

# Singular-value decomposition analysis of source illumination in seismic interferometry by multidimensional deconvolution

Shohei Minato<sup>1</sup>, Toshifumi Matsuoka<sup>2</sup>, and Takeshi Tsuji<sup>3</sup>

## ABSTRACT

We have developed a method to analytically evaluate the relationship between the source-receiver configuration and the retrieved wavefield in seismic interferometry performed by multidimensional deconvolution (MDD). The MDD method retrieves the wavefield with the desired source-receiver configuration from the observed wavefield without source information. We used a singular-value decomposition (SVD) approach to solve the inverse problem of MDD. By introducing SVD into MDD, we obtained quantities that revealed the characteristics of the MDD inverse problem and interpreted the effect of the initial source-receiver configuration for a survey design. We numerically simulated the wavefield with a 2D model and investigated the rank of the incident field matrix of the MDD inverse problem. With a source array of identical length, a sparse and a dense source distribution resulted in an incident field matrix of the same rank and

retrieved the same wavefield. Therefore, the optimum source distribution can be determined by analyzing the rank of the incident field matrix of the inverse problem. In addition, the introduction of scatterers into the model improved the source illumination and effectively increased the rank, enabling MDD to retrieve a better wavefield. We found that the ambiguity of the wavefield inferred from the model resolution matrix was a good measure of the amount of illumination of each receiver by the sources. We used the field data recorded at the two boreholes from the surface sources to support our results of the numerical modeling. We evaluated the rank of incident field matrix with the dense and sparse source distribution. We discovered that these two distributions resulted in an incident field matrix of almost the same rank and retrieved almost the same wavefield as the numerical modeling. This is crucial information for designing seismic experiments using the MDD-based approach.

## INTRODUCTION

Seismic interferometry (SI) is a powerful tool, especially in global and regional seismology, for obtaining new information from noise records, for example, for the retrieval of surface waves to estimate their group velocity (e.g., [Shapiro et al., 2005](#)) and of scattered waves to detect small changes in their propagation velocity caused by earthquakes (e.g., [Wegler et al., 2009](#); [Minato et al., 2012b](#)). The theory can be applied to natural earthquake records ([Ruigrok et al., 2010](#)) as well as exploration seismology, for example, in conjunction with the virtual source (VS) method ([Bakulin and Calvert, 2006](#)).

Multidimensional deconvolution (MDD) (e.g., [Wapenaar et al., 2011b](#)) is an alternative SI method to SI by crosscorrelation

(CC) (e.g., [Curtis et al., 2006](#); [Larose et al., 2006](#); [Schuster, 2009](#)). In the CC method, new seismic wavefields are retrieved by crosscorrelating the wavefields observed by receivers at many positions. The MDD method was developed to overcome limitations of the CC method. Contrary to the CC method, the MDD method is valid for a dissipative medium and compensates for differences in source spectra, and it is also valid for irregular source distributions. Due to these advantages, the MDD method has shown to be superior in some cases, such as in the retrieval of surface waves ([Wapenaar et al., 2011a](#)), in electromagnetic surveys ([Wapenaar et al., 2008](#)), in use with VSs ([van der Neut et al., 2011](#)), and in crosswell seismic data ([Minato et al., 2011](#)). On the other hand, MDD has several disadvantages compared to the CC method; MDD requires receiver arrays and cannot be applied to a single

Manuscript received by the Editor 29 June 2012; revised manuscript received 3 December 2012; published online 10 April 2013.

<sup>1</sup>Delft University of Technology, Department of Geoscience and Engineering, Delft, The Netherlands. E-mail: s.minato-1@tudelft.nl.

<sup>2</sup>Kyoto University, Graduate School of Engineering, Kyoto, Japan. E-mail: matsuoka@earth.kumst.kyoto-u.ac.jp.

<sup>3</sup>Kyushu University, International Institute for Carbon-Neutral Energy Research (WPI-I2CNER), Fukuoka, Japan. E-mail: tsuji@i2cner.kyushu-u.ac.jp.

© 2013 Society of Exploration Geophysicists. All rights reserved.

receiver configuration. MDD further requires an appropriate regularization method to stabilize the solution.

Because the assumptions required for SI are seldom met in field experiments, evaluation of source illumination during survey design is mandatory so that the reliability of the retrieved wavefield can be estimated. For example, when SI is performed by the CC method, it is assumed that sources are homogeneously distributed along a closed surface surrounding the receivers (Wapenaar, 2004). Therefore, the initial source distribution is the most important consideration in the application of SI; the source distribution is the main determinant of the amount of useful information in the observed data for retrieval of the new wavefield by SI. The assumption of homogeneously distributed sources for the CC method can be explained by using a stationary-phase approximation (e.g., Schuster et al., 2004; Snieder, 2004). It states that only initial source positions that satisfy the stationary-phase points of the integral equation of SI contribute to the produced Green's function. Furthermore, it allows us to determine how the objective wavefield can be retrieved from the source distribution in a particular application of SI. For example, Mehta et al. (2008) determine the initial source distribution required to retrieve borehole responses from surface sources by considering the stationary-phase contribution, and Chaput and Bostock (2007) also use the stationary-phase approximation to evaluate the illumination from subsurface noise sources. Minato et al. (2012a) use the stationary-phase approximation to effectively image the plate boundary at the Nankai Trough, Japan, by crosscorrelating locally distributed natural earthquakes.

Whereas the CC method involves CC and summation, the MDD method involves solving an inverse problem. Because the MDD method contains the CC operator (e.g., van der Neut and Thorbecke, 2009; Wapenaar et al., 2011b), it would be possible to apply the stationary-phase argument. To the authors' best knowledge, however, research concerning this issue has not been well studied. Instead of the stationary-phase argument, another method, one that includes a solution to the inverse problem, can be used to evaluate the relationship between the source-receiver distribution and the retrieved wavefield. van der Neut et al. (2011) implement MDD by introducing the interferometric point-spread function. They find that the interferometric point-spread function causes crosscorrelated wavefields to be represented by a blurred wavefield and that inverse filtering makes it possible to obtain an unblurred wavefield. Because the shape of the interferometric point-spread function depends on the initial source distribution (Wapenaar et al., 2011b), it can be used to evaluate the source distribution (van der Neut and Thorbecke, 2009). Curtis and Halliday (2010) show a method to correct the amplitude and phase bias introduced in the Green's function of the CC method due to the amplitude variation of the sources. The operator that they use will also show the source illumination. However, these methods do not reveal quantitative information about the source illumination such as the data redundancy.

The MDD method solves an equation categorized as the discretized Fredholm integral equation of the first kind, and it appears as an ill-conditioned linear inverse problem (e.g., Hansen, 2010). There are a variety of well-known direct and iterative regularization methods for such problems (e.g., Elden, 1977; Graves and Prenter, 1978; O'Leary and Simmons, 1981; Hansen, 1990; van der Sluis and van der Vorst, 1990). The use of the regularization method also gives the information of the ill-conditioned inverse problems (e.g., Hansen, 1992). Singular-value decomposition (SVD), which

is our method in this paper, is known to be useful to estimate and remove the redundancy due to use of the basis conversion (e.g., Kuybeda et al., 2007).

In this paper, we discuss source illumination and the reliability of the retrieved wavefield by introducing the truncated SVD as the regularization for MDD. Recently, we used this method to successfully retrieve a crosswell wavefield from vertical seismic profiling data observed in two vertical wells (Minato et al., 2011). The introduction of SVD into MDD enables us to obtain quantities that represent characteristics, such as data redundancy, of the MDD inverse problem.

First, we briefly review the MDD method, focusing on the inverse problem, and we summarize the relationship between the interferometric point-spread function and our proposed SVD inversion. Then, by numerical modeling, we show the relationships between the initial source distribution, the quality of the retrieved wavefield, and the rank of the kernel matrix of inverse problems solved by SVD. We show that the rank can be improved by introducing point scatterers and that the resolution matrix can help us to evaluate the quality of the retrieved wavefield. Finally, by using field data, we show that the data redundancy and source illumination can be investigated by SVD, which supports our results derived from the numerically modeled data. This is crucial information that can be used in the design of field experiments using SI.

