

## Model-data fusion for spatial and statistical characterization of soil parameters from geophysical measurements



Siddharth S. Parida<sup>a</sup>, Kallol Sett<sup>a,\*</sup>, Puneet Singla<sup>b</sup>

<sup>a</sup> Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

<sup>b</sup> Department of Aerospace Engineering, The Pennsylvania State University, University Park, PA, USA

### ARTICLE INFO

#### Keywords:

Geotechnical site characterization  
Geophysical methods  
Full waveform inversion  
Finite elements  
Uncertainty quantification  
Random field  
Minimum variance estimate  
Stochastic collocation  
Validation

### ABSTRACT

A recently developed PDE-constrained stochastic inverse analysis algorithm for spatial and statistical characterization of soil parameters from geophysical measurements, considering uncertainty due to limited measurements and sensor noise, is exemplified and validated. A 60m × 60 m geotechnical site in Garner Valley, CA is used as the validation testbed. Advanced geophysical test measurements – in terms of velocity waveforms at a few locations on the surface due to surface excitations using a mobile shaker – are available for the site. The algorithm inversely analyzes the available measurements to probabilistically estimate the elastic parameters of the soil at the site up to a depth of 40 m. The algorithm relies on (1) hypothesizing the soil parameters to be heterogeneous, anisotropic random fields, (2) making prior assumptions on them, (3) numerically simulating the geophysical experiment using the finite element method in conjunction with a stochastic collocation approach, and (4) fusing simulated measurements with experimental measurements using a minimum variance framework to update the prior assumptions on the soil parameter random fields. The estimated elastic parameters of the soil are presented in terms of marginal mean and marginal standard deviation profiles of the soil's P- and S-wave velocities as well as their correlation structures in the *x*-, *y*-, and *z*-direction. In ascertaining the accuracy of the inverse analysis algorithm, the geophysical experiment is numerically re-simulated with the estimated P- and S-wave velocity profiles and the model predicted velocity waveforms are compared against the field observations at all the measurement locations. Comments are made at appropriate places regarding several aspects of the algorithm in highlighting the lessons learned through this validation effort towards accurate stochastic full waveform inversion of geophysical measurements.

### 1. Introduction

Estimation of soil parameters at any geotechnical sites using geophysical measurements typically relies on some kind of inverse analysis. Existing analysis techniques range from simplified but widely used spectral analysis of surface waves (SASW; [1]) to high-fidelity partial differential equation (PDE) constrained full waveform inversion (FWI; [2]) technique. The SASW approach analyzes only the surface waves and yields approximate, layered profiles of the S-wave velocity of soils. The FWI technique, on the other hand, analyzes all types of waves that result from a geophysical experiment. It is computationally more expensive, but yields more accurate estimates of the spatial variability of S- as well as P-wave velocities of soils at any sites. All the existing techniques, however, are deterministic in nature and can not account for uncertainty due to limited measurements and any measurement error, both of which are inevitable in characterizing any geotechnical

sites using geophysical measurements.

To overcome the drawbacks of the deterministic analysis techniques, the authors, recently, developed a scalable computational approach to perform PDE-constrained FWIs of geophysical measurements in the probability space by considering the main sources of uncertainties in the soil parameter estimation process [3]. Hypothesizing the soil properties to be three-dimensional, heterogeneous, anisotropic, non-Gaussian random fields, the developed approach utilized a Gaussian mixture model (GMM) in conjunction with the generalized polynomial chaos (gPC) to approximate the random field soil properties with a finite number of random variables. It, then, employed a minimum variance framework to fuse sparse geophysical measurements with predictions from a stochastic finite element model of the geophysical experiment in estimating probabilistic characteristics of the soil properties. A non-product quadrature method known as the conjugate unscented transformation (CUT) technique was used to reduce

\* Corresponding author.

E-mail addresses: [sparida@buffalo.edu](mailto:sparida@buffalo.edu) (S.S. Parida), [kallolse@buffalo.edu](mailto:kallolse@buffalo.edu) (K. Sett), [psingla@psu.edu](mailto:psingla@psu.edu) (P. Singla).

<https://doi.org/10.1016/j.soildyn.2019.05.010>

Received 21 March 2018; Received in revised form 12 March 2019; Accepted 6 May 2019

Available online 28 May 2019

0267-7261/ © 2019 Elsevier Ltd. All rights reserved.

the computational burden associated with the stochastic finite element simulation and with the minimum variance framework for model-data fusion. The algorithm was verified using synthetic measurements from a fictitious geophysical experiment.

While the approach developed in our prior work can be used to efficiently compute the statistical information (such as the mean or covariance) associated with the model output or a quantity of interest, for a real-life model-data fusion one often needs to worry about proper field alignment while computing the weighted average from a finite number of samples. A simple average of many ensembles of a model output at each spatial grid point may not preserve the physical characteristics (e.g., periodicity) of the model output. This issue often arises while conducting ensemble average in meteorological problems. For example, the ensemble average may not show physical flow characteristics (such as the vortex) although each ensemble member include physical flow characteristics [4,5]. To motivate for this point, let us consider the free vibration of a spring mass system. It is well-known that the natural frequency of a spring mass system depends on the spring constant and the mass of the system. For a given value of the spring constant, the displacement of mass is a periodic function and the trajectory of the system in phase plane (displacement vs. velocity) is an ellipse. Fig. 1(a) shows the phase plane trajectories for a spring mass system for different samples of spring constant assumed to be a uniform random variable over prescribed bounds. The ensemble average of these phase plane trajectories at each time is shown in Fig. 1(b) and does not depict the periodic characteristic of each realization of spring-mass system. Although the ensemble average is not required to depict the physical characteristic of individual realization, it is desired to have a mathematical representation in which the ensemble average preserves the physical system behavior. For example, it is a common practice to represent motion in a central force field by using conic section parameters (known as orbit parameters) rather than cartesian coordinates so that ensemble average also leads to a conic section. Similarly, one can preserve the physical characteristic of the spring-mass system output by taking the ensemble average in the frequency domain: Fig. 1(c) shows the phase plane trajectory for the spring-mass system corresponding to the ensemble average in the frequency domain. In general, one needs to take the average of special feature parameter in a model output to preserve the physical characteristic of the model output.

In this work, we have exploited the spectral content of a geophysical experiment output to define the feature vector while computing different statistical information (e.g., mean and variance) to conduct the inverse analysis. We have tested the effectiveness of the developed algorithm when applied to a real geotechnical site in order to qualify and overcome the challenges that such a problem poses.

A 60 m × 60 m parcel of the NEES@UCSB site at Garner Valley, CA

is chosen as the validation site. For the site, measurements, in the form of velocity waveforms, from an advanced geophysical experiment, performed by Bielak et al. [6], are available. We have stochastically inverted the geophysical measurements in probabilistically estimating the P- and S-wave velocities of the soil at the site up to a depth of 40 m. Fig. 2 outlines our basic approach for the soil properties estimation and validation of the estimates at the testbed site. The soil parameter estimation approach, in general, follows our previous work, but with a major difference that the model-data fusion is performed in the frequency domain to preserve the dynamical characteristics of the soil continuum in the ensemble averages of the model output; we have considered the spectral feature vector to be defined by the two dominant frequencies of the velocity waveforms at the measurement locations and their corresponding Fourier amplitudes. Moreover, while making the prior assumptions on the parameters of the GMM, which is used to approximate the soil properties, we have tried two different avenues. We have first tried with assumptions based purely on engineering judgment following local geological information and bore-hole wave velocity measurements available at a few locations close to the site. We then have modified the prior assumptions following a set of correlation information obtained from a probabilistic sensitivity analysis. It is observed that the judicious selection of prior values of the GMM parameters, following the outcome of the sensitivity analysis, yields more accurate estimates of the soil properties. The estimates are validated by numerically simulating the geophysical experiment with the soil properties and the comparing the model predicted velocity waveforms with the experimental observations at all the measurement locations.

## 2. Descriptions of the Geotechnical Site and Geophysical Experiment

We have used a well-studied geotechnical site in Garner Valley, California as the validation testbed. The site primarily consists of soft alluvium – silty sand, sand, and clayey sand – up to a depth of approximately 15–25 m; the soft soil is underlain by gravely sand and weathered granite [7]. The bedrock is located approximately 90 m below the surface [8]. Due to the proximity of the site to Lake Hemet, the ground water table is expected to be close to the surface with some seasonal variation; a couple of PS suspension logging and cone penetration tests, conducted at the site in January 1996, measured it to be approximately 6 m below the ground surface [9].

This study analyzes measurements from an advanced geophysical experiment, conducted by Bielak et al. [6], on a 60 m × 60 m parcel (latitude: 33°40.127' N, longitude: 116°40.427' W) of the site on March 13, 2012. Fig. 3 shows the plan view schematic of the geophysical

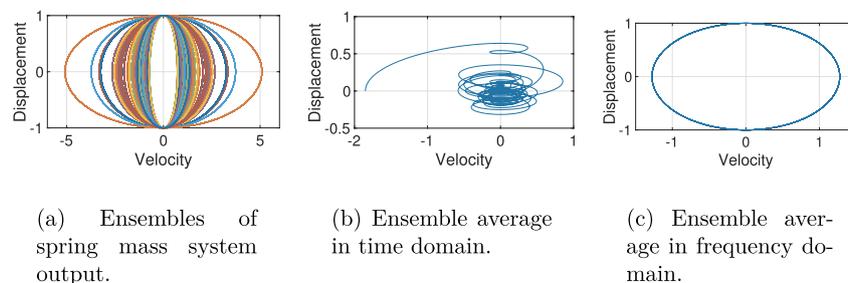


Fig. 1. Ensemble average in time domain vs. frequency domain.

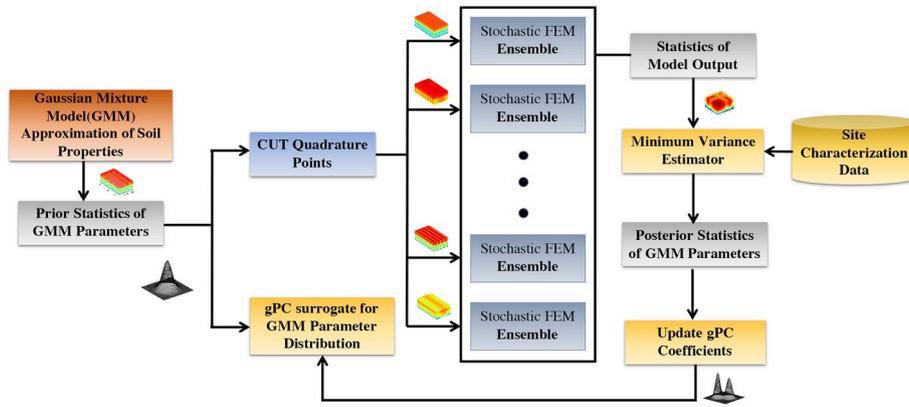


Fig. 2. Framework for soil properties estimation.

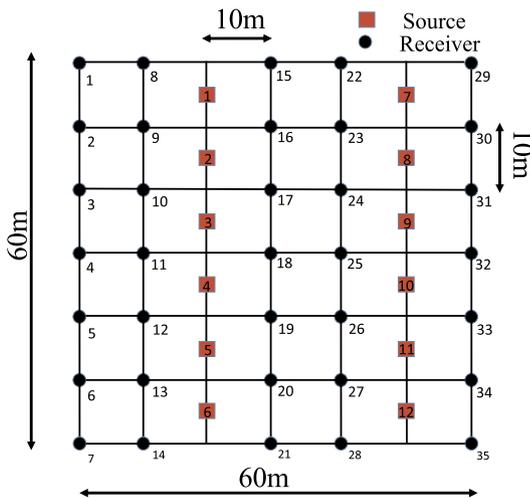


Fig. 3. Plan view schematic of the advanced geophysical experiment performed by Bielak et al. [6] in Garner Valley, CA.

experiment. A NEES@UTexas mobile shaker, T-Rex [10], was used to vertically excite the ground at each of the source locations, represented by red squares in Fig. 3, using a chirp signal with dominant frequencies between 1.5 Hz and 14 Hz for a duration of 2.25s. The force outputs of the T-Rex were recorded using accelerometers installed at its base plate. Fig. 4 shows the forcing time history recorded at Source Location 3 and its Fourier spectrum. The source excitations didn't plastify the soil [11].

The soil responses due to excitations at the source locations were recorded at 35 receiver locations, shown by black circles in Fig. 3, using 1 Hz vertical geophones. Per recommendation of the group that conducted the experiment, the recorded receiver data were filtered using an equiripple finite-impulse-response (FIR) bandpass filter with high and low cuts of 1.5 Hz and 14 Hz, respectively, and high and low slopes of 66.6 dB/Hz and 15 dB/Hz, respectively, to remove any ambient noise [11]. Fig. 5 shows the soil responses, in terms of velocity time histories, at Receiver Locations 10 and 2 due to the ground excitation at Source Location 3; both unprocessed and processed velocity time histories are shown. Note that due to a sensor malfunction, any reliable measurements were not recorded at Receiver Location 29. Accordingly, any measurements from Receiver Location 29 are not included in this study.

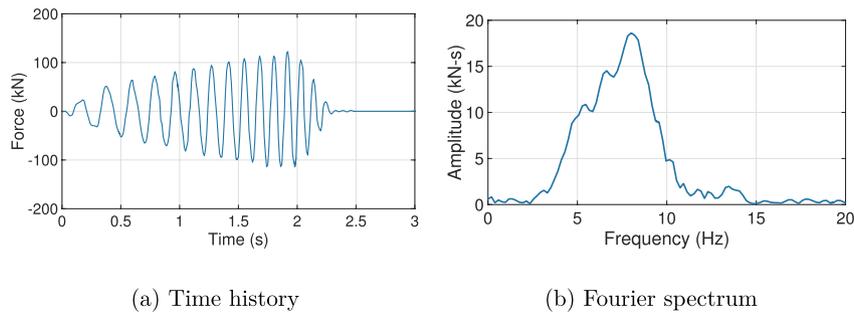


Fig. 4. Chirp signal used to excite the ground at Source Location 3 [6,11,12].

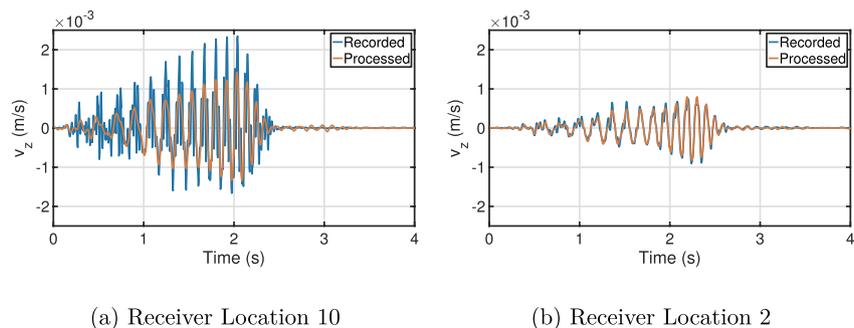


Fig. 5. Soil responses, in terms of velocity time histories, at two receiver locations due to ground excitation at Source Location 3 [6,11,12].

