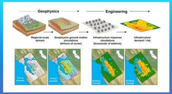


PEER International Pacific Rim Forum
June 16-17, 2021

Spatial and statistical characterization of P- and S-wave velocities of soils from geophysical measurements

Kallol Sett
University at Buffalo

in collaboration with
Siddharth S. Parida (Embry-Riddle) and Puneet Singla (Penn State)

University at Buffalo
The State University of New York

June 17, 2021

PEER

Outline

Objective: To highlight the lessons learned in applying a recently developed stochastic full waveform inversion algorithm to characterize a geotechnical site using geophysical measurements

Stochastic full waveform inversion algorithm

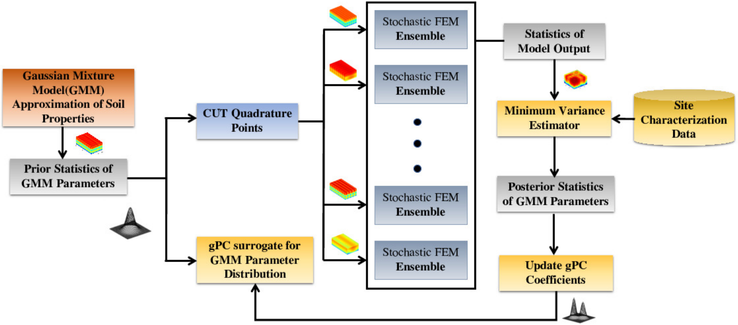
- Minimum variance framework
  - Gaussian mixture model (GMM) for parameterization of soil properties
  - Generalized polynomial chaos (gPC) for representing the PDFs of GMM parameters
  - Non-product quadrature rule for evaluation of high-dimensional expectation integrals

Results and discussions

- Testbed site: a 60m$\times$60m$\times$40m site in Garner Valley, CA
- Lessons learned
- Comments on scalability of the algorithm for larger scale geophysical imaging

A flowchart of the algorithm

It, essentially, fuses sparse experimental measurements with FE simulated measurements using a minimum variance framework



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

$$\mathbf{\Sigma}^+_{\Theta\Theta} = \mathbf{\Sigma}^-_{\Theta\Theta} - \mathbf{K}\mathbf{\Sigma}_{\Theta h}$$

$$\mathbf{K} = \mathbf{\Sigma}^T_{\Theta h}(\mathbf{\Sigma}^-_{hh} + \mathbf{R})^{-1}$$

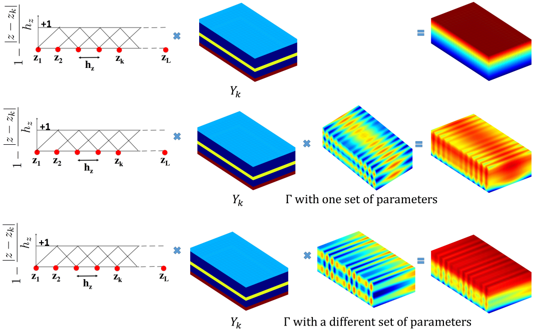
$^+$: Posterior
$^-$: Prior
$\mathcal{E}$: Expected value
$\mathbf{\Sigma}$: Variance-covariance matrix
$\mathbf{\Theta}$: Uncertain soil properties
$\mathcal{Z}$: Experimental measurements
$h(u(\cdot))$: FE simulated measurements
$\mathbf{K}$: Kalman gain matrix
$\mathbf{R}$: Noise covariance matrix

Parameterization of soil properties in the domain of interest

- Soil deposits exhibit spatial variability: nonuniform layers and lenses within layers
  - Uncertain soil properties are increasingly being modeled as random fields
- Conventional random field parameterization approaches use the Karhunen-Loeve expansion for spatial discretization:
  - Mathematically optimal, but..
  - Requires repeated use of very large scale eigenanalysis