## SEISMIC INTERFEROMETRY BY MDD

### Inverse problem

MDD solves the inverse problem arising from the convolution relationship between the observed and objective wavefield being estimated:

$$\hat{G}(\mathbf{x}_R, \mathbf{x}_S, \omega) = \int_{S_0} \hat{G}_d(\mathbf{x}_R, \mathbf{x}, \omega) \hat{G}^{\text{in}}(\mathbf{x}, \mathbf{x}_S, \omega) d^2\mathbf{x}. \quad (1)$$

This equation is an approximated representation of the Green's function derived from convolution-type reciprocity (Wapenaar and van der Neut, 2010), where  $\hat{G}$  is the scalar Green's function in the space-frequency domain. The function on the left side of equation 1 represents the response at  $\mathbf{x}_R$  from an initial source at  $\mathbf{x}_S$  (Figure 1a). The integrand on the right side is the product of the objective wavefield  $\hat{G}_d(\mathbf{x}_R, \mathbf{x}, \omega)$  (Figure 1b) and the actual wavefield  $\hat{G}^{\text{in}}(\mathbf{x}, \mathbf{x}_S, \omega)$  (Figure 1a) in the frequency domain, which results in time convolution. The source positions of the objective wavefield  $\hat{G}_d(\mathbf{x}_R, \mathbf{x}, \omega)$  are distributed along the integration domain  $S_0$ . Note that the derivation of equation 1 rests on several assumptions: namely, the Sommerfeld condition (that domain  $S_0$  has infinite length), the far-field assumption, and the assumption that the superscript "in" of the actual wavefield  $\hat{G}^{\text{in}}$  represents only inward-propagating waves from physical sources (Wapenaar and van der Neut, 2010). The subscript  $d$  of the objective wavefield  $\hat{G}_d(\mathbf{x}_R, \mathbf{x}, \omega)$  denotes that the source at  $\mathbf{x}$  has a dipole character,  $\hat{G}_d \approx \mathbf{n} \cdot \nabla \hat{G}$ .

Equation 1 forms a Fredholm integral equation of the first kind, and we discretized the equation. The objective wavefield  $\hat{G}_d$  of equation 1 holds for different source positions  $\mathbf{x}_S$ . We consider multiple source positions and replace the integral by a summation over the regularly distributed receivers along an integration domain  $S_0$  of finite length. Then, we obtain the following matrix relation:

$$\mathbf{p}_A = \mathbf{P}_B \mathbf{g}, \quad (2)$$

where  $\mathbf{p}_A$  is a column vector containing the observed wavefield  $\hat{G}$  at  $\mathbf{x}_R$  from the initial multiple sources  $\mathbf{x}_S$ . The matrix  $\mathbf{P}_B$  contains the observed wavefield  $\hat{G}^{\text{in}}$  at receivers on the integral surface  $S_0$  from the initial multiple sources, and the column vector  $\mathbf{g}$  contains the objective wavefield  $\hat{G}_d$ . We call the matrix  $\mathbf{P}_B$ , which is created by the observed wavefield on the surface  $S_0$ , the incident field matrix (Minato et al., 2011). We estimate the objective wavefield  $\mathbf{g}$  from the pseudoinverse of the incident field matrix, represented by  $\mathbf{P}_B^{-1}$  as

$$\mathbf{g}^{\text{est}} = \mathbf{P}_B^{-1} \mathbf{p}_A. \quad (3)$$

Equations 2 and 3 imply that the stable and unique estimation of  $\mathbf{g}$  requires observations from initial sources that are sufficiently widely distributed. However, in practical applications, the number of these sources and their spatial distribution are finite. Therefore equation 2 becomes an ill-posed problem.

### Interferometric point-spread function

Here, van der Neut et al. (2011) modify equation 1 to the cross-correlated wavefield relationship:

$$\hat{C}(\mathbf{x}_R, \mathbf{x}', \omega) = \int_{S_0} \hat{G}_d(\mathbf{x}_R, \mathbf{x}, \omega) \hat{\Gamma}(\mathbf{x}, \mathbf{x}', \omega) d^2 \mathbf{x}, \quad (4)$$

where  $\mathbf{x}' \in S_0$  and where  $\hat{C}(\mathbf{x}_R, \mathbf{x}', \omega)$  and  $\hat{\Gamma}(\mathbf{x}, \mathbf{x}', \omega)$  are the correlation function and the interferometric point-spread function, respectively, which obey the following relations:

$$\hat{C}(\mathbf{x}_R, \mathbf{x}', \omega) = \sum_i \hat{G}(\mathbf{x}_R, \mathbf{x}_S^{(i)}, \omega) \{ \hat{G}^{\text{in}}(\mathbf{x}', \mathbf{x}_S^{(i)}, \omega) \}^*, \quad (5)$$

$$\hat{\Gamma}(\mathbf{x}, \mathbf{x}', \omega) = \sum_i \hat{G}^{\text{in}}(\mathbf{x}, \mathbf{x}_S^{(i)}, \omega) \{ \hat{G}^{\text{in}}(\mathbf{x}', \mathbf{x}_S^{(i)}, \omega) \}^*, \quad (6)$$

where the summation of  $i$  sources is performed. Note that we have modified the elastic relation of van der Neut et al. (2011) into an acoustic relation. In equation 4, the crosscorrelated wavefield (left side) is a blurred (by the interferometric point-spread function) version of the objective wavefield. Inverse filtering of the crosscorrelated wavefield is equivalent to conducting MDD. Equation 4 can be rewritten in matrix notation as follows:

$$\mathbf{P}_B^\dagger \mathbf{p}_A = \mathbf{P}_B^\dagger \mathbf{P}_B \mathbf{g}, \quad (7)$$

where the symbol  $\dagger$  indicates the Hermitian conjugation. The cross-correlated incident field matrix ( $\mathbf{P}_B^\dagger \mathbf{P}_B$ ) is a point-spread function matrix whose components are defined by equation 6. The CC function is filtered by the inverse point-spread function, thus making it into a delta function (van der Neut and Thorbecke, 2009; van der Neut et al., 2011):

$$\mathbf{g}^{\text{est}} = (\mathbf{P}_B^\dagger \mathbf{P}_B + \varepsilon^2 \mathbf{I})^{-1} \mathbf{P}_B^\dagger \mathbf{p}_A, \quad (8)$$

where  $\mathbf{I}$  indicates an identity matrix and the symbol  $\varepsilon$  indicates a damping factor (Tikhonov's regularization parameter) to stabilize the inverse matrix in equation 8. This is equivalent to adopting

the damped least-squares solution for our pseudoinverse of the incident field matrix as

$$\mathbf{P}_B^{-1} = (\mathbf{P}_B^\dagger \mathbf{P}_B + \varepsilon^2 \mathbf{I})^{-1} \mathbf{P}_B^\dagger. \quad (9)$$

As mentioned, the interferometric point-spread function blurs the objective wavefield. It is apparent that the interferometric point-spread function (in equation 6) and its matrix form (in equation 7) are dependent on the initial source distribution: If the initial source distribution creates sufficient illumination for unique estimation of the objective wavefield, then it is close to the delta function and its matrix form is close to the identity matrix. Therefore, the shape of the interferometric point-spread function reveals the quality of the initial source distribution (van der Neut and Thorbecke, 2009).

### Singular-value decomposition

Although the interferometric point-spread function is a good representation for estimation of the spatial qualities of the source illumination, it is difficult to determine the characteristics of the inverse problem such as the data redundancy. In many applications, the damping factor in equation 8 is generally not equal to zero (van der Neut et al., 2011), indicating that the point-spread function matrix ( $\mathbf{P}_B^\dagger \mathbf{P}_B$ ) is not close to the identity matrix. Because the point-spread function matrix is represented as the CC of the incident field matrix (equation 6), a point-spread function matrix that is not close to the identity matrix shows that the incident field matrix inherently contains redundancy corresponding to the resemblance of data. Therefore, we introduce SVD into MDD to obtain new information for evaluating qualities of the wavefield, such as the data redundancy due to the initial source distribution.

