Site Characterization at Downhole Arrays by Joint Inversion of Dispersion Data and Acceleration Time Series

Elnaz Seylabi*[1], Andrew M. Stuart[2], and Domniki Asimaki[3]

ABSTRACT

We present a sequential data assimilation algorithm based on the ensemble Kalman inversion to estimate the near-surface shear-wave velocity profile and damping; this is applicable when heterogeneous data and a priori information that can be represented in forms of (physical) equality and inequality constraints in the inverse problem are available. Although noninvasive methods, such as surface-wave testing, are efficient and cost-effective methods for inferring an V_S profile, one should acknowledge that site characterization using inverse analyses can yield erroneous results associated with the lack of inverse problem uniqueness. One viable solution to alleviate the unsuitability of the inverse problem is to enrich the prior knowledge and/or the data space with complementary observations. In the case of noninvasive methods, the pertinent data are the dispersion curve of surface waves, typically resolved by means of active source methods at high frequencies and passive methods at low frequencies. To improve the inverse problem suitability, horizontal-to-vertical spectral ratio data are commonly used jointly with the dispersion data in the inversion. In this article, we show that the joint inversion of dispersion and strong-motion downhole array data can also reduce the margins of uncertainty in the V_S profile estimation. This is because acceleration time series recorded at downhole arrays include both body and surface waves and therefore can enrich the observational data space in the inverse problem setting. We also show how the proposed algorithm can be modified to systematically incorporate physical constraints that further enhance its suitability. We use both synthetic and real data to examine the performance of the proposed framework in estimation of the V_S profile and damping at the Garner Valley downhole array and compare them against the V_S estimations in previous studies.

KEY POINTS

- We present an algorithm for site characterization based on the ensemble Kalman inversion.
- We discuss the effects of using heterogeneous and/or incomplete data on lack of solution uniqueness.
- We show that joint inversion of dispersion and acceleration data can reduce uncertainty in V_S estimations.

INTRODUCTION

Downhole arrays have been extensively used as testbeds by engineers and earth scientists for validation and improvement of predictive models of site response and physics-based ground motions. Strong motions recorded at depth, in particular, are widely sought-after boundary conditions for the validation of 1D wave propagation codes. This is because they minimize the uncertainty that is associated with source and path effects when studying problems of site response. To best serve as validation testbeds for equivalent linear and nonlinear site-response analyses, however, downhole array sites should also be accompanied by reliable estimates of soil profiles. These could include small strain shear modulus (or shear-wave velocity), damping, and nonlinear soil properties, as well as their variability.

Site characterization efforts at strong-motion stations include shear-wave velocity measurements at multiple locations,

---

1. Department of Civil and Environmental Engineering, University of Nevada, Reno, Reno, Nevada, U.S.A.; 2. Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California, U.S.A.; 3. Department of Mechanical and Civil Engineering, California Institute of Technology, Pasadena, California, U.S.A.

*Corresponding author: elnaze@unr.edu

Cite this article as Seylabi, E., A. M. Stuart, and D. Asimaki (2020). Site Characterization at Downhole Arrays by Joint Inversion of Dispersion Data and Acceleration Time Series, *Bull. Seismol. Soc. Am.* **110**, 1323–1337, doi: 10.1785/0120190256

using both invasive (e.g., *PS* logging) and/or noninvasive methods (e.g., spectral analysis; Stokoe, 1994) and multichannel analysis of surface waves; Foti, 2000). Nonlinear soil properties, on the other hand, are most frequently estimated from empirical correlations of published laboratory data (e.g., Darendeli, 2001), or, in rare cases, measured from undisturbed samples retrieved at the site. Laboratory measured properties, however, are not always reliable estimates of in situ soil properties; this is an incompatibility associated with sample disturbance, measurement error, and inherent field-scale spatial variability of soil properties. Downhole array ground-motion records have been used instead by several researchers to infer in situ soil parameters. More specifically, low-amplitude motions have been used to constrain the V_S profile and damping (e.g., Assimaki *et al.*, 2006), whereas high-amplitude motions have been used to parameterize nonlinear soil behavior in moderate-to-large strains (e.g., Assimaki *et al.*, 2011; Chandra *et al.*, 2015; Seylabi and Asimaki, 2018).

Although it is commonly assumed that invasive methods are more reliable than noninvasive methods to retrieve V_S profiles, recent studies have highlighted that the latter are more efficient and cost effective, while yielding uncertainties comparable to the order of invasive methods (Garofalo *et al.*, 2016; Teague and Cox, 2016). Still, one should acknowledge that site characterization using inverse analyses, that is, surface-wave testing or downhole array data, can yield erroneous results associated with a lack of inverse problem uniqueness. Furthermore, recent studies have shown that significantly different V_S profiles that satisfy the error thresholds of the inversion process can still result in similar linear viscoelastic seismic site response and amplification factors (e.g., Foti *et al.*, 2009).

One viable solution to alleviate the inverse problem unsuitability is to enrich the prior information and/or the data space with complementary data. In the case of noninvasive methods, the pertinent data are the dispersion curve of surface waves, typically resolved by means of active source methods at high frequencies and passive methods at low frequencies. Horizontal-to-vertical spectral ratios (HVSRs) have also been used extensively to approximate a site's predominant frequency and therefore constrain the velocity structure at depth. Furthermore, the joint inversion of HVSR with dispersion data has been used successfully to improve the site characterization accuracy (e.g., Arai and Tokimatsu, 2005; Lunedei and Malischewsky, 2015; Piña-Flores *et al.*, 2016; Molnar *et al.*, 2018).

In strong-motion arrays, recorded acceleration time series, which include both body and surface waves, can also be used to complement dispersion data. Because of the complementary characteristics of the body and surface waves (the former carrying information in the form of travel time and the latter in the form of near-surface dispersive characteristics), we anticipate that formulating the inversion problem using both dispersion data and acceleration time series should reduce the margins of uncertainty in the V_S profile estimation.

Inversion methods that are used in site characterization include the stochastic direct search method, which is based on the neighborhood algorithm (e.g., Wathelet *et al.*, 2004); the uniform Monte Carlo method (e.g., Socco and Boiero, 2008), which finds an ensemble of models minimizing the data misfit; and the fully Bayesian Markov chain Monte Carlo method (e.g., Molnar *et al.*, 2010), which provides the most probable and quantitative uncertainty estimates of the V_S profile. In this article, we present a framework based on the ensemble Kalman inversion and use it to examine whether and how the joint inversion of dispersion and downhole array data improves site characterization estimates. We also show how one can systematically incorporate a priori information in terms of physical constraints in the ensemble Kalman inversion to improve the problem suitability.

In the rest of this article, we first explain the ingredients of the framework in the Inverse Problem section. In the Site Characterization: Synthetic Data and Site Characterization: Real Data sections, we perform numerical experiments using synthetic and real data that was simulated or recorded at the Garner Valley downhole array (GVDA) site, which is one of the best-studied and best-instrumented sites in southern California. More specifically, in the Site Characterization: Synthetic Data section, we use synthetic data to study how the combined data sets improve the V_S profile estimation relative to the individual data. Next, in the Site Characterization: Real Data section, we use recorded acceleration time series by the array and experimental dispersion data to invert for the V_S profile and damping, and we compare our results to the inverted V_S profiles from previous studies. Finally, we provide concluding remarks in the Conclusion.

INVERSE PROBLEM

Problem formulation

We consider the problem of finding u from a series of data sets y_i in which:

$$y_i = G_i(u) + \eta_i. \tag{1}$$

$u \in \mathbb{R}^k$ consists of k uncertain parameters, $y_i \in \mathbb{R}^{m_i}$ consists of m_i observations spanning the ith data space, $\eta_i \in \mathbb{R}^{m_i}$ is the noise represented as independent zero-mean Gaussian noise with covariance Γ_i, and G_i is a nonlinear function (referred to as the forward model) that maps the parameter space to the ith data set. In this article, we work with two data sets, the dispersion curve that depicts discrete phase velocity values of surface waves as a function of frequency, and the discrete acceleration time series recorded by the array instruments at different depths. If we combine these two data sets, the inverse problem of equation (1) is modified as follows:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} G_1(u) \\ G_2(u) \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \to y = G(u) + \eta, \qquad \Gamma = \begin{bmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{bmatrix}. \tag{2}$$

We define the covariance matrix of the Gaussian noise as follows:

$$\Gamma_1 = [\beta_1 \mathrm{diag}(\max |y_1| 1)]^2, \qquad \Gamma_2 = [\beta_2 \mathrm{diag}(y_2)]^2, \tag{3}$$

in which β_1 and β_2 determine the noise levels for y_1 and y_2 which are acceleration time series and dispersion data, respectively. $1 \in R^{m_1}$ is a vector of ones.

For the two data sets relevant to the problem in hand, the forward models are described below:

1. For the theoretical dispersion curve, we use the transfer matrix approach originally developed by Thomson (1950) and Haskell (1953) and later modified by Dunkin (1965) and Knopoff (1964). This approach requires the solution of an eigenvalue problem, for which we use the well-known software GEOPSY (Wathelet, 2005). We should mention here that all the examples considered in this study are normally dispersive, so results presented here only require the dispersion curve for the first mode of Rayleigh waves. However, the framework is general enough to allow for higher modes to be incorporated in the inversion process.
2. For the theoretical acceleration time series, we consider wave propagation in a horizontally stratified layered soil of total thickness H and shear-wave velocity $V_S(z)$ varying with depth z. Given an acceleration time series at $z = H$ (i.e., the borehole sensor depth), we compute the soil response numerically using a finite-element model, and we use the extended Rayleigh damping (Phillips and Hashash, 2009) to capture the nearly frequency-independent viscous damping ξ in time-domain analyses. We also assume that V_S below $z = H$ is constant, which corresponds to an elastic bedrock idealization.