GMM for parameterization of random field soil properties

- Hat functions mimic horizontal layering of a soil deposit
- Gaussian functions mimic presence of non-uniformity and lenses within layers



$$Y(x,y,z) = \sum_{i=1}^{N_G} \alpha(z)\Gamma_i(x,y,z)$$

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

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

α: 1-D, piece-wise linear hat function
Γ: 3-D, Gaussian function
L: Number of hat functions
N_G: Number of Gaussian functions
Parameters of α and Γ are random variables

gPC for representing the PDFs of GMM parameters

- Denote the parameters (random variables) of the GMM model by $\boldsymbol{\Theta}$:
  $Y(x, y, z, \boldsymbol{\Theta}) = \sum_{i=1}^{N_G} \alpha(z, \boldsymbol{\Theta}_1)\Gamma_i(x, y, z, \boldsymbol{\Theta}_2)$

  $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2\} = \{\{h_z, Y_1, Y_2, \ldots, Y_k\}, \{l_x, l_y, l_z\}\}$

- Represent $\boldsymbol{\Theta}$ in terms of standardized random vector, $\boldsymbol{\xi}$:
  $\boldsymbol{\Theta} = f(\boldsymbol{\xi})$ where $\boldsymbol{\xi} = [\xi_1, \xi_2, \cdots, \xi_m]^T \in \mathbb{R}^m$

- Write $\boldsymbol{\Theta}$ in terms of orthogonal polynomial basis function, $\phi_k(\boldsymbol{\xi})$ (Xiu and Karniadakis 2003):

$$\Theta_j(\boldsymbol{\xi}) = \sum_{k=0}^{N} c_{jk}\phi_k(\boldsymbol{\xi}), \quad j = 1, 2, \cdots, m \quad \text{where } c_{jk} = \frac{\mathcal{E}[\Theta_j(\boldsymbol{\xi})\phi_k(\boldsymbol{\xi})]}{\mathcal{E}[\phi_k(\boldsymbol{\xi})\phi_k(\boldsymbol{\xi})]}$$

Evaluation of expectation integrals

- The algorithm involves 2 types of integrals:
  - Stochastic collocation integrals for estimating the statistics of FE simulated measurements:

    $$\mathcal{E}[u_{k_i}^N] = \int_{\Theta} u_{k_i}^N p(\Theta) d\Theta = \int_{\xi} u_{k_i}^N p(\xi) d\xi \simeq \sum_{q=1}^{M} w_q u_{k_i}^N(\Theta(\xi^q)), \quad N = 1, 2, \cdots$$

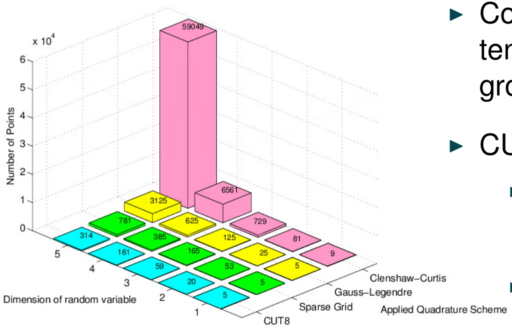
    (u_{k_i}: soil displacement at an FE node i at time step k; N: order of statistical moment; $\Theta(\xi^q)$ and w_q: q^{th} quadrature point and its associated weight)

  - Minimum variance integrals ($\boldsymbol{\Sigma}_{\Theta h}$ and $\boldsymbol{\Sigma}_{hh}$):
    e.g., $\boldsymbol{\Sigma}_{\Theta h} = \mathcal{E}^-[(\Theta - \mathcal{E}^-[\Theta])(h(u(\Theta)) - \mathcal{E}^-[h(u(\Theta))])^T]$

    $$= \sum_{q=1}^{M} w_q \left(\Theta(\xi^q) - \mathcal{E}^-[\Theta]\right) \left(h(u(\Theta(\xi^q))) - \mathcal{E}^-[h(u(\Theta))]\right)^T$$

- Both types suffer from the curse of dimensionality as the number of random variables, ξ, grows

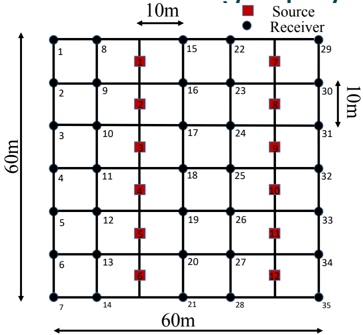
Conjugate Unscented Transformation (CUT) for evaluation of expectation integrals



