SANDIA REPORT
SAND2021-10885
Printed September 2021

Sandia National Laboratories

LDRD #218329: Uncertainty Quantification of Geophysical Inversion Using Stochastic Partial Differential Equations

Leiph A. Preston and Christian Poppeliers

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

This report summarizes work completed under the Laboratory Directed Research and Development (LDRD) project "Uncertainty Quantification of Geophysical Inversion Using Stochastic Differential Equations." Geophysical inversions often require computationally expensive algorithms to find even one solution, let alone propagating uncertainties through to the solution domain. The primary purpose of this project was to find more computationally efficient means to approximate solution uncertainty in geophysical inversions. We found multiple computationally efficient methods of propagating Earth model uncertainty into uncertainties in solutions of full waveform seismic moment tensor inversions. However, the optimum method of approximating the uncertainty in these seismic source solutions was to use the Karhunen-Loève theorem with data misfit residuals. This method was orders of magnitude more computationally efficient than traditional Monte Carlo methods and yielded estimates of uncertainty that closely approximated those of Monte Carlo. We will summarize the various methods we evaluated for estimating uncertainty in seismic source inversions as well as work toward this goal in the realm of 3-D seismic tomographic inversion uncertainty.

ACKNOWLEDGEMENTS

This work was funded by the Laboratory Directed Research and Development program, Geoscience Research Foundation. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

CONTENTS

LIST OF FIGURES

This page left blank

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
1-D	one dimensional
3-D	three dimensional
FDMC	finite-difference Monte Carlo
GF	Green's function
H&G	Hallo and Gallovic
Hz	Hertz
KL	Karhunen-Loève
KLMC	Karhunen-Loève Monte Carlo
km	kilometer
L2	Euclidean norm
LDRD	Laboratory Directed Research and Development
m	meter
MC	Monte Carlo
P-wave	primary (compressional) wave
PCE	polynomial chaos expansion
s	second
S-wave	shear wave
STF	source time function
SVD	singular value decomposition
UQ	uncertainty quantification
Vp	P-wave velocity
Vs	S-wave velocity

This page left blank

1. INTRODUCTION

Many quantities of interest in geophysics are inferred from observed data instead of directly predicted from first principles. For example, the magnitude of an earthquake and how that fault moved during rupture are inferred via geophysical inversion of observed seismograms. Often, however, uncertainties in these inferred quantities of interest are poorly known or not even provided. One reason for this lack of uncertainty quantification (UQ) in geophysical inverse solutions is due to the computational expense of propagating uncertainty of the model and/or data through a complex geomodel to final output. Currently, Monte Carlo (MC) and Bayesian methods are computationally infeasible for many realistic models due to the expense of individual runs and large parameter spaces.

The primary purpose of this LDRD project was to explore methods of propagating uncertainty in geophysical inversion in a more computationally efficient way. We investigated using polynomial chaos expansions (PCEs) of probability distributions due to their succinctness in describing many common probability distributions. We also used the Karhunen-Loève (KL) theorem to represent stochastic processes while minimizing the number of parameters needed to represent that process. In this report we summarize results that have been more fully documented in other SAND reports or a peer-reviewed journal as well as work that we did using polynomial chaos in an attempt to more efficiently parametrize uncertainty in geophysical structural inversion. In the remainder of this chapter, we briefly outline polynomial chaos representations and the KL theorem since they are used throughout this report. In the following chapters we discuss using waveform covariance methods, KL representations of Green's functions, KL reconstructions of inversion residuals, and end with PCE representations for structural inversions.

1.1. Karhunen-Loève Theorem

The Karhunen-Loève theorem states that any second order random process can be decomposed into a series of spatial basis functions with random coefficients (e.g. Le Maitre and Knio, 2010), or:

$$D(\mathbf{x},\theta) = D_0(\mathbf{x}) + \sum_{i>=1} r_i(\theta)\sqrt{\lambda_i}u_i(\mathbf{x}) \tag{1-1}$$

where $D(\mathbf{x},\theta)$ is a realization of a stochastic process, $D_0(\mathbf{x})$ is the mean of the random process, $r_i(\theta)$ are uncorrelated random variables with zero mean and unit standard deviation that span the random space for an event θ; λ_i and $u_i(\mathbf{x})$ are, respectively, the eigenvalue and eigenfunction pairs of the covariance function of the process with respect to (typically) space or time variables $\mathbf{x}$. Thus, if we know or can assume a form for the correlation function among space variables, we can produce a realization that adheres to that correlation function by summing properly normalized randomly scaled eigenfunctions of that correlation function. For continuous variables, analytical expressions for the eigenfunctions and eigenvalues are only known for relatively few covariance functions. However, in the discrete domain, the eigenvectors and eigenvalues can be numerically evaluated for a wide variety of covariance matrices, assuming those matrices actually adhere to the rules of covariance matrices, such as possessing positive definiteness. However, numerically solving for eigenvector and eigenvalues can easily become computationally burdensome if not intractable for multidimensional domain covariance matrices unless some simplifying assumptions can be made, such as wide-sense stationarity or circularity at domain boundaries like is assumed for discrete Fourier transforms. In fact, if the covariance function is a wide-sense circular stationary signal, the eigenfunctions are the Fourier bases and eigenvalues are the discrete Fourier transform of the correlation function (Mallat, 2009). More information on covariance functions and the KL theorem can be found in Preston (2018).

1.2. Polynomial Chaos Expansion

Chaos polynomials are sets of orthogonal polynomials that can be used to compactly represent probability distribution functions. Any probability function for which the chaos polynomial has support can be expanded as a (perhaps infinite) sum of scaled (orthogonal) chaos polynomials of different orders, i.e.

$$v(\mathbf{x}) = \sum_{i=0} V_i(\mathbf{x}, \theta)\Phi_i(\theta)$$

where $v(\mathbf{x})$ is a realization of a random event in the space or time domain $\mathbf{x}$, $V_i(\mathbf{x}, \theta)$ is the coefficient for chaos polynomial $\Phi_i(\theta)$ and random event θ. The coefficients $V_i(\mathbf{x}, \theta)$ depend on the specific probability distribution function that describes the random process. This is called polynomial chaos expansion (PCE), and this can be a means of compactly parameterizing fairly complex probability distributions in certain cases.

There are multiple classes of chaos polynomials and the optimal class depends on the known or expected form of the probability distribution one is trying to represent. Each class has different supports and weight functions that make certain choices optimal for certain distribution functions (Xiu and Karniadakis, 2003). For example, a Gaussian distribution function is best represented with Hermite chaos polynomials since the weight function is a unit Gaussian with support along the entire real line. With Hermite polynomials any Gaussian distribution can be represented by two parameters, which are the coefficients for Hermite polynomials up to first order. A gamma distribution, on the other hand, is optimally represented by Laguerre chaos polynomials since, again, only two parameters are necessary to describe any gamma distribution with up to first-order Laguerre polynomials.

2. WAVEFORM COVARIANCE METHOD

This chapter summarizes the work described in Poppeliers and Preston (2021b), where we propagated model uncertainty into estimates of seismic source time functions using an estimate of model covariance. As with our other methods presented in this report, we employ a Monte Carlo scheme to invert seismic waveforms. The inversion is linear, and assumes a point seismic source that is modeled as a set of six independent source time functions.

For several of the Monte Carlo based inversion schemes that we describe in this report, the basic idea is that the Earth model uncertainty is 'captured' by a probability distribution of the seismic Green's functions (GFs). To propagate the model uncertainty, we simply invert the data N times, where for each inversion we select a set of GFs from the probability distribution of the GFs. In total, we explored three different methods of constructing the probability distribution of GFs:

1. Based on an assumption of Earth model uncertainty, we constructed a suite of stochastic Earth models and used a computationally expensive finite difference calculation to calculate a set of GFs for each realization of the Earth model (details of this work are in Poppeliers and Preston, 2020). Throughout this report, we refer to this method as the finite-difference Monte Carlo (FDMC) approach.
2. Given only a limited set of seismic GFs, estimated using finite difference simulations and stochastic Earth models, we used a Sandia-developed code called Dakota (Adams et al., 2019) to estimate the model covariance. Using the Dakota-estimated covariance matrices, we employed the Karhunen-Loève (KL) theorem to build up a suite of surrogate, stochastic GFs. Using the stochastic GFs, we then inverted data as before (see Chapter 3 in this report; Poppeliers and Preston, 2021a). We refer to this method as the Dakota-KL method.
3. In this chapter we describe an additional method to construct the stochastic GF distribution, but with only a single finite difference simulation. The method, which we term the H&G method, proceeds by constructing approximate covariance matrices and is based on simplifying assumptions of the model and its uncertainty (Hallo and Gallovic, 2016). An attractive aspect of the H&G method is that it assumes that the uncertainty in the model is expressed as variations about the mean wave speed model, which makes it appropriate for stochastic modeling. Using the approximate covariance matrices, we then construct the suite of stochastic GFs using the KL theorem and use those GFs for a Monte Carlo inversion (see Poppeliers and Preston, 2021b, for details).