We directly solve the inverse problem in equation 2 by adopting the Moore-Penrose pseudoinverse (Golub and van Loan, 1983) of the incident field matrix as follows:

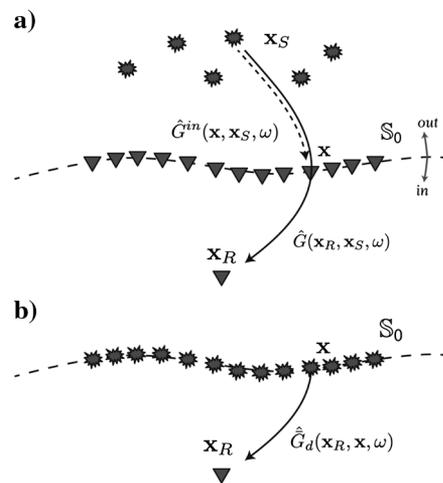


Figure 1. The source-receiver configuration for MDD, where  $\hat{G}^{\text{in}}(\mathbf{x}, \mathbf{x}_S, \omega)$ ,  $\hat{G}(\mathbf{x}_R, \mathbf{x}_S, \omega)$ , and  $\hat{G}_d(\mathbf{x}_R, \mathbf{x}, \omega)$  are the acoustic Green's functions. (a) Actual wavefield with physical sources at  $\mathbf{x}_S$  where  $\hat{G}^{\text{in}}(\mathbf{x}, \mathbf{x}_S, \omega)$  represents only inward-propagating waves recorded at  $\mathbf{x}$  from the physical sources  $\mathbf{x}_S$ ,  $\hat{G}(\mathbf{x}_R, \mathbf{x}_S, \omega)$  represents the wavefield recorded at  $\mathbf{x}_R$  from the physical sources  $\mathbf{x}_S$ , and  $S_0$  represents the domain of integration in equation 1. (b) The objective wavefield  $\hat{G}_d(\mathbf{x}_R, \mathbf{x}, \omega)$  with VSs at  $\mathbf{x}$ .

$$\mathbf{g}^{\text{est}} = \mathbf{P}_B^+ \mathbf{p}_A = \mathbf{V} \begin{pmatrix} \Delta_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{U} \mathbf{p}_A. \quad (10)$$

The matrix  $\mathbf{P}_B^+$  is the Moore-Penrose pseudoinverse of the incident field matrix, and the matrices  $\mathbf{V}$ ,  $\mathbf{U}$ , and  $\Delta_r$ , can be obtained by applying SVD to the incident field matrix  $\mathbf{P}_B$ :

$$\mathbf{P}_B = \mathbf{U} \begin{pmatrix} \Delta_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}^\dagger, \quad (11)$$

where  $r$  indicates the rank of the incident field matrix. We define the rank as the number of nonzero singular values  $i$  corresponding to the specific threshold value  $S_i$  (cumulative contribution):

$$S_i = \frac{\sum_j^i \lambda_j}{\sum_j \lambda_j} \times 100, \quad (12)$$

where  $\lambda_j$  indicates the  $j$ th singular value. Here, we define the threshold value ( $S_i$ ) as 99%. This definition indicates that 99% of the energy of the incident field matrix can be reconstructed by using the estimated number of singular values. We chose 99% as the threshold value under the assumption that the remaining 1% of energy consists mostly of noise due to the limitations of computational precision. When real data that includes field noise (e.g., ambient noise) are used, the threshold value can be varied according to the energy of the noise to avoid the amplification effects of noise originated from the inverse of the singular values in the pseudoinverse (see equation 10). Note that there are wide arguments to choose the appropriate rank indicating an appropriate model parameter by means of, e.g., investigating the Picard condition (Hansen, 1990), the

goodness-of-fit of model parameter with Akaike's information criterion (Akaike, 1974; Matsuoka and Ulrych, 1986), and the extended information criterion (Nishizawa and Lei, 1995). Minato et al. (2009) apply Akaike's information criterion to estimate the rank of the incident field matrix.

As we show later, the rank, defined by equation 12, depends on the initial source distribution. By investigating the rank, the value of  $i$  in equation 12, we can evaluate the effect of the source illumination and data redundancy quantitatively.

## SOURCE DISTRIBUTION AND RANK OF THE INCIDENT FIELD MATRIX

To demonstrate our approach, we numerically simulated wavefields and investigated the relationship between the initial source distribution and the rank of the incident field matrix, determined with equation 12 from the singular values. For clarity and simplicity, we considered the 2D case with a homogeneous medium. For a monochromatic response, the asymptotic Green's function is as follows:

$$G(d) = (8\pi kd)^{-0.5} e^{i(kd+\pi/4)}, \quad (13)$$

where  $d$  indicates the distance from the source and  $k$  is the wavenumber. Hereafter, we calculate the response at 50 Hz with a propagation velocity of 1500 m/s.

### Dense source distribution

We established 101 initial sources  $\mathbf{x}_S$  evenly spaced along a horizontal line with a length of 400 m at depth = 0 m (Figure 2a). We assumed that the integration domain  $S_0$  was a line parallel to the horizontal axis with a length of 400 m at depth = 300 m, and we established 41 evenly spaced receivers along  $S_0$ . These receiver positions become the VS positions after SI is performed. Furthermore, we established another 21 receivers  $\mathbf{x}_R$  along a vertical line with a length of 100 m at position = 0 m. Note that by establishing multiple receivers  $\mathbf{x}_R$ , the vectors in the matrix relation, equation 2, are replaced by matrices whose columns represent the different positions of the receivers  $\mathbf{x}_R$ . However, this use of multiple receivers does not change the incident field matrix. In this case, the incident field matrix ( $\mathbf{P}_B$ ) is a  $101 \times 41$  matrix. This source-receiver configuration can be interpreted as several actual observations; for example, a surface receiver array that observes natural earthquakes to retrieve the surface waves, or two boreholes (one horizontal and one vertical) that observe the wavefield from surface sources to retrieve the interborehole response.

We show the point-spread function matrix ( $\mathbf{P}_B^+ \mathbf{P}_B$ ) derived from the constructed incident field matrix in Figure 3. It is apparent that this matrix is not an identity matrix and that the off-diagonal elements have nonzero values. This result indicates that the original data contain similar wavefields, as pointed out by Minato et al. (2011). After constructing the incident field matrix, we applied SVD (complex-value SVD in this case) and determined the rank of the incident field matrix as the number of singular values obtained with equation 12. The estimated rank was 16 (Figure 4), indicating that the incident field matrix with this initial source distribution is a nonfull rank matrix. We applied the Moore-Penrose pseudoinverse solution, equation 10, to retrieve the objective wavefield (Figure 5a) and compared it with the directly modeled

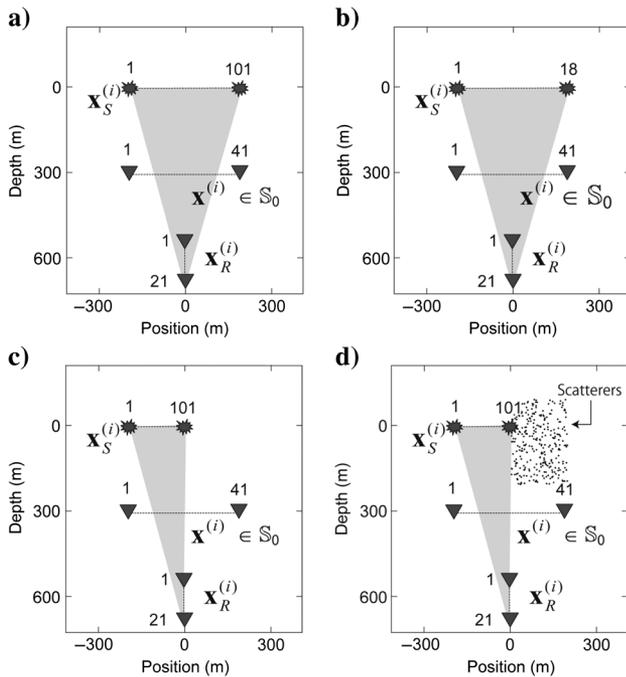


Figure 2. Source-receiver configurations for numerical modeling: (a) dense source distribution, (b) sparse source distribution, (c) localized source distribution, and (d) localized source distribution with scatterers.

wavefield (Figure 5b; only the phases are shown). The matrix in Figure 5b shows the true phase values at the multiple receivers  $\mathbf{x}_R^{(i)}$  from the multiple VSs  $\mathbf{x}^{(i)}$ , and Figure 5a shows the inverted phase values. For example, the true phase values from the VS at  $\mathbf{x}^{(21)}$  (white dotted line in Figure 5b) are shown in Figure 5c. These are equivalent to the time domain data at  $\mathbf{x}_R^{(i)}$  from the VS at  $\mathbf{x}^{(21)}$  (Figure 5d) after extraction of their phases at 50 Hz.

