PROBABILISTIC SITE CHARACTERIZATION FROM INTEGRATED GEOLOGICAL AND GEOPHYSICAL DATA FOR GEOTECHNICAL ANALYSIS

A Dissertation

by

JUNGRAK SON

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Zenon Medina-Cetina
Committee Members,	Charles Aubeny
	Arash Noshadravan
	Niall Slowey
Head of Department,	Robin Autenrieth

December 2019

Major Subject: Civil Engineering

ABSTRACT

Offshore site characterization for geotechnical construction projects relies on geological and geophysical survey data because geotechnical sampling techniques have limitations for offshore environments. However, the seismic images, the observed data from the most popular geophysical surveys for offshore exploration, are only reflected signals in time-domain and still difficult to apply for the geohazards evaluation in space-domain. Thus, offshore geo-site characterization requires the seismic inversion methods, which can convert from observed wave signals into the soil and rock properties along with the depth. However, this spatial inverse problem contains the mathematical non-uniqueness problem, so we need a probabilistic approach to quantify the uncertainty.

The probabilistic site characterization in this study is based on the stochastic process in the Bayesian framework. The spatial Gaussian process and reversible jump Markov chain Monte Carlo (rj-MCMC) methods are developed to utilize the offshore geological drilled borehole data and geophysical seismic survey data. Bayesian inference, which integrates the observed data, model predictions, and expert's beliefs, supports the geospatial analysis to provide probabilistic descriptions of the model parameters. It is required for the inverse problem to understand the quantified uncertainty from all model parameters and widely used for the nonlinear geophysical seismic inverse problems.

This dissertation focuses on the probabilistic inversions with geophysical seismic data and starts from a synthetic offshore shallow case study to integrate geological and

geophysical surveys for better ground model estimation. Geological borehole data provides the initial condition of the stratigraphy information, and geophysical seismic inversion finds out the model parameters, bulk densities, acoustic P-wave velocities, and the depth of layer interfaces. The spatial random field from Gaussian process supports the probabilistic seismic inversion, and this study introduces a new approach to reconstruct the subsurface ground model information in high-resolution.

Practical applications are developed from this case study by applying the rj-MCMC method, which recently showed a great advantage in geophysical studies. The main idea of the rj-MCMC method is to define the unknown subsurface geologic layers as another random variable, and design hierarchical Bayesian priors to support the convergence of the dimensions during the stochastic process. Three case studies in varying dimensions are discussed for the geological offshore stratigraphic modeling, geophysical stochastic post-stack seismic inversion, and geomechanical Bayesian full-waveform inversion (FWI). The rj-MCMC methods in those case studies lead the probabilistic site characterization to find the correct target modeling parameters, which is challenging for the unknown subsurface estimation. One of those studies is also applied to field data near Sigsbee Escarpment, a steep marine slope in the Gulf of Mexico. The results show the two-dimensional subsurface image of the ground model under the seafloor, which can help us to avoid geohazard prone areas.

DEDICATION

To my beloved wife and son.

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Zenon Medina-Cetina, and my committee members, Dr. Auneby, Dr. Noshadravan, and Dr. Slowey, for their guidance and support throughout the course of this research. My background knowledge of Bayesian method, numerical modeling, and seismic data processing and interpretation comes from them, and my advisor has guided me in building up the scientific strategy, the most challenging assignment for my Ph.D.

I sincerely appreciate my colleagues in Stochastic Geomechanics Laboratory (SGL). Especially, my friend, Yichaun Zhu, helped me to understand the basics of the Monte Carlo method and spent a lot of time discussing my research with me. Tam Dung and Patricia Varela, who were the members of Society of Underwater Technology (SUT) TAMU student chapter, introduced me to SGL in civil engineering when I was in the geophysics department for my M.S. Moslem Moradi also gave me lots of support when I needed a colleague to discuss geophysics and geostatistics. Thanks also go to all other SGL members and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my parents for their patience, and to my wife for her encouragement and love, which helped me not to give up on this long journey of my Ph.D. degree.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee chair Dr. Medina-Cetina, a chair of Offshore Site Investigation and Geotechnics (OSIG) of Society of Underwater Technology (SUT) Houston branch. He supported me to attend SUT workshops and Ocean Technology Conference (OTC) in Houston. Those experiences helped me to review the research papers in my field, recognize the problems in the offshore industry, and clarify the objectives of my Ph.D. in Chapter 1.

Special thanks to the International Ocean Drilling Program (IODP) at TAMU campus for their offshore field survey data, which were used for Chapter 2 and Chapter 3. I learned the offshore seismic data interpretation in Chapter 4 from Dr. Slowey, and the discussion with Dr. Noshadravan helped me to understand the Finite-Element Method (FEM) in Chapter 5. Dr. Saba Esmailzaedeh, a former student of Dr. Medina-Cetina's research group, shared a part of MATLAB code of reversible jump Markov chain Monte Carlo method, and it became the foundation to develop my new scientific contribution. All other work conducted for the dissertation was completed by myself independently.

Funding Sources

My graduate study was supported by a research assistantship with Dr. Medina-Cetina from 2016 to 2018, and departmental fellowships in 2015, 2016 & 2019.

NOMENCLATURE

MCMC	Markov chain Monte Carlo method
rj-MCMC	reversible jump Markov chain Monte Carlo method
$\boldsymbol{\theta}$	Random variable as a vector
$\boldsymbol{\theta}^*$	Proposed modeling parameter vector in the MCMC method
k	Dimension parameter of rj-MCMC method
$\boldsymbol{z}$	Depth parameter of layer interfaces as a vector
$\boldsymbol{\beta}$	Material property parameter as a vector
$\boldsymbol{d_{obs}}$	Observed data
$\boldsymbol{d_{pred}}$	Predicted data
ρ	Soil density parameter
ρ_w	Water density parameter
V_P	Acoustic P-wave velocity parameter
V_S	Elastic S-wave velocity parameter
f	Frequency of seismic signals
λ (Chapt.2, 3, 4)	Wavelength of seismic signals
RC	Reflection Coefficient
CMP	Common-Mid-Point
μ (Chapt.2, 3, 4)	Mean of Gaussian distribution
σ	Standard deviation of Gaussian distribution
C_d	Covariance matrix

CV	coefficient of variation
s	Dynamic standard deviation for the likelihood function
v (Chapt.4)	Precise parameter for dynamic proposal distribution
SASW	Spectral Analysis of Surface Waves
MASW	Multichannel Analysis of Surface Waves
FWI	Full Waveform Inversion
SEM	Spectral-Finite Element Method
σ & ϵ (Chapt.5)	Stress & Strain
λ (Chapt.5)	Lame's 1st parameter
μ (Chapt.5)	Lame's 2nd parameter (shear modulus)
E	Young's modulus
K	Bulk modulus
v (Chapt.5)	Poisson's ratio
CFL	Courant-Friedrichs-Lewy

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1. Problem Statement

Geohazards can be defined as a geological state in a large area or local soil conditions having a potential of leading geotechnical failure events to cause loss of life or damage environments and infrastructures. Numerous offshore projects have a large practice to develop deepwater oil fields or install wind turbines for renewable energy, but the marine continental shelves are geohazard prone area with hazardous ground conditions (Lacasse et al., 2013). Abrupt human activities for the offshore projects can be a triggering mechanism for the failure of unstable soil mass on the continental shelve, and this infrequent event becomes a direct threat to subsea infrastructures. A retrogressive slope failure causes the debris flows and turbidity current, which damages the installed offshore foundations of floating platforms for keeping the fixed location. Large scale submarine slides are well known to initiate tsunamis with devastating effects on adjacent coastal areas.

Offshore geohazard studies are important as an essential process to design the geotechnical foundations, but it is particularly challenging because of the possible large uncertainty in soil properties (Jeanjean, Liedtke, Clukey, Hampson, & Evans, 2005). One of the offshore projects in the Gulf of Mexico to develop the Mad Dog oil field in 2005 is a good example to show why we should focus on geohazard studies. The Mad Dog oil field is a deep-water oil field located along the Sigsbee Escarpment in the southeastern part of the Green Canyon. Eleven suction piles arranged in three clusters

were designed for the foundation and mooring system (Berger, Lanier, & Jeanjean, 2006). Two clusters are installed along the lower continental slope, and one cluster is situated along the escarpment. Because of these clusters in geohazard prone area, many slope stability analyses have been studied with both deterministic and probabilistic approaches. Nowacki et al. (2003) explained the detail of deterministic slope stability analyses in this area with the finite-element method. The ground model from the site investigation was used as initial and boundary conditions of the numerical modeling, and this study showed how the vertical and horizontal faults with week layers affect the failure mechanisms based on the quantified factor of safety. Another probabilistic approach is applied for the slope stability analysis in the same area (Nadim, Krunic, & Philippe, 2003). Based on the theory of triggering mechanisms and historical submarine slides data, they used the Bayesian framework to integrate the data and establish the annual probability of slope instability. Even though the triggering mechanisms such as fluid losses in shallow section during drilling, overweight equipment on the seafloor, pressure change from lack of well control or gas hydrate melting, horizontal anchor forces during storm loading (Locat & Lee, 2002) are important for submarine landslide analysis, the information of subsurface ground model is vital for both approaches.

A ground model, an integrated database of subsurface, requires proper site investigations. Offshore geotechnical site investigations include drilling boreholes, core sampling, and in situ testing, and the soil data from these investigations can help us to explain the ground instabilities and heterogeneity under the seafloor. Liedtke, Jeanjean, and Humphrey (2006) presented the salient aspects of the offshore geotechnical

fieldwork performed to characterize the top 150 feet of sediments for the design of the Mad Dog suction anchors. Kvalstad (2007) explained various engineering approaches; the cone penetration tests (CPTs), wireline logging, pore pressure measurements, temperature distribution, and pressure coring for the offshore geohazard investigations. Aubeny et al. (2013) also described various offshore in-situ testing methods, such as the vane shear test, piezo-cone penetrometer test, long core sampling, and continuous PCPT testing. The practical approaches for current industry strongly rely on these geotechnical site investigation based on the Bureau of Safety and Environmental Enforcement (BSEE) regulations and the American Petroleum Institute (API) standards. However, the matter of scale, complexity, and high cost of offshore geotechnical site investigation are still difficult problems, so most of the offshore project engineers suffer from a lack of field data.

1.2. Geophysical Site Investigation

To overcome the lack of site investigation data problem, the geophysical investigations are getting increased for offshore projects. Seismic reflection systems, the most commonly used geophysical techniques, use sound propagation energy generated by a device towed behind a ship (Briaud, 2013), so this survey can explore the wide-area with a relatively lower cost than geotechnical site investigations. Generated acoustic waves travel down to the seafloor or the interfaces between subsurface geological strata and are reflected up to a receiving array. The amount of reflection is affected by acoustic impedance, the product of the soil density (ρ) and the acoustic P-wave velocity (v_p).

The higher contrast of acoustic impedance shows the better chance that the subsurface layers will be detected by seismic reflection. The impedance contrast at the geological strata interface is quantified by the acoustic impedance ratio, defined as the ratio of the acoustic impedances in the lower layer and upper layer.

Empirical equations have been developed to find the two important soil properties, geophysical acoustic impedance and geotechnical shear strength (Brand, Lanier, Berger III, Kasch, & Young, 2003; Hamilton, 1971, 1979; Hamilton & Bachman, 1982). The laboratory test with Multi-Sensor Core Logger (MSCL) measures the geophysical properties of offshore soil samples along with the depth, and the results can be compared with results from the geotechnical measurement. Regression equations based on these measured soil data were suggested (Hamilton & Bachman, 1982) to show the correlation between the geophysical acoustic wave velocity and geotechnical porosity, density, or clay size. The introduced empirical equations were based on three different general marine environments as follows: continental terrace (shelf and slope), abyssal hill (pelagic), and abyssal plain (turbidite). These empirical equations enhanced the advantage of integrated geophysical and geotechnical data. However, general three categorizations of soils are not specific enough because of spatial variation, and geophysical surveys have developed to obtain more high-resolution data.

Autonomous Underwater Vehicle (AUV) survey is one of the seismic reflection systems which collect high-resolution data. This survey was applied to the eastern Sigsbee Escarpment investigation (George, Gee, Hill, Thomson, & Jeanjean, 2002) and supported the engineering problems such as the installation of export pipelines, infield

flowlines and production facilities. Jeanjean, Berger III, Liedtke, and Lanier (2006) described how the results of the geotechnical campaign were integrated with the sub-bottom profiler, high-resolution geophysical data, to define the bathymetric constraints associated with the anchor designing and locations. They overlaid the CPTs results on the sub-bottom profilers to integrate the different types of site investigation from the same site. Based on these integrated data, the geotechnical properties on the soil stratigraphy were considered for the decision of designed anchor locations.

More advanced application of ultra-high resolution 3D seismic data for the offshore geohazard identification is recently introduced (Brookshire Jr, Landers, & Stein, 2015). They applied the P-cable seismic system, which employs multiple short streamers towed from a cross cable diverted between two vanes. Since the heads of the relatively short streamers are fixed along the cross cable, the cross-line streamer spacing can be very short by the system design. Based on this acquisition system, the apparent subsurface horizontal resolution is in the order of 16 meters, and the vertical resolution of the sub-bottom profile is as high as 1.6 meters. This approach supported the likely possibility of applying geophysical data to the geotechnical risk assessment. However, the geophysical data is in still in time-domain and difficult to be compared with geotechnical data in space-domain.

1.2.1. Onshore surface wave inversion

The first application to convert the seismic data in time-domain into space-domain is introduced by (Stokoe, Wright, Bay, & Roesset, 1994) for geotechnical site

investigation. If the velocity of a seismic wave traveling in the soil media depends only on the physical properties, then the wave velocity is constant and independent of frequency. Such a material is called a nondispersive material, and the seismic waves traveling through this media will maintain a constant shape of waves (Stokoe, Joh, & Woods, 2004). This would be the case of waves in the homogeneous soil with uniform properties independent of the depth. However, most of the soils have different properties along with the depth, and they are dispersive materials, because of different deposit history of layering or variations in effective stress. Consequently, the group of waves with different frequencies changes the shapes as the wave travels through the soil media.

The different shapes in the wave group are distinguished by group velocity (v_g) and phase velocity (v_{ph}). The group velocity is the travel speed of the energy carried by the wave group, and the phase velocity is the speed of an individual wave in the group, and it depends on the frequency. In the case of nondispersive material, the group and phase velocities are the same, but in the case of dispersive material like soils, those velocities become different. The SASW technique uses this different frequency content of those velocities in the group of Rayleigh waves because this group of surface waves has the most seismic energy, about two-third, at shallow depth and attenuate a lot less than body waves (Briaud, 2013). The high-frequency waves only penetrate the shallow layers, and low-frequency waves penetrate much deeper layers in the soil deposit, so they give the shear wave velocity profile in shallow and deep layers, respectively. This method has been developed with improved data interpretation method (Joh, Rosenblad, & Stokoe, 1997), multichannel recording device (C. B. Park, Miller, & Xia, 1999; Yuan,

Zhu, & Kim, 2014), and artificial fixed source and receivers on the seafloor (B. Lee, Rosenblad, Wright, & Stokoe, 1997). The analytical case study of SASW along the seafloor used Scholte waves instead of Rayleigh waves on the seafloor between acoustic and elastic material layers and showed the potential benefit for the offshore site investigation. However, most of the offshore seismic data acquisition system is based on the survey system with towed hydrophones, so SASW is still not suitable for the offshore site characterization.

1.2.2. Offshore seismic exploration

Conventional marine operations for offshore seismic data acquisition involve one ship that tows both the sources and hydrophone streamers. When the station, the sources, streamers, and other towed equipment are unreeled from the stern of the large ship, which must maintain its forward motion so as not to lose control over the tow (Sheriff & Geldart, 1995). Navigation of the ship during recording is automatically done by GPS, and normal operations proceed at about 6 knots (11 km/hr or 3 m/s). Hence, about 250 km of CMP data could be recorded in a day with the interval distance about 30 to 45 meters apart when the sources activated every 10 to 15 seconds. A tremendous amount of data is generated from a large number of sensors with these short term shots for the marine survey. Modern marine 3-D seismic survey ships tow multiple sources and streamers with perhaps 500 channels each and generate a huge volume of data from closely spaced parallel 2-D subsurface lines (Yilmaz, 2001).

Common midpoint (CMP) is an important concept in multichannel seismic acquisition, and this represents the point on the surface halfway between the source and receiver that is shared by numerous source-receiver pairs. Such redundant seismic data among source-receiver pairs in the multichannel system suppress the noise signals and enhances the quality of seismic data when the data are stacked. The common-midpoint staking is regarded as the most important data processing application in improving data quality (Sheriff & Geldart, 1995), so the seismic data before and after this processing application are divided into two groups and defined as pre-stack data and post-stack data, respectively. This CMP stacking processing gathers all the raw seismic (pre-stack) data from multiple shots and stacks them as a seismic trace (post-stack) data at the CMP.

After conducting another signal processing applications, the seismic trace becomes the vertical geometry of one point, and a bunch of them as post-stack data shows the 3-dimensional geological formation. In short, most of the offshore seismic data is post-stack data, so the SASW application, which should utilize the pre-stack data, cannot be suitable for the offshore site characterization. Even though the multichannel analysis of surface waves (MASW) has been recently applied for underwater site characterization (Paoletti, Mouton, & Liposcak, 2010; Choon B. Park et al., 2005; Puech, Rivoallan, & Cherel, 2004), the observed data was not conventional offshore seismic data with stratigraphy. They measured the Scholte waves on the boundary where a body of water overlies solid materials, and the location of the towed seismic source and receivers were different from conventional data acquisition system, which requires the extra cost of special surveys.

1.3. Integrated Site Characterization

Since all geophysical site investigation methods have their limitations, integrated studies for the offshore site characterization has been highlighted. The integrated study is to put together all available offshore survey data, geotechnical, geophysical, geological, and oceanographic data from the same location to fully define the subsurface geo-site conditions and geologic processes (Campbell, Quiros, & Young, 1988). Although this concept seems to be simple, this requires to know every data from different site investigation methods and the details of the data can be complex.

One of the practical approaches for an offshore project showed how to use integrated geoscience data to avoid deepwater anchoring problems (Young et al., 2009). They described a proposed site risk assessment that can be used with existing geophysical and geotechnical data to support the evaluation of the anchoring problems, including the soil conditions, spatial soil variability, type of seafloor strength profile, confidence in computed anchor capacity, seafloor topography and so on. This approach showed the potential possibility to develop a particularly useful tool in the early stages of an offshore project before the acquisition of high-resolution geophysical data. This previous study motivated my research to develop more advanced data integration method by using a probabilistic approach for the fully understood geotechnical site characterization with quantified uncertainty.

2. INTEGRATED SEISMIC INVERSION WITH SPATIAL GAUSSIAN PROCESS

2.1. Summary

The stochastic seismic inversion based on Markov chain Monte Carlo (MCMC) method has been widely used for the imaging of unknown geological spatial information. Even though this seismic inversion approach can quantify the uncertainty of the estimated results, but it requires vertical in-situ well log data as the reliable initial accurate information to overcome the non-uniqueness problem in seismic modeling. However, most of the marine seismic surveys for the geohazard evaluation have only a few locations of drilled boreholes, and these data have limitations to apply to the entire target area for accurate site characterization. Recently, the trans-dimensional MCMC method has been used for the 3-D deep oil reservoir monitoring, but we cannot use this approach for the geohazard evaluation in heterogeneous shallow thin soil layers because of the inherent low seismic resolution problem. Therefore, the goal of this study is to apply spatial Gaussian Process (GP) to support the MCMC seismic inversion for the high-resolution near-surface site characterization from conventional offshore post-stack seismic survey data. We used a synthetic marine near-surface ground model in fine-scale grids as the benchmark model and applied the low-frequency seismic source to generate synthetic seismograms from the model. The results show that integrated MCMC seismic inversion with GP can find out thin dense soil layers in high-resolution regardless of the low frequency in observed seismic data.

2.2. Introduction

The most important part of the site characterization for the offshore construction projects is the integration of all available geotechnical, geological, and geophysical marine survey data. Geotechnical surveys from in-situ soil samples show the material properties in vertical depth, but this data is limited to shallow layers. Geological surveys which use drilling boreholes can reach the further depth, but those data shows only the material properties along with the depth, and cannot be extended to the horizontal direction to estimate the 3-dimensional ground model. Furthermore, the number of borehole data is not enough to interpolate accurately because of the high cost of offshore drilling. Thus, geophysical survey data, especially the near-surface seismic reflections, are the only available and important as the offshore site investigation for geotechnical projects, but this time-domain data is challenging to estimate the material properties in space-domain. For the estimation of shallow material properties in-depth, the seismic inversion has been used for the estimation of the P-wave velocity and density. The results from seismic inversion also provide the stratigraphic geometry of geological subsurface layers in-depth, so many offshore projects have relied on not only geotechnical data but also geological and geophysical survey data (Jeanjean et al., 2006; Medina-Cetina, Kang, Esmailzadeh, & Kallivokas, 2013; Medina-Cetina, Son, & Moradi, 2019).

Because of the non-uniqueness problem in seismic inverse modeling, the stochastic approach of Markov chain Monte Carlo (MCMC) method has been used to quantify the uncertainty of modeling parameters. Since Sambridge and Mosegaard

(2002) introduced the generalized MCMC method for geophysical exploration, advanced approaches were developed for Amplitude Verse Offset (AVO) inversion (Gunning & Glinsky, 2004) and Full-Waveform Inversion (FWI) approaches (Peyman P Moghaddam & Herrmann, 2010). The reversible jump Markov chain Monte Carlo (rj-MCMC) method in varying dimension (Green, 1995), which was first applied for the electromagnetic inverse problem (Malinverno, 2002) in applied geophysics, became distinguished and has been widely used for seismic inversion (Thomas Bodin & Sambridge, 2009; S. K. Dadi, 2014; S Esmailzadeh, Medina-Cetina, Kang, & Kallivokas, 2015a). This approach recently showed the analysis of uncertainty quantification from 2- & 3-dimensional seismic data for the deep reservoir monitoring (Cho, Zhu, & Gibson, 2017; Zhu & Gibson, 2016).

Even though those stochastic approaches have shown a great advantage of seismic inversion for the accurate site characterization, it cannot reach to the high-resolution subsurface imaging yet. The high-resolution spatial analysis is a critical issue for marine near-surface characterization because of the three challenging problems. First, we can obtain only a few information of the in-situ soil sample data, because drilling for site investigation may trigger a marine landslide. All the previous seismic inversion research for the deep hydrocarbon reservoir monitoring was possible based on in-situ well-log data, which is important to prevent the non-uniqueness problem in seismic inverse modeling as accurate initial information. However, drilling on shallow marine slope area will make a new path of vertical seawater flow, and may reach to the deep unknown geological fault and trigger an earthquake. Second, even though the

multi-channel system improved the horizontal resolution of seismic surveys, the vertical resolution of seismic data is limited to the seismic source frequency and not good enough to identify thin soil layers. Autonomous underwater vehicle (AUV) used to apply an extremely high-frequency acoustic source for the near-surface site characterization, but the energy loss from attenuation also causes the limitation of data depth problem. Last, the last challenging issues for the near-surface marine site characterization is that the shallow sediments show extremely high heterogeneity, so engineers require high-resolution subsurface imaging results, which can show less than a one-meter thickness of soil layers. In short, we need to use conventional low-frequency seismic data to identify thin soil layers without in-situ well-log data, but the seismic non-uniqueness problem makes it difficult to find the accurate ground model. Consequently, the geotechnical risk assessments still cannot be fully integrated with the geophysical seismic survey data.

In this paper, we introduce a new approach of integrated MCMC seismic inversion based on spatial Gaussian process (GP) for the offshore near-surface site characterization. We hypothesized that spatial random fields from the Gaussian process in anisotropic condition could represent thin horizontal layers between offshore boreholes. The realizations of spatial random fields were used as the proposed candidates of MCMC seismic inversion based on low-frequency observed seismic reflection data. Since the realizations were generated from fine-scale borehole data, the imaging results of seismic inversion could show the high-resolution subsurface ground model. To demonstrate the numerical experiments, we generated a 2-dimensional shallow offshore ground model, which has the soil properties of the Hydrate Ridge on

the West coast of Oregon State. Synthetic seismic traces were generated over the entire ground model by the seismic convolution modeling with 100 *Hz* Ricker wavelet. From those limited observed data, we applied integrated seismic inversion from the spatial Gaussian process and compared to trans-dimensional MCMC seismic inversion, which is recently developed and applied for deep oil reservoir monitoring.

2.3. Methodology

2.3.1. Stochastic seismic inversion

Stochastic inverse modeling based on Markov chain Monte Carlo (MCMC) method uses the modeling parameters as random variables and changes them to maximize the Probability Density Function (PDF) of the posterior ($p(\boldsymbol{\theta}|\boldsymbol{d})$). The calculation of this conditional probability follows the Bayesian paradigm.

$$\boldsymbol{p(\theta|d_{obs})} = \frac{\boldsymbol{p(d_{obs}|\theta)p(\theta)}}{\boldsymbol{p(d_{obs})}} \tag{2.1}$$

The $\boldsymbol{\theta}$ represent the modeling parameters in forward modeling. We use seismic convolution method, whose parameters are the depth of layer interfaces, material properties of soil density, and P-wave velocity at each layer. This forward modeling calculates the seismic impedance in each subsurface layer and the reflection coefficients along with the depth. Those data in space is converted to time domain based on the P-wave velocity information and generate synthetic seismogram based on source wavelet. The seismogram becomes our predicted data ($\boldsymbol{d_{pred}}$) and we compare this to observed data ($\boldsymbol{d_{obs}}$). The important term in the Bayesian framework is the likelihood function ($p(\boldsymbol{d_{obs}}|\boldsymbol{\theta})$), which quantifies the PDF from the data residual ($\epsilon = \left|\boldsymbol{d_{pred}} - \boldsymbol{d_{obs}}\right|$).

$$p(\boldsymbol{d_{obs}}|\boldsymbol{\theta}) = \frac{1}{[(2\pi)^n|C_d|]^{\frac{1}{2}}} * \exp\left[-\frac{1}{2}(\epsilon)^T C_d^{-1}(\epsilon)\right] \quad (2.2)$$

The advantage of this Bayesian paradigm for the stochastic seismic inversion is that we can apply our previous information as priors ($p(\boldsymbol{\theta})$) to support the calculation of the posterior PDF. For the seismic inverse modeling, physical known properties of seawater and subsurface seawater in below list can be used as our priors.

- Density and P-wave velocity of seawater; $\rho_w \sim N(1000,10), Vp_w \sim (1500,10)$
- Depth of seafloor; $z_w \sim N(20,1)$
- Range of material properties in marine sediments; $\rho_s \sim Unif(1500,2200)$

where, ρ_w is the density and acoustic P-wave velocity of seawater, which are known value as 1000 kg/m^3 and 1500 m/s respectively. Since the depth of seafloor (z_w) is relatively clear to estimate in the offshore seismic survey, we applied it as our prior. The minimum and maximum value of soil bulk density properties (kg/m^3) are also available from the measurement of borehole soil samples, and those become the range of uniform distribution as another prior for the inverse modeling.

2.3.2. Spatial regression model

Spatial Gaussian process is a stochastic process to generate the spatial regression model ($Y(s_{ij})$) from Gaussian random fields ($w(s_{ij})$). The spatial model is defined as

$$Y(s_{ij}) = X(s_{ij})^T \beta + w(s_{ij}) + \epsilon(s_{ij}) \quad (2.3)$$

where $X(s_{ij})$ is a set of observed data, and β is the regression parameter to find the deterministic regression model of the spatial target field. The regression remains the

residual, and the residual is partitioned into two terms: $w(s_{ij})$ is stationary Gaussian random process with the covariance matrix, while $\epsilon(s_{ij})$ represents the nugget effect as white noise (Carlin, Gelfand, & Banerjee, 2014; Cressie & Wikle, 2015).

The covariance matrix is calculated by the spatial standard deviation matrix and correlation matrix. The correlation function uses the coordinate information to define the correlation between certain distances, and we chose the correlation function as an exponentially decaying function;

$$y(d) = \exp\left(-\frac{d}{\phi}\right) \tag{2.4}$$

where d is the distance between two points in the spatial domain, and ϕ is the range parameter. Since the anisotropy condition is important for geological stratigraphy estimation, we estimated two range parameters in a horizontal and vertical direction independently. The value of the vertical range parameter ($\phi_{vertical}$) came from the observed well log data, and we decided the horizontal one ($\phi_{horizontal}$) from the seismic data, which only has spatial information in a horizontal direction. The standard deviation is modified to apply the location information of the observed in-situ boreholes. We controlled the standard deviation with respect to the distance from the boreholes to apply higher uncertainty to the far-fields from the drilled wells. All of this spatial Gaussian process is implemented on vertical cross-section, and the calculated the final covariance matrix genets anisotropic geological stratigraphy model, similar to the 2-D seismic image data.

2.3.3. Integrated Markov chain Monte Carlo method

The key point of the integrated MCMC seismic inversion with a spatial Gaussian process is that the candidate modeling parameter ($\boldsymbol{\theta}'$) for MCMC is generated from the spatial regression model ($Y(s_{ij})$). Figure 2.1 shows the workflow of this integrated seismic inversion method. When we obtain all available offshore survey data, we need to sort them into geological borehole data or geophysical seismic data. Both contain spatial information, but the interpretation and extraction methods of valuable information from them are different. From the geological borehole data, we used to apply them to GIS analysis system and apply a statistical approach to interpolate unknown information between borehole data. Geophysical seismic data, which shows the subsurface image in only time-domain, needs seismic inversion to estimate the material property and depth of geological layer interfaces. Those two work processes were independent before, but we apply the anisotropic random fields from spatial Gaussian process as the proposed candidate for the stochastic seismic inversion process. Since these candidate samples came from the high-resolution borehole data and contain the anisotropic geological information, the result of MCMC seismic inversion can estimate the heterogeneous stratigraphy ground model in the fine-scale discretized spatial domain.

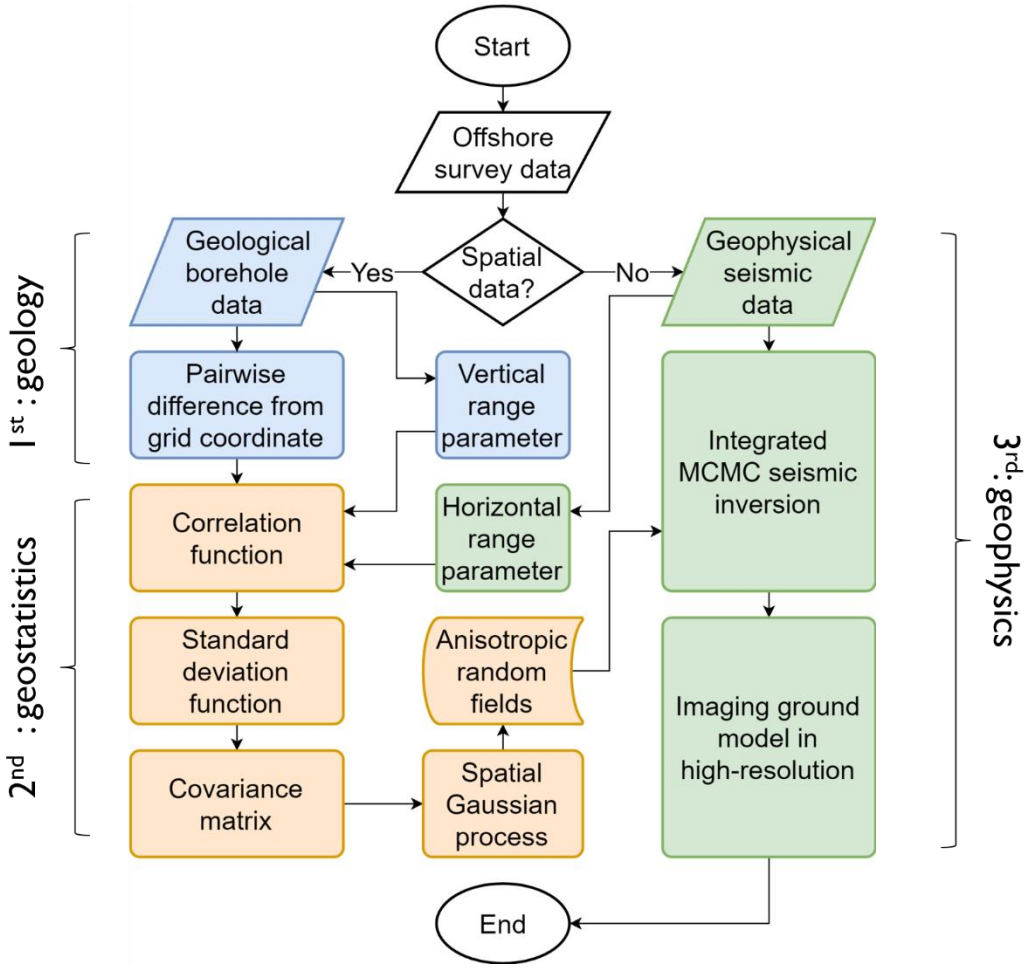

Figure 2.1. Workflow of the integrated MCMC seismic inversion method

2.4. Numerical Experiments

2.4.1. The dataset from the west coast of Oregon State

The west coast of the Oregon State is one of the example locations to discuss how to use our various types of offshore survey data for better site characterization. Since this area contains thin layers of gas hydrate in a shallow depth, many geological and geophysical offshore surveys have been implemented in this area (Bangs, Musgrave, & Trehu, 2005; Kumar, Sen, Bangs, Wang, & Pecher, 2006). Figure 2.2 shows the

locations of drilled boreholes from geological surveys, and the black line on the map indicates the path of a geophysical seismic survey on the same target area. The two boreholes (ODP 204-1246 in red and ODP 204-1250 in yellow) are on the path of this geophysical survey. Since the integration of those two different types of data is challenging, we can see there are other extra boreholes (white) to estimate the accurate subsurface ground model between those two boreholes (red & yellow).

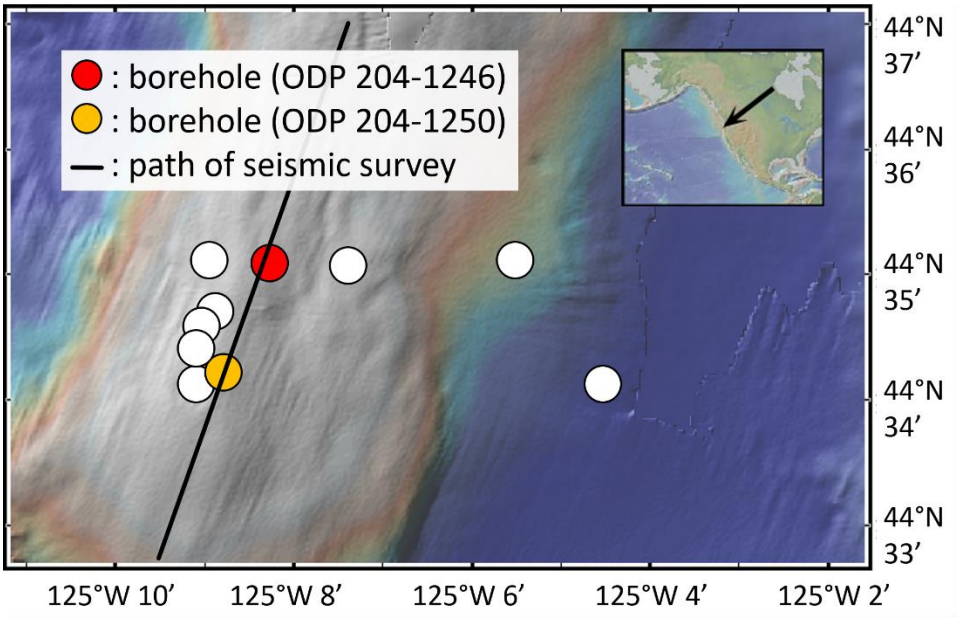

Figure 2.2. Map of geological and geophysical offshore surveys for gas hydrate characterization

All the drilled borehole data are managed in the International Ocean Drilling Projects (IODP), and we obtained in-situ vertical soil bulk density data from those two boreholes (ODP 204-1246 & 1250). Based on those field data, we interpolated these raw data into fine-scale 1-meter discretization in depth from the seafloor to 80 meters below the seafloor and used them to build up our synthetic shallow offshore soil ground model

for the numerical simulation (Figure 2.3). The two boreholes were used as left and right-side boundaries of 1 km width synthetic model, and we interpolated the unknown soil properties inside the model with an assumption of flat seafloor bathymetry. This simple shallow offshore soil ground model in high-resolution shows thin dense soil layers (red in Figure 2.3) at the depth 40 or 60 meters below the seafloor, and we used this as our target model to prove our integrated seismic inverse modeling.

