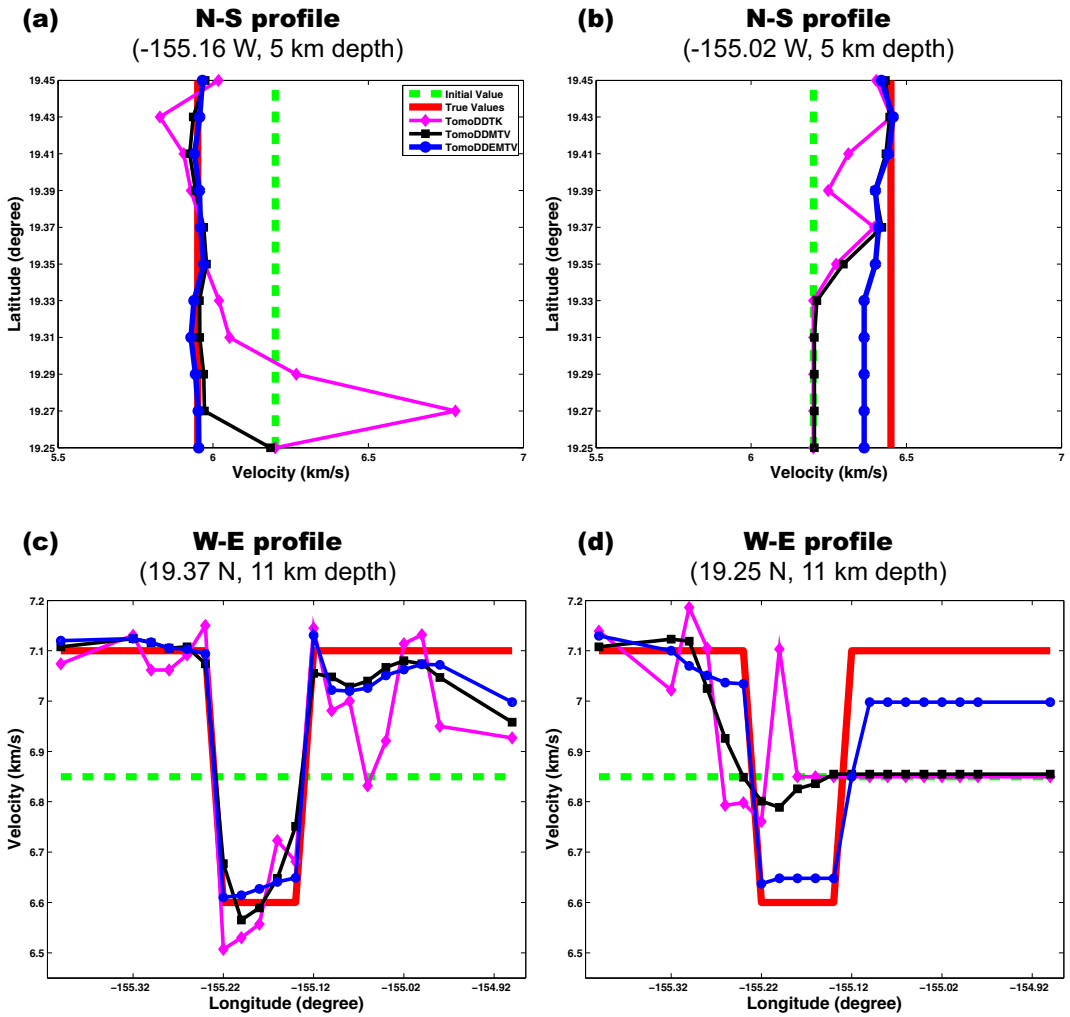

Figure 4. Horizontal profiles through the inversion results shown in Fig. 2 and the synthetic model shown in Fig. 1. (a) A north–south profile at −155.16°W and depth = 5 km; (b) A north–south profile −155.02°W and depth = 5 km; (c) A west–east profile at 19.37°N and depth = 11 km; (d) A west–east profile at 19.25°N and depth = 11 km. Of the three methods, TomoDDEMTV yields the most similar results to the synthetic model. Profile locations are shown by the red lines in Fig. 1. The locations of the earthquakes surrounding these profiles are indicated in Fig. 1.

Table 1. Inversion errors (in m s^{-1}) calculated using eq. (21) for methods of TomoDDTK, TomoDDETV and TomoDDEMTV. Two different data sets are utilized: full data sets and 50 per cent data sets. The velocity value in the parentheses is the degradation value from the results when using the full data set. TomoDDEMTV not only yields the smallest inversion error but also preserves the accuracy of the inversion most consistently among all three methods.

	Methods		
Data usage	TomoDDTK	TomoDDMTV	TomoDDEMTV
Full data sets	719.1	554.0	**340.3**
50 per cent data sets	805.5 (86.4)	684.4 (130.4)	**396.5 (56.2)**

where $(\mathbf{m}_{(i,j)})^{\text{pred}}$ is the calculated model velocity at the (i,j)th node, $(\mathbf{m}_{(i,j)})^{\text{true}}$ is the true model velocity at the (i,j)th node, and (i,j) are the indexes in the latitude and longitude directions. We calculate the sum of the absolute values of the differences between the recovered and true models for the three inversions of TomoDDTK, TomoDDMTV and TomoDDEMTV using eq. (21) and list them in Table 1. This analysis quantitatively shows that TomoDDEMTV yields a more accurate inversion result than those obtained by the other two methods with correct *a priori* interfaces information.

Because the starting locations for the tomographic inversion are the same as those used to calculate the synthetic data and the inversion jointly solves for velocity and hypocentral locations, ideally the final locations should be identical to the starting locations for a well-recovered velocity model. For example, small relocation shifts suggest that the inversion algorithm has no need to move the earthquake hypocentres to fit the data, to accommodate artefacts such as those shown for the TomoDDTK results in Fig. 2. To obtain the relocation error, we calculate the point-to-point Euclidean distance between the location of the initial hypocentre and the relocated hypocentre after inversion. Fig. 5 shows the hypocentre relocations using the three DD traveltime tomography methods. From this figure, it is not easy to identify the best method as the cloud of points suggest comparable relocation errors. We perform three *t*-tests on all pairs possible using the three DD traveltime tomography methods. Each *t*-test provides a two-tailed *p*-value, which indicates the probability, under the null hypothesis, to obtain a data set as extreme as the one actually observed. The null hypothesis in the case of a *t*-test is a non-significant difference between the two data sets. We obtain the following *p*-values from the three tests; *p*-value $= 6.3 \times 10^{-17}$ for TomoDDTK – TomoDDMTV, *p*-value $= 1.3 \times 10^{-20}$ for TomoDDMTV – TomoDDEMTV and *p*-value = 0.11 for

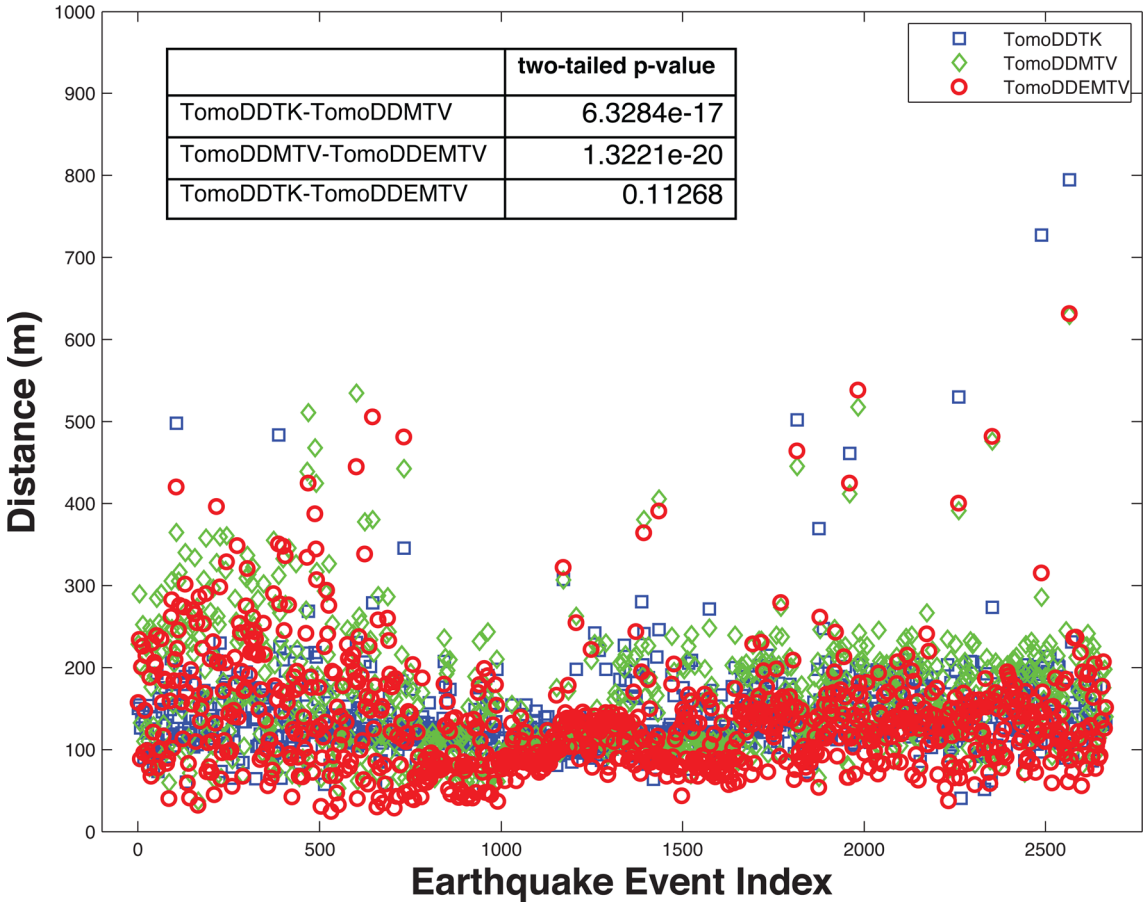

Figure 5. Relocation error of the inversion results using TomoDDTK, TomoDDMTV and TomoDDEMTV. The relocation error is the point-to-point Euclidean distance between the location of the true hypocentre and the relocated hypocentre after inversion.