The phases from the VSs near the center of  $S_0$  (the part of the domain within the dashed white lines in Figure 5a) were better retrieved than those at the edge of  $S_0$  because, given our source distribution, the centrally located receivers likely receive many ray-paths corresponding to the desired wavefield (gray hatched area in Figure 2a).

The fact that the incident field matrix was estimated to be a non-full rank matrix implies that it would be possible to use the basis conversion to construct the incident field matrix from fewer observed data. In other words, the observed data at  $S_0$  contain redundant data. We assumed that the density of the source distribution was one reason for this data redundancy and performed another numerical simulation with fewer sources.

### Sparse source distribution

We established initial sources  $\mathbf{x}_S$  along a line at depth = 0 m with the same length (400 m) as that used for the dense source distribution, but we reduced the number of sources to 18 (Figure 2b). The absolute values of the singular values obtained from the incident field matrix constructed by using this sparse source distribution (Figure 6a) decreased relative to those obtained with the dense source distribution because the total energy in the observation system decreased. Furthermore, the maximum possible rank is 18 because the rank  $r \leq \min(m, n)$ , where  $m$  and  $n$  indicate the number of sources and the number of receivers, respectively. However, the rank estimated from the number of singular values, equation 12, is identical to that of the dense source distribution, 16. Furthermore, the retrieved wavefield (Figure 6b) is almost identical to that retrieved with the dense source distribution (Figure 5a). Therefore, the dense source distribution (101 sources along 400 m) can be replaced by a sparse distribution (18 sources along 400 m). The rank of the incident field matrix obtained by using SVD can thus be used

to determine a more efficient source distribution when planning a field survey by SI.

Here, we decided to reduce the number of sources to 18 by performing calculations for various numbers of sources while keeping the length of the source array and changing only the distance between sources. When we reduced the number of sources to less than

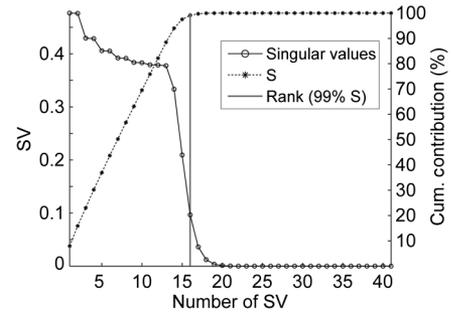


Figure 4. Singular values of the incident field matrix, their cumulative contribution, and the determined rank for the dense source distribution (see Figure 2a).

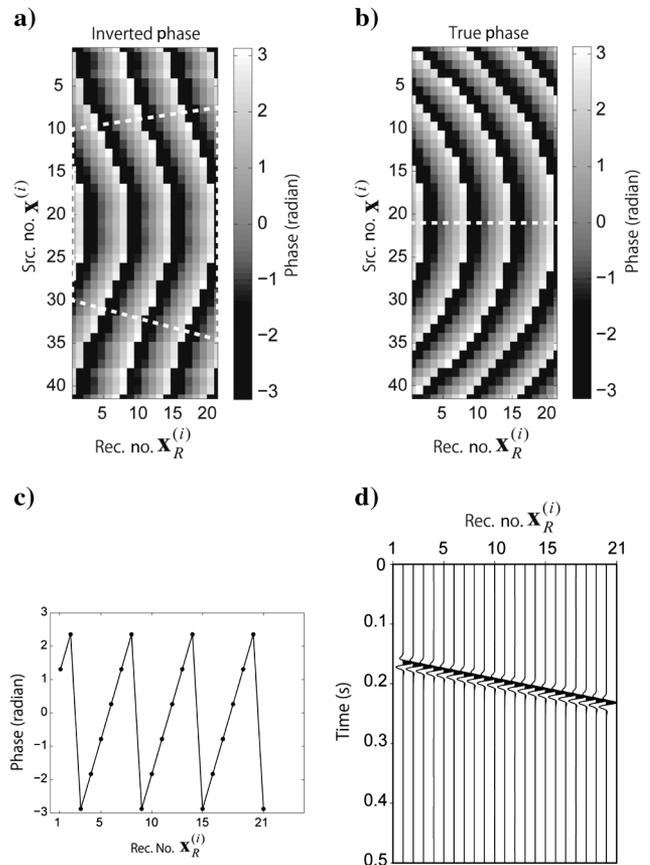


Figure 5. The phase from (a) the wavefield retrieved by MDD and (b) the true wavefield. The vertical and horizontal axes show the (virtual) source number and the receiver number, respectively. The dashed white lines in (a) enclose the area within which the MDD better retrieved the true phase. (c) The phase from the VS at  $\mathbf{x}^{(21)}$  in the true wavefield (white dotted line in [b]). (d) The true wavefield in the time domain for the VS at  $\mathbf{x}^{(21)}$  convolved with the Ricker wavelet (central frequency of 50 Hz).

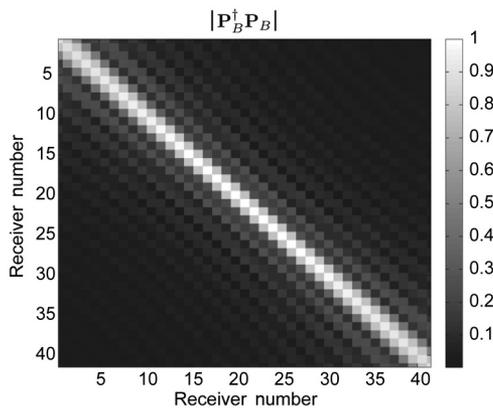


Figure 3. The point-spread function matrix (absolute values). The receiver number corresponds to the receiver's position in the integration domain shown in Figure 2a.

15, the incident field matrix was a full rank matrix, but it retrieved an inferior wavefield. This result indicates that even though we could uniquely estimate the wavefield with a full rank matrix, we failed to estimate the true wavefield because the number of independent data useful for the inversion was less than 16.

The spacing between sources resulting in the maximum rank for the same retrieved wavefield may depend on the wavelength. In the sparse distribution, the sources were 23.5 m apart, a distance that is close to the wavelength of this simulation ( $\lambda = v/f = 30$  m). This result suggests that to retrieve a better wavefield, we should determine the source distribution that results in the maximum rank. This source distribution will well illuminate the receivers. Note that this procedure is possible by introducing SVD into MDD; evaluation using the point-spread function matrix gives no information about the number of independent data.

Figure 6c compares the retrieved wavefield with different source distribution in the time domain for the VS at  $\mathbf{x}^{(21)}$  after convolving the Ricker wavelet (central frequency of 50 Hz). One can see that the wavefield is almost identical for the dense source distribution (red line) and the sparse source distribution (black line). However, the sparse source distribution contains a lower amount of high-frequency noise compared to the dense source distribution (inset of Figure 6c). Because we determined the source distribution at

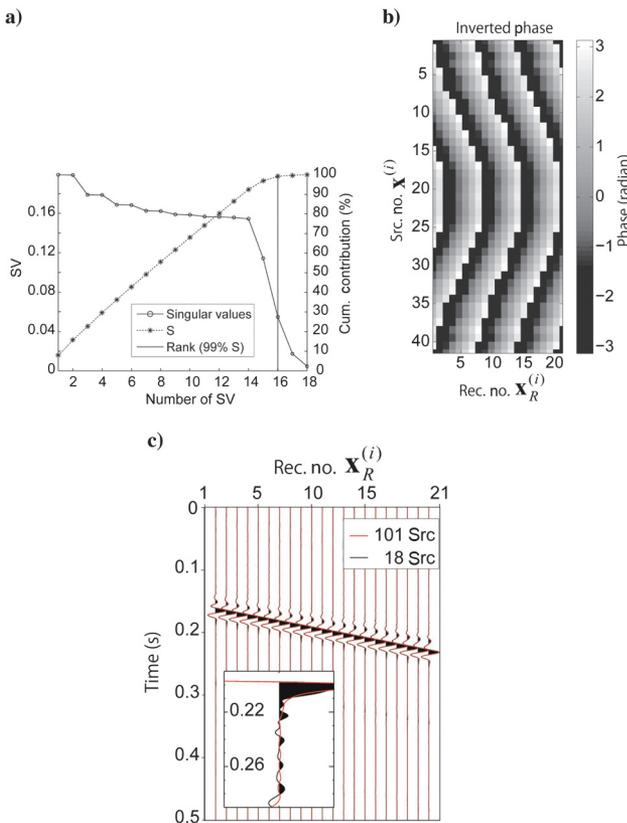


Figure 6. (a) Singular values of the incident field matrix, their cumulative contribution, and the determined rank for the sparse source distribution. (b) Wavefield (phase) estimated by MDD. (c) Retrieved wavefield in the time domain for the VS at  $\mathbf{x}^{(21)}$  convolved with Ricker wavelet (central frequency of 50 Hz) for the dense source distribution (red line) and the sparse source distribution (black line). (c) Inset, enlarged view of the record at  $\mathbf{x}_R^{(7)}$ .