The H&G method assumes that uncertainties in the Earth model lead to time shifts of discrete seismic arrivals. If the uncertainties are Gaussian-distributed, then these time shifts are uniformly distributed, with variance that is directly related to model uncertainty. This information is used to approximate the auto- and cross-covariance functions, which are subsequently used to build up an approximate covariance matrix for each station and GF component (Figure 2-1). The H&G-estimated covariance matrix $\overline{\mathrm{xcov}}(\tau)$ is decomposed using singular value decomposition

$$\left[\mathbf{U}_{\mathrm{sim}}\mathbf{\Sigma}_{\mathrm{sim}}\mathbf{V}_{\mathrm{sim}}\right] = \overline{\mathrm{xcov}}(\tau)$$

where $\mathbf{U}_{\mathrm{sim}}$ is a matrix whose columns contain orthogonal eigenfunction that correspond to the singular values contained on the diagonal of matrix $\mathbf{\Sigma}_{\mathrm{sim}}$. The subscript 'sim' denotes that the eigenfunction/eigenvalue pairs are simulated using the H&G method. Using these eigenpairs, we build up a suite of stochastic GFs using the KL theorem

$$g_{kn}^{\text{stoch}}(\mathbf{x}', t'; \mathbf{x}, t, \theta) = g_{kn}^{\text{base}}(\mathbf{x}', t; \mathbf{x}, t) + \sum_{i=1}^{J} n_i(\theta)\sqrt{\lambda_i}\,\mathbf{u}(t)$$

where $g_{kn}^{\text{stoch}}(\mathbf{x}', t'; \mathbf{x}, t, \theta)$ is a stochastic GF for station k located at $\mathbf{x}'$, from moment tensor component n located at $\mathbf{x}$, with non-random component t and random event θ, $n_i(\theta)$ are uncorrelated random variables with zero mean and unit standard deviation, and $\mathbf{u}_i(t)$ and λ_i are the eigenpairs obtained from $\mathbf{U}$ and $\boldsymbol{\Sigma}$. $g_{kn}^{\text{base}}(\mathbf{x}', t'; \mathbf{x}, t)$ is a 'base' GF, which we assume is the mean GF of the distribution and can be obtained using a single GF estimate through a single realization of a stochastic Earth model, or an Earth model that contains no stochastic heterogeneity.

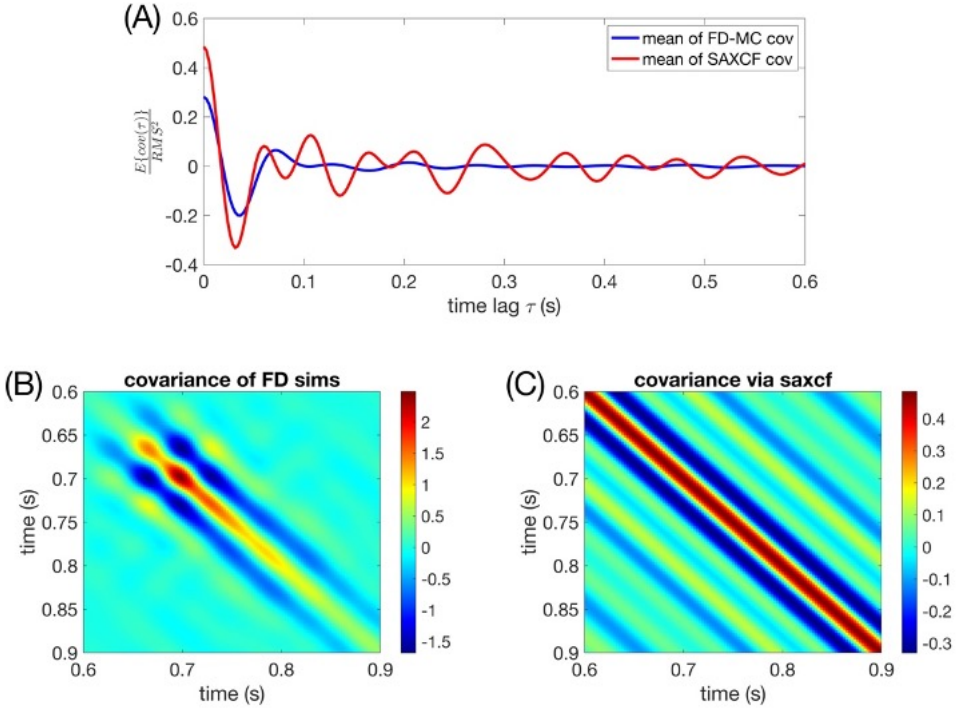

Figure 2-1: Examples of covariances. Panel (A) shows the mean (stationarized) autocovariance functions estimated from 1,000 finite difference simulations (blue) and from the H&G method (red) for an Earth model with stochastic heterogeneities. This result is for a single station and single GF component. Panels (B) and (C) show the covariance matrices for the FDMC simulations and the H&G method, respectively

3. APPROXIMATING GREEN FUNCTION UNCERTAINTY USING AN UNCERTAIN EARTH MODEL

For this work, we devised a computationally efficient method to approximate the probability distribution of seismic Green's functions (GFs) given the uncertainty of an Earth model. The motivation for this work was to be able to propagate the uncertainty of an Earth model into simulated seismic GFs. Seismic GFs are often required for several analysis methods, such as ground motion predictions, estimating seismic source parameters, seismic travel time calculations, or forward modeling to help guide interpretation. The details of this method are described in Poppeliers and Preston (2021a).

Our method is based on the Karhunen-Loève (KL; Section 1.1 and Equation 1-1) theorem and an approximation of the GF (or seismogram) covariance. To obtain the required basis functions, we first estimate the model covariance using a hybrid Monte Carlo approach. Specifically, we begin with an Earth model with known, large scale, geologic structure. Then, for each (of a few tens) Monte Carlo simulations, we superpose a stochastic distribution of impedance heterogeneities that obey a von Karman distribution. The von Karman distribution is assumed to mimic the high wavenumber heterogeneities in the Earth, which we assume is the primary form of uncertainty. Note that the statistics (i.e. the correlation length) of the stochastic heterogeneities are identical for each model,

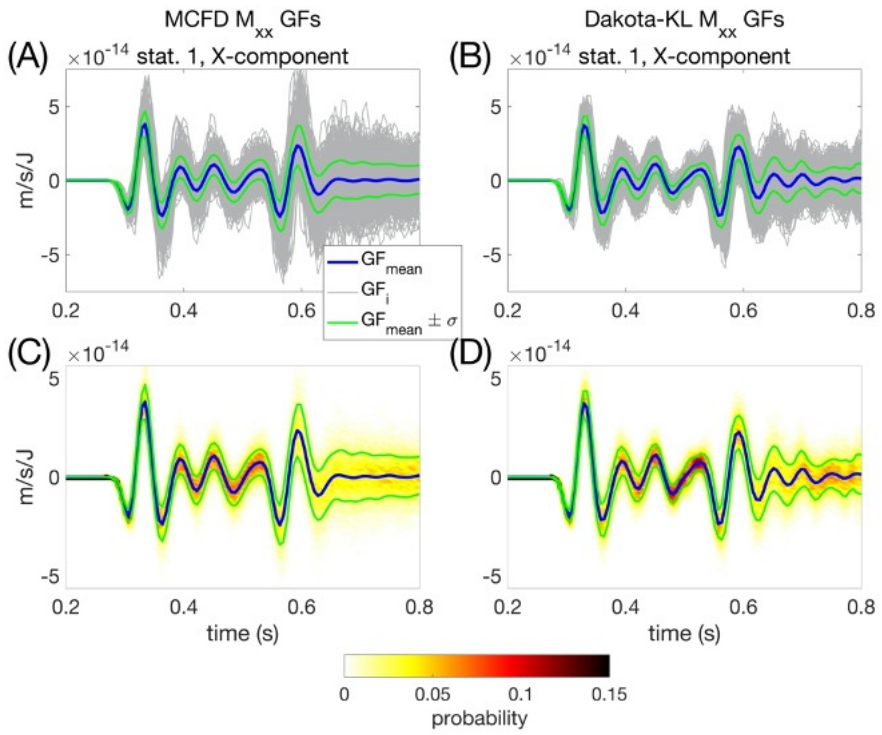

Figure 3-1: Panel A: 1,000 Green's functions (grey) computed using 1,000 stochastic Earth models, where for each model the GFs were estimated using finite differences (i.e. FDMC method), with the mean GF shown in blue. The standard deviation about the mean GF are shown in green. Panel B: 1,000 Green's functions, estimated with the Dakota-KL method. For display purposes, we've filtered the GFs with a 30 Hz lowpass filter. Note that the distribution appears similar to that produced by the Monte Carlo method, but appears more ringy at t > 0.6 seconds. Also, the width of the standard deviation is slightly less than that of the Monte Carlo derived distribution, for this particular Green's function. The probability density functions of the GFs produced by the FDMC method and the Dakota-KL method are shown in (C) and (D), respectively.

but the individual realization of each stochastic realization is unique. Using a finite difference solution to the elastic wave equation, we simulate a complete set of seismic GFs for each model. We then use the suite of GFs as input to the Sandia-developed code Dakota (Adams et al., 2019; dakota.sandia.gov) to estimate the covariance of the GFs. Using singular value decomposition (SVD), we decompose the GF covariance into their eigenfunction/eigenvalue pairs, where we use the eigenvectors as the basis functions, which are scaled by the eigenvalues. We use the SVD-derived basis functions in the KL theorem to estimate a suite of surrogate GFs, which we term a GF probability distribution. We call this approach the Dakota-KL method.

