Geophysical Journal International

Geophys. J. Int. (2019) **219,** 1225–1236
Advance Access publication 2019 August 01
GJI Seismology

doi: 10.1093/gji/ggz356

Balancing unevenly distributed data in seismic tomography: a global adjoint tomography example

Youyi Ruan [1,2] Wenjie Lei,[2] Ryan Modrak,[2,3] Rıdvan Örsvuran,[4] Ebru Bozdağ[5] and Jeroen Tromp [2,6]

[1] *School of Earth Sciences and Engineering, Nanjing University, Nanjing, Jiangsu* 210023*, China. E-mail: youyir@nju.edu.cn*
[2] *Department of Geosciences, Princeton University, Princeton, NJ* 08520*, USA*
[3] *Geophysical Institute, University of Alaska Fairbanks, Fairbanks, AK* 99775*, USA*
[4] *Laboratoire Géoazur, Université Côte d'Azur, Valbonne* 06560*, France*
[5] *Department of Geophysics, Colorado School of Mines, Golden, CO* 80401*, USA*
[6] *Program in Applied & Computational Mathematics, Princeton University, Princeton, NJ* 08520*, USA*

Accepted 2019 July 31. Received 2019 July 9; in original form 2019 February 7

SUMMARY

The uneven distribution of earthquakes and stations in seismic tomography leads to slower convergence of nonlinear inversions and spatial bias in inversion results. Including dense regional arrays, such as USArray or Hi-Net, in global tomography causes severe convergence and spatial bias problems, against which conventional pre-conditioning schemes are ineffective. To save computational cost and reduce model bias, we propose a new strategy based on a geographical weighting of sources and receivers. Unlike approaches based on ray density or the Voronoi tessellation, this method scales to large full-waveform inversion problems and avoids instabilities at the edges of dense receiver or source clusters. We validate our strategy using a 2-D global waveform inversion test and show that the new weighting scheme leads to a nearly twofold reduction in model error and much faster convergence relative to a conventionally pre-conditioned inversion. We implement this geographical weighting strategy for global adjoint tomography.

Key words: Inverse theory; Waveform inversion; Computational seismology; Seismic tomography; Theoretical seismology.

1 INTRODUCTION

The deployment of new global and regional seismographic stations has made more data available for seismic tomography than ever before. The spatial distribution of these stations remains highly lopsided, however, with limited ocean bottom seismometers or ocean island stations, and fewer stations in the Southern Hemisphere than in the Northern Hemisphere. This uneven station coverage, combined with the uneven distribution of earthquakes dictated by plate tectonics, poses major challenges for the inverse problem. In global adjoint tomography in particular, the large number of paths from subduction zones, such as Fiji Tonga, to dense arrays, such as USArray, causes highly oscillatory behaviour in model updates, hindering convergence (Bozdağ *et al.* 2016).

Since the first global tomographic study by Dziewoński *et al.* (1977), uneven data coverage has been an issue of concern. The problem with a cluster of earthquakes or a dense receiver array is that data residuals are correlated, and large portions of data are in some sense redundant. The correlation of data residuals is reflected in the data covariance matrix: its diagonal terms involve the variance in each individual datum and off-diagonal terms involve the covariance between one datum and another.

For any inverse problem, constructing a data covariance matrix is an important yet challenging task. While full measurement of data covariance is not practically feasible, *a priori* knowledge of data covariance is not completely absent. In practice, correlation between data is related to geographical proximity. Two main strategies have been developed based on this idea.

One of the strategies is to merge data that are highly correlated. In a pioneering traveltime tomography study using ray theory (Spakman & Nolet 1988), all the rays from pre-determined event cluster to a single station are forged into a 'composite ray', and the corresponding delay times averaged to one delay time. This composite ray approach was then adopted in many subsequent studies (e.g. Spakman 1991; Bijwaard *et al.* 1998; Bijwaard & Spakman 2000). A slightly different method was proposed by Morelli & Dziewoński (1987) and refined by van der Hilst *et al.* (1997) in which close rays are grouped to form a 'summary ray', that is, the average of a ray bundle. Both 'composite ray' and 'summary ray' methods reduce the number of data and make rays cover the model space more

evenly. In other words, they merge the correlated data to eliminate the associated off-diagonal terms in the data covariance matrix.

Rather than directly merge close rays and reduce the data space, another strategy to deal with uneven data coverage was introduced by Li & Romanowicz (1996) in a global seismic tomography study. They address this issue by weighting the diagonal terms of the data covariance matrix according to how significantly their errors are correlated with the errors of the other data. The weighting function in this study ($\omega = \omega_e\, \omega_n\, \omega_r$) consists of data errors (ω_e—the rms amplitude of a wave packet), data redundancy within a wave packet (ω_n—the inverse of the square root of number of data) and data redundancy among wave packets sampling similar ray paths (ω_r—a geometrical relationship between a given source–receiver pair to all other pairs). The final term addresses data covariance through a path weighting strategy. Down-weighting correlated paths can help homogenize data coverage and reduce the effect of the associated off-diagonal terms in the data covariance matrix, therefore a full data covariance matrix can be approximated as a diagonal matrix. This method has been employed in a few subsequent global tomography studies (e.g. Lebedev & van der Hilst 2008; Schaeffer & Lebedev 2013; Clouzet *et al.* 2018). In short, introducing either 'composite/summary ray' or 'path weighting' strategies to deal with the uneven data distribution provides a proper design of a data covariance matrix.

Besides working directly in data space, effects of uneven data coverage can be dealt with to some extent through model parametrization and regularization. One solution is to introduce an irregular grid or mesh based on ray coverage (Spakman & Bijwaard 2001), or, more recently, based on the Voronoi tessellation (e.g. Debayle & Sambridge 2004; Zhang & Thurber 2005). Sampling-dependent grids can reduce the number of model parameters and allow high resolution wherever warranted by the data. The other solution is strong damping in the inversion. Boschi & Dziewoński (1999) examined the effects of uneven data coverage with differently parametrized and regularized inversions. They parametrized their global model in terms of both blocks and spherical harmonics. Sparse data coverage in the Southern Hemisphere produced fictitious model features, especially in the spherical harmonic model. Applying stronger regularization or damping, however, more or less resolved the issue. In other words, to resolve bias in inversions caused by uneven data coverage, we need to sacrifice resolution in well-covered regions through stronger damping (e.g. Masters *et al.* 1996; Antolik *et al.* 2003; Dalton & Ekström 2006).

Here, we propose a more general approach to handle unevenly distributed data and examine related convergence rate implications. The motivation for reinvestigating this historical topic is that in recent full-waveform inversions (e.g. Zhu *et al.* 2015; Bozdağ *et al.* 2016), uneven data coverage caused oscillatory behaviour in model updates and slowed convergence. The numerical solvers used in full-waveform inversion are computationally very expensive, and so to reduce cost it is crucial to address the issue of slow convergence due to unevenly distributed data. In full-waveform tomography, ray paths are replaced by waveform Fréchet derivatives associated with different phases (e.g. direct body waves, multiply reflected body waves, surface waves) all in one 'event kernel' despite the very different 'paths' associated with each phase (Tromp *et al.* 2005). Untangling these paths, calculating their relative proximity or summing a bundle of them becomes computationally prohibitive because it requires many additional wavefield simulations. Instead, this paper demonstrates an efficient and scalable geographical weighting strategy to address data correlation, applicable not only to full-waveform inversion but also to seismic inverse problems more generally. Using 2-D numerical experiments, we validate this strategy and show its faster convergence rate and larger model error reduction compared with previous methods. Finally, we demonstrate how to apply this weighting strategy in global adjoint tomography.

Table 1. Measurement categories in global adjoint tomography (Bozdağ *et al.* 2016; Lei *et al.* 2019). Seismic waves are categorized in terms of complementary period bands on three components of motion.

