Geophysical Journal International

Geophys. J. Int. (2021) **226,** 1058–1085 doi: 10.1093/gji/ggab131
Advance Access publication 2021 April 05
GJI General Geophysical Methods

Joint Gramian inversion of geophysical data with different resolution capabilities: case study in Yellowstone

Xiaolei Tu [1] and Michael S. Zhdanov [1,2]

[1] *Department of Geology and Geophysics, University of Utah, Salt Lake City, UT* 84112, *USA. E-mail:* tuxl2009@hotmail.com
[2] *TechnoImaging, Salt Lake City, UT* 84107, *USA*

Accepted 2021 April 1. Received 2021 March 28; in original form 2020 September 16

SUMMARY

Joint inversion of multiphysics data is a practical approach to the integration of geophysical data, which produces models of reduced uncertainty and improved resolution. The development of effective methods of joint inversion requires considering different resolutions of different geophysical methods. This paper presents a new framework of joint inversion of multiphysics data, which is based on a novel formulation of Gramian constraints and mitigates the difference in resolution capabilities of different geophysical methods. Our approach enforces structural similarity between different model parameters through minimizing a structural Gramian term, and it also balances the different resolutions of geophysical methods using a multiscale resampling strategy. The effectiveness of the proposed method is demonstrated by synthetic model study of joint inversion of the *P*-wave traveltime and gravity data. We apply a novel method based on Gramian constraints and multiscale resampling to jointly invert the gravity and seismic data collected in Yellowstone national Park to image the crustal magmatic system of the Yellowstone. Our results helped to produce a consistent image of the crustal magmatic system of the Yellowstone expressed both in low-density and low-velocity anomaly just beneath the Yellowstone caldera.

Key words: Gravity anomalies and Earth structure; Joint inversion; Seismic tomography; Physics of magma and magma bodies.

1 INTRODUCTION

Joint inversion of multiphysics data is an effective approach to mitigate the non-uniqueness of geophysical inverse problems and to reduce the uncertainties of inverse models (Zhdanov *et al.* 2012; Capriotti & Li 2014; Moorkamp *et al.* 2016; Giraud *et al.* 2019). It applies constraints through complementing each data set with information derived from the other data sets. Different physical fields are sensitive to separate properties and exhibit distinctive sensitivity patterns as a result of their particular governing physical laws. For example, the sensitivity kernels of seismic wave data have banana-doughnuts shapes (Yoshizawa & Kennett 2005; Liu & Tromp 2008), while the gravity field's sensitivity decreases rapidly with depth (Li & Oldenburg 1998; Zhdanov 2015; Tu & Zhdanov 2020a). We could, therefore, harness the complementary sensitivities to produce geophysical models of the subsurface target with reduced uncertainty.

Several approaches to joint inversion have been developed over the last decades. They can be classified into two categories, namely, petrophysical and structural approaches. The former assumes that there exists a known relationship between different physical parameters and correlates them via a theoretical, empirical, or statistical petrophysical relationship (e.g. Nielsen & Jacobsen 2000; Afnimar *et al.* 2002; Gao *et al.* 2012; Sun & Li 2015, 2016a,b; Giraud *et al.* 2017; Astic & Oldenburg 2019; Astic *et al.* 2020). It requires *a priori* knowledge of the specific form (i.e. linear, polynomial, exponential or logarithmic) of the relationship. The parameters of the relationship could be calculated based on well log or lab data in advance (e.g. Gao *et al.* 2012; Sun & Li 2015, 2016a; Giraud *et al.* 2017; Astic *et al.* 2020) or adaptively determined in the joint inversion (e.g. Zhdanov *et al.* 2012; Sun & Li 2016b; Lin & Zhdanov 2019). In practice, however, the specific form of the relationship might be unknown, and it may even change from one geological formation to another; the uncertainty of petrophysical relationship from existing data might be too high to contain significantly useful information, and we may not have sufficient well log or petrophysical data available. Without necessary *a priori* knowledge of the petrophysical relationship or sufficient petrophysical data, joint inversion based on petrophysical relationship could either be difficult to implement or generate biased geophysical models.

The structural approach enforces the models of different physical properties to have similar spatial structures (e.g. Gallardo & Meju 2004, 2007; Roux *et al.* 2011). It requires the structural resemblance among different parameters to be present everywhere throughout the survey area. Our developed method falls into this category. The geological setting of our case study in the Yellowstone area involves shallow loose sediments, rhyolites, basalts and partial melts. It is not practical to characterize the petrophysics using a simple velocity–density relation with the fixed parameters. In a general case of variable velocity–density relationships, joint inversion assuming structural similarity between density and velocity models is preferred.

The majority of the aforementioned methods, however, do not consider distinctive resolution capabilities of particular geophysical methods while representing models of various physical parameters with the same discretization mesh. The resolution capabilities of geophysical methods often differ significantly from each other due to different factors, such as data coverage, subsurface geology, the physical nature of geophysical field, etc. For instance, active-source seismic reflection data from a dense survey usually have better resolution than the geopotential field data (e.g. Telford *et al.* 1990; Kearey *et al.* 2013). In addition, seismic amplitudes are directly sensitive to density, as are gravity data. Seismic amplitudes can be inverted for the constitutive rock properties on a small distance scale but are unable to recover long-distance trends in those properties. Inversion of gravity data for density has an opposite complementary behaviour. In that sense, the two measurements are sensitive to different bandwidths of the same rock property. Ignoring the resolution issue may lead to (1) bad data fitting for geophysical methods with high-resolution capabilities and data overfitting for methods with low resolution; (2) artefacts of unreasonable fine (i.e. too good to be true) structures for model parameters corresponding to methods with low-resolution capabilities; and (3) slow convergence behaviour of the joint inversion iteration process (e.g. Heincke *et al.* 2017). A challenging but essential issue for joint inversion is to account for the resolution difference in a unified inversion regime without biased model parametrization. For a strict definition of the resolution of geophysical data we refer the interested reader to a large volume of the publications, including Aki *et al.* (1977), Berkhout (1984), Vasco (1989), Schuster (1996), Yao *et al.* (1999), Zhdanov (2002), Spetzler & Snieder (2004), Zhdanov & Tolstaya (2006), An (2012), Huang & Schuster (2014), Schuster (2017) and Tu & Zhdanov (2020a).

We propose a new framework for joint inversion of multiphysics data based on a novel formulation of the Gramian constraint. Our method could (1) enforce structural similarity between different model parameters through minimizing a structural Gramian term formulated as the integration of the point-wise Gramian of the gradients, and (2) honour the distinct resolution capabilities of different geophysical methods with a multiscale resampling strategy. We have demonstrated the effectiveness of the proposed method with synthetic models by jointly inverting the *P*-wave traveltime and gravity data.

We have also applied the developed method to jointly invert the gravity and seismic data collected in Yellowstone national Park to image the crustal magmatic system of the Yellowstone. Our results helped to produce a consistent image of the crustal magmatic system of the Yellowstone expressed both in low-density and low-velocity anomaly just beneath the Yellowstone caldera.

2 JOINT INVERSION METHODOLOGY

2.1 Forward modelling

The forward modelling problems could be represented in matrix form as follows:

$$\mathbf{d}^{(j)} = \mathbf{A}^{(j)}(\mathbf{m}^{(j)}),\ j = 1, 2, \tag{1}$$

where $\mathbf{d}^{(1)}$ and $\mathbf{d}^{(2)}$ represent the column vectors of gravity and seismic traveltime data; $\mathbf{m}^{(1)}$ and $\mathbf{m}^{(2)}$ denote vectors of subsurface density and *P*-wave velocity models; $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ denote the forward operators for gravity field and seismic traveltimes, respectively.

The gravity field is calculated by the integral representation of the gravitational field with the point mass approximation method (Zhdanov 2002, 2015; Čuma *et al.* 2012). The gravity modelling is fully parallelized with message passing interface (MPI) and OpenMP to speed up the computation and to reduce the required computer memory, which is crucial for a regional-scale survey (Čuma *et al.* 2012).

We solve the isotropic eikonal equation for seismic traveltimes. A grid-based solver, namely the multistage fast marching (FMM) method, is employed to solve the eikonal equation. Our implementation of the FMM method is based on the FM3D package (De Kool *et al.* 2006; Rawlinson *et al.* 2006). We also parallelize it by earthquake event with MPI to enable rapid computation on the supercomputer.

The reliability and efficiency of the two forward solvers have been demonstrated in the corresponding references and will not be repeated here. We focus on the joint inversion algorithm in the following sections.

2.2 Parametrization

In order to address the issue of different resolution capabilities, we decouple the forward modelling meshes for different physical properties. Seismic waves usually have a better resolution than the gravity field. We, therefore, set the velocity on the mesh, which is much finer than the density mesh for forward modelling, so that the details of the velocity model could be sampled by seismic waves and that the density model is not oversampled. An oversampled density model may not only increase the computation time of forward modelling but also may introduce fictitious model details that could not be resolved by gravity data. The decoupled forward modelling meshes are not only essential for computational efficiency, but also are necessary for honouring the different resolution capabilities of seismic waves and gravity fields. At the same time, in order to jointly invert seismic and gravity data, the velocity and density models should be coupled at the same scale, especially, if the structural resemblance between them is considered. In this case, we should match only the long-wavelength structures of the velocity model with the density model, as the fine structures of short wavelength are beyond the resolution of the gravity field. Thus, a parametrization scheme enabling multiscale resampling of the model parameters is desirable.

In a general case, we can represent the spatial variations of density and velocity by projecting them to the associated interpolant-type subspace. This representation offers a possibility of describing relatively smooth variations of density and velocity with a few parameters (Rawlinson *et al.* 2008). There are many types of interpolants that can be used for subspace representation, for example, pseudo-linear (Eberhart-Phillips 1986), splines under tension (Neele *et al.* 1993; VanDecar *et al.* 1995) and Fourier spectrum (Dziewonski *et al.* 1977) interpolants. Cubic B-spline interpolant is particularly prevalent in geophysics since it offers C^2 continuity, local control, and the potential for an irregular distribution of knots (Rawlinson *et al.* 2008). Any model, $\mathbf{m}$, in the original high-dimensional space can be represented uniquely as a linear combination of the cubic spline functions (basis) as follows:

$$\mathbf{m} = \mathbf{L}\mathbf{m}_s, \tag{2}$$

where $\mathbf{m}_s$ denotes a vector in the subspace with its elements defined discretely at the knot positions; $\mathbf{L}$ is the matrix constituted of cubic B-spline functions. Matrix $\mathbf{L}$ is controlled by the knots and resampling mesh. The action of $\mathbf{L}$ on vector $\mathbf{m}_s$ would resample the model to the corresponding mesh. The choice of mesh sizes or model discretization and parametrization is very important in the joint inversion approach. For simplicity, we selected constant mesh size for modelling gravity data. In principle, a varying mesh size could be used both for the density mesh and for the structural coupling mesh. It is important to note, however, that, in the framework of the developed approach, all models are characterized by 3-D continuous functions in the subsurface using subspace representation. The smallest modelling feature could be represented by a set of knots controlled by the order of the spline functions and the mesh size. For cubic spline functions, as implemented in this paper, one quarter of the mesh size is approximately the low limit. To determine the mesh size for subspace representation, we need first define the smallest scale we want to recover in the inversion. For example, in our synthetic study, the finest structures we want to recover are ball shaped anomalies with diameter of 2 km. We could therefore safely set the intervals of the knots as 4 km × 4 km × 2 km in the X, Y and Z directions.

As illustrated in Fig. 1, the models are resampled at three scales with three meshes. For forward modelling, we resample the density model with a coarse mesh (a), and the velocity model with a fine mesh (b). The third coarse mesh is employed to resample both the density and velocity models so that they are coupled at the same scale, (d) and (e). The third mesh should not be finer than the coarse mesh (a) so that only the long-wavelength structures resolvable by gravity field are matched between density and velocity models. We set the third mesh the same as mesh (a) in Fig. 1, for simplicity.

2.3 Gramian based structural coupling

A fundamental issue of joint inversion is to define an appropriate coupling term. We propose a novel structural coupling term based on the Gramian constraint (Zhdanov *et al.* 2012; Zhdanov 2015):

$$s(\mathbf{m}_s^{(1)}, \mathbf{m}_s^{(2)}) = \iiint_D G\left(\nabla\left(\mathbf{T}^{(1)}\mathbf{L}_3\mathbf{m}_s^{(1)}(\mathbf{r})\right), \nabla\left(\mathbf{T}^{(2)}\mathbf{L}_3\mathbf{m}_s^{(2)}(\mathbf{r})\right)\right)\mathrm{d}v, \tag{3}$$

where $\nabla = [\nabla_x, \nabla_y, \nabla_z]^T$ denotes the gradient operator. Operators $\mathbf{T}^{(1)}$ and $\mathbf{T}^{(2)}$ transform the models into another space. For example, density and velocity could be transformed into the logarithmic space, as shown by Lin & Zhdanov (2019). In our synthetic study, we transform velocity to the velocity perturbation to promote structural resemblance between velocity perturbation and density. The matrix of cubic B-spline basis, $\mathbf{L_3}$, resamples the models with the same coarse mesh so that the two models are coupled at the same scale. Symbol $G(,)$ represents the Gramian (i.e. the determinant of the Gram matrix) of the gradient at a single point:

$$G = \begin{vmatrix} \left(\nabla\left(\mathbf{T}^{(1)}\mathbf{L}_3\mathbf{m}_s^{(1)}\right), \nabla\left(\mathbf{T}^{(1)}\mathbf{L}_3\mathbf{m}_s^{(1)}\right)\right) & \left(\nabla\left(\mathbf{T}^{(1)}\mathbf{L}_3\mathbf{m}_s^{(1)}\right), \nabla\left(\mathbf{T}^{(2)}\mathbf{L}_3\mathbf{m}_s^{(2)}\right)\right) \\ \left(\nabla\left(\mathbf{T}^{(2)}\mathbf{L}_3\mathbf{m}_s^{(2)}\right), \nabla\left(\mathbf{T}^{(1)}\mathbf{L}_3\mathbf{m}_s^{(1)}\right)\right) & \left(\nabla\left(\mathbf{T}^{(2)}\mathbf{L}_3\mathbf{m}_s^{(2)}\right), \nabla\left(\mathbf{T}^{(2)}\mathbf{L}_3\mathbf{m}_s^{(2)}\right)\right) \end{vmatrix}, \tag{4}$$

where (,) denotes the inner product. As proved in Appendix A, the Gramian, G, is non-negative and it is zero if and only if the two gradient vectors are parallel. Therefore, by minimizing G, we would enforce the gradient vectors of the model parameters to be mutually parallel at a specific subsurface position. The important property of the Gramian constrain presented in formula (3) is that it is formulated as the integration of the point-wise Gramian of the gradient, which enforces two gradients to be parallel everywhere in the subsurface, similar to the cross-gradient approach.

Indeed, it could be demonstrated (see Appendix B) that the Gramian G is equal to the L_2 norm of the cross-product of two gradient terms (Gallardo & Meju 2004, 2007):

$$G\left(\nabla\mathbf{m}_s^{(1)}(\mathbf{r}), \nabla\mathbf{m}_s^{(2)}(\mathbf{r})\right) = \left|\left|\nabla\mathbf{m}_s^{(1)}(\mathbf{r}) \times \nabla\mathbf{m}_s^{(2)}(\mathbf{r})\right|\right|^2. \tag{5}$$

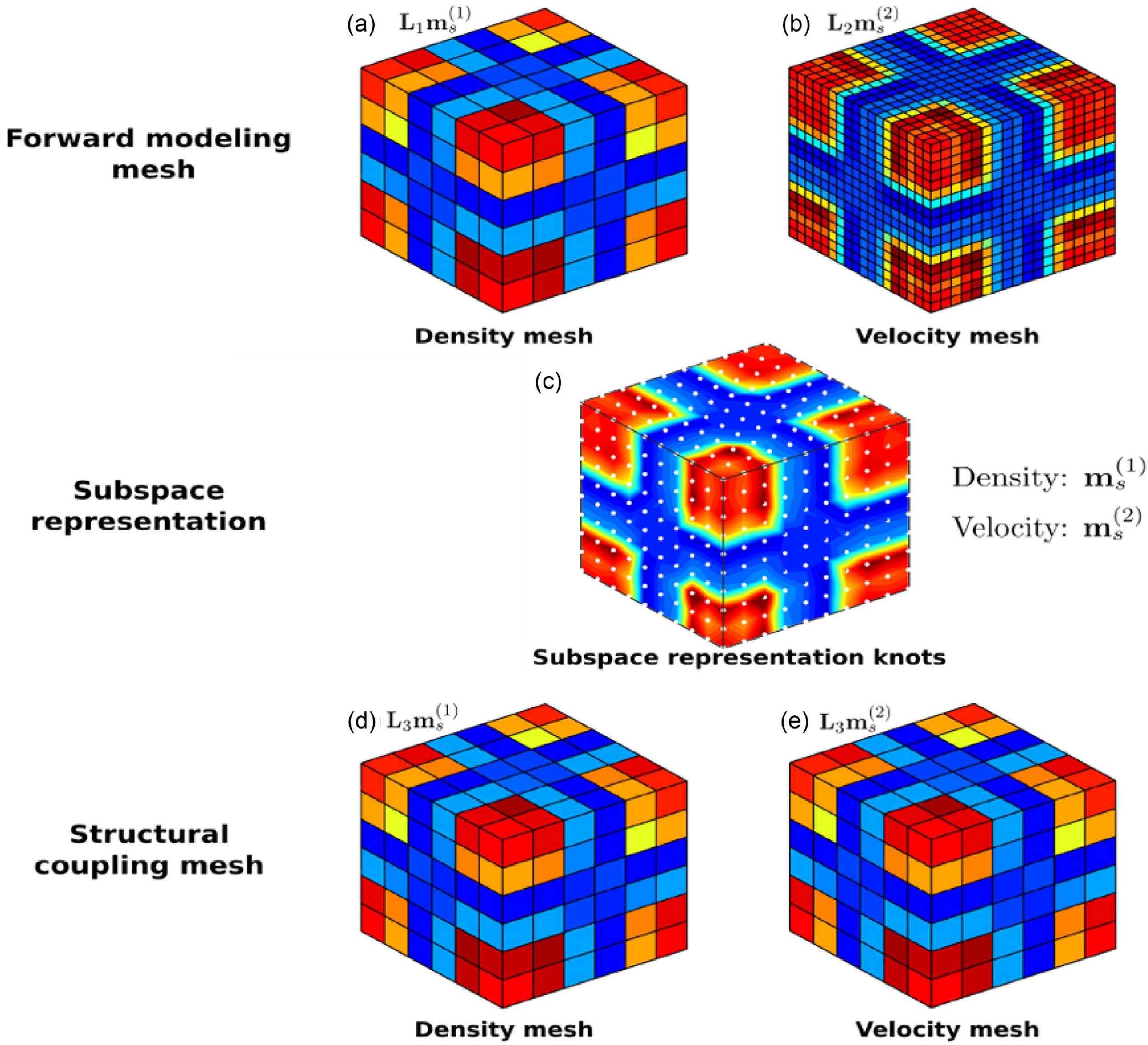

Figure 1. Schematic representation of the multiscale resampling of the density and velocity models for forward modelling and structural coupling. The density and velocity models are assumed to be continuum in a subspace of cubic B-spline functions. The white dots in (c) denote the knots of the subspace representation. The models are resampled by a coarse mesh (a) and a fine mesh (b) to honour the different resolution capabilities of gravity and seismic data. The models are resampled at the same third mesh (d and e) to facilitate the structural coupling between them. This mesh is less finer than or the same as mesh (a) so that only the long wavelength structures of velocity model are preserved and matched with density model.