To test the efficacy of our method, we compare the Dakota-KL derived GF probability distribution to those produced by more traditional Monte Carlo methods. Similar to the method described above, we produce a suite of GFs from 1,000 realizations of a stochastic Earth model and estimate the GFs for each model using finite differences. Comparing the resulting FDMC GFs with those produced using our Dakota-KL method, we show that the GF probability distributions are virtually identical, especially for direct arrival body waves. However the accuracy of the KL-based method generally decreases for later times in the simulated Green's function distribution (Figure 3-1).

4. KL-INVERSION

For this work, we developed a method to efficiently propagate approximate uncertainty when linearly inverting seismic data for the source time functions (STFs) corresponding to the six independent components of the seismic moment tensor. In previous work, we used an estimate of Earth model uncertainty and Monte Carlo tests to propagate model uncertainty for this type of inversion, but this method presented two major challenges. First, we required an estimate of Earth uncertainty to build up a stochastic representation for Monte Carlo simulations. Secondly, for each Monte Carlo simulation, we had to compute a set of seismic Greens functions (GFs) using a computationally expensive finite difference calculation. The resulting suite of GFs were used to invert the data, once for each Monte Carlo step. The final result was a suite of seismic source

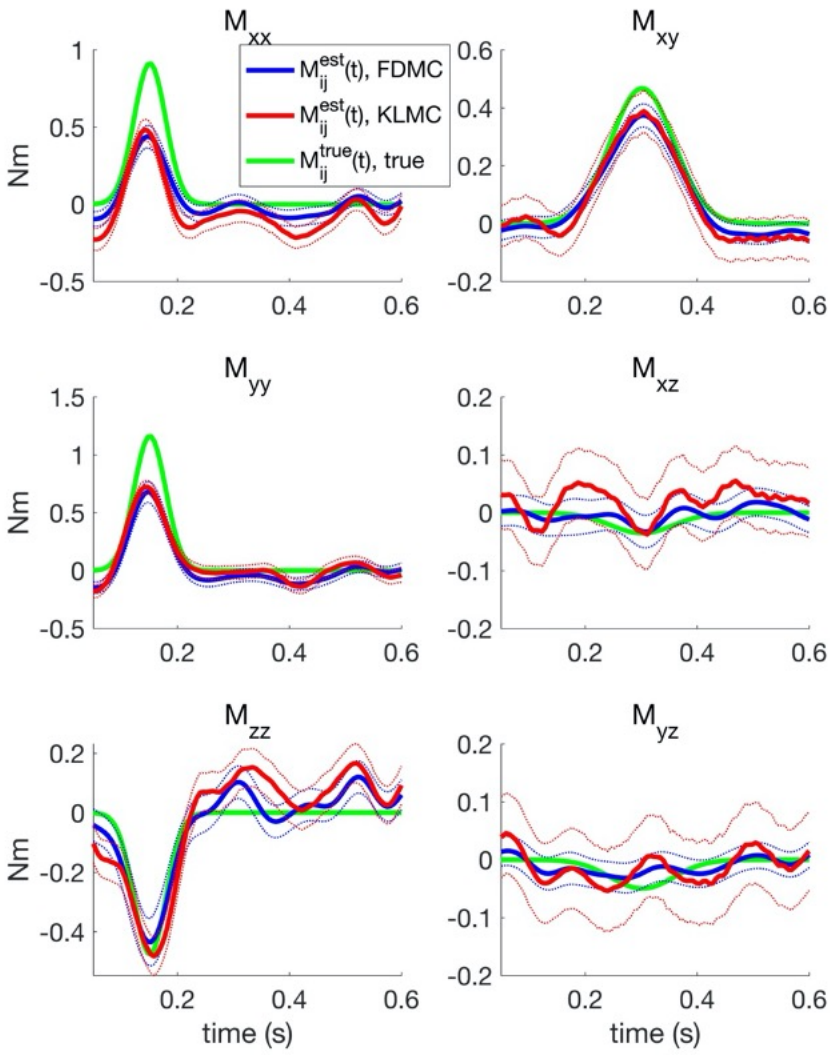

Figure 4-1: Summary of inversion results for both methods. Thick blue shows the mean estimate of the seismic source time functions (STFs) for the FDMC method, with one standard deviation about the mean (thin blue). Thick red shows the mean STFs from the KLMC method, with one standard deviation about the mean (thin red). The green shows the actual STFs used to simulate the data.

parameter estimates in the form of a probability density function. As mentioned previously, we termed this process the Finite Difference Monte Carlo (FDMC) method.

The work described in this chapter has the same goal: propagating model (and data) uncertainty into an estimate of the seismic source parameters. Our method is founded on the idea that the data residual contains the combined effects of model and data uncertainty and therefore we use a stochastic distribution of the residuals to propagate those uncertainties in a pseudo-Monte Carlo scheme. However, the innovation that we present here avoids the expensive finite difference calculations that we used in the FDMC method. Rather, our new method requires only a single estimate of seismic GFs, which we use to build up a suite of surrogate data residuals. Specifically, we invert the data once using a single set of GFs. Then, assuming that the data residual is a stochastic process with zero mean, we construct a stochastic distribution of residuals using the Karhunen-Loève (KL) theorem:

$$\epsilon_k^{\text{stoch}}(\theta, t) = \sum_{f=1}^{F} n_f(\theta)\sqrt{\lambda_{f,k}}\,\mathbf{u}_f(t)$$

where $\epsilon_k^{\text{stoch}}(\theta, t)$ is a suite of random data residuals for channel k with non-random component t and random event θ, $n_i(\theta)$ are uncorrelated random variables with zero mean and unit standard deviation, and $\mathbf{u}_f(t)$ are a set of orthogonal basis functions scaled by $\sqrt{\lambda_{f,k}}$. We construct the basis functions using Fourier series, where $\mathbf{u}_f(t)$ are a pair of sine/cosine functions of frequency f, and $\lambda_{f,k}$ are the (real-valued) amplitude spectral components of the residual's autocorrelation function corresponding to channel k. Then, we invert the data M times, where for each inversion we add a single realization of the stochastic residual distribution to the data. We term this procedure the KLMC method, and show that it produces estimates of the seismic source parameters, along with their uncertainties, that are similar to those estimated using the FDMC method (Figure 4-1). More importantly, the method presented here is computationally several orders of magnitude faster than our previous FDMC method, and requires no a-priori assumptions of model and/or data uncertainty.

The details of this work are, at the time of this writing, undergoing peer review in the journal Geophysical Journal International (Poppeliers and Preston, in review).

5. STRUCTURAL INVERSION

Geophysical structural inversions use either the seismic waveforms themselves or information collected from those waveforms to infer the structure of the Earth. For example, travel times of certain seismic arrivals such as P-waves can be used in a structural inversion to infer the 3-D distribution of seismic P-wave speeds within the Earth. However, each travel time has an associated uncertainty with it. Because of noise from a passing truck, for example, an analyst may be unsure exactly where to pick the P-arrival. These uncertainties in the data will obviously translate into uncertainty in the resultant structure obtained from the inversion. In this section we will attempt to find a more computationally efficient means of estimating the uncertainty in structural inversions using polynomial chaos-based approaches by comparing PCE results to Monte Carlo.

5.1. Data and Model Setup

A subset of the travel time dataset used in Preston et al. (2020) was used for the structural inversion tests in this chapter. For this modeling exercise, we extract absolute P and S travel time data with all sources and receivers within the latitude and longitude bounds from 36.40°N to 36.95°N latitude and from 115.87°W to 116.38°W longitude and to a depth of 15 km from the Preston et al. (2020) model. The model has 2 km horizontal and 1 km vertical node spacing. There are nearly 125,000 absolute travel times, of which about 47,500 are S-picks with the remainder being P arrival times.