	Vertical (Z)	Radial (R)	Transverse (T)
17–40 s	*P–SV* body waves	*P–SV* body waves	*SH* body waves
40–100 s	*P–SV* body waves	*P–SV* body waves	*SH* body waves
40–100 s	Rayleigh waves	Rayleigh waves	Love waves
90–250 s	Rayleigh waves	Rayleigh waves	Love waves

2 WEIGHTING STRATEGY

2.1 Measurements and misfit function

Seismic waves sample different parts of Earth's interior at different dominant frequencies, so it is natural to categorize seismic signals in terms of their type and band. In global adjoint tomography (Bozdağ *et al.* 2016; Lei *et al.* 2019), we currently consider three period bands: 17–40 s, 40–100 s and 90–250 s. Typically, we select body waves in the 17–40 s band, body waves in the 40–100 s band, surface waves in the 40–100 s band and surface waves in the 90–250 s band. Within each band, we consider three-component seismograms rotated into vertical, radial and transverse directions of motion, so in total there are 12 data categories, as summarized in Table 1.

Considering all these data categories, along with all the sources and receivers available for an inversion, we define an overall data misfit

$$\Phi = \sum_{s=1}^{S} \omega_s \sum_{c=1}^{C} \omega_c \sum_{r=1}^{R_{sc}} \omega_{scr} \sum_{w=1}^{N_{scr}} \omega_{scrw}\, \chi_{scrw}. \tag{1}$$

Here, $s = 1, ..., S$ denotes a given source and S the total number of sources. Likewise, $c = 1, ..., C$ denotes a given category and C the total number of categories (in our case, the 12 categories summarized in Table 1). R_{sc} denotes the number of receivers for a given source s and category c. Finally, N_{scr} denotes the number of measurement windows for a given source s, category c and receiver r.

The misfit for a given source s, category c, receiver r and measurement window w is

$$\chi_{scrw} = \left(\frac{\Delta d_{scrw}}{\sigma_{scrw}} \right)^2, \tag{2}$$

where Δd_{scrw} denotes a measurement with associated uncertainties σ_{scrw}. When the model fits the data to within one standard deviation, we expect that

$$\chi_{scrw} \sim 1. \tag{3}$$

In the following sections we discuss various options for the assignment of the source, category, receiver and window weights, ω_s, ω_c, ω_{scr}, and ω_{scrw}, respectively.

2.2 Category-weighting strategy

We start by considering an ideal case where each datum in a certain category is independent and the associated errors are not correlated. (In reality, the measurement error of a datum is difficult to estimate, not to mention the correlation of the errors.) The standard deviation σ_{scrw} in eq. (2) is often set as an *a priori* constant in the practice

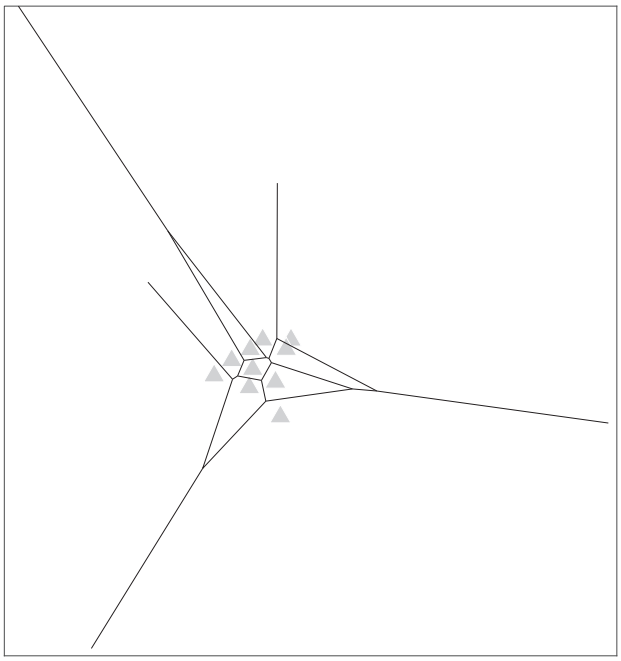

Figure 1. A Voronoi tessellation constructed for a dense array of stations, marked by grey triangles. The area of cells corresponding to stations at the edge of the array is hundreds of times more than stations removed from the edge. We require a different sort of weighting scheme in which all stations from a dense array have roughly similar weights.

of inversion. Under such conditions, a common weighting strategy is to assign a constant weight to all windows, receivers, and sources, that is,

$$\omega_s = 1, \tag{4}$$

$$\omega_{scr} = 1, \tag{5}$$

$$\omega_{scrw} = 1. \tag{6}$$

We seek to define a misfit function Φ such that when the model fits the data to within one standard deviation, $\Phi \sim 1$, thereby making Φ similar to a χ^2/N distribution. The data in each category should contribute equally to the overall misfit, which implies that we should choose a category weight

$$\omega_c = \frac{1}{C}\,\frac{1}{N_c}, \tag{7}$$

where N_c is the number of measurements in category c, that is,

$$N_c = \sum_{s=1}^{S}\sum_{r=1}^{R_{sc}} N_{scr}. \tag{8}$$

Note that the category weight in eq. (7) is independent of source s. Now we see that when the model fits the data to within one standard deviation, that is, $\chi_{scrw} \sim 1$, then

$$\Phi \sim \sum_{s=1}^{S}\omega_s \sum_{c=1}^{C}\omega_c \sum_{r=1}^{R_{sc}}\omega_{scr}\sum_{w=1}^{N_{scr}}\omega_{scrw} \tag{9}$$

$$= \sum_{s=1}^{S}\sum_{c=1}^{C}\omega_c \sum_{r=1}^{R_{sc}} N_{scr} \tag{10}$$

$$= \sum_{c=1}^{C}\frac{1}{C}\,\frac{1}{N_c}\sum_{s=1}^{S}\sum_{r=1}^{R_{sc}} N_{scr} \tag{11}$$

$$= \sum_{c=1}^{C}\frac{1}{C} \tag{12}$$

$$= 1, \tag{13}$$

as desired.

Let us next analyse the contribution of each datum to the misfit function at various levels. At the receiver level, we consider

$$\chi_{scr} = \sum_{w=1}^{N_{scr}}\omega_{scrw}\,\chi_{scrw}. \tag{14}$$

When we assign a weight of one to all windows, that is, $\omega_{scrw} = 1$, thus putting them on the same footing, and when the model fits the data to within one standard deviation, that is, $\chi_{scrw} \sim 1$, we see that

$$\chi_{scr} \sim N_{scr}. \tag{15}$$

Next, we consider the misfit for a given source s and category c, namely,

$$\chi_{sc} = \sum_{r=1}^{R_{sc}}\omega_{scr}\sum_{w=1}^{N_{scr}}\omega_{scrw}\,\chi_{scrw} = \sum_{r=1}^{R_{sc}}\omega_{scr}\,\chi_{scr}. \tag{16}$$

When the model fits the data to within one standard deviation, we find that

$$\chi_{sc} \sim \sum_{r=1}^{R_{sc}} N_{scr} = N_{sc}, \tag{17}$$

where N_{sc} denotes the number of measurements for a given source s in category c. Note that we have used the fact that $\omega_{csr} = 1$, meaning all receivers are weighted equally.

Since category weighting is independent of source weighting, we can change the order of summation. The misfit in a given category c is given by

$$\chi_c = \sum_{s=1}^{S}\omega_s\,\chi_{sc}. \tag{18}$$

Since $\omega_s = 1$, we see that when the model fits the data to within one standard deviation

$$\chi_c \sim \sum_{s=1}^{S} N_{sc} = N_c, \tag{19}$$

as expected.

