Additional file 1

Supplement to

Estimation of source, path, and site factors of S waves recorded at the S-net sites in the Japan Trench area using the spectral inversion technique

By

Yadab P. Dhakal[1], Takashi Kunugi[1], Hiroaki Yamanaka[2], Atsushi Wakai[1], Shin Aoi[1], and Azusa Nishizawa[1]

[1] National Research Institute for Earth Science and Disaster Resilience, Tsukuba, Japan

[2] Tokyo Institute of Technology, Tokyo, Japan

Matrix equation used in the spectral inversion

The observed S-wave acceleration Fourier spectra at far field stations may be expressed

as shown in Eq. S1, which is identical to the Eq. 1 in the main paper.

$$O_{ij}(f) = S_i(f)G_j(f)R_{ij}^{-1}exp\left\{\frac{-\pi R_{ij}f}{Q_s(f)V_s}\right\} \qquad \text{(S1)}$$

where O_{ij} is the observed S-wave Fourier amplitude spectrum at the j^{th} station for the i^{th} earthquake, f is frequency in Hz, S_i is the source spectral amplitude of the i^{th} earthquake, G_j is the site amplification factor at the j^{th} station, Q_s is the average quality factor for S wave along the propagation path, R_{ij} is the distance between the j^{th} site and i^{th} event, and V_s is the average S-wave velocity along the propagation path.

The Eq. S1 may be linearized by taking base-10 logarithms and may be written as shown in Eq. S2. The notation, f, for frequency is skipped for simplicity in the equation.

$$\log_{10}O_{ij} + \log_{10}R_{ij} = \log_{10}S_i + \log_{10}G_j - \frac{\pi R_{ij}f}{Q_s(f)V_s}\log_{10}e \qquad \text{(S2)}$$

where e is the Napier's constant. As the Eq. S1 becomes singular at R = 0, the term $\log_{10}R_{ij}$ in Eq. S2 is undefined at R = 0. Thus, a finite distance is a prerequisite in the above equations. The arrangement of terms as shown in Eq. S2 implies that the value of the source term obtained by solving the equation corresponds to the unit reference distance of 1 km.

A matrix equation relating the unknown parameters with observations may be expressed as shown in Eq. S3.

$$Ax = b \qquad \text{(S3)}$$

where A is the design matrix with known values, x is the vector of unknown parameters, and b is the vector of observed data.

If we assume that the number of observations is n and the number of unknown parameters is m, the matrix A is of size n×m. If the numbers of earthquakes and sites are I and J, respectively, m = I+J+1. The +1 is to account for the path factor.

If the matrix A was square and nonsingular, A^{-1} exists, and the solution vector may be written as $x = A^{-1}b$. However, the number of records is larger than the number of unknown parameters in the present study. As a result, the matrix A is rectangular and its inverse does not exist in the usual sense of inversion of a matrix (e.g., Menke 2012). In the present study, the solution vector was estimated in the least-square sense that minimizes the sum of squares of the residuals between the observed and predicted values, i.e., the L_2-norm was minimized.

The normal equations for the least-square solution may be obtained by multiplying both sides of Eq. S3 by A^T as shown in Eq. S4.

$$A^T A\hat{x} = A^T b \qquad \text{(S4)}$$

where A^T is the transpose of A, and $\hat{x}$ is the estimate of x.

If the matrix $A^T A$ is invertible, the solution vector may be written as shown in Eq. S5.

$$\hat{x} = (A^T A)^{-1} A^T b \qquad \text{(S5)}$$

In the present study, we used the MATLAB R2022a (https://www.mathworks.com/) and obtained the solution vector using the syntax $x = A\backslash b$. The solution set obtained by using the syntax $x = A\backslash b$ was identical to the solution set using the syntax shown in Eq. S5, i.e., inv($A^T A$) $*$ $A^T b$. If the matrix A is large, different approaches may be taken to obtain the values of $\hat{x}$ in the above equations. Different software packages may use different optimization techniques and orders of executions, the descriptions of which are beyond the scope of this paper.