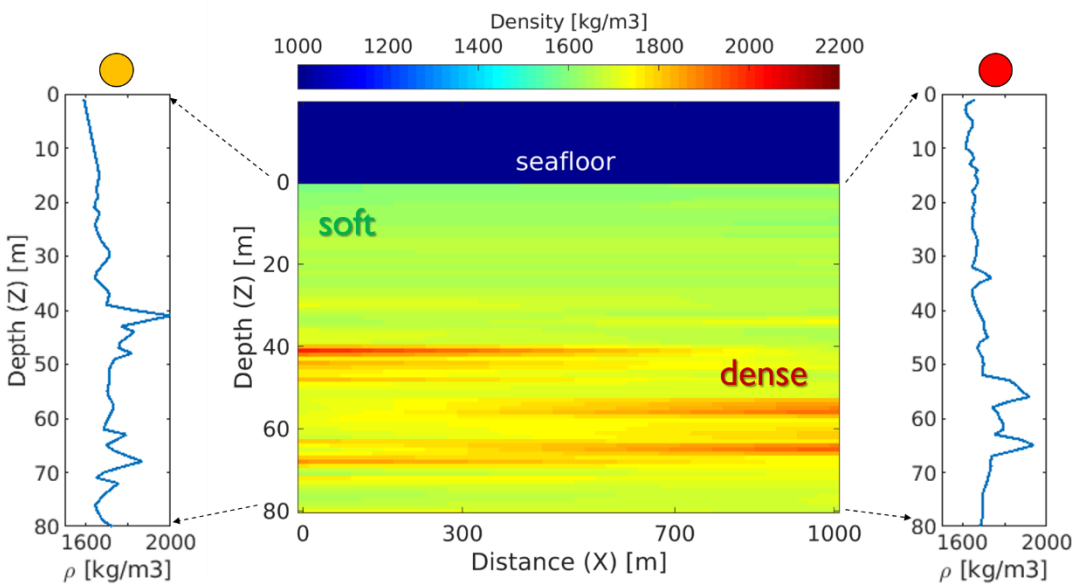

Figure 2.3. Synthetic marine shallow ground model

Since the drilling is very limited in the offshore environment, we assumed that we had only four locations of drilled borehole data in this target area. The locations of them are 0, 300, 700, and 1000 meters in the horizontal distance (Figure 2.4-a). We named those locations A, B, C, and D, and each location has the observed vertical soil profiles (Figure 2.4-b) as accurate in-situ information. The geological soil profiles from the drilled boreholes show two dense thin layers, which has about 3 meters of thickness.

Even though we could get important geological stratigraphy information from the observed borehole data, all other parts between boreholes are still unknown, and it is difficult to estimate the geometry of the subsurface stratigraphy model or properties of the ground soil materials.

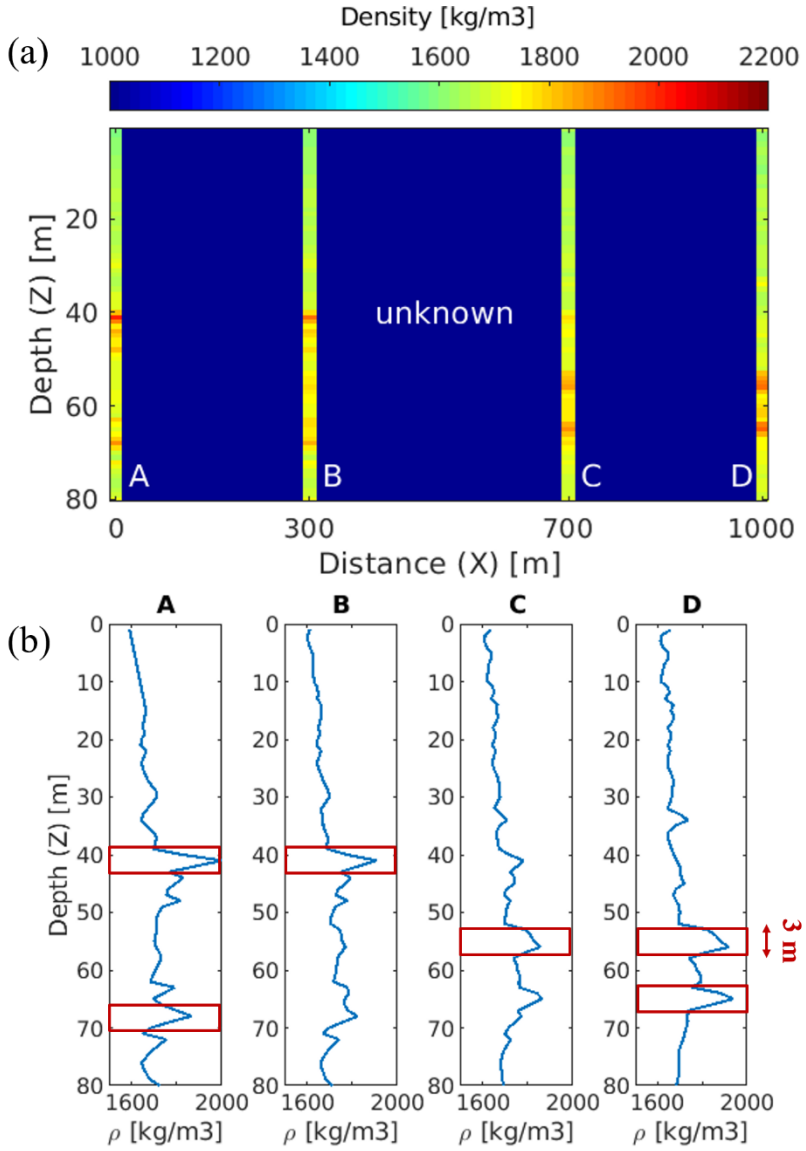

Figure 2.4. Locations of four observed boreholes

Even though the geological data from drilled boreholes are not enough for site characterization, we can apply geophysical survey data for the characterization of this target ground model. We assumed the seismic data are measured every 20 meters in this area; that means 51 shots in 1,000 meters width to get 51 seismic traces as geophysical data. Figure 2.5-a shows the synthetic seismogram, which usually displayed in a grey-scale image. The high amplitude of reflections (white) shows the seafloor and shallow geological strata in the time domain. The P-wave velocity in this shallow synthetic soil ground model was assumed as homogeneous 2,300 m/s, so the heterogeneous bulk density in Figure 2.3 made the different reflection coefficients inside this ground model to generate those seismic reflections. Since we applied the seismic convolution method with 100 Hz frequency Ricker wavelet as the seismic source, the resolution of this seismic data is not high enough for the shallow site characterization to use them for the geohazard evaluation.

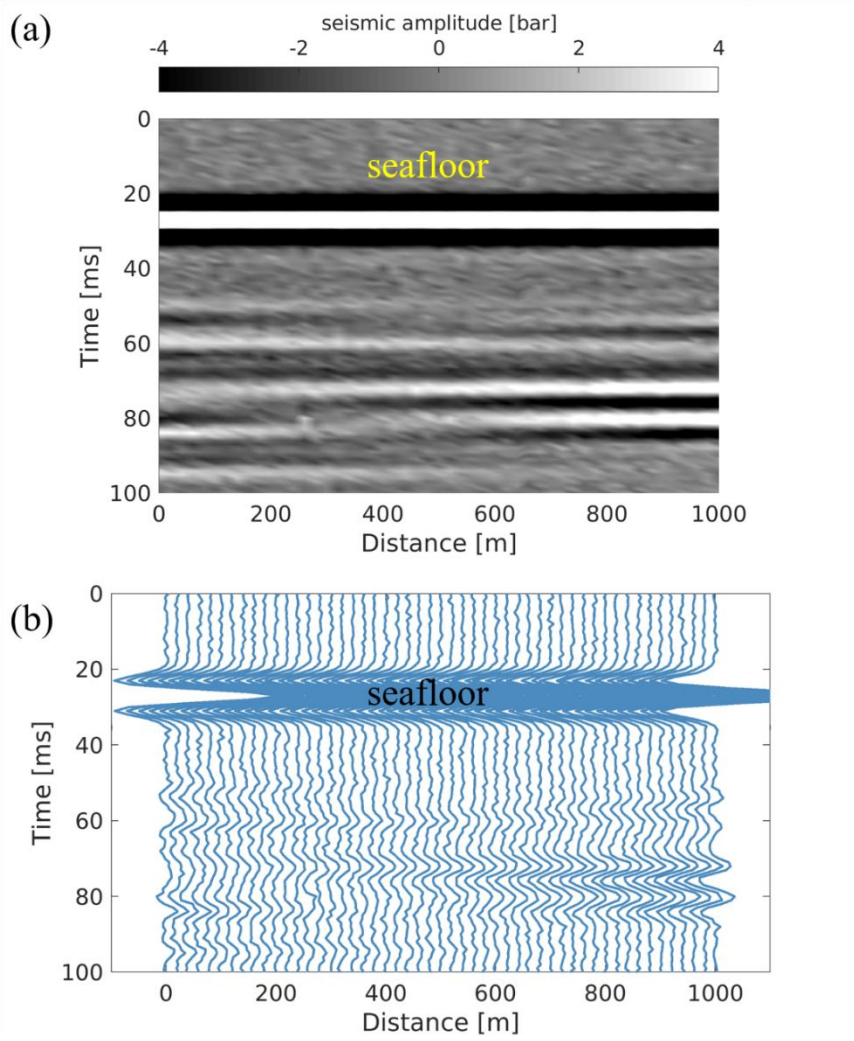

Figure 2.5. Synthetic seismogram with low-frequency source wavelet

2.4.2. Geophysical convolution

Figure 2.5-b shows the same seismic data, as in Figure 2.5-a in a different visualization method. The seismic data is a bunch of wiggling traces, so we applied the convolution as the seismic forward model to capture the features of each seismic trace. Convolution is the mathematical operation of multiplying two-time series signals representing two input data strings. It is one of the most useful and widely used

operations in seismology and digital signal processing because it represents the physical process of combining two seismic signals (Zhou, 2014). A seismogram (D) can be approximated as the convolution ($*$) of a source wavelet (G) with the ground medium function (F). Synthetic seismogram by convolving well-log traces with an appropriate wavelet is often computed to match the in-situ well-log measurements with a seismic reflection profile on the same target area.

$$F(t) * G(t) = D(t) \tag{2.5}$$

Mathematically, multiplication of two-time series vector data leads to another time series vector data as the one convolved data from two original data. We can express the time series data as $a_t = [a_1, a_2, \dots, a_n]$, $b_t = [b_1, b, \dots, b_m]$, and make two matrices from them.

$$A = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ a_2 & a_1 & \ddots & \vdots \\ \vdots & a_2 & \ddots & 0 \\ a_n & \vdots & \ddots & a_1 \\ 0 & a_n & \vdots & a_2 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_n \end{bmatrix} \tag{2.6}$$

where A is the matrix with $N + M - 1$ raws & M columns. Each column contains the time series a_t and it is shifted one element downwards sequentially from the second column. The result of the convolution of a_t and b_t comes from the matrix Ab.

$$c = Ab = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ a_2 & a_1 & \ddots & \vdots \\ \vdots & a_2 & \ddots & 0 \\ a_n & \vdots & \ddots & a_1 \\ 0 & a_n & \vdots & a_2 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \tag{2.7}$$

We need to take the k^{th} row of the above equation, so the convolved signal (c_k) becomes below equation where k goes from 1 to $N + M - 1$, and the index i move through all the non-zero elements of the two input time series.

$$c_k = \sum_i a_{k-i+1} b_i \tag{2.8}$$

2.4.3. Gaussian process

As described in previous part, spatial regression model ($Y(s_{ij})$ is the summation of the deterministic regression model ($X(s_{ij})^T \beta$) and spatial random field ($w(s_{ij})$). Since we already had four locations of observed boreholes and knew the horizontal geometry of the subsurface layers from the obtained seismic data, first we applied the piecewise constant interpolation between boreholes. Figure 2.6 showed the four locations of drilled boreholes (A, B, C, and D), and the four divided subdomains around those boreholes. The horizontal linear regression roughly estimates the material properties between boreholes. For example, the subdomain near the borehole B (distance at 300 m) has a 340-meter width from 160 to 500 m as the left and right side of the boundary. We chose those boundaries at 0, 160, 500, 860, and 1000 meters in the horizontal distance where the farthest points from the observed borehole locations are.

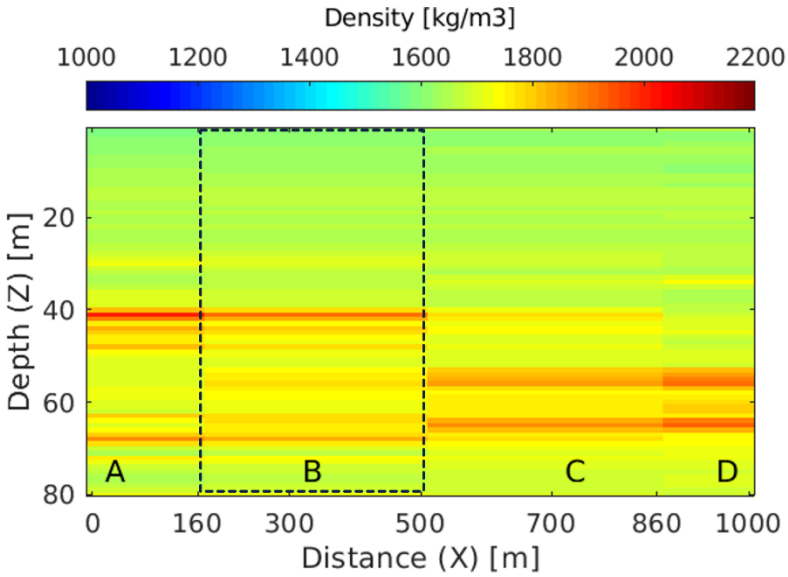

Figure 2.6. Deterministic interpolation from borehole data

Since the rough estimation from Figure 2.6 cannot be accurate enough, we need to apply the spatial Gaussian process (GP) to generate random fields ($w(s_{ij})$) to add on the previous result. The spatial GP relies on the covariance matrix, which is generated from the standard deviation matrix and the correlation matrix. As the first step of the GP, we need to extract the coordinate information from the ground model and calculate the pairwise difference matrix for both horizontal and vertical direction (Figure 2.7). The size of the elements in our synthetic ground model is 1 x 20 meters in the vertical and horizontal direction, respectively. Thus, the ground model consists of 80 elements in vertical and 51 elements in horizontal, so the total number of them is 4080. The pairwise difference matrix has the 4080x4080 size and contains every Euclidean distance between all elements inside the ground model.

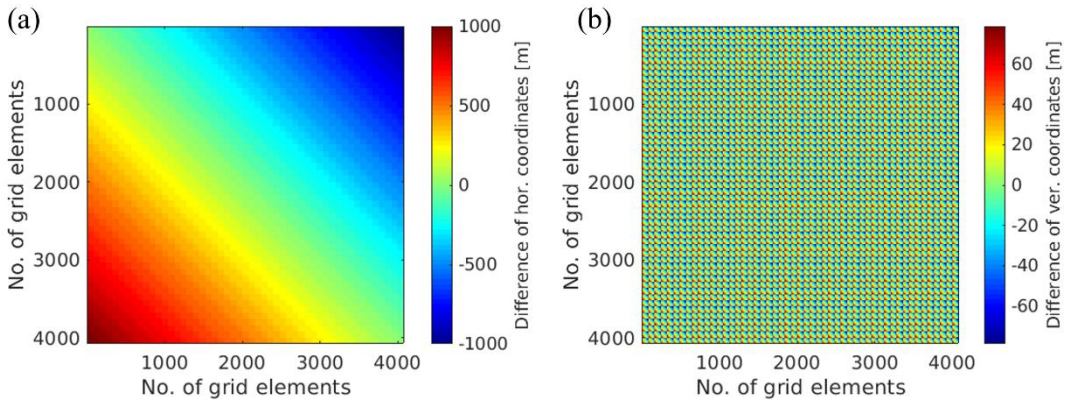

Figure 2.7. Pairwise difference matrix from coordinates

The correlation matrix has the same size as the pairwise difference matrix, and the correlation function uses the distance between each element to calculate the correlation coefficient. The anisotropy condition is the most important part of the correlation function, and this requires two range parameters in horizontal, vertical directions. From the stratigraphy analysis on the observed four boreholes (Figure 2.4-b), we found out the thickness of the thin dense layer was about 3-meters and used it as the vertical range parameter ($\phi_{vertical}$). On the other hand, the horizontal range parameter ($\phi_{horizontal}$) was defined from geophysical seismic interpretation. Even though the seismic data has horizontal information in multichannel streamer accusation system, the length of correlated geological strata was not clear to define, so we roughly chose 300 meters as the horizontal range parameter for the correlation function.

In geostatistics, which can be defined as the branch of statistical science that studies spatial phenomena and capitalizes on spatial relationships to model possible values of variables at unobserved locations (Caers, 2005), variograms are used to describe the geological continuity and quantify the spatial correlation. However, this

study didn't use the variogram, because the thickness of extraordinary thin dense layers were clear from the stratigraphy analysis to define the range parameter. Furthermore, the variograms also has a major limitation of its inability to model realistic geological features. Since the variogram models correlation between only two spatial locations, they are poor discriminators of geological scenarios even with exhaustive data.

The spatial Gaussian process requires not only the correlation matrix but also the standard deviation matrix to define the covariance matrix for the spatial random field generation. One of the main advantages of the Gaussian process in this study is that we can control the amount of uncertainty in a random field by using modified standard deviation function along with the distance (Figure 2.8-a). Since we already know the locations of the observed boreholes, we don't need to allow any uncertainty on those borehole locations. Otherwise, the unknown locations, for example, the center between two boreholes as the far-field, should contain a larger value of standard deviation to allow high uncertainty for the accurate site characterization. Thus, our standard deviation varies with the distance from the observed borehole locations, so the function has the shape of hat function (Figure 2.8-a) to control the uncertainty of the random field generation. The maximum value of the standard deviation for GP is chosen as the same value as the standard deviation in the observed borehole data (Figure 2.4-b). As a result, the allowed uncertainty for the random field generation is almost zero at the observed borehole locations; A, B, C, and D points (0, 300, 700, 1000 meters in horizontal direction), and the center point of the ground model (500 meters in horizontal direction) has the largest uncertainty. The generated standard deviation matrix from this function is

multiplied to the correlation matrix (Figure 2.8-b) to make the covariance matrix (Figure 2.8-c). Since we applied the low standard deviation at the four borehole locations, the points A, B, C, and D in the covariance matrix show almost zero value, which had another value on the diagonal in the previous correlation matrix.

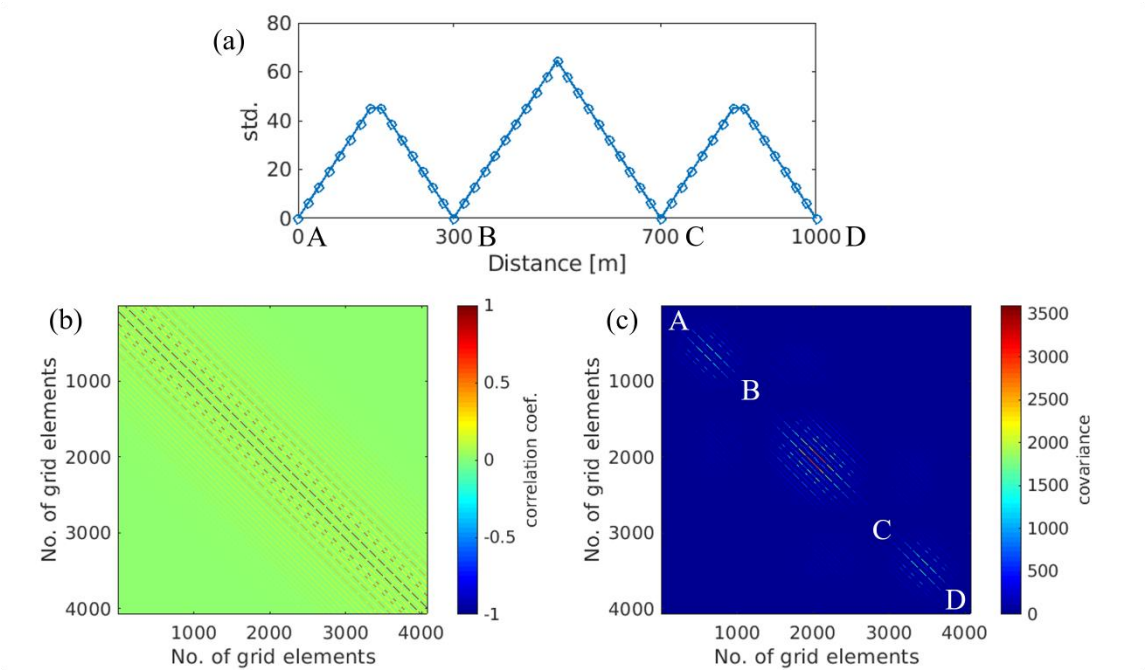

Figure 2.8. (a) Standard deviation for Gaussian process, (b) Correlation matrix, and (c) Covariance matrix

Figure 2.9-a shows one of the generated random fields ($w(s_{ij})$) from the spatial Gaussian process. Since we applied anisotropy condition with uneven range parameters for correlation function, the generated random field shows horizontal long and narrow geological clusters. Furthermore, the random field shows almost zero values in four columns of the observed borehole A, B, C, and D locations because of the modified standard deviation function (Figure 2.8-a). From the covariance matrix, we generated

100,000 samples of random fields and combined them with the previous deterministic regression result ($X(s_{ij})^T \beta$) in the Figure 2.6 to generate the final spatial regression model ($Y(s_{ij})$) in the Figure 2.8-b. Consequently, the generated 100,000 spatial regression models became the geological candidate models, and we proposed those models to the seismic MCMC inversion process, as explained in the previous workflow (Figure 2.1).

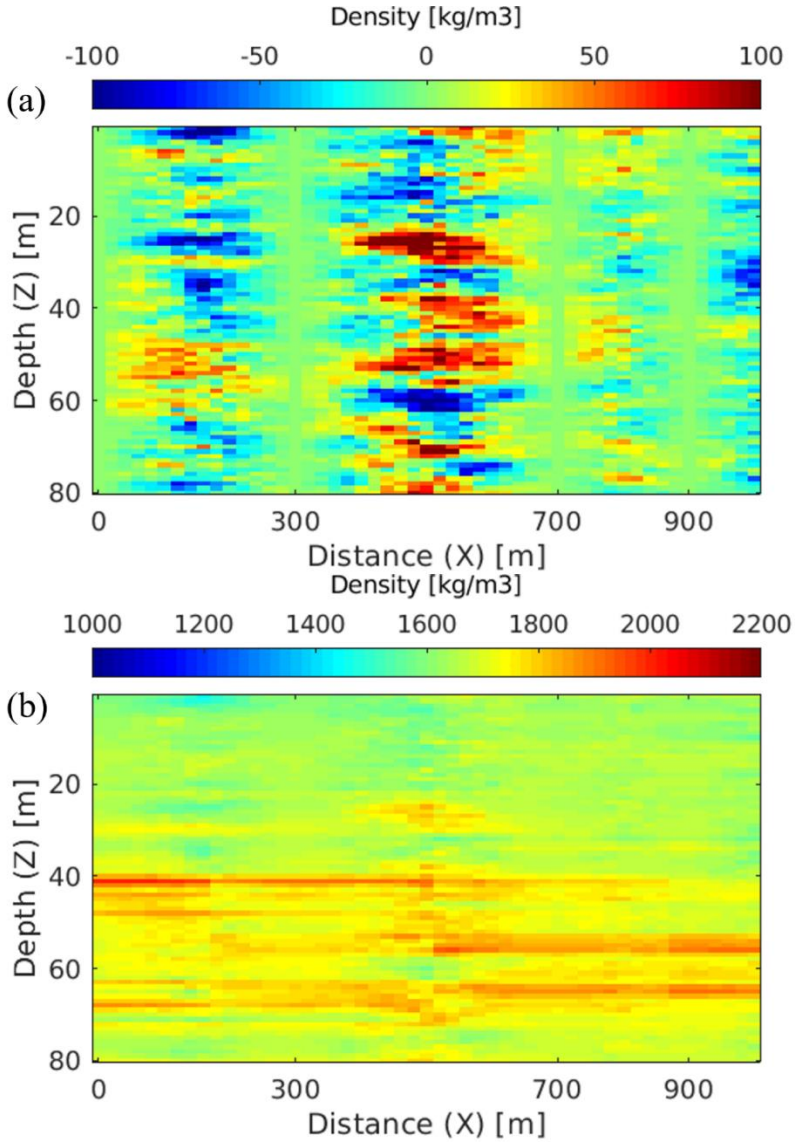

Figure 2.9. (a) One of the spatial random fields from Gaussian process, and (b) one estimated realization as the final spatial regression model

2.4.4. Integrated MCMC seismic inversions

From the spatial regression model with Gaussian process, we got 100,000 samples of 51 columns inside the offshore soil ground model. We added the top seawater layer over the ground model and applied the properties of the soil and seawater in every 1-meter discretized space domain to the seismic convolution forward model to generate synthetic seismograms. Each of 51 predicted seismic-trace signals are compared with observed seismic data at the same location. In short, the reconstruction of the 2-D ground model image requires seismic inverse modeling of 51 independent MCMC simulation to generate each column inside the ground model, and we should gather them together to make the cross-section surface.

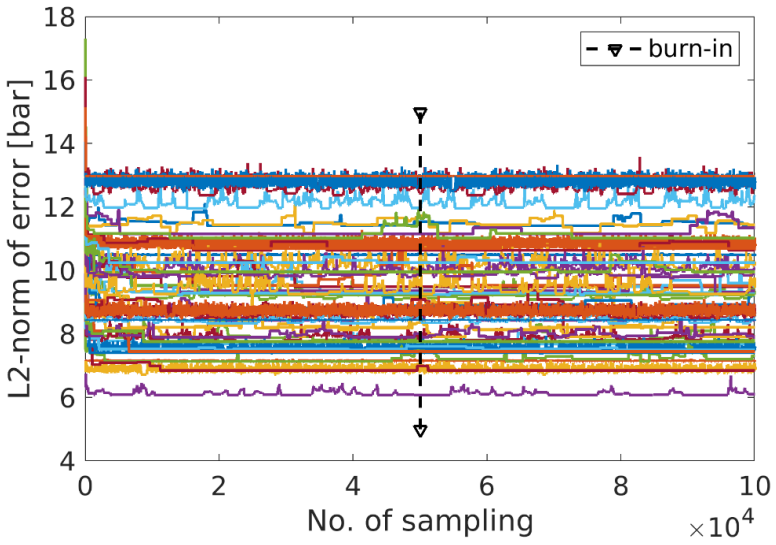

Figure 2.10. Convergence of the 51 MCMC inversions after the burn-in point

Figure 2.10 shows the samplings from 51 independent MCMC inversion at 51 different locations with each observed seismic trace. From the 100,000 samples from the

spatial random fields, MCMC method chose a few of them, and the stochastic process shows a convergence in the L2 norm of residual errors between the predicted and observed seismic signal data. We decided the half point, 50,000 sampling point, as our burn-in point for every 51 simulations, and measure the mean and standard deviation from the chosen samples (Figure 2.11). Since this result comes from the spatial Gaussian regression model from the fine-scale observed vertical borehole data, the estimated final result of the ground model (Figure 2.11-a) shows the heterogeneous soil properties in the high-resolution image. This image is the set of independent columns, and some of those columns are different from the target ground model (Figure 2.3). However, the image can show much better stratigraphy inside the ground model with keeping its high-resolution information than the simple interpolated ground model (Figure 2.6). Furthermore, the quantified standard deviation (Figure 2.11-b) shows less uncertainty near the observed borehole locations (0, 300, 700, and 1000 meters in the horizontal direction), and increasing uncertainty along with the distance from those borehole locations. For example, the uncertainty around 850 meters in Figure 2.11-b has higher value of standard deviation than those at 700 or 1000 meters locations.

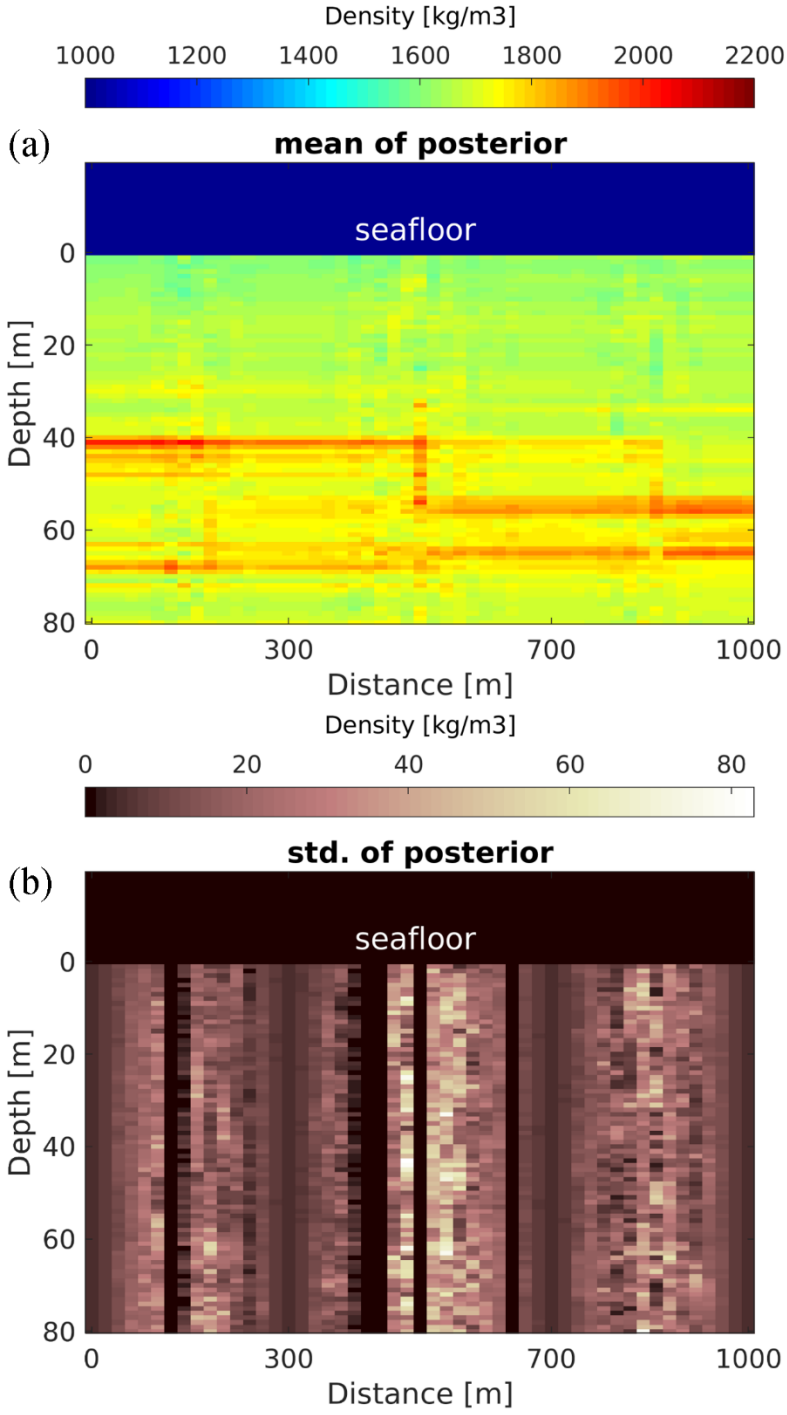

Figure 2.11. Final results of the integrated MCMC seismic inversion using spatial random fields from Gaussian process; (a) mean of the posterior distribution and (b) quantified uncertainty

2.5. Discussion

2.5.1. Random variables in the high-resolution

The quantified uncertainty from the case study showed some columns with an extremely low value of standard deviation (Figure 2.11-b). For example, the location 500 meters, shows almost zero uncertainty because the seismic MCMC inversion couldn't collect any proposed samples from the Gaussian regression model. This zero acceptance happens because this study used neither Bayesian rule nor conventional normal distribution for proposing candidates in seismic inversion. We need to clarify that the only 100,000 realizations, which were generated from the spatial regression model were used for the integrated MCMC seismic inversion.

The conventional stochastic approach, which defines each 81 elements as a fully random variable, is not able to use for this high-resolution seismic imaging because of two reasons. First, if we make all 81 small size elements to be changed in random in the MCMC method, the seismic inversion does not show the convergence. That is the seismic non-uniqueness problem. Since the wavelength is about 23 meters, we have 23 independent elements inside the wavelength. Thus, we cannot let those 23 elements to make one block of seismic signal in a stationary condition. Second, if we apply the fully random perturbation, we will lose the anisotropy condition in our ground model. The seismic inversion is based on each trace for 1-D site characterization, so multi-channel seismic data itself cannot generate the anisotropic random field. In short, the conventional MCMC random variable approach cannot be appropriate for the high-resolution imaging, and spatial Gaussian process is important for this integrated MCMC

method. This also indicates the importance of the appropriate deterministic regression model before the Gaussian process, so this approach also requires more future works in the pre-required regression process for better site characterization.

2.5.2. Trans-dimensional seismic inversion

Since the previous case study of the integrated MCMC seismic inversion method with spatial Gaussian process showed some limitation, we applied the recently developed another method, trans-dimensional MCMC seismic inversion to the same benchmark offshore ground model. This method was introduced for the 3-D monitoring of deep oil reservoirs and showed a great benefit when we know some information about observed in-situ well logging data. This seismic inversion is similar to our previous case study, but the outstanding feature of this rj-MCMC seismic inversion method is that the number of geological layers is an unknown random variable.

The trans-dimensional MCMC is called reversible-jump Markov chain Monte Carlo (rj-MCMC) method, and this approach has a strong point to identify the unknown stratigraphy for the seismic inversion. This approach in varying dimension has been widely used for the geoscience, and the key point of this rj-MCMC method is that the number of unknown geological layers is also another random variable in the stochastic MCMC inversion process. Since the dimension of MCMC becomes a random variable, the model selection of rj-MCMC method becomes a random process to choose one of four types of proposals (S Esmailzadeh et al., 2015a) to generate a new candidate modeling parameter ($\boldsymbol{\theta}'$).

- BIRTH: randomly pick a depth location to create one new layer with the new material property.
- DEATH: randomly pick one of the current layers to remove it and merge to the neighboring layer.
- MOVE: randomly pick one of the current layers and change the depth of the layer interface.
- PERTURB: randomly pick one of the current layers and change the material property.