Finally, the total misfit function is

$$\Phi = \sum_{c=1}^{C}\omega_c\,\chi_c, \tag{20}$$

and thus we see that when the model fits the data to within one standard deviation

$$\Phi \sim \sum_{c=1}^{C}\frac{1}{C}\,\frac{1}{N_c}\,N_c = 1, \tag{21}$$

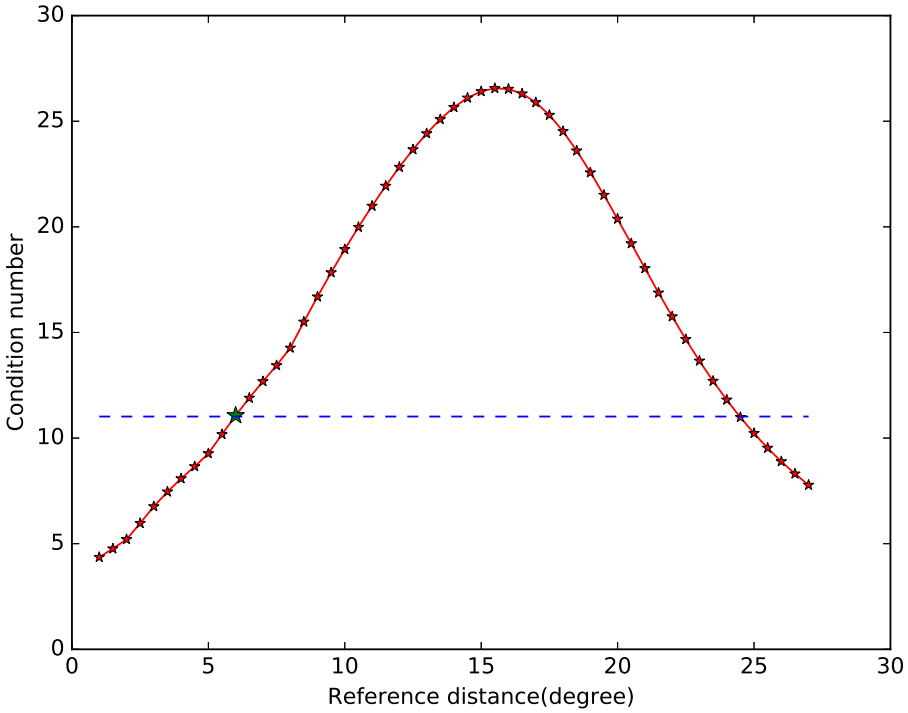

Figure 2. Condition number of the diagonal weighting matrix defined by eq. (22) as a function of the reference distance Δ_0. The chosen value, indicated by the green star, is about one-third of the largest possible value. Since evaluation of eq. (22) for different reference distances adds negligible computational expense, it is possible (and recommended) to repeat this type of analysis for each iteration.

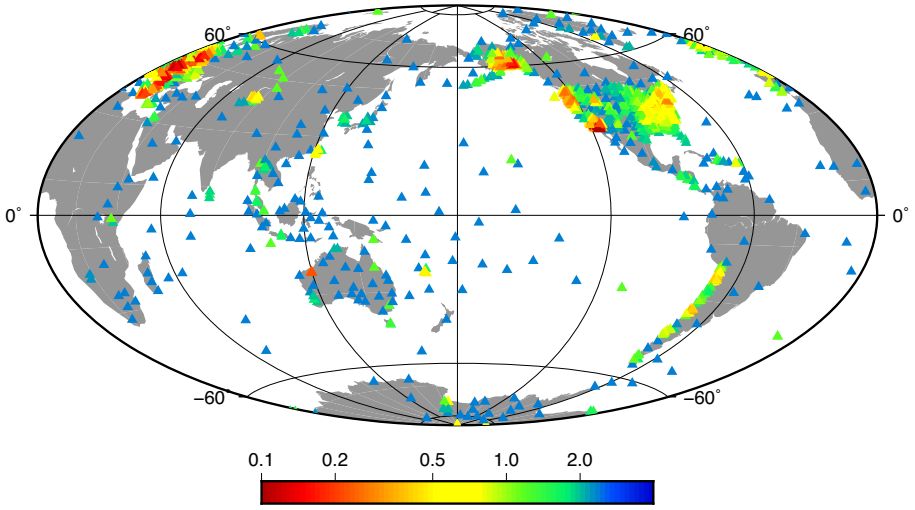

Figure 3. Example of receiver weights for an event C201604131355A at 40–100 s period band and vertical component determined based upon eq. (22) and normalized according to eq. (25). The weights are in logarithmic scale. Note the difference between USArray stations and ocean island stations.

as required. This weighting scheme is widely used in a variety of inversions (e.g. Zhu *et al.* 2015; Bozdağ *et al.* 2016). The behaviour of the misfit function at the various levels demonstrated above will guide the design of source and receiver weights in the next section.

2.3 Geographical-weighting strategy

In the previous section we treated each measurement datum in a given category equally, which means each datum was weighted only by the total number of measurements in its category, regardless of possible correlations between data. Solving the inverse problem based on this 'all data are equal' strategy can fail sometimes because the uneven spatial distribution of sources and receivers on Earth's surface is found to negatively affect convergence rate. In particular, if data from dense regional arrays such as USArray are included, progress of the inversion can be extremely slow. In this section, we add additional weights associated with the geographical distribution of sources and receivers, with the goal of down-weighting densely sampled regions, so that we obtain more uniform spatial sampling and minimize the dominant effects of dense regional arrays.

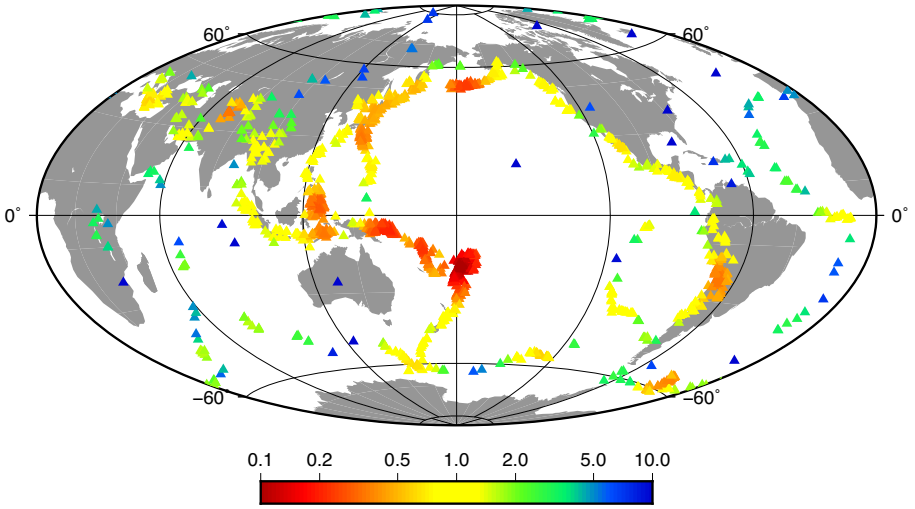

Figure 4. Example of source weights determined based upon eq. (22) and normalized according to eq. (24). Note the difference between dense subduction zone earthquakes and sparse transform fault earthquakes.

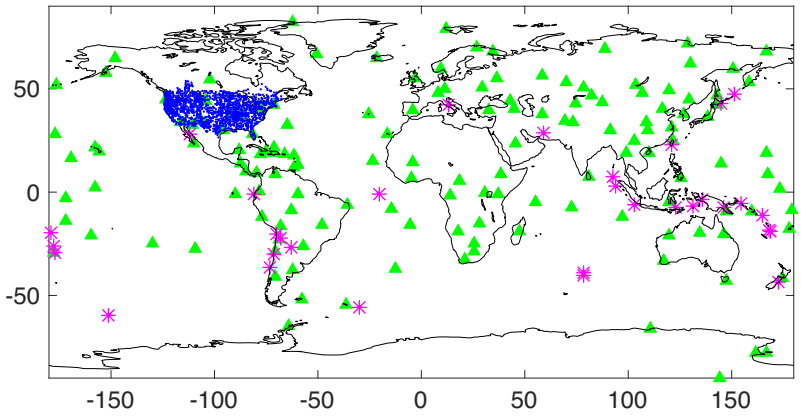

Figure 5. Source–receiver geometry used in synthetic inversions. GSN stations are labelled by green triangles, USArray stations by blue triangles and sources by magenta asterisks.