Algorithm

To solve the inverse problem involving the two data sets just described, alone or in conjunction, we use a sequential data assimilation method (Evensen, 2009) that is based on ensemble Kalman inversion (Iglesias *et al.*, 2013), a methodology pioneered in the oil reservoir community (Chen and Oliver, 2012; Emerick and Reynolds, 2013). In this algorithm, we first define an initial ensemble consisting of N particles. In its most basic form, the ensemble Kalman inversion can regularize unsuitable inverse problems through the subspace property for which the solution found is in the linear span of the initial ensemble employed (Chada *et al.*, 2019). Then, at each iteration j, we use the forward model predictions $G(u_j^{(n)})$ and the observation data y_{j+1} to update these particles sequentially:

$$u_{j+1}^{(n)} = u_j^{(n)} + C_{j+1}^{uw}(C_{j+1}^{ww} + \Gamma)^{-1}(y_{j+1}^{(n)} - G(u_j^{(n)})) \quad \text{for} \quad n = 1, \ldots, N, \tag{4}$$

in which $y_{j+1}^{(n)}$ can be either identical to y_{j+1} (the observation data) or found by adding to y_{j+1} identical and independently distributed zero-mean Gaussian noise $\eta_{j+1}^{(n)}$, with distribution of the same as that of η; the matrices C_{j+1}^{uw} and C_{j+1}^{ww} are empirical covariance matrices that can be computed at each iteration based on predictions and the ensemble mean $\bar{u}_{j+1}$ using the following equations:

$$C_{j+1}^{uw} = \frac{1}{N}\sum_{n=1}^{N}(u_j^{(n)} - \bar{u}_{j+1}) \otimes (G(u_j^{(n)}) - \bar{G}_j), \tag{5}$$

$$C_{j+1}^{ww} = \frac{1}{N}\sum_{n=1}^{N}(G(u_j^{(n)}) - \bar{G}_j) \otimes (G(u_j^{(n)}) - \bar{G}_j), \tag{6}$$

and

$$\bar{u}_{j+1} = \frac{1}{N}\sum_{n=1}^{N} u_j^{(n)}, \qquad \bar{G}_j = \frac{1}{N}\sum_{n=1}^{N} G(u_j^{n}). \tag{7}$$

To enforce physical and prior knowledge systematically, Albers *et al.* (2019) provide an efficient procedure to impose constraints within the ensemble Kalman filtering framework; they use the solution of a constrained quadratic programming to update particles that violate the enforced constraints; in the absence of constraints, the optimization delivers the updated formula that we specified earlier. In the case of linear equality and inequality constraints, the approach is readily implemented using standard optimization algorithms. We briefly summarize the procedure in the case of linear inequality constraints:

$$Au \le a, \tag{8}$$

and when the problem is formulated in the range of the covariance matrix C_{j+1}:

$$C_{j+1} = \begin{bmatrix} C_{j+1}^{uu} & C_{j+1}^{uw} \\ C_{j+1}^{uwT} & C_{j+1}^{ww} \end{bmatrix}, \tag{9}$$

in which:

$$C_{j+1}^{uu} = \frac{1}{N}\sum_{n=1}^{N}(u_j^{(n)} - \bar{u}_{j+1}) \otimes (u_j^{(n)} - \bar{u}_{j+1}). \tag{10}$$

It should be noted that the empirically computed covariance is the sum of rank one matrices and its rank is at most $N - 1$. For this problem setting, $N - 1$ is generally less than $k + m_1 + m_2$; therefore the covariance matrix is not invertible. To overcome this issue, Albers *et al.* (2019) reformulate the problem in the range of the covariance in which they seek the solution as a

ALGORITHM 1

Constrained Ensemble Kalman Inversion Algorithm Formulated in Range of Covariance:

```
1: Choose $\{u_0^{(n)}\}_{n=1}^{N}$, $j = 0$
2: Calculate forward model application $\{G(u_j^{(n)})\}_{n=1}^{N}$
3: Update $\{u_{j+1}^{(n)}\}_{n=1}^{N}$ from (4)
4: for $n = 1 : N$ do
5:     if $u_{j+1}^{(n)}$ violates constraints in (12) then
6:         $b^{(n)}$ & larr; argmin of (11) subject to (12)
7:         Update $\{u_{j+1}^{(n)}\}$ from (14)
8:     end if
9: end for
10: $j \leftarrow j + 1$, go to 2.
```

linear combination of a given set of vectors. For more details, see Albers *et al.* (2019). As such, for each violating particle, we seek a vector $b^{(n)}$ that minimizes the cost function $J_{j,n}(b)$ defined as follows:

$$J_{j,n}(b) := \frac{1}{2}|y_{j+1}^{(n)} - G(u_j^{(n)}) - \frac{1}{N}\sum_{m=1}^{N} b_m(G(u_j^{(m)}) - \bar{G}_j)|_\Gamma^2 + \frac{1}{2N}\sum_{m=1}^{N}(b_m)^2, \tag{11}$$

and subject to:

$$ABb \le a - Au_j^{(n)}, \tag{12}$$

in which:

$$Bb = \frac{1}{N}\sum_{m=1}^{N} b_m(u_j^{(m)} - \bar{u}_{j+1}). \tag{13}$$

Next, we use the computed $b^{(n)}$ to update each violating particle $u_j^{(n)}$ as follows:

$$u_{j+1}^{(n)} = u_j^{(n)} + \frac{1}{N}\sum_{m=1}^{N} b_m^{(n)}(u_j^{(m)} - \bar{u}_{j+1}). \tag{14}$$

We have implemented this algorithm, summarized in Algorithm 1, and we have verified its accuracy in the numerical results section of Albers *et al.* (2019).

SITE CHARACTERIZATION: SYNTHETIC DATA

We first use synthetic data to evaluate the importance of joint inversion of downhole array and dispersion data in near-surface site characterization using the proposed framework. In this numerical experiment, we use the site conditions at the GVDA in southern California, which is also the site where we test the algorithm using recorded ground-motion data in the following sections of this work. GVDA is located in a narrow valley, and the near-surface structure consists of an ancestral lake bed with soft alluvium down to 18–25 m depth, overlaying a layer of weathered granite; the competent granitic bedrock interface is located at a depth of 87 m, according to Bonilla *et al.* (2002), and varies across the valley (Teague *et al.*, 2018). Figure 1 shows the site geology and the layout of sensors at different depths, including accelerometers (red boxes) and pressure transducers (blue boxes). Several invasive and noninvasive V_S measurements have been carried out at this site in the past; the most recent surface-wave measurements were performed in October 2016 by Teague *et al.* (2018), who developed V_S profiles at each of the three surface accelerometer locations (i.e., at the location of sensors 00, 12, and 21 in Fig. 1).

To generate ground-truth downhole array data, we use acceleration time series recorded at GVDA from an event with magnitude 4.28. We also consider the V_S profile shown in Table 1 as the target profile; we also assume mass density $\rho = 1800\ \mathrm{kg/m^3}$, Poisson's ratio $\nu = 0.3$, and damping ratio $\xi = 0.04$ for all layers. We next use these input parameters in the forward models described previously to compute the acceleration time history at the ground surface and the first mode Rayleigh-wave dispersion curve, and we use these simulated data as observations in the inversion process.

When reliable invasive measurements are not available at the site, prior information on the thickness of the soil layers is also unavailable. To overcome this shortcoming, we use a fine discretization of the profile shown in the following equation, with increasing thickness increments Δh with depth, ranging from $\Delta h = 5$ to 25 m. This selection has resulted in $r = 15$ layers for the $H = 150$ m thick profile (from the surface to the depth of borehole sensor 05 shown in Fig. 1). Our forward model also considers elastic bedrock boundary conditions beyond depth 150 m, which we characterize by a thin layer of thickness 1 m in the following equation:

$$\Delta h[\mathrm{m}] = \{5, 5, 5, 5, 5, 5, 10, 10, 10, 10, 15, 15, 25, 24, 1\}. \tag{15}$$

Assuming that the V_S is constant in each layer and the profile is normally dispersive, we enforce monotonic behavior by posing the following linear inequality constraints, as well as lower and upper bounds for the very first and last layers. We assume that the V_S profile can change monotonically between 50 m/s at surface and 5000 m/s at the bedrock, which is a wide enough search space for our inversion. As we mentioned before, enforcing such constraints reduces the velocity model complexity and the inverse problem unsuitability.

$$V_{s,i} \le Vs_{s,i+1} \quad \text{for} \quad i = 1, \ldots, r-1, \qquad V_{s,1} \ge 50, \qquad V_{s,r} \le 5000. \tag{16}$$

For estimation of the small-strain damping, we consider the range $0.001 \le \xi \le 0.1$, whereas the density and Poisson's ratio

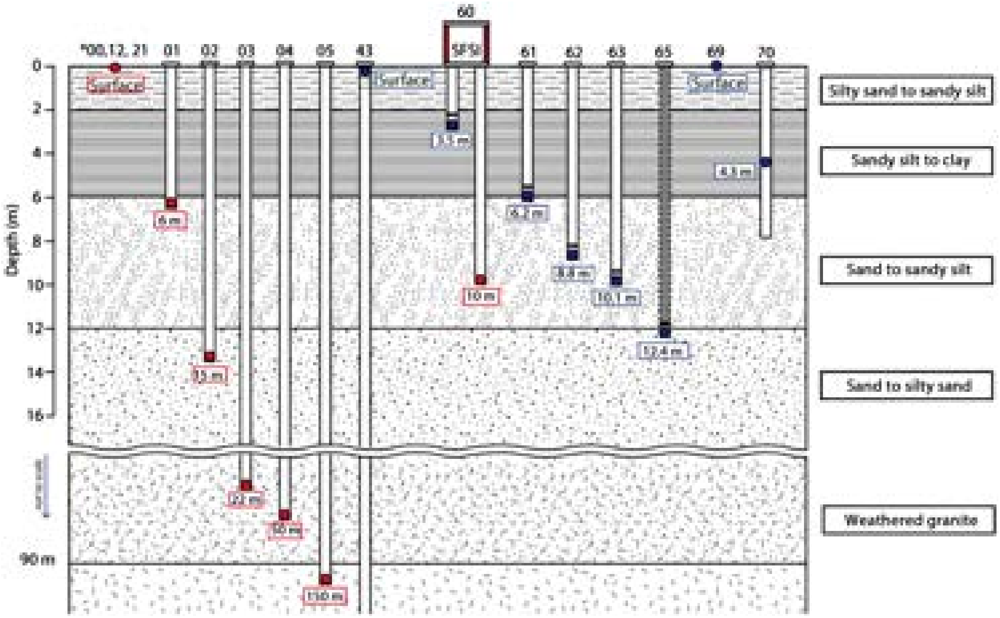


Figure 1. Garner Valley downhole array (GVDA) cross section (see Data and Resources). SFSI, soil-foundation-structure interaction. The color version of this figure is available only in the electronic edition.

for each layer are assumed to be constant and equal to the values we use in our forward model simulation: $\rho = 1800\ \mathrm{kg/m^3}$, Poisson's ratio $\nu = 0.3$.