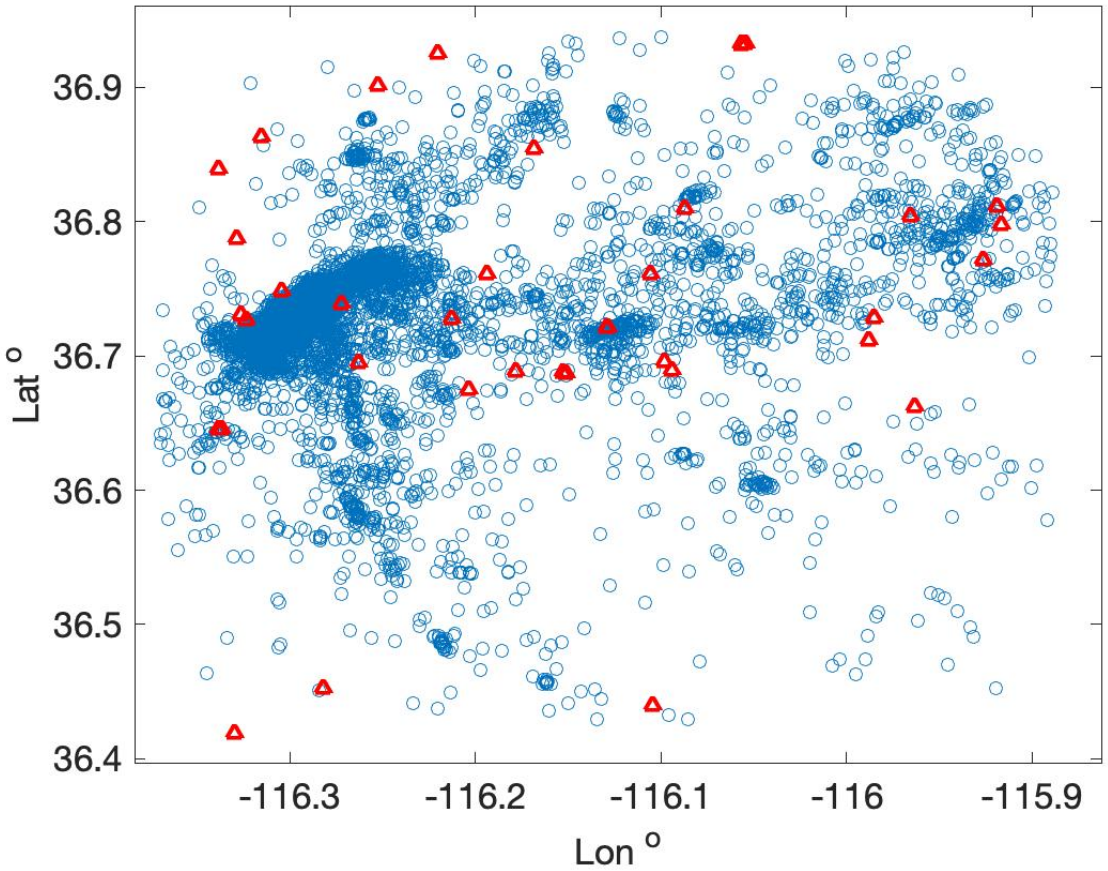

Figure 5-1: Model extent showing seismic stations (red triangles) and starting earthquake source locations (blue circles).

All travel time picks have estimated travel time pick uncertainties associated with them, which were chosen by the analysts when they picked the arrivals. There are a total of 38 stations and over 10,000 sources all of which are earthquakes (Figure 5-1). Earthquake hypocentral parameters are simultaneously inverted for in conjunction with the 3-D Earth structural parameters of P- and S-wave speeds throughout the model domain. In order to ensure reasonable P- and S-wave speeds we impose constraints that are enforced during the inversion process: P-wave speeds must lie between 1 km/s and 7 km/s and the Vp/Vs ratio must lie between 1.6 and 1.9 down to 5 km below sea level and 1.6 and 1.8 below 5 km depth. We use an anisotropic Laplacian operator that allows a different roughness vertically compared to horizontally in order to regularize the solution. The forward problem uses the Vidale-Hole 3-D travel time calculator (Vidale 1990; Hole and Zelt, 1995) to form a linearized sensitivity matrix of travel times with respect to medium parameters, and we use a constrained conjugate gradient least squares algorithm (Preston et al., 2020) to solve the system of equations including regularization. This is a nonlinear inversion problem since the ray paths, and hence sensitivity, depend on the medium parameters we are solving for. Thus, we must re-compute ray paths on each iteration, and we iterate the entire process until changes in the model are small. Note that these images are not meant for interpretation as we did not optimize regularization or other parameters. These models are intended solely to compare and contrast uncertainty quantification methods. See Preston et al. (2020) for further details on the inversion methodology.

5.2. Monte Carlo

The standard with which we will compare PCE-based approaches will be statistics derived from Monte Carlo inversions. For the MC inversions we assumed that the travel time uncertainties were from independent Gaussian distributions with zero mean and standard deviations equal to the estimated pick error. Thus, there are nearly 125,000 independent random variables in this inversion problem. We randomly select error values for each travel time based on these distributions and add these random errors to the observed travel times. We then run the inversion process until convergence. We repeat this procedure 1,000 times in order to obtain estimates of the means and variances of the model parameters. An example of the distribution of medium P-wave speeds for the 1,000 simulations at a point in the model is shown in Figure 5-2, which shows that in this case the distribution of Vp for this well-sampled node in the model has a quasi-Gaussian distribution. The mean and standard deviation maps at a depth of ~300 m below mean sea level derived from the MC results are shown in Figure 5-3ab. For comparison with standard practice, in Figure 5-4 we show the Vp map at the same depth with a standard single inversion of the data using the observed travel time data without any

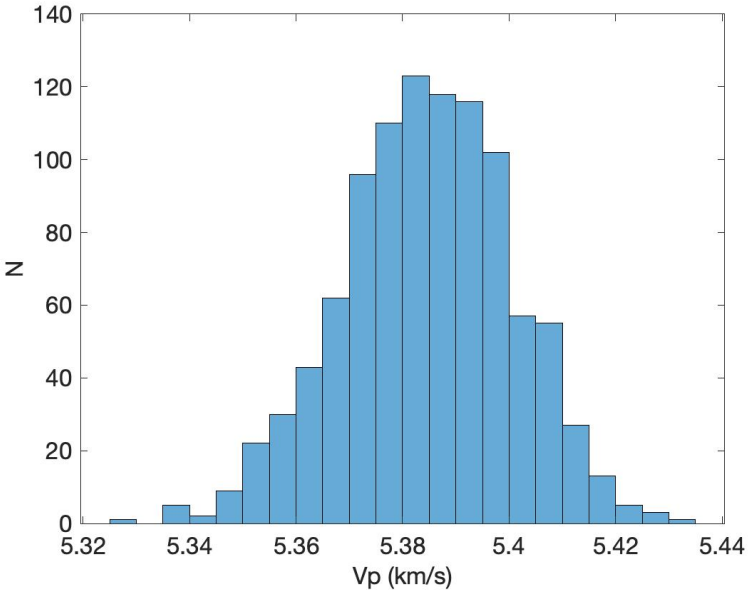

Figure 5-2: Histogram of P-wave speed (Vp) at one node from the 1,000 Monte Carlo inversions.

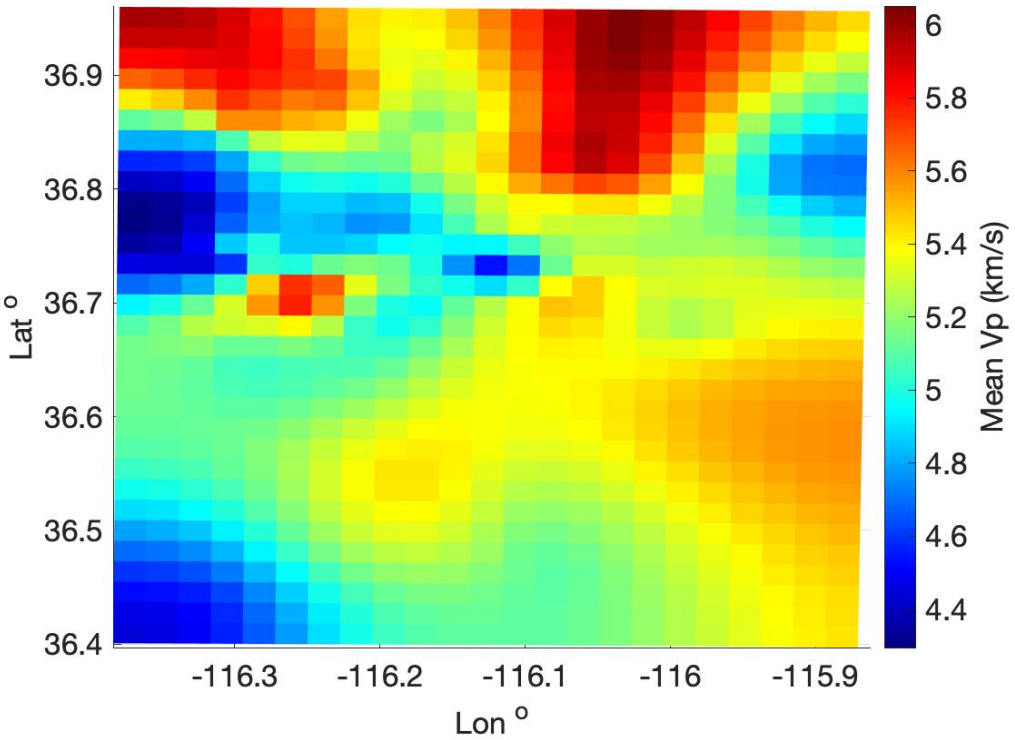

Figure 5-3a: Mean Vp at 300 m depth below sea level from 1,000 Monte Carlo inversions.