Therefore, by minimizing the Gramian G we enforce two gradients to be parallel, similar to the cross-gradient constraint of Gallardo & Meju (2004, 2007). The conventional cross-gradient constraint, however, is based on minimizing the absolute value of the cross-product of the gradients, which is a non-quadratic problem. The Gramian constraint is (1) easier to implement due to its quadratic form, similar to L_2 norm stabilizers; (2) more convenient to be incorporated in optimization; (3) more flexible for manipulation of the model parameters in different transformed spaces; and (4) more flexible to incorporate data of different resolution capabilities.

The structural coupling term, $s(\mathbf{m}_s^{(1)}, \mathbf{m}_s^{(2)})$, in eq. (3) is an integral (superposition) of the non-negative Gramians. It would be minimized only if the Gramians, G, are all minimized at every subsurface position, leading to close to zero Gramians and, consequently, to mutually parallel gradient vectors everywhere in the whole inversion domain, D. Therefore, we can enforce the gradient vectors of the model parameter $\mathbf{m}_s^{(1)}$ and $\mathbf{m}_s^{(2)}$ to be mutually parallel all over the inversion domain by minimizing $s(\mathbf{m}_s^{(1)}, \mathbf{m}_s^{(2)})$. As with the cross-gradient method, aligned gradient orientations of different model parameters necessitate the coincidence of structural boundaries of them. The Gramian constraint, s, is a metric in Gramian space of the structural similarity between the two models (Zhdanov 2015). The smaller the s is, the more structural similarity is between them. The structural similarity of the two model parameters is the strongest when s achieves its minimum.

2.4 Parametric functional

The joint inversion of gravity and seismic traveltime data is performed by minimizing the following parametric functional (Zhdanov *et al.* 2012; Zhdanov 2015):

$$P(\mathbf{m}_s^{(1)}, \mathbf{m}_s^{(2)}, \alpha^{(1)}, \alpha^{(2)}, \beta) = \sum_{j=1}^{2} \phi^{(j)}(\mathbf{m}_s^{(j)}) + \sum_{j=1}^{2} \alpha^{(j)} \psi^{(j)}(\mathbf{m}_s^{(j)}) + \beta s(\mathbf{m}_s^{(1)}, \mathbf{m}_s^{(2)}) \longrightarrow \min, \tag{6}$$

where $\phi^{(j)}$ denotes the misfit functional for the j-th type of data; $\psi^{(j)}$ represents the regularization stabilizer promoting preferred structures of the model; $s(\mathbf{m}_s^{(1)}, \mathbf{m}_s^{(2)})$ is the joint stabilizer enforcing structural coupling. Coefficients $\alpha^{(j)}$ and β are the regularization parameters balancing the misfits and the corresponding stabilizers.

We should note that, the least-squares (L_2 norm) error metric based on Gaussian uncertainty assumption often leads to biased models (Tarantola 2005) because the probability density functions of data and model uncertainties are usually characterized by long-tailed distribution (Claerbout & Muir 1973). This issue is exceptionally severe for seismic traveltimes, which always contain pick errors. The error metrics, which are less sensitive to large measurement errors and are more appropriate to long-tailed probability density functions, could provide a far more stable estimation of the model parameters than L_2 norm (Guitton & Symes 2003). In this paper, we employ the robust norms for the misfit functional:

$$\phi^{(j)}(\mathbf{m}_s^{(j)}) = \left|\left|\mathbf{W}_d^{(j)}\left[\mathbf{A}^{(j)}(\mathbf{L}_j \mathbf{m}_s^{(j)}) - \mathbf{d}_o^{(j)}\right]\right|\right|_\rho^2, \tag{7}$$

where $\mathbf{W}_d^{(j)}$ represents the corresponding data weighting matrix; $\mathbf{d}_o^{(j)}$ is the observed data, that is, gravity data or P-wave first arrivals; $||\cdot||_\rho^2$ denotes the robust norm, for example, Huber or Bisquare norm. The robust norms could be easily represented as quasi-quadratic functionals (Zhdanov 2015; Tu & Zhdanov 2020b), making it convenient to optimize.

We could also incorporate *a priori* information into the inversion using the minimum norm (MN) stabilizer,

$$\psi^{(j)}(\mathbf{m}_s^{(j)}) = \left|\left|\mathbf{W}_m^{(j)}(\mathbf{m}_s^{(j)} - \mathbf{m}_{apr}^{(j)})\right|\right|^2, \tag{8}$$

which would favour a model close to a prior model, $\mathbf{m}_{apr}^{(j)}$, as long as the data are equally fitted. The MN stabilizer requires a relatively smooth behaviour of the recovered model, leading to the simplest model structures. The model weights $\mathbf{W}_m^{(j)}$ in eq. (8) and data weights $\mathbf{W}_d^{(j)}$ in eq. (7) are determined based on the integrated sensitivity (Zhdanov 2002, 2015; Zhu 2017) to provide an equal sensitivity of the different components of observed data to the cells located at different depths and horizontal positions.

The general methods of solving the above inverse problems were developed in Zhdanov (2002, 2015). We base our solution on the reweighted regularized conjugate gradient method (RRCG), which is easier to implement numerically. Implementation details of the RRCG method for joint inversion problems could be found in Zhdanov *et al.* (2012), Zhu (2017) and Lin & Zhdanov (2019). One of the critical aspects of the regularized inversion is the proper choice of regularization parameters (coefficients α and β). As we solve the minimization problem (6) with the RRCG, the regularization parameter is adaptively determined by the RRCG method itself. The basic idea is that it is calculated at the first iteration so that the misfit and regularization terms are equally weighted; and then gradually decreases in the following iterations so that the regularization is not overwhelmed. We describe the computational process in Appendix C. A summary of the RRCG method for joint inversion problem is also included in Appendix C. The first variation of the Gramian constraint (3) is required for the RRCG method. We present the corresponding derivation in Appendix D.

3 NUMERICAL MODEL STUDY

3.1 Model 1

The developed method was tested using computer-generated data. Model 1 represents a homogeneous half-space containing four blocky anomalies with an equal size of 16 km × 16 km × 8 km in east, north and depth. The depths of the anomalies are also the same: 6 km to the top and 14 km to the bottom. Fig. 2 presents horizontal sections of the true density contrast and velocity perturbation at a depth of 10 km. The density distribution inside each anomaly is homogeneous and even. The corresponding velocity perturbation was designed to exhibit fine patterns, which could be resolved by the seismic waves while beyond the resolution capability of the gravity field. Four tiny ball-shaped high-velocity anomalies with radius of 1 km were also set for the velocity model at a depth of 20 km. There is no density anomaly at positions corresponding to the four tiny high-velocity balls. The background P-wave velocity of the half-space is 6 km s^{-1}. A joint gravity and seismic survey consisting of 484 gravity stations and 225 seismic stations were designed to map the anomalies. Eight hundred random but evenly distributed earthquakes were modelled to generate the traveltime data. The distribution of earthquake epicentres and gravity and seismic stations are also shown in Fig. 2. The computer-simulated gravity data were contaminated with random noise with a standard derivation of 0.1 mGal, which corresponds to a relative noise level of 3.6 per cent. The computed traveltimes also contain random noise with a level of 7 ms (i.e. 5 per cent).

We first inverted the gravity and seismic traveltime data separately. The same data were then jointly inverted using the developed method with multiscale resampling set-ups shown in Table 1. The knot/mesh set-ups for the separate inversions are the same as for joint inversion.

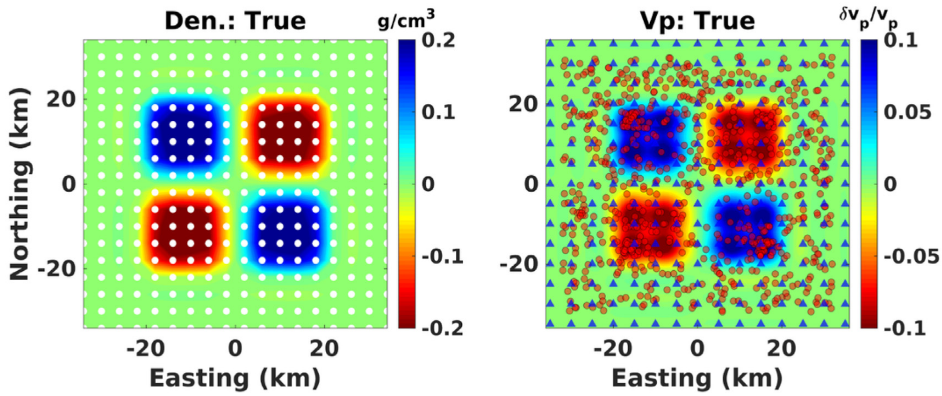

Figure 2. Horizontal sections of true density (left) and velocity perturbation (right) models at a depth of 10 km. The white dots denote the gravity stations. The red dots are earthquake epicentres. The seismic stations are represented by blue triangles.

Table 1. Set-up of knots and meshes with multiscale resampling for Model 1.

Knots/meshes	Intervals/Cell size East × North × Depth	Num. of knots/cells East × North × Depth
Knots of subspace representation	4 km × 4 km × 2 km	21 × 21 × 16
Density mesh forward modelling	4 km × 4 km × 2 km	18 × 18 × 13
Velocity mesh forward modelling	1 km × 1 km × 1 km	73 × 73 × 27
Density and velocity mesh structural coupling	4 km × 4 km × 2 km	18 × 18 × 13

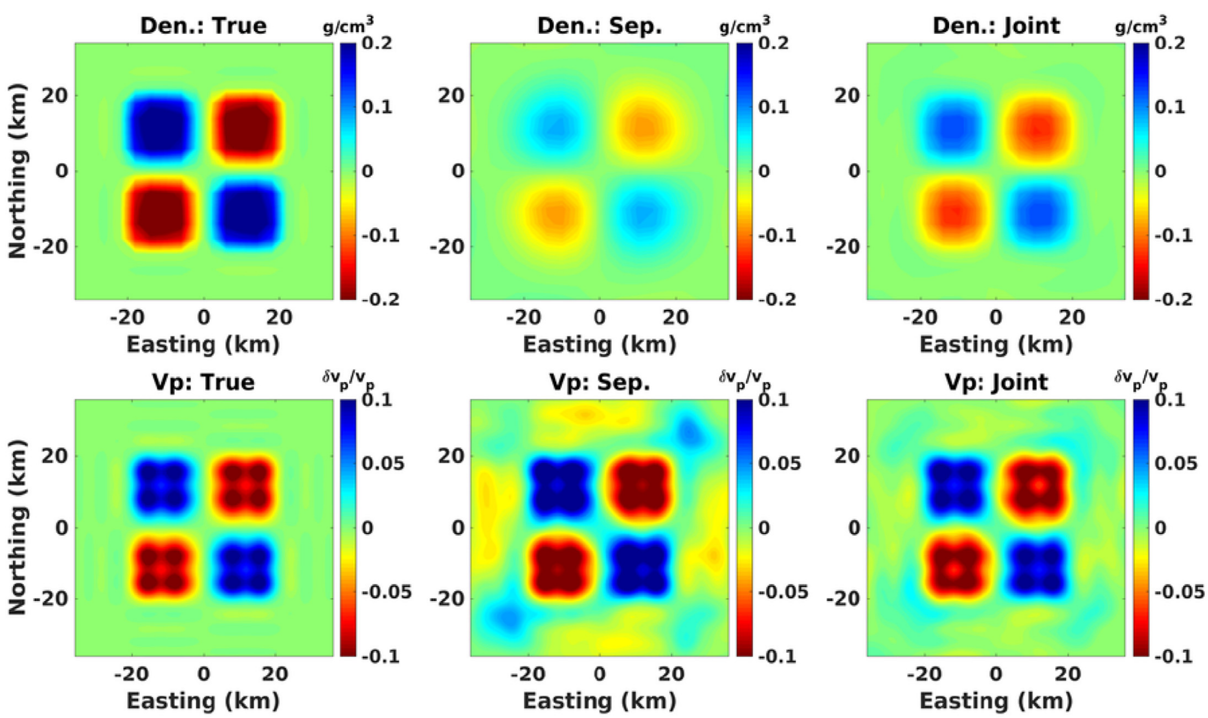

Figure 3. Horizontal sections of the true (left panels), separately inverted (middle panels) and jointly inverted (right panels) density (top panels) and velocity perturbation (bottom panels) models at a depth of 10 km.

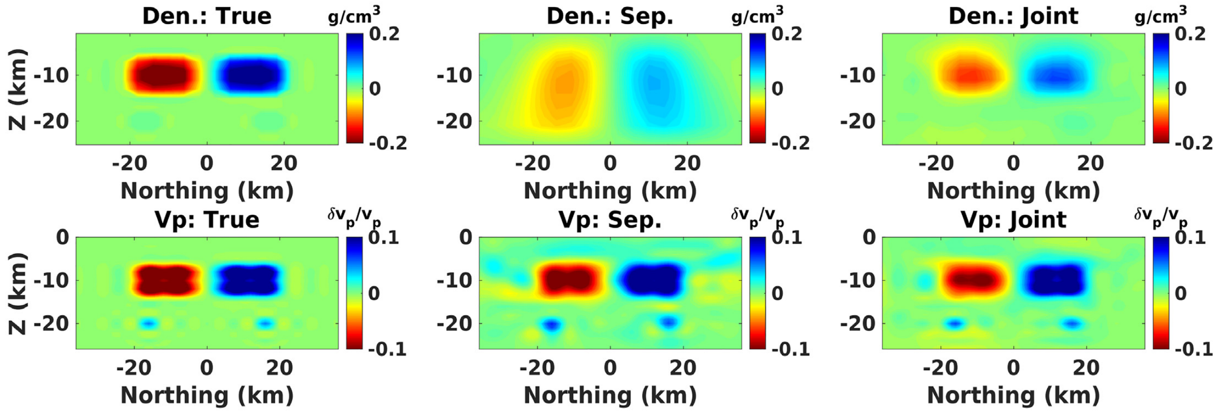

Figure 4. Vertical sections of the true (left panels), separately inverted (middle panels) and jointly inverted (right panels) density (top panels) and velocity perturbation (bottom panels) models corresponding to $X = -16$ km.

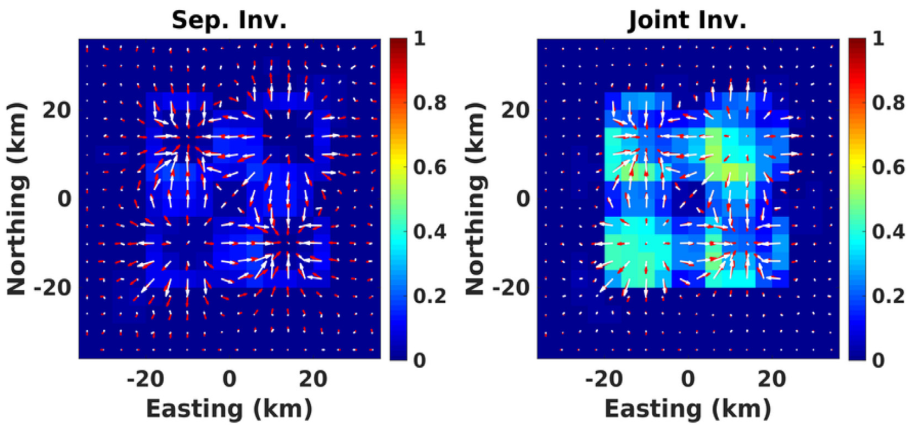

Figure 5. Horizontal sections of structural similarity and gradient vectors of the inverse model parameters at a depth of 10 km. The red and white arrows denote the gradients of density and velocity perturbation, respectively.

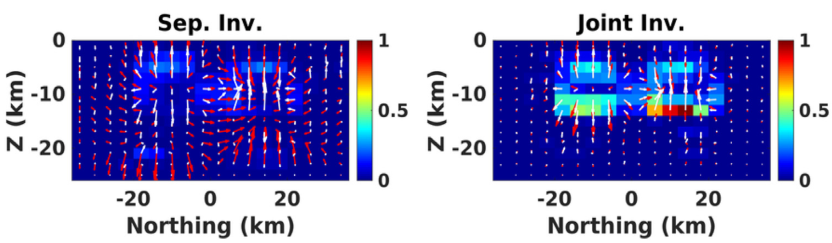

Figure 6. Vertical sections of structural similarity and gradient vectors of the model parameters for a profile at $X = -16$ km.The red and white arrows denote the gradients of density and velocity perturbation, respectively.

Comparisons of the true and inverted models are presented in Figs 3 and 4. The joint inversion recovers the density model much better compared to separate inversion. The artefacts in the velocity model are also significantly reduced in joint inversion. The effectiveness of the joint inversion could also be demonstrated in the maps of structural similarity and gradient vectors of the model parameters shown in Figs 5 and 6. The joint inversion produces models with improved structural similarity and better-aligned density and velocity gradients. Most importantly, by using the multiscale resampling, the joint inversion does not introduce artefact of overcoupling to the density model.

Both the separate and joint inversions converge to the data noise level. The maps of gravity data fitting is shown in Fig. 7. Both separately and jointly inverted models predict the gravity field well below noise level. Fig. 8 present the histograms of the observed traveltimes and the residuals obtained after the inversions. The separate and joint inversions fit the traveltimes equally well.

For comparison, we then ran the same joint inversion with the same structural coupling method but without multiscale resampling. The inversion set-ups are shown in Table 2. The gravity and seismic traveltime data were modelled with the same fine mesh, and the density and velocity models were structurally coupled with the same mesh. The inverted models are shown in Figs 9 and 10. Both jointly inverted density and velocity models show significant improvement compared to the separately inverted ones. The fine structures inside each anomaly are also observed in the jointly recovered density model, which, however, are too good to be true, since such fine-scale structures are beyond the

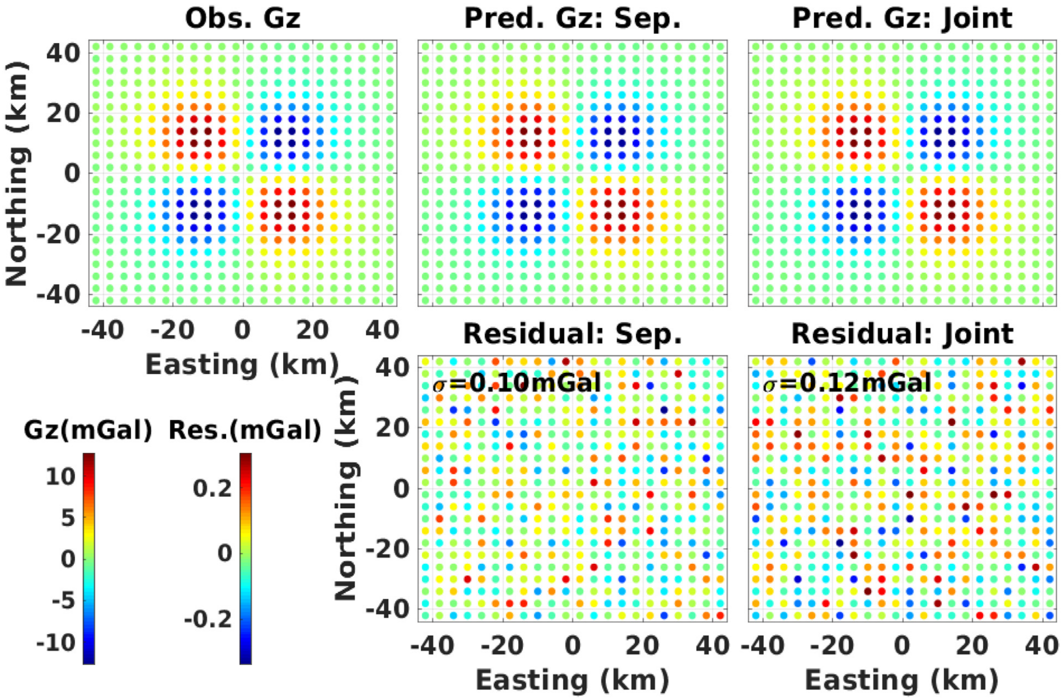

Figure 7. Maps of gravity data fitting. The observed gravity field, and those predicted by separate and joint inversions are shown in the top panels. The residuals between the observed and predicted data are shown in the bottom panels. Note that the gravity data and their residuals (abbreviated as Res.) are shown in different colour scales. We also print the RMS misfit in the top left corner of each residual map.