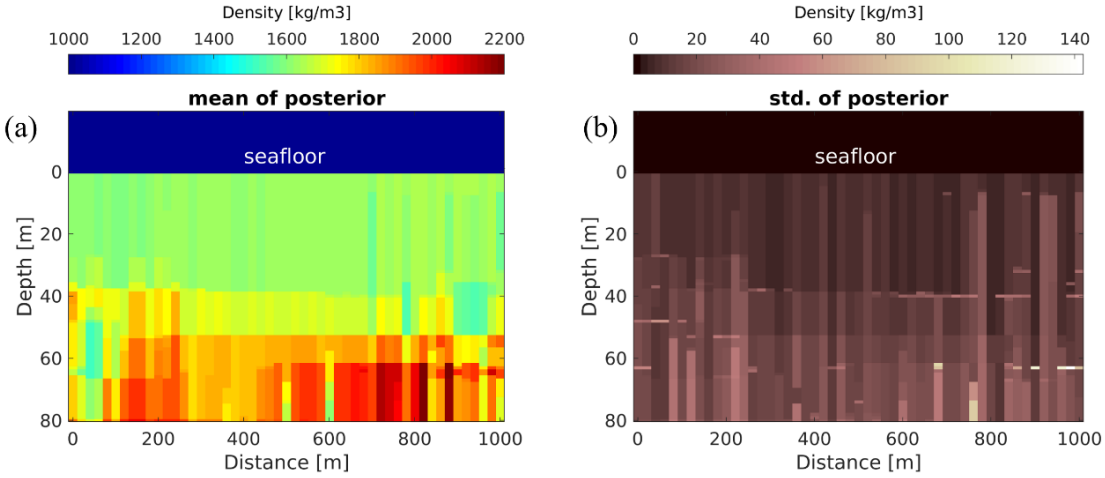

Figure 2.12. Results of the trans-dimensional rj-MCMC seismic inversion; (a) mean of the posterior distribution and (b) quantified uncertainty

The four different types of modeling selection in rj-MCMC seismic inversion estimate the number of geological layers from the observed seismic data. However, the estimated results from this rj-MCMC approach for the shallow site characterization (Figure 2.12) was different from the target ground model (Figure 2.3). The chosen number of layers from this approach is about 5 or 6 and extremely lower than what we

expected in the high-resolution image. Because the seismic wavelength in this 100 Hz frequency simulation was about 23 meters, so the theoretical seismic vertical resolution is about 5-meters. Furthermore, seismic data shows the reflection only at the high contrast layer interfaces, so the 3-meters thickness of the dense soil layers are overestimated to about 10-meters. This inherent seismic low resolution from low-frequency seismic source consequently make the thicker layer estimation and estimate upscaled stratigraphy model under the seafloor. In short, the rj-MCMC seismic inversion method, which is developed recently and showed a great potential possibility for the seismic imaging cannot be appropriate for the shallow site characterization for the high-resolution stratigraphy modeling.

2.6. Conclusion

This paper demonstrated the integrated Markov chain Monte Carlo (MCMC) seismic inversion with spatial Gaussian process (GP) for the offshore shallow site characterization in high-resolution. Spatial regression models are generated from fine-scale geological borehole data and contain the information of thin layers in heterogeneous shallow ground model. The deterministic spatial interpolation and Gaussian process generate anisotropic random fields to represent the aspect of the spatial variation of the geological ground model, and those realizations are used as proposed candidate models to the seismic inversion. Even though the seismic signal frequency is still low as same as the conventional offshore field survey data, the estimated ground model from the posterior distribution of the samples shows the subsurface stratigraphy

model in high-resolution. Some of the columns in the predicted ground model show different values from the target material properties, but this problem can be solved when the various types of deterministic regression models are applied together to support the stochastic approaches. This integrated seismic inversion approach with spatial regression can prevent the extra costs of drilling offshore boreholes, and improve the quality of the offshore ground model estimation when we already have both geological and geophysical survey data on the target area.

3. PROBABILISTIC GEOLOGICAL STRATIGRAPHIC MODELING

3.1. Summary

Stratigraphy estimation from in-situ borehole soil samples is a process to identify the vertical geologic formation under the seafloor. This offshore site characterization process is important for marine slope stability analysis; however, the estimated vertical soil profiles have contained uncertainty problems from the data measuring errors. In this paper, we applied probabilistic Bayesian calibration with a reversible jump Markov chain Monte Carlo (rj-MCMC) method to estimate the depth of layer interfaces and soil properties of each layer in varying dimensions. The spatial average method, based on step function operator, is applied to define the geological strata along with the depth for the stratigraphic modeling. We applied Bayesian calibration to this model with a field shear strength profile from a borehole data at the northeast Gulf of Mexico, where many studies warned the geohazard threat against the marine landslide. The results of probabilistic calibration show the different geological estimation based on different prior information to compare with other data from geophysical or geotechnical surveys. We expect that the results from geological offshore site investigation will be able to support the geomechanical slope stability analysis on this area as reliable in-situ information of subsurface soil layers.

3.2. Introduction

Offshore projects for renewable energy and oil-producing platform have proliferated in the past few years. These projects require to design reliable foundations to support up to the huge size of structural facilities, and many advanced studies about anchors and piles have been studied (Aubeny, 2017; Randolph & Gourvenec, 2017). However, the seafloor ground contains highly heterogeneous and inelastic mud sediments, so we cannot guarantee how the foundation on that unstable soil would behave. Furthermore, the external loads from the foundations may trigger submarine landslide due to the unstable seafloor slope stability condition. Many geomechanical slope stability analysis has been studied (Locat & Lee, 2002; Nadim et al., 2003; Nadim, Lacasse, Choi, & Hadley, 2014; Nowacki et al., 2003) for the geohazards assessment against the marine debris flow threat which would damage the installed geotechnical foundations for offshore infrastructures.

The primary information of soil properties for the marine slope stability analysis should be based on accurate site investigation, and many advanced methods have been studied for offshore site characterization. All those methods rely on the soil properties from in-situ drilled borehole soil samples to show the vertical profile under the seafloor. However, there are three critical problems to estimate this vertical soil profile from the offshore borehole data. First, the soil samples from offshore boreholes are disturbed in the laboratory tests. The offshore sediments are sensitive to the extreme change of pressure and temperature in the indoor laboratory. Furthermore, the measurement of offshore undrained shear strength data is usually measured by using the handheld

penetrometer. That device is appropriate to measure the large scale offshore data in a short time, but it cannot be accurate enough. Thus, the results sometimes show extraordinary outliers, but it is difficult to define them as measurement errors or valuable geological changes. Second, the drilled borehole data may have some missing parts due to losing soil samples during the drilling extraction. Third, we cannot take enough borehole data from our target area because the cost of drilling in offshore is expensive, so the site characterization from boreholes always suffers from a lack of field data.

Because of these challenging issues in stratigraphy analysis, probabilistic calibration method has been widely used for geological analysis with borehole and well log data. The Bayesian inference with Markov chain Monte Carlo (MCMC) method has been used with a forward model, called spatial average method. This forward model is a vertical classification method to make soil blocks with a step function. This function defines the depth of layer interfaces and material properties at each layer for the stratigraphy analysis, but the number of layers was chosen manually and fixed (Richard L. Gibson & Hwang, 2009). Recently, trans-dimensional approach, called reversible jump Markov chain Monte Carlo (rj-MCMC) method (Green, 1995), was applied to the upscaling problems of well logging data (S. Dadi, Gibson Jr, & Wang, 2015). This method defines the number of layers as another random variable and shows reliable stratigraphy result of vertical profiles. Since this method can identify the accurate information of vertical profiles, many studies for the seismic inverse problem has been developed based on this stratigraphy estimation with rj-MCMC method (T Bodin,

Salmon, Kennett, & Sambridge, 2012; Thomas Bodin & Sambridge, 2009; Cho et al., 2017; S Esmailzadeh et al., 2015a; Zhu & Gibson, 2016).

In this study, we introduced our trans-dimensional soil calibration approach with the rj-MCMC method and applied it to the vertical soil profiles from in-situ field borehole data in the northeast Gulf of Mexico. Because of complex geologic formation from the Mississippi sediment deposit, this area has been studied for marine landslide analysis (Sawyer, Flemings, Dugan, & Germaine, 2009). We applied the Bayesian calibration method to the shear strength borehole data, which is important for the geomechanical analysis of slope stability but has not been discussed before. We also explained the importance of Bayesian priors when we need to compare those geological borehole data with other geophysical or geotechnical survey data at the same location. We expect the result of this study could help more advanced risk assessment against the threat of a marine landslide.

3.3. Methodology

3.3.1. Modeling parameters for vertical soil classification

To estimate the depth of layer interfaces from the borehole data, we applied the spatial average method as our forward model, which has been used for the upscaling of well-logging data to classify the fine-scale measurement into course scale layer groups. This method has been widely used for geoscience (S. Dadi et al., 2015; Richard L. Gibson & Hwang, 2009) and petroleum engineering (Mondal, Mallick, Efendiev, & Datta-Gupta, 2014) to estimate the depth of geological formations and thickness of deep

hydrocarbon reservoirs, respectively. Since the subsurface layering relies on the geological deposit process, the soils in one layer can be defined as one homogeneous group which has similar material properties. When this geologic deposit process has significant changes, the sediments property changes with a new layer interface. Figure 3.1 shows an example of spatial block averages overlaid on the scattered shear strength vertical profile as raw data. Since the raw observed data (d_{obs}) shows obvious discontinuities at a certain depth, and layer interfaces can be defined in predicted modeling data (d_{pred}) in dimension five.

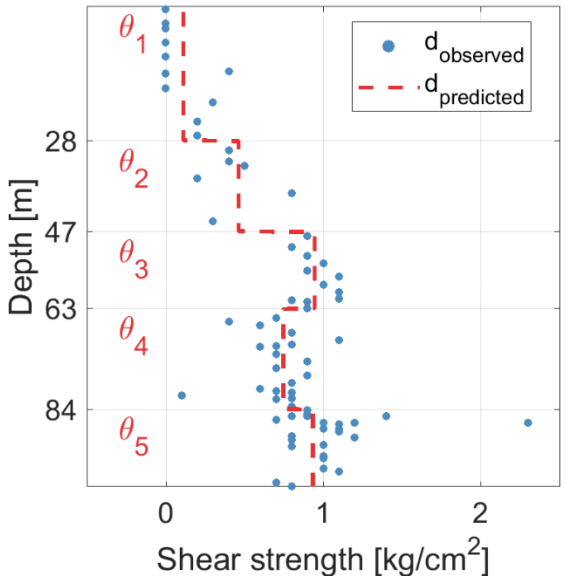

Figure 3.1. Stratigraphic modeling with vertical calibration forward model

Parameters (θ_i) at each layer contains the information of the layer interface depth and the material property along with the depth. The calibration of vertical profile modeling for borehole data analysis has three main parameters, the number of layer (k) within our target depth, depth of layer interfaces (z_i), and the material properties (β_i) for

each layer. The parameter k controls the dimensions of other parameter vectors z_i and β_i, and set of those parameters become the main parameter ($\boldsymbol{\theta}$) in varying dimensions.

$$\boldsymbol{\theta} = [k, z_i, \beta_i] = [k, z_1, \dots, z_k, \beta_1, \dots, \beta_k] \tag{3.1}$$

The spatial average method as a forward model uses the main parameter with the discretized space domain in our target depth to make blocks of soil layers. The Bayesian paradigm for the probabilistic calibration defines those parameters as random variables to estimate the numbers and location of layer blocks from the in-situ borehole data.

3.3.2. The Bayesian paradigm for probabilistic calibration

The main goal of probabilistic calibration is to sample the posterior probability density function (pdf) that is central to Bayesian inference, which updates our prior information of physical properties based on the observed data (D.G.T. Denison, Holmes, Mallick, & Smith, 2002). The posterior pdf of the model parameters can be expressed as

$$\pi(\boldsymbol{\theta}|\boldsymbol{d_{obs}}) = \frac{\pi(\boldsymbol{\theta}) \times f(\boldsymbol{d_{obs}}|\boldsymbol{\theta})}{\int \pi(\boldsymbol{\theta}) \times f(\boldsymbol{d_{obs}}|\boldsymbol{\theta}) d\boldsymbol{\theta}} \tag{3.2}$$

where $\boldsymbol{\theta}$ is the vector of model parameters, and d_obs represent the observed data which features the vertical borehole profile data in this study. The prior knowledge regarding the model is in the prior distribution $\pi(\boldsymbol{\theta})$, while the likelihood function $f(\boldsymbol{d_{obs}}|\boldsymbol{\theta})$ quantifies the probability of obtaining data values from the forward modeling with the given model parameter vector $\boldsymbol{\theta}$. The likelihood pdf compares the observed and predicted data to minimize the difference between those data during the calibration process. The absolute value of this modeling error distribution follows the multivariate

Gaussian distribution with zero mean value. Thus, the likelihood function can be expressed as below.

$$f(\boldsymbol{d_{obs}}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}}|C_d|^{\frac{1}{2}}} \times$$

$$\exp\left[-\frac{1}{2}\,(g(\boldsymbol{\theta}) - \boldsymbol{d_{obs}})^T (C_d)^{-1}(g(\boldsymbol{\theta}) - \boldsymbol{d_{obs}})\right] \tag{3.3}$$

where the $g(\theta)$ is the spatial average method as a forward model, and C_d is the data covariance matrix. We applied the variance times identity matrix ($\sigma^2 * I$) as our covariance matrix without data correlation and n is the number of observations.

Because our model parameters are defined in the different dimensions with the different number of layers, the prior and likelihood functions are updated with the newly proposed dimension (k) for every iteration. Thus, we applied hierarchical Bayesian priors to support the posterior pdf calculation. The number of layers (k) is the fundamental prior, which follows a Normal distribution, and the priors for modeling parameter $\boldsymbol{\theta}$; specified by the parameters z_i and β_i, rely on the dimension prior. We applied uniform distribution for those modeling priors to define the minimum and maximum range of our target depth and material property. We also applied an additional non-informative prior to support the likelihood function. The normal distribution of likelihood has a dynamic standard deviation(s), and the non-informative prior ($p(s)$) has inverse Gamma distribution with hyperparameters, as explained in S Esmailzadeh et al. (2015a).

$$\pi\boldsymbol{\theta}|\boldsymbol{d_{obs}}) \approx f(\boldsymbol{d_{obs}}|\boldsymbol{z}, \boldsymbol{\beta}, k, s)p(s)p(\boldsymbol{z}|k)p(\boldsymbol{\beta}|k)p(k) \tag{3.4}$$

3.3.3. Trans-dimensional stochastic process

To implement the trans-dimensional calibration, we need to sample the proposed candidate with the reversible jump Markov chain Monte Carlo (rj-MCMC) method. We follow the model selection algorithm introduced by S Esmailzadeh et al. (2015a) for the rj-MCMC method and update it for vertical data calibration. This approach defines four types of model selection; BIRTH, DEATH, MOVE, and PERTURB. One of these types is randomly chosen during the stochastic process with four given probabilities; b_k, d_k, m_k, and p_k, respectively. In the BIRTH and DEATH types, the algorithm randomly chooses one current layer and proposes to create another new layer or erase one of the current layers. The MOVE type also randomly choose one of layer interface and allows to change the vertical depth. The PERTURB type doesn't change the layer depth but applies random perturbation of material property for all layers. The moving of layer depth follows uniform distribution inside the depth range of upper and lower interfaces, and perturbing of material property follows a normal distribution with the coefficient of variance.

- BIRTH: a type of proposing to add new parameters (z*,β*) at one newborn layer with the probability b_k at a randomly chosen depth point. The chosen layer is divided into two, and the proposed dimension k* becomes to k+1.
- DEATH: a type of proposing to remove a randomly chosen current layer with the probability d_k. The parameters in the chosen layer become the same values to its upper layer or lower layer. The proposed dimension k* becomes to k-1.

- MOVE: a type of proposing to change the parameters at a randomly chosen one current layer with the probability m_k. The depth of only chosen layer changes within its boundaries. The proposed dimension k* is equal to the current dimension k.
- PERTURB: a type of proposing to change the material properties in every current layer with the probability p_k. The location of layer boundaries doesn't change, so the proposed dimension k* is equal to the current dimension k.

The rj-MCMC method requires generalized Metropolis-Hastings sampling (MH) algorithm (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) that samples from posterior of varying dimension. Procedures of the generalized MH algorithm need to be explained as two steps. The first step is to propose a new candidate model ($\boldsymbol{\theta}^*$) from the current model ($\boldsymbol{\theta}$) and calculate the proposal ratio from the distributions $q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ and $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ for the varying dimensional model selection. When we apply the MOVE and PERTURB types in fixed dimension, those proposal distributions are canceled out, and the proposal ratio becomes one, but this ratio becomes b_k/d_k or d_k/b_k respectively when we apply the BIRTH and DEATH types with probability b_k and d_k. The second step is to apply these ratio to the calculation of acceptance probability (α).

$$\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = min\left\{1, \underbrace{\frac{f(\boldsymbol{D_{obs}}|\boldsymbol{\theta}^*)}{f(\boldsymbol{D_{obs}}|\boldsymbol{\theta})}}_{\text{likelihood ratio}} \times \underbrace{\frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta})}}_{\text{prior ratio}} \times \underbrace{\frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}}_{\text{proposal ratio}} \times |\boldsymbol{J}|\right\} \qquad (3.5)$$

The J is a Jacobian matrix which normalizes the difference in volume between two spaces of different dimensions. The varying dimension of BIRTH and DEATH types requires this term, and the mathematical proof of this is explained in S Esmailzadeh et al. (2015a). The acceptance probability (α) is compared with another uniform random deviate (u). If u is smaller than α, the newly proposed candidate is accepted, and the current model is updated with the proposed one. Otherwise, the proposed candidate is rejected, and the current model remains in the Markov-chain to propose another new candidate.

3.4. Experiments of Calibration Modeling

3.4.1. Field borehole data from the Gulf of Mexico

We applied our Bayesian probabilistic calibration with the rj-MCMC method to an offshore in-situ borehole data. We chose the in-situ field data from Integrated Ocean Drilling Program (IODP) Expedition 308, the site U1322 borehole data in the Ursa basin at the northeast Gulf of Mexico (Figure 3.2), 210 km south-southeast of New Orleans, Louisiana, USA (Reece, Flemings, Dugan, Long, & Germaine, 2012). This location is a margin of the marine continental shelf with the sediments deposits from the Mississippi River (Yamamoto & Sawyer, 2012), and this large mass transport deposits made a complex geological structure in this area (Sawyer et al., 2009; Sawyer, Flemings, Shipp, & Winker, 2007). Direct pore fluid pressure was measured in this area (Fleming et al., 2008), and they found overpressure zones, which may trigger the submarine landslide (Sawyer, Flemings, & Nikolinakou, 2014). The data were used to reveal an active

hydrodynamic environment of this area, and provided insight into the geological process at the deep subsurface layers. Since our goal is to support shallow slope stability analysis from this site investigation data, we decided to use the shallow part of this borehole data. We applied the probabilistic calibration modeling to the shear strength soil profiles to figure out the engineering soil properties at this shallow layers and estimate the layer interfaces and material property at each layer. Observed raw data (Figure 3.3) shows the measurement error problem; meaningless low values at the shallow depth, missing parts, extraordinary outliers as discussed above introduction.

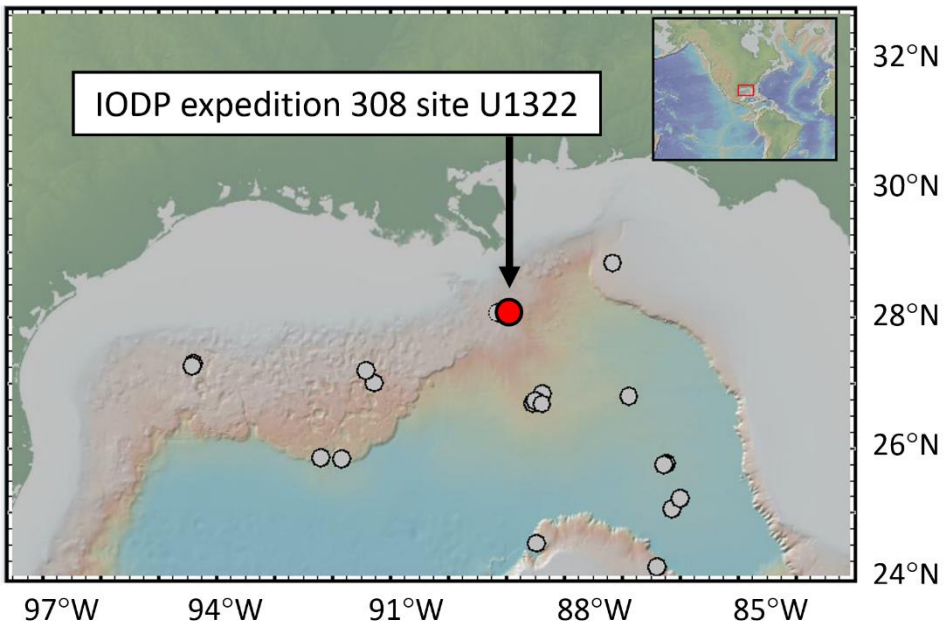

Figure 3.2. Location of the chosen borehole data in the northeast Gulf of Mexico

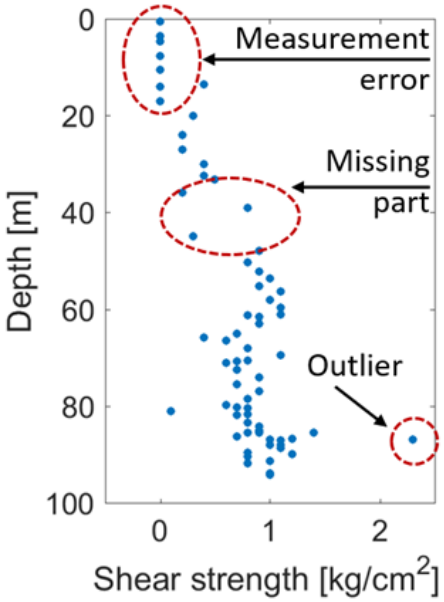

Figure 3.3. Observed in-situ borehole data from ODP 308 U1322 site

3.4.2. Preprocessing and interpolation

Since the in-situ soil data contains undefined errors, we applied the statistical preprocessing and applied the interpolation to this raw data (Figure 3.4). We measured the sampling intervals and checked the cumulative density function and histogram with relative frequency to figure out the best uniform sampling interval. The distribution of sampling intervals has maximum relative frequency near the 0.8 meters, but we chose a much small value of 0.2 meters for the interpolation not to lose any depth information on fine-scale. The figure shows our interpolated data matches well to the raw data, and this process excluded the extreme outlier near the 85 meters depth below the seafloor.

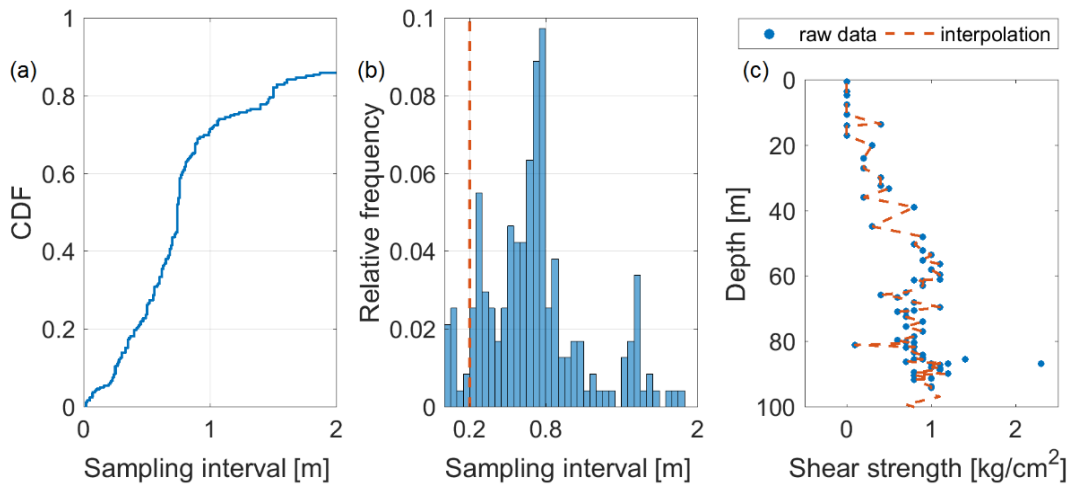

Figure 3.4. Statistical analysis for the raw borehole data and interpolation

3.4.3. Non-informative calibration

Based on the interpolated soil profile data, we applied trans-dimensional calibration with the rj-MCMC method to estimate the number of subsurface layers and the depth of each layer interface. Since we wanted to focus on the in-situ geological data itself without other geophysical or geotechnical data, the calibration was non-informative Bayesian modeling with only the prior of material property; uniform distribution with the range from minimum to maximum observed values. Our target depth is from the seafloor to 100 meters below from the seafloor, and observed range of shear strength material profile is from 0 to 2.5 kg/cm^2. We applied that observed information to set up the priors with uniform distribution.

$$z \sim Unif(0,100) \tag{3.6}$$

$$\beta \sim Unif(0,2.5) \tag{3.7}$$

The probabilities of model selection were all equal; 0.25 for BIRTH (b_k), 0.25 for DEATH (d_k), 0.25 for MOVE (m_k), and 0.25 for PERTURB (p_k), to avoid subjective model parameters. The stochastic process took 200,000 samples during the calibration, and each sample had a different number of layers from one to ten as the dimension of the forward modeling. Figure 3.5 shows the varying dimension during the iterations, which start from only one homogeneous initial condition. The convergence of dimension and histogram shows that most of the chosen samples have 11 layers.

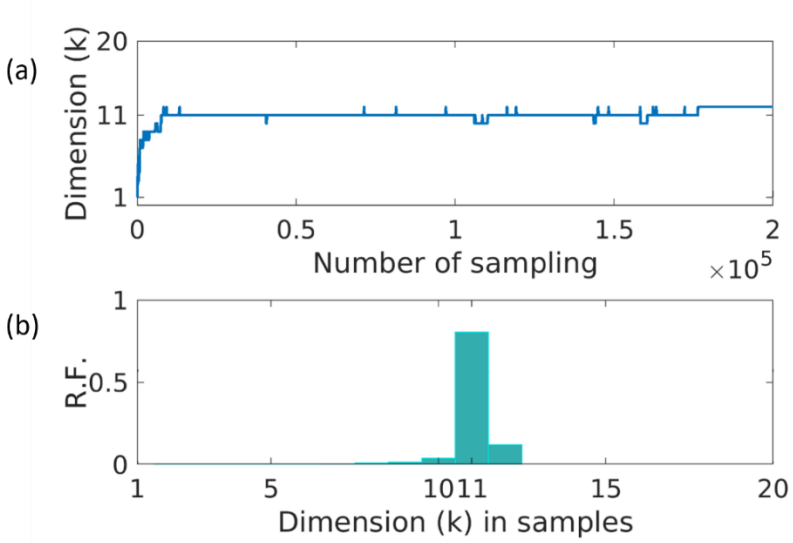

Figure 3.5. Convergence of stochastic process in varying dimensions

To identify the stationary condition in this stochastic process, we chose only samples whose modeling dimension is 11. More than 150,000 samples were collected, and we plot the surface of the modeling parameter (Figure 3.6-a). We defined 75,000 sampling point as our burn-in point to choose the samples in a stationary condition. The calculation of cumulative mean and standard deviation verified that the modeling parameters in the samples after the burn-in point are stationary at each depth. Figure 3.6-

b & c show three chosen depth (25, 50, and 75 meters) to plot the shear strength properties as random variables.

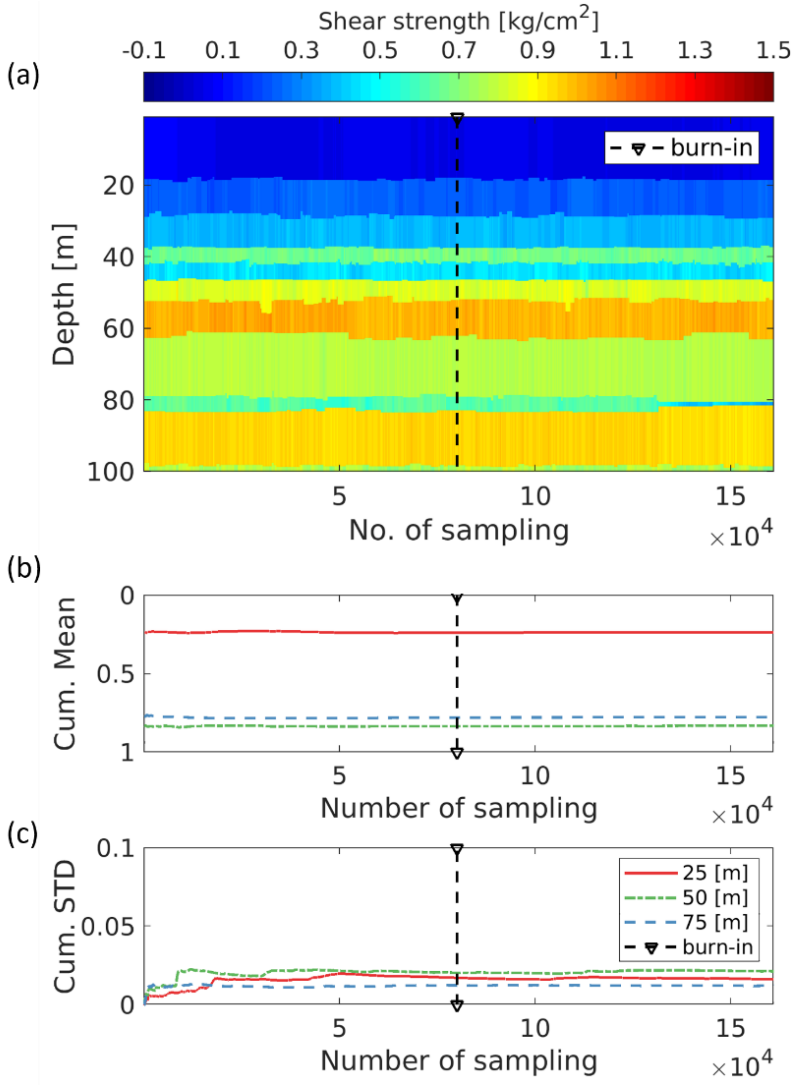

Figure 3.6. Estimation of the burn-in point in the stationary condition

Figure 3.7-a shows the mean of the posterior distribution of the samples (black) as our final result of the estimated vertical soil profile in depth. Even though the initial soil model (red) was homogeneous, the stratigraphic modeling generates vertical random

profiles as realizations (grey) based on the observed borehole data (blue). The probability and uncertainty of those realizations can be quantified (Figure 3.7-b). The soil strata with a yellow indicator have a high probability, which is enough to confirm the soil property. However, others with green indicator mean high uncertainty, where we should consider the potential possibility of soil heterogeneity.

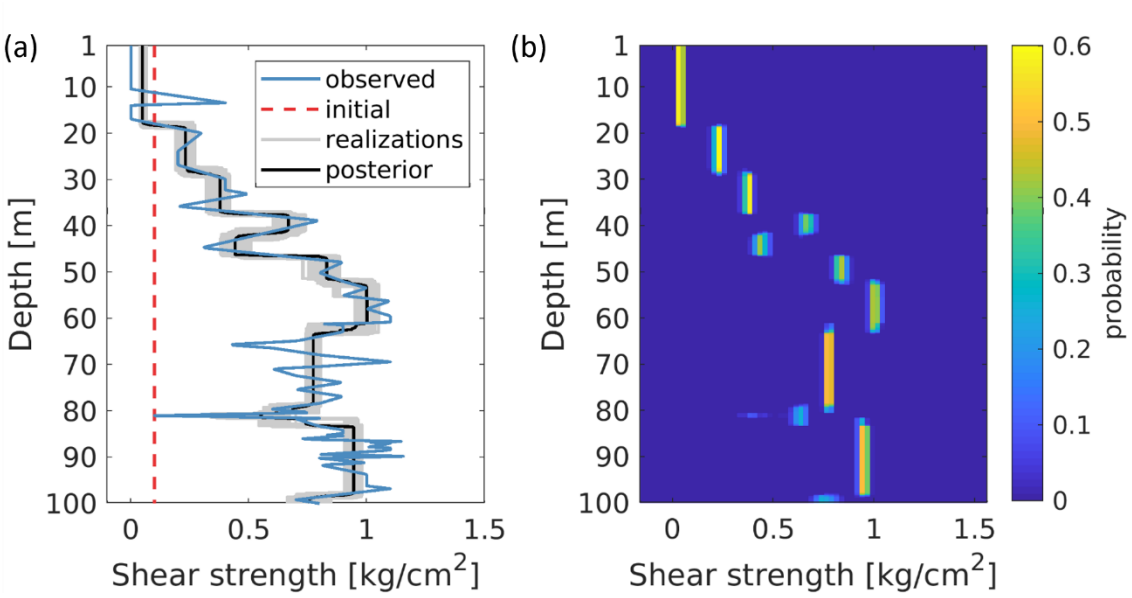

Figure 3.7. Estimated geological stratigraphic model

3.4.4. Informative calibration in coarse-scale

We applied only general priors to the previous non-informative stratigraphic modeling. However, we should consider this approach in course scale when we need to compare the in-situ geological borehole data with geophysical survey data at the same location. Because geophysical seismic data has low-resolution problem due to its inherent frequency limit. For example, if we use 30 Hz air gun source for offshore surveys, the wavelength in seawater becomes about 50 meters. Since the vertical

resolution of seismic data is a quarter of the wavelength, seismic data can show the stratigraphic layers with more than 12.5-meter thickness. Thus, we should apply a new prior of assuming the dimension for our stratigraphic modeling that it would be around 8 layers in our target depth; 100 meters below from the seafloor.

$$k \sim N(8,1) \tag{3.8}$$

After applying this new prior to our previous Bayesian framework, the dimension of stratigraphic modeling from the same in-situ borehole data becomes smaller. The estimated number of layers from the previous non-informative calibration was 11, but the informative calibration with a new prior defined only 6 layers (Figure 3.8).