TomoDDTK – TomoDDEMTV. The first two *p*-values are much less than 1 per cent, and mean that there is a significant difference between the relocations obtained using TomoDDMTV compared to those obtained from TomoDDTK and TomoDDEMTV. This indicates that the larger mean relocation error for TomoDDEMTV is significant, suggesting that the artefacts along the sharp boundaries visible in Fig. 3 affect the relocation more than for the two other methods. Although the *p*-value for TomoDDTK and TomoDDEMTV is too large to reject the null hypothesis (i.e. demonstrate that there is a statistical difference), it must be noted that TomoDDEMTV is the only method that both recovers the synthetic model (Fig. 4) well and produces small relocation errors, with a mean of 140 m. The failure of the *t*-test only means that these two methods produce results that are not statistically different when using this given data set (Box *et al.* 1978) and it must also be noted that earthquake relocations are not an exact proxy to velocity model accuracy because some earthquakes might be poorly located.

4.2 Robustness tests

These tests demonstrate that TomoDDEMTV produces the most accurate inversion results of the methods tested. The *a priori* interface information plays an important role in obtaining accurate results. In this section, we assess the robustness of TomoDDEMTV with respect to the accuracy (Section 4.2.1) and availability (Section 4.2.2) of *a priori* interface information. We also assess the robustness of the methods for a case of sparse data coverage (Section 4.2.3).

4.2.1 Robustness tests on the accuracy of the interfaces

We consider three different interface weighting schemes scenarios to test the robustness of TomoDDEMTV with respect to the accuracy of *a priori* interface information (Fig. 6). Scenario 1 represents a situation where the exact locations of the interfaces are uncertain, and a range of nodes are given equal likelihood of being adjacent to the interface. Within this range of nodes are the actual locations of the interface. Scenario 2 also assumes that the exact interface locations are unknown, but some nodes are thought to be more likely adjacent to the interface. In Scenario 3, the location of the interface is incorrectly identified.

Figs 7 and 8 show the inversion results and the differences between the synthetic model, respectively, for the three scenarios in Fig. 6 using TomoDDEMTV. For Scenario 1, the velocities more than a few nodes from the velocity interface are well recovered, although there are some artefacts near the interfaces due to the broadly identified interfaces. Scenario 2 produces the largest artefacts of the three scenarios tested. Particularly visible in Fig. 8, this scenario produces stripes of large artefacts at the edge of the velocity contrast, where the edge confidence is set to 0.5. In Scenario 3, the velocities within central low-velocity region at some depths

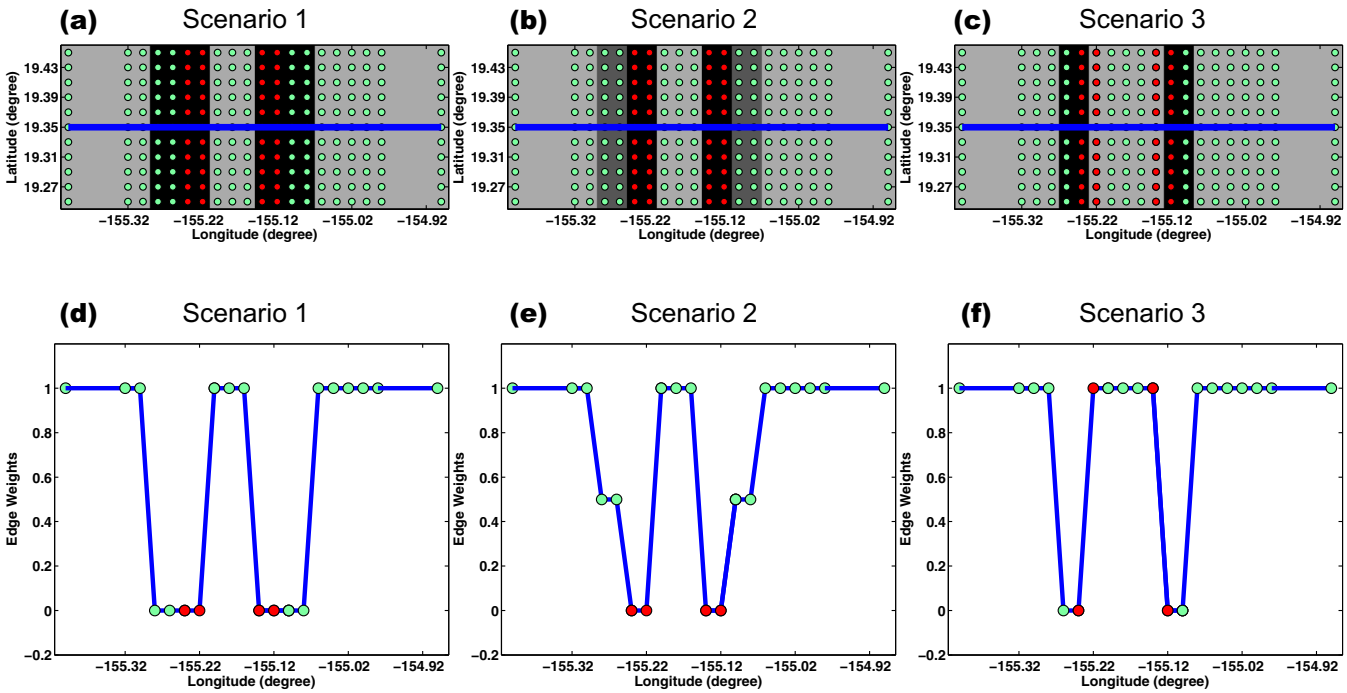

Figure 6. Three types of interface weighting schemes are used to test the robustness of TomoDDEMTV when the interface location is poorly known. The nodes at the true interface locations are plotted in red. Horizontal profiles indicated in the upper panels are plotted in the lower panels. Panels (a) and (d) illustrate Scenario 1, where several nodes are given equal likelihood of being at the interface. Panels (b) and (e) illustrate Scenario 2, where some nodes have a 50 per cent confidence of being at the interface. Panels (c) and (f) illustrate Scenario 3, where the location of the interface is incorrectly identified.

differ slightly from the synthetic velocities, but large artefacts are not introduced in comparison to the results of the TomoDDEMTV inversion.

Comparing the results in Figs 7 and 8 shows that the Scenario 2 yields the most artefacts out of all three scenarios. This is due to the fact that by weighting 0.5 confidence nodes, we actually introduce two additional false interfaces, one to the right and one to the left of the real interfaces. This is inconsistent with the actual data, which causes the stripe-like artefacts. We calculate the inversion errors for these three scenarios using eq. (21) and list them in Table 2, which demonstrates that Scenarios 1 and 3 yield more accurate results than Scenario 2. Based on these three types of parametrization, we conclude that although some artefacts can be generated due to inaccurately identified interfaces, the results of the TomoDDEMTV inversion are still much more accurate than those results obtained using the conventional DD traveltime tomography method of TomoDDTK. If there is some uncertainty in the actual location of a velocity interface, it is preferable to assume a broader interface (Scenario 1) or a narrow, possibly incorrect, interface (Scenario 3), rather than assign a lower confidence to some interface locations (Scenario 2).

4.2.2 *Robustness tests of the availability of interfaces*

In some real problems, an *a priori* interface based on geological data may be available only for near-surface layers. We test this scenario using TomoDDEMTV by incorporating *a priori* interface information only for layers 5 km depth and shallower (Figs 9 and 10). For the shallow layers where *a priori* interface information is included, the results for these layers are as accurate as the results of the inversion where *a priori* interface information is applied for all depths (Fig. 2). For the deeper layers where no *a priori* interface information is included, the accuracy is equivalent to that of the TomoDDMTV results (Fig. 2). The overall results of this scenario when using TomoDDEMTV are still much more accurate than those obtained using conventional DD traveltime tomography method, even when limited *a priori* information is available.

4.2.3 *Robustness tests on the sparse data sets*

We also test the robustness of our method with respect to data set sparsity. To setup the problem, we randomly eliminate 50 per cent of the earthquake events and stations from those in Fig. 1. We compare the inversions using all the three methods and show them in Fig. 11. Their differences to the true velocity model is provided in Fig. 12. As we can visually observe, the inversion result of TomoDDEMTV is comparable to that using the full data sets in Fig. 2, while TomoDDTK and TomoDDMTV result in significant artefacts. Using eq. (21), we calculate the inversion errors for the three methods using the degraded data sets and report them in Table 1. Our method yields the smallest inversion error (396.5 m s^{-1}) out of all three methods. Comparison to the results obtained using the full data set, shown in Fig. 2, the results using both inversions TomoDDTK and TomoDDMTV have degraded significantly, while TomoDDEMTV consistently recovers the velocity model the most accurately. Therefore, given sparse data sets, TomoDDEMTV is robust and superior to the other two methods of TomoDDTK and TomoDDMTV in providing a consistent inversion quality.

