Double Differencing by Demeaning: Applications to Hypocenter Location and Wavespeed Tomography

Steven Roecker*[1], Ariane Maharaj[1], Sean Meyers[1], and Diana Comte[2]

ABSTRACT

Double differencing of body-wave arrival times has proved to be a useful technique for increasing the resolution of earthquake locations and elastic wavespeed images, primarily because (1) differences in arrival times often can be determined with much greater precision than absolute onset times and (2) differencing reduces the effects of unknown, unmodeled, or otherwise unconstrained variables on the arrival times, at least to the extent that those effects are common to the observations in question. A disadvantage of double differencing is that the system of linearized equations that must be iteratively solved generally is much larger than the undifferenced set of equations, in terms of both the number of rows and the number of nonzero elements. In this article, a procedure based on demeaning subsets of the system of equations for hypocenters and wavespeeds that preserves the advantages of double differencing is described; it is significantly more efficient for both wavespeed-only tomography and joint hypocenter location-wavespeed tomography. Tests suggest that such demeaning is more efficient than double differencing for hypocenter location as well, despite double-differencing kernels having fewer nonzeros. When these subsets of the demeaned system are appropriately scaled and simplified estimates of observational uncertainty are used, the least-squares estimate of the perturbations to hypocenters and wavespeeds from demeaning are identical to those obtained by double differencing. This equivalence breaks down in the case of general, observation-specific weighting, but tests suggest that the resulting differences in least-squares estimates are likely to be inconsequential. Hence, demeaning offers clear advantages in efficiency and tractability over double differencing, particularly for wavespeed tomography.

KEY POINTS

- Double differencing of body-wave arrival times is a useful technique, but it can be computationally challenging.
- An approach based on demeaning retains the advantages of double differencing while making it more tractable.
- The demeaning approach extends the range of the application of differential travel-time analysis significantly.

BACKGROUND

Body-wave arrival times are a nonlinear function of hypocenter coordinates and wavespeeds. A standard iterative approach to determining these variables from observations of arrival times involves creating a set of linearized equations by expanding the arrival-time function in a Taylor series and retaining only the first-order terms. Given a set of N events recorded by J stations (and, where convenient, we define a "station" as a particular station-phase combination) we have for a single observation of event i recorded by station j:

$$t_{oij} = t_{cij} + \sum_{l=1}^{4} \frac{\partial t_{cij}}{\partial h_{il}} \Delta h_{il} + \sum_{k=1}^{K} \frac{\partial t_{cij}}{\partial s_k} \Delta s_k, \qquad (1)$$

in which t_{oij} is the observed arrival time, t_{cij} is the arrival time calculated in the current model (hypocenter location and discretized wavespeeds), h_{il} is the lth parameter (x, y, z, t) of hypocenter i, and s_k is the wavespeed (or slowness) of element k in a discretized model containing K elements.

1. Department of Earth and Environmental Sciences, Rensselaer Polytechnic Institute, Troy, New York, U.S.A., https://orcid.org/0000-0002-6231-5812 (SR);
2. Departamento de Geofisica, Advanced Mining Technology Center, Facultad de Ciencias Fisicas y Mátematicas, Universidad de Chile, Santiago, Chile

*Corresponding author: roecks@rpi.edu

Cite this article as Roecker, S., A. Maharaj, S. Meyers, and D. Comte (2021). Double Differencing by Demeaning: Applications to Hypocenter Location and Wavespeed Tomography, *Bull. Seismol. Soc. Am.* **XX**, 1–14, doi: 10.1785/0120210007

© Seismological Society of America

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

The residual r_{ij} is defined as

$$r_{ij} = t_{oij} - t_{cij} = \sum_{l=1}^{4} \frac{\partial t_{cij}}{\partial h_{il}} \Delta h_{il} + \sum_{k=1}^{K} \frac{\partial t_{cij}}{\partial s_k} \Delta s_k. \quad (2)$$

We accumulate these equations as rows (one per observation) in matrix form as follows:

$$\mathbf{G}\Delta\mathbf{x} = \mathbf{R}, \quad (3)$$

in which $\mathbf{G}$ holds the partial derivatives, $\Delta\mathbf{x}$ is the perturbations to the current model, and $\mathbf{R}$ is the residuals. Typically, the perturbations are recovered through some type of regularized least squares estimate, for example:

$$\Delta\mathbf{x} = (\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathbf{dd}}^{-1}\mathbf{G} + \mathbf{C}_{\mathbf{pp}}^{-1})^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathbf{dd}}^{-1}\mathbf{R}, \quad (4)$$

in which $\mathbf{C}_{\mathbf{dd}}^{-1}$ is the inverse data covariance matrix and $\mathbf{C}_{\mathbf{pp}}^{-1}$ is the inverse of the prior model covariance matrix (e.g., Tarantola and Vallette, 1982). If the uncertainty in an observation i is δ_i and the uncertainties for different observations are not correlated, then $\mathbf{C}_{\mathbf{dd}}^{-1} = w_i^2\mathbf{I}$, in which the weight $w_i = \delta_i^{-1}$. Then, we write equation (3) as follows:

$$\hat{\mathbf{G}}\Delta\mathbf{x} = \mathbf{W}\mathbf{G}\Delta\mathbf{x} = \mathbf{W}\mathbf{R} = \hat{\mathbf{R}}, \quad (5)$$

in which $\mathbf{W}$ is a diagonal matrix with elements $W_{ii} = w_i$. Equation (4) then becomes

$$\Delta\mathbf{x} = (\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{G}} + \mathbf{C}_{\mathbf{pp}}^{-1})^{-1}\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{R}}. \quad (6)$$

There are several variations on this theme, but nearly all of them involve the products $\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{G}}$ and $\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{R}}$. There are also iterative ways to achieve this result without actually having to generate $\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{G}}$ and $\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{R}}$ explicitly (e.g., the least-squares conjugate gradient method [LSQR] of Paige and Saunders, 1982), but clearly any approach that involves the same entries for $\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{G}}$ and $\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{R}}$ will generate the same estimate $\Delta\mathbf{x}$.

Instead of simply inverting the set of equations as they appear earlier (as is often done in standard local earthquake tomography), one may consider modifying $\hat{\mathbf{G}}$ by performing some kind of differencing of its rows and the associated elements of $\hat{\mathbf{R}}$. Approaches based on differencing residuals have been used extensively in both hypocenter location (e.g., Douglas, 1967; Evernden, 1969; Frohlich, 1979; Pavlis and Booker, 1983) and wavespeed tomography (e.g., Fitch, 1975; Aki *et al.*, 1977; Roecker, 1985). There are two primary advantages to differencing: first, the difference in time between arrivals from two different events recorded by the same station often can be determined more precisely (e.g., by cross correlation of the waveform data) than absolute onset times, and differencing the rows of $\hat{\mathbf{G}}$ can take direct advantage of these more precise observations. Second, differencing reduces the effects of unknown, unmodeled, or otherwise unconstrained variables on the arrival times, at least to the extent that those effects are common to the observations in question. Because cross correlation of waveform data is most effective when nearby events are recorded at the same station, these two advantages often are realized simultaneously.

Double differencing

Double differencing (Waldhauser and Ellsworth, 2000; Zhang and Thurber, 2003, 2006) combines the linearized equations of equation (5) in a strategic way. Starting with the residual r_{ij} of equation (2), referred to in differencing parlance as a "single difference," we consider a second observation from event n recorded by the same station:

$$r_{nj} = t_{onj} - t_{cnj} = \sum_{l=1}^{4} \frac{\partial t_{cnj}}{\partial h_{nl}} \Delta h_{nl} + \sum_{k=1}^{K} \frac{\partial t_{cnj}}{\partial s_k} \Delta s_k. \quad (7)$$

The difference between r_{ij} and r_{nj} is called the "double difference":

$$\begin{aligned} r_{inj} &= r_{ij} - r_{nj} = t_{oij} - t_{cij} - (t_{onj} - t_{cnj}) \\ &= \sum_{l=1}^{4} \frac{\partial t_{cij}}{\partial h_{il}} \Delta h_{il} - \sum_{l=1}^{4} \frac{\partial t_{cnj}}{\partial h_{nl}} \Delta h_{nl} + \sum_{k=1}^{K} \frac{\partial t_{cij}}{\partial s_k} \Delta s_k - \sum_{k=1}^{K} \frac{\partial t_{cnj}}{\partial s_k} \Delta s_k. \end{aligned} \quad (8)$$

Typically, all combinations of observations from a group of events p recorded by station j are differenced in this way. This operation is repeated for all stations in a network and then, as desired, for different (and usually overlapping) groups of events. An equivalent way of describing this procedure is differencing rows of $\hat{\mathbf{G}}$ and the corresponding elements of $\hat{\mathbf{R}}$ in equation (5) to form expanded versions of $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$. This expansion can increase the size of the original $\hat{\mathbf{G}}$ substantially. Generally, one would not include redundant differences (i.e., both the differences $r_{ij} - r_{nj}$ and $r_{nj} - r_{ij}$), so if N events are recorded by station j, the number of equations created by double differencing for this subset of observations increases from N to $N(N-1)/2$, and the number of nonzero elements for the hypocenter partition of $\hat{\mathbf{G}}$ increases from $4N$ to $4N(N-1)$. The wavespeed partition of $\hat{\mathbf{G}}$ increases proportionally to the number of model parameters as well: for an average number of wavespeed variables K accumulated by each observation, the number of nonzeros increases from approximately KN to $KN(N-1)/2$. Because K can be a large number in a 3D model, double differencing can make wavespeed tomography computationally challenging.