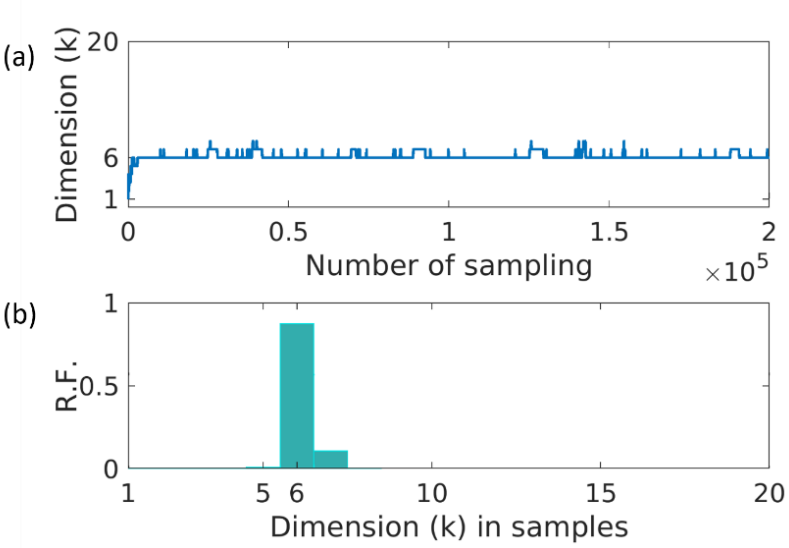

Figure 3.8. Convergence of the stochastic process in varying dimensions with Bayesian prior in coarse-scale

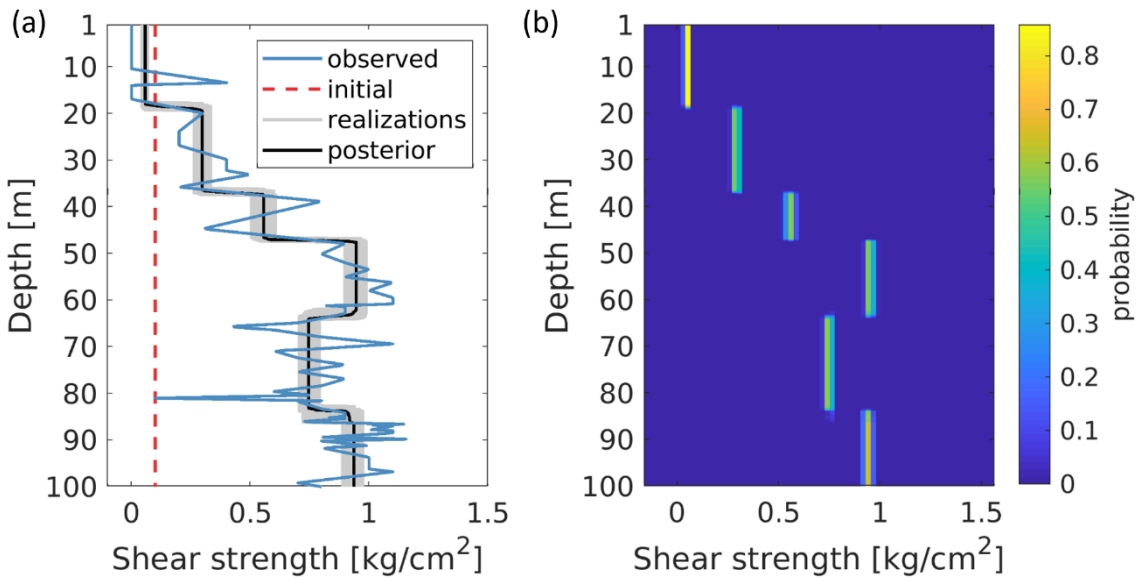

Figure 3.9. Estimated stratigraphic model with Bayesian prior in coarse-scale

3.4.5. Informative calibration in fine-scale

Geophysical data has low-resolution problem, but geotechnical offshore survey data; such as cone penetration test (CPT), box coring, can measure the soil properties in high-resolution. However, those geotechnical investigation methods can measure only shallow depth, so the high-resolution soil data also has to be compared with other geological borehole data or geophysical seismic data to estimate the ground model in subsurface deep soil layers. Since our non-informative calibration showed 11 layers, we decided to assume 20 layers to get fine-scale stratigraphy result.

$$k \sim N(20,1) \tag{3.9}$$

This third calibration with another new prior for fine-scale analysis shows 15 layers. Since this calibration was more accurate to define specific thin layers in the

borehole data, the results show less probability with high uncertainty than previous course scale calibration.

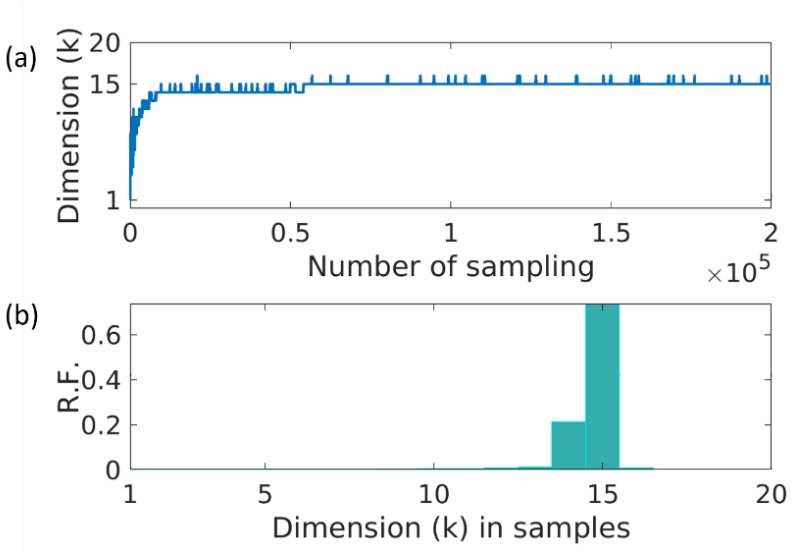

Figure 3.10. Convergence of the stochastic process in varying dimensions with Bayesian prior in fine-scale

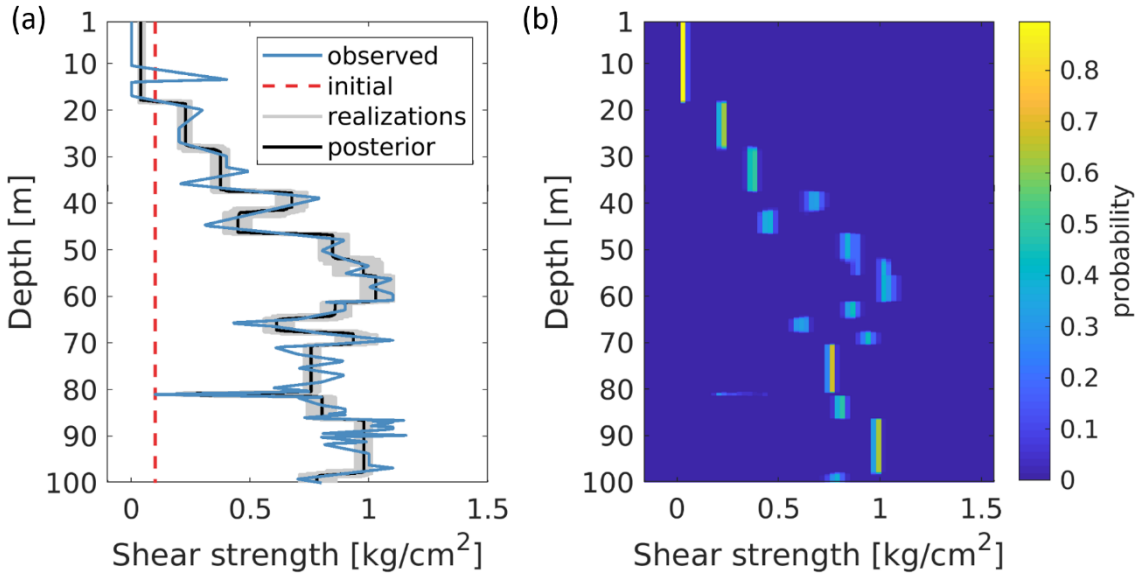

Figure 3.11. Estimated stratigraphic model with Bayesian prior in fine-scale

3.5. Discussion: Quantified Uncertainty

From three experiments with different Bayesian priors, we could estimate the stratigraphy of shallow soil layers in the Northeast Gulf of Mexico. Those calibrations started from the same observed data but showed a different number of layers, and those estimations in different dimensions are important to compare the data with other geophysical or geotechnical survey data at the same location. The integration of all available survey data will improve the quality of the offshore site characterization results with reduced uncertainty.

The vertical calibration with the rj-MCMC method was focused on random modeling in varying dimensions. However, this approach was also available to show the uncertainty with calculated probabilities. We can notice that if we assume less number of layers inside the geological borehole data, then the result shows high probability with less uncertainty. On the other hand, when we applied a bigger number of layers to the calibration modeling, some parts show less probability with a blurry image. For example, the soils between the depths of 50 to 70 meters in Figure 3.11-b shows less probability with a wide range of realization. That means the stratigraphy modeling at that location contains more possibility of uncertainty problem.

To make a more clear comparison of uncertainty, we compared the histogram of estimated soil properties at three-point locations (25, 50, and 75 meters depths) from all those three previous calibration modelings (Figure 3.12). All the samples show the normal distribution in the histogram, and the estimated mean value at three different depths are similar. However, we can notice the standard deviation of the histogram is

getting wider in our third fine-scale calibration, especially at a depth of 50 meters. Table 3.1 shows that the standard deviation at a depth of 50 meters (σ_2) is getting increased from 0.1354 to 0.2726 as increasing the modeling dimension. This result is the same as the estimation from the previous probability calculations, and indicate that the probabilistic Bayesian approach is essential for this stratigraphic modeling work to quantify the uncertainty of site characterization.

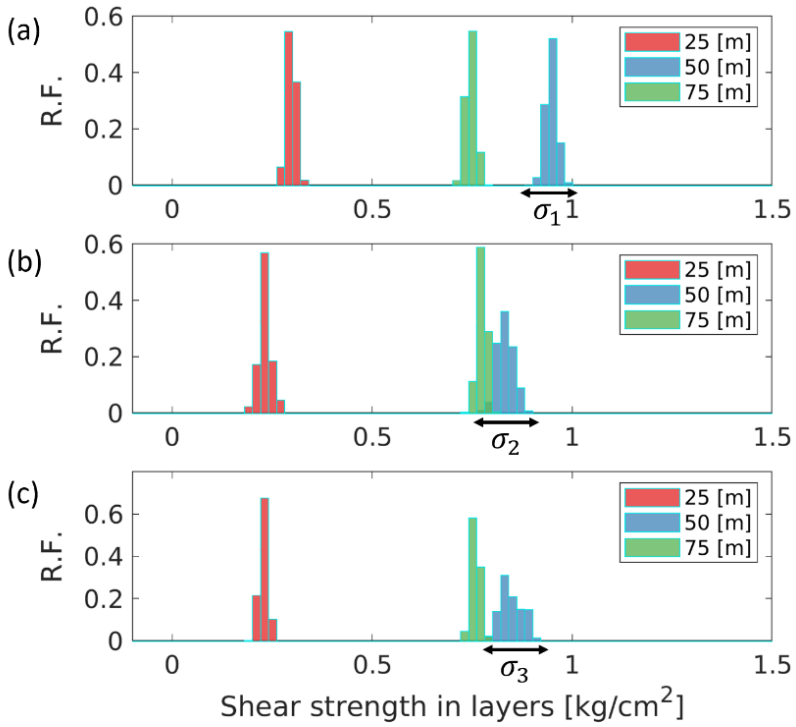

Figure 3.12. Histograms of estimated material properties; (a) dimension 6 in coarse-scale, (b) dimension 11, and (c) dimension 16 in fine-scale

Table 3.1. Uncertainty of the three different Bayesian calibration modeling

Std. / Depth	σ_1 in coarse-scale	σ_2	σ_3 in fine-scale
25 m	0.1144	0.1506	0.0993
50 m	0.1354	0.2213	0.2726
75 m	0.1225	0.1128	0.1071

3.6. Conclusion

We applied the probabilistic Bayesian calibration approach with a trans-dimensional rj-MCMC method to estimate the depth of soil layer interfaces with material properties. This stratigraphic modeling analysis is applied to an offshore in-situ borehole data in the northeast Gulf of Mexico, where it shows potential geohazard of the marine landslide. The observed raw soil data of shear strength profile was not appropriate for the geological interpretation because it suffers from the measurement errors in in-situ soil samples. Our Bayesian calibration with the spatial average method classifies to define the soil layers and quantifies the uncertainty in this soil data. Furthermore, we introduced how to apply Bayesian prior information to control the expected dimension of stochastic modeling to compare the geological borehole data with other geophysical or geotechnical surveys at the same location. This stratigraphic modeling is the fundamental step for the offshore survey data integration, which is important to figure out accurate subsurface information from the limited number of survey data. The estimation of more accurate stratigraphy under the seafloor should be considered for geomechanical slope stability analysis to prevent man-induced marine landslide, tsunami, or ocean pollutions.

4. PROBABILISTIC OFFSHORE SEISMIC INVERSION

4.1. Summary

Geophysical seismic surveys have been widely used for offshore site characterization, and this non-destructive method generates a vertical image of the subsurface geological information under the marine slope, where drilling boreholes may trigger a man-induced landslide. This offshore survey analysis from the seismic data is essential for the geohazard evaluation, but the geophysical data cannot show the depth and material properties because it is reflected wave signals in the time domain. Reconstruction for the information in the space-domain requires the seismic inversion, a method to convert those signals to the material properties in depth by using iterative numerical modeling. In this paper, we introduce a Bayesian seismic inversion as an integrated offshore site characterization method to apply geophysical seismic survey data for geotechnical slope stability analysis. This inversion is based on the reversible jump Markov chain Monte Carlo method, which defines the number of geological strata as a random variable. The forward modeling in this method is the geophysical seismic convolution, whose modeling parameters are the depth of layer interfaces, bulk density and P-wave velocity of the soil and seawater. We applied this method to an offshore multichannel seismic data, obtained near the steep marine slope, Sigsbee Escarpment, close to the deepwater oil fields in the Gulf of Mexico. The image of the ground model result shows the heterogeneous subsurface soil material properties in-depth under the marine slope. This study shows the approach of how to use geophysical survey data to

estimate the soil properties and quantified uncertainty to support the geotechnical slope stability analysis.

4.2. Introduction

Submarine slope stability analysis has been studied for many offshore geotechnical projects to prevent man-induced triggering geohazard disasters. When we install the pipelines and geotechnical foundations under the seawater, we should consider the slope stability condition of the sensitive marine sediments. Because the debris flow, especially the submarine mass movement, may threat various human activities in the offshore environment, appropriate evaluation of geohazard should be developed and implemented (Locat & Lee, 2002). The installation of the offshore geotechnical foundation for Mad Dog spar on a steep slope in the Gulf of Mexico as an example to show the importance of submarine slope stability analysis for the offshore project (Liedtke et al., 2006). Geohazard risk assessment for this project has been discussed in a practical perspective (Jeanjean et al., 2005), and the state of arts in offshore geohazard investigations have been studied for the project (Kvalstad, 2007). The investigations are important for the site characterization; however, geotechnical methods to measure the in-situ soil properties take too much cost and time. Furthermore, these direct investigation approaches can obtain the soil information in only a few meters in shallow depth, and this short range cannot be enough for the slope stability analysis in deep soils.

Offshore geophysical surveys have been applied to support the better offshore site characterization, and the survey data have been integrated with geotechnical investigation data (Jeanjean et al., 2006; Medina-Cetina & Esmailzadeh, 2014; Medina-Cetina et al., 2013; Vanneste et al., 2018; Vanneste et al., 2015). Multi-beam sonar, which can show the high-resolution bathymetry, has been used for marine slope stability estimation (Locat et al., 1999). Since this method was not able to penetrate the seafloor to show the vertical image of the offshore ground model, high-resolution Autonomous Underwater Vehicle (AUV) survey was applied to a marine steep slope area, which is called as Sigsbee Escarpment, in the Gulf of Mexico (George et al., 2002). However, penetration depth was shallow due to lots of attenuation from high-frequency source, so another geophysical survey; multichannel seismic data was used for the site characterization on the slope area again (Jeanjean et al., 2006). Since this seismic survey is based on the acoustic impedance; a geophysical parameter which comes from the multiplication between acoustic wave velocity and bulk density properties, an empirical relationship between geophysical and geotechnical properties have been introduced to utilize the geophysical survey results for offshore projects (Brand et al., 2003). Even though the acquisition system of the geophysical seismic survey has been developed to show higher resolution for better geohazard evaluation (Brookshire Jr et al., 2015), the observed seismic data are reflected wave signals in only time-domain instead of space-domain. This time-domain data cannot be used for the site characterization, and we need to convert those data from time- to space-domain.

The conversion of geophysical seismic data is called seismic inversion, and many deterministic and probabilistic inverse approaches have been developed to show the ground model in space-domain. The Spectral Analysis of Surface Waves (SASW) (Stokoe et al., 1994) is the most popular application of seismic inversion to use geophysical survey data for geotechnical site characterization. This method measures the dispersion of surface waves of the land seismic survey data and converts the observed signals from time- to frequency-domain. Since the surface waves are related to the elastic modulus of the soil media, the chosen dominant signals in frequency-domain are used to estimate the shear modulus property of the onshore ground model. A theoretical approach based on the SASW method was also applied to a synthetic offshore case study and showed a potential possibility for offshore site characterization (B. Lee et al., 1997). However, the data acquisition system was different from conventional offshore seismic surveys, and this approach was difficult to apply for a field case study.

Another geophysical method, which is called Full-Waveform Inversion (FWI), has been developed and introduced for shallow site characterization. This FWI method uses not only the surface waves but also the acoustic and elastic body waves inside the ground model. All of the refraction, reflection, dispersion, and scattering phenomena inside the ground model are used as observed data to predict the unknown soil properties. Recently, the deterministic approach of FWI was compared with SASW on a land site characterization and showed an improved subsurface image with the estimated elastic soil properties (Kallivokas et al., 2013). Many research based on this approach has been studied for onshore case studies (Nguyen, Tran, & McVay, 2016; Tran &

McVay, 2012; Tran, McVay, Faraone, & Horhota, 2013), but offshore case study for geotechnical analysis is not fully described yet. Medina-Cetina et al. (2013) started to introduce a Bayesian geophysical inverse modeling to use the geophysical survey data for the offshore site characterization, and this probabilistic approach has been developed for more accurate stratigraphy estimation under the seafloor (Medina-Cetina et al., 2019)

In this paper, we applied a Bayesian seismic inverse method with the reversible jumping Markov chain Monte Carlo method (rj-MCMC) to multichannel seismic survey field data for offshore site characterization. This trans-dimensional rj-MCMC approach was introduced for a theoretical offshore case study with synthetic electromagnetic and acoustic surveys data (S. Esmailzadeh, Medina-Cetina, Kang, & Kallivokas, 2015), and the estimated ground model with quantified uncertainty showed the potential possibility of the Bayesian inversion method to support the geotechnical site characterization from geophysical surveys. We updated this inversion with another forward model, geophysical seismic convolution, to develop a practical approach and applied this inversion method for the offshore seismic data near the Sigsbee Escarpment in the Gulf of Mexico. Spectral filtering in the frequency-domain is applied as preprocessing for the noisy field seismic data, and the estimated ground model from the Bayesian inversion indicated the locations of dense and soft sediments under the seafloor in the space-domain. Furthermore, the uncertainty of the estimated ground model is quantified and supports to notice the geohazard prone area. This reconstructed spatial ground model information will support the geomechanical slope stability analysis as accurate initial and boundary conditions.

4.3. Methodology

4.3.1. Geophysical seismic convolution method

Since the stochastic inverse modeling requires plenty of sampling from iterative computation, we decided to use a mathematical operator, the geophysical seismic convolution method, as our forward model. Seismic convolution has been widely used for geophysical studies and has many advantages (Zhou, 2014). This convolution operator implements the seismic modeling in fast, and it is appropriate for the iterative inverse modeling. Furthermore, the calculation of reflected signals based on seismic impedance parameters is accurate enough to compare with field data. However, seismic convolution also has some limitations in comparison with numerical simulations, such as Finite-Element method or Finite-Difference Method (Yilmaz, 2001). Since the convolution operator shows the only 1-dimensional seismic trace, the result cannot consider the energy loss from the realistic 2-or 3-dimensional geometrical spreading. This operator is only appropriate for the elastic case study and not able to consider inelastic attenuation, which comes from the seismic energy absorption inside the soil media itself. Furthermore, this theoretical operator cannot generate multiples, which is trapped inside layers and reflected back with small amount of amplitude. Consequently, this seismic convolution is not appropriate for the SASW or FWI methods in land seismic surveys.

However, offshore seismic survey data is different from land seismic data, and the convolution method has been widely used as the best approach for the seismic inversion. Conventional offshore seismic data is post-stack data, which means processed

data based on the acquisition system with multi-channel streamers. The multiple seismic traces from multiple hydrophones are stacked and become a 1-dimensional seismic trace to increase the signal over noise ratio. Another processing technique amplifies the amplitude of the processed seismic traces along with the depth to make a clear vision of the signals in deep layers. Furthermore, most of the identified multiples are removed for better visualization, and the final processed offshore seismic data can show the existence of subsurface geological layers. In short, seismic convolution is appropriate for the seismic inversion with offshore seismic data, because the offshore data already processed and adjusted signals beyond the physical phenomena.

Even though the seismic convolution method is fast and accurate, the application for this forward model is not simple, so we explained the detail process of it with a diagram (Figure 4.1). From the ground model, we need to get geological information in vertical cross-section and define layers with density (ρ) and P-wave velocity (Vp). The acoustic impedance (Imp) can be calculated from them, and the only locations of the layer interfaces will get non-zero values of reflection coefficients (RC_z) in depth.

$$RC_z = \frac{(\rho_2 Vp_2 - \rho_1 Vp_1)}{(\rho_2 Vp_2 + \rho_1 Vp_1)} \tag{4.1}$$

The subscription 2 and 1 for the calculation of reflection coefficients indicate geological lower and upper layers, which define the layer interfaces. Since we know the depths and P-wave velocities at the layer interfaces, the two-way travel time (TWTT) can be calculated in time-domain. We need to convert the reflection coefficients from space- to time-domain and make the reflectivity function in time (g_t). The seismic wavelet (f_t) in time, which comes from the seismic source generation device is applied to the

reflectivity function, and the convolution operator generates the seismic trace data (s_t) from them.

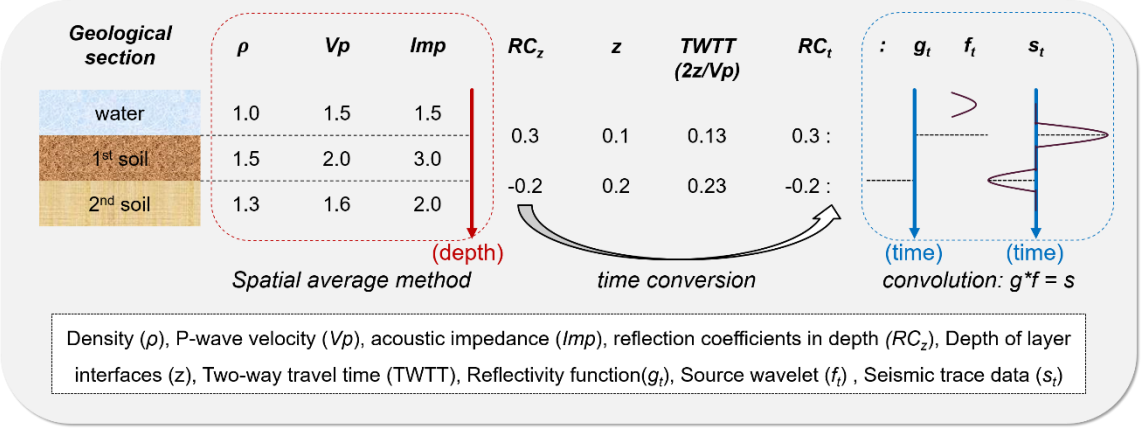

Figure 4.1. Diagram to describe seismic convolution modeling

4.3.2. Geophysical non-uniqueness problem

The most challenging problem for the spatial inverse problem is related to the non-uniqueness of the solution from the inverse modeling. This drawback affects most of the geophysical inverse problems which are often ill-posed from the mathematical point of view (Foti, Comina, Boiero, & Socco, 2009), and it is not possible to prove the uniqueness when we use the seismic convolution method for the inverse problem. A diagram (Figure 4.2) explains the difference between forward and inverse model for the spatial analysis to describe the specific meaning of this non-uniqueness problem. The numerical seismic forward model generates only one reflected wave signals from one input soil ground model. This result is a unique value, so many times of duplicated forward modeling with the same ground model will generate the same seismic data. However, the problem is that some of the other different ground models also can

generate the same seismic data. Thus, inverse modeling from the seismic data as observed data may show us many different ground models as the candidates from one seismic data, and this is called geophysical non-uniqueness problem. Finding the one true target ground model from the observed survey data is challenging, and requires a probabilistic approach to compare all the candidate ground models for the site characterization.

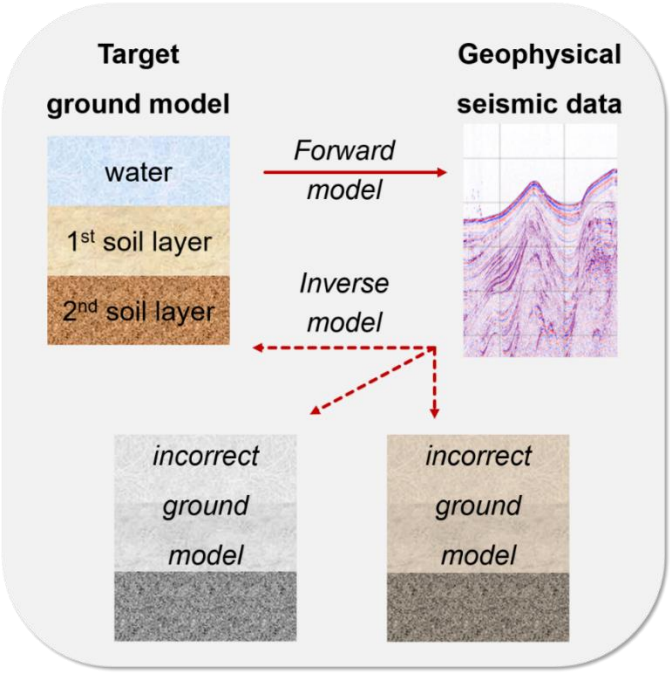

Figure 4.2. Diagram to explain the possibility of incorrect estimation from the spatial inverse model

We tested four ground models with seismic forward modeling and compared the generated seismic signals to clarify the problem of seismic non-uniqueness for the geophysical inverse modeling (Figure 4.3). Since our objective is to focus on offshore shallow sediments, the four different ground models have all the seawater layer at the top and have two model parameters; soil density (Figure 4.3-a) and acoustic P-wave

velocity (Figure 4.3-b). Those ground models have two soil layers under the water, but the material properties and the location depth of layer interfaces are slightly different. The detail information about the modeling parameters is described in Table 4.1. Even though the vertical soil profiles of those four ground models are all different, the generated geophysical seismic signals (Figure 4.3-c) are the same exactly. This result shows an example of the geophysical non-uniqueness problem in seismic inverse modeling and proves the importance of probabilistic inversion approach to consider every possible candidate ground models.

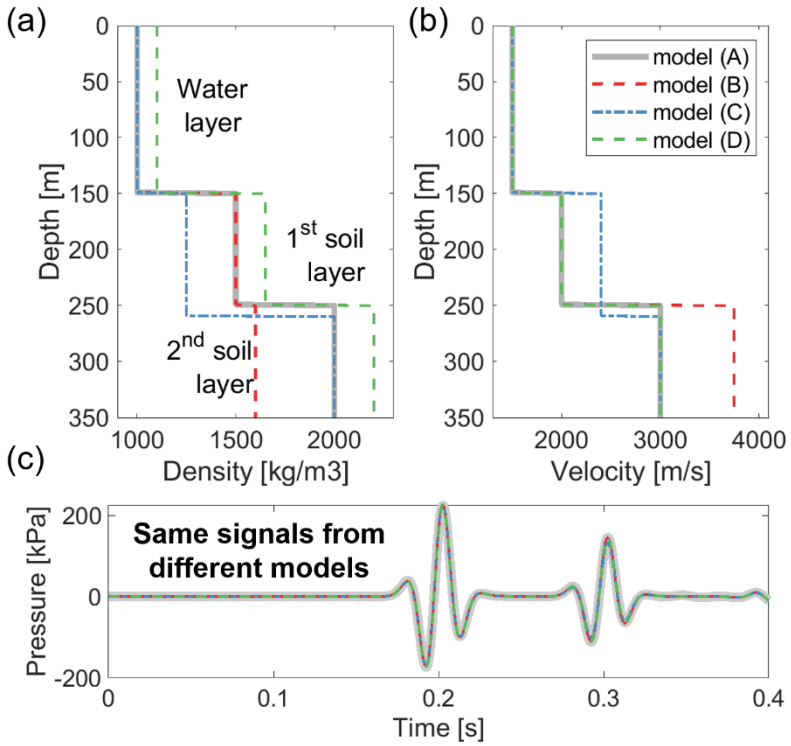

Figure 4.3. Example of the geophysical seismic non-uniqueness problem

Table 4.1. Properties of the four different soil ground models

ground models	Depths of interfaces [m]		Soil density [g/cc]			P-wave velocity [km/s]		
A	150	250	1.00	1.50	2.00	1.50	2.00	3.00
B	150	250	1.00	1.50	1.60	1.50	2.00	3.75
C	150	260	1.00	1.25	2.00	1.50	2.40	3.00
D	150	250	1.10	1.65	2.20	1.50	2.00	3.00

4.3.3. Trans-dimensional Markov chain Monte Carlo method

The probabilistic approach for the geophysical inverse problem requires trans-dimensional Markov chain Monte Carlo (MCMC) method. Because the dimension of the stochastic process, which represents the number of subsurface geological layers, is unknown and has to be another random variable inside the inversion. Medina-Cetina et al. (2013) introduced the theoretical probabilistic approach for the offshore site characterization, and S Esmailzadeh et al. (2015a) explained the details of MCMC approach in varying dimensions, which is called as the reversible jump Markov chain Monte Carlo (rj-MCMC) method. Based on those previous methods, we applied the seismic convolution method as new forward modeling to utilize the conventional offshore seismic survey data with the modeling parameters ($\boldsymbol{\theta}$); depth of layer interfaces (z_i), bulk density (ρ_i) and P-wave velocity (Vp_i). The observed data (d_{obs}) for this inversion is seismic traces from offshore post-stack seismic data, and the likelihood function is defined with the observed data (d_{obs}), seismic forward model ($g(\boldsymbol{\theta})$). The size of the observed data (n) and covariance matrix (C_d) is also defined from the seismic data.

$$f(\boldsymbol{d_{obs}}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}}|C_d|^{\frac{1}{2}}} \times$$

$$\exp\left[-\tfrac{1}{2}\,(g(\boldsymbol{\theta}) - \boldsymbol{d_{obs}})^T (C_d)^{-1}(g(\boldsymbol{\theta}) - \boldsymbol{d_{obs}})\right] \tag{4.2}$$

The proposal model selection in varying dimension is important for the likelihood function with forward modeling. The method to generate candidates from the proposal model, which were introduced for geophysical inversion (S. Esmailzadeh et al., 2015), follows four different types of model selection during the stochastic process; BIRTH, DEATH, MOVE and PERTURB, and each of them has different proposal distribution based on the properties of modeling parameters. We described this method with the extra explanation to clarify the model selection process in detail.

- BIRTH: This creates a new subsurface layer with a new depth of layer interface (z^*) and new material properties (β^*). This new layer is located under a randomly chosen layer depth (z_{idx}).

$$z^* \sim unif(z_{idx}, z_{idx+1}) \tag{4.3.a}$$

$$\beta^* \sim N(\beta_{idx}, v * \sigma) \tag{4.3.b}$$

- DEATH: This removes one of the current subsurface layers. The randomly chosen depth (z_{idx}) is removed, and the material properties at this depth are defined as the average values between the upper and lower layers.
- MOVE: This does not change the dimension of the modeling, and only changes the location of one current layer. One of the current layer depth is randomly chosen, and the new location (z^*) is randomly located between upper (z_{up}) and lower (z_{low}) layer interfaces.

$$z_i^* \sim unif(z_{up}, z_{low}) \tag{4.4}$$

- PERTURB: This does not change the dimension of the modeling, and only changes the material properties of one current layer. One of the current layer material properties (β_i) is randomly chosen, and the new material properties (β^*) is randomly defined with standard deviation (σ) and parameter (v).

$$\beta_i^* \sim N(\beta_i, v * \sigma) \tag{4.5}$$

Since this modeling selection changes the dimension of the modeling in a random process, we need to apply the generalized Metropolis-Hasting sampling algorithm with the acceptance probability (α). The calculation of this probability is based on likelihood, prior and proposal ratios from the current model parameters (θ) and proposed model parameters (θ^*).

$$\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = min\left\{1, \underbrace{\frac{f(\boldsymbol{D_{obs}}|\boldsymbol{\theta}^*)}{f(\boldsymbol{D_{obs}}|\boldsymbol{\theta})}}_{\text{likelihood ratio}} \times \underbrace{\frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta})}}_{\text{prior ratio}} \times \underbrace{\frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}}_{\text{proposal ratio}}\right\} \tag{4.6}$$

The proposal ratio is the key point in the calculation of acceptance probability (α) to control the varying dimension. This ratio should be considered in both forward and reverse movement in varying dimensions. When the rj-MCMC choose the MOVE, PERTURB for model selection, the probability in forward (from θ to θ^*) is equal to the probability in reverse (from θ^* to θ), so the dimension does not change at all. That means the probabilities of proposals are canceled out, and the proposal ratio becomes one to be ignored. However, BIRTH and DEATH types for model selection change their

modeling dimension, so each proposal ratio involves extra calculations. Since the locations of layers are proposed independently from the material properties, the calculation of the proposal ratio can be separated into two terms.

$$\frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} = \frac{q(z|\theta^*)}{q(z^*|\theta)} \times \frac{q(\beta|\theta^*)}{q(\beta^*|\theta)} \tag{4.7}$$

The calculation requires a total number of the discretized elements (N) for the first term, and the second term for the material property simply follows the normal distributions. The specific probabilities are defined in different forms for each BIRTH and DEATH types, and S. Esmailzadeh et al. (2015) explained the details.

- BIRTH: This changes the dimension from k layers to $k + 1$ layers.

$$q(z^*|\theta) = \frac{1}{N-k} \tag{4.8.a}$$

$$q(z|\theta^*) = \frac{1}{k+1} \tag{4.8.b}$$

- DEATH: This changes the dimension from k layers to $k - 1$ layers.

$$q(z^*|\theta) = \frac{1}{k} \tag{4.9.a}$$

$$q(z|\theta^*) = \frac{1}{N-(k-1)} \tag{4.9.b}$$

4.4. Experiments

4.4.1. Synthetic case study

To verify the accuracy of the stochastic seismic inversion, we designed a simple synthetic case study with a shallow offshore ground model. The ground model has five layers with the top seawater and four of soil layers within the depth of 80 meters below the seafloor (Figure 4.4-a). The soil density is getting increased from 1.5 to 2.0 g/cc, but

one layer at the depth around 50 meters shows unexpected soft layer (Figure 4.4-b). Each layer has a different value of soil density, but the P-wave velocities in the seawater and ground model are assumed as constant 1.5 and 2.0 km/s, respectively. This synthetic ground model design tries to catch the features of the offshore shallow sediments, which has a wide range of density and narrow range of P-wave velocity.

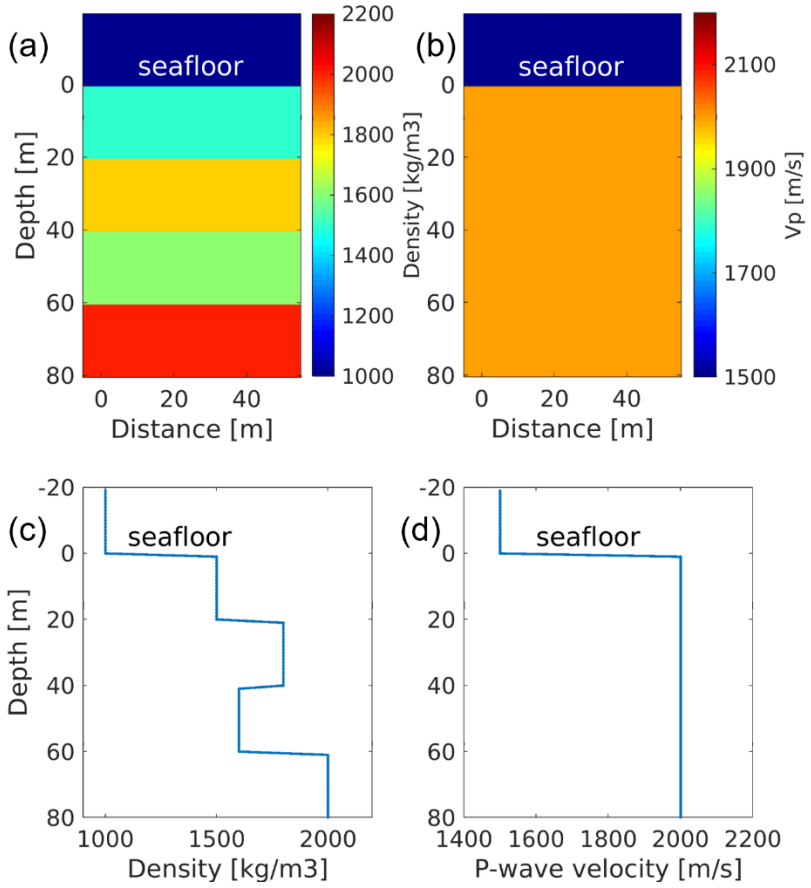

Figure 4.4. Material properties of the synthetic offshore ground model; (a) & (c) soil density, (b) & (d) P-wave velocity

The seismic wave source is located at the top surface of seawater, and four interfaces in the ground model generate the reflected seismic signals. When we apply the

geophysical survey with the multi-channel streamers at every 10 meters, the processed offshore seismic data show the signal traces in the vertical time-domain (Figure 4.5-a). Since the applied source frequency is assumed as low 20 Hz, widely used for the traditional air gun blaster source, the seismic interpretation can only figure out the location of the seafloor and the existence of geologic strata under the seafloor. Furthermore, the observed seismic signal data contains noise signals (Figure 4.5-b), so the direct estimation of ground model properties from the observed seismic data is not possible, and we need the seismic inversion.