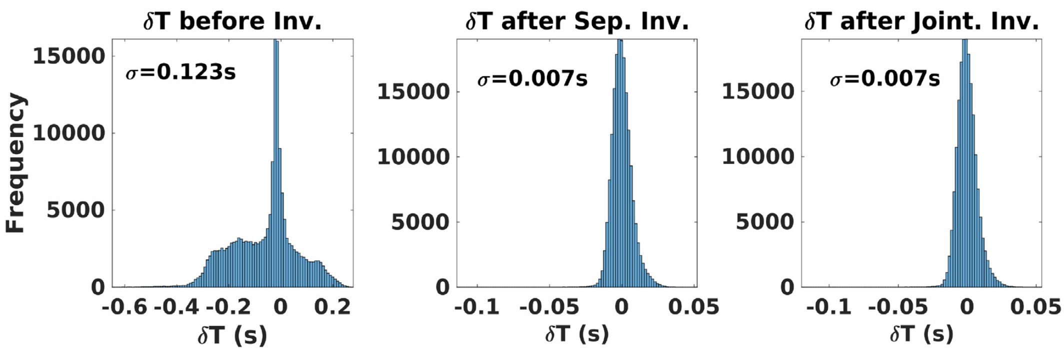

Figure 8. Histograms of the observed traveltime residuals (left panel) and the residuals obtained after the after separate (middle panel) and joint (right panel) inversions. The corresponding RMS misfit is shown in the top left corner of each panel.

Table 2. Set-up of knots and meshes without multiscale resampling for Model 1.

Knots/meshes	Intervals/Cell size East × North × Depth	Num. of knots/cells East × North × Depth
Knots of subspace representation	4 km × 4 km × 2 km	21 × 21 × 16
Density mesh forward modelling	1 km × 1 km × 1 km	73 × 73 × 27
Velocity mesh forward modelling	1 km × 1 *km* × 1 km	73 × 73 × 27
Density and velocity mesh structural coupling	1 km × 1 km × 1 km	73 × 73 × 27

resolution of the gravity data. Besides, the joint inversion also recovered small ball-shaped density anomalies at the depth of 20 km, which are artefacts caused by over coupling between density and velocity models.

As demonstrated in the maps of structural similarity in Figs 11 and 12, the joint inverted density and velocity models are structurally coupled at both long and short wavelength scale, leading to fine-scale artefacts in the density model. Especially at a depth of 20 km, the four tiny ball-shaped anomalies are also presented in the density model as a result of overcoupling. The joint inversion results with multiscale

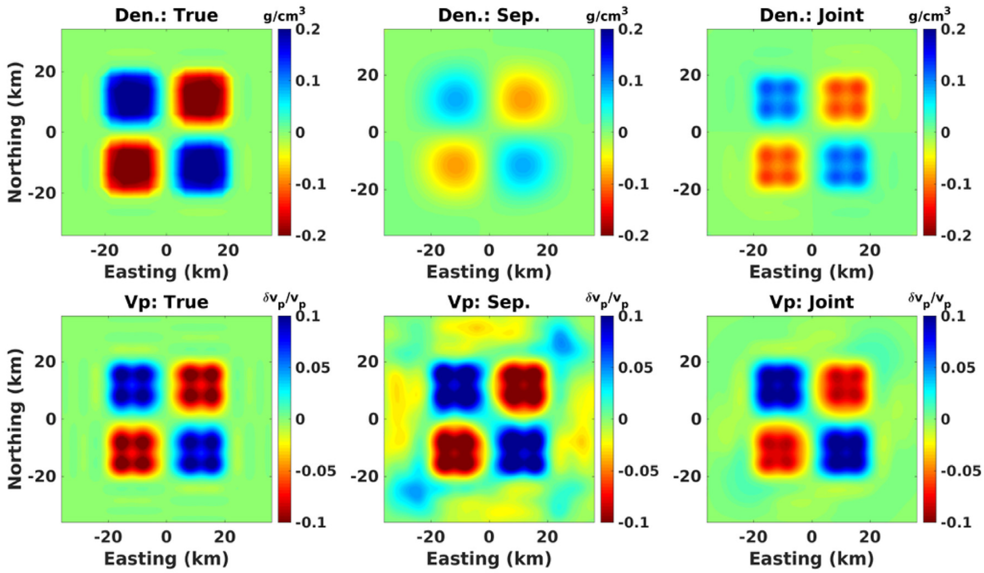

Figure 9. Results of the inversions with the same structural coupling method but without multiscale resampling. Horizontal sections of the true (left panels), separately inverted (middle panels) and jointly inverted (right panels) density (top panels) and velocity perturbation (bottom panels) models at a depth of 10 km.

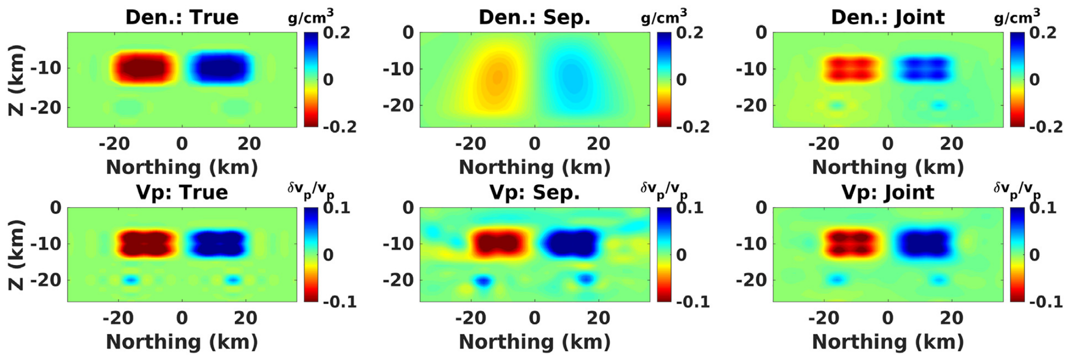

Figure 10. Results of the inversions with the same structural coupling method but without multiscale resampling. Vertical sections of the true (left panels), separately inverted (middle panels) and jointly inverted (right panels) density (top panels) and velocity perturbation (bottom panels) models corresponding to $X = -16$ km.

resampling, however, are free of such over coupling artefacts, which demonstrates the importance of accounting for the difference of resolution capabilities of different geophysical methods in joint inversion.

Note that in these inversions without multiscale resampling, the gravity and traveltime data are fitted appropriately to the same level as that of the inversions with multiscale resampling. We present the data fitting maps of gravity field and histograms of traveltime residuals in Figs 13 and 14.

3.2 Model 2

The second model's *P*-wave velocity distribution is the same as in Model 1; however, Model 2 contains two high density anomalies only. As presented in a horizontal section in Fig. 15, there is no corresponding density anomaly at positions of the low velocity. By designing such decoupled structures, we want to investigate whether the developed joint inversion method would introduce artefact in structurally decoupled targets. The survey configurations are the same as that of Model 1, as also shown in Fig. 15. We have again contaminated the computer-simulated gravity data with 5 per cent of Gaussian noise, which corresponds to a noise level of 0.24 mGal. The traveltimes are the same as those in Model 1.

We ran the separate inversions and joint inversion with multiscale resampling strategy using the same set-up as in Table 1. All inversions converged to the designed noise level. We present the inversion results in Figs 16 and 17. Both density and velocity models are well recovered in separate inversions. The jointly inverted density model is significantly improved in both density values and in the depths of the anomalies. The joint inversion also recovered a velocity model with less artefacts. This results illustrate again the benefits of joint inversion by mutually constraining model parameters with each other.

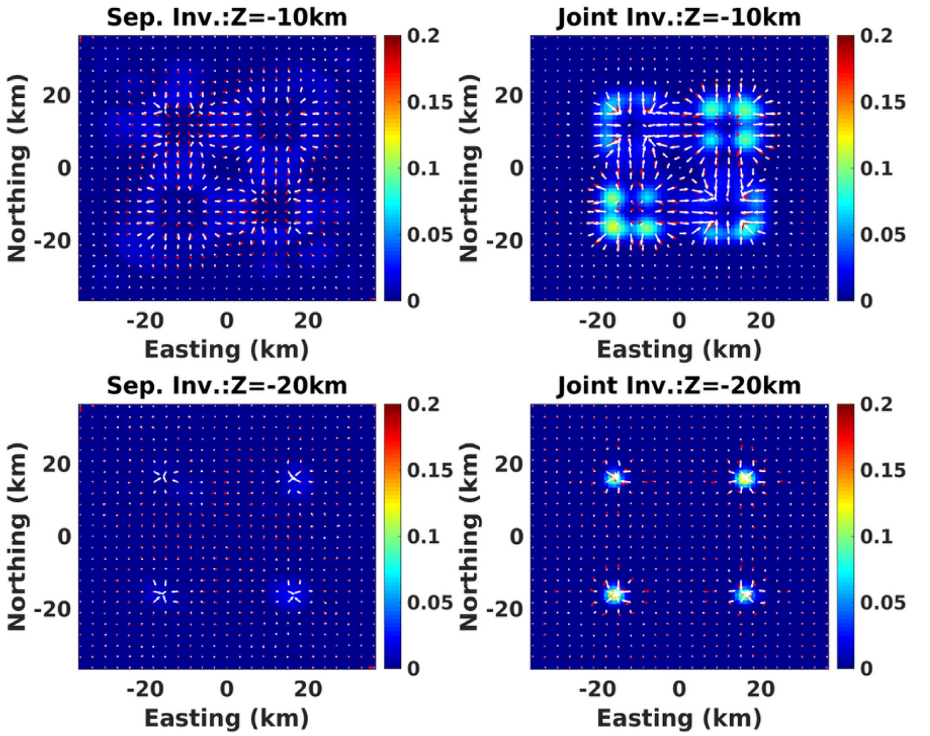

Figure 11. Results of the inversions with the same structural coupling method but without multiscale resampling. Horizontal sections of structural similarity and gradient vectors of model parameters at the depths of 10 and 20 km. The red and white arrows denote the gradients of density and velocity perturbation, respectively.

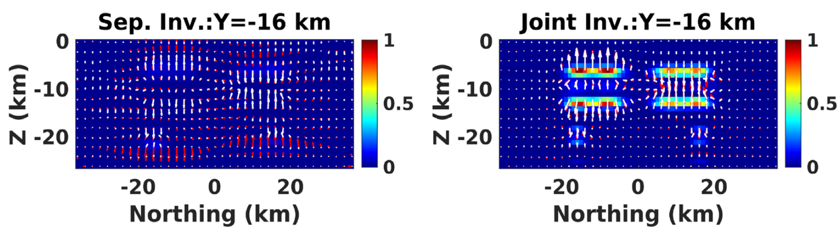

Figure 12. Results of the inversions with the same structural coupling method but without multiscale resampling. Vertical sections of structural similarity and gradient vectors of model parameters for a vertical profile corresponding to $X = -16$ km. The red and white arrows denote the gradients of density and velocity perturbation, respectively.

Another important observation is that there is no density artefact at locations corresponding to the low-velocity anomalies. As further demonstrated in the maps of structural similarity and gradient vectors of model parameters (i.e. Figs 18 and 19), the density and velocity perturbation are decoupled where the structural coupling does not exist; they are only coupled where structural coupling is supported by the data. It shows that the developed joint inversion method is free from coupling artefact when it is properly implemented.

We also present the data fitting maps for gravity field and histograms of traveltime residuals in Figs 20 and 21, which again show a good fitting of both data sets.

4 STUDYING THE YELLOWSTONE CRUSTAL MAGMATIC SYSTEM WITH GRAVITY AND SEISMIC DATA

The Yellowstone volcanic field is located at the eastern end of the Snake River Plain, which was created as the North American Plate moved southwestward across a mantle plume (Schmandt *et al.* 2012; Huang *et al.* 2015). The Yellowstone volcanic field is characterized by hotspots, extensive earthquakes, ground deformation (Farrell *et al.* 2010; Huang *et al.* 2015) and contains approximately 50 per cent of the world's hydrothermal features (Hurwitz & Lowenstern 2014).

Global tomographic studies show that the youthful Yellowstone volcanic field is fed by a west–northwest-dipping plume that extends from the mid mantle to ∼50 km depth (Yuan & Dueker 2005; Smith *et al.* 2009; Schmandt *et al.* 2012; Porritt *et al.* 2014) which in turn provides basaltic magma that fuels the Yellowstone basaltic/rhyolitic crustal magma reservoir (see Fig. 22). Studies using local earthquake tomography reveal that the Yellowstone crustal magma reservoir extends well beyond the Yellowstone Caldera. A low-velocity body (LVB)

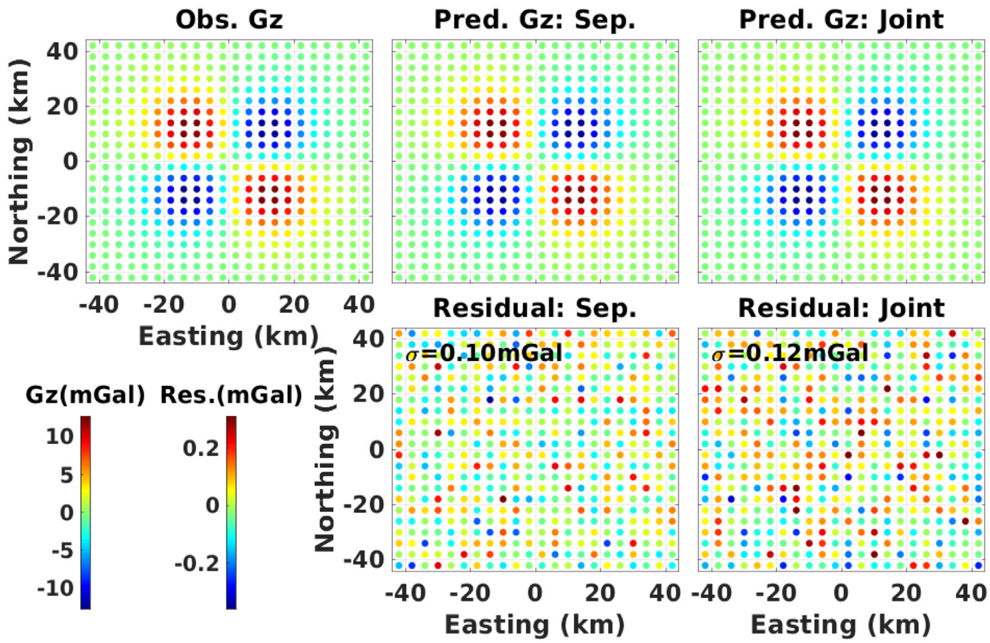

Figure 13. Results of the inversions with the same structural coupling method but without multiscale resampling. Maps of gravity data fitting. The observed gravity field, and those predicted by separate and joint inversions are shown in the top panels. The residuals between the observed and predicted data are shown in the bottom panels. The RMS misfit is also printed in the top left corner of each residual panel.

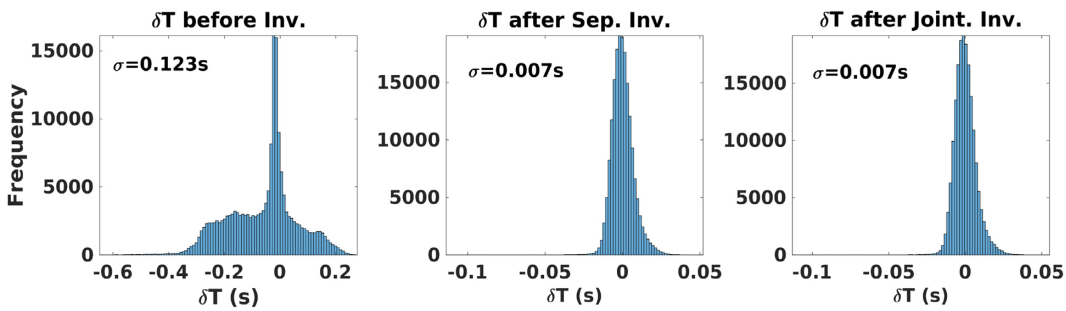

Figure 14. Results of the inversions with the same structural coupling method but without multiscale resampling. Histograms of the observed traveltime residuals (left panel) and the residuals obtained after the after separate (middle panel) and joint (right panel) inversions. The corresponding RMS misfit is shown in the top left corner of each panel.

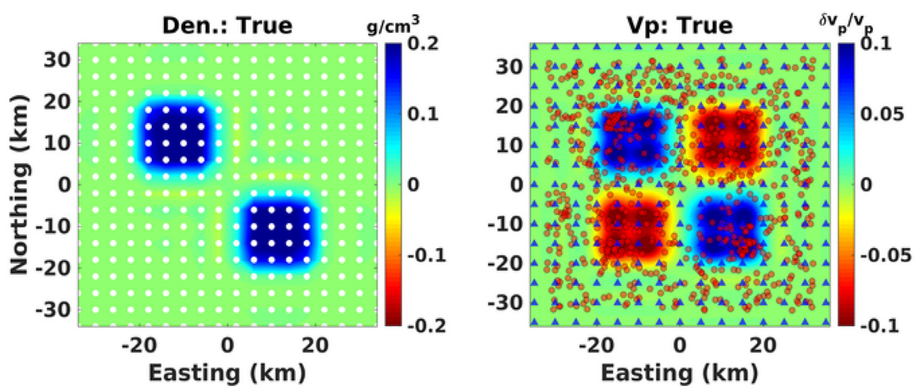

Figure 15. Horizontal sections of true density (left) and velocity perturbation (right) models at a depth of 10 km for Model 2. The white dots denote the gravity stations. The red dots are earthquake epicentres. The seismic stations are represented by blue triangles.

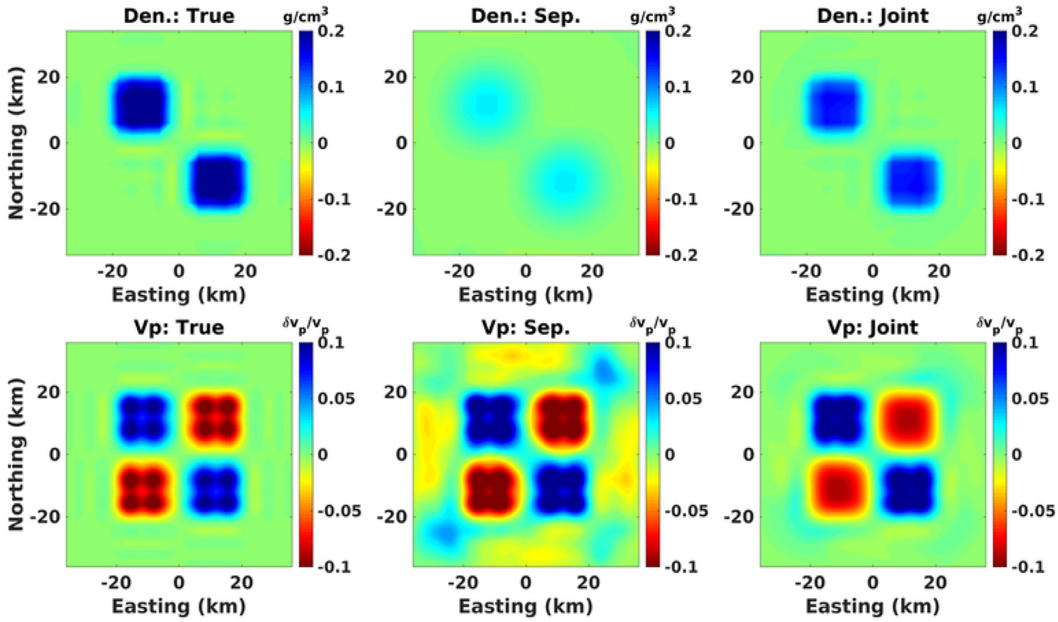

Figure 16. Horizontal sections of the true (left panels), separately inverted (middle panels) and jointly inverted (right panels) density (top panels) and velocity perturbation (bottom panels) models at a depth of 10 km for Model 2.