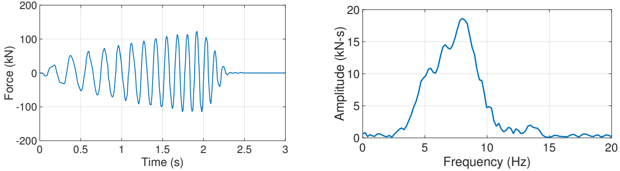
- Conventional quadrature rules (e.g., Gauss) rely on tensor product of 1-D quadrature points: exponential growth in computational cost
- CUT yields a non-product rule:
  - An extension of conventional unscented transformation which satisfies additional higher order moment constraints
  - Alleviates the computational burden associated with evaluation of high-dimensional integrals

Adurthi, N., Singla, P., and Singh, T., "Conjugate Unscented Transformation: Applications to Estimation and Control", *ASME Journal of Dynamic Systems, Measurement, and Control*, Vol. 140, No. 3, pp. 030907-1-22, 2018

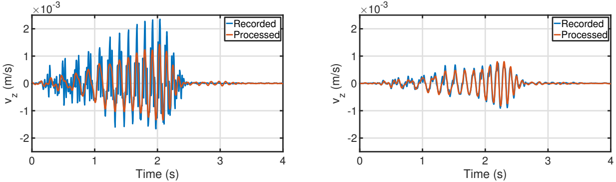
Testbed site and geophysical experiment



- A 60m × 60m parcel of NEES@UCSB Garner Valley site
- Geophysical experiment was performed by UT Austin group



Source # 3 ground excitation: time history and Fourier spectrum

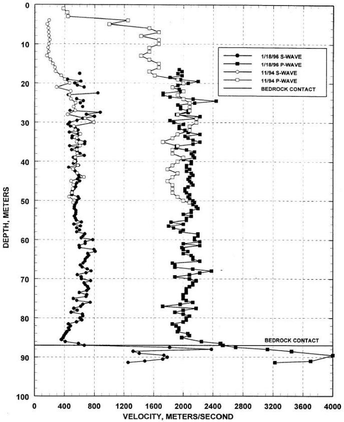


Soil responses at Receivers # 10 (left) and 2 (right)

Testbed site: a 60m × 60m × 40m site in Garner Valley, CA

Prior assumptions for FE simulation

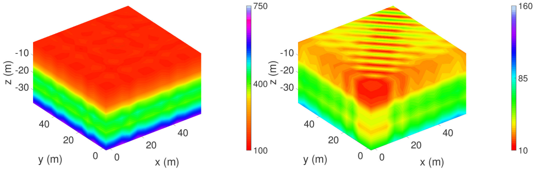
- Based on information on local geology and PS suspension logging available at a nearby site
  - 2 horizontal layers up to a depth of 40m
    - Relatively loose sand layer up to a depth of 20m
    - Denser sand or weathered rock below
  - Proximity to Lake Hemet: shallow water table
- V_S: 3 hat functions with centers at 0m, 20m, and 40m
- V_P: 2 hat function with centers at the surface and at the depth of the water table below which it is represented as a function of ν and V_S



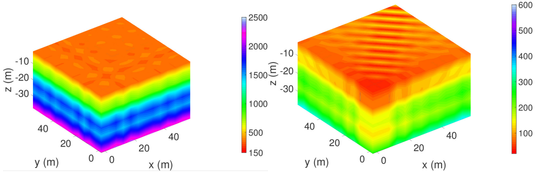
V_S (left) and V_P (right) measurements available at a nearby site

Testbed site: a 60m × 60m × 40m site in Garner Valley, CA

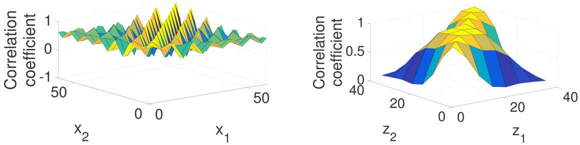
Prior assumptions for FE simulation



V_S: mean (left) and standard deviation (right)