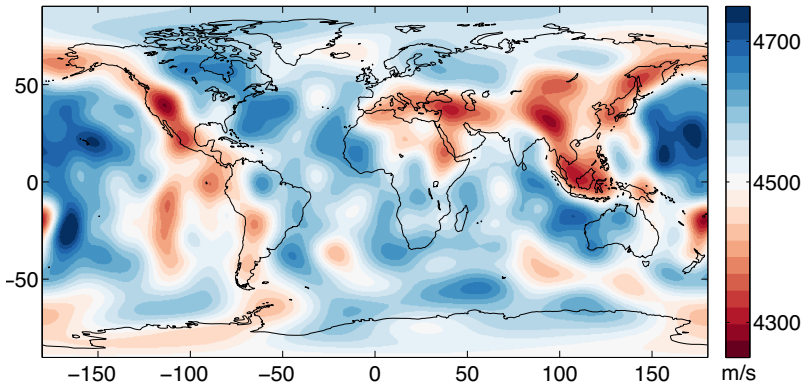

Figure 6. Target model used in the synthetic inversions. Wave speeds are determined by the 40 s Rayleigh wave phase speed model of Trampert & Woodhouse (2003).

2.3.1 Geographical weighting

For closely located sources or dense receiver arrays, measurement errors are correlated for a variety of reasons (e.g. Li & Romanowicz 1996), and the degree of correlation is associated with geographical distance. With these correlations in mind, we seek to define a source and receiver weighting scheme that results in more uniform spatial sampling. In such a scheme, areas with dense sampling, like Japan or North America, should be down-weighted relative to areas which are sparsely sampled, such as the Southern Ocean.

To determine weights for each source and receiver, one option is to construct a Voronoi tessellation (e.g. Du *et al.* 1999) of the sources

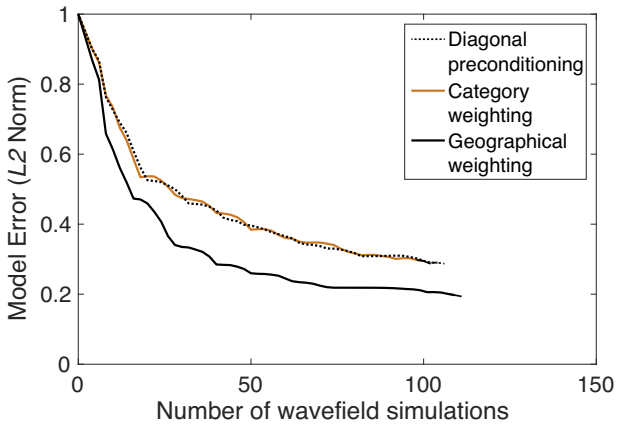

Figure 7. With the category weighting defined in Section 2.2, convergence is extremely slow. With the geographical weighting discussed in Section 2.3, convergence is much faster. Diagonal model-space pre-conditioning, it turns out, is not an effective alternative to weighting when dealing with extremely lopsided source–receiver distributions.

or receivers, and then weight each source or receiver by the area of its corresponding cell. Although the Voronoi weighting has been used successfully in regional and global studies (e.g. Böhm *et al.* 2000; Debayle & Sambridge 2004; Zhang & Thurber 2005), when applied to dense station arrays or event clusters, the approach becomes unstable. In some cases, the stations at the edge of a dense array can be weighted orders of magnitude more than stations slightly removed from the edge, which is undesirable. Fig. 1 illustrates this effect. To avoid convergence problems, we require that stations from a dense array carry roughly equal weights.

A robust alternative approach is as follows. Given a set of N receivers, calculate the epicentral distance Δ_{ij} for each receiver pair. The weight ω_i assigned to each receiver i is calculated via

$$\omega_i^{-1} = \sum_{j=1}^{N} \exp\left[-\left(\frac{\Delta_{ij}}{\Delta_0} \right)^2 \right], \tag{22}$$

where Δ_0 is a reference distance parameter. We note that if a station has few nearby stations, it is assigned a larger weight than if it has many nearby stations. For large values of Δ_0 the scheme reduces to the category weighting in Section 2.2. The adjustable reference distance avoids the issue of overwhelming weights in the Voronoi scheme.

We also note that an equivalent weighting scheme for sources can be obtained by substituting source pairs for receiver pairs in the above expression. In Section 2.3.3 we discuss how weights determined by eq. (22) are normalized to obtain weights ω_c for each event and ω_{csr} for each receiver associated with a certain event.

Interestingly, eq. (22) is just a discretized version of the 'smoothing' method used in many parameter estimation and optimal design studies (Modrak & Tromp 2016). Rather than applying it in the model space, however, we are applying it in the data space to 'smooth' the discrete receiver distribution.

2.3.2 *Choice of reference parameter*

The distribution of weights calculated from eq. (22) depends strongly on the reference distance Δ_0. This parameter needs to be carefully chosen so that the ratio of maximum to minimum weights is not unreasonably small or large. In the example in Fig. 2, we chose Δ_0 so that the ratio is about one-third of the largest possible ratio for all choices of Δ_0. In other words, we chose the reference length parameter to be about one-third of the most aggressive value.

Based on this choice, Figs 3 and 4 illustrate the distribution of weights in a recent global adjoint tomography study (Lei *et al.* 2019). In Fig. 3, weights of USArray and European stations are brought down to about one-tenth of ocean island station's weights. The source weighting is similar, with the contribution of individual Fiji Tonga events brought down to about one-tenth the contribution of intraplate events in Asia. This ratio between the minimum and maximum weights can be adjusted through the reference distance parameter, as informed by practical experience in a given inversion.

2.3.3 *Weighting normalization*

In this section our goal is to introduce geographical source and receiver weighting without changing the event- and category-level behaviour of the misfit function discussed in Section 2.2.

In geometrical ray-based tomography, specific phases are identified and windowed. These phases usually correspond to distinct paths that provide constraints on different parts of Earth's interior, and some phases may be assigned more weight than others depending on the aims of the researcher. In waveform inversion, any part of the wave train can be selected and phase identification is no longer required. We thus assign a uniform weight to all windows in a seismogram,

$$\omega_{scrw} = 1, \tag{23}$$

such that, according to eq. (15), $\chi_{scr} \sim N_{scr}$. The normalization of the geographical source weights from eq. (22) is then determined in a straightforward manner by

$$\sum_{s=1}^{S} \omega_s = S. \tag{24}$$

Because the number of receivers that happen to be online varies from one source to another, including receiver weights in an inversion is not as straightforward. To guide the normalization of receiver weights, we adopt the same type of analysis performed in connection with category-weighting and eq. (16).