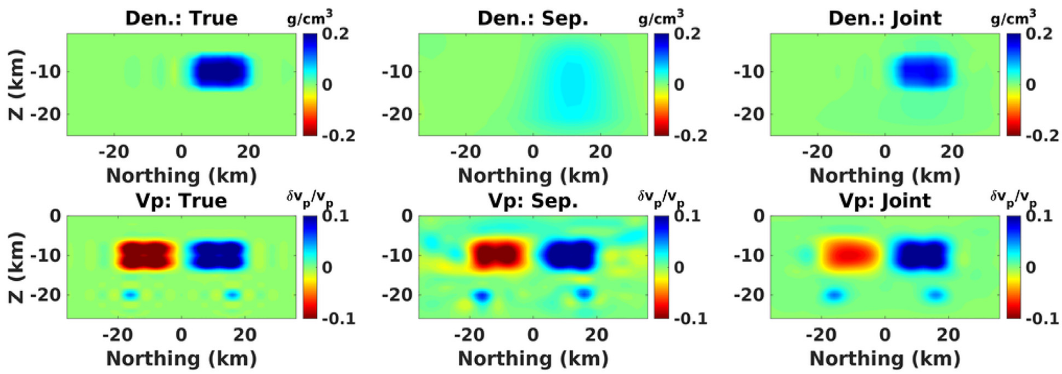

Figure 17. Vertical sections of the true (left panels), separately inverted (middle panels) and jointly inverted (right panels) density (top panels) and velocity perturbation (bottom panels) models corresponding to $X = -16$ km for Model 2.

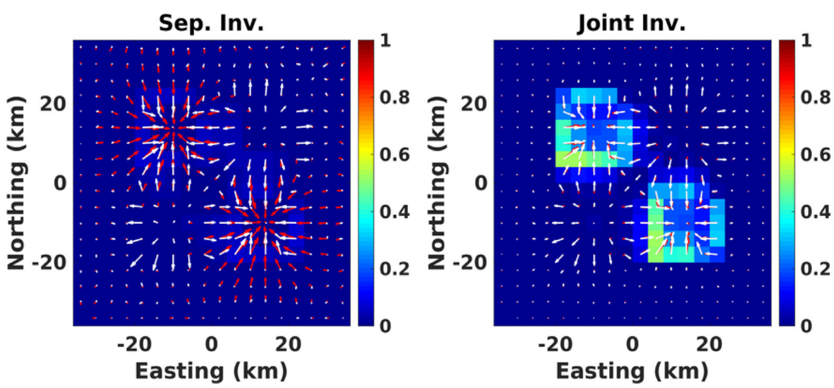

Figure 18. Horizontal sections of structural similarity and gradient vectors of model parameters at a depth of 10 km for Model 2. The red and white arrows denote the gradients of density and velocity perturbation, respectively.

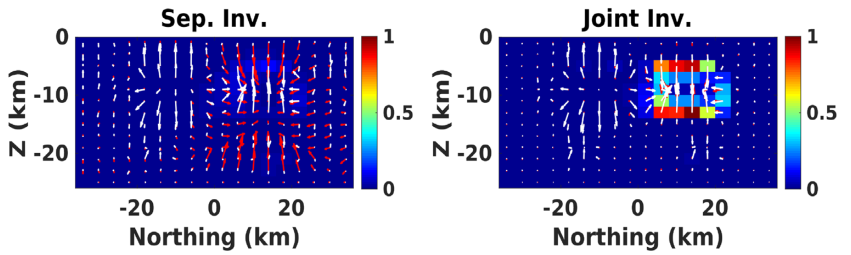

Figure 19. Vertical sections of structural similarity and gradient vectors of model parameters for a vertical profile corresponding to $X = -16$ km for Model 2.

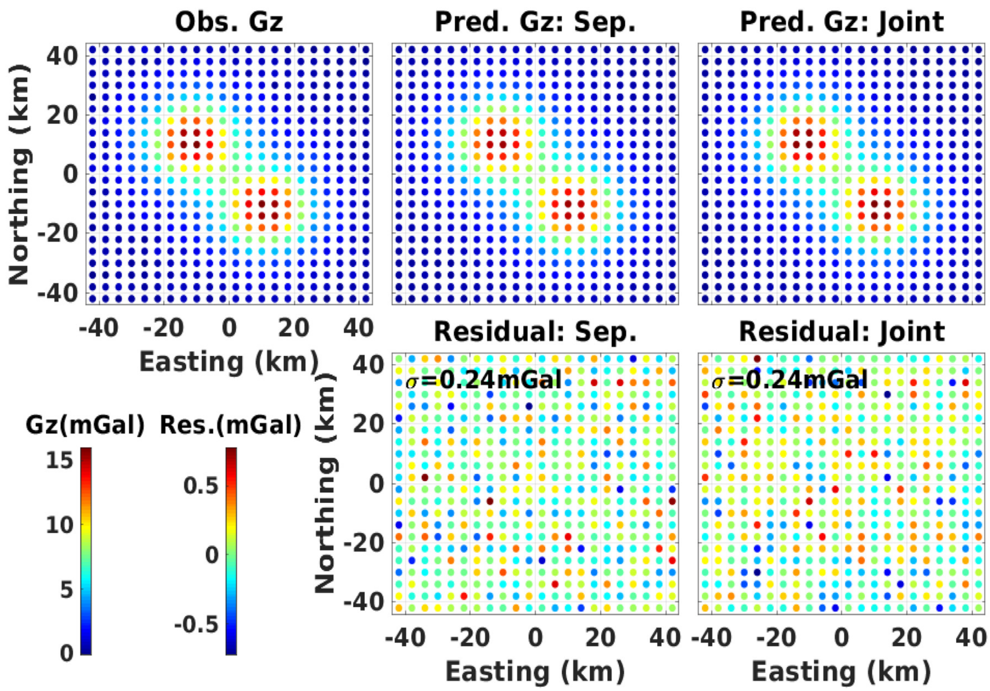

Figure 20. Maps of gravity data fitting for model 2. The observed gravity field, and those predicted by separate and joint inversions are shown in the top panels. The residuals between the observed and predicted data are shown in the bottom panels. The RMS misfit is also printed in the top left corner of each residual panel.

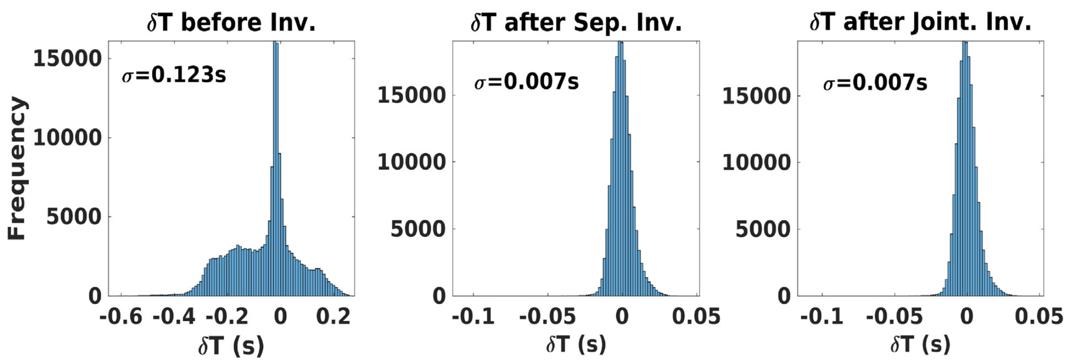

Figure 21. Histograms of the observed traveltime residuals (left panel) and the residuals obtained after the after separate (middle panel) and joint (right panel) inversions for Model 2.

consisting of *P*-wave velocity reduction as significant as 7 per cent at a depth of 5–17 km was believed to have fuelled the volcanism (Husen *et al.* 2004; Farrell *et al.* 2014). By using joint tomographic inversion of local and teleseismic earthquake data, Huang *et al.* (2015) found an upper-crustal rhyolite partial melt magma body and a lower-crustal basaltic partial melt magma reservoir above the Yellowstone mantle plume. The upper-crustal magma reservoir lies between the depths of 5 and 16 km. The shallowest portion of this upper-crustal magma body is responsible for the largest area of hydrothermal activity seen from the earth's surface in Yellowstone.

Due to a lack of adequate seismic stations and earthquakes in the key areas, tomography studies in the Yellowstone area suffered from limited resolution (Farrell *et al.* 2014), leading to increased uncertainty of the velocity models. The model uncertainty, however, could be

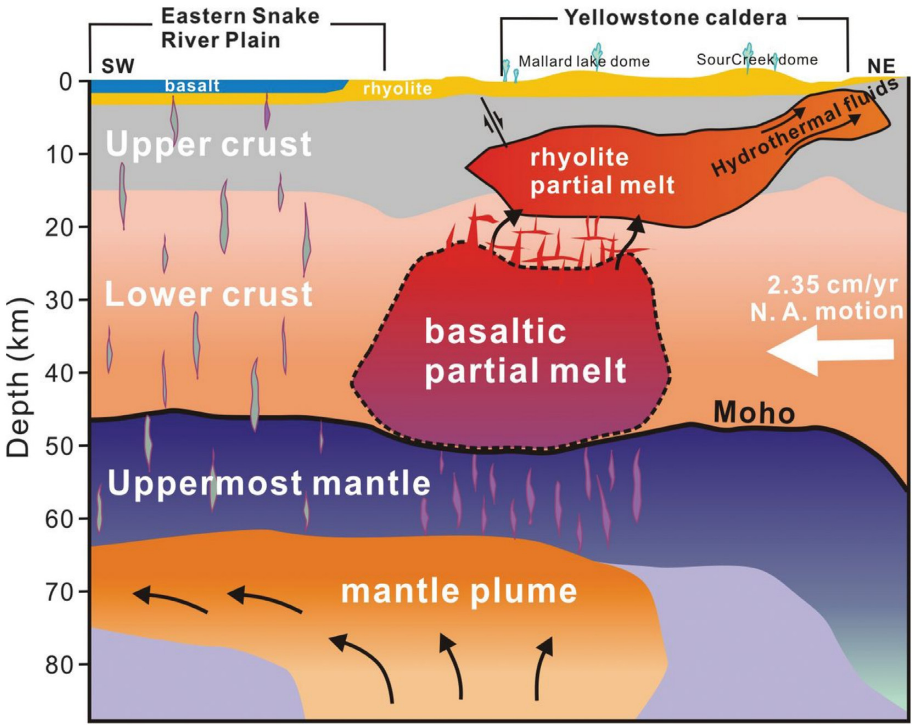

Figure 22. Schematic model for the Yellowstone crust–upper mantle magmatic system (after Huang *et al.* 2015).

reduced by incorporating additional constraints in the inversion. Joint inversion of multiphysics data can be used to apply the constraints which complement each data set with information derived from other data sets (Zhdanov *et al.* 2012; Zhdanov 2015). Different physical fields are sensitive to distinct properties and exhibit particular sensitivity patterns as a result of their specific governing physical laws. We could, therefore, harness the complementary sensitivities of different geophysical data sets to produce geophysical models of the magmatic system with reduced uncertainty.

In order to image the Yellowstone crustal magmatic system, we jointly inverted the gravity data and *P*-wave first arrival traveltimes of local earthquakes recorded over 26 yr from 1984 to 2011 in the Yellowstone area.

4.1 Yellowstone gravity data

The gravity data used in this study were acquired by the United States Geological Survey (USGS), the University of Utah, and the Pan American Center for Earth and Environmental Sciences (PACES). The data were terrain corrected with a USGS elevation model (Hinze *et al.* 2005; DeNosaquo *et al.* 2009). Bouguer correction was also removed with an assumed average crustal density of 2.67 g cm^{-3}. Fig. 23 presents a map of the gravity station distributions and the complete Bouguer anomaly. We observed a strong negative gravity anomaly associated with the Yellowstone Caldera. There are also negative anomalies in the south and northwest valleys, which we ascribe to the less dense sediments. The data have been previously modelled by DeNosaquo *et al.* (2009). They suggested a low density ($\sim$2.52 g cm^{-3}) anomaly beneath the caldera at a depth of 10–20 km. Jorgensen & Zhdanov (2019) jointly inverted the same gravity data with magnetotelluric data and inferred partial melt in both upper and lower crust. However, the density models in those studies are of limited resolution and considerable uncertainty, caused by insufficient gravity station coverage, especially in the east of the survey area, and by non-uniqueness of the gravity inverse problem. Additional constraints could be introduced by jointly inverting the data with seismic traveltime data.

The data were high-pass filtered to remove long-wavelength features associated with deep mantle anomaly as we are to image the crustal magmatic body. The filter passed wavelengths less than 300 km, corresponding to an investigation depth of $\leq$50 km (see Hinze *et al.* 2013). The filtered data were then gridded with a grid spacing of 0.04° × 0.04°. The gridded data should produce roughly equivalent data coverage as the original data in our core study area, that is, the Yellowstone caldera.

4.2 Yellowstone earthquake data

Over 45 000 earthquakes have been recorded by the Yellowstone Seismic Network since 1972 with magnitudes ranging $-1.4 \leq$ MC ≤ 6.1 making it one of the most seismically active areas in the western U.S. Approximately 40 per cent of these earthquakes occur as part of earthquake swarms (Farrell *et al.* 2009). The majority of Yellowstone earthquakes occur in an E–W band of seismicity that extends from the Hebgen Lake, Montana region, west of Yellowstone National Park, to the Norris Geyser Basin area on the northern boundary of the

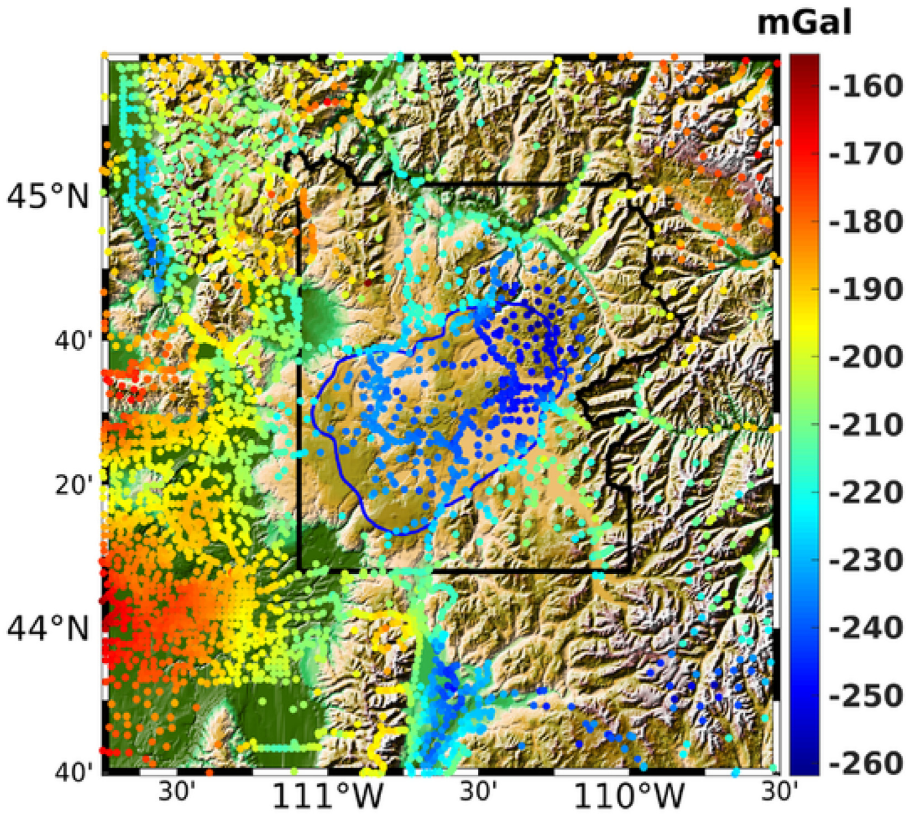

Figure 23. Map showing the distribution of gravity stations in the Yellowstone area. The colour of the stations is scaled by their values of complete Bouguer anomaly. Black solid lines denote the border of Yellowstone National Park. The Late Quaternary Yellowstone caldera was outlined in blue. Data from the online archive of the Pan American Center for Earth and Environmental Sciences (PACES).

Yellowstone caldera (Fig. 24). Since 1995, the Yellowstone area has averaged ∼1600 earthquakes yr^{-1} with magnitudes from $-1.4 \le$ MC ≤ 4.5. The majority of earthquakes in the Yellowstone caldera are less than 5 km deep (Farrell *et al.* 2009; Smith *et al.* 2009). The shallow nature of the maximum focal depths is attributed to high temperatures encountered at shallow depths defining the brittle–ductile transition at $\sim 400\,^{\circ}$C associated with the caldera magma reservoir (Smith *et al.* 2009). Maximum depths of hypocentres deepen from 5 km in the caldera to $\ge$15 km south and north of the caldera.

We employ a record of 26 yr earthquakes in the key areas of Yellowstone from 1984 to 2011 to derive a data set of consistently picked *P* phase first arrival times. To ensure that only the highest-quality data were used for the inversion, we considered earthquakes that had at least eight *P*-wave observations, an azimuthal gap of less than 180° and picks with arrival time uncertainties of less than 0.12 s. The final *P* phase data set consisted of 48 622 high-quality first arrivals from 4520 earthquakes. Fig. 24 is a map of the Yellowstone volcanic field showing the topography. The red dots in Fig. 24 denote the earthquakes selected in this study.

With such a large data set, we would have a good ray coverage at Yellowstone caldera. As shown in Fig. 25, the caldera is well covered by 2-D straight line rays connecting the epicentres and seismic stations. The focal depths of the selected earthquakes are above 20 km depth and the most of them are above 15 km depth. We can thus only image the upper crustal structure above 20 km with this data set.

Farrell *et al.* (2014) inverted the data previously and imaged a 90 km long and 5–17 km deep crustal magma reservoir extending ∼15 km northeast of the caldera. Huang *et al.* (2015) jointly inverted the same data set with teleseismic *P*-wave traveltimes and revealed a lower-crustal magma body feeding the upper-crustal magma reservoir.