Demeaning

We define demeaning as collecting the same set of observations as earlier for a group of N events recorded by the same station j and in some way determining a mean residual, for example:

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

$$\bar{r}_j = \frac{1}{N}\sum_{n=1}^{N} r_{nj} = \frac{1}{N}\sum_{n=1}^{N}(t_{onj} - t_{cnj})$$
$$= \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{l=1}^{4}\frac{\partial t_{cnj}}{\partial h_{nl}}\Delta h_{nl} + \sum_{k=1}^{K}\frac{\partial t_{cnj}}{\partial s_k}\Delta s_k\right), \quad (9)$$

and subtracting this quantity from each row:

$$r_{ij} - \bar{r}_j = \sum_{l=1}^{4}\left(\frac{\partial t_{cij}}{\partial h_{il}}\Delta h_{il} - \frac{1}{N}\sum_{n=1}^{N}\frac{\partial t_{cnj}}{\partial h_{nl}}\Delta h_{nl}\right)$$
$$+ \sum_{k=1}^{K}\left(\frac{\partial t_{cij}}{\partial s_k} - \frac{1}{N}\sum_{n=1}^{N}\frac{\partial t_{cnj}}{\partial s_k}\right)\Delta s_k. \quad (10)$$

Demeaning of residuals by means of station corrections has been shown to be effective in locating hypocenters (e.g., Frohlich, 1979; Richards-Dinger and Shearer, 2000; Lin and Shearer, 2005). Demeaning is also often employed in teleseismic tomography, following the seminal work of Aki *et al.* (1977) who used it as a way to reduce the effects of unresolvable wavespeed and hypocenter variations. Similar to double differencing, demeaning can be envisioned as manipulating rows of $\hat{\mathbf{G}}$, but unlike double differencing, demeaning does not add any additional rows to $\hat{\mathbf{G}}$. This can make a significant difference for wavespeed tomography because that partition of $\hat{\mathbf{G}}$ will have not many more than the original KN nonzeros in $\hat{\mathbf{G}}$. Demeaning increases the number of nonzero elements in the hypocenter partition from $4N$ to $4N^2$, essentially adding a row of nonzeros to the double-difference version for each station-group. Although one might therefore infer a slight advantage of double differencing over demeaning for hypocenter-only applications, tests of these algorithms (discussed subsequently) suggest otherwise. In any event, the significant advantage in wavespeed tomography means that joint hypocenter-tomography applications can be more efficiently done by demeaning.

Although more efficient, it is not immediately clear what negative effects demeaning might have on the quality of the solution obtained. For example, one may suspect that the averaging step in demeaning somehow reduces the amount of information in the original observations and, moreover, that the ability to exploit precise relative times has vanished. In what follows, we show that demeaning can be done in a way that preserves the advantages of double differencing.

COMPARISON OF DEMEANING AND DOUBLE-DIFFERENCING LEAST-SQUARES ESTIMATES

Let $\mathbf{D}$ and $\mathbf{M}$ be the double differenced and demeaned expansions, respectively, of $\hat{\mathbf{G}}$, and $\mathbf{R_D}$ and $\mathbf{R_M}$ be the double differenced and demeaned expansions of $\hat{\mathbf{R}}$. We also define a matrix $\mathbf{H}$that augments $\mathbf{D}$ by appending the $N(N-1)$ additional rows of $-\mathbf{D}$ (these would be the redundant rows ignored in constructing $\mathbf{D}$) and an additional N rows of zeros that represent rows of $\hat{\mathbf{G}}$ differenced with themselves:

$$\mathbf{H} = \begin{bmatrix} \mathbf{D} \\ -\mathbf{D} \\ 0 \end{bmatrix}. \quad (11)$$

The corresponding residual vector $\mathbf{Q}$ would be

$$\mathbf{Q} = \begin{bmatrix} \mathbf{R_D} \\ -\mathbf{R_D} \\ 0 \end{bmatrix}. \quad (12)$$

Note that

$$\mathbf{H^T H} = [\mathbf{D^T} \quad -\mathbf{D^T} \quad 0]\begin{bmatrix} \mathbf{D} \\ -\mathbf{D} \\ 0 \end{bmatrix} = 2\mathbf{D^T D}, \quad (13)$$

and

$$\mathbf{H^T Q} = [\mathbf{D^T} \quad -\mathbf{D^T} \quad 0]\begin{bmatrix} \mathbf{R_D} \\ -\mathbf{R_D} \\ 0 \end{bmatrix} = 2\mathbf{D^T R_D}. \quad (14)$$

Hence the least-squares solution $\Delta\mathbf{x}$ to $\mathbf{D}\Delta\mathbf{x} = \mathbf{R_D}$ is the same as that to $\mathbf{H}\Delta\mathbf{x} = \mathbf{Q}$.

We order the rows of $\mathbf{H}$ and $\mathbf{Q}$ in a way that facilitates comparison between double differencing and demeaning. A general dataset will consist of arrival times, or differences in arrival times, from P groups of events recorded by a network of J stations. Observations from any event may appear in more than one group, the purpose usually being to couple them together. We order the observations first by these individual groups p and then by each station j that reports the arrivals from a given group. This way of ordering, and the designation of a group, follows naturally from how one would normally choose which arrivals to difference. Generally, arrivals are differenced if the events are close enough together so that either (1) their waveforms may be unambiguously correlated or (2) their ray paths are sufficiently similar that differencing will mitigate the effects of structure outside a region of interest near the events. Events associated with arrivals at a given station that satisfy one or both of these criteria would in practice be assigned to the same group. Let $\mathbf{H_{pj}}$ contain the N_{pj} rows of $\mathbf{H}$ that correspond to the observations from group p recorded by station j. We then write

$$\mathbf{H} = \begin{bmatrix} \mathbf{H_{11}} \\ \vdots \\ \mathbf{H_{1J}} \\ \vdots \\ \mathbf{H_{P1}} \\ \vdots \\ \mathbf{H_{PJ}} \end{bmatrix}. \quad (15)$$

The elements of $\mathbf{Q}$ are similarly ordered. Each $\mathbf{H_{pj}}$ will contain those parts of $-\mathbf{D}$ and the row of zeros that pertain to that

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

particular station-group. The products $\mathbf{H^TH}$ and $\mathbf{H^TQ}$ are the sums of the products of these subdivisions, that is:

$$\mathbf{H^TH} = \sum_{p=1}^{P}\sum_{j=1}^{J}\mathbf{H_{pj}^TH_{pj}}, \tag{16}$$

$$\mathbf{H^TQ} = \sum_{p=1}^{P}\sum_{j=1}^{J}\mathbf{H_{pj}^TQ_{pj}}. \tag{17}$$

Organizing $\mathbf{H}$ and $\mathbf{Q}$ in this way simplifies our analysis by reducing it to the case of a single station-group.

Our objective is to construct a "demeaning matrix" $\mathbf{M}$ that will provide the same least-squares estimate as $\mathbf{H}$ and $\mathbf{D}$. As discussed in detail in the Appendix, we can do so by estimating the means of the N_{pj} rows of the $\mathbf{H_{pj}}$ submatrices and the corresponding elements of the $\mathbf{Q_{pj}}$ vector and subtracting their respective elements from them. The resulting $\mathbf{M_{pj}}$ submatrices and $\mathbf{R_{Mpj}}$vector are assembled into a general $\mathbf{M}$ matrix by scaling the ith row of $\mathbf{M_{pj}}$ and ith element of $\mathbf{R_{Mpj}}$ by $\sum_{n=1}^{N_{pj}} w_{inpj}/\sqrt{N_{pj}}$, in which w_{inpj} is a weight based on the uncertainties in the observations of events i and n of group p observed at station j. Making use of Einstein notation as discussed in the Appendix, the general form for the (i, k) element of the demeaning matrix $\mathbf{M}$ in the wavespeed-only case is

$$(\mathbf{M_{pj}})_{ik} = \frac{[w_{inpj}]}{\sqrt{N_{pj}}}\left(\frac{\partial t_{cipj}}{\partial s_k} - \frac{[w_{inpj}\frac{\partial t_{cnpj}}{\partial s_k}]}{[w_{inpj}]}\right), \tag{18}$$

and the corresponding ith element of the demeaned residual is

$$(\mathbf{R_{Mpj}})_i = \frac{[w_{inpj}]}{\sqrt{N_{pj}}}\left(r_{ipj} - \frac{[w_{inpj}r_{npj}]}{[w_{inpj}]}\right), \tag{19}$$

in which square brackets indicate a summation over index n from 1 to N_{pj}. In the case in which weights are identical for a given station-group, that is, $w_{inpj} = \gamma_{pj}$, the appropriate scaling factor is $\sqrt{N_{pj}}\gamma_{pj}$ and the correspondence of least-squares solutions is exact, that is

$$\Delta\mathbf{x} = (\mathbf{D^TD})^{-1}\mathbf{D^TR_D} = (\mathbf{M^TM})^{-1}\mathbf{M^TR_M}. \tag{20}$$

The only structural difference between the wavespeed-only and hypocenter-only systems of equations is that the latter is significantly sparser (a maximum of eight nonzeros per row of $\mathbf{H}$). Moreover, from equations (A17) and (A44), we see that the rows of $\mathbf{M}$ and the data vector $\mathbf{R_M}$ are scaled the same way for both of these cases, so all that is required to generate the joint hypocenter-wavespeed matrix is to append the hypocenter columns to the wavespeed columns of $\mathbf{M}$. Hence, the same equivalences between double differencing and demeaning can be obtained for the hypocenter-only and joint hypocenter-wavespeed matrices by applying the same station-group scaling.