We next initiate the algorithm by creating the initial ensemble of particles. We draw particles from the uniform distributions for both V_S and damping, and we scale them appropriately as follows:

$$P(V_{S,i}) = \sqrt{\frac{z_i}{H}}[5 + 10U(0,1)], \qquad P(\xi) = 0.1 + 0.3U(0,1), \tag{17}$$

in which $U(0,1)$ is the uniform distribution between 0 and 1 and z_i is the depth of the bottom of layer i. Then, we use the projection in the following equation to enforce constraints on each particle $u^{(n)}$ that violates the constraints in the initial ensemble:

$$u_0^{(n)} = \mathrm{argmin}_u \frac{1}{2}|u - u^{(n)}|^2 \text{subject to } Au \le a, \tag{18}$$

in which $u^{(n)}$ is before we enforce the constraints, and $u_0^{(n)}$ is its projected counterpart that satisfies the constraints. Figure 2a shows $N = 50$ realizations of the projected V_S profiles drawn from the above distribution, along with the ensemble mean and the target profile. Figure 2b shows the corresponding dispersion curves for each particle, the mean, and the target profile. In the next sections, we use the ensemble Kalman inversion algorithm that we explained in the Inverse Problem section to estimate the V_S and/or damping profiles of a horizontally layered soil. We specifically consider three test cases to generate the data space y for the inversion. These include: only dispersion data (both a complete and an incomplete set), only downhole array data, and fusion of downhole and dispersion data. We only estimate damping for cases that involve downhole array data.

TABLE 1
V_S Profile at Garner Valley Downhole Array (GVDA) Site Used for Synthetic Data Experiments

Layer	Thickness (m)	Shear-Wave Velocity (V_S) (m/s)
1	0.0–18.0	220
2	18.0–64.5	580
3	64.5–150	1,300
4	Half-space	2,600

The values up to depth 100 m are from Gibbs (1989).

Dispersion data as inversion data space

Here, we assume that both active and passive surface-wave testing results are available; the former is usually used to resolve dispersion data at higher frequencies (shallow layers), and the latter is used to constrain dispersion data at long periods (deep layers). The size of the testing arrays and lateral variability in the depth of the bedrock across these arrays may cause difficulties in resolving the dispersion curve continuously for a wide range of frequencies. To examine how data completeness affects the uncertainty of the inverted profile at strong-motion arrays, we define two data sets: a complete and incomplete dispersion curve. In the incomplete data set, dispersion data corresponding to frequencies smaller than 0.3 Hz, and frequencies between 0.45 and 2.35 Hz, are missing.

We next use Algorithm 1 to iteratively update the particles. Figure 3a shows the estimated V_S profiles that are computed from the ensemble mean in the last iteration, using the complete and incomplete data sets, and Figure 3b shows the computed dispersion curves compared against the data sets used as observations. We should mention here that the stopping criterion for all performed inverse analyses is reaching 100 iterations; this number is more than enough to make it so that the mean of the ensemble is stabilized around the reported final estimates. The lack of solution uniqueness is evident in Figure 3; the dispersion curves of both profiles are in excellent agreement with the curve associated with the target profile, whereas the two

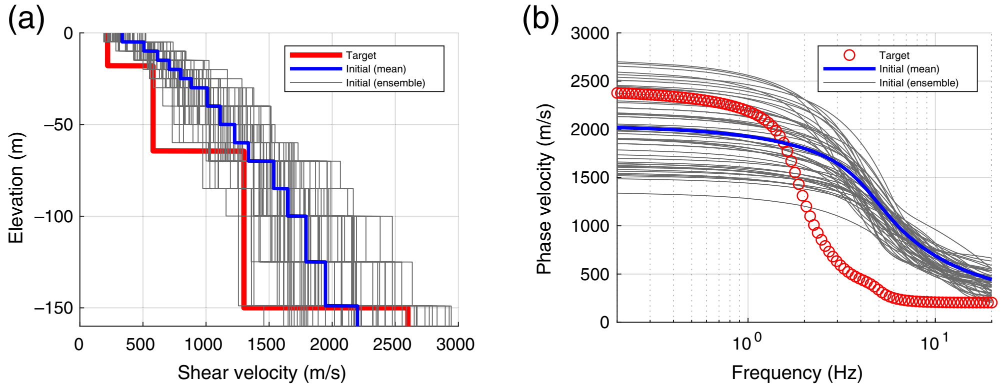


Figure 2. (a) V_S profiles of the initial ensemble, along with the ensemble mean and the target profile—the vertical scale is elevation relative to the ground surface, and (b) dispersion curves computed using the initial ensemble, ensemble mean an target V_S profiles. The color version of this figure is available only in the electronic edition.

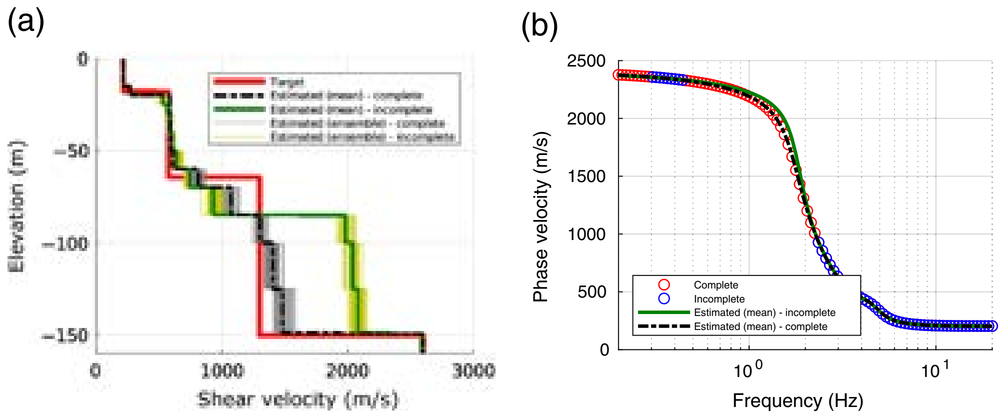


Figure 3. (a) Final estimated ensembles, along with the ensembles' mean and target V_S profile—the vertical scale is elevation relative to the ground surface, and (b) dispersion curves computed using the ensemble mean, along with the complete and incomplete datasets used in the inverse analysis. The color version of this figure is available only in the electronic edition.

inverted profiles show significant differences. Unsurprisingly, for the case of simulated dispersion data, slight deviation is only observed for the dispersion data of the incomplete set, exactly in the frequency range where information is missing.

Downhole array ground-motion recordings as inversion data space

In this section, we repeat the numerical experiment described previously (i.e., estimation of V_S and damping) using only downhole array ground-motion recordings. The forward problem is comprised of propagating the "recorded" (known) motion at depth $z = H$ to the ground surface and minimizing the misfit between the ground surface motion forward predictions and the surface acceleration time series. Downhole arrays are usually instrumented sparsely (e.g., see Fig. 1), which may contribute to the lack of solution uniqueness and certainty. To reflect this issue, we only use the instrument at 150 m depth as input and the ground surface motion as output, which is the most common configuration of a downhole array (e.g., of the Japanese strong-motion network, KiK-net).

To compare with the dispersion data inversion, we first drew the initial particles from the same distribution as in equation (17), but the algorithm was not successful in finding the profile that could reproduce the output data. However, when we slightly shifted the initial ensemble to V_S values closer to the target solution (see the following equation and Fig. 4a), the algorithm successfully converged to a V_S profile and damping ratio that captures the ground surface acceleration. We should mention here that in this case, for which the downhole recorded motion is used as prescribed boundary condition, our forward model is appropriately adjusted to a layered soil on rigid bedrock, in which the thickness of the last layer is 25 m.

$$P(V_{S,i}) = \sqrt{\frac{z_i}{H}}[4 + 10U(0,1)], \qquad P(\xi) = 0.1 + 0.3U(0,1). \tag{19}$$

Figure 4b shows the estimated V_S profile after 100 iterations, and Figure 4c shows the computed surface accelerations using this profile. The final estimate for damping is $\xi = 0.0412$, which is close to the target damping ratio of 0.04 used to generate the synthetic data. Although we see clear differences between the inverted and the target V_S profiles, both predict identical ground surface motion when subjected to the borehole strong-motion record, which once again demonstrates the lack of solution uniqueness when the data space is sparse.