V_P: mean (left) and standard deviation (right)

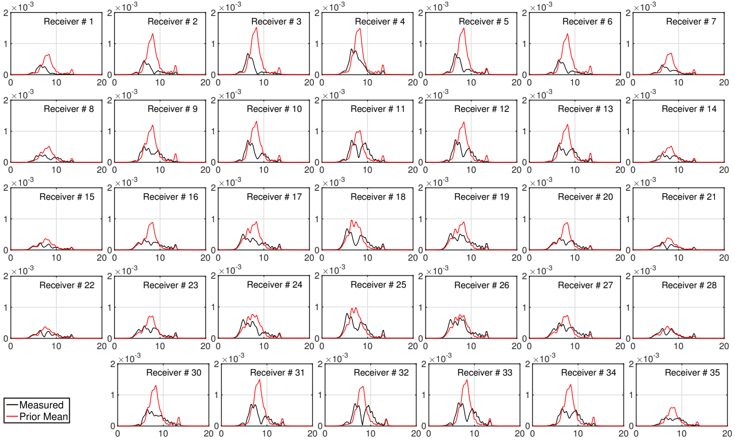


V_S correlation structure: (left) x-direction at (y,z) = (30,0) and (right) z-direction at (x,y) = (20,30)

RV	Type	Mean	COV (%)	Supports
$\Theta_1 = V_{S_0}$	Uniform	200m/s	10	[166, 234]m/s
$\Theta_2 = V_{S_{20}}$	Uniform	500m/s	10	[413, 587]m/s
$\Theta_3 = V_{S_{40}}$	Uniform	700m/s	10	[578, 821]m/s
$\Theta_4 = V_{P_0}$	Uniform	500m/s	10	[413, 587]m/s
$\Theta_5 = H_w$	Uniform	6m	20	[4, 8]m
$\Theta_6 = \nu_{below\ H_w}$	Uniform	0.46	1	[0.452, 0.468]
$\Theta_7 = l_x = l_y = l$	Uniform	9.5m	10	[7.885, 11.115]m

Testbed site: a 60m × 60m × 40m site in Garner Valley, CA

FE simulated **prior mean** versus experimental measurements



- FE simulation was performed using the conventional approach in OpenSees
  - 8-noded brick elements
  - Rayleigh damping (2%)
- Mean was calculated using CUT (only 551 FE runs were needed for this 7 “dimensional” problem)

Posterior estimates

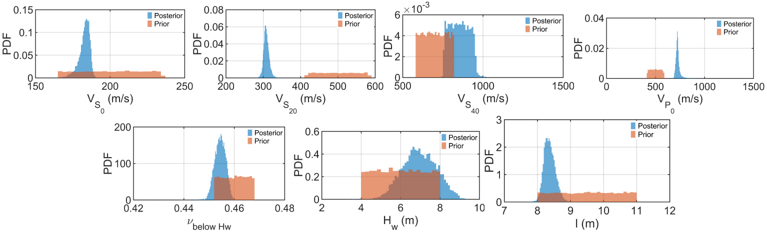
Time domain inversion

RV	Mean	Standard deviation
$\Theta_1 = V_{S_0}$	274.9m/s	0.013m/s
$\Theta_2 = V_{S_{20}}$	488.2m/s	0.047m/s
$\Theta_3 = V_{S_{40}}$	1664.2m/s	0.241m/s
$\Theta_4 = V_{P_0}$	514.6m/s	0.118m/s
$\Theta_5 = H_w$	27.8m	0.014m
$\Theta_6 = \nu_{below\ H_w}$	0.56	0.0001
$\Theta_7 = l$	10.5m	0.0004m