5 CONCLUSIONS

We have developed a DD traveltime tomography method using modified total-variation regularization incorporated with *a priori*

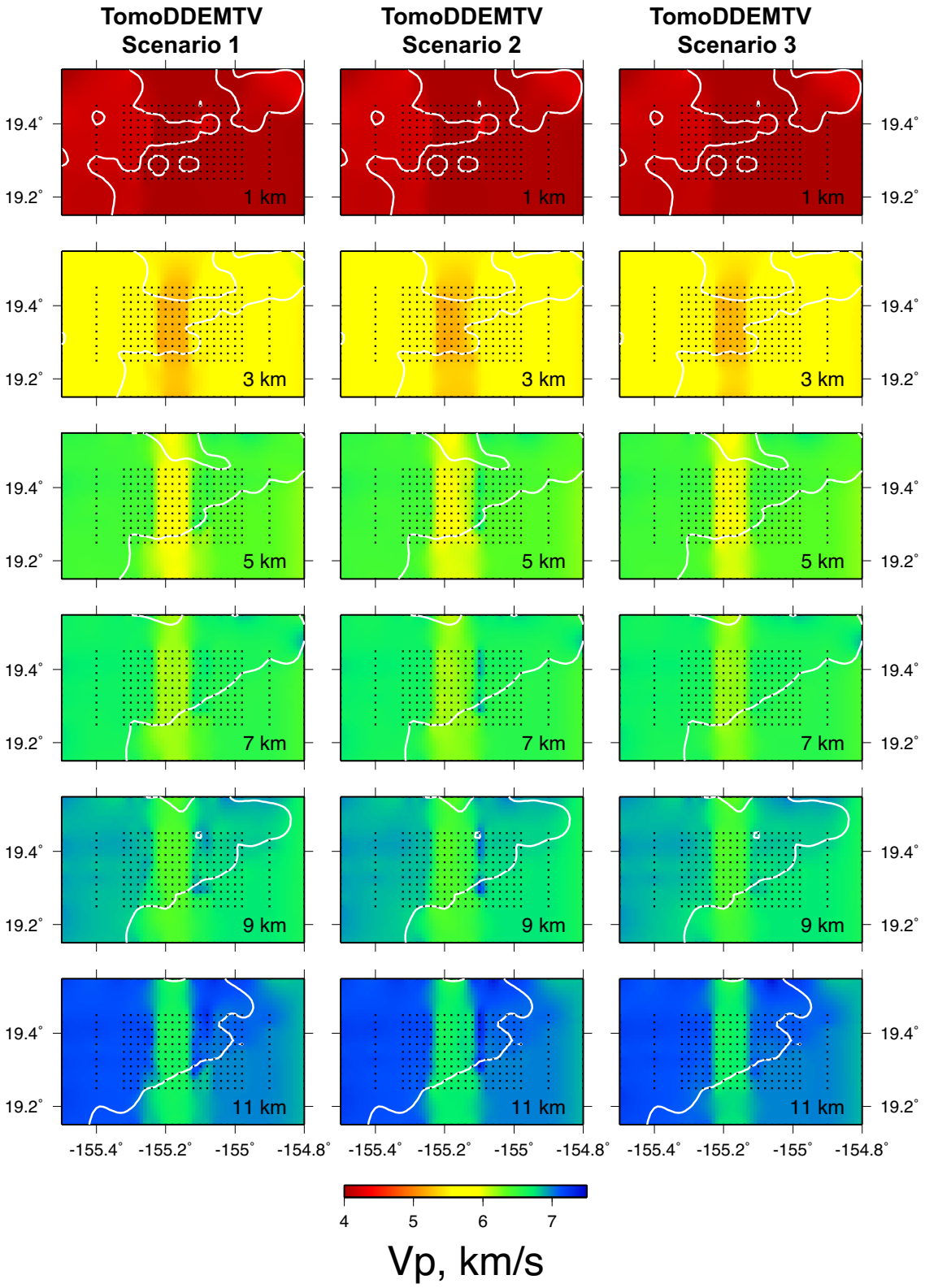

Figure 7. Results for the three scenarios shown in Fig. 6 using TomoDDEMTV: Scenario 1 (left-hand column), Scenario 2 (middle column) and Scenario 3 (right-hand column). The white lines show the area that is well sampled based on DWS values.

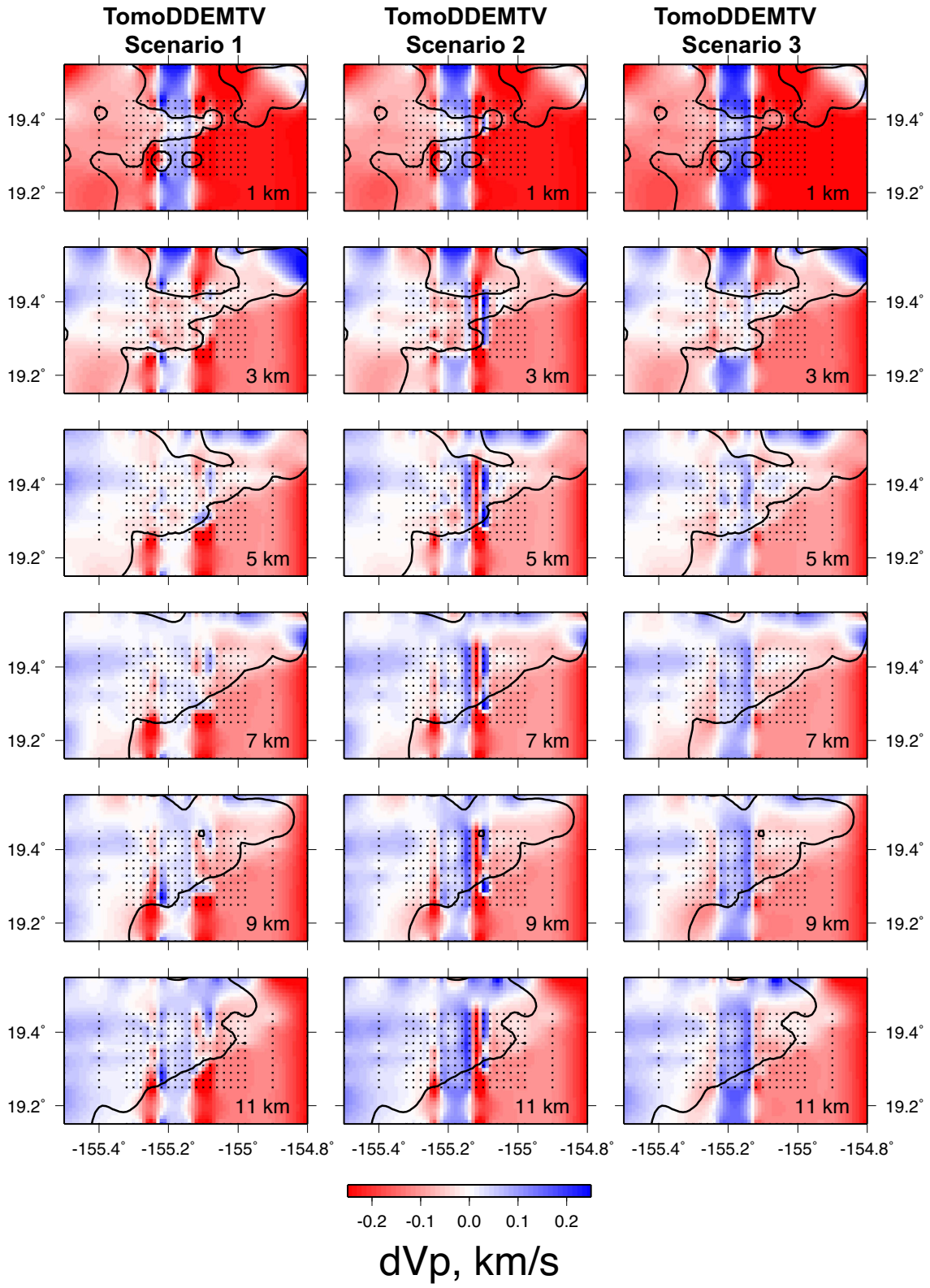

Figure 8. Perturbations from the synthetic model for the three scenarios shown in Fig. 6 using TomoDDEMTV: Scenario 1 (left-hand column), Scenario 2 (middle column), and Scenario 3 (right-hand column). The bold black lines show the area that is well sampled based on DWS values.

Table 2. Inversion errors (in m s^{-1}) calculated using eq. (21) for three different scenarios depicted in Fig. 6. Full data sets are utilized. Scenarios 1 and 3 yield better results than Scenario 2.