### 3. Numerical model of the geophysical experiment

To allow for the model-data fusion, the geophysical experiment, described in Section 2, needs to be numerically simulated. The time-domain dynamic finite element method [13] is employed to this end. A  $60\text{m} \times 60\text{m} \times 40\text{m}$  soil domain is numerically modeled as this study attempts to characterize the soil parameters up to a depth of 40 m. The model is constructed using standard 8-noded brick elements. The effect of radiation damping associated with the wave propagation through soils is captured by utilizing the widely used Lysmer-Kuhlmeyer dashpots [14]. Note that the dashpots are placed 20 m away in each direction from the soil domain of interest to minimize any boundary effects. Moreover, to optimize the dashpot performance, the size of the finite elements ( $5\text{m} \times 5\text{m} \times 5\text{m}$ ) is chosen as per the following equation [14]:

$$\frac{\text{element size}}{\text{minimum wavelength of propagating waves}} \leq \frac{1}{12} \quad (1)$$

As the T-Rex excitations didn't plastify the soil, the soil material is assumed to follow the linear elastic Hooke's law. The elastic parameters of the soil are hypothesized to be heterogeneous, anisotropic random fields to account for the several sources of uncertainty associated with their estimation process, including measurement noise and uncertainty due to limited data which indirectly captures, among others, micro-structural randomness. Note, however, that direct treatment of micro-structural randomness is beyond the scope of this paper. Further note that the "anisotropy" designation does not refer to the soil behavior, which is isotropic, but to the character of the random fields. The random field soil parameters are discretized using a Gaussian mixture model [3] that mimics typical soil formation process:

$$Y(x, y, z, \Theta) = \sum_{i=1}^{N_G} \alpha(z, \Theta_1) \Gamma_i(x, y, z, \Theta_2) \quad (2)$$

with

$$\alpha(z, \Theta_1) = \sum_{k=1}^L \left( 1 - \frac{|z - z_k|}{h_z} \right) Y_k \quad (3)$$

and

$$\Gamma_i(x, y, z, \Theta_2) = e^{-\frac{1}{2} \left( \frac{(x-x_i)^2}{l_x^2} + \frac{(y-y_i)^2}{l_y^2} + \frac{(z-z_i)^2}{l_z^2} \right)} \quad (4)$$

where any random field soil parameter,  $Y$ , defined in  $xyz$ -space, is represented in terms of  $L$  number of hat functions  $\alpha$  that imitate soil layering and  $N_G$  number of Gaussian functions  $\Gamma$  that imitate the mixing of information between layers, making the layers non-uniform and allowing for lenses to appear within the non-uniform layers. Each hat function takes two parameters:  $Y_k$  which is the value of the soil parameter at the center of the function and  $h_z$  which controls the spread of the function. The other arguments of the hat functions are  $z_k$  which are the  $z$ -coordinates of the centers of the hat functions. Each Gaussian function takes three parameters,  $l_x$ ,  $l_y$ , and  $l_z$  which are the spread parameters of the function in the three perpendicular directions. The other arguments of the Gaussian functions are  $x_i$ ,  $y_i$ , and  $z_i$  which are the coordinates of the centers of the functions. The numerical simulation of the geophysical experiment considers all the parameters of the hat and Gaussian functions to be uncertain. Note that symbol  $\Theta = \{\Theta_1, \Theta_2\}$  denotes uncertainty in the soil parameter with  $\Theta_1 = \{h_z, Y_k\}$  and  $\Theta_2 = \{l_x, l_y, l_z\}$  being the uncertain parameters of the hat and Gaussian functions, respectively.

Assuming the statistics of the uncertain parameters of the hat and Gaussian functions, the dynamic finite element simulation of the wave propagation through the soil continuum due to the T-Rex excitations at the source locations can be probabilistically performed to obtain statistical descriptions of simulated geophysical measurements. This study employs a stochastic collocation approach [15] for probabilistic simulation of the finite element model and computes the  $N^{\text{th}}$  order (raw)

statistical moment of soil displacement  $u$ , at any finite element node,  $i$ , at any time step,  $k$  as [3,16]:

$$\mathcal{E}[u_{ki}^N] = \int_{\Theta} u_{ki}^N p(\Theta) d\Theta \quad (5)$$

where  $\mathcal{E}[\cdot]$  denotes the expectation operation and  $p$  denotes the joint probability density function (PDF) of the random variables  $\Theta$  that are used to represent the uncertainty in the parameters of the Gaussian and hat functions. The integral on the r.h.s. of Eq. (5) is evaluated using a recently developed conjugate unscented transform (CUT) quadrature rule [17,18] that efficiently computes multidimensional integrals involving uniform and Gaussian PDFs by constraining the quadrature points to lie on specially defined axes. To this end, the random variables,  $\Theta$ , are first represented in terms of uniform or Gaussian standardized random variables  $\xi = [\xi_1, \xi_2, \dots, \xi_m]^T \in \mathbb{R}^m$  defined by a joint PDF,  $p(\xi)$ , following the theory of the generalized polynomial chaos (gPC; [19]). Eq. (5) is then evaluated as [3,16]:

$$\mathcal{E}[u_{ki}^N] = \int_{\xi} u_{ki}^N p(\xi) d\xi \approx \sum_{q=1}^M w_q u_{ki}^N(\Theta(\xi^q)), \quad N = 1, 2, \dots \quad (6)$$

where  $\Theta(\xi^q) \in \mathbb{R}^{m \times 1}$  represents the  $q^{\text{th}}$  quadrature point,  $w_q$  is the corresponding weight, and  $M$  is the total number of quadrature points. The procedure for evaluating the CUT quadrature points and their corresponding weights are described in Refs. [17,18].

The following are the main steps of the stochastic collocation approach based probabilistic simulation of a finite element model:

- Expansion of the random variables corresponding to the soil properties,  $\Theta$ , as a polynomial of uniform random variables.
- Generation of ensembles of soil properties by using CUT quadrature points for uniform random variable in conjunction with the polynomial expansion for  $\Theta$ .
- Generation of ensembles for quantities of interest (e.g., soil displacement, velocity, acceleration, etc.) by running a deterministic finite element code for each of these soil properties ensembles.
- Computation of the weighted ensemble averages (i.e, statistical moments) of the quantities of interest while using Eq. (6).

### 4. Model-data fusion

In this procedure, the geophysical experiment is numerically simulated using the procedure described in Section 3 after making prior assumptions on the soil parameter random fields. The simulated measurements are then fused with experimental measurements, described in Section 2, using a minimum variance framework to update the prior assumptions on the soil parameter random fields.

#### 4.1. Prior assumptions on the soil parameters

In the field of solid mechanics, an elastic material is conventionally characterized in terms of the Young's modulus ( $E$ ) and Poisson's ratio ( $\nu$ ). However, in the field of geotechnical engineering, an elastic soil is typically characterized in terms of the soil's P-wave velocity ( $V_p$ ) and S-wave velocity ( $V_s$ ). Accordingly, this study will characterize the elastic parameters of soil at the site of advanced geophysical experiment in terms of  $V_p$  and  $V_s$ . Note that these wave velocities are directly related to the fundamental mechanical parameters through the soil's density,  $\rho$ , as [20]:

$$V_p = \sqrt{\frac{E(1-\nu)}{\rho(1+\nu)(1-2\nu)}} \quad (7a)$$

$$V_s = \sqrt{\frac{E}{2\rho(1+\nu)}} \quad (7b)$$

There exist borehole  $V_p$  and  $V_s$  measurements, obtained through PS surface logging and seismic cone tests, at a few locations close to the

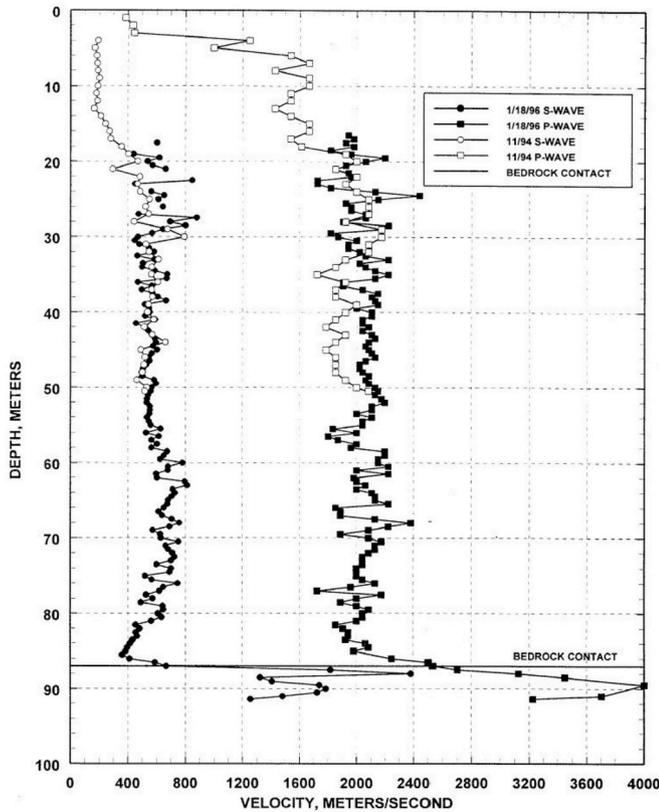


Fig. 6. Borehole  $V_p$  and  $V_s$  measurements at a location close to the site of the advanced geophysical experiment [9].

site of the advanced geophysical experiment [9]. Fig. 6 shows one such set of measurements. The  $V_s$  measurements, when interpreted with the information of the local geology, indicate that the site, up to a depth of 40 m, most likely consists of two distinct soil layers – a relatively loose sand layer up to a depth of approximately 20 m, overlain by a denser sand or weathered rock layer. Accordingly, to discretize the  $V_s$  random field at the site of the advanced geophysical experiment, we have used 3 hat centers – one at the surface, one at 20 m below the surface, and the other at 40 m below the surface – and the spread parameter of the hat functions,  $h_z$ , to be equal to the distance between the hat centers, i.e., 20 m. The values of  $V_s$  at the centers of the hat functions are considered to be random variables and are denoted as  $V_{S0}$ ,  $V_{S20}$ , and  $V_{S40}$ . Without much prior information to rely on, the random variables are assumed to follow uniform distributions with positive supports. Note that the assumed distributions will be updated when the simulated measurements will be fused with the experimental measurements. The prior mean

values of the random variables are chosen to be the values measured at the borehole location at those depths i.e., 200 m/s at the surface, 500 m/s at a depth of 20 m, and 700 m/s at a depth of 40 m. The supports of the uniform random variables are chosen assuming a coefficient of variation (COV) of 10%. This choice of the prior COV is based on the analysis of the borehole  $V_s$  measurements available at a few locations close to the site [9]. Next, to allow for mixing of information between layers, the Gaussian functions are superimposed on the hat functions. In making prior assumptions on the parameters of the Gaussian functions, following typical formation of alluvial soil deposits, it is assumed that the random field  $V_s$  exhibits cross anisotropy with the soil parameter correlated over a larger distance in the horizontal directions than in the vertical direction. In other words, it is assumed that the mixing of information occurs over a longer distance in the horizontal directions than that in the vertical direction. The spread parameters  $l_x$  and of the Gaussian functions that control the mixing of information in the horizontal directions are assumed to be identical, but random variables with a mean value of 9.5 m to mimic the typical horizontal correlation lengths reported in the literature for alluvial soils [21]. Without much prior information to rely on, the random variable, denoted as  $l$ , is, again, considered to follow a uniform distribution with a COV of 10%. The spread parameter  $l_z$  of the Gaussian functions that determines the mixing of information in the vertical direction is assumed to be controlled by the spread parameter  $h_z$  of the hat functions with a value of  $h_z/(L - 1)$  to mimic the typical vertical correlation length reported in the literature for alluvial soils [21]. Figs. 7 and 8 show the prior mean and standard deviation profiles and correlation structures in the  $x$ -,  $y$ -, and  $z$ -direction, respectively, of thus represented  $V_s$  random field.

Next, the other elastic soil parameter,  $V_p$ , is discretized using the Gaussian mixture approach. The borehole  $V_p$  measurements shown in Fig. 6 for a nearby location indicate the presence of water table at a depth of approximately 6 m as below that depth  $V_p$  values have drastically jumped to the range of the typical P-wave velocities of saturated soils. Note that the P-wave velocity of water is 1500 m/s and that of saturated soils typically range from 800 m/s to 2000 m/s depending upon the soil type and degree of saturation. Accordingly, in representing the  $V_p$  random field at the site of the advanced geophysical experiment, we have regarded the depth of water table,  $H_w$ , to be a discretization variable. To account for the uncertainty associated with the seasonal variation of the water table, we have considered  $H_w$  to be a uniform random variable with a mean value of 6 m and a COV of 20%. Above the water table,  $V_p$  is discretized in a similar fashion as  $V_s$ , but with 2 hat centers – one at the surface and the other at the depth of the (uncertain) water table. The value of  $V_p$  at the hat center on the surface, denoted as  $V_{p0}$ , is assumed to be a uniform random variable with a mean value chosen to be the value at the borehole location at the surface, i.e., 500 m/s, and a COV equal to 10%. The value of  $V_p$  at the hat center at the depth of the (uncertain) water table is chosen indirectly from the Poisson's ratio as follows. Below the water table, the soil is expected to

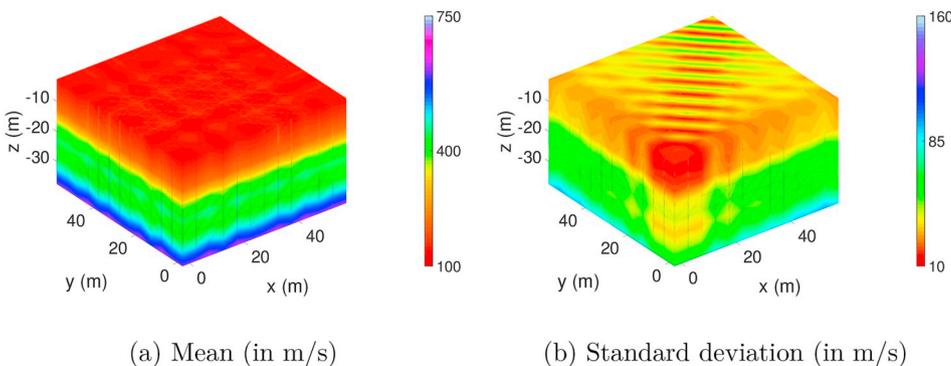


Fig. 7. Prior mean and standard deviation of the soil's S-wave velocity,  $V_s$ .

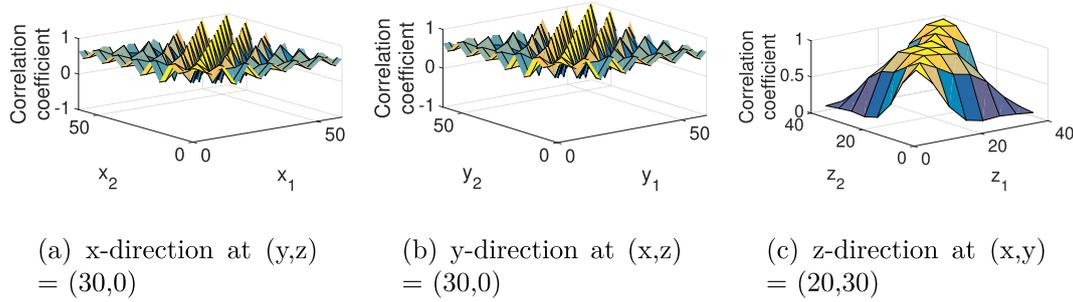


Fig. 8. Prior cross-anisotropic correlation structure of the soil's S-wave velocity,  $V_S$ .