4.3 Inversion set-up

The density and velocity models are represented in the same cubic B-splines subspace, and are resampled with meshes of different scales for forward modelling and structural coupling in the joint inversion. Table 3 presents the set-up of knots and meshes for Yellowstone data. As a result of boundary padding, the numbers of knots or grid cells are not exactly inversely proportional to their intervals. The seismic traveltimes are modelled with a very fine mesh of size $\sim 1\,\mathrm{km} \times 1\,\mathrm{km} \times 1\,\mathrm{km}$ to sample the fine velocity structures. The cell size would guarantee that the traveltime uncertainty of discretization origin for forward modelling is below 0.12 s, which is the estimated data uncertainty. Gravity is calculated with a coarse mesh with a cell size of $\sim 4\,\mathrm{km} \times 4\,\mathrm{km} \times 2\,\mathrm{km}$ to make sure there is at least one corresponding observed datum on the surface above each cell. The decoupled forward modelling meshes would, therefore, honour the resolution capabilities of both gravity and seismic data. In the structural coupling stage of the joint inversion, the density and velocity models are again resampled with a coarse

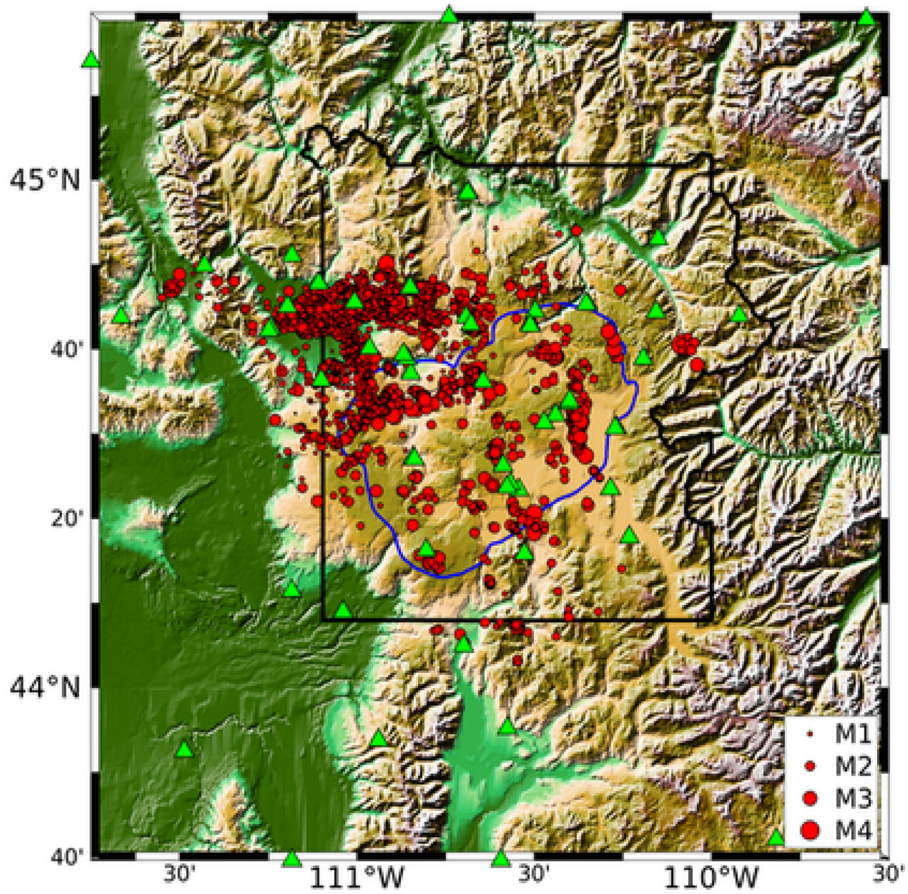

Figure 24. Map of Yellowstone National Park. Seismic stations and earthquakes are denoted by green triangles and red dots, respectively. The area of the red dots represents the magnitude of the earthquakes. Black solid lines denote the border of Yellowstone National Park. The Late Quaternary Yellowstone caldera was outlined in blue.

mesh with a cell size of $\sim 4\,\text{km} \times 4\,\text{km} \times 2\,\text{km}$. Consequently, only long-wavelength (i.e. $\geq$4 km) structures of density and velocity models are enforced to resemble each other. Small (i.e. <4 km) anomalies beyond the resolution capability of gravity data would not be artificially introduced to the gravity model but could still be preserved in the velocity model.

We should note also that both gravity and seismic data are given on irregular grids. Unevenly distribution of surface measurements is a very important practical issue in geophysical applications as well. The presence of irregularly sampled surface measurements or regions of sparse and coarse line spacing may distort the recovered model and leave some area under resolved, since the sensitivity is stronger in area with dense measurements and weaker in area with sparse measurements. Our approach to deal with this issue is to introduce a sensitivity-based model weighting matrix, Wm (eq. 8), to equalize the sensitivity of the data observed on irregular grid to the cells located at different positions. This approach was discussed in details in Zhdanov (2002, 2015).

The initial and *a priori* velocity models for tomographic inversion are the same 1-D layered model as in Farrell *et al.* (2014). It is also used as the reference model to calculate the velocity perturbation. Since we work with the complete Bouguer anomaly, a zero homogeneous background is used for the initial and prior density models. The separate and joint inversions are terminated at the same threshold of the misfits. The thresholds of relative misfits for gravity and seismic traveltime data are 7 and 40 per cent, respectively, which correspond to weighted data root-mean-square misfits of 1.95 mGal and 0.12 s.

4.4 Sensitivity analysis

Sensitivity analysis is a necessary tool to quantify the reliability of the inverted models: model parameters could be of high uncertainties if large variations of them generate a small response in the data only; their uncertainties might be low if the data are very sensitive to their variations. However, any geophysical survey may have limited sensitivity to some sections of the examined subsurface area (Zhdanov 2015), due to factors such as data coverage, survey configuration, and physical laws governing the geophysical fields. Joint inversion of multiphysics data reduces the model uncertainty by complementing the sensitivity of one physical property with others. The usage of complementary sensitivities should enhance the qualities of the inverted models. We calculate the integrated sensitivities of the Yellowstone gravity and

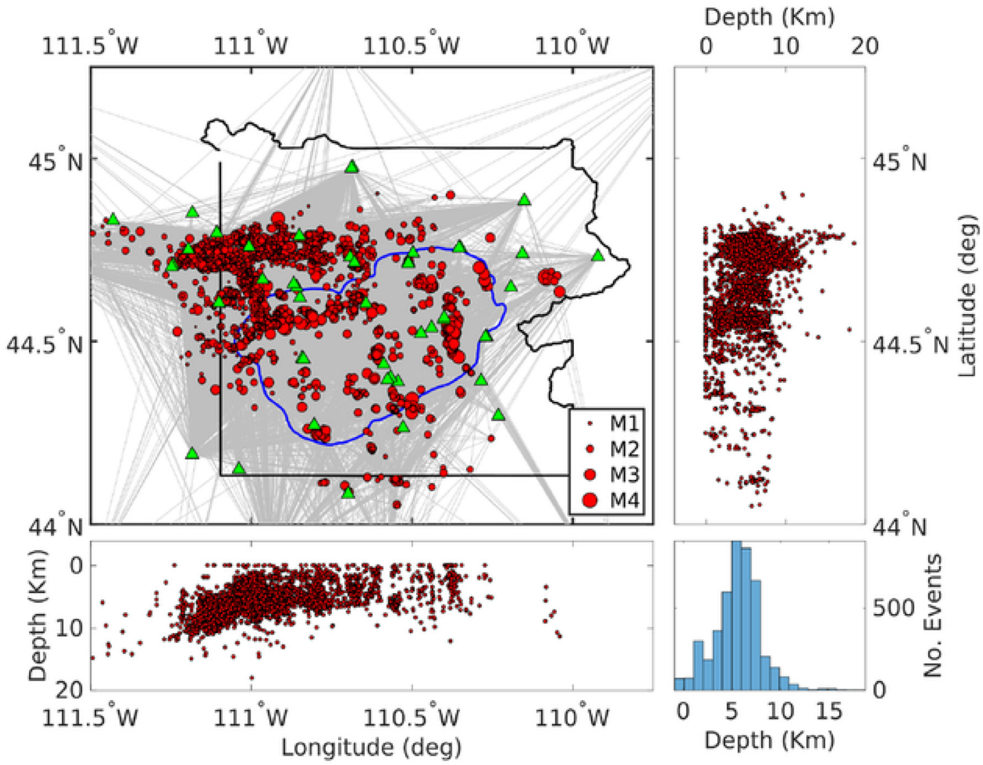

Figure 25. Map of 4520 Yellowstone earthquakes, from 1973 to 2013 for inversion of this study. Epicentres are shown as red dots, with area scaled by earthquake magnitude in the top left panel. Green triangles mark seismic stations. Black and blue lines outline the border of Yellowstone National Park and the 0.64 Ma caldera boundary respectively. 2-D straight ray coverage is shown in grey lines. Bottom left shows longitudinal depth of earthquake events. Top right shows latitudinal depth of earthquake events. Bottom right is histogram of focal depth distribution of 4520 earthquake hypocentre locations. The Yellowstone caldera is well covered by the ray paths.

Table 3. Set-up of knots and meshes for inversion and modelling of Yellowstone data.

Knots/meshes	Intervals/Cell size Long. × Lat. × Depth	Num. of knots/Cells Long. × Lat. × Depth
Knots subspace representation	0.1° × 0.1° × 3 km	27 × 23 × 12
Density mesh forward modelling	0.04° × 0.04° × 2 km	61 × 51 × 14
Velocity mesh forward modelling	0.01° × 0.01° × 1 km	233 × 201 × 28
Density and Velocity mesh structural coupling	0.04° × 0.04° × 2 km	61 × 51 × 14

seismic traveltimes in the subspace with the following formulae (Zhdanov 2015):

$$\mathbf{S}^{(j)} = \frac{\|\delta \mathbf{d}^{(j)}\|}{\delta \mathbf{m}_s^{(j)}} = \sqrt{\mathbf{L}_j^T \mathbf{F}^{(j)^T} \mathbf{F}^{(j)} \mathbf{L}_j}, : j = 1, 2; \tag{9}$$

where T denotes the matrix transpose operator; $\mathbf{F}^{(j)}$ represent the Fréchet matrix of density and seismic P-wave velocity for $j = 1$ and $j = 2$, respectively. The integrated sensitivity provides a measurement of data sensitivity for every subsurface position.

Figs 26 and 27 present the horizontal and vertical sections of the integrated sensitivity. The seismic traveltime sensitivity is calculated based on a 1-D background velocity model as in Farrell *et al.* (2014). The red dots show the locations of the knots of B-splines subspace representation. The gravity sensitivity presents a horizontally uniform pattern, with subtle fluctuations caused by varying elevations of gravity station. The distribution of seismic sensitivity is, comparatively, very sparse. The seismic data have only sufficient sensitivity in the Yellowstone Caldera and nearby areas, leaving vast areas outside the Yellowstone National Park under resolved. There are, therefore, many velocity models that could fit the seismic data. This non-uniqueness of seismic inverse problem could be reduced by additional constraints from gravity data through joint inversion of seismic and gravity data. Besides, the gravity sensitivity decreases rapidly with depth, making it hard to recover the deep density anomaly. The seismic data complement the limited sensitivity of the gravity field at depth, constraining the depth of the density anomaly.

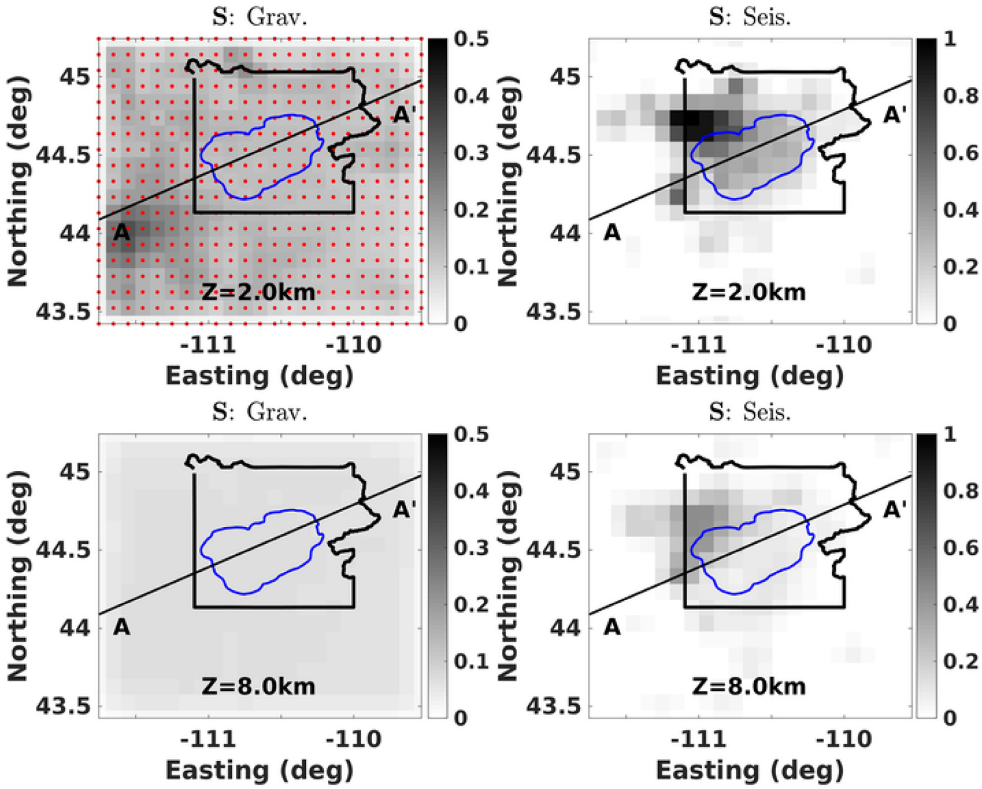

Figure 26. Horizontal sections of normalized integrated sensitivity at depths of 2 and 8 km below sea level. The left-hand column represents sensitivity of the gravity field, whereas the right-hand column denotes that of seismic traveltimes. The red dots show horizontal location of the knots for subspace representation.

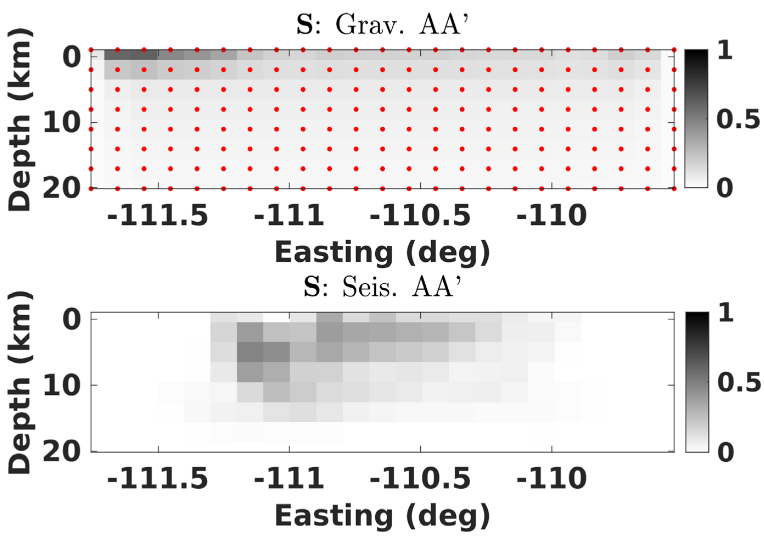

Figure 27. Vertical sections of normalized integrated sensitivity across section AA'. The top panel represents sensitivity of the gravity field; the bottom one denotes that of seismic traveltimes. The red dots show vertical location of the knots for subspace representation.

4.5 Results and discussion

We first applied separate gravity and seismic tomographic inversions to the data and then jointly inverted them. Horizontal sections of the inverted models at four depths (0.7, 4.7, 8.7 and 13.7 km below sea level) are presented in Figs 28–31. The patterns in separately and jointly inverted models are consistent for both density and velocity, as further confirmed by a vertical profile AA' in Fig. 32. However, the depth of the density anomaly is better constrained in the jointly inverted model. The amplitudes of velocity perturbation of jointly inverted model are bounded more tightly compared with those of separately inverted one, leading to a much smoother velocity model, and consequently less

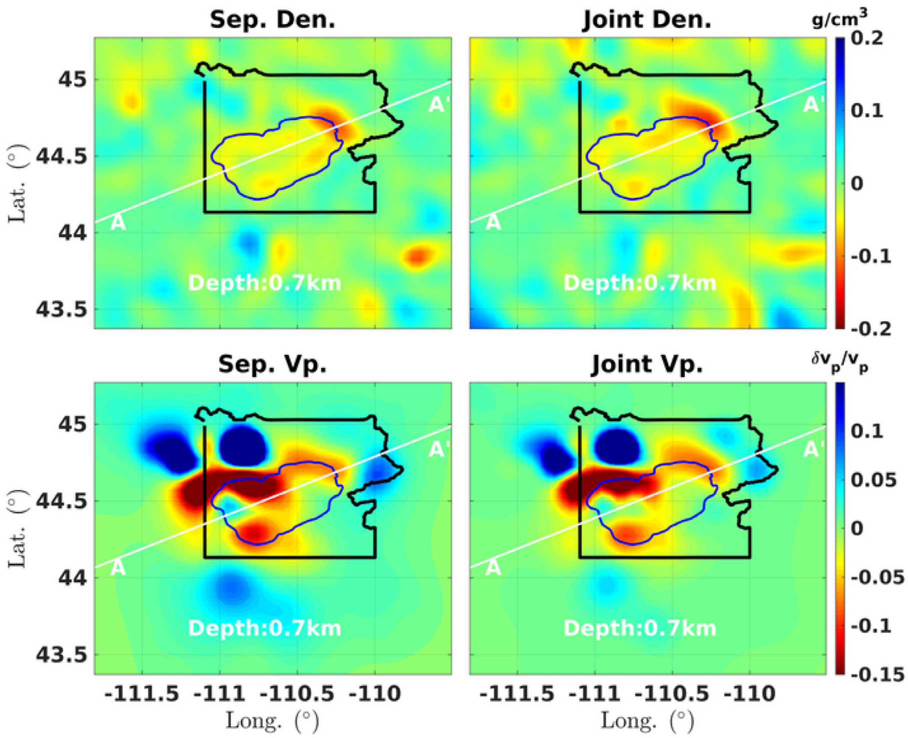

Figure 28. Horizontal sections of separately and jointly inverted density and velocity models at the depth of 0.7 km below sea level. The density (abbreviated to den.) models are on the top row, and *P*-wave velocity (abbreviated to Vp.) models are on the bottom row. The left-hand column present the results of separately (abbreviated to Sep.) inversion, whereas the models of joint inversion are on the right.

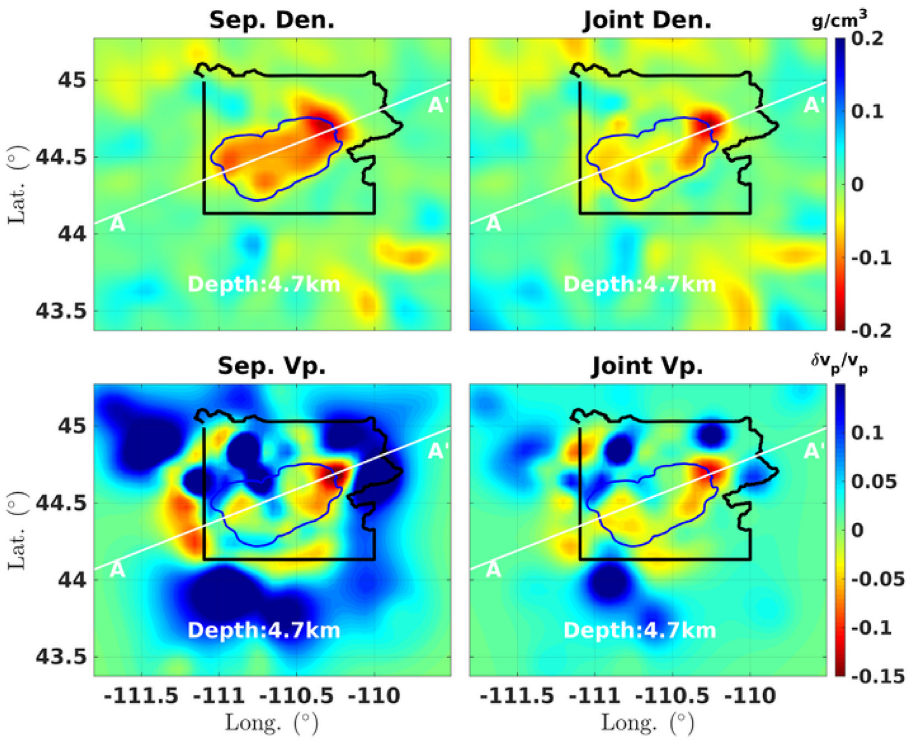

Figure 29. Horizontal sections of separately and jointly inverted density and velocity models at the depth of 4.7 km below sea level.