USING DIFFERENTIAL ARRIVAL TIMES AS DEMEANED OBSERVATIONS

As noted earlier, an advantage of double differencing is that it can directly exploit the precision of correlated relative times rather than having to rely on less precise absolute times. This is also the case for the demeaning approach. From equation (8), if we sum a set of $(N - 1)$ differenced times from a set of N observations (in this case with respect to the first observation r_1):

$$\begin{bmatrix} w_{12}(r_1 - r_2) \\ w_{13}(r_1 - r_3) \\ \vdots \\ w_{1N}(r_1 - r_N) \end{bmatrix} = \begin{bmatrix} w_{12}(t_{o1} - t_{o2}) \\ w_{13}(t_{o1} - t_{o3}) \\ \vdots \\ w_{1N}(t_{o1} - t_{oN}) \end{bmatrix} - \begin{bmatrix} w_{12}(t_{c1} - t_{c2}) \\ w_{13}(t_{c1} - t_{c3}) \\ \vdots \\ w_{1N}(t_{c1} - t_{cN}) \end{bmatrix}, \tag{21}$$

we obtain for any r_i:

$$r_i\sum_{n=1}^{N,n\neq i} w_{in} - \sum_{n=1}^{N,n\neq i} w_{in}r_n = \left(t_{oi}\sum_{n=1}^{N,n\neq i} w_{in} - \sum_{n=1}^{N,n\neq i} w_{in}t_{on}\right) - \left(t_{ci}\sum_{n=1}^{N,n\neq i} w_{in} - \sum_{n=1}^{N,n\neq i} w_{in}t_{cn}\right). \tag{22}$$

Factoring out the sum of the weights, and generalizing equation (A40), the left side is written as

$$\sum_{n=1}^{N,n\neq i} w_{in}\left(r_i - \frac{\sum_{n=1}^{N,n\neq i} w_{in}r_n}{\sum_{n=1}^{N,n\neq i} w_{in}}\right) = [w_{\mathbf{in}}]\left(r_i - \frac{[w_{\mathbf{in}}r_n]}{[w_{\mathbf{n}}]}\right). \tag{23}$$

Hence, we can construct a dataset for the demeaning application that makes direct use of cross-correlated differential times by calculating deviations from the mean value of the differenced residuals. Note that from equations (22) and (23) we have the expression for observed arrival times:

$$[w_{\mathbf{in}}]\left(t_{oi} - \frac{[w_{\mathbf{in}}t_{on}]}{[w_{\mathbf{in}}]}\right), \tag{24}$$

in which the term in the parentheses is essentially a deviation from a weighted stack. This is a common way to derive observations for demeaning applications, for example, when cross-correlated waveforms are used in teleseismic tomography (e.g., Van Decar and Crosson, 1990).

LEAST-SQUARES FORMS

Many tomography or joint location problems of seismological interest involve large numbers of variables, and for reasons of tractability we often resort to iterative techniques such as LSQR

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

operating on equation (5) without actually calculating the normal-equation products in equation (6). Nevertheless, for inversions with small numbers of variables, it can be advantageous to form and solve equation (6) directly. In this case, it can be more efficient computationally to form $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ and $\mathbf{H}^{\mathrm{T}}\mathbf{Q}$ as observations are accumulated as opposed to storing all of $\mathbf{H}$, $\mathbf{D}$, or $\mathbf{M}$ and their data vectors and then multiplying them together. For convenience, these forms are summarized in the Appendix.

TESTS OF EFFICIENCY AND THE EFFECTS OF WEIGHTING

Although the equivalence of double differencing and demeaning can be demonstrated in cases in which weighting exhibits certain regularities for groups of observations, the effects of observation-specific weighting are not obvious. As a test of these effects, we constructed examples that incorporate uncertainties from an actual arrival-time dataset and compare the results using the two approaches.

To mimic the density of activity often employed in differencing studies, we model our example on a dataset consisting of 274 events from a sequence of aftershocks to the 20 January 2019 M_{w} 6.7 Coquimbo-La Serena earthquake (e.g., Comte, Farías, Navarro-Aránguiz, *et al.*, 2019; Ruiz *et al.*, 2019) recorded by a temporary network of 33 stations in Chile (Fig. 1). Both *P* and *S* arrivals are used and are coupled by solving for both V_P and the V_P/V_S ratio (i.e., *S* arrivals will have sensitivities to both V_S and V_P; see Roecker *et al.* (2006) for a discussion of how these sensitives are calculated). The synthetic dataset is generated by calculating times in a checkerboard-like medium using the spherically based eikonal equation solver of Li *et al.* (2009) with the hypocenters located in one of the perturbed blocks (Fig. 1). These times are contaminated with random (Gaussian) noise and given the same uncertainties that were assigned in the original dataset using the regressive estimation autopicking package of Comte, Farias, Roecker, *et al.* (2019) after manual calibration (Fig. 2). The hypocenters are then associated with groups based on their distances from any centroid on a regularly spaced (5 km) grid (Fig. 1). Any hypocenter less than 4.5 km from a centroid is associated with that group, meaning that the maximum distance between events in any particular group is 9 km. This resulted in 30 groups of between 3 and 103 events. Hypocenters can be in multiple groups to couple them together; in this case, the maximum number of groups for any single hypocenter was six. The systems of equations for $\mathbf{M}$, $\mathbf{R_M}$, $\mathbf{D}$, and $\mathbf{R_D}$ were then created by algorithms that use the forms described earlier for the "grouped" dataset, and solved using the parallel least squares algorithm (PLSQR3) of Lee *et al.* (2013).

To assess the dependence of a hypocenter-only inversion on observation-specific weights, we first relocated the events in the 1D background model (preliminary reference Earth model; Dziewonski and Anderson, 1981) to displace starting locations from their true values by a few kilometers. Subsequent relocations of these events by demeaning differ only slightly from those derived from double differencing (Fig. 3); all but two locations were different by less than 100 m (one event was different by 150 m), and 86% were within 40 m of each other. Formal estimates of hypocenter uncertainty are on the order of 100 m; hence, the differences are within the error bars. Remarkably, despite the demeaning matrix being about 50% larger than that for double differencing, PLSQR3 required about the same amount of time to execute the same number of iterative solves when less than four processors were used, with demeaning becoming slightly more efficient as the number of processors increased. Because this was an unexpected result, we repeated this test with a synthetic dataset with about three times as many hypocenters (821 events in 51 groups) in the same region. These results (Table 1) showed that PLSQR3 was significantly, and increasingly, faster with demeaning than double differencing as more processors were used. The probable cause of this difference is the number of vector products performed, which is most likely a result of the double-differenced residual vector containing about 50 times as many elements.

For the wavespeed tomography test, because the objective is to evaluate the effects of weighting on an image as opposed to demonstrating anything about the quality of the image itself, we simplify the analysis by allowing only wavespeeds to vary. Because the uncertainties for *S* arrival times are significantly larger and more variable than those for *P* (Fig. 2), we focus on the results for V_S (the comparisons for V_P are nearly identical). The same regularizations (damping and smoothing) were applied to both cases, and both were iterated four times. As expected for this type of analysis, the sensitivities to wavespeeds are minor in most parts of the model outside the vicinity of the earthquakes. Comparative contours of results in this region (Fig. 4) show them to be very similar, with a maximum difference between models at any grid point of 0.01 km/s in V_S. Other potential sources of solution discrepancy, such as the criteria used by PLSQR3 to terminate the iterative solution to the system of linear equations and the number of calculations done with single precision real numbers, are, by extension, also inconsequential. This is despite the double-differencing kernel having an order of magnitude more nonzeros than the demeaning kernel (317×10^6 vs. 27×10^6), a difference which is largely responsible for the double-differencing solution requiring about 10 times more wall time than demeaning.

This similarity of results is perhaps not surprising; because double differencing and demeaning both involve linear manipulations of the same set of weighted linear equations, one might suspect that the solutions one obtains from these approaches is, if not exactly the same, at least close enough so that they give essentially the same result when used to iteratively solve a nonlinear inverse problem. Although a single test is not a "proof" of this notion, it does provide useful corroboration. Given the savings in memory and processing time achieved

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

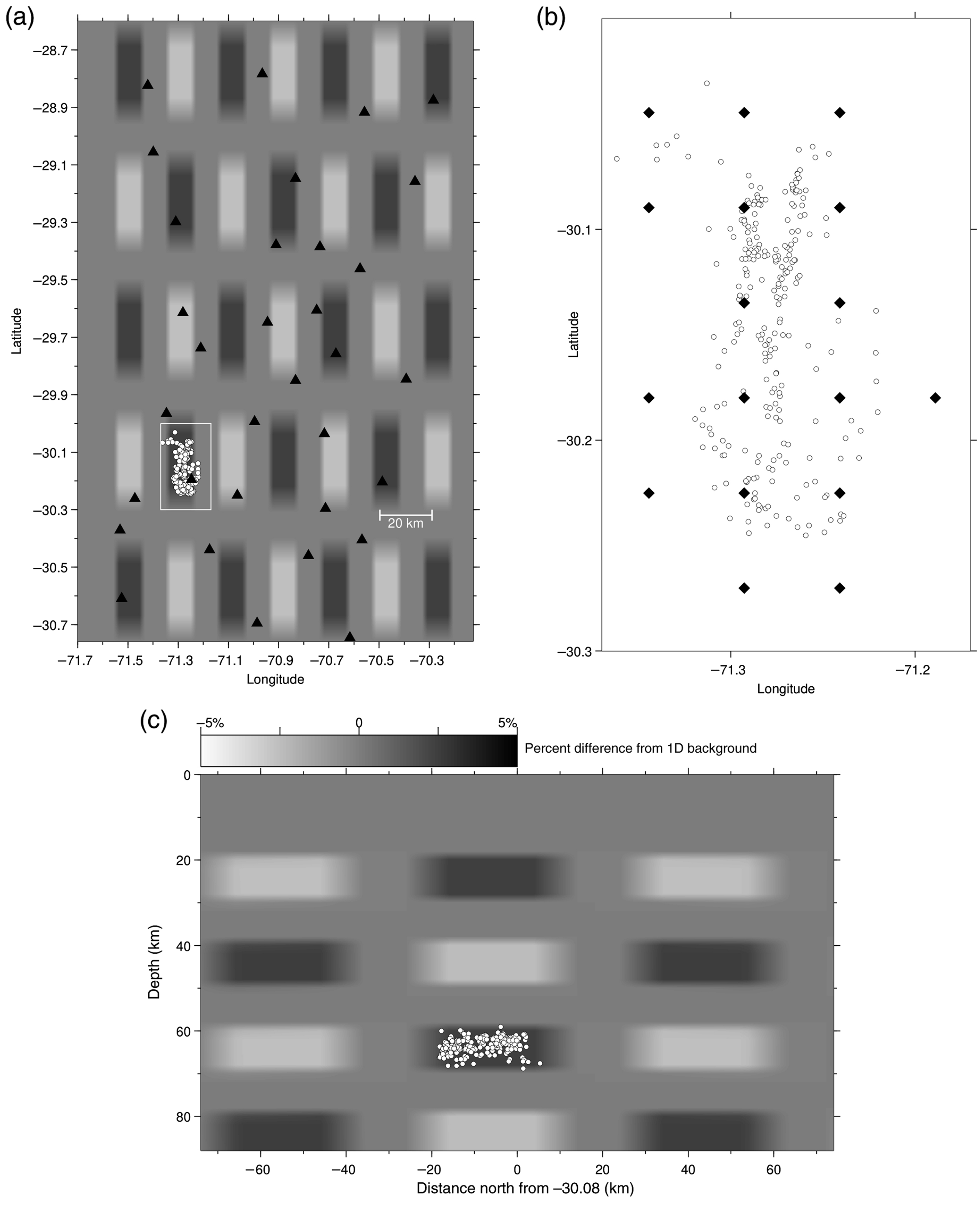

Figure 1. Parameters of tests used to conduct simple comparisons of the double differencing and demeaning techniques. (a) Map view at 64 km depth. Black triangles locate 33 seismic stations. White circles locate the 274 hypocenters of the wavespeed test; the 821 events of the larger hypocenter test are within the same volume. Wavespeed variations are indicated by shades from −5% (white) to +5% (black) relative to a 1D background as shown in the palette at the upper left of the figure. The white rectangle locates the region shown in (b). (b) Enlargement of region shown in the white rectangle in (a), showing epicenters as open circles and locations of centroids used to group events. Centroids are spaced on a grid 5 km apart in each direction starting at 45 km depth. Events within 4.5 km of a centroid are designated as a member of that group. Note that events can be members of more than one group. (c) North–south cross section at 71.3° W of the region near the hypocenters. South is to the left.