	Scenarios		
Data usage	Scenario 1	Scenario 2	Scenario 3
Full data sets	414.3	543.0	401.1

interface information. We employ an alternating minimization algorithm to decouple the tomography into two subproblems that can be solved much more efficiently than solving the original tomography problem. To solve these two subproblems, we use LSQR and split-Bregman iterative solvers. The modified total variation regularization avoids the use of a smoothing parameter or quadratic approximation to the total variation functional, enhancing the robustness of our inversion method. The incorporation of *a priori* interface information further improves the stability and accuracy of the inversion without adding significant additional cost compared to a classical Tikhonov regularization. Our numerical examples demonstrate that this new tomography method yields accurate velocity results and is robust even when constrained by inaccurate *a priori* information. We also demonstrate that our method can be robust when provided with sparse data sets. This can be beneficial for certain real situations when earthquake events and/or stations are limited. However, our method relies on more regularization parameters than those used in the conventional DD

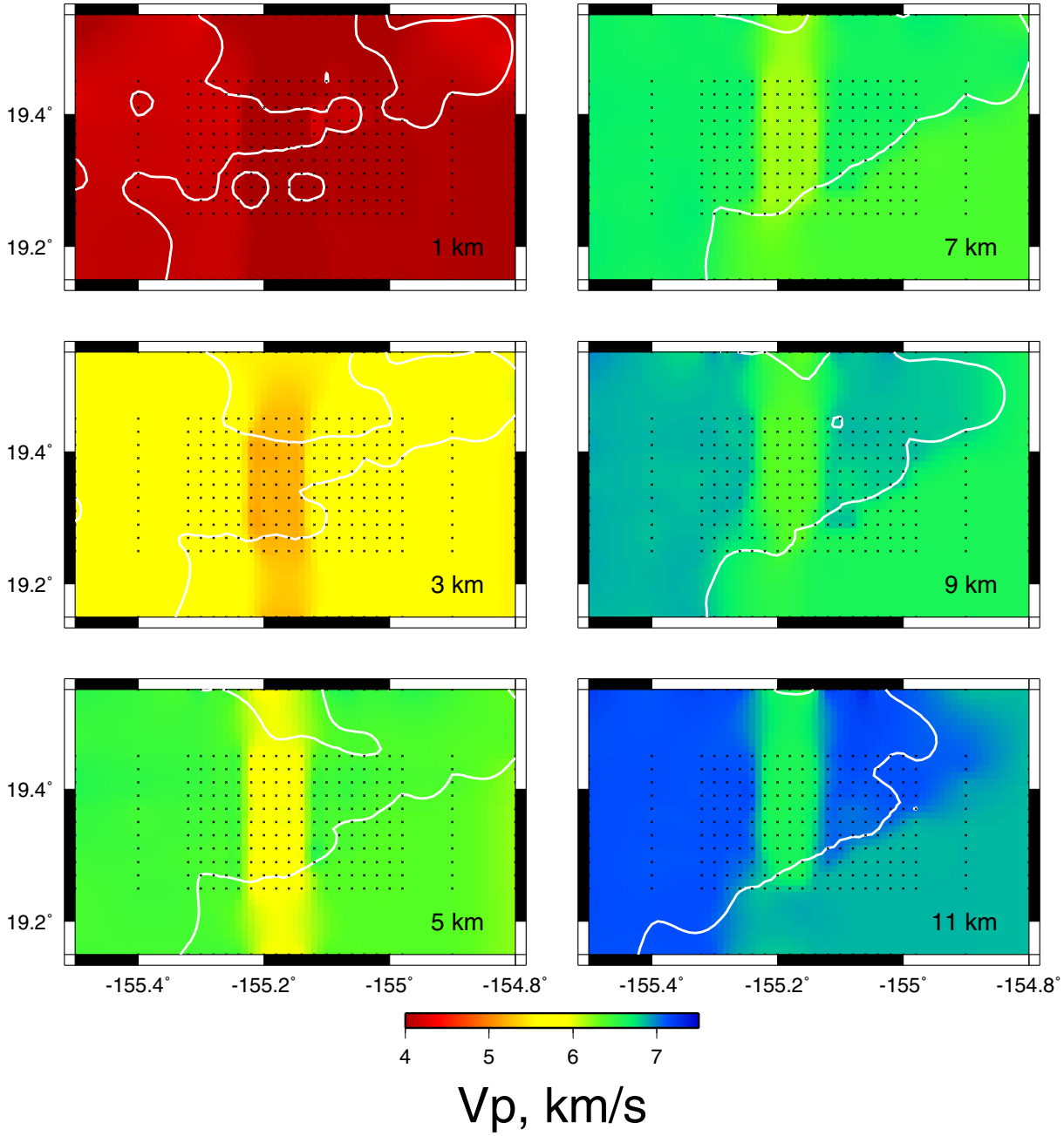

Figure 9. Results of an inversion using TomoDDEMTV where *a priori* information on interface location is only included for layers at 5 km depth and shallower. The white lines show the area that is well sampled based on DWS values.

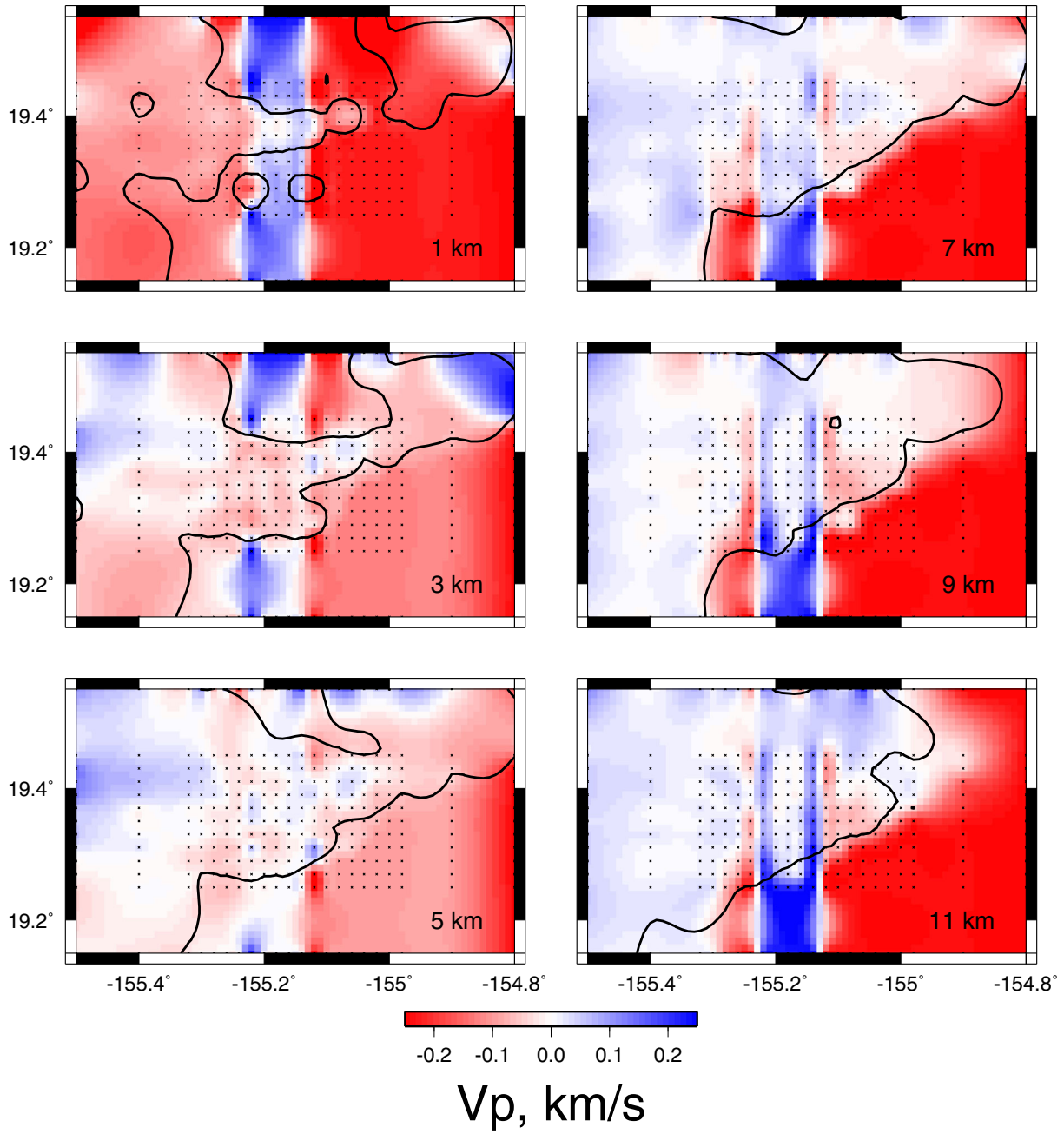

Figure 10. Perturbations from the synthetic model for an inversion using TomoDDEMTV where *a priori* information on interface location is only included for layers at 5 km depth and shallower. The bold black lines show the area that is well sampled based on DWS values.

traveltime tomography with Tikhonov regularization, which can add computational costs to obtain an accurate inversion. Our method can introduce some artefacts due to the false *a priori* interface information, which might add some difficulty to interpretation of the inversion results.

In general, our method has a strong potential to improve the results of inversions of actual seismic data for real-world problems. In future work, we will explore the related computational issues such as convergence rate and behaviour, memory usage, etc. We would also like to consider the incorporation of additional *a priori* information into our current method and evaluate its performance. More over, we will apply our method to a real data set and report results in a future paper.

ACKNOWLEDGEMENTS

This work was supported by the U.S. DOE (Grant LDRD-20120047ER). We thank Dr Charles Ammon from Penn State for his valuable suggestions. Thanks also to the developers of the Generic Mapping Tool (GMT; Wessel *et al.* 2013). We thank the Editor, Dr Egill Hauksson, and both reviewers, Dr Peter Lelièvre and an anonymous reviewer for their valuable comments and suggestions.