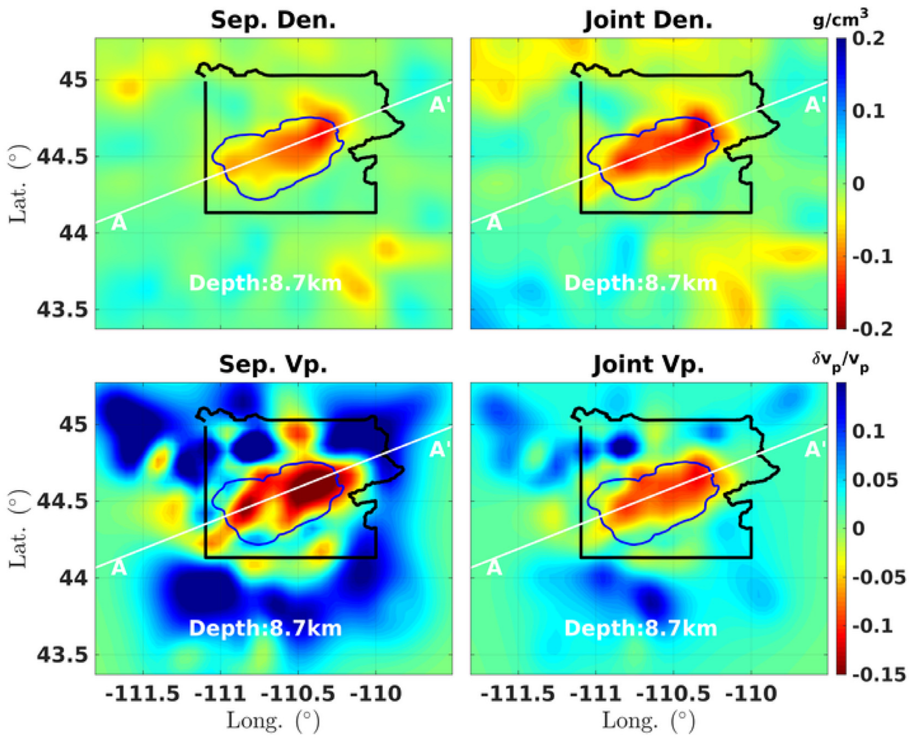

Figure 30. Horizontal sections of separately and jointly inverted density and velocity models at the depth of 8.7 km below sea level.

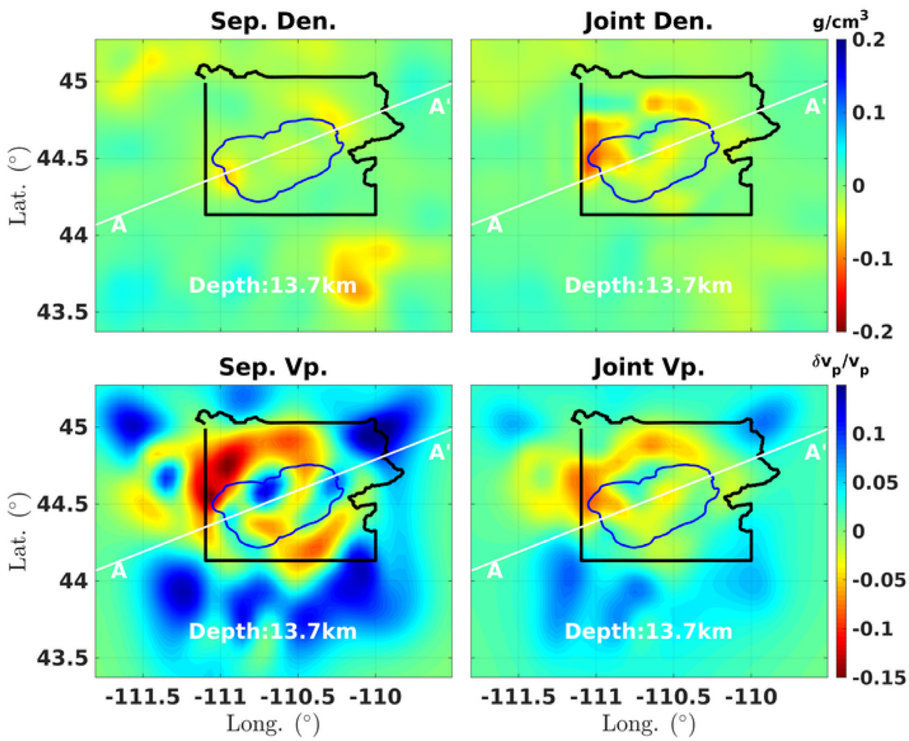

Figure 31. Horizontal sections of separately and jointly inverted density and velocity models at the depth of 13.7 km below sea level.

artefacts. The density and velocity models obtained with joint inversion present a consistent anomaly in the caldera area (see also 3D view of the models in Figs S1 and S2 in the supplementary materials).

Figs 33 and 34 present comparisons of the jointly inverted density and velocity models with published results (Huang *et al.* 2015; Jorgensen & Zhdanov 2019). The pattern of the shallow anomaly is consistent for all models. Jorgensen & Zhdanov's density model, however, resolves a deeper anomaly below 20 km since their model was constrained by a deep conductivity model from magnetotelluric data. Huang's velocity model also presents a deep velocity low anomaly constrained by teleseismic data.

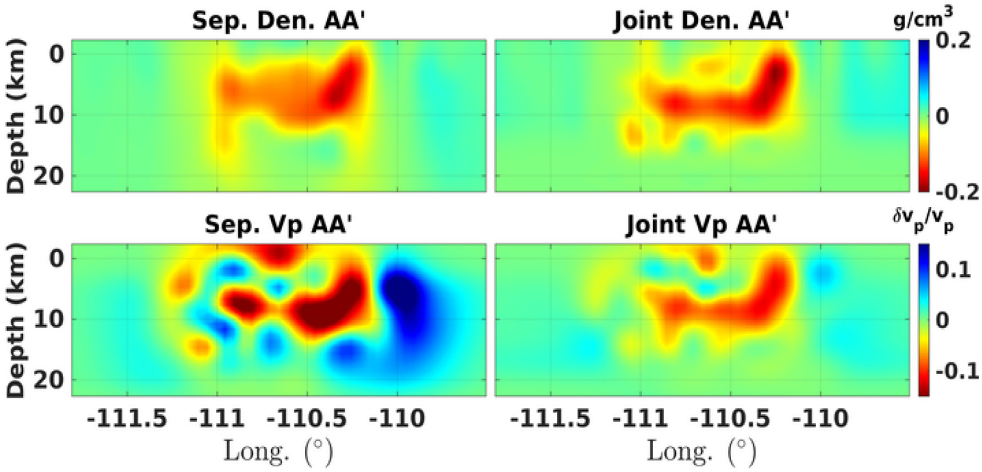

Figure 32. Vertical profile AA' of the separately and jointly inverted density and velocity models. The corresponding location of the profile is marked by white solid line on the horizontal sections.

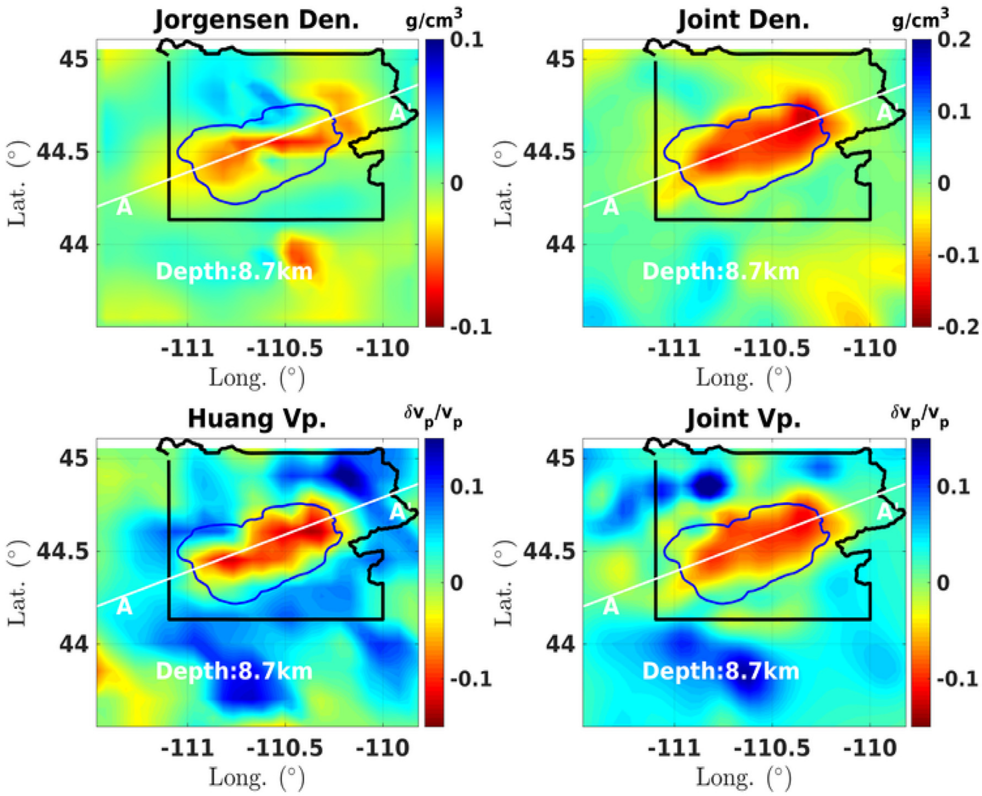

Figure 33. Horizontal sections of the jointly inverted and published density and velocity models at the depth of 8.7 km below sea level.

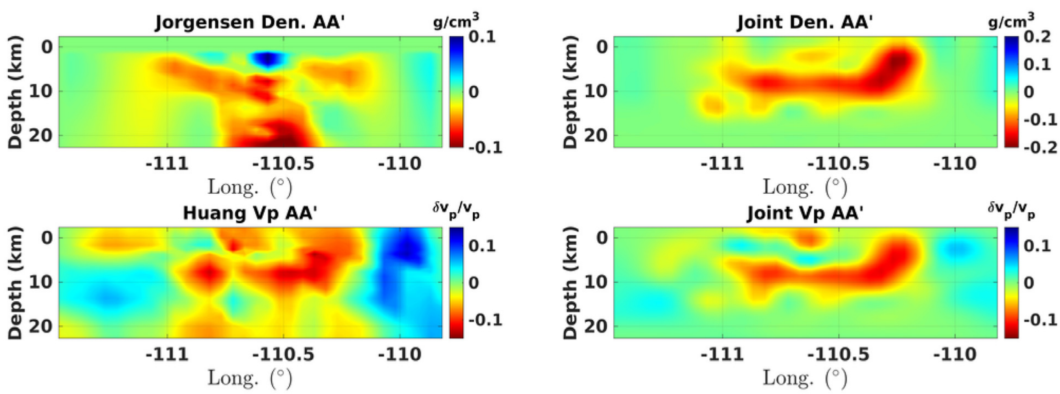

Figure 34. Vertical profile AA' of the jointly inverted and published density and velocity models. The corresponding location of the profile is marked by white solid line on the horizontal sections.

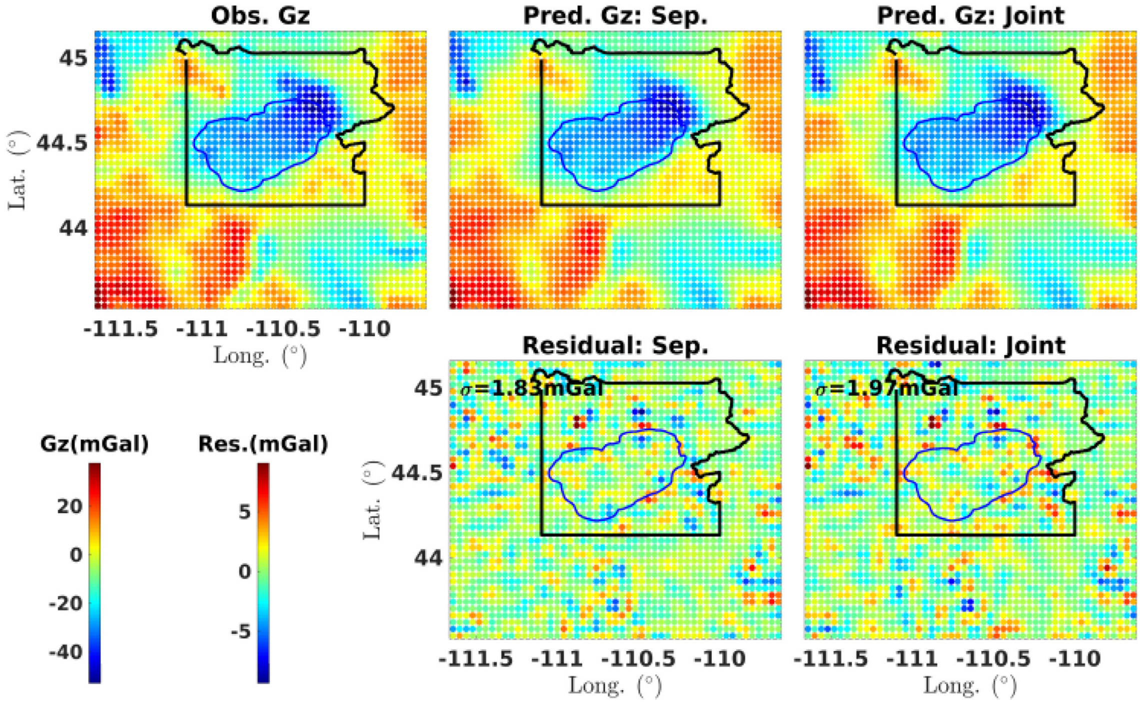

Figure 35. Observed (left) and predicted gravity data with separate and joint inversions. The RMS misfit is also shown in the top left corner of each residual panel. Both separate and joint inversions are converged to a relative data misfit of 7 per cent.

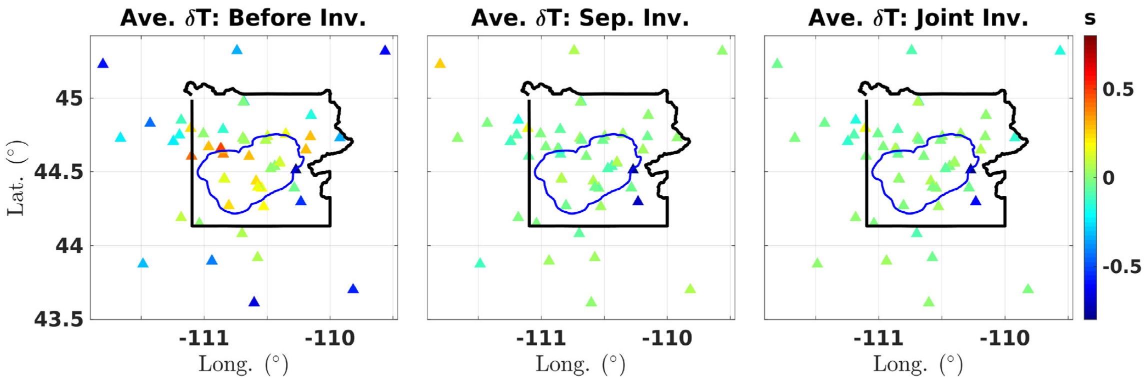

Figure 36. Averaged traveltime residuals at each seismic station before (left), and after separate (middle) and joint (right) inversions. Both separate and joint inversions are converged to a relative data misfit of 40 per cent, that is, 60 per cent residual reduction. The residuals of the YPLB and YTC stations did not change much before and after the inversions.

The data fitting for gravity field is presented in Fig. 35. The separately and jointly inverted density models predict the gravity data equally well, with a relative misfit of 7 per cent. Note that, the observed gravity data uncertainty is larger than that of a typical gravity survey. The gravity data were acquired around 1999, and the seismic data were from 1984 to 2011. According to benchmark gravimeter network data, the gravity had undergone strong temporal variations (0.1–0.3 mGal yr^{-1}) in the Yellowstone area (Farrell 2013). Since the seismic and gravity data were not measured exactly at the same time, we needed to take these temporal variations into consideration. We also calculated the averaged traveltime residuals for each seismic station, as illustrated in Fig. 36. The data again are fitted equally well to the noise level. Two seismic stations, YPLB and YTC, however, have persistent residuals without significant decrease after both separate and joint inversions. We identified one unreasonably large traveltime pick at station YPLB, and two at station YTC after carefully examining the traveltime data, which might be pick errors. It demonstrates that the robust norms for misfit functional use in this inversion appropriately treat the outliers and errors in the observed data.

We should also note that, including gravity data in joint inversion is beneficial for velocity model. As shown in both the model and case studies and discussed in the paper, constrains from gravity data help reducing the artefacts in the velocity model because of the complimentary sensitivities. For the Yellowstone case study, we have limited seismic ray coverage outside of Yellowstone national park and, therefore, have limited sensitivity to the velocity model. A separate seismic tomography tends to put velocity highs (denoted as blue) in the region. However, we have a good gravity station coverage outside of the Yellowstone national park and therefore a reasonable gravity sensitivity. By incorporating the gravity data, we could therefore reduce the uncertainty of the seismic tomography and reduce the artefacts.

5 CONCLUSIONS

We have developed a framework for joint inversion of multiphysics data based on the Gramian constraint. Our method promotes structural similarity between different physical parameters by minimizing a Gramian based structural coupling term. These structural constraints could be considered as a generalization of the cross-gradient method. The quadratic nature of the Gramian structural coupling term makes it numerical implementation similar to conventional L2 norm stabilizing functionals.

The developed method could also consider the differences in resolution capabilities of different geophysical methods using a multiscale resampling strategy. The effectiveness of the method was examined by jointly inverting the *P*-wave traveltime and gravity data computer simulated for synthetic models. This numerical study demonstrated that the developed method successfully avoids the over coupling problem.

We have also applied the developed method to geophysical data collected over the Yellowstone area and jointly inverted the gravity data and *P*-wave traveltimes of local earthquakes to image the crustal magmatic system. Our results have revealed an L-shaped low-density and low-velocity anomaly just beneath the Yellowstone caldera. This result is consistent with previous studies in the same area, but apparently provides a better delineation of the anomalous zone associated with the crustal magma reservoir.

ACKNOWLEDGEMENTS

We are thankful to the Consortium for Electromagnetic Modeling and Inversion at the University of Utah, and TechnoImaging for supporting this research. We thank Dr Ed Biegert and two anonymous reviewers for their valuable comments that helped improving the manuscript.

We also thank the University of Utah Seismograph Station and USGS for providing the earthquake seismic and gravity data.

DATA AVAILABILITY

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

REFERENCES

Afnimar, A., Koketsu, K. & Nakagawa, K., 2002. Joint inversion of refraction and gravity data for the three-dimensional topography of a sediment–basement interface, *Geophys. J. Int.,* **151**(1), 243–254.

Aki, K., Christoffersson, A. & Husebye, E.S., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *J. geophys. Res.,* **82**(2), 277–296.