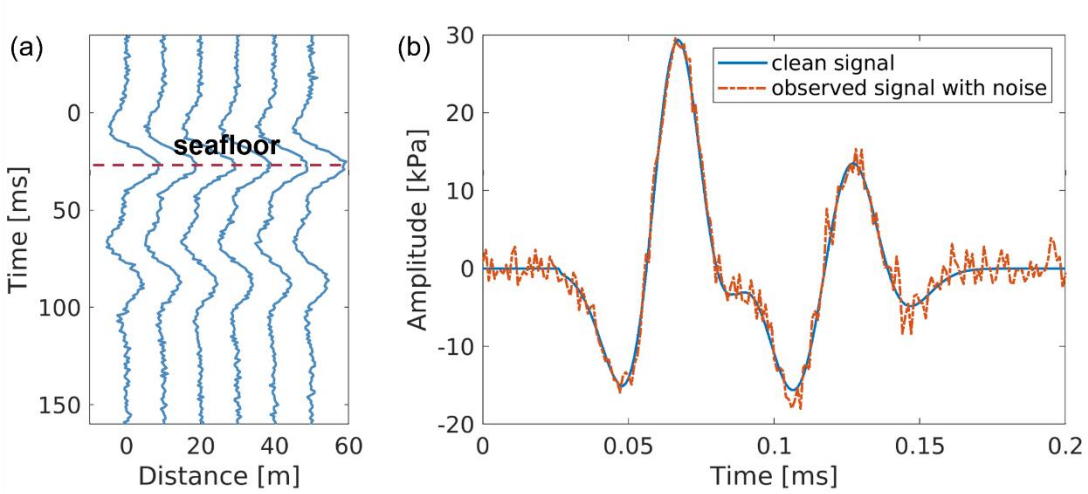

Figure 4.5. Geophysical seismic data; (a) seismic traces from multi-channel streamers, and (b) observed signal with and without noise

We applied the seismic convolution and hierarchical Bayesian model for the stochastic seismic inversion with the synthetic observed data. The Bayesian model based on rj-MCMC method implements the inverse modeling in varying dimensions, so the predicted number of geological layers inside the ground model is a random variable. The stochastic process (Figure 4.6-a) shows the convergence of this random dimension

during the iterative samplings, and histogram (Figure 4.6-b) also verifies the predicted number of layers is five; one for seawater and four for soils in the ground model. Thus, we only collected the samples in five dimensions and discarded other samples from the total 800,000 accepted samples.

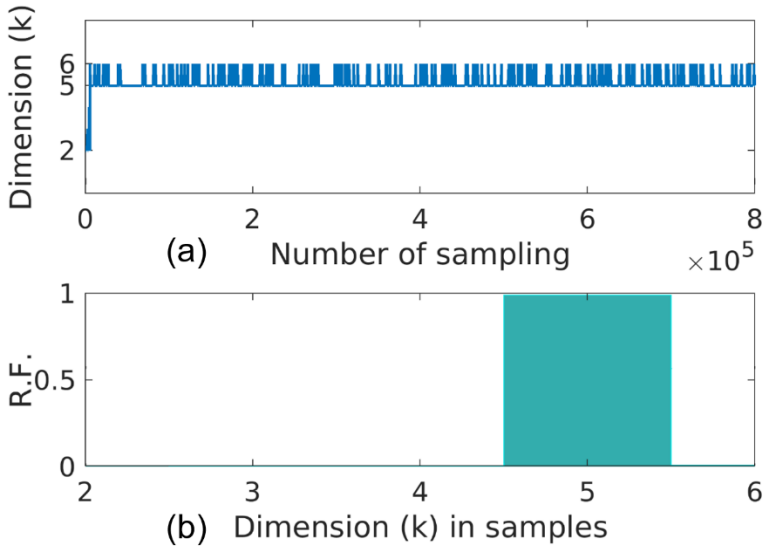

Figure 4.6. Stochastic process in varying dimensions; (a) plot to show the convergence, and (b) histogram

Figure 4.7-a shows the random process of the chosen five dimension samples in a surface with the depth and number of samples in the vertical and horizontal axis, respectively. The cumulative means and standard deviation of the random variables in the ground model, which means the predicted properties of each soil layer, show the convergence and stationary condition. (Figure 4.7-b). We decided the 400,000 as the burn-in point to use the only stationery samples for the posterior distribution to estimate the final value of the ground model. The stochastic process also shows that predicted soil

profiles during the iterative inversion is getting closer to the unknown target values (Figure 4.8).

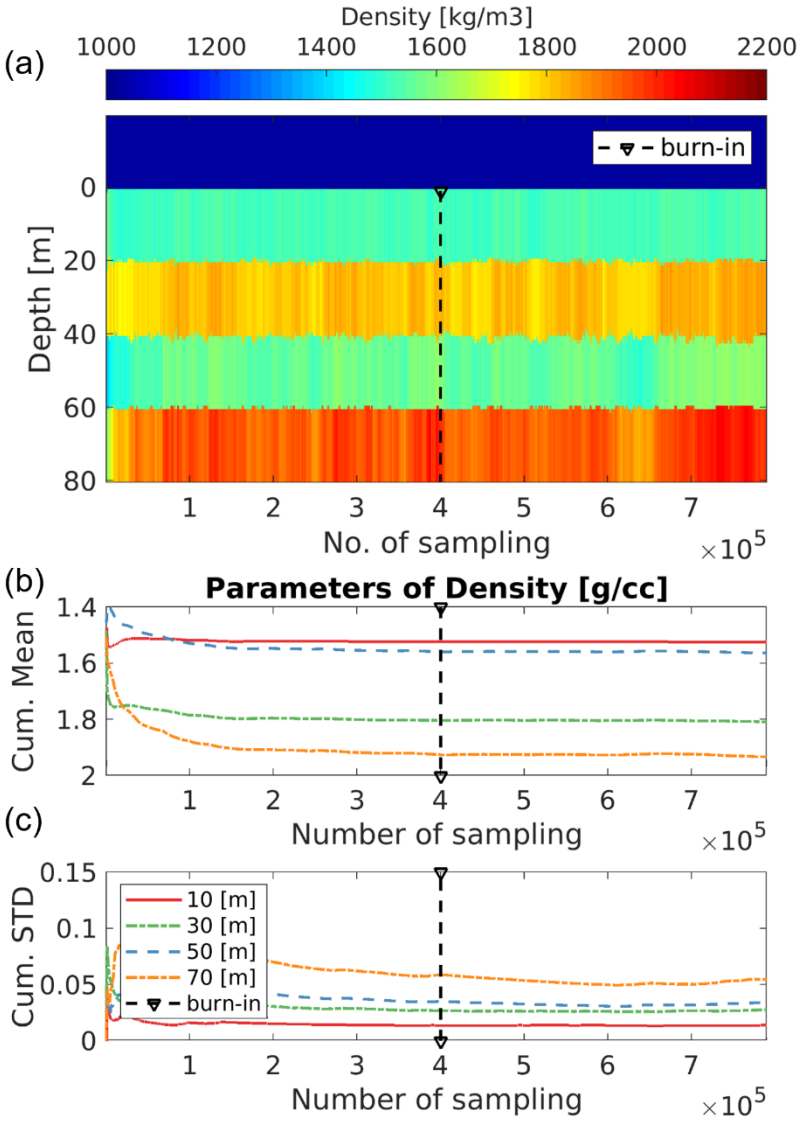

Figure 4.7. Stationary condition after the burn-in point; (a) surface of random variables in-depth, and (b) stationary condition after the burn-in point

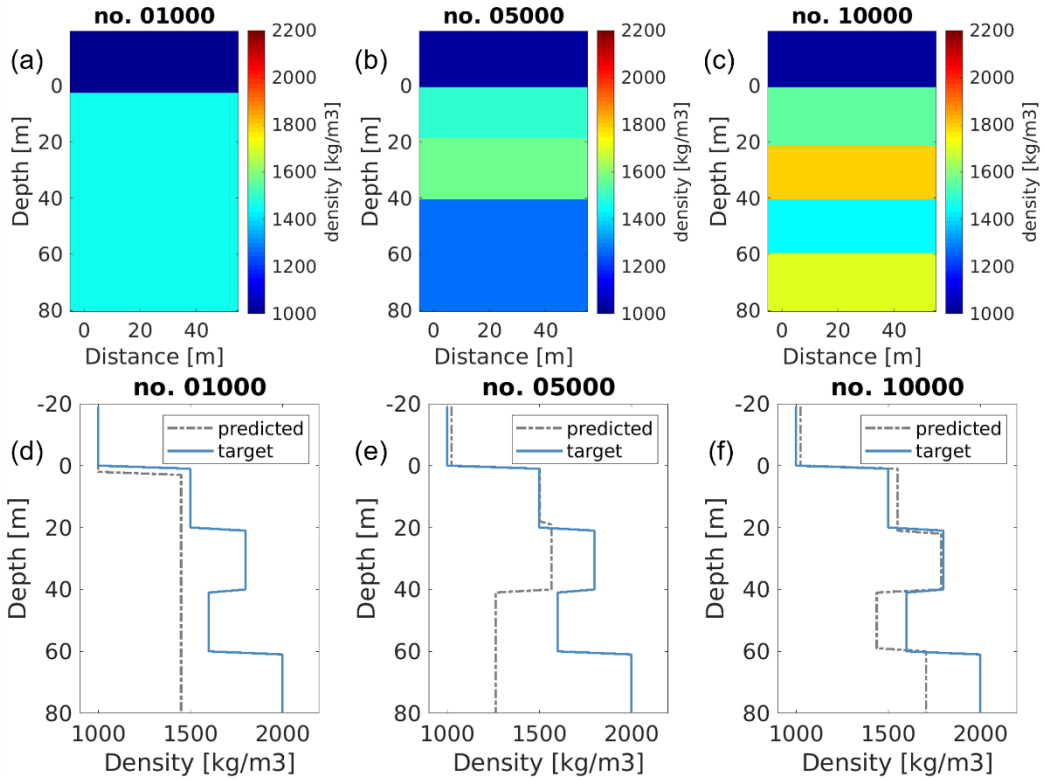

Figure 4.8. Predicted soil profiles from accepted candidates during the stochastic process. (a)(b)(c) images of the ground models, and (d)(e)(f) vertical profiles

The final estimation from the posterior distribution of the samples is accurate regardless of the initial condition of the stochastic seismic inversion was incorrect homogeneous soil. The initial seismic data and initial vertical soil profiles, the red dashed lines in Figure 4.9-a & -b, are different from the observed data. However, the realizations in the time-domain from the chosen samples shows an accurate match to the observed seismic data regardless of the signal noise (Figure 4.9-a). The final estimation of the vertical soil profile, the black line in Figure 4.9-b, is the mean of realizations from the samples after the burn-in point, and it also matches well to the target ground model.

Figure 4.9-c indicates the probability of the estimated results, and this shows the uncertainty is getting increased along with the depth.

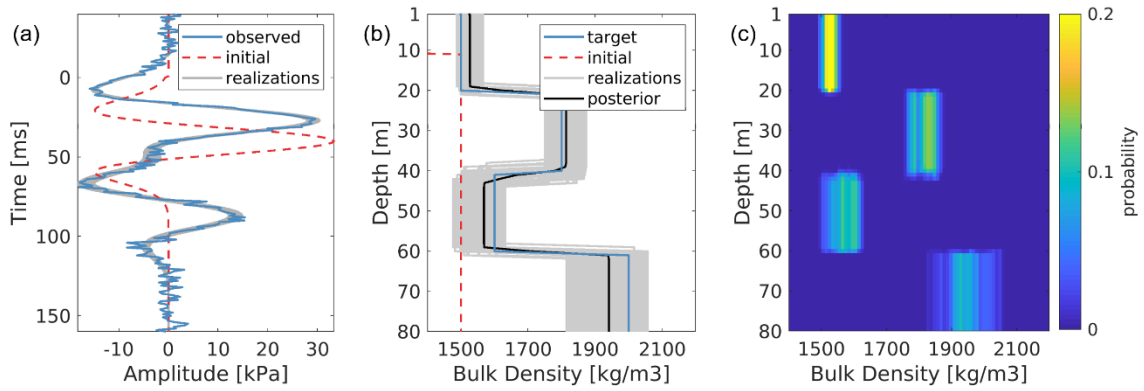

Figure 4.9. Results; (a) realizations in the time-domain, (b) realizations in-depth, and (c) probability in the vertical profile

To compare the uncertainty from each soil layer, we plot the histograms of the posterior distribution of the samples (Figure 4.10). The black vertical lines indicate the unknown true values of the material properties, and the histogram shows normal distribution around the true values. Since the bottom layer in this synthetic ground model cannot calculate the seismic reflection coefficient, the histogram at 70 meters depth shows the widest range of standard deviation. Even though the results contain high uncertainty at the bottom layer of the ground model, this stochastic seismic inversion method can find the accurate soil material properties (Figure 4.10) and depth locations of geological layers (Figure 4.11) from the observed seismic data itself.

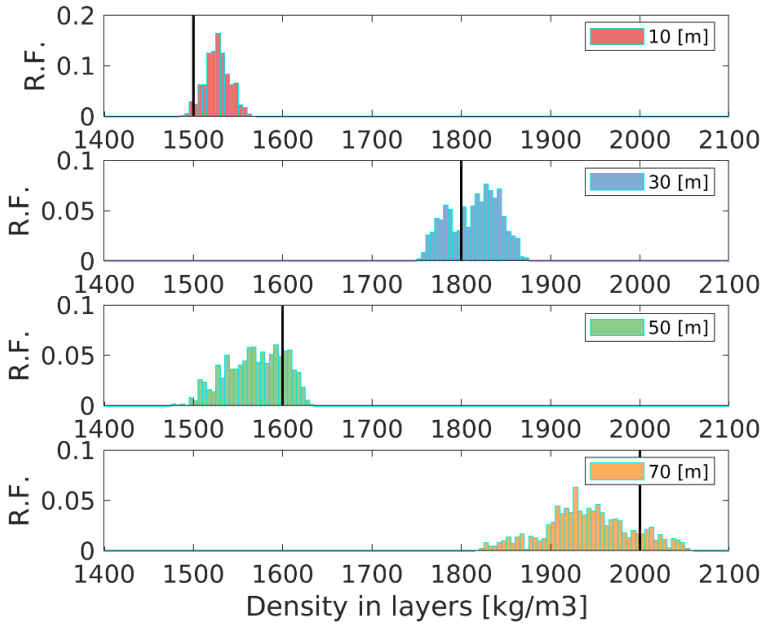

Figure 4.10. Histogram to show the distributions of density parameters in four different soil layers

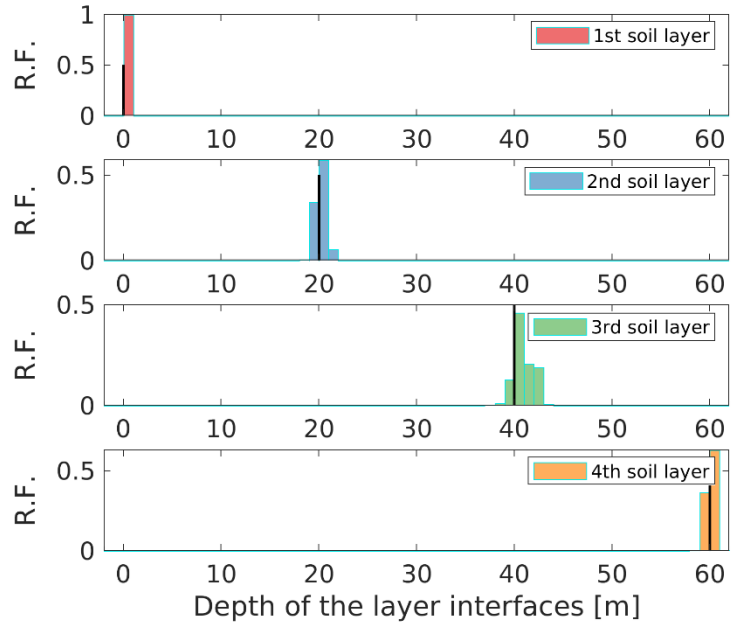

Figure 4.11. Histogram to show the distributions of depth parameters in four different soil layers

4.4.2. Field case study

From the synthetic case study, we verified the stochastic seismic inversion method could find out the subsurface location of geological layers and the soil material properties inside the ground model. We applied this approach to an offshore field seismic survey data from the Gulf of Mexico. We chose the marine steep slope area, called Sigsbee escarpment, near offshore platforms on the deepwater reservoir fields (Figure 4.12). Because the suction pile installation on this area has been issued since the installation of the Mad Dog spar (Berger et al., 2006; Jeanjean et al., 2006; Liedtke et al., 2006). The geophysical seismic survey data is obtained from the Marine Geoscience Data System (MGDS), and Figure 4.13 shows the vertical image of geological stratigraphy under the seafloor as the result of this geophysical survey. We chose only the location of the steep slope data inside the black box on the left side of the figure and first applied the spectral analysis in the frequency-domain.

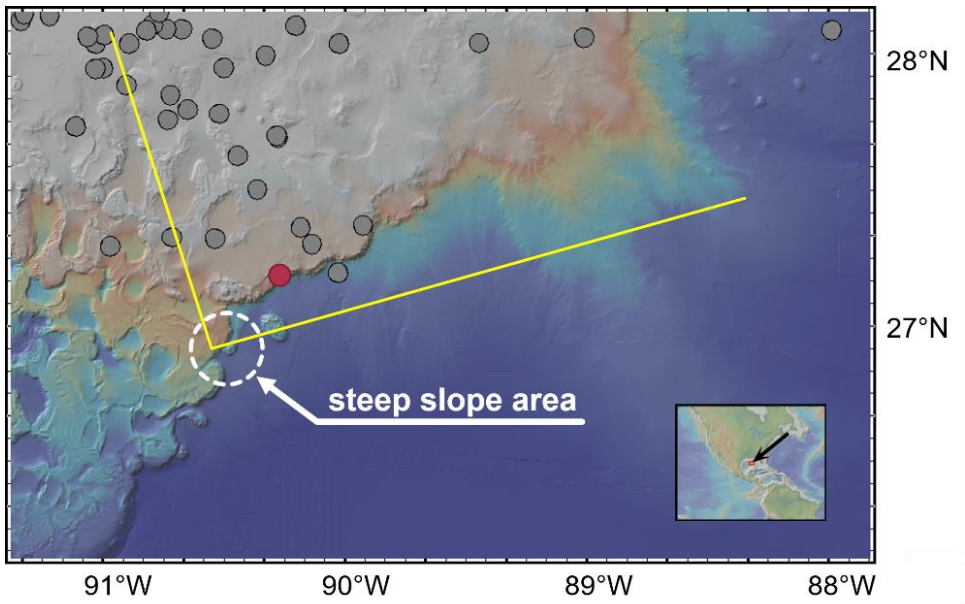

Figure 4.12. Locations of platforms (grey) and the path of the geophysical seismic survey (yellow) near the Mad dog platform (red) in the Gulf of Mexico

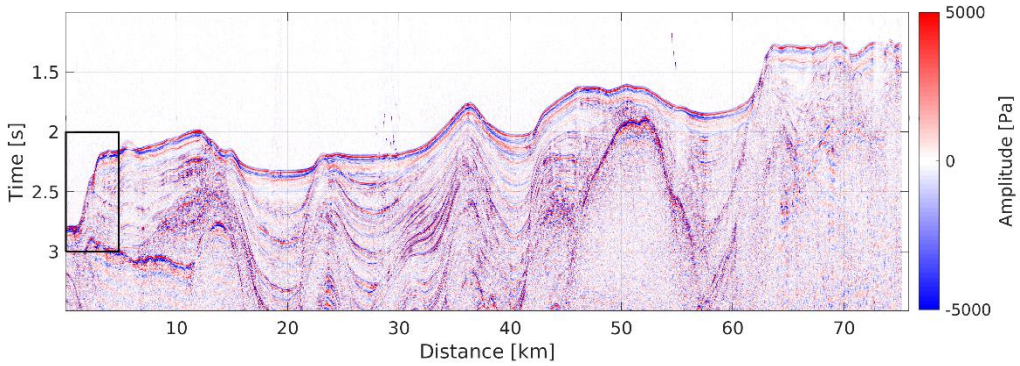

Figure 4.13. Observed seismic data to show the image of geological stratigraphy under the seafloor

Figure 4.14 shows a single seismic trace signal from the chosen seismic data. Even though the seismic data is processed signals, this raw data (Figure 4.14-a) still contains lots of noise in high-frequency (Figure 4.14-c). We applied the bandpass filtering (Menke & Menke, 2016) to focus on the frequency range from 10 to 40 Hz and

removed the noise signals in over this range (Figure 4.14-c). This spectral filtered data (Figure 4.14-b) becomes smoother than raw data and shows clear variations in the observed signals. We applied this spectral filtering process to all seismic traces in the steep slope area, so the previous vertical image of the ground model under the slope (Figure 4.15-a) became more clear to show the existence of geological layers under the slope (Figure 4.15-b).

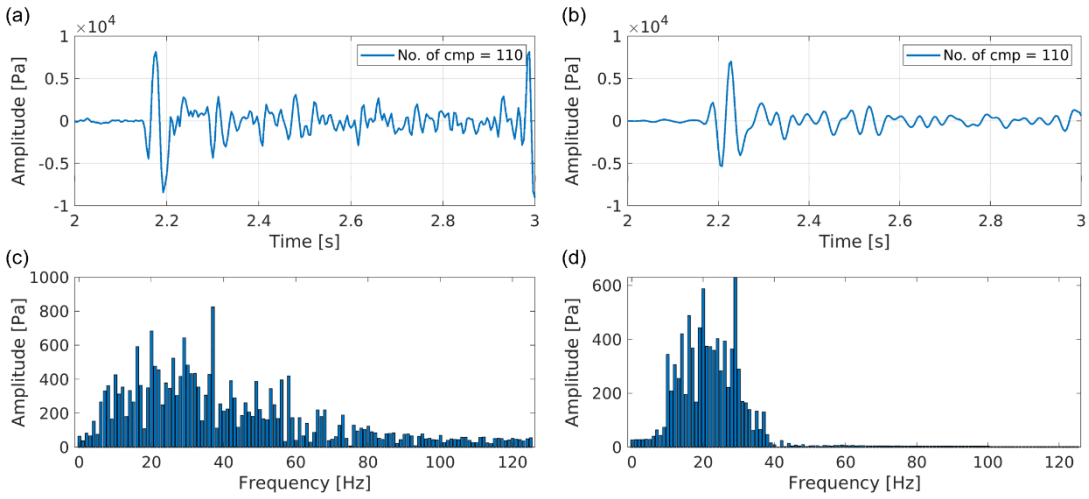

Figure 4.14. Fourier transform and spectral filtering; (a) & (c) raw data, (b) & (d) in time- and frequency-domain, respectively

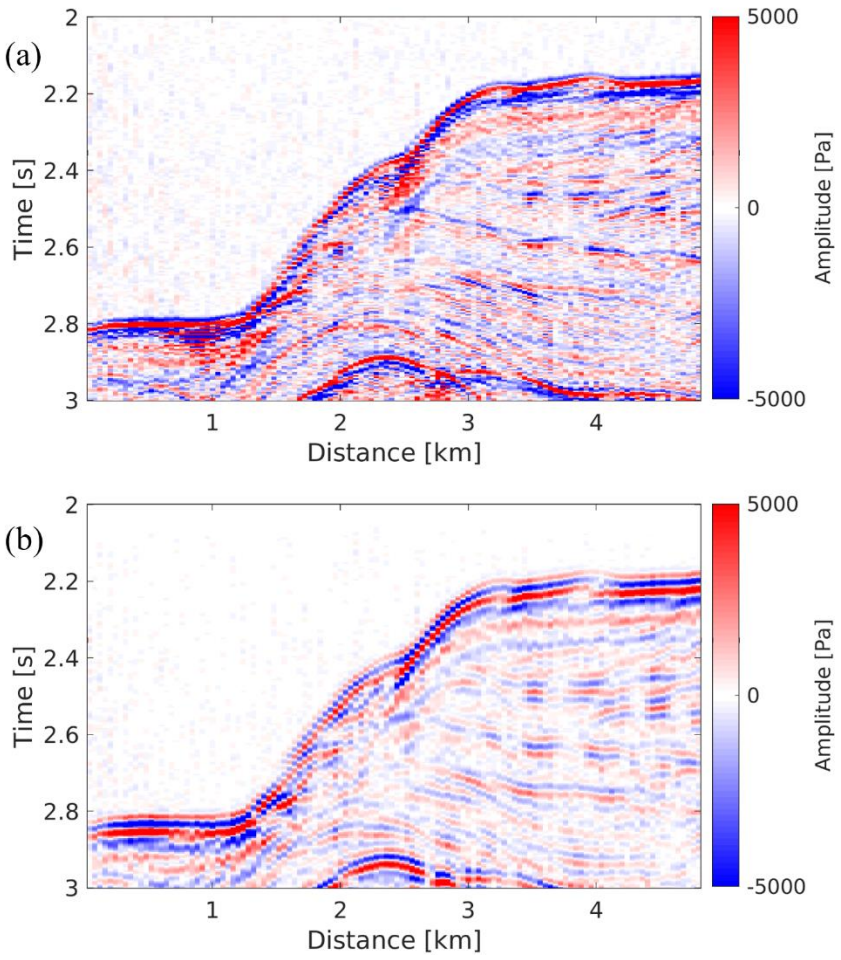

Figure 4.15. Images of the steep marine slope in seismic data; (a) raw data before spectral filtering, and (b) processed data after spectral filtering

Before we apply the stochastic seismic inversion to convert the seismic data from time to space-domain, we need to decide the wavelet for the seismic forward modeling. We chose the Ricker wavelet, which shows the shape of a Mexican hat (Figure 4.16-a) in time-domain. The amplitude (A) of the Ricker wavelet is computed

$$A = (1 - 2\pi^2 f^2 \boldsymbol{t}^2) \exp(-\pi^2 f^2 \boldsymbol{t}^2) \tag{4.10}$$

where f and $\boldsymbol{t}$ indicate the value of dominant frequency and time vector, respectively. The dominant frequency of this theoretical source wavelet function is 20 Hz, and the range of the frequency in this wavelet matches well to the previous spectral filtered data (Figure 4.16-b). Consequently, our spectral filtered seismic data is ready to apply for the stochastic seismic inversion with seismic convolution method.

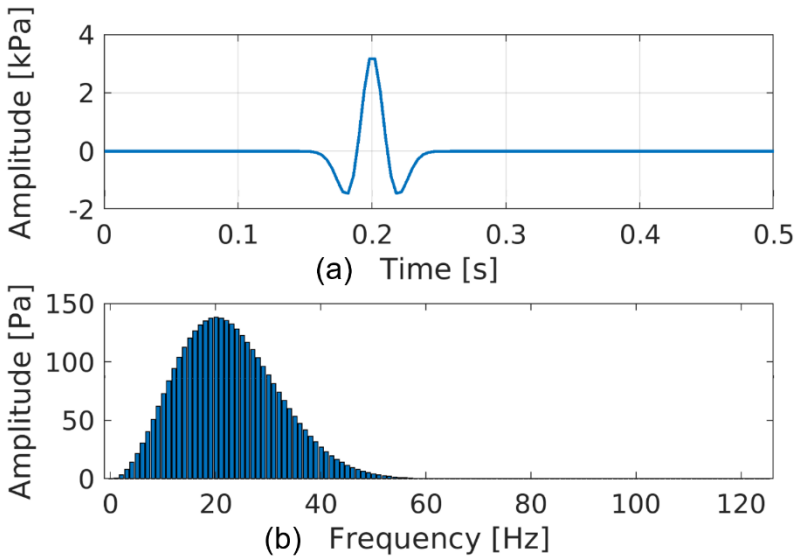

Figure 4.16. Ricker wavelet for seismic forward modeling in time-domain (a) and frequency-domain (b)

The chosen seismic data (Figure 4.15) had 120 vertical traces, and we applied the stochastic seismic inversion with rj-MCMC method to every signal independently. We took 400,000 samples and decided the 250,000 as out burn-in point (Figure 4.17). For every independent stochastic inverse modeling, all the parameters were the same, but one of the important setup was that the depth of the seafloor was controlled to follow the slope and adjust the initial predicted seismic signal to the observed signal. Otherwise, the

inversion may decide an incorrect location of the seafloor and converges to incorrect soil material properties in every layer under the seafloor.

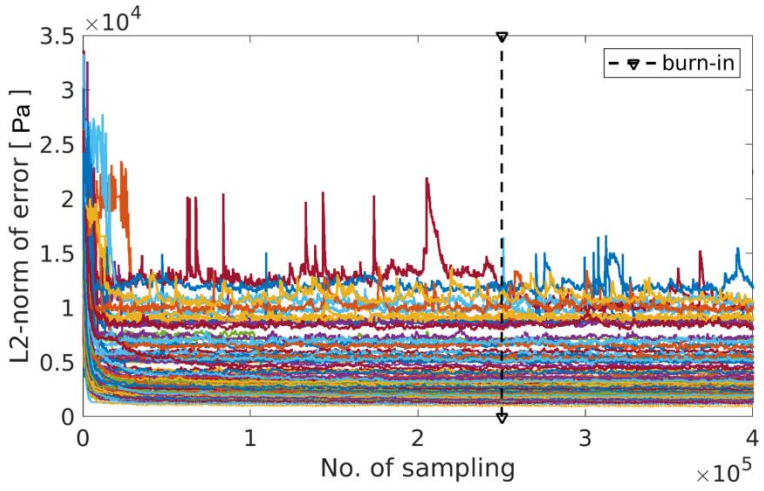

Figure 4.17. Stochastic sampling process of the seismic traces

The final results of the inversion are estimated from the samples after the burn-in point. We calculated the mean (Figure 4.18-a) and standard deviation (Figure 4.18-b) of the samples to estimate the vertical image of the ground model under the marine slope and quantified uncertainty, respectively. From the mean of posterior (Figure 4.18-a), we can figure out the location of dense soils (red) and soft soils (green) under the seawater (blue). Since we roughly assumed the property of P-wave velocity in the ground model as a constant value 2,000 m/s, it is difficult to say that the estimation results would match to the in-situ material properties in this area. However, these independent vertical inversions show continuous horizontal layers under the seafloor, and this approach has a potential possibility to improve our offshore site characterization if we apply more realistic values for the P-wave velocities in the ground model.

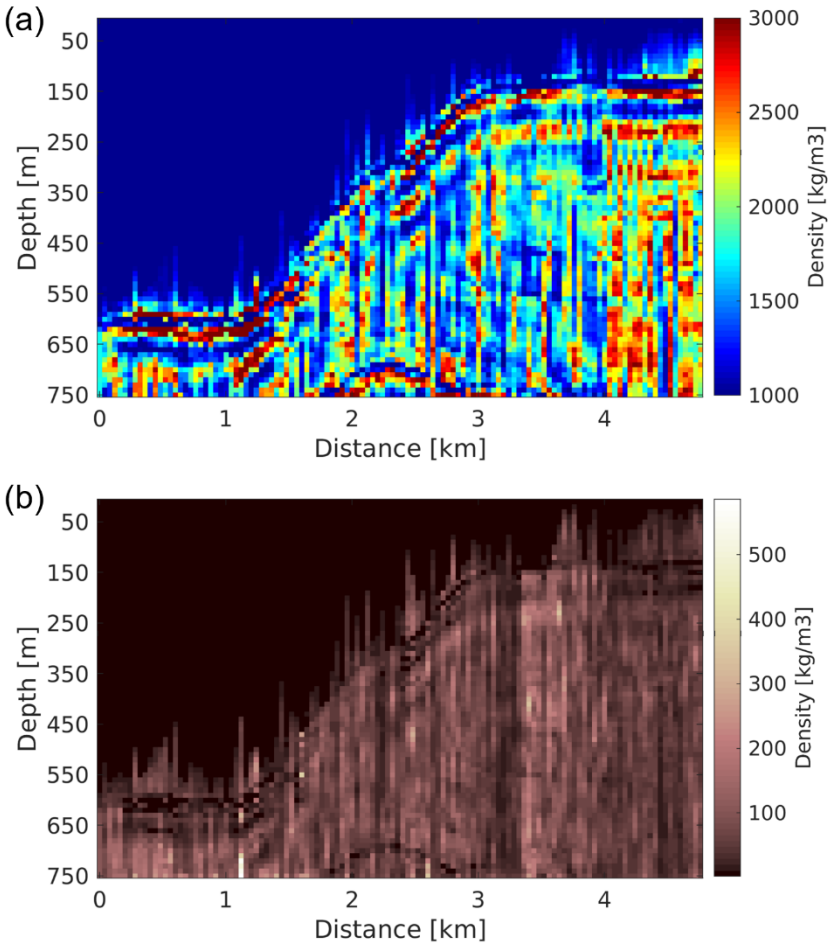

Figure 4.18. Final results from posterior distribution; (a) mean as an estimated ground model, and (b) standard deviation for uncertainty quantification

4.5. Discussion

From the field case study in the Gulf of Mexico, we realized the importance of the appropriate P-wave velocity assumption to improve the quality of the geophysical seismic inversion. Even though many studies introduced the empirical equations to define the relationship between bulk density and P-wave velocity (Gardner, Gardner, & Gregory, 1974; Hamilton, 1971; Hamilton & Bachman, 1982), we need to estimate the

best relationships on different target area to represent each geological environment with respect to the locations.

Another feature we can find from the seismic inversion result is that the image of the bathymetry is not a clear continuous line. The image is still blurry, and this does not relate to the assumption of the P-wave velocity. The observed seismic data near the Sigsbee Escarpment was measured in 1976 from the cruise survey for the project IG1904 at Marine Geoscience Data System (MGDS) (http://www-udc.ig.utexas.edu/sdc/cruise.php?cruiseIn=IG1904#segy_proc). The seismic data and cores were collected crossing the Louisiana and Florida shelf edge on its way into deeper water. The project at that time couldn't fully conduct the seismic data processing, so the signal still contains lots of noise and induces blurry image with high uncertainty at the geological layer interfaces, including the seafloor. Since the Bayesian stochastic approach captured all the possible candidates from the sampling, the bathymetric line cannot become a clear, vivid line in the estimated ground model. If we apply this inversion method to a recent offshore data, which comes after more advanced geophysical seismic processing, we can expect the quality of the estimated ground model will be more clear and accurate for better site characterization.

4.6. Conclusion

We developed a Bayesian seismic inversion method with geophysical seismic convolution method and reversible jump Markov chain Monte Carlo (rj-MCMC) method in varying dimensions for the offshore site characterization. This method is appropriate

to apply for the conventional offshore multichannel seismic survey data, which comes after the geophysical signal processing to remove the noise and show the images of subsurface geological stratigraphy. Since the site characterization near the Sigsbee Escarpment, a submarine steep slope area in the northern Gulf of Mexico, is important for the geohazard assessment of offshore platform foundations, we applied seismic inversion method to an offshore field data and generated a 2-dimensional vertical cross-section image of the marine slope area. Dense and soft sediments under the seafloor were indicated with depth information, which was not able to identify from the offshore geophysical data. Since the result of the offshore ground model contains the soil information in discretized mesh in the space-domain, estimated soil properties can be used for the geomechanical slope stability analysis. This geophysical inversion method can be applied for conventional offshore seismic survey data, and we expect this can support to get better information from geophysical survey data for offshore geotechnical projects.