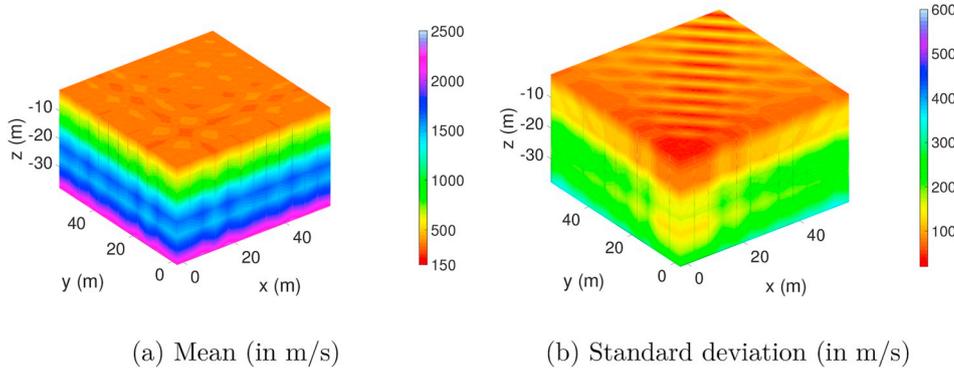


Fig. 9. Prior mean and standard deviation of the soil's P-wave velocity,  $V_P$ .

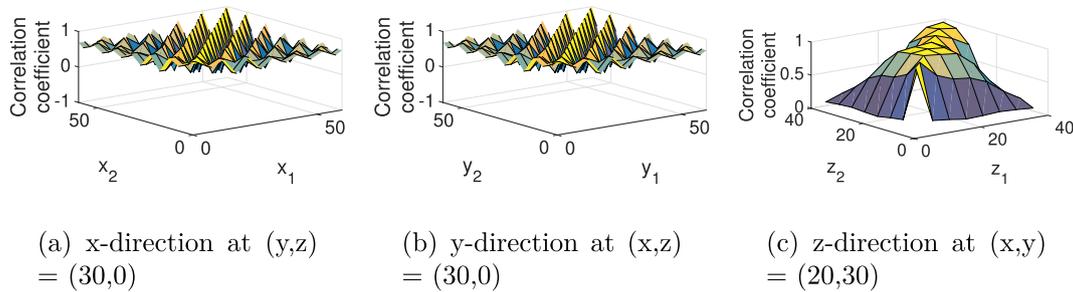


Fig. 10. Prior cross-anisotropic correlation structure of the soil's P-wave velocity,  $V_P$ .

be partially to fully saturated. Such a soil is typically almost incompressible and exhibits a Poisson's ratio close to 0.5 with a very small variability. Accordingly, to represent  $V_P$  at and below the water table, we have assumed a (probabilistic) Poisson's ratio and computed  $V_P$  as per the following equation:

$$V_P = \sqrt{\frac{2(1 - \nu)}{1 - 2\nu}} V_S \tag{8}$$

Note that the above equation may be obtained by combining Eqs. (7a) and (7b). The Poisson's ratio, denoted as  $\nu_{below H_w}$ , is assumed to be a uniform random variable with a mean value of 0.46 which is obtained by analyzing the measured values of  $V_S$  and  $V_P$  at the borehole location, shown in Fig. 6, using Eq. (8). Since, the Poisson's ratio of saturated soil doesn't exhibit much variability, the COV of  $\nu_{below H_w}$  is assumed to be 1%. The Gaussian function parameters of  $V_P$  above the water table are assumed to be the same as the ones used to represent  $V_S$ . This is a reasonable assumption as the pattern of information mixing for all the parameters of an alluvial soil deposit is expected to be very similar due to the formation process of such soil deposits. Note that, below the water table, we have not used any hat and Gaussian functions for  $V_P$  and have let the variation in  $V_P$  to be governed by  $V_S$  and  $\nu_{below H_w}$ , following Eq. (8). Further note that the probabilistic nature of  $\nu_{below H_w}$  will ensure the correlation structure of  $V_P$  to be different from that of  $V_S$ . It will

allow for capturing the effect of partially saturated zones below the water table. Thus represented  $V_P$  random field is shown in Figs. 9 and 10 in terms of mean and standard deviation profiles and correlation structures, respectively.

Table 1 summarizes the prior assumptions on all the random variables  $\Theta$  that are used to represent uncertainty in the soil's S-wave velocity,  $V_S$ , and P-wave velocity,  $V_P$ . The next subsection will probabilistically simulate the geophysical experiment using these prior assumptions. Later, the simulated measurements will be fused with experimental measurements to update the prior assumptions.

Table 1

Prior assumptions on the random variables  $\Theta$  that are used to represent uncertainty in the soil's S-wave velocity,  $V_S$ , and P-wave velocity,  $V_P$ .

RV	Type	Mean	COV (%)	Supports
$\Theta_1 = V_{S0}$	Uniform	200 m/s	10	[166, 234]m/s
$\Theta_2 = V_{S20}$	Uniform	500 m/s	10	[413, 587]m/s
$\Theta_3 = V_{S40}$	Uniform	700 m/s	10	[578, 821]m/s
$\Theta_4 = V_{P0}$	Uniform	500 m/s	10	[413, 587]m/s
$\Theta_5 = H_w$	Uniform	6 m	20	[4,8]m
$\Theta_6 = \nu_{below H_w}$	Uniform	0.46	1	[0.452, 0.468]
$\Theta_7 = l$	Uniform	9.5 m	10	[7.885, 11.115]m

4.2. Simulated geophysical measurements

This subsection numerically simulates the geophysical experiment, described in Section 2, by hypothesizing the soil's  $V_s$  and  $V_p$  to be

random fields with the random field parameters given in Table 1. To this end, the time-domain dynamic finite element method, in conjunction with a CUT-based stochastic collocation approach, as described in Section 3, is employed. Accordingly, following Ref. [3], the

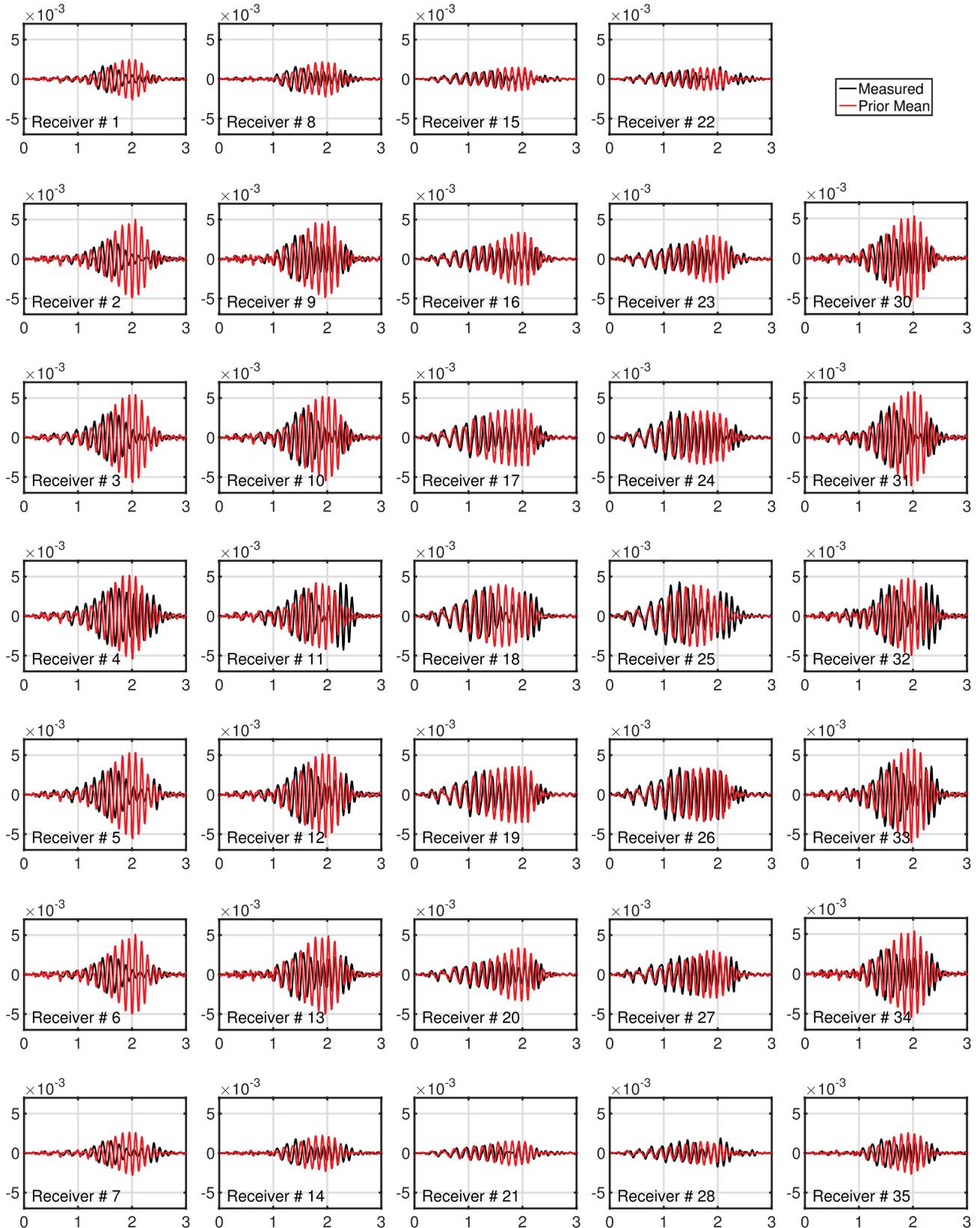


Fig. 11. Time histories of the vertical component of the simulated prior mean velocity at the receiver locations. The x axis represents time (in s) and the y axis represents velocity (in m/s).

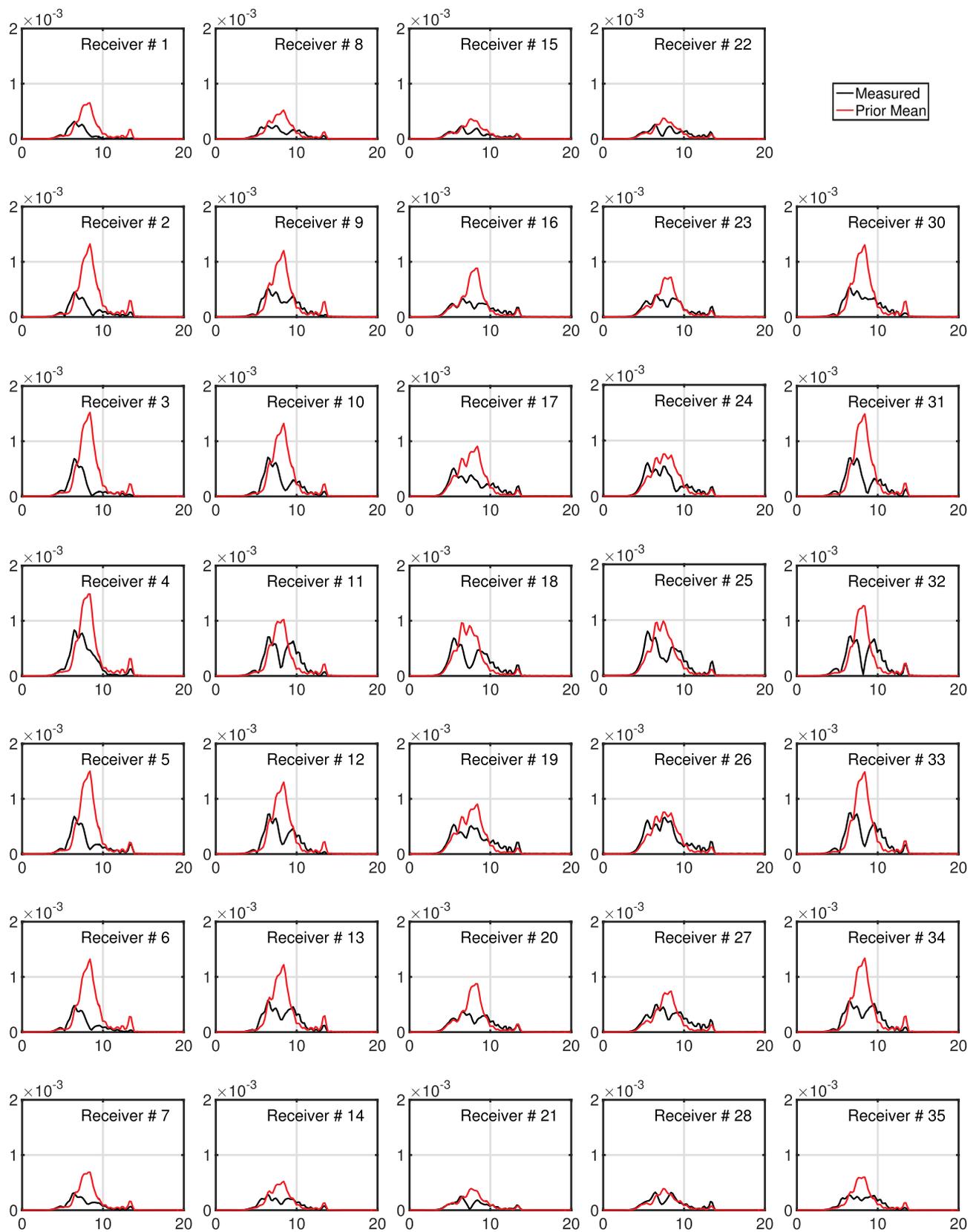


Fig. 12. Fourier spectra of the vertical component of the simulated prior mean velocity at the receiver locations. The x axis represents frequency (in Hz) and the y axis represents Fourier amplitude (in m/s-s).

CUT realizations of the random field soil parameters are first generated, using a 6th order CUT quadrature scheme. For this 7-dimensional stochastic problem, a 6th order CUT quadrature scheme has resulted in 551 CUT realizations. The finite element model of the geophysical experiment – constructed using an open source finite element code OpenSees [22] – is then simulated for each of the 551 CUT realizations to obtain 551 sets of velocity waveforms at the receiver locations. As the soil medium is lossy, damping is introduced while solving the finite element system of equations. The damping is approximated through the Rayleigh formulation [23]. As the soil is expected to remain elastic throughout the finite element simulation process, only a small amount of damping is used: the Rayleigh damping parameters are assumed to be given by 2% critical damping for the first two natural modes of vibration. Again, owing to the linearity of the problem, all the 12 loads are applied together. The simulated responses are appropriately integrated using the CUT quadrature scheme to obtain the statistics – in terms of marginal mean, marginal standard deviation and spatio-temporal correlation structures – of the velocity waveforms at the receiver locations. Figs. 11 and 12 show the vertical component of the simulated mean velocity at all the 34 receiver locations in terms of the time histories and their Fourier spectra, respectively. The experimentally measured time histories and their Fourier spectra are also shown for comparison. As can be observed that they do not match well. There exist significant phase differences between the experimental measurements and the mean of simulated measurements. The simulated measurements also have failed to capture the dominant frequencies at most of the receiver locations. Such mismatches mean that the prior assumptions on the  $V_S$  and  $V_P$  random fields are off. The next subsection will fuse the simulated measurements with the experimental measurements in updating the prior assumptions on the soil parameter random fields.