50 Hz, the retrieved wavefield at a higher frequency with the shorter wavelength includes noise arising from the subsampling of the sources. In Figure 8c, however, convolving the Ricker wavelet (central frequency of 50 Hz) effectively suppresses these noises. Note that the use of this Ricker wavelet also suppresses the higher frequency signals. This observation suggests that one can estimate the optimal source distribution with the central frequency or the highest frequency at the certain bandwidth that one would like to consider for the survey design.

### Localized source distribution

To investigate inhomogeneous illumination by sources, we also modeled a localized source distribution (Figure 2c). We established 101 sources evenly spaced along a horizontal line of length 200 m at depth = 0 m. The singular values obtained from the incident field matrix (Figure 7a) show that the rank determined from the singular values by the criteria of equation 12 ( $r = 11$ ) is less than that obtained with the sources distributed along a longer line ( $r = 16$ ; Figure 4). Furthermore, the wavefields retrieved by MDD (Figure 7b) show that the true response is retrieved from only a part of the domain (between the dashed white lines in Figure 7b). These results reflect the narrower area illuminated by the initial sources (shaded area in Figure 2c) and the greater data redundancy, compared with the dense source distribution described above.

### Effect of introducing scatterers

We consider the slightly complex medium rather than the homogeneous medium by introducing scatterers in the model. The introduction of scatterers into the model increases illumination by the sources, and it is expected to retrieve a better wavefield. In this section, we show that these observations are confirmed by the improvement of the rank of the incident field matrix.

We used the same source distribution as in the previous section but distributed 301 point scatterers randomly in a  $200 \times 300$ -m area on the right of the source position (Figure 2d). Here, we simulated the wavefield as the summation of the direct wave with

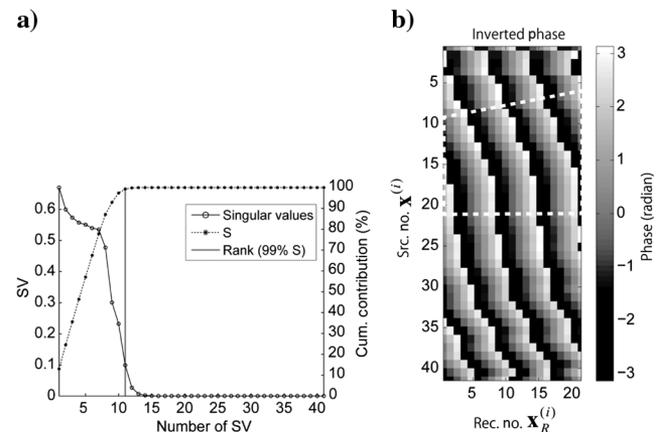


Figure 7. (a) Singular values of the incident field matrix, their cumulative contribution, and the determined rank for a localized source distribution without scatterers (see Figure 2c). (b) Wavefield (phase) estimated by MDD. The dashed white lines in panel (b) enclose the area within which the MDD better retrieved the true phase.

background velocity and the primary scattering wave without considering the interaction between scatterers. The MDD result showed that the rank of the incident field matrix was increased slightly to 12 (Figure 8a), and the wavefield was better retrieved around the center of the VSs (Figure 8b) compared with that retrieved without scatterers (Figure 7b). Furthermore, the retrieved wavefield in the time domain for the VS at  $\mathbf{x}^{(25)}$  and the receivers between  $\mathbf{x}_R^{(1)}$  to  $\mathbf{x}_R^{(6)}$  is shown in Figure 8c. Figure 8c contains the true wavefield (black line) and the retrieved wavefield from the localized source distribution without scatterers (blue line) and those with scatterers (red line). One can see that the results of the interferometry are noisy due to the insufficient illumination of the sources. However, the arrival time of the peak amplitude for the result of the scatterers (red line) is closer to the true wavefield (black line) compared to the result without scatterers (blue line). This result indicates that the presence of scatterers improved the illumination of the receivers and reduced the data redundancy, which improved the rank of the incident field matrix.

### EVALUATION OF THE RETRIEVED WAVEFIELD BY SVD MATRICES

Let us consider matrices  $\mathbf{V}$ ,  $\mathbf{U}$ , and  $\Delta_r$  obtained by SVD of the incident field matrix (equation 11). The model resolution matrix (Menke, 1989), which is a weighted matrix applied to the true model, is related to the SVD matrices as follows:

$$\mathbf{R} = \mathbf{V}\mathbf{V}^\dagger. \quad (14)$$

If we have a sufficient number of observations to uniquely retrieve the true wavefield  $\mathbf{g}^{\text{true}}$  from the data, then the estimated wavefield  $\mathbf{g}^{\text{est}}$  conforms to  $\mathbf{g}^{\text{true}}$ . Therefore, the model resolution matrix  $\mathbf{R}$  is close to an identity matrix when the estimated wavefield is close to the true wavefield, as can be inferred from the fact that the representation of the model resolution matrix is identical to the projection matrix between the data ( $\mathbf{p}_A$ ) space and the model ( $\mathbf{g}$ ) space (Lawson and Hanson, 1974).

We used the observation system with a localized source distribution (Figure 2c) without scatterers and calculated the absolute values of  $\mathbf{R}$  with equation 14 (Figure 9a). The diagonal component became less sharply defined as the number of receivers increased (Figure 9a), indicating that the wavefields whose VSs are on the right side in Figure 2c were blurred. This result is also consistent with the result for the localized source distribution, where the estimated wavefield did not reflect the VSs located on the right side (did not well illuminate the receivers with higher numbers; Figure 7b). Furthermore, the diagonal component values of  $\mathbf{R}$  (Figure 9b), which show the portion of the true model reflected in the estimated model, also indicate that the receivers with higher numbers contributed less. Although the  $\mathbf{R}$  value of the receivers at each end of the array is large (Figure 9b), we cannot evaluate the adequacy of the model at the ends because we assumed the length of  $S_0$  to be finite.

We also calculated  $\mathbf{R}$  (Figure 9c) for a wider source distribution (over 400 m; Figure 2a). Because this source distribution results in a larger rank than the localized source distribution,  $\mathbf{R}$  is close to the identity matrix. Furthermore, the diagonal component values (Figure 9d) of the central receivers were larger, reflecting the fact that the central receivers were well illuminated by the sources (Figure 2a).

### REAL DATA EXAMPLE

We showed that the data redundancy and the source illumination can be analyzed by introducing SVD into MDD. However, the results are derived only from the numerically modeled data using a homogeneous velocity model. In this section, therefore, we show several observations that support the results of our numerical modeling using real data that contain a more complicated velocity structure.

The real data are from vertical seismic profiling (VSP) consisting of seismic records observed at two vertical boreholes due to the surface sources (Figure 10a). The borehole offset is 50 m. For demonstration purposes, we use 51 receivers (hydrophones) at well-2 locating from 28- to 128-m depth, corresponding to the  $\mathbf{x}^{(i)}$  for the integration domain  $S_0$ , and one receiver at well-1 locating 140-m depth for  $\mathbf{x}_R$  in equation 1. The 13 explosive surface sources are installed to the right of well-2 at approximately 5-m intervals, and the total source array length is 65 m (Figure 10a). The upper 200 m of the survey area are composed mainly of horizontal alternating layers of sandstone and tuff. Figure 10b shows the P-wave velocity log from well-2.