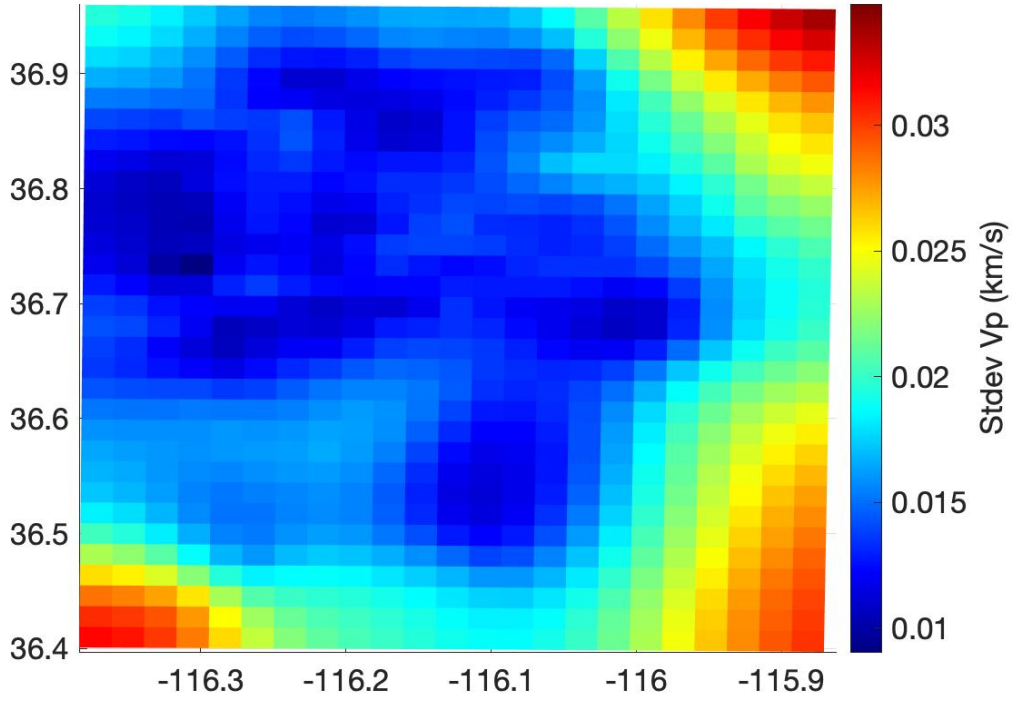

Figure 5-3b: Standard deviation of Vp at 300 m depth below sea level from 1,000 Monte Carlo inversions.

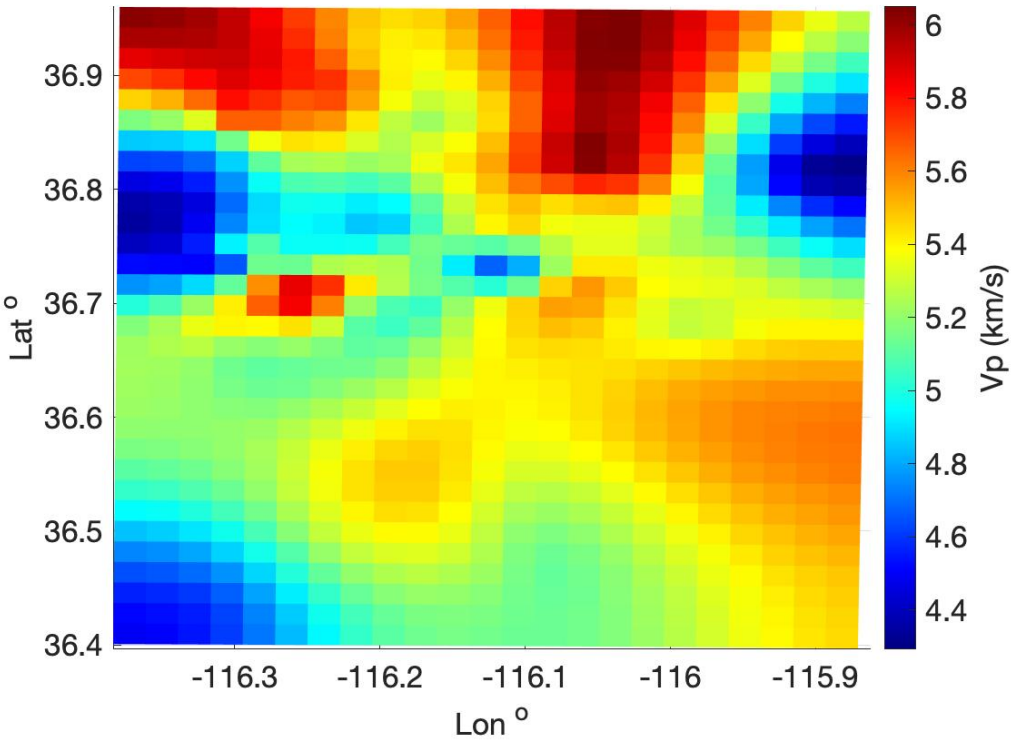

Figure 5-4: Vp at 300 m depth below sea level from single inversion of data without any randomized data error.

randomized data error added. The mean MC Vp in Figure 5-3a is very similar to that of Vp in Figure 5-4. Most differences between the two are in the details, but a notable difference is in the low Vp region in the northeast portion of the models, where the imaged Vp is slower in the single inversion compared to the MC mean.

5.3. Subsample of Monte Carlo

For a computational reference point, we desired to know how well the mean and standard deviations of the full 1,000 runs of MC were matched by just using 20 random models and computing means and standard deviations from those 20 models. Figures 5-5ab shows the results using 20 models. There are only very subtle differences in the mean Vp's between the 1,000 and 20 MC runs, which is not overly surprising. Although the overall pattern of high and low standard deviations are the same between the two, there are more differences in the standard deviation estimate images with more high-wavenumber structure in the standard deviations when using 20 models instead of 1,000. A mean-corrected cross-correlation coefficient (i.e. removing the mean Vp in this depth section based on the 1,000 MC runs from both models before cross-correlating) between the two yields 0.89. None of this is surprising since it is well-known that Monte Carlo estimates converge as $1/\sqrt{N}$, where N is the number of samples. This provides a benchmark that has computational cost of 20 inversion runs so that we can see how well PCE approaches, which we design to have roughly the same cost, perform in relation to both the 1,000 and 20 MC estimates.

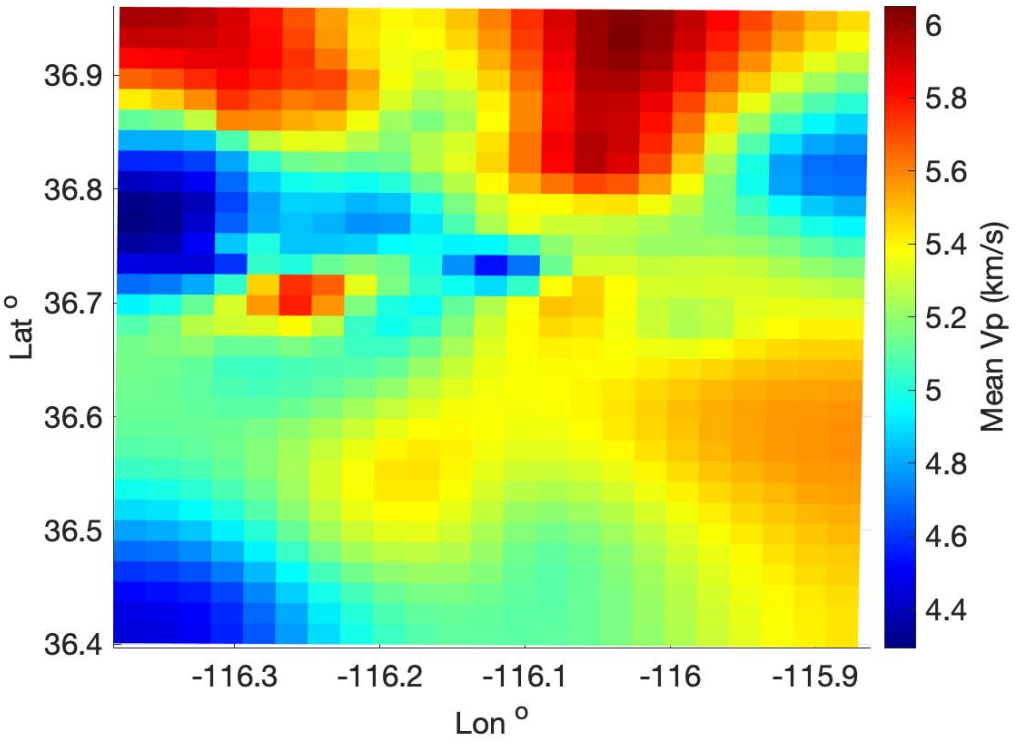

Figure 5-5a: Mean Vp at 300 m depth below sea level from 20 Monte Carlo simulations.