#### 4.3. Posterior estimates of the soil parameters

This subsection employs a minimum variance framework [3,24] to update the prior assumptions on the random variables  $\Theta$  through fusion of the statistics of the simulated measurements with experimental measurements. Following Ref. [3], let's consider that the experimental geophysical measurements,  $\mathcal{Z}$ , are given by:

$$\mathcal{Z} = h(u(\Theta)) + \kappa \quad (9)$$

where  $\kappa$  is the sensor noise and  $h(\cdot)$  represents the mapping from the output,  $u(\Theta)$ , of the numerical model of the geophysical experiment to the measurement variable. The best linear unbiased estimates of the posterior statistics of the random variables  $\Theta$  may, then, be obtained as [3,24]:

$$\mathcal{E}^+[\Theta] = \mathcal{E}^-[\Theta] + \mathbf{K}[\mathcal{Z} - \mathcal{E}^-[h(u(\Theta))]] \quad (10a)$$

$$\Sigma_{\Theta\Theta}^+ = \Sigma_{\Theta\Theta}^- - \mathbf{K}\Sigma_{\Theta h} \quad (10b)$$

where

$$\mathbf{K} = \Sigma_{\Theta h}^T (\Sigma_{hh}^- + \mathbf{R})^{-1} \quad (10c)$$

The matrix  $\mathbf{K}$  is the Kalman gain matrix and the superscripts, + and -, represent posterior and prior, respectively. The matrices  $\Sigma_{\Theta\Theta}$ ,  $\mathbf{R}$ ,  $\Sigma_{hh}$ , and  $\Sigma_{\Theta h}$  are the variance-covariance matrices of the random variables  $\Theta$ , sensor noise  $\kappa$ , simulated measurements  $h$ , and between  $h$  and the random variables  $\Theta$ , respectively. The matrix  $\Sigma_{\Theta\Theta}$  is given by the prior assumptions on the random field soil properties, while the matrix  $\mathbf{R}$ ,

**Table 2**  
Time domain model-data fusion results.

RV	Mean	Standard deviation
$\Theta_1 = V_{S0}$	274.9 m/s	0.013 m/s
$\Theta_2 = V_{S20}$	488.2 m/s	0.047 m/s
$\Theta_3 = V_{S40}$	1664.2 m/s	0.241 m/s
$\Theta_4 = V_{P0}$	514.6 m/s	0.118 m/s
$\Theta_5 = H_w$	27.8 m	0.014 m
$\Theta_6 = \nu_{below H_w}$	0.56	0.0001
$\Theta_7 = l$	10.5 m	0.0004 m

which is related to the noise to signal ratio, may be obtained from sensor calibration. On the other hand, the matrix  $\Sigma_{hh}$  given by the probabilistic solution of the numerical model of the geophysical experiment and the matrix  $\Sigma_{\Theta h}$  may be computed by integrating the simulated measurement variables  $h(\cdot)$  with the soil parameter random variables  $\Theta$  using the CUT quadrature scheme as:

$$\begin{aligned} \Sigma_{\Theta h} &= \mathcal{E}^-[(\Theta - \mathcal{E}^-[\Theta])(h(u(\Theta)) - \mathcal{E}^-[h(u(\Theta))])^T] \\ &= \sum_{q=1}^M w_q (\Theta_q - \mathcal{E}^-[\Theta])(h(u(\Theta_q)) - \mathcal{E}^-[h(u(\Theta))])^T \end{aligned} \quad (11)$$

where  $M$  denotes the total number of CUT quadrature points and  $w_q$  corresponds to the weight allotted to the  $q^{th}$  CUT quadrature realization,  $\Theta_q$ , of the random variables  $\Theta$ .

We have first attempted to solve Eqs. (10a)–(10c) in the time domain, assuming a noise to signal ratio of 1%. Note that since the measurements are already processed through a bandpass filter, following recommendation of the group that conducted the geophysical experiment, the noise to signal ratio is kept low. The resulting posterior estimates are shown in Table 2. As can be observed that the time-domain model-data fusion has yielded poor posterior estimates – impermissible value of the Poisson's ratio and an unrealistic depth of the water table. These poor estimates are attributed to the failure of the ensemble averages of the model output – that define the vector  $\mathcal{E}^-[h(u(\Theta))]$  and matrices  $\Sigma_{hh}^-$  and  $\Sigma_{\Theta h}$  – to preserve the dynamical characteristics of the soil continuum.

In order to overcome the above-mentioned limitation of a time-domain analysis, we have resorted to the frequency domain. We have considered the measurement variables to be the two dominant frequencies of the soil velocities at the receiver locations and their

**Table 3**  
Posterior marginal statistics of the random variables  $\Theta$  that are used to represent uncertainty in the soil's S-wave velocity,  $V_S$ , and P-wave velocity,  $V_P$ .

RV	Mean	Standard deviation	COV (%)	Supports
$\Theta_1 = V_{S0}$	182.6 m/s	3.9 m/s	2.1	[160.5,192.2]m/s
$\Theta_2 = V_{S20}$	308.3 m/s	7.8 m/s	2.5	[274,348.5]m/s
$\Theta_3 = V_{S40}$	854.2 m/s	58.7 m/s	6.8	[760.9,1031]m/s
$\Theta_4 = V_{P0}$	728.4 m/s	29.3 m/s	4.0	[676.8,855]m/s
$\Theta_5 = H_w$	6.9 m	0.8 m	12.51	[4.4,9.6]m
$\Theta_6 = \nu_{below H_w}$	0.454	0.002	0.5	[0.44,0.46]
$\Theta_7 = l$	8.3 m	0.2 m	2.1	[7.8,9.2]m

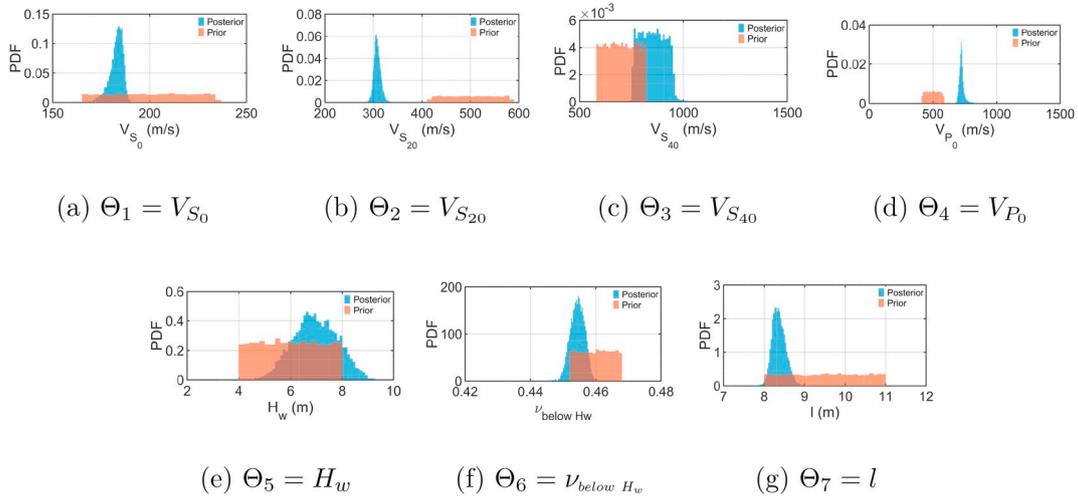


Fig. 13. Posterior marginal probability densities of the random variables  $\Theta$  that are used to represent uncertainty in the soil's S-wave velocity,  $V_s$ , and P-wave velocity,  $V_p$ .

corresponding Fourier amplitudes. The reason behind selecting only two frequencies stems from the observation that the measured velocity at majority of the receiver locations showed two distinct dominant frequencies (Fig. 12). With such a choice of the feature vector, our goal is to capture the predominant dynamic characteristics of the measurements (i.e., soil velocities) at the receiver locations. Of course, one could go with a feature vector comprised of the entire Fourier spectrum of soil velocities at all the receiver locations. However, such treatment will require much larger computational effort as it will necessitate a higher fidelity finite element model along with finer discretizations of the soil parameter random fields with additional random variables to accurately capture the scattering complexity.

The outcome of the above-described frequency domain model-data fusion – again, assuming a noise to signal ratio of 1% – is shown in Table 3. As can be observed that realistic values of the posterior statistics of the random variables  $\Theta$  are obtained through the frequency-domain analysis. The posterior marginal PDFs of the random variables  $\Theta$  are also estimated (Fig. 13). The marginal PDFs are approximately estimated using the theory of gPC. To this end, the random variables  $\Theta$  are first represented in terms of the Legendre basis functions of standardized uniform random variables  $\xi = [\xi_1, \xi_2, \dots, \xi_m]^T$  and the gPC coefficients are estimated from the posterior statistics,  $\mathcal{E}^+[\Theta]$  and  $\Sigma_{\Theta}^+$ , of the random variables  $\Theta$  by solving an underdetermined system of equations as below [3]:

$$\Theta_j(\xi) = \sum_{k=0}^N c_{jk} \phi_k(\xi), \quad j = 1, 2, \dots, m \quad (12a)$$

with the gPC coefficients,  $c_{jk}$ , are given by:

$$c_{j0} = \mathcal{E}^+[\Theta_j], \quad j = 1, 2, \dots, m \quad (12b)$$

$$\sum_{k=1}^Q c_{ik} c_{jk} \mathcal{E}^+[\phi_k(\xi) \phi_k(\xi)] = \Sigma_{\Theta_{ij}}^+, \quad i, j = 1, 2, \dots, m \quad (12c)$$

where  $Q = 1 + \sum_{s=1}^q \frac{1}{s!} \prod_{r=0}^{s-1} (m+r)$  with  $m$  being the length of the vector  $\Theta$ , and  $q$  being the order of the basis functions. This study uses a 6th order gPC basis functions. Normalized histograms are then generated through the Monte Carlo simulation of thus obtained orthogonal polynomial representations of the random variables  $\Theta$ . As can be observed that the frequency-domain model-data fusion through the minimum variance framework has updated both the statistics and distribution characteristics of the random variables  $\Theta$ . Compared to the prior assumptions, the marginal mean values have moved and the marginal COVs have reduced for all the constituent random variables of  $\Theta$ . However, since the true values of the soil parameters at the site of the geophysical experiment are not known, the accuracy of the posterior estimates  $\Theta$  may not be ascertained from these comparisons.

To check the accuracy of the posterior estimates of  $\Theta$ , we have re-simulated the geophysical experiment, probabilistically with the posterior estimates and have compared the simulated posterior measurements with the experimental measurements. Figs. 14 and 15 compare the vertical component of the simulated posterior mean soil velocity time histories at all the 34 receiver locations and their Fourier spectra, respectively, with the experimentally measured soil velocity time histories and their Fourier spectra. As can be observed that the simulated posterior mean velocities match reasonably well, in both the time and frequency domains, with the experimentally measured velocities at all the 34 receiver locations. The large phase differences, that were observed in Fig. 11 between the simulated prior mean and experimentally measured velocity time histories at all the receiver locations, are reduced. Also, the dominant frequencies are well captured. However, there still exist some differences, in particular, in the magnitude of velocities in the time domain at a few receiver locations. The rest of the paper will focus on understanding the reason(s) behind such differences and on improving the posterior estimates.

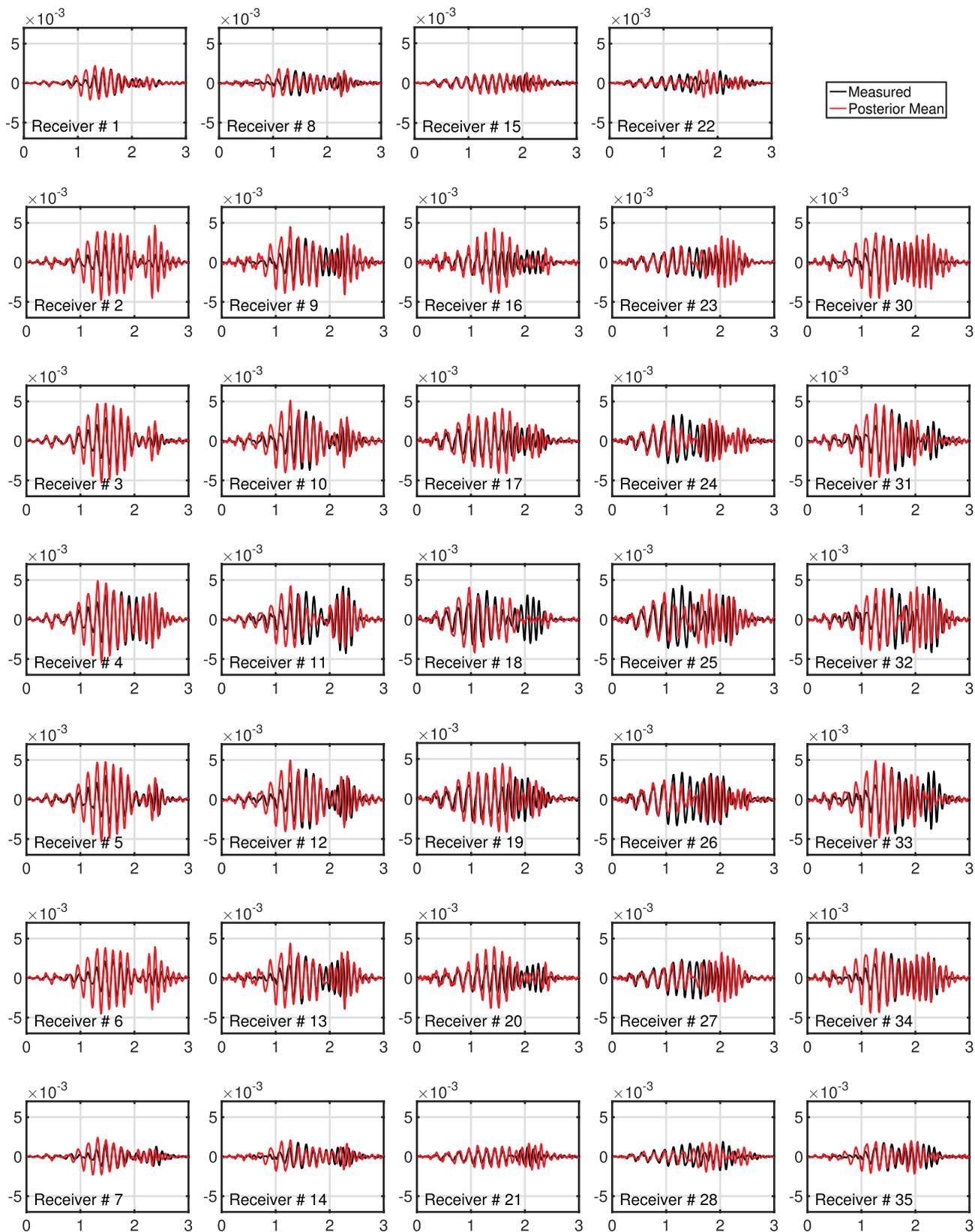


Fig. 14. Time histories of the vertical component of the simulated posterior mean velocity at the receiver locations. The x axis represents time (in s) and the y axis represents velocity (in m/s).