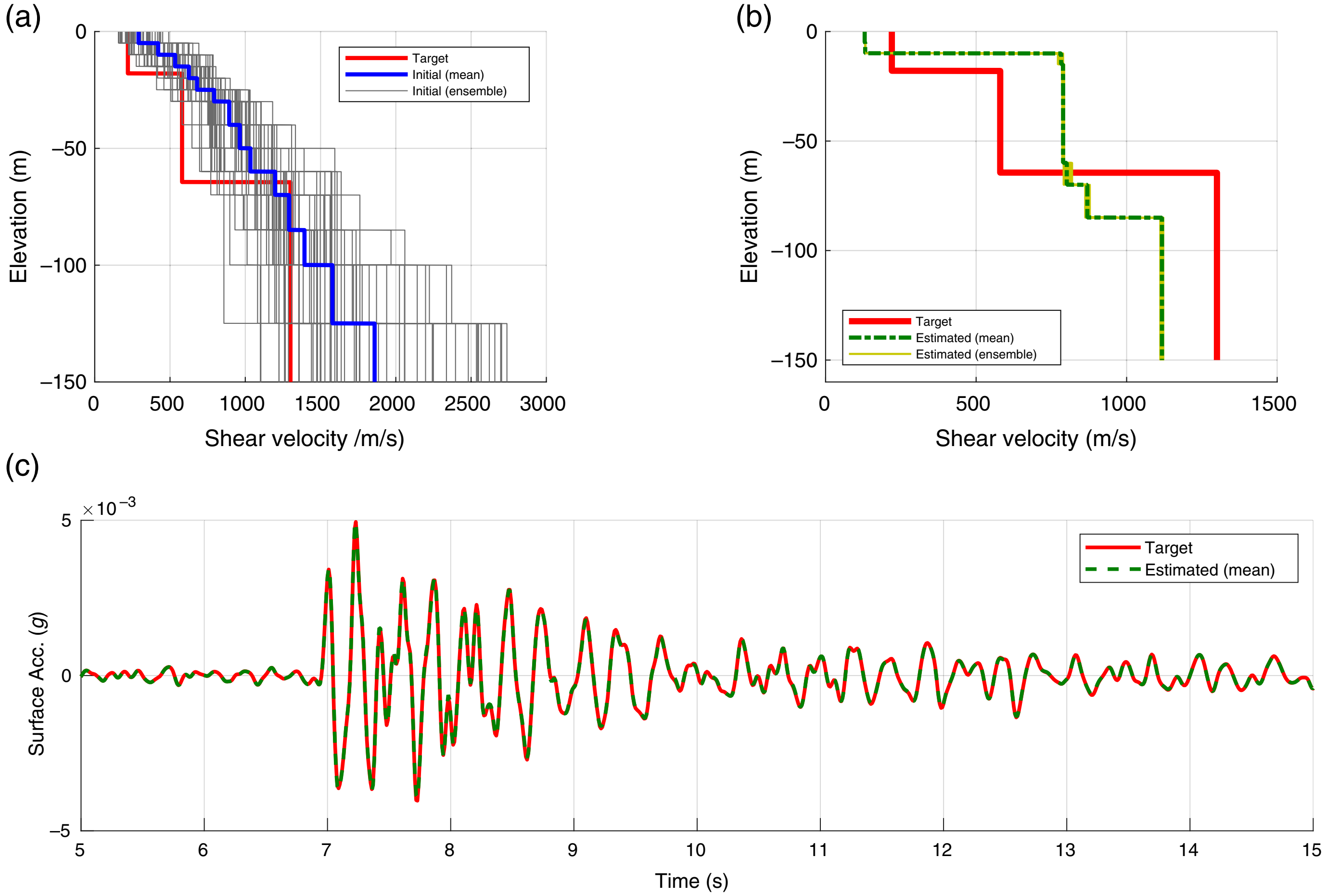


Figure 4. (a) Initial ensemble of particles along with the ensemble mean and target profile—the vertical scale is elevation relative to the ground surface, (b) estimated ensemble of particles along with the ensemble mean and target profile, and (c) computed surface acceleration time series using the target V_S profile and damping, as well as the mean of the estimated ensemble. The color version of this figure is available only in the electronic edition.

Dispersion and downhole array data as joint-inversion data space

So far, we have tried using the two heterogeneous data sets separately. First, we used the dispersion data that provided constraints on the dispersive characteristics of surface waves; then, we used downhole array data, which provided constraints on the travel time of body (shear) waves through the soil layers. As shown in the last two examples, the inversion algorithm could not recover the target profile in either case, whereas all forward models captured the target observations exceptionally well.

In this section, we study the effects of fusing these two complementary data sets. We create the initial ensemble from equation (17), and consider y to be the combination of the incomplete dispersion data and surface acceleration time series. We use $\beta_1 = 0.01$ and $\beta_2 = 0.01$ to define the Gaussian noise covariance matrix. Using the same inversion algorithm, Figure 5 shows the estimated V_S profile and the computed dispersion results. As can be readily seen, by combining the two data sets, we are able to recover the V_S profile with depth. As expected, the profile matches both the incomplete dispersion curve and the acceleration time series on ground surface. The latter is almost identical to the time series shown in Figure 4c and therefore is not repeated here. Furthermore, the algorithm recovers a constant damping ratio of $\xi = 0.0413$; recall that our synthetic example uses a constant damping $\xi = 0.04$ for all layers. This example shows the effectiveness of joint inversion to successfully recover the target V_S profile and damping ratio among possible solutions that can fit the individual observations well in the inverse problem setting. In the next section, we test the algorithm effectiveness for more complex profiles and noise-contaminated synthetic data before proceeding to an example using measured dispersion data and recorded downhole array ground motions.

Profile complexity and noise effects on inverted V_S profiles

As shown in the previous section, joint inversion of downhole array and dispersion data helps to improve the estimation of the V_S profile. So far, however, we have only used noise-free

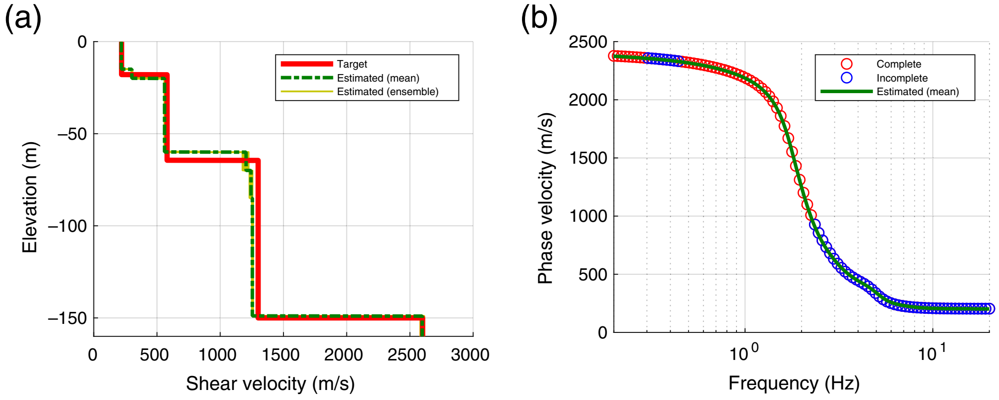


Figure 5. (a) Final estimated ensemble, along with the ensemble mean and target profile—the vertical scale is elevation relative to the ground surface, and (b) dispersion curve using the estimated ensemble mean, along with both complete and incomplete datasets used in the inverse analysis. The color version of this figure is available only in the electronic edition.

synthetic data. Prior to introducing field recorded data, we test the performance of the framework using more complex profiles and noise contaminated synthetic data. It should be noted that the complex profile that we will use in the next example is not intended to be representative of the V_S profile at GVDA; rather, we generate it to assess the framework's capability in dealing with more complex cases.

Profile complexity effects. Here, we use the same framework to estimate the V_S profile for a more complex case. To capture the complex V_S, we use a more refined discretization of the profile; that is, $\Delta h = 5$ m, namely 30 unknown V_S parameters. We again consider four types of data sets:

- complete dispersion data (case 1);
- incomplete dispersion data for which phase velocity values for frequencies $f \leq 0.3$ Hz and $0.46 \leq f \leq 2.83$ Hz are missing (case 2);
- downhole array data only at the surface $z = 0$ (case 3); and
- fusing data sets in cases 2 and 3 (case 4).

For cases 1 and 2, we estimate 31 parameters, including the elastic bedrock V_S and the V_S from zero to H. For case 3, we estimate 31 parameters, including the V_S from zero to H and damping. Then, for case 4, we estimate 32 parameters, including the V_S from zero to H, elastic bedrock V_S, and damping. Figure 6a,b shows the estimated V_S profiles after 100 iterations, along with the resulting dispersion curves. On the other hand, Figure 6c shows the smoothed relative error of the surface acceleration that was computed using different estimated profiles. It should be noted that the absolute error for cases 3 and 4 is very small. Again, we notice that joint inversion of dispersion and downhole array data can significantly improve the estimated profile. Furthermore, increasing the number of soil layers does not affect the performance of the algorithm presented here, and it can correctly resolve the impedance contrasts in the target V_S profile.

Noise effects. To assess the framework robustness when the recorded data is noisy, we next consider case 4, for which we use the combined data sets to estimate V_S profile and damping. For generating the noise contaminated data sets y_1 and y_2, we add a zero mean Gaussian noise, where we use equation (3) to define the covariance matrix considering $\beta_1 = 0.05$ and $\beta_2 = 0.05$. Then, within the ensemble Kalman inversion iterations, we consider $\beta_1 = 0.025$ and $\beta_2 = 0.025$ to define Γ. Figure 7a shows the estimated V_S profile, and Figure 7b,c shows the theoretically computed dispersion curve and surface acceleration time series compared to the data (both clean and noisy); this exercise again shows the successful recovering of the V_S profile and damping in presence of the noisy data.

SITE CHARACTERIZATION: REAL DATA

In this section, we use the same inversion algorithm to estimate the V_S profile and damping at GVDA, using field dispersion data and recorded acceleration time series. For the downhole array data, we use a curated dataset (see Data and Resources), comprising 30 events recorded from 2006 to 2016 with ground surface peak ground acceleration (PGA) greater than 10 Gal. The events are recorded by vertical and aligned horizontal accelerometers from 501 m depth to the surface. From these records, here, we focus on those with magnitude $M_w \leq 5$ to minimize the likelihood of any nonlinear response contaminating the recorded ground motions. In total, we consider 23 events in what follows. Figure 8 shows the epicenter of the events, along with the PGA and peak ground velocity (PGV) as a function of the epicentral distance.