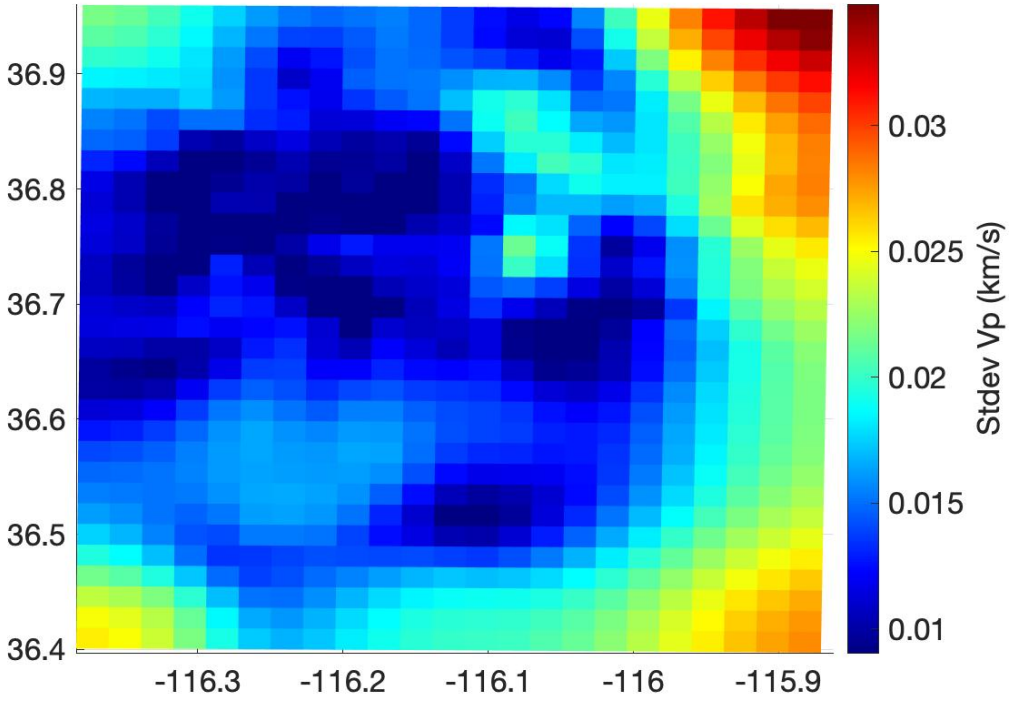

Figure 5-5b: Standard deviation of Vp at 300 m depth below sea level from 20 Monte Carlo inversions.

5.4. Constructing PCE estimates

The first issue is constructing the PCE and then how to use this to estimate statistical quantities such as mean and standard deviation. Instead of using quadrature to estimate the PCE coefficients like is typically done (e.g. in Dakota; Adams et al., 2019, dakota.sandia.gov), we use an L2 norm (Euclidean norm) fit to the values of the random variable at specific points on the standard error axis. For reference, the standard error axis (or value) specifies how many standard deviation units one is away from the mean with direction. So, for example, a point is at -1.5 on this axis if it is 1.5 standard deviations smaller than the mean. Typically, one plots the value on the standard error axis as the x-value and the output value as the y-value. Using the L2 norm allows us to fit any probability distribution with PCE no matter where the samples lie on the standard error axis. Also, even if one uses the quadrature points, since they well-sample the standard error axis, we found that the L2 norm was more stable and consistent with a variety of input and output values. We use Python 3.8 with numpy's implementation of an L2 norm fit to probabilist Hermite polynomials (https://numpy.org/doc/stable/reference/generated/numpy.polynomial.hermite_e.hermefit.html). When the input standard error axis values are known, then one can use this numpy function directly. However, when the input standard error axis values are unknown, we must find an estimate of their values before we can use this function. If it is known or can be reasonably assumed that the input standard error points are from a quasi-Gaussian distribution, we find that sorting the output values and using scipy's normal distribution percent point function (https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html) works well. This function returns the value on the standard error axis that corresponds to a given cumulative distribution probability value. We assign the smallest output value as having a cumulative probability of 1/(N+1) and the largest output value as having a probability of 1-1/(N+1) and all ordered values being equally spaced in cumulative probability space between the two with spacing $\frac{1-\frac{2}{N+1}}{N-1}$.

As an example, in Figure 5-6a we show a Monte Carlo-derived non-Gaussian output distribution with 1,000 samples. In Figure 5-6b we fit a 10th order Hermite polynomial to this random input using the above method. We see a very good fit to the MC samples in this space. The PCE

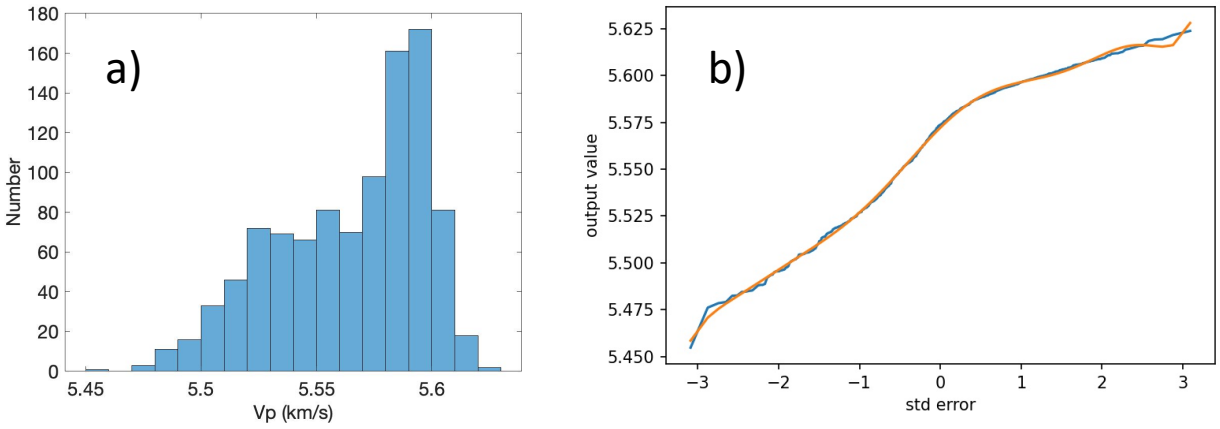

Figure 5-6: a) Non-Gaussian distribution histogram from 1,000 MC simulations. b) Blue line is the data in a) plotted in standard error units; orange line is L2 10th order PCE fit.

coefficients can then be used to rapidly simulate the expensive MC-derived distribution as long as the PCE sufficiently captures the distribution properties.

Another use of PCE coefficients is that they can provide an estimate of the mean and standard deviation and higher moments of the distribution. The mean of the distribution is simply the 0th order coefficient. The standard deviation is:

$$\sigma_{est} = \sqrt{\sum_{i=1}^{P} p_i^2} \tag{5-1}$$

where p_i is the ith order polynomial coefficient and P is the maximum order of the PCE. Note that this sum does not include the 0th order term. See Preston (2017) for a derivation.

5.5. PCE Minimized Number of Random Variables

For any of the PCE approaches to have an opportunity to be faster than MC, the roughly 125,000 random variables must somehow be reduced to a small number of random variables. This is due to the fact that each random variable is expanded via Equation 1-2, so a 4th order PCE expansion (5 terms) with 10 random variables would require 5^{10} terms. There are methods to reduce this, such as sparse PCE expansion methods like Smolyak sparse grids (Smolyak, 1963; see also the Dakota theory manual for further possibilities: Adams et al, 2019). However, PCE, even with these approaches, the "curse of dimensionality" would be problematic with 125,000 random variables.

5.5.1 Large Number Approximation

We attempted multiple approaches, only some of which will be outlined here. The first approach attempts to reduce this dimensionality to the extreme by conglomerating all random variables into one single representative random variable. We do this by using the fact that for a vary large number N of samples, the sum of the log of the probabilities, p_i, of each random sample or:

$$\sum_{i=1}^{N} \ln p_i \tag{5-2}$$

is a quasi-Gaussian distribution (it is not a true Gaussian because the values can only be positive) with a peak at $N \ln\left(p_G(1)\right)$ where $p_G(1)$ is the probability of a standard Gaussian at one standard deviation, and a standard deviation of $\sqrt{N/2}$. What this means is that for a large number of samplings, a particular random sample will very likely be near one standard deviation away from its mean. We then sample this distribution at the roots of a 20th degree chaos probabilist Hermite polynomial and run tomographic inversions at each of these 20 quadrature points. Since there is only one random variable, all data points will be using the same value on the standard error axis for a given inversion (e.g. all will be sampling at -2.5 standard deviations, $\sqrt{N/2}$, from the mean $N \ln\left(p_G(1)\right)$). Since the large number approximation assumption would only produce positive perturbations, the only variability is that half the data points are randomly selected to have negative perturbations, so that the overall mean perturbation is zero. Given this very restrictive sampling and variability, it is not surprising that the estimated standard deviations from this method are roughly 3-10 times too small compared to full MC and has relatively poor correlation with the MC results (mean corrected cross-correlation coefficient of 0.12; Figure 5-7a).

5.5.2 Using Full Standard Error per Observation

In order to increase the variability, the next step is to assume that each observation has its own Gaussian probability distribution with a standard deviation equal to its individual estimated standard

pick error. Again, since there is only one random variable here, all observations are sampling at the same point on their standard error axes for a given inversion (e.g. all are sampling at -2.5 standard deviations). As before, in order to keep the mean perturbation zero regardless of the current value of the standard error axis being sampled, half of the samples are randomly selected to sample from the opposite side of the Gaussian distribution. For example, if the current inversion is sampling at -2.5 standard deviations, then half of the data would sample at -2.5 and the other half at +2.5 standard deviations (the latter set we refer to as the "negated" set). We performed 10 inversions using the 10-point Gauss-Hermite quadrature abscissa values for the standard axis evaluation points. These results were much better in terms of bias in the standard deviation estimates and mean corrected cross-correlation coefficient with a bias of about ~50-75% of the full MC results and cross-correlation coefficient of 0.31 (Figure 5-7b).