If the model fits the data to within one standard deviation, the misfit in a certain source and category χ_{sc} approaches N_{sc}, and the receiver-level misfit χ_{scr} approaches N_{scr} (eqs 15–17). These properties of the misfit imply a normalization requirement for the receiver weights ω_{scr}, determined by eq. (22) for a given category c and source s:

$$\sum_{r=1}^{R_{sc}} \omega_{scr}\, N_{scr} = N_{sc}. \tag{25}$$

When $\omega_{scr} = 1$, as in the simple weighting strategy, this normalization condition is naturally satisfied because

$$\sum_{r=1}^{R_{sc}} N_{scr} = N_{sc}. \tag{26}$$

After we determine the source and receiver weights, what is left is to examine the category weights ω_c. For the misfit function given by eq. (1), we want the contributions from each category to be balanced, and this implies that

$$\omega_c \sum_{s=1}^{S} \omega_s \sum_{r=1}^{R_{sc}} \omega_{scr}\, N_{scr} = \omega_c \sum_{s=1}^{S} \omega_s\, N_{sc} = \frac{1}{C}, \tag{27}$$

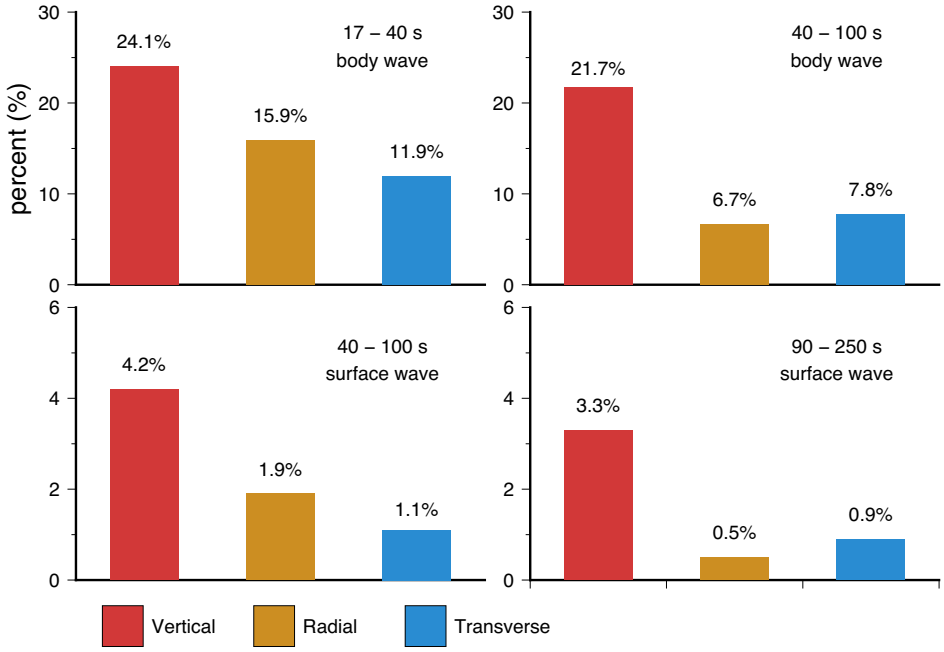

Figure 8. Percentage of data (window count) in each of the 12 categories (see Table 1). Period band and wave type are labelled in the upper right corner of each panel. Note the dramatic differences between body-wave (top) and surface-wave data (bottom). The large variations in data count from category to category illustrate the need for balancing.

and thus

$$\omega_c = \frac{1}{C}\frac{1}{\sum_{s=1}^{S}\omega_s N_{sc}}. \tag{28}$$

Note that when $\omega_s = 1$, the weighting reduce to the category-weighting strategy, namely,

$$\omega_c = \frac{1}{C}\frac{1}{N_c}. \tag{29}$$

From the normalization of geographical weights described above, we see that when the model fits the data to within one standard deviation, the source-level misfit χ_{sc} approaches N_{sc} and category-level misfit χ_c approaches N_c.

In theory, data correlation changes the degrees of freedom in the data space of an inversion. Loosely speaking, the above normalization can be thought of as changing the degrees of freedom of the data set for a given category c and source s, as well as the degrees of freedom of the data set in a given category c. Geographical weighting can be viewed as an approximation of the complete data covariance matrix.

3 NUMERICAL VALIDATION: 2-D GLOBAL ADJOINT TOMOGRAPHY

Data space weighting and model space pre-conditioning are two common approaches to balance uneven coverage in the inverse problem. In this section we investigate performance of the geographical weighting strategy through comparisons with the category weighting strategy and pre-conditioning in terms of model error reduction and convergence rate.

In full-waveform inversion, model space pre-conditioning is accomplished through a change of variables in the nonlinear conjugate gradient, quasi-Newton, or other nonlinear optimization methods used to minimize the objective function (e.g. Modrak & Tromp 2016).

Model space pre-conditioning differs from data space pre-conditioning in that the former is applied to the gradient and the latter to the Jacobian or its action. Because the expense of working with the Jacobian generally far outweighs any increase in convergence rate, data space pre-conditioning becomes impractical for large-scale problems. Data space weighting, in turn, differs from both types of pre-conditioning in that, rather than just a change of variables in the optimization algorithm, it effectively changes the objective function.

To validate the geographical weighting strategy, we performed 2-D inversions with a global test problem. In these numerical experiments, we used both Global Seismographic Network (GSN) stations, which are sparsely distributed at the global scale, and USArray stations, which densely cover the North American continent, as shown in Fig. 5. For the target model, we employed the acoustic test case shown in Fig. 6. We generated synthetic data for this model using periodic boundary conditions at the edges to approximate a spherical Earth. Finally, we inverted these data with the workflow described by Modrak *et al.* (2018).

To test the geographical weighting strategy, we performed 2-D inversions with a global test problem. In these experiments, we used both GSN stations, which are sparsely distributed at the global scale, and USArray stations, which densely cover the North American continent, as shown in Fig. 5. For the target model, we employed the acoustic test case shown in Fig. 6. We generated synthetic data for this model using periodic boundary conditions at the edges to approximate a spherical Earth. Finally, we inverted these data with the workflow described by Modrak *et al.* (2018).

Starting from a homogeneous model, we tracked the reduction in model error as a function of the number of wavefield simulations in