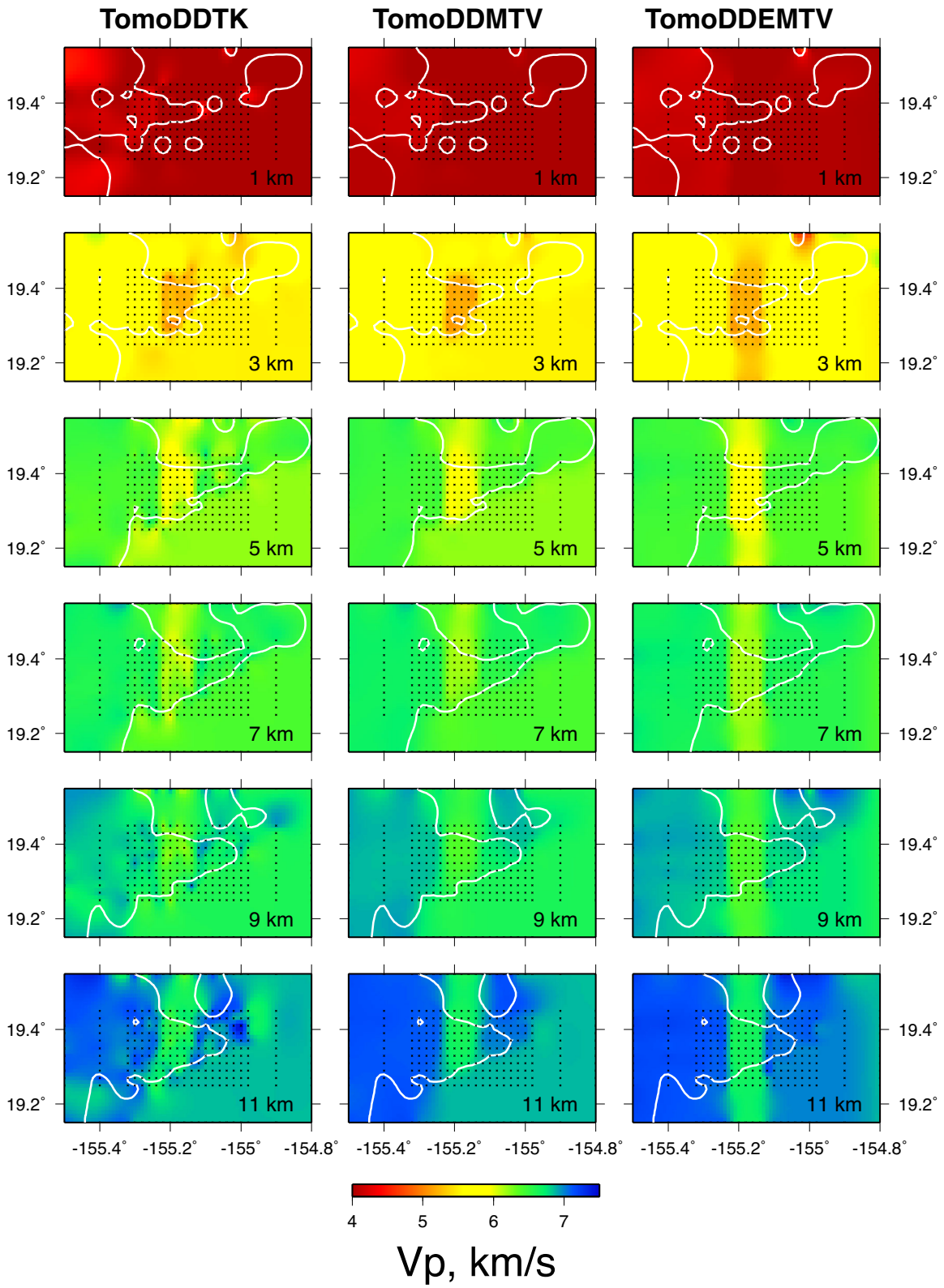

Figure 11. Inversion results produced using Tikhonov regularization (left-hand column), MTV regularization (middle column), and MTV regularization with *a priori* interfaces (right-hand column). The data sets consists of 50 per cent of the total earthquakes and stations shown in Fig. 1. The white lines show the area that is well sampled based on DWS values. The inversion result of TomoDDEMTV is comparable to that using the full data set in Fig. 2, while significant artefacts occur in the results using TomoDDTK and TomoDDMTV.

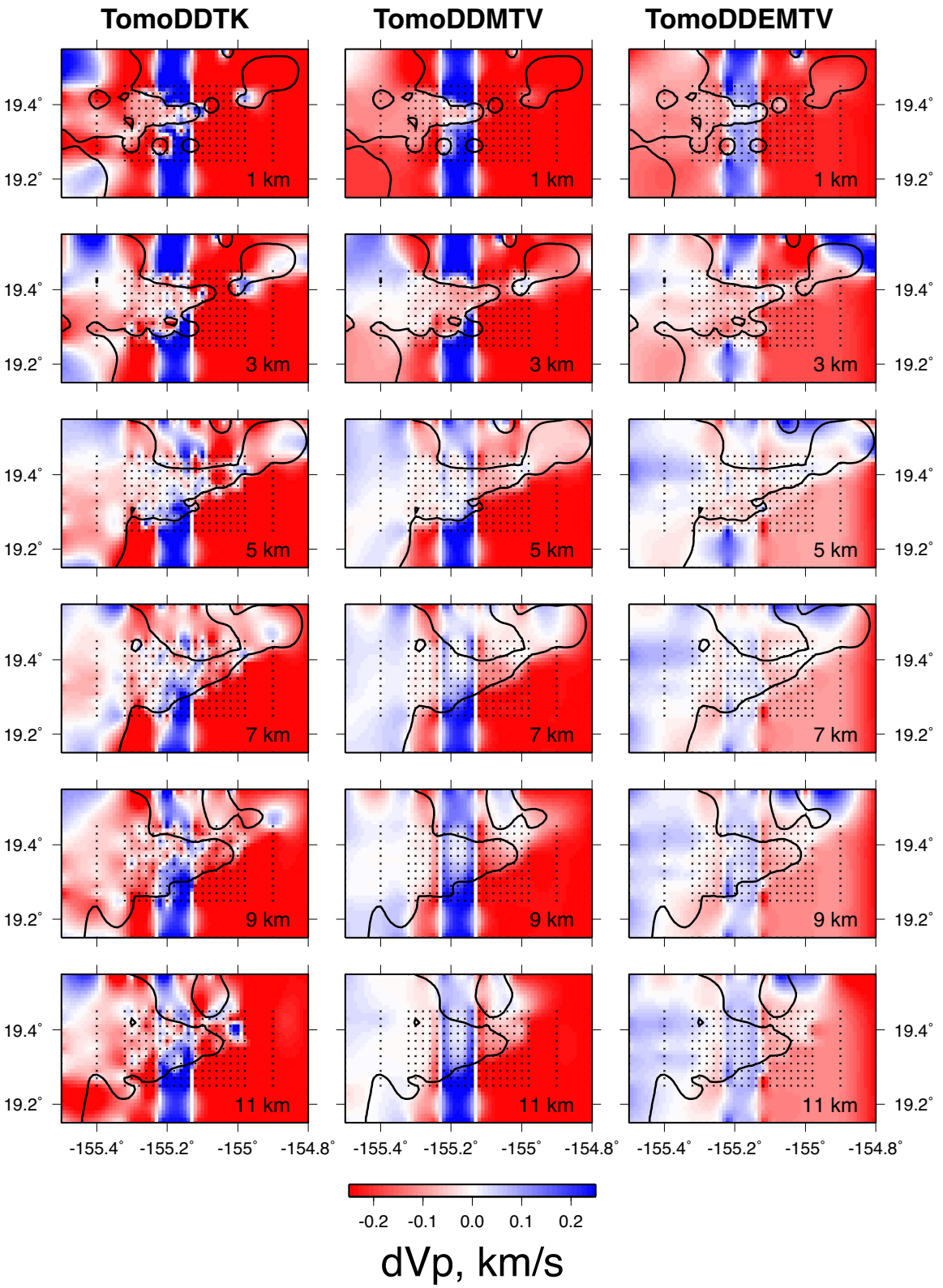

Figure 12. Perturbations from the synthetic model for inversions using Tikhonov regularization (left-hand column), MTV regularization (middle column), and MTV regularization with *a priori* interfaces (right-hand column). Blue indicates that the recovered model is faster than the synthetic model, red indicates that it is slower than the synthetic model. The data sets consists of 50 per cent of the total earthquakes and stations shown in Fig. 1. The bold black lines show the area that is well sampled based on DWS values. The inversion result of TomoDDEMTV is comparable to that using the full data set in Fig. 2, while significant artefacts occur in the results using TomoDDTK and TomoDDMTV.

REFERENCES

Ammon, C.J. & Patton, H.J., 1992. Lateral shear-velocity variations beneath the western United States, *EOS, Trans. Am. geophys. Un.,* **73,** 369.

Aster, R., Borchers, B. & Thurber, C., 2005. *Parameter Estimation and Inverse Problems,* Elsevier.

Barbosa, V.C.F. & Silva, J.B.C., 2006. Interactive 2D magnetic inversion: a tool for aiding forward modeling and testing geologic hypotheses, *Geophysics,* **71**(5), 43–50.

Berryman, J.G., 1989. Weighted least-squares criteria for seismic traveltime tomography, *IEEE Trans. Geosci. Remote Sens.,* **27,** 302–309.