As dispersion data set, we use the mean field measurements associated with the north accelerometer location 00 by Teague *et al.* (2018), who performed both active source multichannel analysis of surface waves and passive source microtremor array measurement testing at GVDA. Using the downhole array data $a_e = (a^e_{00}, a^e_{01}, a^e_{02}, a^e_{03}, a^e_{04})$ for each event e and dispersion data V_r, the observation can be formed as $y_e = (a_e, V_r)$ for $e = 1, \ldots, 23$. For each event, we use the inversion algorithm to estimate both $V_{S,e}$ and damping ξ_e. Please note that a_{00}, a_{01}, a_{02}, a_{03}, a_{04}, a_{05} are accelerations recorded at $z = 0$, 6,

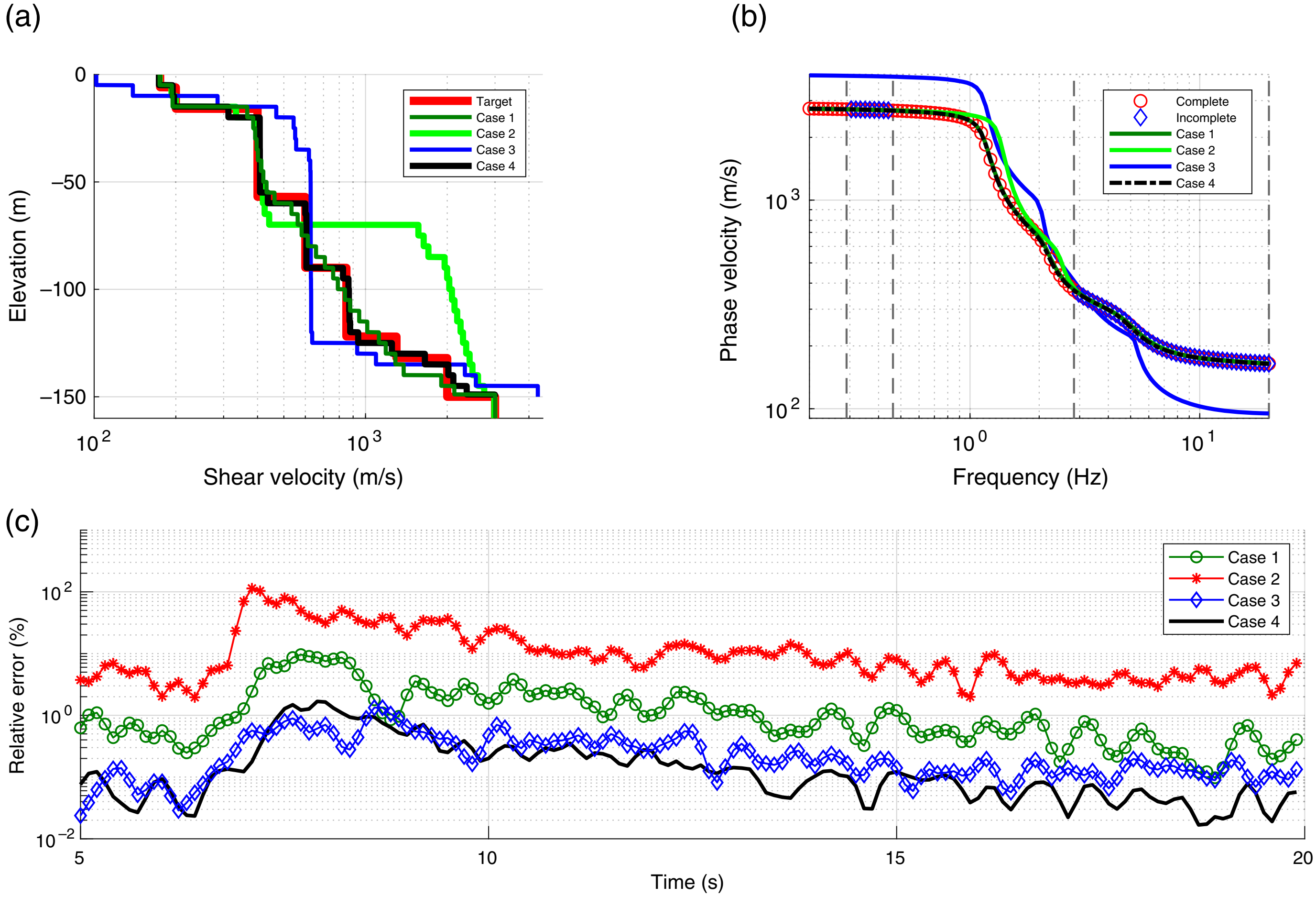


Figure 6. (a) Final estimate of the ensembles' mean, along with the target V_S profile—the vertical scale is elevation relative to the ground surface, (b) dispersion curves computed using the ensembles' mean, along with the complete and incomplete datasets used in the inverse analysis, and (c) smoothed relative error between the surface acceleration dataset and those computed using the estimated ensembles' mean. The color version of this figure is available only in the electronic edition.

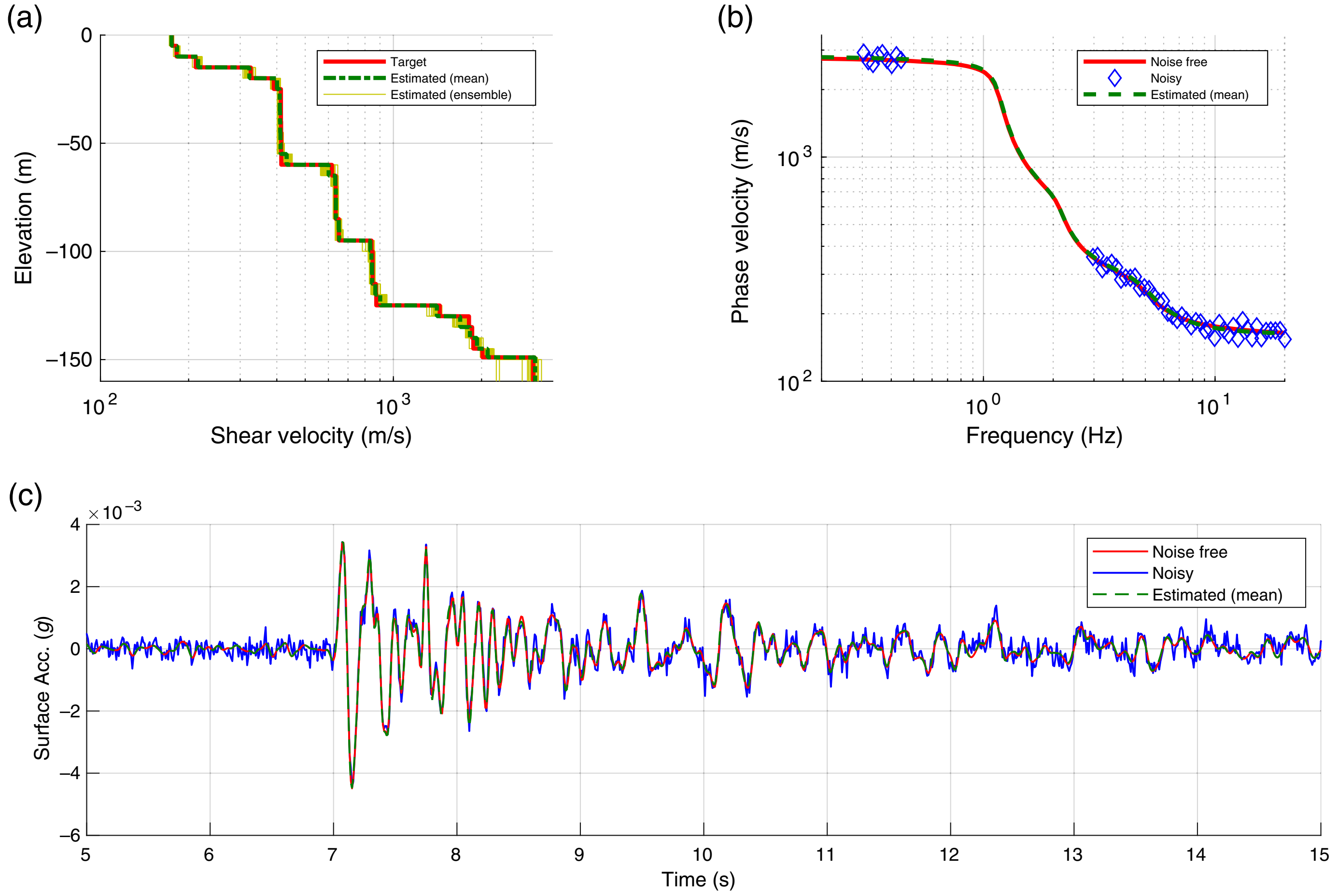
15, 22, 50, and 150 m, respectively (see Fig. 1). All acceleration time series were filtered using second-order band-pass Butterworth filter with frequency range [0.1, 10] Hz.

To start the inversion, we consider three initial ensembles with 50 particles each (see Fig. 9a–c), and we seek to estimate the V_S and damping for $23 \times 3 = 69$ cases. Doing so will allow us to assess the effects of using different prior distributions on the resulting profiles. For profile discretization, we consider uniform layers with thickness of 5 m. It is worth mentioning that in what follows, we will also study the effects of considering thicker layers for profile discretization. Figure 9d–f shows the final estimated ensemble means for each event e and each initial ensemble i, along with the average V_S profile $V_S^{\text{avg},i}$.

$$V_S^{\text{avg},i} = \frac{1}{23}\sum_{e=1}^{23} V_{S,e}^{i}, \qquad i = 1, 2, 3. \tag{20}$$

With the exception of small discrepancies at depths larger than 100 m, the profiles recovered through consideration of each prior and the corresponding ensemble are almost identical. Figure 10 shows the box plot of the estimated V_S at each layer, one for each of the prior distributions that we used to select our particle ensembles. Again, the trends in recovered V_S and associated uncertainty are nearly identical for all three prior V_S distributions. The resulting damping ratios (i.e., the ensemble mean for each e and i) are shown in Figure 11 as a function of the event magnitude M_w, PGA, and PGV; the average damping is $\xi^{\text{avg}} = 0.049$ over 69 events, and it appears that higher damping estimation is associated with higher ground-motion intensities, which is very likely the manifestation of nonlinear response.

To evaluate the effectiveness of joint inversion, we next use the second initial ensemble ($i = 2$) to estimate the V_S profile, using only the dispersion data and only the downhole array data. Figure 12a shows the estimated V_S profiles V_S^{dsp} and V_S^{dh} for dispersion and downhole array data, respectively; V_S^{avg} is the average over all 69 events that we used to



search the fused data space. As shown, V_S^{dh} closely follows the V_S^{avg} whereas V_S^{dsp} starts deviating after z around 30 m. Figure 12b,c shows the effects of these inverted V_S profiles on the site signatures, that is, dispersion curves and site -ransfer functions (TFs). The empirical TF for each event is obtained by dividing the Fourier transform of the recorded acceleration signal on the surface $a_{00}^{e,\dagger}$ by that of sensor $a_{05}^{e,\dagger}$ at depth $z = H = 150$ m. Figure 12c shows the median TF

Figure 7. (a) Final estimated ensemble, along with the ensemble mean and target V_S profile—the vertical scale is elevation relative to the ground surface, (b) dispersion curve computed using the estimated ensemble mean, along with the noise free and incomplete noisy dispersion data used in the inverse analysis, and (c) surface acceleration time series computed using the estimated ensemble mean, along with the noise free and noise contaminated acceleration data used in the inverse analysis. The color version of this figure is available only in the electronic edition.