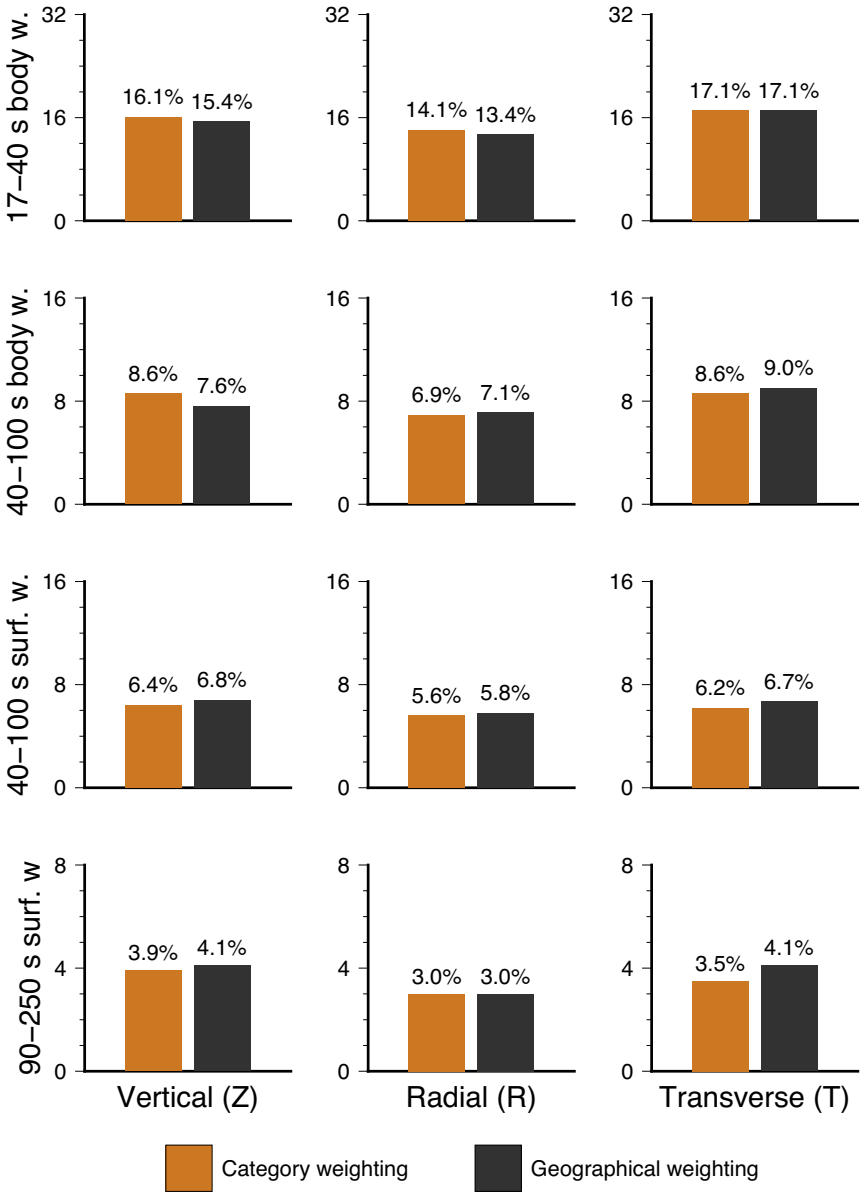

Figure 9. Percentage of the misfit in each of the 12 categories (see Table 1) using the category-weighting strategy (orange) and geographical-weighting strategy (black). The percentage contribution in each category is labelled above the bars. Note that each category's contribution to the misfit barely changes with geographical weighting because weight rebalancing only happens within each category.

three separate inversions. In the first inversion, we employed model-space diagonal pre-conditioning, using the best-performing pre-conditioner of all the variants tested by Modrak & Tromp (2016). In the second inversion, we employed the category-weighting strategy discussed in Section 2.2. In the third inversion, we used the geographical-weighting strategy described in Section 2.3. The performance of the three methods is shown in Fig. 7.

Compared with category weighting, pre-conditioning fails to provide an effective improvement, while the geographical-weighting strategy provides a much faster convergence rate. Considering the high cost of large-scale inverse problems like global adjoint tomography, where one iteration can require millions of core hours, the saving could be significant. In addition to the performance improvement, the geographically weighted inversion demonstrates larger model error reduction than the other approaches.

4 APPLICATION TO 3-D GLOBAL ADJOINT TOMOGRAPHY: MISFIT STATISTICS

After testing the geographical-weighting strategy through synthetic experiments, we deployed it in our ongoing global adjoint tomography study (Lei *et al.* 2019) with the goal of obtaining faster convergence and a better model. In this section we illustrate various aspects of the above category- and geographical-weighting strategies.

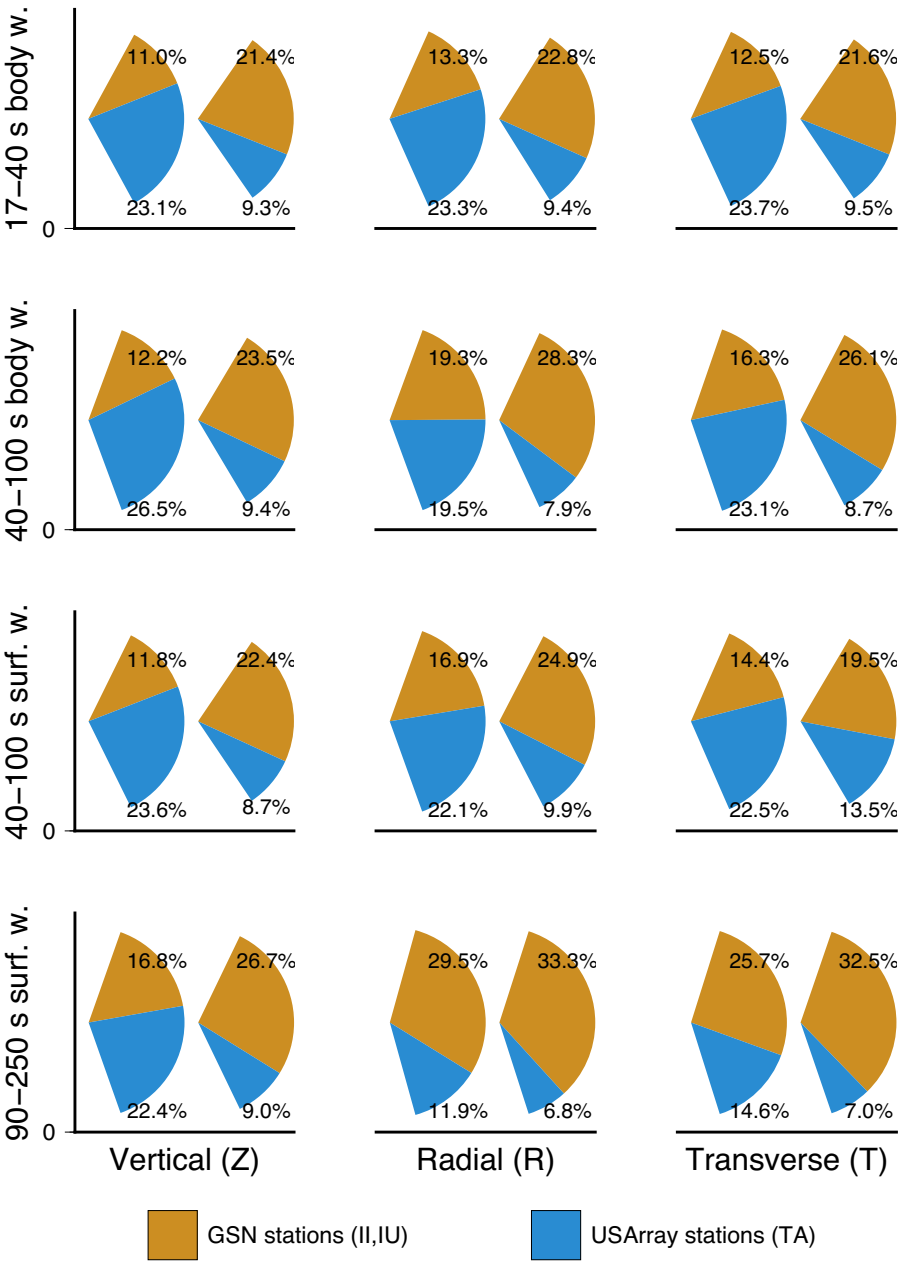

Figure 10. Percentage of the misfit of GSN stations (II, IU) and USArray stations (TA) in each category using category-weighting (left wedges in each panel) and geographical-weighting strategies (right wedges in panel). Compared with the category-weighting strategy, the geographical-weighting strategy assigns more weight to GSN stations and down weights USArray stations in all categories.

As described in Section 2, the weight we assign to each measurement is the product of source, category, receiver, and window weights: $\omega_s\,\omega_c\,\omega_{scr}\,\omega_{scrw}$. In Figs 3 and 4, we plotted weights ω_s assigned to sources and weights ω_{scr} assigned to receivers. Next, we examine the misfit when the product of all four weights is applied.

In total, we picked more than 17 million windows from 1480 sources and 12 categories (Lei *et al.* 2019). As shown in Fig. 8, the contribution from each category is far from balanced. Short period body-wave data (17–40 s) account for more than 50 per cent of the total number of windows while long period surface waves (90–250 s) contribute less than 5 per cent. Across all three periods bands, more than 80 per cent of windows correspond to body-wave data. For a given period band, the vertical component always provides more data than the horizontals. If not balanced, these differences between categories can cause regions sensitive to body waves to be updated more than regions sensitive to surface waves and slow down the overall convergence rate.