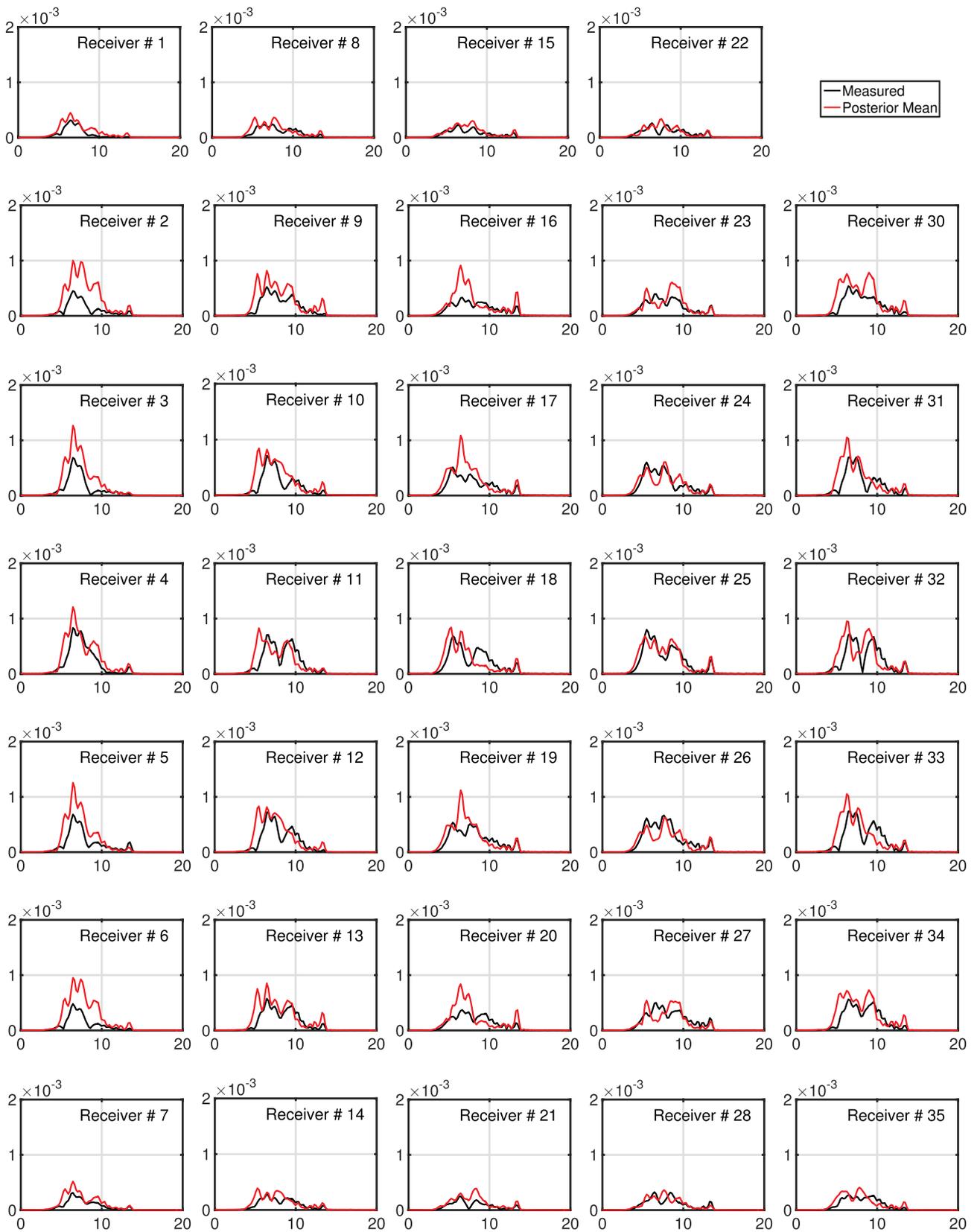


Fig. 15. Fourier spectra of the vertical component of the simulated posterior mean velocity at the receiver locations. The x axis represents frequency (in Hz) and the y axis represents Fourier amplitude (in m/s-s).

4.4. Sensitivity of the simulated measurements on the prior assumptions

It's a well known fact that the success of any inversion process depends on how good the prior assumptions are. The prior assumptions

restrict the space in which the inversion algorithm searches for possible solutions to the problem. In the previous subsection, the inverse algorithm has tried to minimize the error between the two dominant frequency–Fourier amplitude pairs measured experimentally and

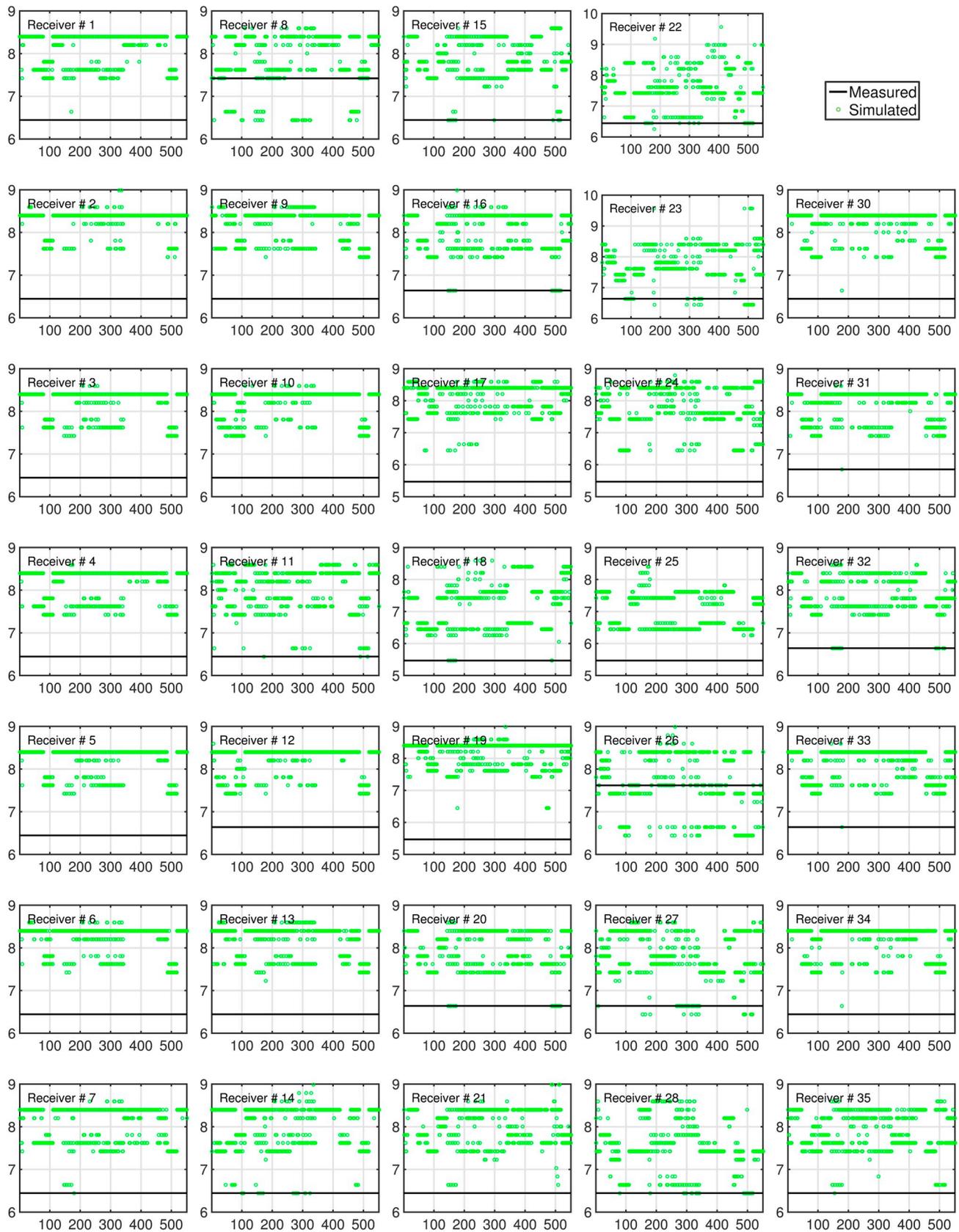


Fig. 16. Dominant frequency of velocity simulated with each of the 551 prior CUT realizations of  $\Theta$  in comparison to the experimentally measured dominant frequency of velocity. The x axis represents the CUT realization number and the y axis represents frequency (in Hz).

simulated with the prior assumptions on the random variables  $\Theta$  at each receiver location. So, it is logical to assume that if the simulated dominant frequency–Fourier amplitude pairs are not too far from the

experimentally measured dominant frequency–Fourier amplitude pairs, then the inverse algorithm will be less “stressed” to find a solution. Accordingly, in Figs. 16 and 17, we have plotted the dominant

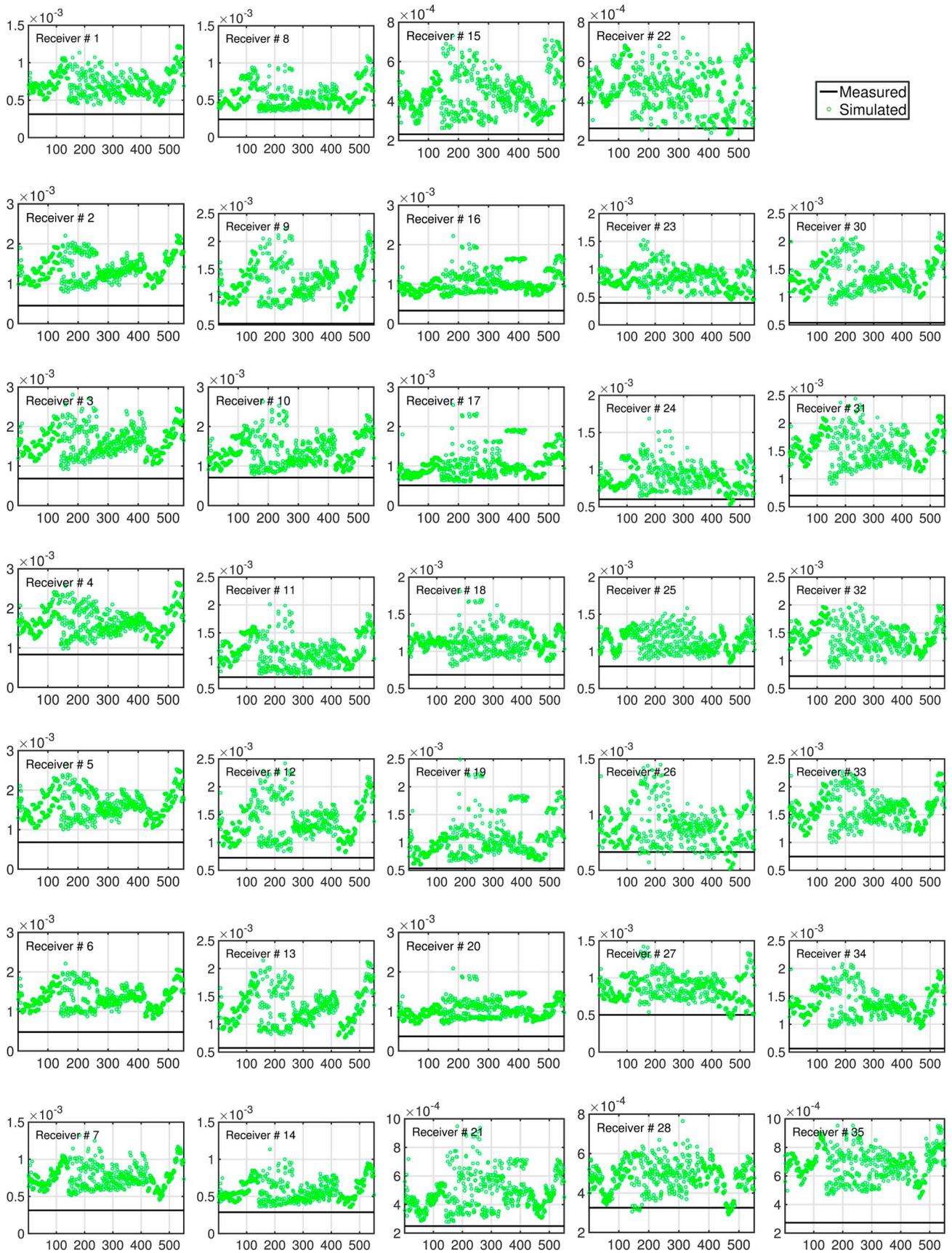


Fig. 17. Fourier amplitude corresponding to the dominant frequency of velocity simulated with each of the 551 prior CUT realizations of  $\Theta$  in comparison to the Fourier amplitude of experimentally measured dominant frequency of velocity. The x axis represents the CUT realization number and the y axis represents Fourier amplitude (in m/s-s).

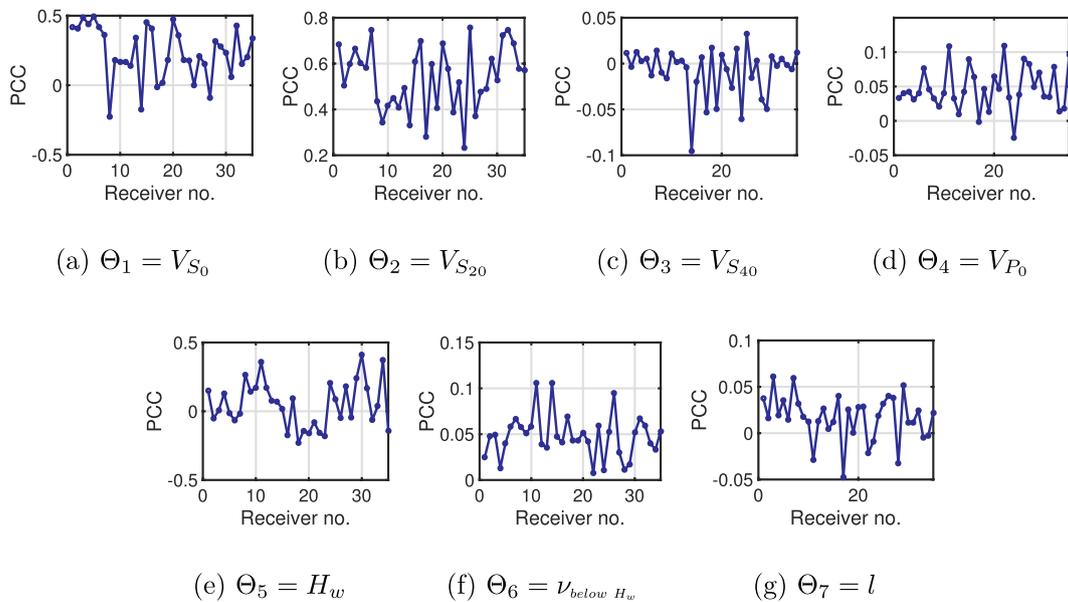


Fig. 18. Pearson's correlation coefficient (PCC) between prior random variables  $\Theta$  and simulated dominant frequency of velocity.