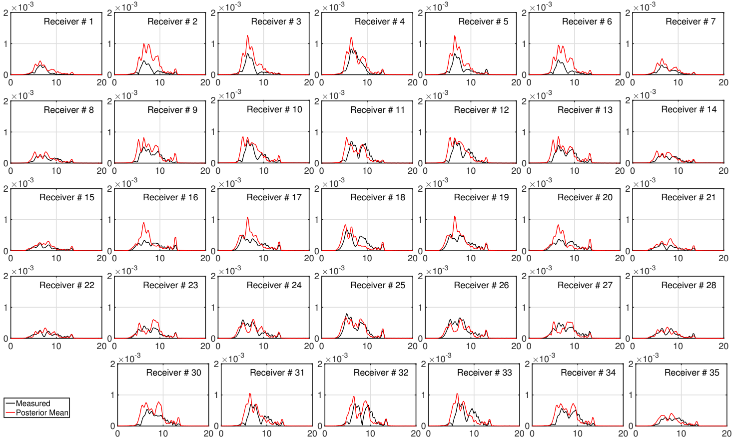
- Time domain inversion didn't work due to failure of the ensemble averages of the model output to preserve the physical characteristics of dynamical systems

Frequency domain inversion

RV	Mean	Standard deviation	COV (%)	Supports
$\Theta_1 = V_{S_0}$	182.6m/s	3.9m/s	2.1	[160.5,192.2]m/s
$\Theta_2 = V_{S_{20}}$	308.3m/s	7.8m/s	2.5	[274,348.5]m/s
$\Theta_3 = V_{S_{40}}$	854.2m/s	58.7m/s	6.8	[760.9,1031]m/s
$\Theta_4 = V_{P_0}$	728.4m/s	29.3m/s	4.0	[676.8,855]m/s
$\Theta_5 = H_w$	6.9m	0.8m	12.51	[4.4,9.6]m
$\Theta_6 = \nu_{below\ H_w}$	0.454	0.002	0.5	[0.44,0.46]
$\Theta_7 = l$	8.3m	0.2m	2.1	[7.8,9.2]m

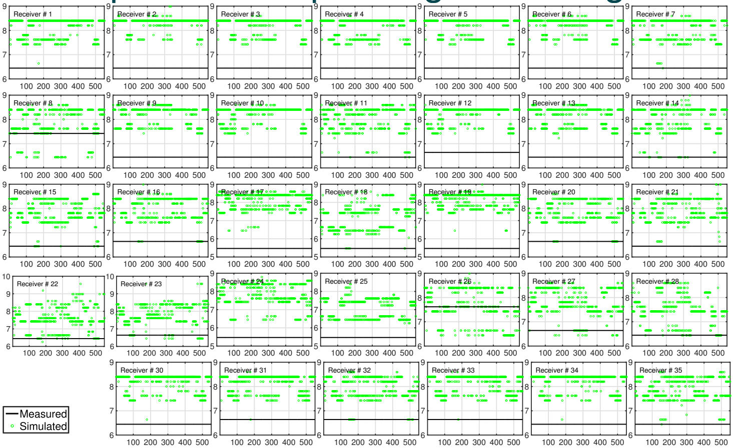


FE simulated **posterior mean** versus experimental measurements



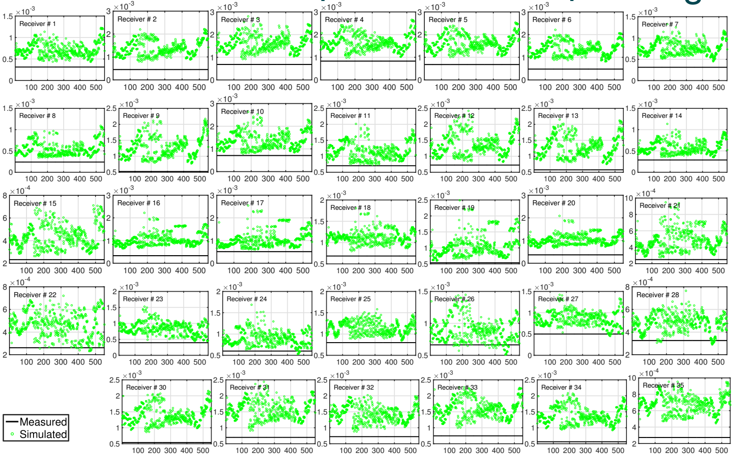
- Reasonable match at most (but not all) of the receiver locations
- Can we do better?

Were prior assumptions good enough?