5. PROBABILISTIC SEISMIC FULL WAVEFORM INVERSION

5.1. Summary

Seismic full-waveform inversion (FWI) has been widely used for the geophysical site characterization to estimate the subsurface geomechanical soil properties. Since FWI requires high computational resources, most of FWI studies rely on uniform rectangular elements for the 2-dimensional regular mesh with the flat free surface condition. However, near-surface FWI for land survey requires the irregular free surface condition to generate the accurate surface waves, which cause high amplitude noise signals on the ground, and this has been challenging in conventional FWI approaches. This study shows Bayesian FWI case studies with the elastic seismic forward model based on Spectral Finite-Element Method (SEM), which is appropriate to apply irregular mesh as a high-order Finite-Element Method (FEM). Reversible jump Markov chain Monte Carlo (rj-MCMC) method is applied with SEM for the iterative inversion process, and we introduce supervised classification to support this rj-MCMC method for mesh partitioning to define geologic layers as classified groups. Synthetic shallow layered soil models, whose material properties are the same as the known soil profiles in GeoPark, Florida, are used with different topography conditions for the numerical experiments. The modeling results show that seismic modeling errors due to the topography cannot be neglected, and the case studies of inversion prove the importance of accurate modeling for near-surface site characterization.

5.2. Introduction

The goal of geophysical inverse modeling is to find the earth model parameters which can explain the observed complex data (Sen & Stoffa, 2013). The model-based seismic inversions attempt to infer the parameters by iteratively fitting the predicted data to the observed data (Albert Tarantola, 2005). This process minimizes the misfit function by using optimization algorithms, and predicted earth model could be used to find out unknown anomalies under the surface. The seismic inversion methods for near-surface land site characterization can be categorized into two; using body waves and using surface waves.

Seismic Full Waveform Inversion (FWI) methods which use body waves are capable of imaging the arbitrarily heterogeneous velocity profiles in high-resolution because the modeling is based on fine-scale discretized mesh model. However, because of demanding computation, early attempts of FWI was based on only the acoustic wave equation (Lailly, 1983; Pan, Phinney, & Odom, 1988; R. G. Pratt & Worthington, 1988; A. Tarantola, 1984). As computational power increases, elastic FWI has been widely used for the deep subsurface imaging (Fang, Herrmann, & Silva, 2014; Fichtner, 2011; Peyman P Moghaddam & Herrmann, 2010; P. P. Moghaddam, Keers, Herrmann, & Mulder, 2013; Shin & Ha, 2008; Shin & Min, 2006; Sirgue & Pratt, 2004; Virieux & Operto, 2009).

Another approach; Spectral Analysis of Surface Waves (SASW), which uses only surface waves, is the most well-known application for near-surface site characterizations (Joh et al., 1997; Stokoe et al., 1994). SASW uses the dispersion of

surface waves in shallow heterogeneous media, and this method has been developed to MASW with multi-channel streamers (C. B. Park et al., 1999). MASW has been applied for various near-surface land surveys (Choon B Park, Miller, Xia, & Ivanov, 2007; Xia, Miller, Park, Hunter, & Harris, 2000; Yuan et al., 2014), but the estimated material properties from this method is not accurate enough because this method relies on subjective decision for the dispersion range.

Even though FWI can image more accurate values of the soil model, FWI was not considered to be appropriate for near-surface land surveys because complex surface waves from irregular free surface generate unwanted ground roll noise. Recently, deterministic FWI study was compared with the SASW method at the same target site in Austin, TX (Kallivokas et al., 2013) and showed a better subsurface image than SASW. Tran et al. (2013) also showed the possibility of applying FWI to image near-surface anomalies to detect the shallow sinkholes. However, all those approaches were still limited to flat topography earth model, and the effect of topography in FWI has not been fully discussed yet.

In this study, we introduce a new approach of seismic FWI to overcome the irregular topography problem of near-surface site characterization. We decided to use the Spectral Finite-Element Method (SEM) as our forward modeling to generate the accurate both body waves and surface waves. For the iterative inversion, we applied the reversible jump Markov chain Monte Carlo (rj-MCMC) method, which has the number of subsurface layers as a random variable. This approach has been widely used for geophysical inverse problems (Thomas Bodin & Sambridge, 2009; Malinverno, 2002),

and this approach was recently developed for the inversion of post-processing seismic data (Cho et al., 2017; S. K. Dadi, 2014; Zhu & Gibson, 2016). We applied this rj-MCMC method for pre-processing seismic data with mesh partitioning from supervised classification. Synthetic shallow earth model, which is the same as the known soil profile in GeoPark, Florida, has been used to prove this application. Since this earth model contains irregular free surface, this study explains how important to consider the effect of topography in the near-surface FWI for land surveys.

5.3. Methodology

5.3.1. Numerical algorithms for seismology

The subsurface soil and rock layers are complex heterogeneous material, so analytical methods cannot provide solutions of wave propagations, and approximate numerical methods are required. The numerical methods transform the original partial differential problems into a system of algebraic equations, and the forward modeling notates the numerical modeling of the full seismic wavefield. The primary purpose of the forward modeling has been the reduction of the numerical dispersion errors and represent the realistic reflections, refractions, scattering with attenuation to simulate the accurate elastic dynamics in heterogeneous anisotropic media (Aki & Richards, 2002).

The partial differential equation of the seismic wavefield modeling in time-domain is as below:

$$M(x)\frac{d^2u(x,t)}{dt^2} = A(x)u(x,t) + s(x,t) \tag{5.1}$$

where M and A are the mass matrix and the stiffness matrix (Marfurt, 1984). The source term is denoted by $s(x,t)$ and the seismic wave field is $u(x,t)$. In the acoustic forward modeling, which considers only P-wave velocity, the wavefield represents the pressure. This should be replaced into horizontal and vertical particle velocities or displacements for elastic forward modeling, which requires considering both P- and S-wave velocities simultaneously. The t represents the time, and x does the spatial coordinates. Equation is solved with an explicit time-marching algorithm that the value of the wavefield at time step $(n+1)$ is inferred from the value at the previous time step (n) for each spatial position. Implicit time marching algorithms are not recommended because they require expensive computational memory (Virieux & Operto, 2009).

The equation can be transformed into the frequency domain, and the wave equation reduces to a system of the linear equation:

$$B(x,w)u(x,w) = s(x,w) \tag{5.2}$$

The $s(x,w)$ is the source term, $B(x,w)$ is the impedance matrix, and $u(x,w)$ is the seismic wavefield as the solution of this equation with the angular frequency (w) (Marfurt, 1984). The impedance matrix has a symmetric pattern, but it becomes non-symmetric when we apply the absorbing boundary conditions(Hustedt, Operto, & Virieux, 2004). Thus, the equation can be solved by decomposition of the impedance matrix into lower and upper (LU) triangular decomposition. This decomposition leads equation to direct solver approach whose advantage is the efficient solving for multiple sources using forward and backward substitutions. The frequency domain forward modeling approach has been widely used for efficient 2-D modeling (R. G. Pratt, 1990,

1999; R Gerhard Pratt, Shin, & Hick, 1998; R. G. Pratt & Shipp, 1999; R. G. Pratt & Worthington, 1988; Sirgue & Pratt, 2004), but its scalability for large scale 3-D problems is limited because of the time and memory complexities of LU factorization (Operto et al., 2007).

For the seismic forward modeling, the time domain is the most appropriate to identify the specific type of arriving signals, whereas the frequency domain is better to select only one or a few frequencies for waveform inversion (Virieux & Operto, 2009). Even though the frequency domain is efficient to solve the wave equation, the useful time windowing based on wavefield snapshots cannot be applied in frequency domain forward modeling, because only a few frequencies are modeled at a time. Since seismic waves in the real soil and rock media always convert to other phases, for example, P- to S-waves or S- to P-waves, arriving signal identification in time-domain is essential for the correct signal interpretation. Consequently, we need to consider both time and frequency domains on how to apply our forward modeling appropriately for each problem with different conditions.

For the accurate elastic wavefield generation in the time domain, the finite-difference method (FDM) has been used as the most common and popular approach to simulate seismic wave propagation. The FDM belongs to grid-point methods, whose computational domain is covered by space-time grids, and each function has its values at grid points. A derivative of wave equations is approximated and uses the values of a specified set of grid points (Moczo, Kristeck, & Halada, 2004). FDM represents the spatial domain with uniformly structured grids. This approach was suitable for the

acoustic wave equation and applied since the early beginning of seismology (Kelly & Alford, 1973; Kelly, Ward, Treitel, & Alford, 1976; Marfurt, 1985). The application based on FDM for wave equations allows the direct and accurate modeling through the complex and heterogeneous geological structures, and it is the most widely used approach for seismic modeling until now.

One of the big steps for making FDM be widely used is the staggered grid finite-difference method (SG-FDM) (Virieux, 1986). Conventional grid, which all displacement and body-force components are located at each spatial grid point, is the most natural choice of the FDM. The particle motion of seismic wave propagation on the grids represents the stress-strain relation based on Hooke's law and constitutive law in continuum mechanics with initial and boundary conditions (Moczo, Kristek, & Bystricky, 2000). However, the displacement on the conventional grids had problems with instabilities and grid dispersion error with high Poisson's ratio. However, with the staggered grid, different components of one physical parameter are defined at different staggered points. Thus, all displacement components share the same grid position, but each stress-tensor component has its own grid position. The staggered grid scheme solved the numerical instabilities due to the vertical derivatives over the free surface and it was also appropriate to handle the acoustic and elastic material coupling problem at the fluid-solid interfaces. The staggered grid was induced for earthquake rupture dynamic modeling first (Madariaga, 1976), and Virieux (1986) made major progress in the seismic forward modeling. Advanced study, the 4th order staggered grid FDM, is also introduced by (Levander, 1988) and this scheme showed less memory requirement

of four times for 2-D and eight times for 3-D modeling than traditional SG-FDM. Thus, this 4th order SG-FDM has become the dominant type of the forward model schemes in the large scale seismic wave propagation models (Graves, 1996; Y. Liu & Sen, 2009a, 2009b; Moczo, Robertsson, & Eisner, 2007).

The FDM has many advantages in practical exploration seismology, but many users worried about the inherent limitations of this method. Even though FDM is relatively easy for implementation and it requires relatively less expensive computational resources, FDM suffers from grid flexibility, so the implementation of irregular boundary conditions is challenging problem (Moczo et al., 2004). Since FDM relies on uniformly structured grids, representation of complex geological structures such as structural faults, salt dome, or complex surface topography for land seismic cannot be correct, and FDM still requires much more elaboration.

The first step of the finite-element method (FEM) dates back to the beginning of the 20th century with remarkable publications (Galerkin, 1923; Ritz, 1909). This method is also known as the method of weighted residuals (Clough, 1980), and the first naming as finite-element was applied by Turner, Clough, Martin, and Topp (1956). The FEM has been developed as the most frequently used technique for many engineering problems to solve the partial differential equations. In structural mechanics, scalar wave propagation was implemented by using FEM (Kallivokas & Lee, 2004b), and this approach extended into 3-dimensional solutions (Kallivokas & Lee, 2004a). The perfectly matched layers (PML) as absorbing boundary conditions are also applied on it (Fathi, Poursartip, & Kallivokas, 2015; Kang & Kallivokas, 2010) and this application

has been used for geotechnical site investigation (Fathi, Poursartip, Stokoe, & Kallivokas, 2016; Kallivokas et al., 2013; Pakravan, Kang, Newtson, & Kallivokas, 2014).

FEM, which based on the interchanging between global and local coordinate system, has a strong advantage to solve the waveform equations on the unstructured grids, compared with the FDM (Marfurt, 1984). Curved layer interfaces and free surface boundary are very important for seismology, so the FEM is applied to an early time (Marfurt, 1977) and developed to improve the accuracy in elastic media with frequency-domain solutions (Marfurt, 1985). However, FEM requires a more expensive cost of computational resources than FDM to implement the complex assembling of global matrices, and FEM approach is recognized as a quite inefficient method (Gao, Chung, Gibson, Fu, & Efendiev, 2015). Thus, a lot of large scale geophysical waveform projects, especially in the oil and gas industry, are developed with FDM, and FEM has rarely been used in contrast to FDM (Virieux, Calandra, & Plessix, 2011).

The spectral finite-element method (SEM) belongs to the continuous Galerkin formulation of FEM (Schuberth, 2003). The basis functions are piecewise continuous Lagrange polynomial within the variable nodes. Therefore, the seismic wavefield solution is supposed to be smoothly continuous. The difference between SEM and the traditional CG-FEM is the choice of appropriate integration points in calculating the mass and stiffness matrices on the local elements. Traditional CG-FEM uses Gaussian integration points; otherwise, SEM uses the Gauss-Lobatto-Legendre (GLL) integration

points, which makes the strictly diagonal global mass matrix and be efficient to calculate.

The SEM was first introduced to the field of computational fluid dynamics (Patera, 1984). This method is based on a weak formulation of the wave equation, just like other FEM approaches, and this SEM approach was first used in geophysics in the 1990s (Priolo, Carcione, & Seriani, 1994; Seriani & Priolo, 1994). Based on these studies, Komatitsch and Vilotte (1998) developed SEM as an efficient tool for accurate seismic wave simulation. Many numerical validation for complex problems such as mesh refinement in 2-D and 3-D problems (Komatitsch & Tromp, 1999), fluid-solid interface (Komatitsch, Barnes, & Tromp, 2000), global seismology (Komatitsch & Tromp, 2002a, 2002b), perfectly matched layer (PML) as absorbing boundary condition (Komatitsch & Tromp, 2003). Since the SEM application for seismology was introduced in a mature level (Komatitsch, Tsuboi, & Tromp, 2005), the grid dispersion and stability criteria for SEM were proved (J. D. De Basabe & Sen, 2007; T. Liu, Sen, Hu, De Basabe, & Li, 2012). The research groups in seismology keep developing this SEM approach and many practical research based on SEM with real site information were published (Casarotti et al., 2008; S. J. Lee et al., 2008; Luo, Tromp, Denel, & Calandra, 2013; Tape, Liu, Maggi, & Tromp, 2010; Tromp, Komatitsch, & Liu, 2008).

The efficiency of Continuous Galerkin FEM (CG-FEM) for waveform modeling is limited by the continuity requirements in a medium when high contrasts of material properties exist. Furthermore, CG-FEM is difficult to apply to mesh discretization, which includes non-conforming element boundaries. These problems are naturally

solved with the discontinuous Galerkin finite-element method (DG-FEM), which allows discontinuous solutions across the cell boundaries. This approach was first proposed for the neutron transport equation (Reed & Hill, 1973), and applied to wave propagation solutions since the 2000s (Chung & Engquist, 2006, 2009; Grote, Schneebeli, & Schotzau, 2006; Riviere & Wheeler, 2003). DG-FEM approaches have the advantage that the global mass matrix is the block diagonal and dramatically reduce computational costs. Besides, the parallelization of the algorithm, which is difficult with CG-FEM due to the overlapping in the supports of element nodes, becomes easier. Many advanced studies are published recently to maximize these attractive advantages of DG-FEM (Chan, Chung, & Cohen, 2013; Chung, Lam, & Qian, 2015; Wilcox, Stadler, Burstedde, & Ghattas, 2010; Ye, de Hoop, Petrovitch, Pyrak-Nolte, & Wilcox, 2016). However, the discussions about the efficiency and accuracy of this DG-FEM is still ongoing compared with another algorithm, multiscale finite-element method. Multiscale finite-element method (MsFEM) is also proposed for the elliptic problems in porous media (Hou & Wu, 1997). In this approach, the partial differential equation is solved on an only course mesh instead of fine mesh. The course mesh represents the fine heterogeneous material properties on a fine scale with multiscale basis functions (Efendiev & Hou, 2009). The idea of multiscale basis functions was applied to wave propagation modeling based on pressure velocity mixed formation for the acoustic wavefield (R. L. Gibson, Gao, Chung, & Efendiev, 2014). The improved algorithm with better stability was developed as a generalized multiscale finite-element method (GMsFEM) (Efendiev, Galvis, & Hou, 2013) and applied to elastic wavefield modeling (Chung, Efendiev, & Leung, 2014). The

multiscale basis functions are built directly from the smallest eigenvalues and efficiently be able to relate the wavefield with a coarse mesh to the finer scale heterogeneities (Gao, Fu, Gibson, Chung, & Efendiev, 2015).

Since the seismic forward modeling is one of the key points for better seismic imaging and inversion methods, we need to compare them and be aware of the strength and weakness of each algorithm. Three different approaches, SEM, DG-FEM, and FDM, for acoustic and elastic wave propagation, were reviewed to compare their accuracy and efficiency (Jonas D De Basabe & Sen, 2009). FEM approaches are suitable for accurately approximating waves with free surface and fluid-solid boundary conditions. However, they are still considerably slower than the FDM because FDM uses only a few of the surrounding nodes. J. D. De Basabe and Sen (2015) developed their previous research more complex fluid-solid interface conditions. They found that classical FDM and low-order FEM have a significant amount of grid dispersion error in the fluid region, which renders them impractical to use for offshore research. On the other hand, SG-FDM and SEM show accurate results for the body waves in both fluid and solid regions.

Even though lots of advanced numerical approaches were developed, Virieux et al. (2011) summarized that finite-difference approaches with fourth-order compact schemes with second-order time integration show the strongest advantages in a marine environment. Because FDM, working with relatively smooth earth parameters at the wavelength scale, is often sufficient in velocity model building with acoustic P-waves. However, with the regaining of the importance of land seismic exploration, they concluded that the SEM might become crucial to better model the seismic wave

propagation because it has reached a mature level of precise modeling including surface waves and around complex interfaces.

5.3.2. Spectral Finite-Element Method (SEM)

We decided to use the SEM algorithm for near-surface seismic FWI because SEM has two main benefits for seismic forward modeling. First, SEM can handle complex geometries with great accuracy. Although the traditional FEM has obvious advantages based on flexible meshes for complex geological structures, its accuracy was lacking (Komatitsch & Vilotte, 1998). Otherwise, the pseudo-spectral method shows high accuracy, but cannot handle the complex geometries. SEM is a combined method between FEM and pseudo-spectral method to obtain both advantages; flexibility and accuracy (Komatitsch et al., 2005). Second, SEM is a much efficient algorithm than the FEM for the forward modeling with high-order basis functions. Traditional high-order FEM, which uses Gauss-Legendre integration, have suffered from expensive computational costs. However, the discretized elements in SEM is based on high-degree Lagrange interpolation, and integration over each elements is accomplished based on the Gauss-Lobatto-Legendre integration (GLL) rule (Komatitsch & Tromp, 1999). This approach leads to an exactly diagonal mass matrix, and the algorithm can be drastically simplified in matrix inversion calculation.

SEM is an efficient algorithm than the FEM for the seismic forward modeling with high-order basis functions. Traditional high-order FEM, which uses Gauss-Legendre integration, have suffered from expensive computational costs. However, the

discretized elements in SEM is based on high-degree Lagrange interpolation, and integration over each elements is accomplished based on the Gauss-Lobatto-Legendre integration (GLL) rule (Komatitsch & Tromp, 1999). This approach leads to an exactly diagonal mass matrix, and the algorithm can be simplified in matrix inversion calculation.

Furthermore, these GLL points help numerical modeling to apply accurate boundary conditions. For the elastic seismic modeling, the free surface, the top boundary of the earth model, is important to generate strong surface waves, so this SEM method is appropriate for the near-surface seismic modeling with irregular topography. Furthermore, the other boundaries on the left, right, and bottom are assigned as Perfect Match Layers (PML), the artificial absorbing boundaries to avoid using a huge mesh (Komatitsch & Tromp, 2003). Figure 5.1 shows an example of the two-dimensional mesh in vertical cross-section with irregular topography, surrounded by PMLs with the source (×) & receivers (◊) on the irregular free surface. All the elements with different colors indicate the independent random variables in the FWI.

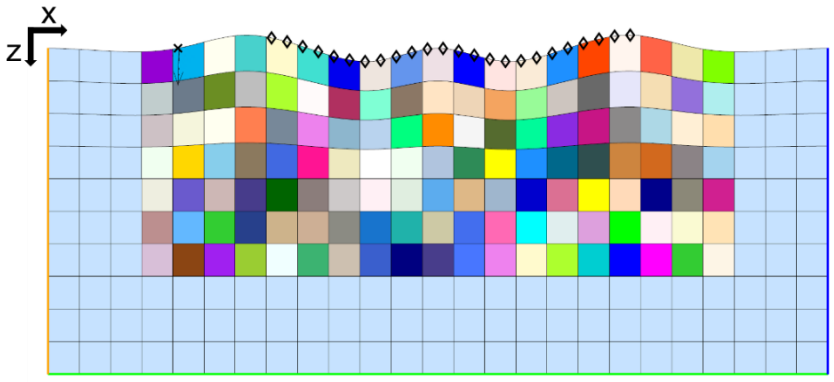

Figure 5.1. Two-dimensional mesh with irregular free surface

Details of the elastic seismic wavefield in SEM is described well in Komatitsch et al. (2005), and the fundamental governing partial differential equations can be defined as below equations in both velocity and stress fields in the 2-D spatial domain.

$$\dot{v_x} = \frac{1}{\rho}\left(\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xz}}{\partial z}\right) + f_x \tag{5.3.a}$$

$$\dot{v_z} = \frac{1}{\rho}\left(\frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{zz}}{\partial z}\right) + f_z \tag{5.3.b}$$

$$\dot{\sigma_{xx}} = (\lambda + 2\mu)\frac{\partial v_x}{\partial x} + \lambda\frac{\partial v_z}{\partial z} \tag{5.3.c}$$

$$\dot{\sigma_{zz}} = (\lambda + 2\mu)\frac{\partial v_z}{\partial z} + \lambda\frac{\partial v_x}{\partial x} \tag{5.3.d}$$

$$\dot{\sigma_{xz}} = \mu\left(\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x}\right) \tag{5.3.e}$$

where $v, f, \sigma,$ and ρ represent the particle velocity vector, body force vector, stress tensor, and density of materials, respectively. The λ and μ are the lame parameters, important to calculate the elastic parameters in the seismic forward modeling, and important free surface boundary condition is implemented as the stresses with vertical component (σ_{zz}, σ_{xz}) on the irregular top surface become to zero. We used SPECFEM2D package for elastic near-surface seismic forward modeling based on SEM, and the validations of this package were shown from theoretical case studies (Komatitsch et al., 2000; Komatitsch & Tromp, 2002a).

5.3.3. Mesh partitioning by supervised classification

We need to define the modeling parameters in discretized mesh to implement forward modeling. However, fine-scale mesh, whose element size is less than seismic resolution, cannot guarantee to find out subsurface layers. Thus, we applied supervised

classification for the mesh partitioning to make a group of elements as geologic strata. The classification is a technique, which has been often used for the quantitative analysis of remote sensing image data, to classify the pixel data into categorized groups with unique values (Richards, 2013). For example, if all the discretized elements in Figure 5.1 had a homogeneous soil property, then this model is defined as one group. To increase the number of soil groups, we need to spread random seeds in the 2-D discretized mesh. Figure 5.2 shows an example of mesh partitioning from the random seeds spreading. The location with horizontal and vertical coordinates of the seeds are random, but each seed contains a unique value for the classification. Based on the locations of those random seeds with a unique value, supervised classification distributes those unique values of seeds to the entire 2-D spatial domain. Based on the coordinates of those random seeds, the classification method automatically generate the interfaces between different groups and defined every discretized element in the whole mesh.

This mesh partitioning is important to support the rj-MCMC method, and it has been studied as Bayesian Partitioning modelling (D. G. T. Denison, Adams, Holmes, & Hand, 2002). The main idea of this method is to use the Voronoi diagram. This is a partitioning of a plane into sub-regions based on distance to points in a specific subset of the plane, and also called as a Voronoi tessellation or decomposition. The set of points can be defined by rj-MCMC approach and applied for geophysical two-dimensional imaging studies (T Bodin et al., 2012; S Esmailzadeh, Medina-Cetina, Kang, & Kallivokas, 2015b). However, this approach was clear to apply for tetrahedral mesh elements, but the calculation with quadrilateral mesh elements takes much more

computational time. Because this method has to make the projection from tetrahedral mesh to quadrilateral mesh, and also requires the information outside of the mesh domain to draw the polygons of the Voronoi cells. On the other hand, the supervised classification from deep learning package in MATLAB library showed fast and accurate mesh partitioning, though we cannot explain everything inside the machine learning algorithm.

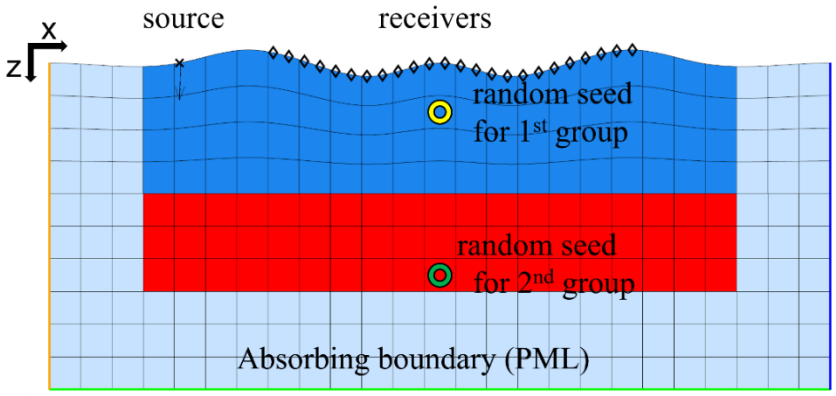

Figure 5.2. Mesh partitioning by supervised classification with random seed points on discretized 2-D mesh

5.3.4. Reversible jump MCMC inversion

Since the number of soil groups is unknown for the inverse modeling, we applied the reversible jump Markov chain Monte Carlo (rj-MCMC) inversion algorithm to use the dimension of modeling parameters as a random variable. This rj-MCMC maximizes the Probability Density Function (PDF) of the posterior, denoted as $p(\boldsymbol{\theta}|\boldsymbol{d})$.

$$\boldsymbol{p(\theta|d)} = \frac{\boldsymbol{p(d|\theta)p(\theta)}}{\boldsymbol{p(d)}} \tag{5.4}$$

The $\boldsymbol{\theta}$ represent the modeling parameter, and d is the observed data. The likelihood function ($p(\boldsymbol{d}|\boldsymbol{\theta})$) quantifies the PDF based on the residual between the observed data and predicted data, where $G(\boldsymbol{\theta})$ is the elastic seismic forward modeling based on SEM algorithm and C_d is the covariance matrix.

$$p(\boldsymbol{d}|\boldsymbol{\theta}) = \frac{1}{[(2\pi)^n |C_d|]^{\frac{1}{2}}} \times \exp\left[-\frac{1}{2}(G(\boldsymbol{\theta}) - \boldsymbol{d})^T C_d^{-1}(G(\boldsymbol{\theta}) - \boldsymbol{d})\right] \qquad (5.5)$$

We applied two Bayesian priors to control this inverse modeling. The first is the range of soil material property. Since we can define the minimum and maximum value of soil properties from 1-D in-situ data, we apply uniform distribution for the PDF of prior. This is important to control the stability condition in seismic forward modeling, and we will explain this in the next forward modeling chapter. The second is a geological prior that we assumed geologic structure has a high correlation in the only horizontal direction. Most of the near-surface soil layers don't have a history of the tectonic force, so the formation of the layer shows a horizontal layered model. That means the difference of material property in the horizontal direction is about to zero, and we applied Gaussian distribution to this feature as the equation.

The key point in this study is the rj-MCMC method defines the modeling dimension in random and spreads random seeds with unique values inside the irregular mesh. Based on the positions of the seeds, classification defines the partitions in mesh and allocates the modeling parameters on every element for SEM forward modeling. This model selection process randomly comes from one of four proposals below to generate a new candidate modeling parameter ($\boldsymbol{\theta}'$).

- BIRTH: randomly pick a new seeding point in 2-d mesh to classify as a new group before new partitioning.
- DEATH: Randomly pick two of the current groups and merge them into one. The new location of random seed become the center of the previous two seeds.
- MOVE: Randomly pick one of the current groups and change the location of the seeding point in the 2-d mesh. The stochastic movement of the new location follows Gaussian distribution; $x' \sim N(x, 1)$ & $z' \sim N(z, 1)$.
- PERTURB: Randomly pick one of the current groups and change the material property in a fixed location. The stochastic movement of the new material flows Gaussian distribution; $\beta' \sim N(\beta, coefficient\ variance * \beta)$

5.4. Modeling of Seismic Wavefield

5.4.1. Dataset from GeoPark in Florida

Our numerical modeling has been proved with a synthetic shallow earth model whose material properties are the same as the soil profile data from GeoPark in Tampa, West Florida. Many geophysical studies have shown that the Florida state may have possibly the most sinkhole areas because of the statewide dominance of Karst geology (Kruse, Grasmueck, Weiss, & Viggiano, 2006). Various geologic, geotechnical, and geophysical investigations have been implemented to compile the cross-section of the buried sinkhole area (Carpenter, Doll, & Kaufmann, 1998; Stewart & Parker, 1991).

The range of S-wave velocity (V_S) in shallow depth was defined from 150 to 450 *m/s*, and we assumed the P-wave velocity (V_P) was double of V_S with constant 2 *g/cc*

density. All the seismic acquisition parameters in this numerical modeling follow the MASW field survey system that uses 24 channel land streamers spaced every 1.2 meters. Since the seismic source from a sledgehammer generates low-frequency energy, 15 *Hz* of vertical point stress was used as the seismic source on the free surface.

Figure 5.3 shows a 2-D soil vertical profile from the GeoPark field MASW survey. This multichannel SASW approach uses the dispersion of surface waves in the frequency domain (Figure 5.3-a) and finds out the dominant wave mode to figure out the S-wave velocity along with the depth (Figure 5.3-b). The survey with multiple shots generate the 2-D cross-section of the subsurface map and shows an abrupt change at a depth of 25 *ft*. Based on the geometry and properties from this analysis, we generated our synthetic shallow soil model for the numerical experiments.

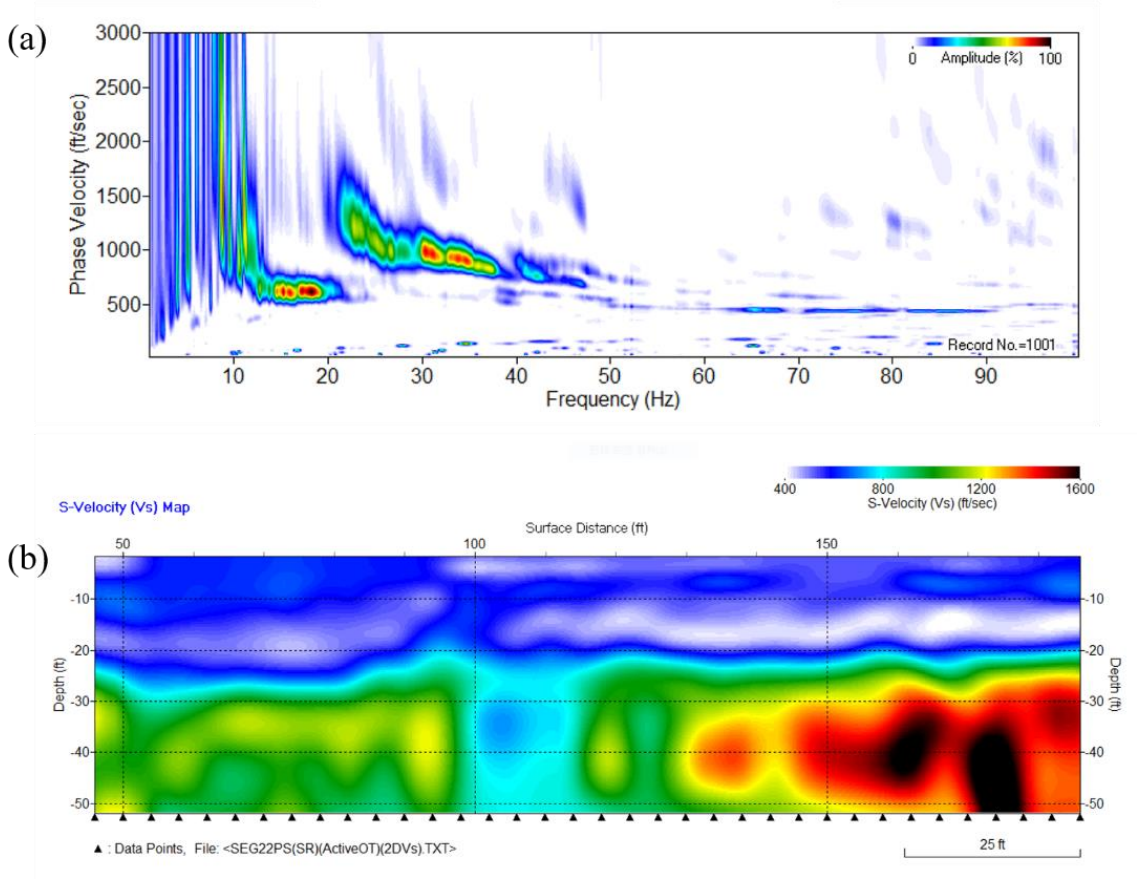

Figure 5.3. SASW analysis of GeoPark field data

5.4.2. Parameter analysis in geomechanics

When we apply the elastic seismic inverse modeling, we can convert the geophysical parameters to the geomechanical parameters. Lame parameters, denoted by λ and μ, are two material dependent quantities derived from strain-stress relationships in continuum mechanics. If we use the homogeneous and isotropic materials, Lame parameters define the Hookes's law;

$$\sigma = 2\mu\epsilon + \lambda\, tr(\epsilon) I \tag{5.6}$$

where σ, ϵ, I are the stress, strain tensor, and identity matrix, respectively. Since this stress was defined with P-wave velocity (V_P), S-wave velocity (V_S) in previous soil dynamics with FEM, the relationship between the Lame parameters and geophysical parameters can be defined as below with the material density (ρ).

$$V_P = \sqrt{\frac{\lambda+2\mu}{\rho}} \tag{5.7.a}$$

$$V_S = \sqrt{\mu/\rho} \tag{5.7.b}$$

$$\lambda = \rho(V_P^2 - 2V_S^2) \tag{5.7.c}$$

$$\mu = \rho V_S^2 \tag{5.7.d}$$

Consequently, we can define the geomechanical parameters; Poisson's ratio (v), Bulk modulus (K), and Elastic modulus (E), which we need to infer from the geophysical seismic non-destructive surveys data before the construction projects.

$$v = \frac{\lambda}{2(\lambda+\mu)} = \frac{V_P^2 - 2V_S^2}{2\left(V_P^2 - V_S^2\right)} \tag{5.8.a}$$

$$K = \lambda + \frac{2}{3}\mu = \rho\left(V_P^2 - \frac{4}{3}V_S^2\right) \tag{5.8.b}$$

$$E = \frac{\mu(3\lambda+2\mu)}{\lambda+\mu} = \frac{\rho V_S^2\left(3V_P^2 - 4V_S^2\right)}{V_P^2 - V_S^2} \tag{5.8.c}$$

Based on those equations above, our synthetic dataset can be used to estimate the geomechanical properties in the shallow layered soil model (Table 5.1). Since the ratio of V_P/V_S is 2, the Poisson's ratio become 0.33, which is similar to the realistic field data

of shallow land soils. The elastic Young's modulus of soil represents weathered soil, and the bedrock can be estimated as a stiff rock.