To demonstrate the same analysis as in the numerical modeling section, we retrieve the direct waves from the VSs at well-2 to the

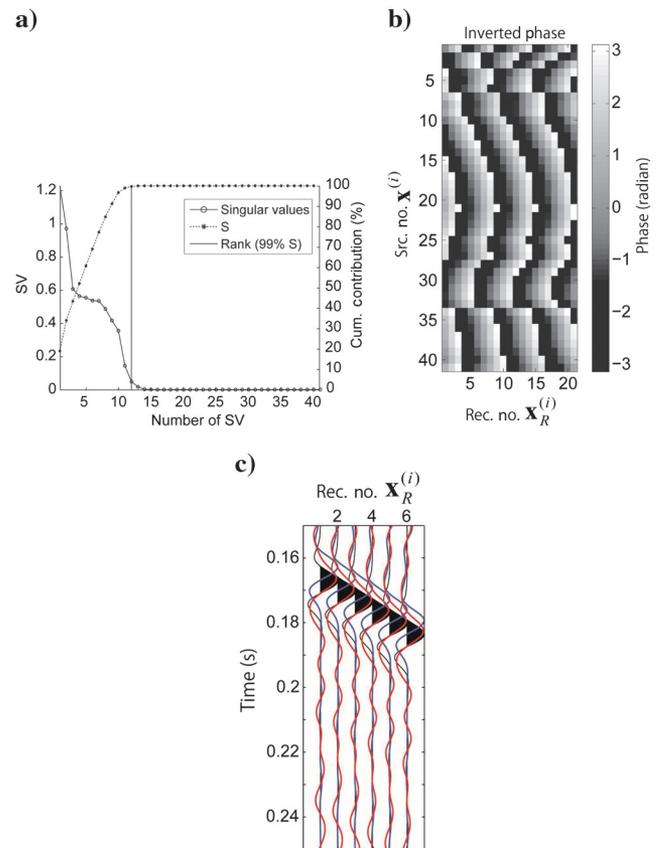


Figure 8. (a) Singular values of the incident field matrix, their cumulative contribution, and the determined rank for a localized source distribution with scatterers (see Figure 2d). (b) Wavefield (phase) estimated by MDD. (c) Retrieved wavefield in the time domain for the VS at  $\mathbf{x}^{(25)}$  without scatterers (blue line) and with scatterers (red line). Only the waveform around the direct arrivals is shown in this figure. The black line indicates the true wavefield.

receiver at well-1 using MDD. Note that one can refer to [Minato et al. \(2011\)](#) for the retrieval of the reflection responses between boreholes using the same data. The temporal recording length is 0.4 s at a sampling rate of 0.25 ms. Figure 11 shows an example of the data recorded at well-2 from the surface sources. Here we use only direct arrivals for the input data and mute the rest of the wave-

forms to check the source illumination corresponding to the numerical modeling section.

Figure 12a shows the calculated singular values and the rank at each frequency from 0 to 300 Hz for all sources (13 sources). The frequency interval is 2.5 Hz. The rank was determined using 95% of the singular value of the global maximum ([Minato et al., 2011](#)). For each frequency, the rank is smaller than the maximum possible rank (13 in this case) indicating that the data contain redundancy. Similar to the numerical modeling section, we reduce the number of the sources and perform SVD again. Figure 12b is similar to Figure 12a, but we reduce the number of sources to seven (approximately 10-m intervals) in Figure 12b. One can see that the rank of a small number of sources (Figure 12b) is almost identical to that of a large number of sources (Figure 12a). Figure 13 show the retrieved receiver gathers in the time domain using these two different source distributions. The result is obtained after applying MDD (equation 10) followed by the convolution with Ricker wavelet (central frequency of 100 Hz). One can see that the result of the 13 sources (red line) has the almost same waveforms of the result of the seven sources (black line) in Figure 13. This result demonstrates that the 13 sources can be effectively replaced by the seven sources similar to the numerical modeling section.

The green line in Figure 13 shows the travel-times of the first arrivals estimated using the P-wave velocity log (Figure 10b) assuming a horizontally layered structure. One can see that the MDD results better retrieved the direct arrivals for the shallower VS positions than the deeper positions. Furthermore, the retrieved record for the deeper VS contains noise that appeared

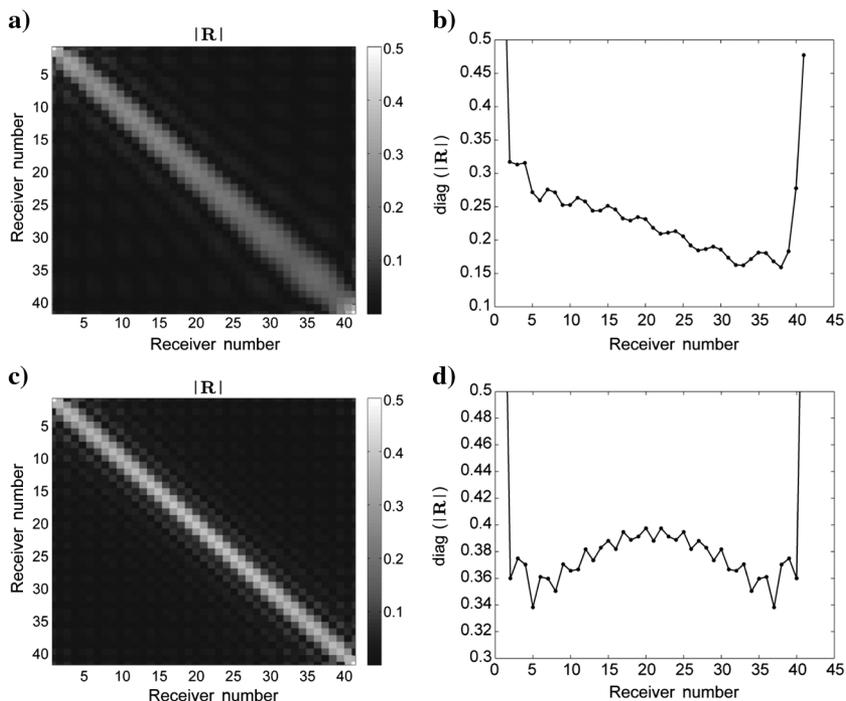


Figure 9. (a) The absolute values of the model resolution matrix  $\mathbf{V}\mathbf{V}^\dagger$  for a localized source distribution (see Figure 2c). (b) Diagonal components of (a). (c) The absolute values of the model resolution matrix for a wider source distribution (see Figure 2a). (d) Diagonal components of (c). The receiver number corresponds to the receiver position in the integration domain  $S_0$  shown in Figures 2a and 2c.

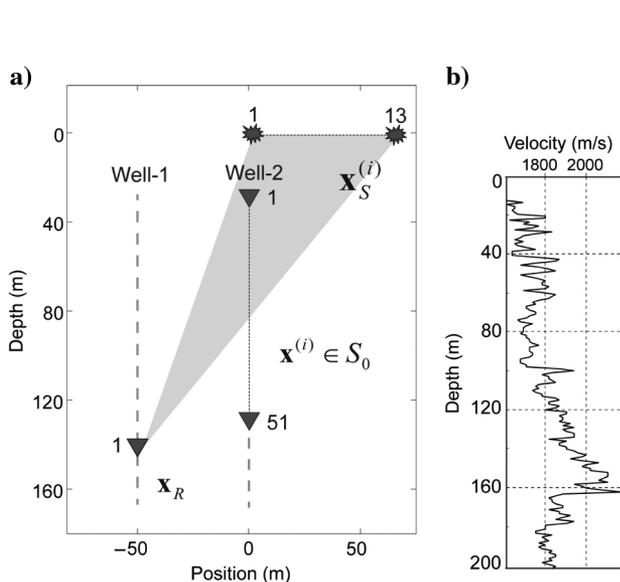


Figure 10. (a) Source-receiver configurations for the real data example. (b) The logged P-wave velocity at well-2.