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

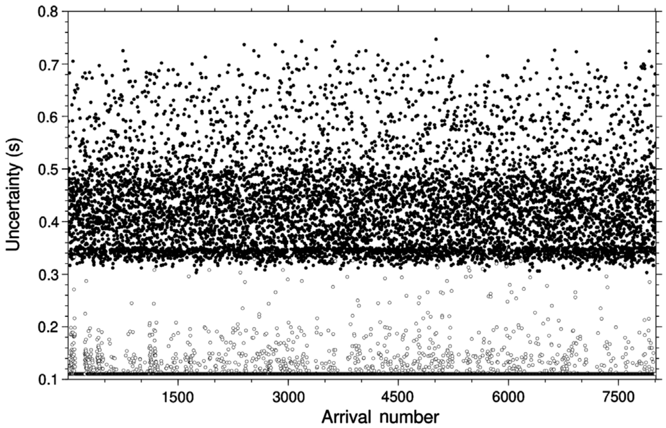

Figure 2. Scatter plot of uncertainties in *P*- (open circles) and *S*- (closed circles) arrival times used to test the effects of observation-specific uncertainties on the least-squares estimates. The *X* axis is a sequential arrival number as it occurs in the dataset. The line of open circles near 0.1 s reflects the minimum uncertainty assigned to impulsive *P* arrivals.

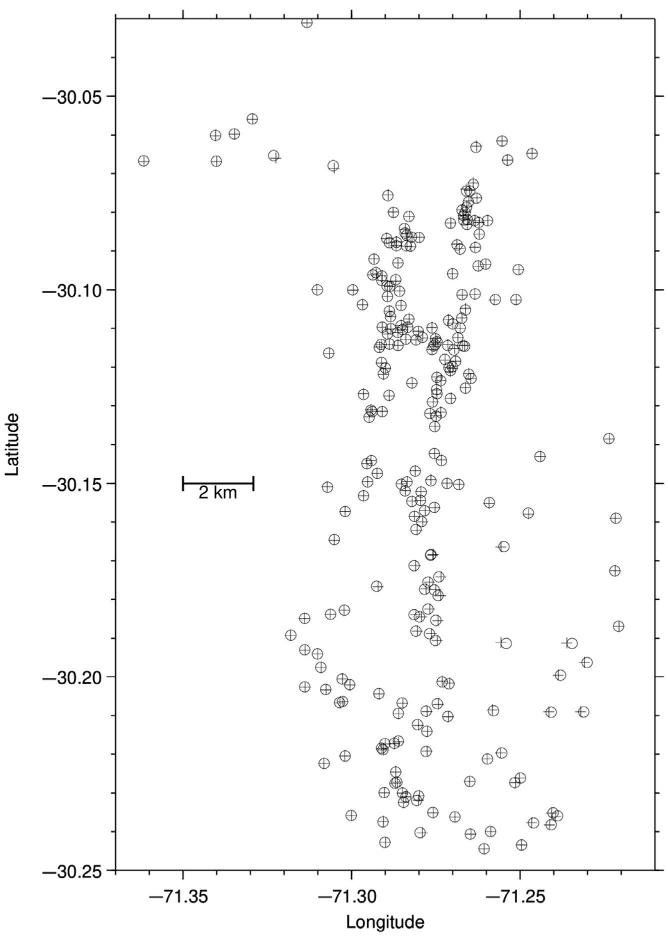

Figure 3. Plot of epicenters obtained by demeaning (crosses) and double differencing (open circles). For scale, the radius of the circles is about 100 m. Note that, with the exception of a few events in the north (near 30.06° S) and southeast (near 30.20° S, 71.25° W), the epicenters are nearly identical.

by demeaning, these slight differences in results would in most cases be considered inconsequential.

CONCLUSIONS

The many successful applications of double differencing over the past several years in both earthquake relocation and wavespeed tomography testify to its advantages. The alternate approach of demeaning offers the same advantages at significantly less computational cost in terms of both memory and time required to solve the system of linear equations at any iteration of a nonlinear inversion. This is especially true for tomographic applications. The main question addressed in this article is the potential for some other hidden cost, such as loss of information, in the demeaning approach that might lead to an inferior least-squares estimate.

The results discussed earlier show that, in the case in which the uncertainties in the observations from a group of events at a given station are the same, the demeaning estimate and the double-difference estimate are identical, provided that the rows of the demeaning matrix corresponding to a given station-group are properly weighted by $\sqrt{N}\gamma$, in which N is the number of observations in the station-group and γ is the inverse of the uncertainty. When these uncertainties are not the same, these rows should be weighted by $[w_{\boldsymbol{n}}]/\sqrt{N}$, in which w_n is the reciprocal of the uncertainty of the nth observation in the station-group. Although the breakdown in symmetry caused by observation-specific uncertainty means that the two estimates are no longer identical, this scaling brings them closer into alignment. Spatially clustered events recorded by the same station will have similar ray paths, so estimates of uncertainty for arrivals from a given station-group may be expected to be similar. But even if they are not, tests suggest that the resulting differences in least-squares estimates are minor. Moreover,

TABLE 1
Comparison of Results of the Hypocenter-Only Inversions Using the Larger Hypocenter Dataset (821 Events, 51 Groups)

Processing Parameter	Double Difference	Demeaning	Ratio (DM/DD)
Number of columns	3,276	3,276	100%
Number of rows	6,943,638	148,274	2%
Number of nonzeros	55,552,348	83,074,752	150%
One processor	8.872	5.150	58%
Four processors	5.438	2.765	51%
Eight processors	4.091	1.501	37%
12 processors	3.808	0.819	22%

First three rows summarize the sizes of the hypocenter-only kernel, and the last four rows show the amount of time (in seconds) required to execute 10 iterations of PLSQR3 as a function of the number of parallel processes used on a Mac Pro desktop with a 24 Core 2.7 GHz Intel Xeon processor. The last column shows the percent ratio of demeaning (DM) to double-differencing (DD) results. Note that, despite having approximately 50% more nonzeros in the hypocenter kernel, iterations on the demeaning solution required significantly less time.