We define the weighted misfit for each category as

$$\Phi_c = \sum_{s=1}^{S}\sum_{r=1}^{R_{sc}}\sum_{w=1}^{N_{scr}} \omega_s\,\omega_c\,\omega_{scr}\,\omega_{scrw}\,\chi_{scr}\,. \tag{30}$$

In Fig. 9, we calculate the percentage of the summed misfit for each category using two weighting strategies: (1) orange bars correspond to category-only weighting and (2) black bars correspond to the full category- and geographical-weighting strategy.

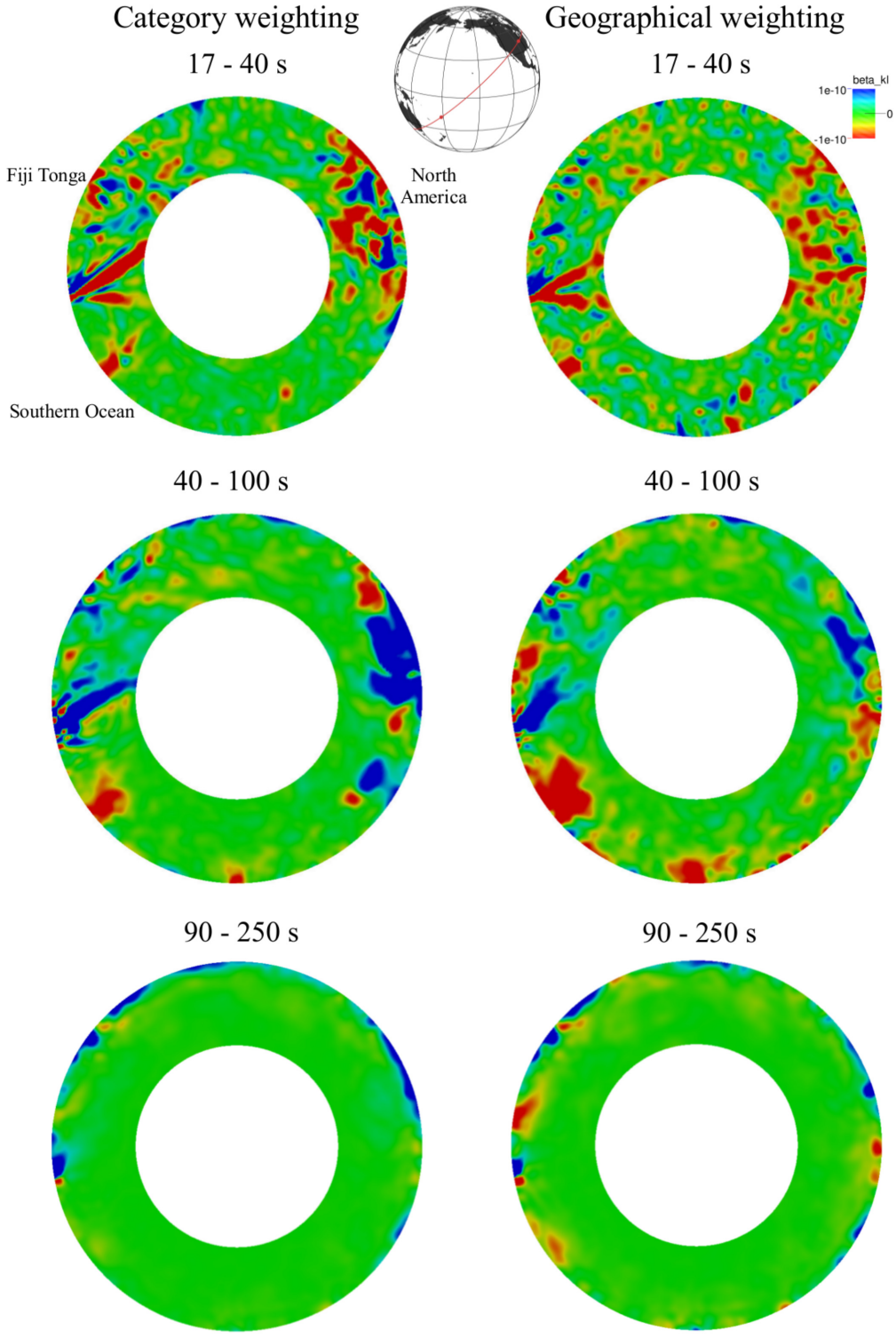

Figure 11. Smoothed shear wave-speed gradient contributions using data from 42 earthquakes without (left-hand column) and with source and receiver weighting (right-hand column) in three period bands: 17–40 s (top row), 40–100 s (middle row) and 90–250 s (bottom row). The top centre map shows the cross-section with a red line for reference. The isotropic smoothing length scale is 100 km.

In the first case, after employing the category-weighting strategy in which the misfit is normalized by the total number of data in each category (Fig. 8), misfits from different categories are better balanced and do not vary dramatically from category to category. Although the summed weights themselves are equal in each category, the weighted misfits from each category are not. The weighted misfits of short-period body waves are on average two to three times larger than the misfits of long-period body and surface waves. We attribute this to the larger number of updates required to fit

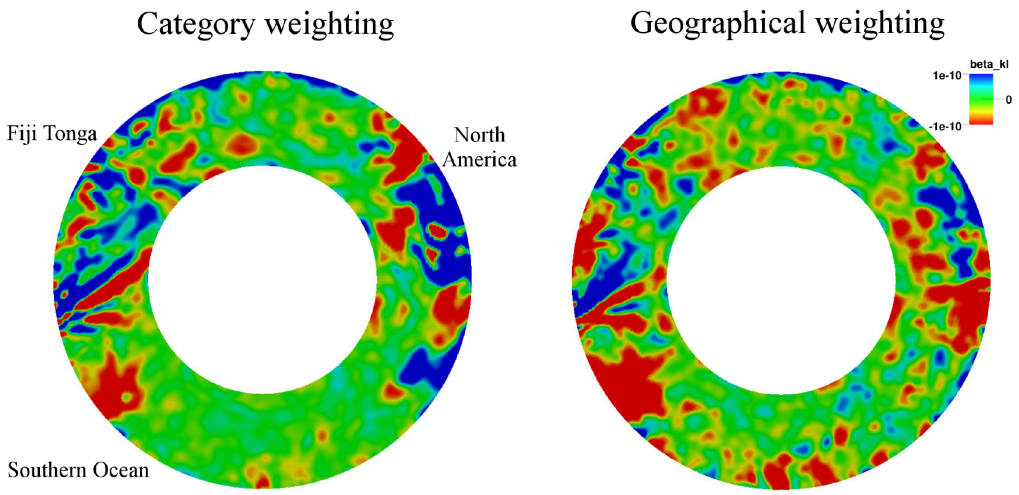

Figure 12. Example of a smoothed shear wave-speed gradient (weighted summation of the contributions of the three period bands shown in Fig. 11) using data from 42 earthquakes without (left) and with source and receiver weighting (right). The isotropic smoothing length scale is 100 km.

short-period phases compared with long-period phases, and we expect the body-wave misfits to decrease as the inversion progresses.

In the second case, when geographical weighting is applied, the distribution of misfits in each category does not significantly change, meaning that the re-balancing happens only within each category, as desired.

To further probe the overall re-balancing within each category, we compare the contribution to the total misfit from two seismographic networks: USArray (network ID TA) and GSN (network IDs II and IU). USArray stations are densely distributed across North America and GSN stations are sparsely distributed across the globe.

The total misfit from USArray or GSN stations in each category is given by

$$\Phi_c^{\text{USArray / GSN}} = \sum_{s=1}^{S} \sum_{r=1}^{R_{sc}} \sum_{w=1}^{N_{scr}} \omega_s \, \omega_c \, \omega_{scr} \, \omega_{scrw} \, \chi_{scr}^{\text{USArray / GSN}}. \qquad (31)$$

Fig. 10 shows the percentage of Φ_c^{USArray} and Φ_c^{GSN} in each category under different weighting strategies. In the category-weighing strategy, GSN stations contribute much less to the overall misfit than USArray stations, reflecting the proportionality to the amount of data, as expected. In the geographical-weighting strategy, the contribution of GSN stations is enhanced due to the re-balancing.