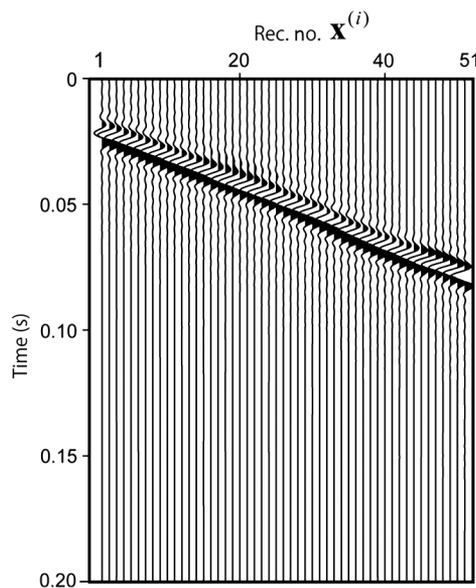


Figure 11. The example of the direct arrival data recorded in well-2 from the fifth surface source at 20 m apart from well-2.

before the first arrivals. These observations can be due to the lack of the desired wavepath in the observed records from the surface sources as shown by the shaded area in Figure 10a.

We conclude this section by showing the model resolution matrix  $\mathbf{R}$  as the measure of the source illumination. Figure 14a shows the model resolution matrix (absolute values) at 100 Hz, and Figure 14b

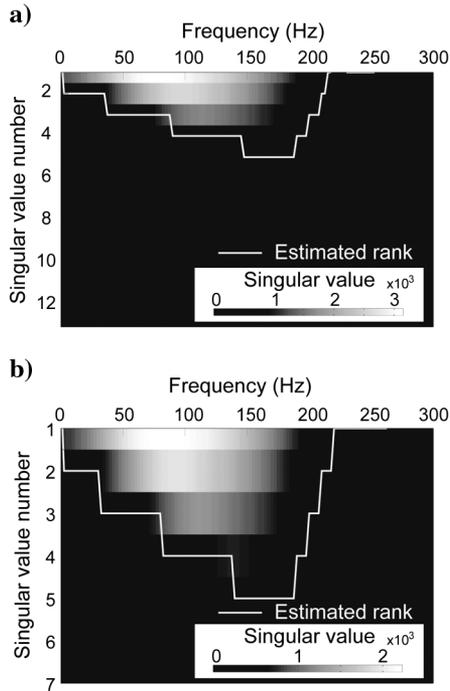


Figure 12. Singular values of the incident field matrix and the determined rank at each frequency for (a) the dense source distribution (13 sources) and (b) the sparse source distribution (seven sources).

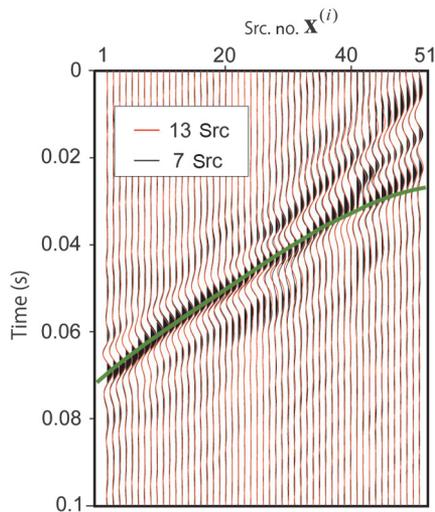


Figure 13. Retrieved receiver gather using MDD for the dense source distribution (13 sources, red line) and the sparse source distribution (seven sources, black line). Note that these two results have almost the same waveforms. The green line indicates the traveltime of the first arrival estimated from the logged data (Figure 10b).

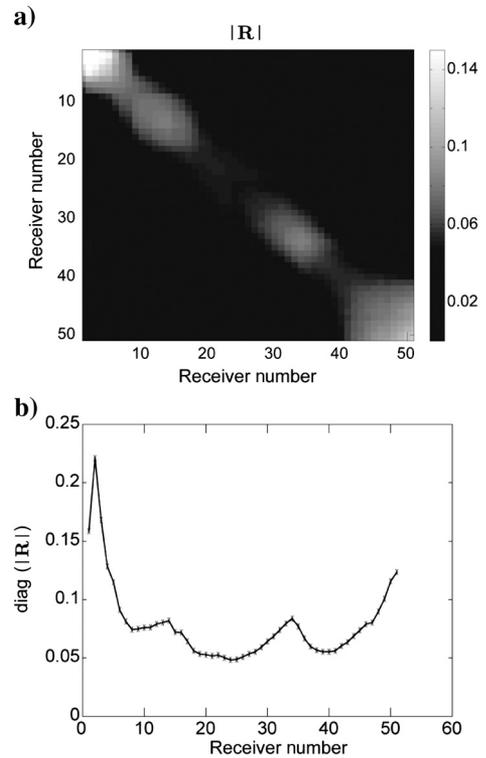


Figure 14. (a) The absolute values of the model resolution matrix  $\mathbf{V}\mathbf{V}^\dagger$  for 13 sources at 100 Hz. (b) Diagonal components of (a). The receiver number corresponds to the receiver position in well-2 (see Figure 10a).

shows its diagonal component. Due to the complexity of the recorded waveforms, the diagonal components showed a complex curvature compared to those from the numerically modeled data. However, it shows that the smaller receiver number tends to have a larger value, which is reflecting the shallow receivers were well illuminated by the sources. Note that in the field data example, the computation time to solve the MDD inverse problems including SVD was short ( $\sim 1.4$  s) using an Intel Core 2 (2.00 GHz) PC with 4 GB memory.

## CONCLUSION

We showed that the introduction of SVD into MDD is an analytic method to estimate the relationship between the source-receiver configuration (data redundancy) and the retrieved wavefield due to the initial source distribution using the quantity that represents the characteristics of the inverse problem. We numerically simulated a 2D scalar wavefield and investigated the relationship between the rank of the observation matrix (incident field matrix) and the density of the initial source distribution. The result showed that the use of a sparse source distribution did not reduce the rank of the incident field matrix compared to a dense source distribution along a source array of constant length in our example. Furthermore, it did not change the reliability of the wavefield retrieved by MDD. This indicates that a denser source distribution does not necessarily improve the retrieval of the wavefield. Therefore, a more efficient source distribution can be determined by investigating the rank of the incident field matrix. Also, we could evaluate the projection matrix constructed using the SVD matrices (the

model resolution matrix) by using the data from spatially localized sources. The ambiguity of the model inferred from the model resolution matrix showed the same trend as the discrepancy between the estimated wavefield and true wavefield. Therefore, the reliability of the estimated model can be evaluated by the matrices obtained by SVD. Furthermore, we demonstrated that the same discussions of the source-receiver configuration using SVD are successfully derived with the real data (vertical seismic profiling data), which supports our results of the numerical modeling. This is crucial information for designing seismic experiments with an MDD-based approach.

### ACKNOWLEDGMENTS

We thank the Recyclable-Fuel Storage Company and OYO Corporation for providing the field data used in this study. We thank I. Vasconcelos, D. Halliday, and an anonymous reviewer for their very useful comments and suggestions that improved the quality of the paper. S. Minato thanks J. van der Neut for helpful discussions. S. Minato is grateful for support from a Grant-in-Aid for JSPS Fellows (212666). This work was partly supported by JST/JICA, SATREPS, and Kyoto University GCOE.