Table 5.1. Converting parameters from geophysics to geomechanics

	Parameters	soil	bedrock
Geophysics	Density [kg/m^3]	2000	2000
	P-wave velocity [m/s]	400	800
	S-wave velocity [m/s]	200	400
Continuum mechanics	Lame's 1st parameter [MPa]	160	640
	Lame's 2nd parameter [MPa]	80	320
Geomechanics	Poisson's ratio	0.33	0.33
	Bulk modulus [MPa]	240	960
	Young's modulus [MPa]	216	864

5.4.3. Parameter analysis for numerical computation

Based on the range of target material properties, numerical forward modeling to generate seismic wavefield requires a specific setup for accurate, stable, and efficient computation. First of all, all of the numerical solvers for seismic wave modeling based on partial differential equations should follow an important computational rule to solve the accurate wavefield. Otherwise, the results would be corrupted by artificial errors, which is called grid dispersion. The grid dispersion is a numerical noise related to grid spacing, which has a detrimental effect on accuracy (J. D. De Basabe & Sen, 2007). It

occurs because the actual velocity of high-frequency waves in the grid is different from the true velocity, and it is dependent on the grid spacing and the size of the time step for the numerical seismic wavefield modeling. We need to avoid this grid dispersion error, and the first step should be the decision of the maximum spatial grid point distance (discretized element size) for the correct sampling of the wavefield. The maximum size of the element should be smaller than the wavelength over the required element number (n_e) based on the order of basis function in the numerical algorithm. When we choose the 4th order of basis function in SEM, then the number of elements in one wavelength should be larger than five (Komatitsch et al., 2005). Since the minimum wavelength is the minimum S-wave velocity divided by the maximum dominant frequency, the minimum wavelength is about 200/15 = 13.3 meters. Thus, the maximum grid spacing becomes 13.3/5 = 2.6 meters, and we chose 2.5 meters as our element size of the 2-D mesh generation in both horizontal and vertical directions.

$$dh \leq \frac{\lambda_{min}}{n_e} = \frac{\min(V_S)}{n_e \times \max(f)} \tag{5.9}$$

The discretization in time domain also has to satisfy the sampling criterion to ensure the numerical modeling in stable condition. Since the wave propagates in the discretized mesh in spatial domain with the chosen element size (dh), the time step (dt) should be less than the time to travel between two adjacent grid points for the waves. In mathematically, this limit is called Courant-Friedrichs-Lewy (CFL) condition (Courant, Friedrichs, & Lewy, 1967), and this is an essential condition while solving any partial differential equations with numerical approaches. The Courant number (C) and the maximum size of the time step (dt) can be calculated from the CFL condition. We

already decided to use grid point distance (dh) as 2.5 meters, and the maximum range of P-wave velocity is assumed as 800 *m/s*. From the maximum Courant number for the elastic seismic modeling in SEM, the maximum suggested value of stable time step (dt) becomes about 0.259 *ms*. We chose 0.2 *ms* as the time step for the seismic forward modeling, so the maximum CFL stability condition becomes 0.386, which should be lower than 0.5.

$$C = \frac{\max(V_P) * dt}{dh} \leq C_{max} \tag{5.10.a}$$

$$dt \leq \frac{dh}{C_{max} \times \max(V_P)} \tag{5.10.b}$$

Based on the chosen time step, we need to define the total number of time iteration for one cycle of seismic forward modeling. The bottom depth of the spatial domain is 25 meters, and the distance between the source and the farthest boundary on the right side is about 50 meters. Thus, the diagonal distance of this spatial mesh is about 55 meters, and the maximum travel distance for reflected waves for going down and back up is about 80 meters. This total travel distance over the minimum range of S-wave velocity (150 *m/s*) becomes the minimum required time length (530 *ms*). Furthermore, the numerical delay time when we use the Ricker wavelet with 15 Hz is inverse of the dominant frequency, and it becomes about 70 *ms* of additional time length. Thus, we should generate the wavefield during at least 600 *ms* with 0.2 *ms* time steps, and the minimum required the number of time iteration for seismic modeling becomes 3000. We chose 4000-time iterations for one cycle of the forward modeling to capture other multiple reflections inside the shallow anomalies.

For the efficient forward modeling with this demanding time iteration, we applied artificial absorbing boundary condition to reduce the total number of elements in the mesh. Perfectly matched layer (PML) is a boundary condition which has the remarkable property of having a zero reflection coefficient for all angles of incidence and frequencies before discretization (Komatitsch & Tromp, 2003). For the land seismic data acquisition, the top boundary is set up as the free surface, which means the interface between the elastic medium and air to vanish all stresses in the normal direction. However, all other left, right, and bottom side of boundaries has PMLs to make the model as a semi-finite spatial 2-D model. The PMLs doesn't show reflection when the wave equation is solved by a coordinate stretch of the wave equations in the frequency domain and creating exponential decaying the amplitude of wave solutions (Komatitsch & Martin, 2007). This artificial boundary condition makes the spatial mesh as small as possible to save the computational resources for the wave modeling, and this efficient artificial boundary condition has been widely used for seismic wavefield simulations (Kang & Kallivokas, 2010; Pakravan et al., 2014).

5.4.4. Effect of irregular topography

Based on our chosen modeling parameters, we generated two synthetic seismograms by SEM forward modeling from different 2-D soil models to verify the effect of the irregular free surface before the stochastic FWI. One soil model has a flat free-surface (Figure 5.4-a), and the other has a smooth irregular free-surface (Figure 5.4-b). All other conditions; the property of shallow soil layer, the property of deep bedrock,

the shape of bedrock interface, are the same in those two synthetic soil models except the topography. One single seismic source (x) is used on the left side of the domain surface, and receivers (◊) measure the propagation of seismic energy from left to right. Even though the elevation of smooth free-surface was less than one-meter, Figure 5.4-c shows the modeling error from the wrong assumption of flat topography. The red traces from the irregular surface model were supposed to be the same as the blue traces from a flat surface model, so the difference between them causes modeling error in near-surface seismic FWI. Thus, near-surface FWI should be able to consider the effect of topography to get accurate results.

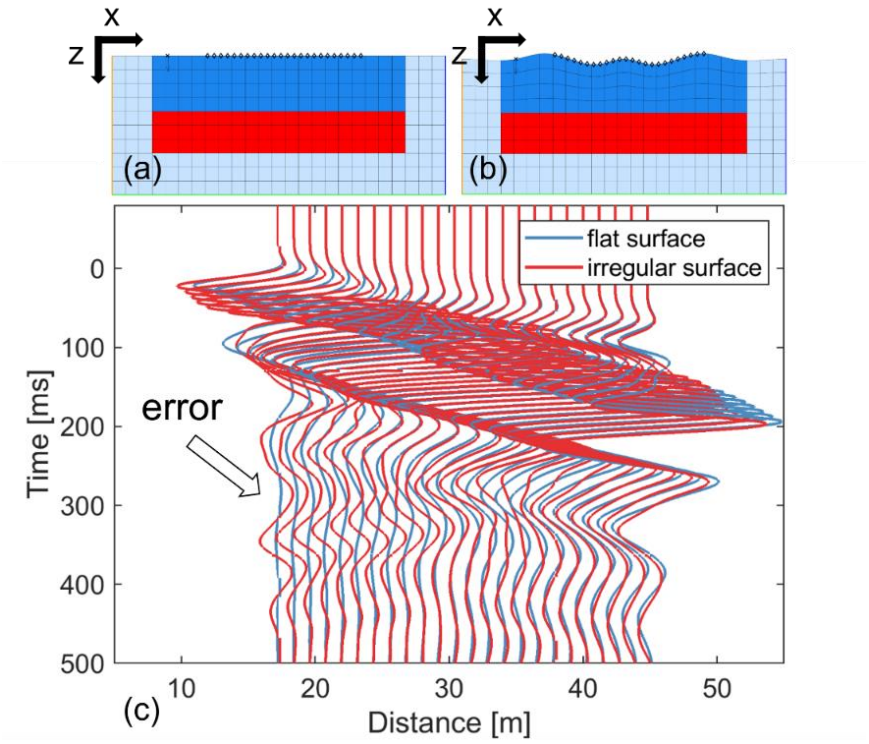

Figure 5.4. Soil models with flat (a) & irregular (b) free surface, and the modeling error of the seismogram due to the topography effect

5.5. Numerical Experiments

5.5.1. Soil model with a flat free-surface

Based on the SEM forward model, which generates an accurate wavefield in shallow soil media, we decided to verify the Bayesian inversion method first on the simple soil model without the topography condition. Our target model is a two-layered soil model (Figure 5.5-a), but the initial model is a simple homogeneous soil (Figure 5.5-b). The condition of source and receivers are the same to the previous Figure 5.4, and the seismogram signal from one of the receivers (the most far from the source location) from two soil models are compared in Figure 5.5-c. The seismogram from the initial homogeneous soil model shows only direct P and S waves. However, the observed signals from the layered target soil model have direct waves and refraction (P-P') and reflection (P-P*) waves from the subsurface bedrock. Those signals in multi-channel acquisition system become complex in near-surface wavefield. Even though the interpretation of those signals is challenging, the benefit of FWI method is that this approach uses the signals as observed data itself for the iterative inverse modeling without complex interpretation.

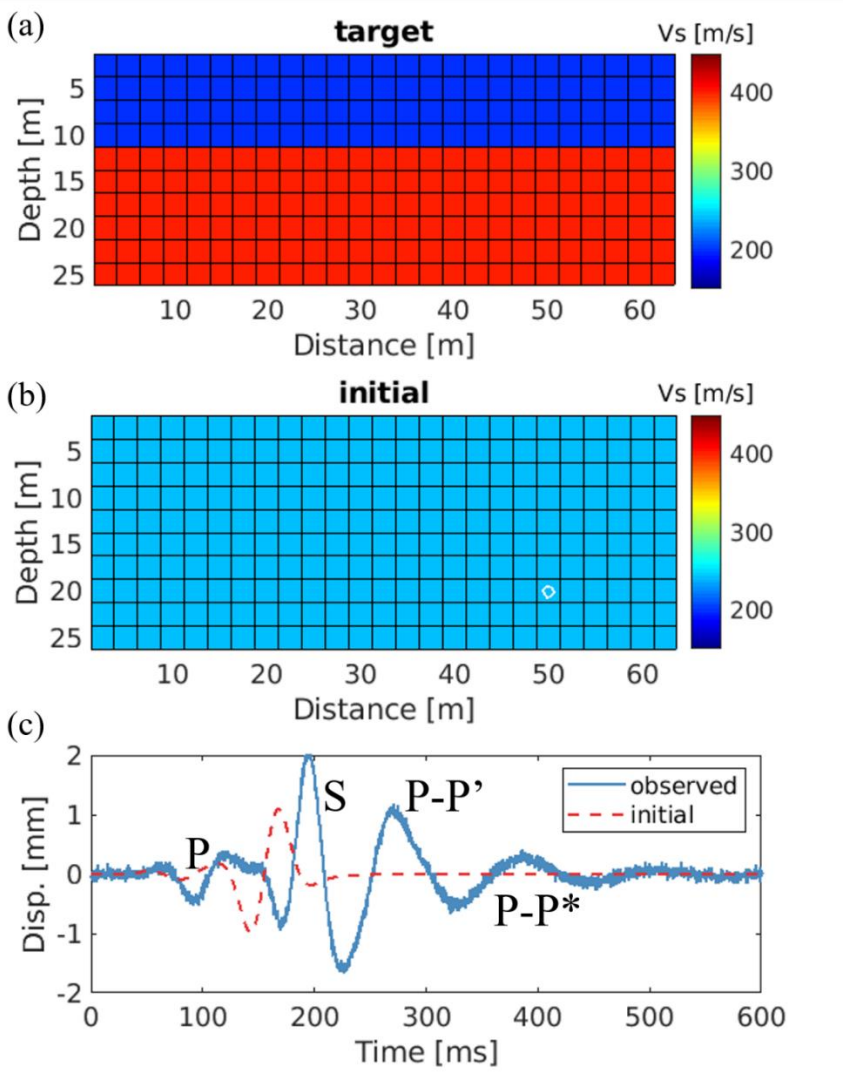

Figure 5.5. (a) unknown target model, (b) initial guess model, and (c) seismograms from (a) and (b)

Figure 5.6 shows the process of Bayesian FWI method. Based on the signal difference in seismograms, the rj-MCMC method randomly changes the dimension of the inverse modeling to increase or decrease the number of subdomains in the mesh. Not only the number of subdomains but also the locations are random, so mesh partitioning in this stochastic process eventually converge to the same geometry as the target layered

soil model. The material properties inside those partitioned subdomains are randomly changed based on the comparison of seismic signal amplitudes. As the number of accepted samples increasing, the predicted soil model shows the values of shallow soil layer (blue), and bedrock layer (red) as similar to the target model.

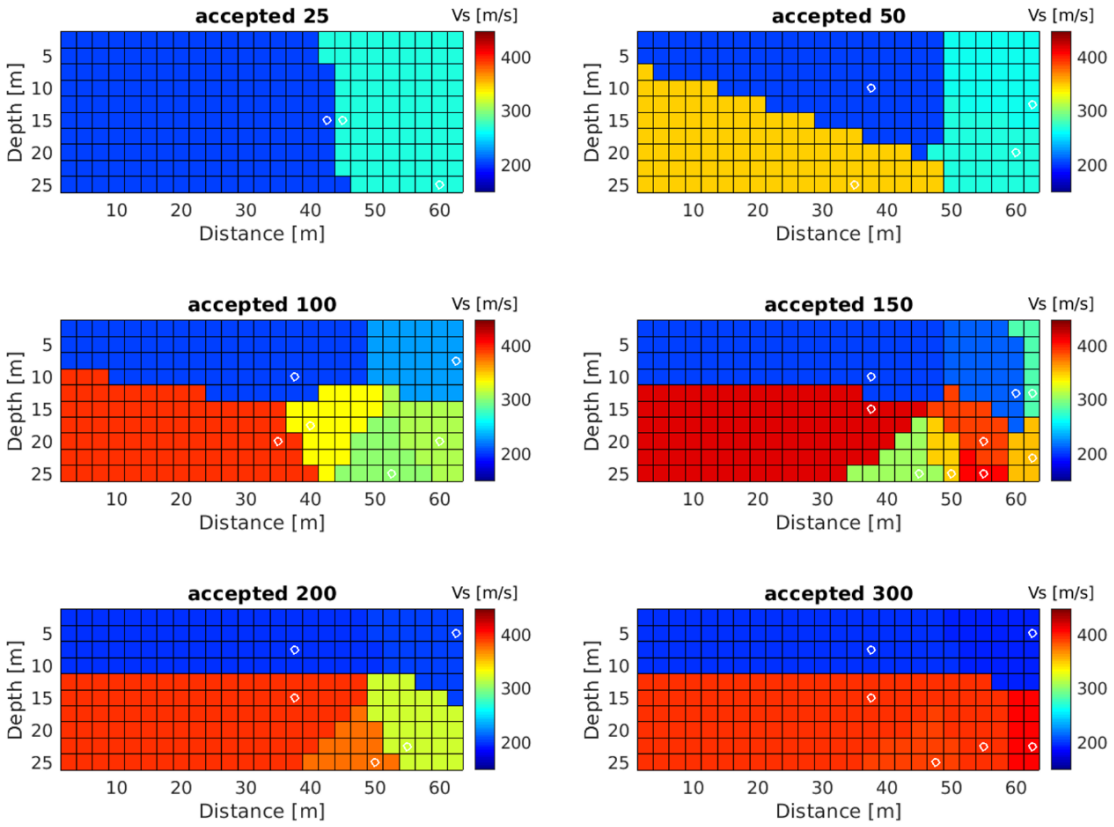

Figure 5.6. Stochastic process in the rj-MCMC method. Random seeds (white) define subdomains with random material properties.

The iterative sampling from rj-MCMC method shows the decrease of the residual error between observed and predicted data (Figure 5.7). Since the dimension of model parameters are in random, most of the rj-MCMC approaches use L2-norm calculation of this error to decide the burn-in point instead of the mean and standard deviation of

random parameters. This L2-norm calculation quantifies Euclidean distance between two different vector data, and the amount of the error converges after the 200,000 sampling point.

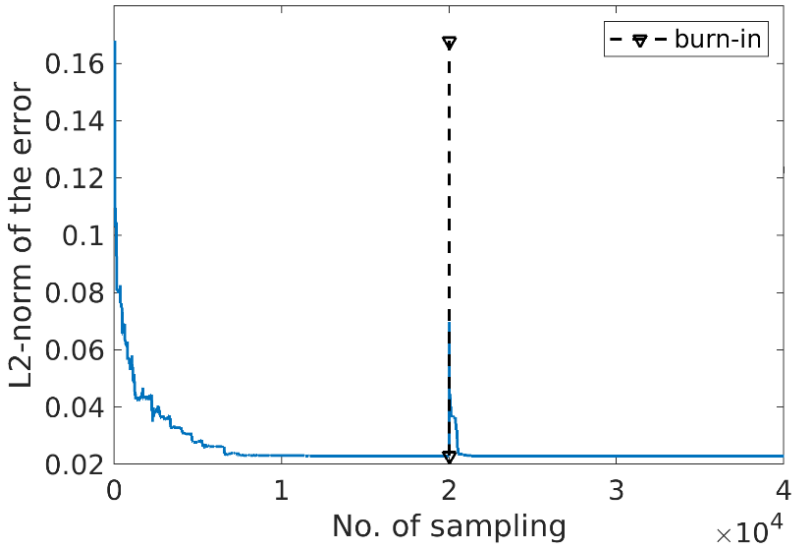

Figure 5.7. Burn-in point from the residual error estimation of simulation with a simple flat soil model

From the 200,000 samples after this burn-in point, we measured the mean and standard deviation of the posterior distribution on every discretized 2-dimensional mesh elements. The mean of parameters show the final results of the estimated soil model (Figure 5.8-a), and the standard deviation quantifies the uncertainty of this inversion result (Figure 5.8-b). From these results, we can find that this inversion method shows three features to discuss more.

- The estimated results show the accurate soil properties in-depth, and it can show the location of subsurface bedrock.

- Since the seismic forward modeling uses absorbing boundary conditions; PMLs on the left, right, and bottom edges of the soil model, the results show high uncertainty inside the PMLs.
- The location of the seismic source was on the left side of the soil model. Multi-channel receivers measure lots of signals from left, but they couldn't have enough signals from the right side. Thus, the results show less uncertainty near the seismic source, and the uncertainty is propagated and increasing along the distance from the source.

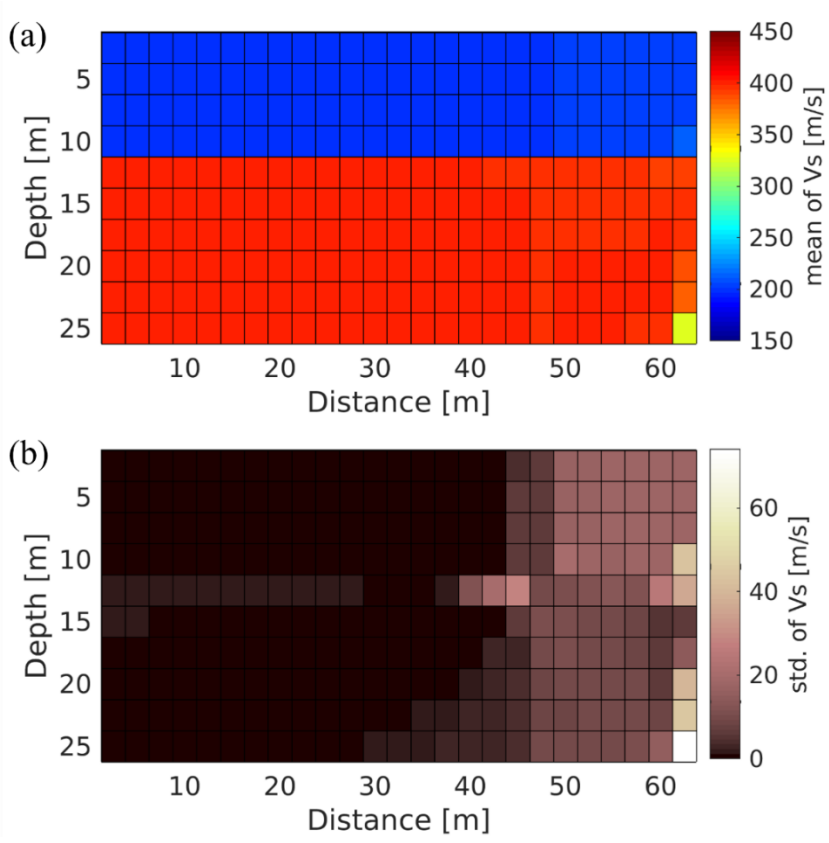

Figure 5.8. Mean (a) and standard deviation (b) of the posterior distribution from the inversion with a flat free surface model

5.5.2. Soil model with an irregular free-surface

After the verification of the simulation with a simple model, we implemented the same inversion with an irregular free-surface model to apply the topography condition. The topography follows the smooth curved shape in the previous Figure 5.1. Since the elevation of topography on the ground surface is easy to measure, we assumed that we already know the information of topography as our priors in the inverse modeling. All other conditions in soil model are the same as the previous model (Figure 5.5-a), and the process of stochastic sampling shows very similar steps to the previous simulation without topography effects.

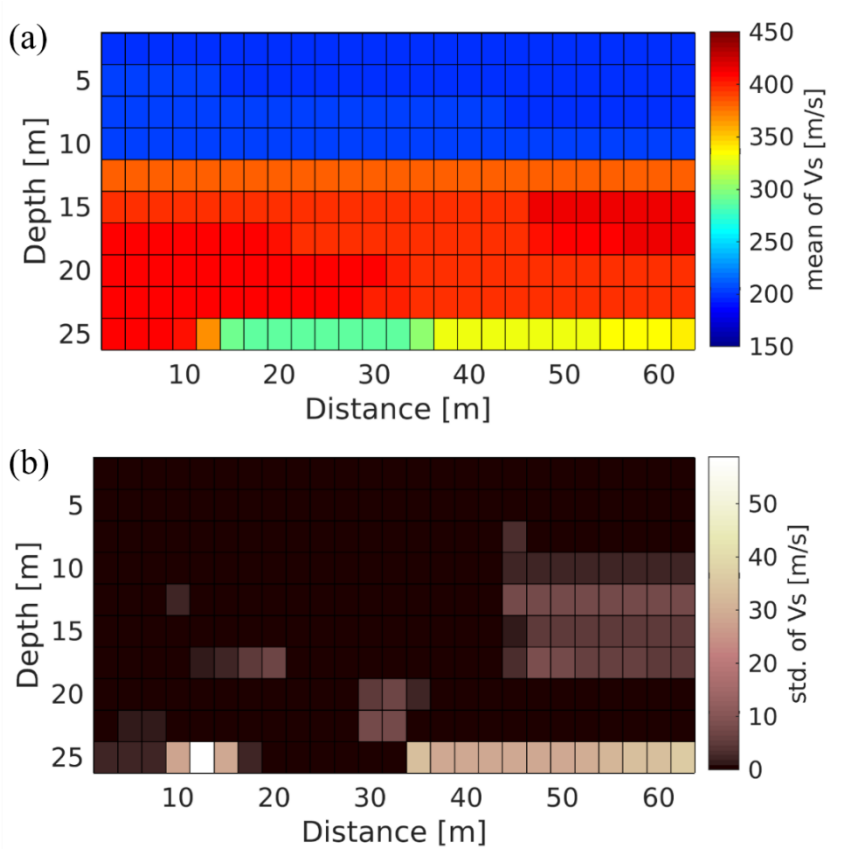

Figure 5.9. Mean (a) and standard deviation (b) of the posteriors from the inversion with the irregular free surface model (no visualization of the irregular surface)

The final result (Figure 5.9-a) is similar to the previous inversion, but we can see that the estimation inside the PML boundary is not accurate enough. Because of the high uncertainty in those boundary elements, the uncertainty propagation from the seismic source is also difficult to find out (Figure 5.9-b). This simulation took 200,000 samples and chose 100,000 as our burn-in point, as same as the previous inversion in Figure 5.8. The only difference between them was applying the irregular shape to the free-surface, and the observed data from the seismogram becomes more complex due to the topography. In short, this simulation shows that the Bayesian FWI approach with classification method can estimate the accurate location of bedrock layer and material properties under the irregular topography. However, the accurate estimation requires more samples as the shape of free-surface becomes more complex.

5.6. Discussion: Inversion with Ignored Topography

As we described in the introduction part, the main purpose of this study is to show the effect of modeling errors due to the irregular topography in geophysical seismic inversion method for the accurate near-surface site characterization of elastic soil properties. From the previous two inverse modelings, we proved that this new approach of Bayesian FWI could reconstruct the accurate soil profile, and the complex geometry of the topography model requires more enough number of sampling. In our 3rd case study, we generated the observed data with irregular the topography model (Figure 5.4-b) and implemented our inversion with the flat free-surface model (Figure 5.4-a). This is the common mistake when we apply near-surface seismic inverse modeling to

observed field data. Even though the elevation of topography is relatively smaller than the size of the whole spatial domain, this inappropriate modeling condition causes a wrong estimation of soil properties in different shape of bedrock.

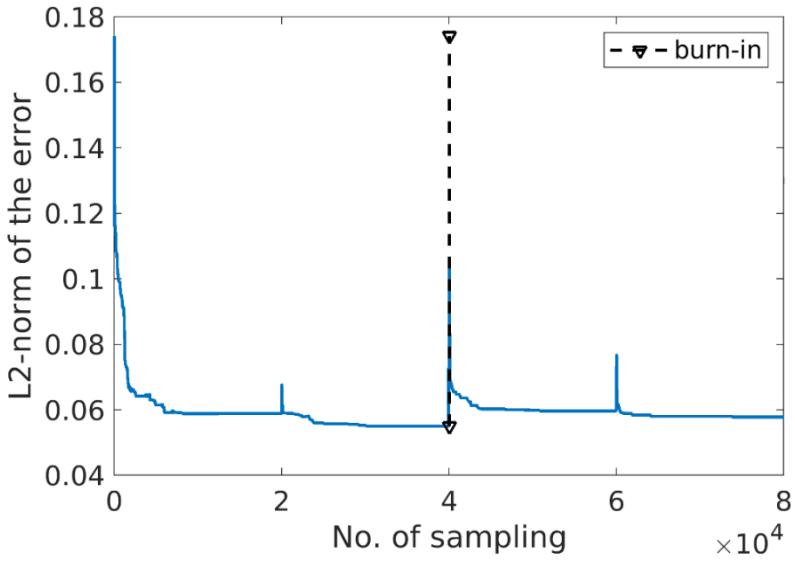

Figure 5.10. Burn-in point from the residual error estimation of simulation with the wrong assumption as the flat soil model

Since the samplings during the stochastic process were not easy to find the stationary condition, we took 800,000 samples, which was four times more than the previous two case studies. We chose the half-point 400,000 as our burn-in point and calculated the mean and standard deviation from the chosen samples. The final result from this wrong assumption inverse modeling is in Figure 5.11. The target layered soil model was the same as the previous. However, the estimated result from the mean of posterior shows a different wrong soil profile (Figure 5.11-a). The most important point of this wrong site characterization is that the modeling cannot find out the depth and shape of the subsurface bedrock. Furthermore, the soil properties inside the bedrock are

underestimated than true values in the target layered soil model. Only the center part, from 30 to 40 meters in the horizontal distance, shows the accurate depth of interface, flat geometry, and similar material properties of the bedrock layer. We think this result comes from the geometry of our synthetic soil model, which has a relatively flat surface only at the center point. In other words, gentle slopes outside of this center part cause critical errors from the irregular free surface, and the estimated soil profile becomes useless. The importance of this topography effect cannot be emphasized enough for accurate near-surface characterization.

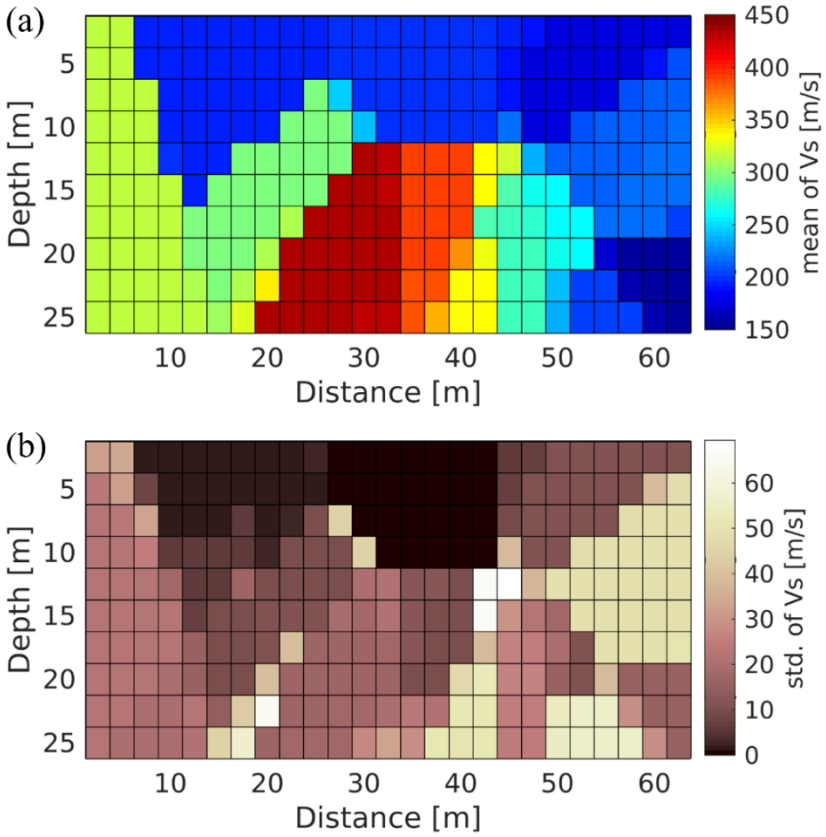

Figure 5.11. Mean (a) and standard deviation (b) of the posterior distribution from the wrong assumption inversion with the flat free surface model

5.7. Conclusion

We introduce a Bayesian seismic full waveform inversion (FWI) method for shallow site characterization to overcome the irregular free-surface topography problem. Spectral finite-element method (SEM) is used as the seismic forward model because this algorithm is suitable to capture the complex geometry of the model and able to generate the accurate elastic wavefield, including challenging surface waves. The stochastic inversion process is based on the trans-dimensional rj-MCMC method to define not only the material properties but also the number of unknown modeling parameters. To support the mesh partitioning for the rj-MCMC method, we applied a supervised classification method and generated random seeding points to define subdomains inside the two-dimensional mesh model. This approach is verified with a simple synthetic layered soil model with a flat free surface condition, and we applied it to an irregular surface topography model again. Even though the seismogram showed significant different observed data due to the topography, the inversion shows an accurate estimation of subsurface soil profiles. This study also shows a case study of the wrong assumption that assumed flat surface condition in modeling when observed data was measured from irregular surface. The results of these studies show the importance of accurate topography condition for the seismic full-waveform inversion to image the accurate subsurface geomechanical soil properties from the geophysical seismic survey on land.

REFERENCES

Aki, K., & Richards, P. G. (2002). *Quantitative Seismology*: University Science Books.

Aubeny, C. (2017). *Geomechanics of Marine Anchors*: CRC Press.

Aubeny, C., Dutt, R., Young, R. G., Murff, J. D., Gilbert, R. B., & Doyle, E. (2013). An Expert Panel Review of Geotechnical Site Investigation Regulations and Current Industry State of Practice. *Offshore Technology Conference*.

Bangs, N. L. B., Musgrave, R. J., & Trehu, A. M. (2005). Upward shifts in the southern Hydrate Ridge gas hydrate stability zone following postglacial warming, offshore Oregon. *Journal of Geophysical Research-Solid Earth, 110*(B3).

Berger, W., Lanier, D., & Jeanjean, P. (2006). Geologic setting of the mad dog mooring system. *Offshore Technology Conference*.

Bodin, T., Salmon, M., Kennett, B., & Sambridge, M. (2012). Probabilistic surface reconstruction from multiple data sets: An example for the Australian Moho. *Journal of Geophysical Research: Solid Earth, 117*(B10).

Bodin, T., & Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. *Geophysical Journal International, 178*(3), 1411-1436.

Brand, J., Lanier, D., Berger III, W., Kasch, V., & Young, A. (2003). Relationship between Near Seafloor Seismic Amplitude, Impedance, and Soil Shear Strength Properties and Use in Prediction of Shallow Seated Slope Failure. *Offshore Technology Conference*.

Briaud, J.-L. (2013). *Geotechnical engineering: unsaturated and saturated soils*: John Wiley & Sons.

Brookshire Jr, B. N., Landers, F. P., & Stein, J. A. (2015). Applicability of ultra-high-resolution 3D seismic data for geohazard identification at mid-slope depths in the Gulf of Mexico: Initial results. *Underwater Technology, 32*(4), 271-278.

Caers, J. (2005). *Petroleum geostatistics*. Richardson: Society of Petroleum Engineers.

Campbell, K. J., Quiros, G. W., & Young, A. G. (1988). *The Importance Of Integrated Studies To Deepwater Site Investigation*. Paper presented at the Offshore Technology Conference, Houston, Texas.

Carlin, B. P., Gelfand, A. E., & Banerjee, S. (2014). *Hierarchical modeling and analysis for spatial data*: Chapman and Hall/CRC.

Carpenter, P. J., Doll, W. E., & Kaufmann, R. D. (1998). Geophysical Character of Buried Sinkholes on the Oak Ridge Reservation, Tennessee (Presented at SAGEEP95 in Orlando, FL). *Journal of environmental and Engineering Geophysics, 3*(3), 133-145.

Casarotti, E., Stupazzini, M., Lee, S. J., Komatitsch, D., Piersanti, A., & Tromp, J. (2008). CUBIT and seismic wave propagation based upon the spectral-element method: An advanced unstructured mesher for complex 3D geological media. *Proceedings of the 16th International Meshing Roundtable*, 579-597.

Chan, H. N., Chung, E. T., & Cohen, G. (2013). Stability and Dispersion Analysis of the Staggered Discontinuous Galerkin Method for Wave Propagation. *International Journal of Numerical Analysis and Modeling, 10*(1), 233-256.

Cho, Y., Zhu, D., & Gibson, R. (2017). 3D transdimensional Markov-chain Monte Carlo seismic inversion with uncertainty analysis. In *SEG Technical Program Expanded Abstracts 2017* (pp. 607-611).

Chung, E. T., Efendiev, Y., & Leung, W. T. (2014). Generalized Multiscale Finite Element Methods for Wave Propagation in Heterogeneous Media. *Multiscale Modeling & Simulation, 12*(4), 1691-1721.

Chung, E. T., & Engquist, B. (2006). Optimal discontinuous Galerkin methods for wave propagation. *Siam Journal on Numerical Analysis, 44*(5), 2131-2158.

Chung, E. T., & Engquist, B. (2009). Optimal Discontinuous Galerkin Methods for the Acoustic Wave Equation in Higher Dimensions. *Siam Journal on Numerical Analysis, 47*(5), 3820-3848.

Chung, E. T., Lam, C. Y., & Qian, J. L. (2015). A staggered discontinuous Galerkin method for the simulation of seismic waves with surface topography. *Geophysics, 80*(4), T119-T135.

Clough, R. W. (1980). The Finite-Element Method after 25 Years - a Personal View. *Computers & Structures, 12*(4), 361-370.

Courant, R., Friedrichs, K., & Lewy, H. (1967). On Partial Difference Equations of Mathematical Physics. *Ibm Journal of Research and Development, 11*(2), 215-+.

Cressie, N., & Wikle, C. K. (2015). *Statistics for spatio-temporal data*: John Wiley & Sons.

Dadi, S., Gibson Jr, R., & Wang, K. (2015). Quantification of uncertainty in velocity log upscaling using reversible jump MCMC. In *SEG Technical Program Expanded Abstracts 2015* (pp. 3476-3481): Society of Exploration Geophysicists.