An, M., 2012. A simple method for determining the spatial resolution of a general inverse problem, *Geophys. J. Int.,* **191**(2), 849–864.

Astic, T. & Oldenburg, D.W., 2019. A framework for petrophysically and geologically guided geophysical inversion using a dynamic Gaussian mixture model prior, *Geophys. J. Int.,* **219**(3), 1989–2012.

Astic, T., Heagy, L.J. & Oldenburg, D.W., 2020. Petrophysically and geologically guided multi-physics inversion using a dynamic Gaussian mixture model, *Geophys. J. Int.,* **224**(1), 40–68.

Berkhout, A., 1984. *Seismic Resolution: A Quantitative Analysis of Resolving Power of Acoustical Echo Techniques,* Handbook of Geophysical Exploration, Geophysical Press.

Capriotti, J. & Li, Y., 2014. Gravity and gravity gradient data: understanding their information content through joint inversions, in *SEG Technical Program Expanded Abstracts 2014,* pp. 1329–1333, doi:10.1190/segam2014-1581.1.

Claerbout, J.F. & Muir, F., 1973. Robust modeling with erratic data, *Geophysics,* **38**(5), 826–844.

Čuma, M., Wilson, G.A. & Zhdanov, M.S., 2012. Large-scale 3D inversion of potential field data, *Geophys. Prospect.,* **60**(6), 1186–1199.

De Kool, M., Rawlinson, N. & Sambridge, M., 2006. A practical grid-based method for tracking multiple refraction and reflection phases in three-dimensional heterogeneous media, *Geophys. J. Int.,* **167**(1), 253–270.

DeNosaquo, K.R., Smith, R.B. & Lowry, A.R., 2009. Density and lithospheric strength models of the Yellowstone–Snake River Plain volcanic system from gravity and heat flow data, *J. Volc. Geotherm. Res.,* **188**(1), 108–127.

Dziewonski, A.M., Hager, B.H. & O'Connell, R.J., 1977. Large-scale heterogeneities in the lower mantle, *J. geophys. Res.,* **82**(2), 239–255.

Eberhart-Phillips, D., 1986. Three-dimensional velocity structure in northern California Coast Ranges from inversion of local earthquake arrival times, *Bull. seism. Soc. Am.,* **76**(4), 1025–1052.

Farrell, J., Husen, S. & Smith, R.B., 2009. Earthquake swarm and *b*-value characterization of the Yellowstone volcano-tectonic system, *J. Volc. Geotherm. Res.,* **188**(1), 260–276.

Farrell, J., Smith, R.B., Taira, T., Chang, W.-L. & Puskas, C.M., 2010. Dynamics and rapid migration of the energetic 2008–2009 Yellowstone Lake earthquake swarm, *Geophys. Res. Lett.,* **37**(19), doi:10.1029/2010GL044605.

Farrell, J., Smith, R.B., Husen, S. & Diehl, T., 2014. Tomography from 26 years of seismicity revealing that the spatial extent of the Yellowstone crustal magma reservoir extends well beyond the Yellowstone caldera, *Geophys. Res. Lett.,* **41**(9), 3068–3073.

Farrell, J.M., 2013. Seismicity and tomographic imaging of the Yellowstone crustal magmatic-tectonic system, *PhD thesis*, The University of Utah.

Gallardo, L.A. & Meju, M.A., 2004. Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints, *J. geophys. Res.,* **109**(B3), doi:10.1029/2003JB002716.

Gallardo, L.A. & Meju, M.A., 2007. Joint two-dimensional cross-gradient imaging of magnetotelluric and seismic traveltime data for structural and lithological classification, *Geophys. J. Int.,* **169**(3), 1261–1272.

Gao, G., Abubakar, A. & Habashy, T.M., 2012. Joint petrophysical inversion of electromagnetic and full-waveform seismic data, *Geophysics,* **77**(3), WA3–WA18.

Giraud, J., Pakyuz-Charrier, E., Jessell, M., Lindsay, M., Martin, R. & Ogarko, V., 2017. Uncertainty reduction through geologically conditioned petrophysical constraints in joint inversion, *Geophysics,* **82**(6), ID19–ID34.

Giraud, J., Ogarko, V., Lindsay, M., Pakyuz-Charrier, E., Jessell, M. & Martin, R., 2019. Sensitivity of constrained joint inversions to geological and petrophysical input data uncertainties with posterior geological analysis, *Geophys. J. Int.,* **218**(1), 666–688.

Guitton, A. & Symes, W.W., 2003. Robust inversion of seismic data using the Huber norm, *Geophysics,* **68**(4), 1310–1319.

Heincke, B., Jegen, M., Moorkamp, M., Hobbs, R.W. & Chen, J., 2017. An adaptive coupling strategy for joint inversions that use petrophysical information as constraints, *J. Appl. Geophys.,* **136,** 279–297.

Hinze, W.J., *et al.*, 2005. New standards for reducing gravity data: the north American gravity database, *Geophysics,* **70**(4), J25–J32.

Hinze, W.J., von Frese, R., Von Frese, R. & Saad, A., 2013. *Gravity and Magnetic Exploration: Principles, Practices, and Applications,* Cambridge Univ. Press.

Huang, H.-H., Lin, F.-C., Schmandt, B., Farrell, J., Smith, R.B. & Tsai, V.C., 2015. The Yellowstone magmatic system from the mantle plume to the upper crust, *Science,* **348**(6236), 773–776.

Huang, Y. & Schuster, G.T., 2014. Resolution limits for wave equation imaging, *J. Appl. Geophys.,* **107,** 137–148.

Hurwitz, S. & Lowenstern, J.B., 2014. Dynamics of the Yellowstone hydrothermal system, *Rev. Geophys.,* **52**(3), 375–411.

Husen, S., Smith, R. & Waite, G., 2004. Evidence for gas and magmatic sources beneath the Yellowstone volcanic field from seismic tomography imaging, *J. Volc. Geotherm. Res.,* **131,** 397–410.

Jorgensen, M. & Zhdanov, M.S., 2019. Imaging Yellowstone magmatic system by the joint Gramian inversion of gravity and magnetotelluric data, *Phys. Earth planet. Inter.,* **292,** 12–20.

Kearey, P., Brooks, M. & Hill, I., 2013. *An Introduction to Geophysical Exploration,* Wiley.

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

Lin, W. & Zhdanov, M.S., 2019. The Gramian method of joint inversion of the gravity gradiometry and seismic data, *Pure appl. Geophys.,* **176**(4), 1659–1672.

Liu, Q. & Tromp, J., 2008. Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.,* **174**(1), 265–286.

Moorkamp, M., Lelièvre, P., Linde, N. & Khan, A., 2016. *Integrated Imaging of the Earth: Theory and Applications,* Geophysical Monograph Series, Wiley.

Neele, F., VanDecar, J. & Snieder, R., 1993. The use of *P* wave amplitude data in a joint inversion with travel times for upper mantle velocity structure, *J. geophys. Res.,* **98**(B7), 12 033–12054.

Nielsen, L. & Jacobsen, B.H., 2000. Integrated gravity and wide-angle seismic inversion for two-dimensional crustal modelling, *Geophys. J. Int.,* **140**(1), 222–232.

Porritt, R.W., Allen, R.M. & Pollitz, F.F., 2014. Seismic imaging east of the Rocky Mountains with USArray, *Earth planet. Sci. Lett.,* **402,** 16–25.

Rawlinson, N., de Kool, M. & Sambridge, M., 2006. Seismic wavefront tracking in 3D heterogeneous media: applications with multiple data classes, *Explor. Geophys.,* **37**(4), 322–330.

Rawlinson, N., Hauser, J. & Sambridge, M., 2008, Seismic ray tracing and wavefront tracking in laterally heterogeneous media, in *Advances in Geophysics,* Vol. 49, pp. 203–273, ed. Dmowska, R., Elsevier.

Roux, E., Moorkamp, M., Jones, A.G., Bischoff, M., Endrun, B., Lebedev, S. & Meier, T., 2011. Joint inversion of long-period magnetotelluric data and surface-wave dispersion curves for anisotropic structure: application to data from central Germany, *Geophys. Res. Lett.,* **38**(5), doi:10.1029/2010GL046358.

Schmandt, B., Dueker, K., Humphreys, E. & Hansen, S., 2012. Hot mantle upwelling across the 660 beneath Yellowstone, *Earth planet. Sci. Lett.,* **331–332,** 224–236.

Schuster, G., 2017. *Seismic Inversion,* Investigations in Geophysics, Society of Exploration Geophysicists.

Schuster, G.T., 1996. Resolution limits for crosswell migration and traveltime tomography, *Geophys. J. Int.,* **127**(2), 427–440.

Smith, R., *et al.*, 2009. Geodynamics of the Yellowstone hotspot and mantle plume: seismic and GPS imaging, kinematics, and mantle flow, *J. Volc. Geotherm. Res.,* **188,** 26–56.

Spetzler, J. & Snieder, R., 2004. The Fresnel volume and transmitted waves, *Geophysics,* **69**(3), 653–663.

Sun, J. & Li, Y., 2015. Multidomain petrophysically constrained inversion and geology differentiation using guided fuzzy *c*-means clustering, *Geophysics,* **80**(4), ID1–ID18.

Sun, J. & Li, Y., 2016a. Joint inversion of multiple geophysical data using guided fuzzy *c*-means clustering, *Geophysics,* **81**(3), ID37–ID57.

Sun, J. & Li, Y., 2016b. Joint inversion of multiple geophysical and petrophysical data using generalized fuzzy clustering algorithms, *Geophys. J. Int.,* **208**(2), 1201–1216.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* Society for Industrial and Applied Mathematics.

Telford, W., Telford, W., Geldart, L., Sheriff, R. & Sheriff, R., 1990. *Applied Geophysics,* Cambridge Univ. Press.

Tu, X. & Zhdanov, M.S., 2020a. Enhancement and sharpening the migration images of the gravity field and its gradients, *Pure appl. Geophys.,* **177,** 2853–2870.

Tu, X. & Zhdanov, M.S., 2020b. Robust synthetic aperture imaging of marine controlled-source electromagnetic data, *IEEE Trans. Geosci. Remote Sens.,* **58,** 5527–5539.

VanDecar, J.C., James, D.E. & Assumpção, M., 1995. Seismic evidence for a fossil mantle plume beneath South America and implications for plate driving forces, *Nature,* **378**(6552), 25–31.

Vasco, D.W., 1989. Resolution and variance operators of gravity and gravity gradiometry, *Geophysics,* **54**(7), 889–899.

Yao, Z.S., Roberts, R.G. & Tryggvason, A., 1999. Calculating resolution and covariance matrices for seismic tomography with the LSQR method, *Geophys. J. Int.,* **138**(3), 886–894.

Yoshizawa, K. & Kennett, B. L.N., 2005. Sensitivity kernels for finite-frequency surface waves, *Geophys. J. Int.,* **162**(3), 910–926.

Yuan, H. & Dueker, K., 2005. Teleseismic *P*-wave tomogram of the Yellowstone plume, *Geophys. Res. Lett.,* **32**(7), doi:10.1029/2004GL022056.

Zhdanov, M., 2002. *Geophysical Inverse Theory and Regularization Problems,* Methods in Geochemistry and Geophysics, Elsevier Science.

Zhdanov, M.S., 2015. *Inverse Theory and Applications in Geophysics,* 2nd edn, Elsevier.

Zhdanov, M.S. & Tolstaya, E., 2006. A novel approach to the model appraisal and resolution analysis of regularized geophysical inversion, *Geophysics,* **71**(6), R79–R90.

Zhdanov, M.S., Gribenko, A. & Wilson, G., 2012. Generalized joint inversion of multimodal geophysical data using Gramian constraints, *Geophys. Res. Lett.,* **39**(9), doi:10.1029/2012GL051233.

Zhu, Y., 2017. *Joint inversion of potential field and electromagnetic data using Gramian constraints, PhD thesis*, The University of Utah.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. A 3-D view of the joint inverted density model for Yellowstone area. The cut-off value of the density contrast is approximately -0.1 g cm^{-3}.

Figure S2. A 3-D view of the joint inverted *P*-wave velocity model for Yellowstone area. The cut-off value of the velocity perturbation is approximately -7.5 per cent.

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

APPENDIX A: NON-NEGATIVITY OF THE GRAMIAN TERM

For simplicity of notation, we denote model parameters in the transformed space as, $\mathbf{m}^{(1)} = \mathbf{T}^{(1)}\mathbf{L}_3\mathbf{m}_s^{(1)}$, and, $\mathbf{m}^{(2)} = \mathbf{T}^{(2)}\mathbf{L}_3\mathbf{m}_s^{(2)}$. The Gramian G in eq. (4) could be computed as follows:

$$G = \|\nabla\mathbf{m}^{(1)}(\mathbf{r})\|^2\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2\left[1-\eta^2\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}),\nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\right], \tag{A1}$$

where $\eta\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}),\nabla\mathbf{m}^{(2)}(\mathbf{r})\right)$ represents the correlation coefficient between the two gradient vectors, $\nabla\mathbf{m}^{(1)}(\mathbf{r})$ and $\nabla\mathbf{m}^{(2)}(\mathbf{r})$, at the current subsurface position $\mathbf{r}$:

$$\eta\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}),\nabla\mathbf{m}^{(2)}(\mathbf{r})\right) = \frac{\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}),\nabla\mathbf{m}^{(2)}(\mathbf{r})\right)}{\|\nabla\mathbf{m}^{(1)}(\mathbf{r})\|\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|}, \tag{A2}$$

and (,) denotes the inner product; $\|\ldots\|$ represents the norm of the vector.

According to the Cauchy–Schwarz inequality, the following inequality holds (Zhdanov, 2002):

$$0 \le \eta^2\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}),\nabla\mathbf{m}^{(2)}(\mathbf{r})\right) \le 1, \tag{A3}$$

with the equality, $\eta^2\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}),\nabla\mathbf{m}^{(2)}(\mathbf{r})\right) = 1$, holding if and only if two gradient vectors, $\nabla\mathbf{m}^{(1)}(\mathbf{r})$ and $\nabla\mathbf{m}^{(2)}(\mathbf{r})$, are linearly dependent. The Gramian is therefore non-negative,

$$G\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}),\nabla\mathbf{m}^{(2)}(\mathbf{r})\right) \ge 0. \tag{A4}$$

The two sides are equal if and only if the two gradient vectors are linearly dependent; or equivalently, we could enforce the gradient vectors of two model parameters to be mutually parallel by minimizing the Gramian G.

APPENDIX B: STRUCTURAL GRAMIAN VERSUS CROSS-GRADIENT

We will prove in this section that the proposed structural Gramian is a generalization of the cross-gradient (Gallardo & Meju 2004, 2007). Assuming the model parameters are structurally coupled in the original model space and that the resolution capabilities of the two geophysical methods are the same, the general form of Gramian in eq. (4) could be simplified by setting the transforming and resampling operators to identity matrices,

$$\mathbf{T}^{(1)} = \mathbf{I};:\mathbf{T}^{(2)} = \mathbf{I};:\mathbf{L}_3 = \mathbf{I}. \tag{B1}$$

The Gramian G could, therefore, be computed as

$$G = \|\nabla\mathbf{m}_s^{(1)}(\mathbf{r})\|^2\|\nabla\mathbf{m}_s^{(2)}(\mathbf{r})\|^2 - \left(\nabla\mathbf{m}_s^{(1)}(\mathbf{r}),\nabla\mathbf{m}_s^{(2)}(\mathbf{r})\right)^2. \tag{B2}$$

Substituting the gradient operators, $\nabla\mathbf{m}_s^{(j)}(\mathbf{r}) = [\nabla_x m_s^{(j)},\nabla_y m_s^{(j)},\nabla_z m_s^{(j)}]^T,\ \ j=1,2$, into the above equation, we obtain the expression for the Gramian G,

$$\begin{aligned}G = {} & \nabla_x^2 m_s^{(1)}\nabla_y^2 m_s^{(2)} + \nabla_x^2 m_s^{(1)}\nabla_z^2 m_s^{(2)} + \nabla_y^2 m_s^{(1)}\nabla_x^2 m_s^{(2)} + \nabla_y^2 m_s^{(1)}\nabla_z^2 m_s^{(2)}\\ & +\nabla_z^2 m_s^{(1)}\nabla_x^2 m_s^{(2)} + \nabla_z^2 m_s^{(1)}\nabla_y^2 m_s^{(2)} - 2\nabla_x m_s^{(1)}\nabla_x m_s^{(2)}\nabla_y m_s^{(1)}\nabla_y m_s^{(2)}\\ & -2\nabla_x m_s^{(1)}\nabla_x m_s^{(2)}\nabla_z m_s^{(1)}\nabla_z m_s^{(2)} - 2\nabla_y m_s^{(1)}\nabla_y m_s^{(2)}\nabla_z m_s^{(1)}\nabla_z m_s^{(2)}.\end{aligned} \tag{B3}$$

The cross-gradient term is given by Gallardo & Meju (2004) as

$$\mathbf{x}_g(\mathbf{r}) = \nabla\mathbf{m}_s^{(1)}(\mathbf{r}) \times \nabla\mathbf{m}_s^{(2)}(\mathbf{r}). \tag{B4}$$

We substitute the three gradient components into the above equation and calculate its L_2 norm,

$$\begin{aligned}\|\mathbf{x}_g(\mathbf{r})\|^2 = {} & \left(\nabla_y m_s^{(1)}\nabla_z m_s^{(2)} - \nabla_z m_s^{(1)}\nabla_y m_s^{(2)}\right)^2 + \left(\nabla_z m_s^{(1)}\nabla_x m_s^{(2)} - \nabla_x m_s^{(1)}\nabla_z m_s^{(2)}\right)^2\\ & +\left(\nabla_x m_s^{(1)}\nabla_y m_s^{(2)} - \nabla_y m_s^{(1)}\nabla_x m_s^{(2)}\right)^2\\ = {} & \nabla_x^2 m_s^{(1)}\nabla_y^2 m_s^{(2)} + \nabla_x^2 m_s^{(1)}\nabla_z^2 m_s^{(2)} + \nabla_y^2 m_s^{(1)}\nabla_x^2 m_s^{(2)} + \nabla_y^2 m_s^{(1)}\nabla_z^2 m_s^{(2)}\\ & +\nabla_z^2 m_s^{(1)}\nabla_x^2 m_s^{(2)} + \nabla_z^2 m_s^{(1)}\nabla_y^2 m_s^{(2)} - 2\nabla_x m_s^{(1)}\nabla_x m_s^{(2)}\nabla_y m_s^{(1)}\nabla_y m_s^{(2)}\\ & -2\nabla_x m_s^{(1)}\nabla_x m_s^{(2)}\nabla_z m_s^{(1)}\nabla_z m_s^{(2)} - 2\nabla_y m_s^{(1)}\nabla_y m_s^{(2)}\nabla_z m_s^{(1)}\nabla_z m_s^{(2)}.\end{aligned} \tag{B5}$$

Therefore, $G = \|\mathbf{x}_g(\mathbf{r})\|^2$, that is, the structural Gramian is equivalent to the L_2 norm of the cross-gradient. The equivalence between them could be derived more intuitively, since both sides equal the square of the area of the parallelogram with the two gradient vectors for sides. Joint inversions based on them should produce similar models if they are appropriately incorporated in the inversion. The non-negativeness and quadratic form for the Gramian, however, make it more convenient to optimize. Moreover, it is more flexible to incorporate model space transformation and resampling in the structural Gramian.