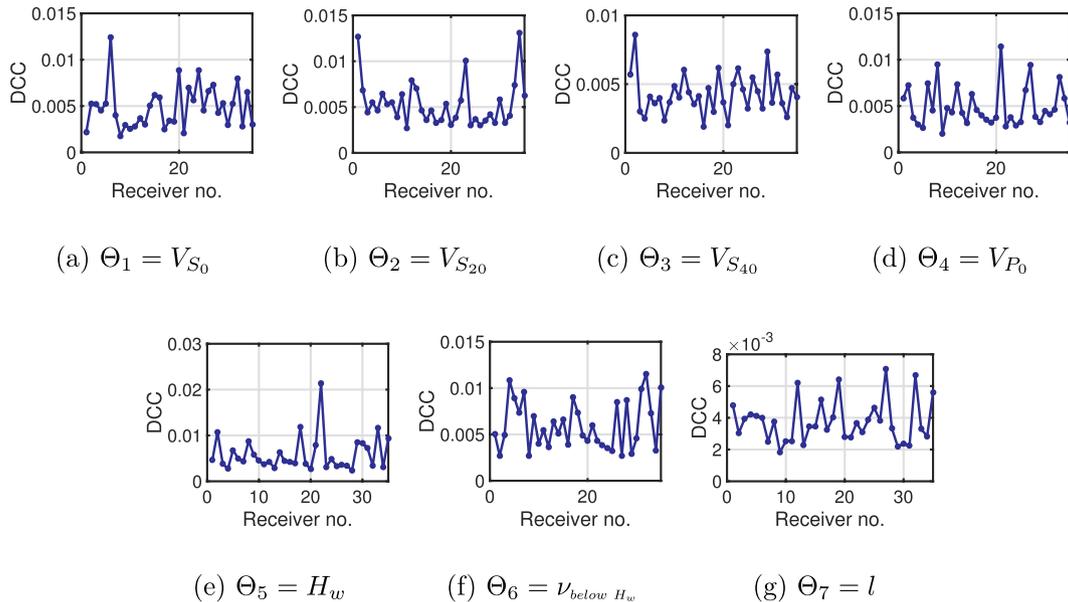


Fig. 19. Distance correlation coefficient (DCC) between prior random variables  $\Theta$  and simulated dominant frequency of velocity.

frequency and its Fourier amplitude, respectively, simulated with each of the 551 prior CUT realizations of  $\Theta$  at each of the receiver locations. We have compared them with the experimentally measured dominant frequency and its Fourier amplitude. As can be observed that at most of the receiver locations, the prior assumptions on the random variables  $\Theta$  have failed to yield the dominant frequency and its Fourier amplitude in the range of the experimental measurement. Ideally, the prior assumptions on the random variables  $\Theta$  should be such that the (CUT) realizations of the quantities of interest of the simulated measurements center around that of the experimental measurements.

Accordingly, with the goal of making better prior assumptions on the random variables  $\Theta$ , we have conducted a sensitivity analysis to determine the relative sensitiveness of the dominant frequency-Fourier amplitude pair on the constituent random variables of  $\Theta$ . To this end, we have looked at the Pearson's [25] and distance [26] correlation coefficients between the random variables  $\Theta$  and simulated dominant frequency and between the random variables  $\Theta$  and Fourier amplitude corresponding to the simulated dominant frequency. Note that while the Pearson's correlation coefficient is

a measure of linear dependence between two random variables, the distance correlation coefficient can also capture nonlinear dependence between two random variables. The approach described in Ref. [27] is used to compute both Pearson's and distance correlation metrics from CUT ensembles. Figs. 18 and 19 show the Pearson's and distance correlation coefficients, respectively, between the random variables  $\Theta$  and the simulated dominant frequency, while Figs. 20 and 21 show the Pearson's and distance correlation coefficients, respectively, between the random variables  $\Theta$  and the Fourier amplitude corresponding to the simulated dominant frequency. Note that the theoretical range of the Pearson's correlation coefficient is between  $-1$  and  $1$ . A value of zero signifies no linear dependence between the random variables, while the larger is the absolute magnitude, the more linearly dependent are the random variables. On the other hand, the theoretical range of the distance correlation coefficient is between  $0$  and  $1$ . A value of zero signifies no dependence (linear or nonlinear) between the random variables, while the larger is the magnitude, the more dependent are the random variables, either linearly or nonlinearly. Accordingly, it can be interpreted from Figs. 18–21 that the dominant

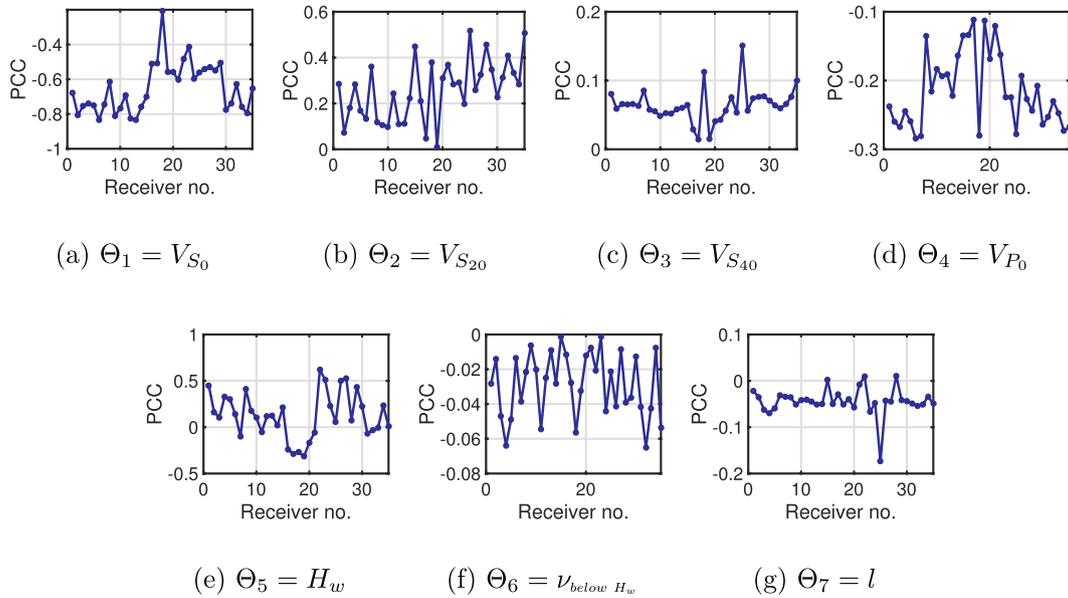


Fig. 20. Pearson's correlation coefficient (PCC) between prior random variables  $\Theta$  and Fourier amplitude corresponding to the simulated dominant frequency of velocity.

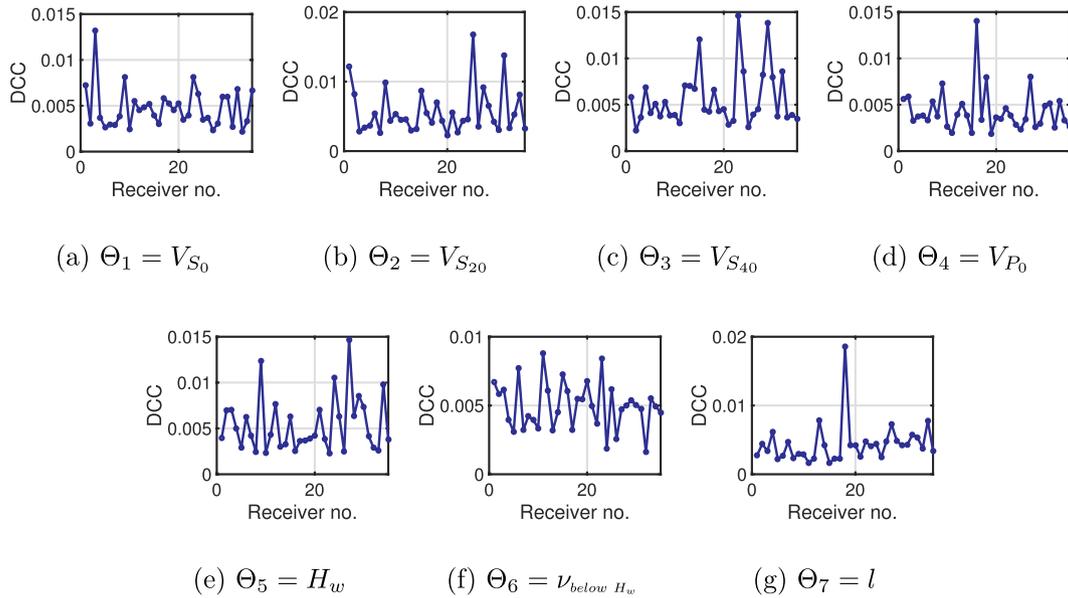


Fig. 21. Distance correlation coefficient (DCC) between prior random variables  $\Theta$  and Fourier amplitude corresponding to the simulated dominant frequency of velocity.

frequency and its Fourier amplitude are more sensitive to  $V_{S0}$ ,  $V_{S20}$ , and  $H_w$  when compared to the other constituent random variables of  $\Theta$ . With this information, in the next subsection, we have modified the prior assumptions on the random variables  $\Theta$  to bring the simulated dominant frequency and its Fourier amplitude at each receiver location closer to the experimentally measured dominant frequency and its Fourier amplitude.

#### 4.5. New prior assumptions on the soil parameters

The simulated dominant frequency of velocity and its Fourier amplitude are positively correlated to the random variable  $V_{S20}$  at all the receiver locations. The random variable  $H_w$  is also positively correlated to both the dominant frequency of velocity and its Fourier amplitude at most of the receiver locations. On the contrary, the random variable  $V_{S0}$ , while slightly positively correlated with the dominant frequency, is highly negatively correlated with the Fourier amplitude corresponding to the dominant frequency of velocity at most of the receiver locations. Accordingly, to bring down the simulated realizations of the dominant

frequency of the velocity and its Fourier amplitude closer to their experimentally measured values, we have decreased the mean values of  $V_{S20}$  and  $H_w$  and have increased the mean value of  $V_{S0}$ . The COVs and distribution types of those random variables are kept the same as before. Table 4 lists the new prior assumptions on the random variables  $\Theta$ .

Table 4

New prior assumptions on the random variables  $\Theta$  that are used to represent uncertainty in the soil's S-wave velocity,  $V_S$ , and P-wave velocity,  $V_P$ .

RV	Type	Mean	COV (%)	Supports
$\Theta_1 = V_{S0}$	Uniform	230 m/s	10	[190.9, 269.1]m/s
$\Theta_2 = V_{S20}$	Uniform	400 m/s	10	[332, 468]m/s
$\Theta_3 = V_{S40}$	Uniform	700 m/s	10	[581, 819]m/s
$\Theta_4 = V_{P0}$	Uniform	500 m/s	10	[406.7, 573.3]m/s
$\Theta_5 = H_w$	Uniform	5 m	20	[3.36, 8.04]m
$\Theta_6 = \nu_{below H_w}$	Uniform	0.46	1	[0.452, 0.468]
$\Theta_7 = l$	Uniform	9.5 m	10	[7.885, 11.115]m

4.6. Re-simulated geophysical measurements

To check the effects of such adjustments to the prior assumptions on the random variables  $\Theta$ , we have re-simulated the geophysical

experiment using the new prior assumptions on the random variables  $\Theta$  following the same procedure described in Section 4.2. The simulation results are presented in Figs. 22 and 23 in terms of the dominant frequency of velocity and its Fourier amplitude, respectively, obtained for

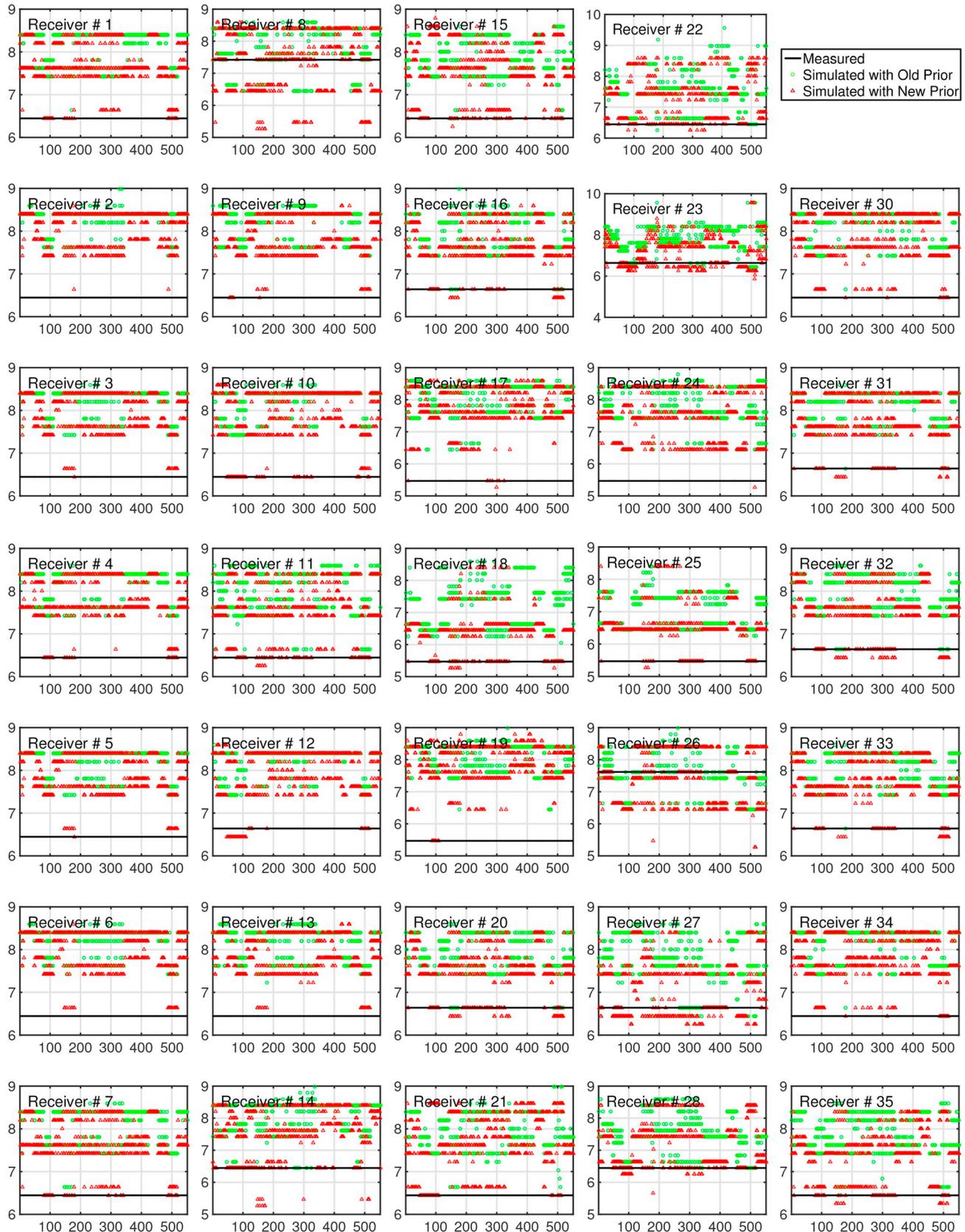


Fig. 22. Dominant frequency of velocity simulated with each of the 551 new prior CUT realizations of  $\Theta$  in comparison to the experimentally measured dominant frequency of velocity. The x axis represents the CUT realization number and the y axis represents frequency (in Hz).