Björck, A., 1996. *Numerical Methods for Least Squares Problems,* SIAM.

Box, G.E., Hunter, W.G. & Hunter, J.S., 1978. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building,* Series in Probability and Statistics, Wiley.

Bregman, L., 1967. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex optimization, *USSR Comput. Math. Math. Phys.,* **7,** 200–217.

Chasseriau, P. & Chouteau, M., 2003. 3D gravity inversion using a model of parameter covariance, *J. appl. Geophys.,* **52**(1), 59–74.

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

Delbos, F., Gilbert, J.C., Glowinski, R. & Sinoquet, D., 2006. Constrained optimization in seismic reflection tomography: a Gauss-Newton augmented lagrangian approach, *Geophys. J. Int.,* **164,** 670–684.

Elad, M., 2010. *Sparse and Redundant Representations,* Springer.

Farquharson, C.G., 2008. Constructing piecewise-constant models in multi-dimensional minimum-structure inversions, *Geophysics,* **73**(1), K1–K9.

Farquharson, C.G., Ash, M.R. & Miller, H.G., 2008. Geologically constrained gravity inversion for the Voisey's Bay ovoid deposit, *Leading Edge,* **27**(1), 64–69.

Goldstein, T. & Osher, S., 2009. The split Bregman method for l1 regularized problems, *SIAM J. Imag. Sci.,* **2,** 323–343.

Golub, G.H. & Van Loan, C.F., 1996. *Matrix Computations,* 3 edn, The Johns Hopkins University Press.

Hansen, P.C., 1998. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion,* SIAM.

Hansen, P.C. & O'Leary, D.P., 1993. The use of the l-curve in the regularization of discrete ill-posed problems, *ACM Trans. Math. Software,* **14,** 1487–1503.

Huang, Y., NG, M. & Wen, Y., 2008. A fast total variation minimization method for image restoration, *SIAM J. Multiscale Model. Simulat.,* **7**(2), 774–795.

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,* **116**(B6), doi:10.1029/2010JB008190.

Lelièvre, P.G., 2009. Integrating geologic and geophysical data through advanced constrained inversions, *PhD thesis,* University of British Columbia.

Lelièvre, P.G. & Farquharson, C.G., 2013. Gradient and smoothness regularization operators for geophysical inversion on unstructured meshes, *Geophys. J. Int.,* **195**(1), 330–341.

Lelièvre, P.G. & Oldenburg, D.W., 2009a. A 3D total magnetization inversion applicable when significant complicated remanence is present, *Geophysics,* **74**(3), L21–L90.

Lelièvre, P.G. & Oldenburg, D.W., 2009b. A comprehensive study of including structural orientation information in geophysical inversions, *Geophys. J. Int.,* **178**(2), 623–637.

Lelièvre, P.G., Oldenburg, D.W. & Williams, N.C., 2009. Integrating geological and geophysical data through advanced constrained inversions, *Explor. Geophys.,* **40,** 334–341.

Li, Y. & Oldenburg, D.W., 1996. 3-D inversion of magnetic data, *Geophysics,* **61**(2), 394–408.

Li, Y. & Oldenburg, D.W., 1998. 3-D inversion of gravity data, *Geophysics,* **63**(1), 109–119.

Li, Y. & Oldenburg, D.W., 2000. Incorporating geological dip information into geophysical inversions, *Geophysics,* **65,** 148–157.

Li, Y. & Oldenburg, D.W., 2003. Fast inversion of large-scale magnetic data using wavelet transforms and a logarithmic barrier method, *Geophys. J. Int.,* **152**(2), 251–265.

Lin, Y., Wohlberg, B. & Guo, H., 2010. UPRE method for total variation parameter selection, *Signal Process.,* **90**(8), 2546–2551.

Loris, I. & Verhoeven, C., 2012. Iterative algorithms for total variation-like reconstructions in seismic tomography, *Int. J. Geomath.,* **3**(2), 179–208.

Loris, I., Douman, H., Nolet, G., Daubechies, I. & Regone, C., 2010. Nonlinear regularization techniques for seismic tomography, *J. Comput. Phys.,* **229,** 890–905.

Maceira, M. & Ammon, J., 2009. Joint inversion of surface wave velocity and gravity observations and its application to central Asian basins shear velocity structure, *J. geophys. Res.: Solid Earth,* **114**(B2), 1–18.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press.

Micchelli, C.A., Shen, L., Xu, Y. & Zeng, X., 2013. Proximity algorithms for the L1/TV image denoising model, *Adv. Comput. Math.,* **38,** 401–426.

Nolet, G., 1987. *Seismic Tomography: With Applications in Global Seismology and Exploration,* D. Reidel Publishing Company.

Osher, S., Burger, M., Goldfarb, D., Xu, J. & Yin, W., 2005. An iterative regularization method for total variation-based image restoration, *SIAM J. Multiscale Model. Simulat.,* **4**(2), 460–489.

Paige, C.C. & Saunders, M.A., 1982a. LSQR: an algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Software,* **8**(1), 43–71.

Paige, C.C. & Saunders, M.A., 1982b. LSQR: sparse linear equations and least squares problems, *ACM Trans. Math. Software,* **8** (2), 195–209.

Rodriguez, P., 2014. Total variation regularization algorithms for images corrupted with different noise models: a review, *J. Electr. Comput. Eng.,* **2013,** 1–18.

Rudin, L., Osher, S. & Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms, *Physica D.,* **60,** 259–268.

Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithmi. Searching a parameter space, *Geophys. J. Int.,* **138,** 479–494.

Syracuse, E.M., Thurber, C.A., Wolfe, C.J., Okubo, P.G., Foster, J.H. & Brooks, B.A., 2010. High-resolution locations of triggered earthquakes and tomographic imaging of Kilauea Volcanos south flank, *J. geophys. Res.,* **115**(B10), doi:10.1029/2010JB007554.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49**(8), 1259–1266.

Tarantola, A., 2005. *Inverse Problem Theory,* SIAM.

Thurber, C. & Eberhart-Phillips, D., 1999. Local earthquake tomography with flexible gridding, *Comput. Geosci.,* **25**(1), 809–818.

Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V. & Yagola, A.G., 1995. *Numerical Methods for the Solution of Ill-Posed Problems,* Kluwer Academic Publishers.

Vogel, C., 2002. *Computational Methods for Inverse Problems,* SIAM.

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

Wang, Y., Yang, J., Yin, W. & Zhang, Y., 2008. A new alternating minimization algorithm for total variation image reconstruction, *SIAM J. Imag. Sci.,* **1**(3), 248–272.

Wessel, P. W., Smith, H.F., Scharroo, R., Luis, J. & Wobbe, F., 2013. Generic mapping tools: improved version released, *EOS, Trans. Am. geophys. Un.,* **94,** 409–410.

Wijns, C. & Kowalczyk, P., 2007. Interactive geophysical inversion using qualitative geological constraints, *Geophysics,* **38**(3), 208–212.

Wohlberg, B. & Rodríguez, P., 2007. An iteratively reweighted norm algorithm for minimization of total variation functionals, *IEEE Signal Process. Lett.,* **14**(12), 948–951.

Ye, X., Chen, Y., Lin, W. & Huang, F., 2011. Fast MR image reconstruction for partially parallel imaging with arbitrary k-space trajectories, *IEEE Trans. Med. Imag.,* **30**(3), 575–585.

Yin, W., Osher, S., Goldfarb, D. & Darbon, J., 2008. Bregman iterative algorithms for l_1-minimization with applications to compressed sensing, *SIAM J. Imag. Sci.,* **1,** 143–168.

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.H., 2006. Development and applications of double-difference seismic tomography, *Pure appl. Geophys.,* **163,** 373–403.

Zhang, H., Maceira, M., Roux, P. & Thurber, C., 2014. Joint inversion of body-wave arrival times and surface-wave dispersion for three-dimensional seismic structure around SAFOD, *Pure appl. Geophys.,* **171,** 3013–3022.

APPENDIX A: COMPUTATIONAL METHODS

Our computational method for solving the optimization in eq. (18) is composed of two parts: solving the subproblem in eq. (19) and solving the subproblem in eq. (20).

Similar to the original DD traveltime tomography with Tikhonov regularization in eq. (4), eq. (19) can be equivalently written as:

$$\begin{bmatrix} \tilde{G} \\ \sqrt{\mu}\, I \end{bmatrix} \tilde{\mathbf{m}} = \begin{bmatrix} \tilde{\mathbf{d}} \\ \sqrt{\mu}\, \tilde{\mathbf{u}}^{(k)} \end{bmatrix}. \tag{A1}$$

Hence, we employ LSQR (Paige & Saunders 1982a,b) to solve eq. (A1) due to its stability and efficiency in solving large scale ill-posed linear systems in comparison to other linear solvers.

To solve eq. (20), we employ the split-Bregman iterative method (Goldstein & Osher 2009), due to its robustness. The details of the split-Bregman iterative method are described in Section A1.

We begin the iteration with an initial model $\tilde{\mathbf{u}}^{(0)} = \tilde{\mathbf{m}}^{(0)}$. By solving the two subproblems in eqs (19) and (20), respectively, we are able to generate a sequence of iterations

$$\tilde{\mathbf{u}}^{(0)} \rightarrow \tilde{\mathbf{m}}^{(1)} \rightarrow \tilde{\mathbf{u}}^{(1)} \rightarrow \tilde{\mathbf{m}}^{(2)} \rightarrow \tilde{\mathbf{u}}^{(2)} \rightarrow \tilde{\mathbf{m}}^{(3)} \rightarrow \tilde{\mathbf{u}}^{(3)} \rightarrow \cdots \rightarrow \tilde{\mathbf{m}}^{(k)} \rightarrow \tilde{\mathbf{u}}^{(k)} \rightarrow \cdots.$$

that converges on the inversion result.