An example formulation of the design matrix and related vectors for two earthquakes, each recorded at four sites, is given below with constrained site amplification factors at two sites. Note that the design matrix can be constructed in multiple ways, and the example presented here (see Eq. S6) is simply for our convenience.

$$\mathrm{A} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & -\pi f R_{11} \log_{10} e/Vs \\ 1 & 0 & 0 & 1 & 0 & 0 & -\pi f R_{12} \log_{10} e/Vs \\ 1 & 0 & 0 & 0 & 0 & 0 & -\pi f R_{13} log_{10} e/Vs \\ 1 & 0 & 0 & 0 & 0 & 0 & -\pi f R_{14} \log_{10} e/Vs \\ 0 & 1 & 1 & 0 & 0 & 0 & -\pi f R_{21} \log_{10} e/Vs \\ 0 & 1 & 0 & 1 & 0 & 0 & -\pi f R_{22} \log_{10} e/Vs \\ 0 & 1 & 0 & 0 & 0 & 0 & -\pi f R_{23} \log_{10} e/Vs \\ 0 & 1 & 0 & 0 & 0 & 0 & -\pi f R_{24} \log_{10} e/Vs \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad \text{(S6)}$$

where the entries in the first two columns correspond to two earthquakes and the entries in the next four columns correspond to four sites. Note that the entries in the fifth and sixth columns are zeroes except for the entries in the last two rows. This is due to that the observed spectral amplitudes were corrected by the predetermined site amplification factors (i.e., the constraints) at the two sites. The last column shows the path term. R_{11} and R_{12} are the source-to-site distances from the earthquake 1 to sites 1 and 2, respectively, and so on. Vs is the average S wave velocity along the propagation path between source and site.

The x vector for the present example may be written as shown in Eq. S7.

$$x = [S_1\ S_2\ G_1\ G_2\ G_3\ G_4\ 1/Qs]^T \qquad \text{(S7)}$$

where S_1 and S_2 are the source-factor variables, G_1, G_2 etc. are the site factor variables, $1/Qs$ is the path factor variable (reciprocal of quality factor, Qs, for S wave), and T is the transpose operation. The b vector may be written as shown below.

$$b = \begin{bmatrix} \log_{10}O_{11} + \log_{10}R_{11} \\ \log_{10}O_{12} + \log_{10}R_{12} \\ \log_{10}O_{13} + \log_{10}R_{13} - \log_{10}SiteAmp_3 \\ \log_{10}O_{14} + \log_{10}R_{14} - \log_{10}SiteAmp_4 \\ \log_{10}O_{21} + \log_{10}R_{21} \\ \log_{10}O_{22} + \log_{10}R_{22} \\ \log_{10}O_{23} + \log_{10}R_{23} - \log_{10}SiteAmp_3 \\ \log_{10}O_{24} + \log_{10}R_{24} - \log_{10}SiteAmp_4 \\ \log_{10}SiteAmp_3 \\ \log_{10}SiteAmp_4 \end{bmatrix} \qquad \text{(S8)}$$

where O_{11}, O_{12} etc. are the observed spectra of earthquake 1 at sites 1 and 2, respectively, and so on. R_{11}, R_{12} etc. are the source-to-site distances as defined previously. $SiteAmp_3$ and $SiteAmp_4$ are the known values of site amplification factors at the Sites 3 and 4, respectively. By plugging the Eqs. S6 and S8 in the Eqs. S4 or S5, the solution vector $\hat{x}$ (estimate of x) may be obtained using an appropriate solver (e.g., Menke 2012). In the above matrix formulations, it is evident that the solution set for the site amplification factors for the Sites 3 and 4 (i.e., G_3 and G_4) is identical to the predefined set of values used as constraints. Please, see the main paper for details about the source, path, and site factors obtained in the present study.

Reference

Menke, W (2012) Geophysical data analysis: discrete inverse theory. Matlab Edition. Elsevier Inc.