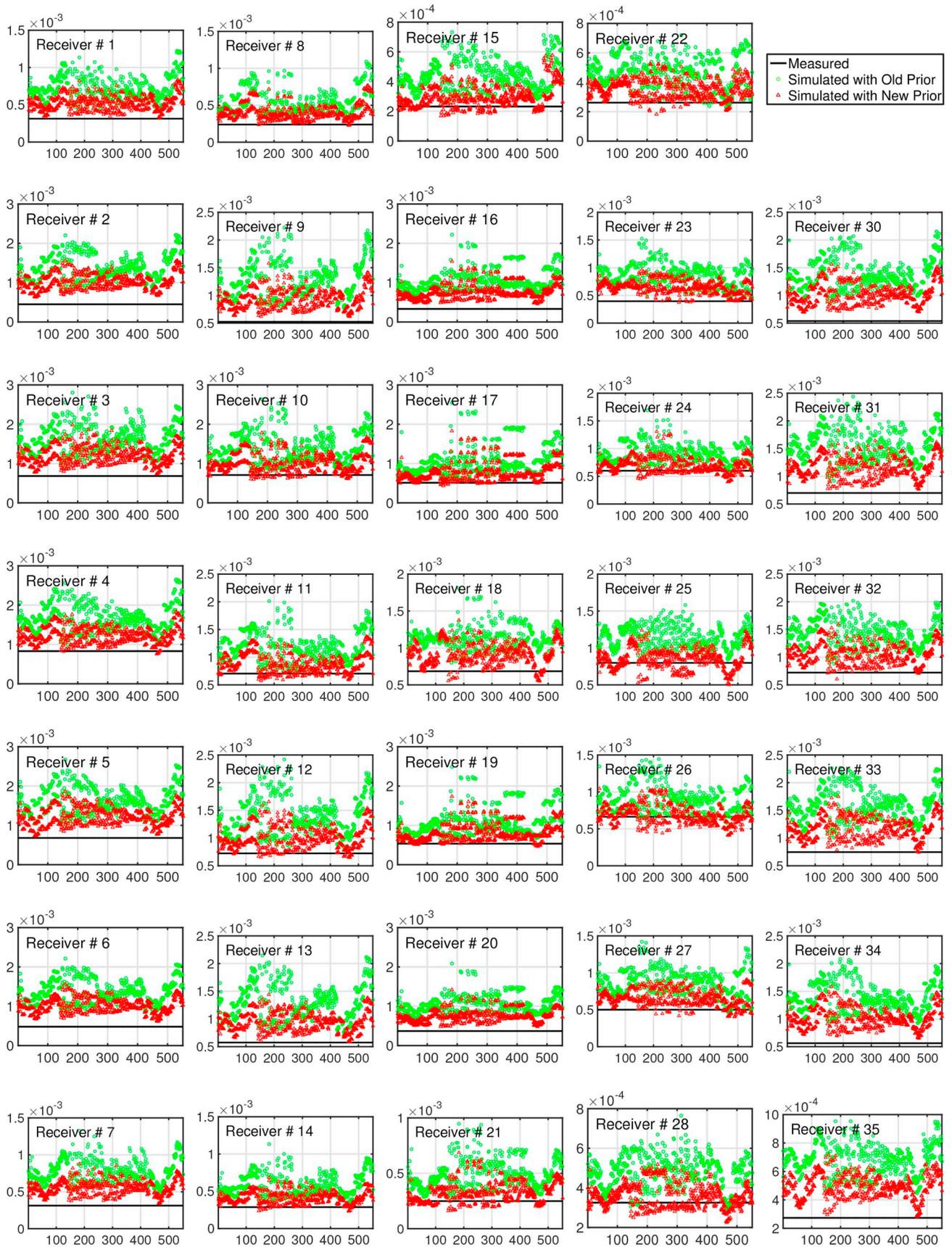


Fig. 23. Fourier amplitude corresponding to the dominant frequency of velocity simulated with each of the 551 new prior CUT realizations of  $\Theta$  in comparison to the Fourier amplitude of experimentally measured dominant frequency of velocity. The x axis represents the CUT realization number and the y axis represents Fourier amplitude (in m/s-s).

each of the 551 new prior CUT realizations of  $\Theta$  at each of the receiver locations. We have compared them with the same obtained for the old prior realizations of  $\Theta$  and with experimentally measured dominant frequency of velocity and its Fourier amplitude. As can be observed that the new prior assumptions on the random variables  $\Theta$  have faired much better than the old assumptions on the same in terms of capturing the experimentally measured dominant frequency of velocity and its Fourier amplitude.

4.7. Improved posterior estimates of the soil parameters

The simulated measurements obtained through the new prior assumptions on the random variables  $\Theta$  are fused with the experimental measurements, following the same procedure described in Section 4.3, for new estimates of the posterior statistics of the random variables  $\Theta$ . Table 5 lists the new posterior statistics, while Fig. 24 shows the new posterior distributions of the random variables  $\Theta$ . As before, the marginal mean values have moved and marginal COVs have reduced for all random variables as compared to the prior assumptions.

To check the accuracy of the new posterior estimates of  $\Theta$ , we have re-simulated the geophysical experiment probabilistically with the new posterior estimates. The simulated new posterior measurements are presented in Figs. 25 and 26 in terms of the vertical component of the mean soil velocity time histories at all the 34 receiver locations and their frequency spectra, respectively. They are compared with the same of the simulated old posterior measurements and with measured soil velocity time histories and their Fourier spectra. Indeed, the new posterior estimates of the random variables  $\Theta$  are more accurate than their previous posterior estimates. The difference between the magnitudes of the mean of the simulated velocities and experimentally measured ve-

locities are reduced at most of the receiver locations. In the frequency domain, too, the new posterior simulations have compared better with the experimental measurements than the previous posterior simulations. The improvement in inversion results can be attributed to the improved prior assumptions on parameters for the Gaussian mixture model representing soil properties. The minimum variance estimator updates the prior mean of random variables proportion to the mean error between measurement data and model predicted measurements. Furthermore, the minimum variance estimator performs better when actual measurement data corresponds to one of the realization of predicted measurement ensembles. The updated prior for soil parameters leads to the stochastic FEM ensembles, which better capture measurement data as their subset (cf. Figs. 22 and 23) and hence leads to more accurate soil parameter estimates.

Finally, the new posterior estimates of the random variables  $\Theta$  are post-processed using Eqs. (2)–(4) to generate different statistics of the soil parameter random fields,  $V_S$  and  $V_P$ . Figs. 27 and 28 shows the estimated three-dimensional marginal mean and marginal standard deviation profiles of  $V_S$  and  $V_P$ , respectively. The estimated correlation structures of  $V_S$  and  $V_P$  are shown in Figs. 29 and 30, respectively. The estimated mean profile of  $V_S$  indicates that both the soft alluvium and gravely sand/weathered granite layers at the site of the geophysical experiment are very nonuniform with presence of many lenses of different shapes and sizes. Note that the layer non-uniformity and lenses are due to intra- and inter-layer mixing of information of varying degrees. The fluctuating horizontal correlations of  $V_S$  from highly positive to highly negative over a long distance suggests intra-layer mixing of at least two different materials over the entire domain in the horizontal directions. The estimated vertical correlations of  $V_S$ , however, are always positive; the estimated correlation structure is approximately exponential (close to surface) to square exponential (at larger depth) with “equivalent correlation lengths” varying between 10 and 15 m. Note that the correlation length is a measure of the extent of correlation and is quantified as the distance at which the value of exponential or square exponential correlation curve drops to  $1/e$  [28]. These characteristics of the vertical correlation structure indicate only limited inter-layer mixing of information in the vertical direction. Also, comparing the marginal mean and standard deviation values of estimated  $V_S$  throughout the soil domain, it can be interpreted that its marginal COVs are larger (around 20%) at and close to the surface and they become smaller with depth. It signifies that the confidence in the  $V_S$  estimate is smaller close to the surface but it increases with depth. The marginal mean profile of  $V_P$  indicates pockets of larger values, consistent with that of partially saturated soils, at and just below the surface. Presence

Table 5

New posterior marginal statistics of the random variables  $\Theta$  that are used to represent uncertainty in the soil's S-wave velocity,  $V_S$ , and P-wave velocity,  $V_P$ .

RV	Mean	Standard deviation	COV (%)	Supports
$\Theta_1 = V_{S_0}$	226.6 m/s	10.9 m/s	4.8	[186,274]m/s
$\Theta_2 = V_{S_{20}}$	265.5 m/s	14 m/s	5.3	[208,322]m/s
$\Theta_3 = V_{S_{40}}$	749.8 m/s	55.5 m/s	7.4	[610,5,882]m/s
$\Theta_4 = V_{P_0}$	511.5 m/s	40.7 m/s	8.0	[442,713]m/s
$\Theta_5 = H_w$	5.9 m	1.1 m	13.8	[3.1,8.9]m
$\Theta_6 = \nu_{below H_w}$	0.455	0.0026	0.8	[0.444,0.461]
$\Theta_7 = l$	7.9 m	0.4 m	4.9	[6.8,9.7]m

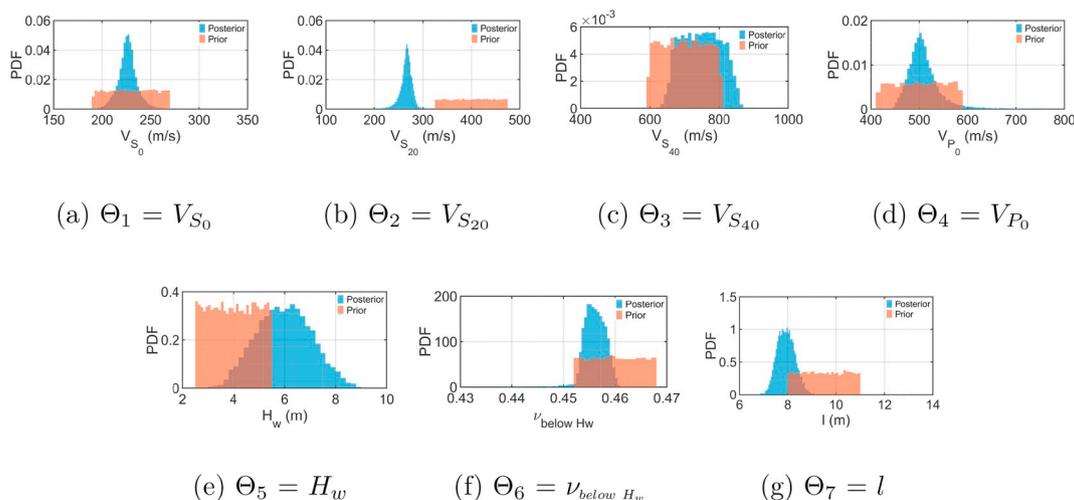


Fig. 24. New posterior marginal probability densities of the random variables  $\Theta$  that are used to represent uncertainty in the soil's S-wave velocity,  $V_S$ , and P-wave velocity,  $V_P$ .

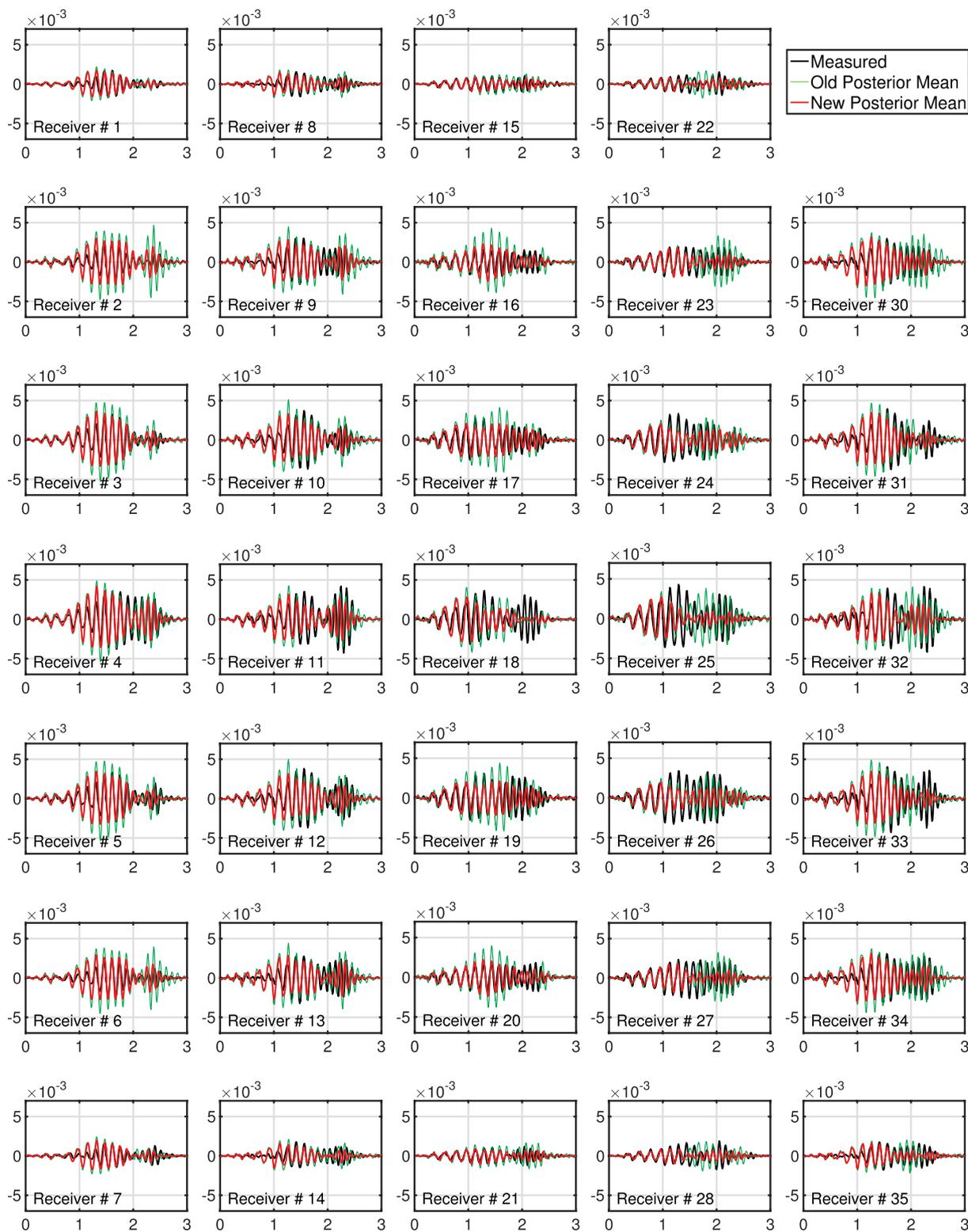
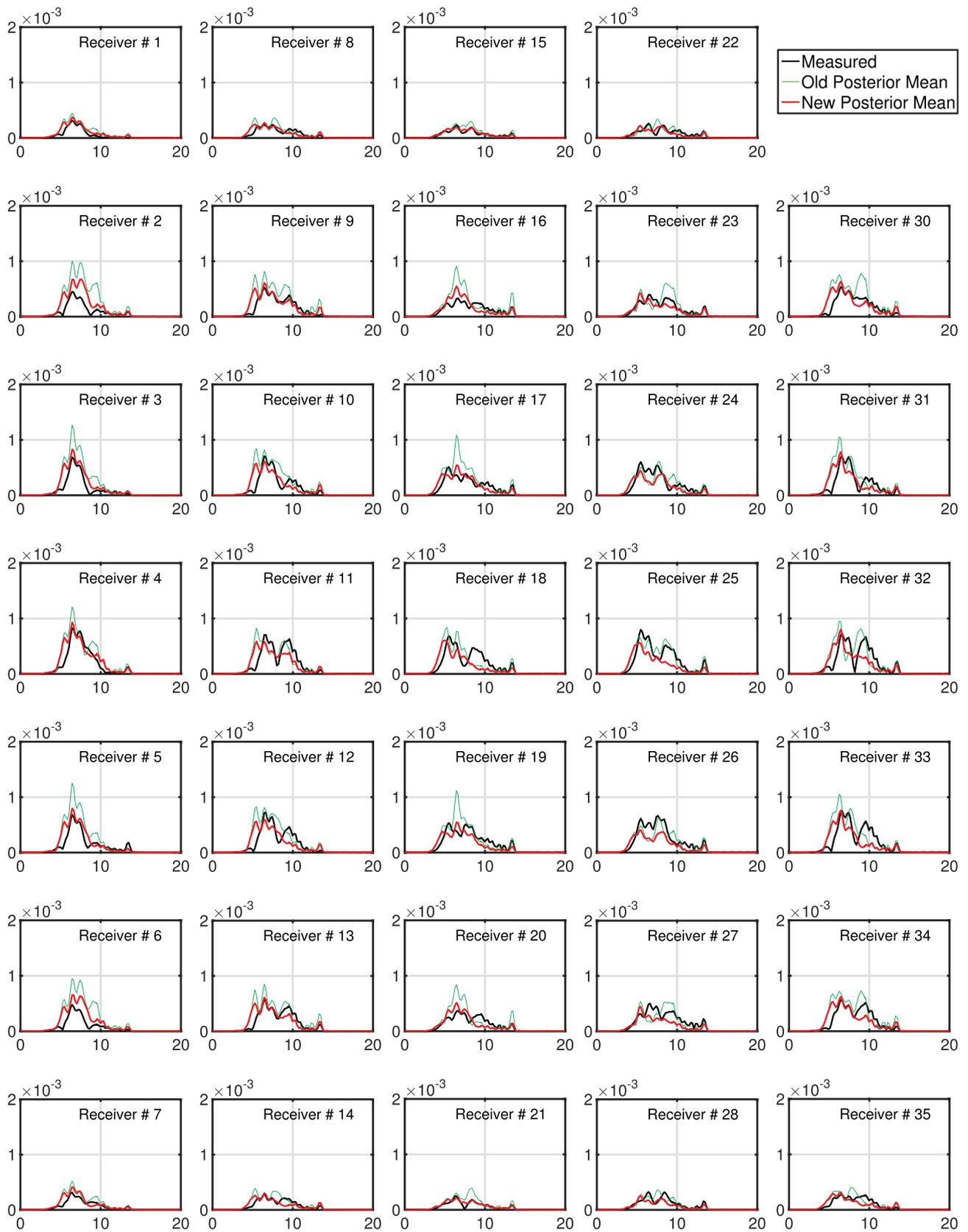