5.5.3 Fixed Sample for Negated Set

One final test using one random variable for PCE evaluation is identical to the previous case except in how the observations are selected that are in the negated set. Unlike the previous case, which makes the random selection independently for each standard error value, in this case once the set of

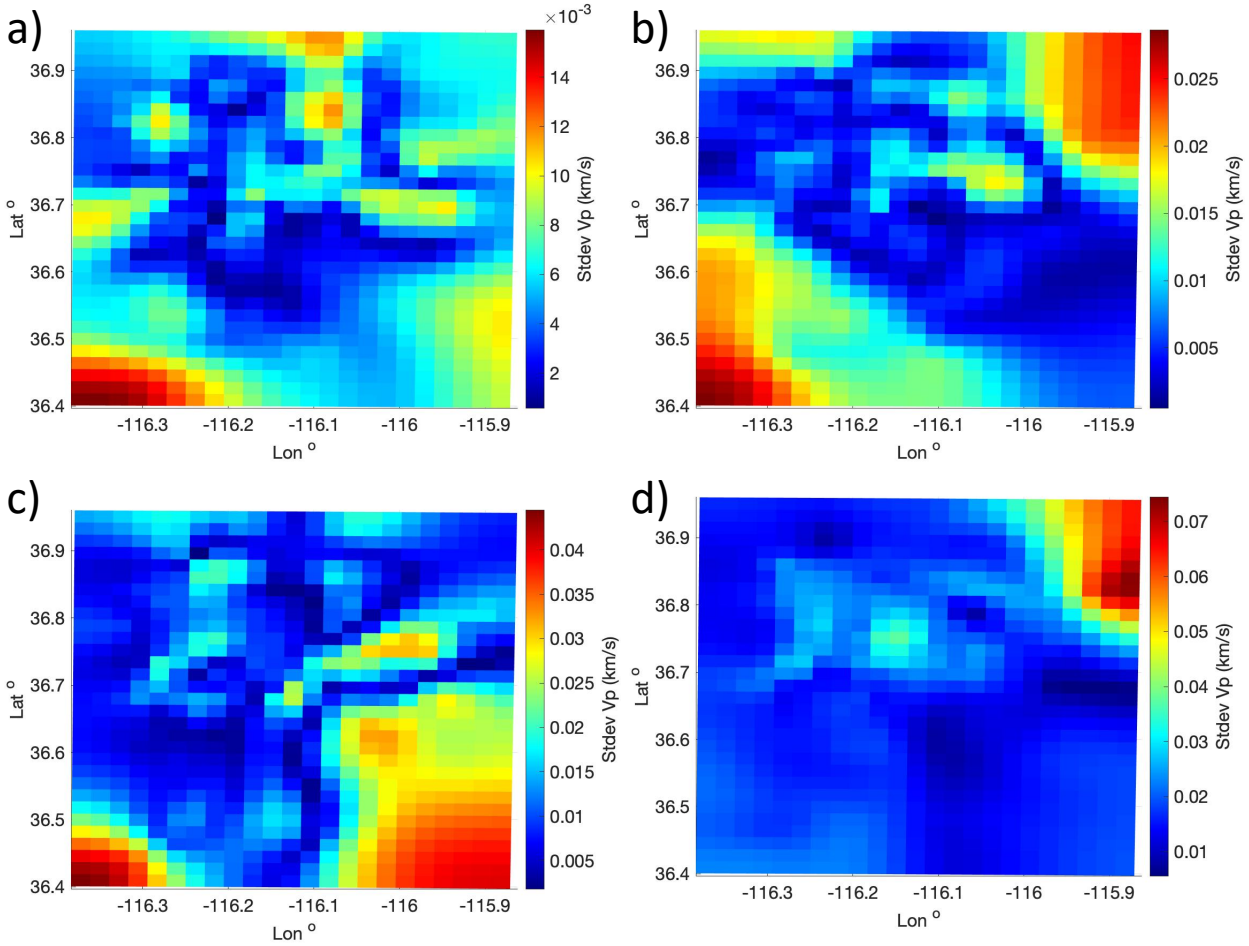

Figure 5-7: Standard deviation estimates for a) 5.5.1 Large number approximation, b) 5.5.2 Using full standard error per observation, c) 5.5.3 Fixed sample for negated set, d) 5.6.2 Using full standard error per observation with multiple groups. Note that each panel has an independent colorbar.

observations are selected as being in the negated set, they remain so for all standard error values. This further improves the approximation to the full MC standard deviation estimate. Although on average there is virtually no bias in the results, it has both over- and underestimates of the standard deviation with a mean corrected cross-correlation coefficient of 0.49 (Figure 5-7c).

5.6. PCE Two-Tier Randomization

In an attempt to improve over the results in the previous section, we desired to explore more combinations of which observation are in the negated set. To do this we run multiple groups with each group being evaluated at the same standard error values but with different observations being in the negated set in each group. Like in the last test, once we randomly choose which observations will be in the negated set, those observation will always use the negative standard error value for all evaluated standard error values in that particular group. Unlike the last test, however, we redo the full suite of standard error values multiple times, with, at the beginning of each group, a new randomly selected set of observations using the negative standard error value (i.e. negated set). For example, if we run 5 groups using 5-point Gauss-Hermite quadrature points for the standard error axis values, the initial group will first select which observations will be in that group's negated set, and then it runs through all five standard error values for 5-point quadrature. Next, the second group begins by independently and randomly selecting different observations to be in its negated set, and then it proceeds through all five standard error values for 5-point quadrature. This is repeated for the third, fourth, and fifth groups: a total of 25 inversions. To process these results into overall standard deviations, the standard deviation for each group is independently computed using Equation 5-1 and then the standard deviations from all five groups are averaged.

5.6.1 Large Number Approximation

We show two test cases using this method. First, we try the large number approximation approach as presented in Section 5.5.1, where the mean standard deviation of a large number of random samplings is equal to unit standard error for that observation. We compute 10 groups using 3-point Gauss-Hermite quadrature for a total of 30 inversions. The results are in line with those of Section 5.5.1 with a mean corrected cross-correlation coefficient of 0.12, once again demonstrating that there is too little variability in the large number approximation approach to adequately represent the full MC standard deviation estimate.

5.6.2 Using Full Standard Error per Observation

In this case we use quadrature values for 5-point Gauss-Hermite quadrature in five groups, so 25 total inversion runs. Like in Sections 5.5.2 and 5.5.3, we use the estimated pick error for each observation as being one standard deviation from the picked observed time. Using this approach greatly improved the results relative to those of Section 5.6.1 and is on par, but slightly better than, the results of Section 5.5.3. We obtain a mean corrected cross-correlation coefficient in this case of 0.51, but with generally overestimated standard deviations by up to a factor of 2 (Figure 5-7d).

5.7. PCE Randomly Mixed Standard Error Values

The final case evaluated in this report further builds on what we discovered in the previous two sections. We extend the successes we found in Section 5.6.2 by randomizing the order of the fixed set of standard error values on a per observation basis. For example, for 5-point Gauss-Hermite quadrature, there will be 5 standard error values that will be sampled for each group. For 5 point quadrature for Hermite chaos polynomials, the quadrature points are at standard error values of approximately -2.86, -1.36, 0.00, +1.36, +2.86. Since there are five values, there will be five inversions in a particular group. On the first inversion in a group, each observation is randomly

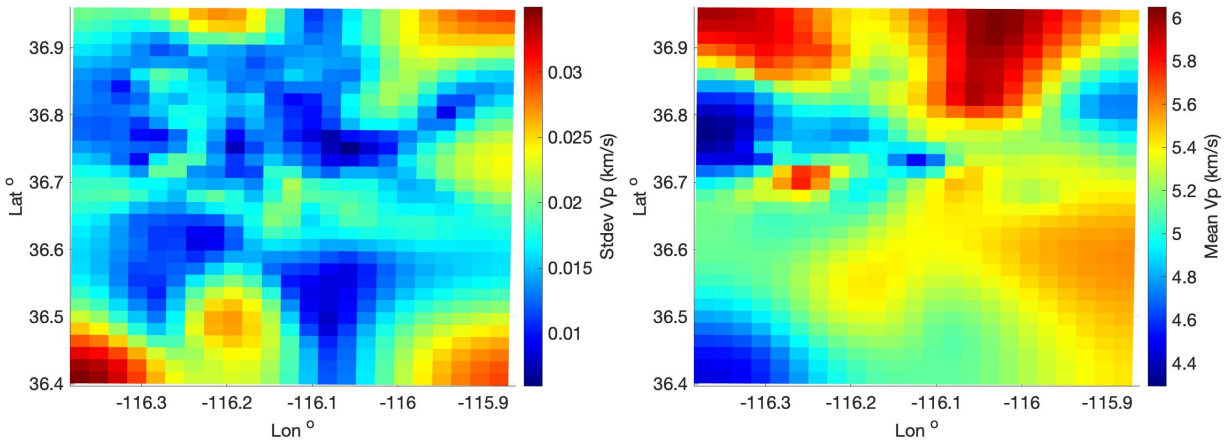

Figure 5-8: a) Standard deviation estimate from test in Section 5.7, b) Mean estimate from Section 5.7.

evaluated at one of the five possible standard error values. On the second inversion in the group, each observation is evaluated at a different randomly selected value of the four standard error values that remain for that observation. This continues on the third through fifth inversions in that group, with the fifth inversion just using whatever of the standard error values each observation hasn't evaluated yet. This way each observation will be evaluated at each of the five possible standard error values but the ordering of each is random. Also, as in Section 5.6.2, we redo a full group multiple times independently, so once the first group of five inversions are completed, the process begins again with a new group of five inversions starting afresh with all five standard error values being evaluated in random orders per observation.