APPENDIX C: RRCG METHOD FOR JOINT INVERSION

We follow Zhu's (2017) implementation of the RRCG method for joint inversion (Zhdanov 2015). We concatenate the vectors, matrices and operators associated with different geophysical methods in a general matrix notation, as follows:

$$\begin{aligned}
&\mathbf{m} = \begin{bmatrix} \mathbf{m}_s^{(1)} \\ \mathbf{m}_s^{(2)} \end{bmatrix};\ \mathbf{d}_o = \begin{bmatrix} \mathbf{d}_o^{(1)} \\ \mathbf{d}_o^{(2)} \end{bmatrix};\ \mathbf{A}(\mathbf{m}) = \begin{bmatrix} \mathbf{A}^{(1)}(\mathbf{L}_1\mathbf{m}_s^{(1)}) \\ \mathbf{A}^{(2)}(\mathbf{L}_2\mathbf{m}_s^{(2)}) \end{bmatrix};\\
&\mathbf{W}_d = \begin{bmatrix} \mathbf{W}_d^{(1)} & \vdots \\ \vdots & \mathbf{W}_d^{(2)} \end{bmatrix};\ \mathbf{W}_m = \begin{bmatrix} \mathbf{W}_m^{(1)} & \vdots \\ \vdots & \mathbf{W}_m^{(2)} \end{bmatrix};\ \alpha = \begin{bmatrix} \alpha^{(1)}\mathbf{I}^{(1)} & \vdots \\ \vdots & \alpha^{(2)}\mathbf{I}^{(2)} \end{bmatrix}.
\end{aligned} \tag{C1}$$

The concatenated Fréchet matrix $\mathbf{F}$ is constructed from the separate ones, $\mathbf{F}^{(1)}$ and $\mathbf{F}^{(2)}$, associated with the forward operators, $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$, respectively,

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}^{(1)}\mathbf{L}_1 & \\ & \mathbf{F}^{(2)}\mathbf{L}_2 \end{bmatrix}. \tag{C2}$$

We denote the concatenated descent gradient directions for the objective functional (6) and the coupling term (3) as $\mathbf{l}$ and $\mathbf{l}_g$, respectively. One iteration of the RRCG method could be represented as

$$\begin{aligned}
&\mathbf{r}_n = \mathbf{A}(\mathbf{m}_n);\\
&\mathbf{l}_n = \mathbf{F}_n^*\mathbf{W}_d^2\mathbf{r}_n + \alpha_n\mathbf{W}_m^2(\mathbf{m}_n - \mathbf{m}_{apr}) + \beta_n\mathbf{l}_{g\,n},\\
&\text{with } \alpha_0^{(j)} = 0,\ \alpha_1^{(j)} = \frac{\left|\left|\mathbf{W}_d^{(j)}\left[\mathbf{A}^{(j)}(\mathbf{L}_j\mathbf{m}_s^{(j)}) - \mathbf{d}_o^{(j)}\right]\right|\right|_\rho^2}{\left|\left|\mathbf{W}_m^{(j)}(\mathbf{m}_s^{(j)} - \mathbf{m}_{apr}^{(j)})\right|\right|^2},\\
&\qquad \beta_0 = 0,\ \beta_1 = \frac{\sum_{j=1}^2\left|\left|\mathbf{W}_d^{(j)}\left[\mathbf{A}^{(j)}(\mathbf{L}_j\mathbf{m}_s^{(j)}) - \mathbf{d}_o^{(j)}\right]\right|\right|_\rho^2}{s(\mathbf{m}_s^{(1)}, \mathbf{m}_s^{(2)})};\\
&\tilde{\mathbf{l}}_n = \mathbf{l}_n + \frac{\|\mathbf{l}_n\|^2}{\|\mathbf{l}_{n-1}\|^2}\tilde{\mathbf{l}}_{n-1},\ \tilde{\mathbf{l}}_0 = \mathbf{l}_0;\\
&k_n = \frac{(\tilde{\mathbf{l}}_n, \mathbf{l}_n)}{\|\mathbf{W}_d\mathbf{F}_n\tilde{\mathbf{l}}_n\|^2 + \alpha_n\|\mathbf{W}_m\tilde{\mathbf{l}}_n\|^2 + \beta_n\|\mathbf{H}_{g_n}\tilde{\mathbf{l}}_n\|^2};\\
&\mathbf{m}_{n+1} = \mathbf{m}_n - k_n\tilde{\mathbf{l}}_n;\\
&\alpha_{n+1} = q\alpha_n,\ 0 < q < 1,\ \text{if } \|\alpha_n\mathbf{W}_m^2(\mathbf{m}_{n+1} - \mathbf{m}_{apr})\| > \|\alpha_n\mathbf{W}_m^2(\mathbf{m}_n - \mathbf{m}_{apr})\|;\\
&\alpha_{n+1} = \alpha_n,\ \text{otherwise};\\
&\beta_{n+1} = q\beta_n,\ 0 < q < 1,\ \text{if } s(\mathbf{m}_{s\ n+1}^{(1)}, \mathbf{m}_{s\ n+1}^{(2)}) > s(\mathbf{m}_{s\ n}^{(1)}, \mathbf{m}_{s\ n}^{(2)});\\
&\beta_{n+1} = \beta_n,\ \text{otherwise}.
\end{aligned} \tag{C3}$$

APPENDIX D: THE FIRST VARIATION OF THE STRUCTURAL GRAMIAN

In this section, we calculate the first variation of the proposed structural Gramian term. We still denote $\mathbf{m}^{(j)} = \mathbf{T}^{(j)}\mathbf{L}_3\mathbf{m}_s^{(j)},\ \ j = 1, 2$ for the simplicity of notation. The first variation of the coupling term $s(\mathbf{m}_s^{(1)}, \mathbf{m}_s^{(2)})$ with respect to $\mathbf{m}^{(1)}$ could be computed as:

$$\begin{aligned}
\delta_{\mathbf{m}^{(1)}}s(\mathbf{m}_s^{(1)}, \mathbf{m}_s^{(2)}) &= \delta_{\mathbf{m}^{(1)}}\iiint_D g\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\mathrm{d}v\\
&= \iiint_D \delta_{\mathbf{m}^{(1)}}g\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\mathrm{d}v.
\end{aligned} \tag{D1}$$

Substituting the Gramian (4) yields the first variation of integral component,

$$
\begin{aligned}
\delta_{\mathbf{m}^{(1)}} G = & \begin{vmatrix} \left(\nabla\left(\mathbf{m}^{(1)}(\mathbf{r})+\delta\mathbf{m}^{(1)}(\mathbf{r})\right), \nabla\left(\mathbf{m}^{(1)}(\mathbf{r})+\delta\mathbf{m}^{(1)}(\mathbf{r})\right)\right) & \left(\nabla\left(\mathbf{m}^{(1)}(\mathbf{r})+\delta\mathbf{m}^{(1)}(\mathbf{r})\right), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right) \\ \left(\nabla\mathbf{m}^{(2)}(\mathbf{r}), \nabla\left(\mathbf{m}^{(1)}(\mathbf{r})+\delta\mathbf{m}^{(1)}(\mathbf{r})\right)\right) & \left(\nabla\mathbf{m}^{(2)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right) \end{vmatrix} \\
& - \begin{vmatrix} \left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(1)}(\mathbf{r})\right) & \left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right) \\ \left(\nabla\mathbf{m}^{(2)}(\mathbf{r}), \nabla\mathbf{m}^{(1)}(\mathbf{r})\right) & \left(\nabla\mathbf{m}^{(2)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right) \end{vmatrix} \\
& = \|\nabla\mathbf{m}^{(1)}(\mathbf{r})\|^2\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2 + \|\nabla\delta\mathbf{m}^{(1)}(\mathbf{r})\|^2\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2 \\
& + 2\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\delta\mathbf{m}^{(1)}(\mathbf{r})\right)\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2 - \left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)^2 \\
& - 2\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\left(\nabla\delta\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right) - \left(\nabla\delta\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)^2 \\
& - \left[\|\nabla\mathbf{m}^{(1)}(\mathbf{r})\|^2\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2 - \left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)^2\right] \\
& = \|\nabla\delta\mathbf{m}^{(1)}(\mathbf{r})\|^2\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2 + 2\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\delta\mathbf{m}^{(1)}(\mathbf{r})\right)\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2 \\
& - 2\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\left(\nabla\delta\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right) - \left(\nabla\delta\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)^2.
\end{aligned} \tag{D2}
$$

We then examine each of the four terms in the above equation separately. The first term could be written as

$$
\begin{aligned}
\|\nabla\delta\mathbf{m}^{(1)}(\mathbf{r})\|^2\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2 &= \left(\nabla\delta\mathbf{m}^{(1)}(\mathbf{r}), \nabla\delta\mathbf{m}^{(1)}(\mathbf{r})\right)\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2 \\
&= \nabla\delta\mathbf{m}^{(1)}(\mathbf{r}) \cdot \left[\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2\nabla\delta\mathbf{m}^{(1)}(\mathbf{r})\right].
\end{aligned} \tag{D3}
$$

According to the product rule of divergence (Zhdanov 2015), the following equation holds,

$$
\nabla \cdot (u\mathbf{V}) = \nabla u \cdot \mathbf{V} + u\nabla \cdot \mathbf{V}. \tag{D4}
$$

Setting $u = \delta\mathbf{m}^{(1)}(\mathbf{r})$, and, $\mathbf{V} = \|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2\nabla\delta\mathbf{m}^{(1)}(\mathbf{r})$, we obtain the following expression,

$$
\begin{aligned}
\nabla\delta\mathbf{m}^{(1)}(\mathbf{r}) \cdot \left[\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2\nabla\delta\mathbf{m}^{(1)}(\mathbf{r})\right] = & \nabla \cdot \left[\delta\mathbf{m}^{(1)}(\mathbf{r})\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2\nabla\delta\mathbf{m}^{(1)}(\mathbf{r})\right] \\
& -\delta\mathbf{m}^{(1)}(\mathbf{r})\left[\nabla \cdot \left(\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2\nabla\delta\mathbf{m}^{(1)}(\mathbf{r})\right)\right].
\end{aligned} \tag{D5}
$$

The other three terms in eq. (D2) could be calculated with a similar rule,

$$
\begin{aligned}
\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\delta\mathbf{m}^{(1)}(\mathbf{r})\right)\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2 = & \nabla \cdot \left[\delta\mathbf{m}^{(1)}(\mathbf{r})\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2\nabla\mathbf{m}^{(1)}(\mathbf{r})\right] \\
& -\delta\mathbf{m}^{(1)}(\mathbf{r})\left[\nabla \cdot \left(\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2\nabla\mathbf{m}^{(1)}(\mathbf{r})\right)\right].
\end{aligned} \tag{D6}
$$

$$
\begin{aligned}
& \left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\left(\nabla\delta\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right) = \nabla \cdot \left[\delta\mathbf{m}^{(1)}(\mathbf{r})\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\nabla\mathbf{m}^{(2)}(\mathbf{r})\right] \\
& -\delta\mathbf{m}^{(1)}(\mathbf{r})\left[\nabla \cdot \left(\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\right].
\end{aligned} \tag{D7}
$$

$$
\begin{aligned}
\left(\nabla\delta\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)^2 = & \nabla \cdot \left[\delta\mathbf{m}^{(1)}(\mathbf{r})\left(\nabla\delta\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\nabla\mathbf{m}^{(2)}(\mathbf{r})\right] \\
& -\delta\mathbf{m}^{(1)}(\mathbf{r})\left[\nabla \cdot \left(\left(\nabla\delta\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\right].
\end{aligned} \tag{D8}
$$

Substituting the four terms into eq. (D2) and omitting the terms of order $O(\|\delta\mathbf{m}^{(1)}(\mathbf{r})\|^2)$, we obtain the following equation,

$$
\begin{aligned}
\delta_{\mathbf{m}^{(1)}} G = & 2\nabla \cdot \left[\delta\mathbf{m}^{(1)}(\mathbf{r})\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2\nabla\mathbf{m}^{(1)}(\mathbf{r})\right] - 2\delta\mathbf{m}^{(1)}(\mathbf{r})\left[\nabla \cdot \left(\|\nabla\mathbf{m}^{(2)}(\mathbf{r})\|^2\nabla\mathbf{m}^{(1)}(\mathbf{r})\right)\right] \\
& -2\nabla \cdot \left[\delta\mathbf{m}^{(1)}(\mathbf{r})\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\nabla\mathbf{m}^{(2)}(\mathbf{r})\right] \\
& +2\delta\mathbf{m}^{(1)}(\mathbf{r})\left[\nabla \cdot \left(\left(\nabla\mathbf{m}^{(1)}(\mathbf{r}), \nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\nabla\mathbf{m}^{(2)}(\mathbf{r})\right)\right] + O(\|\delta\mathbf{m}^{(1)}(\mathbf{r})\|^2),
\end{aligned} \tag{D9}
$$

which is again substituted to eq. (D1),

$$\delta_{\mathbf{m}^{(1)}} s(\mathbf{m}_s^{(1)}, \mathbf{m}_s^{(2)}) = 2 \iiint_D \nabla \cdot \left[\delta \mathbf{m}^{(1)}(\mathbf{r}) \|\nabla \mathbf{m}^{(2)}(\mathbf{r})\|^2 \nabla \mathbf{m}^{(1)}(\mathbf{r})\right] \mathrm{d}v$$
$$-2 \iiint_D \delta \mathbf{m}^{(1)}(\mathbf{r}) \left[\nabla \cdot \left(\|\nabla \mathbf{m}^{(2)}(\mathbf{r})\|^2 \nabla \mathbf{m}^{(1)}(\mathbf{r})\right)\right] \mathrm{d}v$$
$$-2 \iiint_D \nabla \cdot \left[\delta \mathbf{m}^{(1)}(\mathbf{r}) \left(\nabla \mathbf{m}^{(1)}(\mathbf{r}), \nabla \mathbf{m}^{(2)}(\mathbf{r})\right) \nabla \mathbf{m}^{(2)}(\mathbf{r})\right] \mathrm{d}v$$
$$+2 \iiint_D \delta \mathbf{m}^{(1)}(\mathbf{r}) \left[\nabla \cdot \left(\left(\nabla \mathbf{m}^{(1)}(\mathbf{r}), \nabla \mathbf{m}^{(2)}(\mathbf{r})\right) \nabla \mathbf{m}^{(2)}(\mathbf{r})\right)\right] \mathrm{d}v. \quad \text{(D10)}$$

Applying mean value theorem and Gauss theorem to the first term of the above equation,

$$\iiint_D \nabla \cdot \left[\delta \mathbf{m}^{(1)}(\mathbf{r}) \|\nabla \mathbf{m}^{(2)}(\mathbf{r})\|^2 \nabla \mathbf{m}^{(1)}(\mathbf{r})\right] \mathrm{d}v$$
$$= \iiint_D \nabla \cdot \left[\delta \mathbf{m}^{(1)}(\mathbf{r}) \|\nabla \mathbf{m}^{(2)}(\mathbf{r})\|^2 \nabla \mathbf{m}^{(1)}(\mathbf{r})\right] \mathrm{d}v$$
$$= \oint_{\partial D} \delta \mathbf{m}^{(1)}(\mathbf{r}) \|\nabla \mathbf{m}^{(2)}(\mathbf{r})\|^2 \nabla \mathbf{m}^{(1)}(\mathbf{r}) \mathbf{ds}, \quad \text{(D11)}$$

and assuming $\nabla \mathbf{m}^{(1)}(\mathbf{r})$ and $\nabla \mathbf{m}^{(2)}(\mathbf{r})$ are zeros on the surface enclosed the inversion domain (i.e. $\mathbf{r} \in \partial D$), the above term would vanish,

$$\oint_{\partial D} \delta \mathbf{m}^{(1)}(\mathbf{r}) \|\nabla \mathbf{m}^{(2)}(\mathbf{r})\|^2 \nabla \mathbf{m}^{(1)}(\mathbf{r}) \mathbf{ds} = 0. \quad \text{(D12)}$$

Using a similar approach, we could easily prove that the third term in eq. (D10) also vanishes,

$$\iiint_D \nabla \cdot \left[\delta \mathbf{m}^{(1)}(\mathbf{r}) \left(\nabla \mathbf{m}^{(1)}(\mathbf{r}), \nabla \mathbf{m}^{(2)}(\mathbf{r})\right) \nabla \mathbf{m}^{(2)}(\mathbf{r})\right] \mathrm{d}v$$
$$= \oint_{\partial D} \delta \mathbf{m}^{(1)}(\mathbf{r}) \left(\nabla \mathbf{m}^{(1)}(\mathbf{r}), \nabla \mathbf{m}^{(2)}(\mathbf{r})\right) \nabla \mathbf{m}^{(2)}(\mathbf{r}) \mathbf{ds}$$
$$= 0. \quad \text{(D13)}$$

Therefore, the first variation of the proposed structural Gramian constraint could be computed as,

$$\delta_{\mathbf{m}^{(1)}} s(\mathbf{m}_s^{(1)}, \mathbf{m}_s^{(2)}) = \iiint_D \delta \mathbf{m}^{(1)}(\mathbf{r}) 2 \left[\nabla \cdot \left(\left(\nabla \mathbf{m}^{(1)}(\mathbf{r}), \nabla \mathbf{m}^{(2)}(\mathbf{r})\right) \nabla \mathbf{m}^{(2)}(\mathbf{r})\right)\right.$$
$$\left. - \nabla \cdot \left(\|\nabla \mathbf{m}^{(2)}(\mathbf{r})\|^2 \nabla \mathbf{m}^{(1)}(\mathbf{r})\right)\right] \mathrm{d}v. \quad \text{(D14)}$$

Thus, the ascend direction (i.e. gradient) of the structural Gramian constraint with respect to model parameter $\mathbf{m}^{(1)}$ is

$$\mathbf{l}_g^{(1)} = \left[\nabla \cdot \left(\left(\nabla \mathbf{m}^{(1)}, \nabla \mathbf{m}^{(2)}\right) \nabla \mathbf{m}^{(2)}\right) - \nabla \cdot \left(\|\nabla \mathbf{m}^{(2)}\|^2 \nabla \mathbf{m}^{(1)}\right)\right]. \quad \text{(D15)}$$

Similarly, we could get the ascend direction of s with respect to $\mathbf{m}^{(2)}$,

$$\mathbf{l}_g^{(2)} = \left[\nabla \cdot \left(\left(\nabla \mathbf{m}^{(1)}, \nabla \mathbf{m}^{(2)}\right) \nabla \mathbf{m}^{(1)}\right) - \nabla \cdot \left(\|\nabla \mathbf{m}^{(1)}\|^2 \nabla \mathbf{m}^{(2)}\right)\right]. \quad \text{(D16)}$$

Considering the general form of structural Gramian constraint with transforming and resampling operators, the gradients could be written more generally as

$$\mathbf{l}_g^{(1)} = \mathbf{L}_3^* \mathbf{T}^{(1)*} \left[\nabla \cdot \left(\left(\nabla \mathbf{m}^{(1)}, \nabla \mathbf{m}^{(2)}\right) \nabla \mathbf{m}^{(2)}\right) - \nabla \cdot \left(\|\nabla \mathbf{m}^{(2)}\|^2 \nabla \mathbf{m}^{(1)}\right)\right];$$
$$\mathbf{l}_g^{(2)} = \mathbf{L}_3^* \mathbf{T}^{(2)*} \left[\nabla \cdot \left(\left(\nabla \mathbf{m}^{(1)}, \nabla \mathbf{m}^{(2)}\right) \nabla \mathbf{m}^{(1)}\right) - \nabla \cdot \left(\|\nabla \mathbf{m}^{(1)}\|^2 \nabla \mathbf{m}^{(2)}\right)\right], \quad \text{(D17)}$$

where $\mathbf{L}_3^*$ and $\mathbf{T}^{(j)*}$ represent the corresponding adjoint operators.