Fig. 25. Time histories of the vertical component of the simulated new posterior mean velocity at the receiver locations. The x axis represents time (in s) and the y axis represents velocity (in m/s).



**Fig. 26.** Fourier spectra of the vertical component of the simulated new posterior mean velocity at the receiver locations. The x axis represents frequency (in Hz) and the y axis represents Fourier amplitude (in m/s-s).

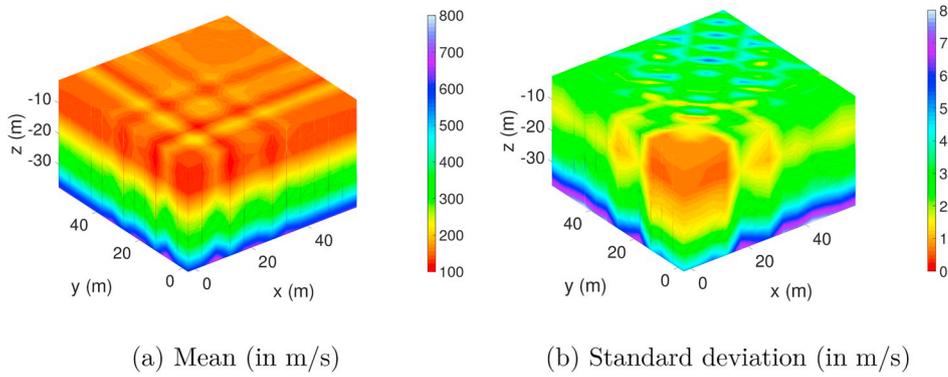


Fig. 27. Estimated mean and standard deviation of the soil's S-wave velocity,  $V_S$ .

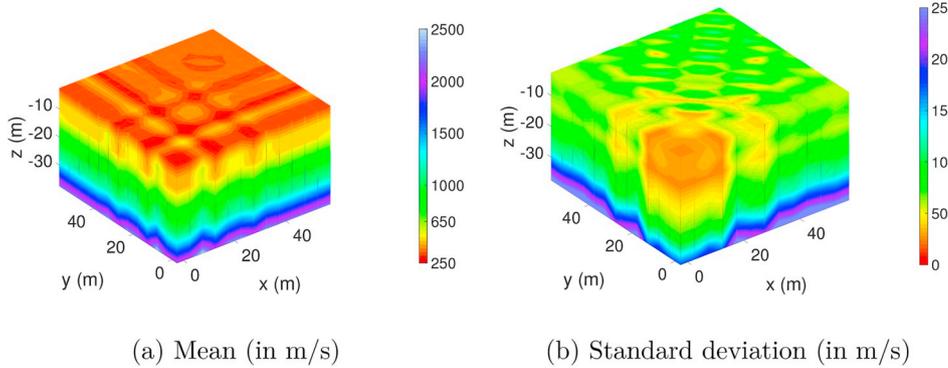


Fig. 28. Estimated mean and standard deviation of the soil's P-wave velocity,  $V_P$ .

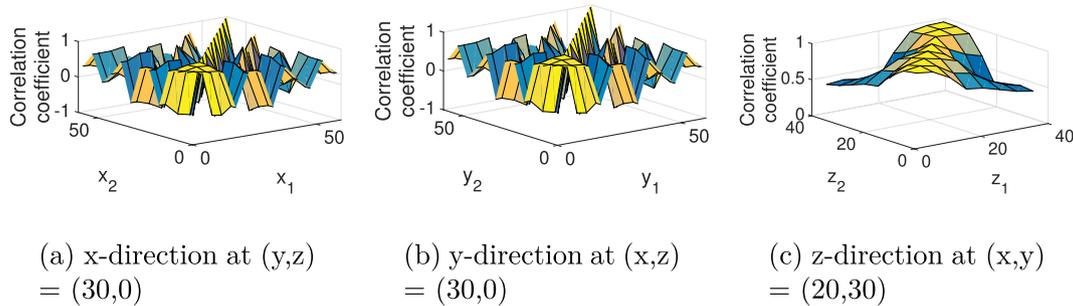


Fig. 29. Estimated correlation structure of the soil's S-wave velocity,  $V_S$ .

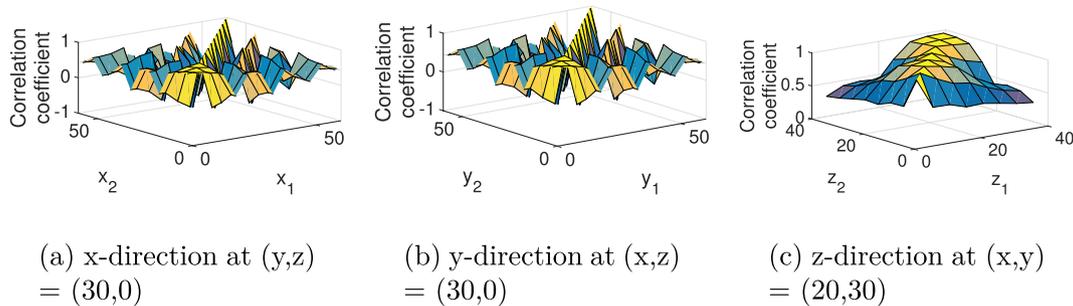


Fig. 30. Estimated correlation structure of the soil's P-wave velocity,  $V_P$ .

of such pockets signifies that the soil above the water table is probably saturated to some degree due to capillarity. The estimated correlation structure of  $V_P$  is similar to that of  $V_S$ , but there exist some differences due to more uniform (more correlated) nature of P-wave velocities of saturated soils; the degree of fluctuations of the correlation structure is less in the horizontal directions and also the “equivalent correlation length” in the vertical direction is larger below the water table.

### 5. Conclusions

A numerical algorithm for high-fidelity probabilistic imaging of soil's P- and S-wave velocities through stochastic full waveform inversion of geophysical measurements is validated using data available from a well-studied geotechnical site in Garner Valley, CA. The algorithm relies on fusing geophysical measurements with predictions from a high-fidelity numerical model of the geophysical experiment using a

minimum variance framework within the concept of Bayesian belief. The validation effort starts by setting up a numerical model of the field geophysical experiment using the finite element method. We, then, have hypothesized the soil properties at the site to be random fields, made prior assumptions on them based on local geological information, and probabilistically predicted the quantities of interest using a stochastic collocation technique. The collocation technique utilizes a Gaussian mixture approach in conjunction with a polynomial surrogate model to represent the random fields in terms of finite dimensional random variables and a non-product CUT quadrature rule to efficiently estimate the ensemble averages of the model predicted quantities of interest. Finally, we have fused the statistics of the model output with the field geophysical measurements using a best linear unbiased estimator to obtain the posterior estimates of the random field soil properties.

The key lessons learned through the validation effort towards accurate stochastic full waveform inversion of geophysical measurements are (a) model-data fusion in the time domain suffers due to failure of the ensemble averages of the model output to preserve the physical characteristics of dynamical systems; a frequency domain fusion, however, can overcome this drawback, and (b) fine-tuning the prior values of the random field soil properties – which are typically assumed using engineering judgment following local geological information – based on a probabilistic sensitivity analysis yields superior posterior estimates.

In closing, due to the use of the Gaussian mixture approach for the parameterization of the random field soil properties and of the non-product CUT technique for the model-data fusion, the stochastic full waveform inversion algorithm, validated herein, is highly scalable and would allow for very high-resolution imaging of soil properties without prohibitive computational cost.

## Acknowledgment

This material is based upon works supported by the National Science Foundation through award nos. NSF-CMMI-1417849 and NSF-CMMI-1054759.

## References

- [1] Heisey JS, Stokoe II KH, Meyer AH. Moduli of pavement systems from spectral analysis of surface wave, Transportation Research Record 852. Washington, D.C: Transportation Research Board; 1982. p. 22–31.
- [2] Kallivokas LF, Fathi A, Kucukcuban S, Stokoe II KH, Bielak J, Ghattas O. Site characterization using full waveform inversion. *Soil Dynam Earthq Eng* 2013;47:62–82.
- [3] Parida SS, Sett K, Singla P. An efficient PDE-constrained stochastic inverse algorithm for probabilistic geotechnical site characterization using geophysical measurements. *Soil Dynam Earthq Eng* 2018;109:132–49.
- [4] Ravela S. Quantifying uncertainty for coherent structures. *Procedia Comput Sci* 2012;9:1187–96.
- [5] Yang CM. Spectral control of viscous alignment for deformation invariant image matching PhD. dissertation Cambridge, MA: Massachusetts Institute of Technology; 2009.
- [6] Bielak J, Civilini F, Crempien J, Fathi A, Gee R, Hegarty P, Kallivokas L, Menq FY, Steidl J, Stokoe II KH. U. D., T-Rex shaking at Garner Valley - toward three-dimensional full-waveform inversion, Dataset. Network for Earthquake Engineering Simulation; 2013. <https://doi.org/10.4231/D3BK16P79>. distributor.
- [7] Steidl JH, Tumarkin AG, Archuleta RJ. What is a reference site? *Bull Seismol Soc Am* 1996;86:1733–48.
- [8] Youd TL, Bartholomew HAJ, Proctor JS. Geotechnical logs and data from permanently instrumented field sites: Garner Valley down hole array (GVDA) and wildlife liquefaction array (WLA), Technical Report. Brigham Young University; 2004 <http://nees.ucsb.edu/sites/eot-dev.nees.ucsb.edu/files/facilities/docs/geotech-data-report.pdf>, Accessed date: January 2018.
- [9] Steller R. New borhole geophysical results at GVDA Technical Report Santa Barbara: University of California; 1996 <http://neesucsb.edu/sites/eot-dev.nees.ucsb.edu/files/facilities/docs/GVDA-Geotech-Stellar1996.pdf>, Accessed date: January 2018.
- [10] Menq FY, Stokoe II KH, Park K, Rosenblad BL, Cox BR. Performance of mobile hydraulic shakers at nees@UTexas for earthquake studies. Proceedings of the 14th World conference on earthquake engineering, Beijing, China. 2008. October 12–17.
- [11] Fathi A, Poursartip B, Stokoe II KH, Kallivokas LF. Three-dimensional P- and S-wave velocity profiling of geotechnical sites using full-waveform inversion driven by field data. *Soil Dynam Earthq Eng* 2016;87:63–81.
- [12] Fathi A. Full-waveform inversion in three-dimensional PML-truncated elastic media: theory, computations, and field experiments, PhD dissertation. Austin, TX: University of Texas at Austin; 2015.
- [13] Bathe K-J. Finite element procedures. New Jersey: Prentice Hall; 1996.
- [14] Lysmer J, Kuhlmeyer RL. Finite dynamic model for infinite media. *J Eng Mech Div* 1969;95:859–77.
- [15] Babu ka IM, Nobile F, Tempone R. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J Numer Anal* 2007;43:1005–34.
- [16] Madankan R, Pouget S, Singla P, Bursik M, Dehn J, Jones M, Patra A, Pavolonis M, Pitman EB, Singh T, et al. Computation of probabilistic hazard maps and source parameter estimation for volcanic ash transport and dispersion. *J Comput Phys* 2014;271:39–59.
- [17] Adurthi N, Singla P. A conjugate unscented transformation based approach for accurate conjunction analysis. *AIAA J Guid Contr Dyn* 2015;38(Sep):1642–58.
- [18] Adurthi N, Singla P, Singh T. Conjugate unscented transformation: applications to estimation and control. *J Dyn Syst Meas Control* 2018;140:030907.
- [19] Xiu D, Karniadakis GE. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J Sci Comput* 2002;24:619–44.
- [20] Mavko G, Mukerji T, Dvorkin J. The rock physics handbook: tools for seismic analysis of porous media. second ed. Cambridge University Press; 2009.
- [21] Phoon K-K, Kulhawy FH. Characterization of geotechnical variability. *Can Geotech J* 1999;36:612–24.
- [22] Mazzoni S, McKenna F, Fenves GL. Open system for earthquake engineering simulation user manual, Pacific earthquake engineering research center. Berkeley, CA; University of California; 2006 <http://opensees.berkeley.edu>.
- [23] Woodward PK, Griffiths DV. Influence of viscous damping in the dynamic analysis of an earth dam using simple constitutive models. *Comput Geotech* 1996;19:245–63.
- [24] Anderson BDO, Moore JB. Optimal filtering. Prentice-Hall; 1979.
- [25] Pearson K. Note on regression and inheritance in the case of two parents. *Proc Roy Soc* 1895;58:240–2.
- [26] Székely GJ, Rizzo ML. Brownian distance covariance. *Ann Appl Stat* 2009;3/4:1233–303.
- [27] Pouget S, Bursik M, Singla P, Singh T. Sensitivity analysis of a one-dimensional model of a volcanic plume with particle fallout and collapse behavior. *J Volcanol Geotherm Res* 2016;326:43–53.
- [28] Baecher GB, Christian JT. Reliability and statistics in geotechnical engineering. second ed. West Sussex PO19 8SQ, England: Wiley; 2003.