A1 Split-Bregman method for solving the second subproblem

Methods based on the Bregman distance (Osher *et al.* 2005; Yin *et al.* 2008; Goldstein & Osher 2009) are reliable approaches for solving the L_2-TV minimization problem in eq. (20). The definition of the Bregman distance is given as (Bregman 1967):

$$D^p_{E_2}(\mathbf{m}_1, \mathbf{m}_2) = E_2(\mathbf{m}_2) - [E_2(\mathbf{m}_1) + \langle p, \mathbf{m}_2 - \mathbf{m}_1 \rangle], \tag{A2}$$

where E_2 is defined in eq. (20). $\langle \cdot, \cdot \rangle$ is an inner product operator and p is the subgradient of E_2 at $\mathbf{m}_2$. Fig. A1 provides a geometric description of the Bregman distance between $\mathbf{m}_2$ and $\mathbf{m}_1$. We observe that the Bregman distance is the vertical distance at $\mathbf{m}_2$ between the original curve E_2 and the line tangent to the original curve at $\mathbf{m}_1$. Two important properties can be found from the definition of the Bregman distance in eq. (A2) (Bregman 1967). First, the Bregman distance is not symmetric, so it is not a distance in the usual sense. Secondly, the Bregman distance measures the closeness of two points in the sense that $D^p_{E_2}(\mathbf{m}_1, \mathbf{m}_2) \geq 0$, and $D^p_{E_2}(\mathbf{m}_1, \mathbf{m}_2) \geq D^p_{E_2}(\mathbf{m}_1, \mathbf{m}_3)$ for $\mathbf{m}_3$ on the line segment between $\mathbf{m}_1$ and $\mathbf{m}_2$. One of the main benefits of using the Bregman distance is that the minimization of $D^p_{E_2}(\mathbf{m}_1, \mathbf{m}_2)$ does not penalize a model containing sharp interfaces. In the work of Osher *et al.* (2005), Yin *et al.* (2008) and Goldstein & Osher (2009), the combination of the Bregman distance with the TV term is shown to be superior to using the conventional TV term, and the split-Bregman iterative method

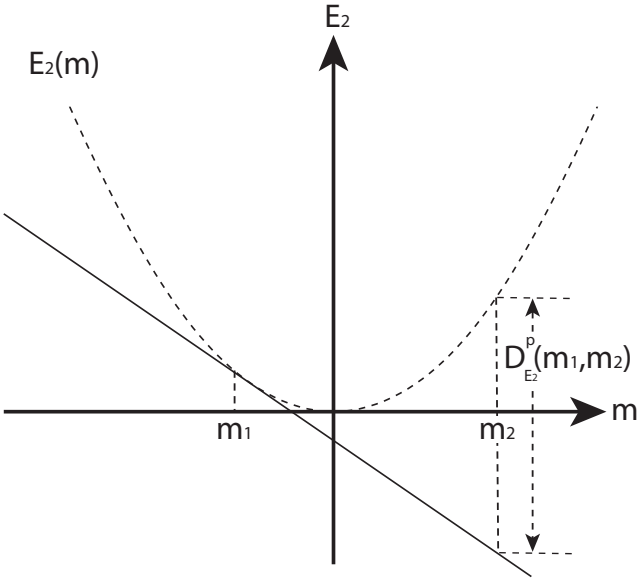

Figure A1. The Bregman distance $D^p_{E_2}$ between two points $\mathbf{m}_1$ and $\mathbf{m}_2$ of the misfit function $E_2(\mathbf{m})$.

is also shown to converge to an accurate solution. Therefore, DD traveltime tomography with MTV regularization and *a priori* interfaces using the split-Bregman iterative method yields a model containing fewer artefacts and sharper interfaces than one obtained using a conventional TV regularization term.

To employ the split-Bregman iterative method in solving our problem, we first introduce two auxiliary variables, $\tilde{\mathbf{w}}_x \approx \nabla_x \tilde{\mathbf{u}}$ and $\tilde{\mathbf{w}}_z \approx \nabla_z \tilde{\mathbf{u}}$, to reformulate eq. (20) as an equivalent minimization problem:

$$\min_{\tilde{\mathbf{u}}, \tilde{\mathbf{w}}_x, \tilde{\mathbf{w}}_z} \{ \|\tilde{\mathbf{u}} - \tilde{\mathbf{m}}^{(k)}\|_2^2 + \lambda_2 \|w \nabla \tilde{\mathbf{u}}\|_1 + \alpha \|\tilde{\mathbf{w}}_x - \nabla_x \tilde{\mathbf{u}}\|_2^2 + \alpha \|\tilde{\mathbf{w}}_z - \nabla_z \tilde{\mathbf{u}}\|_2^2 \}, \tag{A3}$$

where the regularization parameter $\alpha = 2\lambda_2$ following Goldstein & Osher (2009). Applying the Bregman distance to eq. (A3) leads to

$$\min_{\tilde{\mathbf{u}}, \tilde{\mathbf{w}}_x, \tilde{\mathbf{w}}_z} \{ \|\tilde{\mathbf{u}} - \tilde{\mathbf{m}}^{(k)}\|_2^2 + \lambda_2 \|w \nabla \tilde{\mathbf{u}}\|_1 + \alpha \|\tilde{\mathbf{w}}_x - \nabla_x \tilde{\mathbf{u}} - \tilde{\mathbf{b}}_x^{(k)}\|_2^2 + \alpha \|\tilde{\mathbf{w}}_z - \nabla_z \tilde{\mathbf{u}} - \tilde{\mathbf{b}}_z^{(k)}\|_2^2 \}, \tag{A4}$$

where $\tilde{\mathbf{b}}_x^{(k+1)} = \tilde{\mathbf{b}}_x^{(k)} + (\nabla_x \tilde{\mathbf{u}}^{(k+1)} - \tilde{\mathbf{w}}_x^{(k+1)})$ and $\tilde{\mathbf{b}}_z^{(k+1)} = \tilde{\mathbf{b}}_z^{(k)} + (\nabla_z \tilde{\mathbf{u}}^{(k+1)} - \tilde{\mathbf{w}}_z^{(k+1)})$, with $\tilde{\mathbf{b}}_x^{(0)} = \tilde{\mathbf{b}}_z^{(0)} = 0$.

We employ an alternating minimization algorithm to solve the minimization problem in eq. (A4), that is to solve the following two subproblems:

$$\min_{\tilde{\mathbf{u}}} \{ \|\tilde{\mathbf{u}} - \tilde{\mathbf{m}}^{(k)}\|_2^2 + \alpha \|\tilde{\mathbf{w}}_x^{(k)} - \nabla_x \tilde{\mathbf{u}} - \tilde{\mathbf{b}}_x^{(k)}\|_2^2 + \alpha \|\tilde{\mathbf{w}}_z^{(k)} - \nabla_z \tilde{\mathbf{u}} - \tilde{\mathbf{b}}_z^{(k)}\|_2^2 \}, \tag{A5}$$

$$\min_{\tilde{\mathbf{w}}_x, \tilde{\mathbf{w}}_z} \{ \lambda_2 \|w \nabla \tilde{\mathbf{u}}\|_1 + \alpha \|\tilde{\mathbf{w}}_x - \nabla_x \tilde{\mathbf{u}} - \tilde{\mathbf{b}}_x^{(k)}\|_2^2 + \alpha \|\tilde{\mathbf{w}}_z - \nabla_z \tilde{\mathbf{u}} - \tilde{\mathbf{b}}_z^{(k)}\|_2^2 \}. \tag{A6}$$

Eq. (A5) satisfies the optimality condition

$$(I - \alpha \triangle)\tilde{\mathbf{u}}^{(k+1)} = \tilde{\mathbf{m}}^{(k)} + \alpha \nabla_x^T (\tilde{\mathbf{w}}_x^{(k)} - \tilde{\mathbf{b}}_x^{(k)}) + \alpha \nabla_z^T (\tilde{\mathbf{w}}_z^{(k)} - \tilde{\mathbf{b}}_z^{(k)}), \tag{A7}$$

and the solution of eq. (A5) is obtained using the Gauss-Seidel iterative method:

$$\tilde{\mathbf{u}}_{i,j}^{(k)} = \frac{\alpha}{1+4\alpha}\left(\tilde{\mathbf{u}}_{i+1,j}^{(k)} + \tilde{\mathbf{u}}_{i-1,j}^{(k)} + \tilde{\mathbf{u}}_{i,j+1}^{(k)} + \tilde{\mathbf{u}}_{i,j-1}^{(k)} + \tilde{\mathbf{w}}_{x,i-1,j}^{(k)} - \tilde{\mathbf{w}}_{x,i,j}^{(k)} + \tilde{\mathbf{w}}_{z,i,j-1}^{(k)} - \tilde{\mathbf{w}}_{z,i,j}^{(k)} - \tilde{\mathbf{b}}_{x,i-1,j}^{(k)} + \tilde{\mathbf{b}}_{x,i,j}^{(k)} - \tilde{\mathbf{b}}_{z,i,j-1}^{(k)} + \tilde{\mathbf{b}}_{y,i,j}^{(k)}\right) + \frac{1}{1+4\alpha}\tilde{\mathbf{m}}_{i,j}^{(k)}. \tag{A8}$$