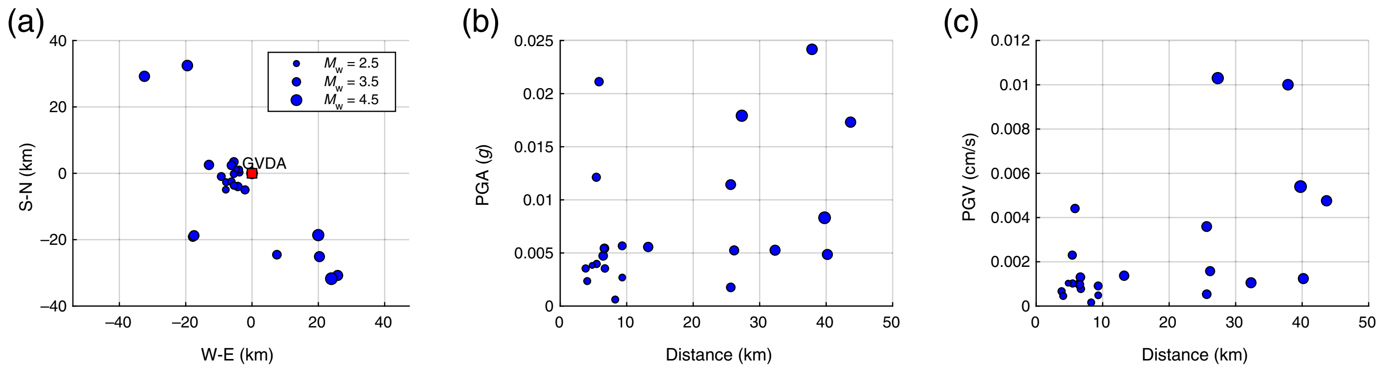


Figure 8. (a) Epicentral distance, (b) peak ground acceleration (PGA), and (c) peak ground velocity (PGV) of the events considered for V_S estimation at GVDA. The color version of this figure is available only in the electronic edition.

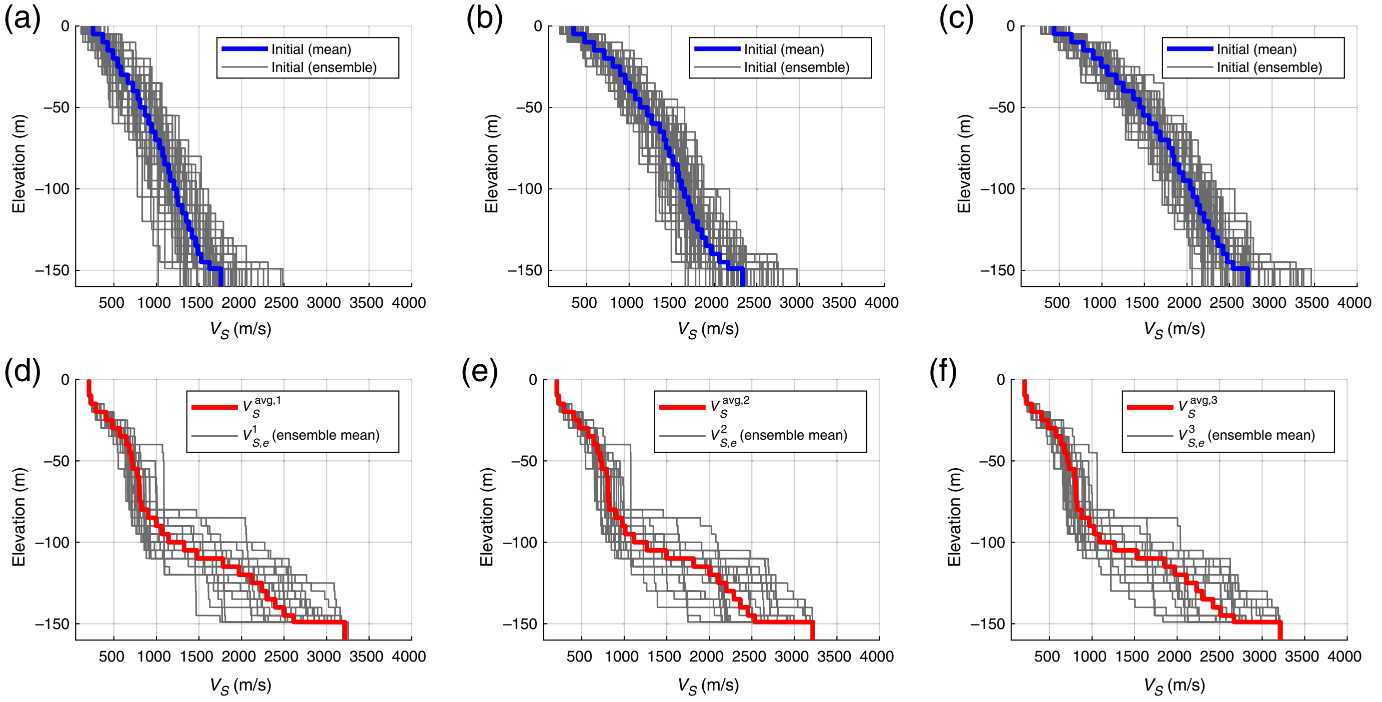


Figure 9. Initial ensemble of particles, along with the ensemble mean: (a) $i = 1$, (b) $i = 2$, and (c) $i = 3$; final estimated ensemble mean for events $e = 1, \ldots, 23$, along with the average V_S profile: (d) $i = 1$, © $i = 2$, and (f) $i = 3$. The vertical scale is elevation relative to the ground surface. The color version of this figure is available only in the electronic edition.

and its standard deviation among the considered 23 events. It is interesting to note that:

1. although V_S^{dh} is following the same trend as V_S^{avg}, the resulting dispersion curve cannot capture the experimental data, while it captures the joint inversion TF quite well.
2. while V_S^{dsp} captures the dispersion data, the associated TF is different from those cases that incorporate downhole array data in the inversion.
3. Joint inversion of the dispersion and downhole acceleration data changes the inverted dispersion data only in the frequency range for which phase velocity data are not available.
4. Although peaks in the theoretical TFs are well aligned with the empirical TF, it is worth noting the possibility that deviations at higher modes are due to the 3D effects.

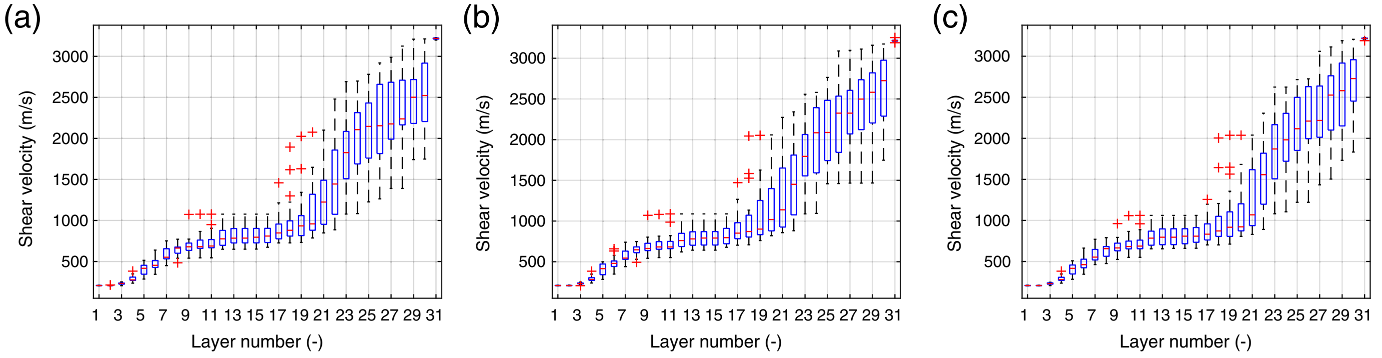


Figure 10. Box plot of the estimated V_S at each layer for the considered a priori distributions: (a) $i = 1$, (b) $i = 2$, and (c) $i = 3$. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the "+" symbol. The color version of this figure is available only in the electronic edition.

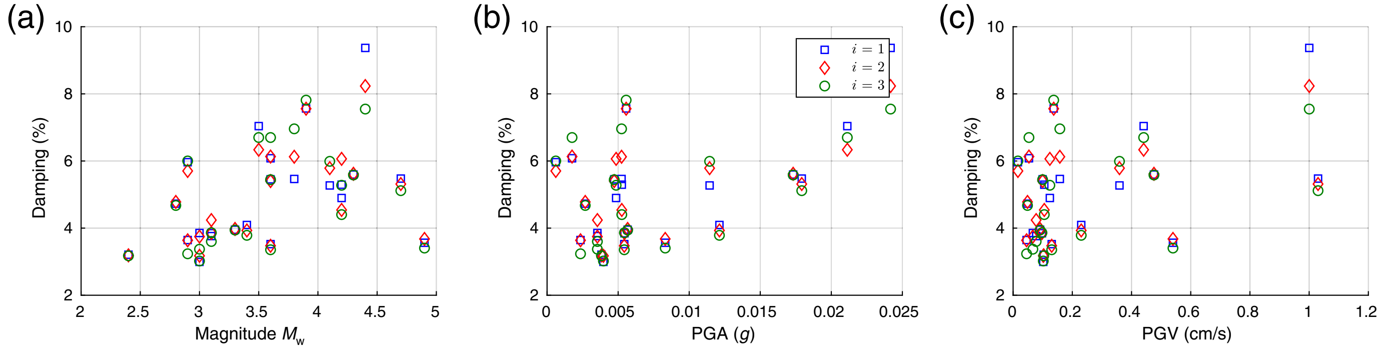


Figure 11. Final estimated ensemble mean of damping ratio for events $e = 1, \ldots, 23$ and as function of (a) event magnitude, (b) event PGA, and (c) event PGV. The color version of this figure is available only in the electronic edition.