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

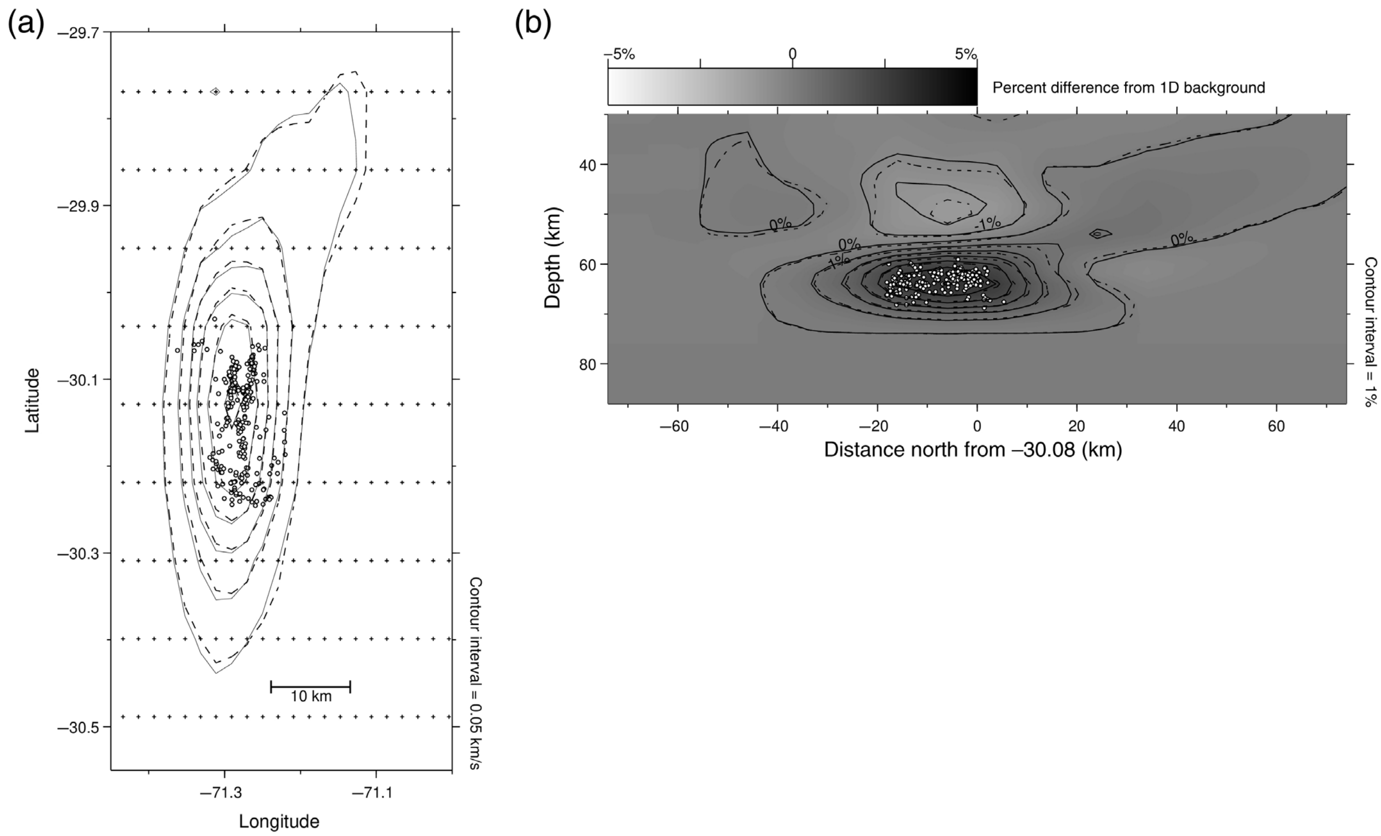

Figure 4. Contour plots of wavespeed images for V_S generated by double differencing (solid contours) and demeaning (dashed contours) in the vicinity of the hypocenters. (a) Map view at 64 km depth. Contour interval is 0.05 km/s with the outermost contour encompassing an area with wavespeeds 0.05 km/s higher than the background of 4.53 km/s. Crosses are grid points in which wavespeeds are specified. (b) North–south cross section taken at 71.3° W and centered at 30.08° S. South is to the left. Contours are percent difference in wavespeed from the 1D background model at an interval of 1%. Shading reflects percent change from the 1D background as indicated in the palette at the upper left of the figure.

because we generally treat these estimates as a single step in an iterative process, small deviations in a least-squares estimate at any particular iteration generally will be inconsequential.

This definition of a weighted mean (e.g., as used in equation 19) differs from the standard form used in least squares in that the weights correspond to the inverse of the uncertainties rather than the variances for the corresponding observations. As shown in equations (21–23), this choice allows for direct use of higher precision cross correlated arrival times in a demeaning approach, and equation (24) suggests that a similar type of weighted demeaning could be advantageously applied when deriving observations from cross-correlated waveforms in teleseismic tomography.

DATA AND RESOURCES

The synthetic data generated for the tests discussed earlier are available from the corresponding author upon request.

DECLARATION OF COMPETING INTERESTS

The authors acknowledge that there are no conflicts of interest recorded.

ACKNOWLEDGMENTS

The idea for this work was motivated by conversations with Keith Priestley. Ben Baker, YoungHee Kim, and an anonymous reviewer provided many useful comments. The authors appreciate their contributions. The data used as a template for the tests were based on observations provided by the Mobile Seismic Array in Central Chile (MSACC-KIT). Portions of this research were supported by the National Science Foundation (NSF) Grant Number EAR-2021040 and the Programa de Investigación Asociativa of the Chilean Comisión Nacional de Investigación Científica y Tecnológica (CONICYT/PIA) Project AFB180004.

REFERENCES

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

Comte, D., M Farias, S Roecker, and R Russo (2019). The nature of the subduction wedge in an erosive margin: Insights from the analysis of aftershocks of the 2015 Mw 8.3 Illapel earthquake beneath the Chilean Coastal Range, *Earth Planet. Sci. Lett.* **520,** 50–62.

Comte, D., M. Farías, A. P. Navarro-Aránguiz, S. W. Roecker, and A. Rietbrock (2019). *Analysis of the Mw 6.7 Coquimbo-La Serena Earthquake of June 14th, 2019: Preliminary Results Determined with a High Density Local Seismic Network*, American Geophysical Union, San Francisco, California, Fall Meeting 2019, abstract #T51F-0338.

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

Douglas, A. (1967). Joint epicentre determination, *Nature*, **215,** 47–48.

Dziewonski, A. M., and D. L. Anderson (1981). Preliminary reference Earth model, *Phys. Earth Planet. In.* **25,** 297–356.

Evernden, J. (1969). Precision of epicenters obtained by small networks of world-wide standard stations, *Bull. Seismol. Soc. Am.* **59,** 3, 1365–1368.

Fitch, T. J., 1975. Compressional velocity in source regions of deep earthquakes: An application of the master earthquake technique, *Earth Planet. Sci. Lett.* **26,** 156–166.

Frohlich, C. (1979). An efficient method for joint hypocenter determination for large groups of earthquakes, *Comput. Geosci.* **5,** 387–389.

Lee, E.-J., H. Huang, J. M. Dennis, P. Chen, and L. Wang (2013). An optimized parallel LSQR algorithm for seismic tomography, *Comput. Geosci.* **61,** 184–197.

Li, Z., S. Roecker, B. Wei, H. Wang, G. Schelochkov, and V. Bragin (2009). Tomographic image of the crust and upper mantle beneath the western Tien Shan from the MANAS broadband deployment: Possible evidence for lithospheric delamination, *Tectonophysics* **477**, nos. 1/2, 49–57.

Lin, G., and P. Shearer (2005). Tests of relative earthquake location techniques using synthetic data, *J. Geophys. Res.* **110,** no. B04304, doi: 10.1029/2004JB003380.

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

Pavlis, G. L., and J. R. Booker (1983). Progressive multiple event location (PMEL), *Bull. Seismol. Soc. Am.* **73,** 1753–1777.

Richards-Dinger, K., and P. Shearer (2000). Earthquake locations in southern California obtained using source-specific station terms, *J. Geophys. Res.* **105,** 10,939–10,960.

Roecker, S. W. (1985). Velocity structure in the Izu-Bonin seismic zone and the depth of the olivine-spinel phase transition in the slab, *J. Geophys. Res.* **90**, no. B9, 7771–7794, doi: 10.1029/JB090iB09p07771.

Roecker, S., C. Thurber, K. Roberts, and L. Powell (2006). Refining the image of the San Andreas Fault near Parkfield, California using a finite difference travel time computation technique, *Tectonophysics* **426,** nos. 1/2, 189–205.

Ruiz, S., J.-B. Ammirati, F. Leyton, L. Cabrera, B. Potin, and R. Madariaga (2019). The January 2019 (Mw 6.7) Coquimbo earthquake: Insights from a seismic sequence within the Nazca Plate, *Seismol. Res. Lett.* **90**, no. 5, 1836–1843.

Tarantola, A., and B. Valette (1982). Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys. Space Phys.* **20,** 219–232

Van Decar, J. C., and R. S. Crosson (1990). Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least-squares, *Bull. Seismol. Soc. Am.* **80,** 150–169.

Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward Fault, California, *Bull. Seismol. Soc. Am.* **90,** 1353–1368.

Zhang, H., and C. H. Thurber (2003). Double-difference tomography: The method and its application to the Hayward fault, California, *Bull. Seismol. Soc. Am.* **93,** 1875–1889.

Zhang, H., and C. H. Thurber (2006). Development and applications of double-difference seismic tomography, *Pure Appl. Geophys.* **163,** 373–403, doi: 10.1007/s00024-005-0021-y.

APPENDIX

In this appendix, we provide a general and detailed derivation of a "demeaning matrix" **M** that can provide the same least-squares estimate as **H** and **D** for the cases of wavespeed-only, hypocenter-only, and joint hypocenter-wavespeed. Because the hypocenter-only matrix is a special case of the wavespeed-only matrix, we begin by comparing the two approaches for the case of wavespeeds-only. Because the scaling and data vectors are the same for the hypocenter-only and wavespeed-only cases, the joint hypocenter-wavespeed matrix can be constructed from a straightforward combination of the two.

A note on notation

Because these derivations involve a significant number of summations, for ease of presentation we use Einstein notation whenever possible. Because the equations also involve multiple dummy indices, for clarity we enclose terms in which a summation over a certain index required (or at least contributes to the clarity of presentation) in square brackets. Unless otherwise stated, implicit sums are from 1 to N, in which N is the number of arrivals in a specific station group. In addition, where more than one dummy variable appears in square brackets, the one over which the sum is taken appears in bold. For example,

$$M_l = \sum_{i=1}^{N} C_{il} \sum_{n=1}^{N} w_{in} = C_{il}[w_{i\mathbf{n}}].$$

Wavespeed tomography

We begin by comparing demeaning and double differencing for the case of wavespeed tomography. Because the location of nonzero elements in the wavespeed partition is arbitrary, hypocenters and joint hypocenter-wavespeed tomography can be viewed as special cases. From equations (5) and (8), a weighted row of **D** corresponding to a difference in observations between events i and n from a given group p recorded at station j is

$$w_{inpj}\left(\sum_{k=1}^{K}\frac{\partial t_{cipj}}{\partial s_k}\Delta s_k - \sum_{k=1}^{K}\frac{\partial t_{cnpj}}{\partial s_k}\Delta s_k\right) = w_{inpj}(r_{ijp} - r_{njp}), \quad \text{(A1)}$$

in which the sum is over all K wavespeed variables in the medium. When differencing absolute arrival times, the weight w_{inpj} would typically be the inverse of the Euclidian norm of the individual uncertainties δ_{ipj} and δ_{npj} associated with residuals r_i and r_{npj}. When differencing relative times, w_{inpj} would be the inverse of the intrinsic uncertainty δ_{injp} associated with the realization of that relative residual. In either case, $w_{inpj} = w_{nipj}$.