Eq. (A6) is solved explicitly using a generalized shrinkage formula (Wang *et al.* 2008)

$$\tilde{\mathbf{w}}_x^{(k+1)} = \max\left(\tilde{\mathbf{q}}^{(k)} - \frac{\lambda_2}{2\alpha}, 0\right)\frac{w\,\nabla_x\tilde{\mathbf{u}}^{(k)} + \tilde{\mathbf{b}}_x^{(k)}}{\tilde{\mathbf{q}}^{(k)}}, \tag{A9}$$

and

$$\tilde{\mathbf{w}}_z^{(k+1)} = \max\left(\tilde{\mathbf{q}}^{(k)} - \frac{\lambda_2}{2\alpha}, 0\right)\frac{w\,\nabla_z\tilde{\mathbf{u}}^{(k)} + \tilde{\mathbf{b}}_z^{(k)}}{\tilde{\mathbf{q}}^{(k)}}, \tag{A10}$$

where $\tilde{\mathbf{q}}^{(k)} = \sqrt{|w\,\nabla_x\tilde{\mathbf{u}}^{(k)} + \tilde{\mathbf{b}}_x^{(k)}|^2 + |w\,\nabla_z\tilde{\mathbf{u}}^{(k)} + \tilde{\mathbf{b}}_z^{(k)}|^2}$.

The numerical algorithm for solving eq. (20) using the split-Bregman iterative method is summarized in Algorithm 1.

Algorithm 1 Split-Bregman iterative method

Input: TOL
Output: $\mathbf{u}^{(k)}$
1: Initialize $j = 0$, $\tilde{\mathbf{w}}_x^{(0)} = \tilde{\mathbf{w}}_z^{(0)} = \tilde{\mathbf{b}}_x^{(0)} = \tilde{\mathbf{b}}_z^{(0)} = 0$;
2: **while** $||\tilde{\mathbf{u}}^{(j)} - \tilde{\mathbf{u}}^{(j-1)}|| >$ TOL **do**
3: Solve eq. (A5) according to eq. (A8);
4: Solve eq. (A6) according to eq. (A9) and eq. (A10);
5: Update $\tilde{\mathbf{b}}_x^{(j+1)} = \tilde{\mathbf{b}}_x^{(j)} + (\nabla_x\tilde{\mathbf{u}}^{(j+1)} - \tilde{\mathbf{w}}_x^{(j+1)})$;
6: Update $\tilde{\mathbf{b}}_z^{(j+1)} = \tilde{\mathbf{b}}_z^{(j)} + (\nabla_z\tilde{\mathbf{u}}^{(j+1)} - \tilde{\mathbf{w}}_z^{(j+1)})$;
7: $j \leftarrow j + 1$;
8: **end while**

In comparison, we discuss the direct application of the split-Bregman iterative method to DD traveltime tomography problem, which leads to two subproblems given by:

$$\tilde{\mathbf{m}}^{(k)} = \underset{\mathbf{m}}{\operatorname{argmin}}\left\{\left\|\tilde{G}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}\right\|_2^2 + \lambda_1\left\|\mathbf{w}_x - \nabla\tilde{\mathbf{m}}^{(k)} - \mathbf{b}_x^{(k)}\right\|_2^2 + \lambda_1\left\|\mathbf{w}_z - \nabla\tilde{\mathbf{m}}^{(k)} - \mathbf{b}_z^{(k)}\right\|_2^2\right\} \tag{A11}$$

$$\left(\mathbf{w}_\mathbf{x}^{(k)}, \mathbf{w}_\mathbf{z}^{(k)}\right) = \underset{\mathbf{w_x},\mathbf{w_z}}{\operatorname{argmin}}\left\{\lambda_1\left\|\mathbf{w}_x - \nabla\tilde{\mathbf{m}}^{(k)} - \mathbf{b}_x^{(k)}\right\|_2^2 + \lambda_1\left\|\mathbf{w}_z - \nabla\tilde{\mathbf{m}}^{(k)} - \mathbf{b}_z^{(k)}\right\|_2^2 + \lambda_2\left\|\tilde{\mathbf{m}}\right\|_{\mathrm{TV}}\right\}. \tag{A12}$$

Eq. (A11) solves DD traveltime tomography with two regularization terms, and eq. (A12) solves a L_2-TV minimization problem. The regularization scheme used in eq. (A11) is different from that in eq. (19) of our method with the MTV regularization scheme. Eq. (A11) uses the spatial derivatives of the model parameters as the prior. Calculations of the spatial derivatives can be inaccurate because of numerical noise and artefacts during inversion, which could lead to an unstable minimization subproblem eq. (A11). By contrast, eq. (19) employs the model inversion result of the previous iteration step $\mathbf{u}^{(k-1)}$ as the *a priori* information, yielding a more robust inversion algorithm than that based on eq. (A11).

APPENDIX B: COMPUTATIONAL COST ANALYSIS

We provide the approximate computational cost of our algorithm for solving the optimization in eq. (18). We first summarize the optimization methods for solving our new DD traveltime tomography method as in Algorithm 2.

Algorithm 2 Computational methods for solving eq. (18)

Input: $\tilde{\mathbf{u}}^{(0)}$, TOL
Output: $\tilde{\mathbf{m}}^{(k)}$
1: Initialize $k = 0$;
2: **while** $||\tilde{\mathbf{m}}^{(k)} - \tilde{\mathbf{m}}^{(k-1)}|| >$ TOL **do**
3: Solve eq. (19) for $\tilde{\mathbf{m}}^{(k)}$ using LSQR method;
4: Solve eq. (20) for $\tilde{\mathbf{u}}^{(k)}$ according to Algorithm 1;
5: $k \leftarrow k + 1$;
6: **end while**

From Algorithm 2, we can observe that the overall computational cost consists of two portions: solving eq. (19) in Step 3 and solving eq. (20) in Step 4. Because the operations in Algorithms 1 and 2 are either matrix or vector operations, we calculate the floating point operations (FLOPS) and use the big $\mathcal{O}$ notation to approximate the computational cost (Golub & Van Loan 1996).

To set up the problem, we assume that the number of ray paths is p and the number of the grid nodes to be inverted is q. The size of model parameters is $\tilde{\mathbf{m}} \in \Re^{q\times 1}$, and the sensitivity matrix $\tilde{G} \in \Re^{p\times q}$. Hence, the left-hand side of the system matrix in eq. (A1) is of size $(p + q) \times q$. The computational cost of employing the LSQR algorithm to solve the linear systems in eq. (A1) is dominated by the Lanczos bidiagonalization, which is composed of two matrix–vector multiplications (Björck 1996):

$$\mathrm{C_{LSQR}} = 2((p+q)\cdot q) + k_0\cdot(4((p+q)\cdot q) + 2(p+q)\cdot 6q) \tag{B1}$$

$$\approx k_0\cdot\mathcal{O}(p\cdot q + q^2), \tag{B2}$$

where k_0 is the number of iterations used in the LSQR algorithm. Eq. (B1) is the exact cost in FLOPS. The first term in eq. (B1) corresponds to the initial steps in the LSQR algorithm, and the second term is the cost when k_0 bidiagonalization is implemented. The cost provided by eq. (B1) is determined by the number of LSQR iterations k_0 and the cost of the matrix–vector multiplications $(p + q) \cdot q$.

In numerical linear algebra, basic linear algebra subprograms (BLAS) are divided into three levels. Level-1 operations involve an amount of data and arithmetic that is linear in the dimension of the operation, and those operations involving a quadratic amount of data and a quadratic amount of work are Level-2 operations (Golub & Van Loan 1996). Following this notation, vector dot product and SAXPY operations are examples of BLAS Level-1 operations (BLAS 1) and the matrix–vector multiplication is the BLAS Level-2 operation (BLAS 2). Considering only the BLAS 2 operations of matrix vector multiplications and ignoring all BLAS 1 operations of vector dot product and SAXPY operations, we obtain the approximate computational cost in eq. (B2) using the big $\mathcal{O}$ notation.

The computational cost of Algorithm 1 is relatively straightforward. It includes the cost from the Gauss–Seidel iteration in

eq. (A8), and the shrinkage formula in eqs (A9) and (A10). We have

$$C_{SB} \approx k_1 \cdot \mathcal{O}(q), \tag{B3}$$

where k_1 is the number of iterations used in the split-Bregman iterative method. Therefore, the overall computational cost of Algorithm 2 is

$$C_{Overall} = C_{LSQR} + C_{SB} \approx k_2 \cdot \left[k_0 \cdot \mathcal{O}(p \cdot q + q^2) + k_1 \cdot \mathcal{O}(q)\right], \tag{B4}$$

where k_2 is the total number of iterations used in Algorithm 2.

In comparison, the overall computational cost of solving the DD traveltime tomography problem with conventional Tikhonov regularization as in eq. (4) is

$$C_{Convension} = k_2 \cdot \left[k_0 \cdot \mathcal{O}(p \cdot q + q^2)\right]. \tag{B5}$$

Comparing the overall cost of using conventional DD traveltime tomography in eq. (B5) to the cost of our new inversion method as in eq. (B4), the costs only differ in the lower-order term of $k_1 \cdot \mathcal{O}(q)$. We therefore conclude that the added cost of using our new method is trivial.