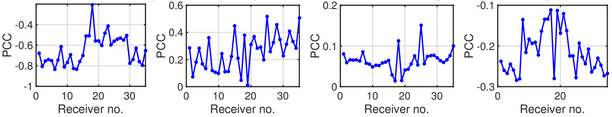
- Realizations (551 in total) of simulated peak frequency are compared with measured peak frequency at each receiver location
- Ideally the realizations should center around the measurement, but at most of the receiver locations they are not.

How about the amplitudes corresponding to the peak frequency?

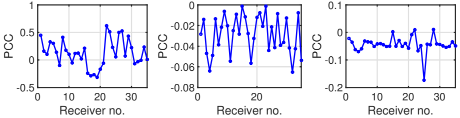


- They didn't behave ideally either!
- Can we do a better job with the prior assumption?
  - Fourier amplitudes (and peak frequencies) of simulated measurements should be smaller
  - But which **Θ** (GMM parameters) to adjust?

Sensitivity analysis for “improving” the prior assumptions



(a) $\Theta_1 = V_{S_0}$ (b) $\Theta_2 = V_{S_{20}}$ (c) $\Theta_3 = V_{S_{40}}$ (d) $\Theta_4 = V_{P_0}$



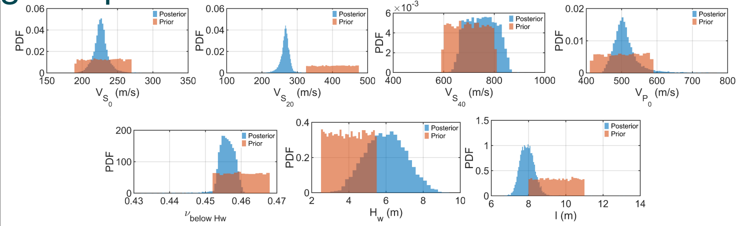
(e) $\Theta_5 = H_w$ (f) $\Theta_6 = \nu_{below\ H_w}$ (g) $\Theta_7 = I$

Pearson correlation coefficient (PCC) between prior **Θ** and FE simulated Fourier amplitude corresponding to the peak frequency at each receiver location

- As prior Θ_1 is negatively correlated with the simulated Fourier amplitude, increasing Θ_1 is expected to reduce the values of the realizations of the simulated Fourier amplitude
- Similarly, decreasing prior Θ_2 and Θ_5 is expected to reduce the values of the realizations of the simulated Fourier amplitude

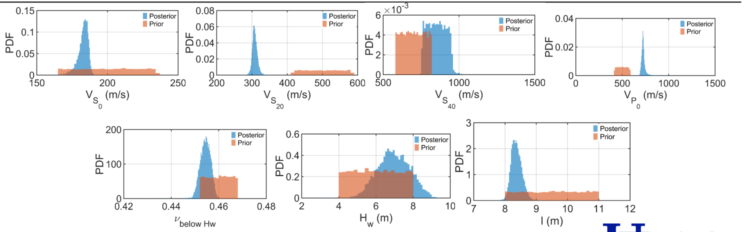
“Improved” versus “original” posterior estimates

“Improved”

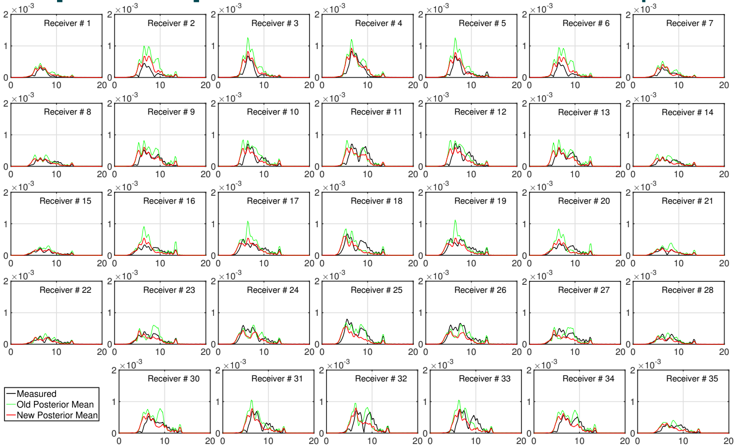


V_{S_0} (↑) and V_{P_0} (↓) moved the most

“Original”