### REFERENCES

- Akaike, H., 1974, A new look at the statistical model identification: *IEEE Transactions on Automatic Control*, **19**, 716–723, doi: [10.1109/TAC.1974.1100705](https://doi.org/10.1109/TAC.1974.1100705).
- Bakulin, A., and R. Calvert, 2006, The virtual source method: Theory and case study: *Geophysics*, **71**, no. 4, S1139–S1150, doi: [10.1190/1.2216190](https://doi.org/10.1190/1.2216190).
- Chaput, J., and M. Bostock, 2007, Seismic interferometry using non-volcanic tremor in Cascadia: *Geophysical Research Letters*, **34**, L07304, doi: [10.1029/2007GL028987](https://doi.org/10.1029/2007GL028987).
- Curtis, A., P. Gerstoft, H. Sato, R. Snieder, and K. Wapenaar, 2006, Seismic interferometry — Turning noise into signal: *The Leading Edge*, **25**, 1082–1092, doi: [10.1190/1.32949814](https://doi.org/10.1190/1.32949814).
- Curtis, A., and D. Halliday, 2010, Directional balancing for seismic and general wavefield interferometry: *Geophysics*, **75**, no. 1, SA1–SA14, doi: [10.1190/1.3298736](https://doi.org/10.1190/1.3298736).
- Elden, L., 1977, Algorithms for the regularization of ill-conditioned least squares problems: *BIT Numerical Mathematics*, **17**, 134–145, doi: [10.1007/BF01932285](https://doi.org/10.1007/BF01932285).
- Golub, G. H., and C. F. van Loan, 1983, *Matrix computations*: The Johns Hopkins University Press.
- Graves, J., and P. Prenter, 1978, Numerical iterative filters applied to first kind Fredholm integral equations: *Numerische Mathematik*, **30**, 281–299, doi: [10.1007/BF01411844](https://doi.org/10.1007/BF01411844).
- Hansen, P., 1990, Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank: *SIAM Journal on Scientific and Statistical Computing*, **11**, 503–518, doi: [10.1137/0911028](https://doi.org/10.1137/0911028).
- Hansen, P., 1992, Analysis of discrete ill-posed problems by means of the *L*-curve: *SIAM Review*, **34**, 561–580, doi: [10.1137/1034115](https://doi.org/10.1137/1034115).
- Hansen, P., 2010, *Discrete inverse problems: Insight and algorithms*: SIAM.
- Kuybeda, O., D. Malah, and M. Barzohar, 2007, Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors: *IEEE Transactions on Signal Processing*, **55**, 5579–5592, doi: [10.1109/TSP.2007.901645](https://doi.org/10.1109/TSP.2007.901645).
- Larose, E., L. Margerin, A. Derode, B. van Tiggelen, M. Campillo, N. Shapiro, A. Paul, L. Stehly, and M. Tanter, 2006, Correlation of random wavefields: An interdisciplinary review: *Geophysics*, **71**, no. 4, S111–S121, doi: [10.1190/1.2213356](https://doi.org/10.1190/1.2213356).
- Lawson, C. L., and R. J. Hanson, 1974, *Solving least squares problems*: Prentice Hall.
- Matsuoka, T., and T. Ulrych, 1986, Information theory measures with application to model identification: *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **34**, 511–517, doi: [10.1109/TASSP.1986.1164851](https://doi.org/10.1109/TASSP.1986.1164851).
- Mehta, K., R. Snieder, R. Calvert, and J. Sheiman, 2008, Acquisition geometry requirements for generating virtual-source data: *The Leading Edge*, **27**, 620–629, doi: [10.1190/1.2919580](https://doi.org/10.1190/1.2919580).
- Menke, W., 1989, *Geophysical data analysis: Discrete inverse theory*: Academic Press.
- Minato, S., T. Matsuoka, T. Tsuji, D. Draganov, J. Hunziker, and K. Wapenaar, 2009, Application of seismic interferometry by multidimensional deconvolution to crosswell seismic reflection using singular-value decomposition: 79th Annual International Meeting, SEG, Expanded Abstracts, 1688–1692.
- Minato, S., T. Matsuoka, T. Tsuji, D. Draganov, J. Hunziker, and K. Wapenaar, 2011, Seismic interferometry using multidimensional deconvolution and crosscorrelation for crosswell seismic reflection data without borehole sources: *Geophysics*, **76**, no. 1, SA19–SA34, doi: [10.1190/1.3511357](https://doi.org/10.1190/1.3511357).
- Minato, S., T. Tsuji, T. Matsuoka, and K. Obana, 2012a, Crosscorrelation of earthquake data using stationary phase evaluation: Insight into reflection structures of oceanic crust surface in the Nankai Trough: *International Journal of Geophysics*, **2012**, 101545, doi: [10.1155/2012/101545](https://doi.org/10.1155/2012/101545).
- Minato, S., T. Tsuji, S. Ohmi, and T. Matsuoka, 2012b, Monitoring seismic velocity change caused by the 2011 Tohoku-oki earthquake using ambient noise records: *Geophysical Research Letters*, **39**, L09369, doi: [10.1029/2012GL051405](https://doi.org/10.1029/2012GL051405).
- Nishizawa, O., and X. Lei, 1995, A numerical study on finding an optimum model in velocity tomography by using the extended information criterion: *Geophysical Research Letters*, **22**, 1313–1316, doi: [10.1029/95GL01092](https://doi.org/10.1029/95GL01092).
- O’Leary, D., and J. Simmons, 1981, A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems: *SIAM Journal on Scientific and Statistical Computing*, **2**, 474–489, doi: [10.1137/0902037](https://doi.org/10.1137/0902037).
- Ruigrok, E., X. Campman, D. Draganov, and K. Wapenaar, 2010, High-resolution lithospheric imaging with seismic interferometry: *Geophysical Journal International*, **183**, 339–357, doi: [10.1111/j.1365-246X.2010.04724.x](https://doi.org/10.1111/j.1365-246X.2010.04724.x).
- Schuster, G., 2009, *Seismic interferometry*: Cambridge University Press.
- Schuster, G., J. Yu, J. Sheng, and J. Rickett, 2004, Interferometric/daylight seismic imaging: *Geophysical Journal International*, **157**, 838–852, doi: [10.1111/j.1365-246X.2004.02251.x](https://doi.org/10.1111/j.1365-246X.2004.02251.x).
- Shapiro, N., M. Campillo, L. Stehly, and M. H. Ritzwoller, 2005, High-resolution surface-wave tomography from ambient seismic noise: *Science*, **307**, 1615–1618, doi: [10.1126/science.1108339](https://doi.org/10.1126/science.1108339).
- Snieder, R., 2004, Extracting the Green’s function from the correlation of coda waves: A derivation based on stationary phase: *Physical Review E*, **69**, 046610, doi: [10.1103/PhysRevE.69.046610](https://doi.org/10.1103/PhysRevE.69.046610).
- van der Neut, J., and J. Thorbecke, 2009, Resolution function for controlled-source seismic interferometry: A data-driven diagnosis: 79th Annual International Meeting, SEG, Expanded Abstracts, 4090–4094.
- van der Neut, J., J. Thorbecke, K. Mehta, E. Slob, and K. Wapenaar, 2011, Controlled source interferometric redatuming by crosscorrelation and multidimensional deconvolution in elastic media: *Geophysics*, **76**, no. 4, SA63–SA76, doi: [10.1190/1.3580633](https://doi.org/10.1190/1.3580633).
- van der Sluis, A., and H. van der Vorst, 1990, SIRT- and CG-type methods for the iterative solution of sparse linear least-squares problems: *Linear Algebra and Its Applications*, **130**, 257–303, doi: [10.1016/0024-3795\(90\)90215-X](https://doi.org/10.1016/0024-3795(90)90215-X).
- Wapenaar, K., 2004, Retrieving the elastodynamic Green’s function of an arbitrary inhomogeneous medium by cross correlation: *Physical Review Letters*, **93**, 254301, doi: [10.1103/PhysRevLett.93.254301](https://doi.org/10.1103/PhysRevLett.93.254301).
- Wapenaar, K., E. Ruigrok, J. van der Neut, and D. Draganov, 2011a, Improved surface-wave retrieval from ambient seismic noise by multidimensional deconvolution: *Geophysical Research Letters*, **38**, L01313, doi: [10.1029/2010GL045523](https://doi.org/10.1029/2010GL045523).
- Wapenaar, K., E. Slob, and R. Snieder, 2008, Seismic and electromagnetic controlled-source interferometry in dissipative media: *Geophysical Prospecting*, **56**, 419–434, doi: [10.1111/j.1365-2478.2007.00686.x](https://doi.org/10.1111/j.1365-2478.2007.00686.x).
- Wapenaar, K., and J. van der Neut, 2010, A representation for Green’s function retrieval by multidimensional deconvolution: *Journal of the Acoustical Society of America*, **128**, EL366–EL371, doi: [10.1121/1.3509797](https://doi.org/10.1121/1.3509797).
- Wapenaar, K., J. van der Neut, E. Ruigrok, D. Draganov, J. Hunziker, E. Slob, J. Thorbecke, and R. Snieder, 2011b, Seismic interferometry by crosscorrelation and by multidimensional deconvolution: A systematic comparison: *Geophysical Journal International*, **185**, 1335–1364, doi: [10.1111/j.1365-246X.2011.05007.x](https://doi.org/10.1111/j.1365-246X.2011.05007.x).
- Wegler, U., H. Nakahara, C. Sens-Schönfelder, M. Korn, and K. Shiomi, 2009, Sudden drop of seismic velocity after the 2004 Mw 6.6 mid-Niigata earthquake, Japan, observed with passive image interferometry: *Journal of Geophysical Research*, **114**, B06305, doi: [10.1029/2008JB005869](https://doi.org/10.1029/2008JB005869).