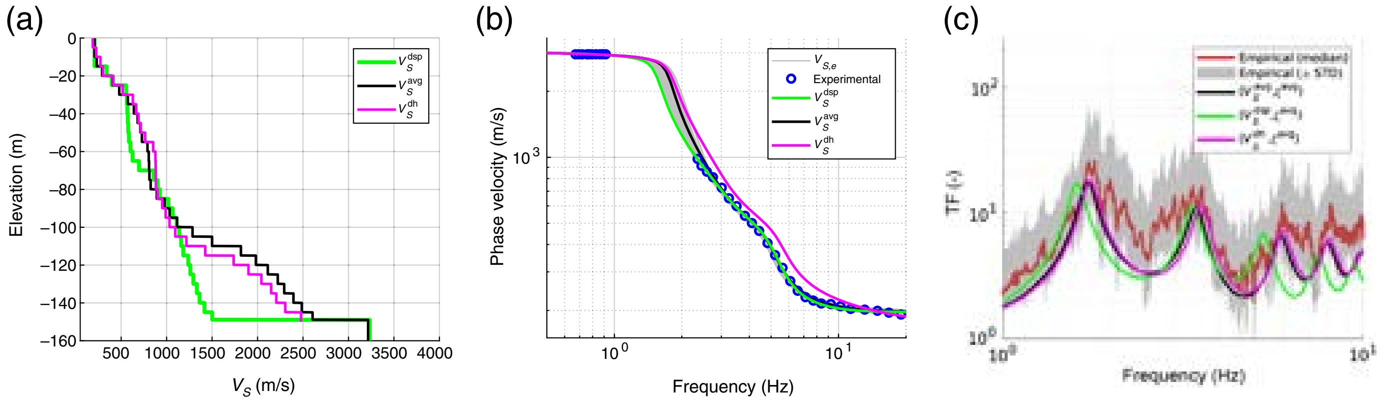


Figure 12. (a) Estimated average V_S profile—the vertical scale is elevation relative to the ground surface, (b) dispersion curves computed using different average V_S profiles, compared against the experimentally available dispersion data used in the inverse analysis, and (c) theoretical transfer functions (TFs) computed using different average V_S profiles, compared against the median of empirical TFs and their standard deviation. The color version of this figure is available only in the electronic edition.

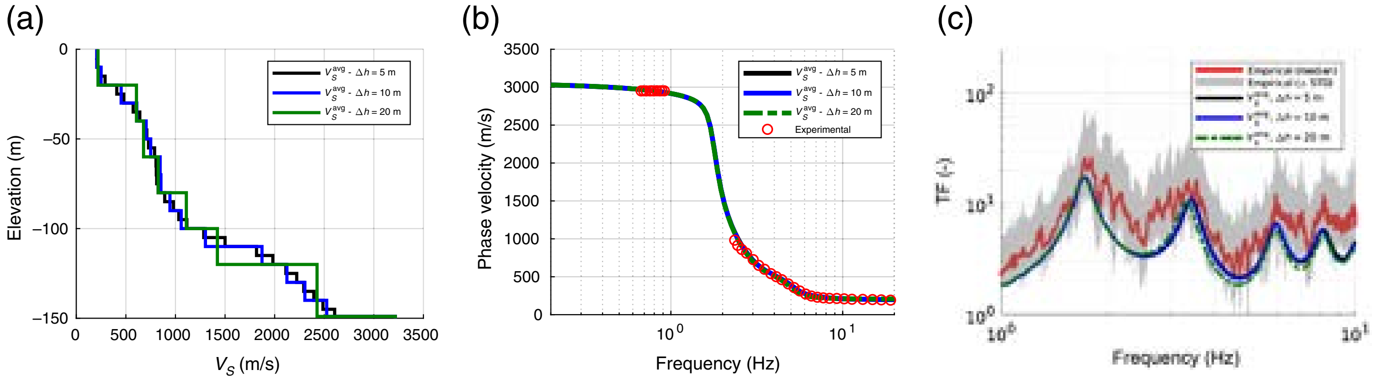


Figure 13. (a) Estimated average V_S profile—the vertical scale is elevation relative to the ground surface, (b) dispersion curves computed using different average V_S profiles, compared against the experimentally available dispersion data used in the inverse analysis, and (c) theoretical TFs computed using different average V_S profiles, compared against the median of empirical TFs and their standard deviation. The color version of this figure is available only in the electronic edition.

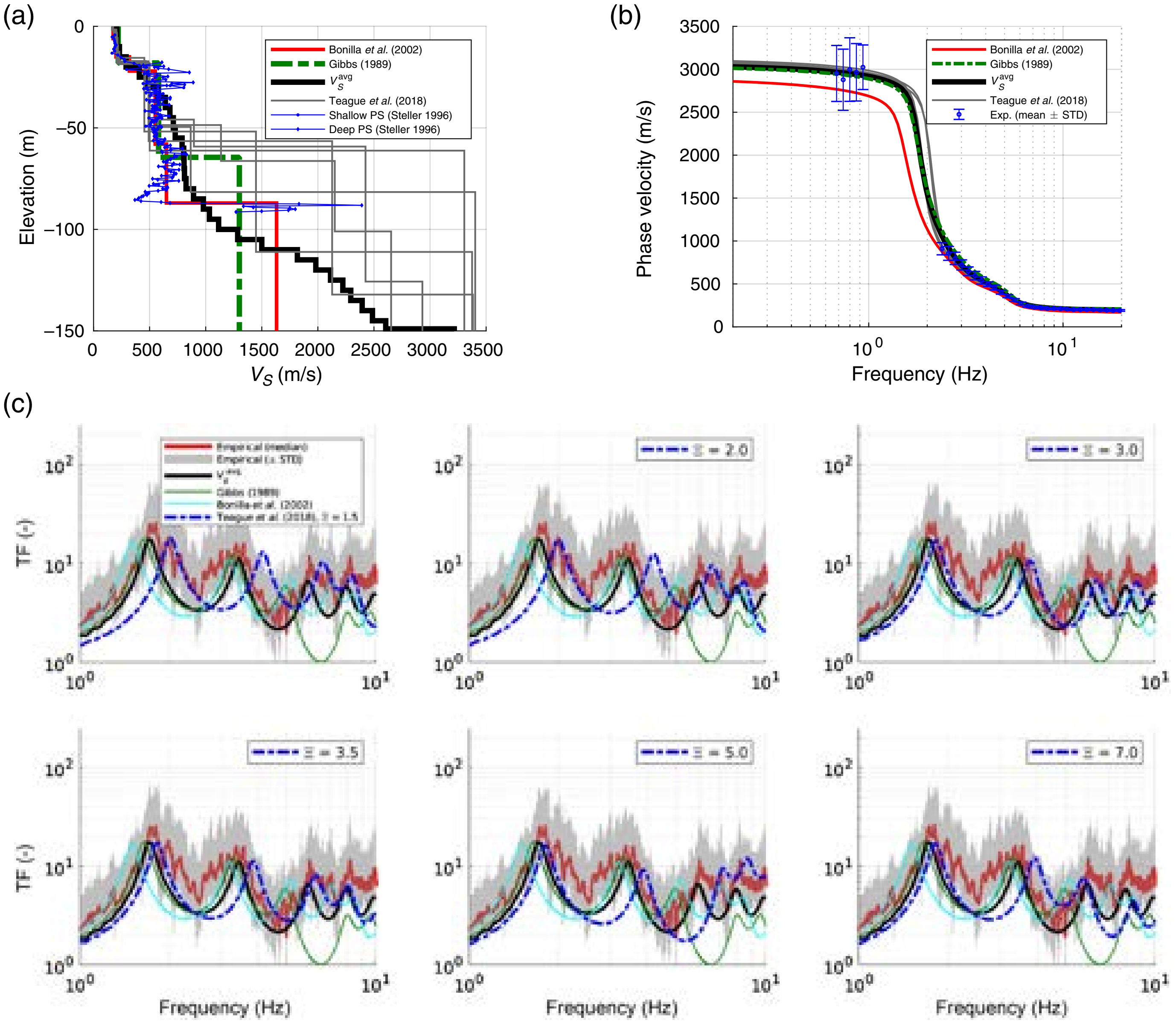


Figure 14. Comparison of (a) V_S profiles (the vertical scale is elevation relative to the ground surface), (b) dispersion curves, and (c) TFs at GVDA. The color version of this figure is available only in the electronic edition.

To assess the effects of the inverse problem parameterization on inverted V_S profiles, we consider two more cases of discretizing the soil height with $\Delta h = 10$ and 20 m and repeating the joint inversion. Figure 13 shows the estimated V_S profiles and associated dispersion curves and TFs. These results suggest that decreasing the layer thickness, that is, increasing the number of V_S values to be estimated, does not necessarily have adverse effects on the suitability of the inverse problem in hand. Furthermore, our previous numerical experiments suggest that fine discretization is successful to capture impedance contrasts, at least when dealing with synthetic data. That being said, it is worth noting that, in many cases, it may not be clear which parameterization produces a "better" answer than the others. Therefore, to fully understand the uncertainties, it may be necessary to consider multiple parameterizations. Also, it is worth noting that our future research is aimed at addressing this issue systematically by estimating both layer thicknesses and V_S values.

Site characterization of Garner Valley has also been the subject of several geotechnical and geophysical studies; therefore, we compare our results to those available in open literature. More specifically, shallow and deep *PS* suspension logging results were provided by Stellar (1996). Gibbs (1989) used a three-component geophone in a 100-m-deep borehole and determined the *P*- and *S*-wave velocities using the conventional methods of travel-time plots and straight-line segments. On the other hand, Bonilla *et al.* (2002) used a trial-and-error approach to find the best V_S profile that minimizes the misfit between the synthetic and real downhole accelerometer time series for weak motions; they used the velocity models by Gibbs (1989); Pecker and Mohammadioum (1993) as the initial guess. Finally, Teague *et al.* (2018) used the active and passive surface-wave measurements to invert for the V_S profile using the dispersion data. They also used the HVSR curves to further constrain the inversion results by comparing the first-mode frequency of theoretical TFs obtained from the inverted V_S profiles against the mean first-mode frequency obtained from the experimental HVSR curves. These results, along with those computed in this study, are shown in Figure 14 in terms of V_S profile, theoretical dispersion curves, and TFs. In all cases except for that of Teague *et al.* (2018), we assume that $\nu = 0.45$ and $\rho = 1800\ \mathrm{kg/m}^3$. For the results attained by Teague *et al.* (2018), we use the data provided by the first author to compute the dispersion curves and TFs considering layering ratio Ξ ranging from 1.5 to 7. Furthermore, for all computed TFs, we assume that the damping ratio is equal to the estimated average damping ratio; that is, $\xi^{\mathrm{avg}} = 0.049$. Again, and as correctly identified by Teague *et al.* (2018), it is evident that a lack of dispersion data in the moderate frequency range has resulted in discrepancies among the interpreted V_S profiles. Moreover, as shown in Figure 14c, TFs of coarser profiles with $\Xi = 3, 3.5, 5, 7$ are in better agreements with those associated with this study.