As discussed in the Comparison of Demeaning and Double-Differencing Least-Squares Estimates section of the main text, we can restrict our attention to a given station-group pair, so in what follows we will drop the p and j subscripts. To further simplify notation, we set

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

$$C_{ik} = \frac{\partial t_{ci}}{\partial s_k}, \quad \text{(A2)}$$

in which case equation (A1) is written as follows:

$$w_{in}\left(\sum_{k=1}^{K} C_{ik}\Delta s_k - \sum_{k=1}^{K} C_{nk}\Delta s_k\right) = w_{in}\sum_{k=1}^{K}(C_{ik} - C_{nk})\Delta s_k = w_{in}(r_i - r_n). \quad \text{(A3)}$$

The submatrix $\mathbf{H_{pj}}$ will have $N^2 = N_{pj}^2$ rows and K columns, with N rows for each event i. For clarity, the first two blocks of equations of $\mathbf{H_{pj}}\Delta\mathbf{s} = \mathbf{Q_{pj}}$ look like

$$\begin{bmatrix} w_{11}(C_{11} - C_{11}) & w_{11}(C_{12} - C_{12}) & \dots & w_{11}(C_{1k} - C_{1k}) \\ w_{12}(C_{11} - C_{21}) & w_{12}(C_{12} - C_{22}) & \dots & w_{12}(C_{1k} - C_{2k}) \\ \vdots & \vdots & \ddots & \vdots \\ w_{1N}(C_{11} - C_{N1}) & w_{1N}(C_{12} - C_{N2}) & \dots & w_{1N}(C_{1N} - C_{NK}) \\ w_{21}(C_{21} - C_{11}) & w_{21}(C_{22} - C_{12}) & \dots & w_{21}(C_{2N} - C_{1K}) \\ w_{22}(C_{21} - C_{21}) & w_{22}(C_{22} - C_{22}) & \dots & w_{22}(C_{2N} - C_{2K}) \\ \vdots & \vdots & \ddots & \vdots \\ w_{2N}(C_{21} - C_{N1}) & w_{2N}(C_{22} - C_{N2}) & \dots & w_{2N}(C_{2N} - C_{NK}) \end{bmatrix} \times \begin{bmatrix} \Delta S_1 \\ \Delta S_2 \\ \vdots \\ \Delta S_K \end{bmatrix} = \begin{bmatrix} w_{11}(r_1 - r_1) \\ w_{12}(r_1 - r_2) \\ \vdots \\ w_{1N}(r_1 - r_N) \\ w_{21}(r_2 - r_1) \\ w_{22}(r_2 - r_2) \\ \vdots \\ w_{2N}(r_2 - r_N) \end{bmatrix}. \quad \text{(A4)}$$

Note that the first row is one of the "zero" rows of $\mathbf{H}$, whereas row $N + 1$ is one of the "redundant" rows of $-\mathbf{D}$. These blocks are repeated for all N events. As the block number increases, the number of redundant rows increases by one, and the position of the zero row descends in order by one.

From equation (A4), we write a general expression for the (l, m) element of $\mathbf{H^T H}$ as

$$\mathbf{H^T H}_{lm} = w_{in}^2(C_{il} - C_{nl})(C_{im} - C_{nm}) = w_{in}^2(C_{il}C_{im} + C_{nl}C_{nm} - C_{il}C_{nm} - C_{nl}C_{im}), \quad \text{(A5)}$$

in which Einstein notation is used to represent implicit summations for indices i and n from 1 to N. Switching the indices on the double sums gives

$$\mathbf{H^T H}_{lm} = 2(C_{il}C_{im}[w_{in}^2] - [w_{in}^2 C_{\mathbf{nm}}]). \quad \text{(A6)}$$

For purposes of comparison between the demeaned and double-difference forms, it will be useful to have an expression for the case in which the uncertainties of observations from a given group p and station j are identical, that is, $w_{in} = \gamma$. In this case, equation (A6) becomes

$$\mathbf{H^T H}_{lm} = 2\gamma^2 C_{il}(NC_{im} - [C_{\mathbf{nm}}]) = 2N\gamma^2 C_{il}(C_{im} - \frac{1}{N}[C_{\mathbf{nm}}]). \quad \text{(A7)}$$

The diagonal term $(l = m)$ is

$$\mathbf{H^T H}_{mm} = 2N\gamma^2([C_{\mathbf{i}m}^2] - \frac{1}{N}[C_{\mathbf{n}m}]^2), \quad \text{(A8)}$$

in which the square brackets imply a sum over index n prior to squaring. The kth element of $\mathbf{H^T Q}$ is

$$\mathbf{H^T Q}_k = w_{in}^2(C_{ik} - C_{nk})(r_i - r_n) = w_{in}^2(C_{ik}r_i + C_{nk}r_n - C_{ik}r_n - C_{nk}r_i). \quad \text{(A9)}$$

Switching the indices on the sums gives

$$\mathbf{H^T Q}_k = 2w_{in}^2[C_{ik}r_i - C_{ik}r_n] = 2[w_{\mathbf{in}}^2]C_{ik}\left(r_i - \frac{[w_{\mathbf{in}}^2 r_{\mathbf{n}}]}{[w_{\mathbf{in}}^2]}\right), \quad \text{(A10)}$$

in which the square brackets imply a sum over the index n. In the case in which $w_{in} = \gamma$,

$$\mathbf{H^T Q}_k = 2N\gamma^2 C_{ik}\left(r_i - \frac{[r_{\mathbf{n}}]}{N}\right). \quad \text{(A11)}$$

To obtain an expression for the demeaning matrix $\mathbf{M}$, we first sum the N rows in $\mathbf{H}$ corresponding to the ith event of group p recorded at station j. From equation (A3):

$$\sum_{n=1}^{N}\sum_{k=1}^{K} w_{in}(C_{ik} - C_{nk})\Delta s_k = \sum_{n=1}^{N} w_{in}(r_i - r_n). \quad \text{(A12)}$$

Reordering the sums

$$\sum_{k=1}^{K}[C_{ik}\sum_{n=1}^{N} w_{in} - \sum_{n=1}^{N} w_{in}C_{nk}]\Delta s_k = r_i\sum_{n=1}^{N} w_{in} - \sum_{n=1}^{N} w_{in}r_n, \quad \text{(A13)}$$

and factoring out the sum of the weights:

$$\sum_{k=1}^{K}[w_{\mathbf{in}}](C_{ik} - \frac{[w_{\mathbf{in}}C_{\mathbf{n}k}]}{[w_{\mathbf{in}}]})\Delta s_k = [w_{\mathbf{in}}]\left(r_i - \frac{[w_{\mathbf{in}}r_{\mathbf{n}}]}{[w_{\mathbf{in}}]}\right). \quad \text{(A14)}$$

The term in the parentheses on the right side of equation (A14) is a standard form for a demeaned residual

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

and suggests a general form for the (i, k) element of the demeaning matrix $\mathbf{M}$:

$$\mathbf{M}_{ik} = C_{ik} - \frac{[w_{in}C_{nk}]}{[w_{in}]}. \quad (A15)$$

The corresponding ith element of the demeaned residual $\hat{r}$ would be

$$\hat{r}_i = r_i - \frac{[w_{in}r_n]}{[w_{in}]}. \quad (A16)$$

For purposes of comparison between the demeaned and double-difference forms, it will be convenient to scale row i of $\mathbf{M}$ by $[w_{in}]/\sqrt{N}$. Note that if the uncertainties are the same for all arrivals in a station-group (p, j), that is, $w_{in} = \gamma$, then this scale would be $N\gamma/\sqrt{N} = \sqrt{N}\gamma$. Because the demeaned observation is the mean of N observations with uncertainty δ_D, the appropriate uncertainty for the demeaned residual would be $\delta_M = \delta_D/\sqrt{N}$, and the corresponding weight for $\mathbf{M}$ would be $w_M = \delta_M^{-1} = \sqrt{N}\delta_D^{-1} = \sqrt{N}\gamma$. Hence, this scaling can be thought of as applying a weight that is appropriate for the uncertainty in the mean. The weighted system of equations matrix $\mathbf{M}$ for an arbitrary station-group would then be as follows:

$$\frac{1}{\sqrt{N}}\begin{bmatrix} C_{11}[w_{1n}] - [w_{1n}C_{n1}] & C_{12}[w_{1n}] - [w_{1n}C_{n2}] & \dots & C_{1K}[w_{1n}] - [w_{1n}C_{nK}] \\ C_{21}[w_{2n}] - [w_{2n}C_{n1}] & C_{22}[w_{2n}] - [w_{2n}C_{n2}] & \dots & C_{2K}[w_{2n}] - [w_{2n}C_{nK}] \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1}[w_{1n}] - [w_{Nn}C_{n1}] & C_{N2}[w_{1n}] - [w_{Nn}C_{n2}] & \dots & C_{NK}[w_{1n}] - [w_{Nn}C_{nK}] \end{bmatrix}\begin{bmatrix} \Delta S_1 \\ \Delta S_2 \\ \vdots \\ \Delta S_K \end{bmatrix} = \frac{1}{\sqrt{N}}\begin{bmatrix} r_1[w_{1n}] - [w_{1n}r_n] \\ r_2[w_{2n}] - [w_{2n}r_n] \\ \vdots \\ r_N[w_{Nn}] - [w_{Nn}r_n] \end{bmatrix}. \quad (A17)$$

From equation (A17), the general expression for the (l, m) element of $\mathbf{M}^\mathrm{T}\mathbf{M}$ is

$$\mathbf{M}^\mathrm{T}\mathbf{M}_{lm} = \frac{1}{N}(w_{in}(C_{il} - C_{nl}))(w_{in}(C_{im} - C_{nm})). \quad (A18)$$

Expanding, and being mindful of the order of summation (indicated by square brackets),

$$\mathbf{M}^\mathrm{T}\mathbf{M}_{lm} = \frac{1}{N}(C_{il}C_{im}[w_{in}]^2 - C_{il}[w_{in}][w_{in}C_{nm}] - [w_{in}C_{im}][w_{in}C_{nl}] + [w_{in}C_{nl}][w_{in}C_{nm}]).$$

Switch the order of summation on the middle two terms to get

$$\mathbf{M}^\mathrm{T}\mathbf{M}_{lm} = \frac{1}{N}(C_{il}C_{im}[w_{in}]^2 - 2C_{il}[w_{in}][w_{in}C_{nm}] + [w_{in}C_{nl}][w_{in}C_{nm}]). \quad (A19)$$

In the case in which $w_{in} = \gamma$, $[w_{in}] = N\gamma$, and

$$\mathbf{M}^\mathrm{T}\mathbf{M}_{lm} = \frac{\gamma^2}{N}(N^2C_{il}C_{im} - 2NC_{il}[C_{nm}] + N[C_{nl}][C_{nm}]). \quad (A20)$$

And because $[C_{il}] = [C_{nl}]$:

$$\mathbf{M}^\mathrm{T}\mathbf{M}_{lm} = N\gamma^2[C_{il}C_{im} - \frac{1}{N}[C_{nl}][C_{nm}]]. \quad (A21)$$