Dadi, S. K. (2014). *Estimation of Impedance Using Seismic Reflection Data Based on Transdimensional Inversion.* (Doctoral dissertation), Texas A & M University.

De Basabe, J. D., & Sen, M. K. (2007). Grid dispersion and stability criteria of some common finite-element methods for acoustic and elastic wave equations. *Geophysics, 72*(6), T81-T95.

De Basabe, J. D., & Sen, M. K. (2009). New developments in the finite-element method for seismic modeling. *The Leading Edge, 28*(5), 562-567.

De Basabe, J. D., & Sen, M. K. (2015). A comparison of finite-difference and spectral-element methods for elastic wave propagation in media with a fluid-solid interface. *Geophysical Journal International, 200*(1), 278-298.

Denison, D. G. T., Adams, N. M., Holmes, C. C., & Hand, D. J. (2002). Bayesian partition modelling. *Computational Statistics & Data Analysis, 38*(4), 475-485.

Denison, D. G. T., Holmes, C. C., Mallick, B. K., & Smith, A. F. M. (2002). *Bayesian Methods for Nonlinear Classification and Regression*: Wiley.

Efendiev, Y., Galvis, J., & Hou, T. Y. (2013). Generalized multiscale finite element methods (GMsFEM). *Journal of Computational Physics, 251*, 116-135.

Efendiev, Y., & Hou, T. Y. (2009). Multiscale Finite Element Methods: Theory and Applications. *Multiscale Finite Element Methods: Theory and Applications, 4*, 1-234.

Esmailzadeh, S., Medina-Cetina, Z., Kang, J., & Kallivokas, L. (2015a). Varying dimensional Bayesian acoustic waveform inversion for 1D semi-infinite heterogeneous media. *Probabilistic Engineering Mechanics*, *39*, 56-68.

Esmailzadeh, S., Medina-Cetina, Z., Kang, J., & Kallivokas, L. F. (2015b). *Bayesian mechanistic imaging of two-dimensional heterogeneous elastic media from seismic geophysical observations.* Paper presented at the Frontiers in Offshore Geotechnics III: Proceedings of the 3rd International Symposium on Frontiers in Offshore Geotechnics (ISFOG 2015).

Fang, Z., Herrmann, F. J., & Silva, C. D. (2014). *Fast uncertainty quantification for 2D full-waveform inversion with randomized source subsampling.* Paper presented at the 76th EAGE Conference and Exhibition 2014.

Fathi, A., Poursartip, B., & Kallivokas, L. F. (2015). Time-domain hybrid formulations for wave simulations in three-dimensional PML-truncated heterogeneous media. *International Journal for Numerical Methods in Engineering*, *101*(3), 165-198.

Fathi, A., Poursartip, B., Stokoe, K. H., & Kallivokas, L. F. (2016). Three-dimensional P- and S-wave velocity profiling of geotechnical sites using full-waveform inversion driven by field data. *Soil Dynamics and Earthquake Engineering, 87*, 63-81.

Fichtner, A. (2011). *Full Seismic Waveform Modelling and Inversion* (1 ed.): Springer-Verlag Berlin Heidelberg.

Fleming, P. B., Long, H., Dugan, B., Germaine, J., John, C., Behrmann, J. H., . . . Scientists, I. E. (2008). Pore pressure penetrometers document high overpressure

near the seafloor where multiple submarine landslides have occurred on the continental slope, offshore Louisiana, Gulf of Mexico. *Earth and Planetary Science Letters, 269*(3-4), 309-324.

Foti, S., Comina, C., Boiero, D., & Socco, L. V. (2009). Non-uniqueness in surface-wave inversion and consequences on seismic site response analyses. *Soil Dynamics and Earthquake Engineering, 29*(6), 982-993.

Galerkin, B. G. (1923). Calculation of free bound elliptical plates in flexion. *Zeitschrift Fur Angewandte Mathematik Und Mechanik, 3*, 113-117.

Gao, K., Chung, E. T., Gibson, R. L., Fu, S. B., & Efendiev, Y. (2015). A numerical homogenization method for heterogeneous, anisotropic elastic media based on multiscale theory. *Geophysics, 80*(4), D385-D401.

Gao, K., Fu, S. B., Gibson, R. L., Chung, E. T., & Efendiev, Y. (2015). Generalized Multiscale Finite-Element Method (GMsFEM) for elastic wave propagation in heterogeneous, anisotropic media. *Journal of Computational Physics, 295*, 161-188.

Gardner, G. H. F., Gardner, L. W., & Gregory, A. R. (1974). Formation Velocity and Density - Diagnostic Basics for Stratigraphic Traps. *Geophysics, 39*(6), 770-780.

George, R. A., Gee, L. A., Hill, A. W., Thomson, J. A., & Jeanjean, P. (2002). High-resolution AUV surveys of the eastern Sigsbee Escarpment. *Offshore Technology Conference*.

Gibson, R. L., Gao, K., Chung, E., & Efendiev, Y. (2014). Multiscale modeling of acoustic wave propagation in 2D media. *Geophysics, 79*(2), T61-T75.

Gibson, R. L., & Hwang, K. (2009). *Quantification of Uncertainty In Velocity Log Upscaling By a Markov Chain Monte Carlo Method*. Paper presented at the 2009 SEG Annual Meeting, Houston, Texas.

Graves, R. W. (1996). Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. *Bulletin of the Seismological Society of America, 86*(4), 1091-1106.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika, 82*(4), 711-732.

Grote, M. J., Schneebeli, A., & Schotzau, D. (2006). Discontinuous Galerkin finite element method for the wave equation. *Siam Journal on Numerical Analysis, 44*(6), 2408-2431.

Gunning, J., & Glinsky, M. E. (2004). Delivery: An open-source model-based Bayesian seismic inversion program. *Computers & Geosciences, 30*(6), 619-636.

Hamilton, E. L. (1971). Prediction of in-Situ Acoustic and Elastic Properties of Marine Sediments. *Geophysics, 36*(2), 266-&.

Hamilton, E. L. (1979). Vp-Vs and Poisson Ratios in Marine-Sediments and Rocks. *Journal of the Acoustical Society of America, 66*(4), 1093-1101.

Hamilton, E. L., & Bachman, R. T. (1982). Sound-Velocity and Related Properties of Marine-Sediments. *Journal of the Acoustical Society of America, 72*(6), 1891-1904.

Hastings, W. K. (1970). Monte-Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika, 57*(1), 97-&.

Hou, T. Y., & Wu, X. H. (1997). A multiscale finite element method for elliptic problems in composite materials and porous media. *Journal of Computational Physics, 134*(1), 169-189.

Hustedt, B., Operto, S., & Virieux, J. (2004). Mixed-grid and staggered-grid finite-difference methods for frequency-domain acoustic wave modelling. *Geophysical Journal International, 157*(3), 1269-1296.

Jeanjean, P., Berger III, W., Liedtke, E., & Lanier, D. (2006). Integrated studies to characterize the Mad Dog Spar anchor locations and plan their installation. *Offshore Technology Conference*.

Jeanjean, P., Liedtke, E., Clukey, E., Hampson, K., & Evans, T. (2005). An operator's perspective on offshore risk assessment and geotechnical design in geohazard-prone areas. *Offshore Technology Conference*, 115-143.

Joh, S. H., Rosenblad, B. L., & Stokoe, K. H. (1997). Improved data interpretation method for SASW tests at complex geotechnical sites. *Proceedings of the Seventh (1997) International Offshore and Polar Engineering Conference, Vol I, 1997*, 875-881.

Kallivokas, L. F., Fathi, A., Kucukcoban, S., Stokoe, K. H., Bielak, J., & Ghattas, O. (2013). Site characterization using full waveform inversion. *Soil Dynamics and Earthquake Engineering, 47*, 62-82.

Kallivokas, L. F., & Lee, S. (2004a). Ellipsoidally-shaped local absorbing boundaries for three-dimensional scalar wave propagation. *Computational Mechanics, 35*(1), 11-23.

Kallivokas, L. F., & Lee, S. (2004b). Local absorbing boundaries of elliptical shape for scalar waves. *Computer Methods in Applied Mechanics and Engineering, 193*(45-47), 4979-5015.

Kang, J. W., & Kallivokas, L. F. (2010). Mixed unsplit-field perfectly matched layers for transient simulations of scalar waves in heterogeneous domains. *Computational Geosciences, 14*(4), 623-648.

Kelly, K. R., & Alford, R. M. (1973). Accuracy of Finite-Difference Modeling of Acoustic-Wave Equation. *Geophysics, 38*(6), 1207-1207.

Kelly, K. R., Ward, R. W., Treitel, S., & Alford, R. M. (1976). Synthetic Seismograms - Finite-Difference Approach. *Geophysics, 41*(1), 2-27.

Komatitsch, D., Barnes, C., & Tromp, J. (2000). Wave propagation near a fluid-solid interface: A spectral-element approach. *Geophysics, 65*(2), 623-631.

Komatitsch, D., & Martin, R. (2007). An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. *Geophysics, 72*(5), Sm155-Sm167.

Komatitsch, D., & Tromp, J. (1999). Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International, 139*(3), 806-822.

Komatitsch, D., & Tromp, J. (2002a). Spectral-element simulations of global seismic wave propagation - I. Validation. *Geophysical Journal International, 149*(2), 390-412.

Komatitsch, D., & Tromp, J. (2002b). Spectral-element simulations of global seismic wave propagation - II. Three-dimensional models, oceans, rotation and self-gravitation. *Geophysical Journal International, 150*(1), 303-318.

Komatitsch, D., & Tromp, J. (2003). A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation. *Geophysical Journal International, 154*(1), 146-153.

Komatitsch, D., Tsuboi, S., & Tromp, J. (2005). The Spectral-Element Method in Seismology. *Seismic Earth: Array Analysis of Broadband Seismograms, 157*, 205-227.

Komatitsch, D., & Vilotte, J. P. (1998). The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of the Seismological Society of America, 88*(2), 368-392.

Kruse, S., Grasmueck, M., Weiss, M., & Viggiano, D. (2006). Sinkhole structure imaging in covered Karst terrain. *Geophysical Research Letters, 33*(16).

Kumar, D., Sen, M. K., Bangs, N. L., Wang, C. S., & Pecher, I. (2006). Seismic anisotropy at Hydrate Ridge. *Geophysical Research Letters, 33*(1).

Kvalstad, T. J. (2007). What is the current" best practice" in offshore geohazard investigations? A State-of-the-art review. *Offshore Technology Conference*.

Lacasse, S., Nadim, F., Vanneste, M., L'Heureux, J.-S., Forsberg, C. F., & Kvalstad, T. J. (2013). *Case studies of offshore slope stability*. Paper presented at the Geo-Congress 2013: Stability and Performance of Slopes and Embankments III.

Lailly, P. (1983). *The seismic inverse problem as a sequence of before stack migrations*.

Lee, B., Rosenblad, B. L., Wright, S. G., & Stokoe, K. H., II. (1997). *Analytical Study of Surface Wave Testing Along the Seafloor*. Paper presented at the Offshore Technology Conference, Houston, Texas.

Lee, S. J., Chen, H. W., Liu, Q. Y., Komatitsch, D., Huang, B. S., & Tromp, J. (2008). Three-dimensional simulations of seismic-wave propagation in the Taipei basin with realistic topography based upon the spectral-element method. *Bulletin of the Seismological Society of America, 98*(1), 253-264.

Levander, A. R. (1988). 4th-Order Finite-Difference P-Sv Seismograms. *Geophysics, 53*(11), 1425-1436.

Liedtke, E., Jeanjean, P., & Humphrey, G. (2006). Geotechnical Site Investigation for the Mad Dog Spar Anchors. *Offshore Technology Conference*.

Liu, T., Sen, M. K., Hu, T. Y., De Basabe, J. D., & Li, L. (2012). Dispersion analysis of the spectral element method using a triangular mesh. *Wave Motion, 49*(4), 474-483.

Liu, Y., & Sen, M. K. (2009a). An implicit staggered-grid finite-difference method for seismic modelling. *Geophysical Journal International, 179*(1), 459-474.

Liu, Y., & Sen, M. K. (2009b). A new time-space domain high-order finite-difference method for the acoustic wave equation. *Journal of Computational Physics, 228*(23), 8779-8806.

Locat, J., Gardner, J. V., Lee, H., Mayer, L., Clarke, J. E. H., & Kammerer, E. (1999). Using multibeam sonar surveys for submarine landslide investigations. *Slope Stability Engineering, Vols 1 & 2*, 127-134.

Locat, J., & Lee, H. J. (2002). Submarine landslides: advances and challenges. *Canadian Geotechnical Journal*, *39*(1), 193-212.

Luo, Y., Tromp, J., Denel, B., & Calandra, H. (2013). 3D coupled acoustic-elastic migration with topography and bathymetry based on spectral-element and adjoint methods. *Geophysics, 78*(4), S193-S202.

Madariaga, R. (1976). Dynamics of an expanding circular fault. *Bulletin of the Seismological Society of America*, *66*(3), 639-666.

Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophysical Journal International*, *151*(3), 675-688.

Marfurt, K. J. (1977). Migration and Inversion of Seismic Data by Finite-Element Method. *Geophysics*, *42*(7), 1523-1523.

Marfurt, K. J. (1984). Accuracy of Finite-Difference and Finite-Element Modeling of the Scalar and Elastic Wave-Equations. *Geophysics*, *49*(5), 533-549.

Marfurt, K. J. (1985). Seismic Modeling - a Frequency-Domain Finite-Element Approach. *Geophysics, 50*(2), 340-340.

Medina-Cetina, Z., & Esmailzadeh, S. (2014). Joint states of information from different probabilistic geo-profile reconstruction methods. *Georisk: assessment and management of risk for engineered systems and geohazards*, *8*(3), 171-191.

Medina-Cetina, Z., Kang, J. W., Esmailzadeh, S., & Kallivokas, L. (2013). *Bayesian inversion of heterogeneous media: Introducing the next generation of integrated*

studies for offshore site investigations. Paper presented at the Offshore Technology Conference.

Medina-Cetina, Z., Son, J., & Moradi, M. (2019). *Bayesian Stratigraphy Integration of Geophysical, Geological, and Geotechnical Surveys Data*. Paper presented at the Offshore Technology Conference, Houston, Texas.

Menke, W., & Menke, J. (2016). *Environmental data analysis with MatLab*: Academic Press.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics, 21*(6), 1087-1092.

Moczo, P., Kristeck, J., & Halada, L. (2004). *The finite-difference method for seismologists: An introduction*: Comenius University Bratislava.

Moczo, P., Kristek, J., & Bystricky, E. (2000). Stability and grid dispersion of the P-SV 4(th)-order staggered-grid finite-difference schemes. *Studia Geophysica Et Geodaetica, 44*(3), 381-402.

Moczo, P., Robertsson, J. O. A., & Eisner, L. (2007). The finite-difference time-domain method for modeling of seismic wave propagation. *Advances in Geophysics, Vol 48, 48*, 421-516.

Moghaddam, P. P., & Herrmann, F. J. (2010). *Randomized full-waveform inversion: a dimenstionality-reduction approach.* Paper presented at the 2010 SEG Annual Meeting.

Moghaddam, P. P., Keers, H., Herrmann, F. J., & Mulder, W. A. (2013). A new optimization approach for source-encoding full-waveform inversion. *Geophysics, 78*(3), R125-R132.

Mondal, A., Mallick, B., Efendiev, Y., & Datta-Gupta, A. (2014). Bayesian Uncertainty Quantification for Subsurface Inversion Using a Multiscale Hierarchical Model. *Technometrics, 56*(3), 381-392.

Nadim, F., Krunic, D., & Philippe, P. (2003). Probabilistic slope stability analyses of the Sigsbee Escarpment. *Offshore Technology Conference*.

Nadim, F., Lacasse, S., Choi, Y. J., & Hadley, C. (2014). Estimation of Temporal Probability in Offshore Geohazards Assessment. *Offshore Technology Conference*.

Nguyen, T. D., Tran, K. T., & McVay, M. (2016). Evaluation of Unknown Foundations Using Surface-Based Full Waveform Tomography. *Journal of Bridge Engineering, 21*(5), 04016013.

Nowacki, F., Solhjell, E., Farrokh Nadim, F., Liedtke, E., Andersen, K. H., & Andresen, L. (2003). Deterministic slope stability analyses of the Sigsbee Escarpment. *Offshore Technology Conference*.

Operto, S., Virieux, J., Amestoy, P., L'Excellent, J. Y., Giraud, L., & Ali, H. B. H. (2007). 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study. *Geophysics, 72*(5), Sm195-Sm211.

Pakravan, A., Kang, J. W., Newtson, C. M., & Kallivokas, L. F. (2014). Hybrid perfectly-matched-layers for transient simulation of scalar elastic waves. *Structural Engineering and Mechanics, 51*(4), 685-705.

Pan, G., Phinney, R. A., & Odom, R. I. (1988). Full-waveform inversion of plane-wave seismograms in stratified acoustic media: Theory and feasibility. *Geophysics, 53*(1), 21-31.

Paoletti, L., Mouton, E., & Liposcak, I. (2010). *Comparison of Underwater MASW, Seismic CPT, and Downhole Methods: Offshore Croatia*.

Park, C. B., Miller, R. D., Xia, J., & Ivanov, J. (2007). Multichannel analysis of surface waves (MASW)—active and passive methods. *The Leading Edge, 26*(1), 60-64.

Park, C. B., Miller, R. D., Xia, J., Ivanov, J., Sonnichsen, G. V., Hunter, J. A., . . . Christian, H. (2005). Underwater MASW to evaluate stiffness of water-bottom sediments. *The Leading Edge, 24*(7), 724-728.

Park, C. B., Miller, R. D., & Xia, J. H. (1999). Multichannel analysis of surface waves. *Geophysics, 64*(3), 800-808.

Patera, A. T. (1984). A Spectral Element Method for Fluid-Dynamics - Laminar-Flow in a Channel Expansion. *Journal of Computational Physics, 54*(3), 468-488.

Pratt, R. G. (1990). Frequency-Domain Elastic Wave Modeling by Finite-Differences - a Tool for Crosshole Seismic Imaging. *Geophysics, 55*(5), 626-632.

Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. *Geophysics, 64*(3), 888-901.

Pratt, R. G., Shin, C., & Hick, G. (1998). Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International, 133*(2), 341-362.

Pratt, R. G., & Shipp, R. M. (1999). Seismic waveform inversion in the frequency domain, Part 2: Fault delineation in sediments using crosshole data. *Geophysics, 64*(3), 902-914.

Pratt, R. G., & Worthington, M. H. (1988). Inversions of Crosshole Seismic Data in the Frequency-Domain. *Geophysical Journal-Oxford, 92*(3), 531-531.

Priolo, E., Carcione, J. M., & Seriani, G. (1994). Numerical-Simulation of Interface Waves by High-Order Spectral Modeling Techniques. *Journal of the Acoustical Society of America, 95*(2), 681-693.

Puech, A., Rivoallan, X., & Cherel, L. (2004). *The use of surface waves in the characterisation of seabed sediments: development of a MASW system for offshore applications.* Paper presented at the Colloque" Caractérisation in situ des fonds marins.

Randolph, M., & Gourvenec, S. (2017). *Offshore Geotechnical Engineering*: CRC Press.

Reece, J. S., Flemings, P. B., Dugan, B., Long, H., & Germaine, J. T. (2012). Permeability-porosity relationships of shallow mudstones in the Ursa Basin, northern deepwater Gulf of Mexico. *Journal of Geophysical Research-Solid Earth, 117*.

Reed, W. H., & Hill, T. (1973). *Triangularmesh methodsfor the neutrontransportequation.* Paper presented at the National topical meeting on

mathematical models and computational techniques for analysis of nuclear systems, Ann Arbor, Michigan, USA.

Richards, J. A. (2013). *Remote sensing digital image analysis : an introduction* (Fifth ed.). Berlin: Springer.

Ritz, W. (1909). On a new method to solve certain variation problems of mathematical physics. *Journal Fur Die Reine Und Angewandte Mathematik, 135*(1/4), 1-5.

Riviere, B., & Wheeler, M. F. (2003). Discontinuous finite element methods for acoustic and elastic wave problems. *Current Trends in Scientific Computing, 329,* 271-282.

Sambridge, M., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics, 40*(3), 3-1-3-29.

Sawyer, D. E., Flemings, P. B., Dugan, B., & Germaine, J. T. (2009). Retrogressive failures recorded in mass transport deposits in the Ursa Basin, Northern Gulf of Mexico. *Journal of Geophysical Research-Solid Earth, 114*.

Sawyer, D. E., Flemings, P. B., & Nikolinakou, M. A. (2014). Continuous deep-seated slope failure recycles sediments and limits levee height in submarine channels. *Geology, 42*(1), 15-18.

Sawyer, D. E., Flemings, P. B., Shipp, R. C., & Winker, C. D. (2007). Seismic geomorphology, lithology, and evolution of the late Pleistocene Mars-Ursa turbidite region, Mississippi Canyon area, northern Gulf of Mexico. *Aapg Bulletin, 91*(2), 215-234.

Schuberth, B. (2003). *The Spectral Element Method for Seismic Wave Propagation - Theory, Implementation and Comparison to Finite Difference Methods.* Ludwig-Maximilians-University,

Sen, M. K., & Stoffa, P. L. (2013). *Global Optimization Methods in Geophysical Inversion*: Cambridge University Press.

Seriani, G., & Priolo, E. (1994). Spectral element method for acoustic wave simulation in heterogeneous media. *Finite Elements in Analysis and Design, 16*(3), 337-348.

Sheriff, R. E., & Geldart, L. P. (1995). *Exploration Seismology*: Cambridge University Press.

Shin, C., & Ha, W. (2008). A comparison between the behavior of objective functions for waveform inversion in the frequency and Laplace domains. *Geophysics, 73*(5), VE119-VE133.

Shin, C., & Min, D.-J. (2006). Waveform inversion using a logarithmic wavefield. *Geophysics, 71*(3), R31-R42.

Sirgue, L., & Pratt, R. G. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics, 69*(1), 231-248.

Stewart, M. T., & Parker, J. (1991). *Localization and seasonal variation in recharge in a covered karst aquifer system, Florida, USA.* Paper presented at the Hydrogeology of Selected Karst Regions.

Stokoe, K. H., Joh, S. H., & Woods, R. D. (2004). Some contributions of in situ geophysical measurements to solving geotechnical engineering problems. *Geotechnical and Geophysical Site Characterization Vols 1 and 2*, 97-132.

Stokoe, K. H., Wright, S. G., Bay, J. A., & Roesset, J. M. (1994). Characterization of Geotechnical Sites by SASW. *Unknown Journal*, 15-25.

Tape, C., Liu, Q. Y., Maggi, A., & Tromp, J. (2010). Seismic tomography of the southern California crust based on spectral-element and adjoint methods. *Geophysical Journal International, 180*(1), 433-462.

Tarantola, A. (1984). Inversion of Seismic-Reflection Data in the Acoustic Approximation. *Geophysics, 49*(8), 1259-1266.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Tran, K. T., & McVay, M. (2012). Site characterization using Gauss–Newton inversion of 2-D full seismic waveform in the time domain. *Soil Dynamics and Earthquake Engineering, 43*, 16-24.

Tran, K. T., McVay, M., Faraone, M., & Horhota, D. (2013). Sinkhole detection using 2D full seismic waveform tomography. *Geophysics, 78*(5), R175-R183.

Tromp, J., Komatitsch, D., & Liu, Q. Y. (2008). Spectral-element and adjoint methods in seismology. *Communications in Computational Physics, 3*(1), 1-32.

Turner, M. J., Clough, R. W., Martin, H. C., & Topp, L. J. (1956). Stiffness and Deflection Analysis of Complex Structures. *Journal of the Aeronautical Sciences, 23*(9), 805-&.

Vanneste, M., Carlton, B., Løvholt, F., Forsberg, C., Sauvin, G., Kjennbakken, H., . . . Degago, S. (2018). *Slope Stability Assessment for Deep Fjord Crossings Using*

VHR Geophysical Data – An Example from Bjørnafjorden, Norway. Paper presented at the 3rd Applied Shallow Marine Geophysics Conference.

Vanneste, M., Forsberg, C., Knudsen, S., Kvalstad, T., L'Heureux, J.-S., Lunne, T., . . . Haflidason, H. (2015). *Integration of very-high-resolution seismic and CPTU data from a coastal area affected by shallow landsliding – the Finneidfjord natural laboratory*. Paper presented at the 3rd International Symposium on Frontiers in Offshore Geotechnics, ISFOG 2015.

Virieux, J. (1986). P-Sv-Wave Propagation in Heterogeneous Media - Velocity-Stress Finite-Difference Method. *Geophysics, 51*(4), 889-901.

Virieux, J., Calandra, H., & Plessix, R. E. (2011). A review of the spectral, pseudo-spectral, finite-difference and finite-element modelling techniques for geophysical imaging. *Geophysical Prospecting, 59*(5), 794-813.

Virieux, J., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics, 74*(6), Wcc1-Wcc26.

Wilcox, L. C., Stadler, G., Burstedde, C., & Ghattas, O. (2010). A high-order discontinuous Galerkin method for wave propagation through coupled elastic-acoustic media. *Journal of Computational Physics, 229*(24), 9373-9396.

Xia, J., Miller, R. D., Park, C. B., Hunter, J. A., & Harris, J. B. (2000). Comparing shear-wave velocity profiles from MASW with borehole measurements in unconsolidated sediments, Fraser River Delta, BC, Canada. *Journal of Environmental & Engineering Geophysics, 5*(3), 1-13.

Yamamoto, Y., & Sawyer, D. E. (2012). Systematic Spatial Variations in the Fabric and Physical Properties of Mass-Transport Deposits in the Ursa Region, Northern Gulf of Mexico. *Submarine Mass Movements and Their Consequences, 31*.

Ye, R. C., de Hoop, M. V., Petrovitch, C. L., Pyrak-Nolte, L. J., & Wilcox, L. C. (2016). A discontinuous Galerkin method with a modified penalty flux for the propagation and scattering of acousto-elastic waves. *Geophysical Journal International, 205*(2), 1267-1289.

Yilmaz, Ö. (2001). *Seismic Data Analysis*: Society of Exploration Geophysicists.

Young, A. G., Phu, D. R., Spikula, D. R., Rivette, J., Lanier, D. L., & Murff, J. (2009). *An Approach for Using Integrated Geoscience Data to Avoid Deepwater Anchoring Problems*. Paper presented at the Offshore Technology Conference, Houston, Texas.

Yuan, J., Zhu, J., & Kim, C. (2014). Comparison of SASW and MASW methods using MSOR approach – a case study. *International Journal of Geotechnical Engineering, 8*(2), 233-238.

Zhou, H.-W. (2014). *Practical Seismic Data Analysis*. Cambridge: Cambridge University Press.

Zhu, D., & Gibson, R. (2016). Seismic inversion and uncertainty analysis using a transdimensional Markov chain Monte Carlo method. In *SEG Technical Program Expanded Abstracts 2016* (pp. 3666-3671): Society of Exploration Geophysicists.

APPENDIX

HIERARCHICAL BAYESION MODELING

This appendix shows the detailed algorithm of the Metropolis-Hastings sampling for trans-dimensional rj-MCMC method, which defines criteria to accept or reject the proposed candidates. The probabilistic approach to integrate multiple data sets requires hierarchical approach, which uses the Bayesian method in multiple levels to estimate the parameters of the posterior distribution.

A. Hierarchical Bayesian priors in varying dimension

Since the reversible jump Markov chain Monte Carlo (rj-MCMC) method uses independent random variables in varying dimensions, we should apply appropriate priors to control the dimension and make convergence to find the stationary condition in stochastic sampling. For the stochastic seismic inversion, the depth of layer interfaces (z) and material properties (β) in each layer are defined as the random input variables. Thus, we can apply two different priors, one is about the material properties inside the layer, and the other is about discretization in the depth of the ground model.

$$p(\boldsymbol{\theta}) = p(k, z, \beta) = p(\beta|k)p(z|k)p(k) \tag{A.1}$$

The first Bayesian prior for the material properties comes from the information about seawater and shallow marine sediments. We can assume the P-wave velocity and density for seawater is around 1,500 m/s and 1.0 g/cc with a small amount of uncertainty. Since those properties in soil layers have large variation, we defined the

only minimum and maximum values and applied uniform distribution for the prior of submarine soil property. This prior has to consider the number of dimensions, and we need to combine every probability based on the dimension parameter (k).

$$p(\beta_i|k) = \begin{cases} \frac{1}{\beta_{max}-\beta_{min}}, if\ \beta \in \beta_{range} \\ 0 \end{cases} \quad \text{(A.2.a)}$$

$$p(\beta|k) = \prod_{i=1}^{k} p(\beta_i|k) \quad \text{(A.2.b)}$$

The second Bayesian prior for the mesh discretization is based on the possible number of the specific event from the given maximum dimension. For example, if we have 100 discretized elements in our target depth, then we can define 1 to 100 layers in this mesh. If we need to define 5 layers from these 100 elements, this becomes a selection problem of 5 from 100 with a combination operator in mathematics, and the probability can be written using factorials.

$$p(\boldsymbol{z}|k) = \left[\frac{N!}{k!(N-k)!}\right]^{-1} \quad \text{(A.3)}$$

The last Bayesian prior for the unknown dimension has only one information, the numbers of minimum and maximum layer interfaces in our target depth. These are subjective, but we can roughly estimate those numbers from the geophysical seismic interpretation.

$$k^* \sim unif(k_{min}, k_{max}) \quad \text{(A.4)}$$

We also applied the non-informative priors to support the dynamic standard deviation of the likelihood function and proposal distribution. These priors use hyperparameters with inverse gamma distribution and increase the acceptance ratio during the stochastic sampling process. The random variable as the standard deviation

(s) supports the likelihood function, and the other random variable (v) supports the proposal distribution. Both non-informative priors use the value of 0.01 for their hyperparameters based on the theoretical background, which were described in other previous trans-dimensional research (S Esmailzadeh et al., 2015a; Mondal et al., 2014)

$$s \sim IG(0.01,0.01) \tag{A.5.a}$$

$$v \sim G(0.01,0.01) \tag{A.5.b}$$

Consequently, the Bayesian priors are defined as below in the hierarchical format based on the information about material properties, mesh discretization, the expected number of layers in the target area, and the non-informative priors.

$$p(\boldsymbol{\theta}) = p(\rho|k,\rho_w)p(\rho_w)p(Vp|k,Vp_w)p(Vp_w) \times p(z|k)p(k)p(s)p(v) \tag{A.6}$$

If we don't apply appropriate priors in this trans-dimensional modeling, the strong likelihood function will keep increasing the number of layers to minimize the misfit between observed and predicted seismic data. This abused optimization in the maximum dimension may cause the over-fitted problem and generate an incorrect ground model. In short, the accurate geophysical seismic interpretation should be implemented as the importance pre-process for the seismic inversion to build up the appropriate Bayesian modeling priors.

B. Proposal distributions for trans-dimensional model selection

The stochastic process in varying dimensions requires to choose one type of model selections at each iterative modeling. This random process is uniformly selected from the four following possibilities.

- BIRTH: This creates a new subsurface layer with a new depth of layer interface (z^*) and new material properties (β^*). This new layer is located under a randomly chosen layer depth (z_{idx}).

$$z^* \sim unif(z_{idx}, z_{idx+1}) \tag{A.7.a}$$

$$\beta^* \sim N(\beta_{idx}, v * \sigma) \tag{A.7.b}$$

$$\{z_1, \dots, z_{idx}, z_{idx+1} \dots, z_k\} \rightarrow \{z_1, \dots, z_{idx}, z^*, z_{idx+1}, \dots, z_k\} \tag{A.7.c}$$

$$\{\beta_1, \dots, \beta_{idx}, \beta_{idx+1} \dots, \beta_k\} \rightarrow \{\beta_1, \dots, \beta_{idx}, \beta^*, \beta_{idx+1}, \dots, \beta_k\} \tag{A.7.d}$$

- DEATH: This removes one of the current subsurface layers. The randomly chosen depth (z_{idx}) is removed, and the material properties at this depth are defined as the average values between the upper and lower layers.

$$\{z, \dots, z_{idx}, z_{idx+1} \dots, z_k\} \rightarrow \{z_1, \dots, z_{idx-1}, z_{idx+1}, \dots, z_k\} \tag{A.8.a}$$

$$\{\beta_1, \dots, \beta_{idx}, \beta_{idx+1} \dots, \beta_k\} \rightarrow \left\{\beta_1, \dots, \frac{\beta_{idx}+\beta_{idx+1}}{2}, \dots, \beta_k\right\} \tag{A.8.b}$$

- MOVE: This does not change the dimension of the modeling, and only changes the location of one current layer. One of the current layer depth is randomly chosen, and the new location (z^*) is randomly located between upper (z_{up}) and lower (z_{low}) layer interfaces.

$$z_i^* \sim unif(z_{up}, z_{low}) \tag{A.9}$$

- PERTURB: This does not change the dimension of the modeling, and only changes the material properties of one current layer. One of the current layer material properties (β_i) is randomly chosen, and the new material properties (β^*) is randomly defined with standard deviation (σ) and parameter (v).

$$\beta_i^* \sim N(\beta_i, v * \sigma) \tag{A.10}$$

C. Proposal ratios

Since this modeling selection changes the dimension of the modeling in a random process, we need to apply the generalized Metropolis-Hasting sampling algorithm with the acceptance probability (α). The calculation of this probability is based on likelihood, prior and proposal ratios from the current model parameters (θ) and proposed model parameters (θ^*).

$$\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = min\left\{1, \underbrace{\frac{p(\boldsymbol{D_{obs}}|\boldsymbol{\theta}^*)}{p(\boldsymbol{D_{obs}}|\boldsymbol{\theta})}}_{\text{likelihood ratio}} \times \underbrace{\frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta})}}_{\text{prior ratio}} \times \underbrace{\frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}}_{\text{proposal ratio}}\right\} \tag{A.11}$$

The proposal ratio is the key point in the calculation of acceptance probability (α) to control the varying dimension. This ratio should be considered in both forward and reverse movement in varying dimensions. When the rj-MCMC choose the MOVE, PERTURB for model selection, the probability in forward (from θ to θ^*) is equal to the probability in reverse (from θ^* to θ), so the dimension does not change at all. That means the probabilities of proposals are canceled out and the proposal ratio becomes one to be ignored. However, BIRTH and DEATH types for model selection change their

modeling dimension, so each proposal ratio involves extra calculations. Since the locations of layers are proposed independently from the material properties, the calculation of the proposal ratio can be separated into two terms.

$$\frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} = \frac{q(z|\theta^*)}{q(z^*|\theta)} \times \frac{q(\beta|\theta^*)}{q(\beta^*|\theta)} \tag{A.12}$$

The calculation requires a total number of the discretized elements (N) for the first term, and the second term for the material property simply follows the normal distributions. The specific probabilities for the BIRTH and DEATH are defined in different forms, respectively.

- BIRTH: This changes the dimension from k layers to $k + 1$ layers.

$$q(z^*|\theta) = \frac{1}{N-k} \tag{A.13.a}$$

$$q(z|\theta^*) = \frac{1}{k+1} \tag{A.13.b}$$

$$q(\beta^*|\theta) = normpdf(\beta, v * \sigma) \tag{A.13.c}$$

$$q(\beta|\theta^*) = 1 \tag{A.13.d}$$

- DEATH: This changes the dimension from k layers to $k - 1$ layers.

$$q(z^*|\theta) = \frac{1}{k} \tag{A.14.a}$$

$$q(z|\theta^*) = \frac{1}{N-(k-1)} \tag{A.14.b}$$

$$q(\beta^*|\theta) = \frac{(\beta_i+\beta_{i+1})}{2} \tag{A.14.c}$$

$$q(\beta|\theta^*) = normpdf(\beta, v * \sigma) \tag{A.14.d}$$