CONCLUSION

In this article, we introduced a sequential data assimilation approach based on ensemble Kalman inversion for near-surface site characterization. Our method was based on heterogeneous data set fusion and assimilation with prior knowledge, which we introduced in terms of inequality constraints to the V_S and/or small strain damping. To characterize the general trend of the V_S profile, we used the piece-wise constant function with a known number of layers. Through a series of synthetic experiments, we demonstrated the lack of inverse problem solution uniqueness when dispersion data or acceleration time series were used in isolation and showed how the joint inversion of these complementary data could improve the V_S estimation. We also showed that increasing the number of layers can help capture more complex profiles without affecting the performance of the algorithm. Finally, we tested the algorithm on real data using the GVDA site as our testbed, and we compared the inverted V_S profile against previous site characterization studies. Our study showed that inversion uncertainties, such as the ones described by Teague *et al.* (2018), may be attributed to incomplete dispersion data in the medium frequency range. Future noninvasive testing that will help complete the available dispersion data across the entire frequency range of interest will help refine inverse algorithms, such as the one presented here.

DATA AND RESOURCES

The Garner Valley downhole array (GVDA) data are available at http://nees.ucsb.edu/curated-datasets, last accessed April 2020, and Figure 1 is downloaded from http://nees.ucsb.edu/facilities/GVDA (last accessed March 2020). The theoretical dispersion curves are computed using the GEOPSY software package gpdc installed from http://www.geopsy.org/download.php?platform=src&branch=testing&release=3.1.1 (last accessed April 2020). Python code for performing ensemble Kalman inversion with constraints will be released at the GitHub account of the first author (https://github.com/elnaz-esmaeilzadeh, last accessed April 2020).

ACKNOWLEDGMENTS

The authors would like to thank Editor-in-Chief Thomas Pratt, Associate Editor Sherly Molnar, and reviewers Jan Dettmer and David Teague for helpful comments and revisions that improved the article. In addition, the authors would like to thank David Teague for making the experimental dispersion data and inverted V_S profiles in Teague *et al.* (2018) available during the review process.

REFERENCES

Albers, D. J., P.-A. Blancquart, M. E. Levine, E. E. Seylabi, and A. M. Stuart (2019). Ensemble Kalman methods with constraints, *Inverse Prob.* **35,** doi: 10.1088/1361-6420/ab1c09.

Arai, H., and K. Tokimatsu (2005). S-wave velocity profiling by joint inversion of microtremor dispersion curve and horizontal-to-vertical (h/v) spectrum, *Bull. Seismol. Soc. Am.* **95,** no. 5, 1766–1778.

Assimaki, D., W. Li, and A. Kalos (2011). A wavelet-based seismogram inversion algorithm for the in situ characterization of nonlinear soil behavior, *Pure Appl. Geophys.* **168,** no. 10, 1669–1691.

Assimaki, D., J. Steidl, and P. C. Liu (2006). Attenuation and velocity structure for site response analyses via downhole seismogram inversion, *Pure Appl. Geophys.* **163,** no. 1, 81–118.

Bonilla, L. F., J. H. Steidl, J.-C. Gariel, and R. J. Archuleta (2002). Borehole response studies at the garner valley downhole array, southern California, *Bull. Seismol. Soc. Am.* **92,** no. 8, 3165–3179.

Chada, N. K., A. M. Stuart, and X. T. Tong (2019). Tikhonov regularization within ensemble Kalman inversion, available at http://arxiv.org/abs/1901.10382 (last accessed April 2020).

Chandra, J., P. Guéguen, J. H. Steidl, and L. F. Bonilla (2015). In situ assessment of the g–γ curve for characterizing the nonlinear response of soil: Application to the garner valley downhole array and the wildlife liquefaction array, *Bull. Seismol. Soc. Am.* **105,** no. 2A, 993–1010.

Chen, Y., and D. S. Oliver (2012). Ensemble randomized maximum likelihood method as an iterative ensemble smoother, *Math. Geosci.* **44,** no. 1, 1–26.

Darendeli, M. B. (2001). Development of a new family of normalized modulus reduction and material damping curves, *Ph.D. Thesis*, University of Texas at Austin, Texas.

Dunkin, J. W. (1965). Computation of modal solutions in layered, elastic media at high frequencies, *Bull. Seismol. Soc. Am.* **55,** no. 2, 335–358.

Emerick, A. A., and A. C. Reynolds (2013). Investigation of the sampling performance of ensemble-based methods with a simple reservoir model, *Comput. Geosci.* **17,** no. 2, 325–350.

Evensen, G. (2009). *Data Assimilation: The Ensemble Kalman Filter*, Springer Science & Business Media, Berlin, Germany.

Foti, S. (2000). Multistation methods for geotechnical characterization using surface waves, *Ph.D. thesis*, Politecnico di Torino, Italy.

Foti, S., C. Comina, D. Boiero, and L. Socco (2009). Non-uniqueness in surface-wave inversion and consequences on seismic site response analyses, *Soil Dynam. Earthq. Eng.* **29,** no. 6, 982–993.

Garofalo, F., S. Foti, F. Hollender, P. Bard, C. Cornou, B. Cox, A. Dechamp, M. Ohrnberger, V. Perron, D. Sicilia, *et al.* (2016). Interpacific project: Comparison of invasive and non-invasive methods for seismic site characterization. Part ii: Inter-comparison between surface-wave and borehole methods, *Soil Dynam. Earthq. Eng.* **82,** 241–254.

Gibbs, J. F. (1989). Near-surface p and s-wave velocities from borehole measurements near lake Hemet, California, *U.S. Geol. Surv. Open-File Rept. 89-630*, doi: 10.3133/ofr89630.

Haskell, N. A. (1953). The dispersion of surface waves on multilayered media, *Bull. Seismol. Soc. Am.* **43,** no. 1, 17–34.

Iglesias, M. A., K. J. Law, and A. M. Stuart (2013). Ensemble Kalman methods for inverse problems, *Inverse Prob.* **29,** no. 4, 045001.

Knopoff, L. (1964). A matrix method for elastic wave problems, *Bull. Seismol. Soc. Am.* **54,** no. 1, 431–438.

Lunedei, E., and P. Malischewsky (2015). A review and some new issues on the theory of the H/V technique for ambient vibrations, in *Perspectives on European Earthquake Engineering and Seismology*, A. Ansal (Editor), Vol. 39, 371–394, Springer, Cham, Switzerland.

Molnar, S., J. Cassidy, S. Castellaro, C. Cornou, H. Crow, J. Hunter, S. Matsushima, F. Sanchez-Sesma, and A. Yong (2018). Application of microtremor horizontal-to-vertical spectral ratio (MHVSR) analysis for site characterization: State of the art, *Surv. Geophys.* **39,** no. 4, 613–631.

Molnar, S., S. E. Dosso, and J. F. Cassidy (2010). Bayesian inversion of microtremor array dispersion data in southwestern British Columbia, *Geophys. J. Int.* **183,** no. 2, 923–940.

Pecker, A., and B. Mohammadioum (1993). Garner valley: Analyse statistique de 218 enregistrements sismiques, *Proc. Of the 3eme Colloque National AFPS*, Saint-Remy-les-Chevreuse, France (in French).

Phillips, C., and Y. M. Hashash (2009). Damping formulation for nonlinear 1d site response analyses, *Soil Dynam. Earthq. Eng.* **29,** no. 7, 1143–1158.

Piña-Flores, J., M. Perton, A. Garca-Jerez, E. Carmona, F. Luzón, J. C. Molina-Villegas, and F. J. Sánchez-Sesma (2016). The inversion of spectral ratio H/V in a layered system using the diffuse field assumption (DFA), *Geophys. J. Int.* **208,** 577–588, doi: 10.1093/gji/ggw416.

Seylabi, E. E., and D. Asimaki (2018). Bayesian estimation of nonlinear soil model parameters: Theory and field-scale validation at KiK-net downhole array sites, *Proc. 16th European Conference on Earthquake Engineering*, Thessaloniki, Greece, 18–21 June 2018.

Socco, L. V., and D. Boiero (2008). Improved Monte Carlo inversion of surface wave data, *Geophys. Prospect.* **56,** no. 3, 357–371.

Stellar, R. (1996). New borehole geophysical results at GVDA, *NEES@UCSB Internal Rept.*, http://nees-dev.nees.ucsb.edu/sites/default/files/facilities/docs/GVDA-Geotech-Stellar1996.pdf (last accessed April 2020).

Stokoe, K., II (1994). Characterization of geotechnical sites by SASW method, in *ISSMFE Technical Committee #10 for XIII ICSMFE, Geophysical Characterization of Sites*, A. A. Balkema Publishers, Rotterdam, The Netherlands, 15–25.

Teague, D. P., and B. R. Cox (2016). Site response implications associated with using non-unique vs profiles from surface wave inversion in comparison with other commonly used methods of accounting for vs uncertainty, *Soil Dynam. Earthq. Eng.* **91,** 87–103.

Teague, D. P., B. R. Cox, and E. M. Rathje (2018). Measured vs. predicted site response at the Garner valley downhole array considering shear wave velocity uncertainty from borehole and surface wave methods, *Soil Dynam Earthq. Eng.* **113,** 339–355.

Thomson, W. T. (1950). Transmission of elastic waves through a stratified solid medium, *J. Appl. Phys.* **21,** no. 2, 89–93.

Wathelet, M. (2005). Array recordings of ambient vibrations: Surface-wave inversion, *Ph.D. Dissertation*, Liége University, 161 pp.

Wathelet, M., D. Jongmans, and M. Ohrnberger (2004). Surface-wave inversion using a direct search algorithm and its application to ambient vibration measurements, *Near Surf. Geophys.* **2,** no. 4, 211–221.

Manuscript received 11 October 2019