The diagonal of $\mathbf{M}^\mathrm{T}\mathbf{M}$ is

$$\mathbf{M}^\mathrm{T}\mathbf{M}_{kk} = N\gamma^2(C_{ik}^2 - \frac{[C_{nk}]^2}{N}). \quad (A22)$$

The kth element of $\mathbf{M}^\mathrm{T}\mathbf{R}_\mathbf{M}$ is

$$(\mathbf{M}^\mathrm{T}\mathbf{R}_\mathbf{M})_k = \frac{1}{N}(w_{in}(C_{ik} - C_{nk}))(w_{in}(r_i - r_n)). \quad (A23)$$

Comparing equation (A23) with the form for $\mathbf{M}^\mathrm{T}\mathbf{M}_{lm}$ in equation (A18), we see that the same result will be obtained if we substitute r_i for C_{im} and r_n for C_{nm} and change the subscript l to k:

$$(\mathbf{M}^\mathrm{T}\mathbf{R}_\mathbf{M})_k = \frac{1}{N}(C_{ik}r_i[w_{in}]^2 - 2C_{ik}[w_{in}][w_{in}r_n] + [w_{in}C_{nk}][w_{in}r_n]). \quad (A24)$$

Hence, in the case that $w_{in} = \gamma$,

$$(\mathbf{M}^\mathrm{T}\mathbf{R}_\mathbf{M})_k = N\gamma^2(C_{ik}r_i - \frac{1}{N}C_{ik}[r_n]) = N\gamma^2C_{ik}\left(r_i - \frac{[r_n]}{N}\right). \quad (A25)$$

Comparing equations (A7), (A8), and (A11) with equations (A21), (A22), and (A25), we conclude that when $w_{in} = \gamma$:

$$\mathbf{H}^\mathrm{T}\mathbf{H} = 2\mathbf{M}^\mathrm{T}\mathbf{M} = 2\mathbf{D}^\mathrm{T}\mathbf{D}, \quad (A26)$$

$$\mathbf{H}^\mathrm{T}\mathbf{Q} = 2\mathbf{M}^\mathrm{T}\mathbf{R}_\mathbf{M} = 2\mathbf{D}^\mathrm{T}\mathbf{R}_\mathbf{D}, \quad (A27)$$

$$\Delta\mathbf{x} = (\mathbf{D}^\mathrm{T}\mathbf{D})^{-1}\mathbf{D}^\mathrm{T}\mathbf{R}_\mathbf{D} = (\mathbf{M}^\mathrm{T}\mathbf{M})^{-1}\mathbf{M}^\mathrm{T}\mathbf{R}_\mathbf{M}. \quad (A28)$$

Hence, when the uncertainties are the same for all observations in each station-group pair, $\mathbf{M}$ can be scaled in such a way that

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

double differencing and demeaning produce identical least-squares estimates.

Demeaning hypocenters

Analogous to wavespeed tomography, a weighted row of **D** for hypocenters corresponding to a difference in residuals between events i and n in a group p recorded at station j is

$$w_{inpj}\left(\sum_{l=1}^{4}\frac{\partial t_{cipj}}{\partial h_{il}}\Delta h_{il}-\sum_{l=1}^{4}\frac{\partial t_{cnpj}}{\partial h_{nl}}\Delta h_{nl}\right)=w_{inpj}(r_{ipj}-r_{npj}). \quad (A29)$$

As with wavespeeds, we drop the p and j subscripts and define the partial derivative of the arrival time from event i with respect to the hypocenter variable l as

$$C_{il}=\frac{\partial t_{ci}}{\partial h_{il}}. \quad (A30)$$

The part of row of **H** corresponding to hypocenter parameter l can then be written as follows:

$$w_{in}(C_{il}\Delta h_{il}-C_{nl}\Delta h_{nl})=w_{in}(r_i-r_n). \quad (A31)$$

The first two N blocks of rows and the first N columns of a subgroup $\mathbf{H_{pj}}$ of **H**, corresponding to a particular variable l of h, are

$$\begin{bmatrix} w_{11}(C_{1l}-C_{1l}) & 0 & 0 & \dots & 0 \\ w_{12}C_{1l} & -w_{12}C_{2l} & 0 & \dots & 0 \\ w_{13}C_{1l} & 0 & -w_{13}C_{3l} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ w_{1N}C_{1l} & 0 & 0 & \dots & -w_{1N}C_{Nl} \\ -w_{21}C_{1l} & w_{21}C_{2l} & 0 & \dots & 0 \\ 0 & w_{22}(C_{2l}-C_{2l}) & 0 & \dots & 0 \\ 0 & w_{23}C_{2l} & -w_{23}C_{3l} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -w_{2N}C_{Nl} \end{bmatrix} \times \begin{bmatrix} \Delta h_{1l} \\ \Delta h_{2l} \\ \Delta h_{3l} \\ \vdots \\ \Delta h_{Nl} \end{bmatrix} = \begin{bmatrix} w_{11}(r_1-r_1) \\ w_{12}(r_1-r_2) \\ w_{13}(r_1-r_3) \\ \vdots \\ w_{1N}(r_1-r_N) \\ w_{21}(r_2-r_1) \\ w_{22}(r_2-r_2) \\ w_{23}(r_2-r_3) \\ \vdots \\ w_{2N}(r_2-r_N) \end{bmatrix}. \quad (A32)$$

To complete the rows of $\mathbf{H_{pj}}$, one would append an analogous N columns three times (one for each additional value of l). From equation (A32), the general expression for elements of $\mathbf{H^TH}$ for variables l and m for the same event i is

$$(\mathbf{H^TH})_{il,im}=2\sum_{n=1}^{N,n\neq i}w_{in}^2C_{il}C_{im}+w_{ii}^2(C_{il}-C_{il})(C_{im}-C_{im}), \quad (A33)$$

or

$$(\mathbf{H^TH})_{il,im}=2w_{in}^2C_{il}C_{im}-2w_{ii}^2C_{il}C_{im}=2C_{il}C_{im}\sum_{n=1}^{N,n\neq i}w_{in}^2. \quad (A34)$$

For the remaining case of different events i and n and variables l and m,

$$(\mathbf{H^TH})_{il,nm}=-2w_{in}^2C_{il}C_{nm}. \quad (A35)$$

We can write $\mathbf{H^TQ}$ as $N-1$ differenced rows of **Q** multiplied by four sets of columns l in **H**, each with $N-1$ rows. The row of $\mathbf{H^TQ}$ corresponding to event i and variable l is

$$\begin{aligned}(\mathbf{H^TQ})_{il}&=2\left[\sum_{n=1}^{N,n\neq i}w_{in}^2C_{il}(r_i-r_n)\right]\\ &=2C_{il}\sum_{n=1}^{N,n\neq i}w_{in}^2\left(r_i-\frac{\sum_{n=1}^{N,n\neq i}w_{in}^2r_n}{\sum_{n=1}^{N,n\neq i}w_{in}^2}\right).\end{aligned} \quad (A36)$$

In the case in which $w_{in}=\gamma$, these equations become

$$(\mathbf{H^TH})_{im}=2\gamma^2C_iC_m(N-1), \quad (A37)$$

$$(\mathbf{H^TH})_{im}=-2\gamma^2C_iC_m, \quad (A38)$$

and

$$\begin{aligned}(\mathbf{H^TQ})_{il}&=2C_{il}\gamma^2\sum_{n=1}^{N,n\neq i}\left(r_i-\frac{\sum_{n=1}^{N,n\neq i}r_n}{N-1}\right)\\ &=2C_{il}\gamma^2(N-1)\left(r_i-\frac{\sum_{n=1}^{n\neq k}r_n}{N-1}\right).\end{aligned} \quad (A39)$$

Note that because

$$\begin{aligned}N[r_i-\frac{[r_n]}{N}]&=Nr_i-[r_n]=Nr_i-\sum_{n=1}^{N,n\neq i}r_n-r_i\\ &=(N-1)r_i-\sum_{n=1}^{N,n\neq i}r_n=(N-1)\left(r_i-\frac{\sum_{n=1}^{N,n\neq i}r_n}{N-1}\right).\end{aligned} \quad (A40)$$

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

We can write equation (A39) as

$$(\mathbf{H}^{\mathrm{T}}\mathbf{Q})_{il} = 2C_{il}\gamma^2 N\left[r_i - \frac{[r_{\boldsymbol{n}}]}{N}\right]. \tag{A41}$$

The form for $\mathbf{M}$ results from summing each of the subsets of the N rows of $\mathbf{H}$ in equation (A32):

$$\sum_{l=1}^{4}\sum_{n=1}^{N} w_{in}(C_{il}\Delta h_{il} - C_{nl}\Delta h_{nl}) = \sum_{n=1}^{N} w_{in}(r_i - r_n), \tag{A42}$$

or

$$[w_{i\boldsymbol{n}}]\left[\sum_{l=1}^{4}\left(C_{il}\Delta h_{il} - \frac{[w_{i\boldsymbol{n}}C_{\boldsymbol{n}l}\Delta h_{\boldsymbol{n}l}]}{[w_{i\boldsymbol{n}}]}\right)\right] = [w_{i\boldsymbol{n}}]\left(r_i - \frac{[w_{i\boldsymbol{n}}r_{\boldsymbol{n}}]}{[w_{i\boldsymbol{n}}]}\right). \tag{A43}$$

As in the tomography case, we scale row i of $\mathbf{M}$ by $[w_{i\boldsymbol{n}}]/\sqrt{N}$. The weighted system of equations $\mathbf{M}\Delta\mathbf{h} = \mathbf{R_M}$ for an arbitrary station-group would then be

$$\frac{1}{\sqrt{N}}\begin{bmatrix} [w_{1\boldsymbol{n}}]C_1 - w_{11}C_1 & -w_{12}C_2 & \cdots & -w_{1N}C_N \\ -w_{21}C_1 & [w_{2\boldsymbol{n}}]C_2 - w_{22}C_2 & \cdots & -w_{2N}C_N \\ \vdots & \vdots & \ddots & \vdots \\ -w_{N1}C_1 & -w_{N2}C_2 & \cdots & [w_{N\boldsymbol{n}}]C_N - w_{NN}C_N \end{bmatrix} \times \begin{bmatrix} \Delta h_1 \\ \Delta h_2 \\ \vdots \\ \Delta h_N \end{bmatrix} = \frac{1}{\sqrt{N}}\begin{bmatrix} r_1[w_{1\boldsymbol{n}}] - [w_{1\boldsymbol{n}}r_{\boldsymbol{n}}] \\ r_2[w_{2\boldsymbol{n}}] - [w_{2\boldsymbol{n}}r_{\boldsymbol{n}}] \\ \vdots \\ r_N[w_{N\boldsymbol{n}}] - [w_{N\boldsymbol{n}}r_{\boldsymbol{n}}] \end{bmatrix}. \tag{A44}$$