To further illustrate the effects of weighting, we selected 42 events for a pilot test and examined the model update gradient. Ideally, we should run forward and adjoint simulations for each of the 12 categories for one weighting strategy, and repeat this process for the other weighting strategy, which would require 2016 simulations in total. To save computational cost while keeping the test meaningful, we considered only three categories, 17–40 s, 40–100 s and 90–250 s, and ignored wave types for the adjoint simulations. Fig. 11 shows cross-sections of the model update gradient. In the shortest period band (17–40 s), the gradient based on category-weighting is dominated by regions beneath Fiji Tonga and North America, with limited updates in the Southern Hemisphere. In contrast, the geographical-weighting approach results in a balanced gradient with more information in the Southern Hemisphere and relatively reduced sensitivity beneath Fiji Tonga and North America. The longer period bands, involving mostly surface waves, demonstrate similar behaviour but more focused on the shallow mantle. Upon combining all three categories, we clearly see that the model update based on the geographical weighting strategy is better balanced, with a more even sampling of the whole mantle (Fig. 12). From this pilot test we conclude that the geographical-weighting strategy effectively balances the inversion and improves the convergence rate, which is necessary for any inversion dealing with the unevenly distributed data.

5 CONCLUSION

We propose a geographic weighting strategy to address uneven data coverage in seismic tomography. To test the approach, we performed synthetic 2-D global adjoint tomography experiments using realistic source and receiver distributions. The results show that geographical weighting performs better than category-only weighting and diagonal model-space pre-conditioning. A 42-event pilot test was used to illustrate how geographical weighting balances densely and sparsely sampled regions. Finally, using a database of 1480 earthquakes, we performed a statistical analysis of 17 million measurements assimilated in the global adjoint tomography inversion of Lei *et al.* (2019), verifying expected effects of the weighting scheme. Although we only demonstrated an example of global full-waveform adjoint tomography, this approach should be applicable to any inverse problem where the sources/receivers are unevenly distributed. As more data from dense regional seismographic networks become available, we expect weighting to play an increasingly important role in scientific studies of Earth's interior (e.g. Örsvuran *et al.* 2019).

ACKNOWLEDGEMENTS

We gratefully acknowledge editor Lapo Boschi, David Al-Attar, and two anonymous reviewers for their constructive reviews which improved the manuscript. This research used resources of the Oak Ridge Leadership Computing Facility (OLCF), which is a DOE Office of Science User Facility supported under contract DE-AC05-00OR22725. This research was supported by the U.S. National Science Foundation under grant 1644826. YR was also supported by the Fundamental Research Fund for Central Universities under grant number 0206-14380080. The 2-D numerical experiments

were performed on 'Tiger' at the Princeton Institute for Computational Science and Engineering, and the 42-event pilot tests were performed on the Cray XK7 supercomputer 'Titan' at the OLCF.

REFERENCES

Antolik, M., Gu, Y.J., Ekström, G. & Dziewoński, A.M., 2003. J362D28: a new joint model of compressional and shear velocity in the Earth's mantle, *Geophys. J. R. astr. Soc.,* **153**(2), 443–466.

Bijwaard, H. & Spakman, W., 2000. Non-linear global *P*-wave tomography by iterated linearized inversion, *Geophys. J. Int.,* **141,** 71–82.

Bijwaard, H., Spakman, W. & Engdahl, E.R., 1998. Closing the gap between regional and global travel time tomography, *J. geophys. Res.,* **103**(B12), 30 055–30 078.

Boschi, L. & Dziewoński, A.M., 1999. High- and low-resolution images of the Earth's mantle: implications of different approaches to tomographic modeling, *J. geophys. Res.,* **104**(B11), 25 567–25 594.

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N. & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophys. J. Int.,* **207**(3), 1739–1766.

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

Clouzet, P., Masson, Y. & Romanowicz, B., 2018. Box tomography: first application to the imaging of upper-mantle shear velocity and radial anisotropy structure beneath the North American continent, *Geophys. J. Int.,* **213,** 1849–1875.

Dalton, C.A. & Ekström, G., 2006. Global models of surface wave attenuation, *J. geophys. Res.,* **111,** B05317.

Debayle, E. & Sambridge, M., 2004. Inversion of massive surface wave data sets: model construction and resolution assessment, *J. geophys. Res.,* **109,** B02316.

Du, Q., Faber, V. & Gunzburger, M., 1999. Centroidal Voronoi tessellations: applications and algorithms, *SIAM Rev.,* **41**(4), 637–676.

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

Lebedev, S. & van der Hilst, R.D., 2008. Global upper-mantle tomography with the automated multimode inversion of surface and *S*-wave forms, *Geophys. J. R. astr. Soc.,* **173**(2), 505–518.

Lei, W. *et al.*, 2019. Global adjoint tomography – Model GLAD-M25, *Geophys. J. Int.,* submitted.

Li, X.-D. & Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *J. geophys. Res.,* **101**(B10), 22 245–22 272.

Masters, G., Johnson, S., Laske, G., Bolton, H. & Davies, J.H., 1996. A shear-velocity model of the mantle, *Phil. Trans. R. Soc. Lond., A,* **354**(1711), 1385–1411.

Modrak, R. & Tromp, J., 2016. Seismic waveform inversion best practices: regional, global and exploration test cases, *Geophys. J. Int.,* **206**(3), 1864–1889.

Modrak, R.T., Borisov, D., Lefebvre, M. & Tromp, J., 2018. Seisflows—flexible waveform inversion software, *Comput. Geosci.,* **115,** 88–95.

Morelli, A. & Dziewoński, A.M., 1987. The harmonic expansion approach to the retrieval of deep earth structure, in *Seismic Tomography,* ed. Nolet, G., pp. 251–274, D. Reidel Publishing Company.

Schaeffer, A.J. & Lebedev, S., 2013. Global shear speed structure of the upper mantle and transition zone, *Geophys. J. Int.,* **194**(1), 417–449.

Spakman, W., 1991. Delay-time tomography of the upper mantle below Europe, the Mediterranean, and Asia Minor, *Geophys. J. Int.,* **64,** 888–901.

Spakman, W. & Bijwaard, H., 2001. Optimation of Cell Parameterations for Tomographyic Inverse Problem, *Pure and Applied Geophysics,* **158,** 1401–1423.

Spakman, W. & Nolet, G., 1988. Imaging algorithms, accuracy and resolution in delay time tomography, in *Mathematical Geophysics,* pp. 155–187, eds Vlaar, N.J., Nolet, G., Wortel, M.J.R. & Cloetingh, S.A.P.L., D. Reidel Publishing Company.

Trampert, J. & Woodhouse, J.H., 2003. Global anisotropic phase velocity maps for fundamental mode surface waves between 40 and 150 s, *Geophys. J. Int.,* **154**(1), 154–165.

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

van der Hilst, R.D., Widiyantoro, S. & Engdahl, E.R., 1997. Evidence for deep mantle circulation from global tomography, *Nature,* **386,** 578–584.

Zhang, H. & Thurber, C., 2005. Adaptive mesh seismic tomography based on tetrahedral and Voronoi diagrams: application to Parkfield, California, *J. geophys. Res.,* **110,** B04303.

Zhu, H., Bozdağ, E. & Tromp, J., 2015. Seismic structure of the European upper mantle based on adjoint tomography, *Geophys. J. Int.,* **201**(1), 18–52.

Örsvuran, R., Bozdağ, E., Modrak, R., Lei, W. & Ruan, Y., 2019. Towards double-difference global adjoint tomography, *Geophys. J. Int*. submitted.