For this test, we use the 5-point quadrature values for Hermite chaos polynomials in five groups for a total of 25 inversions. Just as in Section 5.6, we compute the standard deviation for each group independently and then average the results over the five groups to obtain an overall estimate of the standard deviation of the Vp map. Roughly, the broad-scale pattern and magnitude of the standard deviations correspond to that of the full MC results with a mean corrected cross-correlation coefficient of 0.69 (Figure 5-8a). The mean model output from this test is visually extremely close to that of the full MC mean model with only subtle differences between them; indeed, the mean corrected cross-correlation coefficient of the mean models is > 0.99 (Figure 5-8b).

5.8. Discussion

We showed several different attempts at estimating standard deviations using PCE approaches. Several others were also attempted as variations on these basic schemes, but the ones in this report demonstrate the basic outcomes and results. The large number approximation methods discussed in Sections 5.5.1 and 5.6.1 are clearly inferior to other results. These approaches simply do not allow enough outliers to properly bracket the true variations seen in a large dataset like this. Increasing the variability and sampling to use the full standard errors instead of large number approximations greatly improved results. Allowing each observation to have different standard error values while sampling its full spectrum of standard error values clearly produced the best results.

Although the final test evaluated in Section 5.7 is by far the best model we found using a PCE-based approach, it still was poorer in how well it matched the full MC standard deviation results for Vp compared to just using 20 MC runs. In addition this final test required 25 inversions to obtain a poorer result than the 20 MC inversions, so it is also less computationally efficient.

Overall, the PCE approach failed because of the sheer number of random variables. We were unable to find a means of reducing this random dimensionality to a computationally efficient level while still maintaining sufficient accuracy to the primary statistics we were interested in. In order to more efficiently estimate uncertainties in inversions with high random dimensionality other methods may offer better accuracy and efficiency such as multi-level or multi-fidelity uncertainty quantification models (e.g. Adams et al, 2019). These approaches reduce the computational cost of performing UQ by combining information from a very limited number of high-accuracy, costly simulations with many runs of cheaper, lower fidelity simulations, such as using reduced order model approximations (e.g. simpler physics or making 1-D assumptions) or using lower resolution grids (e.g. Giles, 2013, 2015; Ng, 2013; Peherstorfer et al. 2016; Khan and Elsheikh, 2019; Jakeman et al., 2020).

This page left blank

6. SUMMARY

We discussed two primary topics within this report: full waveform seismic moment tensor inversion and travel time structural inversion of the earth's material properties. Full waveform seismic moment tensor inversion is a simpler inverse problem, being linear in the frequency domain and very fast to compute. However, evaluating the uncertainty in moment tensor solutions due to uncertainty in the 3-D earth model is computationally expensive. We discussed several methods that we attempted that progressively improved on each other and eventually formulated the KLMC method that provides excellent approximations to the mean and variance of moment tensor solutions estimated from full MC simulations at 0.1% of the cost. This enormous cost reduction is made possible because only one expensive 3-D simulation is required in the method, with the remaining computations only requiring computationally inexpensive convolutions. This success means that we can now very efficiently estimate the effects of earth model uncertainty on our source characterization solutions. Consequently, decision makers can have access to high-quality source characterization results with uncertainties attached that account for earth model error that is currently rarely available, grossly approximated, or computationally expensive to compute.

Estimating the uncertainty in earth structure inversions in a computationally efficient manner was a much larger beast. We attempted multiple PCE-based approaches, trying various means to reduce the dimensionality in the random domain. However, even in our test with the best results compared to full MC, the results were subpar and required too much computational effort even to obtain the results we did. Using a relatively small number of MC simulations and estimating the uncertainty from those results were the best and most computationally efficient method that we tested. Other promising methods such as multi-level or multi-fidelity uncertainty estimation techniques may offer a path forward for inverse problems with very high dimensional random spaces like in earth structural inversion problems.

This page left blank

REFERENCES

1. Adams, B.M., M.S. Eldred, G. Geraci, R.W. Hooper, J.D. Jakeman, K.A. Maupin, J.A. Monschke, A.A. Rushdi, J.A. Stephens, L.P. Swiler, T.M. Wildey, W.J. Bohnhoff, K.R. Dalbey, M.S. Ebeida, J.P. Eddy, P.D. Hough, M. Khalil, K.T. Hu, E.M. Ridgway, D.M. Vigil, and J.G. Winokur (Updated 2019). *Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: version 6.11 Theory Manual*, SAND2014-4253, Sandia National Laboratories, Albuquerque, NM.
2. Aki, K., and P.G. Richards (2002). *Quantitative Seismology*, Second Edition, University Science Books, Sausalito, CA.
3. Giles, M.B. (2013). Multilevel Monte Carlo Methods, in *Monte Carlo and Quasi-Monte Carlo Methods 2012*, eds. J.Dick, F. Kuo, G. Peters, and I. Sloan, Springer, Berlin, 83-103.
4. Giles, M.B. (2015). Multilevel Monte Carlo Methods, *Acta Numer.* **24**, 259-328, doi: 10.1017/S096249291500001X.
5. Hallo, M., and F. Gallovic, (2016). Fast and Cheap Approximation of Green Function Uncertainty for Waveform-Based Earthquake Source Inversion, *Geophys. J. Int.*, **207**, 1012-1029.
6. Hole, J., and B. Zelt (1995). 3-D Finite-Difference Reflection Traveltimes, *Geophys. J. Int.* **121**(2), 427-434.
7. Jakeman, J.D., M.S. Eldred, G. Geraci, T.M. Smith, and A.A. Gorodetsky (2020). *LDRD #218317: Learning Hidden Structure in Multi-Fidelity Information Sources for Efficient Uncertainty Quantification*, SAND2020-9657, Sandia National Laboratories, Albuquerque, NM.
8. Khan, N.K.J., and A.H. Elsheikh (2019). A Machine Learning Based Hybrid Multi-Fidelity Multi-Level Monte Carlo method for Uncertainty Quantification, *Front. Environ. Sci.*, doi: 10.3389/fenvs.2019.00105.
9. Le Maître, O.P. and O.M. Knio (2010). *Spectral Methods for Uncertainty Quantification*, Springer, New York, doi: 10.1007/978-90-481-3520-2.
10. Mallat, S. (2009). *A Wavelet Tour of Signal Processing The Sparse Way*. 3rd Ed., Academic Press, Burlington, MA.
11. Ng, L.W.-T. (2013). *Multifidelity Approaches for Design Under Uncertainty*, PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.
12. Peherstorfer, B., K. Wilcox, and M. Gunzburger (2016). Optimal Model Management for Multifidelity Monte Carlo Estimation, *SIAM J. Sci. Comput.* **38**, A3163-A3194, doi: 10.1137/15M1046472.
13. Poppeliers, C., and L. Preston, (2020). The effects of Earth model uncertainty on the inversion of seismic data for seismic source functions, *Geophys. J. Int.*, **224**(1), 100-120.
14. Poppeliers, C., and L. Preston (2021a). *An efficient method to estimate the probability density of seismic Green's functions*. SAND2021-9865, Sandia National Laboratories, Albuquerque, NM.
15. Poppeliers, C., and L. Preston, (2021b). *Approximating and incorporating model uncertainty in an inversion for seismic source functions: Preliminary results*. SAND2021-9866, Sandia National Laboratories, Albuquerque, NM.
16. Poppeliers, C., and L. Preston (in review). An efficient method to propagate model uncertainty when inverting seismic data for time domain seismic moment tensors. Submitted to *Geophys. J. Int.*

17. Preston, L.A. (2017). *Paracousti-UQ: A Stochastic 3-D Acoustic Wave Propagation Algorithm*, SAND2017-10143, Sandia National Laboratories, Albuquerque, NM.
18. Preston, L.A. (2018). *Incorporation of Spatial Stochastic Variability into Paracousti-UQ*. SAND2018-9592, Sandia National Laboratories, Albuquerque, NM.
19. Preston, L., C. Poppeliers, and D.J. Schodt (2019). Seismic Characterization of the Nevada National Security Site Using Joint Body Wave, Surface Wave, and Gravity Inversion, *Bull. Seis. Soc. Am.* **110**, 110-126, doi: 10.1785/0120190151.
20. Smolyak, S.A. (1963). Quadrature and Interpolation Formulas for Tensor Products of Certain Classes of Functions, *Dokl. Akad. Nauk SSSR*, **4**, 240-243.
21. Vidale, J. (1990). Finite-Difference Calculation of Traveltimes in Three Dimensions, *Geophysics* **55**(5), 521-526.
22. Xiu, D. And G.E. Karniadakis (2003). Modeling Uncertainty in Flow Simulations Via Generalized Polynomial Chaos, *J. Comp. Phys.*, **187**, 137-167, doi:10.1016/ S0021-9991(03)00092-5.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Ladonna Martin	1910	ldrd@sandia.gov
Technical Library	01977	sanddocs@sandia.gov

This page left blank

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.