Similar to equation (A32), only the columns for one of the hypocenter variables are shown . To form the rows for the complete $\mathbf{M}$ matrix, we would repeat this construction for each of the remaining three variables. The general expression for elements of $\mathbf{M}^{\mathrm{T}}\mathbf{M}$ for variables l and m and the same event i is

$$(\mathbf{M}^{\mathrm{T}}\mathbf{M})_{il,im} = \frac{1}{N}([w_{i\boldsymbol{n}}]C_{il} - w_{ii}C_{il})([w_{i\boldsymbol{n}}]C_{im} - w_{ii}C_{im}) + \sum_{n=1}^{N,n\neq i} w_{in}^2 C_{il}C_{im}. \tag{A45}$$

In the case in which $w_{in} = \gamma$,

$$(\mathbf{M}^{\mathrm{T}}\mathbf{M})_{il,im} = \frac{\gamma^2}{N}C_{il}C_{im}((N-1)^2 + N - 1) = \gamma^2(N-1)C_{il}C_{im}. \tag{A46}$$

For the remaining case of different events i and n, and variables l and m,

$$\begin{aligned}(\mathbf{M}^{\mathrm{T}}\mathbf{M})_{il,km} &= \frac{1}{N}[-w_{ik}C_{km}([w_{i\boldsymbol{n}}]C_{il} - w_{ii}C_{il}) - w_{ki}C_{il}([w_{i\boldsymbol{n}}]C_{km} - w_{kk}C_{km}) + \sum_{n=1}^{N,n\neq i,N,n\neq k} w_{ni}w_{nk}C_{il}C_{km}] \\ &= \frac{C_{il}C_{km}}{N}(w_{ik}(w_{ii} + w_{kk}) - w_{ik}[w_{i\boldsymbol{n}}] - w_{ki}[w_{k\boldsymbol{n}}] + \sum_{n=1}^{N,n\neq i,N,n\neq k} w_{ni}w_{nk}) = \frac{C_{il}C_{km}}{N}([w_{\boldsymbol{n}i}w_{\boldsymbol{n}k}] - 2w_{ik}[w_{i\boldsymbol{n}}]).\end{aligned} \tag{A47}$$

In the case in which $w_{in} = \gamma$,

$$(\mathbf{M}^{\mathrm{T}}\mathbf{M})_{il,km} = \frac{\gamma^2 C_{il}C_{km}}{N}(N - 2N) = -\gamma^2 C_{il}C_{km}. \tag{A48}$$

The row corresponding to event i and variable l of $\mathbf{M}^{\mathrm{T}}\mathbf{R_M}$ is

$$(\mathbf{M}^{\mathrm{T}}\mathbf{R_M})_{il} = \frac{1}{N}(C_{il}[w_{i\boldsymbol{n}}] - w_{ii}C_{il})(r_{il}[w_{i\boldsymbol{n}}] - [w_{i\boldsymbol{n}}r_{\boldsymbol{n}l}]) - [w_{i\boldsymbol{n}}]C_{il}\sum_{k=1}^{N,k\neq i} w_{ik}\left(r_{kl} - \frac{[w_{i\boldsymbol{n}}r_{\boldsymbol{n}l}]}{[w_{i\boldsymbol{n}}]}\right). \tag{A49}$$

In the case in which $w_{in} = \gamma$,

$$\begin{aligned}(\mathbf{M}^{\mathrm{T}}\mathbf{R_M})_{il} &= \frac{\gamma^2}{N}C_{il}\left((N-1)(N_{ril} - [r_{\boldsymbol{n}l}]) - N\sum_{k=1}^{N,k\neq i}\left(r_{Kl} - \frac{[r_{\boldsymbol{n}l}]}{N}\right)\right) \\ &= \frac{\gamma^2}{N}C_{il}(N^2 r_{il} - Nr_{il} + [r_{\boldsymbol{n}l}] - N[r_{\boldsymbol{n}l}] - N \times \sum_{k=1}^{N,k\neq i} r_{Kl} + (N-1)[r_{\boldsymbol{n}l}]) \\ &= \frac{\gamma^2}{N}C_{il}(N^2 r_{il} - Nr_{il} - N\sum_{k=1}^{N,k\neq i} r_{Kl}) = \frac{\gamma^2}{N}C_{il}(N^2 r_i - N[r_{\boldsymbol{n}l}]) \\ &= \frac{\gamma^2}{N}C_{il}\left(r_{il} - \frac{[r_{\boldsymbol{n}l}]}{N}\right).\end{aligned} \tag{A50}$$

Comparing equations (A7), (A8), and (A11) with equations (A46), (A48), and (A50), we again recover equations (A26)–(A28), which demonstrates the equivalence of double differencing and demeaning when accounting for differences in the uncertainties in the observations per station-group and scaling the rows of $\mathbf{M}$ accordingly.

Summary of least-squares forms

The following is a summary of how the products $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ and $\mathbf{H}^{\mathrm{T}}\mathbf{Q}$ (or, equivalently, $\mathbf{D}^{\mathrm{T}}\mathbf{D}$, $\mathbf{D}^{\mathrm{T}}\mathbf{R}_D$, $\mathbf{M}^{\mathrm{T}}\mathbf{M}$, and $\mathbf{M}^{\mathrm{T}}\mathbf{R}_M$) can be formed for cases in which it may be advantageous to solve equation (6) directly and when $w_{inpj} = \gamma_{pj}$. From the expressions derived earlier, the general form of the square of the submatrix $\mathbf{H_{pj}}$ for wavespeeds is

$$(\mathbf{H}_{\mathbf{pj}}^{\mathrm{T}}\mathbf{H_{pj}})_{ij} = N\gamma_{pj}^2(C_{ii}C_{ij} - \frac{1}{N}[C_{\boldsymbol{n}i}][C_{\boldsymbol{n}j}]). \tag{A51}$$

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929

And the kth component of the data vector $\mathbf{H}_{\mathbf{pj}}^{\mathbf{T}}\mathbf{Q}_{\mathbf{pj}}$ is

$$(\mathbf{H}_{\mathbf{pj}}^{\mathbf{T}}\mathbf{Q}_{\mathbf{pj}})_k = N\gamma_{pj}^2 C_{ik}(r_i - \frac{[r_n]}{N}). \quad \text{(A52)}$$

As shown in equations (16) and (17), $\mathbf{H}^{\mathbf{T}}\mathbf{H}$ and $\mathbf{H}^{\mathbf{T}}\mathbf{Q}$ would then be constructed by summing equations (A51) and (A52) over the entire range of stations J and event groups P.

The equivalent form for hypocenters is simple because most elements are formed by a single product. In general, the submatrix $\mathbf{H}_{\mathbf{pj}}^{\mathbf{T}}\mathbf{H}_{\mathbf{pj}}$ has dimensions $4N \times 4N$, in which N is the number of events in the station-group. If we arrange the columns of $\mathbf{H}_{\mathbf{pj}}^{\mathbf{T}}\mathbf{H}_{\mathbf{pj}}$ by variable (i.e., x, y, z, t), then $\mathbf{H}_{\mathbf{pj}}$ will be composed of 16 (i.e., 4×4) blocks $\mathbf{B}_{\mathbf{lm}}$ of dimension $N \times N$ that have the form

$$\mathbf{B}_{\mathbf{lm}} = \gamma_{pj}^2 \begin{bmatrix} C_{1l}C_{1m}(N-1) & -C_{1l}C_{2m} & \cdots & -C_{1l}C_{Nm} \\ -C_{2l}C_{1m} & -C_{2l}C_{2m}(N-1) & \cdots & -C_{2l}C_{Nm} \\ \vdots & \vdots & \ddots & \vdots \\ -C_{Nl}C_{1m} & C_{Nl}C_{2m} & \cdots & C_{Nl}C_{Nm}(N-1) \end{bmatrix}, \quad \text{(A53)}$$

in which l and m are indices related to the four hypocenter variables. Then

$$\mathbf{H}_{\mathbf{pj}}^{\mathbf{T}}\mathbf{H}_{\mathbf{pj}} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13} & \mathbf{B}_{14} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \mathbf{B}_{23} & \mathbf{B}_{24} \\ \mathbf{B}_{31} & \mathbf{B}_{23} & \mathbf{B}_{33} & \mathbf{B}_{34} \\ \mathbf{B}_{41} & \mathbf{B}_{24} & \mathbf{B}_{43} & \mathbf{B}_{44} \end{bmatrix}. \quad \text{(A54)}$$

Similarly, if we define the kth element of the vector $\mathbf{T}_{\mathrm{l}}$ as

$$(\mathbf{T}_{\mathrm{l}})_k = N\gamma_{pj}^2 C_{kl}(r_k - \frac{[r_n]}{N}), \quad \text{(A55)}$$

then

$$\mathbf{H}_{\mathbf{pj}}^{\mathbf{T}}\mathbf{Q}_{\mathbf{pj}} = \begin{bmatrix} \mathbf{T_1} \\ \mathbf{T_2} \\ \mathbf{T_3} \\ \mathbf{T_4} \end{bmatrix}. \quad \text{(A56)}$$

As with the wavespeed case, the complete $\mathbf{H}^{\mathbf{T}}\mathbf{H}$ and $\mathbf{H}^{\mathbf{T}}\mathbf{Q}$ are formed by the sum of these submatrices over the entire range of station-groups.

Manuscript received 11 January 2021
Published online 11 May 2021

Downloaded from http://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120210007/5303921/bssa-2021007.1.pdf
by 10929