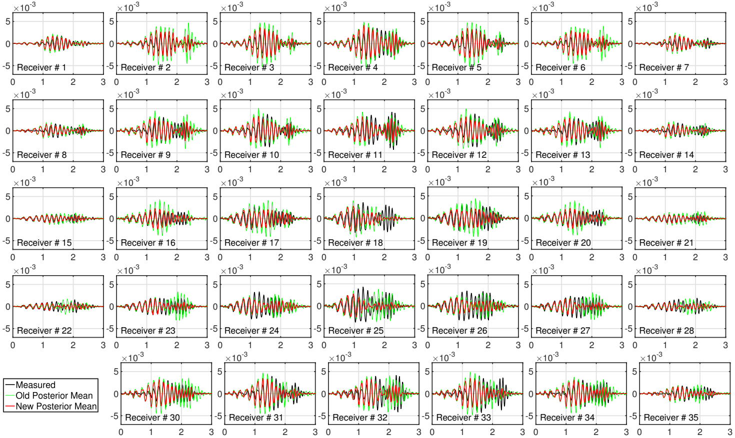


Improved posterior mean versus experimental measurements



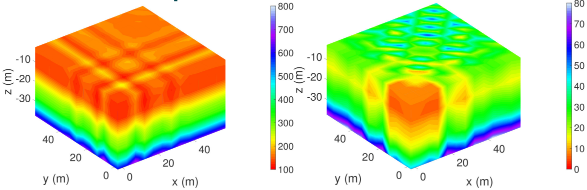
- Frequency domain comparison
- Better match at all the receiver locations compared to the "original" case

Improved posterior mean versus experimental measurements

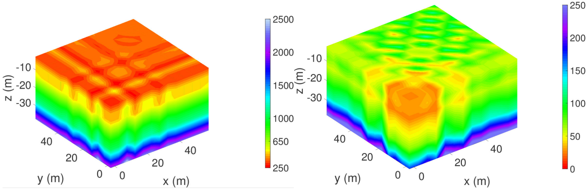


- Time domain comparison
- The updated prior for soil parameters leads to the FE ensembles that better capture measurement data as their subset and hence leads to more accurate soil parameter estimates

Posterior probabilistic 3D profiles



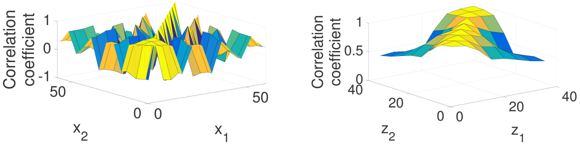
V_S (m/s): mean (left) and standard deviation (right)



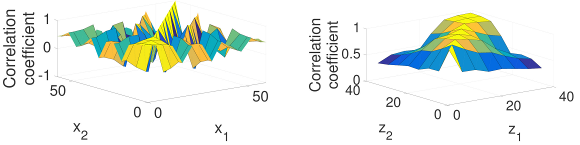
V_P (m/s): mean (left) and standard deviation (right)

- Nonuniform layers and lenses within layers:
  - V_S: Marginal COVs are larger at and close to the surface (around 20%) and they become smaller with depth
  - V_P: Consistent with that of partially saturated soils, pockets of larger values – albeit with large COVs – between the surface and the water table

Posterior correlation structures



V_S correlation structure: (left) x-direction at (y,z) = (30,0) and (right) z-direction at (x,y) = (20,30)



V_P correlation structure: (left) x-direction at (y,z) = (30,0) and (right) z-direction at (x,y) = (20,30)

- The fluctuating horizontal correlations 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 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

Comments on scalability of the algorithm

- Both GMM and CUT are promising for larger scale geophysical imaging without prohibitive computational cost:
  - GMM bypasses large scale eigenanalysis by approximating the eigenfunctions of the KL expansion
  - CUT alleviates the computational burden associated high-dimensional expectation integrals
- However, the algorithm uses the conventional FE method for numerically simulating geophysical measurements
  - An efficient approach is needed for modeling wave propagation through soils; ideally a reduced order model

Thank you!